

Review

# High-Throughput Sequencing is a Crucial Tool to Investigate the Contribution of Human Endogenous Retroviruses (HERVs) to Human Biology and Development

Maria Paola Pisano <sup>1</sup> , Nicole Grandi <sup>1</sup> and Enzo Tramontano <sup>1,2,\*</sup> 

<sup>1</sup> Laboratory of Molecular Virology, Department of Life and Environmental Sciences, University of Cagliari, 09042 Cagliari, Italy; mp.pisano@unica.it (M.P.P.); nicole.grandi@unica.it (N.G.)

<sup>2</sup> Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, 09042 Cagliari, Italy

\* Correspondence: tramon@unica.it; Tel.: +39-070-6754538

Received: 24 April 2020; Accepted: 10 June 2020; Published: 11 June 2020



**Abstract:** Human Endogenous retroviruses (HERVs) are remnants of ancient retroviral infections that represent a large fraction of our genome. Their transcriptional activity is finely regulated in early developmental stages and their expression is modulated in different cell types and tissues. Such activity has an impact on human physiology and pathology that is only partially understood up to date. Novel high-throughput sequencing tools have recently allowed for a great advancement in elucidating the various HERV expression patterns in different tissues as well as the mechanisms controlling their transcription, and overall, have helped in gaining better insights in an all-inclusive understanding of the impact of HERVs in biology of the host.

**Keywords:** HERVs; human genome; high-throughput sequencing; expression

## 1. Introduction

Human Endogenous Retroviruses (HERVs) are repetitive elements derived from infection by ancient retroviruses, now extinct, that integrated within the genome of primate germ line cells and have been transmitted as proviruses through the offspring [1]. These sequences currently constitute a significant portion (about 8%) of human DNA [2,3]. The majority of HERVs have lost their replication and expression activities, mostly due to deletions and recombination events, or for the accumulation of several mutations that inactivated them, while some HERVs have maintained the classical proviral structure—including the main genes flanked by two Long Terminal Repeats (LTRs)—and have, thus, the potential for transcriptional activity [1,4]. In addition, a great proportion of these repetitive elements are present in human chromosomes as single LTRs, termed solo LTRs, which have originated from recombination events. This is evident when comparing HERVs integrations in humans and their orthologs in primates [5–7]. Nonetheless, both proviral and solo LTRs may retain sequences able to control gene expression, such as promoters, enhancers, and polyadenylation signals [8].

Overall, few HERV insertions have been studied to understand their implication in human pathophysiology [9]. The best-described example of HERV involvement in the host physiology is the production of syncytin-1, a retroviral protein coded by the *env* gene of a provirus belonging to the HERV-W group, expressed in human trophoblasts [10–13]. Other HERV proviruses and proteins have been investigated for their possible involvement in pathogenesis, in particular, in both cancer and autoimmunity [13]. In fact, there are several pieces of evidence showing an abnormal increase in HERV expression in tumor cells [14–18]. Probably this HERV upregulation is due to the lack of transcriptional control in cancer. In most cases, the transcripts are not translated and overall,

there are only a few examples of a causative role of HERVs in tumorigenesis [1,14,19]. The complex connection between HERVs and the immune response has been also widely investigated [20,21]. Indeed, some inflammatory settings can induce HERV expression, while some HERV products may trigger the host immune response, and hence, activate the innate immune pathways [20–23]. Importantly, extensive knowledge of the mechanisms of HERV-mediated immune activation would be essential for the understanding of possible HERV implications in inflammatory conditions as well as in autoimmune diseases [21,24].

HERVs can also have an impact on human biology other than through protein production [9,25]. HERV LTRs include enhancers, promoters, polyadenylation signals and splice sites within their sequences, and may influence neighboring cellular gene expression [9,25,26]. In addition, HERV integrations may alter the normal gene functions by providing alternative and aberrant sites for splicing or by interfering, either positively or negatively, with its mRNA transcription through the production of non-coding RNAs [27].

Given that HERVs/LTRs represent a large portion of the human genome and can potentially influence our physiology, it is quite clear that cells should finely control HERV transcriptional and translational activity through various mechanisms such as accumulation of mutations, RNA silencing, and histone/DNA methylation [28–32]. While most HERVs are silenced, some elements are normally expressed in various developmental stages of human embryogenesis, and their activity is regulated in different human tissues [31,33,34]. HERVs could be also activated as a consequence of some pathological conditions, like HIV infections or cancer, characterized by alterations in epigenetic regulation [15,35–37]. Overall, such expression patterns make it difficult to clearly establish a causal association between HERVs and diseases.

Some of the latest integrated members of the HERV-K HML-2 group are insertional polymorphisms in the human population [38]. The identification of insertional polymorphic proviruses might be important in the investigation of the role of HERVs in human biology, as physiological or pathological phenotypic variants may co-occur with or be associated to such polymorphic HERV insertions [39].

The great majority of the studies about the effects of HERVs on human pathophysiology are based on microarrays, hybridization-based approaches, or reverse transcription followed by polymerase chain reaction (RT-PCR). Unfortunately, due to technical limitations, these studies have often failed to explain the complexity of the HERVs impact on host biology in its entirety [40]. However, the sequencing of the human genome, the resulting genomic characterization of HERVs and, finally, the advent of high-throughput technologies has led to a great advancement in this field [41]. In fact, such technologies have allowed one to take into account genome variations, to analyze regulatory elements and the three-dimensional organization of the genome, and to characterize the HERV transcriptome [42,43]. This review focuses on new insights of HERV contribution to human pathophysiology and the development obtained through the application of high-throughput sequencing technologies. We briefly describe the HERV databases currently available, then we discuss the possible applications of specific high-throughput sequencing technologies to the HERV field. Finally, we report an overview of discoveries on the role of HERVs in human biology made through the application of these high-throughput sequencing techniques.

## 2. Identification of HERVs in the Human Genome

The comprehensive sequencing of the human genome has made an important contribution to genetics and HERV research. An initial improvement has been the possibility to identify and classify retroviral sequences through computational methods. One of the most used software for the identification of HERVs and other repetitive elements is RepeatMasker (<http://www.repeatmasker.org>), a program that checks the genomes for interspersed repeats, by making use of a database of repetitive sequences, Repbase (<https://www.girinst.org>). RepeatMasker also makes use of Dfam (<https://www.dfam.org>), another database of repetitive sequences organized by families. Of note, the analysis of RepeatMasker allows for collecting of the majority of repetitive elements, referred to as HERVs and solo LTRs, but it

is not able to predict the retroviral structure of HERV proviruses. A similar database, *hervgdb4*, has been created with the specific aim to detect HERVs through an Affymetrix array (HERV-V3). *Hervgdb4* includes proviral and solo LTR sequences that have been collected by using 42 selected proviral sequences (prototypes) as references for RepeatMasker analyses or, alternatively, by reconstructing proviral structures from data of the Dfam database [44]. Of note, the “prototype” subset of sequences in the *hervgdb4* database also includes gene annotation. Since this database has been created to design the probes of an Affymetrix array, all the sequences are fragmented [44]. A third interesting tool is RetroTector, a software developed for the automated recognition of the best-preserved proviral sequences in the genome of vertebrates [45,46]. The HERV recognition process through RetroTector led to a related study that analyzed the human genome assembly GRCh 37/hg19, identifying, characterizing and classifying a total of 3173 HERV sequences [3]. The classification work, based on a multiple approach, split the HERV proviruses into 39 well-supported phylogenetic groups, belonging to the retroviral classes I (Gamma- and Epsilon-like), II (Beta-like), and III (Spuma-like) [3]. Interestingly, an additional 31 noncanonical HERV groups were identified, revealing a high degree of mosaicism [3]. RetroTector can also predict the sequence of the retroviral genes and a multitude of other retroviral features, like the Primer Binding Site (PBS) or the Poly Purine Tract (PPT), as well as the putative protein products. By contrast, it is unable to detect solo LTRs [3].

