Host Variation in Type I Interferon Signaling Genes (*MX1*), CCR5Δ32, and MHC Class I Alleles in Treated HIV+ Non-Controllers Predict Viral Reservoir Size

David A. Siegel[1], Cassandra Thanh[4], Eunice Wan[2], Rebecca Hoh[1], Kristen Hobbs[4], Tony Pan[4], Erica A. Gibson[4], Deanna L. Kroetz[3], Peter W. Hunt[4], Jeffrey Martin[5], Frederick Hecht[1], Christopher Pilcher[1], Jeffrey Milush[4], Maureen Martin[6], Mary Carrington[6,7], Satish Pillai[8], Michael P. Busch[8], Mars Stone[8], Meei-Li Huang[9,10], Pavitra Roychoudhury[9,10], Claire N. Levy[9,10], Keith R. Jerome[9,10], Hans-Peter Kiem[9,10], Timothy J. Henrich[4], Steven G. Deeks[1*], and Sulggi Lee[1*#]

[1] Department of Medicine, Division of HIV, Infectious Diseases & Global Medicine, University of California San Francisco, 995 Potrero Avenue, San Francisco, CA 94110, USA. [2] Institute of Human Genetics, University of California San Francisco, CA 94143, USA. [3] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, CA 94158. [4] Department of Medicine, Division of Experimental Medicine, University of California San Francisco, CA 94110, USA. [5] Department of Biostatistics & Epidemiology, University of California San Francisco, CA 94158, USA. [6] Basic Science Program, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD and Laboratory of Integrative Cancer Immunology, Center for Cancer Research, National Cancer Institute, Bethesda, MD USA. [7] Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [8] Vitalant Blood Bank, San Francisco, CA 94118 USA. [9] Department of Laboratory Medicine and Pathology, University of Washington, Seattle WA 98195, USA. [10] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle WA 98109, USA.

*Co-correspondence:

sulggi.lee@ucsf.edu, steven.deeks@ucsf.edu

#Lead contact

Abstract

Prior genomewide association studies have identified variation in MHC Class I alleles and *CCR5Δ32* as genetic predictors of viral control, especially in "elite" controllers, individuals who remain virally suppressed in the absence of therapy. We analyzed custom host whole exome sequencing and direct HLA typing from 202 ART-suppressed HIV+ non-controllers in relation to four measures of the peripheral CD4+ T cell reservoir: HIV intact DNA, total (t)DNA, unspliced (us)RNA, and RNA/DNA. Linear mixed models were adjusted for potential covariates including age, sex, nadir CD4+ T cell count, pre-ART HIV RNA, timing of ART initiation from date of detected HIV infection, and duration of ART suppression. Previously reported HLA "protective" B*57:01 (in untreated elite controllers) was significantly associated with lower usRNA ($p=4.2x10^{-4}$, Benjamini-Hochberg false discovery rate [FDR] $q=3.3x10^{-3}$). The favorable *CCR5Δ32* mutation was also associated with lower tDNA and usRNA, primarily in Europeans ($p=4.3x10^{-3}$ and $8.7x10^{-3}$). Several SNPs in interferon stimulated gene, *MX1*, were associated with HIV total DNA ($p=1.3x10^{-7}$; q=0.02), and the direction of these associations paralleled predicted eQTL (expression quantitative trait) *MX1* gene expression. Among Europeans, gene set enrichment analysis identified genes involved in "negative regulation of host cell viral entry" to be significantly associated with usRNA (q=0.03). These findings suggest a potential role for interferon signaling, MHC class I alleles, and CCR5Δ32 in maintaining the reservoir size in HIV non-controllers. These data need to be further validated in functional studies and replication cohorts.

Abstract (max 300 words): 235

Full Title (max 200 characters: 157 with spaces): Host Variation in Type I Interferon Signaling Genes (*MX1*), *CCR5Δ32*, and MHC Class I Alleles, and in Treated HIV+ Non-Controllers Predict Viral Reservoir Size

Short Title (max 70 characters: 69 with spaces): Host Genetic Predictors of HIV Reservoir Size in ART+ Non-Controllers

Keywords: HIV reservoir, host genetics, whole exome sequencing, type I interferon, MHC class I, HLA, CCR5

CONFLICTS: The authors do not have a commercial or other association that might pose a conflict of interest.

FUNDING: This work was supported in part by the National Institutes of Health: K23GM112526 (SAL), the DARE Collaboratory (U19 AI096109; SGD), the Division of Intramural Research of the National Institutes (MC), UM1 AI126623 (HPK, KRJ), and NIH/NIAID R01A141003 (TJH). 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 (MC). 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 (MC). This work was also supported by the amfAR Research Consortium on HIV Eradication a.k.a. ARCHE (108072-50-RGRL; SGD) and a Collaboration for AIDS Vaccine Discovery

(CAVD) grant from the Bill & Melinda Gates Foundation (INV-008500), the Reservoir Assay Validation and Evaluation Network (RAVEN) Study Group. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Reprints: Reprint requests can be directed to Dr. Sulggi Lee, the corresponding author (contact information above).

Acknowledgements

The authors wish to acknowledge the participation of all the study participants who contributed to this work as well as the clinical research staff of the SCOPE and Options who made this research possible. All funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. All authors provided critical feedback in finalizing the report. SAL conceived and designed the study with critical feedback from SGD, PWH, TH, DK, KRJ, and SH. SGD, JM, FH, CP, RH, and SAL coordinated the collection, management, and quality control processes for the cohort clinical data and SGD, JM, FH, CP, MPB, MS provided biospecimens. SAL and EW performed the whole exome sequencing assays. SAL, CT, JM, KB, TP, EAG performed participant sample processing, SAL and EW performed the whole exome sequencing, and DAS and SAL performed quality control analyses and the genomic association analyses for the study. SAL, SAL, CT, TH, KRJ, MLH, PR performed the HIV reservoir measurements, and PR, DAS, TJH, and SAL analyzed these data. MM and MC performed the HLA typing to determine HLA alleles for the analyses. DAS and SAL wrote the report. All authors provided critical feedback in finalizing the manuscript.

1. Background

HIV persistence may contribute to ongoing inflammation, immune activation, and subsequent clinical outcomes, even during effective ART (1-4). However, factors associated with the HIV reservoir size, including the timing of ART initiation (smaller reservoir sizes with starting therapy during acute versus chronic infection (5, 6)), and host factors such as the level of T cell activation and T cell function only explain a fraction of the person-to-person variability in HIV reservoir size (7). Identifying host genetic predictors of the HIV reservoir in ART-suppressed individuals may shed light on additional (potentially modifiable factors) that may contribute to HIV persistence.

Most prior host genetic HIV studies have evaluated untreated HIV-infected individuals to identify variants associated with viral setpoint, especially among "elite controllers", HIV+ individuals able to maintain viral suppression in the absence of therapy (8-17). These studies identified several key single nucleotide polymorphisms (SNPs) in the human Major Histocompatibility Complex (MHC), or human leukocyte antigen (HLA)-B and -C regions as well as deletions in the C-C chemokine receptor type 5 gene (*CCR5Δ32*) (18-21) and a SNP in the HLA complex 5 (*HCP5*) gene (9). It remains unclear if "residual" viral control during treated HIV disease – a.k.a., "the HIV reservoir" – is influenced by the same genetic variants that influence viral control during untreated disease. It is also possible that identified genetic variants may have variable "penetrance" (22) in the population, i.e., lead to varying degrees to which the clinical phenotype is expressed; the same genetic variants enriched in "elite controllers" (e.g., *CCR5Δ32*) might drive the same immune processes influence HIV reservoir size in non-controllers, but to a lesser degree.

Therefore, in order to identify host genetic variation associated with the HIV reservoir size, we performed custom whole exome sequencing and HLA typing among HIV non-controllers, including a range of individuals treated during "early" (within 6 months) and "later" (after 6 months) of HIV infection from the University of California San Francisco SCOPE and

Options cohorts. Our study adds to two recent host genetic studies of the HIV reservoir (23, 24), one of which was published concurrently to ours (24). The first (23) evaluated 797 HIV-infected acute- and chronic-treated individuals from the Swiss Cohort Study, of which 194 had whole exome sequencing performed but observed no association between genetic variants (including HLA or *CCR5Δ32*) and reservoir size. The study included longitudinal measurements of the HIV reservoir, but only HIV total DNA from peripheral blood mononuclear cells (PBMCs) were quantified. The study also imputed HLA alleles (from genotypes), rather than performed direct HLA typing, and it is unclear if the study adjusted for timing of ART initiation as a covariate in the modeling. The second (24) study was published concurrently with our own and included 207 chronic-treated individuals. GWAS microarray, rather than genetic sequencing, was performed, but HIV total DNA and unspliced RNA were quantified from peripheral CD4+ T cells. The study did not report HLA or *CCR5Δ32* findings, but did report a statistically significant association between HIV tDNA and a SNP in *PTDSS2*, encoding an immunosuppressive protein, using a stringent genomewide significance cutoff of $P<5x10^{-8}$ (25). A trend was also reported with HIV RNA/DNA and three additional genes, *IRF7* (Interferon Regulatory Factor 7), *RNH1* (Ribonuclease/Angiogenin Inhibitor 1), and *DEAF1* (DEAF1 Transcription Factor).

Here, we performed custom whole exome sequencing (including intronic, 50 kilobase upstream, and 50 kilobase downstream regulatory regions of select candidate genes) and direct HLA typing of 25 polymorphic exons of HLA-A & B, C, E, DPA1, DPB1, DQA1, DQB1, DRB1, and DRB3, 4, 5. We quantified four measures of the HIV reservoir from peripheral CD4+ T cells: HIV cell-associated “intact” DNA (an estimate of the frequency of potentially “replication-competent” virus with intact HIV genomes) (26), as well as total DNA (approximates intact + defective HIV DNA), unspliced (full length transcript prior to alternative splicing) RNA, and RNA/DNA (previously used as a marker of “transcriptionally active” HIV reservoir) from 202 HIV+ ART-suppressed acute- and chronic-treated non-controllers (**Figure 1**). We performed both “common” (polymorphisms observed at minor allele frequency, MAF, ≥5%) and rare-

variant, gene-based analyses. We observed significant genetic associations with variants previously linked to viral control during untreated HIV disease (e.g., MHC class I alleles (9, 16) and *CCR5Δ32* (19, 20, 27)), to be associated with the HIV reservoir size in our study population. In addition, we identified several novel genes, highlighting potentially important roles for interferon signaling (28) and actively transcribing HIV (29-32) in determining the peripheral HIV reservoir size.

3. Results

Study population

A total of 202 HIV-infected ART-suppressed individuals from the UCSF SCOPE and Options HIV+ cohorts were included in the study. Consistent with our San Francisco-based HIV patient population, participants were mostly male (94%) with median age of 46 (**Table S1**). The study included individuals who initiated ART during early (<6 months) as well as chronic (≥6 months) HIV infection (**Table S1**). HIV "elite controllers," individuals with a history of undetectable viral load in the absence of therapy for at least 1 year prior to the specimen collection date (33-35), were excluded. Participants had a median of 5.1 years of ART suppression at the time of biospecimen sampling and reported a median nadir CD4+ T cell count of 341 cells/mm^3 and maximum pre-ART HIV RNA of 5.1 log_{10}copies/mL (**Table S1**). The majority of study participants reported White/European American ethnicity (63%), and the remainder reported Black/African American (12%), Hispanic/Latino (11%), Mixed Ethnicity/Multiracial (6%), Asian (4%), Pacific Islander (1.5%), Native American (<1%), and Middle Eastern (<1%) ethnicities. Interestingly, upon further inspection of the raw clinical data, we identified 22 individuals with at least one documented HIV RNA value <10,000 prior to ART initiation (which did not occur during primary HIV infection when such values might be

considered within normal range (36)); these individuals were assessed in subanalyses shown in **Figures S1-S2** as "possible partial controllers" to ensure that surreptitious HIV controllers were not included in the analyses, potentially driving our results.

