

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

2 1

3

4 Nicholas Dopkins*¹, Tongyi Fei¹, Stephanie Michael¹, Nicholas Liotta¹, Kejun Guo^{2,3,4}, Kaylee

5 L. Mickens^{2,3,4}, Brad S. Barrett^{2,3,4}, Matthew L. Bendall¹, Stephanie M. Dillon^{2,3,4}, Cara C.

6 Wilson^{2,3,4}, Mario L. Santiago^{2,3,4} and Douglas F. Nixon¹

7

8 Keywords: human endogenous retrovirus (HERV), interferons, human immunodeficiency virus

9 type I (HIV-I), Endogenous Retroelement (ERE), Long interspersed nuclear element (LINE)

10

11 ¹Division of Infectious Diseases, Department of Medicine, Weill Cornell Medicine, New York,

12 NY, United States of America.

13 ²Department of Medicine, University of Colorado School of Medicine, Aurora, CO, United

14 States of America.

15 ³RNA Bioscience Initiative, University of Colorado School of Medicine, Aurora, CO, United

16 States of America.

17 ⁴Department of Immunology and Microbiology, University of Colorado School of Medicine,

18 Aurora, CO, United States of America.

19

20 **Correspondence:**

21 Nicholas Dopkins, PhD

22 Email: nid4009@med.cornell.edu

23

24 **Summary**

25

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

35

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

41

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

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

50

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

55

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

57

58

59

60

61

62

63

64

65

66

67

68

69

70 **Introduction**

71 Human immunodeficiency virus type I (HIV-1) is a retrovirus that infects immune cells
72 by binding to receptors on the cell surface¹. HIV-1 infects host cells by integrating a copy of the
73 viral genome into the host genome, resulting in proviruses that then hijack the host cell's
74 replicative machinery to produce the building blocks of infectious particles¹. The HIV-1 lifecycle
75 may rapidly deplete host CD4+ T cell populations and disrupt global immune dynamics².
76 Disruptions to the host immune system predispose people living with HIV-1 (PLWH) to life-
77 threatening comorbidities³. Since 1996, life expectancy and quality of life for PLWH have risen
78 due to the advent of antiretroviral therapies (ARTs) that pharmacologically inhibit key processes
79 of the retroviral life cycle⁴. Unfortunately, ART merely suppresses retroviral processes,
80 emphasizing the need for cure strategies that rid individuals of latent proviral reservoirs which
81 may resume infectious particle formation after a lapse in treatment^{5,6}. One such reservoir site is
82 the gastrointestinal (GI) associated lymphoid tissue (GALT)^{7,8}, which is particularly rich in
83 CD4+ T cells⁹. HIV-1 induced immunopathologies in the GALT are associated with severe
84 comorbidities in PLWH^{10,11}. Many of these pathologies could be sustained by the accumulation
85 of plasmacytoid dendritic cells (pDCs) that are trafficked to the GI tract during chronic HIV-1
86 infection¹². Models of Simian Immunodeficiency Virus (SIV), a retrovirus closely related to
87 HIV-1, demonstrate that pDCs traffic from the circulation, where they are depleted, to the GALT,
88 where their overabundance contributes to immunopathologies that allow the site to become the
89 primary source of viral replication^{13,14}. pDCs reflect a unique subset of DCs whose primary
90 function is not antigen presentation but is instead immunomodulation through the release of
91 cytokines¹⁵. In the acute stages of HIV-1 infection, pDCs produce type 1 interferons (IFN-Is) as
92 one of the first lines of defense against the virus¹⁶, and are therefore protective. However, the

93 sustainment of their IFN-I production during the chronic stages of infection exhausts and
94 deregulates antiviral immunity, negatively impacting the health of PLWH^{12,16-18}.

95 To better characterize how immunopathologies in PLWH develop, we investigated how
96 endogenous retroelement (ERE) expression may underlie the development of HIV-1 induced
97 immune deregulation. ERs are highly abundant sequences estimated to make up roughly ~40%
98 of the human genome¹⁹. In humans, ERs include long interspersed nuclear elements (LINES),
99 short interspersed nuclear elements (SINEs), and human endogenous retroviruses (HERVs)²⁰.
100 Previous studies have demonstrated that ERE activity is highly susceptible to change during
101 infection with HIV-1²¹⁻³¹. While IFN-I signaling may activate ERs proximal to interferon
102 stimulated genes (ISGs), other ERs, such as HERV elements, can be directly trans-activated by
103 HIV-1 cofactors due to high sequence homology between endogenous and exogenous
104 retroviruses^{21,22,24,32,33}. This homology can be so substantial that certain HERV-K18 and HERV-
105 W products can be incorporated in the assembly of HIV-1 particles³⁴⁻³⁸. ERE activity also
106 functions as a major regulator of transcriptional networks that have been co-opted in organismal
107 physiology^{39,40}. In immunity, examples of ERE activity cooption suggest their intimate
108 association with the host mounted responses^{41,42}. Recurrent observations continually reinforce
109 hypotheses that ERs likely participate in human health and disease through poorly understood
110 mechanisms⁴²⁻⁴⁶. The activity of multiple HERVs²¹⁻²⁸ and LINES^{29,30} is altered during HIV-1
111 infection, which then impacts host gene expression²⁴, packaging of HIV-1 particles³⁴⁻³⁸, and
112 immunity^{25,47}. For the purpose of further defining HIV-1 mediated gut pathogenesis, we sought
113 to characterize how ERE activity at the RNA level is modulated in the gut microenvironment of
114 PLWH. Since ERs have been largely overlooked in genomic approaches to classify underlying

