



## Expression of endogenous retroviruses reflects increased usage of atypical enhancers in T cells

Saliha Azebi, Eric Batsche, Frédérique Michel, Etienne Kornobis, Christian Muchardt

### ► To cite this version:

Saliha Azebi, Eric Batsche, Frédérique Michel, Etienne Kornobis, Christian Muchardt. Expression of endogenous retroviruses reflects increased usage of atypical enhancers in T cells. *EMBO Journal*, 2019, 38 (12), pp.e101107. 10.15252/emboj.2018101107 . hal-02323835

**HAL Id: hal-02323835**

**<https://hal.science/hal-02323835>**

Submitted on 24 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Expression of endogenous retroviruses reflects increased usage of atypical enhancers in T cells

Saliha Azébi <sup>1,2</sup>, Eric Batsché <sup>1</sup>, Frédérique Michel <sup>3</sup>, Etienne Kornobis <sup>1</sup> and Christian Muchardt <sup>\*,1</sup>

<sup>1</sup>Unité de Régulation Epigénétique, UMR3738, CNRS, Institut Pasteur, Paris, France

<sup>2</sup>Ecole Doctorale “Complexité du Vivant” (ED515), Sorbonne Université, Paris, France

<sup>3</sup>Unit of Cytokine Signaling, Department of Immunology, Institut Pasteur, Paris, France

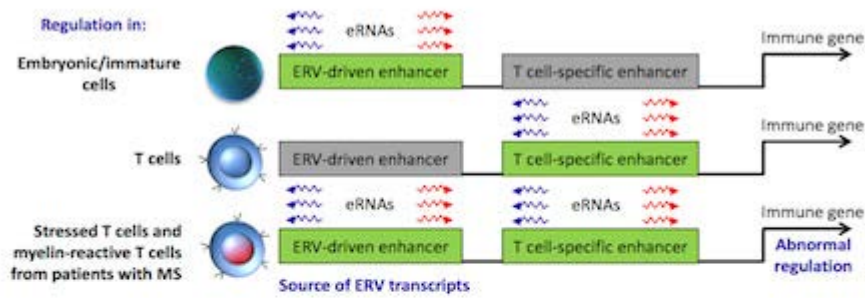
\*Corresponding author. Tel: +33 145688525; E-mail: christian.muchardt@pasteur.fr

EMBO J (2019)38:e101107<https://doi.org/10.15252/embj.2018101107>

## Abstract

Several autoimmune diseases including multiple sclerosis (MS) cause increased transcription of endogenous retroviruses (HERVs) normally repressed by heterochromatin. In parallel, HERV-derived sequences were reported to drive gene expression. Here, we have examined a possible link between promoter and enhancer divergent transcription and the production of HERV transcripts. We find that HERV-derived sequences are in general counter-selected at regulatory regions, a counter-selection that is strongest in brain tissues while very moderate in stem cells. By exposing T cells to the pesticide dieldrin, we further found that a series of HERV-driven enhancers otherwise active only at stem cell stages can be reactivated by stress. This in part relies on peptidylarginine deiminase activity, possibly participating in the reawakening of silenced enhancers. Likewise, usage of HERV-driven enhancers was increased in myelin-reactive T cells from patients with MS, correlating with activation of nearby genes at several sites. Altogether, we propose that HERV-driven enhancers constitute a reservoir of auxiliary enhancers transiently induced by stress while chronically active in diseases like MS.

# Synopsis



Human endogenous retroviruses (HERVs) derived from ancient exogenous viruses can act as *cis*-enhancers. This study shows that increased disease-related gene expression is in part due to a chronic increase of HERV *cis*-enhancer activity in the context of autoimmune diseases.

- Enhancers driven by HERVs are enriched in the neighborhood of immune genes.
- HERV-driven enhancers are mostly active in embryonic cells but silenced in mature adult tissues.
- Extensive stress causes transient reactivation of HERV transcription.
- In multiple sclerosis, some HERV-driven regulatory regions become chronically active.

## Introduction

Several autoimmune diseases including multiple sclerosis (MS), type 1 diabetes mellitus, and rheumatoid arthritis are associated with increased transcription of sequences originating from ancient retroviral infections of the germ line. The possible role of these human endogenous retroviruses (HERVs) in the onset of autoimmune diseases has been extensively investigated, mostly with a focus on their protein products, that trigger immune reactions and occasionally assemble into virus-like particles (Christensen, [2016](#); Nexø *et al.*, [2016](#)). In contrast, the regulatory mechanisms at the root of the increased transcription of HERV-derived sequences in the patients remain essentially unexplored. This is an important issue as the understanding of transcriptional events associated with autoimmune diseases may help unraveling the mechanisms leading to their onset.

Transposable elements are maintained under strict transcriptional control *via* a combination of DNA methylation, histone H3 lysine 9 methylation, and binding of KRAB domain-containing zinc-finger proteins (Karimi *et al.*, [2011](#); Imbeault *et al.*, [2017](#)). It has been suggested that the function of the silencing is to restrain the mobility of the repeated elements and prevent them from damaging the genome. While mobility may be an issue for other transposable elements like LINES that maintain transposition capacity, there is currently no evidence for new somatic or germ line insertions of HERVs in humans, and nearly all have lost coding potential (Bannert & Kurth, [2006](#); Magiorkinis *et al.*, [2015](#)). Another possibility is that HERV sequences require silencing because of their regulatory potential. Indeed, HERVs are an

important source of *cis*-regulatory elements initially necessary for the viral cycle, and several studies have established exaptation of HERV-derived sequences for the transcriptional regulation of host genes, mostly involved in stemness and development, but also in immunity and antiviral defense (Sundaram *et al*, [2014](#); Wang *et al*, [2014](#); Chuong *et al*, [2016](#); Hackett *et al*, [2017](#); Hummel *et al*, [2017](#); Imbeault *et al*, [2017](#)). As an argument in favor of a regulatory role of HERVs, post-translational histone modifications characteristic of promoters and enhancers have been detected on retroviral sequences (Chuong *et al*, [2013](#); Xie *et al*, [2013](#)). These histone modifications include the promoter-specific histone H3 lysine 4 trimethylation and the enhancer-enriched histone H3 lysine 4 monomethylation and lysine 27 acetylation (Calo & Wysocka, [2013](#)). In parallel, HERVs have been reported to contribute significantly more than expected by chance to DNase I accessible regions, characteristic of transcribed DNA sequences (Jacques *et al*, [2013](#)).

The implication of HERV-derived sequences in transcriptional regulation justifies their tight regulation, but it may also provide an explanation for their transcription. Indeed, promoters and enhancers are sites of (mostly) bidirectional transcription and the abundance of product of this transcription (uaRNAs for promoters and eRNAs for enhancers) reflects the activity of the regulatory sequences (Kim *et al*, [2010](#); Melgar *et al*, [2011](#); Hah *et al*, [2013](#)). Thus, promoter or enhancer activity represents an opportunity for HERV-derived sequences to be transcribed, whether they are located at the site of transcription initiation or in its vicinity, and this could be the basis for HERV transcripts detected in autoimmune diseases.

To investigate this possibility, we have here examined the position of HERV-derived sequences relative to enhancers and promoters. For this, we have taken advantage of the systematic mapping of functional domains by the NIH Roadmap Epigenomics Mapping and the Fantom5 consortia, respectively, relying on combinations of histone modifications detected by chromatin immunoprecipitation assays and on Cap Analysis of Gene Expression (CAGE), defining sites of transcription initiation. The approach showed that the bulk of HERV-containing sequences are counter-selected inside promoters and enhancers and even more so at the borders of these regulatory regions. Yet, it also confirmed that a subset of HERV sequences function as *cis*-regulatory elements active mostly in embryonic stem cells and frequently located in the neighborhood of genes involved in innate immune defense. To investigate whether transcriptional activation of such HERV-driven *cis*-regulatory elements could be a source of disease-related HERV transcripts, we identified a small molecule able to induce transcription of several MS-associated HERVs. This allowed us to show that production of disease-related HERV transcripts accompanies an acute stress-induced transcriptional response that involves awakening of *cis*-regulatory elements atypical for mature T cells and enriched in HERV-driven enhancers. Finally, examination of transcriptomic data from patients with MS revealed a chronic disease-associated increase in the usage of HERV-driven enhancers, resulting in the production of HERV transcripts, and correlating with transcriptional activation of nearby genes.

## Results

### HERV-driven enhancers and promoters are rare and tissue-specific

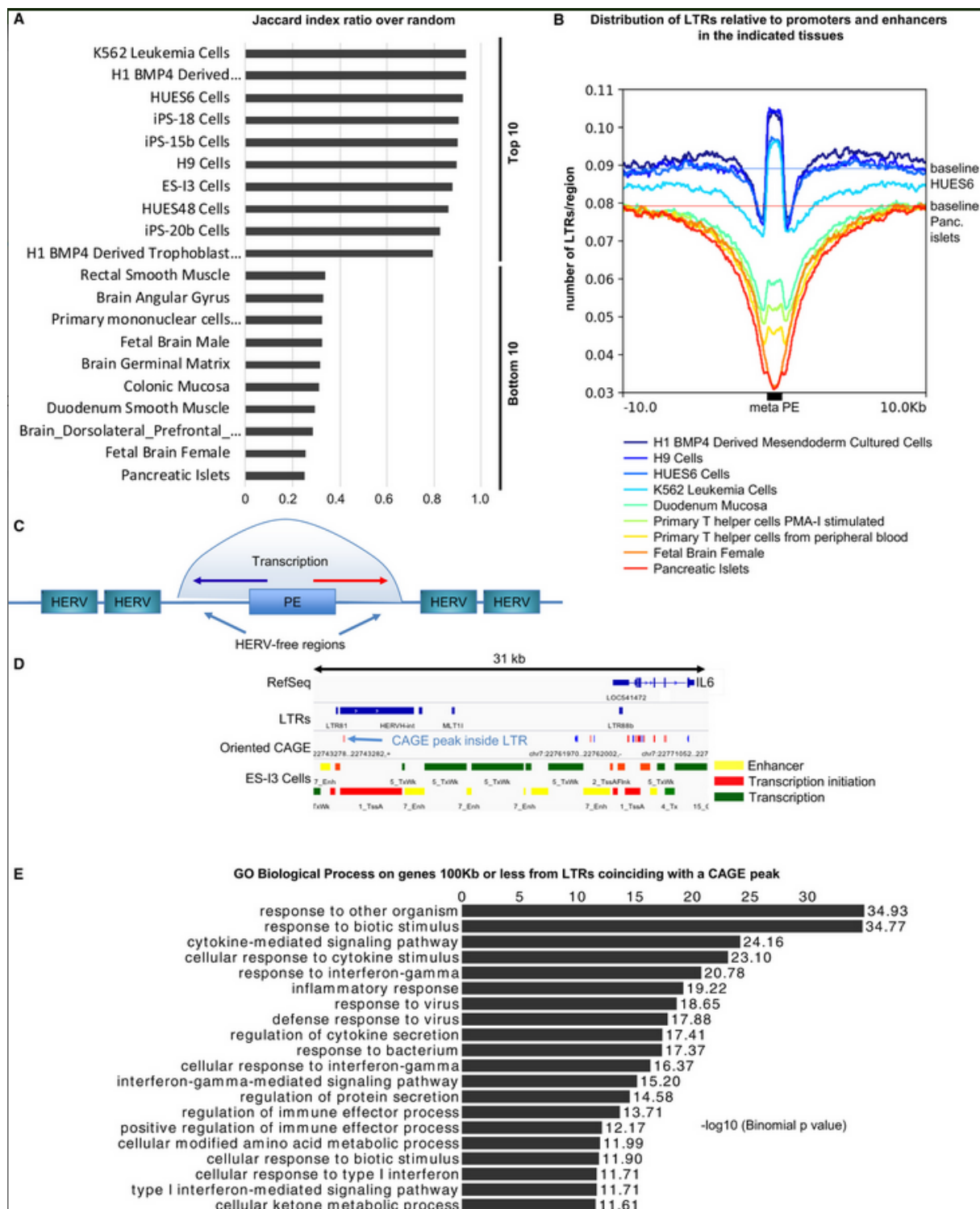
To reach an overview of the participation of HERV-derived sequences in transcription initiation throughout human tissues, we extracted all regions annotated as active promoters or enhancers in the 127 tissues mapped by the NIH Roadmap Epigenomics Mapping Consortium (Roadmap Epigenomics Consortium *et al*, [2015](#)). These regions will be referred to as PEs for



“Promoter or Enhancer” regions. To estimate the similarity between each of the 127 sets of PEs and regions annotated “LTR” in RepeatMasker, we used the Jaccard index defined as the size of the intersection divided by the size of the union of the sample sets. As the total number of PEs varies among the tissues, we also calculated the Jaccard index between the PEs from each tissue and randomly selected non-LTR regions (average of thousand iterations). The score shown for each tissue is the Jaccard index (PEs vs. LTRs) divided by the Jaccard index (PEs vs. random). The approach showed that in all tissues, the overlap between PEs and HERVs is smaller than that expected by chance (all ratios are below 1, Fig 1A, full list [Appendix Fig S1A](#)). It also showed that overlap of HERV sequences with PEs varied extensively from one tissue to the other, with several pluripotent stem and cancer cell lines harboring HERV sequences inside their PEs approximately fourfold more frequently than tissues at the bottom of the list. These last-mentioned tissues notably included pancreatic islets, several brain tissues, and tissues from the gastrointestinal tract. To further explore the tissue specificity of HERV-driven PEs, we also compared LTRs to either ubiquitous or tissue-specific PEs. Likewise, we compared LTRs to either evolutionary conserved or human-specific liver *cis*-regulatory elements ([Appendix Fig S1B](#)). In these tests, tissue-specific and human-specific elements reached the best similarity scores with LTRs, in agreement with earlier observations (Trizzino *et al.*, [2017](#), [2018](#)).

We next examined the position of HERV sequences relative to PEs. In all tissues, the PEs were located at the bottom of a HERV-depleted valley (Fig 1B). This indicates that the neighborhood of PEs is depleted of HERV sequences, possibly to avoid their serendipitous transcription (schematic in Fig 1C). In all tissues except pancreatic islets and fetal female brain, we also noted an upturn of variable size within the core of the PEs. In a vast majority of tissues, this upturn or “HERV-butte” remained below the basal level (considered 10 kb away from the PE—see, e.g., HUES6 and pancreatic islets). However, in a series of embryonic stem and iPS cells and in the K562 leukemia cell line, the HERV-butte reached at or above the baseline, indicative of extensive usage of HERV sequences in the PEs of these cells (Fig 1B and [Appendix Fig S1C–F](#)). We further tested whether the evolutionary age of the LTRs would affect their position relative to PEs ([Appendix Fig S1G–I](#)). Examinations of the profiles suggested that the ERV1 clade (approximately 170,000 copies) is better tolerated in the neighborhood of PEs than is the older ERVL clade (approximately 500,000 copies), in agreement with an earlier study showing that ERV1 is the most abundant LTR within *cis*-regulatory elements (Trizzino *et al.*, [2017](#)). The younger ERVKs were too few in number for the profiles to be meaningful (approximately 10,000 copies).

