Genome Biology

RESEARCH

Open Access

Transposable elements may enhance antiviral resistance in HIV-1 elite controllers

Manvendra Singh[1,4†], Sabrina M. Leddy[1†], Luis Pedro Iñiguez[2], Matthew L. Bendall[3], Douglas F. Nixon[3,5*] and Cédric Feschotte[1*]

†Manvendra Singh and Sabrina M. Leddy are co-first authors—these authors have made an equal contribution and thus share first authorship.

*Correspondence: dnixon1@northwell.edu; cf458@cornell.edu

[1] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA
[2] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain
[3] Division of Infectious Diseases, Department of Medicine, Weill Cornell Medicine, New York, NY, USA
[4] Clinical Neuroscience, Max Planck Institute for Multidisciplinary Sciences, City Campus, Göttingen, Germany
[5] Feinstein Institutes for Medical Research, Manhasset, NY, USA

Abstract

Background: Less than 0.5% of people living with HIV-1 are elite controllers (ECs)—individuals who maintain undetectable plasma viremia without antiretroviral therapy, despite having replication-competent viral reservoirs. While EC CD4$^+$ T cells have been investigated for gene expression signatures associated with HIV-1 resistance, the expression and regulatory activity of transposable elements (TEs) remain unexplored. TEs can directly impact host immune responses to pathogens, including HIV-1, suggesting their activities could contribute to HIV-1 elite control. To begin testing this hypothesis, we conduct a TE-centric analysis of public multi-omics data from ECs and other populations.

Results: We find the CD4$^+$ T cell transcriptome and retrotranscriptome of ECs are distinct from healthy controls, from people living with HIV-1 on antiretroviral therapy, and from viremic progressors. However, there is substantial transcriptomic heterogeneity among ECs. We categorize ECs into four clusters with distinct expression and chromatin accessibility profiles of TEs and antiviral factors. Several TE families with known immuno-regulatory activity are differentially expressed among ECs. Their expression positively correlates with their chromatin accessibility in ECs and negatively correlates with the expression of their KRAB zinc-finger (KZNF) repressors. This coordinated, locus-level variation forms a network of putative *cis*-regulatory elements for genes involved in HIV-1 restriction.

Conclusions: We propose that the EC phenotype is driven in part by reduced KZNF-mediated repression of specific TE-derived *cis*-regulatory elements for antiviral genes, heightening their resistance against HIV-1. Our study reveals heterogeneity in the EC CD4$^+$ T cell transcriptome, including variable expression of TEs and their KZNF controllers, that must be considered when deciphering HIV-1 control mechanisms.

Keywords: Genomics, HIV, Retroviruses, Immunology, Gene regulation, Transposable elements

Background

Human immunodeficiency virus 1 (HIV-1) remains a major viral pandemic with a current estimate of 39 million people living with HIV-1 (PLWH), the majority of whom live in Sub-Saharan Africa [1, 2]. Despite the availability of antiretroviral drug therapy (ART), there were more than 1.3 million new infections in 2022 alone [3]. There have been intense efforts to develop an effective vaccine and cure for HIV-1; in fact, over the past decade, several PLWH appear to have been cured [4]. Some of these individuals received transplants with donor *CCR5Δ32* material, rendering their cells resistant to R5 HIV-1 infection [5–7]. There are also post-ART controllers in whom no proviral integrations were detected for several years, suggesting a natural ability for viral maintenance after treatment interruption [8, 9]. Early in the HIV-1 pandemic, longitudinal studies showed that untreated HIV-1 infection led to a progressive loss of $CD4^+$ T cells, but the rate of $CD4^+$ T cell decline varied from person to person [10]. A small number of PLWH had no clinical symptoms and no significant decline in $CD4^+$ T cell count. These long-term non-progressors were dubbed "elite controllers" (ECs). Some ECs developed low-level viremia and $CD4^+$ T cell loss over time, while others maintained viremic control long-term and became known as "exceptional elite controllers" [11–13]. These exceptional ECs include those who appear to have cleared the virus completely [8, 9]. Thus, the ability to control viral replication in PLWH varies widely between individuals. However, our understanding of the molecular mechanisms underlying this range of antiviral resistance remains limited.

Seminal genome-wide association studies of the EC phenotype identified cell-mediated immunity [14] and the HLA-B region as major determinants of viral control [15, 16]. Certain HLA alleles, such as HLA-B*5701 and HLA-B*2705 in Caucasian populations and HLA-B*81:01 in some African studies, are associated with delayed onset of immunodeficiency [17, 18]. Interestingly, alleles are associated with either relatively low viral loads or high $CD4^+$ T cell count [19]. The presumed mechanism of control involves $CD8^+$ T cell-mediated killing of infected targets. Certain alleles restrict HIV-1-derived peptides, driving responses that are better able to suppress and kill infected cells [20]. Subsequent studies of natural killer (NK) cell receptors identified a number of innate immune responses as additional contributors to low-level viremia [21], and humoral immunity has also been suggested to play a role in elite control [22]. More recent studies have characterized proviral reservoirs in ECs [9, 23, 24]. Compared to PLWH-on-ART, intact proviral sequences from ECs were retained at distinct sites in the human genome, preferentially located in centromeric satellite DNA or in KZNF genes on chromosome 19 [9], which are associated with heterochromatin features [25]. Moreover, the integration sites of intact proviral sequences from ECs appear further from transcriptional start sites and accessible chromatin of the host genome than in PLWH-on-ART and were enriched in repressive chromatin marks [9, 24]. In contrast, defective proviruses in ECs were commonly located in permissive genic euchromatin positions [9, 23]. These differences likely result from post-integration selective pressures shaping the proviral landscape that are either EC-specific or enhanced in ECs. Thus, the EC phenotype likely results from a complex combination of genetic and cellular mechanisms.

Transposable elements (TEs) are mobile DNA elements that can replicate and insert themselves into different locations within the host genome. Half of the human genome

is composed of these elements, though most TE loci have long lost the ability to self-propagate due to mutational decay and/or epigenetic silencing from repressive factors like KRAB-containing zinc finger proteins (KZNFs) [26–28]. Despite the loss of transpositional activity, many human TEs retain transcriptional activity and regulatory capabilities, facilitated by their internal promoters and embedded *cis*-regulatory elements including binding sites for various transcription factors. The biological consequences of TE transcriptional and regulatory activities have been the subject of intense investigation. Multiple studies have documented the co-option of TE-derived *cis*-regulatory elements for host gene regulation, including the control of innate immunity genes [26, 29–33]. For example, multiple elements from the human endogenous retrovirus (HERV) family MER41 function as interferon-inducible enhancers triggering the activation of innate immunity genes upon infection, including some involved in restricting HIV-1 [34]. Furthermore, elements from the LTR12C and LTR7 HERV families are known to serve as promoters or enhancers for human genes encoding restriction factors for HIV-1, specifically *GBP2/5* and *APOBEC3G/H*, respectively [35–38]. Thus, interindividual variation in the *cis*-regulatory activity and expression profiles of these and other TEs has the potential to influence susceptibility to viral infection, including HIV-1 [39–41]. Here, we begin testing the hypothesis that the EC phenotype may be driven, in part, by the increased ability of TE-derived *cis*-regulatory sequences to boost immune gene expression.

Through parallel transcriptomic and epigenomic analyses on publicly available $CD4^+$ T cell data [9, 42–44], we find that ECs have distinct TE and gene expression profiles compared to uninfected healthy controls (HCs) and treatment-naïve viremic progressors (VPs). Furthermore, we find that ECs are transcriptionally heterogenous and can be divided into four clusters, distinguished by the expression of innate immune genes and TE families. In a subset of ECs, we identify an increase in the chromatin accessibility of specific TE loci which correlates with higher expression of nearby HIV-1 restriction factors and other immune-related genes in ECs compared to HCs. Finally, we observe that transcript levels of KZNFs in EC $CD4^+$ T cells negatively correlate with that of the TE families they are predicted to transcriptionally repress, suggesting that interindividual genetic and/or epigenetic variation at KZNFs may underlie differential TE regulation in ECs. We propose that derepression of certain TEs boosts their *cis*-regulatory activity on antiviral genes, thereby enhancing the immune response to HIV-1 infection. Overall, our findings support the notion that the genomic activation of specific TE loci may contribute to HIV-1 resistance.

Results

The EC (retro)transcriptome is distinct from healthy controls, PLWH-on-ART, and viremic progressors

In order to identify unique transcriptomic and retrotranscriptomic (defined as TE transcripts) features of ECs, we first analyzed RNA-Seq data from a study investigating the role of the HIV-1 coreceptor CCR5 in the EC phenotype, which was generated from activated $CD4^+$ T cells of ECs ($n=4$) and HCs ($n=5$) [44]. The differential expression of immune-related genes between ECs and HCs has been documented [43, 45], but the EC retrotranscriptome has yet to be profiled. We analyzed this dataset to identify the genes and TE families that were differentially expressed between the two populations

(Additional File 1: Table S1). The differential expression of genes and TEs was similar in magnitude, with maximum $\log_2$ fold changes of significance around ± 10 (Fig. 1A).

Among the differentially expressed genes, we noted the lower expression of factors known to facilitate HIV-1 entry in ECs, including *CCR2/3/5/6* and *CXCR4* receptors (Fig. 1B), consistent with the study from which the RNA-Seq data originated [44]. Furthermore, multiple HLA genes were more highly expressed in ECs, including *HLA-B*, *-C*, and *-F* (Fig. 1B). We and others have shown differences in antiviral restriction factor expression between PLWH and HCs based on their HLA genotypes [15, 46, 47], and certain *HLA-B* alleles have been strongly associated with elite control in previous GWAS studies [16, 21]. We also observed elevated transcript levels for several restriction factor-encoding genes (*IFIT* and *IFITM*s) and pro-inflammatory factors (*STAT1*, *IRF8*) known to be activated in response to viral infections, including HIV-1 [48–50], which may also contribute to the EC phenotype. For example, the immune transcription factor *IRF8* is highly expressed in ECs relative to HCs (Fig. 1B). Low levels of *IRF8* have been previously associated with adverse neurological outcomes in PLWH-on-ART [51]. Elevated expression of *IRF8* may be yet another protective aspect of elite control. Thus, the distinct immune gene expression profile we observe in ECs is largely consistent with previous studies and brings further support to the idea that enhanced innate immunity contributes to HIV-1 restriction in ECs [52–55].

The most differentially expressed TE families were dominated by primate-specific TEs—specifically SVA and LTR/ERV elements—known to harbor complex *cis*-regulatory sequences (Fig. 1C). For example, numerous previous studies have found that SVA and HERVK elements are frequently co-opted as *cis*-regulatory elements in human gene regulatory networks, including those involved in early embryonic development [56, 57] and cell type identity [58–61]. Similarly, many LTR10 elements carry p53-bound enhancers activated in cancer [62–64]. Finally, MER41, LTR26, and MER57—certain subfamilies of which have higher expression in ECs compared to HCs (Fig. 1C)—are known to be enriched for STAT1 binding sites and frequently behave as interferon-inducible enhancers associated with innate immunity genes [34]. We also note that TE families with lower expression in ECs are enriched for LTR/ERV families as well (Fig. 1C), suggesting broad differences in the regulatory landscape of $CD4^+$ T cells in ECs compared to HCs.

Knowing that ECs are distinct from HCs, we aimed to validate that these differences were not solely attributed to the presence of HIV-1 in one group but not the other. Thus, we next compared ECs with PLWH-on-ART. We compared TE family expression in $CD4^+$ T cell subsets of ECs ($n = 12$) and PLWH-on-ART ($n = 3$) using RNA-Seq data from a previous study that investigated EC-specific proviral integration patterns (Additional File 1: Table S2) [9]. Consistent with our EC vs. HC comparisons, MER41 and SVA families were more highly expressed in ECs than in PLWH-on-ART (Additional File 2: Figure S1), in addition to ERV families LTR12, LTR13, and THE1B. However, it is possible that some of these differences are influenced by the presence of antiretroviral treatment in PLWH-on-ART, rather than being solely driven by EC versus non-controller distinctions.

To ensure that these findings were not confounded by the presence of ART, we performed a more extensive comparison between the (retro)transcriptomes of untreated ECs ($n = 19$) and treatment-naïve VPs ($n = 8$). These data came from a study identifying

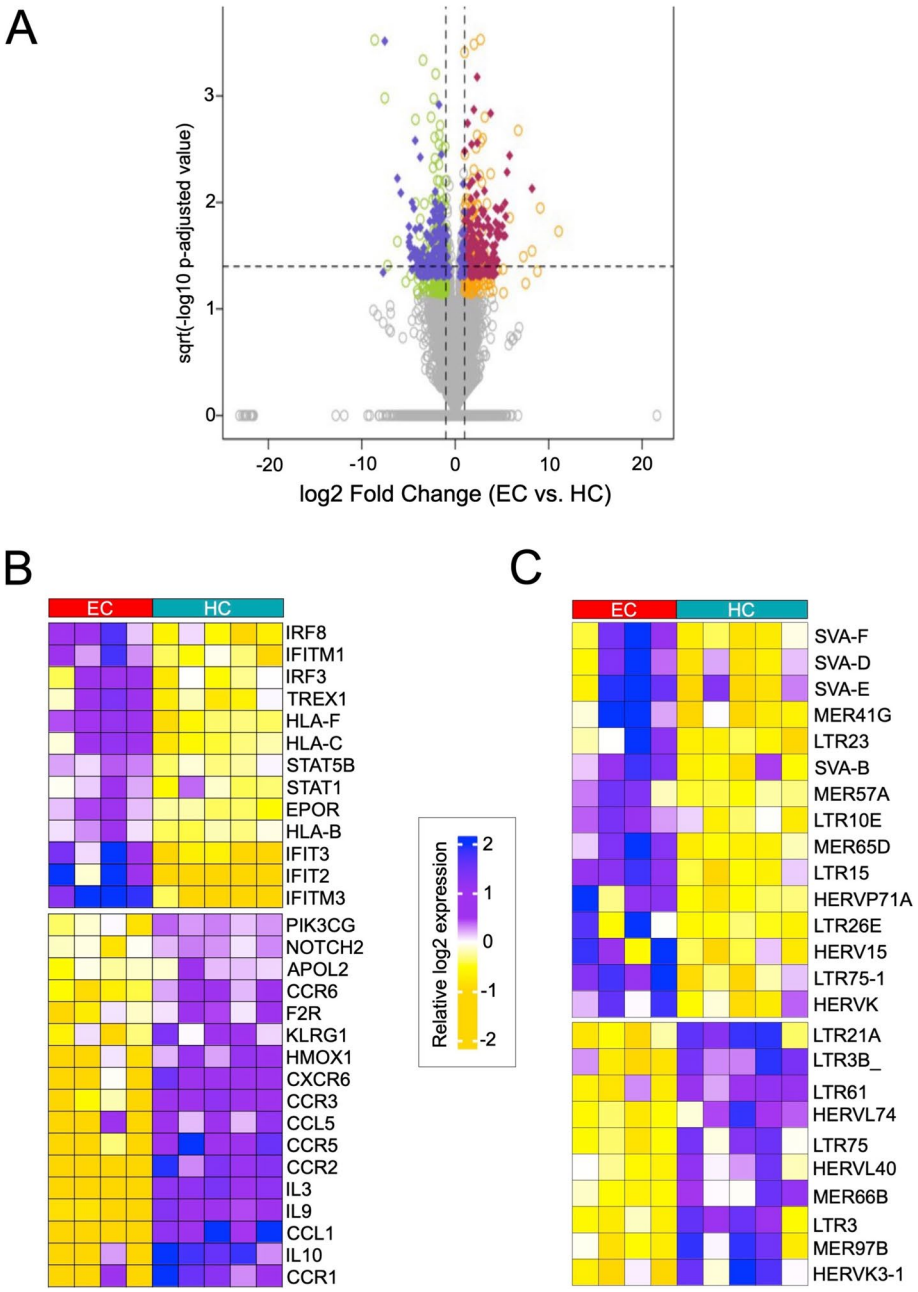

Fig. 1 Differential (retro)transcriptomic profiles in ECs vs. HCs. **A** Volcano plot illustrating the differentially expressed TEs and genes between ECs ($n=4$) and HCs ($n=5$) [44]. Coloration is based on increased or decreased expression of genes (orange and green, respectively) and TEs (red and purple, respectively). Total detected genes and TE loci are plotted by log2-transformed fold change of DESeq2-normalized counts on the x-axis (EC vs. HC). Statistical significance is given in the form of sqrt -log10 adjusted *p*-value on the y-axis, calculated by Wilcoxon rank sum test with Bonferroni correction. **B** Heatmap displaying the expression of the top differentially expressed genes in $CD4^+$ T cells of ECs ($n=4$; red bar) vs. HCs ($n=5$; blue bar) [44]. Relative expression levels are representative of row-wise scaled, log2-transformed expression in transcripts per million (TPM). Heatmap coloration is based on the *z*-score distribution from low (gold) to high (purple) expression. **C** Heatmap displaying the expression of the top differentially expressed TE families in $CD4^+$ T cells of EC ($n=4$; red bar) vs. HCs ($n=5$; blue bar) [44]. Relative expression levels are representative of row-wise scaled, log2-transformed expression in TPM. Heatmap coloration is based on the *z*-score distribution from low (gold) to high (purple) expression