In addition to the mentioned three larger databases, there are also some smaller ones that only collect the HERV sequences belonging to a single group. This kind of database usually starts with the analyses of the human genome with RepeatMasker, RetroTector, or both. The data collected are then manually visualized and inspected, and sometimes implemented by performing BLAT searches. The HERV coordinates provided in these databases are, hence, the most accurate and well-annotated. Example of HERV groups deeply studied are the HERV-W, and several HML subgroups [6,32,47–50].

The abovementioned HERV databases differ especially in the number of HERV elements included, the accuracy of the coordinates, and the availability of gene annotations and other retroviral features. Therefore, a proper application of high-throughput in HERVs research should start with the choice of the best HERV database for that specific analysis. For example, a large and inclusive dataset may be useful for the identification of HERVs as biomarkers. Instead, the identification of putative physiological or pathological functions, or the prediction of coding capability, may need accurate coordinates and possibly, gene annotations.

### 3. Examples of High-Throughput Applications in HERV Research

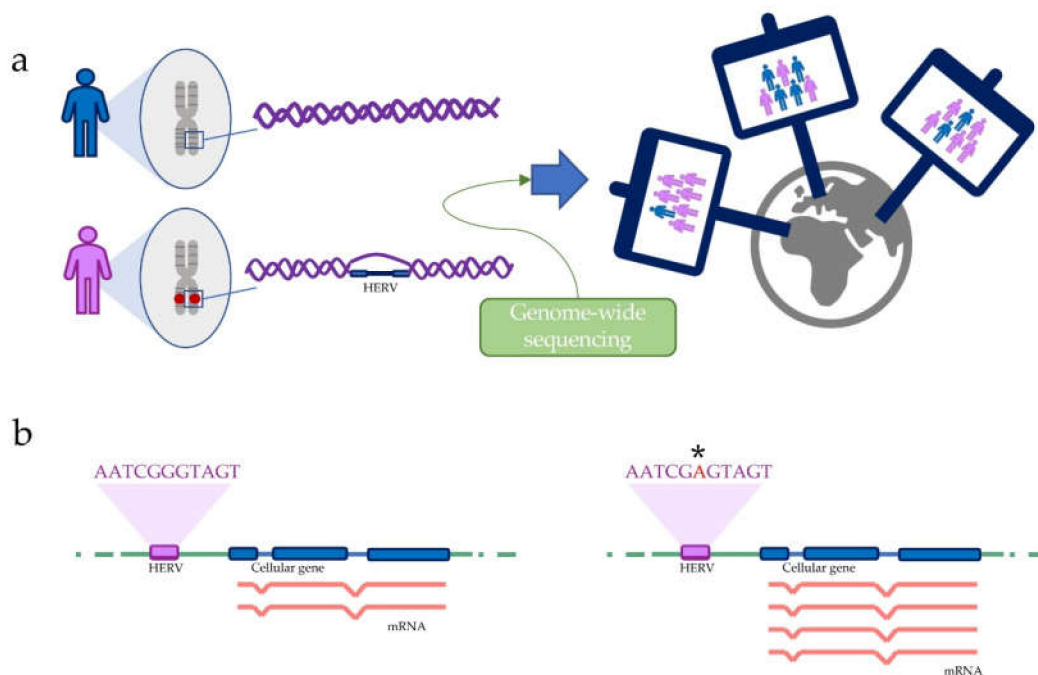
Since the first draft of the human genome sequence was completed in 2001 [51,52], thousands of full genome sequences have become publicly available [53]. High-throughput sequencing technologies have allowed the performance of multiple genome and transcriptome sequencing in parallel. For example, DNA sequencing and RNA sequencing (RNA-seq) can help to evaluate human genomic diversity through the identification of variants and mutations [42]. DNA–protein interactions, such as Chromatin Immunoprecipitation sequencing (ChIP-seq) and Methylation sequencing (Methyl-seq), are useful to explain epigenetic changes [43]. At the transcriptomic level, RNA-seq can be instead used to analyze the transcriptome and identify modulated genes, while Ribosome sequencing (Ribo-seq) can determine mRNA transcripts that are being translated [43].

Before starting the description of some examples of high-throughput applications in HERV research, it is necessary to underline some limitations of such technologies in this field. Indeed, most RNA-seq data analysis software, which have not been specifically designed to identify repetitive elements, may have some difficulties in properly mapping short-reads to repetitive regions of the genome. For this reason, it is necessary to take into account the limitations of the application of this technology to repetitive sequences, including HERVs [54]. Indeed, the number of HERVs identified after mapping the reads to the genome may be underestimated, as the majority of pipelines do not count reads mapping equally well to two or more genomic regions [54]. Moreover, a good practice is the use of

the paired-end and high-quality reads sequences, and to always keep in mind possible complications in de novo assembly of whole genomes and mapping in repetitive regions [55–57].

### 3.1. Genome Sequencing and HERVs Variability

As already mentioned, there is a certain HERV variability in the human genome. A deep knowledge of such variability may effectively help to better understand the real impact of HERVs in human biology. For this reason, one of the most essential challenges in HERVs research is the identification of insertional polymorphisms (Figure 1a). For this purpose, several studies take advantage of data coming from large datasets of whole-genome sequences. Indeed, many projects have provided numerous copies of whole-genome sequences in the form of short DNA fragments, namely reads, collected in databases. Whole-genome sequences sources can include different type of data, e.g., data from healthy donors, like those collected by the 1000 Genomes Project [53,58], or from pathological samples, like those from cancer patients collected by The Cancer Genome Atlas (TCGA) [59] and the International Cancer Genome Consortium (ICGC) [60,61]. Reads from these sources can be mapped to a reference human genome assembly, and variation from the latter may be useful to identify polymorphic insertions or single nucleotide changes [42,43] (Figure 1a). Hence, some bioinformatics tools have been developed to discover insertional polymorphic transposable elements, including one that is specifically designed for HERVs [62]. In this way, it is possible to detect insertional polymorphic HERVs and to study their frequency in different human populations [63–68]. However, it is still difficult to study insertional polymorphisms within the same individual by using whole-genome sequences data. The application of these technologies remains, hence, insufficient to investigate HERV mobility or somatic integrations [41].



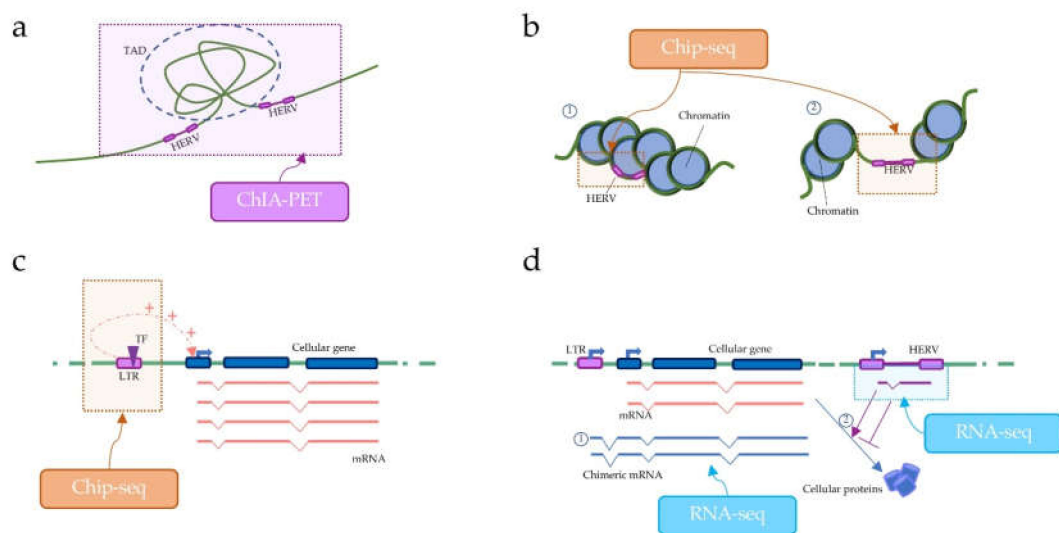
**Figure 1.** Example of possible application of high-throughput sequencing in the study of Human Endogenous Retrovirus (HERV) variability. The high-throughput sequencing of whole genomes allows the identification of insertional polymorphisms and to assess the genome-wide distribution of these polymorphic loci (a). This application also allows the identification of single nucleotide variations associated with expression quantitative trait loci (\*), which explain changes in the host gene expression levels (b).

