

## Gene dysregulation among virally suppressed people living with HIV links to non-AIDS defining cancer pathways

Bryan C. Quach<sup>1,†</sup>, Eric Earley<sup>1</sup>, Linran Zhou<sup>1</sup>, Caryn Willis<sup>1</sup>, Jesse A. Marks<sup>1</sup>, Jeran K. Stratford<sup>1</sup>, Fang Fang<sup>1</sup>, Laura J. Bierut<sup>2</sup>, M-J S. Milloy<sup>3,4</sup>, Kanna Hayashi<sup>4,5</sup>, Kora DeBeck<sup>4,6</sup>, Dana B. Hancock<sup>1</sup>, Ke Xu<sup>7,8</sup>, Bradley E. Aouizerat<sup>9,10</sup>, and Eric O. Johnson<sup>1,11</sup>

1. GenOmics and Translational Research Center, RTI International, Research Triangle Park, North Carolina, USA
2. Department of Psychiatry, Washington University in St. Louis, St. Louis, Missouri, USA
3. Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada
4. British Columbia Centre on Substance Use, Vancouver, British Columbia, Canada
5. Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada
6. School of Public Policy, Simon Fraser University, Burnaby, British Columbia, Canada
7. Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut, USA
8. Connecticut Veteran Healthcare System, West Haven, Connecticut, USA
9. Bluestone Center for Clinical Research, College of Dentistry, New York University, New York, New York, USA
10. Department of Oral and Maxillofacial Surgery, College of Dentistry, New York University, New York, New York, USA
11. Fellows Program, RTI International, Research Triangle Park, North Carolina, USA

<sup>†</sup>Corresponding author

### Abstract

Combination antiretroviral therapy (ART) has changed the landscape of the HIV epidemic by providing an effective means for viral suppression to people living with HIV (PLWH). Understanding living with HIV as a chronic disease requires an improved understanding of how HIV and/or ART impacts susceptibility to and development of co-occurring conditions. Genome-wide gene expression (transcriptome) differences provide a key view into biological dysregulation associated with living with HIV. Here we present the first whole blood transcriptome-wide study comparing gene expression profiles between virally suppressed PLWH and HIV negative individuals (N=555). We identify 566 genes and 5 immune cell types with differential proportions by HIV status, which were significantly enriched for immune function and cancer pathways. Leveraging quantitative trait loci (QTL) for these HIV status-associated genes, partitioned heritability, and colocalization analyses, we observed limited genetic drivers of these relationships. Our findings suggest that gene dysregulation does not return to a pre-infection state for virally suppressed PLWH, and that persistent gene dysregulation is broadly associated with immune function and cancer pathways, highlighting potential biological drivers for these causes of excess mortality and targets for pharmacological preventative treatment among PLWH.

### Introduction

Advancements in HIV treatment have transformed infection with the retrovirus into a chronic disease in much of the Western World. For people living with HIV (PLWH), adherence to an effective combination antiretroviral therapy (ART) regimen suppresses viral replication and reduces the risk of progression to

Acquired Immune Deficiency Syndrome (AIDS)<sup>1</sup>. However, with improved survival, PLWH on ART are experiencing increased rates of non-AIDS defining conditions (e.g., cardiovascular disease, diabetes mellitus, neurocognitive disorders)<sup>2,3</sup>. Among these diseases, non-AIDS defining cancers, such as Hodgkin's lymphoma and lung, liver, kidney, and anal cancers, are key drivers of excess mortality for PLWH<sup>4-6</sup>. Persistent HIV-associated inflammation is hypothesized to contribute to the pathogenesis of non-AIDS defining conditions and chronic, virologically suppressed HIV disease or ART may accelerate cellular aging that influences morbidity and mortality risks for PLWH<sup>7,8</sup>.

Studies of host gene expression changes in HIV have uncovered a wealth of molecular targets associated with HIV acquisition, acute infection, progression, and latency.<sup>9,10</sup> A handful of these studies investigated aberrant gene expression in B-cells or lung biopsy samples for both AIDS-defining and non-AIDS-defining cancers within PLWH<sup>9,10,11</sup>. No studies to date have measured gene expression differences transcriptome-wide in virally non-detectable PLWH compared to HIV negative individuals. The extent to which differential gene expression in PLWH is associated with HIV-associated cancers remains to be studied.

To investigate gene dysregulation associated with successfully treated HIV, we conducted the first transcriptome-wide association study (TWAS) comparing virally non-detectable PLWH (N=194) to HIV-1 seronegative individuals (N=361). By using RNA sequencing of whole blood instead of isolated immune cells or cell lines, this study captures a broad, cross-cell-type signature of gene expression changes *in vivo* for chronic HIV disease and ART: identifying dysregulated genes in PLWH linked to multiple aspects of immune function and cancer-related biological pathways. To examine cell-type specific effects, we applied gene expression deconvolution and tested for differential cell-type proportions by HIV status. Additionally, we assessed cell-type specific expression for HIV-associated genes from our TWAS using expression profiles available in the Database of Immune Cell Expression, Expression quantitative trait loci, and Epigenomics (DICE)<sup>12</sup>. We characterized genetic contributions to dysregulated genes by mapping expression quantitative trait loci (eQTL) and assessed genetic relationships between HIV-associated genes and immune-related traits and multiple cancers using partitioned heritability enrichment and colocalization analyses. This study provides insight into the interplay of chronic HIV infection, ART, and genetically-regulated gene expression as it relates to molecular processes and pathways linked to comorbid conditions, including non-AIDS defining cancers. The identified genes, and drugs targeting them, may be key to developing treatments for co-occurring health conditions in PLWH.

## Results

### *Differential gene expression by HIV status*

Gene expression was measured using paired-end RNA-seq from whole blood samples of 194 PLWH and 361 people seronegative for HIV-1 (Table 1). All PLWH met the Center for Disease Control and Prevention (CDC) definition of virally non-detectable (i.e., <200 viral copies/mL of blood)<sup>13</sup> and were on ART at the time of blood draw. We conducted transcriptome-wide differential expression analysis by HIV status, accounting for potential biological and technical sources of confounding (see *Online Methods*). Of 20,211 genes tested, we identified 566 differentially expressed genes (DEGs; permutation p-value <0.05; Figures 1A and 1B, Supplemental Table 1). Relative to HIV-negative samples, 274 DEGs were up-regulated and 292 were down-regulated in PLWH, and we did not observe a significant bias towards a higher proportion of up-regulated or down-regulated genes (Binomial test  $p=0.48$ ).

Among the HIV-associated DEGs, gene set overrepresentation analysis using the GO Biological Process ontology identified 77 significantly enriched GO terms (FDR <0.05), which we grouped into categories based on semantic similarity (Figure 1C, Supplemental Table 2). The GO terms reflected a central theme of dysregulation in immune system processes related to cell migration, cell adhesion, T cell biology, lymphocyte immune response, calcium signaling, and phosphatidylinositol 3-kinase (PI3K) signaling. Overrepresentation analysis using the MSigDB KEGG canonical pathways collection identified eight significantly enriched pathways (FDR <0.05; Figure 1D, Supplemental Table 3), several of which are related to immune function, including *T cell receptor signaling pathway*, *proteasome*, *hematopoietic cell lineage*, *primary immunodeficiency*, and *cytokine-cytokine receptor interaction*. Furthermore, three cancer pathways exhibited overrepresentation of HIV-associated DEGs: *pathways in cancer*, *chronic myeloid leukemia*, and *non-small cell lung cancer*. We note that the DEGs in *chronic myeloid leukemia* and *non-small cell lung cancer* are subsets of the *pathways in cancer* gene set (Supplemental Figure 1). This gene set, which encompasses several biological pathways, is related to multiple types of cancer and tumorigenesis (Supplemental Figure 2)<sup>14-23</sup>.

### *Cell-type proportion differences by HIV status and cell-type specific expression*

Deconvolution of whole blood data was performed using the LM22 leukocyte reference panel<sup>24</sup> to estimate the cell-type proportions of 22 cell types. Of these, we observed five cell types with significant differences in proportions by HIV status (Bonferroni  $p < 0.05$ ; Figure 2A). The mean proportions of CD8+ T cells and follicular helper T cells were higher in HIV+ samples. In contrast, the mean proportions of naïve CD4+ T cells, resting CD4+ memory T cells, and resting mast cells were lower in HIV+ samples. We further characterized cell-type specific expression across 15 immune cell types for the HIV-associated DEGs. We examined differences in expression profiles using data from DICE, which characterizes immune cell gene expression in a healthy state<sup>12</sup>. For each cell type, the mean of transcripts per million (TPM) values across samples is used to represent the expression level for a gene. To assess which cell types exhibited similar/dissimilar expression profiles based on the overlap between HIV-associated DEGs with data available in DICE, we applied Pearson correlation hierarchical clustering of DICE expression data for the DEGs. The highest divergence in expression was between the T lymphocyte subset and innate immune cell types. Within the T lymphocytes, the cell types further grouped by naïve and memory T cells, with activated CD4+ and CD8+ T cells being an outgroup (Figure 2B). This clustering structure resembles the global unsupervised hierarchical clustering using the top 1000 most variable genes in DICE,<sup>12</sup> suggesting that DEGs in the current study align with overall gene expression landscape in immune cells.

To assess the degree of cell-type specificity for each HIV-associated DEG, we calculated Shannon diversity index values<sup>25</sup> based on the proportion of total expression (sum of cell type mean TPMs) for each cell type. The distribution of Shannon index values was strongly left-skewed, indicating that most DEGs were expressed more uniformly across the 15 cell types (Figure 2C and 2D). When examining genes in the fifth percentile for Shannon index values (i.e., less evenness), most of these DEGs had high mean TPM proportions concentrated within only one or two cell types (Supplemental Figure 3A). For example, the antigen encoded by *KLRD1* is preferentially expressed on natural killer (NK) cells<sup>26</sup>, and we observed that *KLRD1* was disproportionately expressed in CD56<sup>dim</sup> CD16<sup>+</sup> natural killer (NK) cells with low expression levels in other cell types. However, in addition to the disproportionately high expression cell types, some genes still had moderate or high expression levels across several other cell types (Supplemental Figure 3B). *CD40LG*, a gene well-studied in HIV infection and immunity<sup>27</sup>, exhibited

disproportionately strong expression in naïve activated CD4+ T cells but also showed moderate expression in other CD4+ T cell subtypes and activated CD8+ T cells. These findings show that gene dysregulation in virally suppressed PLWH occurs in multiple cell types involved in both innate and adaptive immune response.

### *Characterization of gene dysregulation by ART status*

Since effective ART suppresses viral replication and stalls HIV disease progression, we sought to better understand whether our HIV-associated DEGs were impacted by ART exposure. We used work by Massanella *et al.*<sup>28</sup> to define HIV-associated DEGs as ART-affected or ART-unaffected. Massanella *et al.* reported 4,157 genes with differential expression in peripheral blood mononuclear cell samples from 32 males living with HIV pre-ART and 48 weeks post-ART initiation. Intersecting our HIV-associated DEGs with this gene list resulted in 234 HIV-associated DEGs being classified as ART-affected and 332 DEGs classified as ART-unaffected (Supplemental Table 1). Among ART-affected HIV-associated DEGs, a higher proportion were down-regulated (Binomial test  $p=0.011$ ). No significant enrichment in proportion of up- or down-regulated genes was observed for ART-unaffected, HIV-associated DEGs (Binomial test  $p=0.25$ ; Figure 3A). Comparing fold change values reported by Massanella *et al.* (post-ART treatment vs. pre-ART treatment) to the fold changes for the ART-affected, HIV-associated DEGs (HIV+ vs. HIV-) revealed that almost all DEGs have an inverse direction of expression change between studies (Figure 3B). 137 genes that were upregulated post-ART compared to pre-ART still exhibited lower expression for the HIV+ group relative to HIV- samples. Similarly, the 97 post-ART downregulated DEGs showed higher expression for the HIV+ group relative to the HIV- samples. This discordance in direction of effect suggests that ART and viral suppression shifts expression of most DEGs toward the HIV uninfected state but does not fully restore gene regulation to this state.