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

117 ERE expression at the RNA level was quantified in colon biopsies collected from a
118 cohort of viremic individuals prior to or at the initiation of ART¹⁷ to identify those best
119 correlating with viral load. We next overlaid these findings with IFN-I stimulated gut CD4+ T
120 cells to estimate if the differential expression of EREs in the gut of PLWH is driven by the HIV-1
121 lifecycle or by innate immunopathologies. Collectively, our analyses demonstrate that a subset of
122 ERE loci are impacted in PLWH. The bulk of PLWH-upregulated EREs in the gut are not
123 upregulated by IFN-I stimulation of gut CD4+ T cells, suggesting their activation by other
124 aspects of the HIV-1 life cycle or the host response. Two identified EREs, LTR19_12p13.31 and
125 L1FLnI_1q23.1s, further correlated with myeloid and lymphoid immune cell abundances.
126 Collectively, incorporation of the GI-specific expression of LTR19_12p13.31 and
127 L1FLnI_1q23.1s can better define HIV-1 driven immunopathologies of the GALT, and the
128 expression of these elements may possess underlying roles in pathogenesis.

129

130 **Methods**

131 *Locus-specific quantification of ERE expression at the RNA level*

132 RNA sequencing FASTQ files of colon pinch biopsies and interferome arrays of
133 uninfected CD4+ T cells were obtained from a previously published study¹⁷ under the Bioproject
134 accession numbers “PRJNA558500” and “PRJNA558974”, respectively. RNA sequencing
135 FASTQ files performed on the PBMC fraction from the same participants are available under the
136 Bioproject accession number “PRJNA659373”. FASTQ files were aligned to the human genome
137 build 38 (hg38) using STAR (v2.7.9.a)⁴⁸. STAR alignment was performed using the parameters

138 “--outSAMstrandField intronMotif --outFilterMultimapNmax 200 --winAnchorMultimapNmax
139 200” for the purpose of capturing multimapping ERE reads. We then used Telescope (v1.0.3)⁴⁹,
140 which utilizes an expectation-maximization algorithm to improve upon the definition of
141 ubiquitously mapping ERE transcripts to a custom annotation, for locus-specific approximation
142 of ERE expression. The Telescope assign module was performed using the parameters “--
143 theta_prior 200000 -- max_iter 1000” to align reads to a custom annotation of EREs accessible at
144 https://github.com/mlbendall/telescope_annotation_db. Metadata for the ERE annotation
145 predicting the intronic, exonic, and intergenic status of individual loci in the Telescope
146 annotation relative to coding gene annotations can be found at
147 https://github.com/liniguez/Telescope_MetaAnnotations. Original code produced for analysis
148 can be accessed at <https://github.com/NicholasDopkins/JuneHIVMucosa>.

149

150 *Differential expression analysis*

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

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

169

170 *Integrative Genomics Viewer*

171 National Center for Biotechnology Information (NCBI) Reference Sequence Database
172 (Refseq)⁵⁵, ENCyclopedia Of DNA Elements⁵⁶ (Gencode), and Telescope⁴⁹ annotations of the
173 human genome were loaded into Integrative Genomics Viewer⁵⁷ (IGV) for alignment of
174 StringTie⁵⁸ reconstructed transcriptomes from uninfected controls and PLWH samples.

175

176 *Statistics*

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

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

186

187 *Clinical cohort*

188 ERE expression analyses were conducted using FASTQ files from previously generated RNAseq
189 data collected from colonic tissue and PBMC fractions collected from 19 PLWH and 13 age- and
190 sex-matched controls, and associations with features of HIV-1 pathogenesis were determined
191 using archived datasets of colon and systemic immunological and virological parameters^{17,18,53,54}.
192 On average, PLWH had been infected with HIV-1 (defined by the first HIV-1 seropositive test)
193 for 5.25 ± 1.1 (mean \pm SEM) years. PLWH were ART-naive or were not on ART for more than 7
194 days in the preceding 6 months, and had CD4 T cell counts >200 cells/ μ l within 3 months of
195 clinical visit. Exclusion criteria for both cohorts are extensively detailed elsewhere⁵³. In addition
196 to blood samples, study participants underwent a flexible sigmoidoscopy with multiple colon
197 pinch biopsies obtained. Clinical characteristics for the study participants included in this current
198 study are detailed in Supplemental Table 1.

199

200 *Immune system characterization of PLWH and uninfected controls*

201 Methods used to generate the archived datasets have been extensively detailed^{17,18,53,54,59,60}. In
202 brief, indicators of systemic inflammation were assessed with plasma IL-6 concentrations by
203 ELISA, and microbial translocation was measured by plasma LPS concentrations with the
204 Limulus Amebocyte Lysate assay⁵³. Frequencies of colon mDCs, pDCs, and T cells were
205 determined using multi-color flow cytometry^{53,54} and expressed as an absolute number per gram
206 of tissue based on the frequency within viable CD45+ cells, initial cell counts, and biopsy

207 weights. Colon tissue HIV-1 RNA was quantified by real-time PCR and HIV-1 RNA copy
208 numbers normalized per CD4 T cell within each biopsy calculated by the percent of all viable
209 cells that were CD45+CD3+CD4+ as determined by flow cytometry and weight of each biopsy
210 and reported as HIV-1 RNA per million CD4 T cell⁵³.

