Endogenous Retroelement Expression in the Gut Microenvironment of People Living with HIV-1

Nicholas Dopkins*[1], Tongyi Fei[1], Stephanie Michael[1], Nicholas Liotta[1], Kejun Guo[2,3,4], Kaylee L. Mickens[2,3,4], Brad S. Barrett[2,3,4], Matthew L. Bendall[1], Stephanie M. Dillon[2,3,4], Cara C. Wilson[2,3,4], Mario L. Santiago[2,3,4] and Douglas F. Nixon[1]

Keywords: human endogenous retrovirus (HERV), interferons, human immunodeficiency virus type I (HIV-I), Endogenous Retroelement (ERE), Long interspersed nuclear element (LINE)

[1]Division of Infectious Diseases, Department of Medicine, Weill Cornell Medicine, New York, NY, United States of America.

[2]Department of Medicine, University of Colorado School of Medicine, Aurora, CO, United States of America.

[3]RNA Bioscience Initiative, University of Colorado School of Medicine, Aurora, CO, United States of America.

[4]Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO, United States of America.

Correspondence:

Nicholas Dopkins, PhD

Email: nid4009@med.cornell.edu

Summary

Background: Endogenous retroelements (EREs), including human endogenous retroviruses (HERVs) and long interspersed nuclear elements (LINEs), comprise almost half of the human genome. HIV-1 infects a target cell already possessing ancient retroviral genetic material, and exogenous HIV-1 infection modulates the expression of cell associated EREs. Following initial HIV-1 infection, there is a rapid destruction of CD4+ cells in the gut associated lymphoid tissue (GALT). Our previous studies of the interferome in the gut suggest potential mechanisms regarding how IFNb may drive HIV-1 gut pathogenesis. As ERE activity is suggested to partake in type 1 immune responses and is incredibly sensitive to viral infections, we sought to elucidate underlying interactions between ERE expression and GALT dynamics in PLWH.

Methods: ERE expression profiles from bulk RNA sequencing of colon biopsies and PBMC were compared between a cohort of PLWH prior to or within 7 days of initiating antiretroviral therapy (ART) (n=19) and uninfected controls (n=13). Individual EREs were then compared with the profiling of uninfected gut CD4+ T cells activated with type 1 interferons (IFN-Is) (n=3) to elucidate potential mechanisms for their induction in PLWH.

Findings: 59 EREs were differentially expressed in the colon of PLWH when compared to uninfected controls (Wald's Test with Benjamin-Hochberg correction: padj < 0.05 and FC ≤ -1 or ≥ 1). Of these 59, 12 EREs were downregulated in PLWH and 47 were upregulated. Colonic expression of the ERE loci LTR19_12p13.31 and L1FLnI_1q23.1s showed significant correlations with CD8+ T Cells and dendritic cell subset frequencies in the GI tract (Spearman's

Correlation: p value < 0.05). Furthermore L1FLnI_1q23.1s showed a significant upregulation in the blood of PLWH when compared to uninfected controls (T test: p <0.05) suggesting a common mechanism of differential ERE expression in PBMC and GALT.

Interpretation: ERE activity has been largely understudied in genomic characterizations of human pathologies. We show that the activity of certain EREs in the GI tract of PLWH is deregulated, supporting our hypotheses that their underlying activity could function as (bio)markers and potential mediators of pathogenesis in HIV-1 reservoirs.

Funding: NCI CA260691 (DFN) and NIAID UM1AI164559 (DFN).

Introduction

Human immunodeficiency virus type I (HIV-1) is a retrovirus that infects immune cells by binding to receptors on the cell surface[1]. HIV-1 infects host cells by integrating a copy of the viral genome into the host genome, resulting in proviruses that then hijack the host cell's replicative machinery to produce the building blocks of infectious particles[1]. The HIV-1 lifecycle may rapidly deplete host CD4+ T cell populations and disrupt global immune dynamics[2]. Disruptions to the host immune system predispose people living with HIV-1 (PLWH) to life-threatening comorbidities[3]. Since 1996, life expectancy and quality of life for PLWH have risen due to the advent of antiretroviral therapies (ARTs) that pharmacologically inhibit key processes of the retroviral life cycle[4]. Unfortunately, ART merely suppresses retroviral processes, emphasizing the need for cure strategies that rid individuals of latent proviral reservoirs which may resume infectious particle formation after a lapse in treamtent[5,6]. One such reservoir site is the gastrointestinal (GI) associated lymphoid tissue (GALT)[7,8], which is particularly rich in CD4+ T cells[9]. HIV-1 induced immunopathologies in the GALT are associated with severe comorbidities in PLWH[10,11]. Many of these pathologies could be sustained by the accumulation of plasmacytoid dendritic cells (pDCs) that are trafficked to the GI tract during chronic HIV-1 infection[12]. Models of Simian Immunodeficiency Virus (SIV), a retrovirus closely related to HIV-1, demonstrate that pDCs traffic from the circulation, where they are depleted, to the GALT, where their overabundance contributes to immunopathologies that allow the site to become the primary source of viral replication[13,14]. pDCs reflect a unique subset of DCs whose primary function is not antigen presentation but is instead immunomodulation through the release of cytokines[15]. In the acute stages of HIV-1 infection, pDCs produce type 1 interferons (IFN-Is) as one of the first lines of defense against the virus[16], and are therefore protective. However, the

sustainment of their IFN-I production during the chronic stages of infection exhausts and deregulates antiviral immunity, negatively impacting the health of PLWH[12,16–18].

To better characterize how immunopathologies in PLWH develop, we investigated how endogenous retroelement (ERE) expression may underlie the development of HIV-1 induced immune deregulation. EREs are highly abundant sequences estimated to make up roughly ~40% of the human genome[19]. In humans, EREs include long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and human endogenous retroviruses (HERVs)[20]. Previous studies have demonstrated that ERE activity is highly susceptible to change during infection with HIV-1[21–31]. While IFN-I signaling may activate EREs proximal to interferon stimulated genes (ISGs), other EREs, such as HERV elements, can be directly trans-activated by HIV-1 cofactors due to high sequence homology between endogenous and exogenous retroviruses[21,22,24,32,33]. This homology can be so substantial that certain HERV-K18 and HERV-W products can be incorporated in the assembly of HIV-1 particles[34–38]. ERE activity also functions as a major regulator of transcriptional networks that have been co-opted in organismal physiology[39,40]. In immunity, examples of ERE activity cooption suggest their intimate association with the host mounted responses[41,42]. Recurrent observations continually reinforce hypotheses that EREs likely participate in human health and disease through poorly understood mechanisms[42–46]. The activity of multiple HERVs[21–28] and LINEs[29,30] is altered during HIV-1 infection, which then impacts host gene expression[24], packaging of HIV-1 particles[34–38], and immunity[25,47]. For the purpose of further defining HIV-1 mediated gut pathogenesis, we sought to characterize how ERE activity at the RNA level is modulated in the gut microenvironment of PLWH. Since EREs have been largely overlooked in genomic approaches to classify underlying

components of human pathologies, we hypothesize that their consideration would identify novel markers of HIV-1 pathogenesis in the gut.

ERE expression at the RNA level was quantified in colon biopsies collected from a cohort of viremic individuals prior to or at the initiation of ART[17] to identify those best correlating with viral load. We next overlaid these findings with IFN-I stimulated gut CD4+ T cells to estimate if the differential expression of EREs in the gut of PLWH is driven by the HIV-1 lifecycle or by innate immunopathologies. Collectively, our analyses demonstrate that a subset of ERE loci are impacted in PLWH. The bulk of PLWH-upregulated EREs in the gut are not upregulated by IFN-I stimulation of gut CD4+ T cells, suggesting their activation by other aspects of the HIV-1 life cycle or the host response. Two identified EREs, LTR19_12p13.31 and L1FLnI_1q23.1s, further correlated with myeloid and lymphoid immune cell abundances. Collectively, incorporation of the GI-specific expression of LTR19_12p13.31 and L1FLnI_1q23.1s can better define HIV-1 driven immunopathologies of the GALT, and the expression of these elements may possess underlying roles in pathogenesis.

Methods

Locus-specific quantification of ERE expression at the RNA level

RNA sequencing FASTQ files of colon pinch biopsies and interferome arrays of uninfected CD4+ T cells were obtained from a previously published study[17] under the Bioproject accession numbers “PRJNA558500” and “PRJNA558974”, respectively. RNA sequencing FASTQ files performed on the PBMC fraction from the same participants are available under the Bioproject accession number “PRJNA659373”. FASTQ files were aligned to the human genome build 38 (hg38) using STAR (v2.7.9.a)[48]. STAR alignment was performed using the parameters

"--outSAMstrandField intronMotif --outFilterMultimapNmax 200 --winAnchorMultimapNmax 200" for the purpose of capturing multimapping ERE reads. We then used Telescope (v1.0.3)[49], which utilizes an expectation-maximization algorithm to improve upon the definition of ubiquitously mapping ERE transcripts to a custom annotation, for locus-specific approximation of ERE expression. The Telescope assign module was performed using the parameters "--theta_prior 200000 -- max_iter 1000" to align reads to a custom annotation of EREs accessible at https://github.com/mlbendall/telescope_annotation_db. Metadata for the ERE annotation predicting the intronic, exonic, and intergenic status of individual loci in the Telescope annotation relative to coding gene annotations can be found at https://github.com/liniguez/Telescope_MetaAnnotations. Original code produced for analysis can be accessed at https://github.com/NicholasDopkins/JuneHIVMucosa.