Earlier ART initiation was associated with smaller HIV reservoir sizes

Consistent with prior work (37-39), earlier ART initiation was statistically significantly associated with smaller HIV reservoir sizes: tDNA: R=0.28; $p=7.6x10^{-5}$, usRNA: R=0.28,$p=5.9x10^{-5}$, and intact DNA: R=0.21; $p=3.8x10^{-3}$ (**Figure 2**). Lower nadir CD4+ T cell counts were significantly associated with larger reservoir sizes: tDNA: r=-0.28, $p=7.1x10^{-5}$, usRNA: R=-0.23, $p=9.9x10^{-4}$, RNA/DNA: -0.15, $p=3.3x10^{-2}$, and intact DNA: R=-0.27, $p=2.5x10^{-4}$ (**Figure S3**). Maximum pre-ART viral load and duration of ART suppression were less strongly correlated with the HIV reservoir measures (**Figure S4** and **Figure S5**, respectively).

Measures of the HIV reservoir size were correlated with each other

The HIV reservoir largely consists of "defective" virus that harbors mutations prohibiting the production of infectious virus (39, 40), and there is currently no "gold standard" for measuring the HIV reservoir. Here, we estimated the frequency of HIV "intact" DNA using a droplet digital PCR (ddPCR)-based assay as a proxy for estimating the size of the potential "replication-competent" reservoir (41, 42). We also measured HIV total DNA (which includes both defective and intact HIV) and unspliced RNA (which corresponds to full-length HIV RNA transcript, prior to alternative splicing) with an in-house quantitative polymerase chain reaction (qPCR) TaqMan assay using HIV-1 long terminal repeat (LTR)-specific primers (43). Cell-associated HIV total DNA and unspliced RNA were simultaneously extracted from the same sample aliquot using the dual extraction AllPrep Universal Kit (Qiagen, Hilden, Germany)); tDNA and usRNA results were found to be significantly correlated with one another (Spearman r=0.59, $p=8.3x10^{-20}$, **Figure S6a**). HIV intact DNA was associated with HIV usRNA (Spearman

r=0.25, p=7.0x10^{-4}, **Figure S6b**) but not with HIV tDNA (Spearman r=0.10, p=0.19, **Figure S6c**).

Custom whole exome sequencing enhanced genomic coverage and allowed principal component analysis to minimize population stratification bias

We designed custom primers and probes to sequence an additional 54 megabases of intronic, intergenic, and untranslated regions of putative genes associated with the HIV reservoir, including genes previously associated with HIV integration into the host genome, host HIV antiviral restriction factors, and genes involved in the regulation of cell cycle progression, in addition to performing whole exome sequencing of all exons in the human genome (**Table S2**) (44-47). A total of 1,279,156 variants from 23,733 genes from 199 participants passed quality control metrics, with a total of 46% intronic, 21% intergenic, 19% exonic, 6% untranslated region (UTR), 6% non-coding RNA variants that were included in the final analysis (**Figure 3a**). Among the exonic variants, 53% were nonsynonymous and 42% were synonymous variants (**Figure 3b**).

Principal component analysis (PCA) was performed to investigate the population structure and account for potential population stratification bias (i.e., differences in allele frequencies due to differences in ancestry rather than to clinical phenotype). These analyses generated the PCs used as covariates for ancestry in final multivariate models (**Figure 4)**. Secondary PC analyses were performed among Europeans, the largest cluster, providing a relatively homogenous subgroup in which to perform sensitivity analyses (**Figure 4b**). Final models demonstrated lambda GC (genomic inflation factor (48)) values that were all close to 1 (**Figure 5**), further ensuring minimal systematic bias secondary to population stratification in the analyses.

HLA "protective" B*57:01 and "risk" C*07 alleles from prior studies of untreated HIV disease were associated with smaller and larger HIV reservoir sizes, respectively

We performed direct HLA typing to determine whether HLA alleles known to be associated

with viral control during untreated disease (16) were also associated with the HIV reservoir (**Tables S3-S6**). Using a Benjamini-Hochberg (49) false discovery rate (FDR) adjusted q<0.05 for statistical significance, as well as evaluating a less stringent q<0.25 for suggestive trends (given our smaller genomewide study, as discussed previously (50)), we examined previously reported protective (B*57:01, B*27:05, B*14, C*08:02, B*52, and A*25) and risk (B*35 and C*07) alleles (16). HLA-B*57:01 demonstrated a statistically significant "protective effect" on HIV usRNA (β=-1.5, p=4.2x10^{-4}, q=3.3x10^{-3}), with a similar trend observed for tDNA (β=-1.6, p=0.032, q=0.13). The previously reported HLA-C*07 "risk" allele associated with higher viral loads during untreated disease, which was observed primarily in Europeans, also demonstrated a trend towards larger HIV reservoir size in our European subgroup (tDNA: β=0.76, p=0.012, q=0.072, and usRNA: β=0.41, p=0.034, q=0.10). Previously reported "risk" allele HLA-B*35, in contrast, demonstrated a trend with smaller HIV reservoir size (tDNA: β=-0.81, p=0.021, q=0.13, and usRNA: β=-0.44, p=0.032,q=0.13). Neither HIV RNA/DNA (a frequently used biomarker of the "active" HIV reservoir in older studies (51, 52)) or HIV intact DNA (a more recent assay that estimates the "replication-competent" HIV reservoir (41, 42)) were statistically significantly associated with HLA alleles. Further analyses employing a composite HLA variable (the aggregate of the total number of protective alleles minus the total number of risk alleles was analyzed in relation to each of the HIV reservoir measures), did not identify statistically significant associations (**Tables S3-S6**).

CCR5Δ32 was associated with smaller HIV reservoir sizes

Deletions in the C-C chemokine receptor type 5 gene (*CCR5Δ32*) have previously been shown to be associated with HIV viral control in the absence of therapy (19, 20, 27). We observed a statistically significant association between the presence of *CCR5Δ32* mutation and smaller HIV reservoir size in Europeans (tDNA: β=-1.3, p=4.3x10^{-3}; usRNA: β=-0.78, p=8.7x10^{-}

^{3}), with a similar trend observed in the total population (tDNA: β=-0.86, p=0.045; usRNA: β=-0.41, p=0.12), In addition, the previously reported long noncoding RNA variant which regulates differential CCR5 expression (rs1015164) (21), was found to be significantly associated with smaller HIV reservoir size in Europeans (usRNA: β=-0.39, p=0.027), which reached near-statistical significance in the total population as well (usRNA: β=-0.30, p=0.051),

Several SNPs in the antiviral interferon stimulated gene, *MX1*, were associated with larger and smaller HIV reservoir sizes, paralleling predicted *MX1* gene expression

The strongest genetic associations with HIV reservoir size in the total study population were with total HIV DNA, in particular, with the human interferon-inducible myxovirus resistance 1 gene, *MX1*, also known as *MXA* (53, 54). Again using a Benjamini-Hochberg (49) FDR-adjusted q<0.05 for statistical significance, and a less stringent q<0.25 to evaluate suggestive trends in our smaller genomewide study (50), we identified 44 SNPs in linkage disequilibrium (LD) in *MX1* (all q<0.03) (**Table 1 and Table S7**). Many of these *MX1* SNPs lie in known expression quantitative trait loci (eQTL) (55), genomic regions shown to influence gene (mRNA) expression in experimental and/or in silico studies. *MX1* is closely related to *MX2* (*MXB*), which encodes a well-known potent host restriction factor that inhibits HIV-1 infection (56-58). MX1 has been associated with HIV-1 latency in in latently infected cell lines (59) and been shown to inhibit viral replication for other viruses (e.g., influenza and hepatitis B) (60). Interestingly, per the eQTL data, the top hit *MX1* SNPs associated with upregulated *MX1* expression (55), were the same SNPs that predicted larger HIV reservoir sizes, while the *MX1* SNPs with predicted lower *MX1* expression, were the SNPs associated with smaller HIV reservoir sizes (**Table S7**). These eQTL data might support a model where enhanced *MX1* gene expression reflects a larger HIV reservoir size. Perhaps, for example, that genetically predisposed individuals with robust type I interferon responses may fail to limit the size of the HIV reservoir due to

desensitization and/or selection of interferon-resistant clones over time. This model would fit with a recent study demonstrating that after each round of ART interruption, rebounding virus emerged increasingly from interferon-resistant clones (able to evade host pressure) (28).

We also observed two additional SNPs significantly associated with HIV tDNA, the first in *PPP1CB* ($p=1.5x10^{-7}$, q=0.03), which encodes Protein Phosphatase 1 Catalytic Subunit Beta, a protein shown to reduce the antiviral potency of MX2 against HIV-1 (58), and *LRMP* ($p=4.1x10^{-6}$, q=0.03), which encodes Lymphoid-Restricted Membrane Protein, which plays a critical role in the delivery of peptides to MHC class I molecules (61) (**Table 1**). Additional SNPs that showed trends with HIV tDNA included: *DDX3X* ($p=2.5x10^{-5}$, q=0.17), which encodes DEAD-box helicase 3 X-linked, a protein that regulates the production of type I interferons (62), and *AKAP6* ($p=3.0x10^{-5}$, q=0.20), encoding A-Kinase Anchoring Protein 6, which that binds to regulatory subunits of protein kinase A, a critical signaling pathway that has been associated with HIV latency reversal and T cell proliferation (63, 64). Among Europeans, *OSBP* ($p=1.9x10^{-5}$, q=0.14), encoding oxysterol-binding protein, involved in intracellular lipid transport and previously associated with HIV-1 infection of monocyte-derived macrophages from highly-exposed seronegative individuals (65), also showed a trend with HIV tDNA.

No SNPs met statistical significance in association with HIV intact DNA or RNA/DNA ratio, and one SNP demonstrated a slight trend with HIV usRNA (**Table 1**). This SNP is intronic in *PLAVP* ($p=6.9x10^{-7}$, q=0.21), which encodes a protein regulating lymphocyte migration into lymph nodes (66) and also lies <30 kilobases upstream of *BST2*, encoding for the HIV host restriction factor, tetherin (67).

Genes involved in viral entry into host cell were significantly associated with the putative "replication-competent" HIV reservoir (intact DNA)

We then performed multi-SNP rare variant analyses using the Variant Set Mixed Model

Association Test (SMMAT) (68), a similar rare variant analysis method to SKAT-O (69), and gene set enrichment analyses (GSEA) using the Broad MSigDB database (47, 70), to identify genes and gene sets associated with the HIV reservoir size. No individual genes met statistical significance at an FDR q<0.05 or a trend at q<0.25 (data not shown). GSEA identified one statistically significant association between "Negative Regulation of Viral Entry into Host Cell" and HIV intact DNA levels (NES, normalized enrichment score=1.81, q=0.03) (**Figure 6, Table S8)**. Genes contributing to the gene set's statistical significance included: *IFITM1*, *IFITM3*, *APCS*, *IFITM2*, *FCN3*, *FCN1*, *GSN*, *TRIM59*, *SNX3*, *TRIM25*, *PTX3*, *TRIM11*, *TRIM8*, *MID2*, *TRIM5*, *IFNA2* (**Table S8**).

Several other gene sets showed trends with HIV intact DNA (including 2 gene sets related to the key interferon-induced STAT, signal transducer and activator of transcription, signaling pathway; NES=1.67, q=0.19 and NES=1.63, q=0.25), HIV tDNA (glycosylation gene set; NES=1.76, q=0.12) and usRNA (retroviral transcription gene set; NES=1.73, q=0.16) (**Figure 6, Table S8**).

4. Discussion

We performed custom whole exome sequencing and HLA typing from 202 HIV+ ART-suppressed non-controllers treated during both acute and chronic HIV infection, in relation to four measures of the peripheral CD4+ T cell reservoir (intact DNA, total DNA, unspliced RNA, and RNA/DNA). Our study found that the previously reported "protective" HLA-B*57:01 allele and the favorable *CCR5Δ32* mutation were associated with smaller HIV reservoir size (usRNA for B*57:01; tDNA and usRNA for *CCR5Δ32*), especially among individuals of European ancestry. SNP-based analyses identified *MX1*, a type I interferon stimulated gene, to be significantly associated with tDNA, and these associations correlated with predicted eQTL *MX1* gene expression data. Finally, gene-based analyses identified genes involved in "viral entry into

host cell" to be significantly associated with HIV intact DNA. The gene set again included several genes involved in interferon signaling, potentially reflecting an important role in maintenances of the replication-competent HIV reservoir. Further validation of our findings in other cohorts is needed, especially among non-European and including female participants, to determine whether these genetic variants might serve as potential targets for viral eradication in the majority of people living with HIV (PLWH).

