



Review

# The Role of Transposable Elements of the Human Genome in Neuronal Function and Pathology

Ekaterina Chesnokova \* , Alexander Beletskiy and Peter Kolosov

Laboratory of Cellular Neurobiology of Learning, Institute of Higher Nervous Activity and Neurophysiology of the Russian Academy of Sciences, 117485 Moscow, Russia; apbeletskiy@gmail.com (A.B.); kolosov@ihna.ru (P.K.)

\* Correspondence: chesnokova@ihna.ru

**Abstract:** Transposable elements (TEs) have been extensively studied for decades. In recent years, the introduction of whole-genome and whole-transcriptome approaches, as well as single-cell resolution techniques, provided a breakthrough that uncovered TE involvement in host gene expression regulation underlying multiple normal and pathological processes. Of particular interest is increased TE activity in neuronal tissue, and specifically in the hippocampus, that was repeatedly demonstrated in multiple experiments. On the other hand, numerous neuropathologies are associated with TE dysregulation. Here, we provide a comprehensive review of literature about the role of TEs in neurons published over the last three decades. The first chapter of the present review describes known mechanisms of TE interaction with host genomes in general, with the focus on mammalian and human TEs; the second chapter provides examples of TE exaptation in normal neuronal tissue, including TE involvement in neuronal differentiation and plasticity; and the last chapter lists TE-related neuropathologies. We sought to provide specific molecular mechanisms of TE involvement in neuron-specific processes whenever possible; however, in many cases, only phenomenological reports were available. This underscores the importance of further studies in this area.

**Keywords:** transposons; transposon silencing; transposon exaptation; somatic mosaicism in neurons; neurogenesis; synaptic plasticity; human brain evolution; neuropathology



**Citation:** Chesnokova, E.; Beletskiy, A.; Kolosov, P. The Role of Transposable Elements of the Human Genome in Neuronal Function and Pathology. *Int. J. Mol. Sci.* **2022**, *23*, 5847. <https://doi.org/10.3390/ijms23105847>

Academic Editors: Konstantin Volcho, Michael Ugrumov, Nina K. Popova and Elena Rybnikova

Received: 29 April 2022

Accepted: 19 May 2022

Published: 23 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Mobile Genetic Elements in the Human Genome and Their Regulation

### 1.1. Transposon Classification and Types of Transposable Elements in Our Genome

Transposable elements (TEs), also known as transposons, mobile genetic elements, or “jumping genes”, constitute a large fraction of most eukaryotic genomes. Since their discovery by Barbara McClintock [1], transposons in genomes of multiple species have been thoroughly studied and classified, but their interaction with host genes is so complex and diverse that it is still a subject of many ongoing research projects.

TE mobilization is a powerful source of mutagenesis. Besides insertion mutations, repetitive elements provide a basis for nonallelic homologous recombination, leading to the deletion or duplication of genomic fragments [2,3]. An impressive example of TE-mediated adaptive mutation is the famous industrial melanism of the peppered moth [4]. TE-induced mutagenesis may also affect epigenetic regulation, gene expression rate and mRNA processing (see [3,5] for a review).

TEs differ by their mechanisms of mobilization, structure and size. Two main classes of TEs are retrotransposons (class 1) and DNA transposons (class 2). DNA transposons are more ancient; they do not use RNA intermediates for transposition and most of them do not replicate, but use “cut and paste” mechanisms. Retrotransposons replicate using RNA intermediates and reverse transcription (RT) [3,6]. Although the origin of retrotransposons can be traced to near the split of prokaryotes from eukaryotes, these elements only started to thrive around the beginning of mammalian evolution [7]. Retrotransposons include LTR and non-LTR elements, distinguished by the presence or absence of long terminal

repeats. Non-LTR elements include two major groups: long interspersed nucleotide elements (LINEs), which are a few kilobases long, and short interspersed nucleotide elements (SINEs), which are usually under 600 bp [6,8]. Another essential difference is that LINEs (and LTRs) are transcribed by RNA polymerase II (pol II), while most SINEs are transcribed by RNA polymerase III (pol III) [8,9]. All active SINEs in mammalian genomes are derived from pol III-transcribed genes [10]. Quite obviously, SINEs inserted within pol II-transcribed genes are also transcribed by pol II as a part of a corresponding transcript [11]. SINEs are nonautonomous, meaning they depend on the machinery of LINEs in order to retrotranspose. Nevertheless, SINE transcription is controlled by their own regulatory elements [12]. Most mammalian SINEs have a 3' tail homologous to a "partner" LINE; this LINE-like region allows SINEs to be recognized by LINE retrotransposition machinery [13]. More complex TE classification systems are used to accommodate specific minor TE groups across different taxa [14,15].

The initial sequencing of the human genome estimated that about 45% of our genome is derived from different TEs [16]. More recent TE annotation demonstrated that this fraction is actually higher [17]. By some estimates, TEs and products of their activity may even constitute up to 69% of the human genome [18]—TE insertions that experience no purifying selection inevitably accumulate mutations with time, so the most ancient TE copies may be unrecognizable by homology-based methods [13]. By the current estimation, the human genome contains traces of 10 clades of LINEs, 3 types of SINEs, 1 composite retrotransposon family (SVA), 5 classes of LTR retrotransposons, and 12 superfamilies of DNA transposons [19]. Among these, only three groups of non-LTR retrotransposons (Figure 1): L1 (LINE), Alu (SINE) and SVA, are currently capable of mobilization [3,20]. These three groups of TEs together comprise about a third of the human genome [5,7], with L1 taking up the largest fraction of the genome [5] but Alu being the most numerous [11]. The current estimated rates of germline insertion are about 1 in 20 births for Alu, 1 in 270 births for L1, and 1 in 916 births for SVA [7]. These three TE groups are discussed in detail in Sections 1.3–1.6.

Mechanisms of TE interaction with host cells are summarized in Table 1 and described in detail in Sections 1.7 and 1.8.

**Table 1.** Mechanisms of interaction between TEs and host organisms. The right column shows section numbers in the present review.

<b>TE repression</b>	Epigenetic	DNA methylation	1.7.1	
		Histone methylation	1.7.2	
		KRAB-ZFP transcriptional repressors	1.7.3	
	Epigenetic and post-transcriptional	RNA interference	1.7.4	
	Other mechanisms		1.7.5	
	TE derepression in cell stress conditions		1.7.7	
<b>TE exaptation</b>	Regulation of host gene expression by TE sequences	Gene expression rate tuning	1.8.1	
		Promoters, enhancers, insulators	1.8.2	
		TF binding sites	1.8.3	
		Insertions within exons, coding sequence mobilizations		1.8.4
	TE insertions affecting protein sequence	TE insertions affecting the fate of RNA	Exonization (new splice isoforms)	1.8.5
			Alternative polyadenylation sites	1.8.6
			RNA A-I editing	1.8.7
			Other mechanisms	1.8.8
		TE-encoded proteins and non-coding RNAs used by the host cell		1.8.9
		TEs and DNA repair		1.8.10

Interestingly, TEs are able to horizontally transfer between genomes of diverse eukaryotic species. This happens more often with DNA transposons and LTRs, but horizontal transposon transfer was also reported for some non-LTR retrotransposons [21–23].