The presence of a HERV-butte inside the boundaries of the PEs in most tissues prompted us to specifically examine HERVs likely to be the functional core of promoters or enhancers. To that end, we listed RepeatMasker “LTRs” overlapping with a PE in at least one of the 127 tissues while also matching a Fantom5 Cap Analysis of Gene Expression (CAGE) peak (example in Fig 1D and [Appendix Fig S1M](#)). These CAGE peaks define sites of transcriptional initiation active in at least one of the 975 CAGE libraries examined by the Fantom5 consortium (DGT, [2014](#)). This approach selected approximately 1% of all HERV sequences (6684 “CAGE-LTRs” out of 708210 regions annotated “LTR”). Interestingly, CAGE-LTRs had highest similarity (Jaccard index) with PEs from several hematopoietic tissues, while brain tissues segregated clearly at the bottom of the list ([Appendix Fig S1N](#)). We then identified genes located within 100 kb of the CAGE-LTRs using GREAT (McLean *et al.*, [2010](#)). GO term analysis of these genes showed a clear enrichment in genes involved in cytokine and inflammatory response (Fig 1E).



**Figure 1. Genome-wide localization of HERVs relative to promoters and enhancers**

- Comparison of regions annotated “LTR” in RepeatMasker with the regions annotated TssA, TssAFlnk, Enh, EnhG, TssBiv, or EnhBiv in the 15 core marks model of the Epigenomic Roadmap consortium (referred to as PEs for Promoters and Enhancers). For each tissue, the Jaccard index comparing PEs to LTRs is divided by the average Jaccard index (1,000 iterations) obtained when comparing PEs to randomly selected genomic locations (of the same sizes as the LTRs). Only the 10 tissues with the highest and the lowest score are shown. Full list in [Appendix Fig S1A](#).
- Profile reporting the position of regions annotated “LTR” in RepeatMasker relative to PEs from the indicated tissues. Red and blue horizontal lines materialize the level of LTR density expected in a random sequence from pancreatic islets and human embryonic stem cell line HUES6, respectively.
- A graphic interpretation of the positioning of PEs at the bottom of LTR valleys in panel (A).
- Example of a region annotated “LTR” in RepeatMasker and overlapping with a PE and with a CAGE peak from the FANTOM5 consortium data, indicative of a site of transcription initiation.

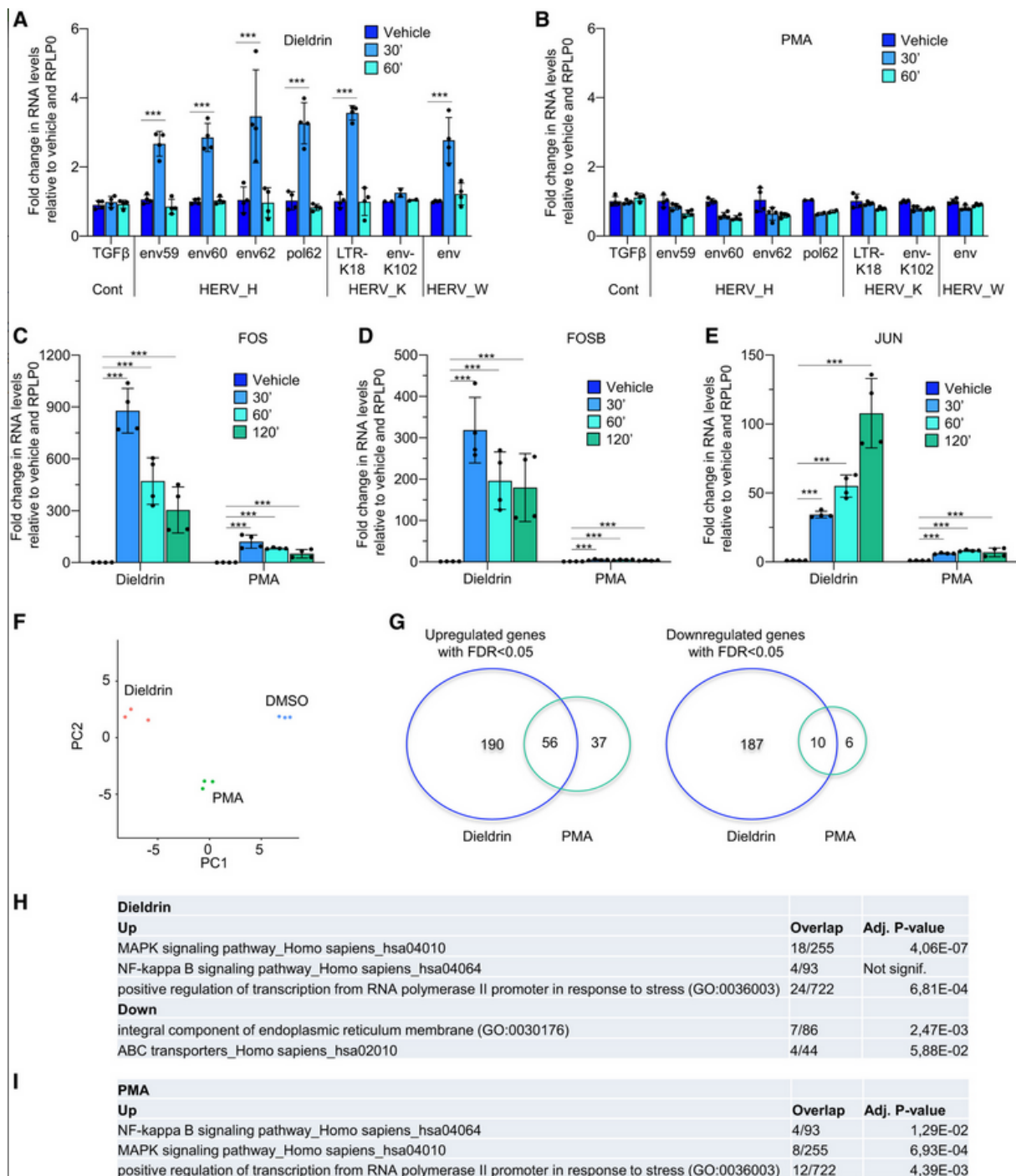
- E. Biological process GO terms associated with genes located 100 kb or less from a region annotated “LTR” in RepeatMasker and overlapping with a PE and with a CAGE peak.

As above, we next stratified the LTRs according to their evolutionary age. The ERV1 clade showed the highest similarity with CAGE peaks ([Appendix Fig S1J](#)) with approximately 2% of the ERV1 copies hosting a CAGE peak; the younger ERVKs showed the lowest similarity. GO term analysis of genes located within 100 kb of CAGE-containing ERV1s essentially recapitulated that reached when considering all CAGE-LTRs ([Appendix Fig S1L](#)). GO term analysis of genes located within 100 kb of CAGE-containing ERVLs also designated pathways related to innate immunity, yet with a focus on the interferon gamma response ([Appendix Fig S1K](#)). In contrast, ERVKs did not appear to have specialized in any specific pathway. We note however here that, due to the difficulty of their single-locus mapping, the impact of these relatively young HERVs on transcriptional regulation may be underestimated.

Altogether, we concluded from this analysis that (i) most HERV sequences are located away from sites of transcriptional initiation and are therefore less likely to be transcribed accidentally upon activation of a neighboring promoter or enhancer, (ii) the usage of HERVs as *cis*-regulatory elements varies extensively from one tissue to the other, being highest in stem and cancer cells and lowest in brain tissues, (iii) *cis*-regulatory elements borne by HERV sequences, while associated with multiple regulatory pathways, play a distinct role in the regulation of immune genes, in line with earlier observations on the role of HERV sequences in cellular defense (Chuong *et al*, [2016](#); Grandi & Tramontano, [2018](#)).

## Dieldrin as a tool to induce transcription of disease-related HERVs

To challenge the bioinformatic analysis described above, we searched for a reliable tool to activate HERVs in a tissue culture cell line. In an initial approach, we chose to monitor some of the HERVs of the H, K, and W clades frequently described as activated in patients with autoimmune diseases ([Appendix Fig S2A](#)). Searching the literature for compounds potentially interfering with heterochromatin-mediated repression brought us to test dieldrin, an organochlorine pesticide described as an inducer of histone acetylation in dopaminergic neuronal N27 cells (Song *et al*, [2010](#)). In our assays, we examined the effect of dieldrin on HERV transcription in the Jurkat CD4<sup>+</sup> T-cell line in which we had previously observed inducible HERV activity (Sharma *et al*, [2012](#)). The sequences of the primers we used to detect HERV transcription were obtained from the literature, yet they all detected more than one locus in both Jurkat and haploid HAP1 cells as determined by digital droplet PCR ([Appendix Fig S2B](#)). Titration experiments showed that Jurkat cells tolerated dieldrin at concentrations up to 100  $\mu$ M for 1 h with no significant cell death (< 2%—[Appendix Fig S2C](#)). Exposure to that dosage resulted in a very transient increase in transcription for 6 of the 7 tested HERVs (twofold to fourfold) peaking at 30 min, while being undetectable at 60 min ([Fig 2A](#)). In comparison, activation of the Jurkat T cells with 100  $\mu$ M of the phorbol ester PMA did not induce HERV expression ([Fig 2B](#)). The concentration of PMA was here intentionally higher than that usually used (40 nM) to ascertain that the absence of HERV stimulation was not a consequence of insufficient dosage. Cell responsiveness to PMA was verified by monitoring activation of two cytokine genes TNF $\alpha$  and IL8 ([Appendix Fig S2D](#)). As a negative control, we monitored TGF $\beta$ , a gene induced neither by PMA nor dieldrin ([Fig 2A](#)).



**Figure 2. Dieldrin is a strong T-cell activator causing transcription of several HERVs**

A–E. Jurkat T cells were treated with either DMSO (vehicle) or with 100  $\mu$ M of either dieldrin or PMA for the indicated times. Abundance of mRNA for the indicated genes or HERVs was assessed by RT–qPCR. Data shown are means  $\pm$  SEM from four independent experiments. Significance (*P*-value) was estimated using the two-sided Student's *t*-test appropriate for small sample numbers. \*\*\**P*-value < 0.001.

F. Jurkat cells were exposed to either DMSO, PMA in DMSO, or dieldrin in DMSO for 30 min in triplicates. RNA-seq was then performed on cDNA libraries prepared with poly(A) selection: principal component analysis on the 485 genes the most affected by DMSO, PMA, or dieldrin.

G. Venn diagrams reporting genes up- or down-regulated by dieldrin or PMA as compared to DMSO.

H, I. Gene ontology analysis on genes up- and down-regulated by either dieldrin or PMA as estimated from RNA-seq data.



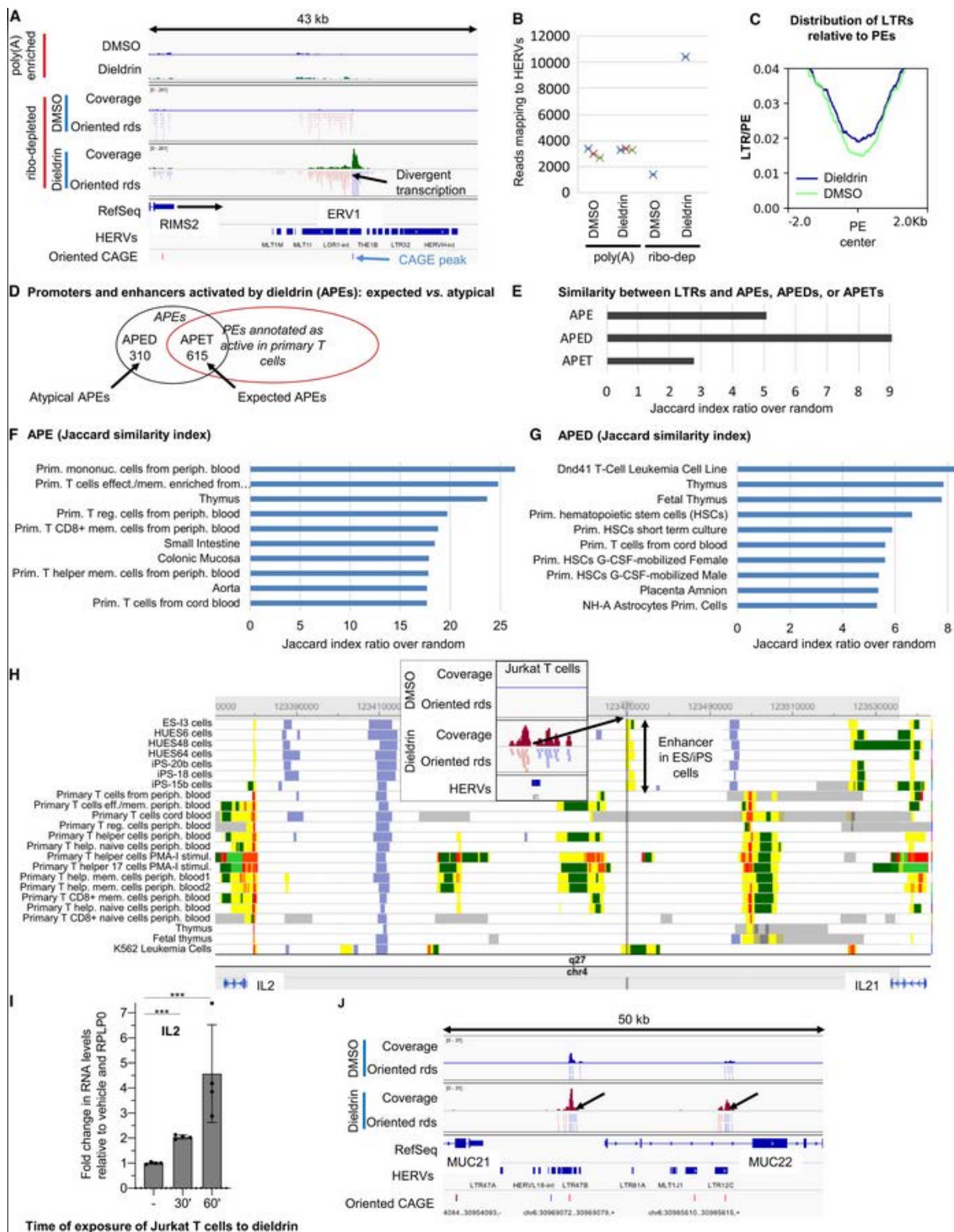
To gain insight into the global effect of dieldrin on gene expression, we used next-generation sequencing of a poly(A)-enriched random-primed cDNA library after a 30-min exposure. The transcriptome was compared to that of untreated cells or cells activated with PMA. Dieldrin caused exceptionally strong activation of multiple markers of T-cell activation, including CD69, SERPINE1, and the immediate early genes JUN, FOS, FOSB, EGR2/3/4, and NR4A1/2/3 ([Table EV1](#), [Fig 2C–E](#), and MA plots [Appendix Fig S2E–F](#)). Yet, principal component analysis positioned the effect of dieldrin away from that of PMA ([Fig 2F](#)). In particular, the effect of dieldrin was more extensive than that of PMA with 246 genes significantly upregulated by dieldrin, while only 93 genes were upregulated by PMA (false discovery rate, FDR < 0.05). KEGG pathway analysis further suggested that dieldrin was a particularly efficient activator of the MAPK pathway ([Fig 2H and I](#)). Finally, dieldrin caused moderate (less than twofold) but significant (FDR < 0.05) transcriptional repression of a large set of genes (197 genes—[Fig 2G](#) and [Table EV1](#)) along with moderate (less than twofold) activation of heat-shock proteins HSPA6, HSPA1A/B, and HSPH1. This was suggestive of an early phase of a heat-shock response antagonizing protein misfolding in the cytosol (Mahat *et al.*, [2016](#)). Dieldrin did not seem to induce other types of stress as confirmed by RT-qPCR analysis of a series of gene markers for environmental, oxidative, osmotic, cytotoxic, hypoxic, endoplasmic reticulum, or metal stress ([Appendix Fig S2G and H](#)).