The identification of Single Nucleotide Polymorphisms (SNPs) into the HERVs sequence is another application that required whole-genome sequences data, especially for its possible implication in diseases [53,58]. Indeed, SNPs may be associated with Expression Quantitative Trait Loci (eQTLs),

which are variants that explain changes in the host gene expression levels (Figure 1b) [39,69,70]. Such variations in gene expression can be accountable for phenotypical or pathological traits, and HERVs, including SNPs associated with eQTLs, can hence be directly linked to physiology and diseases [39,69].

### 3.2. Regulation of HERV Expression and Impact on Human Gene Expression

High-throughput sequencing technologies have profoundly improved our knowledge of the three-dimensional organization of the genome, chromatin state, and chromatin modifications [42]. The 3D chromatin can be analyzed by paired-end tag sequencing (ChIA-PET) and Hi-C [71], which consist of consecutive steps of ligation and sequencing of close crosslinked chromatin portions, so that frequently interacting portions can be visualized in the contact matrix [42,71]. These techniques are particularly useful to identify regions of the same chromosome that preferentially interact with each other due to the particular organization of the chromatin in Topologically Associating Domains (TADs). The main application of these techniques in HERV research is the investigation of possible involvements of HERVs in TADs' formation (Figure 2a) [34,72].



**Figure 2.** Example of possible application of high-throughput sequencing to the study of HERV regulation of gene expression. Paired-end tag sequencing (ChIA-PET) can be used to analyze 3D interactions of the chromatin. It is possible to investigate HERV involvement in the formation of Topologically Associating Domains (TADs) (a). Chromatin Immunoprecipitation sequencing (ChIP-Seq) allows the analysis of DNA–protein interactions. ChIP-Seq can be used to check the chromatin state of HERV loci (b), which may be transcriptionally silenced (1) or transcriptionally active (2). ChIP-Seq can be also used to study the interaction of DNA with other proteins, e.g., Transcription Factors (TFs). TF–LTR interactions can enhance the expression of neighbor genes through cis-regulatory mechanisms (c). RNA-seq technologies allow the analysis of individual HERV loci expression that may influence cellular gene expression (d). For example, HERVs may provide alternative promoters to neighbor genes, resulting in HERV–gene chimeric transcripts (1). HERV non-coding RNAs (ncRNAs) can also be detected through RNA-seq approaches (2).

The chromatin state and modification can also be studied through the application of ChIP-Seq technologies. Indeed, ChIP-Seq allows the analysis of DNA–protein interactions, through the combination of chromatin immunoprecipitation and high-throughput sequencing [42,43]. When applied to the HERV research, ChIP-Seq can be used to check the chromatin state of HERV loci, and its modification in different developmental stages or diseases (Figure 2b) [73,74]. ChIP-Seq can be also used to study the interaction of DNA with proteins other than those composing chromatin, and to predict, for example, the interaction between LTRs and Transcription Factors (TFs). Indeed, due to the presence of enhancers in their LTR

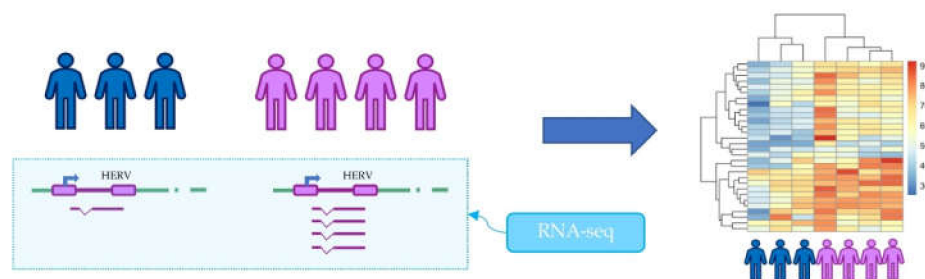


sequences, HERVs and solo LTRs integrated near to a cellular gene can enhance the expression of the gene through cis-regulatory mechanisms (Figure 2c) [21].

HERV integrations may also interfere with the normal gene functions by providing alternative promoters, polyadenylation signals, and sites for splicing [21,75,76]. These mechanisms of interference can result in HERV-gene chimeric transcripts, which may be computationally reconstructed by applying RNA-seq to the transcriptome, and then, by using software tools that detect reads mapping in splice junctions [75]. HERV non-coding RNAs (ncRNAs) can also be detected through RNA-seq approaches, and the results obtained could be used as a starting point for studies of HERV-mediated cis-regulation of cellular genes (Figure 2d) [21,77].

### 3.3. Identification of Expressed and Modulated HERV Loci

The study of HERV role in human pathophysiology has also been performed through the analysis of the expression of each HERV locus and its modulation in different healthy and diseased tissues (Figure 3) [13]. Indeed, the causal relationship between HERVs and diseases is still not clear, and most of the studies have not been able to identify the expression of individual loci, but were limited to the expression of entire HERV families [35,78]. The application of RNA-seq to the HERV transcriptome can provide essential information on the transcriptional contribution of the different families, and especially, the expression levels of the individual HERV loci based on their unequivocal localization in the human genome. For example, RNA-seq pipelines for differential expression have been used to show over- or under-expression of HERVs in tissues [79], diseases [80,81], and medical treatments [82]. The data obtained with this approach can be used as a starting point for understanding the importance of HERV transcripts and their possible translation into proteins, but also to identify HERVs to be used as biomarkers [36]. Finally, bioinformatics pipelines for metagenomics allowed the study of HERV contribution to the human intestinal virome [83].



**Figure 3.** Example of possible application of high-throughput sequencing for the identification of expressed and modulated HERV loci. The application of RNA-seq to the HERV transcriptome can provide the expression levels of the individual loci. Differential expression analyses can show HERVs modulation in different conditions, for example, healthy control versus disease.