ART-affected, HIV-associated DEGs were overrepresented for 90 GO biological process terms (FDR  $<0.05$ ; Figure 3C) and 8 MSigDB KEGG canonical pathway terms (FDR  $<0.05$ ; Figures 3D and 3E). In line with the full set of HIV-associated DEGs, many of the significant GO terms were related to immune function processes, but some processes were distinctly enriched for these ART-affected DEGs such as nucleic acid and nucleoside metabolism, a critical component of reverse-transcriptase inhibition through ART. Other processes targeted by ART were also overrepresented, including terms related to viral infection (e.g., *regulation of viral life cycle* and *viral process*). Many biological processes were also related to NK cell functions such as cell killing, ruffling, inositol lipid-mediated signaling, and chemokine signaling<sup>29,30</sup>. Several KEGG pathway terms in the overrepresentation analysis for all DEGs were also significant when assessing only ART-affected DEGs, with *pathways in cancer* persisting as the term enriched for the most DEGs (Supplemental Table 4). Of the signaling pathways in the *pathways in cancer* gene set, all the pathways described above, including Wnt/ $\beta$ -catenin, PI3K/AKT, MAPK/ERK, p53, and TGF- $\beta$  signaling, had at least one ART-affected HIV-associated DEG (Supplemental Figure 2). ART-unaffected, HIV-associated DEGs were overrepresented for 5 GO biological process and 1 KEGG canonical pathway term (FDR  $<0.05$ ; Supplemental Table 5) more narrowly linked to T cell biology (e.g., *T cell receptor signaling pathway*, *T cell selection*, *T cell co-stimulation*, *cell-substrate adhesion*). However, these findings do not preclude the involvement of ART-unaffected DEGs in the biological processes and pathways with an overrepresentation of ART-affected DEGs. Across the GO and KEGG term gene sets with overrepresentation of ART-affected DEGs, 80 ART-unaffected DEGs are components of 77 of the terms. For the GO biological process term clusters (Figure 3C), each cluster includes at least 1 term with an ART-unaffected DEG. For the KEGG terms, *basal cell carcinoma* is the only term without an ART-

unaffected DEG. (Supplemental Table 6; Figure 3E). These results show that although ART influences a broad group of immune and cancer-related molecular processes, gene dysregulation in these processes are not solely attributable to ART.

### *eQTL mapping for HIV status DEGs, partitioned heritability, and colocalization*

Given the observed immune- and cancer-related biological processes and pathways, we assessed whether heritability of immune- and cancer-related traits was enriched in genomic regions for HIV-associated DEGs. We tested for partitioned heritability enrichment of 30 GWAS traits (see Supplemental Table 7 for GWAS details) using stratified LDSC (sLDSC) and the set of HIV-associated DEGs with a window of 100 kilobases (kb) upstream and downstream around each DEG. Of these traits, eight showed statistically significant heritability enrichment (Bonferroni  $p < 0.05$ ; Figure 4A; Supplemental Table 8, Supplemental Figure 4). Three of the significant traits were blood cell counts (platelet, red blood cell, and white blood cell count). The other significant traits include asthma (estimated proportion of  $h^2$  [prop.  $h^2$ ]=0.1197), Crohn's Disease (prop.  $h^2$ =0.1069), rheumatoid arthritis (prop.  $h^2$ =0.1731), primary biliary cirrhosis (prop.  $h^2$ =0.1259), and primary sclerosing cholangitis (prop.  $h^2$ =0.1594). These five traits are inflammatory diseases that impact several different tissues. Although 12 cancer GWAS traits were tested, none showed significant heritability enrichment for DEG regions. Likewise, enrichment was not observed for HIV acquisition, suggesting that collectively our observed HIV-associated DEGs are not strongly associated with known genetic risk for cancer or HIV acquisition.

The statistically nonsignificant global enrichment results do not preclude some of the DEGs from having shared genetic signal with these traits. We explored this possibility by first conducting cis-expression quantitative trait locus (eQTL) mapping for the 566 HIV status DEGs using gene expression and imputed genotype dosage data from our study cohort. After applying a two-stage multiple testing correction (see *Online Methods*), we identified 22,475 eQTL corresponding to 258 eGenes (DEGs with at least one significantly associated genetic variant; Supplemental Table 9). Of these eGenes, 110 were ART-affected, 148 ART-unaffected, and five overlapped with the KEGG *pathways in cancer* gene set. Sixty-seven eSNPs (genetic variants with at least one significantly associated eGene) corresponding to four eGenes were *pathways in cancer* genes and designated as ART-affected. Fifty-three eSNPs were linked to a single ART-unaffected, *pathways in cancer* eGene, *LAMC3*. These *pathways in cancer* eGenes are components of Wnt/ $\beta$ -catenin, PI3K/AKT, MAPK/ERK, and TGF- $\beta$  signaling (Supplemental Figure 2). These eSNP findings suggest genetic association with cancer risk among a subset of eGenes for both ART-affected and unaffected gene sets.

For the 258 eGenes, we used an 800 kb total *cis* window around the gene body to define regions to test for colocalization with each of the 30 GWAS traits from the partitioned heritability analysis (see Supplemental Table 10 for colocalization test summary statistics). As shown in Figure 4B, 17 traits and 26 eGenes had at least one colocalization event (posterior probability  $>0.75$  of the trait and gene being associated and sharing a causal variant). Despite not observing heritability enrichment for cancer traits, we identified colocalization between genetic signals for pancreatic cancer and *FAM135B* and *RP11-826N14.7*, and lymphocytic leukemia and *CMTM8*. No colocalization tests between eGenes and HIV acquisition met our significance threshold. Additionally, we observed pleiotropic effects for several eGenes including *ANKRD55* (systemic lupus erythematosus, rheumatoid arthritis, red blood cell count, and atopic dermatitis), *CCR6* (rheumatoid arthritis and primary biliary cirrhosis), *FUT11* (type 2 diabetes and primary biliary cirrhosis), *IKZF3* (ulcerative colitis, rheumatoid arthritis, primary biliary cirrhosis, inflammatory bowel disease, and Crohn's disease), *ITGA4* (white blood cell count and primary sclerosing



cholangitis), *LIME1* (platelet count, atopic dermatitis, and Crohn's disease), and *SESN1* (primary sclerosing cholangitis and asthma). Based on cell-type gene expressions in DICE, these pleiotropic eGenes showed varying levels of cell-type specificity (Figure 4C). For example, *ITGA4*, and *SESN1* showed ubiquitous expression, whereas *IKZF3* displayed expression in lymphocyte cell types only, which aligns with its well-established role as a regulator of lymphocyte differentiation<sup>31</sup>. *ANKRD55* exhibited CD4+ T cell specific expression (noted previously in Lopez de Lapuente *et al.*<sup>32</sup>), and *CCR6* showed expression in only naïve B cells and a subset of CD4+ T cell types. Together our partitioned heritability and colocalization analysis results show that the genetic architecture of HIV-associated DEGs primarily overlaps with inflammatory diseases, but shared genetics between these DEGs and the cancers we assessed is very limited or not observed.

### ***Drug target analysis***

To prioritize those DEGs that may have a higher potential for clinical translation, we sought to understand whether the HIV-associated DEGs were known targets for drug compounds in the clinical trials or approved drug development stage. From querying these DEGs against four drug target databases (DrugBank [go.drugbank.com], Open Targets [www.opentargets.org], Pharos [pharos.nih.gov], and Therapeutic Target Database [TTD; db.idrblab.net/ttd]), we observed 111 DEGs that were targets for at least one compound. Of these 111 DEGs, 55 were part of our “ART-affected” classification, and 42 were drug targets for an approved drug. Eleven of the *pathways in cancer* DEGs were drug targets. Three of the 11 were also eGenes, which highlights their potential for genotype-drug interactions (Figure 5; Supplemental Table 11).

Through this analysis, we noted multiple types of potential interaction effects. For example, *PRKCA*, one of the ART-affected eGenes and HIV-associated DEGs involved in non-canonical *Wnt* signaling and MAPK/ERK signaling pathways, is targeted by midostaurin, a drug approved for treatment of acute myeloid leukemia and systemic mastocytosis. *PRKCA* is also targeted by bryostatin 1, which has been in clinical trials for Alzheimer's disease and anti-cancer treatment. Bryostatins have also been observed to mildly inhibit acute HIV infection *in vitro* and activate latent HIV in monocytes and lymphocytes<sup>33,34</sup>.

## **Discussion**

Prior research to understand transcriptomic dynamics within the context of HIV infection has focused on changes during acute infection. In contrast, this study focused on our increasing need to understand HIV disease as a chronic condition, being the first to evaluate transcriptional alterations genome-wide among PLWH that are virally suppressed on ART. This large TWAS in whole blood identified 566 DEGs which show significant enrichment for immune system and cancer pathways – a substantial contributor to excess mortality among PLWH. Identification of eQTLs for DEGs and their assessment in partitioned heritability and colocalization analyses indicated that the observed pathway enrichments are generally not driven by underlying genetic susceptibility, suggesting the state of gene dysregulation related to these outcomes may be malleable with treatment. Intersecting this study's 566 DEGs with results from the singular pre/post study of the gene expression effects of ART among men LWH, we found that approximately 40% of the HIV-associated DEGs may be affected by ART. Evaluation of drug target databases identified many DEGs that are targets of one or more FDA approved drugs and investigational compounds. Overall, these results suggest that gene dysregulation among virally suppressed PLWH may be a key focus for understanding the biological linkage between HIV and excess risk of non-AIDS defining cancers, as well as suggesting studies of potential targets for pharmacological interventions.

Although previous HIV transcriptomics studies exist, they involve only tens of samples which is substantially smaller than our study (N=555). Moreover, there are substantial variations in experimental designs (cell culture models, acute infection studies, differing tissue or cell type of interest, differences in gene expression assay technology, etc.)<sup>35</sup> and statistical approaches which limits the feasibility of using prior studies to replicate the DEGs we identified or to evaluate prior studies finding in our results. However, we applied a robust approach to identifying DEGs by leveraging data-driven estimation of confounding variables and applying bootstrapping for our TWAS with a conservative multiple testing correction threshold, which supports confidence in our results.

Our observations of differential gene expression in chronic HIV disease relate to multiple elements of immune function. This aligns with clinical observations of persistent inflammation and immune dysfunction in PLWH that can affect treatment options and plans of care for co-occurring conditions<sup>3,36</sup>. GO terms and biological pathways with an overrepresentation of HIV-associated DEGs highlight multiple facets of immunity. Chemotaxis and cell migration processes are vital in directing leukocyte localization<sup>37</sup>. Calcium ion transport and signaling is involved in numerous immune processes such as lymphocyte development<sup>38</sup>, T cell and B cell receptor function<sup>39,40</sup>, mast cell activation<sup>41</sup>, and macrophage and dendritic cell phagocytosis<sup>42</sup>. Similarly, cell adhesion is critical to many immune processes including leukocyte homing to sites of inflammation, lymphocyte activation, and pathogen recognition and phagocytosis<sup>43-46</sup>. These clear connections between identified DEGs to expected dysregulation of the immune system under HIV supports confidence in our results.

### *Dysregulated Signaling Pathways*

Our findings also link HIV-associated gene expression to several molecular signaling pathways associated with cancer, e.g., Wnt/ $\beta$ -catenin, PI3K/AKT, MAPK/ERK, p53, and TGF- $\beta$  signaling. Wnt/ $\beta$ -catenin signaling affects cell proliferation and is initiated by extracellular signals and Wnt proteins. Interactions between Wnt and cell membrane proteins regulate  $\beta$ -catenin, which can translocate from the cytoplasm to the nucleus to initiate transcription of target genes<sup>47</sup>. HIV-associated DEGs in the Wnt/ $\beta$ -catenin signaling pathway included *WNT7A*, *WNT10A*, *AXIN2*, *TCF7*, *LEF1*, *PPARD*, and *PRKCA*. *WNT7A* and *WNT10A* produce Wnt proteins; *AXIN2* contributes to a protein complex that degrades  $\beta$ -catenin in the cytoplasm; and *TCF7* and *LEF1* encode transcription factors TCF-1 and LEF-1, respectively, which recruit  $\beta$ -catenin to initiate gene transcription of target genes, including *PPARD*. *PRKCA* encodes protein kinase C alpha, which is involved in non-canonical Wnt signaling that can activate cell proliferation independent of  $\beta$ -catenin<sup>48</sup>. These genes all show lower mean expression among HIV+ samples, which largely indicates a repression of Wnt/ $\beta$ -catenin signaling pathway and cell proliferation. However, decreased expression of *AXIN2* implies decreased  $\beta$ -catenin degradation, allowing  $\beta$ -catenin to move into the nucleus. Knockdown experiments for the TCF-1 and LEF-1 transcription factors have been linked to disrupted T cell development and lymphoma in mice<sup>49,50</sup>.

PI3K/AKT signaling is a pathway related to cell proliferation and apoptosis. PI3K is an enzyme that phosphorylates the cell membrane phospholipid PIP2 to form PIP3 which activates the protein kinase AKT. AKT then interacts with many downstream targets to inhibit or activate various cellular processes. HIV-associated DEGs in this pathway included *LAMC3*, *ITGA6*, *PIK3R1*, *PIK3CB*, *PML*, *FOXO1*, and *CDKN1B*. *LAMC3* and *ITGA6* encode an extracellular matrix laminin subunit and plasma membrane integrin protein subunit alpha 6, respectively. Integrins function as receptors for laminins, and *ITGA6* activates PI3K with overexpression linked to multiple cancers<sup>51,52</sup>. *PIK3R1* and *PIK3CB* are genes for PI3K regulatory and catalytic subunits, respectively. *PML* encodes promyelocytic leukemia proteins that

coordinate with the phosphatase PTEN to inhibit AKT<sup>53</sup>. *FOXO1* and *CDKN1B* are genes for targets of AKT, and their inhibition has been linked to tumorigenesis. AKT inactivates the FOXO1 transcription factor by phosphorylation, subsequently repressing apoptosis<sup>54</sup>. Similarly, AKT inactivates the cyclin-dependent kinase inhibitor p27 (i.e., the *CDKN1B* gene product), which increases cell proliferation<sup>55</sup>. Except for *PML*, the DEGs associated with PI3K/AKT signaling show decreased mean expression in HIV+ samples relative to HIV- samples, suggesting decreased PI3K activity. However, the decreased expression of FOXO1 and p27 genes implies AKT is still being activated, which may be occurring through PI3K-independent processes<sup>56</sup>.