We observed a statistically significant protective association with HLA-B*57:01 and a risk association with HLA-C*07, as previously described during untreated HIV disease (9, 16), on the HIV reservoir size in our study. These associations were observed primarily among individuals with European ancestry and with usRNA measures; however similar trends were observed in the total population, as well as with tDNA levels (**Tables S3-S6**). Similarly, previously reported "risk" allele HLA-B*35, showed a trend with usRNA and tDNA only – but the direction for this allele was in contrast to prior reported associations; here B*35 was associated with a "protective" effect, or smaller HIV reservoir size. If confirmed in other studies, the observed associations between these HLA alleles and tDNA and usRNA (but not intact DNA) may simply reflect the lower frequency of intact DNA compared to these other measures, but might also suggest that even defective provirus can produce proteins (e.g., HIV Nef), making the cells "visible" to the immune system (e.g., cells expressing MHC class I molecules) (71-73).

Similarly, our observation that the *CCR5Δ32* mutation was associated with a smaller HIV reservoir size is consistent with a prior study in ART-suppressed individuals (74). A second study, in contrast, did not report an association between *CCR5Δ32* and reservoir size during treated disease, but the study only included chronic-treated individuals, which may have influenced the differential findings (75). Our results are also in contrast to a recent genomewide study which also reported no association between *CCR5Δ32* and HIV reservoir size (measured solely as total DNA from PBMCs) (23). However, this study as well did not adjust their analyses for timing of ART initiation, one of the strongest clinical predictors of HIV reservoir size (5, 6),

although Table 1 of the paper indicates that at least 140 individuals were labeled as “Infection Stage” as being “acute.” The study also reported approximately 200 (25%) of the participants had “blips” or “low-level” viremia during the study period, which may have also contributed to their findings.

We also performed SNP-based common genetic variant (MAF≥5%) and multi-SNP rare variant gene-based analyses. The most striking finding from our SNP-based analysis was the identification of several SNPs in the human interferon-inducible myxovirus resistance 1 gene, *MX1*, also known as *MXA*, as well as additional genes linked to type I interferon signaling (**Table 1, Table S7**). The encoded protein, MX1, is closely related to MX2 (MXB), a potent host restriction factor that inhibits HIV infection prior to integration (56-58). MX1 itself is long recognized as a potent antiviral factor, inhibiting replication of several RNA viruses, including influenza A and measles viruses, and DNA viruses, such as hepatitis B virus (60). Indeed, *MX1* gene expression has been shown to be upregulated in HIV-infected individuals compared to HIV-uninfected individuals (76), in HIV-infected individuals with high vs low viremia (77), and with HIV-1 latency in latently infected cell lines (59). eQTL function at these loci (55) suggest that elevated levels of MX1 expression may be associated with higher total HIV DNA levels (78, 79). We also observed statistically significant (*PPP1CB*) and suggestive (*DDX3X*) associations with HIV tDNA and these additional interferon-related genes. *PPP1CB* encodes Protein Phosphatase 1 Catalytic Subunit Beta, which regulates the antiviral potency of MX2 (58). PP1CB has been shown to interact with HIV Tat and is required for Tat-induced HIV-1 transcription in cultured cells; suppression of PP1 has been shown to inhibit HIV-1 transcription in cell lines (80). *DDX3X* encodes DEAD Box Helicase 3 X-Linked, which is an X-chromosome-encoded RNA helicase that regulates type I inteferon production (62). Prior work suggests that DDX3X may be involved in HIV-1 translation, delivering HIV-1 RNA from the nucleus to the cytoplasm (81). The other statistically significant SNP associated with tDNA was in *LRMP*, which encodes for Lymphoid-Restricted Membrane Protein, important for the delivery of peptides to MHC class I molecules (61). *LRMP* is

differentially expressed in HIV-1 infected lymph nodes from patients before and after ART initiation, compared with untreated controls (82). Interestingly, HIV-infected patients who underwent ART interruption were found to have differential *LRMP* gene expression between HIV-infected individuals with slow versus rapid CD4 decline during treatment interruption (83). Given its role in peptide delivery to MHC class I molecules, *LRMP* may play an important role in CTL recognition of HIV-infected cells (61). The remaining SNPs showed suggestive trends with tDNA or usRNA (**Table 1**). *AKAP6* encodes for A-Kinase Anchoring Protein 6, which binds to regulatory subunits of protein kinase A and anchors them to the nuclear membrane or sarcoplasmic reticulum. PKA activation is related to many cell cycle functions and has been associated with HIV-1 infection (84, 85), T cell proliferation (86), and HIV persistence (87, 88). Among males with European ancestry, we also observed an association between tDNA ans a SNP in *OSBP*, a gene that encodes oxysterol-binding protein, implicated in cholesterol metabolism and anti-HIV viral defenses HIV (65, 89). *OSBP* blocked HIV entry into host cells when its interaction with another key protein regulating intracellular cholesterol homeostasis, vesicle-associated membrane protein-associated protein A (VAPA) was disrupted (90). In a study comparing highly exposed seronegative (HESN) individuals to “lower risk” HIV-uninfected controls, HESNs demonstrated significantly higher expression of *OSBP*, which correlated with a reduced susceptibility to *in vitro* HIV-1 infection of PBMCs and monocyte-derived macrophages (MDMs) (65). Finally, a SNP in *PLVAP*, which encodes Plasmalemma Vesicle Associated Protein, which regulates the transmigration of lymphocytes into lymph nodes (66) and lies <30kb of *BST2*, (encodes for the host HIV restriction factor, tetherin (67)).

Gene set enrichment analyses identified one pathway, “Negative Regulation of Viral Entry into Host Cells,” that was significantly associated HIV intact DNA (**Figure 6, Table S8**). A non-significant trend was observed between usRNA and a somewhat related gene set, “Single-Stranded Viral RNA Replication via Double-Stranded DNA Intermediate.” The transcriptionally and translationally “active reservoir” has become an increasingly important component of the

HIV reservoir (91). Perhaps these top two gene sets associated with intact DNA and usRNA, respectively, reflect genes reflecting a role in this “active reservoir” (e.g., the genes contributing to statistical significance for the intact DNA-associaged gene set included several interferon stimulated genes and restriction factors, such as *IFITM1, IFITM3, TRIM5,* etc., **Table S8**). Previously, the HIV reservoir was thought to consist mostly of quiescent, latently-infected cells that could reactivate and lead to systemic infection with interruption of ART (32, 92, 93). New methods, combining HIV integration, viral sequencing, and HIV outgrowth assays, now suggest that transcriptionally and translationally active cells might be a major source of residual viremia during suppressive ART (92), and these “active” cells can persist for years (previously thought to be short-lived and cleared by immune recognition/cytotoxic response) (29-32). Several additional gene sets demonstrated a suggestive association with intact DNA – in particular two gene sets associated with interferon-stimulated pathway, STAT signaling (94), “Regulation of Peptidyl Serine Phosphorylation of STAT Protein” and “Serine Phosphorylation of STAT Protein.” Finally, a gene set reflecting glycosylation (“N-glycan processing”), showed a non-significant association with HIV tDNA. There have been recent studies suggesting an important role for glycoprotein metabolism in the maintenance of the HIV reservoir (95, 96). Glycosylation is the process by which proteins with covalently attached glycans (oligosaccharide chains) are (post-translationally) modified. Indeed, the composition of glycans in the IgG’s Fc region has been shown to modify functional and FcγR binding properties in HIV-infected patients (97, 98). N-glycan processing is associated with HIV proliferation, as the envelope (Env) glycoprotein of HIV is densely glycosylated with N-glycans and has been shown to modulate host antibody response to the HIV virus (99, 100). The immunologically and/or clinically relevant associations with HIV persistence were less clear for the remaining gene sets demonstrating a slight trend with reservoir size (all with HIV intact DNA). These included gene sets associated with bone formation, fatty acid metabolism, “cellular response to lead” (genes associated with changes in Th1 and Th2 responses due to lead toxicity (101, 102)).

There are several limitations to our study that deserve mention. Our study design is cross-sectional, and thus our findings may not necessarily reflect genetic variants that predict reservoir decay over time. Thus, although, the HIV reservoir has been shown to be relatively stable over time (38, 103, 104), our cross-sectional design provides a “snapshot” of the HIV reservoir after a median of 5.1 years of ART suppression. Second, the direction of the observed associations between host DNA variants and the HIV reservoir cannot be made from whole exome sequencing and HLA typing data alone; further studies are needed to confirm the functional relevance of the observed genetic variants. The exception to this is *MX1*, for which several SNPs localized to eQTL regions (55), demonstrating parallel direction of the predicted *MX1* gene expression and the size of the HIV reservoir. As is typical for current U.S.-based HIV+ cohorts, our study population consisted primarily of individuals (mostly males) of European ancestry (**Figure 4**). We performed all analyses adjusted for ancestry using established methods (105) to avoid false positive associations secondary to population stratification (106). We also performed secondary analyses within the largest genetically homogeneous group, individuals of European ancestry. Further studies are needed to validate these findings in non-European populations, especially including more female participants. In addition, like most HIV genomewide studies to date, our sample size was smaller than most host genomic cancer or cardiovascular disease registry-based studies. To address this, we carefully selected a range of HIV+ ART-suppressed non-controllers, including those who initiated ART during acute infection, and comprehensively fit separate linear mixed models for each of the four HIV reservoir measures, including covariates to further improve the fit of the models (e.g., timing of ART intiation, nadir CD4+ T cell count, etc.). Although we excluded HIV elite controllers, we included a range of HIV non-controllers, including individuals who may have had at least one documented HIV viral load <10,000 copies/mL prior to ART initiation. These individuals were included in the study in order to maximize the dynamic range of HIV host responses (**Figure S1**). Indeed, the maximum pre-ART viral loads did not correlate with lower HIV reservoir

measures for these individuals. To further ensure that inadvertent outlier bias was not present (e.g., as would be the case if HIV elite controllers were included), we performed two additional SNP-based analyses: a sensitivity analysis removing these individuals (**Figure S2a-b**) and a "case-control" analysis, comparing these individuals to the rest of the study population (**Figure S2c**). The first analysis demonstrated that the top SNP associations were overall unchanged (Spearman R=0.79 for the correlation of top hit SNPs with and without these individuals). The second analysis demonstrated that there were no statistically significantly SNPs by "possible partial controller" status, in association with HIV reservoir size. Interestingly, several of the top (non-significant) SNPs identified from within the "possible partial controller" group were in or within 50 kb of immunologically relevant genes (**Table S9**), such as *BRINP1*, *SNX29*, *IL1RAPL2*, *MX1*, *NFATC3*, *RIPK2*, *MASP2*, *TRIM6-TRIM34*, *HLA-A*, and *RNH1*, the latter of which was reported as the top hit SNP from a recent genomewide association studiy of HIV reservoir size, as noted above (24). Finally, we only measured the peripheral CD4+ T cell reservoir, not the tissue reservoir, where the majority of HIV persists during treatment (107). Although the tissue HIV reservoir has been shown to reflect the peripheral blood HIV reservoir (108-110), and a recent study demonstrated that rebounding virus subsequently seeds tissue reservoirs throughout the body (111), it remains to be seen whether the genetic variants identified in our study hold true for the tissue HIV reservoir.