211

212 *Ethics Statement*

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

217

218 *Role of Funders*

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

221

222 **Results**

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

224 Following preprocessing to remove lowly abundant transcripts, we quantified the
225 expression of 13,706 EREs in colon pinch biopsies collected from PLWH in comparison to age
226 and sex-matched HIV-1-uninfected controls (Fig. 1a). Significance in differential expression
227 between controls and PLWH was found for 59 ERE loci (Fig. 1b and Fig. 1c). Of these 59 loci,
228 47 were upregulated and 12 were downregulated in PLWH. All HERVs differentially expressed
229 were upregulated in PLWH when compared to controls. Overall, ERE activity at the RNA level

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

244

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

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

253 controls (Fig. 2b). In PBMCs collected from the same participants, expression of
254 L1FLnI_1q23.1s is significantly upregulated in concurrence with gut expression of the element
255 (Fig. 2c). Analysis of IFN-I activated gut CD4+ T cells suggests that IFN alpha 1 (IFNa1) and
256 IFN beta (IFNb) may contribute to the upregulated expression of L1FLnI_1q23.1s, however no
257 IFN-Is yielded significant effects (Fig. 2d). IGV demonstrates that the L1FLnI_1q23.1s locus is
258 an ERE within the Pyrin And HIN Domain Family Member 1 (PYHIN1) gene with minor
259 differences in the putative alignments provided by StringTie for uninfected controls and PLWH
260 samples (Fig. 2e). L1FLnI_5q32o possessed a positive association with viral load in the
261 circulation (Fig. 2f) and is upregulated in the colon of PLWH when compared to controls (Fig.
262 2g). In PBMCs collected from the same study participants, expression of L1FLnI_5q32o is
263 significantly upregulated in concurrence with gut expression of the element (Fig. 2h). Analysis
264 of IFN-I activated gut CD4+ T cells suggests that IFN-Is yield no significant effects on the
265 expression of L1FLnI_5q32o (Fig. 2i). IGV demonstrates that the L1FLnI_5q32o locus is an
266 ERE within the Janus Kinase And Microtubule Interacting Protein 2 (JAKMIP2) gene with no
267 differences in the putative alignments provided by StringTie for PLWH and control samples
268 (Fig. 2j). LTR19_12p13.31 possessed a positive association with viral load in the circulation
269 (Fig. 2k) and is upregulated in the colon of PLWH when compared to controls (Fig. 2l). In
270 PBMCs collected from the same study participants, expression of LTR19_12p13.31 shows a
271 non-significant upregulation (Fig. 2m). Analysis of IFN-I activated gut CD4+ T cells suggests
272 that IFN-Is yield no significant effects on the expression of LTR19_12p13.31 (Fig. 2n). IGV
273 demonstrates that LTR19_12p13.31 locus is an ERE that overlaps the Pregnancy Zone Protein
274 (PZP) and Killer Cell Lectin Like Receptor G1 (KLRG1) genes with a unique transcript in the
275 putative alignments of PLWH samples provided by StringTie (Fig. 2o). Collectively, expression

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

279
280 *Correlation Between LTR19_12p13.31 and L1FLnI_1q23.1s expression and immune signatures*
281 *of the gut microenvironment in people living with HIV-1*

282 We next analyzed how expression of L1FLnI_1q23.1s, L1FLnI_5q32o and
283 LTR19_12p13.31 correlate with cellular abundances that may be associated with pathogenesis in
284 the gut of PLWH. There were no significant associations between the number colon CD4+ T
285 cells (measured per gram of gut tissue)with LTR19_12p13.31 (Fig. 3a) and L1FLnI_1q23.1s
286 expression (Fig. 3b). LTR19_12p13.31 demonstrated significant positive correlations with the
287 number of colon CD8+ T cells (Fig. 3c) and IFN gamma (IFN γ)+CD8+ T cells (Fig. 3e),
288 whereas L1FLnI_1q23.1s demonstrated near significant positive correlations with CD8+ (Fig.
289 3d) and IFN γ +CD8+ T cell (Fig. 3f) abundances. Further consideration into dendritic cell (DC)
290 abundances demonstrate that LTR19_12p13.31 (Fig. 3g) and L1FLnI_1q23.1s (Fig. 3h) show
291 near significant positive correlations with CD11c+ myeloid DCs (mDCs), while
292 LTR19_12p13.31 (Fig. 3i) displayed a significant positive correlation and L1FLnI_1q23.1s (Fig.
293 3j) displayed a near significant positive association with plasmacytoid DCs (pDCs). Amongst
294 mDCs, LTR19_12p13.31 (Fig. 3k) and L1FLnI_1q23.1s (Fig. 3l) show significant positive
295 correlations with the CD1c- mDCs, but not with CD1c+ mDCs (Fig. S6), suggesting potential
296 perturbations related to specific mDC subsets⁵⁴. Expression of the ERE L1FLnI_5q32o did not
297 demonstrate positive or negative correlations with CD4+ T cells, CD8+ T cells, IFN γ +CD8+ T
298 cells, CD11c+ mDCs, pDCs, or CD1c- mDCs in PLWH (Fig. S7). In analyses of frequencies of