In MAPK/ERK signaling, an extracellular signal leads to phosphorylation of the small GTPase Ras, resulting in a protein kinase cascade that activates ERK. Activated ERK interacts with transcription factors in the nucleus to promote cell proliferation. Ras can also promote cell proliferation by directly activating PI3K. HIV-associated DEGs involved in this pathway included *TGFA* (downregulated in HIV+ samples), *PRKCA* (downregulated), *RASSF1* (upregulated), and *DAPK3* (upregulated). *TGFA* encodes TGF- $\alpha$ , which binds to the plasma membrane epidermal growth factor receptor (EGFR) to activate Ras. Protein kinase C alpha (product of *PRKCA*) phosphorylates Raf-1, a kinase downstream of Ras in the kinase cascade<sup>57</sup>. *RASSF1* produces multiple RASSF1 protein isoforms that can interact with Ras and mediate apoptosis<sup>58,59</sup>. *DAPK3* belongs to a gene family associated with apoptosis, and knockdown of DAPK3 has been shown to inhibit phosphorylation of ERK. Interestingly, DAPK3 exhibits dual properties as a tumor suppressor in some cell types and a cancer cell proliferator in others<sup>60</sup>.

The p53 signaling pathway is another pathway that regulates cell proliferation and apoptosis<sup>61</sup>. HIV-associated DEGs involved in this pathway included kinase inhibitor genes *CDKN2A* (upregulated) and *CDKN1B* (downregulated), protein kinase gene *CDK6* (upregulated), and *PML* (upregulated). *PML* was described above as an inhibitor of AKT in PI3K/AKT signaling, but in the context of p53 signaling, it functions both upstream and downstream of p53 tumor suppressor activation in the pathway. *PML* recruits p53 to nuclear bodies to enhance transcription of downstream target genes, but also is directly regulated by p53<sup>62</sup>. *CDKN2A* produces multiple protein isoforms, including tumor suppressor protein p14<sup>ARF</sup> and kinase inhibitor p16. p14<sup>ARF</sup> binds to ubiquitin-protein ligase MDM2 which prevents p53 from being degraded<sup>63</sup>. The p16 isoform inhibits the cell cycle progression protein complex CDK4/6 (for which *CDK6* encodes the catalytic subunit) from phosphorylating Rb. Phosphorylated Rb activates the E2F transcription factor, leading to cell proliferation. Similarly, p27 protein encoded by *CDKN1B* inhibits CDK2, another kinase that phosphorylates Rb.

In the TGF- $\beta$  signaling pathway, only *TGFBR2* was identified as an HIV-associated DEG, downregulated in HIV+ samples. *TGFBR2* encodes one component of TGFBR, the transmembrane receptor protein heterodimer for TGF- $\beta$  growth factor. When bound with TGF- $\beta$ , TGFBR phosphorylates SMAD proteins that initiate a signaling cascade promoting cell growth. Notably, Liu et al. demonstrated that depletion of TGFBR2 stopped cancer progression in CD4+ T cells<sup>64</sup>.

### ***DEGs and cancer pathways***

Incidence of non-AIDS defining cancers is higher among PLWH<sup>4,65</sup> which aligns with our observations of HIV-associated DEGs being overrepresented in pathways that impact cell growth, proliferation, and apoptosis function. These processes are associated with many cancer types<sup>15,16,18,66-68</sup>. However, the direction of effect for the DEGs in these cancer-related pathways do not clearly confer a consistently increased or decreased cancer risk. For example, the observed decreased expression of *TCF7* and *LEF1* in



PLWH in Wnt/ $\beta$ -catenin signaling would be expected to increase aberrant cell proliferation and cancer risk<sup>49,50</sup>. However, within the same pathway, the decreased expression of *WNT7A* and *WNT10A* would be expected to decrease cell proliferation, resulting in reduced cancer risk. These seemingly contradicting findings raise two important points. First, the directions of effect of our HIV-associated DEGs and their magnitude are based on mean group differences, so the behavior of these genes and related pathways in an individual may deviate from expectations based on the group means. Second, our findings are based on whole blood RNA-seq data and these DEGs may function differently in other tissues that may be more relevant for a given cancer type. Characterizing the impact of these genes in a specific context requires additional functional studies with refined hypotheses.

### *Intersection of HIV and ART associated DEGs*

Studying transcriptomic differences in PLWH on ART compared to HIV negative controls advances knowledge of persistent biological effects of living with HIV as a chronic condition. However, the confounding of chronic HIV infection and ART poses limitations in interpreting the impact of infection and treatment, separately. Clarifying the transcriptomic impact of antiretroviral drugs versus chronic HIV infection *in vivo* would benefit from peripheral blood RNA-seq studies in individuals using pre-exposure prophylaxis to prevent HIV infection. No publicly available data are available with this experimental design, so we used genes identified by Massanella *et al.*<sup>28</sup> (comparing pre/post ART among PLWH) to categorize HIV-associated DEGs as ART-affected or unaffected. This comes with the caveat that Massanella *et al.* only included males with likely varying combinations of drug classes used for ART among participants. Nevertheless, we observed substantial overlap (40%) between DEGs associated with virally suppress PLWH in our study and DEGs associated with post/pre ART in Massanella *et al.*<sup>28</sup> The predominant opposite direction of effect for these overlapping genes (e.g., upregulation in virally suppress PLWH compared to HIV- individual, compared with down regulation in post- vs. pre-ART among men LWH) suggests the expected drive of gene regulation toward the uninfected state under ART. Although ART-affected genes showed statistically significant enrichment for a number of pathways, including cancer-related pathways, and the ART-unaffected genes did not, a number of ART-unaffected genes also contribute to these significant pathways. Thus, while it appears that DEGs associated with ART influence a multitude of immune and cancer-related molecular processes, gene dysregulation in these processes cannot be exclusively attributable to ART. Follow-up functional studies will be needed to determine the mechanistic attribution of these observed associations.

### *Cell-type deconvolution*

Analysis of our HIV-associated DEGs using gene expression deconvolution reinforces several previously observed changes to leukocytes after HIV infection and ART treatment. An increase in CD8+ T cells and a decrease in CD4+ T cells has been well-documented<sup>69-71</sup>. However, prior literature provides a more complex picture for peripheral Tfh (pTfh) cells. Our observed increase in pTfh cells among PLWH has only been recorded in specific subtypes of pTfh cells (e.g., central memory CD27+CD45RO+ pTfh cells)<sup>72,73</sup>. Due to limitations of the cell type reference panel used for deconvolution of our whole blood gene expression data, estimating pTfh subtype proportions was not possible. The pTfh cell type is a major source of HIV latent reservoir that is highly susceptible to HIV infection<sup>74</sup>, so increases in pTfh cell proportion in virally suppressed PLWH on ART may indicate a larger HIV latent reservoir<sup>75</sup>. In relation to our finding of mast cell proportion differences by HIV status, we note that mature mast cells are tissue-resident, so our estimates from deconvolution may be relying on cell-type gene expression markers that lack the specificity to differentiate between mast cells and basophils or their common progenitors.

Regardless, a change in these cell populations warrants further experimental validation given that mast cells have been noted to harbor HIV reservoirs and basophils are involved in viral transmission to CD4+ T cells<sup>76</sup>. From analyzing the DICE single cell expression data for the HIV-associated DEGs, we noted genes expressed across the spectrum of cell type selectivity (e.g., *KLRD1* expressed predominantly in NK cells versus *CD40LG* expressed ubiquitously across T cell types), indicating that gene dysregulation associated with chronic HIV may occur across many immune cell types. Though informative, these data from DICE do not include HIV status information to directly assess cell-type specific differential gene expression for the bulk tissue DEGs. Single-cell RNA sequencing studies that compare people living with chronic HIV to people without HIV are needed to further dissect chronic HIV impact on gene expression across and within immune cell types. Similar work in acute HIV infection for PLWH not undergoing ART has provided insight regarding variability in immune response across cell types and individuals<sup>77</sup>.

### *DEGs and genetic susceptibility*

We considered whether our data supported gene-level findings from genomic studies of HIV acquisition since those studies also compare PLWH to people without HIV. Powell et al.<sup>78</sup> and Duarte et al.<sup>79</sup> conducted genetic association analyses that reported genes expressed in blood that were related to genetic risk of HIV acquisition (*EFCAB14*, *HERC1*, *HIST1H4L*, *IER3*). None were identified as DEGs in our study, suggesting that these potential HIV susceptibility genes are not persistently differentially expressed in PLWH on ART. Additionally, our partitioned heritability analysis did not show enrichment of genetic associations for HIV acquisition at genomic loci surrounding HIV-associated DEGs, and significant eQTLs for DEGs did not colocalize with genetic signal for HIV acquisition. A lack of heritability enrichment was also noted for all the cancer traits and most immune disorders we evaluated, which supports a hypothesis that the differential gene expression observed in this study is a result of living with chronic HIV infection while on ART and not a shared genetic predisposition with HIV acquisition or various cancers. In contrast, significant heritability enrichment at HIV-associated DEGs for Crohn's Disease, asthma, rheumatoid arthritis, primary biliary cirrhosis, and primary sclerosing cholangitis implies that transcriptional alterations to these DEGs may impact pathogenesis and progression for these diseases. Prior studies have observed differences in incidence or activity of these diseases among PLWH and/or on ART compared to individuals not living with HIV<sup>80-83</sup>. For example, in Crohn's disease, one hypothesis is that decreased CD4 T cell counts resulting from HIV infection attenuates the severity of inflammatory bowel disease<sup>84,85</sup>.

We identified colocalization of genetic signals among several traits and eQTLs for HIV-associated DEGs, despite the general lack of heritability enrichment for most of the traits we evaluated. The genes that exhibited eQTL colocalization with pancreatic cancer, *FAM135B* and *RP11-826N14.7*, are not well-characterized. The latter encodes a long noncoding RNA, which are increasingly studied for their role in cancer development<sup>86</sup>. *CMTM8*, whose eQTL colocalized with lymphocytic leukemia, has been noted in the regulation of apoptosis and cancer inhibition for other cancer types<sup>87</sup>. Several other HIV-associated DEGs had eQTLs colocalized with at least one immune-related trait. *ANKRD55*, whose eQTL colocalized with three autoimmune disorders assessed in this study (systemic lupus erythematosus, rheumatoid arthritis, and atopic dermatitis), has gene expression or variants within and proximal to the gene body previously linked to multiple autoimmune disorders, including rheumatoid arthritis and systemic lupus erythematosus<sup>32</sup>. Similar observations from past studies were made for *CCR6*<sup>88</sup>, *IKZF3*<sup>89,90</sup>, *ITGA4*<sup>91,92</sup>, and *LIME1*<sup>93-95</sup>—additional genes with colocalization signal with immune-related traits in this study. We only observed *SESN1* eQTLs colocalizing with primary sclerosing cholangitis and asthma, but this gene has

been associated with carcinogenesis through its relationship as a p53-mediated mTOR pathway inhibitor<sup>96,97</sup>. Similarly, *FUT11* eQTL colocalization was only observed with type 2 diabetes and primary biliary cirrhosis, but this gene has been linked to regulation of the mTOR pathway, pancreatic cancer, and hepatocellular carcinoma<sup>98,99</sup>.

### *DEGs as targets of pharmacological treatment*

The identification of 111 HIV-associated DEGs as targets for clinical trial or approved drugs provides a survey of potential compounds that could alter the pathways overrepresented for DEGs, such as the cancer-related pathways. The example of midostaurin and bryostatin 1 as drugs that both target the ART-affected, HIV-associated, eGene *PRKCA* showcases the complex interactions to consider when assessing drug efficacy and safety in PLWH. One could hypothesize that midostaurin and bryostatin 1 have a genotype-dependent efficacy that would be further modulated if both drugs were simultaneously introduced within PLWH that are on ART. Ideally the interactions would be synergistic, but even so this would have implications for proper dosing. Furthermore, the antiviral properties of bryostatin 1 and its link to midostaurin through *PRKCA* as a common target suggests midostaurin could have an impact on HIV latency/latency reversal in PLWH. As more PLWH reach older age and polypharmacy becomes more common<sup>100</sup>, identifying and understanding genotype-dependent dosage effects, drug-drug interactions, and how drugs can alter disease risk for common co-occurring conditions becomes increasingly important. This complexity notwithstanding, our results provide a large number of drug – target pairs that bare further study as potential points of intervention to reduce the risk for development of non-AIDS defining cancers.