differentially expressed genes in the donors' peripheral blood mononuclear cells (PBMCs) [43]. After identifying the most variable genes and TEs (Additional File 1: Table S3), a principal component analysis (PCA) was conducted to evaluate the level of transcriptome separation between ECs and VPs (Fig. 2A). Approximately 1900 genes and 100 TE families segregated the EC and VP samples on the first three principal

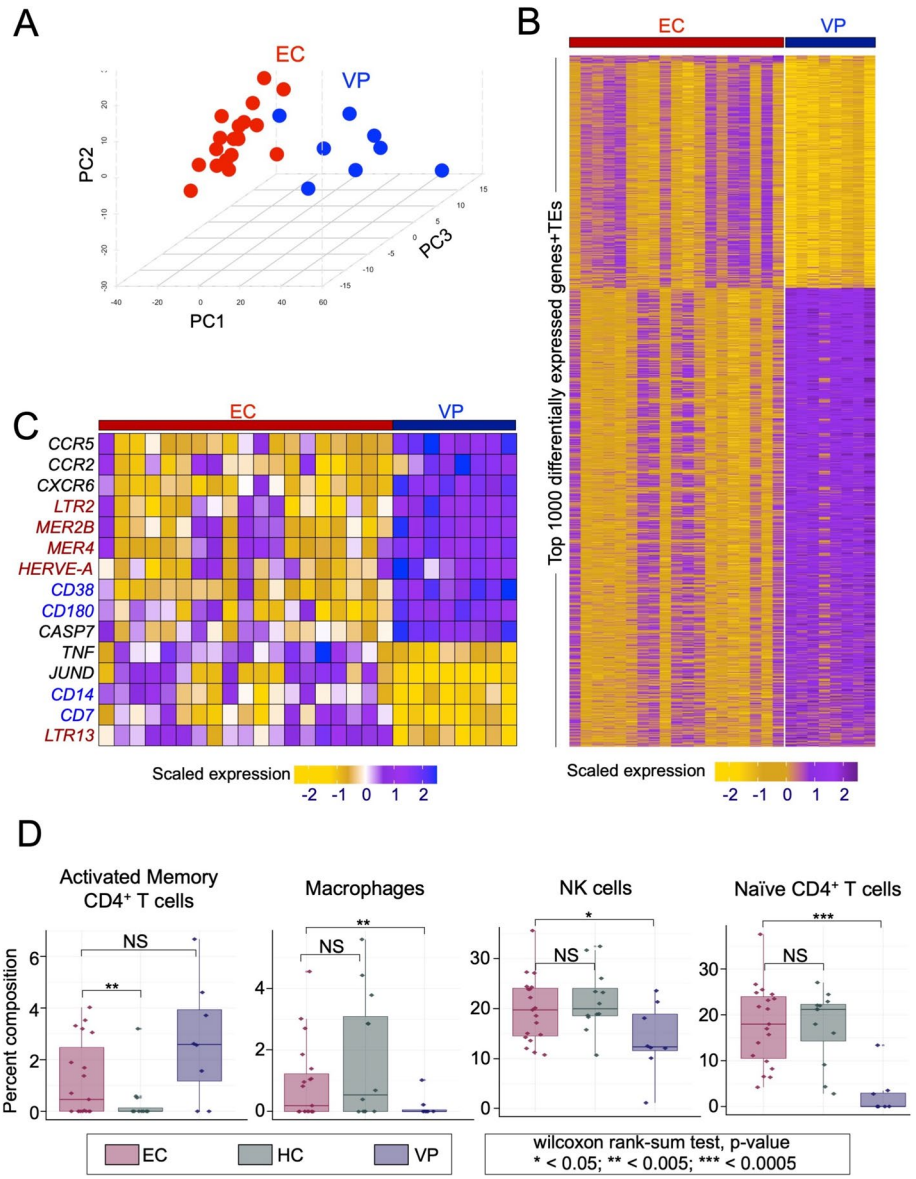

Fig. 2 Differential (retro)transcriptomic and immune cell profiles in ECs versus VPs. **A** PCA triplot from PBMCs of ECs (red) and VPs (blue) [43], based on the most variably expressed genes and TE families ($n = 2000$). **B** Heatmap displaying the row-wise scaled, log2-transformed expression in TPM of genes and TEs distinguishing EC and VP RNA-Seq samples [43]. Heatmap coloration is based on the z-score distribution from low (gold) to high (purple) expression. Every row denotes a gene or TE element. **C** Heatmap displaying the relative expression of selected genes and TEs between ECs and VPs [43]. Columns are EC and VP samples; rows display immune-related genes (black), leukocyte surface markers (blue), and TEs (red). **D** Box plots for percent composition of leukocyte populations of interest between ECs, HCs, and VPs, identified via deconvolution analysis of PBMC RNA-Seq data [43]. Statistical significance was determined by Wilcoxon rank-sum tests

components, indicating that ECs have distinct transcriptomic and retrotranscriptomic profiles from VPs (Fig. 2B). In fact, when conducting a separate PCA that included the HC samples from the same study, the ECs appear to be even more distinct from the VPs than from the HCs (Additional File 2: Figure S2A&B).

Immune signatures in our EC versus VP comparison were generally consistent with previous comparisons of ECs to other cohorts, including the notably lower expression of *CCR5* in ECs (Fig. 2C), even though no EC donor in this cohort harbored the *CCR5Δ32* deletion [43, 44]. Additional notable differences between ECs and VPs included higher expression of *TNF* in ECs, a cytokine with a crucial role in the immune response (Fig. 2C). Furthermore, we observed differential expression of several leukocyte surface markers: *CD14* (macrophage marker) and *CD7* (effector CD8 + T cell marker) both had higher expression in ECs, while expression of activated immune cell markers *CD38* and *CD180* was higher in VPs (Fig. 2C). For the retrotranscriptome, we noted higher expression of the LTR13 family in ECs compared to VPs, consistent with our EC versus PLWH-on-ART analysis and of note because LTR13 is a family known to be enriched for STAT1 binding [34]. Taken together, these analyses suggest that the ECs have a (retro)transcriptomic profile distinct from HCs, VPs, and PLWH-on-ART characterized by elevated levels of specific immune-responsive genes and TEs.

Having focused on identifying bulk transcriptomic differences thus far, we next aimed to investigate whether ECs, VPs, and HCs have distinct immune cell compositions. With the abovementioned PBMC RNA-Seq dataset [43], we conducted deconvolution analyses through which we were able to distinguish transcriptomic signatures of 19 immune cell types (see Methods for more details). Using these data, we could infer that ECs, VPs, and HCs had distinct immune cell compositions (Additional File 2: Figure S3A). Most significantly, ECs had a higher proportion of macrophages, naïve $CD4^+$ T cells, and NK cells compared to VPs, and a higher proportion of activated memory $CD4^+$ T cells compared to HCs (Fig. 2D). Thus, the transcriptomic differences observed between ECs and other populations may be partially driven by variation in immune cell type composition. We were also interested in assessing whether the TE expression profile of ECs was more similar to that of a specific $CD4^+$ T cell state or subtype. To explore this, we compared TE expression of ECs with that of stimulated and unstimulated T cell subsets from healthy donors (Additional File 2: Figure S3B) [65]. Unsupervised clustering of these samples shows that the TE expression pattern of ECs is most similar to that of Th2 progenitor cells, which are associated with HIV-1-specific adaptive immune responses [66]. Still, we observed that, for the majority of families, TE expression was higher on average in all EC $CD4^+$ T cell subsets than in $CD4^+$ T cell subsets from HCs, regardless of stimulation (Additional File 2: Figure S3B). While a subset of TE families exhibited an expression pattern in ECs similar to that of activated $CD4^+$ T cells of HCs (e.g., high expression of L1s and THE1B), multiple TE families appear to have increased transcription uniquely in ECs (e.g., LTR12C and LTR7). Together, these findings underscore the unique immune cell composition, transcriptome, and retrotranscriptome of ECs.

$CD4^+$ T cells from ECs can be split into four functionally distinct clusters

In our initial differential expression analyses, we observed more consistent expression patterns among HCs and VPs than among ECs (Figs. 1B, C, 2B, C, and Additional File

2: Figure S2B). To further explore this apparent transcriptomic heterogeneity, we analyzed a larger group of ECs using two previously published RNA-Seq datasets of $CD4^+$ T cell subtypes [9, 42] including naïve, central memory (CM), effector memory (EM), transitional memory (TM), and total $CD4^+$ T cells. The first study used RNA-Seq to characterize the proviral reservoir of 12 ECs compared to 3 PLWH-on-ART [9]. The second study investigated mechanisms for viral persistence in 15 EC individuals [42]. We integrated the RNA-Seq data from both studies (a cumulative 128 EC samples – Additional File 1: Table S4) to increase our power to classify ECs into subgroups. We then performed unbiased clustering on the scaled, normalized gene and TE family expression data by feeding the first five principal components to the graph-based k-nearest neighbors (KNN) algorithm (see Methods for more details). This analysis revealed four distinct clusters (Fig. 3A & Additional File 2: Figure S4). Importantly, the samples did not cluster by participant cohort or study of origin, thereby ruling out batch effects as the drivers of the clustering (Additional File 2: Figure S5A&B). To examine whether the clustering could be explained by genetic ancestry, we next visualized the samples by ancestry as inferred by variant comparisons with HapMap [67]. All clusters were highly heterogeneous with respect to study participant ancestry (Additional File 2: Figure S5C), suggesting that this is unlikely to be a major driver of clustering. Finally, we visualized the samples by cellular subtype, as identified in the original studies, to assess whether the clustering could be explained by $CD4^+$ T cell subtype composition (Additional File 2: Figure S5D). Clusters 1 and 2 were essentially indistinguishable in cell type composition, whereas Clusters 3 and 4 showed an overrepresentation of TM/EM and naïve/CM cell types, respectively (Fig. 3B). Thus, cell subtype composition could only partially explain the clustering.

To functionally characterize the four EC clusters, we performed pathway and gene ontology analyses on each cluster using gene lists that defined the individual clusters, extracted by differential expression analysis of the cluster's samples versus the samples of the remaining three clusters (Additional File 1: Table S5). Using both the KEGG pathway database [68] and the biological process aspect from the Gene Ontology Consortium [69], we were able to identify distinct pathways and ontological enrichments for each of the four clusters (Fig. 3C & Additional File 1: Table S6). Cluster 1 was enriched for genes related to cell turnover, autophagy, and cell cycle regulation. Cluster 2 was enriched for genes related to RNA processing and splicing. Clusters 3 and 4 had some overlap, as both contained markers of immune function, but with distinctive enrichments. Cluster 3 was enriched for genes regulating T cell activation and proliferation, whereas Cluster 4 was enriched for genes involved in neutrophil–T cell interactions and cytokine production. The enrichment of TM and EM $CD4^+$ T cells in Cluster 3 is consistent with the overrepresentation of pathways related to T cell activation and proliferation. Similarly, the enrichment of naïve and CM $CD4^+$ T cells in Cluster 4 aligns with the overrepresentation of pathways related to neutrophil-T cell interactions and cytokine production, reflecting their involvement in pathogen surveillance and immune cell communication. Thus, we conclude that EC transcriptomes can be classified into four distinct clusters defined by differentially expressed genes enriched for specific biological functions.

