Host variation in type I interferon signaling genes (*MX1*), *CCR5*Δ*32*, 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], Claire N. Levy[9], Meei-Li Huang[10,11], Pavitra Roychoudhury[10,11], Florian Hladik[9,11], Keith R. Jerome[10,11], Hans-Peter Kiem[10,11], 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 Obstetrics and Gynecology, University of Washington, WA, 98105 USA. [10] Department of Laboratory Medicine and Pathology, University of Washington, Seattle WA 98195, USA. [11] 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

Word count (Max 3500): 4502

Abstract

Objective: 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.

Design: Cross-sectional genomewide association study.

Methods: We analyzed custom 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, and duration of ART suppression.

Results: Previously reported "protective" host genetic mutations related to viral setpoint (e.g., among elite controllers) were found to predict smaller HIV reservoir size. The HLA "protective" B*57:01 was associated with significantly lower HIV usRNA ($q=3.3x10^{-3}$), and among the largest subgroup, European ancestry individuals, the *CCR5Δ32* deletion was associated with smaller HIV tDNA ($p=4.3x10^{-3}$) and usRNA ($p=8.7x10^{-3}$). In addition, genomewide analysis identified several SNPs in *MX1* (an interferon stimulated gene) that were significantly associated with HIV tDNA (q=0.02), and the direction of these associations paralleled *MX1* gene eQTL expression.

Conclusions: We observed a significant association between previously reported "protective" MHC class I alleles and *CCR5Δ32* with the HIV reservoir size in non-controllers. We also found a novel association between *MX1* and HIV total DNA (in addition to other interferon signaling relevant genes, *PPP1CB*, *DDX3X*). These findings warrant further investigation in future validation studies.

Abstract (max 250 words): 247

Keywords: HIV reservoir, host genetics, type I interferon, MHC class I, 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 (MB). 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, DLK, KRJ, and SP. 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, CT, and KH performed the qPCR HIV reservoir assays (total DNA, unspliced RNA) in the lab of TH. CNL and MLH performed the ddPCR HIV reservoir assay (intact DNA) in the labs of FH, KRJ, and HPK. PR, DAS, TJH, and SAL analyzed these HIV reservoir data in relation to host genomic and clinical phenotype 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.

Introduction

Although antiretroviral therapy (ART) prolongs life, it does not fully restore health, as evidenced by persistently high levels of immune activation [1] and increased rates of non-AIDS-related mortality [2] observed in HIV-infected compared to uninfected individuals [3-6]. Persistent HIV may contribute to ongoing inflammation, immune activation, and subsequent clinical outcomes, even during effective ART [5-8]. Identifying host genetic predictors of the HIV reservoir in ART-suppressed individuals may shed light on novel (and potentially modifiable) targets to reduce the HIV reservoir and inflammation- and immune activation-associated adverse effects on long-term morbidity and mortality.

Most prior host genetic HIV studies have focused on identifying variants associated with viral setpoint, e.g., among "elite controllers", HIV+ individuals able to maintain viral suppression in the absence of therapy[9-18]. 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*)[19-22] and a SNP in the HLA complex 5 (*HCP5*) gene[10]. However, whether *residual* viral control during *treated* HIV disease – i.e., "the HIV reservoir" – is influenced by the same genetic variants is unknown. We performed custom whole exome sequencing among HIV non-controllers in relation to four measures of the peripheral CD4+ T cell HIV reservoir: cell-associated "intact" DNA[23], total DNA, unspliced RNA, and RNA/DNA (**Figure S1**). We found that previously reported "protective" HLA-B*57:01[10, 17] and *CCR5Δ32*[20, 21, 24] mutation were associated with smaller HIV reservoir size. Genomewide analyses demonstrated several novel associations with SNPs in interferon signaling-associated genes (*MX1*, *PPP1CB*, *DDX3X*) and total HIV DNA reservoir size. Gene set enrichment analysis identified several interferon signaling-associated genes to significantly predict intact HIV DNA levels in the largest subgroup, Europeans.

Methods

Study participants

HIV+ non-controllers who initiated ART during chronic (>2 years) or early (<6 months) HIV infection were sampled from the UCSF SCOPE and Options cohorts (**Table S1**). Inclusion criteria were laboratory-confirmed HIV-1 infection, availability of $10x10^6$ cryopreserved PBMCs, and plasma HIV RNA levels below the limit of assay quantification (<40 copies/mL) for at least 24 months at the time of biospecimen collection. HIV "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[25-27] , were excluded. The estimated date of detected infection (EDDI) was calculated for each study participant to determine recency of infection in relation to ART start date using detailed clinical HIV diagnostic test results, using the Infection Dating Tool (https://tools.incidence-estimation.org/idt/)[28]. Additional exclusion criteria were potential factors that might influence HIV reservoir quantification, 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 UCSF Committee on Human Research (CHR), and all participants provided written informed consent.

Custom whole exome host DNA sequencing

Genomic DNA was extracted (AllPrep Universal Kit, Qiagen, Hilden, Germany) from negatively selected CD4+ T cells from cryopreserved PBMCs (StemCell, Vancouver, Canada). Targeted exome capture was performed with custom addition of 50 Mb regulatory regions (Roche NimbleGen, Wilmington, MA), sequencing libraries were generated and then run on the Illumina HiSeq 2000 system (Illumina, San Diego, CA). The custom regions included 50 kb upstream and 50 kb downstream of 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) (**Table S2**).

HLA typing

Direct HLA typing was performed from extracted genomic DNA following the PCR-SSOP (sequence-specific oligonucleotide probing) typing and PCR-SBT (sequence based typing) protocols recommended by the 13th International Histocompatibility Workshop[29, 30]. Locus-specific primers were used to amplify a total of 25 polymorphic exons of HLA-A & B (exons 1-4), C (exons 1-5), E (exon 3), DPA1 (exon 2), DPB1 (exons 2-4), DQA1 (exon 1-3), DQB1 (exons 2-3), DRB1 (exons 2-3), and DRB3, 4, 5 (exon 2) genes with Fluidigm Access Array (Fluidigm, Singapore) and sequenced on an Illumina MiSeq sequencer (Illumina, San Diego, USA). HLA alleles and genotypes are called using the Omixon HLA Explore (version 2.0.0) software (Omixon, Budapest, Hungary).

