

Review history

First round of review

Reviewer 1

Singh et al. analyze the expression and putative contribution of TEs in CD4+ T cells in HIV elite controllers. Through re-analysis of existing datasets, the authors describe broad differences in expression of TEs in ECs through analysis of RNAseq and ATACseq data, and come up with convincing examples where differentially-expressed innate immune genes correlate with increased accessibility of proximal TEs. Overall, the authors' conclusions are appropriately measured, though the manuscript text should be re-organized for clarity and a few further analyses are needed to support the main message of the paper.

****Major comments:**** The manuscript would benefit from a re-organization of the figures to focus on TEs - in particular, Fig 1B, Fig 2, and Fig 3 reproduce known transcriptional differences between ECs and HCs and serve as quality controls for the authors' computational analysis. Conversely, Supplementary Fig 6 contains very interesting data on KZNF expression and should be included in the main figures.

It remains unclear whether differences in TE expression described are specific to ECs or to EC-like CD4+ T cell states. As there are plenty of datasets available that compare the transcriptome of naïve, activated, exhausted, and regulatory CD4+ T cells, the authors should compare the TE expression patterns observed in ECs to activated CD4+ T cells, particularly those with a Th1 and cytotoxic phenotype analogous to those observed in ECs, from healthy donors.

In Fig 1, the authors demonstrate differential expression of both innate immune genes and TEs, but the link between the two is unclear. Is there any enrichment in differential expression for TEs located proximal to innate immune genes? This type of analysis should be possible using the authors' own software to map TE expression to specific genomic loci.

****Optional:**** In Fig 3, the authors cluster CD4+ T cells based on transcriptomic profiles. It would be interesting to re-cluster these samples based on TE expression alone, given the differences in TE expression described in Fig 5.

The manuscript by Singh et al. describes for the first time the role of TEs in HIV elite controllers, suggesting that TEs may be co-opted for cis-regulatory function. This study builds off prior work demonstrating that HIV-infected CD4+ T cells activate LTR elements that may regulate expression of interferon-inducible genes, demonstrating that ECs show further upregulation of innate immune genes. While these findings will need to be experimentally validated, this study constitutes a useful resource and adds to the growing body of evidence implicating TEs in cis-regulatory control of immune genes. This study will be of interest to basic scientists interested in genetic mechanisms of HIV control, and if further developed may comprise a useful source of biomarkers to predict viral kinetics in HIV-infected individuals.

My expertise is in immunology, TE biology, and viral infection

Reviewer 2

The authors have re-analyzed published RNA-Seq data from CD4 T cells isolated from HIV elite controllers and reference cohorts, including HIV negative persons, viremic progressors and ART-treated persons. Their main finding is that in some of their comparisons, EC have higher levels of interferon-stimulated genes (ISG), paired with distinct expression patterns of transposable elements. The authors suggest that expression of transposable elements may induce altered expression of ISG, presumably due to immune recognition of TE. They also suggest that reduced expression of KZNF genes, which encode for transcription factors that can suppress TE, may be responsible for enhanced expression of TE. I have the following comments:

1. All data included in this manuscript derive from previously published data. A new dataset, specifically designed to focus on a high-resolution analysis of TE expression, would be better suited to address the proposed questions.
2. As the authors acknowledge, the described investigations are exploratory, and do not allow to draw firm conclusions. Mechanistic experiments are recommended to address the authors' hypotheses.
3. An important limitation is that virological data of EC are not considered. For example, I believe it is a lot more likely that the upregulation of ISG in EC relates to ongoing low-level viral replication. The authors could analyze cell-associated HIV RNA and DNA levels and determine how they associate with ISG expression.
4. KZNF genes seem downregulated in EC. Can the authors propose a reason/mechanism for that?

Overall, I think this is an interesting manuscript that proposes a distinct and potentially important mechanisms that may contribute to immune control of HIV. My suggestions to improve the manuscript are complex and cannot be easily addressed through experimental work. I believe a possible option would be to publish the present manuscript without my proposed modifications, but highlight the weaknesses of the current paper more clearly; mechanistic studies could then be deferred to a future study.

Reviewer 3

****Summary:****

This manuscript presents an analysis of published gene expression (RNA-seq and ATAC-seq) data from a couple of cohorts of HIV-infected elite controllers (EC), as compared to uninfected controls, (HC), virological progressors (VP). The authors report that HIV elite controllers may exhibit 4 distinct patterns of TE (and gene) expression and suggest that TE expression may drive some form of antiviral gene expression. Further, they show that heterogeneous TE expression may be determined by differential KZNF gene activity among the different clusters of elite controllers. These results are very interesting, even though the conclusions are very preliminary. It presents intriguing correlations between expression of certain TE groups of LINES and HERVs, and the clustering into 4 gene expression groups in EC and is a novel finding. That said, correlation is not causation, and the authors need to be more cautious in presenting their highly preliminary model in Figure 6.

****Major comments:****

Overall, although preliminary, as the authors note, the results are interesting and worthy of follow-up. At this point, however, a number of issues arise that need further clarification and analysis before I would consider this study complete.

First, the analyses shown in Figures 3-5 based on data from studies on EC of CD4 cells are apparently motivated by the differential TE expression in total PBMCs shown in Fig 1 and 2. Yet, the TE groups (please don't use taxonomic terms like "subfamily") identified in Fig 2 and Fig 4 are completely different, with no overlap. This discrepancy underscores the possibility that the differential expression observed is, at least in part, due to the differences among the groups or clusters in cell type composition, as seen in Fig 2D and 3B which, themselves, could be a consequence of HIV infection and elite control (which has been shown to involve ongoing, albeit low-level, virus replication). This issue must be addressed.

Similarly, of the 12 DE TE groups in EC in Fig 5A, only 3 overlap with the 16 in EC Fig S1.

Second, The introduction points out the strongly supported, association between elite control and immunogenetic determinants, most notably specific HLA-B types, but also innate immunity factors. This cries out for inclusion of these factors in the analyses of this manuscript, in the format of Figure S4, for example, but none is to be found. The relevant genotypes are likely available in the metadata in the references cited, but, if not, could be inferred from the RNA-seq data.

It would also be very useful to present the KZNF data in Figure 5 the same way, since, looking at Fig 5C, the correlation of high and low KZNF expression, while clearly correlated with a that of few groups of elements, with clustering into specific groups does not appear to be well supported.

In general, other than the cell type composition differences, there is no presentation of evidence for any biologically important feature associated with the clusters found.

Third, the figures present values that have been very heavily analyzed, and it is difficult to impossible to infer what the underlying data look like. For example, with the exception of a few selected examples in Figs 4 and 5, individual provirus data are lacking. Nor can we tell how consistent the distribution of expression values within a TE group is, whether the TEs included solo LTRs (which constitute the majority of all ERVs), the possible contribution of other TFs to expression (with the exception of a brief mention of STAT1).

****Other issues****

Figure 1: A) Log2 fold change of what? TPM values? Needs to be specified.

Why Bonferroni correction? Usually BH q values or other less stringent adjustments are used nowadays. B,C): Z-score of what? Scaled, normalized counts? Scaled TPM values?

Figure 2: B) The blue font color is very difficult to see C) This heatmap should demarcate or separate genes versus TE clades. If that's not possible, then the two should be shown separately.

L191: How many groups (NOT Fam1lies) and how many total elements were examined?

L198: 2B, not C

L205: Did the expressed proviruses have STAT1 sites?

L333: 10 kb is very close. Why was it chosen?

L351-352: Again., correlation is not causation. How do the authors know it's not the other way around?

Figure 4 Title: For "induction" Substitute "correlation"

Panel B: Need to show a scale of the genome region, the orientation of both the gene and the TE, whether it is a solo LTR

5 Panel A: Would benefit from also showing HCs

C: Would be helped by showing adjusted p-values, and also should show examples of non-correlating relationships between these KZNF genes and other TEs.

6 Title: should start with "proposed model for.." or some such.

L 537: Again, how do the alleles segregate in the clusters?

General

In the correlation analyses, I'd like to see adjusted p-values and examples of non-correlated results.

This study presents an in depth analysis of the reverse transcriptome in Elite controllers. It will be of interest to both HIV researchers and those interested in the regulation of the human retrotranscriptome and its consequences

- Describe the nature and significance of the advance (e.g. conceptual, technical, clinical) for the field.

Provides an avenue for future explanation into elite controllers and TE involvement in the phenotype.

- Place the work in the context of the existing literature (provide references, where appropriate).

Does a good job of this, synthesizing other papers regarding TEs and immune control.

- State what audience might be interested in and influenced by the reported findings.

Potential immune regulatory involvement of specific HERV clades.

- Define your field of expertise with a few keywords to help the authors contextualize your point of view.

Indicate if there are any parts of the paper that you do not have sufficient expertise to evaluate.

Authors' response to reviewers

Overview of the revisions

This study was submitted to the Review Commons (RC-2023-02304) and received positive responses from three reviewers. We appreciate that the reviewers found its findings to be "interesting" and "novel", remarking that it "constitutes a useful resource and adds to the growing body of evidence implicating TEs in cis-regulatory control of immune genes". The reviewers highlight that this will be of interest to "both HIV researchers and those interested in the regulation of the human retrotranscriptome and its consequences". They also acknowledge the pioneering nature of the work and its potential to catalyze future studies, providing "an avenue for future explanation into elite controllers and TE involvement in the phenotype" and "a useful source of biomarkers to predict viral kinetics in HIV-infected individuals". We also thank the reviewers for their helpful suggestions to improve the study. Below is a summary of the main changes that resulted from these revisions:

1. We added a new analysis comparing TE expression profiles in EC CD4⁺ T cell

subsets to those in unstimulated and stimulated CD4⁺ T cell subsets from HCs (Figure S3B), in order to assess if the EC retrotranscriptome appears to mimic any specific uninfected T cell type. ECs displayed overall higher TE expression across CD4⁺ T cell subsets but were most similar to HC Th2 progenitor cells.