With a better understanding of the functional delineation of the four EC clusters, we next wanted to examine whether cluster-specific genes could reflect

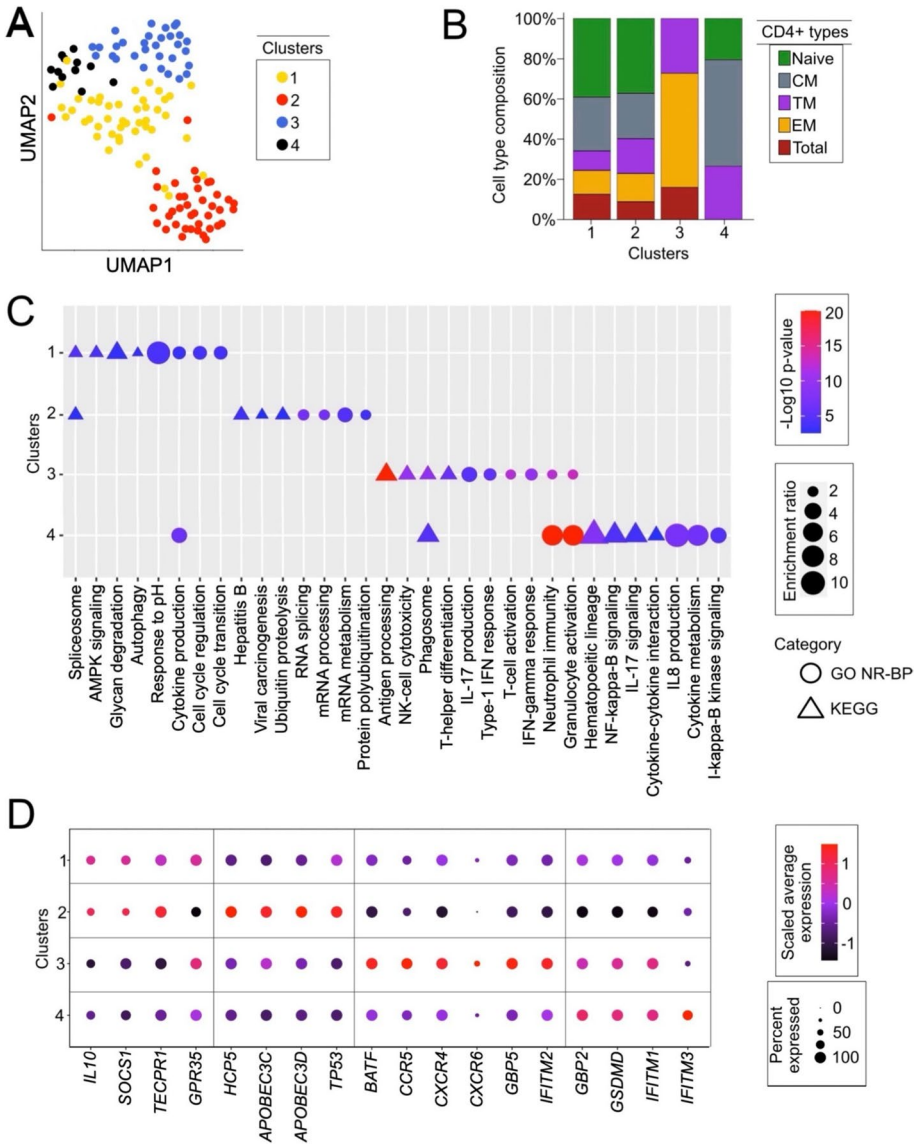

Fig. 3 EC CD4$^+$ T cells can be grouped into four distinct clusters. **A** UMAP plot of the four clusters of EC CD4$^+$ T cell subtypes ($N = 128$) using KNN graph construction on gene and TE family expression from bulk RNA-Seq data [9, 42], based on the Euclidean distance in PCA space (see Methods). Clusters are numbered by descending sample count. Every point is a CD4$^+$ T cell RNA-Seq sample, colored by cluster assignment. **B** Stacked barplot displaying the composition of CD4$^+$ T cell subtypes (naïve, CM, TM, EM, total) in each of the four clusters. **C** Gene ontology biological process (GO NR-BP; ●) and KEGG pathway (KEGG; ▲) delineation of the four EC clusters using WebGestalt [75]. Data derived from differential expression analysis of the EC clusters, using the significant DEGs (p-value < 0.05) as each cluster's respective gene list. For each of the four EC clusters, the highest ranked GO terms and KEGG pathways by adjusted p-value are shown. "Enrichment ratio" refers to the number of observed genes divided by the number of expected genes from each GO or KEGG category in the cluster's gene list. **D** Dot plot illustrating the scaled expression of selected genes related to HIV-1 replication in the four EC clusters. Coloration represents the log2-transformed expression of each gene, scaled to the transcriptome and averaged across each cluster's samples, from lower (blue) to higher (red) expression. The size of the dots is proportional to the percent of samples expressing the given gene in a given cluster

different mechanisms of HIV-1 resistance. To accomplish this, we focused on differentially expressed genes that encode factors known to directly modulate HIV-1 infection. Cluster 2, which is the most distinct in our Uniform Manifold Approximation and Projection (UMAP) (Fig. 3A), accordingly had the most distinct gene expression profile. This cluster was characterized by relatively low expression of HIV-1 coreceptors *CCR5* and *CXCR4*, and high expression of restriction factors like *HCP5* and *APOBEC*s (Fig. 3D) [37, 70, 71]. Additionally, Cluster 2 showed high expression of *TP53*, which is involved in multiple mechanisms opposing HIV-1 replication [72–74]. Within the other clusters, we found that Clusters 3 and 4 expressed at least one restriction factor (*IFITM1*, *IFITM2*, *IFITM3*, *GBP2*, *GBP5*, etc.) at a significantly higher level than Clusters 1 and 2 (Fig. 3D). In sum, each EC cluster expresses a unique "cocktail" of proviral and antiviral factors.

To exclude the possibility that these gene expression signatures in ECs are associated with viremia, we quantified HIV-1 transcript levels in deconvoluted $CD4^+$ T cell RNA-Seq samples from ECs and PLWH-on-ART for comparison. In the original studies, all samples were reported to have undetected viremia by blood tests [9, 42–44]. Consistent with this, we found that the vast majority of the EC and ART samples taken from PBMCs exhibited very low HIV-1 transcript levels, with TPM values generally below 1. However, in samples originating from the lymph nodes of EC individuals ($n = 22$) [42], we detected HIV-1 expression in some subsets (Additional File 2: Figure S6A&B). In agreement with the corresponding study [42], we found elevated HIV-1 transcript levels in germinal center and non-germinal center T follicular helper cells (GC Tfh & nGC Tfh, not included in our clustering analyses)—and to a lesser extent in T effector memory (EM) cells (average TPM < 2 – Additional File 2: Figure S6A). However, we observed no significant association between HIV-1 transcript levels and the expression of interferon-stimulated genes (ISGs) across all EC samples (Additional File 2: Figure S6C-D). These analyses confirm that the unique immune gene expression profile observed in ECs is not associated with viremia.

Restriction factor expression in ECs may be driven by more accessible retroelements

Previous studies have implicated TEs as direct regulators of interferon-stimulated gene expression upon viral infection, including antiviral factors [34, 36]. Having determined that both TEs and HIV-1 restriction factors are differentially expressed in ECs compared to HCs, VPs, and PLWH-on-ART, we next aimed to determine whether changes in TE expression and *cis*-regulatory activity could be correlated to changes in innate immune gene expression.

We used paired ATAC-Seq—which measures chromatin accessibility—and RNA-Seq datasets from the $CD4^+$ T cells of ECs ($n = 4$) and HCs ($n = 4$) [44] to create a list of TE-gene pairs where the TE locus and gene show significantly increased accessibility and expression, respectively, in ECs compared to HCs (Additional File 1: Table S7, see Methods for details). These loci and genes were paired based on proximity, with a maximum distance of 10 kb between the TE locus and the gene's transcription start site, in order to increase the likelihood of a direct *cis*-regulatory influence of the TE over the nearby gene. Subsequent gene set enrichment analysis revealed that these genes were predominantly involved in cellular activation, cytokine production, and immune response regulation (Fig. 4A). The enrichment for differential accessibility of TE loci near genes

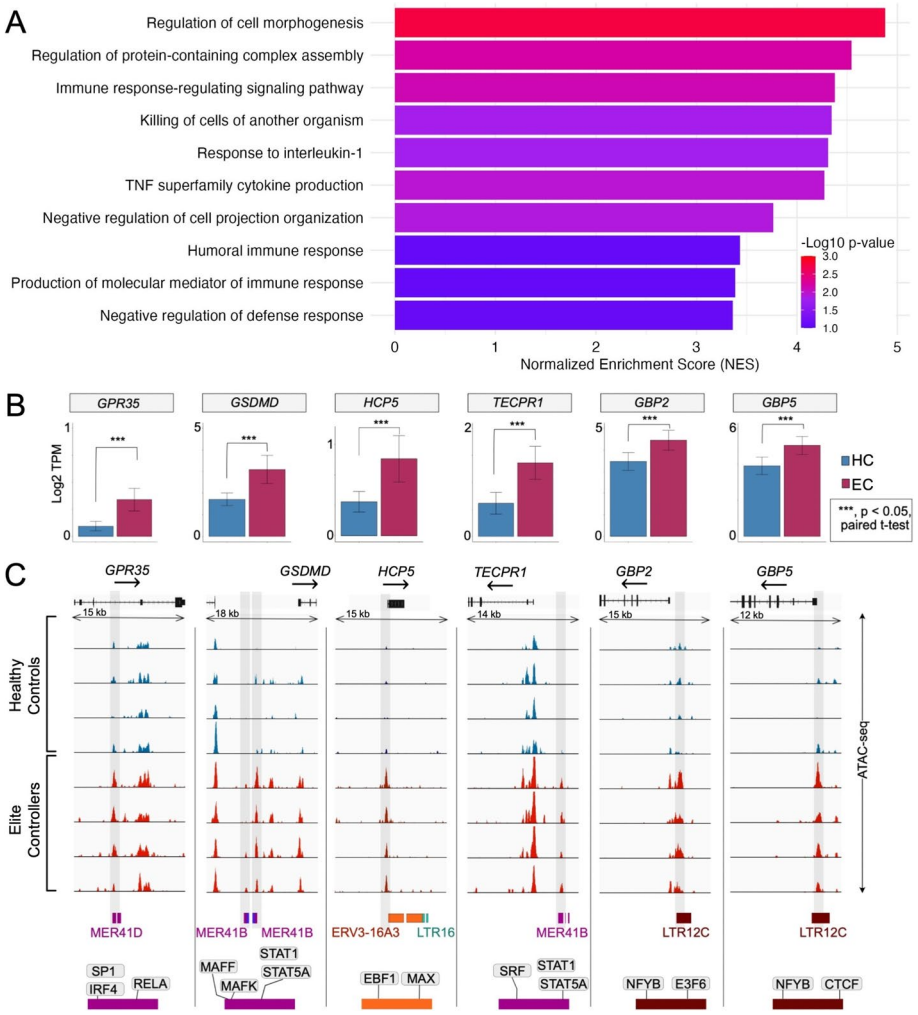

Fig. 4 Induction of innate immune gene expression by proximal LTRs. **A** Bar plot displaying the top ten gene ontology terms enriched within the list of differentially accessible and expressed TE-gene pairs in ECs compared to HCs [44]. Gene sets are listed in descending order by their normalized enrichment score (NES) and colored by -log10-transformed *p*-value ranging from blue (low) to red (high). **B** Barplots showing the expression in log2-transformed TPM of selected differentially expressed innate immune genes in HC ($n=5$) and EC ($n=4$) $CD4^+$ T cells [44]. *P*-value is calculated by paired Student's *t* test. **C** Integrative genome visualization (IGV) of normalized ATAC-Seq signal around the selected DEGs in Fig. 4B between HC ($n=4$) and EC ($n=4$) $CD4^+$ T cells [44]. ATAC-Seq peaks of interest are shaded in light gray. The proximal TE integrants are shown below the IGV graph, under which the encoded transcription factor binding over the corresponding TE integrant(s) is also shown

involved in these pathways suggests that the distinct TE landscape observed in ECs may contribute to a unique immune regulome in these individuals.

In Fig. 4B, C, we have highlighted six of the TE-gene pairs from Table S7 based on the gene's function in HIV-1 restriction and the TE family's known contribution to immune gene regulation. *GPR35* [76], *GSDMD* [77], and *TECPR1* [78] are three antiviral genes more highly expressed in ECs compared to HCs (Fig. 4B); all three are flanked by MER41 elements bound by STAT1 and STAT5 and marked by more accessible chromatin in ECs (Fig. 4C). Similarly, *GBP2* and *GBP5* are two guanylate-binding proteins known to restrict HIV-1 entry, previously shown to be transcribed from LTR12C elements upon

HIV-1 infection of primary CD4$^+$ T cells [36]. We found that transcript levels for both *GBP2* and *GBP5* are moderately but significantly elevated in ECs compared to HCs and observed that their respective LTR12C-derived promoters are marked by more accessible chromatin in ECs (Fig. 4B, C). The final case highlighted here is *HCP5*, a lncRNA gene derived from an ERV3-16 element, which has been implicated in the innate immune response to several pathogens, including HIV-1 [18, 79]. We found that *HCP5* expression is higher in ECs compared to HCs and correlates with the accessibility of its ERV3-16-derived promoter. These examples suggest that the upregulation of some HIV-1 restriction factors and other antiviral genes in ECs may be driven by increased *cis*-regulatory activity of the nearby TEs controlling their expression.

EC clusters are characterized by increased expression of specific TE families

Building upon the idea that the cis-regulatory activity of certain TE loci may enhance the expression of antiviral genes in ECs (Fig. 4) and the knowledge that many of these antiviral factors are differentially expressed across the four EC clusters identified (Fig. 3), we hypothesized that the expression of TE families varies across the clusters. To examine this, we first compared TE transcript levels across the RNA-Seq datasets used to define the four EC clusters [9, 42]. For the samples in each cluster, we calculated TE family expression by averaging the transcript levels of all loci from a given family and comparing these average values between clusters. We found that each EC cluster was characterized by a distinct TE expression profile. Figure 5A highlights a subset of differentially expressed TE families of particular interest. Cluster 1 was characterized by high expression of the youngest LINE1 families in the human genome (L1HS, L1PA2, L1PA3) [80]. Cluster 2 was marked by high levels of LTR7/HERVH and MER41B, two families previously implicated in the regulation of antiviral factors [34, 38]. Cluster 3 was marked by high LTR12C expression, consistent with the high level of *GBP2* and *GBP5* observed in some EC individuals (Fig. 4B) including those falling within Cluster 3 (Fig. 3D). Finally, Cluster 4 was characterized by a high level of HERVL40 expression, a TE family previously observed to be downregulated upon HIV-1 latency reversal [81]. Thus, each cluster displayed a unique profile of gene and TE family expression.