HIV reservoir quantification from peripheral CD4+ T cells

The HIV reservoir largely consists of “defective” virus that harbors mutations prohibiting the production of infectious virus[31, 32]. There is currently no “gold standard” for measuring the HIV reservoir. Therefore, we estimated the frequency of HIV “intact” DNA using a ddPCR-based assay to quantify the size of the potentially “replication-competent” reservoir[23, 33, 34]. We also measured HIV total DNA (quantifies both defective and intact HIV) and unspliced RNA (quantifies full-length HIV RNA) using an HIV-1 LTR-specific quantitative polymerase chain reaction (qPCR) TaqMan assay[35]. DNA and RNA were simultaneously dual extracted using the AllPrep Universal Kit (Qiagen, Hilden, Germany). HIV tDNA and usRNA were then quantified in triplicate reaction wells using a 7-point standard curve (1–10,000 copies/second). To estimate the frequency of “intact” HIV DNA, five regions on the HIV genome were interrogated in a

multiplex ddPCR assay[23]. Droplet generation and thermocycling were performed according to manufacturer instructions. To determine potentially replication-competent ("intact") HIV genomes, the number of positive droplets for 3 targets per assay were quantified. 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, to normalize the HIV copy numbers/10^6 CD4+ T cells. A DNA shearing index (DSI) (using *RPP30*) was then used to calculate the estimated number of intact HIV genomes after correcting for shearing.

Data processing and quality control

The bcbio bioinformatics pipeline[36] was used to perform DNA alignment, which included the Burroughs-Welcome Aligner (BWA) tool[37] and the GenomeAnalysisToolkit (GATK) HaplotypeCaller joint variant calling method[38]. Reads were initially mapped to reference genome b37, then transposed to human genome assembly GRCh38 using Picard tools[39]. SNPs and insertions or deletions (indels) were then filtered by variant quality score recalibration (VQSR) using GATK[40]. The whole genome data analysis toolset, PLINK[41], 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 $1x10^{-8}$. 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[42].

The GENESIS analysis pipeline[43] 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 (**Figure 1**). We accounted for population stratification in the total population by (1) including a genetic effects term with a

genetic relatedness matrix (GRM), (2) by including the first five PCs as covariates in the multivariate models, and by (3) performing sensitivity analyses among the largest subgroup, Europeans.

Single SNP common variant analyses

Individual 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. Final covariates in multivariate models were sex, timing of ART initiation (**Figure S2**), nadir CD4+ T cell count (**Figure S3**), and the first 5 PCs. Pre-ART viral load (**Figure S4**) and duration of ART suppression (**Figure S5**) were not associated with HIV reservoir size nor improved the fit of the final models. A Gaussian link function was used, and a GRM was included with results filtered for SNPs with MAF $\geq$5%. SNP annotations were then obtained using Annovar[44].

Gene-based rare variant analyses

Gene level multi-SNP associations were calculated with the GENESIS software package "assocTestAggregate" function implementing the variant Set Mixed Model Association Test (SMMAT)[45] for alleles with MAF<5% with weights following the beta distribution parameters of a_1=1 and a_2=25[46]. The same covariates, GRM, and regression family were used as for the individual SNP associations. Outcomes were quantile-normalized to follow a normal distribution. Gene regions were defined according to UCSC hg38 assembly[47].

Gene set enrichment analyses (GSEA) were performed using the Molecular Signatures Database (MSigDB)[48, 49]. 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[50] with covariates for sex, timing of ART initiation, nadir CD4+ T cell count, and 3 genetic PCs.

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**). Participants had a median of 5.1 years of ART suppression, a median nadir CD4+ T cell count=341 cells/mm^3, and pre-ART HIV RNA=5.1 $\log_{10}$copies/mL. 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. Most study participants (N=147) had highly detailed clinical test results to be able to calculate their estimated date of detected infection (EDDI), but a subset of 55 study participants only had self-reported data regarding date of ART initiation in relation to date of HIV seroconversion. For these individuals (all of whom initiated ART prior to widespread guidelines for initiating ART at the time of HIV diagnosis[51]), we mean-imputed values assuming ART initiation starting after 2

years from infection. This estimation is supported by prior data from our cohort and others demonstrating that the HIV reservoir size is relatively stable after 2 years of infection[52-56]. Overall results for all final models were unchanged when performing sensitivity analyses excluding those with imputed values for timing of ART initiation.

Earlier ART initiation and lower nadir CD4+ T cell count were associated with smaller HIV reservoir, and HIV reservoir measures were correlated with each other

Consistent with prior work[32, 57, 58], earlier ART initiation was associated with significantly smaller HIV reservoirs (tDNA, usRNA, intact DNA) (**Figure S2**), while lower nadir CD4+ T cell count was associated with larger HIV reservoir (tDNA, usRNA, intact DNA, RNA/DNA) (**Figure S3**). Pre-ART viral load (**Figure S4**) and duration of ART suppression (**Figure S5**) were not associated with HIV reservoir size. Although usRNA was correlated with both tDNA intact DNA (**Figure S6a-b**), tDNA was not associated with intact DNA (**Figure S6c**).

HLA "protective" B*57:01 and "risk" C*07 alleles were associated with smaller and larger HIV reservoir sizes, respectively

Using a Benjamini-Hochberg false discovery rate (FDR) adjusted $q<0.05$ threshold[59], 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 for viral setpoint in untreated HIV+ controllers[17] and found a "protective" association with HLA-B*57:01 and usRNA ($\beta=-1.5$, $q=3.3\times10^{-3}$), with a similar trend observed with tDNA ($\beta=-1.6$, $q=0.13$). Similarly, previously reported HLA-C*07 "risk" allele also demonstrated a "risk" trend (larger reservoir size) in our European subgroup (tDNA: $\beta=0.76$, $q=0.072$, and usRNA: $\beta=0.41$, $q=0.10$). Further analyses employing a composite HLA variable did not identify statistically significant associations (**Tables S3-S6**).

