RNA-seq transcriptome analysis reveals LTR-retrotransposons modulation in Human Peripheral Blood Mononuclear Cells (PBMCs) after *in vivo* Lipopolysaccharides (LPS) injection

Maria Paola Pisano¹, Olivier Tabone², Maxime Bodinier², Nicole Grandi¹, Julien Textoris², François Mallet² and Enzo Tramontano^{1,3} #

¹Laboratory of Molecular Virology, Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy.

²EA 7426 Pathophysiology of Injury-induced Immunosuppression, University of Lyon1-Hospices Civils de Lyon-bioMérieux, Hôpital Edouard Herriot, Lyon, France.

³Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Cagliari, Italy.

#Address correspondence to Enzo Tramontano, tramon@unica.it.

Word count

Abstract: 261

Body text: 3771

Abstract: Human Endogenous Retroviruses (HERVs) and Mammalian apparent LTR-retrotransposons (MaLRs) are retroviral sequences that integrated into the germline cells millions year ago. Transcripts of these LTR-retrotransposons are present in several tissues, and their expression is modulated in pathological conditions, although their function remains often far from being understood. In this work, we focused on the HERVs/MaLRs expression and modulation in a scenario of immune system activation. We used a public dataset of Human Peripheral Blood Mononuclear Cells (PBMCs) RNA-seq from 15 healthy participants to a clinical trial before and after the exposure to Lipopolysaccharide (LPS), for which we established an RNA-seq workflow for the identification of expressed and modulated cellular genes and LTR-retrotransposon elements.

Importance: We described the HERV and MaLR transcriptome in PBMCs, finding that about 8.4 % of the LTR-retrotransposons loci were expressed, and identifying the betaretrovirus-like HERVs as those with the highest percentage of expressed loci. We found 4,607 HERVs and MaLRs loci that were modulated as a result of *in vivo* stimulation with LPS. The HERV-H group showed the highest number of differentially expressed most intact proviruses. We characterized the HERV and MaLR loci differentially expressed checking their genomic context of insertion and, interestingly, we found a general co-localization with genes that are involved and modulated in the immune response, as consequence of LPS stimulation. The analyses of HERVs and MaLRs expression and modulation show that this LTR-retrotransposons are expressed in PBMCs and regulated in inflammatory settings. The similar regulation of HERVs/MaLRs and genes after LPS stimulation suggests possible interactions of LTR-retrotransposons and the immune host response.

Keywords: HERV, MaLR, PBMCs, LPS, immunity, RNA-seq, human

Background

A large proportion of the human genome consists of repeated elements, including Human Endogenous Retroviruses (HERVs) and Mammalian Apparent LTRs Retrotransposons (MaLRs) (1). These elements are remnants of ancestral and independent infections within the germline cells that took place million years ago (2). At the time of integration, the HERV genome was initially composed of four retroviral genes (*gag*, *pro*, *pol*, and *env*) flanked by two Long Terminal Repeats (LTRs). Similarly, the genomic structure of MaLRs, more ancient retroviral elements, is similar but characterized by the absence of the *env* gene (3, 4). During the time, these elements have mostly accumulated abundant mutations often compromising their coding capability, and a great number of solitary LTRs were obtained by recombination occurrences (2, 4), as clearly observably when comparing human LTR-retrotransposons integrations with their orthologs in primates (5, 6). The potential involvement of these elements in the human biology has been investigated in several studies (7) which have shown the presence of few preserved retroviral Open Reading Frames (ORFs) expressed in human tissues (8–10), including Syncytin-1, a functional HERV-W envelope protein that is involved into the trophoblastic-cells fusion in placenta development (11, 12). Another important function of HERVs and MaLRs insertions is linked to species evolutionary developments. For example, data suggest that the high density of HERV integrations had a role in promoting and maintaining the diversity of multigenic regions, as in the case of human Major Histocompatibility Complex (MCH) (13–15). Furthermore, the promoters, enhancers and polyadenylation signals within the LTR sequence allow the proviral and solitary LTRs to play a role into the expression of the host genes (16). More specifically, the integration of newly provided regulatory signals at one of the ends of the host gene sequence could result in the modulation of its expression (16), while the presence of viral promoters or polyadenylation signals near to the gene or within its intron could result in altered RNA splicing (16–19). In

71 this respect, the HERVs and MaLRs contribution in shaping and influencing the human innate
72 immunity is an argument of particular interest (20). Indeed, in some cases LTR-
73 retrotransposon derived antigens could be recognized as pathogen-associated molecular
74 patterns (PAMPs) by Pattern Recognition Receptors (PRRs) such as the transmembrane
75 proteins Toll-Like Receptors (TLRs) (21, 22). In these cases, the activation of PRRs evokes
76 complex cellular signaling pathways altering gene expression for transducing pro-
77 inflammatory signals (20). Despite the fact that on the one hand these interactions with PRRs
78 are hypothesized to be positively involved in shaping the evolution of the immune response
79 (21, 23), on the other hand, these mechanisms have been investigated for their possible
80 contribution to the development of autoimmunity and inflammatory diseases (9, 24–26), like
81 multiple sclerosis (27, 28). Interestingly, the activation of immune response through
82 treatments with LPS or TNF- α can lead to an increase of HERVs and MaLRs expression (29,
83 30). For instance, a recent microarray-based study revealed the *in vitro* modulation of
84 HERV/MaLR expression in PBMCs, after high- and low-dose LPS and Interferon- γ (IFN- γ)
85 stimulations (29). A similar approach allowed to observe HERV/MaLR *in vivo* modulation in
86 samples of blood in various contexts of injuries, also introducing a possible role of
87 HERVs/MaLRs close to immunity-related genes in the regulation of their expression (31). In
88 any case, many intriguing questions about the effective role of HERV and MaLR expression in
89 immunity are still unsolved, and, in this respect, identifying the individual LTR-
90 retrotransposon genomic localization and coding capacity could help to understand of their
91 potential effects in this context (32). In this context, with the aim of investigating possible
92 interactions between LTR-retrotransposons and the immune host response, we performed a
93 comprehensive description of HERV and MaLR transcriptome observed in the analysis of a
94 public RNA-seq dataset of PBMCs samples from patients *in vivo* stimulated with LPS. This
95 experimental design has allowed mimicking an inflammation gram-negative induced. We

focused our attention on the identification of specific expressed loci, as well as of those that were modulated as a result of stimulation with LPS. At last, we checked for similar regulation of HERVs/MaLRs and genes expression.