Differential expression analysis

Lowly abundant EREs were filtered from downstream analysis by ensuring that all EREs possessed at least 2 reads within 10% of the total samples. Differential expression analysis was performed between the ERE expression profiles of uninfected control and PLWH colon biopsies using standard DESEQ2 (v1.30.1)[50] parameters. Differential expression analysis was performed between the ERE expression profiles of IFN-I stimulated gut CD4+ T cells using DESEQ2 with the parameters "parallel = T" and "betaPrior = T". Results were extracted as DESEQ objects, with a numbered contrast of each group compared against all others. Differentially expressed EREs were visualized with the packages pheatmap (v1.0.12), ggVennDiagram (v1.2.2)[51] and EnhancedVolcano (v1.8.0). Individual EREs were visualized with a custom function that overlays geom_jitter and geom_boxplot functions provided by ggplot2 v(3.3.5)[52] to produce a

Tukey box and whisker plot that displays individual replicates. Clinical parameters of immunopathologies were provided by previously published metadata[17,18,53,54] and a table summary is available at https://github.com/NicholasDopkins/JuneHIVMucosa/blob/main/Metadata.csv. Principal coordinate analyses (PCA) were performed on variance stabilizing transformed DESEQ outputs in order to visualize if untested confounding factors explained observations pertaining to differential expression analyses. PCA generation was performed using PCAtools (v2.2.0) with the variable parameter set to "removeVar = 0.1".

Integrative Genomics Viewer

National Center for Biotechnology Information (NCBI) Reference Sequence Database (Refseq)[55], ENCyclopedia Of DNA Elements[56] (Gencode), and Telescope[49] annotations of the human genome were loaded into Integrative Genomics Viewer[57] (IGV) for alignment of StringTie[58] reconstructed transcriptomes from uninfected controls and PLWH samples.

Statistics

Standard DESEQ2 statistics of Wald's Test with Benjamin-Hochberg correction of p values were utilized for the identification of significantly deregulated EREs between uninfected controls and PLWH samples. EREs a log2 fold change of >1 or <-1 and an adjusted p value (padj) of <0.05 were deemed significantly differentially expressed. Statistical values for correlations between clinical parameters and ERE expression were conducted using Pearson's correlation coefficient values provided by ggplot2. To elucidate values of significance regarding of ERE expression between experimental groups, T tests were used to gather significance values

when stated. For significance of post hoc tests the following scale is used to indicate p values when applicable: ns, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$

Clinical cohort

ERE expression analyses were conducted using FASTQ files from previously generated RNAseq data collected from colonic tissue and PBMC fractions collected from 19 PLWH and 13 age- and sex-matched controls, and associations with features of HIV-1 pathogenesis were determined using archived datasets of colon and systemic immunological and virological parameters[17,18,53,54]. On average, PLWH had been infected with HIV-1 (defined by the first HIV-1 seropositive test) for 5.25±1.1 (mean±SEM) years. PLWH were ART-naive or were not on ART for more than 7 days in the preceding 6 months, and had CD4 T cell counts >200 cells/μl within 3 months of clinical visit. Exclusion criteria for both cohorts are extensively detailed elsewhere[53]. In addition to blood samples, study participants underwent a flexible sigmoidoscopy with multiple colon pinch biopsies obtained. Clinical characteristics for the study participants included in this current study are detailed in Supplemental Table 1.

Immune system characterization of PLWH and uninfected controls

Methods used to generate the archived datasets have been extensively detailed[17,18,53,54,59,60]. In brief, indicators of systemic inflammation were assessed with plasma IL-6 concentrations by ELISA, and microbial translocation was measured by plasma LPS concentrations with the Limulus Amebocyte Lysate assay[53]. Frequencies of colon mDCs, pDCs, and T cells were determined using multi-color flow cytometry[53,54] and expressed as an absolute number per gram of tissue based on the frequency within viable CD45+ cells, initial cell counts, and biopsy

weights. Colon tissue HIV-1 RNA was quantified by real-time PCR and HIV-1 RNA copy numbers normalized per CD4 T cell within each biopsy calculated by the percent of all viable cells that were CD45+CD3+CD4+ as determined by flow cytometry and weight of each biopsy and reported as HIV-1 RNA per million CD4 T cell[53].

Ethics Statement

Analyses performed in this study were performed on deidentified and publicly available or previously archived data. The original clinical study was approved by the Colorado Multiple Institutional Review Board and all study participants voluntarily gave written, informed consent[17,18,53,54].

Role of Funders

The funding bodies had no influence on the planning, conduction and analysis of data performed during this study.

Results

HIV-1 status influences endogenous retroelement expression in the gut microenvironment

Following preprocessing to remove lowly abundant transcripts, we quantified the expression of 13,706 EREs in colon pinch biopsies collected from PLWH in comparison to age and sex-matched HIV-1-uninfected controls (Fig. 1a). Significance in differential expression between controls and PLWH was found for 59 ERE loci (Fig. 1b and Fig. 1c). Of these 59 loci, 47 were upregulated and 12 were downregulated in PLWH. All HERVs differentially expressed were upregulated in PLWH when compared to controls. Overall, ERE activity at the RNA level

displays discrete changes in the gut microenvironment in viremic individuals prior to the initiation of ART. Further consideration of the phylogenies and chromosomal origin of ERE expression profiles demonstrated no distinctive alterations in PLWH when compared to controls, suggesting no large-scale changes in ERE expression in the gut microenvironment of PLWH (Fig. S1). To infer mechanisms by which HIV-1 infection induces the differential expression of EREs in the gut, we next quantified changes in expression resulting from IFN-I stimulation of uninfected gut-derived CD4+ T cells. IFN-Is do not shift the global landscape of ERE activity at the RNA level (Fig. S2). Discrete changes in locus-specific expression demonstrate that IFN-I activation of uninfected gut CD4+ T cells share upregulated expression of 23 EREs, suggesting that they are IFN-inducible (Fig. S3 and Fig. S4). 7 of these 23 EREs were upregulated in the gut of PLWH, suggesting immune-dependent mechanisms for their in HIV-1 infection, while 35 of the 42 upregulated EREs in PLWH showed no significant induction by IFN-I stimulation (Fig. S5). Collectively, HIV-1 infection upregulates specific EREs in the gut of PLWH through both IFN-dependent and IFN-independent mechanisms.

Correlation between gut endogenous retroelement expression and viral load in people living with HIV-1

Pearson's correlation coefficients were calculated for all differentially expressed EREs in the gut of PLWH to identify any potential links with their expression and HIV-1 infection severity as determined by circulating HIV-1 viral load. Of the 59 differentially expressed EREs, three showed near significant ($p<0.10$) positive correlations between colonic expression and viral load in PLWH. L1FLnI_1q23.1s possessed a positive association with viral load in the circulation (Fig. 2a) and is upregulated in the colon of PLWH when compared to uninfected

controls (Fig. 2b). In PBMCs collected from the same participants, expression of L1FLnI_1q23.1s is significantly upregulated in concurrence with gut expression of the element (Fig. 2c). Analysis of IFN-I activated gut CD4+ T cells suggests that IFN alpha 1 (IFNa1) and IFN beta (IFNb) may contribute to the upregulated expression of L1FLnI_1q23.1s, however no IFN-Is yielded significant effects (Fig. 2d). IGV demonstrates that the L1FLnI_1q23.1s locus is an ERE within the Pyrin And HIN Domain Family Member 1 (PYHIN1) gene with minor differences in the putative alignments provided by StringTie for uninfected controls and PLWH samples (Fig. 2e). L1FLnI_5q32o possessed a positive association with viral load in the circulation (Fig. 2f) and is upregulated in the colon of PLWH when compared to controls (Fig. 2g). In PBMCs collected from the same study participants, expression of L1FLnI_5q32o is significantly upregulated in concurrence with gut expression of the element (Fig. 2h). Analysis of IFN-I activated gut CD4+ T cells suggests that IFN-Is yield no significant effects on the expression of L1FLnI_5q32o (Fig. 2i). IGV demonstrates that the L1FLnI_5q32o locus is an ERE within the Janus Kinase And Microtubule Interacting Protein 2 (JAKMIP2) gene with no differences in the putative alignments provided by StringTie for PLWH and control samples (Fig. 2j). LTR19_12p13.31 possessed a positive association with viral load in the circulation (Fig. 2k) and is upregulated in the colon of PLWH when compared to controls (Fig. 2l). In PBMCs collected from the same study participants, expression of LTR19_12p13.31 shows a non-significant upregulation (Fig. 2m). Analysis of IFN-I activated gut CD4+ T cells suggests that IFN-Is yield no significant effects on the expression of LTR19_12p13.31 (Fig. 2n). IGV demonstrates that LTR19_12p13.31 locus is an ERE that overlaps the Pregnancy Zone Protein (PZP) and Killer Cell Lectin Like Receptor G1 (KLRG1) genes with a unique transcript in the putative alignments of PLWH samples provided by StringTie (Fig. 2o). Collectively, expression

of the EREs L1FLnI_1q23.1s, L1FLnI_5q32o, and LTR19_12p13.31 show associations with blood HIV-1 RNA levels,, and their induction in the gut may likely be independent of IFN-I mediated pathologies.