CCR5Δ32 was associated with smaller HIV reservoir size

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[20, 21, 24]. Among individuals of European ancestry (where *CCR5Δ32* is more commonly observed), the presence *CCR5Δ32* was associated with smaller HIV reservoir size (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)[22], 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),

Genomewide association analysis identified several SNPs in *MX1* associated with larger and smaller HIV reservoir sizes, paralleling predicted *MX1* gene expression

A total of 1,279,156 variants from 23,733 genes were included in the final analysis from 199 study participants whose sequencing data passed quality control metrics (**Figure S7**). Final models demonstrated lambda genomic inflation factor[60] values close to 1, supporting adequate adjustment for possible bias due to population stratification (ancestry) (**Figure 2**).

The strongest genomewide associations were observed with HIV total DNA reservoir measures (**Tables 1 and S7**). In particular, 44 SNPs in linkage disequilbrium (LD) in the human interferon-inducible myxovirus resistance 1 gene, *MX1*, also known as *MXA*[61, 62] were significantly associated with tDNA (all q<0.03). *MX1* is closely related to *MX2* (*MXB*), which encodes a well-known potent host restriction factor that inhibits HIV-1 infection[63-65]. We then compared the directionality of the SNP hits with previously reported whole blood eQTL data at these loci[66-68] and found the *MX1* SNPs associated with larger total HIV DNA reservoir sizes seemed to be in

eQTL regions predicting increased *MX1* expression and vice versa (**Table S7**). We also observed two additional SNPs significantly associated with HIV tDNA, the first in *PPP1CB* (encodes Protein Phosphatase 1 Catalytic Subunit Beta, which reduces antiviral potency of MX2 against HIV-1[65], q=0.03) and the second in *LRMP* (encodes Lymphoid-Restricted Membrane Protein, which plays a critical role in the delivery of peptides to MHC class I molecules[69], q=0.03) (**Table 1**). Additional SNPs that showed non-statistically significant trends with HIV tDNA were in *DDX3X* (DEAD-box helicase 3 X-linked, regulates the production of type I interferons[70], q=0.17) and *AKAP6* (A-Kinase Anchoring Protein 6, binds to protein kinase A regulatory subunits, a critical signaling pathway associated with HIV latency reversal and T cell proliferation[71, 72], q=0.20). Among Europeans, *OSBP* (oxysterol-binding protein, associated with HIV-1 infection of monocyte-derived macrophages from highly-exposed seronegative individuals[73], q=0.14), showed a non-significant trend with HIV tDNA.

Although not statistically significant, a SNP in *PLAVP* (protein regulating lymphocyte migration into lymph nodes[74], q=0.21), lying <30 kilobases upstream of *BST2* (tetherin, an HIV host restriction factor[75]) demonstrated a non-significant trend with usRNA (**Table 1**). No SNPs met statistical significance in association with HIV intact DNA or RNA/DNA ratio.

Gene set enrichment analysis demonstrated several interferon signaling-associated genes associated with intact HIV DNA

We then performed multi-SNP analyses to identify genes associated with HIV reservoir size. GSEA identified several interferon signaling-associated genes (e.g., *IFITM1*, *IFITM3*, *APCS*, *IFITM2*, *FCN3*, *FCN1*, *GSN*, *TRIM59*, *SNX3*, *TRIM25*, *PTX3*, *TRIM11*, *TRIM8*, *MID2*, *TRIM5*, *IFNA2*) in the gene set called “Negative Regulation of Viral Entry into Host Cell,” to significantly predict HIV intact DNA (q=0.03) (**Figure 3, Table S8**). Several other gene sets showed non-significant trends with HIV reservoir size (**Figure 3, Table S8**), including gene sets related to interferon-induced STAT signaling and intact DNA, glycosylation and tDNA, and retroviral

transcription and usRNA.

Discussion

HIV eradication remains a critical goal in reducing long-term morbidity and mortality among all PLWH since life-long viral suppression does not appear to fully restore health, as evidenced by persistently high levels of immune activation[2] and high rates of mortality[8] in HIV-infected compared to healthy individuals. ART is also expensive, carries long-term risk of toxicity, and poses major challenges in being able to be accessible to a global population for decades[76]. HIV cure clinical trials to date have yielded disappointing results[77-82]. Novel approaches are needed to better target potential immunologic pathways that help maintain the HIV reservoir.

Our study is the first host genomic study to evaluate several measures of the peripheral HIV reservoir in HIV+ non-controllers, including quantification of HIV intact DNA, an estimate of the putative "replication-competent" reservoir by droplet digital PCR[23, 33, 83]. We also performed direct HLA typing of 25 polymorphic exons of HLA-A & B, C, E, DPA1, DPB1, DQA1, DQB1, DRB1, and DRB3,4,5. We performed individual SNP and gene-based analyses including detailed clinical covariate data such as timing of ART initiation, one of the strongest clinical predictors of HIV reservoir size, which enhanced the fit of our final multivariate models, potentially allowing us to detect otherwise difficult-to-discern genetic effects. Unlike prior genomic studies that have primarily focused on the ~1% of the HIV+ population able to suppress virus in the absence of therapy ("elite controllers"), we focused on HIV+ ART-suppressed non-controllers, which make up the majority of people living with HIV (PLWH). We found that prior significant HLA and *CCR5Δ32* genetic associations predicting viral setpoint among HIV+ elite controllers[10, 19-22] in our study, predicted the HIV reservoir size. We also identified several additional (uninvestigated) host genetic variants associated with the HIV

reservoir (signals that might have been attenuated in a study population enriched for "stronger" genetic effects, such as HLA and/or *CCR5Δ32*).