Results

Description of the LTR-retrotransposon transcriptome in PBMCs

With the aim to detect the HERV/MaLR transcriptome in PBMCs, we defined an RNA-seq pipeline to be used on a public RNA-seq (GEO:GSE87290) that included the PBMC transcriptome of 15 healthy volunteers before and after *in vivo* stimulation with 1ng/Kg LPS. Using this pipeline we could discriminate 424,515 loci from the *hervgdb4* database, including those of 197,341 HERVs and those of 227,174 MaLRs (33). Importantly, as the *hervgdb4* database has been created as part of the design of Affymetrix HERV-V3 array probes (33), the HERV and MaLR loci are included in the database as 881,603 *hervgdb4* fragments (Fig. S1) (33). Moreover, according to the annotation accuracy of the fragments and to their source, the loci were part of two different subsets, i) the highly informative HERV_prototypes and ii) the roughly annotated HERV_Dfam and MaLR_Dfam both collected from the Dfam database (Fig. S1) (34). Our analysis showed that 18,633 HERV *hervgdb4* loci and 17,053 MaLR *hervgdb4* loci were expressed in either stimulated and un-stimulated PBMCs samples, for a total of 35,686 loci, accounting for ~ 8.4% of the LTR-retrotransposons in the human genome (Fig. S1). The majority of the LTR-retrotransposons expressed were part of the roughly-annotated HERV_Dfam and MaLR_Dfam subset, for which no information about the groups was available. Among the well-annotated HERV_prototypes we found expressed, 2084 were members of the class I gamma-like, 527 of the class II beta-like and 310 of the class III spuma/epsilon-like (Fig. 1a). When considering only the absolute number of expressed loci, just ahead from the HERV-H group, the PRIMA41 group was the one most represented. Instead, when considering

121 the proportion of expressed loci among the HERV_prototypes groups, the most recently
 122 integrated group, HML-2, was the one showing the highest transcriptional activity, with more
 123 than 30% of its loci expressed. HML-10, HML-8 and HML-9 groups were also very active, with
 124 a proportion of 25.8%, 26.4% and 25% expressed loci, respectively. In general, the II beta-like
 125 groups were those showing overall the greatest percentage of activation. The proportions of
 126 expressed proviruses compared to solitary LTRs within the 3 classes were 55%, 61% and
 127 64%, respectively (data not shown). Furthermore, in order to better define the HERV
 128 transcriptome in PBMCs, we decided to analyze the most intact proviruses that have been
 129 identified, classified and characterized in Vargiu *et al.* (4). These sequences have been
 130 collected by Vargiu *et al.* after the analysis of the software RetroTector (ReTe) on the human
 131 genome assembly GRCh 37/hg19, which searched for the presence of conserved motif hits
 132 and from these allowed to reconstruct the coordinates of 3173 HERVs satisfying distance
 133 constraints (4, 35). We identified 723 expressed ReTe proviruses and also in this case we
 134 found a large proportion of expressed beta-like elements. HML-4 (6 out of 12 ReTe
 135 proviruses) and HML-2 (43 out of 92 ReTe proviruses) were the most active groups (Fig. 1b).
 136 The group with the highest number of expressed ReTe proviruses was the class I gamma-like
 137 HERV-H, with 241 loci, representing the 23% of the whole group.

138 **Distinct transcriptional patterns induced by LPS stimulation**

139 Next we wanted to assess the variability across the samples and to detect specific signatures
 140 induced by LPS stimulation. Hence, we analyzed the expression data of *hervgdb4* fragments in
 141 the 30 PBMC samples using the unsupervised Principal Component Analysis (PCA) (Fig. 2).
 142 The first Principal Component (PC1) explained the 49% of the variance across samples and
 143 revealed clustering specifically related to LPS response, dividing almost the totality of LPS-
 144 stimulated from not LPS-stimulated samples. Indeed, these data suggested that LPS response
 145 is the principal determinant for defining the inter-sample variability, showing *hervgdb4*

146 fragments activation pattern specific of each of the two conditions. Moreover, we further
 147 investigated the transcriptional signature of the samples by performing hierarchical
 148 clustering on the 1,000 *hervgdb4* fragments with the highest mean of reads counts across
 149 samples (Fig. 3a). Results showed clear differences in expression of the *hervgdb4* fragments
 150 across samples, related to LPS stimulation. However, both PCA and hierarchical clustering
 151 analyses showed that the LTR-retrotransposon expression profile of samples 7, 9, 14 and 15
 152 after LPS-stimulation was similar to the profiles of the LPS-unstimulated samples. This was
 153 further confirmed by the hierarchical clustering performed on the 1000 human genes with
 154 the highest mean among samples, and, also in this case, the four samples showed the typical
 155 profiles of the LPS-unstimulated ones (data not shown). As we observed differences among
 156 samples depending on inter-individual variability to immune response, we tried to obtain
 157 more information on the interpersonal reaction to LPS by describing the pattern of expression
 158 of a subset of 44 genes that have been previously reported to be a specific signature of
 159 induced cytokine response with a heatmap of the sample-to-sample Euclidean distances (36).
 160 While in the absence of stimulation no evident differences between samples were observed,
 161 after LPS injection, considering these 44 genes two clusters were observed that reflected the
 162 traits of the inflammatory response, with high-responders clearly divided from low-
 163 responders (Fig. 3b). Of note, since these 44 genes have been identified to be able of
 164 deconvoluting complex responses to immune stimulation, we expected the variability
 165 between low- and high-responders to be well-defined. Subsequently, we similarly analyzed
 166 the Euclidean sample-to-sample distances defined by the expression of all the *hervgdb4*
 167 fragments (Fig. 3b). Interestingly, as shown for the 44 immunity genes, also for *hervgdb4*
 168 fragments no sample clustering was observed in LPS-unstimulated sample, while in LPS-
 169 stimulated samples two clusters were formed, roughly corresponding to low- and high-
 170 responders to immune stimulation. However, the three low-responder samples clustering

171 with the high-responder ones do not coincide with the four LPS-stimulated samples that
172 showed a pattern of *hervgdb4* fragments expression similar to the LPS-unstimulated samples
173 in Fig. 3A, suggesting that factors other than the severity of inflammatory response may
174 contribute rather than interfere with the inter-individual variability for *hervgdb4* fragments
175 expression.