Correlation Between LTR19_12p13.31 and L1FLnI_1q23.1s expression and immune signatures of the gut microenvironment in people living with HIV-1

We next analyzed how expression of L1FLnI_1q23.1s, L1FLnI_5q32o and LTR19_12p13.31 correlate with cellular abundances that may be associated with pathogenesis in the gut of PLWH. There were no significant associations between the number colon CD4+ T cells (measured per gram of gut tissue)with LTR19_12p13.31 (Fig. 3a) and L1FLnI_1q23.1s expression (Fig. 3b). LTR19_12p13.31 demonstrated significant positive correlations with the number of colon CD8+ T cells (Fig. 3c) and IFN gamma (IFNy)+CD8+ T cells (Fig. 3e), whereas L1FLnI_1q23.1s demonstrated near significant positive correlations with CD8+ (Fig. 3d) and IFNy+CD8+ T cell (Fig. 3f) abundances. Further consideration into dendritic cell (DC) abundances demonstrate that LTR19_12p13.31 (Fig. 3g) and L1FLnI_1q23.1s (Fig. 3h) show near significant positive correlations with CD11c+ myeloid DCs (mDCs), while LTR19_12p13.31 (Fig. 3i) displayed a significant positive correlation and L1FLnI_1q23.1s (Fig. 3j) displayed a near significant positive association with plasmacytoid DCs (pDCs). Amongst mDCs, LTR19_12p13.31 (Fig. 3k) and L1FLnI_1q23.1s (Fig. 3l) show significant positive correlations with the CD1c- mDCs, but not with CD1c+ mDCs (Fig. S6), suggesting potential perturbations related to specific mDC subsets[54]. Expression of the ERE L1FLnI_5q32o did not demonstrate positive or negative correlations with CD4+ T cells, CD8+ T cells, IFNy+CD8+ T cells, CD11c+ mDCs, pDCs, or CD1c- mDCs in PLWH (Fig. S7). In analyses of frequencies of

CD4 T helper (Th) cell subsets, L1FLnI_1q23.1s significantly positively correlated with IL22+IFNy-IL17- CD4+ T cells (i.e. Th22) and IFNy-IL17+ CD4+ T cells (i.e Th17), and trended towards a significant positive association with IFNy+IL17+ CD4+ T cells (i.e. inflammatory Th17 cells) (Fig. S8). L1FLnI_5q32o (Fig. S9) and LTR19_12p13.31 (Fig. S10) demonstrated no significant correlations with analyzed CD4+ T cell subset abundances in the gut. Furthermore, systemic markers of microbial translocation and inflammation including as plasma lipopolysaccharide (LPS) and interleukin (IL)-6, were analyzed and demonstrated insignificant associations with LTR19_12p13.31, L1FLnI_1q23.1s, or L1FLnI_5q32o expression (Fig. S11). Collectively, LTR19_12p13.31 and L1FLnI_1q23.1s, but not L1FLnI_5q32o, show significant associations with the various innate and adaptive immune cell frequencies in the gut of PLWH.

Discussion

In this study, we leverage modulations in ERE activity at the RNA level in the GI tract of untreated PLWH with frequencies of immune cell populations that are likely critical players in driving HIV-1 gut pathogenesis[11]. As the derepression of EREs may possess immunomodulatory roles in human diseases[42,61], our results suggest further investigation into their activity in HIV-1 pathogenesis. By analysis of RNA sequencing from colon biopsies of uninfected controls vs untreated PLWH, we identified 59 differentially expressed EREs associated with HIV. By finding correlations between their colonic expression and circulating viral load, we were able to narrow further investigations to focus on 3 EREs upregulated in PLWH that might indicate the severity of HIV-1 pathogenesis in the gut, or function as markers of immune deregulation. The EREs LTR19_12p13.31 and L1FLnI_1q23.1s best correlated with cellular abundances including

higher frequencies of pDCs, CTLs, and CD1c- mDCs. L1FLnI_5q32o demonstrated no associations with these cellular markers, however, did show associations with differential abundances of CD4+ T cell Th subsets that were not found with LTR19_12p13.31 and L1FLnI_1q23.1s expression. By considering their expression in PBMCs from the same study participants, we observed a significant upregulation in L1FLnI_1q23.1s expression and a non-significant upregulation in LTR19_12p13.31. In accordance with previous studies[22], we observe significant alterations in ERE activity at the RNA level in the PBMC fraction of PLWH and uninfected controls (Fig. S12). This suggests future use of L1FLnI_1q23.1s as a potential circulating biomarker for deregulations of immunity in the gut microenvironment of PLWH, while LTR19_12p13.31 is more likely an intrinsic marker specific to the GI tract. Collectively, we find significant associations between LTR19_12p13.31 and L1FLnI_1q23.1s and cellular dynamics of the GALT in PLWH, further emphasizing their identity as biomarkers or potential mediators of immune deregulation.

Previous studies have demonstrated that ERE activity is widely changed in response to the HIV-1 lifecycle[21,22,27,29,31] and certain HIV-1 associated comorbidities[62,63]. ERE expression is also susceptible to changes the in epigenetics and immunity that are elicited during viral infections and the host response[64]. While it remains unknown if the changes in expression for most EREs result from a bystander effect or by a cooption of host immunity in HIV-1 infection, their activity can be incorporated to better characterize poorly understood immunopathologies in viral infections. Here, we demonstrate that significant correlations can be established between differentially expressed EREs and cellular markers that likely contribute to GI immunopathologies of untreated HIV-1 infection. Some limitations of this study concern bulk RNA sequencing approaches utilized, leaving the data unable to discern what cell types are

responsible for the overexpression of differentially expressed EREs. As the cellular composition of the sequenced tissues fluctuates with disease severity, it is unclear whether the phenotype or mere abundance of disproportionate cell types, such as pDCs and CTLs, is responsible for their expression. Future studies that accurately quantify ERE expression from single cell RNA sequencing data that were not available from this cohort would provide further insight into the activation state of these EREs in the GI tract of PLWH. Additionally, due to the poorly defined molecular characteristics and low copy number of LTR19 HERV sequences, exceedingly little is known about their potential roles in health and disease. LTR19 sequences are a member of the HERVFA clade and are estimated to only possess roughly ~15 copies in the human genome[65].

In conclusion, this study provides the first correlative link between ERE expression and the deregulations of GALT observed in untreated HIV-1 infection. While EREs can be modulated in context dependent manners by various aspects of HIV-1 infection, their potential roles in systemic immunity in the GI tract have remained undetermined. We also demonstrate that the IFN-I response mediates ERE expression in uninfected gut CD4+ T cells, however this only appears to be responsible for a distinct subset of roughly 7 differentially expressed elements. It is thus unclear if the HIV-1 lifecycle or IFN-I activity on CD4- cell types in the GALT are responsible for changes in ERE activity. We find that LTR19_12p13.31 and L1FLnI_1q23.1s are promising markers of HIV-1 pathogenesis in the gut and may possess undetermined roles in determining cellular composition. This study also provides one of the first observations of a differentially expressed LTR19 elements in a human disease, thus emphasizing the importance of their further study.

Contributors

ND co-designed the study, co-prepared the first manuscript draft, co-performed data analyses, co-curated code, led statistical analyses, and curated figures. TF co-prepared the first manuscript draft, co-performed data analyses, co-performed statistical analysis, co-curated code, and provided intellectual input critical to analysis. SM co-designed the study, co-prepared the first manuscript draft, co-performed data analyses, and provided intellectual input critical to data interpretation. NL co-performed analyses and co-curated code. KG provided intellectual input critical to hypothesis generation and co-provided data access. KLM provided intellectual input critical to hypothesis generation and co-provided data access. BSB provided intellectual input critical to hypothesis generation and co-provided data access. MLB provided intellectual input critical to hypothesis testing. SMD provided intellectual input critical to hypothesis generation and testing, co-provided data access, and performed immunological analysis of the gut microenvironment. CCW provided intellectual input critical to hypothesis generation and co-provided data access. DFN co-designed the study and co-prepared the first manuscript draft. MLS co-designed the study, co-prepared the first manuscript draft, and co-provided data access. All authors contributed to data interpretation, discussion regarding significances, and editing of the manuscript. All authors have read and approved the final manuscript.