### 1.2. Inactive TEs in the Human Genome

Our genome harbors multiple copies of TEs that are no longer able to transpose but may still be transcribed and/or used by the host cell as regulatory elements.

Human LTR elements are represented by endogenous retroviruses (ERVs) or their remnants (solitary LTRs), which together account for about 8% of our genome [24]. The genomic organization of intact ERVs is similar to the provirus of their exogenous ancestors [3]. A typical replication-competent provirus is 7–11 kb in size. Its 3' and 5' ends have identical LTRs containing a pol II promoter and binding sites for cellular transcription factors (TFs). The canonical ERV genome in its simplest form contains four genes (more accurately, four ORFs): *gag*, *pro*, *pol*, and *env* (Figure 1). More complex ERVs have additional ORFs. In the ERV genome, *gag* encodes structural components of the virion capsid, *pro* encodes protease, *pol* encodes reverse transcriptase/integrase, and *env* encodes a glycoprotein that binds with the receptor on the host cell surface, allowing the virus to enter the cell (*pro* and *pol* sequences are often considered as parts of one gene). Very few ERVs still retain the ability to produce infectious extracellular particles [9,25,26]. In the case of human ERVs (HERVs), such particles were found in pathologic tissues and cancer cell lines [27–29]. HERVs may still be transcribed in both healthy and pathologic tissues [9,30–32], but the ability of mobilization in HERVs is either nonexistent or very limited [33–35]. Out of at least 31 HERV subfamilies, only HERV-K has some indications of recent retrotransposition activity [3]. About 90% of human LTR elements are solitary LTRs [3,24]—products of homologous recombination between two LTRs that excludes viral genes and leaves one LTR at the locus [3,9,25,35]. DNA transposons are apparently extinct in mammals, with very few exceptions [22,36]; in the human genome, they constitute about 3% but are unable to mobilize [3,5]. DNA transposons have a single ORF encoding a transposase enzyme (Figure 1); this ORF is usually rendered inactive by nonsense or frameshift mutations. However, some of these are still transcribed and present a potential threat to the genome [37].

There are also multiple non-LTR retrotransposon families that are currently unable to mobilize but have many copies in the human genome [19]. A prominent example is mammalian-wide interspersed repeats (MIRs, also called CORE-SINEs [38]) that were actively propagating prior to the radiation of mammals and still represent the second most numerous SINE subfamily in humans [9].

### 1.3. L1 Elements: Their Structure and Mechanism of Mobilization

L1 elements (also known as LINE-1) are a subset of long interspersed nucleotide elements. L1s are the only autonomous human retrotransposons because they encode their own reverse transcriptase. They constitute about 17% of our genome [5], with roughly 500,000 L1 copies per haploid genome [33] (by some estimations, there are ~900,000 copies of L1-derived sequences, including numerous fragmented elements [39]). About 79% of human genes contain at least one segment of L1 in their transcription unit, mostly in introns [40]. However, the overwhelming majority of L1 copies are degenerated—they are incapable of mobilization because of 5' truncations, internal rearrangements, and other mutations. The number of L1 copies currently capable of retrotransposition is less than 100, and just a few of these were shown to be highly active, or “hot” [41]. Nevertheless, more recent analysis of newly inserted L1 elements from geographically diverse individuals demonstrated that “hot” L1s must be more abundant in the human population than was previously estimated; Beck et al. identified dozens of polymorphic active L1s with low allele frequencies [42].

A full-size L1 element is about 6000 bp long [34]; however, L1 truncation during retrotransposition is in fact so common that, in the human genome, an average size of

all L1 copies is only 900 bp [16]. The L1 sequence includes an internal pol II promoter, which allows every full-size L1 copy to be transcribed independently of its location in the genome, and two open reading frames, ORF1 and ORF2, which encode proteins necessary for retrotransposition [34,43]. The L1 sequence also contains some elements that seem to be functional, but their role in L1 mobilization is not obvious. Two additional promoters reading “outwards” into the flanking genomic sequence were identified within L1: an antisense promoter in the 5'-UTR [44,45], and a sense promoter in the 3'-UTR [46,47]. These promoters were found in both human and mouse L1 elements [46,48]. There is also an antisense open reading frame, ORF0, that is only found in primate-specific L1s [49].

The strength of the main L1 promoter strongly depends on upstream flanking genome sequences that may either repress or enhance the promoter activity [50]. ORF1 encodes an RNA-binding chaperone ORF1p, and ORF2 encodes ORF2p, an enzyme with endonuclease and reverse transcriptase activities [5,34]. These proteins form a complex with L1 RNA (L1 ribonucleoprotein particle, or L1 RNP) in cytoplasm. L1 RNP includes one L1 mRNA, many ORF1p molecules, and one or two ORF2p molecules, so ORF1p expression level is much higher than that of ORF2p, and ORF1p is much easier to detect in experiments [51]. After being assembled in the cytoplasm, L1 RNP must enter the nucleus, and the exact mechanism of this is still not well understood. However, a noncanonical nuclear localization signal (NLS) was identified within ORF1p [52]. Host proteins that supposedly participate in L1 RNP nuclear import are transportin 1 (TNPO1) [53] and hnRNPA1, which associates with poly(A)<sup>+</sup> RNAs [54]. The mechanism of L1 retrotransposition is called target-site primed reverse transcription (TPRT): ORF2p generates a single-strand nick in genomic DNA to expose a 3'-OH. The free DNA 3' end is then used as a primer for ORF2p to initiate cDNA synthesis using the L1 mRNA as a template [55]. However, the processes of second-strand target site DNA cleavage and second-strand L1 cDNA synthesis are not well-understood yet [56]. Host DNA repair mechanisms are then necessary to restore dsDNA. DNA repair proteins may facilitate or prevent L1 integration, and experiments studying specific proteins produced inconsistent results in different models [57–60]. L1 endonuclease has relatively weak target-site specificity, preferentially cleaving a loosely defined motif 5'-TTTT/A-3' (where “/” denotes the cleavage site) [61]. TPRT typically results in L1 insertions with recognizable features: insertion at an L1 endonuclease motif, target site duplications and a poly(A) tail. The resulting L1 insertion is very often truncated at the 5' end. The exact reason of truncation is not clear, although there are some data suggesting that host DNA repair systems may intervene in the RT before it is complete [58,59]. In addition, during the insertion, a 5' segment of an L1 may become inverted with respect to its 3' end [62]. The supposed mechanism of such inversion is “twin priming”, when one overhang of the nicked host-DNA strand anneals to the poly(A) tail of the L1 RNA, while the other one folds and anneals somewhere in the middle of the L1 RNA [63]. A minor fraction of L1s insertions happens not by TPRT but rather by an endonuclease-independent mechanism. These insertions likely occur at pre-existing DNA lesions and lack target site duplications [61,64]. There is also evidence that L1 endonuclease activity may cause DNA double-strand breaks (DSBs) without subsequent insertion of the retrotransposed sequence [57].