176 **Differential LTR-retrotransposon expression in PBMCs**

177 Once assessed that the variability across samples mostly fitted with LPS stimulation, we
178 evaluated the *hervgdb4* fragments for differential expression. After applying a filter ($FDR \leq$
179 0.01 and absolute values of $\log_2$ Fold Change ≥ 1) we identified a total of 6,452 (11% of the
180 total expressed) Differentially Expressed (DE) *hervgdb4* fragments (Table S1). We
181 represented all the expressed *hervgdb4* fragments in a volcano-plot, where they were
182 indicated as points that spread according to the $\log_2$ Fold Change on the x-axes and to the
183 adjusted p-value on the y-axes (Fig. 4). It is worth noting that the great majority of *hervgdb4*
184 fragments were positively modulated, showing a general trend of LTR-retrotransposon up-
185 regulation in PBMCs after LPS stimulation. In fact, among the 6,452 DE *hervgdb4* fragments,
186 5,383(83%) were up-modulated while 1,069 (17%) were down-modulated after stimulation
187 (Table 1). Subsequently, we checked how many *hervgdb4* DE loci were present, observing that
188 4,607 loci (13% of the total expressed) were modulated. Interestingly, results showed that
189 3,688 loci (80%) were up-regulated, while 919 loci (20%) were down-regulated. We then
190 focused on the mostly intact ReTe HERV proviruses, observing that 115 of them were DE
191 (17% of the ReTe HERV expressed proviruses). In particular, 86 were up-regulated while 29
192 were down-regulated. In general, the majority of the DE HERVs was over-expressed: out of
193 the 55 HERV groups, 6 included only up-modulated proviruses, 13 included both up- and
194 down-regulated proviruses and 5 only down-regulated proviruses (Fig. 5). Importantly, we

found 23 groups that were constitutively expressed in PBMCs but were not modulated by the stimulation.

Concordant modulation of LTR-retrotransposons and co-localized immunity-related genes

To gain more insights into the LTR-retrotransposon modulation, we focused on the 15 *hervgdb4* fragments with the highest DE according to their adjusted p-value (Table 2, and Fig. 4). Importantly, all these 15 high DE *hervgdb4* fragments were up-regulated after LPS-stimulation. We used Transcript Per Million (TPM) normalization to have information on the expression levels of the fragments before and after the LPS stimulation (Table 2) and, subsequently, we investigated their context of insertion. Interestingly, we found that, among the 15 most-highly modulated LTR-retrotransposons, 10 were neighbor integrations (within a 10-kb window of distance) with human coding genes. In particular, 6 of them were inside, 3 were downstream and 1 was upstream the colocalized gene. We hence analyzed the modulation of these human genes, observing that all of them were up-regulated as a consequence of LPS stimulation, as summarized in Fig. 6. The fragment with the lowest adjusted p-value ($6.99\text{E-}36$) was 081524901ERVH5LU3I, at coordinates chr8:103002077-103002365. This fragment is the U3 region of a 5'LTR belonging to an HERVH provirus (chr8:103002064-103004587), and its expression levels were increased from an average TPM value of 2.4 to an average TPM value of 124.5 after LPS stimulation. It is worth to note that, even if this LTR sequence is not co-localized with coding genes, it is integrated in a promoter-flanking region that is affected by copy number variation according to ENSEMBL annotations (data not shown), possibly suggesting a potential transcriptional control role. Interestingly, also fragment 081524801-HERV0183cL, part of a solo LTR (chr8:103001306-103001748) within the same region as 081524901ERVH5LU3I, increased its average TPM value from 2.4 to 74.2. Fragments 021629602-MALR1014uL and 021629601-MALR1014uL

220 were part of the same solo LTR at coordinates chr2:113131173-113131620. This solo LTR is
 221 integrated within the intron of the Interleukin 1 Receptor Antagonist gene (IL1RA), which
 222 codes for a protein that is known to have an anti-inflammatory role (37). Both gene and solo
 223 LTR significantly increased their expression levels after LPS administration, showing high
 224 average TPM values in stimulated samples (Table 2). Similarly, fragments 022246101-
 225 HERV0758cL and 022246102-HERV0758cL, part of the same solo LTR at coordinates
 226 chr2:113131173-113131620, showed a pattern of up-regulation comparable with their
 227 neighbor gene, namely TNF alpha induced protein 6 (TNFAIP6). Thus, also in this case, the
 228 solo LTR is co-localized with a gene that is involved in immunity, having a known regulatory
 229 function (38). Instead, fragments 081524501PRM4pol and 081524402-HERV0948cL, which
 230 are portions of the internal regions in proviral loci chr8:102984015-102986529 and
 231 chr8:102980120-102991770, respectively, were found to be intergenic integration. The basal
 232 expression levels of both fragments were 0.1 TPM, increasing to 3.6 and 5.5 TPM after
 233 stimulation, respectively. Next, we found that fragment 062308001-HERV0741cL, a solo LTR
 234 in chr6:159677828-159678891, is integrated within the intron of Superoxide dismutase 2
 235 (SOD2) gene. Of note, fragments 150538602-MALR1003uL and 150538601-MALR1003uL,
 236 both part of a solo LTR in chr15:63906995-63907370, showed an average TPM increased
 237 from 3.7 and 1.0 to 49.7 and 14.9 after stimulation, respectively, are integrated within the 3'
 238 untranslated region (UTR) of the Death-associated Protein Kinase 2 (DAPK2) gene. Fragment
 239 120084901-HERV0406uL, a solo LTR in locus chr12:8537686-8538696, is integrated inside
 240 the C-type lectin domain family 4 member E (CLEC4E) gene, and increased its average TPM
 241 values from 1.0 to 87.2. The fragments 120085903MR52env (chr12:8566648-8568185),
 242 representing an intergenic integration, and 040314601-HERV0741uL (chr4:15677257-
 243 15678021), being inserted upstream the Family With Sequence Similarity 200 Member B
 244 gene (FAM200B), showed more than 5-folds increase in average TPM after LPS stimulation.

Finally, the fragment at coordinates chr19:35887015-35887532, which increased its average TPM values from 5.7 to 28.9, is integrated downstream the NFKB inhibitor delta gene.