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

310

311 **Discussion**

312 In this study, we leverage modulations in ERE activity at the RNA level in the GI tract of
313 untreated PLWH with frequencies of immune cell populations that are likely critical players in
314 driving HIV-1 gut pathogenesis¹¹. As the derepression of EREs may possess immunomodulatory
315 roles in human diseases^{42,61}, our results suggest further investigation into their activity in HIV-1
316 pathogenesis. By analysis of RNA sequencing from colon biopsies of uninfected controls vs
317 untreated PLWH, we identified 59 differentially expressed EREs associated with HIV. By
318 finding correlations between their colonic expression and circulating viral load, we were able to
319 narrow further investigations to focus on 3 EREs upregulated in PLWH that might indicate the
320 severity of HIV-1 pathogenesis in the gut, or function as markers of immune deregulation. The
321 EREs LTR19_12p13.31 and L1FLnI_1q23.1s best correlated with cellular abundances including

322 higher frequencies of pDCs, CTLs, and CD1c- mDCs. L1FLnI_5q32o demonstrated no
323 associations with these cellular markers, however, did show associations with differential
324 abundances of CD4+ T cell Th subsets that were not found with LTR19_12p13.31 and
325 L1FLnI_1q23.1s expression. By considering their expression in PBMCs from the same study
326 participants, we observed a significant upregulation in L1FLnI_1q23.1s expression and a non-
327 significant upregulation in LTR19_12p13.31. In accordance with previous studies²², we observe
328 significant alterations in ERE activity at the RNA level in the PBMC fraction of PLWH and
329 uninfected controls (Fig. S12). This suggests future use of L1FLnI_1q23.1s as a potential
330 circulating biomarker for deregulations of immunity in the gut microenvironment of PLWH,
331 while LTR19_12p13.31 is more likely an intrinsic marker specific to the GI tract. Collectively,
332 we find significant associations between LTR19_12p13.31 and L1FLnI_1q23.1s and cellular
333 dynamics of the GALT in PLWH, further emphasizing their identity as biomarkers or potential
334 mediators of immune deregulation.

335 Previous studies have demonstrated that ERE activity is widely changed in response to
336 the HIV-1 lifecycle^{21,22,27,29,31} and certain HIV-1 associated comorbidities^{62,63}. ERE expression is
337 also susceptible to changes in epigenetics and immunity that are elicited during viral
338 infections and the host response⁶⁴. While it remains unknown if the changes in expression for
339 most EREs result from a bystander effect or by a cooption of host immunity in HIV-1 infection,
340 their activity can be incorporated to better characterize poorly understood immunopathologies in
341 viral infections. Here, we demonstrate that significant correlations can be established between
342 differentially expressed EREs and cellular markers that likely contribute to GI
343 immunopathologies of untreated HIV-1 infection. Some limitations of this study concern bulk
344 RNA sequencing approaches utilized, leaving the data unable to discern what cell types are

345 responsible for the overexpression of differentially expressed EREs. As the cellular composition
346 of the sequenced tissues fluctuates with disease severity, it is unclear whether the phenotype or
347 mere abundance of disproportionate cell types, such as pDCs and CTLs, is responsible for their
348 expression. Future studies that accurately quantify ERE expression from single cell RNA
349 sequencing data that were not available from this cohort would provide further insight into the
350 activation state of these EREs in the GI tract of PLWH. Additionally, due to the poorly defined
351 molecular characteristics and low copy number of LTR19 HERV sequences, exceedingly little is
352 known about their potential roles in health and disease. LTR19 sequences are a member of the
353 HERVFA clade and are estimated to only possess roughly ~15 copies in the human genome⁶⁵.

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

366

367 **Contributors**

368 ND co-designed the study, co-prepared the first manuscript draft, co-performed data analyses,
369 co-curated code, led statistical analyses, and curated figures. TF co-prepared the first manuscript
370 draft, co-performed data analyses, co-performed statistical analysis, co-curated code, and
371 provided intellectual input critical to analysis. SM co-designed the study, co-prepared the first
372 manuscript draft, co-performed data analyses, and provided intellectual input critical to data
373 interpretation. NL co-performed analyses and co-curated code. KG provided intellectual input
374 critical to hypothesis generation and co-provided data access. KLM provided intellectual input
375 critical to hypothesis generation and co-provided data access. BSB provided intellectual input
376 critical to hypothesis generation and co-provided data access. MLB provided intellectual input
377 critical to hypothesis testing. SMD provided intellectual input critical to hypothesis generation
378 and testing, co-provided data access, and performed immunological analysis of the gut
379 microenvironment. CCW provided intellectual input critical to hypothesis generation and co-
380 provided data access. DFN co-designed the study and co-prepared the first manuscript draft.
381 MLS co-designed the study, co-prepared the first manuscript draft, and co-provided data access.
382 All authors contributed to data interpretation, discussion regarding significances, and editing of
383 the manuscript. All authors have read and approved the final manuscript.