2. We added a gene set enrichment analysis of TE-gene pairs that are differentially expressed and accessible in ECs compared to HCs to assess functional enrichments in the gene list (Figure 4A). These genes were predominantly involved cellular activation, cytokine production, and immune response regulation, revealing a potential TE-driven branch of the EC immune regulome.
3. We re-clustered the 128 RNA-seq samples from deconvoluted EC CD4⁺ T cells using only locus-level TE expression to compare to the four clusters resolved from gene and TE family expression (Figure S7). We found four distinct EC clusters that aligned well with the original clusters but revealed novel patterns of variable TE expression within families, underscoring the complexity of TE activity and the need for high-resolution analyses.
4. We quantified HIV-1 transcription in all RNA-seq samples and, in ECs, correlated this with ISG expression to assess whether HIV-1 transcription is a driving factor in the observed high expression of ISGs in ECs compared to ART-treated PLWH (Figure S6). Low to undetectable HIV-1 transcript levels in EC CD4⁺ T cells and no correlation with ISG expression confirm that the unique immune gene expression profiles in ECs are not the result of viremia.
5. We assessed enrichment for transcription factor (TF) binding to TE families with increased expression in ECs using public ChIP-seq data for TFs expressed in CD4⁺ T cells, followed by a pairwise-ranked expression correlation analysis between TE families and TFs to identify positively correlated pairs (Figure S8). Together, these analyses identified TFs which may contribute to the increased expression of specific TEs in ECs through transcriptional activation.

In the following pages, we copy the three reviews as we received them (black font) and answer each of their comments point-by-point (red font) with references to new analyses resulting from these revisions.

Reviewer #1

Evidence, reproducibility, and clarity

Singh et al. analyze the expression and putative contribution of TEs in CD4⁺ T cells in HIV elite controllers. Through re-analysis of existing datasets, the authors describe broad differences in expression of TEs in ECs through analysis of RNA-seq and ATAC-seq data and come up with convincing examples where differentially-expressed innate immune genes correlate with increased accessibility of proximal TEs. Overall, the authors' conclusions are appropriately measured, though the manuscript text should be re-organized for clarity and a few further analyses are needed to support the main message of the paper.

Major comments:

The manuscript would benefit from a re-organization of the figures to focus on TEs - in particular, Fig 1B, Fig 2, and Fig 3 reproduce known transcriptional differences between ECs and HCs and serve as quality controls for the authors' computational analysis. Conversely, Supplementary Fig 6 contains very interesting data on KZNF expression and should be included in the main figures.

Authors: Thank you for the suggestion. We agree that Figure S6 should be featured more prominently in the manuscript. Accordingly, we have now incorporated it into the main text as Figure 6. The TE-KZNF correlation plots, previously Figure 5C, have been relocated to this new figure to provide a cohesive presentation of all KZNF-related data within the same figure.

We've chosen to keep Figures 1B, 2, and 3 in their original places. We contend that they provide a foundational view of transcriptional variances in gene expression between patient groups, encompassing both previously identified and novel DEGs, which we believe warrants their placement in the main text. Furthermore, they serve as robust quality control measures for subsequent TE-centric transcriptional analyses. Given that there is no limitation in the number of figures in Genome Biology articles, we think it's adequate to retain them as main figures.

It remains unclear whether differences in TE expression described are specific to ECs or to EC-like CD4⁺ T cell states. As there are plenty of datasets available that compare the transcriptome of naïve, activated, exhausted, and regulatory CD4⁺ T cells, the authors should compare the TE expression patterns observed in ECs to activated CD4⁺ T cells, particularly those with a Th1 and cytotoxic phenotype analogous to those observed in ECs, from healthy donors.

Authors: We thank the reviewer for this constructive suggestion to further study the foundations of HIV-1 elite control. In our initial study, we demonstrate that PBMCs from elite controllers (ECs) exhibit a heightened proportion of activated CD4⁺ T cells compared to PBMCs of healthy controls (HCs) and a heightened proportion of macrophages, naïve CD4⁺ T cells, and NK cells compared to PBMCs of treatment-naïve viremic progressors (VPs) (Figure 2D). Additionally, through clustering analysis of deconvoluted CD4⁺ T cell samples from elite controllers, we ascertain that the clustering

pattern is not predicated on the CD4⁺ T cell subtype (Figure 3B). To further explore the reviewer's inquiry, we compared the TE expression profile of ECs with that of unstimulated and stimulated CD4⁺ T cell subsets from HCs (data source: PMID 31570894), integrated into the revised manuscript as Figure S3B.

"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 (61). 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 (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 be upregulated in an EC-specific way (e.g., LTR12C and LTR7). Together, these findings underscore the unique immune cell composition, transcriptome, and retrotranscriptome of ECs." [pg.13-14, L224-233]

While these observations are interesting, pursuing this question further falls beyond the scope of our study, as we note in the Discussion of the revised manuscript. We believe the reviewer's inquiry pertains to a distinct research question, namely whether the potential for elite control of HIV-1 infection manifests as a detectable phenotype pre-infection within healthy CD4 T cell subsets (i.e., EC-like CD4⁺ T cell states) or is a unique phenotype that emerges solely after HIV-1 infection.

"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 here evidence of common TE signatures between EC CD4⁺ T cells and Th2 progenitors from HCs (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 can serve as predictive variables for whether an individual displays enhanced viral control." [pg. 38, L661-669]

Therefore, while we appreciate the reviewer's suggestion and offer the addition of these preliminary findings, we believe that further investigation would be better suited for future studies specifically designed to address that question. Our manuscript aims to provide insight into the retrotranscriptome dynamics in ECs and their potential implications for natural HIV-1 control.

In Fig 1, the authors demonstrate differential expression of both innate immune genes and TEs, but the link between the two is unclear. Is there any enrichment in differential expression for TEs located proximal to innate immune genes? This type of analysis should be possible using the authors' own software to map TE expression to specific genomic loci.

Authors: Thank you for this excellent question. To answer this inquiry, we used the paired ATAC-seq and RNA-seq datasets for from ECs and HCs (used in Figures 1 and 4) to produce a new list of TE-gene pairs on which we could perform gene set

enrichment analysis, the results of which have been integrated into the revised manuscript as Figure 4A.

“We used paired ATAC-seq – which measures chromatin accessibility – and RNA-seq datasets for ECs (n=4) and HCs (n=4) to create a list of TE-gene pairs where the TE locus and gene show increased accessibility and expression, respectively, in ECs compared to HCs (Table S7, see Methods for details). These loci and genes were paired based on proximity, with a maximum distance of 10kb between the TE locus and the gene’s transcription start site, 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 (Figure 4A). The enrichment for differential accessibility of TE loci near genes involved in these pathways suggests that the distinct TE landscape observed in ECs may contribute significantly to a unique immune regulome in these individuals.” [pg. 21, L355-366]

Thus, we conclude that yes, there is an enrichment for immune-related genes with higher expression in ECs, proximal to differentially accessible TEs. We highlight six of these TE-gene pairs in Figure 4B-C. While we have high confidence in our analyses, future experimental validation is needed to confirm these regulatory relationships.

Optional: In Fig 3, the authors cluster CD4⁺ T cells based on transcriptomic profiles. It would be interesting to re-cluster these samples based on TE expression alone, given the differences in TE expression described in Fig 5.

Authors: Thank you for the suggestion. We agree that it would be valuable to assess how the EC clustering is altered when considering TE expression alone, as opposed to combining gene and TE family expression. To address this, we used the same graph-based k-nearest neighbors method to re-cluster the EC CD4⁺ T cell RNA-seq samples based only on locus-level TE expression, integrated into the revised manuscript as Figure S7.

“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 (Figure S7A), which interestingly appeared even more distinct than those identified by gene and TE family expression (Figure 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 (Figure S7B), indicating high consistency (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 (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 (Figure S7C, Table S8). 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 (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 (Figure S7D). These signatures were heterogeneous across families, with subsets of variable loci from one TE family marking separate clusters (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 L1s 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 sequencing to fully capture the complexity of TE expression.” [pg. 27-28, L460-486]

We believe these findings not only validate the distinct clustering patterns observed but also highlight the potential of locus-level TE analysis to reveal additional layers of retrotranscriptomic diversity in EC CD4⁺ T cells.

Significance

The manuscript by Singh et al. describes for the first time the role of TEs in HIV elite controllers, suggesting that TEs may be co-opted for cis-regulatory function. This study builds off prior work demonstrating that HIV-infected CD4⁺ T cells activate LTR elements that may regulate the expression of interferon-inducible genes, demonstrating that ECs show further upregulation of innate immune genes. While these findings will need to be experimentally validated, this study constitutes a useful resource and adds to the growing body of evidence implicating TEs in cis-regulatory control of immune genes. This study will be of interest to basic scientists interested in genetic mechanisms of HIV control, and if further developed may comprise a useful source of biomarkers to predict viral kinetics in HIV-infected individuals. My expertise is in immunology, TE biology, and viral infection.

Authors: We greatly appreciate this positive evaluation of our manuscript and recognition of its significance in uncovering novel evidence of TE co-option for immune regulatory function in HIV-1 elite control, as well as the suggestion of promising avenues for future research in this field.

Reviewer #2

Evidence, reproducibility and clarity

The authors have re-analyzed published RNA-Seq data from CD4 T cells isolated from HIV elite controllers and reference cohorts, including HIV negative persons, viremic progressors and ART-treated persons. Their main finding is that in some of their comparisons, EC have higher levels of interferon-stimulated genes (ISG), paired with distinct expression patterns of transposable elements. The authors suggest that expression of transposable elements may induce altered expression of ISG, presumably due to immune recognition of TE. They also suggest that reduced expression of KZNF genes, which encode for transcription factors that can suppress TE, may be responsible for enhanced expression of TE. I have the following comments:

1. All data included in this manuscript derive from previously published data. A new dataset, specifically designed to focus on a high-resolution analysis of TE expression, would be better suited to address the proposed questions.