### *Conclusion*

In this first transcriptome-wide study of virally suppressed PLWH on ART, these findings highlight widespread gene dysregulation. Among these genes modulated by living with HIV and ART are numerous links to altered function in various cancers or immune-related diseases. These genes, and potentially drugs that target them, may be key to developing treatments for co-occurring health conditions in PLWH. Until a cure for HIV infection is developed, understanding the health implications of transcriptomic changes, genetics, drug-drug interactions, and ART while living with HIV will remain a critical component to ensuring PLWH receive appropriately tailored healthcare and to reducing co-occurring conditions associated with living with HIV.

## Online Methods

### *Cohort description*

Phenotypic and demographic data, RNA samples, and DNA samples were obtained from individuals of European ancestry originating from two cohorts with harmonized data for PLWH and people seronegative for HIV-1 who inject drugs in Vancouver, Canada (The AIDS Care Cohort to evaluate Exposure to Survival Services [ACCESS] and the Vancouver Injection Drug Users Study [VIDUS]). Details of cohort design and recruiting for the ACCESS and VIDUS cohorts have been described in detail elsewhere<sup>101</sup>. Phenotype, genotype, and gene expression data from 361 PLWH classified as virally non-detectable (<200 viral copies/ml of blood sample) and 194 individuals without HIV were used for this study. We refer to the combined studies as the Vancouver People Who Inject Drugs Study (VPWIDS). Biological samples obtained from VPWIDS are denoted as “HIV+ samples” when obtained from PLWH and “HIV- samples” when obtained from study participants living without HIV. Participants provided

informed consent using methods approved by the Institutional Review Boards of University of British Columbia – Vancouver and RTI International.

### *RNA sequencing and data processing*

RNA was extracted from whole blood samples of VPWIDS participants and prepared for sequencing using the NuGEN Universal Plus mRNA-seq library preparation kit with human globin AnyDeplete primers. RNA samples were sequenced using paired-end 75 bp reads on an Illumina NextSeq by the Rutgers University Cell and DNA Repository (RUCDR Infinite Biologics). Following RNA sequencing (RNA-seq), read adapters were trimmed using cutadapt v1.15<sup>102</sup>, and reads less than 20 bp were removed using sickle v1.33 (<https://github.com/najoshi/sickle>). To generate quality control metrics, adapter trimmed RNA-seq reads were mapped to the GRCh38 human reference transcriptome (GENCODE v28) using HISAT2 v2.1.0<sup>103</sup>. Transcript-level quantifications were estimated using Salmon v0.11.2<sup>104</sup>. Gene-level quantifications were derived from the transcript-level estimates using the R package, *tximport*<sup>105</sup>.

To identify potential sample swaps, genotype correlations derived from genotyping array and RNA-seq data were calculated. Genotypes were called from RNA-seq alignments using the *mpileup* utility from *SAMTools*<sup>106</sup>. Problematic samples were identified by calculating pairwise genotype correlations (Pearson's *r*) between the RNA-seq and genotype datasets. Samples were either excluded or mismatches resolved if  $r > 0.8$  for non-matching sample IDs or  $r < 0.6$  for matching sample IDs.

Sex discrepancies were assessed following the resolution of sample swaps. Males and females formed two distinct chromosome Y gene expression clusters, resulting in a bimodal distribution of normalized per-sample mean count values across chromosome Y genes. Consequently, discrepant sex assignments based on k-means (k=2) clustering of the mean count values of chromosome Y genes were excluded. The following sample-level exclusions were applied: low transcript diversity (unique transcripts detected per-sample) defined as the lower quartile - 1.5\*IQR; read duplication rate  $\geq 0.75$ ; missing genotype data; problematic based on sample swap analysis; sex discrepancies. RIN  $> 5$  and ribosomal RNA mapping rate were also considered as quality filters but did not result in any additional sample exclusions.

### *Differential gene expression analysis*

Differential gene expression by HIV status was tested using RNA-seq data from 361 PLWH and 194 people living without HIV. Prior to model fitting, genes considered lowly expressed were excluded ( $\leq 10$  counts in  $\leq 30\%$  of samples [the approximate proportion of the smaller HIV status group]). Using the 20,211 remaining genes, surrogate variables (SVs) were estimated with *SVA* v3.34.0<sup>107</sup> on variance-stabilized, median-of-ratios normalized counts<sup>108</sup> to capture sources of technical and biological gene expression variance orthogonal to variation by HIV status.

A bootstrap sampling approach was taken to identify differentially expressed genes (DEGs) by HIV status, because no whole blood data set with a comparable sample size and experimental design exists to conduct a suitable replication analysis. For each bootstrap iteration, 80% of the HIV+ and HIV- samples were randomly selected (i.e., stratified sampling) and analyzed with *DESeq2* v1.26.0<sup>109</sup> using a negative binomial generalized linear model with HIV status and 30 SVs as explanatory variables. Shrinkage was applied to fold change estimates using *apeglm*<sup>110</sup>. After 1000 bootstrap iterations, we used the p-values across bootstrap iterations for a given gene to determine statistical significance. We expect robust signals to have p-value distributions with values mostly or exclusively below a stringent threshold. Given a p-value significance threshold  $\alpha = 0.05$ , we applied Bonferroni correction to obtain an

adjusted alpha,  $\alpha_{\text{adjusted}}=2.47\text{e-}06$  ( $\alpha/20211$ ) and calculated a permutation p-value from how often a gene's p-value is smaller than  $\alpha_{\text{adjusted}}$ . We classified a gene as a DEG if  $\geq 95\%$  of its bootstrap p-values are less than  $\alpha_{\text{adjusted}}$  (permutation p-value  $\leq 0.05$ ).

### ***Gene set overrepresentation analysis***

Gene set overrepresentation analysis was conducted using the R package *clusterProfiler* (v4.2.0)<sup>111</sup> with two gene set databases: Gene Ontology (GO) biological process terms and the Kyoto Encyclopedia of Genes and Genomes (KEGG) canonical pathway terms. GENCODE v28 IDs for DEGs were converted to ENSEMBL IDs for compatibility with GO and KEGG database IDs used by *clusterProfiler*. The semantic similarity measure proposed by Wang *et al.*<sup>112</sup> was used to calculate pairwise semantic distances between significant GO terms (FDR  $< 0.05$ ). The semantic relationships among GO terms were represented in a two-dimensional coordinate space using UMAP<sup>113</sup> followed by clustering using HDBSCAN<sup>114</sup>. These discrete groupings do not modify the GO term results but organize terms to aid in biological interpretation and categorization.

### ***Bulk gene expression deconvolution and differential cell-type proportion testing***

To estimate cell-type proportions, median-of-ratios normalized transcripts per million (TPM) values were obtained and genes considered lowly expressed were excluded ( $\leq 10$  counts in  $\leq 30\%$  of samples [the approximate proportion of the smaller HIV status group]). Expression data for genes that constitute that CIBERSORTx LM22 cell-type reference panel were input to CIBERSORTx<sup>115</sup> to obtain cell-type proportion estimates for 22 blood cell types. CIBERSORTx was run with the following parameter specifications: B-mode batch correction enabled, quantile normalization disabled, absolute run mode specified, and 1000 permutation iterations. Expression data for genes overlapping with the LM22 source gene expression profile matrix (11,724 of 11,845 genes) was provided for use in B-mode batch correction.

Prior to differential cell-type proportion testing, we identified sample outliers using a CIBERSORTx quality metric for cell-type proportion estimation. This metric equates to the correlation between the original bulk tissue sample gene expression and the reconstituted bulk gene expression from the estimated cell-type proportions. Any sample with a quality metric smaller than three standard deviations below the mean were considered outliers and excluded (seven sample exclusions). To identify cell-type proportion estimates that differed by HIV status, a linear regression model was fit for each cell type with the cell type proportions as the outcome variables. The explanatory variables for all models included HIV status, sex, ever vs. never sexually transmitted disease status, opioid use status (whether heroin was used within the last six months), age at blood draw, and five genotype PCs. Two-sided *t*-tests were conducted to identify significant, non-zero coefficient estimates for HIV status. Cell-type proportions were considered significantly different by HIV status for Bonferroni adjusted  $p < 0.05$ .

### ***Genotyping and imputation***

DNA was extracted from whole blood samples and genotyped on Illumina Infinium OmniExpress-24 BeadChip arrays by The Genomics Center at Rutgers University. Samples were removed during quality control using the following criteria: genotype missing rate  $> 3\%$ , sample duplication (identity-by-state  $> 90\%$ ), first-degree relatedness (identity-by-descent  $> 40\%$ ), discordance between self-reported sex and genotype-derived sex ( $F_{\text{st}} < 0.2$  for chromosome X single nucleotide polymorphisms to confirm females and  $F_{\text{st}} > 0.8$  to confirm males, where  $F_{\text{st}}$  denotes the coefficient of inbreeding), excessive homozygosity ( $F_{\text{st}} > 0.5$  or  $F_{\text{st}} < -0.2$ ), discordance between self-reported race and genotype-derived



ancestry (AFR  $\geq 25\%$  and EAS  $\geq 25\%$  when using STRUCTURE<sup>116</sup> with 1000 Genomes Phase 3 reference panel AFR, EUR, and EAS superpopulations<sup>117</sup>) or chromosomal anomalies. Genetic variants were removed due to missing rate  $>3\%$  across samples or Hardy-Weinberg equilibrium  $P < 1 \times 10^{-4}$ . Following quality control, phasing and imputation were performed with the TOPMed Imputation Server<sup>118</sup> using Eagle v2.4<sup>119</sup>, Minimac4<sup>120</sup>, and the TOPMed reference panel (genome build GRCh38)<sup>121</sup>. Imputed variants with R-squared  $<0.8$  and minor allele frequency  $\leq 1\%$  were removed.

### Genotype PCA

Genotype principal components were derived using linkage disequilibrium (LD)-pruned observed genotypes (i.e., pre-imputation). Regions of high LD (GRCh37 coordinates chr5:43964243-51464243; chr6:24892021-33392022; chr8:7962590-11962591; chr11:45043424-57243424) were first removed, followed by LD-pruning using PLINK v1.9<sup>122</sup> with the option “--indep-pairwise 1500 150 0.2”. The top 10 principal components were then computed using EIGENSTRAT v7.2.1<sup>123</sup>.

### Expression quantitative trait loci mapping

For HIV-associated DEGs, we tested whether imputed genetic variants in *cis* were expression quantitative trait loci (eQTL) using the following steps: 1) latent factor estimation and selection, 2) *cis* window size selection, 3) linear regression model fitting and association testing, and 4) multiple testing correction. Across steps, we used the following linear model structure with Matrix eQTL<sup>124</sup> for variant-gene association testing:

$$Y = \beta_0 + \gamma * g + \sum_1^k \beta_i * l_i + \varepsilon \quad (1)$$

Here,  $Y$  is the median-of-ratios normalized counts for a given gene (additionally transformed using a rank-inverse normal transformation in step 3),  $g$  is the imputed genotype dosage, and  $l_i$  are latent factors estimated by PEER<sup>125</sup> to capture sources of technical and biological gene expression variance not attributable to HIV status or genetic factors. The  $\gamma$  and  $\beta_i$  terms are the effect size estimates for the genotype dosage and latent factors, respectively, and  $\varepsilon$  is the error term. Matrix eQTL tests and reports for the significance of  $\gamma$ , with the null hypothesis  $\gamma = 0$  and the alternative  $\gamma \neq 0$ .

Latent factors were selected similarly to the “standard approach” as described by Fan *et al.*<sup>126</sup> Briefly, a variance-stabilizing transformation was applied to  $Y$ . These transformed counts and an HIV status indicator variable were provided as inputs to PEER for estimating 50 latent factors, ordered by decreasing variance. For variants within a 2 megabase (Mb) total window (i.e., 1 Mb upstream and downstream) around chromosome 1 genes, eQTL were mapped for 50 linear models based on equation 1. Each model used a different number of latent factors,  $k$ , with  $k \in \{1, 2, \dots, 50\}$ . For each of the models, gene-level p-values were assigned by choosing the smallest p-value across association tests for a given gene. The number of significant eQTL genes (eGenes) from each model was determined through permutation testing. Genotype dosages were permuted 25 times for each gene for the 50 models to create null, gene-level p-value distributions. We classified eGenes as any gene with  $\geq 95\%$  permutation p-values larger than the non-permutation p-value. By assessing the number of eGenes discovered for each of the 50 models (i.e., eGenes discovered as a function of the numbers of latent factors), 13 factors were selected to balance model complexity with maximizing eGene discovery (Supplemental Figure 5A).

To select a *cis* window size for association testing with the final model, we first applied equation 1 with  $k = 13$  for eQTL mapping of genetic variants within a 2 Mb total window around an HIV status DEG. We observed a systematic decrease in  $-\log_{10}$  p-values with increasing variant distance from a gene

(Supplemental Figure 5B). Using the approximate inflection point for  $-\log_{10}$  p-values as a function of variant distance from a gene, we selected a final model total *cis* window size of 800 kb.