Discussion

We used an RNA-seq approach and the *hervgdb4* database (33) to assess an overview of the LTR-retrotransposons transcriptome in PBMCs. This analysis has allowed, for the first time, to investigate HERVs and MaLRs proviruses and solo LTRs expression in human patients by mimicking gram-negative bacterial infection, revealing a total of 35686 LTR-retrotransposon elements expressed in PBMCs.

Previously, the same database has been used by Mommert *et al.* (29) to identify expressed and modulated loci in an *ex vivo* system of LPS stimulation and endotoxin tolerance, through microarray analyses (29). Interestingly, the percentage of expressed elements in our *in vivo* model (6.8% of *hervgdb4* fragments and 8.4% of *hervgdb4* loci expressed) was slightly higher than the one measured in *in vitro* experiments (5.6% of *hervgdb4* fragments). Among all groups, class II groups are those most active, while groups belonging to class III were, in general, less active. In particular, HML-2 is the most expressed in PBMCs. This group is also one of the most investigated, due to recent HML-2 integrations and the hypothesized implication of some of active loci in several diseases (39–42). In comparison to Mommert *et al.* (29), who reported instead an abundant activation of several class I and of all class III groups, the LTR-retrotransposon expression among classes showed some differences. These differences can be explained by i) the technologies and methods that have been used (*in vivo* vs *ex vivo*, microarray vs RNA-seq), and ii) the great differences in the basal transcriptional activity of each individual, which have been already observed in PBMCs (43). Moreover, the use of the Vargiu *et al.* database (4) has allowed us to analyze the expression of the most intact HERV proviruses for which we can hypothesize a higher probability of protein

270 production. The pattern of counts distribution among classes of the most intact HERV
 271 proviruses is similar to the one obtained when considering the *hervgdb4* loci, and it is mostly
 272 in agreement with the information on HERV-H, HML-2, HERV-E and HERV-W transcriptional
 273 activity in PBMCs already reported in previous studies (37). Importantly, this study has given
 274 information on the overall expression of the groups, but not on the single loci. Moreover, the
 275 expression of HERV-E has been reported to be characteristic of only a small percentage of the
 276 subjects analyzed, while a large portion of both *hervgdb4* loci (the 24% of the group) and
 277 most intact HERVs (the 28% of the group) is expressed in PBMCs, in the present study.

278 The analysis of the variability among the 30 samples analyzed showed differences between
 279 non-stimulated samples and LPS-stimulated samples. Indeed, in both PCA analysis and
 280 hierarchical clustering the not-stimulated and stimulated samples spread in two distinct
 281 clusters. However, a great interpersonal variability on the response of patients to LPS is also
 282 evident. This heterogeneity in the HERVs and MaLRs expression, as a consequence of LPS
 283 stimulus, is in line with the already observed strong inter-individual variability of gene
 284 expression in response to microbial agents (36, 44). Importantly, the definition of such
 285 variability should be a necessary foundation for better understanding a possible implication
 286 of HERVs and MaLRs expression in immunity and disease.

287 The analysis of *in vivo* HERVs and MaLRs modulation reveals a total of 6,452 *hervgdb4*
 288 fragments modulated in PBMCs, with a clear trend of up-regulation induced by LPS response
 289 (81%). In comparison to Mommert *et al.* (29), in which 111 DE HERVs and MaLRs (44 down-
 290 regulated and 67 up-regulated) have been identified, a larger number of DE elements was
 291 observed, which can be explain by the different sensitivity of the methods used. Specifically,
 292 only 18 of the DE *hervgdb4* fragments identified by Mommert *et al.* are included in the present
 293 list of DE *hervgdb4* fragments, suggesting that the two models are not equal but
 294 complementary. Moreover, another study has identified a list of LTRs-retrotransposon INF-

induced (45) and a proportion of them (about the 16%) was observed in this study to be also LPS-induced (data not shown). This highlights the importance of investigating the LTRs-retrotransposon modulation in several conditions, but it also suggests that distinctive patterns of LTRs-retrotransposon induction, if used as biomarker, may become indicative to separate physiological from pathological conditions. The LPS stimulus also provoked a strong activation of the most intact ReTe proviruses. Importantly, not all the groups respond in the same way to the LPS stimulation.

Among the 15 most modulated *hervgdb4* fragments, 10 were co-localized with human genes. We used the average expression level of the most DE human gene, IL1A (ENSG00000115008), which showed an average TMP of 19.4 after LPS stimulation (data not shown), as a comparison to understand a possible biological meaning of the expression of the fragments we found modulated. In general, all the fragments are not only modulated but also generally medium/highly expressed. Among 12 *hervgdb4* loci, 10 are LTRs, including 9 solo LTRs. This kind of information might be integrated by other studies on transcription factor binding sites (TFBS) provided by HERVs and MaLRs (14), including those for NF- κ B, IRFs and STATs, suggesting a possible HERVs and MaLRs activation in immunity (46, 47). Moreover, 10 *hervgdb4* fragments are neighbor integrations of human genes that are also activated after inflammatory response. Of particular interest is the identification of 3 solo LTRs localized outside the modulated genes, due to the possible presence of promoters or polyadenylation signals that may play a role in the regulation of the nearby gene. Specifically, the solo LTRs in chr2:151381023-151381962 and chr19:35887015-35887532 are integrated downstream the TNF alpha induced protein 6 gene (TNFAIP6) and NF-kappa-B inhibitor delta (NFKBID), respectively. Interestingly, while the most DE genes identified in this model are mainly those coding for proteins that are positive regulators of immunity, such as IL1A and IL1B (data not shown), these data showed a strong up-regulation of LTRs-retrotransposon co-localized with

genes coding for cellular inhibitors of these proteins. Hence, if on the one hand it has been suggested that a subset of HERVs hold TFBSs (47) that may increase their activation in immunity, on the other hand the genes products of TNFAIP6 and NFKBID are potentially able to inhibit this phenomena(48, 49). For this reason, this data underlines the complexity of the relationship between the HERVs/MaLRs modulation and the immune response, especially if hypothesizing their active role on the regulation of the co-localized immune-related genes. Certainly, this data may give the bases to clarify the LTRs-retrotransposon involvement in the regulation of immune functions, but further studies are needed to clarify these mechanisms and may help to understand how HERVs and MaLRs contribute to the complexity of immune response.