### Atypical T-cell enhancers are enriched in HERVs

We next questioned whether the transcription of HERV-derived sequences induced by dieldrin could be an indicator of increased usage of HERV-driven enhancers. HERV transcription induced by dieldrin was poorly detected in the poly(A)-enriched data (see example of an ERV1 downstream of the RIMS2 gene, [Fig 3A](#), top lanes). We therefore repeated the RNA-seq on one set of samples using cDNA libraries constructed from ribo-depleted total RNA. This protocol allowed us to detect and quantify non-coding RNAs, including those resulting from transcription of HERV-derived sequences (see example [Fig 3A](#), bottom). The inadequacy of the poly(A)-enriched data for the detection of HERV transcription was confirmed by examining a series of 4686 HERVs transcribed twofold or more in the ribo-depleted data. On average, the read count at these sequences was not affected by dieldrin in the data from the poly(A)-enriched transcripts in any of the replicates ([Fig 3B](#)). These observations suggested that HERV transcripts in their majority are not stably polyadenylated and therefore unlikely to be messenger RNAs (mRNAs).

We next mapped putative promoters and enhancers based on detection of sites of divergent transcription in both DMSO- (solvent) and dieldrin-treated cells. To increase robustness, we retained only positions previously annotated as sites of transcription initiation in at least one of the 975 CAGE libraries of the Fantom5 consortium. Examination of the distribution of RepeatMasker LTRs relative to the sites of divergent transcription revealed that dieldrin did not allow overcoming the counter-selection of HERV sequences in PEs, yet it increased the usage of HERV-containing PEs by approximately 25% ([Fig 3C](#), compare blue and green lines).

We next concentrated on sites of divergent transcription activated twofold or more in the presence of dieldrin (referred to as APEs for activated promoters and enhancers—[Table EV2](#)).



**Figure 3. Dieldrin causes transcription of HERV-containing PEs atypical for activated T cells**

Jurkat cells were exposed to either DMSO or dieldrin in DMSO for 30 min. RNA-seq was then performed on cDNA libraries prepared after depletion of ribosomal RNAs.

- Screen captures from IGV showing an example of a HERV becoming a site of divergent transcription upon exposure of the Jurkat cells to dieldrin. Top tracks show coverage from sequencing of poly(A)-enriched libraries for comparison. All coverage tracks are at the same scale, and all the sequencing result files are of comparable size. Blue and red reads (rds) are in opposite orientation. CAGE peaks use the same color code.
- Reads mapping inside a series of 4,686 HERVs located away (30 kb) from protein-coding genes were quantified both in the poly(A)-enriched and the ribo-depleted data. Read counts from each replicate were then summed and plotted.

- C. Distribution of regions annotated “LTR” in RepeatMasker relative to sites of divergent transcription in Jurkat cells treated with either DMSO or diethylstilbestrol in DMSO.
- D. Schematic showing that only a fraction of the APEs (sites of divergent transcription activated twofold or more by diethylstilbestrol) match promoters and enhancers annotated as active in T cells (APETs). The remaining APEs are designated APEDs (specific to the diethylstilbestrol treatment).
- E. Comparison of regions annotated “LTR” in RepeatMasker with the APEs, APETs, and APEDs. The graph provides the ratio between the Jaccard index of the indicated category over the average of Jaccard indexes calculated for 1,000 series of randomly selected PEs.
- F. , G. Comparison of either APEs or APEDs as indicated with the regions annotated TssA, TssAFlnk, Enh, EnhG, TssBiv, or EnhBiv in the 15 core marks model of the Epigenomic Roadmap consortium as in Fig 1A.
- H. Region upstream of IL2 as a screen capture from the genome browser of the Epigenomic Roadmap consortium. Promoters and enhancers are annotated in red and yellow, respectively. Green indicates transcription. Insert is a detail from [Appendix Fig S3A](#) showing the site of divergent transcription induced by diethylstilbestrol.
- I. Jurkat T cells were treated with either DMSO (vehicle) or 100  $\mu$ M diethylstilbestrol for the indicated times. Abundance of IL2 mRNA was assessed by RT-qPCR. Data shown are means  $\pm$  SEM from four independent experiments. Significance (*P*-value) was estimated using the two-sided Student's *t*-test appropriate for small sample numbers. \*\*\**P*-value < 0.001.
- J. Screen captures from IGV showing increased divergent transcription inside a HERV sequence upon exposure of the Jurkat cells to diethylstilbestrol in the intergenic region between MUC21 and MUC22. Color code as in (A).

The Jaccard similarity index between APEs and RepeatMasker LTRs was fivefold higher than expected for randomly selected sites of divergent transcription, indicating a further enrichment in HERV sequences among sites of divergent transcription strongly activated by diethylstilbestrol (Fig 3E, APE). We next separated APEs annotated as active in T-cell epigenomes (APETs) from those atypical for T cells and more likely to be diethylstilbestrol-specific (termed APED for diethylstilbestrol-specific, Fig 3D). This latter group was threefold richer in HERV sequences than the APETs as shown by the Jaccard similarity index (Fig 3E, compare APET and APED). Furthermore, comparing APEs and APEDs with the PEs from the 127 epigenomes (Jaccard index) showed that filtering away T-cell PEs from the APEs, generated a list of regulatory sites sharing similarity with PEs from the cancer cell line Dnd41, from fetal thymus, and from several hematopoietic stem cell lines (Fig 3F and G). This suggested that APEDs included regulatory sequences active in less differentiated hematopoietic cells.

Taken together, this analysis suggested that diethylstilbestrol causes activation of regulatory sequences required for T-cell activation, but also of more atypical regulatory regions enriched in HERV sequences. To identify examples of such atypical regulatory regions, we examined individually some of the sites of divergent transcription overlapping with HERV sequences and activated by diethylstilbestrol. We particularly noted the presence of a site located upstream of IL2, annotated as an active enhancer in stem cells and in the K562 leukemia-derived cell line, but not in primary hematopoietic cells (Fig 3H and [Appendix Fig S3A](#)). At the time point examined by RNA-seq (30 min of exposure to diethylstilbestrol), the IL2 gene is still in a phase of progressive activation (Fig 3I). Remarkably, at this time, the HERV-driven enhancer had a transcriptional activity stronger than that observed at the two enhancers annotated as active in T cells (arrows, [Appendix Fig S3A](#)). This observation suggested that activation of the IL2 gene by diethylstilbestrol largely relied on the activity of a HERV-driven enhancer atypical for T cells. The atypical nature of enhancers overlapping with HERV sequences was also well illustrated by a site located upstream of the gene encoding the Fc receptor-like protein 4 (FCRL4) on chromosome 1, annotated as an enhancer in T cells from cord blood and in the Dnd41 T-cell leukemia cell line, while annotated as Polycomb-repressed, heterochromatinized, or quiescent in all other T-cell epigenomes ([Appendix Fig S3B](#)). Finally, we noted two unannotated HERV-based enhancers in the neighborhood of mucin genes MUC21 and MUC22, further suggesting that HERV sequences provide alternative enhancers for genes involved in cellular defense (Fig 3J and [Appendix Fig S3C](#)).

## Atypical activation of multiple pathways by dieldrin is required for HERV transcription

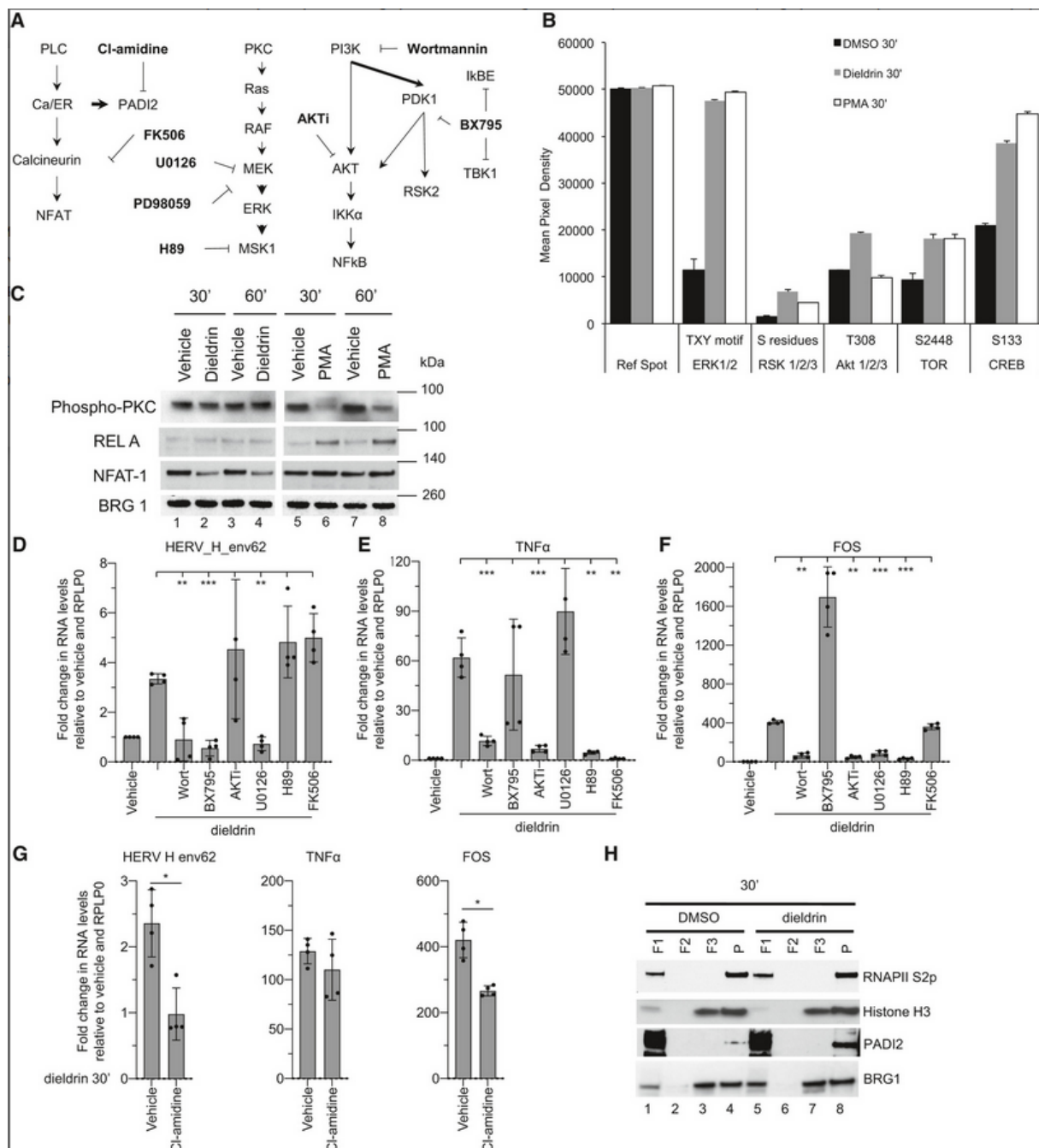
To gain a better insight into how dieldrin may drive Jurkat cells into using atypical enhancers, we next investigated the signal transduction pathways stimulated by this molecule in comparison with PMA. A human phospho-kinase array showed that both dieldrin and PMA treatments activated the MAP kinase pathway as demonstrated by the phosphorylation of the TXY motif in ERK1/2 proteins and of the SRF kinases RSK1/2/3 (Fig 4A and B, and [Appendix Fig S4A and B](#)). In this context, we also examined phosphorylation of PKC, the primary target of PMA. Surprisingly, dieldrin, unlike PMA, did not cause any detectable change in PKC phosphorylation when examined by Western blot with a pan phospho-PKC antibody (Fig 4C phospho-PKC, compare lanes 2 and 6), suggesting that dieldrin activates the MAP kinase pathway independently of the PKC pathway.

Jurkat cells fail to express PTEN and SHIP lipid phosphatases and therefore have elevated basal activity of the PI3K pathway. This activity accounted for the high levels of AKT S473 phosphorylation detected on the phospho-kinase arrays ([Appendix Fig S4A](#), arrow). Nevertheless, both PMA and dieldrin caused increased TOR phosphorylation at S2448, suggestive of a further activation of the pathway (Fig 4B). However, only PMA induces detectable phosphorylation of REL A, a downstream target of the PI3K pathway (Fig 4C REL A, compare lanes 2 and 6), again suggesting an atypical effect of dieldrin on this kinase cascade.

Finally, we examined pathways activated by increased calcium concentrations. On the kinase array, both dieldrin and PMA caused phosphorylation of CREB at S133, a modification catalyzed either by PKA or by CamK in the presence of  $\text{Ca}^{2+}$ . In contrast, dieldrin, but not PMA, caused NFATc2 de-phosphorylation, an indicator of the activation of the calcium/calmodulin pathway (Fig 4C NFAT-1, compare lanes 2 and 6).

Together, these observations documented a simultaneous activation by dieldrin of the MAPK, the PI3K, and the calcium/calmodulin pathways, possibly via atypical routes. To investigate the importance of these pathways in the activation of HERV transcription, we tested several small-molecule inhibitors. For these experiments, we monitored transcription using the HERVH env62 primer set. Indeed, sequencing of the resulting PCR product indicated that it largely originated from a locus downstream of GALNT3, a gene transcribed with the same kinetic as HERVH env62 in the presence of dieldrin ([Appendix Fig S4C and D](#)). ERK inhibitors U0126 and PD98059 caused strong inhibition of transcription of this HERV (Fig 4D–F and [Appendix Fig S4E–G](#)). In contrast H89, an inhibitor of PKA and MSK1 inhibited the transcription of several AP1 genes, but had no effect on the HERV. Wortmannin very efficiently inhibited the HERV, documenting the implication of the PI3K pathway. Yet, the Akt inhibitor, efficient on the AP1 genes, had no effect on the HERV. To investigate this further, we followed up on the fact that dieldrin but not PMA caused phosphorylation of Akt at T308 (Fig 4B), a residue modified by PDK1 (Manning & Toker, 2017). BX795, an inhibitor PDK1, resulted in repression of the HERV, without affecting any of the other genes we tested. As BX795 also targets TBK1 and IKKε, it is difficult to draw definitive conclusions on the pathway at play, but this molecule may provide a potential therapeutic lead in diseases associated with increased HERV expression.





**Figure 4. Dieldrin activates multiple signal transduction pathways**

**A.** Schematic of the signal transduction pathways under scrutiny. Used inhibitors are indicated in bold font.

**B.** A human phospho-kinase antibody array was probed with extracts from Jurkat cells treated with either DMSO, dieldrin, or PMA for 30 min as indicated. The array processed as a Western blot and revealed by ECL. Histogram shows selected antibodies. Values shown are means  $\pm$  SEM from two measures. The full experiment is shown in [Appendix Fig S4A and B](#).