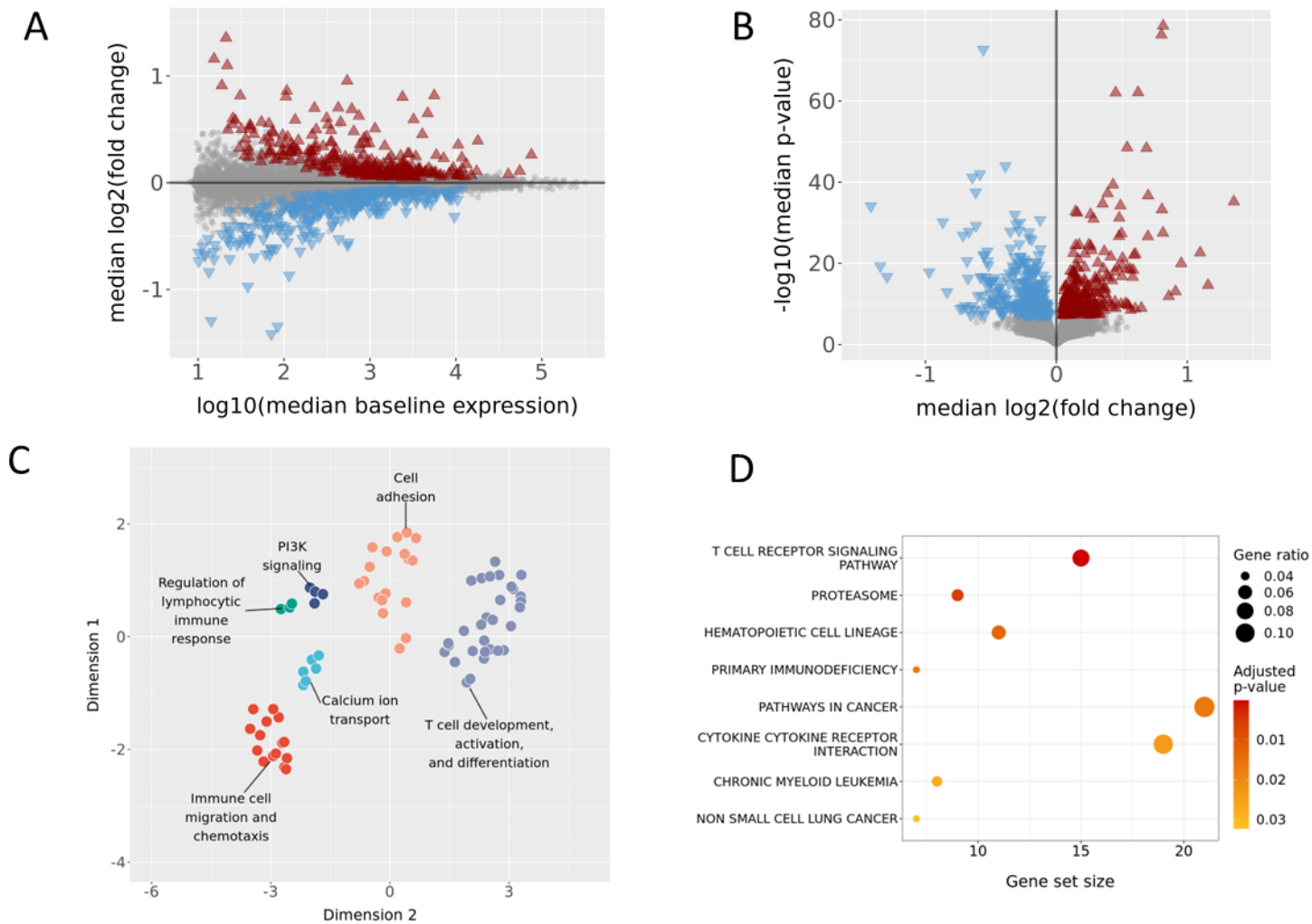
After *cis*-eQTL mapping using equation 1 with  $k = 13$  and a window size of 800 kb, p-values were corrected for multiple hypothesis testing using a two-stage approach. This hierarchical approach first accounts for association tests across variants for a given gene, then corrects for tests across all DEGs. In the first stage, all nominal p-values for a given gene were adjusted using EigenMT<sup>127</sup> with parameters `--var_thresh 0.99`, `--window 200`, and `--cis_dist 400000`. In the second stage, EigenMT adjusted p-values were multiplied by the number of genes tested as a final correction. A two-stage adjusted p-value cutoff of 0.05 was used to identify statistically significant eQTL.

### *Partitioned heritability analysis*

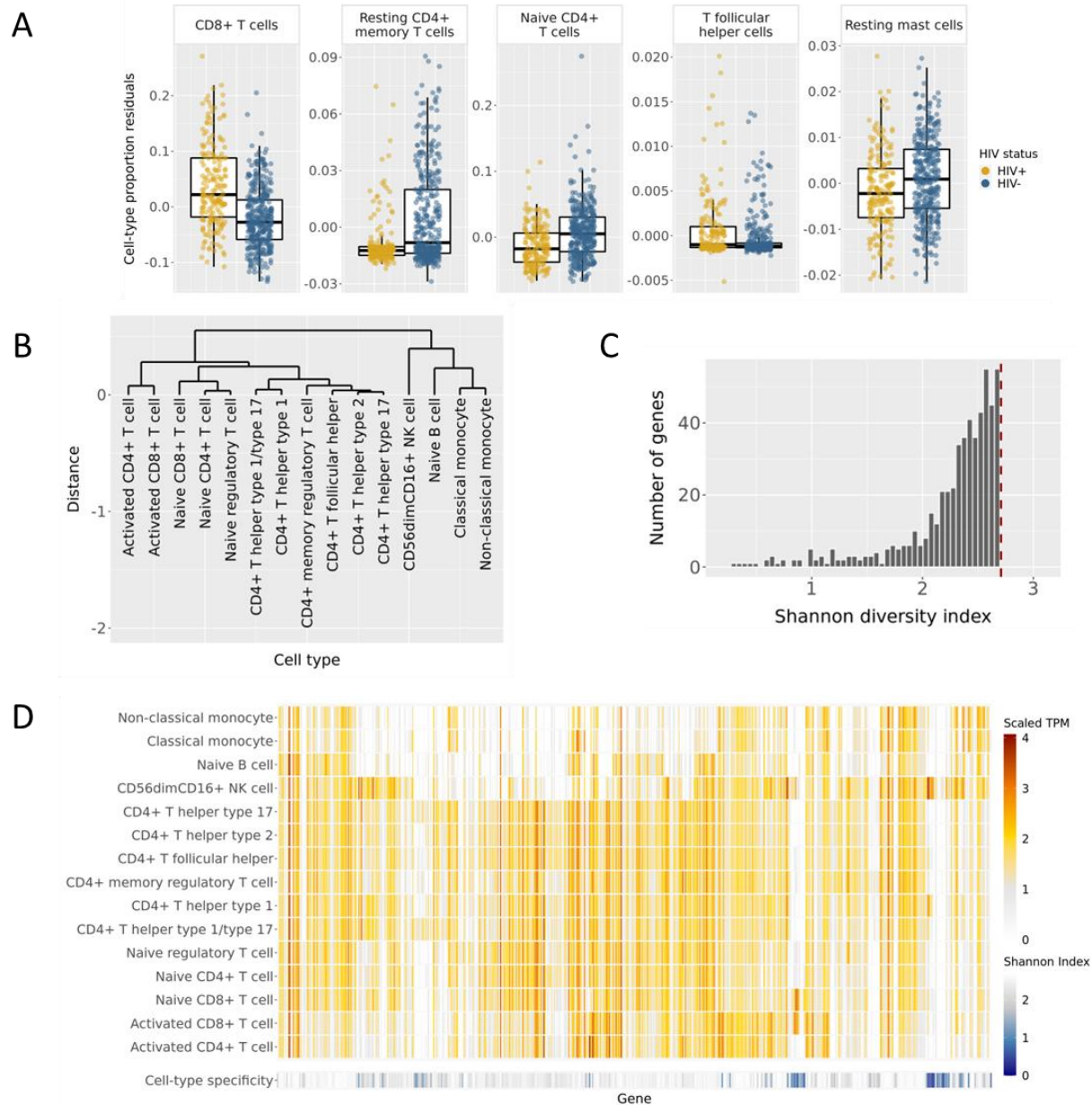
We used stratified LD score regression (s-LDSC) v1.0.1<sup>128</sup> with summary statistics from 30 genome-wide association studies (GWAS) (see Supplemental Table 7 for GWAS details) to test for trait heritability enrichment of genetic variants proximal (up to 100 kb total *cis* window size) to DEGs by HIV status. Aside from GRCh37 genomic coordinates for HIV status DEGs and the 50 kb up and downstream regions around them, several input files for s-LDSC are required that can be obtained from the developers at <https://console.cloud.google.com/storage/browser/broad-alkesgroup-public-requester-pays/LDSCORE>. Specifically, the `--bimfile` and `--bfile` parameters (data for variants used in LD score calculations) used files from `1000G_Phase3_plinkfiles.tgz`, `--ref-ld-chr` (LD scores) used files from `1000G_Phase3_baseline_ldscores.tgz`, `--w-ld-chr` (LD score regression weights) used files from `weights_hm3_no_hla.tgz`, and `--freqfile-chr` (allele frequencies) used files from `1000G_Phase3_freq.tgz`. These files contain data from the 1000 Genomes project Phase 3 EUR superpopulation.

### *Colocalization of GWAS signals and HIV eQTL*

For each HIV-associated DEG with a significant eQTL, colocalization was tested using the `coloc v5.1.0 R` package<sup>129</sup> with eQTL summary statistics for variants proximal (800 kb total *cis* window size) to the DEG and summary statistics from the same 30 GWAS used in partitioned heritability analysis. Each DEG in the eQTL data was specified as a quantitative trait for the `coloc.abf` function, with the sample standard deviation parameter calculated from the RINT and median-of-ratios normalized gene counts. GWAS traits were specified for `coloc.abf` as case-control if the outcome variable was categorical and binary and quantitative otherwise. Proximal variants were subset to only those present in both the eQTL and GWAS data and minor allele frequencies for these variants were calculated based on the imputed genotype dosages for the study cohort. A posterior probability of colocalization (*coloc* hypothesis H4)  $>0.75$  was used to denote a colocalization event between a given eQTL and GWAS locus.

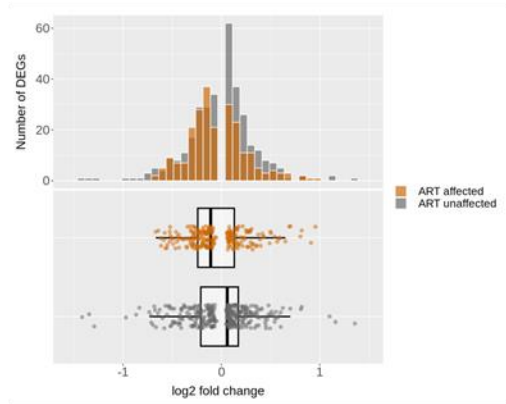


**Figure 1. Differentially expressed genes by HIV status are overrepresented in immune processes and cancer pathways.** A) The median of the average gene expression level and fold change values across 1000 bootstrap sampling iterations for each gene tested for differential expression by HIV status. Significantly up-regulated and down-regulated genes at FDR <0.05 are color-coded red and blue, respectively. Genes not passing the significance threshold are shown as gray points. B) The median of p-values and fold change values across 1000 bootstrap sampling iterations for each gene tested for differential expression by HIV status. Significantly up-regulated and down-regulated genes at FDR <0.05 are color-coded red and blue, respectively. Genes not passing the significance threshold are shown as gray points. C) UMAP projection of semantic distances between GO biological process terms with overrepresentation of HIV status DEGs. GO term clusters identified by HDBSCAN (distinguished by color) are manually assigned labels denoting an underlying theme for terms within a cluster. D) Dotplot of significant (FDR <0.05) KEGG canonical pathway terms from gene set overrepresentation analysis for all HIV status DEGs. The gene ratio indicates the proportion of DEGs in the gene set for a KEGG term relative to the gene set size.