330

331 **Conclusions**

We used RNA-seq transcriptome data from 15 healthy participants to a clinical trial injected with LPS, with the aim to identify expressed and modulated LTR-retrotransposon elements. This RNA-seq based approach revealed the basal expression of 18,633 HERV *hervgdb4* loci, including those of 723 most intact ReTe proviruses, and 17,053 MaLR *hervgdb4* loci, in PBMCs. To assess the specific signatures induced by LPS we investigated the sample-to-sample differences, which showed a specific pattern of *hervgdb4* fragments expression as a consequence of inflammation. We hence evaluated the *hervgdb4* fragments modulation, identifying a total of 6,452 differentially expressed elements. Moreover, we observed a general trend of activation of the LTR-retrotransposons after LPS stimulation. Interestingly, the HERVs/MaLRs regulation was similar to the cellular gens regulation, suggesting possible interactions between LTR-retrotransposons and the immune response. However, further analyses are required to evaluate whether such modulation is involved in the regulation of

immune response and to better assess the LTR-retrotransposons expression significance in immunity.

Methods

RNA-seq data and quality control

We used a RNA-seq dataset public available (GEO:GSE87290), including the transcriptome of PBMCs from healthy humans (n=15) before and after 1ng/kg LPS exposure. We checked the quality of the RNA sequences by using FastQC Galaxy Version 0.72 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Low quality reads were trimmed with Trim Galore! V.0.4.3.1 (<https://github.com/FelixKrueger/TrimGalore>).

All the mentioned analyses were done on Galaxy release_17.09 locally installed (<http://galaxyproject.org/>).

Reads mapping and counting

HISAT2 Galaxy Version 2.1.0 was used with default parameters to map reads to genome assembly hg38. We assessed the quality of the alignments by using the stats function of bamtools Galaxy Version 2.0.1 (50). We counted the reads mapping to each *hervgdb4* fragment and human gene by using the “union” mode in htseq-count Galaxy Version 0.6.1galaxy3 (51), and *hervgdb4* database (33) and gencode.v27 (52) for respectively LTR-retrotransposon and gene annotations. All the mentioned analyses were done on Galaxy release_17.09 locally installed (<http://galaxyproject.org/>). We calculated the expression values of the expressed *hervgdb4* fragments and genes as Transcripts Per Million (TPM).

Raw data filtering and transcriptome analysis

We selected all the genes and *hervgdb4* fragments with at least 1 count in at least 10 samples, conventionally considering them as expressed in at least one experimental condition and we used the obtained data for the analysis of HERVs/MaLRs transcriptome in PBMCs. We used

369 the GenomicRanges v.1.30.3 (53) R (version 3.4.4) package to obtain the coordinates of the
370 most intact HERV proviruses (4) from those of the expressed *hervgdb4* fragments.

371 **Hierarchical clustering**

372 The DESeq2.v.1.18.1 R (version 3.4.4) package (54) has been used to perform rlog
373 normalization on human genes and HERVs/MaLRs raw counts. From the output of the
374 normalization we extracted the HERVs/MaLRs rlog counts and we assessed the interpersonal
375 variability through PCA and Heatmap. The PCA was built with the function plotPCA in
376 DESeq2.v.1.18.1 and visualized by using ggplot 3.0.0 in R (version 3.4.4). The Heatmap of the
377 1000 *hervgdb4* fragments with the higher average rlog counts across samples was built
378 through the pheatmap 1.0.10 R (version 3.4.4) package, considering the correlation distance
379 across samples in column. The dist function in DESeq2.v.1.18.1 was applied to the transpose
380 of the rlog transformed count matrix to get Euclidean sample-to-sample distances; the
381 heatmaps were built through the pheatmap 1.0.10 R (version 3.4.4) package.

382 **Differential gene and *hervgdb4* fragment expression analysis**

383 We performed a differential expression analysis of the genes and *hervgdb4* fragments by using
384 the DESeq2.v.1.18.1 R (version 3.4.4) package (54) and the filtered raw counts as input.
385 During the analysis, false discovery rate/adjusted p-value were used for multiple test
386 comparison according to the Benjamini-Hochberg procedure (55). We used a threshold (FDR
387 ≤ 0.01 and absolute values of $\log_2$ Fold Change ≥ 1) to identify the modulated elements. The
388 differentially expressed *hervgdb4* fragments were visualized in a volcano plot built by using
389 ggplot 3.0.0 in R (version 3.4.4). We sorted the *hervgdb4* fragments by adjusted p-values to
390 recognize the mostly differentially expressed. We compared the TPM expression values with
391 the filtered adjusted p-values (≤ 0.01) to see the relative distributions, as describe in Fig. S1.
392 We manually checked for the presence of neighbor genes on ENSEMBL (56), within a 10-kb
393 window of distance from the nearest-neighbor gene.