The most striking finding from our SNP-based analysis was the identification of several SNPs in *MX1*, which encodes for a potent antiviral factor which inhibit replication of several RNA viruses, including influenza A and measles, and DNA viruses, such as hepatitis B[84]. *MX1* expression has also been shown to be upregulated in HIV+ vs. HIV-uninfected individuals[85], in HIV+ individuals with high vs. low viremia[86], and with HIV-1 latency in latently-infected cell lines[87]. Genomewide (i.e., DNA-based) results cannot directly infer directionality of gene function without further functional studies. However, for our top hit SNPs in *MX1*, we compared the directionality of the SNP hits with previously reported whole blood eQTL data at these loci[66-68] and found that the *MX1* SNPs associated with larger total HIV DNA reservoir sizes seemed to be in eQTL regions predicting increased *MX1* expression and vice versa. However, determining whether a single variant is responsible for both genomewide and eQTL signals in a locus can be challenging. Nonetheless, as a further query, we performed colocalization analysis, an *in silico* method to integrate GWAS and eQTL results to calculate a *probability* of whether a SNP is causal for both an eQTL and disease trait[88], but only found a 1% probability that the *MX1* top SNPs are causally linked to both gene expression and HIV reservoir size. Gene-based analyses also identified several interferon signaling-associated genes (within the "negative regulation of viral entry into host cell" gene set) that significantly predict intact HIV DNA (**Table S8**), but as these genes were not in eQTL regions, the putative directionality of these associations could not be further queried. Additional functional genomic validation, e.g., CRISPR-Cas9 editing of primary human T cells[89], is needed to further investigate the potential role of *MX1* (and other interferon signaling genes) in HIV persistence.

We also found that the previously reported "protective" HLA-B*57:01 and *CCR5Δ32* mutation were associated with smaller HIV reservoir size in our study. These findings suggest

that immune pathways that control viral setpoint during untreated disease may also play a role in the maintenance of the HIV reservoir during treated infection. It is also possible that identified genetic variants may have variable "penetrance"[90] – e.g., the genetic variants that may drive "elite control" might similarly, but less obviously, influence HIV persistence in treated non-controllers.

There are limitations to our study that deserve mention. Although the HIV reservoir has been shown to be relatively stable over time[58, 91, 92], our cross-sectional design provides a "snapshot" of the HIV reservoir after a median of 5.1 years of ART suppression and may not reflect genetic associations with reservoir decay. Second, as is characteristic of many U.S.-based HIV+ cohorts, our San Francisco-based population consisted mostly of males of European ancestry. Population stratification is a critically important potential bias in any multiethnic genomic study. Thus, we approached this in least three ways using well-established methods to adjust for population stratification bias[43, 93]: first by calculating principal components and including these as covariates in the final models, second by including a genetic relatedness matrix (GRM) in the models, and finally by performing stratified analyses, focusing on the largest homogenous subpopulation (individuals with European ancestry). Overall, the findings observed in the European ancestral group did not overlap with the non-European (e.g., African-American) subgroup (**Table S9**). Thus, it is critical that these results be replicated in larger studies, especially those including women and individuals from different ethnic backgrounds. Third, the majority of the HIV reservoir persists in lymphoid tissues, not in the periphery[94]. Although recent data suggests that the tissue compartment largely reflects (and is the likely source of) the peripheral compartment[52, 95, 96], it will be important to determine whether the results from our study are generalizable to the tissue HIV reservoir. Fourth, intact HIV DNA represents the putative replication-competent reservoir. Although we observed several genes that were significantly associated with intact HIV DNA in the gene set enrichment analyses, individual genes did not meet statistical significance. This may be due to the challenge in

estimate the frequency of intact and/or replication-competent HIV when in fact, the majority of the HIV reservoir consists of defective HIV. For these reasons, quantitative outgrowth assays and assays to measure intact HIV DNA often result in many low/ zero values compared to total HIV DNA, which has a larger dynamic range[83, 97, 98]. In our study, HIV intact DNA was undetectable in over half of our measured samples, while for example, total DNA was measurable in 95% of samples (**Figure S6**). With so many samples below the limit of detection for intact DNA, the statistical power to detect SNP associations is much lower for this assay than for the other HIV reservoir assays included in our study. However, when we were able to enhance the ability to detect an association by performing the gene set enrichment analyses (essentially, a method that aggregates several rare variants into immunologically relevant "gene sets" to test for an association with HIV reservoir size), we observed several statistically significant associations with HIV intact DNA in the total population (STAT signaling, critical for regulating the innate and adaptive immune responses) and among individuals of European ancestry, the largest subgroup with the greatest statistical power ("negative regulation of viral entry into host cell" – which included several interferon signaling genes) (**Figure 3, Table S8**).

Our findings are in contrast to two recent genomewide studies of the HIV reservoir, which did not identify an association with *MX1*, HLA-B*57:01, or *CCR5Δ32*, nor reported similar findings to each other[99, 100]. The first study performed GWAS microarray genotyping from 797 HIV+ treated individuals (194 with whole exome sequencing data), included several longitudinal measures of HIV total DNA from peripheral blood mononuclear cells (PBMCs), and imputed HLA alleles (from genotypes), but did not observe any significant associations with HLA alleles, *CCR5Δ32*, or SNPs[99]. The second study included 207 HIV+ treated individuals and performed GWAS microarray (no HLA or *CCR5Δ32* typing) and included measures of HIV tDNA and usRNA from peripheral CD4+ T cells. They reported a significant association between tDNA and a SNP in *PTDSS2* (phosphatidylserine synthase 2) at genomewide $p<5x10^{-8}$, which was not statistically significant in our analysis.[100] These differences highlight a particular challenge in

performing genomic studies, let alone for studies of the HIV reservoir size (which can be measured in several different ways as total HIV DNA, HIV unspliced RNA, HIV intact DNA, etc.) from different biospecimens (PBMCs, CD4+ T cells). Add to this the use of different study designs (untreated HIV+ controllers, treated HIV+ non-controllers – and those treated during chronic versus acute infection), and analytic methods (multivariate models with or without key clinical covariates such as timing of ART initiation, nadir CD4+ T cell count, etc.) and a lack of understanding regarding the exact mechanism by which the genetic code is expressed from DNA to RNA to protein (which also varies by cell type and within different tissues[101]) – then differences between these three small studies might fall within the expected range of variability. Given the polygenic nature of the host immune response, the contribution of host genetics in predicting HIV reservoir size might vary widely, leading to variable results when comparing small genomic studies. Prior genomewide association studies of HIV progression during untreated disease explain ~13% of the variability in viral load, with strong genetic predictors such as HLA and *CCR5Δ32*[102]. Using a tool for genome-wide complex trait analysis (GCTA)[103], we calculated the heritability of our total HIV DNA phenotype to be up to 0.78, but the error bars were large (+/-0.94). This suggests that unknown host genetic loci might play a significant role in determining the size of the HIV reservoir – but that there is a high degree of variability in that estimate. For these reasons, we are careful to describe our study results within the limits of a genomewide association study identifying potential novel DNA variants related to HIV persistence (e.g., only highlighting *MX1* as it lies with an eQTL region but the others for which there are no functional data, we do not) and again emphasize the need for functional studies to pursue the novel hypotheses identified from our discovery-based study. Our findings may vary from the two prior published studies due to several differences in (1) study design (cross-sectional, only including HIV+ ART-suppressed non-controllers), (2) statistical modeling (detailed clinical covariates for timing of ART initiation, nadir CD4+ T cell count, etc.), (3) HIV reservoir quantification (e.g., HIV total DNA, unspliced RNA, and intact DNA), (4) sampling