Declaration of Interests

All authors have no potential conflicts of interests to disclose.

Acknowledgments

These works are supported by US NIH grants NCI CA260691 (DFN) and NIAID UM1AI164559 (DFN).

Data Sharing Statement

FASTQ from the bulk RNA sequencing reanalyzed by this study can be accessed at the NCBI sequence read archive (SRA) under PRJNA558974 and PRJNA558500. Code for analysis can be accessed at https://github.com/NicholasDopkins/JuneHIVMucosa. Details on running the Telescope pipeline can be accessed at https://github.com/mlbendall/telescope.

References

1. Justiz Vaillant, A. A. & Gulick, P. G. HIV Disease Current Practice. in *StatPearls* (StatPearls Publishing, 2023).
2. Boasso, A., Shearer, G. M. & Chougnet, C. Immune dysregulation in human immunodeficiency virus infection: know it, fix it, prevent it? *J. Intern. Med.* **265**, 78–96 (2009).
3. Deeks, S. G., Overbaugh, J., Phillips, A. & Buchbinder, S. HIV infection. *Nat. Rev. Dis. Primer* **1**, 1–22 (2015).
4. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet* **372**, 293–299 (2008).
5. Sengupta, S. & Siliciano, R. F. Targeting the latent reservoir for HIV-1. *Immunity* **48**, 872–895 (2018).
6. Veenhuis, R. T. *et al.* Monocyte-derived macrophages contain persistent latent HIV reservoirs. *Nat. Microbiol.* **8**, 833–844 (2023).
7. Thompson, C. G., Gay, C. L. & Kashuba, A. D. M. HIV Persistence in Gut-Associated Lymphoid Tissues: Pharmacological Challenges and Opportunities. *AIDS Res. Hum. Retroviruses* **33**, 513–523 (2017).
8. Kelley, C. F. *et al.* HIV-1 RNA Rectal Shedding Is Reduced in Men With Low Plasma HIV-1 RNA Viral Loads and Is Not Enhanced by Sexually Transmitted Bacterial Infections of the Rectum. *J. Infect. Dis.* **204**, 761–767 (2011).
9. Mowat, A. McI. & Viney, J. L. The anatomical basis of intestinal immunity. *Immunol. Rev.* **156**, 145–166 (1997).

436 10. Guadalupe, M. *et al.* Severe CD4 $^{+}$ T-Cell Depletion in Gut Lymphoid Tissue during
437 Primary Human Immunodeficiency Virus Type 1 Infection and Substantial Delay in
438 Restoration following Highly Active Antiretroviral Therapy. *J. Virol.* **77**, 11708–11717 (2003).
439 11. Dillon, S. M. & Wilson, C. C. Gut Innate Immunity and HIV Pathogenesis. *Curr.*
440 *HIV/AIDS Rep.* **18**, 128–138 (2021).
441 12. Lehmann, C. *et al.* Longitudinal Analysis of Distribution and Function of Plasmacytoid
442 Dendritic Cells in Peripheral Blood and Gut Mucosa of HIV Infected Patients. *J. Infect. Dis.*
443 **209**, 940–949 (2014).
444 13. Ansari, A. A. *et al.* Blocking of α4β7 Gut-Homing Integrin during Acute Infection Leads
445 to Decreased Plasma and Gastrointestinal Tissue Viral Loads in Simian Immunodeficiency
446 Virus-Infected Rhesus Macaques. *J. Immunol.* **186**, 1044–1059 (2011).
447 14. Reeves, R. K. *et al.* SIV Infection Induces Accumulation of Plasmacytoid Dendritic Cells
448 in the Gut Mucosa. *J. Infect. Dis.* **206**, 1462–1468 (2012).
449 15. Villadangos, J. A. & Young, L. Antigen-Presentation Properties of Plasmacytoid
450 Dendritic Cells. *Immunity* **29**, 352–361 (2008).
451 16. O'Brien, M., Manches, O. & Bhardwaj, N. Plasmacytoid Dendritic Cells in HIV
452 Infection. in *HIV Interactions with Dendritic Cells* (eds. Wu, L. & Schwartz, O.) vol. 762 71–
453 107 (Springer New York, 2012).
454 17. Guo, K. *et al.* Qualitative Differences Between the IFNα subtypes and IFNβ Influence
455 Chronic Mucosal HIV-1 Pathogenesis. *PLoS Pathog.* **16**, e1008986 (2020).
456 18. DILLON, S. M. *et al.* A Compartmentalized Type I Interferon Response in the Gut
457 During Chronic HIV-1 Infection is Associated with Immunopathogenesis. *AIDS Lond. Engl.*
458 **32**, 1599–1611 (2018).

19. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

20. Bourque, G. *et al*. Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).

21. Bhardwaj, N., Maldarelli, F., Mellors, J. & Coffin, J. M. HIV-1 Infection Leads to Increased Transcription of Human Endogenous Retrovirus HERV-K (HML-2) Proviruses *In Vivo* but Not to Increased Virion Production. *J. Virol.* **88**, 11108–11120 (2014).

22. Contreras-Galindo, R., López, P., Vélez, R. & Yamamura, Y. HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) in vitro. *AIDS Res. Hum. Retroviruses* **23**, 116–122 (2007).

23. Vincendeau, M. *et al.* Modulation of human endogenous retrovirus (HERV) transcription during persistent and de novo HIV-1 infection. *Retrovirology* **12**, 27 (2015).

24. Srinivasachar Badarinarayan, S. *et al*. HIV-1 infection activates endogenous retroviral promoters regulating antiviral gene expression. *Nucleic Acids Res.* **48**, 10890–10908 (2020).

25. Jones, R. B. *et al.* HERV-K–specific T cells eliminate diverse HIV-1/2 and SIV primary isolates. *J. Clin. Invest.* **122**, 4473–4489 (2012).

26. Contreras-Galindo, R., Almodóvar-Camacho, S., González-Ramírez, S., Lorenzo, E. & Yamamura, Y. *Short Communication:* Comparative Longitudinal Studies of HERV-K and HIV-1 RNA Titers in HIV-1-Infected Patients Receiving Successful versus Unsuccessful Highly Active Antiretroviral Therapy. *AIDS Res. Hum. Retroviruses* **23**, 1083–1086 (2007).

27. Gonzalez-Hernandez, M. J. *et al.* Expression of Human Endogenous Retrovirus Type K (HML-2) Is Activated by the Tat Protein of HIV-1. *J. Virol.* **86**, 7790–7805 (2012).

28. Contreras-Galindo, R. *et al.* HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses. *Genome Res.* **23**, 1505–1513 (2013).

29. Jones, R. B. *et al.* LINE-1 retrotransposable element DNA accumulates in HIV-1-infected cells. *J. Virol.* **87**, 13307–13320 (2013).

30. Jones, R. B. *et al.* HIV-1 infection induces retrotransposition of LINE-1 elements. *Retrovirology* **6**, P43, 1742-4690-6-S2-P43 (2009).

31. O'Carroll, I. P. *et al.* Structural Mimicry Drives HIV-1 Rev-Mediated HERV-K Expression. *J. Mol. Biol.* **432**, 166711 (2020).

32. van der Kuyl, A. C. HIV infection and HERV expression: a review. *Retrovirology* **9**, 6 (2012).

33. Kyriakou, E. & Magiorkinis, G. Interplay between endogenous and exogenous human retroviruses. *Trends Microbiol.* **31**, 933–946 (2023).

34. Brinzevich, D. *et al.* HIV-1 Interacts with Human Endogenous Retrovirus K (HML-2) Envelopes Derived from Human Primary Lymphocytes. *J. Virol.* **88**, 6213–6223 (2014).

35. An, D. S., Xie, Y. & Chen, I. S. Y. Envelope Gene of the Human Endogenous Retrovirus HERV-W Encodes a Functional Retrovirus Envelope. *J. Virol.* **75**, 3488–3489 (2001).

36. Ogata, T., Okui, N., Sakuma, R., Kobayashi, N. & Kitamura, Y. Integrase of human endogenous retrovirus K-10 supports the replication of replication-incompetent Int- human immunodeficiency virus type 1 mutant. *Jpn. J. Infect. Dis.* **52**, 251–252 (1999).

37. Monde, K., Contreras-Galindo, R., Kaplan, M. H., Markovitz, D. M. & Ono, A. Human Endogenous Retrovirus K Gag Coassembles with HIV-1 Gag and Reduces the Release Efficiency and Infectivity of HIV-1. *J. Virol.* **86**, 11194–11208 (2012).