Our study also has several strengths. It is the first host genomic study to evaluate HIV intact DNA, an estimate of the putative "replication-competent" reservoir (41, 42), as well as HIV total DNA, unspliced RNA, and RNA/DNA from peripheral CD4+ T cells. Here, we estimated the frequency of HIV intact DNA which quantified absolute copies of HIV intact DNA using a droplet digital PCR-based assay. Our study performed direct HLA typing via targeted next generation sequencing of 25 polymorphic exons of HLA-A, B, C, E, DPA1, DPB1, DQA1, DQB1, DRB1, and DRB3, 4, 5 genes on the Illumina MiSeq platform, allowing highly accurate HLA characterization. We also performed custom whole exome sequencing to include rare and

commonly observed variants, which included over 50 megabases of intronic and regulatory region sequencing of putative candidate genes involved in HIV persistence (**Table S2**). The UCSF SCOPE and Options cohort also collects detailed clinical data, including key variables known to impact the size of the HIV reservoir, such as estimated date of HIV infection (adjudicated clinical test results – such as date of last HIV negative test and date and type of positive incidence assay tests), date of ART initiation, and nadir CD4+ T cell count. Inclusion of these covariates in the final regression models may have enhanced model fit, enhancing the ability to detect the oberved genetic effects associated with HIV reservoir size. We also performed simultaneous RNA and DNA extraction from CD4+ T cells from a single PBMC aliquot. The relatively higher dynamic range of usRNA versus tDNA levels in our study population might be due to this dual extraction, in contrast to prior studies where HIV RNA and DNA extractions were performed from separate aliquots and often shown higher tDNA compared to usRNA values (**Figure S6**) (112).

We performed a comprehensive host genetic analysis of variants associated with the peripheral CD4+ T cell reservoir in treated HIV non-controllers. Unlike in a recent genomewide association study (23), we observed statistical significant associations between MHC class I alleles and *CCR5Δ32* with the HIV reservoir, which recapitulated protective effects previously described in untreated HIV elite controllers (9, 16, 19, 20, 27). We further identified several SNPs in the interferon stimulated gene, *MX1*, which were associated with the HIV reservoir (tDNA), paralleling predicted gene expression eQTL data. Gene set enrichment analyses highlighted potential genes associated with HIV intact DNA, including genes reflecting viral entry into host cell (e.g., interferon signaling genes). These findings provide further support for a potential role for the innate immune response (i.e., antiviral interferon signaling (29-32)) in maintaining the “active” HIV reservoir that persists during therapy (29-32). Additional studies are needed to functionally validate these findings, especially in non-European and female HIV-infected individuals.

2. Methods

Study participants

Participants were sampled from the San Francisco-based SCOPE and Options cohorts, which include HIV-infected individuals who initiated ART during chronic (>2 years) or early (<6 months) HIV infection. Inclusion criteria were laboratory-confirmed HIV-1 infection, availability of at least one aliquot of cryopreserved 10^6 PBMCs, and plasma HIV RNA levels below the limit of assay quantification for at least 24 months at the time of biospecimen collection. HIV "elite controllers," individuals with a history of undetectable viral load in the absence of therapy for at least 1 year prior to the specimen collection date (33-35), were excluded. Additional exclusion criteria included factors that could potentially influence HIV reservoir measures, including recent hospitalization, infection requiring antibiotics, vaccination, or exposure to immunomodulatory drugs in the six months prior to sampling timepoint. The research was approved by the University of California Committee on Human Research (CHR), and all participants provided written informed consent.

Custom whole exome host DNA sequencing

Cryopreserved PBMCs were enriched for CD4+ T cells (StemCell, Vancouver, Canada), and genomic DNA was extracted from CD4+ T cells (AllPrep Universal Kit, Qiagen, Hilden, Germany). Whole exome sequencing was using the extracted genomic host DNA. Using targeted exome capture with the SeqCap EZ Exome + UTR Library (Roche NimbleGen, Wilmington, MA) with custom addition of 50 Mb regulatory regions, exon-enriched sequencing libraries were generated and then sequenced on the Illumina HiSeq 2000 system (Illumina, Inc., San Diego, CA). Additional custom regions included intronic and regulatory regions (50 kilobases upstream and 50 kilobases downstream of each gene) for the analysis of potential

variants influencing key gene expression quantitative trait loci (113). The custom region included 442 candidate genes related to cell cycle regulation, HIV host restriction factors, and HIV-host integration, which were selected based on Gene Ontology (GO) Consortium experimental evidence codes (EXP, IDA, IPI, IMP, IGI, IEP) indicating experimental evidence directly supporting the annotations (http://geneontology.org/docs/guide-go-evidence-codes/) (**Table S2**), as well as previously associated HLA variants associated with viral setpoint and/or HIV disease progression (114, 115).

HLA typing

HLA typing was performed from extracted genomic DNA from cryopreserved PBMCs, following the PCR-SSOP (sequence-specific oligonucleotide probing) typing and PCR-SBT (sequence based typing) protocols recommended by the 13th International Histocompatibility Workshop (116, 117). Targeted sequencing using locus-specific primers were used to amplify a total of 25 polymorphic exons of HLA-A & B (exons 1 to 4), C (exons 1 to 5), E (exon 3), DPA1 (exon 2), DPB1 (exons 2 to 4), DQA1 (exon 1 to 3), DQB1 (exons 2 & 3), DRB1 (exons 2 & 3), and DRB3, 4, 5 (exon 2) genes with Fluidigm Access Array (Fluidigm Singapore PTE Ltd, Singapore). The 25 Fluidigm PCR amplicons are pooled and subjected to sequencing on an Illumina MiSeq sequencer (Illumina, San Diego, CA 92122 USA). HLA alleles and genotypes are called using the Omixon HLA Explore (version 2.0.0) software (Omixon, Budapest, Hungary).

HIV total DNA and unspliced RNA from peripheral CD4+ T cells

Cryopreserved PBMCs were enriched for CD4+ T cells using magnetic bead separation (StemCell Technologies) and host genomic DNA extracted from purified CD4+ T cells as described above (AllPrep Universal Kit, Qiagen, Hilden, Germany). Participant specimens were

assayed with up to 800 ng total cellular RNA or DNA in replicate reaction wells and copy number determined by extrapolation against a 7-point standard curve (1–10,000 copies/second) performed in triplicate.

HIV intact DNA from peripheral CD4+ T cells

Cell-associated HIV intact DNA was quantified from purified CD4+ T cells as above. Five regions on the HIV genome, including highly conserved regions, and positions that are frequently deleted or hypermutated, were interrogated to quantify potentially intact provirus (26). Optimized restriction enzyme digestion was used to prepare the genomic DNA for droplet formation while minimizing the amount of shearing within the viral genome. The protocol targeted 5 regions in the HIV genome across two droplet digital (dd)PCR assays. Droplet generation and thermocycling were performed according to manufacturer instructions. This multiplex ddPCR assay allowed the analysis of potentially replication-competent ("intact") proviral genomes by quantifying the number of droplets positive for 3 targets per assay. Two targets in a housekeeping gene (*RPP30)* were used to quantify all cells, and a target in the T cell receptor D gene (*TRD)* was used to identify all non-T cells, in order to normalize the HIV copy numbers per $1x10^6$ CD4+ T cells. A DNA shearing index (DSI) was then calculated and mathematically corrected for residual DNA shearing as measured by *RPP30* targets to calculate the estimated number of intact proviral genomes per million CD4+ T cells after correcting for shearing.

Data processing and quality control

The bcbio bioinformatics pipeline (118) was used to perform DNA alignment, which included the Burroughs-Welcome Aligner (BWA) tool (119) and the GenomeAnalysisToolkit (GATK) HaplotypeCaller joint variant calling method (120). Reads were initially mapped to reference

genome b37, then transposed to human genome assembly GRCh38 using Picard tools (121). SNPs and insertions or deletions (indels) were then filtered by variant quality score recalibration (VQSR) using GATK (122). The whole genome data analysis toolset, PLINK (123), was then used to validate the chromosomal sex of each individual, filter out individuals with excessive heterozygosity, and SNPS violating Hardy-Weinberg equilibrium (HWE) at a p-value threshold of 1×10^{-8}. The VCFtools suite of functions were then used to summarize data, run calculations, convert data, and filter out data, and convert data, and filter out relevant SNPs (124).

The GENESIS analysis pipeline (105) was used to analyze the relatedness and ancestries of the individuals in the study. All individuals were determined to be unrelated (kinship estimates < 0.05) aside from one pair of siblings, so one sibling was randomly removed from the study. The remaining 199 unrelated individuals had diverse and mixed ancestries, as demonstrated in **Figure 4a**. To attempt to increase the power of our study, we identified the largest homogeneous subgroup from the first two principal components (PCs), a group of 126 men with European ancestry (**Figure 4b**). We performed the rest of our analysis in parallel on the full 199 individuals from diverse ancestries and genders and the 126 males with European ancestry. We used principal component analysis (PCA) to account for genetic heterogeneity in two more ways: First, our model includes a genetic effects term with a genetic relatedness matrix (GRM) to account for relatedness. Second, we included the first five PCs as covariates in our models to account for population-level effects.

Single SNP common variant analyses

Single SNP associations were calculated with GENESIS "assocTestSingle". For HIV total DNA, unspliced RNA, RNA/DNA, and intact DNA, respectively, the outcome variables were: $\log_{10}$((DNA copies/10^6 CD4+ T cells + offset)); $\log_{10}$((RNA copies/10^6 CD4+ T cells + offset));

$\log_{10}$((RNA copies/10^6 CD4+ T cells + offset) / (DNA copies/10^6 CD4+ T cells + offset)); $\log_{10}$((Intact DNA copies/10^6 CD4+ T cells + offset)). The offsets for RNA and DNA counts were given by the smallest nonzero measured values of RNA and DNA, respectively, to avoid divergences in the logarithm. Covariates include sex for the full population, time to antiretroviral therapy (ART) initiation, nadir CD4 count, and 5 genetic principal components (PCs). A Gaussian link function was used, and a genetic relatedness matrix (GRM) was included as calculated by PC-AiR and PC-Relate. Results were filtered to allele frequencies ≥5%. SNP annotations were then obtained using Annovar, a bioinformatics software tool to interpret SNPS, insertions, deletions, and copy number variants in a genome (125).

Gene-based rare variant analyses

Gene level rare variant multi-SNP associations were calculated with the GENESIS software package "assocTestAggregate" function implementing the variant Set Mixed Model Association Test (SMMAT), an adjusted gene-based test combined with a burden test (68), for alleles with frequencies < 5% with weights following the beta distribution parameters of a_1=1 and a_2=25 as suggested by Ref (126). Genes with a total of fewer than 5 alternate alleles were omitted. The same covariates, GRM, and regression family were used as for the single SNP associations. Outcomes were quantile-normalized to follow a normal distribution. Gene regions were defined according to the intersection of Bioconductor's UCSC hg38 genes(127). Exons for genes with custom sequencing were defined according to (Roche NimbleGen, Wilmington, MA). Exons for genes without custom sequencing were defined according to the intersection of exonic regions defined by Bioconductor and (Roche NimbleGen, Wilmington, MA).

Gene set enrichment analyses (GSEA) were performed using the Molecular Signatures Database (MSigDB) (47, 128). We used unsigned SMMAT p-values and focused on Gene

Ontology (GO) Biological Processes (BP) gene sets that were up-regulated for low p-values. For all gene set analyses, introns and flanking regions of ±50kb were included in the SMMAT p-value calculations for each gene to account for potential regulatory regions and SNPs with smaller effects. GSEAPreranked was run with default parameters on the SMMAT gene-level $-log_{10}(P)$.

HLA analysis

Multivariate regression models were fit using the python statsmodels OLS function (129) with covariates for sex, timing of ART initiation, nadir CD4+ T cell count, and 3 genetic PCs.

FIGURE LEGENDS

Figure 1. Study participant sample selection flowchart. Specific inclusion and exclusion criteria are listed for each selection step and for HIV reservoir measure analysis.

Figure 2. HIV total DNA, unspliced RNA, RNA/DNA, and intact DNA were associated with timing of ART initiation. Spearman correlation and p-values are shown in the inset boxes.

Figure 3. Distribution of the total number of host variants (a), as well as all exonic variants (b) sequenced in the study population.

Figure 4. Genetic principal component analysis (PCA) plots of the full study population (with legend with self-identified race/ethnicity) (a) and of the largest homogenous ancestral population, males of mostly European ancestry (b). Recalculated European ancestry male PCA plot is shown in panel b, from lower left dashed box in panel a.

Figure 5. Quantile-quantile (QQ) plots (a, c) and Manhattan plots (b, d) of the total study population (a-b) and of European ancestry (c-d). QQ plots: the blue line represents the expected -log10 p-values while the black lines denote the expected error bars. Manhattan plots: the horizontal black line delineates a traditional conservative genome-wide significance of p-value of $5x10^{-8}$, while less conservative Benjamini-Hochberg false discovery rate (FDR) statistical significance q-values are shown as blue (q=0.05) and grey (q=0.25) lines.