(CD4+ T cells, not bulk PBMCs), (5) host genomic assays (custom whole exome sequencing instead of GWAS microarray, and direct HLA typing instead of imputed HLA alleles), and (6) approach to handling potential population stratification bias (at least three methods - principal component analysis, genetic related matrix methods, and sensitivity analyses restricted to ancestral subgroups).

Using carefully selected ART-suppressed HIV non-controllers, we performed custom whole exome sequencing and direct HLA typing, quantified several measures of the peripheral HIV reservoir, and fit multivariate models adjusted for clinical and demographic covariates that influenced the size of the HIV reservoir, and found that the previously reported “protective” HLA-B*57:01 and the favorable *CCR5Δ32* (during untreated disease) were associated with smaller HIV reservoir size. Genomewide analyses identified several SNPs in *MX1*, a type I interferon stimulated gene, were significantly associated with total HIV DNA, which correlated with predicted *MX1* eQTL gene expression, and HIV intact DNA were associated with several interferon signaling-associated genes in gene set enrichment analyses. Our findings support a surprising role of the innate immune response (e.g., genes involved in interferon signaling[104-107]) in maintaining the HIV reservoir during long-term suppressive ART. These findings support recent studies demonstrating that measures of cell-associated HIV RNA correlate with time to viral rebound[106, 108-111]. Perhaps host genes driving immediate antiviral responses play a major role in maintaining the HIV reservoir if the “transcriptionally active” reservoir is indeed a major source of the “rebound-competent” reservoir.” Additional studies are needed to functionally validate these findings, especially among more diverse patient populations, including female and non-European HIV+ patients.

FIGURE LEGENDS

Figure 1. 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 2. 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 of q=0.05 is shown as the horizontal blue line (q=0.25 is shown in grey).

Figure 3. 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. Yukl SA, Gianella S, Sinclair E, Epling L, Li Q, Duan L, et al. **Differences in HIV burden and immune activation within the gut of HIV-positive patients receiving suppressive antiretroviral therapy**. *J Infect Dis* 2010; 202(10):1553-1561.

2. Hunt PW, Martin JN, Sinclair E, Bredt B, Hagos E, Lampiris H, et al. **T cell activation is associated with lower CD4+ T cell gains in human immunodeficiency virus-infected patients with sustained viral suppression during antiretroviral therapy**. *J Infect Dis* 2003; 187(10):1534-1543.

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**. *Ann Intern Med* 2007; 146(2):87-95.

4. Antiretroviral Therapy Cohort C. **Causes of death in HIV-1-infected patients treated with antiretroviral therapy, 1996-2006: collaborative analysis of 13 HIV cohort studies**. *Clin Infect Dis* 2010; 50(10):1387-1396.

5. Eisele E, Siliciano RF. **Redefining the viral reservoirs that prevent HIV-1 eradication**. *Immunity* 2012; 37(3):377-388.

6. Volberding PA, Deeks SG. **Antiretroviral therapy and management of HIV infection**. *Lancet* 2010; 376(9734):49-62.

7. **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):1387-1396.

8. 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.

9. 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.

10. 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):944-947.

11. 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):773-779.

12. 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.

13. 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):618-626.

14. 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):1194-1201.

15. 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):419-426.

16. 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):1141-1149.

17. 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):1551-1557.

18. 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):513-518.

19. 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):882-890.

20. 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):1856-1862.

21. 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):922-923.

22. 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):824-834.

23. 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.

24. 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):882-890.

25. 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):1553-1562.

26. 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.

27. 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):5398-5407.

28. Grebe E, Facente SN, Bingham J, Pilcher CD, Powrie A, Gerber J, et al. **Interpreting HIV diagnostic histories into infection time estimates: analytical framework and online tool**. *BMC Infect Dis* 2019; 19(1):894.

29. 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):429-434.

30. 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):733-740.

31. 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):540-551.

32. 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):1043-1049.

33. 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):120-125.

34. Levy CN, Hughes SM, Roychoudhury P, Amstuz C, Zhu H, Huang ML, et al. **HIV reservoir quantification by five-target multiplex droplet digital PCR**. *STAR Protoc* 2021; 2(4):100885.

35. 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):1240-1248.

36. Chapman Bea. **bcbio/bcbio-nextgen**. In. v.1.2.8 ed: Zenodo; 2021.

37. Li H. **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. *arXiv:13033997* 2013.

38. Poplin R, et al. **Scaling accurate genetic variant discovery to tens of thousands of samples**. *bioRxiv* 2017; 201178.

39. **Picard toolkit**. In; 2019.

40. 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):491-498.

41. 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):559-575.

42. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. **The variant call format and VCFtools**. *Bioinformatics* 2011; 27(15):2156-2158.

43. 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):5346-5348.

44. Wang K, Li M, Hakonarson H. **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data**. *Nucleic Acids Res* 2010; 38(16):e164.

45. 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):260-274.

46. 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.

47. Bioconductor Core Team BPMMBO. **Txdb.hsapiens.ucsc.hg38.knowngene**. In; 2017.

48. 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):267-273.

49. 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):15545-15550.

50. Seabold; Skipper; Perktold J. **statsmodels: Econometric and statistical modeling with python**. In: *Proceedings of the 9th Python in Science Conference*; 2010.

51. (OARAC). DPoAGfAaAAWGotOoARAC. **Guidelines for the use of antiretroviral agents in adults and adolescents with HIV. AIDSinfo. 2019 Dec 18.** 2018.

52. 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):E4987-4996.