To examine whether the differential expression of TEs across the four EC clusters (Fig. 3A) correlated with the chromatin accessibility of their canonical promoters, we analyzed CD4$^+$ T cell ATAC-Seq data produced in parallel with the RNA-Seq data used in the EC clustering, available for a subset of the ECs from one study ($n = 60$) [9]. For the samples in each cluster, we calculated the averaged, normalized ATAC-Seq signal over all loci of the TE family of interest (highlighted in Fig. 5A) and compared the average signal for each cluster. For most TE families, we found that the average chromatin accessibility profile across clusters correlated with their RNA expression profile (Fig. 5B). For example, the promoter region of young L1 elements was more accessible in individuals from Cluster 1 where these elements are most highly expressed. Likewise, LTR7 elements were on average more accessible in individuals from Clusters 2 and 4 in which these elements are also most highly expressed (Fig. 5A,B). Of those highlighted in Fig. 5A, there were four TE families for which we could not correlate family-wide average chromatin accessibility and cluster-specific expression: MER41B, LTR10B1, THE1B, and LTR8A. We found that the higher expression of these families

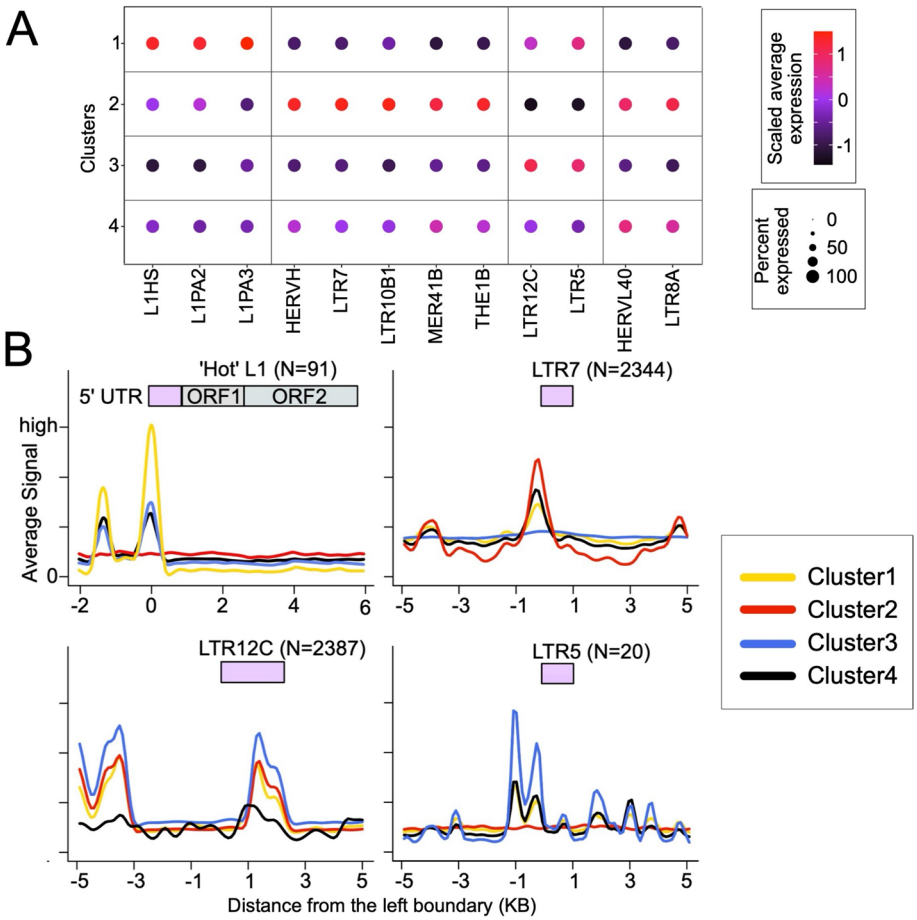

Fig. 5 Characterization of TE expression and accessibility in specific EC clusters. **A** Dot plot depicting the prevalence of TE family expression across the four EC clusters [9, 42]. Coloration represents log2-transformed expression for each TE family, derived from locus-level calculations aggregated to the family level, scaled to the transcriptome, and averaged across the samples within each cluster, ranging from lower (blue) to higher (red) expression. The size of each point corresponds to the percentage of samples within the cluster that express the TE family. **B** Line plots showing the distribution of averaged, normalized ATAC-Seq [9] signal over all loci in the selected TE families for the samples in each cluster. The ATAC-Seq signal counts are calculated and normalized as mappable reads per million per 100 bp bins, in a +/− 5-kb genomic window at the elements' left boundary in the human genome. For the LINE1 elements, only those that are intact and full-length are considered, denoted as "hot" LINE1s for their recent transpositional activity. These were analyzed in a -2-kb to 6-kb genomic window from their left boundary. Lines are color coded by cluster.

in a certain cluster was driven by a small subset of particularly active loci, likely limiting detection of more open chromatin at the family level.

To further explore locus-level expression patterns, we re-clustered the same EC samples ($n = 128$) using only locus-level TE expression. This again resolved four EC clusters (Additional File 2: Figure S7A), which interestingly appeared even more distinct than those identified by gene and TE family expression combined (Fig. 3A). The TE locus-based clusters (TL-Cs) aligned well with the gene and TE family clusters (GT-Cs), with an average 70% overlap in samples between each GT-C and its corresponding TL-C (Additional File 2: Figure S7B), indicating high consistency (Additional File 1: Table S8). The remaining 30% of samples that shifted between clusters did so consistently within individuals, not cohorts, maintaining heterogeneous TL-C compositions

similar to the GT-Cs (Additional File 2: Figures S7C & S5A). An exception to this heterogeneity was TL-C4, comprising 22 samples from GT-C1 that were almost entirely from the $CD4^+$ T cell subsets of only four participants in the Jiang cohort (Additional File 1: Table S8 & Additional File 2: Figure S7C). No other samples from the Jiang cohort shifted to this cluster from other GT-Cs, suggesting that these patterns reflect individual variation rather than cohort bias. Like the GT-Cs, each TL-C included samples from all five $CD4^+$ T cell subsets and was largely heterogeneous (Additional File 2: Figure S7C). Notably, TL-C2 mirrored corresponding GT-C3 in its overrepresentation of EM and TM cells, while TL-C1 uniquely showed an overrepresentation of naïve $CD4^+$ T cells. Beyond sample composition, each TL-C was characterized by a unique pattern of expressed TE loci (Additional File 2: Figure S7D). These signatures were heterogeneous across families, with subsets of variable loci from one TE family marking separate clusters (Additional File 2: Figure S7E), some of which did not reach the threshold of significance in earlier analyses when analyzed at the family level, like SVA-D. Many families maintained their cluster-specific signatures, like THE1B (a marker of GT-C2), for which the majority of variable loci were found in corresponding TL-C1. However, some TE families, like the LINE1s that marked GT-C1, showed more heterogeneous signatures with variable loci marking multiple TL-Cs. These findings underscore the need for future locus-level investigations with high-depth or long-read RNA sequencing to fully capture the complexity of TE expression.

Variable TE expression in ECs may be mediated by KZNFs

What could drive the differential accessibility and expression of TEs that we observe across ECs? We hypothesized that KRAB-containing zinc finger proteins (KZNFs) may be involved since they are known to bind directly to specific TE families through their zinc finger DNA binding domain and recruit the KAP1 corepressor through their KRAB domain, which in turn attracts silencing factors such as histone deacetylases and methyltransferases to nucleate repressive chromatin at the bound TE loci [82]. To test this idea, we performed a pairwise correlation analysis of TE families and KZNFs across the transcriptomes of the EC samples used for our gene and TE family clustering analysis. With over 350 annotated KZNFs in the human genome, the repressive interactions between these proteins and TEs is thought to be highly specific and dynamic [83, 84], and as such we predicted that KZNFs and their target TE families would show a negative correlation in expression across individuals.

Visualizing these correlation analyses by EC cluster, we observed distinct anti-correlative KZNF-TE expression patterns (Fig. 6A, shown in blue), suggesting that each cluster is characterized by a unique combination of anti-correlated KZNF-TE pairs. Focusing on the TE families that stood out in our earlier analyses, we found multiple KZNFs whose expression was anti-correlated to that of the TE families they are known to target (Fig. 6B). For example, expression of *ZNF84*—which is known to bind and repress young LINE1 families [84]—was significantly anti-correlated to the expression of these LINE1 families across EC individuals. Similarly, *ZNF534* was anti-correlated with HERVH/LTR7, *ZNF430* with THE1B, and *ZNF2* with LTR12C. Targeting of the aforementioned TE families by their respective KZNFs was confirmed by re-analysis of public ChIP-exo data (Fig. 6C) [85, 86]. Furthermore, we found that these KZNFs have significantly lower

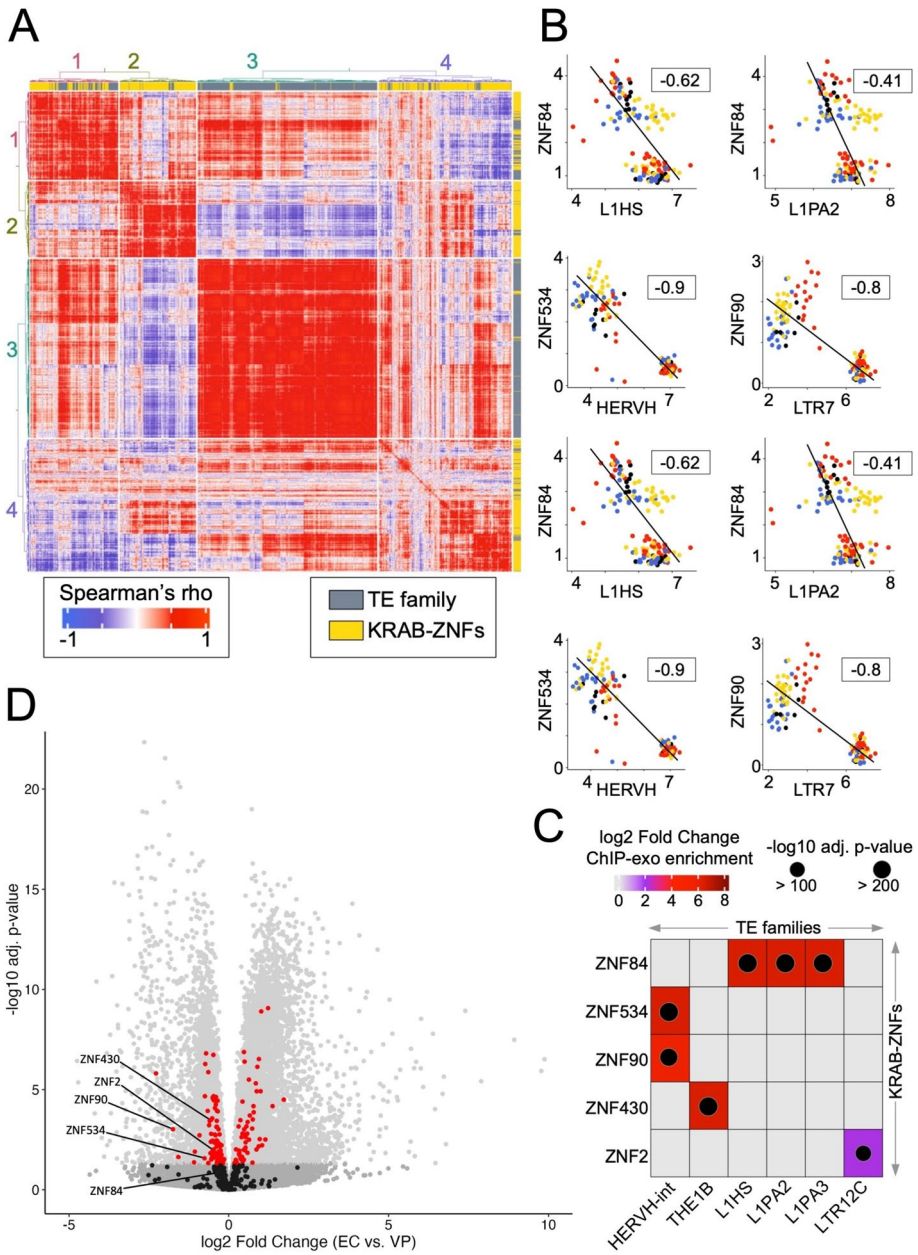

Fig. 6 Dynamic KZNF expression and regulation of TEs in ECs. **A** Heatmap showing the correlation matrix of TE family expression with KZNF expression across the analyzed EC samples [9, 42]. Clustered pairwise correlation matrix generated by weighted gene co-expression network analysis, using Spearman's rank correlation and Euclidean distance. **B** Scatter plot showing scaled, log2-transformed expression in TPM of previously implicated TE families and KZNFs in the analyzed RNA-Seq data of EC $CD4^+$ T cell subtypes [9, 42], cluster classified in Fig. 3A. Linear regression analysis (black line) indicates the correlation between TE families and their targeting KZNF's expression in EC samples (adjusted p-value < 0.01). The rho value is obtained from pairwise-ranked correlation analysis. **C** Heatmap showing the enrichment for binding of a subset of KZNFs to TE families with increased expression in ECs. Coloration is scaled from lower (gray) to higher (red) ChIP-exo enrichment in HEK cells [86]. Point size is proportional to the -log10-transformed adjusted p-value. **D** Volcano plot displaying the differential expression of genes between EC and VP RNA-Seq samples [43], plotted by log2-transformed fold change of DESeq2-normalized counts on x-axis and -log10-transformed adjusted p-values on y-axis. KZNF genes are in red (adjusted p-value < 0.05) or black (adjusted p-value > 0.05). All other genes are in gray

expression in ECs compared to treatment-naïve VPs (Fig. 6D). Thus, differential expression of KZNFs may explain the variable expression and cis-regulatory activity of TEs across individuals.

In addition to KZNF repressors, transcriptional activators may also drive the differential expression of specific TE families across ECs [87]. To investigate this, we focused on transcription factors (TFs) expressed in $CD4^+$ T cells and mined ChIP-Seq data from the ENCODE Consortium [88] to identify TFs with binding enrichment to TE families of interest, selected for their elevated, cluster-specific expression in ECs (highlighted in Figs. 4 and 5, and Additional File 2: Figure S4). We then examined the correlation between TF and target TE expression in the deconvoluted $CD4^+$ T cell samples from ECs used for our clustering analysis (Fig. 3) [9, 42]. We observed several significant positive correlations between TF and TE expression across ECs (Additional File 2: Figure S8). Thus, differential expression of immune-related TFs may also contribute to the variation in TE expression and cis-regulatory activity across ECs, in tandem with the repressive activities of KZNFs.

Discussion

In this study, we explore the hypothesis that TEs, which are known to modulate the innate immune response and to directly regulate restriction factors, could contribute to the elite control of HIV-1. Through analyses of $CD4^+$ T cell RNA-Seq data, we determined that the TE expression profile of ECs is distinct from that of PLWH-on-ART, VPs, and HCs. There was also considerable transcriptomic heterogeneity among the set of EC individuals analyzed here, despite being a relatively small cohort ($n=49$) combining four different studies [9, 42–44]. Unsupervised clustering and principal component analysis of the most variable genes and TEs revealed four EC clusters, each characterized by different gene ontologies and pathway enrichments, differential expression of a subset of HIV-1 pro- and anti-viral factors, and unique TE expression profiles. Further analyses integrating ATAC-Seq data revealed that a number of innate immune genes with increased expression in ECs were flanked by *cis*-regulatory TE loci marked with higher chromatin accessibility in ECs. Thus, changes in the chromatin states of these elements may contribute to the upregulation of these factors in some ECs, which in turn would contribute to their HIV-1 resistance phenotype. To illuminate the mechanisms that may underlie chromatin changes at these TEs, we investigated the expression of KZNFs across the same EC cohort—genes which are known to encode proteins that repress TE expression and *cis*-regulatory activity [27, 89]. We found extensive variation in KZNF transcript levels across $CD4^+$ T cells of EC individuals and striking anti-correlation with the TE families they target, suggesting that interindividual variation in KZNF expression may underlie the observed variation in TE chromatin accessibility. Taken together, these data converge to a preliminary model (Fig. 7) in which decreased expression of some KZNFs leads to elevated basal expression of innate immune genes and restriction factors in EC individuals through an increase in the *cis*-regulatory activity of TEs serving as promoters or enhancers for these genes. This model introduces a new possible mechanism underlying the EC phenotype.