- 394 1. Hattori M. 2005. Finishing the euchromatic sequence of the human genome.
395 Tanpakushitsu Kakusan Koso 50:162–168.
- 396 2. Bannert N, Kurth R. 2006. The evolutionary dynamics of human endogenous retroviral
397 families. *Annu Rev Genomics Hum Genet* 7:149–173.
- 398 3. Smit AFA. 1993. Identification of a new , abundant superfamily of mammalian LTR-
399 transposons 21:1863–1872.
- 400 4. Vargiu L, Rodriguez-Tomé P, Sperber GO, Cadeddu M, Grandi N, Blikstad V, Tramontano
401 E, Blomberg J. 2016. Classification and characterization of human endogenous
402 retroviruses; mosaic forms are common. *Retrovirology* 13:2–29.
- 403 5. Grandi N, Cadeddu M, Blomberg J, Mayer J, Tramontano E. 2018. HERV-W group
404 evolutionary history in non-human primates : characterization of ERV-W orthologs in
405 Catarrhini and related ERV groups in Platyrrhini. *BMC Evol Biol* 18:1–14.
- 406 6. Grandi N, Cadeddu M, Blomberg J, Tramontano E. 2016. Contribution of type W human
407 endogenous retrovirus to the human genome: characterization of HERV-W proviral
408 insertions and processed pseudogenes. *Retrovirology* 13:1–25.
- 409 7. Grandi N, Pisano MP, Tramontano E. 2019. The emerging field of human endogenous
410 retroviruses : understanding their physiological role and contribution to diseases.
411 *Future Virol* 14:441–444.
- 412 8. Seifarth W, Frank O, Zeilfelder U. 2005. Comprehensive analysis of human endogenous
413 retrovirus transcriptional activity in human tissues with a retrovirus-specific
414 microarray. *J Virol* 79:341–352.
- 415 9. Grandi N, Tramontano E. 2018. HERV envelope proteins: Physiological role and
416 pathogenic potential in cancer and autoimmunity. *Front Microbiol* 9:1–26.
- 417 10. Schiavetti F, Thonnard J, Colau D, Boon T, Coulie PG. 2002. A human endogenous
418 retroviral sequence encoding an antigen recognized on melanoma by cytolytic T
419 lymphocytes. *Cancer Res* 62:5510–5516.
- 420 11. Cheynet RIE, Bouton O, Blond J, Lavillette D. 2000. An Envelope Glycoprotein of the
421 Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses
422 Cells Expressing the Type D Mammalian Retrovirus Receptor. *J Virol* 4:3321–3329.
- 423 12. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, Lavallie E, Tang X, Edouard P, Howes
424 S, Jr JCK, McCoy JM. 2000. Syncytin is a captive retroviral envelope protein involved
425 403:785–789.
- 426 13. Doxiadis GGM, de Groot N, Bontrop RE. 2008. Impact of Endogenous Intronic
427 Retroviruses on Major Histocompatibility Complex Class II Diversity and Stability. *J*
428 *Virol* 82:6667–6677.
- 429 14. Pisano MP, Grandi N, Cadeddu M, Blomberg J, Tramontano E. 2019. Comprehensive
430 Characterization of the Human Endogenous Retrovirus HERV-K(HML-6) Group:
431 Overview of Structure, Phylogeny, and Contribution to the Human Genome. *J Virol*
432 93:1–19.
- 433 15. Grandi N, Cadeddu M, Pisano MP, Esposito F, Blomberg J, Tramontano E. 2017.
434 Identification of a novel HERV-K(HML10): Comprehensive characterization and
435 comparative analysis in non-human primates provide insights about HML10 proviruses
436 structure and diffusion. *Mob DNA* 8.
- 437 16. van de Lagemaat LN, Medstrand P, Mager DL. 2006. Multiple effects govern endogenous
438 retrovirus survival patterns in human gene introns. *Genome Biol* 7:R86.1-R86.14.
- 439 17. Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J. 2010. Long-range function of an
440 intergenic retrotransposon. *PNAs* 107:12992–12997.
- 441 18. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer
442 O, Trono D, Pfaff SL. 2013. ES cell potency fluctuates with endogenous retrovirus
443 activity 487:57–63.

- 444 19. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wesche D, Martin L, Ware C, Blish CA. 2015.
445 Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent
446 cells. *Nature* 522:221–225.
- 447 20. Grandi N, Tramontano E. 2018. Human Endogenous Retroviruses Are Ancient Acquired
448 Elements Still Shaping Innate Immune Responses. *Front Immunol* 9:1–16.
- 449 21. Hurst TP, Magiorkinis G. 2015. Activation of the innate immune response by
450 endogenous retroviruses. *J Gen Virol* 96:1207–1218.
- 451 22. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS,
452 Cope LM, Snyder A, Makarov V, Buhu S, Slamon DJ, Wolchok JD, Pardoll DM, Beckmann
453 MW, Zahnow CA, Mergoub T, Chan TA, Baylin SB, Strick R. 2015. Inhibiting DNA
454 Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous
455 Retroviruses. *Cell* 162:974–986.
- 456 23. Aswad A, Katzourakis A. 2012. Paleovirology and virally derived immunity. *Trends Ecol*
457 *Evol* 27:627–636.
- 458 24. Bouallaga I, Heidmann O, Letzelter C, Mangeney M, Renard M, Ducos B, Heidmann T.
459 2007. Placental syncytins : Genetic disjunction between the fusogenic and
460 immunosuppressive activity of retroviral envelope proteins.
- 461 25. Smedt M De, Leclercq G, Plum J, Leupin O, Marguerat S, Conrad B. 2019. Negative
462 thymocyte selection to HERV-K18 superantigens in humans *Franc* 105:4377–4383.
- 463 26. Hummel J, Ulrike K. 2015. Human endogenous retrovirus envelope proteins target
464 dendritic cells to suppress T-cell activation 1748–1759.
- 465 27. Madeira A, Burgelin I, Perron H, Curtin F, Lang AB, Faucard R. 2016. MSRV envelope
466 protein is a potent , endogenous and pathogenic agonist of human toll-like receptor 4 :
467 Relevance of GNBAC1 in multiple sclerosis treatment. *J Neuroimmunol* 291:29–38.
- 468 28. Hera D, Varade J, Garc M, Alcina A, Izquierdo G, Alloza I, Astobiza I, Leyva L, Ferna O,
469 Arroyo R, Roberto A, Matesanz F, Urcelay E. 2014. Human Endogenous Retrovirus
470 HERV-Fc1 Association with Multiple Sclerosis Susceptibility : A Meta-Analysis 9:1–6.
- 471 29. Mommert M, Tabone O, Oriol G, Cerrato E, Guichard A, Naville M, Fournier P, Volff JN,
472 Pachot A, Monneret G, Venet F, Brengel-Pesce K, Textoris J, Mallet F. 2018. LTR-
473 retrotransposon transcriptome modulation in response to endotoxin-induced stress in
474 PBMCs. *BMC Genomics* 19:1–17.
- 475 30. Johnston JB, Silva C, Holden J, Warren KG, Clark AW, Power C. 2001. Monocyte
476 Activation and Differentiation Augment Human Endogenous Retrovirus Expression :
477 Implications for Inflammatory Brain Diseases 434–442.
- 478 31. Tabone O, Mommert M, Jourdan C, Cerrato E, Legrand M, Lepape A, Allaouchiche B,
479 Rimmel T, Pachot A, Monneret G, Venet F, Mallet F. 2019. Endogenous Retroviruses
480 Transcriptional Modulation After Severe Infection , Trauma and Burn 9:1–12.
- 481 32. Grandi N. 2017. Integrations and Their Mobilization by L1 Machinery : Contribution to
482 the Human Transcriptome and Impact on the Host Physiopathology. *Viruses* 9:1–37.
- 483 33. Becker J, Prot P, Cheynet V, Oriol G, Mugnier N, Mommert M, Tabone O, Textoris J,
484 Veyrieras J, Mallet F. 2017. A comprehensive hybridization model allows whole HERV
485 transcriptome profiling using high density microarray 1–14.
- 486 34. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AFA, Wheeler TJ. 2016.
487 The Dfam database of repetitive DNA families. *Nucleic Acids Res* 44:D81–D89.
- 488 35. Sperber G, Lvgren A, Eriksson N, Benachenh F, Blomberg J. 2009. RetroTector online
489 , a rational tool for analysis of retroviral elements in small and medium size vertebrate
490 genomic sequences. *BMC Bioinformatics* 4:4–7.
- 491 36. Urrutia A, Duffy D, Rouilly V, Quintana-murci L, Albert ML, Inte M, Djebali R, Illanes G.
492 2016. Standardized Whole-Blood Transcriptional Profiling Enables the Deconvolution
493 of Complex Induced Standardized Whole-Blood Transcriptional Profiling Enables the