503 38. Terry, S. N. *et al.* Expression of HERV-K108 envelope interferes with HIV-1 production.
504 *Virology* **509**, 52–59 (2017).
505 39. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the
506 regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23**, 481–497 (2022).
507 40. Du, A. Y., Chobirko, J. D., Zhuo, X., Feschotte, C. & Wang, T. *Regulatory Transposable*
508 *Elements in the Encyclopedia of DNA Elements*.
509 http://biorxiv.org/lookup/doi/10.1101/2023.09.05.556380 (2023)
510 doi:10.1101/2023.09.05.556380.
511 41. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity
512 through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 (2016).
513 42. Kassiotis, G. & Stoye, J. P. Immune responses to endogenous retroelements: taking the
514 bad with the good. *Nat. Rev. Immunol.* **16**, 207–219 (2016).
515 43. Payer, L. M. & Burns, K. H. Transposable elements in human genetic disease. *Nat. Rev.*
516 *Genet.* **20**, 760–772 (2019).
517 44. Küry, P. *et al.* Human Endogenous Retroviruses in Neurological Diseases. *Trends Mol.*
518 *Med.* **24**, 379–394 (2018).
519 45. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of
520 cancer. *Genome Biol.* **22**, 147 (2021).
521 46. Dopkins, N. & Nixon, D. F. Activation of human endogenous retroviruses and its
522 physiological consequences. *Nat. Rev. Mol. Cell Biol.* (2023) doi:10.1038/s41580-023-00674-
523 z.
524 47. De Mulder, M. *et al.* Anti-HERV-K (HML-2) capsid antibody responses in HIV elite
525 controllers. *Retrovirology* **14**, 41 (2017).

48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

49. Bendall, M. L. *et al.* Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* **15**, e1006453 (2019).

50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

51. Gao, C.-H., Yu, G. & Cai, P. ggVennDiagram: An Intuitive, Easy-to-Use, and Highly Customizable R Package to Generate Venn Diagram. *Front. Genet.* **12**, 706907 (2021).

52. Ito, K. & Murphy, D. Application of ggplot2 to Pharmacometric Graphics. *CPT Pharmacomet. Syst. Pharmacol.* **2**, e79 (2013).

53. Dillon, S. M. *et al.* An altered intestinal mucosal microbiome in HIV-1 infection is associated with mucosal and systemic immune activation and endotoxemia. *Mucosal Immunol.* **7**, 983–994 (2014).

54. Dillon, S. M. *et al.* Gut dendritic cell activation links an altered colonic microbiome to mucosal and systemic T-cell activation in untreated HIV-1 infection. *Mucosal Immunol.* **9**, 24–37 (2016).

55. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

56. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

57. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

549 58. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from
550 RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
551 59. Jones, S. T. *et al*. Altered Immunoglobulin Repertoire and Decreased IgA Somatic
552 Hypermutation in the Gut during Chronic HIV-1 Infection. *J. Virol*. **96**, e0097622 (2022).
553 60. Dillon, S. M. *et al.* Granzyme B+ CD4 T cells accumulate in the colon during chronic
554 HIV-1 infection. *Gut Microbes* **14**, 2045852 (2022).
555 61. Kassiotis, G. The Immunological Conundrum of Endogenous Retroelements. *Annu. Rev.*
556 *Immunol.* **41**, annurev-immunol-101721-033341 (2023).
557 62. Curty, G. *et al*. Human Endogenous Retrovirus Expression Is Upregulated in the Breast
558 Cancer Microenvironment of HIV Infected Women: A Pilot Study. *Front. Oncol.* **10**, 553983
559 (2020).
560 63. Dai, L. *et al*. Transactivation of human endogenous retrovirus K (HERV-K) by KSHV
561 promotes Kaposi's sarcoma development. *Oncogene* **37**, 4534–4545 (2018).
562 64. Macchietto, M. G., Langlois, R. A. & Shen, S. S. Virus-induced transposable element
563 expression up-regulation in human and mouse host cells. *Life Sci. Alliance* **3**, e201900536
564 (2020).
565 65. Vargiu, L. *et al.* Classification and characterization of human endogenous retroviruses;
566 mosaic forms are common. *Retrovirology* **13**, 7 (2016).
567
568
569
570
571

Figure Legends

Figure 1. HIV-1 status influences endogenous retroelement expression in the gut microenvironment

Heatmap demonstrating the normalized per sample abundances of EREs following the preprocessing filtering of reads in uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19) individuals. ERE reads were filtered to possess at least 2 reads in 10% of the total samples for quality assurance (A). Heatmap demonstrating the normalized per sample abundances of EREs differentially expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between uninfected controls (red) (n = 13) and PLWH (blue) (n = 19) gut biopsies. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (B). Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between uninfected controls (n = 13) and PLWH (n = 19) gut biopsies. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (C).

Figure 2. Correlation between gut endogenous retroelement expression and viral load in people living with HIV-1

Scatterplot demonstrating the viral load from serum on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis. uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19) (A). Normalized expression of L1FLnI_1q23.1s in colon biopsies from uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19). Post hoc significance values

595 determined with unpaired T test (B). Normalized expression of L1FLnI_1q23.1s in PBMCs from
596 uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19). Post hoc significance values
597 determined with unpaired T test (C). Interferome array of L1FLnI_1q23.1s expression following
598 IFN-a1 (n = 3), -a2 (n = 3), -a5 (n = 3), -a8 (n = 3), -a14 (n = 3), and -b (n = 3) treated gut CD4+
599 T cells when compared to mock (m) treatment (n=3). ANOVA performed for significance with
600 post hoc significance values determined via unpaired T test (D). IGV representation of
601 L1FLnI_1q23.1s locus provided by the Telescope annotation in the Refseq and Gencode
602 annotations. StringTie reconstructed transcriptomes demonstrate putative transcripts present in
603 sequencing collected from the colon of uninfected controls and PLWH individuals (E).
604 Scatterplot demonstrating the viral load from serum on the X axis and normalized expression of
605 L1FLnI_5q32o in colon biopsies on the Y axis. Uninfected controls (red) (n = 13) vs PLWH
606 (blue) (n = 19) (F). Normalized expression of L1FLnI_5q32o in colon biopsies from uninfected
607 controls (red) (n = 13) vs PLWH (blue) (n = 19). Post hoc significance values determined with
608 unpaired T test (G). Normalized expression of L1FLnI_5q32o in PBMCs from uninfected
609 controls (red) (n = 13) vs PLWH (blue) (n = 19). Post hoc significance values determined with
610 unpaired T test (H). Interferome array of L1FLnI_5q32o expression following IFN-a1 (n = 3), -
611 a2 (n = 3), -a5 (n = 3), -a8 (n = 3), -a14 (n = 3), and -b (n = 3) treated gut CD4+ T cells when
612 compared to mock (m) treatment (n=3). ANOVA performed for significance with post hoc
613 significance values determined via unpaired T test (I). IGV representation of L1FLnI_5q32o
614 locus provided by the Telescope annotation in the Refseq and Gencode annotations. StringTie
615 reconstructed transcriptomes demonstrate putative transcripts present in sequencing collected
616 from the colon of uninfected controls and PLWH individuals (J). Scatterplot demonstrating the
617 viral load from serum on the X axis and normalized expression of LTR19_12p13.31 in colon

biopsies on the Y axis. uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19) (K). Normalized expression of LTR19_12p13.31 in colon biopsies from uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19). Post hoc significance values determined with unpaired T test (L). Normalized expression of LTR19_12p13.31 in PBMCs from uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19). Post hoc significance values determined with unpaired T test (M). Interferome array of L1FLnI_1q23.1s expression following IFN-a1 (n = 3), -a2 (n = 3), -a5 (n = 3), -a8 (n = 3), -a14 (n = 3), and -b (n = 3) treated gut CD4+ T cells when compared to mock (m) treatment (n=3). ANOVA performed for significance with post hoc significance values determined via unpaired T test (N). IGV representation of LTR19_12p13.31 locus provided by the Telescope annotation in the Refseq and Gencode annotations. StringTie reconstructed transcriptomes demonstrate putative transcripts present in sequencing collected from the colon of uninfected controls and PLWH individuals (O).

Figure 3. Correlation Between LTR19_12p13.31 and L1FLnI_1q23.1s expression and immune signatures of the gut microenvironment in people living with HIV-1.

For all samples uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19). Scatterplot demonstrating the CD4+ T Cells per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (A). Scatterplot demonstrating the CD4+ T Cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (B). Scatterplot demonstrating the CD8+ T Cells per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (C). Scatterplot demonstrating the CD8+ T Cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (D). Scatterplot demonstrating the IFNy+CD8+

T Cells per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (E). Scatterplot demonstrating the IFNy+CD8+ T Cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (F). Scatterplot demonstrating the CD11c+ mDCs per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (G). Scatterplot demonstrating the CD11c+ mDCs per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (H). Scatterplot demonstrating the pDCs per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (I). Scatterplot demonstrating the pDCs per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (J). Scatterplot demonstrating the CD1c- mDCs per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (K). Scatterplot demonstrating the CD1c- mDCs per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (L).