**C.** Jurkat cells were treated with either DMSO (vehicle), dieldrin or PMA for the indicated times. Cell extracts were resolved by PAGE, and Western blotting was carried out with the indicated antibodies. Nearest size marker is indicated to the right (kDa).

**D–G.** Jurkat cells were cultured in the absence or in the presence of the indicated small molecule inhibitors for 1 h then exposed to 100  $\mu$ M dieldrin for 30 min. Abundance of mRNA for the indicated genes or HERVs was assessed by RT-qPCR. Data shown are means  $\pm$  SEM from four independent experiments. Significance (*P*-value) was estimated using the two-sided Student's *t*-test appropriate for small sample numbers. \*\*\**P*-value < 0.001, \*\*0.001 < *P*-value < 0.01, \*0.01 < *P*-value < 0.05.

**H.** Jurkat cells treated as in (C) were subject to fractionation as indicated in the [Materials and Methods](#) section to distinguish a fraction with mostly RNAPII S2p (F1), one with mostly histone H3 (F3), and one with both RNAPII S2p and histone H3 (P). Fractions were resolved by PAGE, and Western blots were carried out with the indicated antibodies.

Surprisingly, the calcineurin inhibitor FK506 did not affect transcription of HERVH env62 in the presence of dieldrin, while suppressing that of TNF $\alpha$  (Fig 4D–F). Therefore, we envisioned the implication of peptidylarginine deiminases (PADIs) that are highly dependent on intracellular Ca<sup>2+</sup> concentrations (Arita *et al*, 2006), and shown to positively regulate several HERVs (Sharma *et al*, 2012). The PADI inhibitor Cl-amidine caused a twofold reduction in the effect of dieldrin on HERVH env62, while not affecting expression of TNF $\alpha$  (Fig 4G). To further document an implication of PADI enzymes in the response to dieldrin, we used a previously described cell fractionation protocol (Reyes *et al*, 1997), involving isolation of a very soluble fraction, rich in RNAPII and poor in chromatin (as determined by histone H3 content—Fig 4H Fraction F1), a fraction rich in chromatin but containing no RNAPII (F3, regarded as containing inactive chromatin), and a pellet containing both chromatin and RNAPII (P, regarded as transcribed chromatin). We then examined in these fractions the distribution of PADI2, the only PADI enzyme detected in unstimulated and dieldrin-stimulated Jurkat cells. Western blots revealed an increased accumulation of this enzyme in transcribed chromatin in cells exposed to dieldrin (Fig 4H PADI2, compare lanes 4 and 8).

Altogether, these experiments suggest that the specificity of dieldrin resides in its ability to simultaneously activate several pathways, including MAPK, PI3K, and calcium-dependent pathways. Inhibition of any of the three pathways resulted in the loss of HERVH env62 transcription, suggesting that the cumulative effect of these pathways is required. Possibly, the atypical routes including PKC-uncoupled activation of ERK and a likely activation of PDK1 downstream of PI3K may also play a role. Importantly, our data also suggest that the effect of calcium on both the HERVs and the AP1 genes is mediated by PADI enzymes rather than calcineurin and NFAT. This is consistent with our earlier observations showing that citrullination of histone H3 at arginine 8 by these enzymes relieves repression by H3K9me3, a modification frequently detected at repressed HERVs and at inactive enhancers (Sharma *et al*, 2012; Zhu *et al*, 2012).

### Enhancer activity of HERV-containing regulatory regions in patients with MS

Our observations above indicated that HERV transcription in dieldrin-treated Jurkat cells is at least in part the consequence of the activation of enhancers harboring HERV sequences and participating in the immune response. To extend this observation to a disease associated with HERV transcription and chronic activation of the immune system, we examined existing RNA-seq data from patients with multiple sclerosis (Cao *et al*, 2015). In this report, transcriptomes of sorted myelin-reactive and non-myelin-reactive CCR6<sup>+</sup> CD4<sup>+</sup> memory T cells (respectively referred to as myelin<sup>+</sup> and myelin<sup>−</sup>) were compared between MS patients and healthy controls (Fig 5A). Manual examination of these data indicated that multiple HERV-derived sequences annotated as active enhancers in several hematopoietic tissues were transcribed abundantly only in the myelin<sup>+</sup> T cells from patients with MS (see example in Appendix Fig S5A). We next wished to quantify the phenomenon. As the data were not stranded, we relied on non-exonic annotated CAGE peaks to approach putative regulatory regions (CAGE regions defined as a 1-kb interval framing a CAGE peak—Table EV3). Overlap between RepeatMasker LTRs and CAGE regions upregulated ( $P < 0.05$ ) in patient with MS (as evaluated by the Jaccard index) was eightfold lower than that of randomly selected CAGE regions, indicating that patients with MS maintain a counter-selection of HERV sequences within regulatory regions activated by the disease. Yet, the LTR to CAGE similarity index increased at least threefold in the myelin<sup>+</sup> T cells, indicating an increased usage of LTRs as *cis*-regulatory elements in these cells (Fig 5B).



### Figure 5.HERV transcripts originate from activated enhancers

- A. Schematic of the experiment.
- B. Comparison of RepeatMasker regions annotated “LTR” or “SINE”, “LINE”, or “SVA” with upregulated CAGE regions ( $P < 0.05$ ) as defined in the text based on the Jaccard similarity index. The graph also provides the average and the standard deviation (error bars) of Jaccard indexes calculated for 1,000 series of randomly selected CAGE regions.
- C. Genes activated in myelin<sup>+</sup> patient T cells as compared to myelin<sup>+</sup> T cells from healthy controls and annotated in GeneHancer as interacting with a regulatory element containing an LTR sequence and activated twofold or more ( $P < 0.05$ ) in myelin<sup>+</sup> patient T cells as compared to myelin<sup>+</sup> T cells from healthy controls: Histogram shows the log<sub>2</sub> of fold change in read counts within the indicated genes. Error bar corresponds to the log<sub>2</sub> fold change standard error as reported by DESeq2.
- D–F. Screen captures from UCSC genome browser showing regulatory elements containing HERV-derived sequences with transcription correlating with that of a nearby gene annotated as linked in the GeneHancer database.

As a control, we also examined the similarity of the CAGE regions with SINEs and LINEs, two other families of repeats. SINEs were not counter-selected inside the regulatory regions activated in patients with MS, and there was no enrichment in SINE-containing regulatory regions in the myelin<sup>+</sup> T cells when compared to the myelin<sup>−</sup> cells. For the LINEs, we observed some counter-selection within regulatory regions activated in patients with MS (−58% for the myelin<sup>−</sup> cells) and there was a moderate (28%) increase in usage of these repeats in the myelin<sup>+</sup> cells. These observations suggest that MS has a greater effect on regulatory elements containing HERV sequences than on those containing other types of repeats.

To document that transcription of regulatory regions harboring HERV-derived sequences actually reflected enhancer activity relevant for MS, we next crossed the list of RepeatMasker LTRs with GeneHancer Regulatory Elements for which target genes have been identified by chromosome conformation capture techniques (Fishilevich *et al.*, 2017). This allowed us to identify a series of genes that (i) were significantly activated in myelin<sup>+</sup> cells from patients with MS as compared to myelin<sup>+</sup> cells from healthy controls and (ii) were linked to a HERV-containing enhancer showing at least twofold increased transcription ( $P < 0.05$ ) in patients with MS compared to healthy controls (Fig 5C and Table EV3). Among these, we particularly noted activated transcription of a HERV sequence upstream of CCL20, the gene encoding the ligand for CCR6. Seemingly, this HERV was the only enhancer-like activity in the neighborhood of CCL20, activated only in the myelin-reactive patient cells (Fig 5D). An example of an intragenic enhancer is shown in Fig 5E. Finally, we noted that the isoform of IL1R2 expressed exclusively in myelin-positive patient cells used a HERV sequence as a promoter (Fig 5F).

Altogether, these observations show that in patients with MS, activation of HERV-driven enhancers and promoters can be a source of HERV transcripts. Furthermore, the observation brings strong evidence for an implication of chronically active HERV-driven enhancers in the transcriptional regulation of disease-associated genes.

## Discussion

Our objective in this study was to investigate some of the transcriptional aspects behind the activity of endogenous retroviral sequences. Our initial hypothesis was that HERV-derived sequences located within or in the vicinity of promoters and enhancers might be transcribed as uaRNAs and eRNAs. Examination of the average position of enhancers and promoters relative to HERV-derived sequences revealed that the periphery of regulatory regions is in general depleted of retroviral sequences and located at the bottom of a HERV valley. We interpret this depletion as a safeguard against serendipitous transcription of HERVs when the



regulatory element becomes active. This stands out as a remarkable case of shaping of the genome sequence by constraints of transcription.

Our approach also revealed that, with the exception of some embryonic stem cell lines, all tissues counter-select HERV sequences inside their regulatory elements. In other words, any randomly selected sequence is more likely to be a promoter or an enhancer than is a HERV-derived sequence. When compared to a related analysis based on ENCODE DNase I accessibility data rather than chromatin states (Jacques *et al*, [2013](#)), our analysis confirms the cell type-specific activation of enhancers containing HERV sequences, but it also suggests a less widespread usage of HERV sequences inside regulatory elements. Chromatin states as calculated by the Epigenomic Roadmap consortium combine data on DNA methylation, histone modifications, and chromatin accessibility, and are therefore likely to provide more accurate boundaries of regulatory regions than would nuclease-accessibility assays. This may possibly explain some of the differences between our and earlier studies.

The counter-selection of HERV sequences inside regulatory sequences is strongest in pancreatic islets, in several brain tissues, and in the gastrointestinal tract. Likewise, an earlier study has reported that transposons are not an abundant source of enhancers in the mammalian neocortex (Emera *et al*, [2016](#)). We speculate that keeping HERV sequences silent in brain may be of particular importance because of the several neurological diseases associated with HERV expression in that tissue, including multiple sclerosis, amyotrophic lateral sclerosis, and chronic inflammatory demyelinating polyradiculoneuropathy (Küry *et al*, [2018](#)). We note also that the low usage of HERVs in pancreatic islets is yet another similarity between this tissue and brain tissues, which in numerous studies have been reported to share transcriptional programs (Arntfield & van der Kooy, [2011](#)).

At the other end of the spectrum, stem cells evade the negative selection of HERVs inside their regulatory regions. This is compatible with studies reporting the influence of endogenous retrovirus-driven transcription on the maintenance of pluripotency (Wang *et al*, [2014](#); Hackett *et al*, [2017](#)). In that context, we note that GO term analysis of genes located in the vicinity of HERV-based regulatory elements suggests a role in cellular defense. Possibly, defense genes may be more exposed to viral integration because they are active during acute phases of infection. We note also that several HERV-derived sequences located in the neighborhood of cytokine genes were annotated as enhancers only in embryonic stem cells ([Appendix Figs S1M and S3A](#)). This may suggest that innate immunity of embryonic stem cells, which by many aspects is different from that of differentiated cells (D'Angelo *et al*, [2017](#)), relies more than other tissues on HERV-driven transcription.

While T cells score relatively low in the usage of HERV sequences inside enhancers, they seem to be amenable to a limited level of HERV transcription. In the present study, we reached this effect by treating the cells with the pesticide dieldrin that induces a strong and, by several aspects, atypical T-cell activation. In particular, the effect of dieldrin stands out from classical T-cell activation by the amplitude of the activation and the number of genes that this pesticide induces. The MAPK, PI3K, and calcium-dependent pathways that get induced are standard for T-cell activation, but there seem to be divergences in the routes taken by these pathways in the presence of dieldrin. For example, we observe that activation of ERK in the MAPK pathway does not occur through activation of PKC and that NF- $\kappa$ B does not seem to be a major player downstream of the pathway.

Monitoring the activity of HERV Env62 showed that inhibition of just one of three induced pathways interfered with expression of the viral sequence in the presence of dieldrin. Dieldrin also caused a re-localization of PADI2 to chromatin where we have previously shown that peptidyl arginine deiminase activity interferes with H3 lysine 9 trimethyl (H3K9me3)-mediated silencing and thereby favors transcription of HERV-derived sequences (Sharma *et al*, [2012](#)). The H3K9me3 histone mark is also involved in regulating enhancer activity (Zhu *et al*, [2012](#)). It is therefore possible that PADI activity, beyond its effect on HERV transcription, may allow unlocking of enhancers from earlier stages of cell differentiation and thereby increase the range of regulatory regions available for activation of defense genes in cells exposed to stress. This scenario is compatible with dieldrin-treated Jurkat cells relying on more embryonic and HERV-enriched enhancers to confront the stress engendered by the pesticide. An activation of dormant enhancers together with induction of multiple signal transduction pathways is also consistent with earlier studies indicating that transcription of HERV-derived sequences is the consequence of both de-repression and activation (Gerdes *et al*, [2016](#)), possibly with the MAPK and the PI3K pathways providing the transcription factors for the activation of enhancers de-repressed by PADI or other H3K9me3-neutralizing activities.

In the Jurkat T cells exposed to dieldrin, the time window for transcription of the HERV sequences we monitored by RT-qPCR was very short and, in the RNA-seq data, transcription of HERV-derived sequences was difficult to correlate with activation of nearby genes. Possibly, the duration of the exposure to dieldrin (30 min) may have been too short to allow detectable activation of the target genes of these putative enhancers. Alternatively, transcription of HERV-driven enhancers may play a more general role in early phases of chromatin reprogramming and nuclear reorganization associated with T-cell activation (Robson *et al*, [2017](#)).

While the kinetic of the HERV transcription in the dieldrin-treated Jurkat cells is likely to reflect the behavior of these sequences in healthy T cells, the data from patients with MS provided us with an opportunity to examine the viral sequences in the context of a chronic disease-associated condition. Re-analysis of this high-quality data showed that MS does not result in the loss of the counter-selection of enhancers containing HERV-derived sequences. Yet, in the myelin-autoreactive CD4<sup>+</sup> T cells, this counter-selection is reduced. In these cells, we could also observe an increased activity of a series of enhancers containing HERVs, many of which were located in the neighborhood of activated immune genes. This is a strong argument in favor of the contribution of HERV-driven transcriptional activity in the immune mechanisms of MS pathogenesis. It also defines enhancer activity as a very likely source of HERV transcripts in patients with MS.

## Materials and Methods

### Antibodies and chemicals

Antibodies used for Western blot: anti-Phospho-PKC (Cell Signaling 9371T), anti-REL A (Cell Signaling 3033T), anti-NFAT-1 (BD Transduction Laboratories 610702), anti-BRG1 (Euromedex 2E12), anti-RNA Pol 2 pS2 (Abcam 5095), anti-H3 (Abcam 1791), anti-PADI2 (Abcam 56928), and anti-BRG1 (Euromedex 2E12). Each antibody was profiled by the vendor as appropriate for Western blots. Chemicals used in this study: Dieldrin and PMA were purchased from SIGMA (Ref 291218, and P1585, respectively). The inhibitors used for the pathways determination: wortmannin (Sigma W3144), BX795 (Sigma SML0694), Akt1/2

kinase inhibitor (Sigma A6730), H89 (Sigma B1427), U0126 (EMD Millipore 19147), FK506 (Enzo Life Sciences ALX 380 008), and PD98059 (Selleckchem S1177). Cl-amidine was purchased from EMD Millipore (reference 506282).