- Deconvolution of Complex Induced Immune Responses 1–15.
37. Arend WP, Gabay C. 2000. Physiologic role of interleukin-1 receptor antagonist. *Arthritis Res* 2:2–5.
38. Lee TH, Wisniewski H, Vildek J. 1992. A Novel Secretory Tumor Necrosis Factor-inducible Protein (TSG-6) Is a Member of the Family of Hyaluronate Binding Proteins , Closely Related to the Adhesion Receptor CD44. *J Cell Biol J Cell Biol* 116:545–557.
39. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. 2011. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* 8:1–22.
40. George R. Young, Sandra N. Terry, Lara Manganaro, Alvaro Cuesta-Dominguez GD, Dabeiba B-R, Laura Campisi, Ana Fernandez-Sesma RS, Viviana Simon LCFM. 2018. HIV-1 Infection of Primary CD4+ T Cells Regulates the Expression of Specific Human Endogenous Retrovirus HERV-K (HML-2) Elements. *J Virol* 92:1–13.
41. Mayer J, Harz C, Sanchez L, Pereira GC, Maldener E, Heras SR, Ostrow LW, Ravits J, Batra R, Meese E, García-pérez JL, Goodier JL. 2018. Transcriptional profiling of HERV-K (HML-2) in amyotrophic lateral sclerosis and potential implications for expression of HML-2 proteins. *Mol Neurodegener* 1–25.
42. Ma W, Hong Z, Liu H, Chen X, Ding L, Liu Z, Zhou F, Yuan Y. 2016. Human Endogenous retroviruses-k (HML-2) expression is correlated with prognosis and progress of hepatocellular carcinoma. *Biomed Res Int* 2016.
43. Balestrieri E, Pica F, Matteucci C, Zenobi R, Sorrentino R, Argaw-denboba A, Cipriani C, Bucci I, Sinibaldi-vallebona P. 2015. Transcriptional Activity of Human Endogenous Retroviruses in Human Peripheral Blood Mononuclear Cells 2015.
44. Ferguson JF, Patel PN, Shah RY, Mulvey CK, Gadi R, Nijjar PS, Usman HM, Mehta NN, Shah R, Master SR, Propert KJ, Reilly MP. 2013. Race and gender variation in response to evoked inflammation. *J Transl Med* 2013, 11:1–9.
45. Attig J, Young GR, Stoye JP, Kassiotis G. 2017. Physiological and pathological transcriptional activation of endogenous retroelements assessed by RNA-sequencing of B lymphocytes. *Front Microbiol* 8:1–11.
46. Douville R, Liu J, Rothstein J, Nath A. 2011. Identificaiton of Active Loci of a Human Endogenous Retrovirus in Neurons of Patients with Amyotrophic Lateral Sclerosis. *Ann Neurol* 69:141–151.
47. Chuong EB, Elde NC. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses 1083–1088.
48. Dyer DP, Salanga CL, Johns SC, Valdambrini E, Fuster MM, Milner CM, Day AJ, Handel TM. 2016. The Anti-inflammatory Protein TSG-6 Regulates Chemokine Function by Inhibiting Chemokine / Glycosaminoglycan. *J Biol Chem* 291:12627–12640.
49. Tao Z, Fusco A, Huang D, Gupta K, Young D, Ware CF. 2014. p100 / I κ B δ sequesters and inhibits NF- κ B through kappaBsosome formation. *PNAS* 111:15946–15951.
50. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT, Api T. 2011. BamTools : a C ++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692.
51. Anders S, Pyl PT, Huber W. 2015. Genome analysis HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.
52. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F. 2012. GENCODE: The Reference Human Genome Annotation for The ENCODE Project. *Genome Res* 22:1760–1774.
53. Aboyoun P, Carlson M, Lawrence M, Huber W, Gentleman R, Morgan MT, Carey VJ. 2013. Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* 9:1–10.
54. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion

544 for RNA-seq data with DESeq2. *Genome Biol* 15:1–21.

545 55. Hochberg YB and Y. 1995. Controlling the False Discovery Rate : A Practical and

546 Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef

547 Hochberg Source : *Journal of the Royal Statistical Society . Series B (Methodological)*,

548 Vol . 57 , No . 1 Published by : J R Stat Soc 57:289–300.

549 56. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K,

550 García Girón C, Hourlier T, Howe K, Kähäri A, Kokocinski F, Martin FJ, Murphy DN, Nag

551 R, Ruffier M, Schuster M, Tang YA, Vogel JH, White S, Zadissa A, Flicek P, Searle SMJ.

552 2016. The Ensembl gene annotation system. *Database (Oxford)* 2016:1–19.

553

Fig.1

a

b

555

556 **Fig. 1 The LTR-retrotransposon transcriptome in PBMCs**

557 Basal expression of the *hervgdb4* loci (a) and of the mostly intact HERV loci reported in Vargiu
558 *et al.* (4) (b) in PBMCs. All the expressed elements are grouped by retroviral classes. and
559 groups.

560

561

Fig.2

563 **Fig. 2 Principal Component Analysis (PCA) of samples**

564 PCA on rlog-normalized *hervgdb4* fragments expression data. It is possible to see the division
565 between not-stimulated and stimulated samples according to the PC1.

566

567

Fig.3

569 **Fig. 3 Heatmaps of the overall similarity between samples**

570 (a) Hierarchical clustering of the top 1000 *hervgdb4* fragments with the highest average of
571 rlog-normalized counts, correlation distance measure used in clustering columns. The two
572 clusters highlight specific signatures induced by LPS.