Supplemental Legends

Figure S1: Characterization of ERE expression profiles in the colonic microenvironment of PLWH. Principal component analysis (PCA) plot based solely on endogenous retroelement (ERE) abundances demonstrate variance in sample distribution between uninfected controls (red) (n = 13) and PLWH (blue) (n = 19) gut biopsies (A). Stacked bar charts demonstrating mean abundance of exonic, intronic, and intergenic EREs as the average proportion of ERE denoted reads from gut biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (B). Stacked bar charts demonstrating mean abundance of LINE1 and HERV elements as the average proportion of ERE denoted reads from gut biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (C). Stacked bar charts demonstrating the proportional abundance of chromosomes of origin for ERE loci from gut biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (D). Stacked bar charts demonstrating category of EREs the average proportion of TE denoted reads from gut biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (E). Stacked bar charts demonstrating mean abundance of transposable element (TE) families as the average proportion of ERE denoted reads from gut biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (F).

Figure S2: Characterization of ERE expression profiles in uninfected gut CD4+ T cells activated with IFN-Is. PCA plot based solely on ERE abundances demonstrate variance in sample distribution between mock (m) (n=3), IFNα1 (a1) (n = 3), IFNα2 (a2) (n = 3), IFNα5 (a5) (n = 3), IFNα8 (a8) (n = 3), IFNα14 (a14) (n = 3), and IFNβ (b) (n =3) treated gut CD4+ T cells

(A). Stacked bar charts demonstrating mean abundance of exonic, intronic, and intergenic TEs as the average proportion of ERE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (B). Stacked bar charts demonstrating mean abundance of LINE1 and HERV elements as the average proportion of ERE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (C). Stacked bar charts demonstrating the chromosomal location of ERE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (D). Stacked bar charts demonstrating ERE category as the average proportion of TE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (E). Stacked bar charts demonstrating mean abundance of TE families as the average proportion of TE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (F).

Figure S3: IFN-Is induce the differential expression of EREs in gut CD4+ T cells. Heatmap demonstrating the normalized per sample abundances of TEs following the preprocessing filtering of reads in m (gold) (n=3), a1 (purple) (n = 3), a2 (pink) (n = 3), a5 (green) (n = 3), a8 (blue) (n = 3), a14 (red) (n = 3), and b (teal) (n =3) treated gut CD4+ T Cells. ERE reads were filtered to possess at least 2 reads in 10% of the total samples for quality assurance (A). Heatmap demonstrating the normalized per sample abundances of EREs differentially expressed with an adjusted p value of ≤ 0.05 and a LFC ≥ 1 or ≤ -1 between m (red) (n=3) and a1 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (B). Heatmap demonstrating the normalized per sample abundances of EREs differentially

expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between m (red) (n=3) and a2 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (C). Heatmap demonstrating the normalized per sample abundances of EREs differentially expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between m (red) (n=3) and a5 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (D). Heatmap demonstrating the normalized per sample abundances of EREs differentially expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between m (red) (n=3) and a8 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (E). Heatmap demonstrating the normalized per sample abundances of EREs differentially expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between m (red) (n=3) and a14 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (F). Heatmap demonstrating the normalized per sample abundances of EREs differentially expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between m (red) (n=3) and b (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (G). Venn diagram of shared vs uniquely downregulated ERE loci across a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n = 3) treated gut CD4+ T cells when compared to m (H). Venn diagram of shared vs

uniquely upregulated ERE loci across a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n = 3) treated gut CD4+ T cells when compared to m (I).

Figure S4: Upregulation of ERE loci in gut CD4+ T cells activated with IFN-Is. Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between m (n = 3) and a1 (n = 3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (A). Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between m (n = 3) and a2 (n = 3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (B). Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between m (n = 3) and a5 (n = 3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (C). Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between m (n = 3) and a8 (n = 3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (D). Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between m (n = 3) and a14 (n = 3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (E). Volcano plot demonstrating the average fold change ≥1 or≤-1 and an adjusted p value of ≤0.05 and a LFC between m (n = 3) and b (n = 3) treated gut CD4+

T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction (F).

Figure S5: HIV-1 status and IFN-Is induce 7 shared EREs in the gut microenvironment. Venn diagram of shared vs uniquely downregulated TE loci across PLWH (n = 19) vs uninfected controls (n = 13) gut biopsies, and a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n = 3) treated gut CD4+ T cells when compared to m (A). Venn diagram of shared vs uniquely upregulated TE loci across PLWH (n = 19) vs uninfected controls (n = 13) gut biopsies, and a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n = 3) treated gut CD4+ T cells when compared to m (B). Gut biopsies from uninfected controls (n = 13) and PLWH (n = 19) individuals (left) as well interferome assay responses to type 1 interferons (n = 3) (right) show significant upregulation in HARLEQUIN_17q12 (C), HERV4_4q22.1 (D), MER101_2p25.2 (E), MER4_7q21.2 (F), L1FLnI_1p22.2l (G), L1FLnI_9p21.1ha (H), and L1FLnI_17q12k (I). All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg correction. Tukey box and whisker plots for with the line demonstrating the median of expression while the large dot represents the mean.

Figure S6: EREs of interest in the gut microenvironment show no association with CD1c+ mDC abundances. Gut expression of L1FLnI_1q23.1s (A), L1FLnI_5q32o (B), and LTR19_12p13.31 (C) show no association with CD1c+ mDC abundances in the gut microenvironment.

Figure S7: L1FLnI_5q32o is not associated with major cellular parameters of immune deregulation in PLWH. Scatterplot demonstrating the CD4+ T Cells per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (A). Scatterplot demonstrating the CD8+ T Cells per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (B). Scatterplot demonstrating the IFNy+CD8+ T Cells per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (C). Scatterplot demonstrating the CD11c+ mDCs per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (D). Scatterplot demonstrating the pDCs per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (E). Scatterplot demonstrating the CD1c- mDCs per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (F).

Figure S8: Association between colonic L1FLnI_1q23.1s expression and CD4+ T Cell subsets. Scatterplot demonstrating the IFNy-IL17+CD4+ T cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (A). Scatterplot demonstrating the IFNy+IL17-CD4+ T cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (B). Scatterplot demonstrating the IFNy+IL17+CD4+ T cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (C). Scatterplot demonstrating the IL22+IFNy-IL17-CD4+ T cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (D).

799 **Figure S9: Association between colonic L1FLnI_5q32o expression and CD4+ T Cell subsets**.
800 Scatterplot demonstrating the IFNy-IL17+CD4+ T cells per gram on the X axis and normalized
801 expression of L1FLnI_5q32o in colon biopsies on the Y axis (A). Scatterplot demonstrating the
802 IFNy+IL17-CD4+ T cells per gram on the X axis and normalized expression of L1FLnI_5q32o
803 in colon biopsies on the Y axis (B). Scatterplot demonstrating the IFNy+IL17+CD4+ T cells per
804 gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis
805 (C). Scatterplot demonstrating the IL22+IFNy-IL17-CD4+ T cells per gram on the X axis and
806 normalized expression of L1FLnI_5q32o 1s in colon biopsies on the Y axis (D).
807
808 **Figure S10: Association between colonic LTR19_12p13.31 expression and CD4+ T Cell**
809 **subsets**. Scatterplot demonstrating the IFNy-IL17+CD4+ T cells per gram on the X axis and
810 normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (A). Scatterplot
811 demonstrating the IFNy+IL17-CD4+ T cells per gram on the X axis and normalized expression
812 of LTR19_12p13.31 in colon biopsies on the Y axis (B). Scatterplot demonstrating the
813 IFNy+IL17+CD4+ T cells per gram on the X axis and normalized expression of
814 LTR19_12p13.31 in colon biopsies on the Y axis (C). Scatterplot demonstrating the IL22+IFNy-
815 IL17-CD4+ T cells per gram on the X axis and normalized expression of LTR19_12p13.31 in
816 colon biopsies on the Y axis (D).
817
818 **Figure S11: Colonic expression of EREs of interest produce negligible correlations with**
819 **serum IL-6 and LPS**. Scatterplots demonstrating serum lipopolysaccharide (LPS)
820 concentrations on the X axis and normalized expression of L1FLnI_1q23.1s (A), L1FLnI_5q32o
821 (B), LTR19_12p13.31 (C) in colon biopsies on the Y axis. Scatterplots demonstrating serum IL-6

822 concentrations on the X axis and normalized expression of L1FLnI_1q23.1s (D), L1FLnI_5q32o
823 (E), LTR19_12p13.31 (F) in colon biopsies on the Y axis.
824
825 **Figure S12: HIV-1 status influences endogenous retroelement expression in the PBMC**
826 **fraction.** Heatmap demonstrating the normalized per sample abundances of EREs following the
827 preprocessing filtering of reads in uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19)
828 individuals. ERE reads were filtered to possess at least 2 reads in 10% of the total samples for
829 quality assurance (A). Heatmap demonstrating the normalized per sample abundances of EREs
830 differentially expressed with an adjusted p value of ≤0.05 and a LFC ≥1 or≤-1 between
831 uninfected controls (red) (n = 13) and PLWH (blue) (n = 19) gut biopsies. All statistics were
832 performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default
833 parameters for a Benjamini-Hochberg correction (B). PCA plot based solely on ERE abundances
834 demonstrate no variance in sample distribution between uninfected controls (red) (n = 13) and
835 PLWH (blue) (n = 19) PBMCs (C).
836
837
838
839
840
841
842
843
844