Each step in the model will require experimental work to be validated. First and foremost, it will be important to confirm that the TEs exhibiting increased transcript

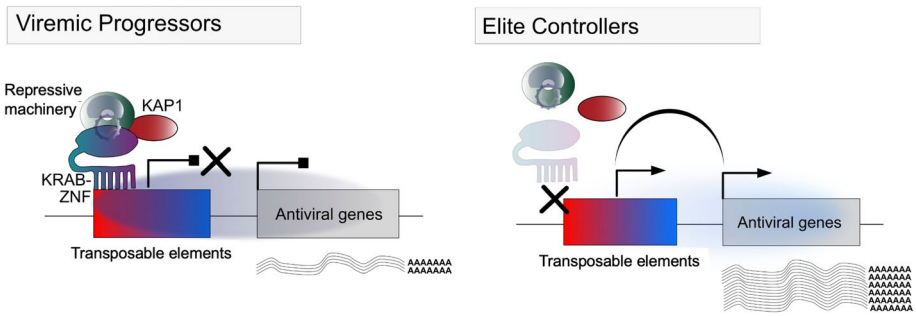

Fig. 7 Proposed model for the interplay of KZNFs and TEs regulating proximal antiviral gene expression in elite controllers of HIV-1

levels and accessibility in ECs are indeed boosting the innate immune response and control of HIV-1 in these individuals. We note that these TEs include HERV families that have been previously implicated in controlling the human innate immune response. For example, we found that the MER41 family has increased expression in a subset of ECs (Fig. 5A), and we identified several MER41 elements marked by higher chromatin accessibility in EC individuals relative to HCs. This correlated with higher transcript levels of adjacent immunity genes such as *GPR35*, *GSDMD*, and *TEPCR1* (Fig. 4B) [76–78]. We previously reported that MER41 elements carry the hallmarks of interferon-inducible enhancers and are frequently bound by the transcription factor STAT1, a key transcriptional activator of the innate immune response. CRISPR-Cas9 editing was used in cell lines to demonstrate that a subset of MER41 elements function as enhancers driving the interferon-inducibility of several innate immune genes [34]. However, the specific MER41 loci we identified here as differentially active in ECs have not been tested experimentally for enhancer activity. Thus, further work is warranted to confirm the regulatory function of these loci under the control of STAT1 or other immune TFs, as well as other TE families identified as targets of immune-related TFs (Additional File 2: Figure S8).

LTR7 and LTR12C are two other HERV families with increased expression in ECs (Fig. 5A), whose members have been previously implicated in the activation of HIV-1 restriction factors. For example, a recent study found that an LTR7 element controls the transcription of *APOBEC3G* and *APOBEC3H* [38], two cytidine deaminases that inhibit HIV-1 replication [37, 90–94]. Similarly, LTR12C elements are known to be activated upon HIV-1 infection, and two of these elements act as promoters for *GBP2* and *GBP5* [36], which encode proteins inhibiting furin-mediated processing of the HIV-1 envelope and other viral glycoproteins [35]. Our work expands upon these findings by showing that the chromatin at these two LTR12C loci is more accessible in ECs than in HCs, which correlates with increased *GBP2* and *GBP5* expression in these individuals (Fig. 4B,C). We observed a similar pattern of increased accessibility and expression in ECs of *HCP5*, a lncRNA gene entirely derived from an ERV3-16 element [79]. *HCP5* is located in the HLA class I region and harbors multiple single-nucleotide variants that are associated with HIV-1 control [95, 96], yet the potential relationship of *HCP5* expression levels to HIV-1 resistance has not been examined [79]. Here, we find increased accessibility of the ERV3-16 promoter and elevated

expression of *HCP5* in ECs compared to HCs (Fig. 4B,C), suggesting that heightened *HCP5* expression may contribute to HIV-1 control. Overall, our results reinforce the concept that TEs are important players in the human antiviral response [30, 97] and uncover specific candidate elements for boosting cellular defenses against HIV-1 in ECs. We acknowledge that these associations are drawn from correlative patterns and manipulative experiments are needed to infer causality between chromatin changes at these TEs and increased expression of nearby immunity genes. High-resolution approaches, such as long-read RNA-Seq, will be crucial to exploring the role of these TE loci and others in driving alternative promoter usage and splicing events in ECs.

To our knowledge, our study is unique for integrating TEs in the transcriptomic analysis of $CD4^+$ T cells of ECs. A major takeaway of our analyses is that there is substantial heterogeneity across the transcriptomes of the 128 EC samples examined, which can be resolved into four distinct clusters, each with unique gene and TE expression profiles (Figs. 3 and 5 and Additional File 2: Figures S4, S5). Importantly, these clusters are driven neither by the study of origin or participant cohort, nor by the genetic ancestry of the individuals (Additional File 2: Figure S5), and they cannot be explained solely by variation in $CD4^+$ T cell subtype composition (Fig. 3B & Additional File 2: Figure S5). The transcriptomic heterogeneity we observe is consistent with the known phenotypic plasticity of ECs [98, 99]. For example, it has been reported that viremic control and T cell population maintenance vary substantially among ECs [11]. Previous studies have identified multiple genetic determinants of viral control in a subset of ECs, including multiple *HLA-B* alleles identified through GWAS [15, 16] and the well-documented *CCR5Δ32* deletion blocking HIV-1 cellular entry [100]. Increased cytotoxicity of $CD8^+$ T cells and natural killer cells have also been implicated in the EC phenotype [101–103], along with superior HIV-1 antibody and cytokine patterns [104, 105]. Finally, genomic profiling of EC reservoir T cells has identified unique chromosomal distribution of retained proviruses in which defective proviruses are preferentially retained in permissive, euchromatin-rich genomic regions while functional proviruses are relegated to heterochromatin-rich regions [9, 12, 23, 24]. Thus, no single factor can universally explain the EC phenotype and one can expect that multiple factors are likely to act in concert within an individual to achieve and sustain HIV-1 resistance. Our study focuses on processes in $CD4^+$ T cells that may operate alongside well-established factors like $CD8^+$ T cell cytotoxicity. Together, these mechanisms illustrate the complex interplay of genetic, immunological, and epigenetic contributions to viral control. For example, a recent study noted that the *CCR5Δ32* allele occurred in a heterozygous state in only one out of five ECs, while no individual in that cohort held the homozygous deletion [44]. Similarly, *CCR5* mRNA downregulation, another proposed mechanism for elite control of HIV-1 [44], was only found to be a feature of a subset of ECs (Cluster 2 in our analysis; Fig. 3D). Thus, our study is useful in defining a minimum of four major "transcriptome types" among ECs that may reflect different combinations of resistance factors, including newly recognized mechanisms involving TEs.

We found that distinct sets of innate immunity genes and restriction factors are upregulated in different EC clusters even in the absence of active viremia, suggesting that elevated basal expression of these factors plays a previously underappreciated role in the EC phenotype. Further studies will be necessary to cement this idea and would

especially benefit from the integration of single-cell omics to dissect TE regulation and clustering in deconvoluted $CD4^+$ T cells of ECs. We also acknowledge that our study is limited by the small number of EC individuals with available omics data, which likely limited our ability to identify significant relationships between transcriptome clustering and available participant metadata (Additional File 2: Figure S5). While the rarity of ECs in the seropositive population makes it challenging to study this phenotype, the transcriptomic heterogeneity revealed by our analyses underscores the need for surveying larger and more diverse EC cohorts.

Another outstanding question is whether the gene and TE signatures revealed by our analysis of ECs exist in the general population independent of HIV-1 infection or if they are driven by the initial infection. While this inquiry is beyond the scope of this study, we have presented evidence of common TE signatures between EC $CD4^+$ T cells and Th2 progenitors from HCs (Additional File 2: Figure S3B) and established that ECs possess a unique CD4 + T cell retrotranscriptome with potential implications for natural HIV-1 control. Future studies designed to assess elite control prediction should explore whether these TE profiles are predictive of an individual's enhanced viral control. In a relevant study [106] investigating the contribution of TEs to variable immune responses in influenza infection, a strong negative correlation was found between the level of TE transcription pre-infection and viral load post-infection, suggesting that high TE expression may be protective. When integrated into a predictive model, the activity of TEs, KZNFs, and KZNF-recruited SETDB1 improved the ability to forecast the outcome of an infection [106]. These findings suggest that TE expression profiles may be predictive of immune potency and further support the need for their inclusion in predictive studies. TE profiles may also be predictive of other rare pathological trajectories, including the inverse of the EC phenotype: HIV-1 rapid progressors [107, 108].

Our study confirms that the epigenetic state of TEs is a key factor modulating their activity as alternative promoters or enhancers for proximal genes. Prior studies have documented this relationship by comparing the epigenetic states (e.g., DNA methylation, chromatin accessibility) of TE loci across healthy tissues [33, 109, 110] or between cancer and non-cancer cells [111–114], but very few studies have explored this relationship across human individuals. The effects of interindividual epigenetic variation at TEs (i.e., metastable epialleles) on adjacent gene expression are well documented in mice and best exemplified by the *viable yellow agouti* allele, which is controlled by a retroviral LTR variably methylated across individuals [115, 116]. But what controls this type of epigenetic variability? There is growing evidence that KZNF proteins are important trans-regulators of these effects through their ability to bind to TEs in a sequence-specific fashion and recruit the co-repressor KAP1, which in turn recruits chromatin modifying factors like histone and DNA methyltransferases to nucleate local heterochromatin at the bound TEs [32, 82]. Indeed, recent studies show that genetic variation within TE sequences and/or variation in KZNF expression levels result in variation in chromatin accessibility and *cis*-regulatory activity at different TE loci, with occasional effects on adjacent gene expression [86, 117, 118].

Motivated by these recent insights, we explored the potential role of KZNFs in mediating the variable expression and chromatin states of TEs among ECs. We found a striking anti-correlation between the expression level of several KZNFs and the TE families they

target (Fig. 6A,B). For instance, transcript levels of *ZNF534* and *ZNF430* were strongly anti-correlated with that of HERVH/LTR7 and THE1B elements, respectively (Fig. 6B), which are the genomic targets of these KZNFs (Fig. 6C) [56, 85, 86, 119, 120]. Notably, the expression of these KZNFs is lower in ECs compared to VPs (Fig. 6D). These observations suggest that interindividual variation in KZNF expression in $CD4^+$ T cells could explain why certain TEs are variably expressed and accessible across ECs. But what are the mechanisms underlying variation in KZNF expression? It is possible that TE-KZNF regulatory loops are involved, in which a copy of the TE family targeted by a KZNF is inserted near and regulates the KZNF gene, thereby introducing a negative feedback loop. This phenomenon has been documented in prior studies of KZNF activity in embryogenesis [56] and in cancer [121]. Thus far, there have been few investigations of TE-KZNF interactions in T cells and of their contribution to immunity [122–124]. Our results point to a direct interplay between TEs and KZNFs in $CD4^+$ T cells, which may have important implications for T cell biology and immune responses. Further work is needed to validate TE-KZNF regulatory interactions in T cells, probe their connection to epigenetic variation at individual TE loci, and explore their repercussions on gene expression variation in $CD4^+$ T cells, with and without HIV-1 infection.

Conclusions

Our study culminates in a testable model (Fig. 7) in which TEs are regulatory agents contributing to HIV-1 elite control by altering the expression of host immune genes in $CD4^+$ T cells, and points to KZNFs as important controllers of these activities. As single-copy genes often exclusively dedicated to regulating a specific TE family, KZNFs represent attractive targets to manipulate TEs for research and therapeutic prospects. Our findings open new avenues for understanding the molecular mechanisms of viral control and developing targeted interventions for HIV-1 treatment.

Methods

Data availability

The datasets supporting the conclusions of this article are publicly available and can be found in NCBI and ENA repositories under the accession numbers GSE122323 [125], PRJNA420459 [126], GSE144334 [127], and GSE83482 [128]. ChIP-Seq data was sourced from the ENCODE consortium: https://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/. The relevant scripts from our analyses are available on our GitHub repository [129].

RNA-Seq data processing

We mined published RNA-Seq datasets of human immune cells isolated from EC, HC, PLWH-on-ART, and VP individuals ranging from PBMCs to deconvoluted $CD4^+$ T cell subsets. Quality control was conducted with *FastQC* on the raw fastq sequencing files, with two nucleotides removed from the ends due to variable quality scores. To facilitate comprehensive expression quantification, we curated a reference transcriptome by combining gene, TEs, and HIV-1 genomic sequences. This was achieved by integrating the locus-level TE classification from *RepeatMasker*, the hg19 gene annotation (GenCode), and the HXB2 reference HIV-1 annotation. For the TEs, we removed simple repeats,

SINE elements, and DNA transposons, including retro-transpositionally active Alu elements. The latter were excluded from our annotation due to the extreme challenges in accurately assessing Alu expression. First, Alus are short and extremely abundant in the human genome with an especially high level of sequence similarity to one another, making short-read mapping unreliable even at the subfamily level. Additionally, standalone Alu transcripts are produced by RNA Pol III, not polyadenylated, and therefore depleted from polyA-enriched RNA-Seq libraries like those analyzed here. Finally, Alus often reside in UTRs or alternative exons, making them co-transcribed in gene transcripts. With short-read sequencing, it is nearly impossible to disentangle Alu expression originating from their own promoters from that of the genes they reside in. As we aim to study the expression of TEs themselves, independent from nearby genes, our pipeline is designed to maximize this. We also removed any other TE loci within gene exons/UTRs. In our final TE annotation, we retained intronic and intergenic LINE and HERV loci, including all solo LTRs. The remaining sequences were appended in fasta format, and all sequences were annotated with their respective gene, TE locus, or HIV subunit and modeled in GTF format. Next, we indexed the concatenated gene and TE genomic sequences using salmon and STAR. Next, we aligned the trimmed sequencing reads against our curated hg19 reference genome using STAR [130]. Salmon [131] was used to quantify the counts and normalize expression (transcripts per million, TPM) for each RNA-Seq sample, enabling simultaneous calculation of gene, TE, and HIV-1 expression using expectation–maximization (EM) algorithms. Data integration, logarithmic normalization, and scaling of the obtained count matrices were performed using custom R scripts (see GitHub for details). DESeq2 [132] was used to perform differential expression analysis between populations. All statistics and visualization of RNA-Seq data were performed using R.

Clustering of EC RNA-Seq samples

With a subset of EC samples for which deconvoluted $CD4^+$ T cell RNA-Seq data was available, we clustered the scaled data by feeding the first five principal components to the graph-based KNN method (see GitHub for details). We calculated differential expression and tested their significance level using the Kruskal–Wallis test by comparing the cluster of interest with the remaining three clusters. The Benjamini–Hochberg method further adjusted the obtained *p*-values to determine the adjusted false discovery rate (FDR).

ATAC-Seq data processing

The ATAC-Seq reads were aligned to the hg19 reference genome by Bowtie 2 (version 2.2.2) [133] under the parameters *–very-sensitive-local*. All unmapped reads, non-uniquely mapped reads, and PCR duplicates were removed. MACS2 [134] was used for peak calling. For downstream analysis, we normalized the read counts by computing the numbers of reads per kilobase of bin per million of reads sequenced (RPKM). RPKM values were then averaged for each bin across replicates. To minimize batch and cell type variation, the RPKM values were further normalized by *Z*-score transformation. To visualize the ATAC-Seq signal in the UCSC genome browser [135], we extended each read by 100 bp and counted the coverage for each base. To compare ATAC-Seq signal over

specific TE families between EC clusters, the genome-wide signal was concatenated per cluster and then normalized as mappable reads per million per 100-bp bins. We then visualized each family of interest within a +/− 5 kb genomic window at the elements' left boundaries according to the hg19 RepeatMasker annotation.