Figure 6. Gene set enrichment analysis (GSEA) was used to identify associations between genes, ordered by p-value under a host multi-SNP rare variant (minor allele frequency, MAF, <5%) analysis using HIV total DNA, unspliced RNA, RNA/DNA, and intact DNA as outcomes;

and biological processes gene sets, in the total study population (a) and among European ancestry (b). Horizontal dashed lines represent the GSEA Benjamini-Hochberg false discovery rate (FDR) statistical significance level of q=0.05 (blue) and q=0.025 (grey), respectively.

REFERENCES

1. Causes of death in HIV-1-infected patients treated with antiretroviral therapy, 1996-2006: collaborative analysis of 13 HIV cohort studies. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2010;50(10):1387-96.
2. Eisele E, Siliciano RF. Redefining the viral reservoirs that prevent HIV-1 eradication. Immunity. 2012;37(3):377-88.
3. Lohse N, Hansen AB, Pedersen G, Kronborg G, Gerstoft J, Sorensen HT, et al. Survival of persons with and without HIV infection in Denmark, 1995-2005. Annals of internal medicine. 2007;146(2):87-95.
4. Volberding PA, Deeks SG. Antiretroviral therapy and management of HIV infection. Lancet. 2010;376(9734):49-62.
5. Chun TW, Justement JS, Moir S, Hallahan CW, Maenza J, Mullins JI, et al. Decay of the HIV reservoir in patients receiving antiretroviral therapy for extended periods: implications for eradication of virus. J Infect Dis. 2007;195(12):1762-4.
6. Hocqueloux L, Avettand-Fenoel V, Jacquot S, Prazuck T, Legac E, Melard A, et al. Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts. J Antimicrob Chemother. 2013;68(5):1169-78.
7. Cohn LB, Chomont N, Deeks SG. The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. Cell Host Microbe. 2020;27(4):519-30.
8. 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(12):e3907.
9. 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(5840):944-7.
10. Fellay J, Frahm N, Shianna KV, Cirulli ET, Casimiro DR, Robertson MN, et al. Host genetic determinants of T cell responses to the MRKAd5 HIV-1 gag/pol/nef vaccine in the step trial. J Infect Dis. 2011;203(6):773-9.
11. 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(12):e1000791.
12. Herbeck JT, Gottlieb GS, Winkler CA, Nelson GW, An P, Maust BS, et al. Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. J Infect Dis. 2010;201(4):618-26.
13. Le Clerc S, Limou S, Coulonges C, Carpentier W, Dina C, Taing L, et al. Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). J Infect Dis. 2009;200(8):1194-201.
14. 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(3):419-26.
15. Pelak K, Goldstein DB, Walley NM, Fellay J, Ge D, Shianna KV, et al. Host determinants of HIV-1 control in African Americans. J Infect Dis. 2010;201(8):1141-9.
16. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010;330(6010):1551-7.
17. Petrovski S, Fellay J, Shianna KV, Carpenetti N, Kumwenda J, Kamanga G, et al. Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. AIDS. 2011;25(4):513-8.
18. de Roda Husman AM, Koot M, Cornelissen M, Keet IP, Brouwer M, Broersen SM, et al. Association between CCR5 genotype and the clinical course of HIV-1 infection. Ann Intern Med. 1997;127(10):882-90.
19. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic

restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science. 1996;273(5283):1856-62.

20. Rappaport J, Cho YY, Hendel H, Schwartz EJ, Schachter F, Zagury JF. 32 bp CCR-5 gene deletion and resistance to fast progression in HIV-1 infected heterozygotes. Lancet. 1997;349(9056):922-3.

21. Kulkarni S, Lied A, Kulkarni V, Rucevic M, Martin MP, Walker-Sperling V, et al. CCR5AS lncRNA variation differentially regulates CCR5, influencing HIV disease outcome. Nat Immunol. 2019;20(7):824-34.

22. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013;132(10):1077-130.

23. Thorball CW, Borghesi A, Bachmann N, Von Siebenthal C, Vongrad V, Turk T, et al. Host Genomics of the HIV-1 Reservoir Size and Its Decay Rate During Suppressive Antiretroviral Treatment. J Acquir Immune Defic Syndr. 2020;85(4):517-24.

24. Zhang Z, Trypsteen W, Blaauw M, Chu X, Rutsaert S, Vandekerckhove L, et al. IRF7 and RNH1 are modifying factors of HIV-1 reservoirs: a genome-wide association analysis. BMC Med. 2021;19(1):282.

25. Birge RB, Boeltz S, Kumar S, Carlson J, Wanderley J, Calianese D, et al. Phosphatidylserine is a global immunosuppressive signal in efferocytosis, infectious disease, and cancer. Cell Death Differ. 2016;23(6):962-78.

26. Levy CH, S.; Roychoudhury, P.; Reeves, D. B.; Amstuz, C.; Zhu, H.; et al. A highly multiplexed droplet digital PCR assay to measure the intact HIV-1 proviral reservoir. Cell Reports Medicine. 2021;2.

27. de Roda Husman AM, Koot M, Cornelissen M, Keet IP, Brouwer M, Broersen SM, et al. Association between CCR5 genotype and the clinical course of HIV-1 infection. Annals of internal medicine. 1997;127(10):882-90.

28. Gondim MVP, Sherrill-Mix S, Bibollet-Ruche F, Russell RM, Trimboli S, Smith AG, et al. Heightened resistance to host type 1 interferons characterizes HIV-1 at transmission and after antiretroviral therapy interruption. Sci Transl Med. 2021;13(576).

29. Pace MJ, Agosto L, Graf EH, O'Doherty U. HIV reservoirs and latency models. Virology. 2011;411(2):344-54.

30. McManus WR, Bale MJ, Spindler J, Wiegand A, Musick A, Patro SC, et al. HIV-1 in lymph nodes is maintained by cellular proliferation during antiretroviral therapy. J Clin Invest. 2019;129(11):4629-42.

31. Kearney MF, Wiegand A, Shao W, Coffin JM, Mellors JW, Lederman M, et al. Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. J Virol. 2016;90(3):1369-76.

32. Wiegand A, Spindler J, Hong FF, Shao W, Cyktor JC, Cillo AR, et al. Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. Proc Natl Acad Sci U S A. 2017;114(18):E3659-E68.

33. Ramirez CM, Sinclair E, Epling L, Lee SA, Jain V, Hsue PY, et al. Immunologic profiles distinguish aviremic HIV-infected adults. AIDS. 2016;30(10):1553-62.

34. Boufassa F, Lechenadec J, Meyer L, Costagliola D, Hunt PW, Pereyra F, et al. Blunted response to combination antiretroviral therapy in HIV elite controllers: an international HIV controller collaboration. PLoS One. 2014;9(1):e85516.

35. Emu B, Sinclair E, Hatano H, Ferre A, Shacklett B, Martin JN, et al. HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. J Virol. 2008;82(11):5398-407.

36. Robb ML, Eller LA, Kibuuka H, Rono K, Maganga L, Nitayaphan S, et al. Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand. N Engl J Med. 2016;374(22):2120-30.

37. Eriksson S, Graf EH, Dahl V, Strain MC, Yukl SA, Lysenko ES, et al. Comparative analysis of

measures of viral reservoirs in HIV-1 eradication studies. PLoS pathogens. 2013;9(2):e1003174.
38. Bachmann N, von Siebenthal C, Vongrad V, Turk T, Neumann K, Beerenwinkel N, et al. Determinants of HIV-1 reservoir size and long-term dynamics during suppressive ART. Nat Commun. 2019;10(1):3193.
39. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat Med. 2016;22(9):1043-9.
40. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155(3):540-51.
41. Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. Nature. 2019;566(7742):120-5.
42. Levy CN, Hughes SM, Roychoudhury P, Reeves DB, Amstuz C, Zhu H, et al. A highly multiplexed droplet digital PCR assay to measure the intact HIV-1 proviral reservoir. Cell Rep Med. 2021;2(4):100243.
43. Malnati MS, Scarlatti G, Gatto F, Salvatori F, Cassina G, Rutigliano T, et al. A universal real-time PCR assay for the quantification of group-M HIV-1 proviral load. Nat Protoc. 2008;3(7):1240-8.
44. Kulpa DA, Lawani M, Cooper A, Peretz Y, Ahlers J, Sekaly RP. PD-1 coinhibitory signals: the link between pathogenesis and protection. Semin Immunol. 2013;25(3):219-27.
45. Abdel-Mohsen M, Wang C, Strain MC, Lada SM, Deng X, Cockerham LR, et al. Select host restriction factors are associated with HIV persistence during antiretroviral therapy. AIDS. 2015;29(4):411-20.
46. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science. 2014;345(6193):179-83.
47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
48. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011;19(7):807-12.
49. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1-2):279-84.
50. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-5.
51. Yukl S, Wong JK. Blood and guts and HIV: preferential HIV persistence in GI mucosa. J Infect Dis. 2008;197(5):640-2.
52. Telwatte S, Lee S, Somsouk M, Hatano H, Baker C, Kaiser P, et al. Gut and blood differ in constitutive blocks to HIV transcription, suggesting tissue-specific differences in the mechanisms that govern HIV latency. PLoS Pathog. 2018;14(11):e1007357.
53. Staeheli P, Haller O. Human MX2/MxB: a Potent Interferon-Induced Postentry Inhibitor of Herpesviruses and HIV-1. J Virol. 2018;92(24).
54. Haller O, Kochs G. Human MxA protein: an interferon-induced dynamin-like GTPase with broad antiviral activity. J Interferon Cytokine Res. 2011;31(1):79-87.
55. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-13.
56. Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature. 2013;502(7472):559-62.
57. Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, et al. The interferon-inducible MxB protein inhibits HIV-1 infection. Cell Host Microbe. 2013;14(4):398-410.
58. Betancor G, Jimenez-Guardeno JM, Lynham S, Antrobus R, Khan H, Sobala A, et al. MX2-mediated innate immunity against HIV-1 is regulated by serine phosphorylation. Nat Microbiol. 2021;6(8):1031-42.