845 **Supplemental Table**

846

	Controls	PLWH	P value
Number of study participants	13	19	
Age			
Median (min, max) (years)	33 (23, 54)	33 (22, 58)	0.96
Sex			
Male	8	13	0.72
Female	5	6	
Ethnicity			
Non-Hispanic	11	14	0.67
Hispanic	2	5	
Race			
White/Caucasian	9	15	0.62
Black/African American	2	3	
Asian	2	1	
Men who have sex with men (MSM)			
Yes	1	12	0.003
No	12	7	
Reported intravenous drug use (IDU)*			
Yes	0	4	0.11
No	14	14	
Blood CD4 count			
Median (min, max) (cells/μl)	720 (468,1071)	429 (221, 836)	0.002
Plasma *HIV RNA copies/ml*			

Median (min, max)	N/A	26000 (2880, 196000)
Years since first HIV-1 seropositive test		
Median (min, max)	N/A	3.5 (0.17, 15.0)

Study participants were selected from a completed clinical study[53]. Values are shown as median (range) for continuous measures and numbers for categorical measures. Statistical analyses were performed using Mann-Whitney test for continuous characteristics and Fishers exact test to compare categorical characteristics between control and PWH groups. *not reported for one person in PWH cohort. N/A: not applicable.

Research in Context

Evidence before this study

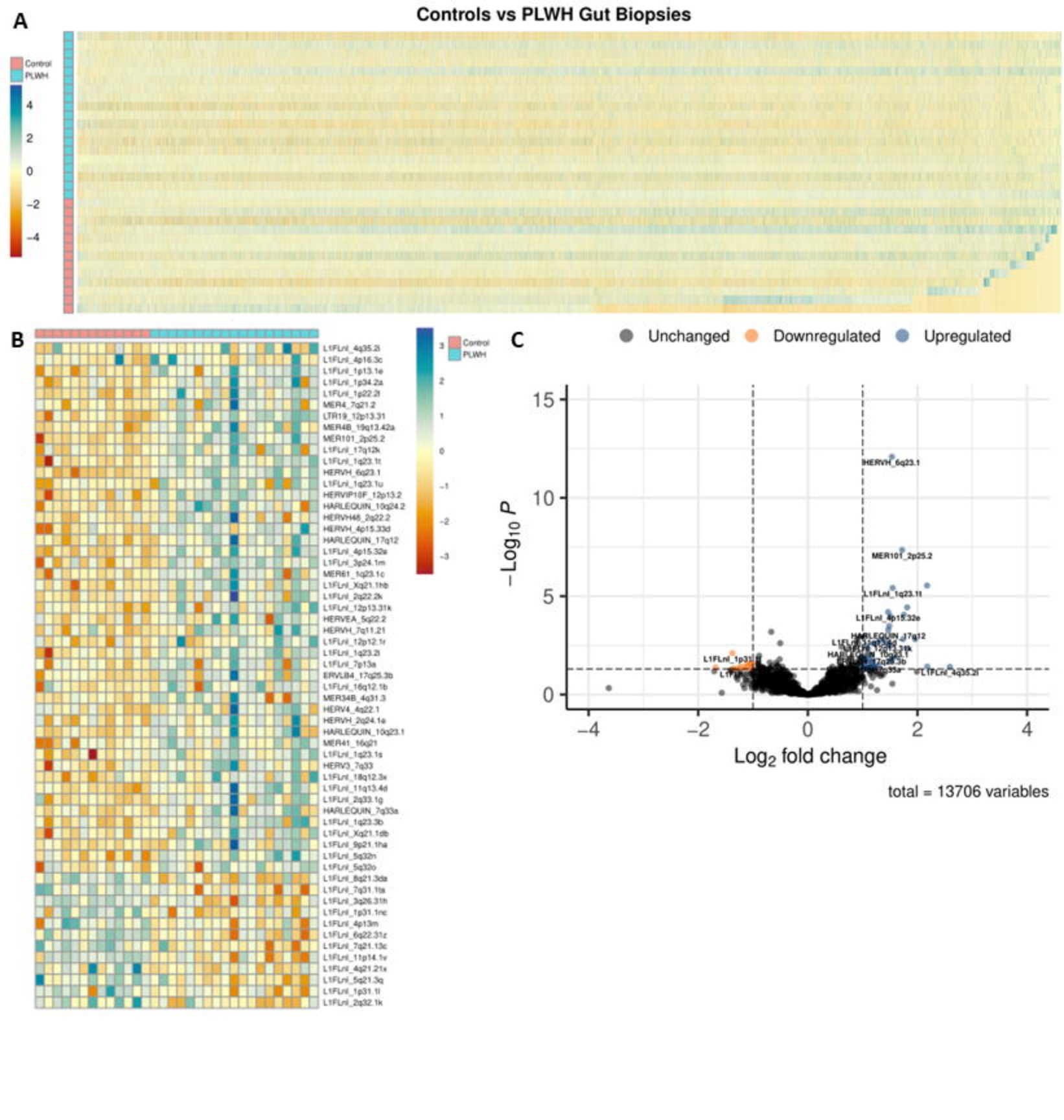
Previous studies have demonstrated that HIV-1 infection alters the physiology of the GI tissues, where it forms quiescent reservoirs. In chronic HIV-1, the sustainment of antiviral immune responses through poorly understood mechanisms likely contributes to pathogenesis and prohibits effective clearance of the viral reservoirs. EREs comprise a substantial portion of the human genome that has been implicated in HIV-1 pathogenesis but remains understudied in comparison to coding genes. Therefore, studying the activity of EREs in tissues such as the GI tract may help further define the physiology of immunopathologies that contribute to HIV-1 reservoir persistence.