Authors: We agree that a new dataset tailored specifically for high-resolution analysis of TE expression would be optimal for addressing the proposed inquiries, and we emphasize this point in the Discussion of the revised manuscript.

“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 (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.” [pg. 37-38, L649-660]

Regrettably, we do not have access to elite controller samples (which are exceedingly rare), and as such the addition of a novel dataset was not feasible within the scope of this revision. Nevertheless, we assert that the publicly available sequencing data analyzed here is robust and suitable for locus- and family-level TE analysis. All sequencing runs were paired-end and of high depth, ensuring proper alignment to and high coverage of TEs at a locus-specific resolution. Additionally, we use in-house pipelines curated for TE analysis, to optimize the accuracy and quantity of TE-assigned reads (see Methods and [our GitHub repository](#) for more details).

2. As the authors acknowledge, the described investigations are exploratory, and do not allow to draw firm conclusions. Mechanistic experiments are recommended to address the authors' hypotheses.

Authors: We agree and have duly acknowledged throughout the Discussion the exploratory nature of our investigations and the need for future mechanistic experiments to validate our model. Below are passages from the revised manuscript which we've added to emphasize these points.

"These findings underscore the need for future locus-level investigations with high-depth sequencing to fully capture the complexity of TE expression." [pg. 28, L484-486]

"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 levels and accessibility in ECs are indeed boosting the innate immune response and control of HIV-1 in these individuals." [pg. 34, L581-584]

"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. 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 (Figure S8)." [pg. 35, L592-598]

"Overall, our results reinforce the concept that TEs are important players in the human antiviral response (25,93) 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." [pg. 36, L616-621]

"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." [pg. 40, L714-716]

Thus, while we appreciate and agree with the suggestion of experimental validation, we contend that these experiments fall beyond the scope of the present study, which is a computational investigation providing insight into the EC retrotranscriptome and its potential implications for natural HIV-1 control.

3. An important limitation is that virological data of EC are not considered. For example, I believe it is a lot more likely that the upregulation of ISG in EC relates to ongoing low-level viral replication. The authors could analyze cell-associated HIV RNA and DNA levels and determine how they associate with ISG expression.

Authors: Thank you for bringing up this important consideration. It's worth noting that the public datasets used in our study reported undetectable viremia in the EC volunteers (PMIDs 30964004, 29269040, 32848246, 27453467). Nonetheless, we sought to address this limitation and explore the potential association between ISG expression and viremia as recommended by the reviewer. These analyses were integrated into the revised manuscript as Figure S6.

“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 ART-treated PLWH for comparison. In the original studies, all samples were reported to have undetected viremia by blood tests (9,37-39). 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) (37), we detected HIV-1 expression in some subsets (Figure S6A&B). In agreement with the corresponding study (37), 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 (Figure S6A, average TPM < 2). However, we observed no significant association between HIV-1 transcript levels and the expression of interferon-stimulated genes (ISGs) across all EC samples (Figure S6C-D). These analyses confirm that the unique immune gene expression profile observed in ECs is not associated with viremia.” [pg. 18, L307-321]

This added analysis confirms that the increased expression of ISGs in ECs is not correlated with virological transcription and is therefore likely not to be driven by viremia.

4. KZNF genes seem downregulated in EC. Can the authors propose a reason/mechanism for that?

Authors: There is the possibility that KZNF regulatory loops are the cause of their transcriptional downregulation, which has been documented in embryogenesis (PMID 31006620) and cancer (PMID 33087347). We’ve incorporated this hypothesis into the Discussion as an additional consideration for the reader.

“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 ZNF 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 (51) and cancer (115).” [pg. 39-40, L703-709]

While we believe this is a viable hypothesis, it requires further experimentation to confirm the existence of this phenomenon and its impacts in the context of immune cells.

Significance

Overall, I think this is an interesting manuscript that proposes distinct and potentially important mechanisms that may contribute to immune control of HIV. My suggestions to improve the manuscript are complex and cannot be easily addressed through experimental work. I believe a possible option would be to publish the present manuscript without my proposed modifications but highlight the weaknesses of the current paper more clearly; mechanistic studies could then be deferred to a future study.

Authors: We appreciate the reviewer's positive assessment of our manuscript and their recognition of its significance in elucidating novel TE-derived mechanisms that may contribute to natural HIV-1 control. We agree that mechanistic studies are required to test our predictions. As the reviewer suggests, these would be complex experiments that we feel fall beyond the scope of this study. With the additions detailed above in response to the reviewer's point #2, we believe that we have clearly highlighted the limitations of our work and emphasized the need for future experimentation to validate our findings.

Reviewer #3

Evidence, reproducibility, and clarity

Summary: This manuscript presents an analysis of published gene expression (RNA-seq and ATAC-seq) data from a couple of cohorts of HIV-infected elite controllers (EC), as compared to uninfected controls, (HC), virological progressors (VP). The authors report that HIV elite controllers may exhibit 4 distinct patterns of TE (and gene) expression and suggest that TE expression may drive some form of antiviral gene expression. Further, they show that heterogeneous TE expression may be determined by differential KZNF gene activity among the different clusters of elite controllers. These results are very interesting, even though the conclusions are very preliminary. It presents intriguing correlations between expression of certain TE groups of LINES and HERVs, and the clustering into 4 gene expression groups in EC and is a novel finding. That said, correlation is not causation, and the authors need to be more cautious in presenting their highly preliminary model in Figure 6.

Authors: We are grateful for the reviewer's insightful assessment of our manuscript, acknowledging the novelty and interest of our findings regarding TE expression patterns in HIV-1 elite controllers. We also appreciate their constructive feedback regarding the cautious interpretation of preliminary conclusions. In the revised manuscript, we have underscored the exploratory nature of our investigations and the need for future mechanistic experiments to validate our model.

“These findings underscore the need for future locus-level investigations with high-depth sequencing to fully capture the complexity of TE expression.” [pg. 28, L484-486]

“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 levels and accessibility in ECs are indeed boosting the innate immune response and control of HIV-1 in these individuals.” [pg. 34, L581-584]

“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. 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 (Figure S8).” [pg. 35, L592-598]

“Overall, our results reinforce the concept that TEs are important players in the human antiviral response (25,93) 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.” [pg. 36, L615-621]

“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.” [pg. 40, L713-716]

We hope these passages provide sufficient caution and clarity in the presentation of our scientific inquiry.

Major comments:

Overall, although preliminary, as the authors note, the results are interesting and worthy of follow-up. At this point, however, a number of issues arise that need further clarification and analysis before I would consider this study complete.

First, the analyses shown in Figures 3-5 based on data from studies on EC of CD4 cells are apparently motivated by the differential TE expression in total PBMCs shown in Fig 1 and 2. Yet, the TE groups (please don't use taxonomic terms like "subfamily") identified in Fig 2 and Fig 4 are completely different, with no overlap. This discrepancy underscores the possibility that the differential expression observed is, at least in part, due to the differences among the groups or clusters in cell type composition, as seen in Fig 2D and 3B which, themselves, could be a consequence of HIV infection and elite control (which has been shown to involve ongoing, albeit low-level, virus replication). This issue must be addressed.

Authors: Thank you for the suggestion. First, we'd like to clarify that the data used in Figures 1 and 2 were not both derived from PBMCs. Figures 1 and S1 examine the differential expression of TEs in EC CD4⁺ T cells compared to HCs and ART-treated PLWH, respectively. Figure 2 examines differential expression of TEs in EC PBMCs compared to treatment-naïve VPs. Second, regarding Figure 4B-C, the TE loci that we chose to highlight were not based on our results from the PBMC analysis in Figure 2, which is why there is no overlap in the TE families presented. Instead, we selected those TE-gene pairs based on 1) known function of the genes in immunity and/or HIV-1 restriction, 2) known contribution of the TE families to immunity, and 3) differential accessibility and expression of the TEs and genes respectively in ECs compared to HCs. Thus, Figure 4B-C represents select examples that we deemed particularly relevant to the EC phenotype. We have revised the manuscript to better explain the process of TE-gene pair identification and the rationale behind our selection for Figure 4B-C.

“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) (39) to create a list of TE-gene pairs where the TE locus and gene show increased accessibility and expression, respectively, in ECs compared to HCs (Table S7, see Methods for details). These loci and genes were paired based on proximity, with a maximum distance of 10kb between the TE locus and the gene's transcription start site, to increase the likelihood of a direct cis-regulatory influence of the TE over the nearby gene.” [pg. 21, L355-361]

“In Figure 4B & 4C, 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.” [pg. 21, L367-369]

Regarding cell type composition, we acknowledge that the differences observed in the proportion of immune cell subtypes may contribute to the differential expression between ECs, VPs, and HCs (Figures 2D and S3A). However, we provide evidence that cell type composition cannot be the sole driver for the clustering of deconvoluted CD4⁺ T cell RNA-seq samples (Figure 3B and S5D). Cell subtype alone could not explain the observed clustering of EC samples by gene and TE family expression. Clusters 1 and 2, for example, had nearly identical subtype compositions, but were clearly separated on the UMAP (Figures 3A, 3B, and S5D). We remark on this in the Results of the revised manuscript.

“[W]e 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 (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 (Figure 3B). Thus, cell subtype composition could only partially explain the clustering.” [pg. 16, L269-274]

The EC CD4⁺ T cell clusters also had unique gene ontology, gene & TE expression, and TE accessibility profiles (Figures 3C, 3D, 5). Moreover, while we do not have parallel RNA- and ATAC-seq data from similarly deconvoluted CD4⁺ T cells of ECs like those used in the clustering analysis (PMIDs 32848246 & 27453467), the original article from which we sourced the parallel RNA- and ATAC-seq data used in Figures 1 and 4 reported that these samples are predominantly effector memory CD4⁺ T cells (PMID 30964004). If new deconvoluted, multi-omic datasets from ECs become available, we would be interested in further exploring the contribution of cell type composition. However, the current data indicate that it is not a major contributor to the differential TE expression identified in our analyses.