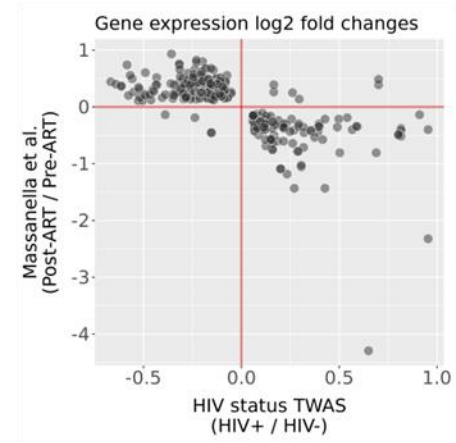


**Figure 2. Gene dysregulation in PLWH impacts multiple immune cell types.** A) Residuals for estimated proportions from the LM22 cell-type reference panel after removing potential confounding effects of biological variables (see Methods). Only cell types with a significant difference by HIV status (Bonferroni adjusted  $p < 0.05$ ) are shown. See Supplemental Figure X for all LM22 cell types. B) Hierarchical clustering of DICE cell types based on Pearson correlations between cell types using expression levels for the HIV-associated DEGs. C) Distribution of Shannon diversity index values for HIV-associated DEGs. The Shannon index values measure uniformity of gene expression levels across 15 immune cell types available from the DICE data repository. Higher values indicate more uniform expression levels across cell types. The red dashed horizontal line indicates the theoretical maximum value for these data (i.e., uniform gene expression levels across all cell types). D) Gene expression levels in 15 DICE immune cell types for HIV status DEGs. Gene expression levels are displayed as  $\log_{10}(\text{mean TPM} + 1)$  values. Shannon diversity index values are displayed for each gene to denote the degree of cell-type specificity (larger values equate to more uniform expression across cell types).

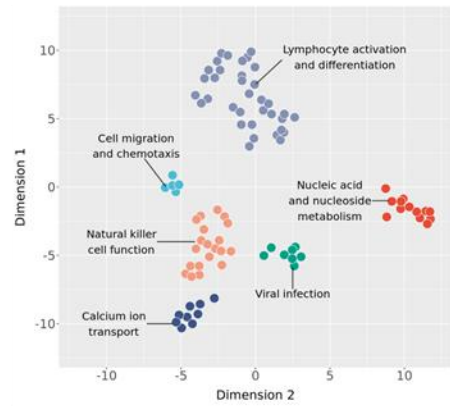
A



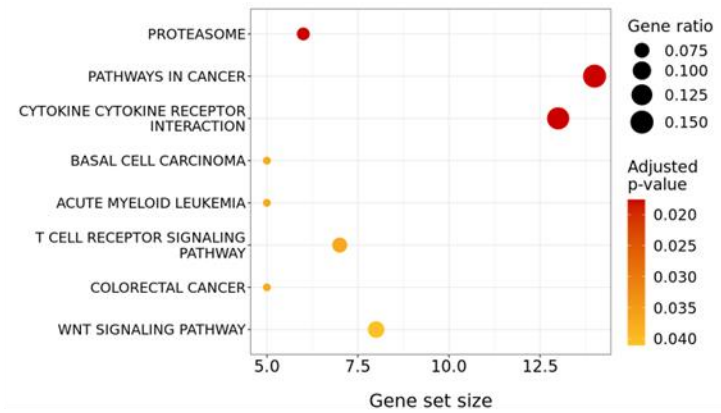
B



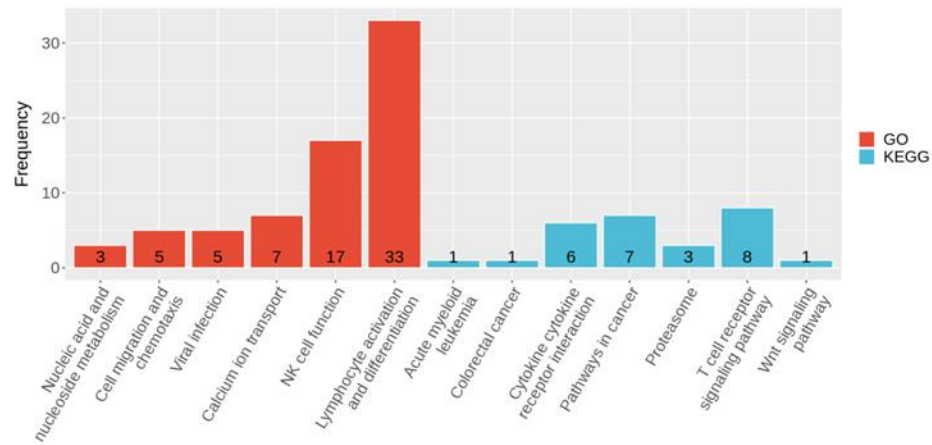
C



D



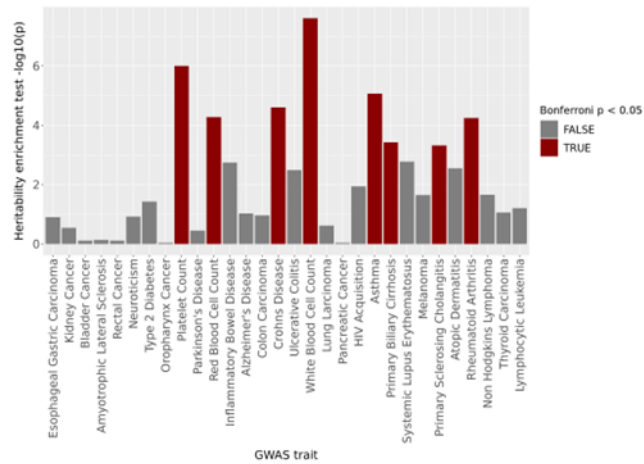
E



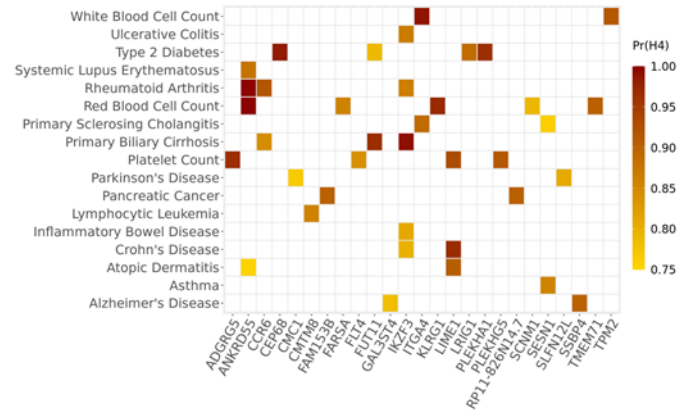


**Figure 3. HIV-associated genes are affected by ART.** A) Distribution of log<sub>2</sub> fold change values for HIV-associated DEGs partitioned by ART status. B) Comparison of differential expression analysis log<sub>2</sub> fold changes between this study and Massanella et al. for ART-affected, HIV-associated DEGs. The log<sub>2</sub> fold change values plotted from this study are the medians of the log<sub>2</sub> fold change values for each gene across 1000 bootstrapping iterations using 80% of the sample size while preserving the HIV+ to HIV- ratio. C) UMAP projection of semantic distances between GO biological process terms with overrepresentation of ART-affected, HIV-associated DEGs. GO term clusters identified by HDBSCAN (distinguished by color) are manually assigned labels denoting an underlying theme for terms within a cluster. D) Dotplot of significant (FDR <0.05) KEGG canonical pathway terms from gene set overrepresentation analysis for ART-affected, HIV-associated DEGs. The gene ratio indicates the proportion of DEGs in the gene set for a KEGG term relative to the gene set size. E) Number of ART-unaffected, HIV-associated DEGs belonging to GO biological process and KEGG canonical pathway terms from gene set overrepresentation analysis for ART-affected, HIV-associated DEGs.

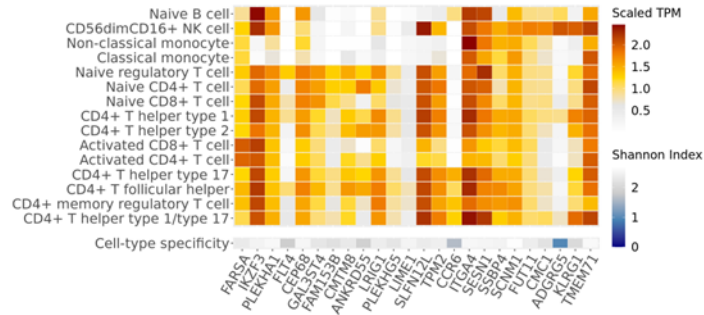
A



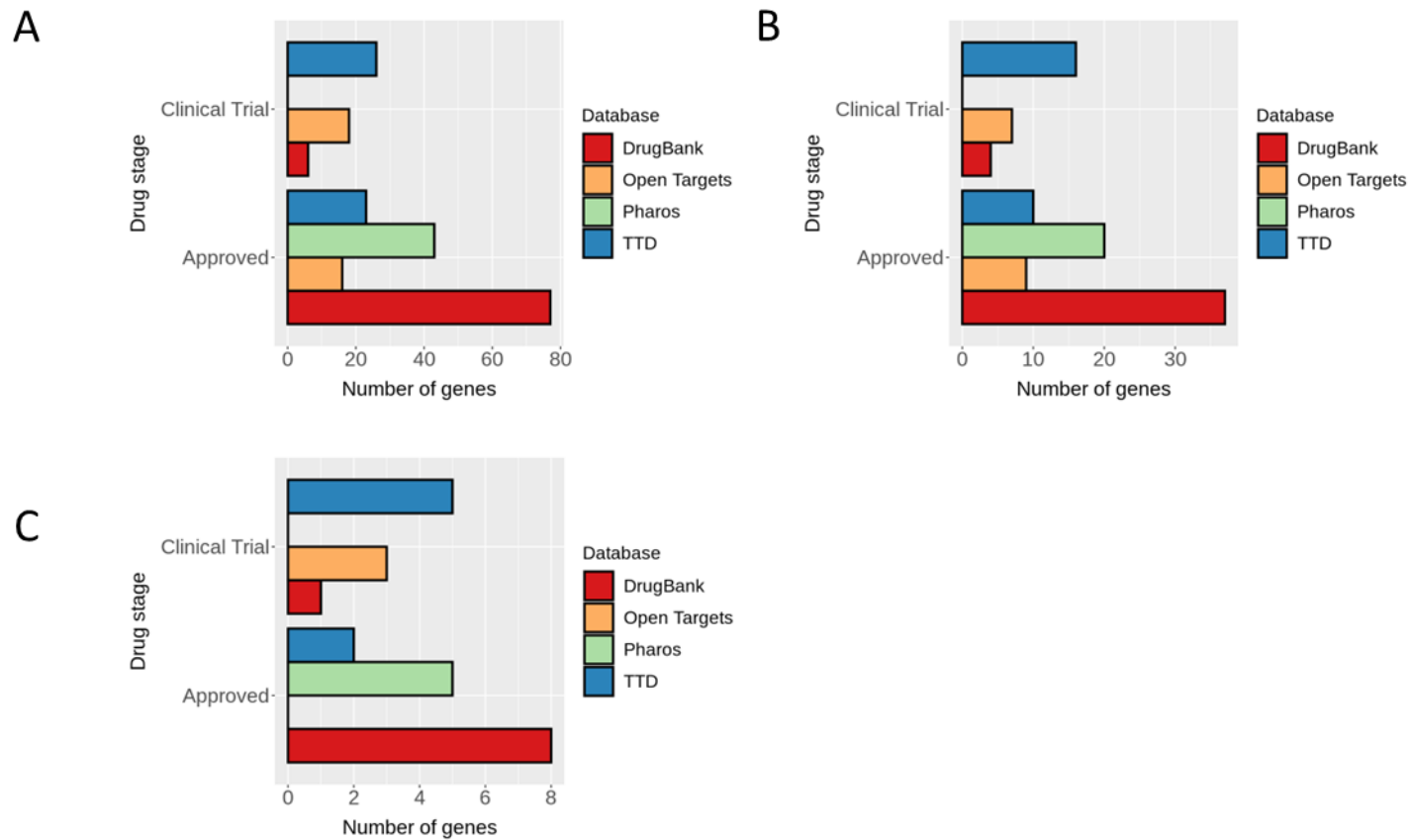
B



C



**Figure 4. HIV-associated DEGs are not strongly associated with known genetic risk for cancer or HIV acquisition.** A) Partitioned heritability enrichment test  $-\log_{10}$  transformed p-values using sLDS for 30 GWAS traits. B) Posterior probabilities for genes (x-axis) that colocalize with at least one of thirty immune-related and cancer GWAS traits. Only gene-trait pairs with a colocalization posterior probability  $>0.75$  are shown here. The posterior probabilities correspond to the colocalization hypothesis that both a gene and trait are associated and share a causal variant. C) Gene expression levels in 15 DICE immune cell types for HIV-associated DEGs in Figure 4B. Gene expression levels are displayed as  $\log_{10}(\text{mean TPM} + 1)$  values. Shannon diversity index values are displayed for each gene to denote the degree of cell-type specificity (larger values equate to more uniform expression across cell types). Gene *RP11-826N14.7* is not included due to not being available in DICE.



**Figure 5. HIV-associated DEGs are targets for FDA-approved drugs or compounds in clinical trials.** Most advanced drug development stage in drug target databases for drug-targeted, HIV status DEGs. A DEG may be present across multiple databases. Consequently, each bar does not represent an independent set of genes. A) All DEGs that are targets for at least one drug (N=111). B) ART-affected DEGs that are targets for at least one drug (N=55). C) KEGG *pathways in cancer* DEGs that are targets for at least one drug (N=11).