53. Buzon MJ, Martin-Gayo E, Pereyra F, Ouyang Z, Sun H, Li JZ, et al. **Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells**. *J Virol* 2014; 88(17):10056-10065.

54. Archin NM, Vaidya NK, Kuruc JD, Liberty AL, Wiegand A, Kearney MF, et al. **Immediate antiviral therapy appears to restrict resting CD4+ cell HIV-1 infection without accelerating the decay of latent infection**. *Proc Natl Acad Sci U S A* 2012; 109(24):9523-9528.

55. 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):1169-1178.

56. Jain V, Hartogensis W, Bacchetti P, Hunt PW, Hatano H, Sinclair E, et al. **Antiretroviral Therapy Initiated Within 6 Months of HIV Infection Is Associated With Lower T-Cell Activation and Smaller HIV Reservoir Size**. *The Journal of infectious diseases* 2013;

208(8):1202-1211.

57. 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.

58. 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.

59. 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):279-284.

60. 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):807-812.

61. Staeheli P, Haller O. **Human MX2/MxB: a Potent Interferon-Induced Postentry Inhibitor of Herpesviruses and HIV-1**. *J Virol* 2018; 92(24).

62. 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.

63. 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):559-562.

64. 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.

65. 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):1031-1042.

66. 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):204-213.

67. 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.

68. 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):2333-2347.

69. Collins KL. **How HIV evades CTL recognition**. *Curr HIV Res* 2003; 1(1):31-40.

70. 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.

71. 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):629-640.

72. Troger J, Moutty MC, Skroblin P, Klussmann E. **A-kinase anchoring proteins as potential drug targets**. *Br J Pharmacol* 2012; 166(2):420-433.

73. 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):1593-1602.

74. 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):386-396.

75. 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.

76. Deeks SG, Archin N, Cannon P, Collins S, Jones RB, de Jong M, et al. **Research priorities for an HIV cure: International AIDS Society Global Scientific Strategy 2021**. *Nat Med* 2021; 27(12):2085-2098.

77. Gay CL, Kuruc JD, Falcinelli SD, Warren JA, Reifeis SA, Kirchherr JL, et al. **Assessing the**

impact of AGS-004, a dendritic cell-based immunotherapy, and vorinostat on persistent HIV-1 Infection. *Sci Rep* 2020; 10(1):5134.

78. Fidler S, Stohr W, Pace M, Dorrell L, Lever A, Pett S, et al. **Antiretroviral therapy alone versus antiretroviral therapy with a kick and kill approach, on measures of the HIV reservoir in participants with recent HIV infection (the RIVER trial): a phase 2, randomised trial**. *Lancet* 2020; 395(10227):888-898.

79. Gutierrez C, Serrano-Villar S, Madrid-Elena N, Perez-Elias MJ, Martin ME, Barbas C, et al. **Bryostatin-1 for latent virus reactivation in HIV-infected patients on antiretroviral therapy**. *AIDS* 2016; 30(9):1385-1392.

80. Vibholm L, Schleimann MH, Hojen JF, Benfield T, Offersen R, Rasmussen K, et al. **Short-Course Toll-Like Receptor 9 Agonist Treatment Impacts Innate Immunity and Plasma Viremia in Individuals With Human Immunodeficiency Virus Infection**. *Clin Infect Dis* 2017; 64(12):1686-1695.

81. Riddler SA, Para M, Benson CA, Mills A, Ramgopal M, DeJesus E, et al. **Vesatolimod, a Toll-like Receptor 7 Agonist, Induces Immune Activation in Virally Suppressed Adults Living With Human Immunodeficiency Virus-1**. *Clin Infect Dis* 2021; 72(11):e815-e824.

82. Elliott JH, McMahon JH, Chang CC, Lee SA, Hartogensis W, Bumpus N, et al. **Short-term administration of disulfiram for reversal of latent HIV infection: a phase 2 dose-escalation study**. *Lancet HIV* 2015; 2(12):e520-529.

83. 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.

84. Verhelst J, Hulpiau P, Saelens X. **Mx proteins: antiviral gatekeepers that restrain the uninvited**. *Microbiol Mol Biol Rev* 2013; 77(4):551-566.

85. 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.

86. 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; 41:443-454.

87. 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):327-333.

88. Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M, et al. **Colocalization of GWAS and eQTL Signals Detects Target Genes**. *Am J Hum Genet* 2016; 99(6):1245-1260.

89. Shifrut E, Carnevale J, Tobin V, Roth TL, Woo JM, Bui CT, et al. **Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function**. *Cell* 2018; 175(7):1958-1971 e1915.

90. 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):1077-1130.

91. 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).

92. 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.

93. 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):904-909.

94. 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):9838-9842.

95. 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.

96. 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):309-314.

97. Procopio FA, Fromentin R, Kulpa DA, Brehm JH, Bebin AG, Strain MC, et al. **A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals**. *EBioMedicine* 2015; 2(8):872-881.

98. 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 Pathog* 2013; 9(2):e1003174.

99. 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):517-524.

100. 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.

101. Dolcemascolo R, Goiriz L, Montagud-Martinez R, Rodrigo G. **Gene regulation by a protein translation factor at the single-cell level**. *PLoS Comput Biol* 2022; 18(5):e1010087.

102. McLaren PJ, Carrington M. **The impact of host genetic variation on infection with HIV-1**. *Nat Immunol* 2015; 16(6):577-583.

103. Yang J, Lee SH, Goddard ME, Visscher PM. **GCTA: a tool for genome-wide complex trait analysis**. *Am J Hum Genet* 2011; 88(1):76-82.

104. Pace MJ, Agosto L, Graf EH, O'Doherty U. **HIV reservoirs and latency models**. *Virology*

2011; 411(2):344-354.

105. 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):4629-4642.

106. 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):1369-1376.

107. 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-E3668.

108. Fischer M, Joos B, Hirschel B, Bleiber G, Weber R, Gunthard HF, et al. **Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef**. *J Infect Dis* 2004; 190(11):1979-1988.

109. Li JZ, Etemad B, Ahmed H, Aga E, Bosch RJ, Mellors JW, et al. **The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption**. *AIDS* 2016; 30(3):343-353.