L1s are the major source of insertional mutagenesis in the human genome [65]. Each individual has L1 copies not present in the reference genome, which contribute to as much as 20% of all structural variants in humans [66]. Moreover, in some cases, L1 insertions may cause significant deletions of genomic sequence at the target site [67]. Even chromosome rearrangements may potentially result from L1 activity [68]. L1 insertions are interspersed throughout the genome [36]. However, the analysis of de-novo L1 insertions and of recently integrated elements in the human genome revealed the presence of insertional hotspots [69]. Recent studies provide evidence that L1 integration events do not target expressed genes, open chromatin or transcribed regions but instead associate with DNA replication [66]. Another aspect of L1 endonuclease specificity is that this enzyme preferentially attacks pre-existing (germline) L1 elements, which makes these regions particularly prone to DSBs [70].

Nearly all of the current L1 activity in the human genome is due to one particular subfamily, Ta1 within the L1HS family. This family is also human-specific, hence the name [42,71,72].

#### 1.4. *Trans RNA Targets Mobilized by L1 Elements and Gene Retrocopying*

L1 elements are not just able to copy themselves, but their enzymatic machinery may also be used to mobilize other DNA sequences if the latter are transcribed. Non-autonomous retroelements, such as Alu and SVA, “hijack” L1s to be retrotransposed. Similar hijacking is possible for some host genes: both non-coding and messenger RNA transcripts may be used as templates for TPRT. This causes formation of retrocopies with a set of specific features—loss of introns and promoter, acquisition of a poly(A) on the 3' end and target-site duplications of varying lengths [34,65]. It must be noted that there is some confusion about the usage of “pseudogene”, “processed gene”, “retrogene” and similar terms [73], especially in some early works. These terms are used in different papers, sometimes interchangeably, to describe a gene retrocopy generated by L1 reverse transcriptase from a host mRNA; for a long time, all retrocopies were routinely annotated as pseudogenes [74]. Nevertheless, the term “pseudogene” means any genomic sequence that is similar to a specific gene but is somehow defective, and while most pseudogenes are indeed formed by retrocopying, there are also TE-independent mechanisms of pseudogene formation [75,76] and retrocopies that are functional genes [77]. Most pseudogenes originate from genes highly expressed in the germline, such as housekeeping genes and ribosomal protein genes [73,75]. In the human genome, 3391 fully functional “parent” genes were associated with pseudogenes, and it was found that almost two thirds of parent genes have only given rise to a single pseudogene, while a small fraction of parent genes produced dozens of pseudogenes each [75].

L1 retrotransposition machinery has a strong *cis* preference, meaning that ORF1p and ORF2p preferentially associate with their own mRNA, so L1-mediated retrocopy formation is quite rare. Experiments with retrotransposition assay in cultured cells showed that mobilization of other cellular RNAs happens in less than 0.05% of all L1 transposition events [78]. Nevertheless, it is suggested that *trans* retrotransposition events may account for at least an additional 10% of human DNA [79]. The very low efficiency of *trans* retrocopy formation observed by Wei et al. in cell cultures [78] may reflect only the retrotransposition of two specific *trans* targets used in this experiment. Indeed, there are reports that some *trans* sequences are retrocopied by the L1 machinery more effectively. Based on the hallmarks of L1-mediated retrotransposition in the human genome, most retrotransposed *trans* RNAs (that are not Alu and SVA elements) are pol III transcripts associated with the ribosome and nucleolus. L1 preferential targets include some highly structured small RNAs and some mRNAs encoding housekeeping proteins. The suggestion that L1 machinery has a preference for some specific *trans* RNA targets based on the the high copy number of these sequences in the genome was confirmed in vitro in cells with ectopically expressed ORF2p: analysis of RNAs associated with L1 RNPs showed the presence of Alu, SVA, U snRNAs and hYRNAs. Surprisingly, it was shown that RPLP1, GAPDH and  $\beta$ -actin mRNAs are not only present in the L1 RNPs, but are more abundant there than the L1 transcript itself [80]. Similarly, a recent study showed that, among all RNA species bound by ORF1p in prostate cancer cells, only a small percentage is represented by L1 RNA [81]. It is therefore possible that the *cis* preference of L1 machinery should be explained not by preferential RNA binding but by some other factors. However, the secondary structure of RNA seems to be important for succesful retrotransposition by L1 [82]. L1 RNPs binding *trans* RNA targets is not the only mechanism of retrocopy formation: there is some evidence that ORF2p may switch from L1 RNA to other RNA templates after TPRT has already started. Such template switching was confirmed to mobilize some small non-coding RNAs [83]. The total number of retrocopies in the human genome is estimated to be from 8000 to 17,000 [84]. Ewing et al., by comparing individual whole-genome sequence data with corresponding reference genome assemblies, estimated the rate of novel germline retrocopy insertions in humans as about 1 insertion per 6000 individuals [73]. Very few

retrocopies are expressed because they generally lack a functional promoter [34]. As in other cases of gene duplication, retrocopies have a functional counterpart elsewhere in the genome, so they are free from selective pressure and thus can accumulate mutations and acquire novel functions [2,74]. Alternatively, retrocopies may substitute for their parental genes. In mammals, many housekeeping X chromosome genes have given rise to functional retrocopies located on autosomes. Autosomal retrocopies are specifically expressed in testis during and after the meiotic stages of spermatogenesis, when all X chromosome genes are transcriptionally silenced due to male meiotic sex chromosome inactivation [74,77]. Lastly, functional retrocopies may simply provide more resources for producing similar proteins, an example of this being genes encoding APOBEC3 antiviral proteins [85].

Exogenous RNA species may also be retrotransposed. cDNA copies of nonretroviral RNA virus sequences have been detected in genomes of many vertebrate species, and some of these elements have signs of integration assisted by LINE retrotransposons. A very recent study showed that SARS-CoV-2 sequences can integrate into genomes of cultured human cells by an L1-mediated mechanism [86].