## 4. New Insights on HERV Contribution to Human Pathophysiology and Development

### 4.1. HERV Variability in Human Population

The HERV-K (HML-2) group is known to include some young elements integrated into the genome of modern humans after the divergence from the lineage of chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) [47]. Moreover, the HML-2 group may have been active in archaic hominids also after the divergence from the lineage of modern humans [84,85]. Indeed, some studies have analyzed the genome sequences of Neanderthal and Denisovan, identifying some HML-2 insertions not included in the modern human genome assembly [84,85]. Specifically, Agoni et al. identified 14 HML-2 loci in Neanderthal and Denisovan, including 8 for which coordinates have been clearly detected. Only seven of these insertions in archaic hominids were also found in the genome of certain modern humans, as polymorphic insertions [58,65,84]. Moreover, Lee et al. found six HML-2 integrations that appear to be unique to Neanderthal and Denisovan genomes [84]. Hence, the recent insertional activity of

the HML-2 group is of particular interest for the presence of polymorphic proviruses [38,47]. In total, there are 36 HML-2 proviruses in the human population that are not included in the human reference genome [39,63–65,84,85]. Furthermore, also some of the proviruses present in the human genome assemblies hg19 and hg38 are known to be unfixed among the population [39,47,64]. The analysis of data from the 1000 Genomes Project has revealed differences in the frequency of HML-2 insertions among the five super-populations: African, East Asian, Admixed American, European, and South Asian. Not all the HML-2 insertions occur in the different populations with the same frequency [63,64], and the state of presence or absence of the totality of the proviruses is sufficient to distinguish the five super-populations [63,64]. Many of the unfixed insertions are rare, and the East Asian population is the one with the lowest prevalence of HML-2 insertional polymorphisms [63,64]. About the potential role of these loci, interesting data derive from the insertions that have significant Single Nucleotide Polymorphism (SNP) association enriched for eQTLs. Indeed, such information tries to establish a relationship between a single nucleotide variant for HML-2 polymorphic and tissue-specific gene expression [39,69]. Interestingly, 46 insertional polymorphisms have SNPs enriched for eQTLs across 44 human tissues [39]. Moreover, 15 of them have SNPs associated with specific neurologic and immunologic traits, including Parkinson's disease and other autoimmune diseases [39].

Polymorphisms in HERVs that are not polymorphic insertions may also be very informative. For example, polymorphisms occurring in transcription factor binding sites may explain the differential HERV expression among individuals and in cancer [86]. Moreover, many HERV elements have been found to be enriched for somatic mutations (hotspots) in cancer [69]. Among these hotspots, the mutation C2270G in ZNF99 is associated with a lower survival rate in kidney cancer patients, and it might be potentially used as a biomarker [69].

#### 4.2. HERV Expression is Regulated during Human Development

HERVs are regularly expressed in the germline, but the epigenetic regulation is essential to finely control their expression since the first steps of embryogenesis [31,33,87–89]. The HML-2 group was reported to be expressed during the early embryonic stages, in cells from morula and pre-implantation blastocysts [89]. Such proviral RNAs result in the production of retroviral products, such as Gag proteins and the HML-2 accessory protein Rec [89]. The expression pattern of HERV-H elements is a key component of pluripotency, and a useful application is to use their expression as a marker for capturing the human naïve pluripotent state in vitro [74,90]. HERV-H elements provide functional binding sites for transcription factors driving the production of chimeric transcripts that modulate pluripotency acting as long ncRNAs [87,91,92]. For example, HERV-H LTRs can include binding sites for the LTR-binding protein 9 (LBP9) [91,93], NANOG protein [74,88], and the Octamer-Binding Transcription Factor 4 (OCT4) [88].

Among the HERV-H group, several proviruses are also transcribed. The classical RNA structure from expressed proviruses is 5'LTR-gag-pro-3'LTR, but seems not to include intact open reading frames. Of note, HERV-H RNA constitutes about 2% of all poly-A RNA in embryonic cells [67].

HERV-H elements actively contribute to the creation of transcriptionally active and self-interacting compartments, TADs [34,94]. Importantly, the creation of HERV-mediated TAD boundaries suggests that these elements can have an important impact on gene regulation [72,95].

The repression of HERV expression is then established in the pre-implantation embryo and is maintained in most developed tissues [96]. The mechanisms of HERV silencing vary, and many data refer to murine Endogenous Retroviruses (mERVs) in a mouse model. Firstly, DNA methylation, catalyzed in mice by DNA methyltransferases, is required to repress mERVs in differentiated cells [96,97]. Moreover, the protein Histone-lysine N-methyltransferase SETDB1 may have a critical role in inhibiting mERVs expression, as it is evident in SETDB1 knockout embryonic cells, where several mERVs are de-repressed [98]. Indeed, HERVs may take advantage of the Ten-eleven translocation methylcytosine dioxygenase (TET) class of proteins to evade DNA methylation and transcriptional repression. For this reason, cells may have evolved methylation-independent silencing pathways, like histone modification,

during developmental stages or tissues when DNA methylation is compromised [97]. KRAB zinc finger proteins (KZFPs) are also involved in HERV silencing by targeting repressive chromatin states [99,100]. Finally, a microRNA (miR-34a) can repress mERV expression by restraining some transcription factor binding proteins [101].

#### 4.3. HERVs Contribute to Somatic Cell Physiology and Disease

HERVs are transcribed also in somatic cells [66,97]. A study analyzed the HERV expression in RNA-Seq samples from the ENCODE project, finding that HERVs are active in a cell line-specific manner [73,79,102]. Of course, as the consequences of HERV activity in somatic cells may be deleterious, the host makes a great effort to efficiently repress the great majority of HERV expression [103,104]. Indeed, 794,972 LTR sequences have Transcription Factor Binding Sites (TFBSs)—most of which co-localized with genes involved in the immune response—that may potentially interfere with neighboring genes [105]. The activity of HERV proviruses and solo LTRs is modulated in response to stress and immune activation [15,106,107]. For example, some MER41 elements include STAT1- and IRF1-binding sites and mediate the activation of the response to pathogens [72,105]. The high number of regulatory elements linked to the immune response is not accidental [108]. In fact, by introducing and amplifying interferon-sensitive enhancers, HERV integrations have shaped the evolution of transcriptional pathways that define the interferon response [108]. In the hippocampus, an area of the brain particularly susceptible to stress, the acute stress is correlated with the silencing of HERVs [107]. HERVs were shown to be also modulated by the histone deacetylases inhibitor vorinostat, which reactivates HIV in latently infected cells [82].

HERVs are known to be de-silenced, and hence, transcriptionally activated in cancer. Accordingly, HERV transcriptional activity has been found to be significantly upregulated in cancer cells than in controls [73,80,86,109–111].

Patterns of HERV modulation are evident also in other diseases. For instance, some HERV-W and HERV-H loci have been found to be expressed in postmortem brain samples from schizophrenic and bipolar patients [112]. The HERV-W group has been often tentatively correlated to multiple sclerosis [22,23]. Nonetheless, transcript levels of HERV-W loci were similar in healthy brain samples and in multiple sclerosis lesions, suggesting a lack of HERV-W modulation in the presence of the disease [113]. Furthermore, analyses of HERV expression have led to contrasting results in PBMCs from patients with systemic lupus erythematosus. A first work observed in fact a general trend of HERV downregulation [114], while a second one identified 124 significantly upregulated HERV loci and no downregulated ones [57]. HERVs have also been recently associated with drug addiction. This is the case of a polymorphic HML-2 solo LTR found in antisense orientation within the sequence of a gene that affects dopaminergic activity [77]. The expression of this antisense LTR can modulate the expression of the neighbor gene, and this HML2 integration is more frequently present in drug-addicted individuals as compared to the general population [70].

## 5. Concluding Remarks

The application of novel high-throughput sequencing technologies has allowed great improvements in the understandings of the HERV contribution to human physiology and pathology. The application of these technologies gives essential insights about HERV variability in the population, HERV expression and silencing during human development, their involvement in the transcriptional activity of the host, and last but not least, HERV expression and modulation in several diseases. Importantly, different from other techniques, such an approach allows the obtaining of important information about the individual HERV loci, being either polymorphic or fixed in the human population. Of course, the analysis of the transcription allows for the obtaining of data on the HERV expression, and further analysis of translation and proteomics are needed to identify retroviral proteins expressed and their functions.



Overall, high-throughput sequencing has, hence, the great potential to provide all-inclusive bases to understand the impact of HERV in the host biology.