### Cell culture

Jurkat E6-1 cells purchased from ATCC were cultured in RPMI-1640 with 10% decomplemented fetal bovine serum (FBS) and 100 U ml<sup>-1</sup> penicillin–streptomycin at 37°C in a 5% CO<sub>2</sub> incubator. Cells were treated with phorbol myristate acetate (PMA) or dieldrin at 100 µM for the indicated times, with wortmannin (10 mM), BX795 (6 µM), Akt1/2 kinase inhibitor (40 µM), H89 (40 µM), U0126 (12 µM), FK506 (100 nM), and PD98059 (25 µM) for 1 h and with 200 µM Cl-amidine for exactly 16 h in complete cell culture medium.

### mRNA, protein quantification, and cell fractionation

Total RNA from Jurkat cells was extracted with a phenol–chloroform-based method and quantified with an ND-1000 (Nanodrop). After DNase treatment (Roche), reverse transcription was performed using SuperScript III (Invitrogen) and random hexanucleotides according to the manufacturer's instructions. Complementary DNAs were quantified by RT–qPCR (Mx3005P, Stratagene) using SYBR Green PCR master mix (Applied Biosystems). PCR primers are listed in [Appendix Table S1](#). Data are shown relative to RPLP0 considered invariant. Total proteins were separated by electrophoresis on 4–12% gradient PAGE gels (Bio-Rad, #345–0124) and transferred on nitrocellulose membrane (Bio-Rad, #1620115) for Western blot with indicated antibodies. Cell fractionation was carried out as previously described (Reyes *et al*, [1997](#)).

### High-throughput RNA sequencing and bioinformatics

Stranded libraries of cDNA were prepared by random priming followed by either enrichment in poly(dA) with an oligo(dT) resin or depletion from ribosomal RNA. A minimum of 30-Gb sequencing was obtained for each sample with reads of 150 bases. All poly(dA)-enriched libraries were sequenced on a same flow cell. Reads were aligned with STAR 2.5.0a (Dobin *et al*, [2013](#)) on human genome version Hg19 primary assembly without patches, allowing only single alignments. Differential expression analysis on poly(dA)-enriched libraries was performed with DESeq2 1.14.1 (Love *et al*, [2014](#)), and raw counts were corrected with a TMM normalization. Divergent transcription sites were extracted from the previous alignments by identifying region in between groups of reads mapping on opposite strands separated by < 500 bp, developing a python module (available at [https://github.com/khourhin/divtrans\\_azebi](https://github.com/khourhin/divtrans_azebi) [2019](#)) wrapping bedtools 2.27.1 and samtools 1.9. Reads inside HERV sequences and chromatin states were quantified with featureCount 1.6.0 from the Subread package (Liao *et al*, [2014](#)). Identification of genes in the neighborhood of HERVs was carried out with GREAT (McLean *et al*, [2010](#)). Jaccard indexes were calculated with bedtools 2.27.1, while profiles were generated with deepTools3.0.1. Panels with chromatin states were screen captures from NIH Roadmap Epigenomics Mapping Consortium web site ([https://egg2.wustl.edu/roadmap/web\\_portal/chr\\_state\\_learning.html#core\\_15state](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state)).

## Statistics

GO term analysis with GREAT shows  $-\log_{10}$  of the binomial  $P$ -value. The binomial test is potentially biased by a large number of genomic regions being associated with a small set of genes. We therefore ensured that for each indicated pathway, the hypergeometric test also indicated significant enrichment. In the transcriptomic analysis, significance was estimated based on the  $P$ -value adjusted according to the Benjamini and Hochberg correction. Two-sided Student's  $t$ -test was used when sample numbers were small.

## Human subjects

The CD4<sup>+</sup> T-cell data from patient with MS have been deposited under the BioProject accession no. [GSE66763](#). The original publication stated that peripheral blood was drawn from healthy individuals and patients with MS after informed consent and approval by the Institutional Review Board at Yale University School of Medicine. All patients had relapsing-remitting MS and were aged  $42 \pm 12$  years. All patients were immunotherapeutic-naïve (untreated) or had not received treatment for at least 12 months. Healthy donors were age- and sex-matched with the MS patients, with no history of autoimmune disease or malignancies and no acute or chronic infections (Cao *et al*, [2015](#)).

## Data availability

All RNA-seq data were deposited on GEO with the record number [GSE111167](#).

## Acknowledgements

We thank S. El Messaoudi-Aubert for valuable discussion and J.S. Seeler for critical reading of the manuscript. The work was supported by grants from Institut Pasteur, “La Fondation pour la Recherche Médicale”, LABEX REVIVE, and the Global Care initiative of the Institut Carnot.

## Author contributions

SA performed most experiments. EB examined HERV copy numbers and contributed qPCR analysis tools. FM contributed with reagents and provided expertise in immunology. EK carried out bioinformatic analysis of the RNA-seq on poly(A)-enriched libraries and coded the Python module for detection of sites of divergent transcription. CM designed most of the experiments, carried out the bioinformatic analysis on regulatory elements, and wrote the manuscript with input from all authors.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Supporting Information

- [Appendix](#) (PDF document, 8.7 MB)



- [Table EV1](#) (MS Excel, 529.2 KB)
- [Table EV2](#) (MS Excel, 31.1 KB)
- [Table EV3](#) (MS Excel, 18.5 KB)

## References

- Arita K, Shimizu T, Hashimoto H, Hidaka Y, Yamada M, Sato M (2006) Structural basis for histone N-terminal recognition by human peptidylarginine deiminase 4. *Proc Natl Acad Sci USA* **103**: 5291–5296 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Arntfield ME, van der Kooy D (2011)  $\beta$ -Cell evolution: how the pancreas borrowed from the brain. *BioEssays* **33**: 582–587 [Wiley Online Library](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Bannert N, Kurth R (2006) The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet* **7**: 149–173 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Calo E, Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825–837 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Cao Y, Goods BA, Raddassi K, Nepom GT, Kwok WW, Love JC, Hafler DA (2015) Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. *Sci Transl Med* **7**: 287ra74 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Christensen T (2016) Human endogenous retroviruses in neurologic disease. *APMIS* **124**: 116–126 [Wiley Online Library](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Chuong EB, Rumi MAK, Soares MJ, Baker JC (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45**: 325–329 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Chuong EB, Elde NC, Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**: 1083–1087 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- D'Angelo W, Gurung C, Acharya D, Chen B, Ortolano N, Gama V, Bai F, Guo Y-L (2017) The molecular basis for the lack of inflammatory responses in mouse embryonic stem cells and their differentiated cells. *J Immunol* **198**: 2147–2155 [Crossref](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- DGT TFCATRPAC (2014) A promoter-level mammalian expression atlas. *Nature* **507**: 462–470 [Crossref](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Emera D, Yin J, Reilly SK, Gockley J, Noonan JP (2016) Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc Natl Acad Sci USA* **113**: E2617–E2626 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M *et al* (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**: 1217 [Crossref](#)[Google Scholar](#)
- Gerdes P, Richardson SR, Mager DL, Faulkner GJ (2016) Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol* **17**: 100 [Crossref](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Grandi N, Tramontano E (2018) Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front Immunol* **9**: 2039 [Crossref](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)

- Hackett JA, Kobayashi T, Dietmann S, Surani MA (2017) Activation of lineage regulators and transposable elements across a pluripotent spectrum. *Stem Cell Reports* **8**: 1645–1658 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Hah N, Murakami S, Nagari A, Danko CG, Kraus WL (2013) Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* **23**: 1210–1223 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Hummel B, Hansen EC, Yoveva A, Aprile-Garcia F, Hussong R, Sawarkar R (2017) The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. *Nat Struct Mol Biol* **24**: 234–242 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Imbeault M, Helleboid P-Y, Trono D (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**: 550–554 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Jacques P-É, Jeyakani J, Bourque G (2013) The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9**: e1003504–e1003512 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M *et al* (2011) DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**: 676–687 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S *et al* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Kürty P, Nath A, Créange A, Dolei A, Marche P, Gold J, Giovannoni G, Hartung H-P, Perron H (2018) Human endogenous retroviruses in neurological diseases. *Trends Mol Med* **24**: 379–394 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550. [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Magiorkinis G, Blanco-Melo D, Belshaw R (2015) The decline of human endogenous retroviruses: extinction and survival. *Retrovirology* **12**: 8 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Mahat DB, Salamanca HH, Duarte FM, Danko CG, Lis JT (2016) Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Mol Cell* **62**: 63–78 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Manning BD, Toker A (2017) AKT/PKB signaling: navigating the network. *Cell* **169**: 381–405 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nature Publishing Group* **28**: 495–501 [CAS](#)[Google Scholar](#)
- Melgar MF, Collins FS, Sethupathy P (2011) Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol* **12**: R113 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Nexø BA, Villesen P, Nissen KK, Lindegaard HM, Rossing P, Petersen T, Tarnow L, Hansen B, Lorenzen T, Hørslev-Petersen K *et al* (2016) Are human endogenous retroviruses triggers of autoimmune diseases? Unveiling associations of three diseases and viral loci. *Immunol Res* **64**: 55–63 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)

- Reyes JC, Muchardt C, Yaniv M (1997) Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. *J Cell Biol* **137**: 263–274 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J *et al* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Robson MI, de Las Heras JI, Czapiewski R, Sivakumar A, Kerr ARW, Schirmer EC (2017) Constrained release of lamina-associated enhancers and genes from the nuclear envelope during T-cell activation facilitates their association in chromosome compartments. *Genome Res* **27**: 1126–1138 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Sharma P, Azebi S, England P, Christensen T, Moller-Larsen A, Petersen T, Batsché E, Muchardt C (2012) Citrullination of histone H3 interferes with HP1-mediated transcriptional repression. *PLoS Genet* **8**: e1002934 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Song C, Kanthasamy A, Anantharam V, Sun F, Kanthasamy AG (2010) Environmental neurotoxic pesticide increases histone acetylation to promote apoptosis in dopaminergic neuronal cells: relevance to epigenetic mechanisms of neurodegeneration. *Mol Pharmacol* **77**: 621–632 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* **24**: 1963–1976 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD (2017) Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res* **27**: 1623–1633 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Trizzino M, Kapusta A, Brown CD (2018) Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genom* **19**: 468 [Crossref](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, Cai H, Besser D, Prigione A, Fuchs NV *et al* (2014) Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**: 405–409 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL *et al* (2013) ng.2649(1). *Nat Genet* **45**: 836–841 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)
- Zhu Y, van Essen D, Sacconi S (2012) Cell-type-specific control of enhancer activity by H3K9 trimethylation. *Mol Cell* **46**: 408–423 [Crossref](#) [CAS](#) [PubMed](#) [Web of Science](#)®[Google Scholar](#)

## APPENDIX

### TABLE OF CONTENT

**Figure S1:** Genome-wide localization of HERVs relative to promoters and enhancers

**Figure S2:** Dieldrin is a strong T cell activator causing transcription of several HERVs

**Figure S3:** Dieldrin causes transcription of HERV-containing PEs atypical for activated T cells

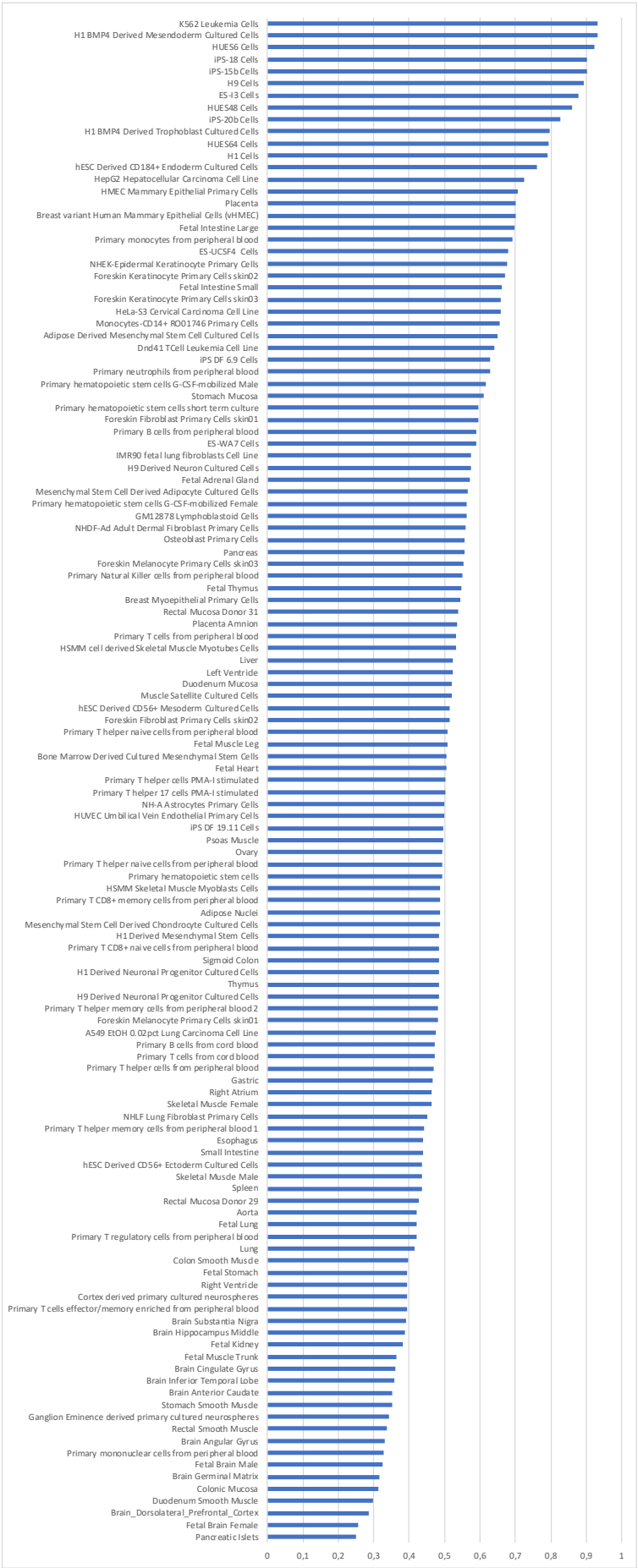
**Figure S4:** Dieldrin activates multiple signal transduction pathways.