The ability of L1 machinery to process *trans* RNA templates was used to create a reporter assay that is now a common method for estimating L1 retrotransposition rate. Ostertag et al. designed an L1-EGFP retrotransposition reporter cassette that is capable to express EGFP only after being retrotransposed. The cassette containing the EGFP coding sequence and regulatory elements (a promoter and a poly(A) signal) was inserted into the L1 3'-UTR in the antisense orientation. Importantly, the EGFP gene within this cassette is disrupted by an intron (containing splice donor and splice acceptor sites) in a sense orientation. The L1 element tagged with the EGFP cassette was then cloned into a pCEP-based mammalian expression vector. Thus, cells transfected with this construct may only express EGFP after an L1 transcript containing the antisense EGFP marker has undergone splicing, RT and integration into chromosomal DNA [87]. This method was used in many of the articles cited below.

### 1.5. Alu Elements

Alu elements constitute about 11% of the human genome [5,88], with about one million copies of Alu per haploid genome [33], which makes them the most successful retrotransposons in humans [11]. Three quarters of human genes have Alu insertions [88].

The name of these elements originally comes from AluI, an *Arthrobacter luteus* endonuclease, because most of these TEs bear an AluI restriction site [89,90]. Alus are about 280 bp long [88,91]. Alu elements are derived from the 7SL RNA gene [90,92,93]. The RNA encoded by this gene is a component of the protein signal recognition complex that directs translating ribosomes to the endoplasmic reticulum membrane [94]. Alu elements contain two monomeric sequences derived from 7SL RNA and end in an A-rich tail. The left monomer contains an internal pol III promoter and is separated from the right monomer by an adenosine-rich sequence [93]. Left and right Alu monomers are imperfect repeats of each other. A number of independent left and right Alu monomers have also been identified, the most important of them is a part of the human BC200 RNA gene [95] (discussed in the Section 2.7.3). The internal pol III promoter is too weak to support efficient transcription by itself; genes of pol III transcripts require additional *cis*-acting elements to stimulate their transcription. In the case of Alus, their ability to be transcribed depends on their flanking sequences, so each of one million Alu members might be, to some extent, uniquely regulated [93,96].

Alus are not autonomous and may only be mobilized by hijacking L1 molecular machinery [90]. However, Alus do not have a specific region homologous to L1 or any other "partner" LINE; L1 ORF2p binds tightly to the A-rich tail and the adjacent RNA sequence is not very important for this [97,98]. Still, L1 machinery somehow discriminates between Alu RNAs and mRNAs that also have a poly-A tail. Grechishnikova and Poptsova identified a highly conserved 3'-UTR stem-loop structure in human L1 and Alu RNAs and suggested that this secondary structure is necessary for binding ORF2p [82]. L1 ORF2p also

has relatively weak target-site specificity as an endonuclease, and this explains why more recent Alu and L1 insertions exhibit similar interspersed integration patterns. However, it was repeatedly demonstrated that evolutionarily older L1 and Alu insertions show distinct genomic distributions: older L1s are found mostly in gene-poor AT-rich sequences, whereas older Alus are common in gene-rich GC-rich regions of the genome. This difference likely results from post-integration selection pressure. Some researchers suggest that Alus may possibly have some useful function in gene-rich regions of the genome; the other version is that L1 insertions into genic regions are more deleterious compared with Alu insertions [36]. Indeed, even while human Alus are enriched in genes compared with intergenic regions, they are rarely found in exons [88], suggesting that exon disruption by Alus is harmful.

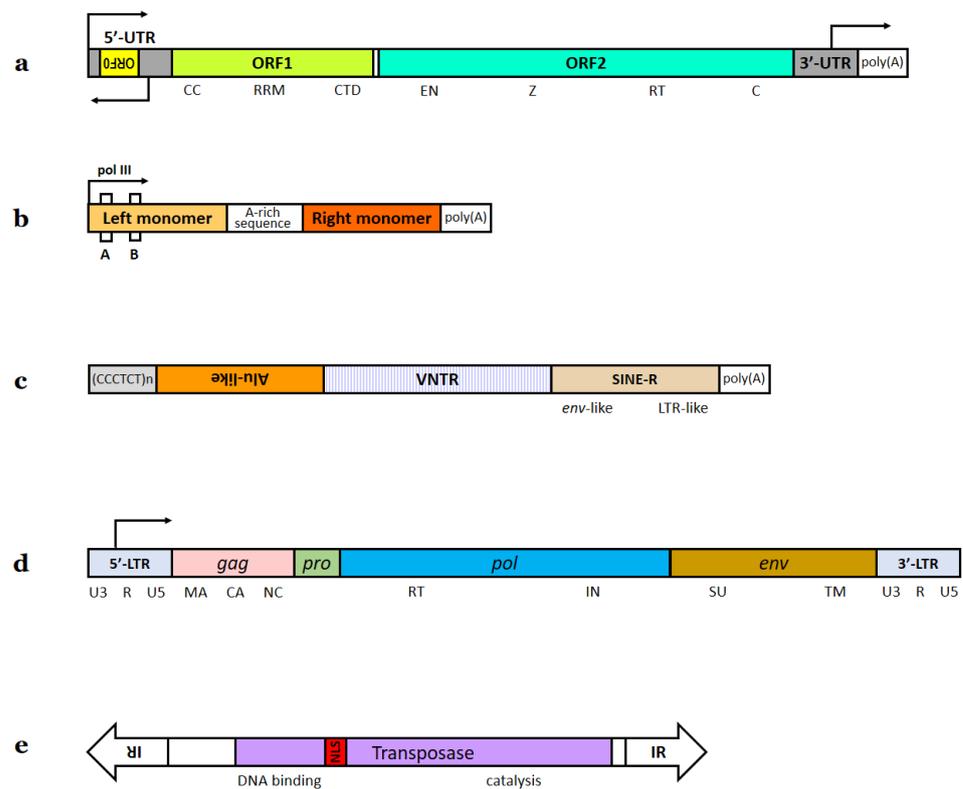
Alu elements are primate-specific [91], even though there are 7SL RNA-related SINES in other species. Rodent B1 SINES are the most well-known example. The B1 sequence is very similar to the left Alu monomer [99]. It must be noted that, in RepeatMasker, a commonly used bioinformatics tool, rodent B1 elements are classified as “Alu family repeats” [100], which sometimes leads to confusion. Rodents also have B2 SINES that bear no structural homology with Alus, being tRNA-related, but may be regarded as Alu counterparts by their functions. Rodent B2s, just like primate Alus, participate in RNA editing and cell stress response (discussed in Sections 1.7.7 and 1.8.7, respectively) [11]. Alu-insertion polymorphisms are very useful for studying human phylogeny and population genetics [90]. The most active Alu subfamilies are AluYa5 and AluYb8 [72].