Regarding the impact of ongoing HIV-1 replication upon the unique expression patterns in the EC participants, it's worth noting that the public datasets used in our study reported undetectable viremia in the EC volunteers (PMIDs 30964004, 29269040, 32848246, 27453467). Nonetheless, we sought to address this by quantifying HIV-1 transcription and exploring its potential association with interferon-stimulated gene (ISG) expression, a group of genes that we know would be reactive to active viremia. These analyses were integrated into the revised manuscript as Figure S6.

“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 ART-treated PLWH for comparison. In the original studies, all samples were reported to have undetected viremia by blood tests (9,37-39). 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) (37), we detected HIV-1 expression in some subsets (Figure S6A&B). In agreement with the corresponding study (37), 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 (Figure S6A, average TPM < 2). However, we observed no significant association between HIV-

1 transcript levels and the expression of interferon-stimulated genes (ISGs) across all EC samples (Figure S6C-D). These analyses confirm that the unique immune gene expression profile observed in ECs is not associated with viremia.” [pg. 18, L307-321]

Based on these results, we have concluded that the differential expression of genes and TEs in the EC clusters are not a consequence of low-level viral transcription in ECs.

Finally, a remark on TE nomenclature: The reviewer suggests that we use the term “TE groups” as opposed to taxonomic terms such as TE subfamily or TE family. We respectfully disagree. This nomenclature of TEs has been well defined (PMIDs 26612867, 26612867, 17984973) and is widely used in TE literature. Throughout the manuscript, we have conformed to the nomenclature used to annotate the human genome. One can debate the way TE families and subfamilies have been classified in Dfam (the database through which repetitive elements in the human genome have been annotated), but it is outside the scope of this study to revisit that nomenclature.

Similarly, of the 12 DE TE groups in EC in Fig 5A, only 3 overlap with the 16 in EC Fig S1.

Authors: This is correct, but we don’t believe it’s concerning. In Figure 5A, we are comparing the expression of TE families between separate EC clusters. In Figure S1, we are comparing the expression of TE families in ECs compared to ART-treated PLWH. These are fundamentally different comparisons and thus the differences in the identified DE-TEs between the two figures reflect the distinct biological contexts being investigated in each analysis.

Second, the introduction points out the strongly supported association between elite control and immunogenetic determinants, most notably specific HLA-B types, but also innate immunity factors. This cries out for inclusion of these factors in the analyses of this manuscript, in the format of Figure S4, for example, but none is to be found. The relevant genotypes are likely available in the metadata in the references cited, but, if not, could be inferred from the RNA-seq data.

Authors: Thank you for the recommendation. While our project’s primary focus is on the transcriptomic and epigenomic signatures, we agree that studying the HLA-B genotypes of all EC participants could provide valuable context for understanding the clustering of elite controllers. To explore this, we inferred the HLA-B alleles in each EC participant whose RNA-seq data was included in the clustering analysis, utilizing the arcasHLA tool (PMID: 31173059) on the total CD4⁺ T cell samples. We then validated these inferred HLA-B alleles against the available metadata from one of the source studies (PMID 27453467) and found that they matched for all participants. This strengthened our confidence in the accuracy of the HLA-B genotype inferences for the other samples where comprehensive HLA-B data was not provided.

In order to assess how these protective HLA-B alleles segregated between the four EC clusters derived from gene and TE family expression, we chose to visualize three of the most common alleles associated with HIV-1 elite control: HLA-B*27:03, *57:01, and *57:03 (PMIDs 30964004, 25119688, 21051598) (Figure R1).

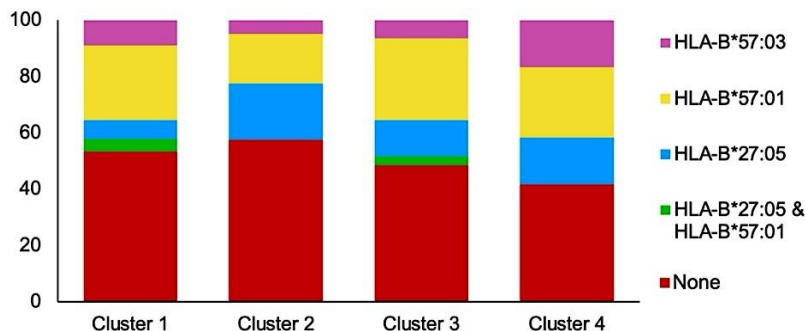


Figure R1. Protective HLA-B alleles in EC clusters. Boxplot displaying the proportion of EC CD4⁺ T cell samples with a given HLA-B genotype in each cluster. Colored by HLA-B genotype.

Data sources: [Jiang et al., 2020](#) and [Boritz et al., 2016](#).

Our analysis revealed that these major protective alleles were not significantly overrepresented in any particular cluster. Consequently, we believe that HLA-B genotype does not have a major impact on the clustering observed in Figure 3.

It would also be very useful to present the KZNF data in Figure 5 the same way, since, looking at Fig 5C, the correlation of high and low KZNF expression, while clearly correlated with a that of few groups of elements, with clustering into specific groups does not appear to be well supported.

Authors: Thank you for the insightful suggestion. While the KZNF genes are included in the gene set used for the clustering analysis in Figure 3, we agree that clustering based solely on KZNF expression and displaying it as we have in Figures 3A and S5 could provide valuable insights. However, when we attempted to cluster the EC RNA-seq samples using only KZNF expression data, we were limited by the relatively low number of KZNF genes that showed sufficient variability across samples ($n = 120$). For robust statistical power, we require at least 200 features to reliably cluster the 128 EC CD4⁺ T cell samples. We believe this limitation does not diminish the relevance of KZNFs in the observed clustering patterns but rather highlights the nuanced role each KZNF plays in the regulation of the transcriptome. Each individual KZNF is responsible for the regulation of hundreds to thousands of TE loci (PMID 37730438). Thus, while a clustering approach based solely on KZNF expression may not be feasible, the integral role of KZNFs in modulating the transcriptome through TE regulation remains evident and supports their inclusion in Figure 6 of the revised manuscript.

In general, other than the cell type composition differences, there is no presentation of evidence for any biologically important feature associated with the clusters found.

Authors: We agree that the root cause of the transcriptomic differences between the EC clusters is hard to pin down but we do identify several distinctive features of the clusters that we believe are biologically significant. First, having extracted the lists of genes whose differential expression defined the four EC clusters, gene set enrichment analysis revealed that the clusters were functionally distinct, each characterized by a unique list of top GO terms (Figure 3C). Second, we provide evidence that KZNFs expressed in CD4⁺ T cells significantly bind to the candidate TE families whose expression defines each of these clusters (Figure 6D) and have significantly decreased expression in ECs compared to VPs (Figure 6C). This is corroborated by pairwise correlation analysis that revealed cluster-specific anticorrelation patterns between these KZNFs and their target TEs (Figure 6A). We present this data in support of our

hypothesized KZNF-based mechanism for TE co-option in viral immunity. We do not yet have data indicative of the mechanism by which KZNF expression is in turn regulated. However, we speculate that negative feedback loops may be contributing to changes in KZNF expression.

“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 ZNF 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 (51) and cancer (115).” [pg. 39-40, L703-709]

Overall, our study presents preliminary evidence that the four EC clusters derived from gene & TE family expression may be distinguished by complex interplay of activators (Figure S8) and repressors (Figure 6) altering the activity of infection-responsive TE families to co-opt specific elements for immune regulatory function. While not yet validated in an experimental setting, we believe these results are of biological significance.

Third, the figures present values that have been very heavily analyzed, and it is difficult to impossible to infer what the underlying data look like. For example, with the exception of a few selected examples in Figs 4 and 5, individual provirus data are lacking. Nor can we tell how consistent the distribution of expression values within a TE group is, whether the TEs included solo LTRs (which constitute the majority of all ERVs), the possible contribution of other TFs to expression (with the exception of a brief mention of STAT1).

Authors: We respectfully disagree that the values presented in our figures are heavily analyzed. As this manuscript represents the first investigation of TEs’ role in HIV-1 elite control, we believe the most reasonable initial approach was to compile and visualize the data at the family level, rather than at the level of individual loci, which is harder to interpret due to mapping issues, commonly low transcription, and often idiosyncratic behavior of individual loci. Nonetheless, we did not limit our analysis to full-length HERVs (proviruses) and thus retain all solo LTR data in our analyses. This was added to the Methods of the revised manuscript.

“To facilitate comprehensive expression quantification, we curated a reference transcriptome by combining gene, TE, 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, retaining LINE and HERV loci, including all solo LTRs. We also removed any loci within gene exons/UTRs. 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.” [pg. 54, L851-860]

For the sake of transparency, all relevant details on sequencing data analysis and the corresponding scripts are available in the Methods and [our GitHub repository](#).

Additionally, while most of our figures make comparisons at the family level, we do visualize multiple TE loci (Figure 4C) and provide a list of putative locus-level TE-gene pairs from which those shown in Figure 4C were selected (Table S7). In our revisions, we also re-clustered the 128 EC CD4⁺ T cell RNA-seq samples based only on locus-level TE expression, using the same graph-based k-nearest neighbors method as in Figure 3. The results of this new analysis have been integrated into the revised manuscript as Figure S7.

“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 (Figure S7A), which interestingly appeared even more distinct than those identified by gene and TE family expression (Figure 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 (Figure S7B), indicating high consistency (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 (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 (Figure S7C, Table S8). 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 (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 (Figure S7D). These signatures were heterogeneous across families, with subsets of variable loci from one TE family marking separate clusters (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 L1s 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 sequencing to fully capture the complexity of TE expression.” [pg. 27-28, L460-486]

With this addition, we include significantly more data analyzed at the locus level, which we believe not only validate the distinct clustering observed in Figure 3, but also underscore the potential for locus resolution analysis to reveal additional layers of retrotranscriptomic diversity in EC CD4⁺ T cells.