**Author Contributions:** M.P.P., writing—original draft preparation; N.G. and E.T., writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** M.P.P. was funded by Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR), grant number CCI 2014IT16M2OP005”.

**Acknowledgments:** Nothing to declare.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jern, P.; Coffin, J.M. Effects of Retroviruses on Host Genome Function. *Annu. Rev. Genet.* **2008**, *42*, 709–732. [[CrossRef](#)]
2. International Human Genome Sequencing Consortium Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945. [[CrossRef](#)]
3. Vargiu, L.; Rodriguez-Tomé, P.; Sperber, G.O.; Cadeddu, M.; Grandi, N.; Blikstad, V.; Tramontano, E.; Blomberg, J. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* **2016**, *13*, 2–29. [[CrossRef](#)]
4. Bannert, N.; Kurth, R. The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genom. Hum. Genet.* **2006**, *7*, 149–173. [[CrossRef](#)]
5. Grandi, N.; Cadeddu, M.; Blomberg, J.; Mayer, J.; Tramontano, E. HERV-W group evolutionary history in non-human primates: Characterization of ERV-W orthologs in Catarrhini and related ERV groups in Platyrrhini. *BMC Evol. Biol.* **2018**, *18*, 1–14. [[CrossRef](#)]
6. Grandi, N.; Cadeddu, M.; Blomberg, J.; Tramontano, E. Contribution of type W human endogenous retrovirus to the human genome: Characterization of HERV-W proviral insertions and processed pseudogenes. *Retrovirology* **2016**, *13*, 1–25. [[CrossRef](#)]
7. Grandi, N.; Pisano, M.P.; Demurtas, M.; Blomberg, J.; Magiorkinis, G.; Mayer, J.; Tramontano, E. Identification and characterization of ERV-W-like sequences in Platyrrhini species provides new insights into the evolutionary history of ERV-W in primates. *Mob. DNA* **2020**, *11*, 1–16. [[CrossRef](#)]
8. Van de Lagemaat, L.N.; Medstrand, P.; Mager, D.L. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biol.* **2006**, *7*, R86.1–R86.14. [[CrossRef](#)]
9. Grandi, N.; Pisano, M.P.; Tramontano, E. The emerging field of human endogenous retroviruses: Understanding their physiological role and contribution to diseases. *Future Virol.* **2019**, *14*, 441–444. [[CrossRef](#)]
10. Blond, J.L.; Lavillette, D.; Cheynet, V.; Bouton, O.; Oriol, G.; Chapel-Fernandes, S.; Mandrand, B.; Mallet, F.; Cosset, F.L. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* **2000**, *74*, 3321–3329.
11. Sha, M.; Lee, X.; Li, X.; Veldman, G.M.; Finnerty, H.; Racie, L.; LaVallie, E.; Tang, X.Y.; Edouard, P.; Howes, S.; et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **2000**, *403*, 785–789. [[CrossRef](#)]
12. Lavalie, C.; Cornelis, G.; Dupressoir, A.; Esnault, C.; Heidmann, O.; Vernochet, C.; Heidmann, T. Paleovirology of “syncytins”, retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2013**, *368*, 20120507. [[CrossRef](#)]
13. Grandi, N.; Tramontano, E. HERV envelope proteins: Physiological role and pathogenic potential in cancer and autoimmunity. *Front. Microbiol.* **2018**, *9*. [[CrossRef](#)]
14. Cegolon, L.; Salata, C.; Weiderpass, E.; Vineis, P.; Palù, G.; Mastrangelo, G. Human endogenous retroviruses and cancer prevention: Evidence and prospects. *BMC Cancer* **2013**, *13*, 4. [[CrossRef](#)]
15. Attig, J.; Young, G.R.; Stoye, J.P.; Kassiotis, G. Physiological and pathological transcriptional activation of endogenous retroelements assessed by RNA-sequencing of B lymphocytes. *Front. Microbiol.* **2017**, *8*, 1–11. [[CrossRef](#)]

16. Ruprecht, K.; Gronen, F.; Sauter, M.; Best, B.; Rieckmann, P.; Mueller-Lantsch, N. Lack of immune responses against multiple sclerosis—Associated retrovirus/human endogenous retrovirus W in patients with multiple sclerosis. *J. Neurovirol.* **2008**, *14*, 143–151. [[CrossRef](#)]
17. Chen, T.; Meng, Z.; Gan, Y.; Wang, X.; Xu, F.; Gu, Y.; Xu, X.; Tang, J.; Zhou, H.; Zhang, X.; et al. The viral oncogene Np9 acts as a critical molecular switch for co-activating  $\beta$ -catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. *Leukemia* **2013**, *27*, 1469–1478. [[CrossRef](#)]
18. Gross, H.; Barth, S.; Pfuhl, T.; Willnecker, V.; Spurrk, A.; Gurtsevitch, V.; Sauter, M.; Hu, B.; Noessner, E.; Mueller-Lantsch, N.; et al. The NP9 protein encoded by the human endogenous retrovirus HERV-K (HML-2) negatively regulates gene activation of the Epstein-Barr virus nuclear antigen 2 (EBNA2). *Int. J. Cancer* **2011**, *129*, 1105–1115. [[CrossRef](#)]
19. Lamprecht, B.; Walter, K.; Kreher, S.; Kumar, R.; Hummel, M.; Lenze, D.; Köchert, K.; Bouhleb, M.A.; Richter, J.; Soler, E.; et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **2010**, *16*, 571–579. [[CrossRef](#)]
20. Hurst, T.P.; Magiorkinis, G. Activation of the innate immune response by endogenous retroviruses. *J. Gen. Virol.* **2015**, *96*, 1207–1218. [[CrossRef](#)]
21. Grandi, N.; Tramontano, E. Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Front. Immunol.* **2018**, *9*, 1–16. [[CrossRef](#)]
22. Dolei, A.; Perron, H. The multiple sclerosis-associated retrovirus and its HERV-W endogenous family: A biological interface between virology, genetics, and immunology in human physiology and disease. *J. Neurovirol.* **2009**, *15*, 4–13. [[CrossRef](#)]
23. Ruprecht, K.; Mayer, J. On the origin of a pathogenic HERV-W envelope protein present in multiple sclerosis lesions. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19791–19792. [[CrossRef](#)]
24. Curtin, F.; Perron, H.; Kromminga, A.; Porchet, H.; Lang, A.B. Preclinical and early clinical development of GNBAC1, a humanized IgG4 monoclonal antibody targeting endogenous retroviral MSRV-Env protein. *MAbs* **2015**, *7*, 265–275. [[CrossRef](#)]
25. Grandi, N.; Tramontano, E. Type W human endogenous retrovirus (HERV-W) integrations and their mobilization by L1 machinery: Contribution to the human transcriptome and impact on the host physiopathology. *Viruses* **2017**, *9*. [[CrossRef](#)]
26. Samuelson, L.C.; Wiebauer, K.; Snow, C.M.; Meisler, M.H. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol. Cell. Biol.* **1990**, *10*, 2513–2520.
27. Kamp, C. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.* **2000**, *9*, 2563–2572. [[CrossRef](#)]
28. Chuma, S.; Pillai, R.S. Retrotransposon silencing by piRNAs: Ping-pong players mark their sub-cellular boundaries. *PLoS Genet.* **2009**, *5*, 1–3. [[CrossRef](#)]
29. Armitage, A.E.; Katzourakis, A.; de Oliveira, T.; Welch, J.J.; Belshaw, R.; Bishop, K.N.; Kramer, B.; McMichael, A.J.; Rambaut, A.; Iversen, A.K.N. Conserved Footprints of APOBEC3G on Hypermutated Human Immunodeficiency Virus Type 1 and Human Endogenous Retrovirus HERV-K(HML2) Sequences. *J. Virol.* **2008**, *82*, 8743–8761. [[CrossRef](#)]
30. Ito, J.; Gifford, R.J.; Sato, K. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 610–618. [[CrossRef](#)]
31. Rowe, H.M.; Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **2011**, *411*, 273–287. [[CrossRef](#)]
32. Lavie, L.; Medstrand, P.; Schempp, W.; Meese, E.; Mayer, J. Human Endogenous Retrovirus Family HERV-K (HML-5): Status, Evolution, and Reconstruction of an Ancient Betaretrovirus in the Human Genome. *J. Virol.* **2004**, *78*, 8788–8798. [[CrossRef](#)]
33. Hurst, T.P.; Magiorkinis, G. Epigenetic control of human endogenous retrovirus expression: Focus on regulation of long-terminal repeats (LTRs). *Viruses* **2017**, *9*, 1–13. [[CrossRef](#)]
34. Zhang, Y.; Li, T.; Preissl, S.; Amaral, M.L.; Grinstein, J.D.; Farah, E.N.; Destici, E.; Qiu, Y.; Hu, R.; Lee, A.Y.; et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **2019**, *51*, 1380–1388. [[CrossRef](#)]