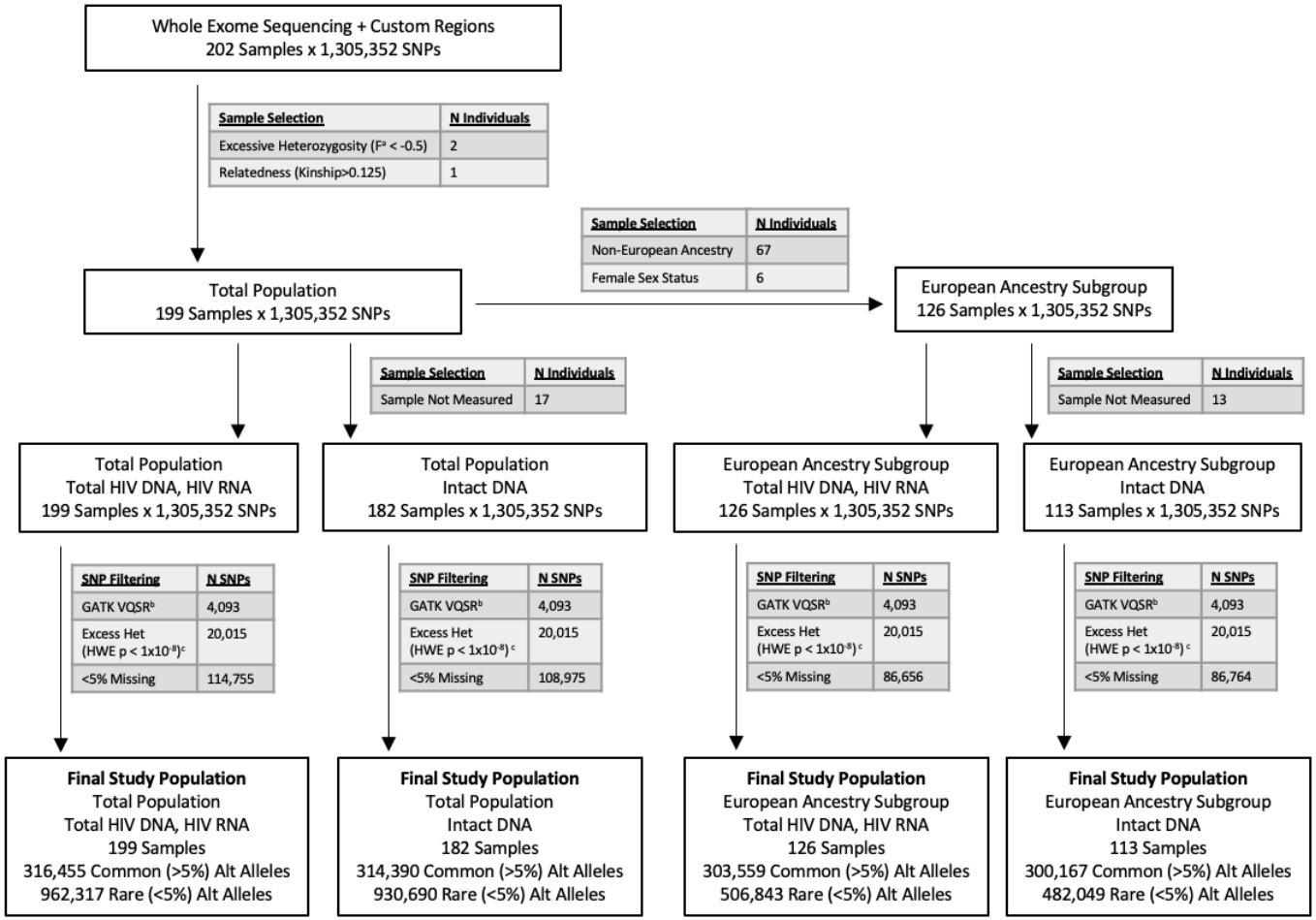
59. Lee SY, Choi BS, Yoon CH, Kang C, Kim K, Kim KC. Selection of biomarkers for HIV-1 latency by integrated analysis. Genomics. 2019;111(3):327-33.
60. Verhelst J, Hulpiau P, Saelens X. Mx proteins: antiviral gatekeepers that restrain the uninvited. Microbiol Mol Biol Rev. 2013;77(4):551-66.
61. Collins KL. How HIV evades CTL recognition. Curr HIV Res. 2003;1(1):31-40.
62. Szappanos D, Tschismarov R, Perlot T, Westermayer S, Fischer K, Platanitis E, et al. The RNA helicase DDX3X is an essential mediator of innate antimicrobial immunity. PLoS Pathog. 2018;14(11):e1007397.
63. Perfettini JL, Roumier T, Castedo M, Larochette N, Boya P, Raynal B, et al. NF-kappaB and p53 are the dominant apoptosis-inducing transcription factors elicited by the HIV-1 envelope. J Exp Med. 2004;199(5):629-40.
64. Troger J, Moutty MC, Skroblin P, Klussmann E. A-kinase anchoring proteins as potential drug targets. Br J Pharmacol. 2012;166(2):420-33.
65. Saulle I, Ibba SV, Vittori C, Fenizia C, Mercurio V, Vichi F, et al. Sterol metabolism modulates susceptibility to HIV-1 Infection. AIDS. 2020;34(11):1593-602.
66. Rantakari P, Auvinen K, Jappinen N, Kapraali M, Valtonen J, Karikoski M, et al. The endothelial protein PLVAP in lymphatics controls the entry of lymphocytes and antigens into lymph nodes. Nat Immunol. 2015;16(4):386-96.
67. Perez-Caballero D, Zang T, Ebrahimi A, McNatt MW, Gregory DA, Johnson MC, et al. Tetherin inhibits HIV-1 release by directly tethering virions to cells. Cell. 2009;139(3):499-511.
68. Chen H, Huffman JE, Brody JA, Wang C, Lee S, Li Z, et al. Efficient Variant Set Mixed Model Association Tests for Continuous and Binary Traits in Large-Scale Whole-Genome Sequencing Studies. Am J Hum Genet. 2019;104(2):260-74.
69. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012;91(2):224-37.
70. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739-40.
71. Pollack RA, Jones RB, Pertea M, Bruner KM, Martin AR, Thomas AS, et al. Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. Cell Host Microbe. 2017;21(4):494-506 e4.
72. Thomas AS, Jones KL, Gandhi RT, McMahon DK, Cyktor JC, Chan D, et al. T-cell responses targeting HIV Nef uniquely correlate with infected cell frequencies after long-term antiretroviral therapy. PLoS Pathog. 2017;13(9):e1006629.
73. Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. Cell Rep. 2017;21(3):813-22.
74. Wang C, Abdel-Mohsen M, Strain MC, Lada SM, Yukl S, Cockerham LR, et al. Decreased HIV type 1 transcription in CCR5-Delta32 heterozygotes during suppressive antiretroviral therapy. J Infect Dis. 2014;210(11):1838-43.
75. Henrich TJ, Hanhauser E, Harrison LJ, Palmer CD, Romero-Tejeda M, Jost S, et al. CCR5-Delta32 Heterozygosity, HIV-1 Reservoir Size, and Lymphocyte Activation in Individuals Receiving Long-term Suppressive Antiretroviral Therapy. J Infect Dis. 2016;213(5):766-70.
76. Zhao G, Liu L, Su B, Zhang T, Chen P, Li W, et al. The dynamic changes of interferon lambdas related genes and proteins in JAK/STAT pathway in both acute and chronic HIV-1 infected patients. AIDS Res Ther. 2017;14(1):31.
77. Van Hecke C, Trypsteen W, Malatinkova E, De Spiegelaere W, Vervisch K, Rutsaert S, et al. Early treated HIV-1 positive individuals demonstrate similar restriction factor expression profile as long-term non-progressors. EBioMedicine. 2019;41:443-54.
78. Szabo PA, Levitin HM, Miron M, Snyder ME, Senda T, Yuan J, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat Commun.

2019;10(1):4706.
79. Wang S, Zhang Q, Hui H, Agrawal K, Karris MAY, Rana TM. An atlas of immune cell exhaustion in HIV-infected individuals revealed by single-cell transcriptomics. Emerg Microbes Infect. 2020;9(1):2333-47.
80. Ammosova T, Jerebtsova M, Beullens M, Voloshin Y, Ray PE, Kumar A, et al. Nuclear protein phosphatase-1 regulates HIV-1 transcription. J Biol Chem. 2003;278(34):32189-94.
81. Yedavalli VS, Neuveut C, Chi YH, Kleiman L, Jeang KT. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. Cell. 2004;119(3):381-92.
82. Li Q, Schacker T, Carlis J, Beilman G, Nguyen P, Haase AT. Functional genomic analysis of the response of HIV-1-infected lymphatic tissue to antiretroviral therapy. J Infect Dis. 2004;189(4):572-82.
83. Vahey MT, Wang Z, Su Z, Nau ME, Krambrink A, Skiest DJ, et al. CD4+ T-cell decline after the interruption of antiretroviral therapy in ACTG A5170 is predicted by differential expression of genes in the ras signaling pathway. AIDS Res Hum Retroviruses. 2008;24(8):1047-66.
84. Cartier C, Hemonnot B, Gay B, Bardy M, Sanchiz C, Devaux C, et al. Active cAMP-dependent protein kinase incorporated within highly purified HIV-1 particles is required for viral infectivity and interacts with viral capsid protein. J Biol Chem. 2003;278(37):35211-9.
85. Barnitz RA, Wan F, Tripuraneni V, Bolton DL, Lenardo MJ. Protein kinase A phosphorylation activates Vpr-induced cell cycle arrest during human immunodeficiency virus type 1 infection. J Virol. 2010;84(13):6410-24.
86. Skalhegg BS, Johansen AK, Levy FO, Andersson KB, Aandahl EM, Blomhoff HK, et al. Isozymes of cyclic AMP-dependent protein kinases (PKA) in human lymphoid cell lines: levels of endogenous cAMP influence levels of PKA subunits and growth in lymphoid cell lines. J Cell Physiol. 1998;177(1):85-93.
87. Lim H, Kim KC, Son J, Shin Y, Yoon CH, Kang C, et al. Synergistic reactivation of latent HIV-1 provirus by PKA activator dibutyryl-cAMP in combination with an HDAC inhibitor. Virus Res. 2017;227:1-5.
88. Giroud C, Chazal N, Gay B, Eldin P, Brun S, Briant L. HIV-1-associated PKA acts as a cofactor for genome reverse transcription. Retrovirology. 2013;10:157.
89. Roberts BL, Severance ZC, Bensen RC, Le-McClain AT, Malinky CA, Mettenbrink EM, et al. Differing activities of oxysterol-binding protein (OSBP) targeting anti-viral compounds. Antiviral Res. 2019;170:104548.
90. Amini-Bavil-Olyaee S, Choi YJ, Lee JH, Shi M, Huang IC, Farzan M, et al. The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. Cell Host Microbe. 2013;13(4):452-64.
91. Groebner J, Pérez L, Sklutuis R, Bale M, Shao W, Wiegand A, et al., editors. Resting and activated CD4+ T cells both have silent and active HIV proviruses in vivo. Abstract # 304. Spotlight E4. Conference on Retroviruses and Opportunistic Infections 2021; Boston, MA.
92. Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, et al. Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell. 2022;185(2):266-82 e15.
93. Yukl SA, Kaiser P, Kim P, Telwatte S, Joshi SK, Vu M, et al. HIV latency in isolated patient CD4(+) T cells may be due to blocks in HIV transcriptional elongation, completion, and splicing. Sci Transl Med. 2018;10(430).
94. Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. Nat Rev Immunol. 2005;5(5):375-86.
95. Vadrevu SK, Trbojevic-Akmacic I, Kossenkov AV, Colomb F, Giron LB, Anzurez A, et al. Frontline Science: Plasma and immunoglobulin G galactosylation associate with HIV persistence during antiretroviral therapy. J Leukoc Biol. 2018;104(3):461-71.
96. Bartsch YC, Loos C, Rossignol E, Fajnzylber JM, Yuan D, Avihingsanon A, et al. Viral Rebound Kinetics Correlate with Distinct HIV Antibody Features. mBio. 2021;12(2).
97. Ackerman ME, Mikhailova A, Brown EP, Dowell KG, Walker BD, Bailey-Kellogg C, et al. Polyfunctional HIV-Specific Antibody Responses Are Associated with Spontaneous HIV Control. PLoS

Pathog. 2016;12(1):e1005315.
98. Chung AW, Kumar MP, Arnold KB, Yu WH, Schoen MK, Dunphy LJ, et al. Dissecting Polyclonal Vaccine-Induced Humoral Immunity against HIV Using Systems Serology. Cell. 2015;163(4):988-98.
99. Quinones-Kochs MI, Buonocore L, Rose JK. Role of N-linked glycans in a human immunodeficiency virus envelope glycoprotein: effects on protein function and the neutralizing antibody response. J Virol. 2002;76(9):4199-211.
100. Shen R, Raska M, Bimczok D, Novak J, Smith PD. HIV-1 envelope glycan moieties modulate HIV-1 transmission. J Virol. 2014;88(24):14258-67.
101. Heo Y, Lee WT, Lawrence DA. Differential effects of lead and cAMP on development and activities of Th1- and Th2-lymphocytes. Toxicol Sci. 1998;43(2):172-85.
102. Hsiao CL, Wu KH, Wan KS. Effects of environmental lead exposure on T-helper cell-specific cytokines in children. J Immunotoxicol. 2011;8(4):284-7.
103. Peluso MJ, Bacchetti P, Ritter KD, Beg S, Lai J, Martin JN, et al. Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. JCI Insight. 2020;5(4).
104. Falcinelli SD, Kilpatrick KW, Read J, Murtagh R, Allard B, Ghofrani S, et al. Longitudinal Dynamics of Intact HIV Proviral DNA and Outgrowth Virus Frequencies in a Cohort of Individuals Receiving Antiretroviral Therapy. J Infect Dis. 2021;224(1):92-100.
105. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35(24):5346-8.
106. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006;38(8):904-9.
107. Pantaleo G, Graziosi C, Butini L, Pizzo PA, Schnittman SM, Kotler DP, et al. Lymphoid organs function as major reservoirs for human immunodeficiency virus. Proc Natl Acad Sci U S A. 1991;88(21):9838-42.
108. Evering TH, Mehandru S, Racz P, Tenner-Racz K, Poles MA, Figueroa A, et al. Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiviral therapy initiated during primary infection. PLoS Pathog. 2012;8(2):e1002506.
109. Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, et al. The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. Proc Natl Acad Sci U S A. 2013;110(51):E4987-96.
110. Imamichi H, Degray G, Dewar RL, Mannon P, Yao M, Chairez C, et al. Lack of compartmentalization of HIV-1 quasispecies between the gut and peripheral blood compartments. J Infect Dis. 2011;204(2):309-14.
111. Chaillon A, Gianella S, Dellicour S, Rawlings SA, Schlub TE, De Oliveira MF, et al. HIV persists throughout deep tissues with repopulation from multiple anatomical sources. J Clin Invest. 2020;130(4):1699-712.
112. Lee SA, Bacchetti P, Chomont N, Fromentin R, Lewin SR, O'Doherty U, et al. Anti-HIV Antibody Responses and the HIV Reservoir Size during Antiretroviral Therapy. PLoS One. 2016;11(8):e0160192.
113. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010;6(4):e1000888.
114. Martin MP, Carrington M. Immunogenetics of HIV disease. Immunol Rev. 2013;254(1):245-64.
115. Martin MP, Naranbhai V, Shea PR, Qi Y, Ramsuran V, Vince N, et al. Killer cell immunoglobulin-like receptor 3DL1 variation modifies HLA-B*57 protection against HIV-1. J Clin Invest. 2018;128(5):1903-12.
116. 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(4):429-34.
117. 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(6):733-40.