Finally, we agree with the reviewer that TFs other than STAT1 may contribute to the observed changes in TE expression. To investigate this, we analyzed several TFs expressed in CD4⁺ T cells and, for TFs enriched over TEs of interest, subsequently examined the correlation between TF and target TE expression in the deconvoluted EC CD4⁺ T cell samples used for the clustering. The results of this analysis have been integrated into the revised manuscript at Figure S8.

“In addition to KZNF repressors, transcriptional activators may also drive the differential expression of specific TE families across ECs (83). To investigate this, we focused on transcription factors (TFs) expressed in CD4+ T cells and mined ChIP-seq data from the ENCODE Consortium (84) to identify TFs with binding enrichment to TE families of interest, selected for their elevated, cluster-specific expression in ECs (highlighted in Figures 4, 5, and 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 (Figure 3) (9,37). We observed several significant positive correlations between TF and TE expression across ECs (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.” [pg. 30, L515-525]

This evidence supports the reviewer’s suggestion that other TFs may be contributing to the unique EC retrotranscriptome we profile in this study. These added analyses, mimicking those conducted for KZNFs in Figure 6B & 6D, demonstrate that transcriptional activators may indeed play a crucial role in shaping the TE landscape in ECs.

Other issues

Figure 1:

A) Log2 fold change of what? TPM values? Needs to be specified.

Authors: Thank you for pointing out this ambiguity. The log2-transformed fold change values plotted in Figure 1A refer to DESeq2-normalized expression. They were extracted from the results of the DESeq2 pipeline, which we applied to the raw count expression matrix (see our Methods for more details). Following your suggestion, we have clarified this point in the figure legend in the revised manuscript.

“Total detected genes and TE loci are plotted by log2-transformed fold change of DESeq2-normalized counts (EC vs. HC).” [pg. 10, L161-162]

We have similarly made these changes to any figure legend which was ambiguous in its description of the expression data.

Why Bonferroni correction? Usually BH q values or other less stringent adjustments are used nowadays.

Authors: In our analysis, we opted for the Bonferroni correction due to its well-established reliability and stringent control of the family-wise error rate when conducting multiple tests. Given the exploratory nature of our investigation and the desire to minimize the risk of false positive findings, we chose to employ this traditional correction method within our analytical pipelines.

B,C): Z-score of what? Scaled, normalized counts? Scaled TPM values?

Authors: Thank you again for highlighting this point of uncertainty. We now clarify this in the figure legend in the revised manuscript.

“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). 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.” [pg. 11, L165-169]

Figure 2:

B) The blue font color is very difficult to see.

Authors: We have changed the blue font color to make it more easily distinguishable from the black.

C) This heatmap should demarcate or separate genes versus TE clades. If that's not possible, then the two should be shown separately.

Authors: We appreciate your suggestion regarding the heatmap presentation. While we understand the rationale for demarcating genes versus TE clades, we have chosen to retain the original figure layout. In this analysis, TEs were analyzed simultaneously with genes. The order in which they are shown was obtained by default clustering of the expression matrix using the hclust function. We chose to present them together and in this order to provide a comprehensive visualization of the differential expression patterns between the two groups and highlight the homogenous nature of gene and TE expression across VPs.

L191: How many groups (NOT families) and how many total elements were examined?

Authors: We begin with the RepeatMasker annotation of the hg19 assembly and filter out the SINE elements, DNA transposons, simple repeats, and all loci within gene exons/UTRs. These details are provided in the Methods of the revised manuscript, as was quoted above.

“To facilitate comprehensive expression quantification, we curated a reference transcriptome by combining gene, TE, 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, retaining LINE and HERV loci, including all solo LTRs. We also removed any loci within gene exons/UTRs. 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.” [pg. 54, L851-860]

In total, our analyses examine 1,104,828 loci from 603 TE groups (which we refer to as families). We apologize if this figure is not accurate to a separate classification of TEs into groups, rather than families. Any such method of grouping TEs is unfamiliar to us and outside of the Dfam annotation.

L198: 2B, not C

Authors: Thank you for catching this. The figures labelled were swapped in error and have been changed to reflect in Figure 2 to match the in-text references.

L205: Did the expressed proviruses have STAT1 sites?

Authors: Thank you for your question. The identification of LTR13's increased expression in ECs compared to VPs was the result of a family level analysis which considered expression additively across the LTR13 loci in our annotation. To answer your question, we analyzed STAT1 ChIP-seq data from the [ENCODE Consortium](#) to characterize which LTR13 loci were bound by STAT1 (corroborated by motif prediction calls). We then integrated the EC RNA-seq data and found that the expressed LTR13 proviruses significantly overrepresented those with bound STAT1 sites (Figure R2).

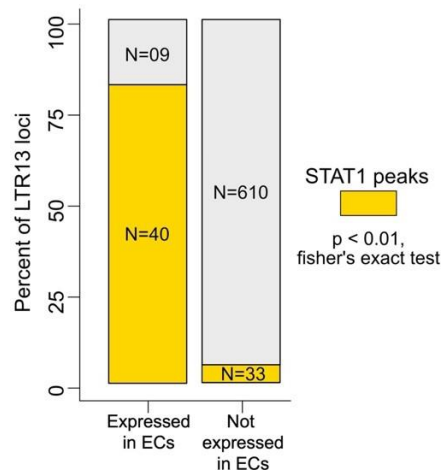


Figure R2. STAT1 binding enrichment in EC-expressed LTR13 loci. Boxplot displaying the proportion of LTR13 loci bound by STAT1, separated column-wise by their expression in ECs. Significance between the two groups was determined by Fisher's exact test. Data Sources: [Zhang et al., 2018](#) and the [ENCODE Consortium](#).

These data suggest that STAT1 binding may play a critical role in the transcriptional regulation of LTR13 in ECs, contributing to their differential expression profile. Further exploration into the contribution of activating, immune-related TFs is explored in Figure S8 in the revised manuscript.

L333: 10 kb is very close. Why was it chosen?

Authors: We chose 10 kb as our cutoff for selection because it allowed for very high confidence in the TE loci's cis-regulatory capacity over the nearby genes. For transparency, we have made this clearer in the Results text of the revised manuscript.

"These loci and genes were paired based on proximity, with a maximum distance of 10kb between the TE locus and the gene's transcription start site, to increase the likelihood of a direct *cis*-regulatory influence of the TE over the nearby gene." [pg. 21, L358-361].

However, if desired, a less stringent cutoff could also be used with relative confidence (e.g., 50 kb).

L351-352: Again, correlation is not causation. How do the authors know it's not the other way around?

Authors: The candidates that we chose to display in Figure 4 (the figure to which these lines refers) are from MER41, ERV3-16, and LTR12C. Our lab and others have shown that these specific loci or other loci in these TE families are capable of regulating neighboring genes' expression, with specific evidence in the context of immunity (PMID

Smitha, Ed, APOBEC, etc.). Based on this knowledge, we believe that it's most likely that TE-derived regulatory sequences are the cause of the increased restriction factor expression, rather than TE accessibility being a consequence of the transcriptional activation of the neighboring genes. However, we recognize that these results are correlative, as the reviewer notes, and we emphasize this in the revised manuscript. Most notably:

"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." [pg. 36, L618-621]

Figure 4

B) Need to show a scale of the genome region, the orientation of both the gene and the TE, whether it is a solo LTR

Authors: Thank you for the suggestion. Genomic scale and orientation have been added to Figure 4C. All loci visualized were solo LTRs, save for *HCP5*, which is a lncRNA derived from a full-length ERV3 element.

Figure 5

A) Would benefit from also showing HCs

Authors: Thank you for the recommendation. The RNA-seq datasets used in this analysis do not include HC samples. Additionally, this analysis is meant to highlight differences in TE expression between the four EC clusters. Thus, we have chosen to keep Figure 5A as it appears in the original manuscript.

C) Would be helped by showing adjusted p-values, and also should show examples of non-correlating relationships between these KZNF genes and other TEs.

Authors: Thank you for the suggestion. All correlation analyses had adjusted p-values below 0.01, derived from `corr.test` in R. We've added this to the figure legends of Figure 6B [pg. 32, L537] and S8B [pg. 53, L836]. However, we have chosen not to integrate non-correlating examples into the revised manuscript for the sake of space.

Figure 6

Title: should start with "proposed model for.." or some such.

Authors: Thank you for the suggestion. The title has been changed to "Proposed model for the interplay of KZNFs and TEs regulating proximal antiviral gene expression in elite controllers of HIV-1" in the revised manuscript [pg. 34, L578-579].

L 537: Again, how do the alleles segregate in the clusters?

Authors: This question has been addressed in response to an earlier comment.

Generally, in the correlation analyses, I'd like to see adjusted p-values and examples of non-correlated results.

Authors: Thank you for the suggestion. As mentioned above, all correlation analyses have been annotated with the adjusted p-value threshold. Additionally, below we've included examples of non-correlated results from two analyses. First, we show a TE-gene pair whose increased TE accessibility in HCs compared to ECs does not correlate with increased expression of the proximal gene (Figure R3). Notably, this gene does not play a role in HIV-1 infection response. Here, we show that genes with proximal (<10kb from TSS), differentially accessible TEs do not always correlate with increased gene expression.

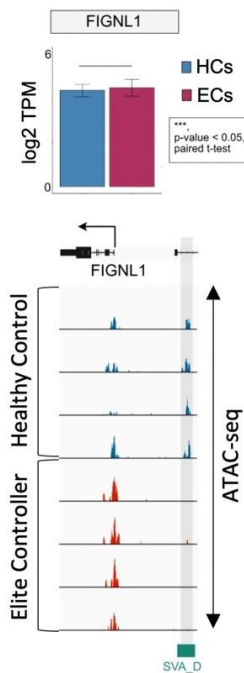


Figure R3. Non-correlated TE-gene pair.

Top: Barplots of FIGNL1 expression (in log2-transformed TPM) in HCs (n=5) and ECs (n=4). P-value is calculated by paired student t-test.