#### 1.6. SVA Elements

SVA elements only comprise about 0.13% of the human genome, with about 2700 copies. This is the smallest of the known retrotransposon families [20]. Notably, ~1000 of SVA repeats reside within introns of genes [101].

SVAs stand out because they are composite elements. The acronym “SVA” means “SINE-R, VNTR, and Alu” since these elements contain a SINE-R sequence, a variable-number-of-tandem-repeats (VNTR) locus and an Alu-like sequence. The SINE-R sequence had been historically described as a SINE transposon; however, it is related to the endogenous retrovirus HERV-K10, sharing high homology with its 3′ part that includes *env* gene and the 3′-LTR [20,102]. A “canonical” SVA sequence is about 2000 bp long, but different SVA insertions may have sizes from 700 to 4000 bp [103]. This length variation is due to differing numbers of hexamer repeats in the SVA 5′ region and tandem repeats in the VNTR part [104]. Just like Alus, for their mobilization, SVAs need L1 molecular machinery. Endogenous SVAs are very likely transcribed by pol II [105]. However, SVAs do not have any obvious promoters at their 5′ end. Despite the lack of promoters and small copy number, SVAs are retrotranspositionally quite active; a few pathogenic mutations caused by SVA insertions have been described [106]. It was shown that SVA elements can use external heterologous promoters for their own transcription [107]. SVA sequence contains an internal enhancer element which was suggested to recruit TFs to such promoters [103]. GC content in SVAs is generally around 60%, with >70% within the VNTR region; this means that SVA insertions may be considered CpG islands and may thus function as transcriptional regulators, binding TFs, altering the local chromatin structure or forming alternative DNA structures that affect transcription. SVAs make up nearly 2% of genome regions predicted to form G-quadruplex structures [108]. Experiments with reporter gene expression confirmed the ability of SVA elements to act as transcriptional regulators in vitro and in vivo; the effect was different in different models. Notably, different parts of the SVA sequence displayed distinct regulatory properties [20,109].

SVA elements are hominid-specific and represent the youngest family of hominid non-LTR retrotransposons [7,104,107]. SVA insertions are valuable markers for human phylogenetic and population genetic studies. The most active SVA subfamilies are SVA E, F and F1 [72].



**Figure 1.** Types of human transposons. (a). L1 element. Two ORFs of human L1 are separated by a non-coding spacer region containing a stop codon [48]. The L1 5'-UTR has promoter activity in both sense and antisense directions. A small primate-specific antisense ORF0 was identified within L1 5'-UTR downstream of the antisense promoter [49]. An additional sense promoter was identified in 3'-UTR [46]. ORF1 encodes a protein that contains a coiled-coil domain (CC), a non-canonical RNA recognition motif domain (RRM) and a basic C-terminal domain (CTD). ORF2 encodes a protein that has endonuclease (EN) and reverse transcriptase (RT) activities. Functions of the Z domain and the cysteine-rich motif (C) are unknown [101]. (b). Alu element. These elements contain two monomeric sequences derived from 7SL RNA and end in an A-rich tail. The left monomer contains an internal pol III promoter and is separated from the right monomer by an adenosine-rich sequence [34]. Alu pol III promoter elements (Box A and Box B) may only provide efficient transcription if an upstream pol III enhancer is present at the site of integration [93]. (c). SVA element. These elements include CCCTCT hexameric repeats, the Alu-like domain consisting of two antisense Alu fragments separated by a unique sequence, a GC-rich VNTR domain consisting of tandem repeats, and the SINE-R domain sharing homology to the 3' end of the HERV-K10 element (a fragment of its *env* gene and right LTR) [103]. (d). ERV. The order of the structural genes and the arrangement of their major cleavage products are completely conserved among all retroviruses and are necessary for correct virion assembly. Domains within Gag protein: matrix (MA), capsid (CA), nucleocapsid (NC). Domains within Pol protein: reverse transcriptase (RT), integrase (IN). Domains within Env protein: surface protein (SU), transmembrane protein (TM). Env is translated from a spliced subgenomic RNA and is then cleaved to generate SU and TM subunits. Each of the LTRs is composed of the unique U3 and U5 regions separated by a segment (R) repeated at each end of the viral RNA. U3 contains TF binding sites. Transcription from the provirus starts at the left U3-R junction, and the right R-U5 junction provides the site of 3'-polyadenylation. The left U3 and the right U5 are restored during dsDNA formation in the virus replication cycle [25]. (e). DNA transposon (Tc1/*mariner* family). The terminal inverted repeats (IR) contain binding sites for the transposase. The N-terminal part of the transposase contains a DNA binding domain. The C-terminal part of the protein performs target DNA site cleavage and transposon integration. NLS—nuclear localization signal [110]. The elements in this figure are not drawn to scale.

### 1.7. Molecular Mechanisms Suppressing TE Activity and Specific Cases of TE Unsilencing

TE mobilization or expression can have detrimental consequences for the cell, including genome instability, activation of DNA damage stress response, or toxic effects from accumulation of TE-derived RNAs and proteins. For example, ectopic expression of full-length L1 in cultured cells was shown to cause high levels of DSBs, cell cycle arrest, induction of a senescence-like cellular state and apoptosis; moreover, somatic L1 insertions within genes are associated with cancer [111]. Due to this, host organisms have evolved multiple elaborate mechanisms to control transposition events [112]. Systems that suppress TE activity overlap with mechanisms regulating host chromatin structure and DNA repair [36]. TE repression includes epigenetic silencing, RNA interference and some other mechanisms discussed in Sections 1.7.1–1.7.5.

There are specific cases when TE activity is permitted. Notably, TEs play an integral role in early mammalian embryonic development. Regulation of TEs is stage-specific, and on some stages TE repression is temporarily lifted [113–115]. Strikingly, L1 RNA in embryonic stem cells (ESCs) is critical in recruiting necessary machinery to direct a transcriptional program specific to the two-cell stage [115]. L1 epigenetic repression is usually established during early gastrulation and maintained thereafter [116]. Other cases of TE derepression include cell stress response and neuronal differentiation; we discuss these processes below in Sections 1.7.7 and 2.4, respectively.

It must be noted that some regions in vertebrate genomes seem to be absolutely intolerant of TE insertions. Human and mouse genomes each contain almost 1000 transposon-free regions (TFRs) longer than 10 kb. Human, mouse, opossum, *Xenopus tropicalis* and zebrafish genomes have many orthologous TFRs; however, the sequence of most TFRs is not highly conserved. Over 90% of the TFR sequence is noncoding, and most of these regions do not show unusual nucleotide composition. Apparently, these regions are kept TE-free because most of them are significantly associated with genes encoding vertebrate developmental regulators, such as members of the *HOX*, *SOX*, *FOX* and *TBX* gene families [117,118].