118. Chapman Bea. bcbio/bcbio-nextgen. v.1.2.8 ed: Zenodo; 2021.
119. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997. 2013.
120. Poplin R, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2017;201178.
121. Picard toolkit 2019 [Available from: http://broadinstitute.github.io/picard/.
122. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8.
123. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.
124. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8.
125. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
126. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82-93.
127. Bioconductor Core Team BPMMBO. Txdb.hsapiens.ucsc.hg38.knowngene. 2017.
128. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34(3):267-73.
129. Seabold; Skipper; Perktold J, editor statsmodels: Econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference; 2010.

Figure 1. Study participant sample selection flowchart. Specific inclusion and exclusion criteria are listed for each selection step and for HIV reservoir measure analysis.



[a] F = inbreeding coefficient
[b] VQSR = Variant Quality Score Recalibration
[c] HWE = Hardy-Weinberg equilibrium

Figure 2. HIV total DNA, unspliced RNA, RNA/DNA, and intact DNA were associated with timing of ART initiation. Spearman correlation and p-values are shown in the inset boxes.

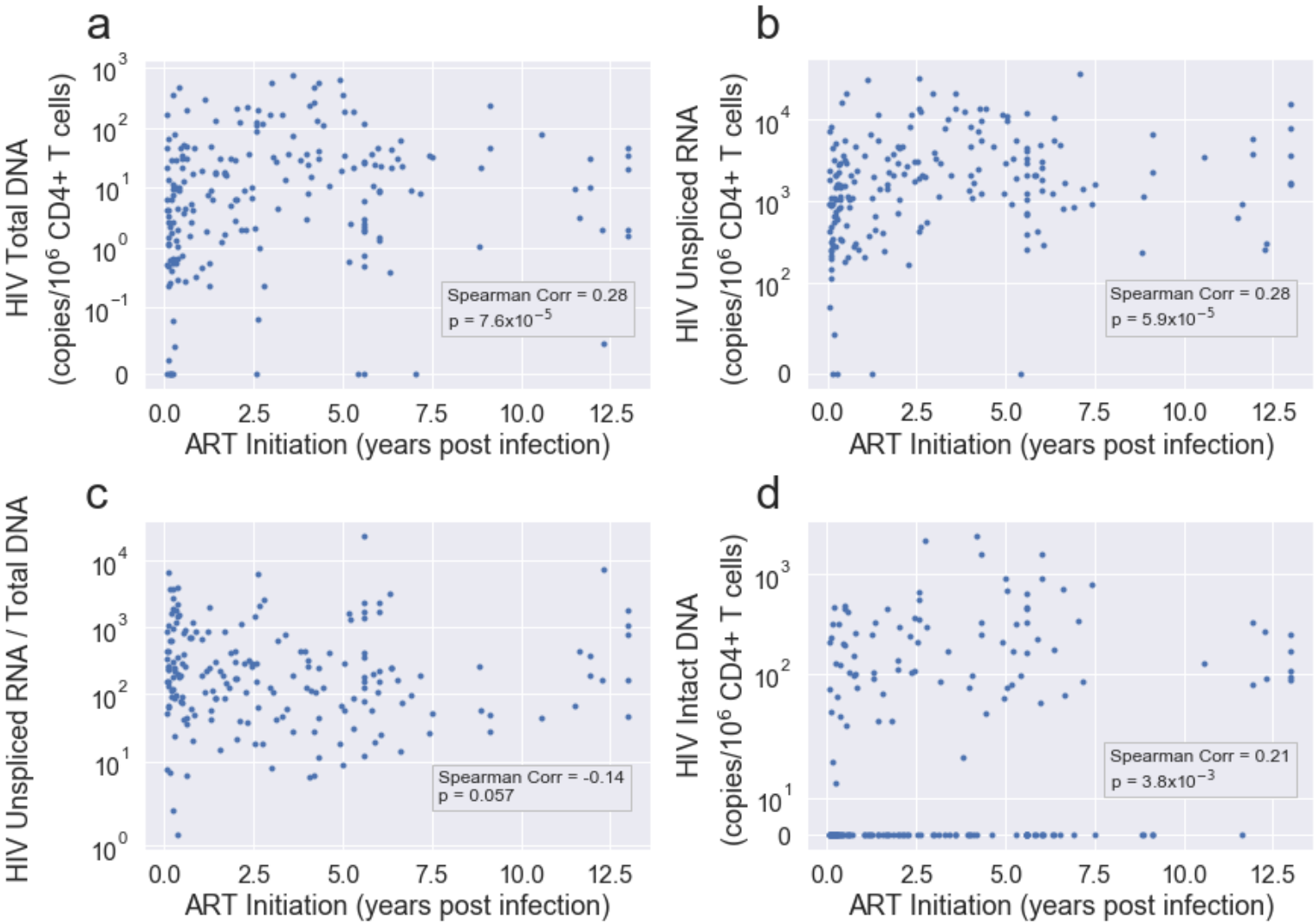


Figure 3. Distribution of the total number of host variants (a), as well as all exonic variants (b) sequenced in the study population.

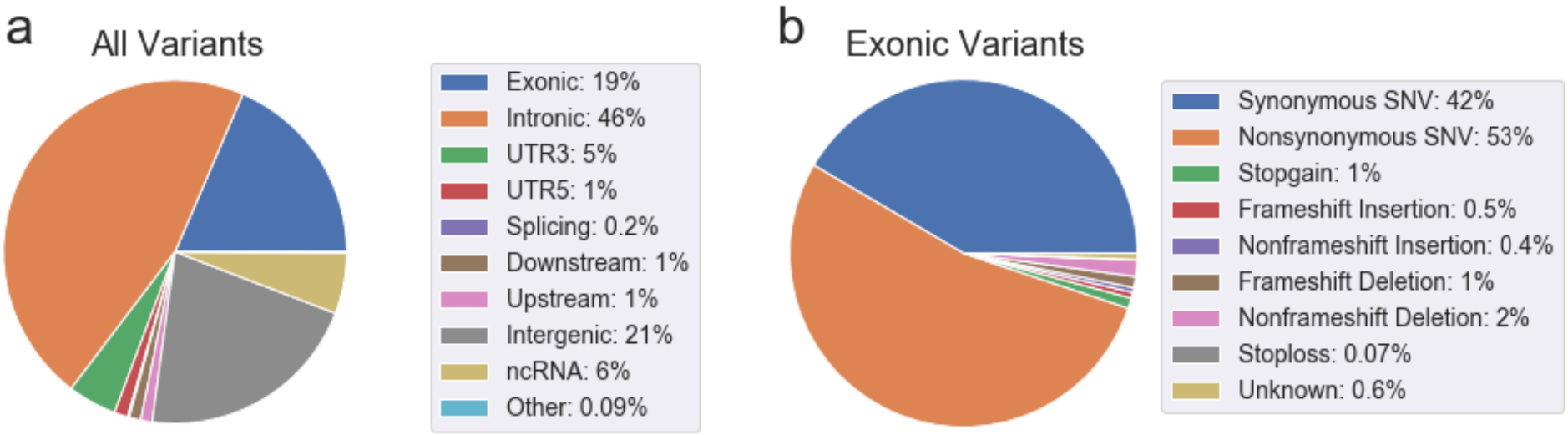


Figure 4. Genetic principal component analysis (PCA) plots of the full study population (with legend with self-identified race/ethnicity) (a) and of the largest homogenous ancestral population, participants of mostly European ancestry (b). Recalculated European ancestry PCA plot is shown in panel b, from lower left dashed box in panel a.

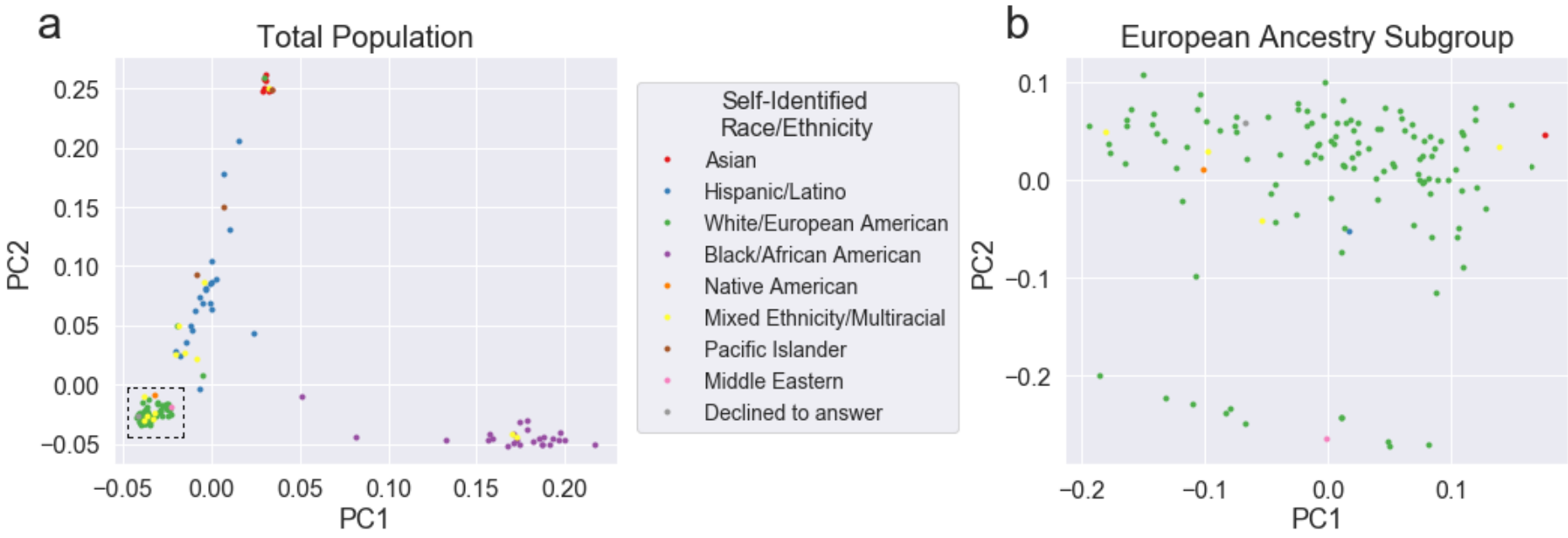


Figure 5. Quantile-quantile (QQ) plots (a, c) and Manhattan plots (b, d) of the total study population (a-b) and of European ancestry (c-d). QQ plots: the blue line represents the expected -$\log_{10}$ p-values while the black lines denote the expected error bars. Manhattan plots: the horizontal black line delineates a traditional conservative genome-wide significance of p-value of 5×10^{-8}, while less conservative Benjamini-Hochberg false discovery rate (FDR) statistical significance q-values are shown as blue (q=0.05) and grey (q=0.25) lines.