110. Pasternak AO, Grijsen ML, Wit FW, Bakker M, Jurriaans S, Prins JM, et al. **Cell-associated HIV-1 RNA predicts viral rebound and disease progression after discontinuation of temporary early ART**. *JCI Insight* 2020; 5(6).

111. Moron-Lopez S, Kim P, Sogaard OS, Tolstrup M, Wong JK, Yukl SA. **Characterization of the HIV-1 transcription profile after romidepsin administration in ART-suppressed individuals**. *AIDS* 2019; 33(3):425-431.

Figure 1. 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.

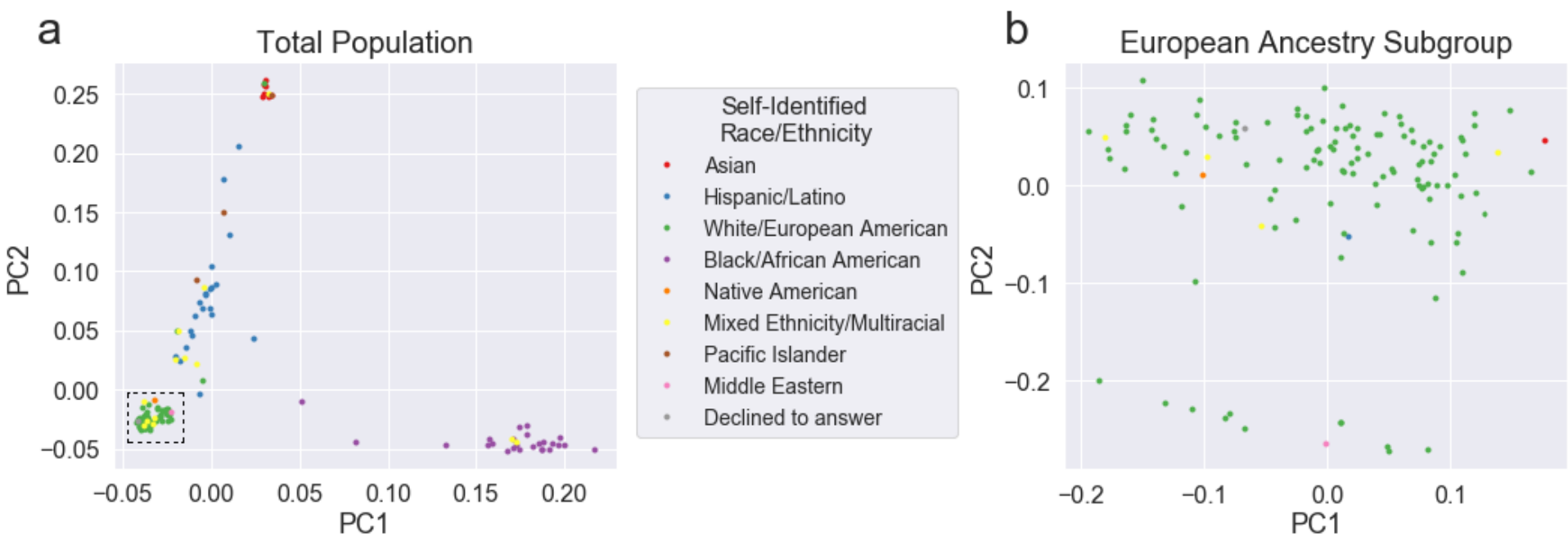

Figure 2. Quantile-quantile (QQ) plots (a, c) and Manhattan plots (b, d) of the total study population (a-b) and of European ancestry (c-d). QQ plots: the blue line represents the expected -log10 p-values while the black lines denote the expected error bars. Manhattan plots: the horizontal black line delineates a traditional conservative genome-wide significance of p-value of $5x10^{-8}$, while a less conservative Benjamini-Hochberg false discovery rate (FDR) statistical significance of q=0.05 is shown as the horizontal blue line (q=0.25 is shown in grey).