#### 1.7.1. DNA Methylation

Most TEs are normally silenced by DNA methylation. The majority of cytosine methylation in mammals occurs within repetitive elements [36]. TE methylation in the human genome is in fact so thorough that, considering the large portion of the genome occupied by TEs, measuring overall TE methylation level may be used as a simplified approach to estimate global DNA methylation level [119]. This is a potentially heritable epigenetic modification, so it may suppress deleterious TE insertions in the germline [120]. In mammalian cells, DNA methyltransferases DNMT3A and DNMT3B perform de-novo methylation, while DNMT1 shows preference towards hemimethylated substrates and maintains the methylation landscape. Recently, it was shown that DNMT1 also has de-novo methylation activity targeted specifically at retrotransposons [121]. Lack of TE methylation strongly reactivates their transcription: knock out (KO) of *Dnmt3L* (a testis-specific DNA methyltransferase regulator) in mice prevents de-novo methylation of L1 and LTR elements in male germ cells, leading to TE derepression, meiotic catastrophe and cell death [122]. DNA hypomethylation and transcriptional reactivation of replication-competent L1 copies happens in some tumors, causing increased rate of transposition [24,37]. Hypomethylation of different human TEs (usually HERVs or L1s) was also reported in some autoimmune diseases [123] and after exposure to various toxic substances or radiation [124,125]. Overall methylation of human TEs, especially Alus, decreases with age. Different TEs have different age-dependent demethylation rates, with L1 methylation being largely unaffected by aging; L1 demethylation occurring in cancer likely has different mechanisms [126,127]. However, it seems like highly specific TE demethylation happens in normal tissues and is important for their function. Examination of genome-wide TE DNA methylation status in 11 human cell types revealed tissue-specific hypomethylation signatures. Many genes proximal to hypomethylated TEs were important for tissue functions, and expression of these genes correlated strongly with TE hypomethylation. Moreover, some of these TEs contained

tissue-specific TF binding motifs. For 26 of such TEs, enhancer activity was confirmed in a reporter gene assay [128]. Similar results were obtained by Philippe et al., in a study where transcriptional and epigenetic signatures of specific L1 copies were identified in 12 human somatic cell lines. The authors found that the bulk of L1 expression happens in a very restricted subset of L1 loci, and these loci are differentially expressed between distinct cell lines. Strikingly, even in cancer cell lines exhibiting high L1 expression, only selected L1 copies were unsilenced [129]. Using different strategies, Muotri et al. have shown that L1 retrotransposition can be modulated by MeCP2, a protein involved in global DNA methylation. They demonstrated that knock down (KD) of MeCP2 increases L1 promoter activity; that removal of DNA methylation with 5-azacytidine reduces MeCP2 association to L1 promoter and increases L1 expression; that MeCP2 KO in mice causes higher retrotransposition in the brain; and that MeCP2 KO neuroepithelial cells have more L1 copies in their genome than WT cells [130].

### 1.7.2. Histone Methylation

Modifications of histone *N*-terminal tails alter the binding of proteins with DNA sequences. Nucleosomes associated with TEs are enriched for methylation of histone H3 at lysine 9 (H3K9), marking transcriptionally repressive and inactive chromatin [131]. There are at least two pathways of H3K9me<sub>3</sub>-mediated TE silencing in human cells: by the KZNF/KAP1 system (described in the next section), or by the human silencing hub (HUSH) complex partnering with the ATPase MORC2. HUSH preferentially binds young, full-length L1s within euchromatic regions, indicating that this is a targeted mechanism of host defense with the capacity to respond to novel insertions [132]. Another important histone modification involved in (less stringent) TE silencing is H3K27 tri-methylation (H3K27me<sub>3</sub>) deposited by the Polycomb repressive complex 2 (PRC2). H3K27me<sub>3</sub> was previously considered to be specific for the silencing of protein-coding genes (facultative heterochromatin), but there is increasing evidence that this modification also targets TEs in diverse eukaryotic species [133]. Disruption of TE silencing by histone methylation may be seen in various cancer types overexpressing histone demethylase KDM4B able to demethylate H3K9me<sub>3</sub>. This results in altered gene expression and genome instability. Ectopic overexpression of KDM4B in cultured cell lines enhanced L1 retrotransposition efficacy, copy number, and associated DNA damage [134]. Histone modification is especially important for silencing SINE elements transcribed by pol III. ChIP assays in cultured cells showed that the DNA sequence of human (Alu) and mouse (B1 and B2) SINEs is heavily methylated and occupied by methyl-CpG-binding proteins. Despite this, pol III was also detected at these SINEs, as were the pol III-specific TFs TFIIB and TFIIC. Global genomic demethylation (achieved by treating cells with 5-azacytidine or *Dnmt1* KO) had little effect on the accessibility of SINEs to pol III transcription machinery or their expression levels. In contrast, pol III loading and expression of SINEs increased significantly in cells treated with chaetocin, an inhibitor of SUV39H1 methyltransferase that deposits H3K9me<sub>3</sub> mark [135].

### 1.7.3. Krüppel-Associated Box Domain Zinc Finger Proteins

Krüppel-associated box domain zinc finger proteins (called KRAB-ZFPs in mice and KZNFs in humans) constitute the largest family of TFs in vertebrates and are one of the most rapidly evolving families of proteins. KRAB-ZFPs have an *N*-terminal KRAB domain and a variable number of tandem C2H2 zinc fingers at the *C*-terminus. Zinc fingers bind with DNA, and different KRAB-ZFPs recognize different sequences. The KRAB domain is a strong transcriptional repressor; it works by tethering the KRAB-associated protein 1 (KAP1, also known as TRIM28) to the DNA sequence recognized by the zinc fingers. KAP1 acts as a cofactor and scaffold protein for the recruitment of multiple proteins participating in chromatin silencing, the most important of them is SETDB1, a methyltransferase that deposits H3K9me<sub>2/3</sub> [136,137]. Several lines of evidence demonstrate that TEs are the primary targets of most KRAB-ZFPs, and it is now considered that the rapid expansion of KRAB-ZFP genes is being driven primarily by the coevolution of retrotransposons. The

number of C2H2 zinc finger genes correlates with the number of ERVs in different species of tetrapods. In human ESCs (hESCs), about three quarters of KAP1-binding sites reside within retroelements [136]. KAP1 deletion leads to a marked upregulation of a range of ERVs in mouse ESCs and early embryos [138]. ChIP-seq experiments revealed that, in hESCs, KAP1 predominantly associates with active primate-specific retrotransposons. A reporter gene system in mouse cells used to screen multiple primate-specific KZNF genes allowed to identify two genes, *ZNF91* and *ZNF93*, encoding TFs that recognize and silence SVA and L1PA elements, respectively. Both genes emerged in the last common ancestor of humans and Old World monkeys and have rapidly undergone dramatic structural changes that parallel the evolution of their TE targets [139].