384

385 **Declaration of Interests**

386 All authors have no potential conflicts of interests to disclose.

387

388 **Acknowledgments**

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

391

392 **Data Sharing Statement**

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

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414 **References**

- 415 1. Justiz Vaillant, A. A. & Gulick, P. G. HIV Disease Current Practice. in *StatPearls* (StatPearls
416 Publishing, 2023).
- 417 2. Boasso, A., Shearer, G. M. & Chougnnet, C. Immune dysregulation in human
418 immunodeficiency virus infection: know it, fix it, prevent it? *J. Intern. Med.* **265**, 78–96
419 (2009).
- 420 3. Deeks, S. G., Overbaugh, J., Phillips, A. & Buchbinder, S. HIV infection. *Nat. Rev. Dis.*
421 *Primer* **1**, 1–22 (2015).
- 422 4. Life expectancy of individuals on combination antiretroviral therapy in high-income
423 countries: a collaborative analysis of 14 cohort studies. *Lancet* **372**, 293–299 (2008).
- 424 5. Sengupta, S. & Siliciano, R. F. Targeting the latent reservoir for HIV-1. *Immunity* **48**, 872–895
425 (2018).
- 426 6. Veenhuis, R. T. *et al.* Monocyte-derived macrophages contain persistent latent HIV reservoirs.
427 *Nat. Microbiol.* **8**, 833–844 (2023).
- 428 7. Thompson, C. G., Gay, C. L. & Kashuba, A. D. M. HIV Persistence in Gut-Associated
429 Lymphoid Tissues: Pharmacological Challenges and Opportunities. *AIDS Res. Hum.*
430 *Retroviruses* **33**, 513–523 (2017).
- 431 8. Kelley, C. F. *et al.* HIV-1 RNA Rectal Shedding Is Reduced in Men With Low Plasma HIV-1
432 RNA Viral Loads and Is Not Enhanced by Sexually Transmitted Bacterial Infections of the
433 Rectum. *J. Infect. Dis.* **204**, 761–767 (2011).
- 434 9. Mowat, A. McI. & Viney, J. L. The anatomical basis of intestinal immunity. *Immunol. Rev.*
435 **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 $\alpha 4\beta 7$ 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 IFN α subtypes and IFN β 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).

- 459 19. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**,
460 860–921 (2001).
- 461 20. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome*
462 *Biol.* **19**, 199 (2018).
- 463 21. Bhardwaj, N., Maldarelli, F., Mellors, J. & Coffin, J. M. HIV-1 Infection Leads to
464 Increased Transcription of Human Endogenous Retrovirus HERV-K (HML-2) Proviruses *In*
465 *Vivo* but Not to Increased Virion Production. *J. Virol.* **88**, 11108–11120 (2014).
- 466 22. Contreras-Galindo, R., López, P., Vélez, R. & Yamamura, Y. HIV-1 infection increases
467 the expression of human endogenous retroviruses type K (HERV-K) in vitro. *AIDS Res. Hum.*
468 *Retroviruses* **23**, 116–122 (2007).
- 469 23. Vincendeau, M. *et al.* Modulation of human endogenous retrovirus (HERV) transcription
470 during persistent and de novo HIV-1 infection. *Retrovirology* **12**, 27 (2015).
- 471 24. Srinivasachar Badarinarayan, S. *et al.* HIV-1 infection activates endogenous retroviral
472 promoters regulating antiviral gene expression. *Nucleic Acids Res.* **48**, 10890–10908 (2020).
- 473 25. Jones, R. B. *et al.* HERV-K-specific T cells eliminate diverse HIV-1/2 and SIV primary
474 isolates. *J. Clin. Invest.* **122**, 4473–4489 (2012).
- 475 26. Contreras-Galindo, R., Almodóvar-Camacho, S., González-Ramírez, S., Lorenzo, E. &
476 Yamamura, Y. *Short Communication: Comparative Longitudinal Studies of HERV-K and*
477 *HIV-1 RNA Titers in HIV-1-Infected Patients Receiving Successful versus Unsuccessful*
478 *Highly Active Antiretroviral Therapy.* *AIDS Res. Hum. Retroviruses* **23**, 1083–1086 (2007).
- 479 27. Gonzalez-Hernandez, M. J. *et al.* Expression of Human Endogenous Retrovirus Type K
480 (HML-2) Is Activated by the Tat Protein of HIV-1. *J. Virol.* **86**, 7790–7805 (2012).

- 481 28. Contreras-Galindo, R. *et al.* HIV infection reveals widespread expansion of novel
482 centromeric human endogenous retroviruses. *Genome Res.* **23**, 1505–1513 (2013).
- 483 29. Jones, R. B. *et al.* LINE-1 retrotransposable element DNA accumulates in HIV-1-infected
484 cells. *J. Virol.* **87**, 13307–13320 (2013).
- 485 30. Jones, R. B. *et al.* HIV-1 infection induces retrotransposition of LINE-1 elements.
486 *Retrovirology* **6**, P43, 1742-4690-6-S2-P43 (2009).
- 487 31. O’Carroll, I. P. *et al.* Structural Mimicry Drives HIV-1 Rev-Mediated HERV-K
488 Expression. *J. Mol. Biol.* **432**, 166711 (2020).
- 489 32. van der Kuyl, A. C. HIV infection and HERV expression: a review. *Retrovirology* **9**, 6
490 (2012).
- 491 33. Kyriakou, E. & Magiorkinis, G. Interplay between endogenous and exogenous human
492 retroviruses. *Trends Microbiol.* **31**, 933–946 (2023).
- 493 34. Brinzevich, D. *et al.* HIV-1 Interacts with Human Endogenous Retrovirus K (HML-2)
494 Envelopes Derived from Human Primary Lymphocytes. *J. Virol.* **88**, 6213–6223 (2014).
- 495 35. An, D. S., Xie, Y. & Chen, I. S. Y. Envelope Gene of the Human Endogenous Retrovirus
496 HERV-W Encodes a Functional Retrovirus Envelope. *J. Virol.* **75**, 3488–3489 (2001).
- 497 36. Ogata, T., Okui, N., Sakuma, R., Kobayashi, N. & Kitamura, Y. Integrase of human
498 endogenous retrovirus K-10 supports the replication of replication-incompetent Int- human
499 immunodeficiency virus type 1 mutant. *Jpn. J. Infect. Dis.* **52**, 251–252 (1999).
- 500 37. Monde, K., Contreras-Galindo, R., Kaplan, M. H., Markovitz, D. M. & Ono, A. Human
501 Endogenous Retrovirus K Gag Coassembles with HIV-1 Gag and Reduces the Release
502 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).