Multiomic annotation of putative cis-regulatory TE-gene pairs

To identify TEs with putative *cis*-regulatory potential in ECs, we first used intersectBed from bedtools [136] with default parameters to calculate the overlap between the ATAC-Seq peaks identified in ECs and HCs with the locations of annotated repeats in the hg19 RepeatMasker annotation. The summit output files provided by MACS2 were used to establish a 50% overlapping threshold between each ATAC-Seq peak and TE locus. Before the *cis*-regulatory element annotation, we calculated the enrichment of ATAC-Seq peaks within each TE family for each population (EC and HC). As repeats of different classes vary greatly in numbers, a random set of ATAC-Seq peaks was used as a control, generated by selecting random regions in the genome size-matched to each individual TE-associated ATAC-Seq peak. The number of observed peaks that overlap with retroelements was compared to the number of random peaks exhibiting the same overlap, and a log2-transformed ratio value was generated as the "observed/expected" enrichment. This TE family-wide significance score was calculated using the R package enrichR [137], while employing in-house codes (see GitHub for details). Finally, we focused on TE ATAC-Seq peaks within 10 kb of annotated transcription start sites (TSS) in hg19. Among these, the peaks which were differentially accessible in ECs compared to HCs (ATAC-Seq) and for which the corresponding gene was differentially expressed in ECs compared to HCs (RNA-Seq), were annotated as putative EC-specific *cis*-regulatory TE-gene pairs.

ChIP-Seq data processing

ChIP-Seq datasets for immune-related TFs of interest were retrieved from the ENCODE database. The fastq reads were mapped against the hg19 reference genome using bowtie2 with the –very-sensitive-local parameter for optimal sensitivity. All unmapped reads, PCR duplicates, and reads with a mapping quality score below 10 were excluded from further analysis. MACS2 was used for peak calling using the parameters –nomodel, -q 0.01, and -B. Peaks overlapping with blacklisted regions were removed to reduce false positives (annotation here). The resulting peak files were merged using mergeBed to generate a set of unique peaks. TE family-wide enrichment scores were calculated as described above (see GitHub for details). The significance of ChIP-Seq enrichment was determined by comparing the observed vs. expected ratios across 1000 permutations, followed by Bonferroni correction to account for multiple testing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03484-y.

Additional file 1: Tables S1-S8. Table S1. Differential expression' analysis of genes and TE families, EC vs. HC $CD4^+$ T cells. Table S2. Differential expression analysis of TE families, EC vs. PLWH-on-ART $CD4^+$ T cells. Table S3. Differential expression analysis of genes and TE families, EC vs. VP PBMCs. Table S4. EC $CD4^+$ T cell cluster sample metadata from clustering by gene and TE family expression. Table S5. Most variable genes and TE families in each of the four EC $CD4^+$T cell clusters. Table S6. Enriched gene ontologies and KEGG pathways for each EC cluster. Table S7. TE-gene

pairs with a predicted cis-regulatory relationship in ECs. Table S8. EC $CD4^+$ T cell cluster sample metadata from clustering by TE locus expression.

Additional file 2: Figures S1-S8 with legends. Figure S1. Unique retrotranscriptome of ECs vs PLWH-on-ART. Figure S2. Heterogeneity of EC (retro)transcriptome. Figure S3. Immune cell profiles of ECs, VPs, and HCs. Figure S4. Broad (retro) transcriptomic profile of the EC clusters. Figure S5. Characterization of the EC clusters. Figure S6. HIV-1 transcription in $CD4^+$ T cells of ECs and PLWH-on-ART and its correlation to ISG expression. Figure S7. Clustering of EC $CD4^+$ T cell samples by locus-level TE expression. Figure S8. Immune-related TF dynamics in ECs.

Additional file 3. Review history.

Acknowledgements
We would like to thank Jez Marston, Melanie Ott, Susana Valente, Lishomwa Ndhlovu, Warner Greene, Ulrike Lange, Zichong Li, Miguel de Mulder Rougvie, Nicholas Dopkins, and members of the Nixon and Feschotte labs for their helpful notes and discussions. We also extend our sincere appreciation to Eli A. Boritz, Daniel C. Douek, Chenyang Jiang, and Xu G. Yu—whose sequencing data was utilized in our analyses—for their open communication throughout this study.

Peer review information
Juergen Schmitz and Tim Sands were the primary editors of this article at Genome Biology and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history
This article was first peer reviewed at Review Commons and reviewer reports and authors' response are available online [138–141]. The full review history is also available as Additional file 3.

Authors' contributions
M.S., S.M.L., L.P.I., M.L.B., D.F.N., and C.F. designed the study. M.S., S.M.L., D.F.N., and C.F. wrote the manuscript. M.S., S.M.L., L.P.I., and M.L.B. performed the data analysis. All authors edited and approved the manuscript.

Funding
This work was supported in part by a Cornell Presidential Postdoctoral Fellowship to MS and funds from the National Institutes of Health to CF (U01HG009391; R35GM122550), to DFN and CF (UM1AI164559), to DFN (R01CA260691), and a seed grant from the Cornell Office of Academic Integration to CF and DFN. MLB is supported in part by the Department of Medicine Fund for the Future program at Weill Cornell Medicine, sponsored by the Elsa Miller Foundation.

Data availability
The datasets supporting the conclusions of this article are publicly available and can be found in NCBI repositories under the accession numbers GSE122323 [125], PRJNA420459 [126], GSE144334 [127], and GSE83482 [128]. ChIP-Seq data was sourced from the ENCODE consortium: https://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/. The source code supporting this study is available in our GitHub repository [129] and archived on Zenodo [DOI: 10.5281/zenodo.14625191] [142]. This code is distributed under the MIT License, details of which can be found in the license file included in the GitHub repository.

Declarations

Ethics approval and consent to participate
Not applicable. This study involves the use of publicly available sequencing data which was collected from human subjects who provided informed consent in the original studies. No new data from human participants was collected or used.

Competing interests
The authors declare that they have no competing interests.

Received: 30 August 2024 Accepted: 24 January 2025
Published online: 24 February 2025

References

1. The path that ends AIDS: UNAIDS Global AIDS Update 2023. (n.d.). Retrieved November 8, 2023, fromhttps://www.unaids.org/en/resources/documents/2023/global-aids-update-2023.
2. HIV and AIDS. (n.d.). Retrieved May 31, 2023, fromhttps://www.who.int/news-room/fact-sheets/detail/hiv-aids.
3. Global HIV & AIDS statistics—Fact sheet. (n.d.). Retrieved January 19, 2023, fromhttps://www.unaids.org/en/resources/fact-sheet.
4. Research priorities for an HIV cure: International AIDS Society Global Scientific Strategy 2021 | Nature Medicine. (n.d.). Retrieved August 18, 2023, fromhttps://www.nature.com/articles/s41591-021-01590-5.
5. Hütter G, Nowak D, Mossner M, Ganepola S, Müssig A, Allers K, Schneider T, Hofmann J, Kücherer C, Blau O, Blau IW, Hofmann WK, Thiel E. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med. 2009;360(7):692–8. https://doi.org/10.1056/NEJMoa0802905.
6. Gupta, R. K., Abdul-jawad, S., McCoy, L. E., Mok, H. P., Peppa, D., Salgado, M., Martinez-Picado, J., Nijhuis, M., Wensing, A. M. J., Lee, H., Grant, P., Nastouli, E., Lambert, J., Pace, M., Salasc, F., Monit, C., Innes, A., Muir, L., Waters, L., …

Olavarria, E. (2019). HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem cell transplantation. Nature, 568(7751), 244–248.https://doi.org/10.1038/s41586-019-1027-4.
7. Psomas, C. K., & Kinloch, S. (n.d.). Highlights of the Conference on Retroviruses and Opportunistic Infections, 4–9 March 2019, Seattle, WA, USA. Journal of Virus Eradication, 5(2), 125–131.
8. Turk, G., Seiger, K., Lian, X., Sun, W., Parsons, E. M., Gao, C., Rassadkina, Y., Polo, M. L., Czernikier, A., Ghiglione, Y., Vellicce, A., Varriale, J., Lai, J., Yuki, Y., Martin, M., Rhodes, A., Lewin, S. R., Walker, B. D., Carrington, M., … Yu, X. G. (2022). A Possible Sterilizing Cure of HIV-1 Infection Without Stem Cell Transplantation. Annals of Internal Medicine, 175(1), 95–100.https://doi.org/10.7326/L21-0297.
9. Jiang, C., Lian, X., Gao, C., Sun, X., Einkauf, K. B., Chevalier, J. M., Chen, S. M. Y., Hua, S., Rhee, B., Chang, K., Blackmer, J. E., Osborn, M., Peluso, M. J., Hoh, R., Somsouk, M., Milush, J., Bertagnolli, L. N., Sweet, S. E., Varriale, J. A., … Yu, X. G. (2020). Distinct viral reservoirs in individuals with spontaneous control of HIV-1. Nature, 585(7824), 261–267.https://doi.org/10.1038/s41586-020-2651-8.
10. Lifson AR, Rutherford GW, Jaffe HW. The natural history of human immunodeficiency virus infection. J Infect Dis. 1988;158(6):1360–7. https://doi.org/10.1093/infdis/158.6.1360.
11. Navarrete-Muñoz MA, Restrepo C, Benito JM, Rallón N. Elite controllers: A heterogeneous group of HIV-infected patients. Virulence. 2020;11(1):889–97. https://doi.org/10.1080/21505594.2020.1788887.
12. Woldemeskel BA, Kwaa AK, Blankson JN. Viral reservoirs in elite controllers of HIV-1 infection: Implications for HIV cure strategies. EBioMedicine. 2020;62: 103118. https://doi.org/10.1016/j.ebiom.2020.103118.
13. Casado, C., Galvez, C., Pernas, M., Tarancon-Diez, L., Rodriguez, C., Sanchez-Merino, V., Vera, M., Olivares, I., De Pablo-Bernal, R., Merino-Mansilla, A., Del Romero, J., Lorenzo-Redondo, R., Ruiz-Mateos, E., Salgado, M., Martinez-Picado, J., & Lopez-Galindez, C. (2020). Permanent control of HIV-1 pathogenesis in exceptional elite controllers: A model of spontaneous cure. Scientific Reports, 10(1), Article 1.https://doi.org/10.1038/s41598-020-58696-y.
14. Gaiha, G. D., Rossin, E. J., Urbach, J., Landeros, C., Collins, D. R., Nwonu, C., Muzhingi, I., Anahtar, M. N., Waring, O. M., Piechocka-Trocha, A., Waring, M., Worrall, D. P., Ghebremichael, M. S., Newman, R. M., Power, K. A., Allen, T. M., Chodosh, J., & Walker, B. D. (2019). Structural topology defines protective CD8+ T cell epitopes in the HIV proteome. Science (New York, N.Y.), 364(6439), 480–484.https://doi.org/10.1126/science.aav5095.
15. Rotger M, Dang KK, Fellay J, Heinzen EL, Feng S, Descombes P, Shianna KV, Ge D, Günthard HF, Goldstein DB, Telenti A, Immunology TSHCS, the C. for H. V. Genome-Wide mRNA Expression Correlates of Viral Control in CD4+ T-Cells from HIV-1-Infected Individuals. PLoS Pathog. 2010;6(2):e1000781. https://doi.org/10.1371/journal.ppat.1000781.
16. International HIV Controllers Study, Pereyra, F., Jia, X., McLaren, P. J., Telenti, A., de Bakker, P. I. W., Walker, B. D., Ripke, S., Brumme, C. J., Pulit, S. L., Carrington, M., Kadie, C. M., Carlson, J. M., Heckerman, D., Graham, R. R., Plenge, R. M., Deeks, S. G., Gianniny, L., Crawford, G., … Zhao, M. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science (New York, N.Y.), 330(6010), 1551–1557.https://doi.org/10.1126/science.1195271.
17. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci. 2000;97(6):2709–14. https://doi.org/10.1073/pnas.050567397.
18. Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., … Goldstein, D. B. (2007). A whole-genome association study of major determinants for host control of HIV-1. Science (New York, N.Y.), 317(5840), 944–947.https://doi.org/10.1126/science.1143767.
19. Prentice, H. A., Porter, T. R., Price, M. A., Cormier, E., He, D., Farmer, P. K., Kamali, A., Karita, E., Lakhi, S., Sanders, E. J., Anzala, O., Amornkul, P. N., Allen, S., Hunter, E., Kaslow, R. A., Gilmour, J., Tang, J., & for the IAVI African HIV Research Network. HLA-B*57 versus HLA-B*81 in HIV-1 Infection: Slow and Steady Wins the Race? J Virol. 2013;87(7):4043–51. https://doi.org/10.1128/jvi.03302-12.
20. Moosa Y, Tanko RF, Ramsuran V, Singh R, Madzivhandila M, Yende-Zuma N, Abrahams M-R, Selhorst P, Gounder K, Moore PL, Williamson C, Abdool Karim SS, Garrett NJ, Burgers WA. Case report: Mechanisms of HIV elite control in two African women. BMC Infect Dis. 2018;18:54. https://doi.org/10.1186/s12879-018-2961-8.
21. Carrington M, Bashirova AA, McLaren PJ. On stand by: Host genetics of HIV control. AIDS (London, England). 2013;27(18):2831–9. https://doi.org/10.1097/01.aids.0000432536.85335.c8.
22. Kwaa AK, Blankson JN. Immune Responses in Controllers of HIV Infection. Annu Rev Immunol. 2024;42(1):21–33. https://doi.org/10.1146/annurev-immunol-083122-035233.
23. Lian, X., Gao, C., Sun, X., Jiang, C., Einkauf, K. B., Seiger, K. W., Chevalier, J. M., Yuki, Y., Martin, M., Hoh, R., Peluso, M. J., Carrington, M., Ruiz-Mateos, E., Deeks, S. G., Rosenberg, E. S., Walker, B. D., Lichterfeld, M., & Yu, X. G. (2021). Signatures of immune selection in intact and defective proviruses distinguish HIV-1 elite controllers. Science Translational Medicine, 13(624), eabl4097.https://doi.org/10.1126/scitranslmed.abl4097.
24. Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, Parsons EM, Gladkov GT, Seiger KW, Blackmer JE, Jiang C, Yukl SA, Rosenberg ES, Yu XG, Lichterfeld M. Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell. 2022;185(2):266-282.e15. https://doi.org/10.1016/j.cell.2021.12.011.
25. Huang AS, Ramos V, Oliveira TY, Gaebler C, Jankovic M, Nussenzweig MC, Cohn LB. Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. J Exp Med. 2021;218(12):e20211427. https://doi.org/10.1084/jem.20211427.
26. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. Ten things you should know about transposable elements. Genome Biol. 2018;19(1):199. https://doi.org/10.1186/s13059-018-1577-z.
27. Rosspopoff O, Trono D. Take a walk on the KRAB side. Trends Genet. 2023;39(11):844–57. https://doi.org/10.1016/j.tig.2023.08.003.
28. Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Biol. 2015;31:429–51. https://doi.org/10.1146/annurev-cellbio-100814-125514.

29. Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. Nature Reviews Genetics, 18(2), Article 2.https://doi.org/10.1038/nrg.2016.139.
30. Frank JA, Singh M, Cullen HB, Kirou RA, Benkaddour-Boumzaouad M, Cortes JL, Garcia Pérez J, Coyne CB, Feschotte C. Evolution and antiviral activity of a human protein of retroviral origin. Science. 2022;378(6618):422–8. https://doi.org/10.1126/science.abq7871.
31. Feschotte C. The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet. 2008;9(5):397–405. https://doi.org/10.1038/nrg2337.
32. Cosby RL, Chang N-C, Feschotte C. Host-transposon interactions: Conflict, cooperation, and cooption. Genes Dev. 2019;33(17–18):1098–116. https://doi.org/10.1101/gad.327312.119.
33. Fueyo, R., Judd, J., Feschotte, C., & Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. Nature Reviews Molecular Cell Biology, 23(7), Article 7.https://doi.org/10.1038/s41580-022-00457-y.
34. Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science (New York, N.Y.), 351(6277), 1083–1087.https://doi.org/10.1126/science.aad5497.
35. Braun, E., Hotter, D., Koepke, L., Zech, F., Groß, R., Sparrer, K. M. J., Müller, J. A., Pfaller, C. K., Heusinger, E., Wombacher, R., Sutter, K., Dittmer, U., Winkler, M., Simmons, G., Jakobsen, M. R., Conzelmann, K.-K., Pöhlmann, S., Münch, J., Fackler, O. T., … Sauter, D. (2019). Guanylate-Binding Proteins 2 and 5 Exert Broad Antiviral Activity by Inhibiting Furin-Mediated Processing of Viral Envelope Proteins. Cell Reports, 27(7), 2092–2104.e10.https://doi.org/10.1016/j.celrep.2019.04.063.
36. Srinivasachar Badarinarayan S, Shcherbakova I, Langer S, Koepke L, Preising A, Hotter D, Kirchhoff F, Sparrer KMJ, Schotta G, Sauter D. HIV-1 infection activates endogenous retroviral promoters regulating antiviral gene expression. Nucleic Acids Res. 2020;48(19):10890–908. https://doi.org/10.1093/nar/gkaa832.
37. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 2003;424(6944):99–103. https://doi.org/10.1038/nature01709.
38. Singh M, Kondrashkina AM, Widmann TJ, Cortes JL, Bansal V, Wang J, Römer C, Garcia-Canadas M, Garcia-Perez JL, Hurst LD, Izsvák Z. A new human embryonic cell type associated with activity of young transposable elements allows definition of the inner cell mass. PLoS Biol. 2023;21(6):e3002162. https://doi.org/10.1371/journal.pbio.3002162.
39. Kyriakou E, Magiorkinis G. Interplay between endogenous and exogenous human retroviruses. Trends Microbiol. 2023;31(9):933–46. https://doi.org/10.1016/j.tim.2023.03.008.
40. Jones, R. B., Garrison, K. E., Mujib, S., Mihajlovic, V., Aidarus, N., Hunter, D. V., Martin, E., John, V. M., Zhan, W., Faruk, N. F., Gyenes, G., Sheppard, N. C., Priumboom-Brees, I. M., Goodwin, D. A., Chen, L., Rieger, M., Muscat-King, S., Loudon, P. T., Stanley, C., … Ostrowski, M. A. (2012). HERV-K-specific T cells eliminate diverse HIV-1/2 and SIV primary isolates. The Journal of Clinical Investigation, 122(12), 4473–4489.https://doi.org/10.1172/JCI64560.
41. Liu C-H, Grandi N, Palanivelu L, Tramontano E, Lin L-T. Contribution of Human Retroviruses to Disease Development-A Focus on the HIV- and HERV-Cancer Relationships and Treatment Strategies. Viruses. 2020;12(8):852. https://doi.org/10.3390/v12080852.
42. Boritz, E. A., Darko, S., Swaszek, L., Wolf, G., Wells, D., Wu, X., Henry, A. R., Laboune, F., Hu, J., Ambrozak, D., Hughes, M. S., Hoh, R., Casazza, J. P., Vostal, A., Bunis, D., Nganou-Makamdop, K., Lee, J. S., Migueles, S. A., Koup, R. A., … Douek, D. C. (2016). Multiple Origins of Virus Persistence during Natural Control of HIV Infection. Cell, 166(4), 1004–1015.https://doi.org/10.1016/j.cell.2016.06.039.
43. Zhang W, Ambikan AT, Sperk M, van Domselaar R, Nowak P, Noyan K, Russom A, Sönnerborg A, Neogi U. Transcriptomics and Targeted Proteomics Analysis to Gain Insights Into the Immune-control Mechanisms of HIV-1 Infected Elite Controllers. EBioMedicine. 2018;27:40–50. https://doi.org/10.1016/j.ebiom.2017.11.031.
44. Gonzalo-Gil, E., Rapuano, P. B., Ikediobi, U., Leibowitz, R., Mehta, S., Coskun, A. K., Porterfield, J. Z., Lampkin, T. D., Marconi, V. C., Rimland, D., Walker, B. D., Deeks, S., & Sutton, R. E. (2019). Transcriptional down-regulation of ccr5 in a subset of HIV+ controllers and their family members. eLife, 8, e44360.https://doi.org/10.7554/eLife.44360.
45. Walker WE, Kurscheid S, Joshi S, Lopez CA, Goh G, Choi M, Barakat L, Francis J, Fisher A, Kozal M, Zapata H, Shaw A, Lifton R, Sutton RE, Fikrig E. Increased Levels of Macrophage Inflammatory Proteins Result in Resistance to R5-Tropic HIV-1 in a Subset of Elite Controllers. J Virol. 2015;89(10):5502–14. https://doi.org/10.1128/JVI.00118-15.
46. Raposo RAS, Abdel-Mohsen M, Holditch SJ, Kuebler PJ, Cheng RG, Eriksson EM, Liao W, Pillai SK, Nixon DF. Increased expression of intrinsic antiviral genes in HLA-B*57-positive individuals. J Leukoc Biol. 2013;94(5):1051–9. https://doi.org/10.1189/jlb.0313150.
47. McLaren, P. J., & Fellay, J. (2021). HIV-1 and human genetic variation. Nature Reviews Genetics, 22(10), Article 10.https://doi.org/10.1038/s41576-021-00378-0.
48. Foster TL, Wilson H, Iyer SS, Coss K, Doores K, Smith S, Kellam P, Finzi A, Borrow P, Hahn BH, Neil SJD. Resistance of Transmitted Founder HIV-1 to IFITM-Mediated Restriction. Cell Host Microbe. 2016;20(4):429–42. https://doi.org/10.1016/j.chom.2016.08.006.
49. Chaudhuri A, Yang B, Gendelman HE, Persidsky Y, Kanmogne GD. STAT1 signaling modulates HIV-1–induced inflammatory responses and leukocyte transmigration across the blood-brain barrier. Blood. 2008;111(4):2062–72. https://doi.org/10.1182/blood-2007-05-091207.
50. Chen J, He Y, Zhong H, Hu F, Li Y, Zhang Y, Zhang X, Lin W, Li Q, Xu F, Chen S, Zhang H, Cai W, Li L. Transcriptome analysis of CD4+ T cells from HIV-infected individuals receiving ART with LLV revealed novel transcription factors regulating HIV-1 promoter activity. Virologica Sinica. 2023;38(3):398–408. https://doi.org/10.1016/j.virs.2023.03.001.
51. D'Antoni ML, Kallianpur KJ, Premeaux TA, Corley MJ, Fujita T, Laws EI, Ogata-Arakaki D, Chow DC, Khadka VS, Shikuma CM, Ndhlovu LC. Lower Interferon Regulatory Factor-8 Expression in Peripheral Myeloid Cells Tracks With Adverse Central Nervous System Outcomes in Treated HIV Infection. Front Immunol. 2019;10:2789. https://doi.org/10.3389/fimmu.2019.02789.

52. Sugawara S, Reeves RK, Jost S. Learning to Be Elite: Lessons From HIV-1 Controllers and Animal Models on Trained Innate Immunity and Virus Suppression. Front Immunol. 2022;13:858383. https://doi.org/10.3389/fimmu.2022.858383.
53. Morou, A., Brunet-Ratnasingham, E., Dubé, M., Charlebois, R., Mercier, E., Darko, S., Brassard, N., Nganou-Makamdop, K., Arumugam, S., Gendron-Lepage, G., Yang, L., Niessl, J., Baxter, A. E., Billingsley, J. M., Rajakumar, P. A., Lefebvre, F., Johnson, R. P., Tremblay, C., Routy, J.-P., … Kaufmann, D. E. (2019). Altered differentiation is central to HIV-specific CD4+ T cell dysfunction in progressive disease. Nature Immunology, 20(8), 1059–1070.https://doi.org/10.1038/s41590-019-0418-x.
54. Krishnan S, Wilson EMP, Sheikh V, Rupert A, Mendoza D, Yang J, Lempicki R, Migueles SA, Sereti I. Evidence for Innate Immune System Activation in HIV Type 1–Infected Elite Controllers. J Infect Dis. 2014;209(6):931–9. https://doi.org/10.1093/infdis/jit581.
55. Díez-Fuertes F, De La Torre-Tarazona HE, Calonge E, Pernas M, del Alonso-Socas MM, Capa L, García-Pérez J, Sakuntabhai A, Alcamí J. Transcriptome Sequencing of Peripheral Blood Mononuclear Cells from Elite Controller-Long Term Non Progressors. Scientific Reports. 2019;9:14265. https://doi.org/10.1038/s41598-019-50642-x.
56. Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell. 2019;24(5):724-735.e5. https://doi.org/10.1016/j.stem.2019.03.012.
57. Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell. 2015;161(6):1453–67. https://doi.org/10.1016/j.cell.2015.04.053.
58. Barnada SM, Isopi A, Tejada-Martinez D, Goubert C, Patoori S, Pagliaroli L, Tracewell M, Trizzino M. Genomic features underlie the co-option of SVA transposons as cis-regulatory elements in human pluripotent stem cells. PLoS Genet. 2022;18(6):e1010225. https://doi.org/10.1371/journal.pgen.1010225.
59. Patoori, S., Barnada, S. M., Large, C., Murray, J. I., & Trizzino, M. (2022). Young transposable elements rewired gene regulatory networks in human and chimpanzee hippocampal intermediate progenitors. Development, 149(19), dev200413.https://doi.org/10.1242/dev.200413.
60. Trizzino M, Kapusta A, Brown CD. Transposable elements generate regulatory novelty in a tissue-specific fashion. BMC Genomics. 2018;19:468. https://doi.org/10.1186/s12864-018-4850-3.
61. Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. Transposable elements are the primary source of novelty in primate gene regulation. Genome Res. 2017;27(10):1623–33. https://doi.org/10.1101/gr.218149.116.
62. Karttunen K, Patel D, Xia J, Fei L, Palin K, Aaltonen L, Sahu B. Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. Nat Commun. 2023;14(1):5313. https://doi.org/10.1038/s41467-023-41081-4.
63. Ivancevic, A., Simpson, D. M., & Chuong, E. B. (2021). Endogenous retroviruses mediate transcriptional rewiring in response to oncogenic signaling in colorectal cancer (p. 2021.10.28.466196). bioRxiv.https://doi.org/10.1101/2021.10.28.466196.
64. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci USA. 2007;104(47):18613–8. https://doi.org/10.1073/pnas.0703637104.
65. Calderon D, Nguyen MLT, Mezger A, Kathiria A, Müller F, Nguyen V, Lescano N, Wu B, Trombetta J, Ribado JV, Knowles DA, Gao Z, Blaeschke F, Parent AV, Burt TD, Anderson MS, Criswell LA, Greenleaf WJ, Marson A, Pritchard JK. Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat Genet. 2019;51(10):1494–505. https://doi.org/10.1038/s41588-019-0505-9.
66. Hokello, J., Tyagi, K., Owor, R. O., Sharma, A. L., Bhushan, A., Daniel, R., & Tyagi, M. (2024). New Insights into HIV Life Cycle, Th1/Th2 Shift during HIV Infection and Preferential Virus Infection of Th2 Cells: Implications of Early HIV Treatment Initiation and Care. Life, 14(1), Article 1.https://doi.org/10.3390/life14010104.
67. Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., Ch'ang, L.-Y., Huang, W., Liu, B., Shen, Y., Tam, P. K.-H., Tsui, L.-C., Waye, M. M. Y., Wong, J. T.-F., Zeng, C., Zhang, Q., Chee, M. S., Galver, L. M., Kruglyak, S., … Methods Group. (2003). The International HapMap Project. Nature, 426(6968), Article 6968.https://doi.org/10.1038/nature02168.
68. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–92. https://doi.org/10.1093/nar/gkac963.
69. The Gene Ontology Consortium, Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. Genetics, 224(1), iyad031.https://doi.org/10.1093/genetics/iyad031.
70. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell. 2002;10(5):1247–53. https://doi.org/10.1016/s1097-2765(02)00742-6.
71. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. DNA deamination mediates innate immunity to retroviral infection. Cell. 2003;113(6):803–9. https://doi.org/10.1016/s0092-8674(03)00423-9.
72. Yaseen MM, Abuharfeil NM, Darmani H. The Role of p53 in HIV Infection. Curr HIV/AIDS Rep. 2023;20(6):419–27. https://doi.org/10.1007/s11904-023-00684-8.
73. Duan L, Ozaki I, Oakes JW, Taylor JP, Khalili K, Pomerantz RJ. The tumor suppressor protein p53 strongly alters human immunodeficiency virus type 1 replication. J Virol. 1994;68(7):4302–13.
74. Gualberto A, Baldwin AS. P53 and Sp1 interact and cooperate in the tumor necrosis factor-induced transcriptional activation of the HIV-1 long terminal repeat. J Biol Chem. 1995;270(34):19680–3. https://doi.org/10.1074/jbc.270.34.19680.
75. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199–205. https://doi.org/10.1093/nar/gkz401.