35. Vincendeau, M.; Göttesdorfer, I.; Schreml, J.M.H.; Wetie, A.G.N.; Mayer, J.; Greenwood, A.D.; Helfer, M.; Kramer, S.; Seifarth, W.; Hadian, K.; et al. Modulation of human endogenous retrovirus (HERV) transcription during persistent and de novo HIV-1 infection. *Retrovirology* **2015**, *12*, 1–17. [[CrossRef](#)]
36. Li, M.; Radvanyi, L.; Yin, B.; Li, J.; Chivukula, R.; Lin, K.; Lu, Y.; Shen, J.; Chang, D.Z.; Li, D.; et al. Down-regulation of human endogenous retrovirus type K (HERV-K) viral env RNA in pancreatic cancer cells decreases cell proliferation and tumor growth. *Clin. Cancer Res.* **2018**, *23*, 5892–5911. [[CrossRef](#)]
37. Tabone, O.; Mommert, M.; Jourdan, C.; Cerrato, E.; Legrand, M.; Lepape, A.; Allaouchiche, B.; Rimmelé, T.; Pachot, A.; Monneret, G.; et al. Endogenous retroviruses transcriptional modulation after severe infection, trauma and burn. *Front. Immunol.* **2019**, *10*, 1–12. [[CrossRef](#)]
38. Turner, G.; Barbulescu, M.; Su, M.; Jensen-Seaman, M.I.; Kidd, K.K.; Lenz, J. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* **2001**, *11*, 1531–1535. [[CrossRef](#)]
39. Wallace, A.D.; Wendt, G.A.; Barcellos, L.F.; de Smith, A.J.; Walsh, K.M.; Metayer, C.; Costello, J.F.; Wiemels, J.L.; Francis, S.S. To ERV is human: A phenotype-wide scan linking polymorphic human endogenous retrovirus-K insertions to complex phenotypes. *Front. Genet.* **2018**, *9*, 1–14. [[CrossRef](#)]
40. Voisset, C.; Weiss, R.A.; Griffiths, D.J. Human RNA “Rumor” Viruses: The Search for Novel Human Retroviruses in Chronic Disease. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 157–196. [[CrossRef](#)]
41. Magiorkinis, G.; Belshaw, R.; Katzourakis, A. There and back again: Revisiting the pathophysiological roles of human endogenous retroviruses in the post-genomic era. *Philos. R. Soc. B* **2013**, *368*, 1–12. [[CrossRef](#)]
42. Reuter, J.A.; Spacek, D.; Snyder, M.P. High-Throughput Sequencing Technologies. *Mol. Cell.* **2015**, *58*, 586–597. [[CrossRef](#)]
43. Churko, J.M.; Mantalas, G.L.; Snyder, M.P.; Wu, J.C. Overview of High Throughput sequencing. *Clin. Res.* **2013**, *112*, 1–26. [[CrossRef](#)]
44. Becker, J.; Pérot, P.; Cheynet, V.; Oriol, G.; Mugnier, N.; Mommert, M.; Tabone, O.; Textoris, J.; Veyrieras, J.B.; Mallet, F. A comprehensive hybridization model allows whole HERV transcriptome profiling using high density microarray. *BMC Genom.* **2017**, *18*, 1–14. [[CrossRef](#)]
45. Sperber, G.; Lövgren, A.; Eriksson, N.; Benachenhou, F.; Blomberg, J. RetroTector online, a rational tool for analysis of retroviral elements in small and medium size vertebrate genomic sequences. *BMC Bioinform.* **2009**, *4*, 4–7. [[CrossRef](#)]
46. Sperber, G.O.; Airola, T.; Jern, P.; Blomberg, J. Automated recognition of retroviral sequences in genomic data—RetroTector©. *Nucleic Acids Res.* **2007**, *35*, 4964–4976. [[CrossRef](#)]
47. Subramanian, R.P.; Wildschutte, J.H.; Russo, C.; Coffin, J.M. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* **2011**, *8*, 1–22. [[CrossRef](#)]
48. Flockerzi, A.; Burkhardt, S.; Schempp, W.; Meese, E.; Mayer, J. Human endogenous retrovirus HERV-K14 families: Status, variants, evolution, and mobilization of other cellular sequences. *J. Virol.* **2005**, *79*, 2941–2949. [[CrossRef](#)]
49. Pisano, M.P.; Grandi, N.; Cadeddu, M.; Blomberg, J.; Tramontano, E. Comprehensive Characterization of the Human Endogenous Retrovirus HERV-K (HML-6) Group: Overview of Structure, Phylogeny, and Contribution to the Human Genome. *J. Virol.* **2019**, *93*, 1–19. [[CrossRef](#)]
50. Grandi, N.; Cadeddu, M.; Pisano, M.P.; Esposito, F.; Blomberg, J.; Tramontano, E. Identification of a novel HERV-K(HML10): Comprehensive characterization and comparative analysis in non-human primates provide insights about HML10 proviruses structure and diffusion. *Mob. DNA* **2017**, *8*, 1–18. [[CrossRef](#)]
51. Craig Venter, J.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351. [[CrossRef](#)]
52. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)]
53. Auton, A.; Abecasis, G.R.; Altshuler, D.M.; Durbin, R.M.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; Flicek, P.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [[CrossRef](#)]
54. Jin, Y.; Tam, O.H.; Paniagua, E.; Hammell, M. TETranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **2015**, *31*, 3593–3599. [[CrossRef](#)]
55. Pačes, J.; Huang, Y.-T.; Pačes, V.; Rídl, J.; Chang, C.-M. New insight into transcription of human endogenous retroviral elements. *New Biotechnol.* **2013**, *30*, 314–318. [[CrossRef](#)]