573 (b) Hierarchical clustering of the Euclidean sample-to-sample distance before (NS samples)
574 and after LPS injection (ST samples). We searched for difference related to the pattern of
575 expression of the 44 genes that captured the diversity of complex innate immune responses
576 (36) (44 immune-genes) and of the *hervgdb4* fragments. High- and low-responders showed
577 different response to inflammation.

578

579

Fig.4

Fig.4 Differential HERV/MaLR expression analysis

Volcano-plot of the differentially expressed *hervgbd4* fragments. Each point represents *hervgbd4* fragments, which spread according to the $\log_2$ fold change (x-axis), and the $\log_{10}$ adjusted p-values (y-axis). Red points are the significantly modulated *hervgbd4* fragments. For the 15 *hervgbd4* fragments with lowest adjusted p-values, the names are indicated.

Fig.5

Fig.5 Differentially expressed ReTe HERV proviruses

Bar-plot of the distribution of differentially expressed ReTe HERV proviruses across retroviral classes and groups. Different colors show up- and down-regulation.

Fig.6

595 **Fig.6 Co-regulation of *hervgbd4* fragments and human genes**

596 Boxplot of the TMP expression values of the *hervgbd4* fragments co-localized with human
 597 genes and the neighbor genes. The red boxes indicated values from non-stimulated samples;
 598 blue boxes indicated values from stimulated samples. Significant modulations according to
 599 the DEseq2 analysis ($\text{padj} < 0.01$) are marked with stars.

600

601

	Expressed		Modulated	
	Non DE	DE (%*)	Up-modulated	Down-modulated
<i>hervgdb4</i> fragments	54,061	6,452 (11%)	5,383	1,069
<i>hervgdb4</i> loci	31,079	4,607 (13%)	3,688	919
HERV proviruses	561	115 (17%)	86	29

* Proportion of expressed elements that are modulated

Table 1 LTR-retrotransposon modulation

Different proportion of modulated, up- and down-regulated *hervgdb4* fragments (a), *hervgdb4* loci (b) and ReTe proviruses (c) in PBMCs after *in vivo* LPS stimulation

Name	Description	Orientation	Locus coordinates	Host gene	Gene coordinates	Gene function	Position	paf1	log2FC	Avg TPM Pre-LPS	Avg TPM Post-LPS
081524901 ERVHSLU31	LTR (HERV)	NA	chr8:103002064-103004587	NA	NA	NA	NA	6.99E-36	6.3	2.3	99.3
081524801- HERV0183cL	solo LTR (HERV)	NA	chr8:103001306-103001748	NA	NA	NA	NA	6.89E-35	5.9	2.4	74.2
021629602- MALR1014uL	solo LTR (MALR)	-	chr2:113131173-113131620	ENSG00000136689	chr2:113099365-113134016	Interleukin 1 receptor antagonist	Inside	3.05E-30	7.2	2.2	156.7
021629601- MALR1014uL	solo LTR (MALR)	-	chr2:113131173-113131620	ENSG00000136689	chr2:113099365-113134016	Interleukin 1 receptor antagonist	Inside	3.38E-30	7.1	1.4	99.5
022246102- HERV0758dL	solo LTR (HERV)	+	chr2:151381023-151381962	ENSG00000123610	chr2:151357583-151381245	TNF alpha induced protein 6	Downstream	9.52E-22	7.9	0.0	2.0
081524501 PRM4pol	INTERNAL (HERV)	NA	chr8:102984015-102986529	NA	NA	NA	NA	1.89E-20	6.9	0.1	3.6
081524402- HERV0948dL	INTERNAL (HERV)	NA	chr8:102980120-102991770	NA	NA	NA	NA	3.84E-20	7.2	0.1	5.5
022246101- HERV0758dL	solo LTR (HERV)	+	chr2:151381023-151381962	ENSG00000123610	chr2:151357583-151381245	TNF alpha induced protein 6	Downstream	1.15E-19	6.7	0.1	5.6
062308001- HERV0741dL	solo LTR (HERV)	+	chr6:159677828-159678891	ENSG00000112096	chr6:159679064-159693234	Superoxide dismutase 2	Inside	3.45E-19	4.2	0.5	5.5
150538602- MALR1003uL	solo LTR (MALR)	+	chr15:63906995-63907370	ENSG00000035664	chr15:63907036-64040267	Death-associated protein kinase 2	Inside	5.86E-19	4.4	4.0	49.8
120084901- HERV0406uL	solo LTR (HERV)	-	chr12:8537686-8538696	ENSG00000166523	chr12:8533305-8540963	C-type lectin domain family 4 member E	Inside	1.41E-18	3.6	1.0	82.7
150538601- MALR1003uL	solo LTR (MALR)	+	chr15:63906995-63907370	ENSG00000035664	chr15:63907036-64040267	Death-associated protein kinase 2	Inside	2.73E-18	4.5	1.1	14.0
120085903 MR52env	INTERNAL (HERV)	NA	chr12:856648-8568185	NA	NA	NA	NA	6.20E-18	3.4	5.2	30.7
040314601- HERV0741uL	solo LTR (HERV)	+	chr4:15677257-15678021	ENSG0000023765	chr4:15604539-15681679	Family With Sequence Similarity 200 Member B	Upstream	1.54E-17	3.1	4.4	22.9
190407702- HERV0425uL	solo LTR (HERV)	+	chr19:35887015-35887532	ENSG00000167604	chr19:35887653-35896259	NFKB inhibitor delta	Downstream	1.54E-17	3.2	5.7	28.9

612 Same colors indicated fragments of the same locus
613

614

615 **Table 2** Top 15 most DE *hervgdb4* fragments

616 Description of the most 15 most DE *hervgdb4* fragments, including their expression levels
617 indicated as TPM values, and their context of insertion.

Expressed elements	hervgdb4 fragments	hervgdb4 loci	HERV proviruses
HERV	31751	18633	676
MaLR	28762	17053	NA
tot	60513	35686	676

Fig. S1 Experimental design of Differential Expression analysis.

RNA-seq workflow for the identification of modulated HERVs and MaLRs (a). The input files used are in blue boxes. The composition of hervgdb4 database is schematized in (b). The amount of expressed hervgdb4 fragments and loci have been obtained by filtering the raw counts, and are summarized in the table. The coordinates of the expressed ReTE most intact proviruses have been obtained by using the findOverlaps function from package "Iranges" and the coordinates of expressed hervgdb4 fragments.