76. Boleij A, Fathi P, Dalton W, Park B, Wu X, Huso D, Allen J, Besharati S, Anders RA, Housseau F, Mackenzie AE, Jenkins L, Milligan G, Wu S, Sears CL. G-protein coupled receptor 35 (GPR35) regulates the colonic epithelial cell response to enterotoxigenic Bacteroides fragilis. Communications Biology. 2021;4:585. https://doi.org/10.1038/s42003-021-02014-3.
77. Lao X, Mei X, Zou J, Xiao Q, Ning Q, Xu X, Zhang C, Ji L, Deng S, Lu B, Chen M. Pyroptosis associated with immune reconstruction failure in HIV-1- infected patients receiving antiretroviral therapy: A cross-sectional study. BMC Infect Dis. 2022;22(1):867. https://doi.org/10.1186/s12879-022-07818-0.
78. Tang S-W, Ducroux A, Jeang K-T, Neuveut C. Impact of cellular autophagy on viruses: Insights from hepatitis B virus and human retroviruses. J Biomed Sci. 2012;19(1):92. https://doi.org/10.1186/1423-0127-19-92.
79. Kulski JK. Long Noncoding RNA HCP5, a Hybrid HLA Class I Endogenous Retroviral Gene: Structure, Expression, and Disease Associations. Cells. 2019;8(5):480. https://doi.org/10.3390/cells8050480.
80. Jiang J-C, Rothnagel JA, Upton KR. Widespread Exaptation of L1 Transposons for Transcription Factor Binding in Breast Cancer. Int J Mol Sci. 2021;22(11):5625. https://doi.org/10.3390/ijms22115625.
81. Curty, G., Iñiguez, L. P., Nixon, D. F., Soares, M. A., & de Mulder Rougvie, M. (2021). Hallmarks of Retroelement Expression in T-Cells Treated With HDAC Inhibitors. Frontiers in Virology, 1. https://www.frontiersin.org/articles/10.3389/fviro.2021.756635.
82. Yang P, Wang Y, Macfarlan TS. The role of KRAB-ZFPs in transposable element repression and mammalian evolution. Trends in Genetics : TIG. 2017;33(11):871–81. https://doi.org/10.1016/j.tig.2017.08.006.
83. Jacobs, F. M. J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A. D., Katzman, S., Paten, B., Salama, S. R., & Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature, 516(7530), Article 7530.https://doi.org/10.1038/nature13760.
84. de Tribolet-Hardy J, Thorball CW, Forey R, Planet E, Duc J, Coudray A, Khubieh B, Offner S, Pulver C, Fellay J, Imbeault M, Turelli P, Trono D. Genetic features and genomic targets of human KRAB-zinc finger proteins. Genome Res. 2023;33(8):1409–23. https://doi.org/10.1101/gr.277722.123.
85. de Tribolet-Hardy J, Thorball CW, Forey R, Planet E, Duc J, Coudray A, Khubieh B, Offner S, Pulver C, Fellay J, Imbeault M, Turelli P, Trono D. Genetic features and genomic targets of human KRAB-zinc finger proteins. Genome Res. 2023. https://doi.org/10.1101/gr.277722.123.
86. Imbeault, M., Helleboid, P.-Y., & Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature, 543(7646), Article 7646.https://doi.org/10.1038/nature21683.
87. Hermant C, Torres-Padilla M-E. TFs for TEs: The transcription factor repertoire of mammalian transposable elements. Genes Dev. 2021;35(1–2):22–39. https://doi.org/10.1101/gad.344473.120.
88. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. https://doi.org/10.1038/nature11247.
89. Bruno M, Mahgoub M, Macfarlan TS. The Arms Race Between KRAB–Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. Annu Rev Genet. 2019;53(1):393–416. https://doi.org/10.1146/annurev-genet-112618-043717.
90. Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature. 2005;433(7024):430–3. https://doi.org/10.1038/nature03238.
91. Lecossier, D., Bouchonnet, F., Clavel, F., & Hance, A. J. (2003). Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science (New York, N.Y.), 300(5622), 1112.https://doi.org/10.1126/science.1083338.
92. Wang X, Ao Z, Chen L, Kobinger G, Peng J, Yao X. The Cellular Antiviral Protein APOBEC3G Interacts with HIV-1 Reverse Transcriptase and Inhibits Its Function during Viral Replication. J Virol. 2012;86(7):3777–86. https://doi.org/10.1128/JVI.06594-11.
93. Dang Y, Siew LM, Wang X, Han Y, Lampen R, Zheng Y-H. Human Cytidine Deaminase APOBEC3H Restricts HIV-1 Replication. J Biol Chem. 2008;283(17):11606–14. https://doi.org/10.1074/jbc.M707586200.
94. Yang, H., Ito, F., Wolfe, A. D., Li, S., Mohammadzadeh, N., Love, R. P., Yan, M., Zirkle, B., Gaba, A., Chelico, L., & Chen, X. S. (2020). Understanding the structural basis of HIV-1 restriction by the full length double-domain APOBEC3G. Nature Communications, 11(1), Article 1.https://doi.org/10.1038/s41467-020-14377-y.
95. Yoon W, Ma B-J, Fellay J, Huang W, Xia S-M, Zhang R, Shianna KV, Liao H-X, Haynes BF, Goldstein DB. A polymorphism in the HCP5 gene associated with HLA-B*5701 does not restrict HIV-1 in vitro. AIDS (London, England). 2010;24(1):155–7. https://doi.org/10.1097/QAD.0b013e32833202f5.
96. Thørner LW, Erikstrup C, Harritshøj LH, Larsen MH, Kronborg G, Pedersen C, Larsen CS, Pedersen G, Gerstoft J, Obel N, Ullum H. Impact of polymorphisms in the HCP5 and HLA-C, and ZNRD1 genes on HIV viral load. Infect Genet Evol. 2016;41:185–90. https://doi.org/10.1016/j.meegid.2016.03.037.
97. Buttler CA, Chuong EB. Emerging roles for endogenous retroviruses in immune epigenetic regulation. Immunol Rev. 2022;305(1):165–78. https://doi.org/10.1111/imr.13042.
98. Lopez-Galindez C, Pernas M, Casado C, Olivares I, Lorenzo-Redondo R. Elite controllers and lessons learned for HIV-1 cure. Curr Opin Virol. 2019;38:31–6. https://doi.org/10.1016/j.coviro.2019.05.010.
99. De La Torre-Tarazona E, Ayala-Suárez R, Díez-Fuertes F, Alcamí J. Omic Technologies in HIV: Searching Transcriptional Signatures Involved in Long-Term Non-Progressor and HIV Controller Phenotypes. Front Immunol. 2022;13:926499. https://doi.org/10.3389/fimmu.2022.926499.
100. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell. 1996;86(3):367–77. https://doi.org/10.1016/s0092-8674(00)80110-5.
101. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, Lederman MM, Benito JM, Goepfert PA, Connors M, Roederer M, Koup RA. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. Blood. 2006;107(12):4781–9. https://doi.org/10.1182/blood-2005-12-4818.
102. Hersperger AR, Martin JN, Shin LY, Sheth PM, Kovacs CM, Cosma GL, Makedonas G, Pereyra F, Walker BD, Kaul R, Deeks SG, Betts MR. Increased HIV-specific CD8+ T-cell cytotoxic potential in HIV elite controllers is associated with T-bet expression. Blood. 2011;117(14):3799–808. https://doi.org/10.1182/blood-2010-12-322727.

103. Genovese L, Nebuloni M, Alfano M. Cell-Mediated Immunity in Elite Controllers Naturally Controlling HIV Viral Load. Front Immunol. 2013;4:86. https://doi.org/10.3389/fimmu.2013.00086.
104. Nabi R, Moldoveanu Z, Wei Q, Golub ET, Durkin HG, Greenblatt RM, Herold BC, Nowicki MJ, Kassaye S, Cho MW, Pinter A, Landay AL, Mestecky J, Kozlowski PA. Differences in serum IgA responses to HIV-1 gp41 in elite controllers compared to viral suppressors on highly active antiretroviral therapy. PLoS ONE. 2017;12(7):e0180245. https://doi.org/10.1371/journal.pone.0180245.
105. Jacobs, E. S., Keating, S. M., Abdel-Mohsen, M., Gibb, S. L., Heitman, J. W., Inglis, H. C., Martin, J. N., Zhang, J., Kaidarova, Z., Deng, X., Wu, S., Anastos, K., Crystal, H., Villacres, M. C., Young, M., Greenblatt, R. M., Landay, A. L., Gange, S. J., Deeks, S. G., … Norris, P. J. (2017). Cytokines Elevated in HIV Elite Controllers Reduce HIV Replication In Vitro and Modulate HIV Restriction Factor Expression. Journal of Virology, 91(6), e02051–16.https://doi.org/10.1128/JVI.02051-16.
106. Chen X, Pacis A, Aracena KA, Gona S, Kwan T, Groza C, Lin YL, Sindeaux R, Yotova V, Pramatarova A, Simon M-M, Pastinen T, Barreiro LB, Bourque G. Transposable elements are associated with the variable response to influenza infection. Cell Genomics. 2023;3(5):100292. https://doi.org/10.1016/j.xgen.2023.100292.
107. McLaren PJ, Carrington M. The impact of host genetic variation on infection with HIV-1. Nat Immunol. 2015;16(6):577–83. https://doi.org/10.1038/ni.3147.
108. Zhang Z-N, Xu J-J, Fu Y-J, Liu J, Jiang Y-J, Cui H-L, Zhao B, Sun H, He Y-W, Li Q-J, Shang H. Transcriptomic analysis of peripheral blood mononuclear cells in rapid progressors in early HIV infection identifies a signature closely correlated with disease progression. Clin Chem. 2013;59(8):1175–86. https://doi.org/10.1373/clinchem.2012.197335.
109. Xie, M., Hong, C., Zhang, B., Lowdon, R. F., Xing, X., Li, D., Zhou, X., Lee, H. J., Maire, C. L., Ligon, K. L., Gascard, P., Sigaroudinia, M., Tlsty, T. D., Kadlecek, T., Weiss, A., O'Geen, H., Farnham, P. J., Madden, P. A. F., Mungall, A. J., … Wang, T. (2013). DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nature Genetics, 45(7), 836–841.https://doi.org/10.1038/ng.2649.
110. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Dev Cell. 2016;36(6):611–23. https://doi.org/10.1016/j.devcel.2016.02.024.
111. Leung A, Trac C, Kato H, Costello KR, Chen Z, Natarajan R, Schones DE. LTRs activated by Epstein-Barr virus–induced transformation of B cells alter the transcriptome. Genome Res. 2018;28(12):1791–8. https://doi.org/10.1101/gr.233585.117.
112. Babaian A, Romanish MT, Gagnier L, Kuo LY, Karimi MM, Steidl C, Mager DL. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. Oncogene. 2016;35(19):2542–6. https://doi.org/10.1038/onc.2015.308.
113. Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, Babaian A, Sistiaga-Poveda M, Lai CB, Nemirovsky O, Serrano I, Steidl C, Karimi MM, Mager DL. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. Proc Natl Acad Sci USA. 2014;111(34):E3534–43. https://doi.org/10.1073/pnas.1405507111.
114. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M. A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K. D., Callen, D. F., Harte, M. F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I., … Mathas, S. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nature Medicine, 16(5), 571–579, 1p following 579.https://doi.org/10.1038/nm.2129.
115. Bertozzi TM, Ferguson-Smith AC. Metastable epialleles and their contribution to epigenetic inheritance in mammals. Semin Cell Dev Biol. 2020;97:93–105. https://doi.org/10.1016/j.semcdb.2019.08.002.
116. Bertozzi TM, Becker JL, Blake GET, Bansal A, Nguyen DK, Fernandez-Twinn DS, Ozanne SE, Bartolomei MS, Simmons RA, Watson ED, Ferguson-Smith AC. Variably methylated retrotransposons are refractory to a range of environmental perturbations. Nat Genet. 2021;53(8):1233–42. https://doi.org/10.1038/s41588-021-00898-9.
117. Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. Epigenetics Chromatin. 2018;11(1):7. https://doi.org/10.1186/s13072-018-0177-1.
118. Bertozzi TM, Elmer JL, Macfarlan TS, Ferguson-Smith AC. KRAB zinc finger protein diversification drives mammalian interindividual methylation variability. Proc Natl Acad Sci. 2020;117(49):31290–300. https://doi.org/10.1073/pnas.2017053117.
119. Carter, T. A., Singh, M., Dumbović, G., Chobirko, J. D., Rinn, J. L., & Feschotte, C. (2022). Mosaic cis-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. eLife, 11, e76257.https://doi.org/10.7554/eLife.76257.
120. Altemose, N., Noor, N., Bitoun, E., Tumian, A., Imbeault, M., Chapman, J. R., Aricescu, A. R., & Myers, S. R. (2017). A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. eLife, 6, e28383.https://doi.org/10.7554/eLife.28383.
121. Ito, J., Kimura, I., Soper, A., Coudray, A., Koyanagi, Y., Nakaoka, H., Inoue, I., Turelli, P., Trono, D., & Sato, K. (2020). Endogenous retroviruses drive KRAB zinc-finger protein family expression for tumor suppression. Science Advances, 6(43), eabc3020.https://doi.org/10.1126/sciadv.abc3020.
122. Tie, C. H., Fernandes, L., Conde, L., Robbez-Masson, L., Sumner, R. P., Peacock, T., Rodriguez-Plata, M. T., Mickute, G., Gifford, R., Towers, G. J., Herrero, J., & Rowe, H. M. (2018). KAP1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. EMBO Reports, 19(10).https://doi.org/10.15252/embr.201745000.
123. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. 2014;24(8):1260–70. https://doi.org/10.1101/gr.172833.114.
124. Marzetta, F., Simó-Riudalbas, L., Duc, J., Planet, E., Verp, S., Turelli, P., & Trono, D. (2019). The KZFP/KAP1 system controls transposable elements-embedded regulatory sequences in adult T cells (p. 523597). bioRxiv.https://doi.org/10.1101/523597.

125. Gonzalo-Gil E, Rapuano PB, Ikediobi U, Leibowitz R et al. Transcriptional Down-regulation of CCR5 in a Subset of HIV+ Controllers. GSE122323. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122323. (2019).
126. Zhang W et al. Transcriptomic analysis to map mechanisms of viral replication control in HIV-1 positive Elite Controllers. PRJNA420459. BioProject. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA420459. (2017).
127. Jiang C, Lian X, Gao C, Sun X et al. A unique viral reservoir landscape associated with durable natural control of HIV-1 infection. GSE144334. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144334. (2020).
128. Douek D, Boritz E, Hu J. Multiple Origins of Virus Persistence during Natural Control of HIV Infection. GSE83482. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE83482. (2016).
129. Singh, Manvendra, Leddy, Sabrina M., Iñiguez, Luis Pedro, Bendall, Matthew L., Nixon, Douglas F., & Feschotte, Cedric. (2024). Transposable elements may enhance antiviral resistance in HIV-1 elite controllers, code repository. GitHub.https://github.com/Manu-1512/EndogenousRetroelements-in-HIV-infection.
130. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.
131. Salmon provides fast and bias-aware quantification of transcript expression | Nature Methods. (n.d.). Retrieved December 11, 2023, fromhttps://www.nature.com/articles/nmeth.4197.
132. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8.
133. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9. https://doi.org/10.1038/nmeth.1923.
134. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.
135. The Human Genome Browser at UCSC. (n.d.). Retrieved December 11, 2023, fromhttps://genome.cshlp.org/content/12/6/996.abstract.
136. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033.
137. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research, 44(Web Server issue), W90–W97.https://doi.org/10.1093/nar/gkw377.
138. Review Commons Report 1. Early Evidence Base. (2024).https://doi.org/10.15252/rc.2024874864.
139. Review Commons Report 2. Early Evidence Base. (2024).https://doi.org/10.15252/rc.2024309412.
140. Review Commons Report 3. Early Evidence Base. (2024).https://doi.org/10.15252/rc.2024025927.
141. Review Commons Response 1. Early Evidence Base. (2024).https://doi.org/10.15252/rc.2024679388.
142. Singh, M., Leddy, Sabrina M., Iñiguez, Luis Pedro, Bendall, Matthew L., Nixon, Douglas F., & Feschotte, Cedric. (2025). Transposable elements may enhance antiviral resistance in HIV-1 elite controllers (Version 1.0.0) [Computer software]. Zenodo.https://doi.org/10.5281/zenodo.14625191.

Publisher's Note