Bottom: IGV of normalized ATAC-seq signal around FIGNL1 between HC (n=4) and EC (n=4) CD4⁺ T cells. ATAC-seq peaks of interest are shaded in light gray. The proximal TE integrants are shown below the IGV graph. RNA- & ATAC-seq data source: [Gonzalo-Gil et al., 2019](#).

Second, we show the pairwise correlation and linear regression results of L1PA6 and ZNF2 (Figure R4). ZNF2 is one of the KZNFs highlighted in Figure 6 for its low expression in ECs, anticorrelated to its repressive target LTR12C. On the other hand, L1PA6 is active in ECs, with variably high expression across samples. ZNF2 ChIP-exo revealed that ZNF2 has no capacity to bind to L1PA6 loci (adj. p-value = 1; PMID 37730438). Thus, even though both genes are variable across samples, we observe no significant (anti)correlation between the two variables ($\rho = 0.051$ & p-value = 0.866).

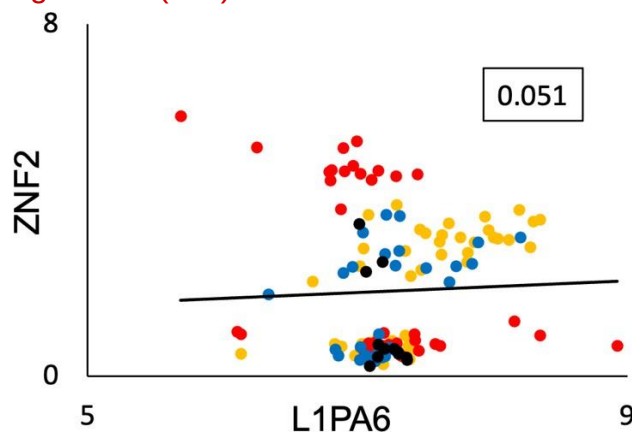


Figure R4. Correlation analysis between ZNF2 and L1PA6 expression. Scatter plot showing scaled, log2-transformed TPM in the CD4⁺ T cell subtype EC RNA-seq data, cluster classified in Figure 3A. Linear regression analysis (black line) indicates the correlation ZNF2 and L1PA6. The rho value is obtained from pairwise- ranked correlation analysis. Data sources: [Jiang et al., 2020](#) and [Boritz et al., 2016](#).

While we have not integrated these results into the revised manuscript for the sake of space, we hope that the provided examples satisfactorily demonstrate the presence of non-correlated results in our analyses, further reinforcing the specificity and robustness of our significant findings.

Significance:

This study presents an in-depth analysis of the reverse transcriptome in Elite controllers. It will be of interest to both HIV researchers and those interested in the regulation of the human retrotranscriptome and its consequences.

Provides an avenue for future explanation into elite controllers and TE involvement in the phenotype.

Does a good job of placing the work in the context of existing lit, synthesizing other papers regarding TEs and immune control.

Potential immune regulatory involvement of specific HERV clades.

Authors: We'd like to thank the reviewer for their encouraging feedback. We're pleased that they found our analysis of the EC retrotranscriptome to be of broad interest and appreciate their recognition of our efforts to synthesize existing literature, contextualizing our findings within the broader field. We agree that our study opens new avenues for exploring the role of TEs, particularly specific HERV clades, in not only the EC phenotype but immune regulation as a whole.

Second round of review

Reviewer 1

This manuscript by Singh, Leddy et al. analyses the expression of TEs in CD4+ T cells in HIV elite controllers, describing broad dysregulation of TE families and suggesting that TEs may be co-opted for cis-regulatory function. Mechanisms of TE dysregulation through KZNF expression and innate immune gene expression through increased accessibility of neighbouring TEs are implied rather than experimentally demonstrated, but the conclusions drawn are appropriately measured. I have a few minor suggestions that would improve the clarity of the paper; if such analyses are not included in the paper they should at least be discussed in the text.

1. The TE expression differences shown in Fig 1C are shown at the family level. This analysis (or a subset of TEs, such as the STAT1-inducible ones) should also be conducted at the locus level in order to differentiate broad epigenetic deregulation from selective re-expression of certain key TEs as implied in the following figures.
2. The CD4 clustering data largely reproduces known transcriptional phenotypes. Does clustering solely based on TE expression reproduce these transcriptional clusters, or do they simply reflect the transcriptional state of the cell? The reverse analysis is shown in Fig 5, but it would be interesting to see the clustering based on TE expression alone.
3. ATACseq data is shown for selected genes in Fig 4; it would be useful to see RNAseq pileups for those samples in order to see whether the highlighted TEs drive alternative promoter/splicing.

Reviewer 3

In this revised manuscript, the authors formulate the hypothesis that differential expression of endogenous retroelements may be of importance to the phenomenon of HIV elite control (EC). For this reason, they reanalyze several existing datasets to look for correlations between EC and expression and related chromatin structure of both LTR and non LTR elements and uncover some interesting relationship patterns, in particular the clustering of expression of groups of elements, which correlates well with reduced expression of likely suppressive KZNF genes.

While the authors have fixed many of the issues raised in my prior review, there are still a some that need to be addressed.

First, as the authors as the authors themselves point out, conclusions drawn the meta-analysis they present must be considered as provisional until tested by direct experimentation.

Second, there appears to be an unstated assumption that any correlation seen relative to "controls" must be somehow causally involved in the EC phenotype. On the contrary, there are several other possibilities that must also be discussed.

1. It is equally possible that the correlations observed are a consequence of elite control. By far the most likely mechanism for EC, not discussed in the manuscript, is a chance match between a specific MHC type and a virally coded epitope (or epitopes) that the virus cannot readily escape by mutation, leading to an especially potent CTL response capable of reducing virus replication 4 or more orders of magnitude (but, unlike with cART, not stopping it altogether). For example, the different pattern of HIV integration in ECs can only be due to potent selection, not integration site targeting. Given the likely importance of the CTL response to EC, the authors are looking at the wrong cells.

2. The CD4 cells are not only important as targets for HIV infection, they are also an important part of the immune response against it. How much of the difference observed between EC and VP is due to the substantial difference in the overall immune response and generalized immune activation due to the huge difference in HIV replication (and possibly opportunistic infection) in the two cases? An additional important control that might shed some light on this issue would be to analyze a dataset from individuals undergoing active or chronic infection with other agents.

3. The possible background due to variation in gene expression from one individual to another is not well accounted for. Rather than being somehow related to EC, it might well reflect natural genetic or epigenetic variability. This issue is particularly important in the interesting analysis presented in Figure 3, in which the authors identify 4 well separated clusters based on differences among ECs in gene and TE expression. The interpretation that these clusters are an EC specific cannot be supported without a similar analysis using data from a similar number of HC donors.

In sum, Additional control analyses would greatly help to better tease out effects related specifically to elite control and those due to other immune responses and to background individual variation.

Also, the sophisticated discussion of TE element expression needs to be accompanied by an equally sophisticated discussion of elite control.

Other issues:

Lines 76-79. This observation can only be explained by selection against cells with intact proviruses It makes no sense to infer the that intact and defective proviruses differ in their integration targeting.

Figure 1 legend and elsewhere. It is nice that the authors supply references for the datasets used, but PLEASE give them numbers so the reader can find them in the list.

Line 179-180 The possible effect of ART on the TE expression profile should be considered.

Figure 3 legend. Please tell us up front you are looking at gene expression.

Line 359 Please tell the reader why you chose 10 kb.

Line 382 "...some HIV restriction..."

Figure 4. The light grey is invisible when printed.

Line 602. Is there any evidence for excessive C-T hypermutation in EC proviruses?

Figure 7. Must include the role of the CTL response to make any sense.

Lines 636-639. Again, integration site differences can only be sensibly explained as a consequence, not a cause, of elite control.

Authors' response to reviewers

Response to Reviewers

From Editor, Juergen Schmitz: At some point in the text, it would be helpful to elaborate on the distinction between transcribed TE elements and fully transpositionally active elements. Additionally, a brief note explaining why certain retrotranspositionally active Alu elements were absent from your proposed regulatory network would help to understand the influence of different TE types.

AUTHORS: Thank you for these suggestions. In the Background, we expanded the text to further explain that very few TEs in the human genome remain capable of transposition, but many do retain the capacity to be transcribed and exert regulatory activities [L89-96].

Regarding Alu elements, we added the following explanation in the Methods to clarify why Alus were excluded from our analysis. "The latter [Alus] 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." [L851-861].

Reviewer #1: This manuscript by Singh, Leddy et al. analyses the expression of TEs in CD4+ T cells in HIV elite controllers, describing broad dysregulation of TE families and suggesting that TEs may be co-opted for cis-regulatory function. Mechanisms of TE dysregulation through KZNF expression and innate immune gene expression through increased accessibility of neighbouring TEs are implied rather than experimentally demonstrated, but the conclusions drawn are appropriately measured. I have a few minor suggestions that would improve the clarity of the paper; if such analyses are not included in the paper they should at least be discussed in the text.

1. The TE expression differences shown in Fig 1C are shown at the family level. This analysis (or a subset of TEs, such as the STAT1-inducible ones) should also be conducted at the locus level in order to differentiate broad epigenetic deregulation from selective re-expression of certain key TEs as implied in the following figures.

AUTHORS: We thank the reviewer for this suggestion. Locus-level analyses such as this were discussed in our previous revisions with the addition of Figure S8, in which we found that TEs with elevated, cluster-specific expression in elite controller (EC) CD4+ T cells were enriched for binding of multiple immune-related transcription factors, reinforcing these differentially expressed TEs' (DE-TEs) potential relevance to antiviral immunity. Additionally, we provided examples of

specific loci of interest that are differentially expressed and accessible between ECs and healthy controls (HCs) later in the manuscript (Figure 4),

including a list of specific TE locus-gene pairs that meet our criteria for potential regulatory activity heightened in ECs (Table S7). To further address the reviewer's suggestion, we repeated the analysis from Figure 1C at the locus-level, identifying differentially expressed TE loci in EC CD4+ T cells compared to HCs [PMID 30964004]. This analysis revealed 870 DE-TE loci (**Figure R1A**). These loci predominantly belonged to TE families such as THE1B, L1PA2, LTR12C, and SVA_D (**Figure R1B**), whose differential activity in ECs was highlighted in our manuscript's initial figures (Figures 5, 6, S1, S7, S8). Notably, the TE loci with higher expression in ECs were significantly enriched for binding motifs of transcription factors NFKB1, STAT1, FOXP2, and NFYB, many of which are key regulators in antiviral responses identified in our previous analyses (Figure S8). This underscores the potential role of TEs in shaping EC transcriptomes by providing binding sites for immune-responsive TFs.