**Figure S5:** HERV transcripts originate from activated enhancers.

**Table S1:** List of PCR primers used in this study

A

**Jaccard index of similarity between enhancers/promoters and HERV-derived sequences in the indicated tissues  
(normalized to the Jaccard index between enhancers/promoters and a random set of intervals not overlapping with HERVs)**

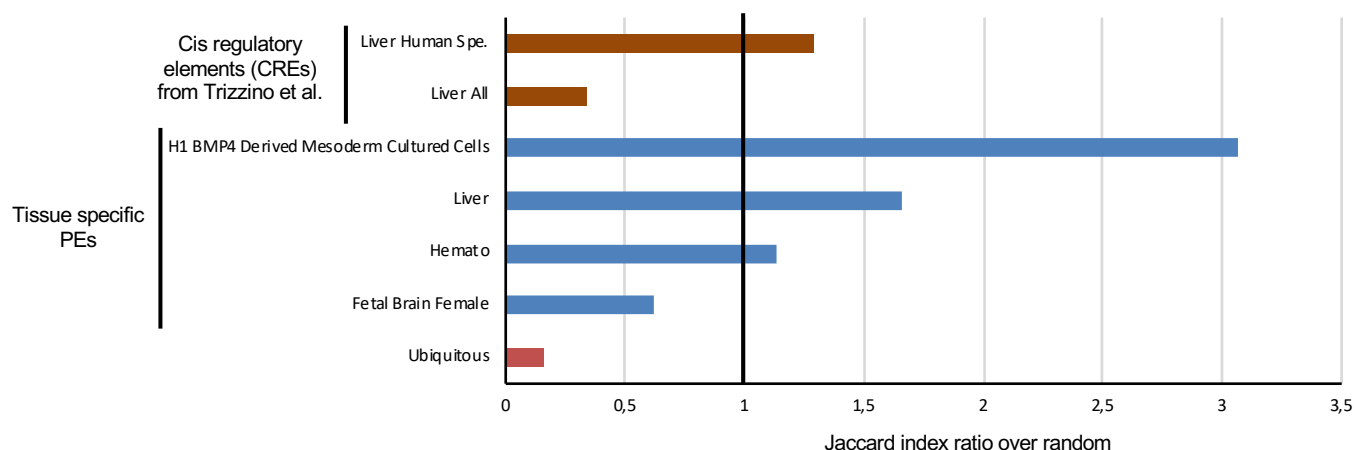


Jaccard index ratio over random

Figure S1

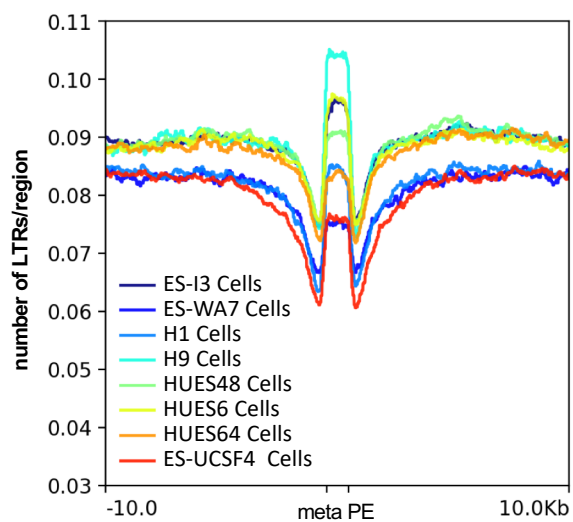
**B**

### Comparison of LTRs with either ubiquitous or tissue specific PEs



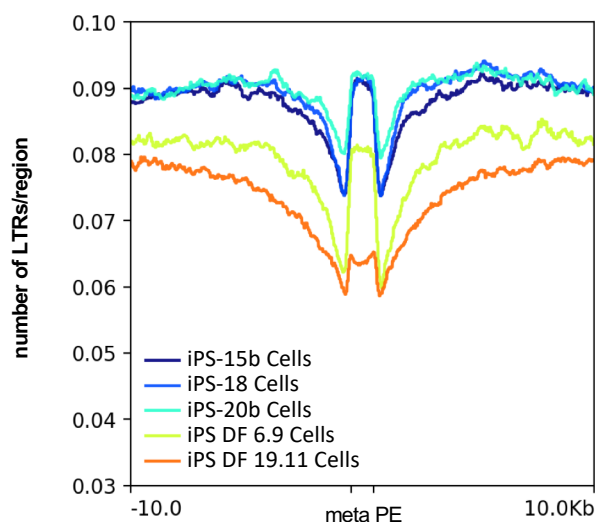
**C**

### Embryonic stem cells



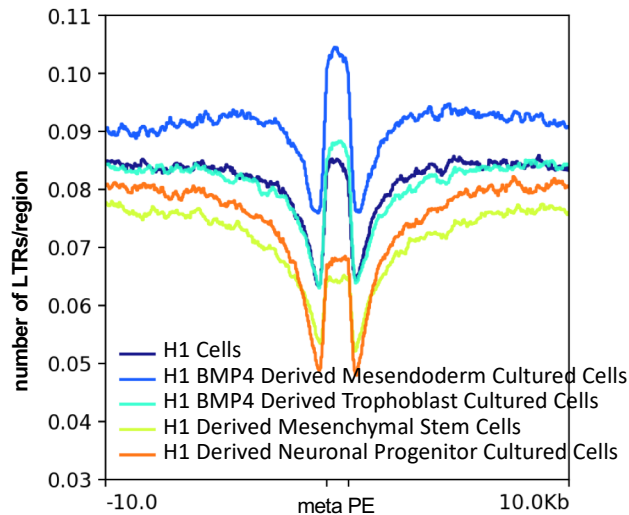
**D**

### iPS cells



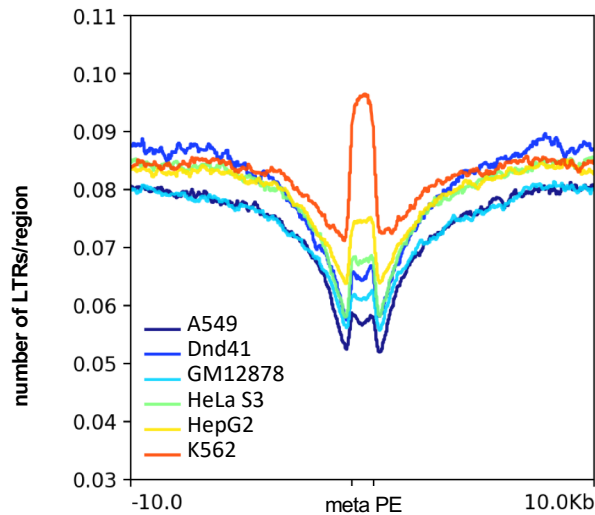
**E**

### H1-derived cells



**F**

### Transformed cell lines



**Figure S1**

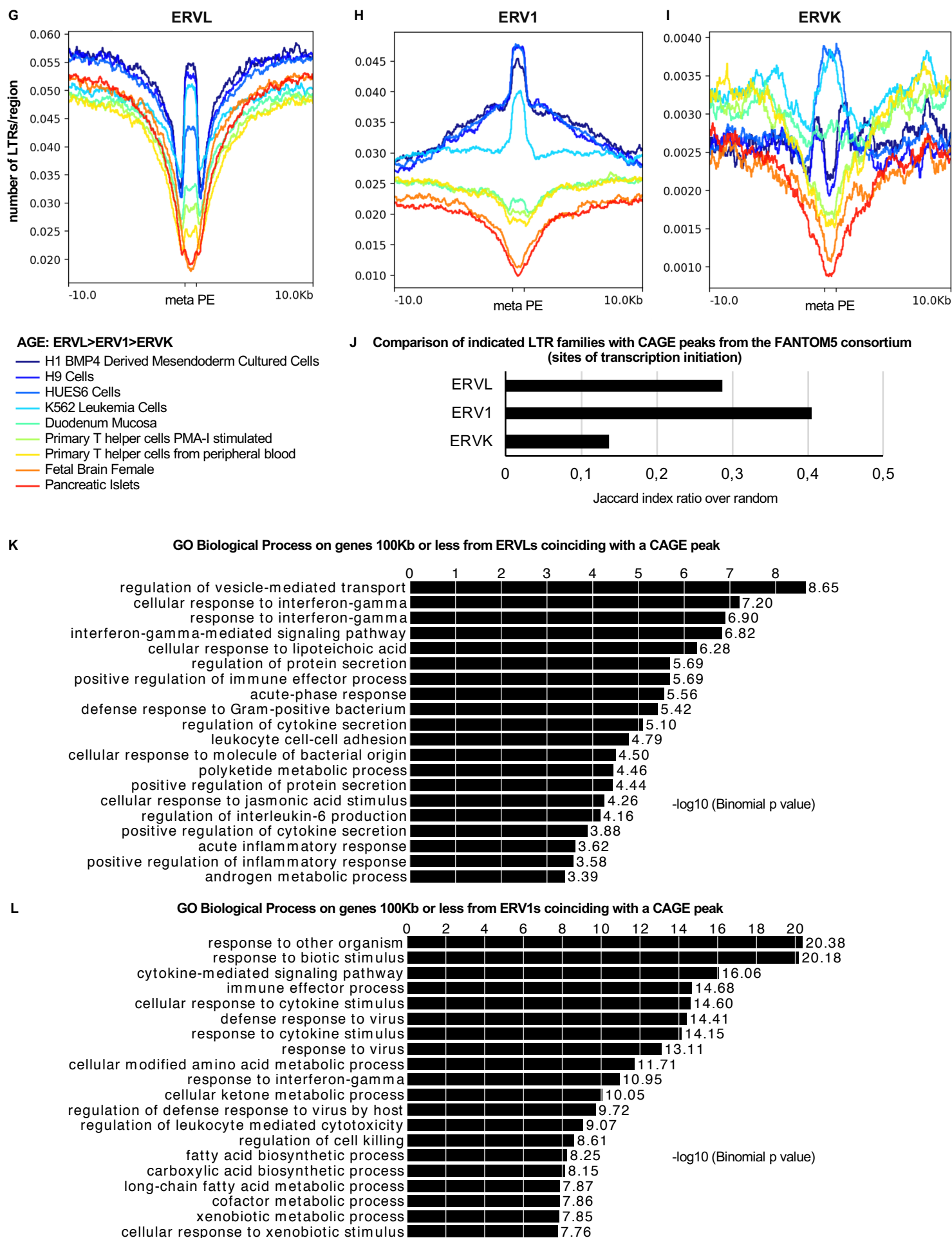


Figure S1



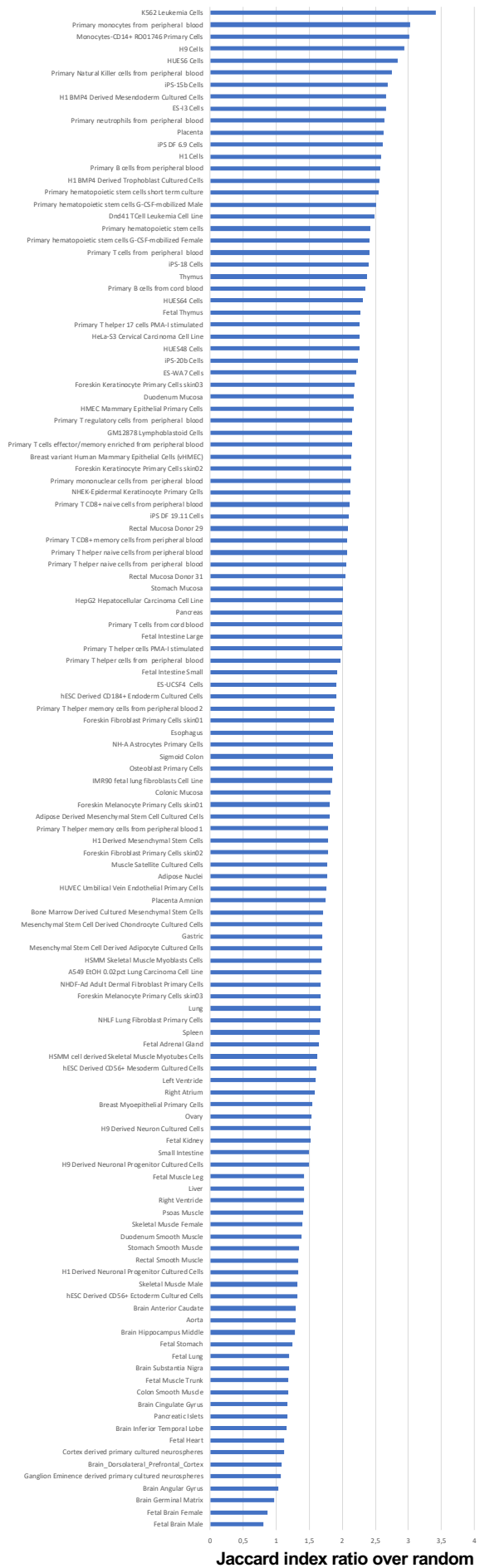
M



Figure S1



**N** Jaccard index of similarity between localization of enhancers/promoters and HERV-derived sequences overlapping with CAGE peaks in the indicated tissue (normalized to the Jaccard index between enhancers/promoters and a random set of intervals not overlapping with HERVs)



**Figure S1**

**Appendix Figure S1:** Genome-wide localization of HERVs relative to promoters and enhancers

- (A) Comparison of regions annotated “LTR” in RepeatMasker with the regions annotated TssA, TssAFlnk, Enh, EnhG, TssBiv, or EnhBiv in the 15 core marks model of the Epigenomic Roadmap consortium (referred to as PEs for Promoters and Enhancers). For each tissue, the Jaccard index comparing PEs to LTRs is divided by the average Jaccard index (1000 iterations) obtained when comparing PEs to randomly selected genomic locations (of the same sizes as the LTRs).
- (B) Regions annotated “LTR” in RepeatMasker compared with the indicated regions. For each set of regions, the Jaccard index comparing PEs or CREs to LTRs is divided by the average Jaccard index (1000 iterations) obtained when comparing the PEs or CREs to randomly selected genomic locations (of the same sizes as the LTRs). Tissue-specific PEs were selected as PEs present in the indicated tissue and not overlapping with PEs from other tissues. Ubiquitous PEs were selected as PEs overlapping with a PE in all tissues. “H1 BMP4 Derived Mesoderm Cultured Cells” and “Fetal Brain Female” provide min. and max. similarity of tissue-specific PEs with LTRs.
- (C-F) Profile reporting the position of regions annotated “LTR” in RepeatMasker relative to PEs from the indicated tissues. Each panel reports profiles from related tissues as indicated.
- (G-I) Profile reporting the position of regions annotated “ERV1”, “ERV2”, or “ERV3” in RepeatMasker relative to PEs from the indicated tissues.
- (J) Comparison regions annotated “ERV1”, “ERV2”, or “ERV3” in RepeatMasker with CAGE peaks from the FANTOM5 consortium. For each HERV clade, the Jaccard index comparing LTRs to CAGE peaks is divided by the average Jaccard index (1000 iterations) obtained when comparing CAGE peaks to randomly selected genomic locations (of the same sizes as each of the HERV clades).
- (K-L) Biological process GO terms associated with genes located 100 kb or less from a region annotated “ERV1” or “ERV2” in RepeatMasker and overlapping with a CAGE peak.
- (M) Position of the LTR upstream of IL6 relative to functional annotations in the 127 tissues of the Epigenomic Roadmap consortium. Color code: red, active TSS, orange red, flanking active TSS, lime green, transcription at gene 5' and 3', green, strong transcription, dark green, weak transcription, green yellow, genic enhancers, yellow, enhancers, medium aquamarine, ZNF genes & repeats, pale turquoise, heterochromatin, Indian red, bivalent/poised TSS, darksalmon, flanking bivalent TSS/Enh, dark khaki, bivalent enhancer, silver, repressed polycomb, Gainesboro, weak repressed polycomb, white, quiescent/low.
- (N) For each of the 127 tissues mapped by Epigenomic Roadmap consortium, regions annotated “LTR” in RepeatMasker and overlapping with a CAGE peak from the FANTOM5 consortium data (CAGE-LTRs) are compared with PEs. For each tissue, the Jaccard index comparing PEs to CAGE-LTRs is divided by the average Jaccard index (1000 iterations) obtained when comparing PEs to randomly selected genomic locations (of the same sizes as the CAGE-LTRs).