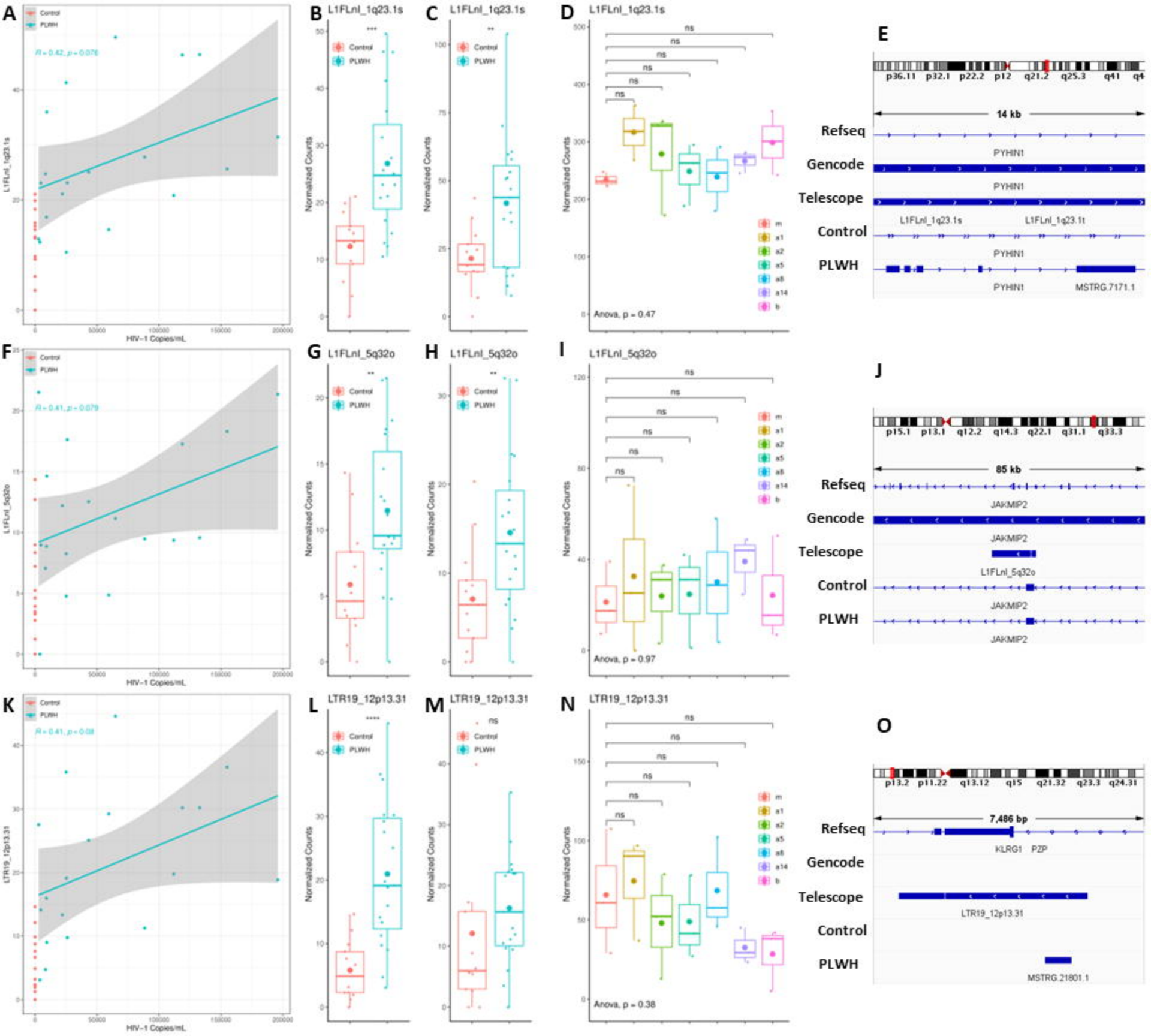
Added value of this study

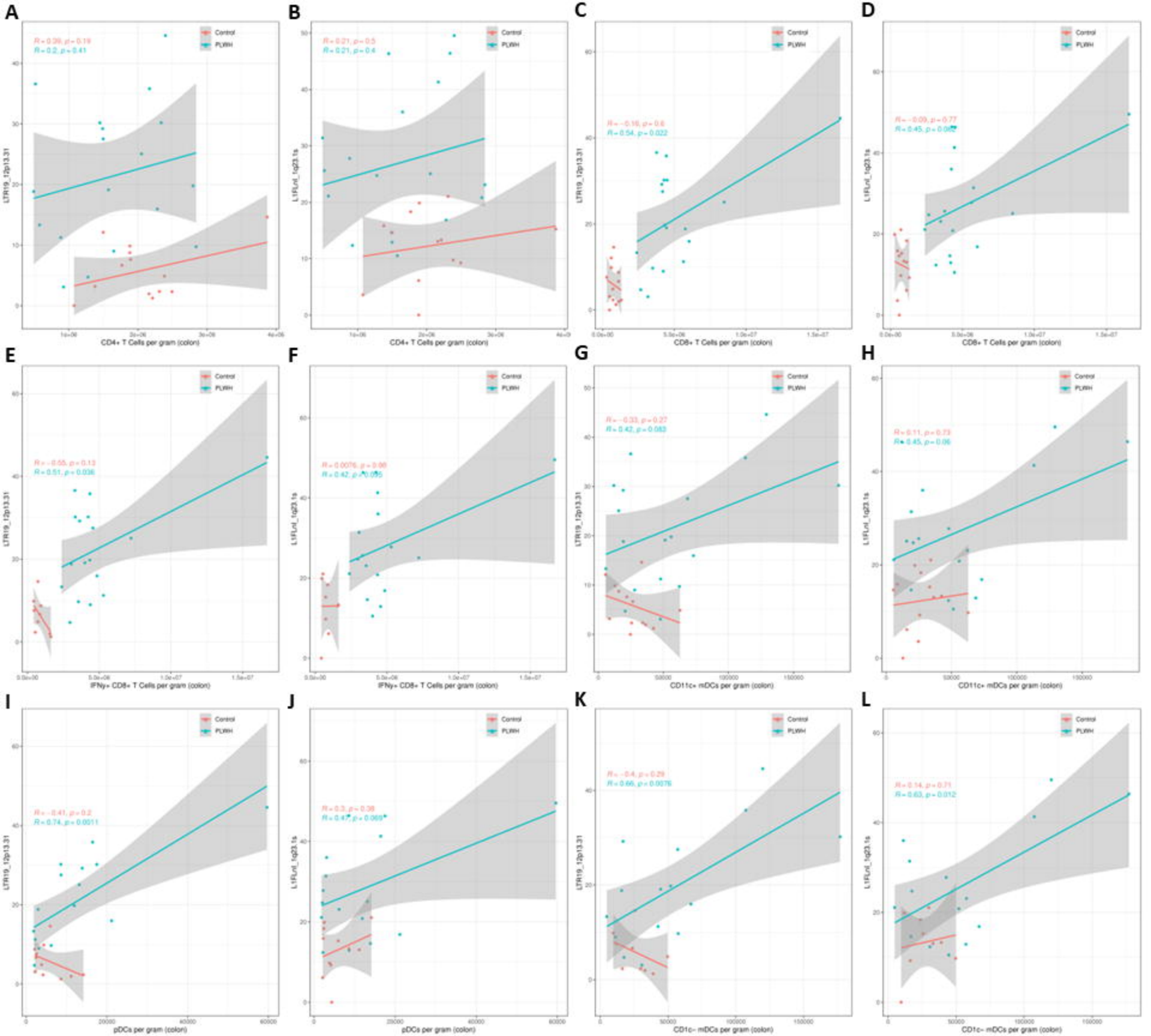
Colon pinch biopsies from PLWH demonstrate differential expression of 59 EREs when compared to uninfected controls. Of these 59, 3 possess near-significant correlations with viral load. Further analysis into cellular composition of the gut microenvironment demonstrates that 2 the EREs LTR19_12p13.31 and L1FLnI_1q23.1s possess significant associations with immune cell phenotypes indicative of localized HIV-1 pathogenesis. Therefore, high expression of LTR19_12p13.31 and L1FLnI_1q23.1s can function as markers of GI HIV-1 reservoirs, as well as potentially possess undefined roles in human health.

Implications of all the available evidence

Recurrent observations suggest pathophysiological roles for differentially expressed EREs in diseased tissues. In HIV-1 infection, EREs can be over expressed due to trans-activation of HERV LTRs and shifts in the regulome that permit the expression of typically silent elements. However, it has not been established how EREs may be deregulated in the gut where immunopathologies persist surrounding viral reservoirs. Correlations between ERE expression and cellular composition demonstrate that certain EREs may be indicative of cell type abundances thought to drive GI-specific immunopathologies resultant from HIV-1 infection.


A
Controls vs PLWH Gut Biopsies
Control
PLWH
4
2
0
−2
−4
B
Control
PLWH
3
2
1
0
−1
−2
−3
L1FLnI_4q35.2i
L1FLnI_4p16.3c
L1FLnI_1p13.1e
L1FLnI_1p34.2a
L1FLnI_1p22.2i
MER4_7q21.2
LTR19_12p13.31
MER4B_19q13.42a
MER101_2p25.2
L1FLnI_17q12k
L1FLnI_1q23.1t
HERVH_6q23.1
L1FLnI_1q23.1u
HERVIP10F_12p13.2
HARLEQUIN_10q24.2
HERVH48_2q22.2
HERVH_4p15.33d
HARLEQUIN_17q12
L1FLnI_4p15.32e
L1FLnI_3p24.1m
MER61_1q23.1c
L1FLnI_Xq21.1hb
L1FLnI_2q22.2k
L1FLnI_12p13.31k
HERVEA_5q22.2
HERVH_7q11.21
L1FLnI_12p12.1r
L1FLnI_1q23.2i
L1FLnI_7p13a
ERVLB4_17q25.3b
L1FLnI_16q12.1b
MER34B_4q31.3
HERV4_4q22.1
HERVH_2q24.1e
HARLEQUIN_10q23.1
MER41_16q21
L1FLnI_1q23.1s
HERV3_7q33
L1FLnI_18q12.3x
L1FLnI_11q13.4d
L1FLnI_2q33.1g
HARLEQUIN_7q33a
L1FLnI_1q23.3b
L1FLnI_Xq21.1db
L1FLnI_9p21.1ha
L1FLnI_5q32n
L1FLnI_5q32o
L1FLnI_8q21.3da
L1FLnI_7q31.1ta
L1FLnI_3q26.31h
L1FLnI_1p31.1nc
L1FLnI_4p13m
L1FLnI_6q22.31z
L1FLnI_7q21.13c
L1FLnI_11p14.1v
L1FLnI_4q21.21x
L1FLnI_5q21.3q
L1FLnI_1p31.1l
L1FLnI_2q32.1k
C
Unchanged
Downregulated
Upregulated
15
10
5
0
−4
−2
0
2
4
−Log10 P
Log2 fold change
HERVH_6q23.1
MER101_2p25.2
L1FLnI_1q23.1t
L1FLnI_4p15.32e
HARLEQUIN_17q12
L1FLnI_1p31.1nc
L1FLnI_4q35.2i
total = 13706 variables


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
Control
PLWH
HIV−1 Copies/mL
Normalized Counts
L1FLnI_1q23.1s
L1FLnI_5q32o
LTR19_12p13.31
Anova, p = 0.47
Anova, p = 0.97
Anova, p = 0.38
Refseq
Gencode
Telescope
Control
PLWH
14 kb
85 kb
7,486 bp
PYHIN1
MSTRG.7171.1
JAKMIP2
KLRG1
PZP
MSTRG.21801.1


A
B
C
D
E
F
G
H
I
J
K
L
Control
PLWH
LTR19_12p13.31
L1FLnI_1q23.1s
CD4+ T Cells per gram (colon)
CD8+ T Cells per gram (colon)
IFNγ+ CD8+ T Cells per gram (colon)
CD11c+ mDCs per gram (colon)
pDCs per gram (colon)
CD1c- mDCs per gram (colon)