- 526 48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
527 (2013).
- 528 49. Bendall, M. L. *et al.* Telescope: Characterization of the retrotranscriptome by accurate
529 estimation of transposable element expression. *PLoS Comput. Biol.* **15**, e1006453 (2019).
- 530 50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
531 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 532 51. Gao, C.-H., Yu, G. & Cai, P. ggVennDiagram: An Intuitive, Easy-to-Use, and Highly
533 Customizable R Package to Generate Venn Diagram. *Front. Genet.* **12**, 706907 (2021).
- 534 52. Ito, K. & Murphy, D. Application of ggplot2 to Pharmacometric Graphics. *CPT*
535 *Pharmacomet. Syst. Pharmacol.* **2**, e79 (2013).
- 536 53. Dillon, S. M. *et al.* An altered intestinal mucosal microbiome in HIV-1 infection is
537 associated with mucosal and systemic immune activation and endotoxemia. *Mucosal*
538 *Immunol.* **7**, 983–994 (2014).
- 539 54. Dillon, S. M. *et al.* Gut dendritic cell activation links an altered colonic microbiome to
540 mucosal and systemic T-cell activation in untreated HIV-1 infection. *Mucosal Immunol.* **9**, 24–
541 37 (2016).
- 542 55. O’Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status,
543 taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
- 544 56. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE
545 Project. *Genome Res.* **22**, 1760–1774 (2012).
- 546 57. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV):
547 high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192
548 (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

572 **Figure Legends**

573

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

575 **microenvironment**

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

588

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

591 Scatterplot demonstrating the viral load from serum on the X axis and normalized expression of
592 L1FLnI_1q23.1s in colon biopsies on the Y axis. uninfected controls (red) (n = 13) vs PLWH
593 (blue) (n = 19) (A). Normalized expression of L1FLnI_1q23.1s in colon biopsies from
594 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

618 biopsies on the Y axis. uninfected controls (red) (n = 13) vs PLWH (blue) (n = 19) (K).
619 Normalized expression of LTR19_12p13.31 in colon biopsies from uninfected controls (red) (n =
620 13) vs PLWH (blue) (n = 19). Post hoc significance values determined with unpaired T test (L).
621 Normalized expression of LTR19_12p13.31 in PBMCs from uninfected controls (red) (n = 13)
622 vs PLWH (blue) (n = 19). Post hoc significance values determined with unpaired T test (M).
623 Interferome array of L1FLnI_1q23.1s expression following IFN- α 1 (n = 3), - α 2 (n = 3), - α 5 (n =
624 3), - α 8 (n = 3), - α 14 (n = 3), and - β (n = 3) treated gut CD4+ T cells when compared to mock (m)
625 treatment (n=3). ANOVA performed for significance with post hoc significance values
626 determined via unpaired T test (N). IGV representation of LTR19_12p13.31 locus provided by
627 the Telescope annotation in the Refseq and Gencode annotations. StringTie reconstructed
628 transcriptomes demonstrate putative transcripts present in sequencing collected from the colon of
629 uninfected controls and PLWH individuals (O).

630

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

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

641 T Cells per gram on the X axis and normalized expression of LTR19_12p13.31 in colon biopsies
642 on the Y axis (E). Scatterplot demonstrating the IFN γ +CD8 $^+$ T Cells per gram on the X axis and
643 normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (F). Scatterplot
644 demonstrating the CD11c $^+$ mDCs per gram on the X axis and normalized expression of
645 LTR19_12p13.31 in colon biopsies on the Y axis (G). Scatterplot demonstrating the CD11c $^+$
646 mDCs per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies
647 on the Y axis (H). Scatterplot demonstrating the pDCs per gram on the X axis and normalized
648 expression of LTR19_12p13.31 in colon biopsies on the Y axis (I). Scatterplot demonstrating the
649 pDCs per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on
650 the Y axis (J). Scatterplot demonstrating the CD1c $^-$ mDCs per gram on the X axis and
651 normalized expression of LTR19_12p13.31 in colon biopsies on the Y axis (K). Scatterplot
652 demonstrating the CD1c $^-$ mDCs per gram on the X axis and normalized expression of
653 L1FLnI_1q23.1s in colon biopsies on the Y axis (L).

654

655

656

657

658

659

660

661

662

663

664 **Supplemental Legends**

665

666 **Figure S1: Characterization of ERE expression profiles in the colonic microenvironment of**

667 **PLWH.** Principal component analysis (PCA) plot based solely on endogenous retroelement

668 (ERE) abundances demonstrate variance in sample distribution between uninfected controls (red)

669 (n = 13) and PLWH (blue) (n = 19) gut biopsies (A). Stacked bar charts demonstrating mean

670 abundance of exonic, intronic, and intergenic EREs as the average proportion of ERE denoted

671 reads from gut biopsies collected from uninfected controls (n = 13) and PLWH (n = 19)

672 individuals (B). Stacked bar charts demonstrating mean abundance of LINE1 and HERV

673 elements as the average proportion of ERE denoted reads from gut biopsies collected from

674 uninfected controls (n = 13) and PLWH (n = 19) individuals (C). Stacked bar charts

675 demonstrating the proportional abundance of chromosomes of origin for ERE loci from gut

676 biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (D). Stacked

677 bar charts demonstrating category of EREs the average proportion of TE denoted reads from gut

678 biopsies collected from uninfected controls (n = 13) and PLWH (n = 19) individuals (E). Stacked

679 bar charts demonstrating mean abundance of transposable element (TE) families as the average

680 proportion of ERE denoted reads from gut biopsies collected from uninfected controls (n = 13)

681 and PLWH (n = 19) individuals (F).