56. Bendall, M.L.; de Mulder, M.; Iñiguez, L.P.; Lecanda-Sánchez, A.; Pérez-Losada, M.; Ostrowski, M.A.; Jones, R.B.; Mulder, L.C.F.; Reyes-Terán, G.; Crandall, K.A.; et al. Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* **2019**, *15*, 1–24. [[CrossRef](#)]
57. Tokuyama, M.; Kong, Y.; Song, E.; Jayewickreme, T.; Kang, I.; Iwasaki, A. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *PNAs* **2018**, *115*, 12565–12572. [[CrossRef](#)]
58. Sudmant, P.H.; Rausch, T.; Gardner, E.J.; Handsaker, R.E.; Abyzov, A.; Huddleston, J.; Zhang, Y.; Ye, K.; Jun, G.; Fritz, M.H.Y.; et al. An integrated map of structural variation in 2504 human genomes. *Nature* **2015**, *526*, 75–81. [[CrossRef](#)]
59. Sanchez-Vega, F.; Mina, M.; Armenia, J.; Schultz, N. Oncogenic Signaling Pathways in the Cancer Genome Atlas. *Cell* **2018**, *173*, 321–337. [[CrossRef](#)]
60. Campbell, P.J.; Getz, G.; Korbel, J.O.; Stuart, J.M.; Jennings, J.L.; Stein, L.D.; Perry, M.D.; Nahal-Bose, H.K.; Ouellette, B.F.F.; Li, C.C.H.; et al. Pan-cancer analysis of whole genomes. *Nature* **2020**, *578*, 82–93. [[CrossRef](#)]
61. The International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **2010**, *464*, 993–998. [[CrossRef](#)]
62. Chen, X.; Li, D.; Birol, I. ERVcaller: Identifying polymorphic endogenous retrovirus and other transposable element insertions using whole-genome sequencing data. *Bioinformatics* **2019**, *35*, 3913–3922. [[CrossRef](#)]
63. Wildschutte, J.H.; Williams, Z.H.; Montesion, M.; Subramanian, R.P.; Kidd, J.M.; Coffin, J.M. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2326–E2334. [[CrossRef](#)]
64. Li, W.; Id, L.L.; Malhotra, R.; Yang, L.; Acharya, R.; Id, M.P. A computational framework to assess genome-wide distribution of polymorphic human endogenous retrovirus-K In human populations. *PLoS Comput. Biol.* **2019**, *15*, 1–21. [[CrossRef](#)]
65. Marchi, E.; Kanapin, A.; Magiorkinis, G.; Belshaw, R. Unfixed Endogenous Retroviral Insertions in the Human Population. *J. Virol.* **2014**, *88*, 9529–9537. [[CrossRef](#)]
66. Thomas, J.; Perron, H.; Feschotte, C. Variation in proviral content among human genomes mediated by LTR recombination. *Mob. DNA* **2018**, *9*, 36. [[CrossRef](#)]
67. Diehl, A.G.; Ouyang, N.; Boyle, A.P. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat. Commun.* **2020**, *11*, 1796. [[CrossRef](#)]
68. Sundaram, V.; Cheng, Y.; Ma, Z.; Li, D.; Xing, X.; Edge, P.; Snyder, M.P.; Wang, T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **2014**, *24*, 1963–1976. [[CrossRef](#)]
69. Chang, T.C.; Goud, S.; Torcivia-Rodriguez, J.; Hu, Y.; Pan, Q.; Kahsay, R.; Blomberg, J.; Mazumder, R. Investigation of somatic single nucleotide variations in human endogenous retrovirus elements and their potential association with cancer. *PLoS ONE* **2019**, *14*. [[CrossRef](#)]
70. Sun, W.; Hu, Y. eQTL Mapping Using RNA-seq Data. *Stat. Biosci.* **2013**, *5*, 198–219. [[CrossRef](#)]
71. Lieberman-Aiden, E.; van Berkum, N.L.; Dekker, J. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)]
72. Raviram, R.; Rocha, P.P.; Luo, V.M.; Swanzey, E.; Miraldi, E.R.; Chuong, E.B.; Feschotte, C.; Bonneau, R.; Skok, J.A. Analysis of 3D genomic interactions identifies candidate host genes that transposable elements potentially regulate. *Genome Biol.* **2018**, *19*, 1–19. [[CrossRef](#)]
73. Criscione, S.W.; Zhang, Y.; Thompson, W.; Sedivy, J.M.; Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genom.* **2014**, *15*, 1–17. [[CrossRef](#)]
74. Santoni, F.A.; Guerra, J.; Luban, J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **2012**, *9*, 1. [[CrossRef](#)]
75. Sokol, M.; Jessen, K.M.; Pedersen, F.S. Utility of next-generation RNA-sequencing in identifying chimeric transcription involving human endogenous retroviruses. *Apmis* **2016**, *124*, 127–139. [[CrossRef](#)]
76. Sokol, M.; Jessen, K.M.; Pedersen, F.S. Human endogenous retroviruses sustain complex and cooperative regulation of gene-containing loci and unannotated megabase-sized regions. *Retrovirology* **2015**, *12*, 1–11. [[CrossRef](#)]
77. Karamitros, T.; Hurst, T.; Marchi, E.; Karamichali, E.; Georgopoulou, U.; Mentis, A.; Riepsaame, J.; Lin, A.; Paraskevis, D.; Hatzakis, A.; et al. Human endogenous retrovirus-K HML-2 integration within RASGRF2 is

- associated with intravenous drug abuse and modulates transcription in a cell-line model. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 10343–10439. [[CrossRef](#)]
78. Gosenca, D.; Gabriel, U.; Steidler, A.; Mayer, J.; Diem, O.; Erben, P.; Hofmann, W.; Seifarth, W.; Fabarius, A.; Leib-mo, C. HERV-E-Mediated Modulation of PLA2G4A Transcription in Urothelial Carcinoma. *PLoS ONE* **2012**, *7*. [[CrossRef](#)]
  79. Haase, K.; Mösch, A.; Frishman, D. Differential expression analysis of human endogenous retroviruses based on ENCODE RNA-seq data. *BMC Med. Genom.* **2015**, *8*, 1–12. [[CrossRef](#)]
  80. Johanning, G.L.; Malouf, G.G.; Zheng, X.; Esteva, F.J. Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. *Sci. Rep.* **2017**, *7*, 1–11. [[CrossRef](#)]
  81. Grandi, N.; Pisano, M.P.; Scognamiglio, S.; Pessiu, E. Comprehensive Analysis of HERV Transcriptome in HIV + Cells: Absence of HML2 Activation and General Downregulation of Individual HERV Loci. *Viruses* **2020**, *12*, 1–19. [[CrossRef](#)]
  82. White, C.H.; Beliakova-Bethell, N.; Lada, S.M.; Breen, M.S.; Hurst, T.P.; Spina, C.A.; Richman, D.D.; Frater, J.; Magiorkinis, G.; Woelk, C.H. Transcriptional modulation of human endogenous retroviruses in primary CD4+T cells following vorinostat treatment. *Front. Immunol.* **2018**, *9*, 1–10. [[CrossRef](#)]
  83. Carding, S.R.; Hoyles, N.D.L. Review article: The human intestinal virome in health and disease. *Aliment. Pharmacol. Ther.* **2017**, *46*, 800–815. [[CrossRef](#)]
  84. Lee, A.; Huntley, D.; Aiewsakun, P.; Kanda, R.K.; Lynn, C.; Tristem, M. Novel Denisovan and Neanderthal Retroviruses. *J. Virol.* **2014**, *88*, 12907–12909. [[CrossRef](#)]
  85. Agoni, L.; Golden, A.; Guha, C.; Lenz, J. Neandertal and Denisovan retroviruses. *Curr. Biol.* **2012**, *22*, R437–R438. [[CrossRef](#)]
  86. Montesion, M.; Williams, Z.H.; Subramanian, R.P.; Kuperwasser, C.; Coffin, J.M. Promoter expression of HERV-K (HML-2) provirus-derived sequences is related to LTR sequence variation and polymorphic transcription factor binding sites. *Retrovirology* **2018**, *15*, 1–16. [[CrossRef](#)]
  87. Fort, A.; Hashimoto, K.; Yamada, D.; Salimullah, M.; Keya, C.A.; Saxena, A.; Bonetti, A.; Voineagu, I.; Bertin, N.; Kratz, A.; et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* **2014**, *46*, 558–566. [[CrossRef](#)]
  88. Lu, X.; Sachs, F.; Ramsay, L.A.; Jacques, P.É.; Göke, J.; Bourque, G.; Ng, H.H. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **2014**. [[CrossRef](#)]
  89. Grow, E.J.; Flynn, R.A.; Chavez, S.L.; Bayless, N.L.; Wesche, D.; Martin, L.; Ware, C.; Blish, C.A. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **2015**, *522*, 221–225. [[CrossRef](#)]
  90. Ruiz-Gonzalez, I.; Xu, J.; Wang, X.; Burghardt, R.C.; Dunlap, K.A.; Bazer, F.W. Exosomes, endogenous retroviruses and toll-like receptors: Pregnancy recognition in ewes. *Reproduction* **2015**, *149*, 281–291. [[CrossRef](#)]
  91. Wang, J.; Xie, G.; Singh, M.; Ghanbarian, A.T.; Raskó, T.; Szvetnik, A.; Cai, H.; Besser, D.; Prigione, A.; Fuchs, N.V.; et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **2014**, *516*, 405–409. [[CrossRef](#)]
  92. Macfarlan, T.S.; Gifford, W.D.; Driscoll, S.; Lettieri, K.; Rowe, H.M.; Bonanomi, D.; Firth, A.; Singer, O.; Trono, D.; Pfaff, S.L. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **2012**, *487*, 57–63. [[CrossRef](#)]
  93. Dodsworth, B.T.; Flynn, R.; Cowley, S.A. The current state of Naïve human pluripotency. *Stem Cells* **2015**, *33*, 3181–3186. [[CrossRef](#)]
  94. Lengronne, A.; Katou, Y.; Mori, S.; Yokobayashi, S.; Kelly, G.P.; Itoh, T.; Watanabe, Y.; Shirahige, K.; Uhlmann, F. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **2004**, *430*, 573–578. [[CrossRef](#)]
  95. Schmidt, D.; Schwalie, P.C.; Wilson, M.D.; Ballester, B.; Goncalves, Â.; Kutter, C.; Brown, G.D.; Marshall, A.; Flicek, P.; Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **2012**, *148*, 335–348. [[CrossRef](#)]
  96. Friedli, M.; Trono, D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annu. Rev. Cell Dev. Biol.* **2015**, *31*, 429–451. [[CrossRef](#)]
  97. Leung, D.C.; Lorincz, M.C. Silencing of endogenous retroviruses: When and why do histone marks predominate? *Trends Biochem. Sci.* **2012**, *37*, 127–133. [[CrossRef](#)]