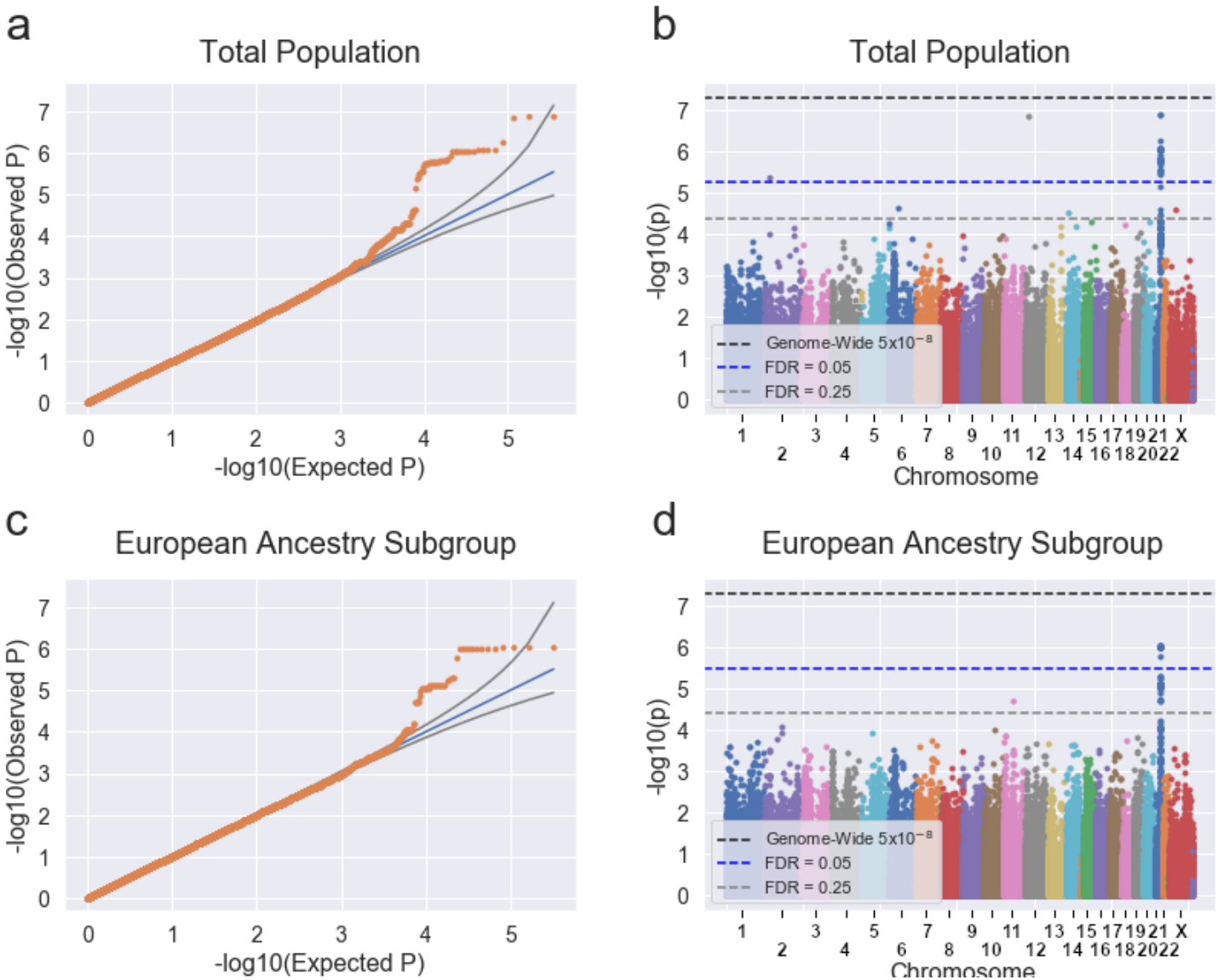

Table 1. Single nucleotide polymorphisms (SNPs) associated with HIV total DNA (upper table) and unspliced RNA (lower table) in the total study population. Two additional SNPs were significantly associated with total HIV DNA in the subpopulation of European ancestry (middle table). For genes in which 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[69] and associated with higher viremia among HIV+ individuals[70]. Associated with HIV-1 latency[71]. MX2, but not MX1, was shown to be capable of directly acting against HIV-1 virus[52, 53].
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[58]. Differentially expressed in lymphatic tissue[86] and peripheral blood mononuclear cells (PBMCs)[87] 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[54]. PP1 is involved in transcription of HIV-1; inhibition of PP1 inhibits HIV-1 transcription[88].
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[59] 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[89]. Also plays a key role in innate antimicrobial immunity[59].
rs17506750	chr14	32599402	*AKAP6*	Intronic	0.07	-1.8	0.09	$3.0x10^{-5}$	0.20	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[60, 61].
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[69] and associated with higher viremia among HIV+ individuals[70]. Associated with HIV-1 latency[71]. MX2, but not MX1, was shown to be capable of directly acting against HIV-1 virus[52, 53].
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)[62]. Is required for the replication of several human viruses such as hepatitis C (HCV), encephalomyocarditis (EMCV), Zika,

										etc.[90].
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[63]. 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[64].

[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.
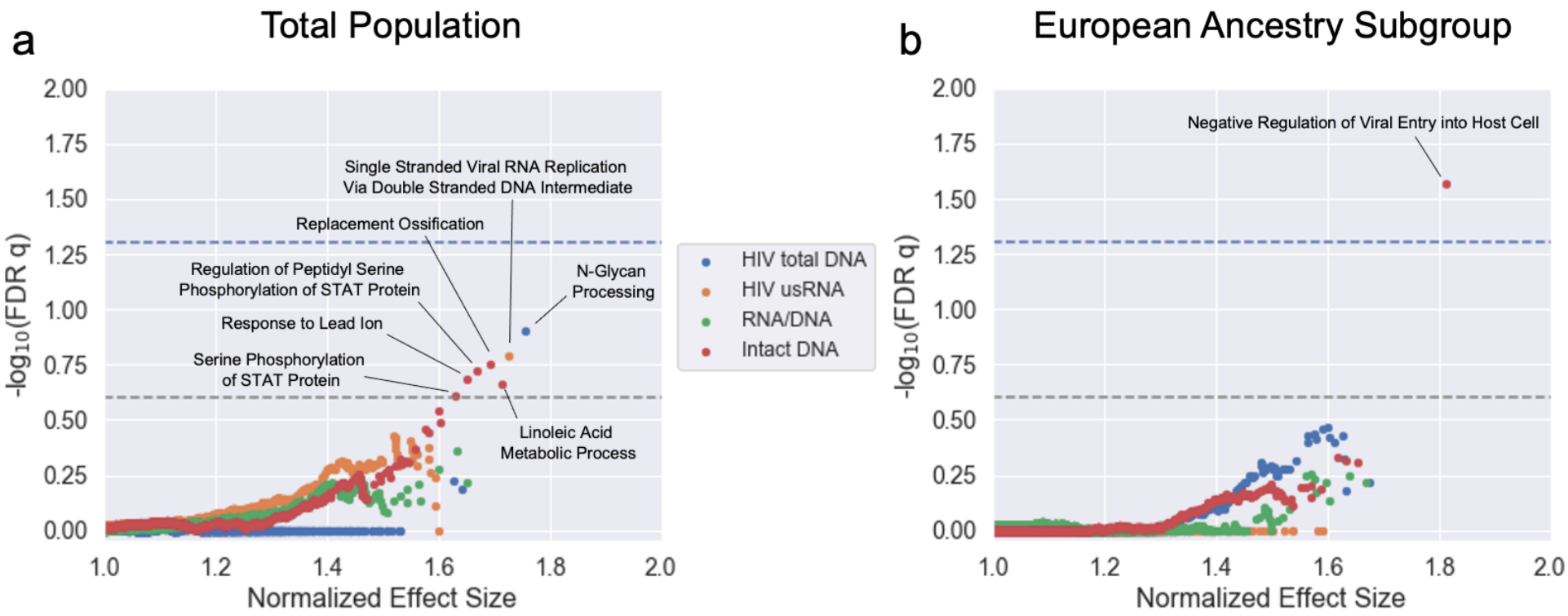
Figure 3. Gene set enrichment analysis (GSEA) was used to identify associations between genes, ordered by p-value under a host multi-SNP rare variant (minor allele frequency, MAF, <5%) analysis using HIV total DNA, unspliced RNA, RNA/DNA, and intact DNA as outcomes; and biological processes gene sets, 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.