## Tables

**Table 1. Sample characteristics and demographics of study participants.**

|  |                      | HIV-          | HIV+         |
|--|----------------------|---------------|--------------|
| <b>Total N</b>                               |                      | 361           | 194          |
| <b>Sex</b>                                   |                      |               |              |
|  | <i>Male, N (%)</i>   | 265 (73.4)    | 148 (76.3)   |
|  | <i>Female, N (%)</i> | 96 (26.6)     | 46 (23.7)    |
| <b>Mean age (SD)</b>                         |                      | 48.96 (10.25) | 51.07 (8.37) |
| <b>STD status</b>                            |                      |               |              |
|  | <i>Never, N (%)</i>  | 90 (24.9)     | 27 (13.9)    |
|  | <i>Ever, N (%)</i>   | 271 (75.1)    | 167 (86.1)   |
| <b>Mean RNA sample RIN (SD)</b>              |                      | 7.13 (0.52)   | 7.07 (0.55)  |
| <b>log<sub>10</sub> viral copies/mL (SD)</b> |                      | 1 (0)         | 1.58 (0.12)  |
| <b>Total (%)</b>                             |                      | 361 (65.05)   | 194 (34.95)  |

Percentages and standard deviations (SD) for specific characteristics are in relation to the respective HIV status group for a given table cell.

## Supplemental Table Legends

**Supplemental Table 1.** Summary statistics for differential gene expression analysis by HIV status.

Bootstrapping was applied to fit 1000 differential expression regression models, resulting in 1000 sets of summary statistics that are summarized in this table as median/median absolute deviation and mean/standard deviation values. The ART affection status is only provided for significant DEGs. “NA” values indicate missing values.

**Supplemental Table 2.** GO Biological Process terms with significant overrepresentation of HIV status DEGs (FDR <0.05) determined using the R package *clusterProfiler*. The gene ratio denotes the number of HIV status DEGs overlapping a given GO term divided by the number of DEGs present across all GO terms. The background ratio indicates the number of genes linked to a given GO term divided by the total number of genes represented across all GO terms. The *Overlapping DEGs* values include ENSEMBL IDs delimited by “/” for DEGs overlapping a given GO term.

**Supplemental Table 3.** MSigDB KEGG canonical pathway terms with significant overrepresentation of HIV status DEGs (FDR <0.05) determined using the R package *clusterProfiler*. The gene ratio denotes the number of HIV status DEGs overlapping a given KEGG term divided by the number of DEGs present across all KEGG terms. The background ratio indicates the number of genes linked to a given KEGG term divided by the total number of genes represented across all KEGG terms. The *Overlapping DEGs* values include ENSEMBL IDs delimited by “/” for DEGs overlapping a given KEGG term.

**Supplemental Table 4.** GO Biological Process and MSigDB KEGG canonical pathway terms with significant overrepresentation of ART-affected, HIV status DEGs (FDR <0.05) determined using the R package *clusterProfiler*. The gene ratio denotes the number of HIV status DEGs overlapping a given GO/KEGG term divided by the number of DEGs present across all GO/KEGG terms. The background ratio indicates the number of genes linked to a given GO/KEGG term divided by the total number of genes

represented across all GO/KEGG terms. The *Overlapping DEGs* values include ENSEMBL IDs delimited by “/” for DEGs overlapping a given GO/KEGG term.

**Supplemental Table 5.** GO Biological Process and MSigDB KEGG canonical pathway terms with significant overrepresentation of ART-unaffected, HIV status DEGs (FDR <0.05) determined using the R package *clusterProfiler*. The gene ratio denotes the number of HIV status DEGs overlapping a given GO/KEGG term divided by the number of DEGs present across all GO/KEGG terms. The background ratio indicates the number of genes linked to a given GO/KEGG term divided by the total number of genes represented across all GO/KEGG terms. The *Overlapping DEGs* values include ENSEMBL IDs delimited by “/” for DEGs overlapping a given GO/KEGG term.

**Supplemental Table 6.** ART-unaffected, HIV status DEGs that overlap with GO Biological Process and MSigDB KEGG canonical pathway terms that had significant overrepresentation of ART-affected DEGs (FDR <0.05).

**Supplemental Table 7.** GWAS traits included in partitioned heritability enrichment and colocalization analysis.

**Supplemental Table 8.** Partitioned heritability enrichment analysis summary statistics output from s-LDSC. Results are shown for testing heritability enrichment of 30 GWAS traits at regions within and 100 kb proximal to HIV status DEGs.

**Supplemental Table 9.** Summary statistics for significant variant-gene associations (adjusted p-value <0.05) from eQTL mapping for HIV status DEGs. Indicator variables are included for a DEG to denote ART affection status and overlap with the MSigDB KEGG canonical pathways term *pathways in cancer*.

**Supplemental Table 10.** Colocalization analysis results for HIV status DEGs and 30 GWAS traits. Each value in the table corresponds to the *coloc* hypothesis 4 (H4) posterior probability for a given DEG and GWAS trait. The H4 posterior probability is defined as the probability that both traits (i.e., the GWAS trait and gene expression for the DEG) are associated and share a single causal variant.

**Supplemental Table 11.** Drugs from DrugBank, PHAROS, Open Targets, or TTD that target an HIV status DEG. The drug ID corresponds to the ID used by the corresponding drug database for a given record. Some drugs have only an ID and no name. These are denoted with the value “NA” for the drug name. Indicator variables are included for a DEG to denote ART affection status, overlap with the MSigDB KEGG canonical pathways term *pathways in cancer*, and whether the DEG has a significant eQTL.

## Author Contributions

Author contributions are categorized using the CRediT taxonomy ([credit.niso.org](https://credit.niso.org)).

BCQ: Formal analysis, methodology, software, visualization, writing – original draft, writing – review & editing

EE: Formal analysis, methodology, software, writing – review & editing

LZ: Formal analysis, software, writing – review & editing

CW: Formal analysis, software, writing – review & editing

JAM: Formal analysis, software, writing – review & editing

JKS: Formal analysis, writing – review & editing

FF: Writing – review & editing



LJB: Writing – review & editing

MJSM: Resources, writing – review & editing

KH: Resources, writing – review & editing

KD: Resources, writing – review & editing

DBH: Conceptualization, writing – review & editing

KX: Conceptualization, funding acquisition, writing – review & editing

BEA: Conceptualization, funding acquisition, writing – review & editing

EOJ: Conceptualization, funding acquisition, project administration, supervision, writing – review & editing

## Conflicts of Interest

None of the authors have conflicts of interest to declare.

## Acknowledgements

This work was funded by the National Institutes of Health grants U01DA038886, U01DA021525, R61/R33DA047011, R01DA051908, and R01DA051908-03S1.

## References

1. Cihlar, T. & Fordyce, M. Current status and prospects of HIV treatment. *Curr Opin Virol* **18**, 50-6 (2016).
2. Guaraldi, G. *et al.* Premature age-related comorbidities among HIV-infected persons compared with the general population. *Clin Infect Dis* **53**, 1120-6 (2011).
3. Chu, C. & Selwyn, P.A. An epidemic in evolution: the need for new models of HIV care in the chronic disease era. *J Urban Health* **88**, 556-66 (2011).
4. Yuan, T. *et al.* Incidence and mortality of non-AIDS-defining cancers among people living with HIV: A systematic review and meta-analysis. *EClinicalMedicine* **52**, 101613 (2022).
5. Deeken, J.F. *et al.* The rising challenge of non-AIDS-defining cancers in HIV-infected patients. *Clin Infect Dis* **55**, 1228-35 (2012).
6. Wang, C.C., Silverberg, M.J. & Abrams, D.I. Non-AIDS-Defining Malignancies in the HIV-Infected Population. *Curr Infect Dis Rep* **16**, 406 (2014).
7. Deeks, S.G., Tracy, R. & Douek, D.C. Systemic effects of inflammation on health during chronic HIV infection. *Immunity* **39**, 633-45 (2013).
8. Lagathu, C. *et al.* Basic science and pathogenesis of ageing with HIV: potential mechanisms and biomarkers. *AIDS* **31 Suppl 2**, S105-S119 (2017).
9. Zheng, J. *et al.* Molecular Changes of Lung Malignancy in HIV Infection. *Sci Rep* **8**, 13128 (2018).
10. Piccaluga, P.P. *et al.* Gene expression analysis uncovers similarity and differences among Burkitt lymphoma subtypes. *Blood* **117**, 3596-608 (2011).
11. Deffenbacher, K.E., Iqbal, J., Liu, Z., Fu, K. & Chan, W.C. Recurrent chromosomal alterations in molecularly classified AIDS-related lymphomas: an integrated analysis of DNA copy number and gene expression. *J Acquir Immune Defic Syndr* **54**, 18-26 (2010).
12. Schmiedel, B.J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715 e16 (2018).
13. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2019. in *HIV Surveillance Supplemental Report 2021 Vol. 26* (2021).

14. Guo, Y.J. *et al.* ERK/MAPK signalling pathway and tumorigenesis. *Exp Ther Med* **19**, 1997-2007 (2020).
15. Yang, J. *et al.* Targeting PI3K in cancer: mechanisms and advances in clinical trials. *Mol Cancer* **18**, 26 (2019).
16. Zhan, T., Rindtorff, N. & Boutros, M. Wnt signaling in cancer. *Oncogene* **36**, 1461-1473 (2017).
17. Aster, J.C., Pear, W.S. & Blacklow, S.C. The Varied Roles of Notch in Cancer. *Annu Rev Pathol* **12**, 245-275 (2017).
18. Yu, Y. & Feng, X.H. TGF-beta signaling in cell fate control and cancer. *Curr Opin Cell Biol* **61**, 56-63 (2019).
19. Owen, K.L., Brockwell, N.K. & Parker, B.S. JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression. *Cancers (Basel)* **11**(2019).
20. Roberts-Thomson, S.J., Chalmers, S.B. & Monteith, G.R. The Calcium-Signaling Toolkit in Cancer: Remodeling and Targeting. *Cold Spring Harb Perspect Biol* **11**(2019).
21. Soni, S. & Padwad, Y.S. HIF-1 in cancer therapy: two decade long story of a transcription factor. *Acta Oncol* **56**, 503-515 (2017).
22. Mitsuishi, Y., Motohashi, H. & Yamamoto, M. The Keap1-Nrf2 system in cancers: stress response and anabolic metabolism. *Front Oncol* **2**, 200 (2012).
23. Grossmann, M.E., Huang, H. & Tindall, D.J. Androgen receptor signaling in androgen-refractory prostate cancer. *J Natl Cancer Inst* **93**, 1687-97 (2001).
24. Newman, A.M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-7 (2015).
25. Shannon, C.E. The mathematical theory of communication. 1963. *MD Comput* **14**, 306-17 (1997).
26. Aramburu, J. *et al.* A novel functional cell surface dimer (Kp43) expressed by natural killer cells and T cell receptor-gamma/delta+ T lymphocytes. I. Inhibition of the IL-2-dependent proliferation by anti-Kp43 monoclonal antibody. *J Immunol* **144**, 3238-47 (1990).
27. Kornbluth, R.S. The emerging role of CD40 ligand in HIV infection. *J Leukoc Biol* **68**, 373-82 (2000).
28. Massanella, M. *et al.* Differential gene expression in HIV-infected individuals following ART. *Antiviral Res* **100**, 420-8 (2013).
29. Pace, M. *et al.* Impact of antiretroviral therapy in primary HIV infection on natural killer cell function and the association with viral rebound and HIV DNA following treatment interruption. *Front Immunol* **13**, 878743 (2022).
30. Gumbleton, M. & Kerr, W.G. Role of inositol phospholipid signaling in natural killer cell biology. *Front Immunol* **4**, 47 (2013).
31. Hosokawa, Y., Maeda, Y., Takahashi, E., Suzuki, M. & Seto, M. Human aiolos, an ikaros-related zinc finger DNA binding protein: cDNA cloning, tissue expression pattern, and chromosomal mapping. *Genomics* **61**, 326-9 (1999).
32. Lopez de Lapuente, A. *et al.* Novel Insights into the Multiple Sclerosis Risk Gene ANKRD55. *J Immunol* **196**, 4553-65 (2016).
33. Mehla, R. *et al.* Bryostatins modulates latent HIV-1 infection via PKC and AMPK signaling but inhibits acute infection in a receptor independent manner. *PLoS One* **5**, e11160 (2010).
34. Darcis, G. *et al.* An In-Depth Comparison of Latency-Reversing Agent Combinations in Various In Vitro and Ex Vivo HIV-1 Latency Models Identified Bryostatins-1+JQ1 and Ingenol-B+JQ1 to Potently Reactivate Viral Gene Expression. *PLoS Pathog* **11**, e1005063 (2015).
35. Le Clerc, S., Limou, S. & Zagury, J.F. Large-Scale "OMICS" Studies to Explore the Physiopathology of HIV-1 Infection. *Front Genet* **10**, 799 (2019).
36. Erlandson, K.M. & Campbell, T.B. Inflammation in chronic HIV infection: what can we do? Vol. 212 339-342 (Oxford University Press, 2015).