98. Karimi, M.M.; Goyal, P.; Maksakova, I.A.; Bilenky, M.; Leung, D. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements and chimaeric transcripts in mouse ES cells. *Cell Stem Cell* **2011**, *8*, 676–687. [[CrossRef](#)]
99. Shi, H.; Strogantsev, R.; Takahashi, N.; Kazachenka, A.; Lorincz, M.C.; Hemberger, M.; Smith, A.C.F. ZFP57 regulation of transposable elements and gene expression within and beyond imprinted domains. *Epigenetics Chromatin* **2019**, *12*, 1–13. [[CrossRef](#)]
100. Turelli, P.; Castro-Diaz, N.; Marzetta, F.; Kapopoulou, A.; Raclot, C.; Duc, J.; Tieng, V.; Quenneville, S.; Trono, D. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **2014**, *24*, 1260–1270. [[CrossRef](#)]
101. Choi, Y.J.; Lin, C.; Risso, D.; Chen, S.; Tan, M.H.; Li, J.B.; Wu, Y.; Chen, C.; Xuan, Z.; Macfarlan, T.; et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **2017**, *355*, 1–26. [[CrossRef](#)]
102. Id, T.N.; Autio, A.; Mishra, B.H.; Marttila, S. Aging-associated patterns in the expression of human endogenous retroviruses. *PLoS ONE* **2018**, *13*. [[CrossRef](#)]
103. Brocks, D.; Schmidt, C.R.; Daskalakis, M.; Jang, H.S.; Shah, N.M.; Li, D.; Li, J.; Zhang, B.; Hou, Y.; Laudato, S.; et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* **2017**, *49*, 1052–1060. [[CrossRef](#)]
104. Tobiasson, M.; Abdulkadir, H.; Lennartsson, A.; Marabita, F.; De Paepe, A.; Karimi, M.; Einarsdottir, E.; Grövdal, M.; Jansson, M.; Azenkoud, B.; et al. Comprehensive mapping of the effects of azacitidine on DNA methylation, repressive/permmissive histone marks and gene expression in primary cells from patients with MDS and MDS-related disease. *Oncotarget* **2017**, *8*, 28812–28825.
105. Ito, J.; Sugimoto, R.; Nakaoka, H.; Yamada, S.; Kimura, T.; Hayano, T.; Inoue, I. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet.* **2017**, *13*. [[CrossRef](#)]
106. Schmidt, N.; Domingues, P.; Golebiowski, F.; Patzina, C.; Tatham, M.H.; Hay, R.T. An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *PNAS* **2019**, *116*, 17399–17408. [[CrossRef](#)]
107. Hunter, R.G.; Murakami, G.; Dewell, S.; Baker, M.E.R.; Datson, N.A. Acute stress and hippocampal histone H3 lysine 9 trimethylation, a retrotransposon silencing response. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1–6. [[CrossRef](#)]
108. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **2016**, *351*, 1083–1087. [[CrossRef](#)]
109. Rajagopalan, D.; Tirado-Magallanes, R.; Bhatia, S.S.; Teo, W.S.; Sian, S.; Hora, S.; Lee, K.K.; Zhang, Y.; Jadhav, S.P.; Wu, Y.; et al. TIP60 represses activation of endogenous retroviral elements. *Nucleic Acids Res.* **2018**, *46*, 9456–9470. [[CrossRef](#)]
110. Lock, F.E.; Babaian, A.; Zhang, Y.; Gagnier, L.; Kuah, S.; Weberling, A.; Karimi, M.M.; Mager, D.L. A novel isoform of IL-33 revealed by screening for transposable element promoted genes in human colorectal cancer. *PLoS ONE* **2017**, *12*. [[CrossRef](#)]
111. Siebenthal, K.T.; Miller, C.P.; Vierstra, J.D.; Mathieu, J.; Tretiakova, M.; Reynolds, A.; Sandstrom, R.; Rynes, E.; Haugen, E.; Johnson, A.; et al. Integrated epigenomic profiling reveals endogenous retrovirus reactivation in renal cell carcinoma. *EBioMedicine* **2019**, *41*, 427–442. [[CrossRef](#)]
112. Id, F.L.; Sabuncian, S.; Yolken, R.H.; Lee, D.; Kim, S. Transcription of human endogenous retroviruses in human brain by RNA-seq analysis. *PLoS ONE* **2019**, *14*. [[CrossRef](#)]
113. Schmitt, K.; Richter, C.; Backes, C.; Meese, E.; Ruprecht, K.; Mayer, J. Comprehensive Analysis of Human Endogenous Retrovirus Group HERV-W Locus Transcription in Multiple Sclerosis Brain Lesions by High-Throughput Amplicon Sequencing. *J. Virol.* **2013**, *87*, 13837–13852. [[CrossRef](#)]
114. Reis, C.; Song, L.; Petri, M.; Sullivan, K.E. The SLE Transcriptome Exhibits Evidence of Chronic Endotoxin Exposure and Has Widespread Dysregulation of Non-Coding and Coding RNAs. *PLoS ONE* **2014**, *9*. [[CrossRef](#)]