#### 1.7.4. RNA Interference

RNA interference is a mechanism of silencing transcripts when small silencing RNAs form a complex with an Argonaute family protein and then hybridize with their target RNAs, which eventually leads to the target being degraded or translationally repressed by specific proteins. There are three classes of small silencing RNAs: miRNAs (that mostly target cell mRNAs), siRNAs (targeting mostly viral RNAs) and piRNAs (targeting mostly TE RNAs). These classes have different lengths, and distinct mechanisms of formation and function (see [140–143] for review). Both piRNAs and siRNAs participate in TE silencing; earlier it was considered that piRNAs silence TEs exclusively in germline cells, while siRNAs do the same in somatic cells [144,145], but this distinction seems oversimplified in the light of more recent findings [146–150].

piRNAs interact with PIWI proteins. The *piwi* gene was originally discovered as an essential factor for germline stem cell self-renewal in *Drosophila*, and it was demonstrated that mutations in this gene cause hyperactivation of retrotransposons [145]. Both *Drosophila* and mouse have multiple PIWI proteins, and mutations in these proteins lead to defects in germ cell development [151]. The PIWI–piRNA pathway maintaining germline genomic integrity is conserved in animals. However, piRNAs lack conserved structural motifs and sequence homology across species, being the largest and most heterogeneous class of small non-coding RNAs [152]. piRNAs are mainly derived from intergenic regions containing TEs, TE remnants, and other repetitive elements [145]. TE silencing by piRNAs and associated proteins happens co-transcriptionally or post-transcriptionally. Co-transcriptional silencing involves heterochromatinization and deposition of repressive chromatin marks at target loci. Post-transcriptionally, TE RNAs are cleaved by PIWI [152–155]. Until recently, piRNAs were thought to be restricted to germline cells, but a few years ago they were also detected in neurons [156] (this is discussed in more detail in the Section 2.7.3).

The siRNA pathway (involving Ago2 protein) was initially discovered to function using exogenous short RNAs in *Drosophila* cells [157,158]. Mutant flies with dysfunctional siRNA pathway components are hypersensitive to viral infections [159], and it was shown that Ago2 binds small RNAs derived from viral RNAs [160]. Endogenous Ago2 partners, endosRNAs, were identified later. These RNAs originate from retrotransposons or other repetitive sequences in *Drosophila* genome. Some endosRNAs overlap with piRNA sequences; however, the transcripts have different sizes and bind different proteins [144,157]. Endogenous siRNA derived from retrotransposons were also found in mammalian cells, and KO of *Dicer1* (encoding a key endonuclease generating siRNAs) caused TE derepression in mouse oocytes [149,150,161]. Young and Kazazian demonstrated that the L1 antisense promoter can be used to form L1-specific siRNAs [161]. Berrens et al. performed an acute deletion of *Dnmt1* in mouse ESCs to cause global DNA demethylation and showed that endosRNAs are involved in an immediate response to TE derepression. Small RNA-seq detected a substantial increase in small RNAs mapping to some TEs. Ago2 immunoprecipitation confirmed that a subset of Ago2-bound endosRNAs mapped to TEs. KD of *Dicer* or Ago2 in *Dnmt1*-deficient cells caused strong upregulation of transcription of some TEs. Long-term TE silencing by histone methylation in *Dnmt1*-deficient cells was observed at a later time point [162].

miRNA silencing was confirmed for L1 elements. It was shown that miR-128 directly targets L1 RNA for degradation. The same research group found additional mechanisms of miR-128 indirectly inhibiting L1 transposition: miR-128 binds mRNAs encoding both TNPO1 and hnRNPA1 proteins that are necessary for nuclear import of L1 RNPs [53,54].

#### 1.7.5. Other Mechanisms

Some other proteins interacting with nucleic acids, such as DNA repair factors and proteins binding virus-derived molecules, may also contribute to suppressing TE mobility.

It is well-known that proteins that limit the replication of retroviruses may also inhibit retrotransposons [163]. Proteins of the APOBEC3 family are an important example. These proteins repress retroviruses and retrotransposons by deaminating cytidines (C-U editing) in their RNA [164] or cDNA [164,165]. It was also shown that TE repression by some of the APOBEC3 proteins happens by a deamination-independent mechanism which is not well understood [165–167]. Antiviral proteins involved in transposon silencing also include TREN1, a DNase that breaks down ssDNA and dsDNA in the cytoplasm [168,169], and SAMHD1, which depletes dNTP levels and also silences L1 by a specific mechanism [170].

ERCC2 is a DNA helicase participating in nucleotide excision repair. Deficiency of ERCC2 was reported to increase retrocopying. The mechanism of this is unclear, but it is suggested that ERCC2 prevents TE-mediated insertions [171]. Mov10 is an RNA helicase that participates in miRNA-mediated mRNA silencing. However, it was also shown that, in murine neurons, Mov10 suppresses L1 retrotransposition in the nucleus by directly inhibiting cDNA synthesis. Apparently, the N-terminal region of Mov10 binds the reverse transcriptase domain of ORF2p, and the C-terminal part of Mov10 possessing the helicase activity unwinds the L1 RNA [172]. Another important RNA-binding protein that is involved in TE silencing is TDP-43 [112]; this protein is involved in pathogenesis of some neurodegenerative disorders, and so this mechanism is discussed at length in the Section 3.2.6.

#### 1.7.6. “Self-Restriction” Inherent for Many TEs Is Absent in L1 Elements

TE insertions disrupting host genes jeopardize TE survival in the evolutionary perspective. Thus, many TEs have developed their own highly specific mechanisms that direct their integration to intergenic regions or introns. Even TEs that preferably integrate into gene-rich regions use targeting mechanisms that prevent disruption of ORFs [36,97]. However, L1s do not have evident mechanisms preventing their retrotransposition into genes. Experiments with HeLa cells confirmed that new L1 insertions broadly target all regions of the human genome, being insensitive to transcriptional activity, although a bias for early-replicating genomic domains was noted [66]. Conversely, germline L1 copies are enriched in intergenic regions, which likely reflects negative selection [2,36]. An interesting consequence of purifying selection against L1 insertions is that the density of L1s on the human X chromosome is almost twice as much as on the autosomes [62]; in autosomes, L1 insertions may be cleared by recombination, while sex chromosomes have restricted recombination capacity [173].