682

683 **Figure S2: Characterization of ERE expression profiles in uninfected gut CD4+ T cells**

684 **activated with IFN-Is.** PCA plot based solely on ERE abundances demonstrate variance in

685 sample distribution between mock (m) (n=3), IFN α 1 (a1) (n = 3), IFN α 2 (a2) (n = 3), IFN α 5 (a5)

686 (n = 3), IFN α 8 (a8) (n = 3), IFN α 14 (a14) (n = 3), and IFN β (b) (n =3) treated gut CD4+ T cells

687 (A). Stacked bar charts demonstrating mean abundance of exonic, intronic, and intergenic TEs as
688 the average proportion of ERE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8
689 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (B). Stacked bar charts demonstrating
690 mean abundance of LINE1 and HERV elements as the average proportion of ERE denoted reads
691 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
692 CD4+ T cells (C). Stacked bar charts demonstrating the chromosomal location of ERE denoted
693 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
694 gut CD4+ T cells (D). Stacked bar charts demonstrating ERE category as the average proportion
695 of TE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and
696 b (n =3) treated gut CD4+ T cells (E). Stacked bar charts demonstrating mean abundance of TE
697 families as the average proportion of TE denoted reads from m (n=3), a1 (n = 3), a2 (n = 3), a5 (n
698 = 3), a8 (n = 3), a14 (n = 3), and b (n =3) treated gut CD4+ T cells (F).

699

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

710 expressed with an adjusted p value of ≤ 0.05 and a $LFC \geq 1$ or ≤ -1 between m (red) (n=3) and a2
711 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ using the Wald's
712 Test. Adjusted p values were calculated using default parameters for a Benjamini-Hochberg
713 correction (C). Heatmap demonstrating the normalized per sample abundances of EREs
714 differentially expressed with an adjusted p value of ≤ 0.05 and a $LFC \geq 1$ or ≤ -1 between m (red)
715 (n=3) and a5 (blue) (n=3) treated gut CD4+ T cells. All statistics were performed in DESEQ
716 using the Wald's Test. Adjusted p values were calculated using default parameters for a
717 Benjamini-Hochberg correction (D). Heatmap demonstrating the normalized per sample
718 abundances of EREs differentially expressed with an adjusted p value of ≤ 0.05 and a $LFC \geq 1$
719 or ≤ -1 between m (red) (n=3) and a8 (blue) (n=3) treated gut CD4+ T cells. All statistics were
720 performed in DESEQ using the Wald's Test. Adjusted p values were calculated using default
721 parameters for a Benjamini-Hochberg correction (E). Heatmap demonstrating the normalized per
722 sample abundances of EREs differentially expressed with an adjusted p value of ≤ 0.05 and a
723 $LFC \geq 1$ or ≤ -1 between m (red) (n=3) and a14 (blue) (n=3) treated gut CD4+ T cells. All
724 statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated
725 using default parameters for a Benjamini-Hochberg correction (F). Heatmap demonstrating the
726 normalized per sample abundances of EREs differentially expressed with an adjusted p value of
727 ≤ 0.05 and a $LFC \geq 1$ or ≤ -1 between m (red) (n=3) and b (blue) (n=3) treated gut CD4+ T cells.
728 All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated
729 using default parameters for a Benjamini-Hochberg correction (G). Venn diagram of shared vs
730 uniquely downregulated ERE loci across a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3),
731 and b (n = 3) treated gut CD4+ T cells when compared to m (H). Venn diagram of shared vs

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

734

735 **Figure S4: Upregulation of ERE loci in gut CD4+ T cells activated with IFN-Is.** Volcano plot

736 demonstrating the average fold change ≥ 1 or ≤ -1 and an adjusted p value of ≤ 0.05 and a LFC

737 between m (n = 3) and a1 (n = 3) treated gut CD4+ T cells. All statistics were performed in

738 DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a

739 Benjamini-Hochberg correction (A). Volcano plot demonstrating the average fold change ≥ 1 or $\leq -$

740 1 and an adjusted p value of ≤ 0.05 and a LFC between m (n = 3) and a2 (n = 3) treated gut CD4+

741 T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were

742 calculated using default parameters for a Benjamini-Hochberg correction (B). Volcano plot

743 demonstrating the average fold change ≥ 1 or ≤ -1 and an adjusted p value of ≤ 0.05 and a LFC

744 between m (n = 3) and a5 (n = 3) treated gut CD4+ T cells. All statistics were performed in

745 DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a

746 Benjamini-Hochberg correction (C). Volcano plot demonstrating the average fold change ≥ 1 or $\leq -$

747 1 and an adjusted p value of ≤ 0.05 and a LFC between m (n = 3) and a8 (n = 3) treated gut CD4+

748 T cells. All statistics were performed in DESEQ using the Wald's Test. Adjusted p values were

749 calculated using default parameters for a Benjamini-Hochberg correction (D). Volcano plot

750 demonstrating the average fold change ≥ 1 or ≤ -1 and an adjusted p value of ≤ 0.05 and a LFC