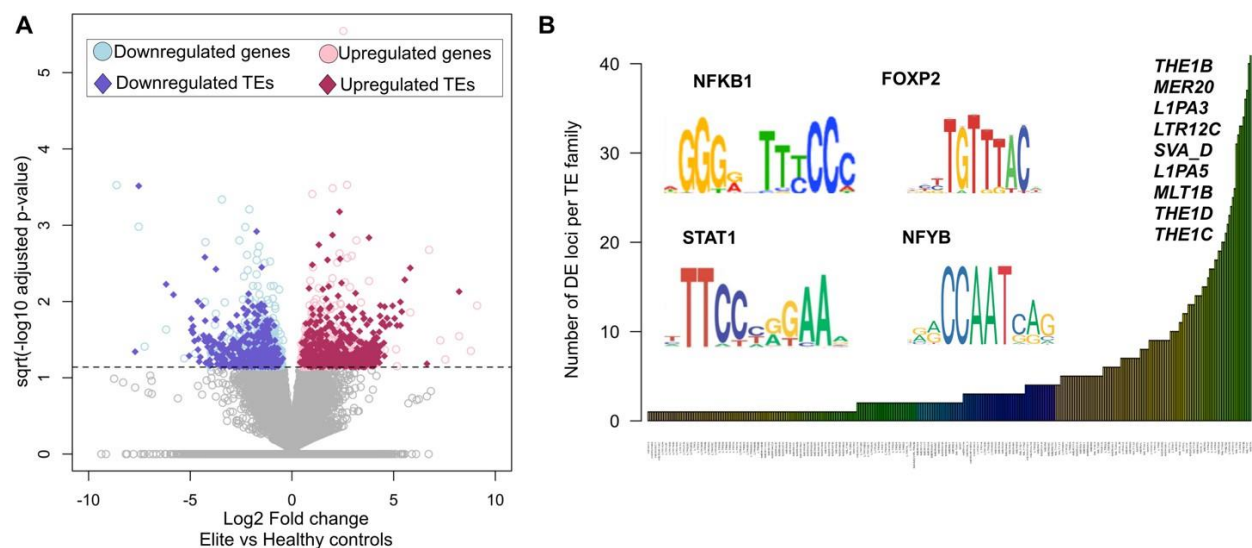


Figure R1. A. Volcano plot illustrating the differentially expressed genes and TE loci between EC (n = 4) and HC (n = 5) CD4+ T cells [PMID 30964004]. The horizontal dashed line indicates the significance cutoff (FDR = 0.05). Coloration is based on increased or decreased expression of genes (pink and light blue, respectively) and TEs (maroon and purple, respectively). **B.** Bar plot showing the number of differentially expressed TE loci per TE family from (A). TE families with the most DE-TEs are listed in bold. Enriched transcription factor (TF) binding motifs are also shown within the plot.

In summary, this additional analysis further substantiates our hypothesis that immune-related, TF motif-enriched TEs contribute significantly to the unique transcriptomic landscape of ECs. While this aligns with our study's conclusions, we have chosen not to integrate it into the revised manuscript because we believe that these findings are redundant with the previously cited figures and tables present in the manuscript.

Additionally, we believe that incorporating locus-level data for each comparative analysis in our study would require extensive effort (and deeper sequencing) to unravel that is beyond the scope of this work. We previously underscored the need for further locus-level investigations into TE expression with high-depth sequencing in the manuscript [L489-491].

2. The CD4 clustering data largely reproduces known transcriptional phenotypes. Does clustering solely based on TE expression reproduce these transcriptional clusters, or do they simply reflect the transcriptional state of the cell? The reverse analysis is shown in Fig 5, but it would be interesting to see the clustering based on TE expression alone. **AUTHORS:** Although all four EC clusters were heterogeneous in terms of CD4+ T cell subset composition, we agree that clustering based on TE expression alone is an important addition to our manuscript to ensure their separation was not driven solely by gene expression. In our previous revisions, we re-clustered the EC CD4+ T cell RNA-seq samples used in Figure 3, based solely on TE locus-level expression. The results of this clustering were previously incorporated as Figure S7 and remain in the final revised version of the manuscript [L465-491].

3. ATACseq data is shown for selected genes in Fig 4; it would be useful to see RNAseq pileups for those samples in order to see whether the highlighted TEs drive alternative promoter/splicing.

AUTHORS: We appreciate the reviewer's suggestion to include RNA-seq pileups to assess whether the highlighted TEs drive alternative promoter usage or splicing. Of the six examples in Figure 4, GBP2, GBP5, and HCP5 are already known to involve TE-derived promoters. GBP2 uses an alternative TSS within its nearby LTR12C locus upon HIV-1 infection, resulting in a novel transcript with an additional exon [PMID 33021676]. On the other hand, GBP5's canonical TSS originates from its nearby LTR12C locus, which is upregulated upon HIV-1 infection, driving gene expression [PMID 33021676]. Finally, HCP5 is a lncRNA driven by an ERV3-16A3 element and its associated LTR16 [PMID 31137555]. Our analyses revealed greater activity of these elements and genes in elite controllers compared to non-controllers. To validate that these transcripts are TE-derived, and to investigate the remaining three examples in Figure 4 (along with the TE-gene pairs in Table S7), high-resolution investigation of alternative promoter and splicing activity would be better served by long-read RNA-seq analysis, which we plan to use in future studies. Thus, given the existing evidence for TE-driven alternative promoters for three of these genes and our plans for more detailed analyses of others in a separate manuscript, we have opted not to include short-read RNA-seq pileups in the current figure. Instead, we have highlighted in the Discussion the importance of exploring non-canonical promoter usage and splicing events in ECs in the future [L623-625].

Reviewer #2: In this revised manuscript, the authors formulate the hypothesis that differential expression of endogenous retroelements may be of importance to the phenomenon of HIV elite control (EC). For this reason, they reanalyze several existing datasets to look for correlations between EC and expression and related chromatin structure of both LTR and non LTR elements and uncover some interesting relationship patterns, in particular the clustering of expression of groups of elements, which correlates well with reduced expression of likely suppressive KZNF genes.

While the authors have fixed many of the issues raised in my prior review, there are still a some that need to be addressed.

First, as the authors as the authors themselves point out, conclusions drawn the meta- analysis they present must be considered as provisional until tested by direct experimentation.

Second, there appears to be an unstated assumption that any correlation seen relative to "controls" must be somehow causally involved in the EC phenotype. On the contrary, there are several other possibilities that must also be discussed.

1. It is equally possible that the correlations observed are a consequence of elite control. By far the most likely mechanism for EC, not discussed in the manuscript, is a chance match between a specific MHC type and a virally coded epitope (or epitopes) that the virus cannot readily escape by mutation, leading to an especially potent CTL response capable of reducing virus replication 4 or more orders of magnitude (but, unlike with cART, not stopping it altogether). For example, the different pattern of HIV integration in ECs can only be due to potent selection, not integration site targeting. Given the likely importance of the CTL response to EC, the authors are looking at the wrong cells.

AUTHORS: We agree that the relationship between observed traits in ECs can reflect either causative or consequential mechanisms. This is a subject of ongoing discussion in the EC literature, as discussed for example in the review "Immune Responses in Controllers of HIV Infection" [PMID 37827174]. While specific MHC-viral epitope interactions leading to potent CTL responses are well-documented as a major factor in the EC phenotype, our work focuses on complementary mechanisms involving CD4+ T cells which can further delay or prevent the onset of immunodeficiency. Many excellent studies have investigated the important role of CD8+ T cells in control, but this is not the focus of our work, and we believe that an investigation of the HIV-targeted CD4+ T cells as major players in immune regulation and viral restriction offers an interesting perspective on EC biology. To ensure the dimensions of the EC phenotype mentioned by the reviewer are sufficiently referenced in our work, we added further detail on the MHC and CD8+ T cell dimensions to the Background of the revised manuscript [L67-76]. We also revised the Discussion to clarify the complementary focus of our study [L644-649].

2. The CD4 cells are not only important as targets for HIV infection, they are also an important part of the immune response against it. How much of the difference observed between EC and VP is due to the substantial difference in the overall immune response and generalized immune activation due to the huge difference in HIV replication (and possibly opportunistic infection) in the two cases? An additional important control that might shed some light on this issue would be to analyze a dataset from individuals undergoing active or chronic infection with other agents.