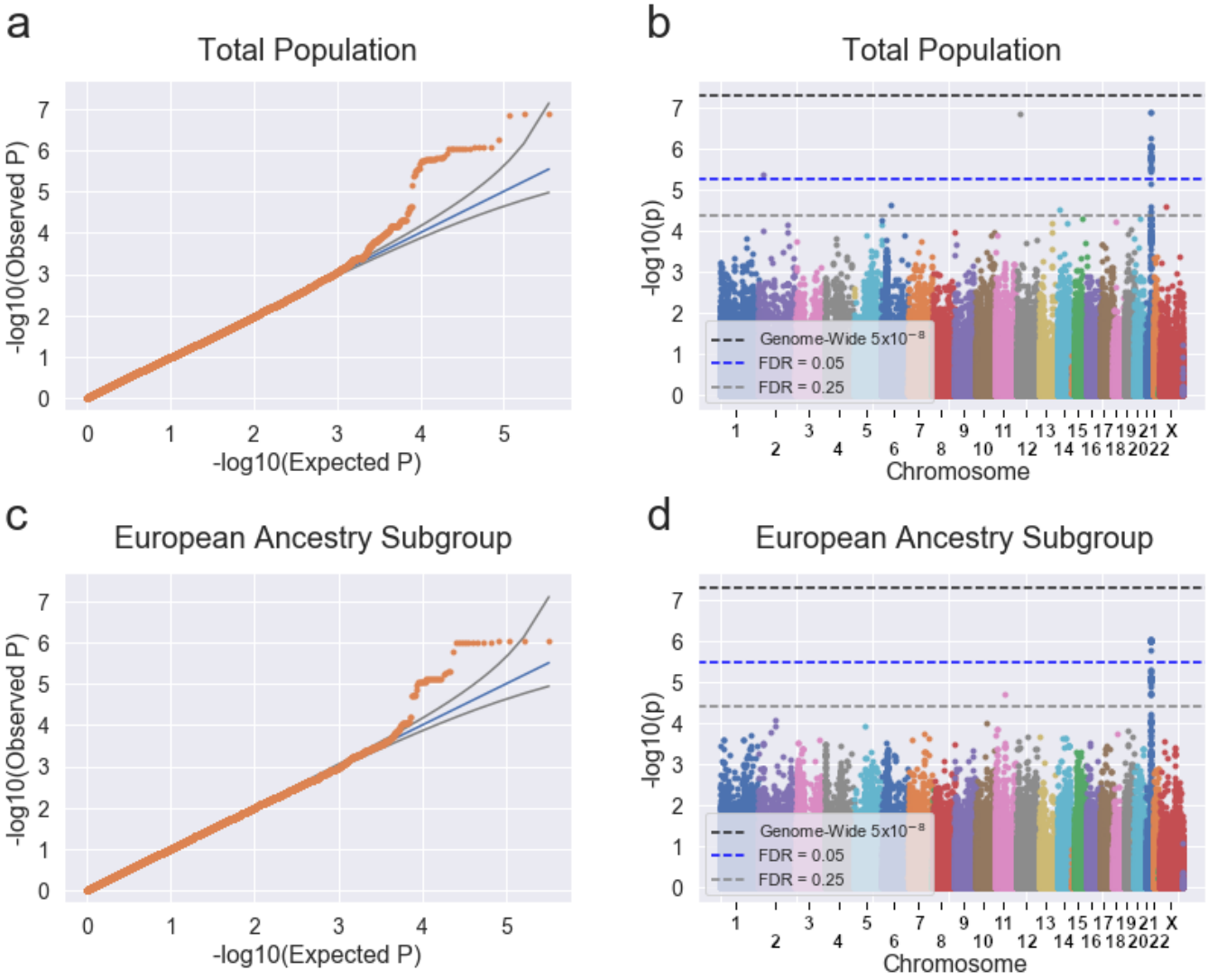


Table 1. Single nucleotide polymorphisms (SNPs) associated with HIV total DNA (upper table) and unspliced RNA (lower table) in the total study population. Two additional SNPs were significantly associated with total HIV DNA in the subpopulation of European ancestry (middle table). For genes in which there were several SNPs in linkage disequilibrium (LD), the top SNP for each gene is shown, with the full list of SNPs shown in **Table S7**.

SNP	Chrom	Position	Nearest Gene	Gene Location	MAF[a]	Beta[b]	PVE[c]	p[d]	q[e]	Description
HIV TOTAL DNA										
Full Cohort										
rs10670165	chr21	41421873	*MX1*	Intron/Exon/ 5'UTR[f]	0.45	-1.2	0.15	$1.3x10^{-7}$	0.02	Antiviral. Upregulated in HIV+ compared to uninfected individuals (76) and associated with higher viremia among HIV+ individuals (77). Associated with HIV-1 latency (59). MX2, but not MX1, was shown to be capable of directly acting against HIV-1 virus (56-57).
rs74867009	chr12	25063777	*LRMP*	5'UTR[f]	0.06	-2.5	0.15	$1.5x10^{-7}$	0.02	Delivers peptides to MHC class I molecules. (61) Differentially expressed in lymphatic tissue (82) and peripheral blood mononuclear cells (PBMCs) (83) in HIV+ individuals in response to ART initiation and cessation, respectively.
rs751660317	chr2	28786774	*PPP1CB*	Intronic	0.07	-1.8	0.11	$4.1x10^{-6}$	0.03	Encodes a subunit of PP1. Depletion of PPP1CB was shown to reduce the antiviral potency of MX2 against HIV (58). PP1 is involved in transcription of HIV-1; inhibition of PP1 inhibits HIV-1 transcription (80).
rs776025235	chr6	51638799	*PKHD1*	Intronic	0.09	-1.7	0.09	$2.5x10^{-5}$	0.17	Predicted to have a transmembrane-spanning domain and an immunoglobulin-like plexin-transcription-factor domain. We could not find a direct relationship with HIV in the literature.
N/A	chrX	41382082	*DDX3X*; *NYX*	Intergenic	0.18	-1.4	0.09	$2.5x10^{-5}$	0.17	*DDX3X* regulates the production of type I IFNs (62) and encodes a protein that shuttles HIV-1 RNA from the nucleus to the cytoplasm. Is upregulated in HIV-infected cells; knockdown of *DDX3X* suppresses HIV-1 replication (81). Also plays a key role in innate antimicrobial immunity (62).
rs17506750	chr14	32599402	*AKAP6*	Intronic	0.07	-1.8	0.09	$3.0x10^{-5}$	0.2	Binds to regulatory subunits of protein kinase A (PKA) and anchors them to the nuclear membrane. PKA activation has been associated with HIV-1 infection, T cell proliferation, and dysfunction (63-64).
European Ancestry Subgroup										
rs469390	chr21	41446003	*MX1*	Intron/Exon/ 5'UTR[f]	0.54	1.3	0.2	$1.0x10^{-6}$	0.03	Antiviral. Upregulated in HIV+ compared to uninfected individuals (76) and associated with higher viremia among HIV+ individuals (77). Associated with HIV-1 latency (59). MX2, but not MX1, was shown to be capable of directly acting against HIV-1 virus (56-57).
N/A	chr11	59574427	*OSBP*	3'UTR[f]	0.15	-1.8	0.15	$1.9x10^{-5}$	0.14	Oxysterol binding protein involved in intracellular lipid transport. Associated with *in vitro* HIV-1 infection of monocyte-derived macrophages (MDMs) from highly (HIV)-exposed seronegatives (HESNs) (65). Is required for the replication of several human viruses such as hepatitis C (HCV), encephalomyocarditis (EMCV), Zika, etc. (89).

HIV UNSPLICED RNA										
Full Cohort										
N/A	chr19	17374631	*PLVAP*	Intronic	0.1	-1.2	0.14	6.9×10^{-7}	0.21	Endothelial membrane protein, also controls the entry of lymphocytes and antigens into lymph nodes (66). Located within 30kb of *BST2*, an interferon stimulated gene encoding the host restriction factor tetherin, which is known to inhibit HIV-1 release by directly tethering virions to cells (67).

[a] MAF = minor allele frequency in the study population.
[b] Beta = estimate from multivariate linear mixed model adjusted for age, sex, nadir CD4+ T cell count, timing of ART initiation, and ancestry.
[c] PVE = proportion of phenotype variance explained.
[d] p = two sided p-value.
[e] q = two-sided false discovery rate (FDR) Benjamini-Hochberg q-value.
[f] UTR= untranslated region

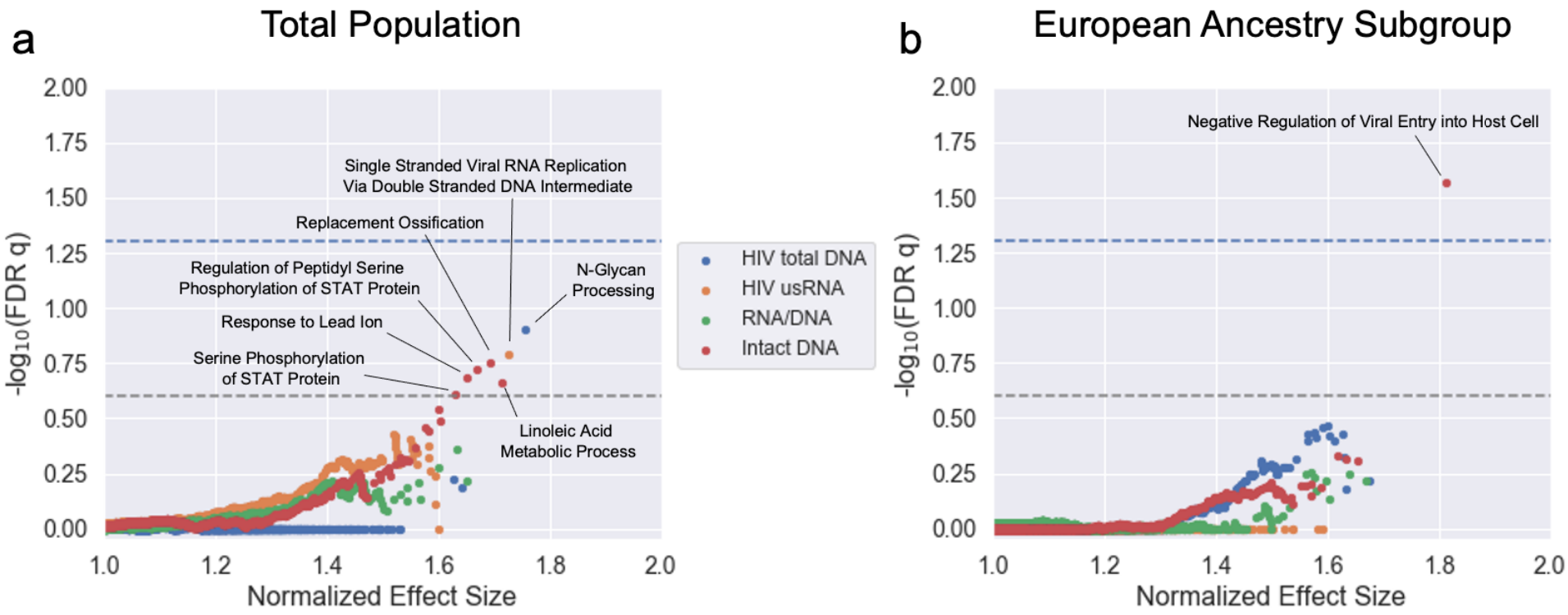



Figure 6. Gene set enrichment analysis (GSEA) was used to identify associations between genes, ordered by p-value under a host multi-SNP rare variant (minor allele frequency, MAF, <5%) analysis using HIV total DNA, unspliced RNA, RNA/DNA, and intact DNA as outcomes; and biological processes gene sets, in the total study population (a) and among European subgroup (b). Horizontal dashed lines represent the GSEA Benjamini-Hochberg false discovery rate (FDR) statistical significance level of q=0.05 (blue) and q=0.025 (grey), respectively.