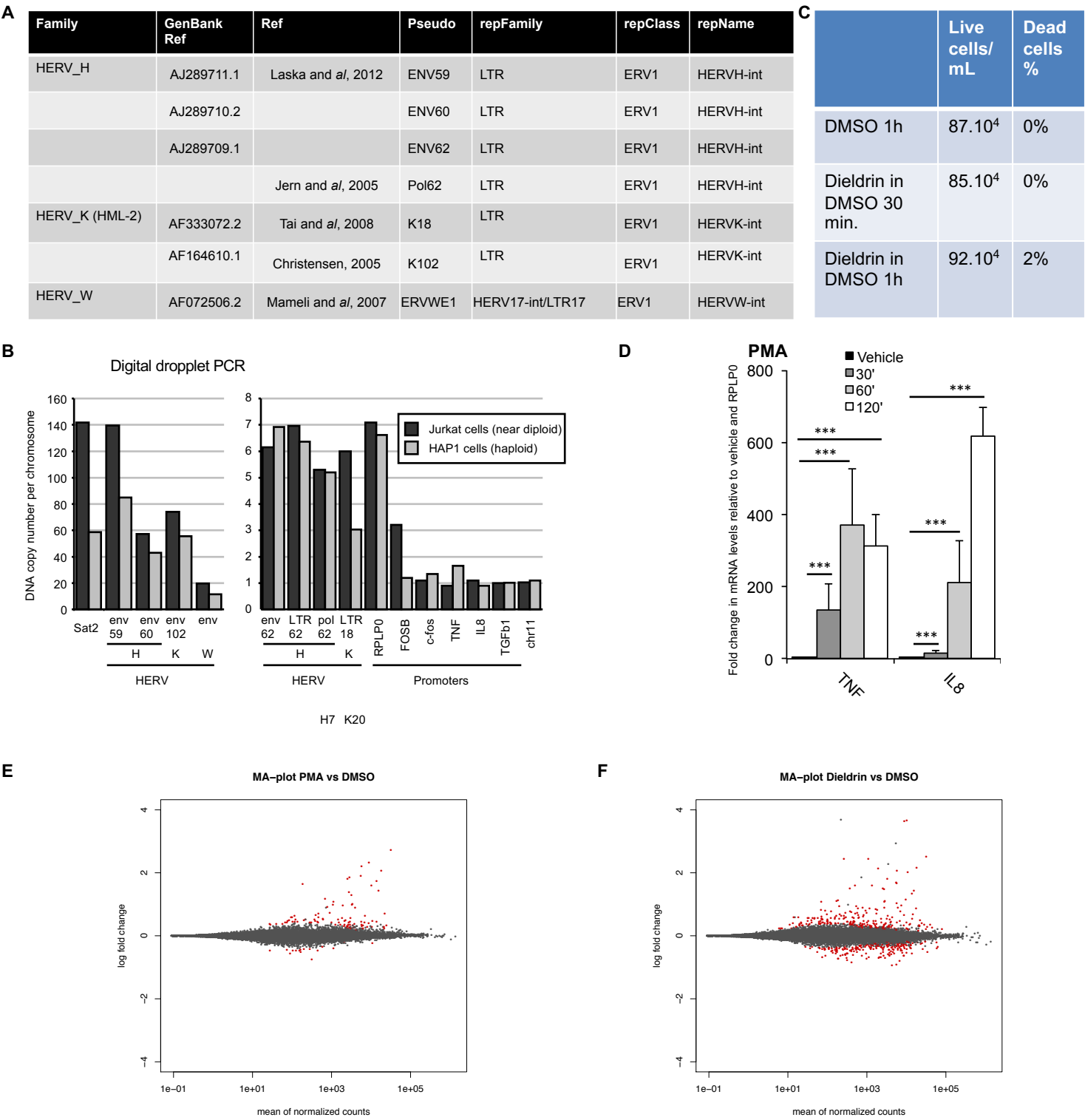
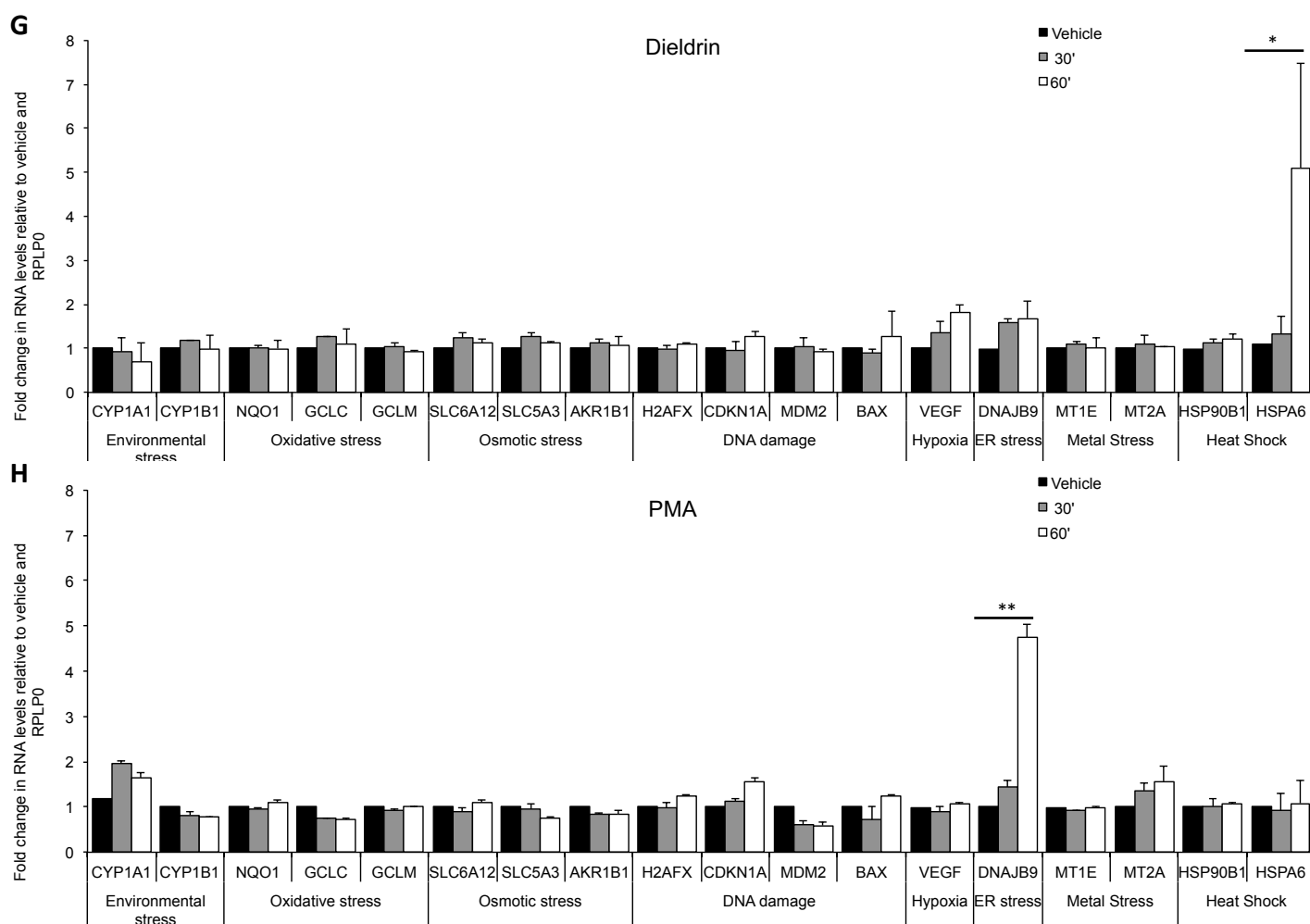


Figure S2



**Appendix Figure S2: Dieldrin is a strong T cell activator causing transcription of several HERVs**

(A) Description and references linking HERVs to MS.

(B) To estimate the number of loci amplified by the primers, digital droplet PCR was carried out on Jurkat cell DNA, using DNA from a haploid cell line as a reference.

(C) Impact of dieldrin on cell viability at the indicated times.

(D) Jurkat T cells were treated with either DMSO (vehicle) or with 100 $\mu$ M of either Dieldrin or PMA for the indicated times. Abundance of mRNA for the indicated genes was assessed by RT-qPCR. Data shown are means  $\pm$  SEM from three independent experiments. Significance (p-value) was estimated using the two-sided student t test.

(E, F) MA-plot recapitulating changes in gene expression in the presence of PMA or dieldrin as indicated. Every dot represents a gene. Red color signals significant change (FDR<0.05).

(G, H) Jurkat cells were treated for indicated times with either DMSO (vehicle) or 100 $\mu$ M of either dieldrin or PMA. Abundance of mRNA for the indicated genes was assessed by RT-qPCR. Data shown are means  $\pm$  SEM from three independent experiments. Significance (p-value) was estimated using the two-sided student t test.

\*\*\*: p-value<0.001, \*\*: 0.001<p-value<0.01, \*: 0.01<p-value<0.05.

A



Figure S3

B



Figure S3

**Figure S3**

**Appendix Figure S3:** Dieldrin causes transcription of HERV-containing PEs atypical for activated T cells

(A-C) Position of LTRs described in the result section relative to functional annotations in the 127 tissues of the Epigenomic Roadmap consortium. Color code as in Appendix Figure S1. Arrow heads indicate remarkable PEs (read arrow indicates PEs active in embryonic or cancer cells; black arrows indicate PEs active in T cells).