751 between m (n = 3) and a14 (n = 3) treated gut CD4+ T cells. All statistics were performed in

752 DESEQ using the Wald's Test. Adjusted p values were calculated using default parameters for a

753 Benjamini-Hochberg correction (E). Volcano plot demonstrating the average fold change ≥ 1 or $\leq -$

754 1 and an adjusted p value of ≤ 0.05 and a LFC between m (n = 3) and b (n = 3) treated gut CD4+

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

757

758 **Figure S5: HIV-1 status and IFN-Is induce 7 shared EREs in the gut microenvironment.**

759 Venn diagram of shared vs uniquely downregulated TE loci across PLWH (n = 19) vs uninfected
760 controls (n = 13) gut biopsies, and a1 (n = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b
761 (n = 3) treated gut CD4+ T cells when compared to m (A). Venn diagram of shared vs uniquely
762 upregulated TE loci across PLWH (n = 19) vs uninfected controls (n = 13) gut biopsies, and a1 (n
763 = 3), a2 (n = 3), a5 (n = 3), a8 (n = 3), a14 (n = 3), and b (n = 3) treated gut CD4+ T cells when
764 compared to m (B). Gut biopsies from uninfected controls (n = 13) and PLWH (n = 19)
765 individuals (left) as well interferome assay responses to type 1 interferons (n = 3) (right) show
766 significant upregulation in HARLEQUIN_17q12 (C), HERV4_4q22.1 (D), MER101_2p25.2 (E),
767 MER4_7q21.2 (F), L1FLnI_1p22.21 (G), L1FLnI_9p21.1ha (H), and L1FLnI_17q12k (I). All
768 statistics were performed in DESEQ using the Wald's Test. Adjusted p values were calculated
769 using default parameters for a Benjamini-Hochberg correction. Tukey box and whisker plots for
770 with the line demonstrating the median of expression while the large dot represents the mean.

771

772 **Figure S6: EREs of interest in the gut microenvironment show no association with CD1c+**

773 **mDC abundances.** Gut expression of L1FLnI_1q23.1s (A), L1FLnI_5q32o (B), and
774 LTR19_12p13.31 (C) show no association with CD1c+ mDC abundances in the gut
775 microenvironment.

776

777 **Figure S7: L1FLnI_5q32o is not associated with major cellular parameters of immune**
778 **deregulation in PLWH.** Scatterplot demonstrating the CD4+ T Cells per gram on the X axis and
779 normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (A). Scatterplot
780 demonstrating the CD8+ T Cells per gram on the X axis and normalized expression of
781 L1FLnI_5q32o in colon biopsies on the Y axis (B). Scatterplot demonstrating the IFN γ +CD8+ T
782 Cells per gram on the X axis and normalized expression of L1FLnI_5q32o in colon biopsies on
783 the Y axis (C). Scatterplot demonstrating the CD11c+ mDCs per gram on the X axis and
784 normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (D). Scatterplot
785 demonstrating the pDCs per gram on the X axis and normalized expression of L1FLnI_5q32o in
786 colon biopsies on the Y axis (E). Scatterplot demonstrating the CD1c- mDCs per gram on the X
787 axis and normalized expression of L1FLnI_5q32o in colon biopsies on the Y axis (F).

788
789 **Figure S8: Association between colonic L1FLnI_1q23.1s expression and CD4+ T Cell**
790 **subsets.** Scatterplot demonstrating the IFN γ -IL17+CD4+ T cells per gram on the X axis and
791 normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y axis (A). Scatterplot
792 demonstrating the IFN γ +IL17-CD4+ T cells per gram on the X axis and normalized expression
793 of L1FLnI_1q23.1s in colon biopsies on the Y axis (B). Scatterplot demonstrating the
794 IFN γ +IL17+CD4+ T cells per gram on the X axis and normalized expression of L1FLnI_1q23.1s
795 in colon biopsies on the Y axis (C). Scatterplot demonstrating the IL22+IFN γ -IL17-CD4+ T cells
796 per gram on the X axis and normalized expression of L1FLnI_1q23.1s in colon biopsies on the Y
797 axis (D).

798

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

<i>Median (min, max)</i>	N/A	26000 (2880, 196000)
--------------------------	-----	----------------------

Years since first HIV-1 seropositive test

<i>Median (min, max)</i>	N/A	3.5 (0.17, 15.0)
--------------------------	-----	------------------

847 Study participants were selected from a completed clinical study⁵³. Values are shown as median
848 (range) for continuous measures and numbers for categorical measures. Statistical analyses were
849 performed using Mann-Whitney test for continuous characteristics and Fishers exact test to
850 compare categorical characteristics between control and PWH groups. *not reported for one
851 person in PWH cohort. N/A: not applicable.

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868 **Research in Context**

869

870 *Evidence before this study*

871

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

880

881 *Added value of this study*

882

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

890

891 *Implications of all the available evidence*

892

893 Recurrent observations suggest pathophysiological roles for differentially expressed EREs in

894 diseased tissues. In HIV-1 infection, EREs can be over expressed due to trans-activation of

895 HERV LTRs and shifts in the regulome that permit the expression of typically silent elements.

896 However, it has not been established how EREs may be deregulated in the gut where

897 immunopathologies persist surrounding viral reservoirs. Correlations between ERE expression

898 and cellular composition demonstrate that certain EREs may be indicative of cell type

899 abundances thought to drive GI-specific immunopathologies resultant from HIV-1 infection.