AUTHORS: We agree that there are major differences between ECs and VPs in both overall immune activation and HIV replication which may confound observed patterns if not properly controlled for. This limitation is inherent to many observational studies of EC cohorts, especially given sample size constraints. We began to address this concern in the previous revisions with the addition of Figure S6 [L317-331], in which we quantify HIV-1 expression in the EC RNA-seq samples and correlate it with ISG expression as a measure of

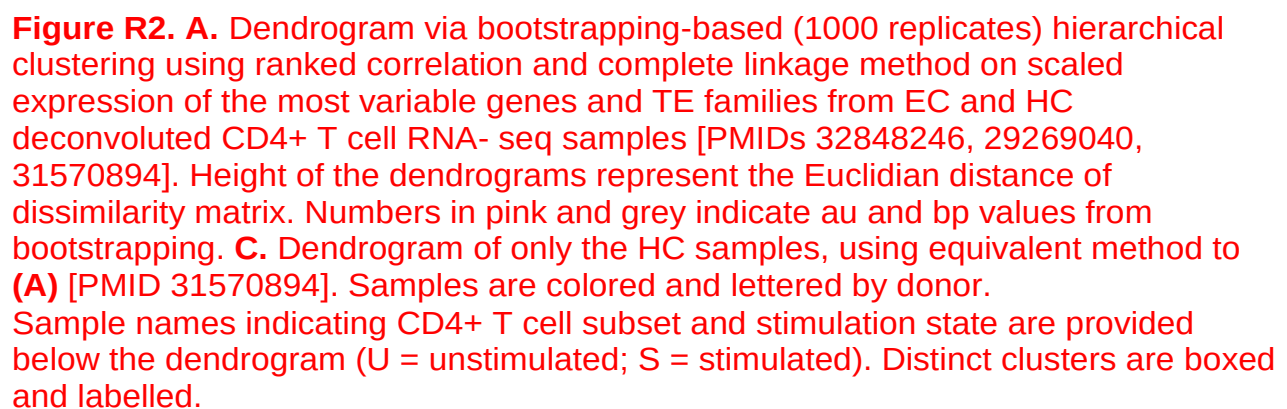
immune activation, to assess whether the gene expression signature that we define for ECs is a consequence of HIV replication. While we did observe variation in the level of HIV-1 expression between donors and donor sites (Figure S6A), there was no significant association between HIV-1 expression and the expression of ISGs across deconvoluted EC samples (Figure S6C-D). From this, we conclude that the unique immune gene expression profile observed in ECs is not directly associated with viremia. We are intrigued by the reviewer's suggestion to analyze datasets from individuals living with HIV as well as other active or chronic infections. This represents an exciting avenue for future work but is beyond the scope of this manuscript.

3. The possible background due to variation in gene expression from one individual to another is not well accounted for. Rather than being somehow related to EC, it might well reflect natural genetic or epigenetic variability. This issue is particularly important in the interesting analysis presented in Figure 3, in which the authors identify 4 well separated clusters based on differences among ECs in gene and TE expression. The interpretation that these clusters are an EC specific cannot be supported without a similar analysis using data from a similar number of HC donors.

AUTHORS: We acknowledge the importance of considering natural interindividual (epi)genetic variability when interpreting gene and TE expression patterns, and our findings do not discount the possibility that this variability may play a role in shaping antiviral responses. Our clustering analysis in Figure 3 highlights variation among ECs, and we agree that comparisons to a larger cohort of HCs than those we include in Figures 1, 2D, S2, and 4 would strengthen the conclusions drawn from these findings. In support of addressing this, we first note that in Figures 2B-C and S2, EC CD4+ T cells exhibit more inherent variability in both gene and TE expression compared to intra-cohort variability of HCs and VPs. Furthermore, we observed that the four EC clusters do not exhibit significant biases related to ethnicity (Figure S5C). While these begin to address the point raised by the reviewer, we have conducted an expanded analysis as suggested, incorporating an RNA-seq dataset of unstimulated and stimulated, deconvoluted CD4+ T cells from HCs (n = 5 donors, totaling 38 samples) [PMID 31570894]. Using these samples, we performed 1,000 bootstrapping iterations on hierarchical clusters derived from Spearman's correlation of gene and TE family expression profiles. This analysis validated the distinction of the four EC clusters from both the unstimulated and stimulated HC samples (**Figure R2A**). To further assess whether similar donor-based clustering would emerge among HCs, we repeated this analysis using only the 38 RNA-seq samples from deconvoluted HC CD4+ T cells (**Figure R2B**). We observed that clustering within the HC samples was not driven primarily by donor-specific factors, but rather by a mixture of stimulation state and CD4+ T cell subset. For example, the two boxed clusters on the left were enriched for stimulated samples, while the group on the far right was enriched for unstimulated samples.

Moreover, all individual donors with multiple sequenced subsets appeared in multiple clusters, further indicating that donor identity was not the primary driver of clustering in the HC cohort. These results suggest that while interindividual transcriptomic variability is present among HCs, the heterogeneity does not appear sufficient to drive the clustering

observed in ECs, not driven by CD4+ subset or ancestry. We believe this further supports our conclusion that – at least for the donors included in our analyses – ECs are distinctly heterogeneous, characterized by unique gene and TE expression profiles described in our study. However, we acknowledge that this HC dataset is significantly smaller than that used for our EC clustering analysis (n = 128 CD4+ T cell samples) and thus are conclusions here are preliminary. As part of future work, we intend to expand this analysis to incorporate a private RNA-seq dataset of deconvoluted CD4+ T cells from nearly 100 healthy donors [PMID 30449622], pending approval of our request for access to the raw sequencing data. Additionally, we plan to investigate interindividual variability in greater depth, expanding beyond transcriptomic variation to include genetic variation at the DNA level. For this reason, while we believe the additional analyses we performed strengthen the robustness of our conclusions, we have decided not to integrate them into our current manuscript as we aim to address interindividual variability in a more comprehensive future study.



AUTHORS: We thank the reviewer for their suggestion. Our intention was to focus on the novel roles of TEs identified here rather than extensively revisiting elite control, which has been well reported in the literature. However, we agree that providing more context

enhances the manuscript, and we have added some key references on elite control in the Background to help further orient the reader [L67-76].

Other issues:

Lines 76-79. This observation can only be explained by selection against cells with intact proviruses. It makes no sense to infer that intact and defective proviruses differ in their integration targeting.

AUTHORS: Our intention was not to imply that these differences in the proviral reservoir are a result of integration targeting, but we understand why the reviewer was misled to think so. We have revised the text accordingly to better align with the current understanding of HIV integration biology, clarifying that the observed differences between intact and defective integration sites are most likely a result of post-integration selective pressures shaping the proviral landscape [L77, 84-85].

Figure 1 legend and elsewhere. It is nice that the authors supply references for the datasets used, but PLEASE give them numbers so the reader can find them in the list.

AUTHORS: We appreciate the reviewer pointing out this inconsistency in citation formatting. We have corrected it in the revised version.

Line 179-180 The possible effect of ART on the TE expression profile should be considered. **AUTHORS:** We agree with the reviewer that ART could influence TE expression. For this reason, we positioned the comparative analysis between ECs and PLWH-on-ART as a supplementary figure (Figure S1), instead emphasizing the main comparisons of ECs to treatment-naïve VPs and HCs in our study. We noted in the Results that we chose to prioritize VP vs. EC analyses to ensure our “findings were not confounded by the presence of antiretroviral treatment” [L196-197], highlighting that we focus on likely more biologically relevant comparisons for our hypothesis. Additionally, we have further revised the paragraph in which we present Figure S1 to explicitly mention this possible confounding variable [L193-195].

Figure 3 legend. Please tell us up front you are looking at gene expression.

AUTHORS: We have revised the legend to explicitly state that the RNA-seq samples are clustered by gene and TE family expression [L335-336].

Line 359 Please tell the reader why you chose 10 kb.

AUTHORS: We appreciate the reviewer’s suggestion to clarify our choice of the 10-kb cutoff. This justification was added in our previous revisions, explaining that “loci and genes were paired based on proximity, with a maximum distance of 10kb between the TE locus and the gene’s transcription start site, to increase the likelihood of a direct cis- regulatory influence of the TE over the nearby gene.” [L367-370]. This proximity-based criterion reflects common practice in cis-regulatory predictions, balancing inclusivity and plausibility of regulation.

Line 382 "...some HIV restriction..."

AUTHORS: We have added the recommended clarification to this line [L391], emphasizing that some but not all HIV-1 restriction factors may be driven by increased cis-regulatory activity of nearby TEs.

Figure 4. The light grey is invisible when printed.

AUTHORS: We have darkened the color of these bars accordingly to improve visualization. Line 602. Is there any evidence for excessive C-T hypermutation in EC proviruses?

AUTHORS: The reviewer raises an interesting question regarding the extent of C-T hypermutation in EC proviruses, which could reflect higher APOBEC3 cytidine deaminase activity. Previous studies have noted that hypermutated sequences are prevalent in proviral DNA from ECs [PMIDs 27453467, 12890627, 24131498]. However, they are not necessarily more abundant in ECs compared to other PLWH cohorts. For example, hypermutated sequences accounted for a high proportion of proviruses in ECs but were found in equal or greater total numbers in non-controllers [PMID 27453467]. Similarly, while some cases suggest APOBEC3-mediated hypermutation significantly contributes to proviral inactivation in ECs [PMID 19733595], this pattern is not consistently observed across cohorts, nor is hypermutation universally correlated with the EC phenotype [PMIDs 19733595, 24131498]. Our findings align with this since we found that only one of our four EC clusters exhibited high expression of APOBEC3s and other restriction factors (Figure 3D) [L310]. Thus, a subset of ECs may have enhanced APOBEC activity and proviral hypermutation.

Figure 7. Must include the role of the CTL response to make any sense.

AUTHORS: Figure 7 is a hypothetical model derived from our findings in CD4+ T cells and focuses specifically on the processes we analyzed in this study. Since our work does not examine the CTL response directly, we have opted not to integrate additional components to the figure. For a broader and more comprehensive visualization of elite control, including the role of CD8+ T cells, NK cells, HLA alleles, and other immune mechanisms, we refer readers to existing reviews on the subject [PMIDs 31252326, 35844607], which have been integrated as new references in the revised manuscript [L634].

Lines 636-639. Again, integration site differences can only be sensibly explained as a consequence, not a cause, of elite control.

AUTHORS: We agree that the observed differences in proviral integration sites are most sensibly explained as a consequence of selective pressures unique to ECs. In this section of the manuscript, we do not claim that the preferential retention of defective proviruses in euchromatin-rich regions is causative. Rather, we cite studies on integration site patterns as an additional characteristic of EC biology that contributes to a more complete understanding of the phenotype. To be clearer in this sentence, we have changed “proviral integration patterns” to “chromosomal distribution of retained proviruses” [L641-642].