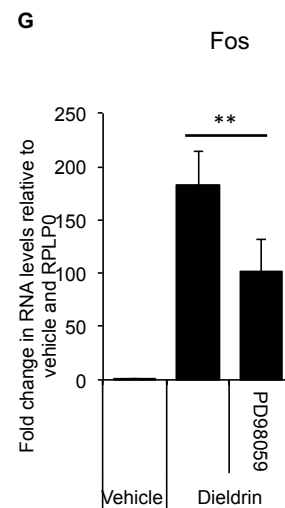
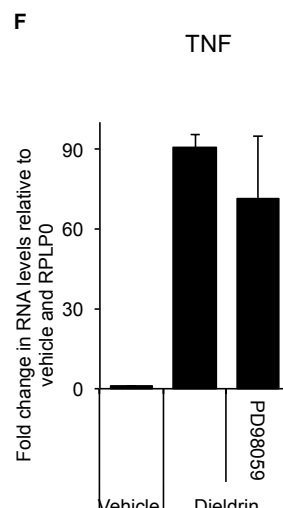
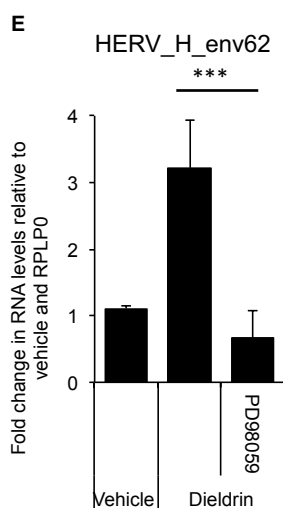
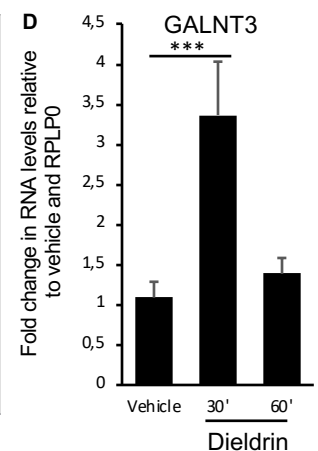
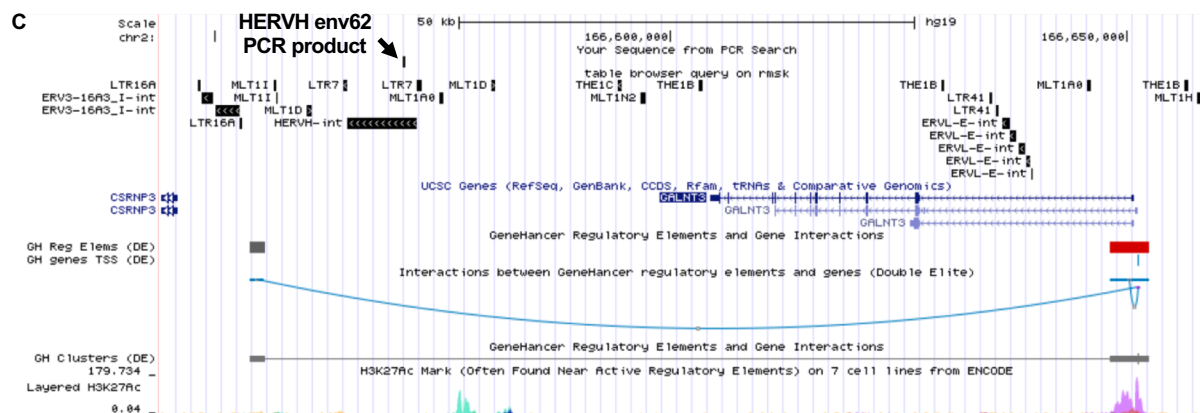
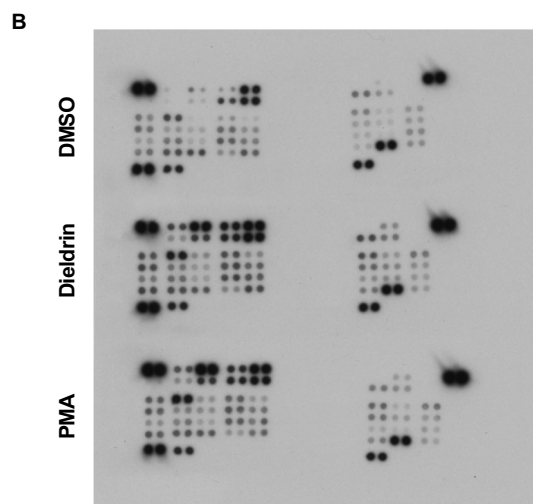
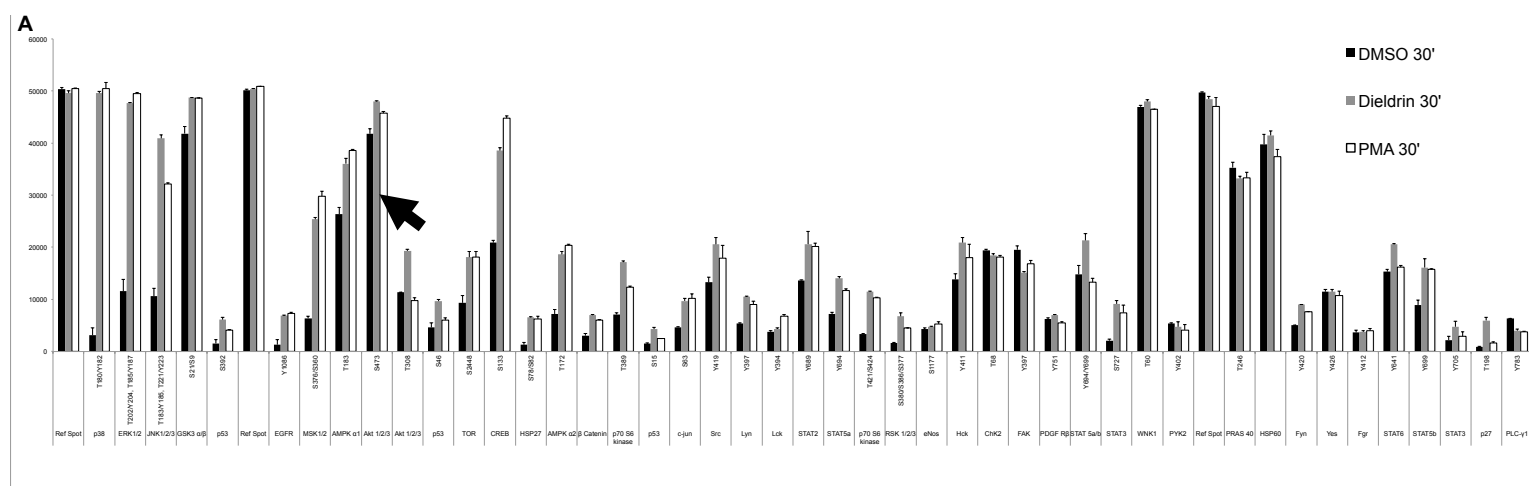
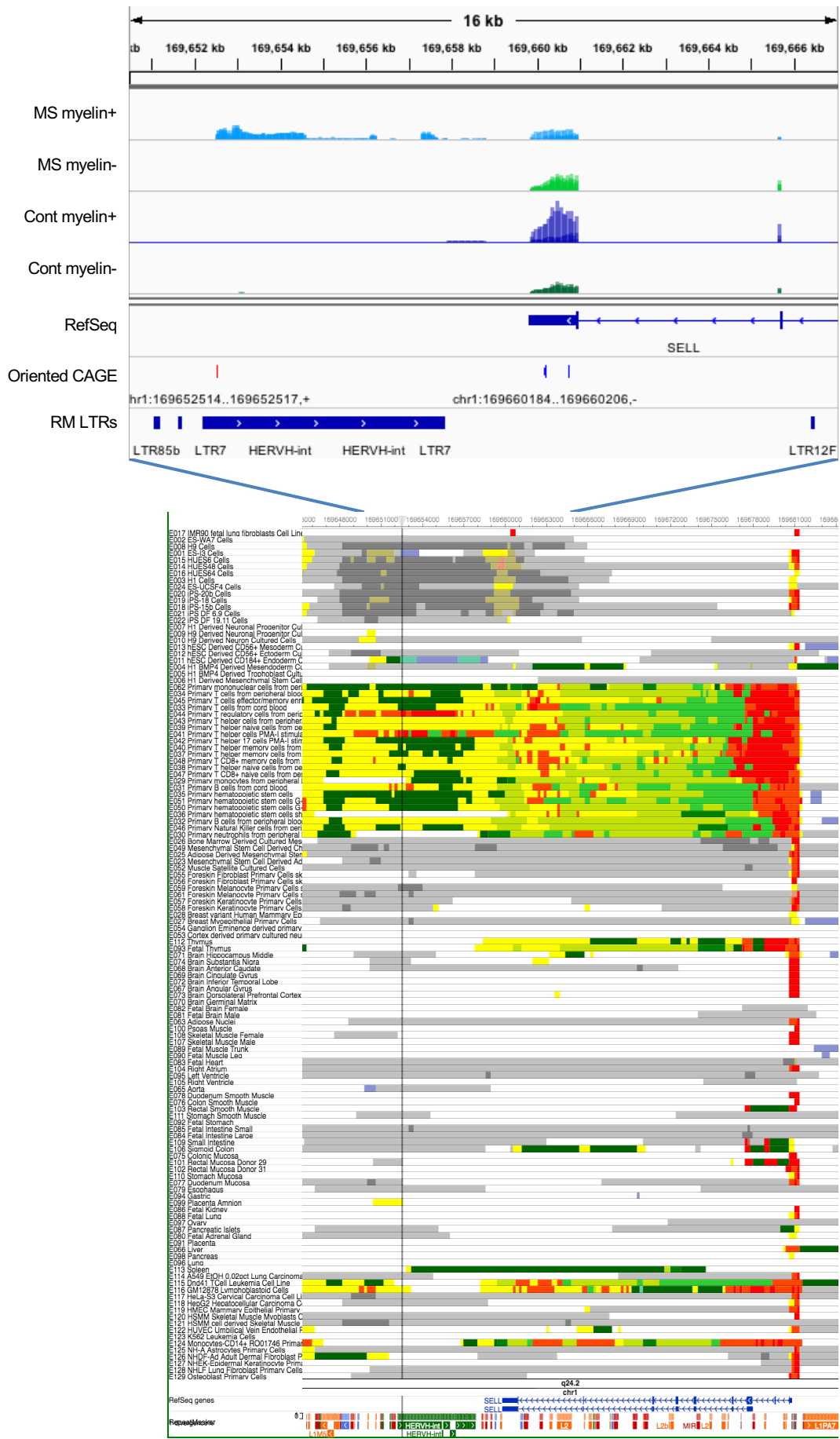


Figure S4

**Appendix Figure S4:** Dieldrin activates multiple signal transduction pathways.

- (A) A human phospho-kinase antibody array was probed with extracts from Jurkat cells treated with either DMSO, dieldrin, or PMA for 30 min as indicated. On the array, each antibody is spotted twice. The array was processed as a western and revealed by ECL. Spots were quantified using Fiji; Values shown are means  $\pm$  SEM from measures of the two spots for each antibody. Arrow points at AKT S473 phosphorylation.
- (B) Spots in the array quantified in A are distributed as indicated by the provider of the array (<https://resources.rndsystems.com/pdfs/datasheets/ary003b.pdf>).
- (C) Screen capture from the UCSC genome browser locating the HERVH env62 PCR product downstream of the GALNT3 gene. The link annotated in the GeneHancer data and connecting the regulatory region to the promoter of the GALNT3 gene is also indicated.
- (D) Jurkat cells were exposed to 100 $\mu$ M dieldrin for the indicated times. Abundance of GALNT3 mRNA was assessed by RT-qPCR. Data shown are means  $\pm$  SEM from three independent experiments. Significance (p-value) was estimated using the two-sided student t test.
- (E-G) Jurkat cells were cultured in the absence or in the presence of the MEK inhibitor PD98059 for 1h then exposed to 100 $\mu$ M dieldrin for 30 min. Abundance of mRNA for the indicated genes or HERVs was assessed by RT-qPCR. Data shown are means  $\pm$  SEM from three independent experiments. Significance (p-value) was estimated using the two-sided student t test.
- \*\*\*: p-value<0.001, \*\*: 0.001<p-value<0.01.

**A**



**Figure S5**

B

Target gene	log2FoldChange	pvalue	GeneHancer item	Fold Change T test	Overlapping LTRs
FASLG	10,41	3,5756E-21	<b>GH01J172637</b>	6543,00	0,044 MLT1B
CCL20	3,87	3,231E-15	<b>GH02J227805</b>	135,95	0,004 LTR16A
ZNF267	5,93	2,8757E-11	<b>GH16J031879</b>	379,20	0,019 MER21B,MLT1H1
AIM2	6,60	4,7344E-06	<b>GH01J159067</b>	55,56	0,017 ERV3-16A3_LTR,LTR81C,MLT1A0
KLHL6	4,08	6,9746E-06	<b>GH03J183552</b>	35,73	0,045 LTR85c,MER21A,MER51B,MLT1L,PABL_A
PRNP	5,08	9,9981E-06	<b>GH20J004664</b>	230,00	0,028 LTR70
CD44	3,02	1,4447E-05	<b>GH11J035076</b>	982,00	0,028 MLT1H2,MSTB
TRAF1	3,11	1,5481E-05	<b>GH09J120916</b>	74,75	0,058 LTR16E1,LTR2C
RNF19A	3,42	4,5298E-05	<b>GH08J100296</b>	18,40	0,019 ERVL-E-int
LRR8C	4,16	7,387E-05	<b>GH01J089755</b>	577,00	0,016 LTR1B,LTR1C,LTR8,MLT1A
CD96	2,51	0,00124374	<b>GH03J111607</b>	34,59	0,027 MLT1B
CMAHP	3,45	0,01209303	<b>GH06J025130</b>	12,08	0,019 MLT1I
NABP1	2,55	0,01707621	<b>GH02J191695</b>	10,83	0,047 MLT1F1
NEDD9	2,12	0,02569855	<b>GH06J011349</b>	45,53	0,034 MLT2B5
SYS1	1,85	0,03610297	<b>GH20J044965</b>	3,26	0,009 MER39
JARID2	2,13	0,03816523	<b>GH06J015322</b>	2902,00	0,021 MLT1B
DUSP28	2,68	0,04287648	<b>GH02J240564</b>	55,35	0,007 MER4-int
VHL	2,04	0,0482829	<b>GH03J010188</b>	82,89	0,015 MLT1D,MLT1J2
CDK17	2,36	0,04837823	<b>GH12J096419</b>	10,62	0,014 MLT1B,MLT1D

**Appendix Figure S5:** HERV transcripts originate from activated enhancers.

- (A) Top panel: screen captures from IGV showing a HERV overlapping with a region annotated as an enhancer in several tissues but showing increased transcription only in the data from myelin+ patient T cells. Bottom panel: position of the LTR downstream relative to functional annotations in the 127 tissues of the Epigenomic Roadmap consortium. Color code as in Appendix Figure S1.
- (B) Table listing the genes described in Figure 5C and indicating the connected GeneHancer item, its fold activation in myelin+ patient T cells as compared to myelin+ T cells from healthy controls, and the annotation of the overlapping HERV-derived sequence from RepeatMasker.

Primer name	Forward (5'-3')	Revers (5'-3')
HERV-H/env62 (AJ289709.1)	TGGCCGCTCCTTTGTGTATCTCTT	ACAGGGCAAGTGTCTTCCTAAGCA
HERV-H/env59 (AJ289711.1)	TTTATCAGTCCTCCAGGCCCAAGT	AAGAATTGGGAGGACCCAGGACAT
HERV-H/env60 (AJ289710.2)	AGGCCTTGACTTACTCACTGCTGA	TTGGAAGGACCCAGGACATCCAAT
HERV-H/pol62	CCTTCAGCATTTATTCTCCAAAGG	GAAGAATTATGCCGAGGTAGGGTAA
HERV-K18/LTR (AF333072)	GGCGGTGCAAGATGTGCTTTGTTA	ACCAGCGTTCAGCATATGGAGGAT
HERV-K102/env	TGCCAAACCTGAGGAAGAAGGGAT	TGCAGGCATTAAACATCCTGGTGC
HERV-W/env	TCAGCTTAACCTCCCTAGCAGCAGT	TTCGAAGCTCCTCTGCTCTACGTT
RPLP0	AGGTGTTTCGACAATGGCAGCAT	TGCAGACAGACACTGGCAACAT
TGFbeta1	ATGAACCGGCCTTTCTGCTTC	AAGCAATAGTTGGTGTCCAGGGCT
TNFalpha	GCCCATGTTGTAGCAAACCCTCAA	AGTAGATGAGGTACAGGCCCTCTGAT
IL8	CCCACTGCGCCAACACAGAAATTA	AAACTTCTCCACAACCCTCTGCAC
c-FOS	GAGAATCCGAAGGGAAAGGAATAA	TTCTCATCTTCTAGTTGGTCTGTC
FOSB	CTCGGCCTAGGTCACGTT	GCCAGAGTTTCTAGAAGCAGTTT
c-JUN	ATCAAGGCGGAGAGGAAGCG	TGAGCATGTTGGCCGTGGAC
IL2	CCAAACTCACCAGGATGCTCACAT	AGCACTTCTCCAGAGGTTTGAGT
GALNT3	CTCTATGTCTGGATGTTGGAGAAA	GAGCAGAGTATTCAAAGTACTGGT
CYP1A1	TCTTCCTTCGTCCCTTCAC	TGGTTGATCTGCCACTGGTT
CYP1B1	GCCTGTCACTATTCCTCATGCCA	ATCAAAGTTCTCCGGGTTAGGCCA
NQO1	TCCCCCTGCAGTGGTTTGGAGT	ACTGCCCTCTTACTCCGGAAGGGT
GCLC	GAACAATGTCCGAGTTCAATACAG	GCTTGTTTATTGCTTCATCTGGAA
GCLM	TCAACCCAGATTTGGTCAGGGAGT	TCCAGCTGTGCAACTCCAAGGA
SLC6A12	ATAGTGCCCTGTATTTTATCCTG	ATCCCAGAAGTTGTAAAAGCAGTA
SLC5A3	ATAGTGCCCTGTATTTTATCCTG	ATCCCAGAAGTTGTAAAAGCAGTA
AKR1B1	AGAACGCATTGCTGAGAACTTTAAGGTC	GTAATC CTTGTGGGAGGTACAGCTCAA
CDKN1A	AGGTGGACCTGGAGACTCTCAG	TCCTCTTGGAGAAGATCAGCCG
MDM2	GGAGCAGGCAAATGTGCAATACCA	TTCCGAAGCTGGAATCTGTGAGGT
BAX	AGGG TTTCATCCAGGATCGAGCAG	ATCTTCTTCCAGATGGT GAGCGAG
VEGF	CTACCTCCACCATGCCAAGT	GCAGTAGCTGCGCTGATAGA
DNAJB9	GTAAAGGACAAAGAGGTAGTGGA	TGATTTTCAAACGCTTCTTGAT
MT1E	GACCCCAACTGCTCTTGC	GCATTTGCACTCTTTGCACTT
MT2A	GCGTGCAACCTGTCCCGACTC	TGGGATCCATGGCGTGCT
HSP90B1	GACGGATGATGAAGTAGTACAGAG	GTGATGCATTTAATCCATCCAAC
HSPA6	CAAGGTGCGCGTATGCTAC	GCTCATTGATGATCCGCAACAC
RPLP0 promoter	ACAGAGCGACACTCCGTCTCAAA	ACCTGGCGAGCTCAGCAAACATAA
FOSB promoter	AGAATAAACCCCAACCCCA	AACATCTCCGACTCTTTGC
c-FOS promoter	AGCAGTTCCCGTCAATCCC	AGGGCTACAGGGAAAGGC
TNFalpha promoter	AAACACAGGCCTCAGGACTCAACA	ACCAGGTCTGTGGTCTGTTTCTT
IL8 promoter	AAGAAAACCTTTCGTCACTCCG	TGGCTTTTTATATCATCACCTAC
TGFbeta1 promoter	ATGGCACCGCTTCTGTCTTTCTA	TAAGCAGCCTCTGTCACTCAACA
chr11	TGGAATAGTCGTGATTGAGGGA	CCAGCAAGTATTTCTCACCAA