37. Rodriguez-Fernandez, J.L. & Cabañas, L.G. Chemoattraction: Basic Concepts and Role in the Immune Response. *eLS* (2013).
38. Feske, S. Calcium signalling in lymphocyte activation and disease. *Nat Rev Immunol* **7**, 690-702 (2007).
39. Omilusik, K. *et al.* The CaV1. 4 calcium channel is a critical regulator of T cell receptor signaling and naive T cell homeostasis. *Immunity* **35**, 349-360 (2011).
40. Matsumoto, M. *et al.* The calcium sensors STIM1 and STIM2 control B cell regulatory function through interleukin-10 production. *Immunity* **34**, 703-14 (2011).
41. Baba, Y. *et al.* Essential function for the calcium sensor STIM1 in mast cell activation and anaphylactic responses. *Nat Immunol* **9**, 81-8 (2008).
42. Vaeth, M. *et al.* Ca<sup>2+</sup> Signaling but Not Store-Operated Ca<sup>2+</sup> Entry Is Required for the Function of Macrophages and Dendritic Cells. *J Immunol* **195**, 1202-17 (2015).
43. Dustin, M.L. Integrins and Their Role in Immune Cell Adhesion. *Cell* **177**, 499-501 (2019).
44. O'Rourke, A.M. & Mescher, M.F. Cytotoxic T-lymphocyte activation involves a cascade of signalling and adhesion events. *Nature* **358**, 253-5 (1992).
45. Bierer, B.E. & Burakoff, S.J. T cell adhesion molecules. *FASEB J* **2**, 2584-90 (1988).
46. Greenberg, S. & Grinstein, S. Phagocytosis and innate immunity. *Curr Opin Immunol* **14**, 136-45 (2002).
47. Liu, J. *et al.* Wnt/beta-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct Target Ther* **7**, 3 (2022).
48. Luna-Ulloa, L.B., Hernandez-Maqueda, J.G., Castaneda-Patlan, M.C. & Robles-Flores, M. Protein kinase C in Wnt signaling: implications in cancer initiation and progression. *IUBMB Life* **63**, 915-21 (2011).
49. Okamura, R.M. *et al.* Redundant regulation of T cell differentiation and TCRalpha gene expression by the transcription factors LEF-1 and TCF-1. *Immunity* **8**, 11-20 (1998).
50. Wang, F. *et al.* Exploring the stage-specific roles of Tcf-1 in T cell development and malignancy at single-cell resolution. *Cell Mol Immunol* **18**, 644-659 (2021).
51. Shaw, L.M., Rabinovitz, I., Wang, H.H., Toker, A. & Mercurio, A.M. Activation of phosphoinositide 3-OH kinase by the alpha6beta4 integrin promotes carcinoma invasion. *Cell* **91**, 949-60 (1997).
52. Kariya, Y., Kariya, Y. & Gu, J. Roles of Integrin alpha6beta4 Glycosylation in Cancer. *Cancers (Basel)* **9**(2017).
53. Trotman, L.C. *et al.* Identification of a tumour suppressor network opposing nuclear Akt function. *Nature* **441**, 523-7 (2006).
54. Lu, H. & Huang, H. FOXO1: a potential target for human diseases. *Curr Drug Targets* **12**, 1235-44 (2011).
55. Chu, I.M., Hengst, L. & Slingerland, J.M. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer* **8**, 253-67 (2008).
56. Mahajan, K. & Mahajan, N.P. PI3K-independent AKT activation in cancers: a treasure trove for novel therapeutics. *J Cell Physiol* **227**, 3178-84 (2012).
57. Kolch, W. *et al.* Protein kinase C alpha activates RAF-1 by direct phosphorylation. *Nature* **364**, 249-52 (1993).
58. Vos, M.D., Ellis, C.A., Bell, A., Birrer, M.J. & Clark, G.J. Ras uses the novel tumor suppressor RASSF1 as an effector to mediate apoptosis. *J Biol Chem* **275**, 35669-72 (2000).
59. Dubois, F., Bergot, E., Zalzman, G. & Levallet, G. RASSF1A, puppeteer of cellular homeostasis, fights tumorigenesis, and metastasis-an updated review. *Cell Death Dis* **10**, 928 (2019).
60. Kake, S., Usui, T., Ohama, T., Yamawaki, H. & Sato, K. Death-associated protein kinase 3 controls the tumor progression of A549 cells through ERK MAPK/c-Myc signaling. *Oncol Rep* **37**, 1100-1106 (2017).

61. Weber, H.O., Samuel, T., Rauch, P. & Funk, J.O. Human p14(ARF)-mediated cell cycle arrest strictly depends on intact p53 signaling pathways. *Oncogene* **21**, 3207-12 (2002).
62. de Stanchina, E. *et al.* PML is a direct p53 target that modulates p53 effector functions. *Mol Cell* **13**, 523-35 (2004).
63. Ozenne, P., Eymin, B., Brambilla, E. & Gazzeri, S. The ARF tumor suppressor: structure, functions and status in cancer. *Int J Cancer* **127**, 2239-47 (2010).
64. Liu, M. *et al.* TGF-beta suppresses type 2 immunity to cancer. *Nature* **587**, 115-120 (2020).
65. Ceccarelli, M. *et al.* Non-AIDS defining cancers: a comprehensive update on diagnosis and management. *Eur Rev Med Pharmacol Sci* **24**, 3849-3875 (2020).
66. Garcia-Gomez, R., Bustelo, X.R. & Crespo, P. Protein-Protein Interactions: Emerging Oncotargets in the RAS-ERK Pathway. *Trends Cancer* **4**, 616-633 (2018).
67. Burotto, M., Chiou, V.L., Lee, J.M. & Kohn, E.C. The MAPK pathway across different malignancies: a new perspective. *Cancer* **120**, 3446-56 (2014).
68. Chen, X. *et al.* Mutant p53 in cancer: from molecular mechanism to therapeutic modulation. *Cell Death Dis* **13**, 974 (2022).
69. Giorgi, J.V. & Detels, R. T-cell subset alterations in HIV-infected homosexual men: NIAID Multicenter AIDS cohort study. *Clin Immunol Immunopathol* **52**, 10-8 (1989).
70. Giorgi, J.V. *et al.* Selective alterations in immunoregulatory lymphocyte subsets in early HIV (human T-lymphotropic virus type III/lymphadenopathy-associated virus) infection. *J Clin Immunol* **7**, 140-50 (1987).
71. Hunt, P.W. *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* **187**, 1534-43 (2003).
72. McCarty, B. *et al.* Low Peripheral T Follicular Helper Cells in Perinatally HIV-Infected Children Correlate With Advancing HIV Disease. *Front Immunol* **9**, 1901 (2018).
73. Pallikkuth, S. *et al.* Impaired peripheral blood T-follicular helper cell function in HIV-infected nonresponders to the 2009 H1N1/09 vaccine. *Blood* **120**, 985-93 (2012).
74. Pallikkuth, S. *et al.* Peripheral T Follicular Helper Cells Are the Major HIV Reservoir within Central Memory CD4 T Cells in Peripheral Blood from Chronically HIV-Infected Individuals on Combination Antiretroviral Therapy. *J Virol* **90**, 2718-28 (2015).
75. Garcia, M. *et al.* Peripheral T follicular helper Cells Make a Difference in HIV Reservoir Size between Elite Controllers and Patients on Successful cART. *Sci Rep* **7**, 16799 (2017).
76. Espindola, M.S. *et al.* HIV infection: focus on the innate immune cells. *Immunol Res* **64**, 1118-1132 (2016).
77. Kazer, S.W. *et al.* Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat Med* **26**, 511-518 (2020).
78. Powell, T.R. *et al.* The behavioral, cellular and immune mediators of HIV-1 acquisition: New insights from population genetics. *Sci Rep* **10**, 3304 (2020).
79. Duarte, R.R.R., Pain, O., Furler, R.L., Nixon, D.F. & Powell, T.R. Transcriptome-wide association study of HIV-1 acquisition identifies HERC1 as a susceptibility gene. *iScience* **25**, 104854 (2022).
80. Adizie, T., Moots, R.J., Hodgkinson, B., French, N. & Adebajo, A.O. Inflammatory arthritis in HIV positive patients: A practical guide. *BMC Infect Dis* **16**, 100 (2016).
81. Foster, S.B. *et al.* Increased incidence of asthma in HIV-infected children treated with highly active antiretroviral therapy in the National Institutes of Health Women and Infants Transmission Study. *J Allergy Clin Immunol* **122**, 159-65 (2008).
82. Mahajani, R.V. & Uzer, M.F. Cholangiopathy in HIV-infected patients. *Clin Liver Dis* **3**, 669-84, x (1999).

83. Puoti, M., Torti, C., Bruno, R., Filice, G. & Carosi, G. Natural history of chronic hepatitis B in co-infected patients. *J Hepatol* **44**, S65-70 (2006).
84. Skamnelos, A. *et al.* CD4 count remission hypothesis in patients with inflammatory bowel disease and human immunodeficiency virus infection: a systematic review of the literature. *Ann Gastroenterol* **28**, 337-346 (2015).
85. Guillo, L. *et al.* Impact of HIV Infection on the Course of Inflammatory Bowel Disease and Drug Safety Profile: A Multicenter GETAID Study. *Clin Gastroenterol Hepatol* **20**, 787-797 e2 (2022).
86. Fang, Y. & Fullwood, M.J. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics* **14**, 42-54 (2016).
87. Wu, J., Li, L., Wu, S. & Xu, B. CMTM family proteins 1-8: roles in cancer biological processes and potential clinical value. *Cancer Biol Med* **17**, 528-542 (2020).
88. Meitei, H.T., Jadhav, N. & Lal, G. CCR6-CCL20 axis as a therapeutic target for autoimmune diseases. *Autoimmun Rev* **20**, 102846 (2021).
89. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).
90. Lessard, C.J. *et al.* Identification of IRF8, TMEM39A, and IKZF3-ZBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. *Am J Hum Genet* **90**, 648-60 (2012).
91. de Krijger, M., Wildenberg, M.E., de Jonge, W.J. & Ponsioen, C.Y. Return to sender: Lymphocyte trafficking mechanisms as contributors to primary sclerosing cholangitis. *J Hepatol* **71**, 603-615 (2019).
92. Garner, C.P. *et al.* Replication of celiac disease UK genome-wide association study results in a US population. *Hum Mol Genet* **18**, 4219-25 (2009).
93. Van Horebeek, L. & Goris, A. Transcript-specific regulation in T-cells in multiple sclerosis susceptibility. *Eur J Hum Genet* **28**, 849-850 (2020).
94. Kugathasan, S. *et al.* Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet* **40**, 1211-5 (2008).
95. Sobczyk, M.K. *et al.* Triangulating Molecular Evidence to Prioritize Candidate Causal Genes at Established Atopic Dermatitis Loci. *J Invest Dermatol* **141**, 2620-2629 (2021).
96. Budanov, A.V. & Karin, M. p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell* **134**, 451-60 (2008).
97. Chen, Y. *et al.* The functions and roles of sestrins in regulating human diseases. *Cell Mol Biol Lett* **27**, 2 (2022).
98. Cao, W. *et al.* Hypoxia-Related Gene FUT11 Promotes Pancreatic Cancer Progression by Maintaining the Stability of PDK1. *Front Oncol* **11**, 675991 (2021).
99. Ruan, W. *et al.* FUT11 is a target gene of HIF1alpha that promotes the progression of hepatocellular carcinoma. *Cell Biol Int* **45**, 2275-2286 (2021).
100. Back, D. & Marzolini, C. The challenge of HIV treatment in an era of polypharmacy. *J Int AIDS Soc* **23**, e25449 (2020).
101. Strathdee, S.A. *et al.* Barriers to use of free antiretroviral therapy in injection drug users. *JAMA* **280**, 547-9 (1998).
102. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).
103. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019).
104. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).



105. Sonesson, C., Love, M.I. & Robinson, M.D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
106. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
107. Leek, J.T. & Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* **3**, 1724-35 (2007).
108. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
109. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
110. Zhu, A., Ibrahim, J.G. & Love, M.I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084-2092 (2019).
111. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
112. Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S. & Chen, C.F. A new method to measure the semantic similarity of GO terms. *Bioinformatics* **23**, 1274-81 (2007).
113. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).
114. Campello, R.J., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. in *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II* 17 160-172 (Springer, 2013).
115. Newman, A.M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019).
116. Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* **9**, 1322-32 (2009).
117. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
118. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
119. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
120. Fuchsberger, C., Abecasis, G.R. & Hinds, D.A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-4 (2015).
121. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
122. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
123. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
124. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-8 (2012).
125. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
126. Fan, Y., Zhu, H., Song, Y., Peng, Q. & Zhou, X. Efficient and effective control of confounding in eQTL mapping studies through joint differential expression and Mendelian randomization analyses. *Bioinformatics* **37**, 296-302 (2021).



127. Davis, J.R. *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am J Hum Genet* **98**, 216-24 (2016).
128. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
129. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).