Whole-genome profiling of somatic L1 insertions in a few brain regions and the myocardium [174] showed that the rate of such insertions within genes and promoters is, in fact, even higher than predicted for random distribution in all studied samples. No explanation of this observation was proposed in this paper, but it suggests that a specific mechanism directing L1 insertions into genes might exist. The in-vitro results obtained by Sultana et al. [66] suggest that L1 integration does not depend on transcription. Nevertheless, Baillie et al., studying somatic L1 insertions in the brain, found that protein-coding loci are disproportionately affected, and that many genes with intronic insertions are differentially overexpressed in the brain [175]. This leaves the question of L1 insertions targeting transcriptionally active regions open.

### 1.7.7. TE Derepression during Cell Stress and Its Possible Explanations

TE activation (increased transcription or mobilization) during cell stress was registered in multiple eukaryote species [36]. In experiments in cell cultures and in vivo, TE activation was shown to be triggered by various stressors, including oxidative stress, heat shock, exposure to toxic substances or viral infections [11,33,36,96,176–179]. Stress-related TE activation may be a part of an adaptive response; examples of this are known in yeast, plants and insects [36,180–182]. The role of mammalian TEs in cell stress response is not absolutely clear.

Overexpression of L1 in tumors [36,111] may indicate that L1 activation in stressed cells is primarily a side effect of disrupted chromatin maintenance. In works where L1 activation in tumors and/or cells exposed to genotoxic agents was studied, it is considered either a consequence [183] or even a driver [184–186] of genomic instability. The only suggested mechanism by which L1 activation in stress conditions may benefit the host is by increasing genetic diversity [176], which is just a very general idea. Still, Okudaira et al. showed that the cellular cascades involved in inducing L1 retrotransposition after exposure to methamphetamine or cocaine are different from those that trigger L1 activity after DNA damage [187]; therefore, presumably there are multiple mechanisms involved in L1 activation during cell stress.

On the other hand, upregulation of SINEs during cell stress most likely has adaptive functions. It was repeatedly demonstrated that SINE expression in mammalian cells significantly and transiently increases after stress-inducing treatments such as heat shock, viral infection, ethanol application, and UV exposure [11]. This effect was observed in different mammalian species and even in silkworm [177,188]. SINE activation after heat shock was also demonstrated in vivo in experiments with hyperthermia in mice [179]. Notably, other mammalian pol III transcripts are unaffected after heat shock [188] or exposure to DNA-damaging agents [189]; therefore, SINE transcription upregulation must be specific. Mariner et al. showed that, in heat-shock conditions, human Alu RNA acts as a *trans*-acting transcriptional repressor by directly binding pol II [190]. A similar mechanism was shown for murine B2 RNA [191]. Another suggested role for Alu in cell stress is the regulation of stress-induced kinase PKR. PKR normally reacts to viral infection and is activated by viral dsRNA. However, some highly structured RNA species may block PKR activation. Chu et al. demonstrated that Alu RNA forms stable complexes with stress-induced kinase PKR and prevents its activation. The authors studied the consequences of heat-shock recovery, viral infection or cycloheximide treatment in cell cultures. In all three cases, an increase in full-length Alu RNA level was observed. PKR activity decreased in a dose-dependent manner with the level of Alu overexpression [192]. Still, some authors [189] suggest that SINE overexpression after a damaging influence may be detrimental if it causes SINE mobilization and associated mutagenesis.

TEs are not always activated by stress; there are also examples of TE repression under stress conditions in different species [193]. A mechanism of stress-related L1 silencing has been described in human cell cultures. Goodier et al. reported colocalization of ORF1p foci with markers of stress granules and polyadenylated mRNA. The amount of mRNA present in these foci was low under normal conditions and increased as a result of stress induction [194]. It was later shown that artificial inhibition of stress granule formation increases L1 retrotransposition, and overexpression of the viral restriction factor SAMHD1 promotes the accumulation of L1 RNP in stress granules [195]. An example of TE silencing by histone methylation in rats after stress exposure in vivo was reported by Hunter et al.; interestingly, this silencing apparently does not apply to L1 elements [196].

### 1.8. Mechanisms and the Most Well-Known Examples of TE Exaptation

Exaptation is a shift in the function of a trait during evolution. Mobile genetic elements are essentially parasitic, but the host cell may repurpose TE regulatory sequences or RNA and protein products encoded by TEs. A frequently used metaphor for this is “domestication”. Notably, some TEs exhibit tightly regulated activity in somatic tissues,

which is likely used by the host organism because it does not ensure trans-generational TE propagation [24].

TE-derived *cis*-regulatory sites are present in promoter and enhancer elements. TE sequences are found in multiple non-coding regulatory RNAs [24,197]. There are many examples of TE-derived protein-coding genes [198–200].

#### 1.8.1. TEs as Modulators of Gene Expression Rate

At least three mechanisms of TEs inhibiting transcription of nearby coding sequences were proposed.

Han et al. demonstrated that insertion of an L1 sequence inside a gene reduces its expression rate. They established that this is caused by poor transcription elongation of the ORF2 sequence. An intragenic L1 insertion may also cause premature polyadenylation, since L1 sequence has cryptic poly(A) signals. The authors then analyzed the human transcriptome data and showed that highly expressed genes had 4–5 times fewer L1 nucleotides within their sequence compared with poorly expressed ones [40].

Host gene expression hindering was also showed for SINEs. Estécio et al. demonstrated that B1 insertions can repress the activity of downstream gene promoters *in vitro*. The mechanism of repression includes acquisition of DNA methylation and loss of activating histone marks. It was established that not all genes are equally sensitive to such repression [201].

Another supposed mechanism by which TE may decrease expression rate of nearby genes is trapping TFs, since many TEs have TF binding sites [2]. However, these sites may actually be used to promote proximal gene expression (discussed in the next Sections 1.8.2 and 1.8.3).

#### 1.8.2. TEs as a Source of Regulatory DNA Sequences

TEs rely on the host machinery for their transcription, and so they evolved regulatory sequences that mimic host *cis*-regulatory elements [12,24,197]. Analysis of chromatin accessibility maps for 41 human cell types derived from normal, embryonic, and cancer tissues showed that 44.1% of DNase I-hypersensitive sites (DHS), representing open chromatin, overlapped with TEs (LTRs, DNA transposons, LINEs and SINEs). DHS-associated TEs were enriched in chromatin states corresponding to promoters, enhancers, and insulators, defined previously by histone marks profiling [202].

Kellner and Makałowski scanned the data from the ENCODE project for active TF binding sites located in TE-originated parts of pol II promoters. More than 35,000 promoters in six tissues were analyzed, and over 26,000 of them harbored TEs [203]. Criscione et al. identified 988 sites within the human genome that may generate chimeric transcripts originating from the L1 antisense promoter, suggesting that this promoter contributes to the transcription of up to 4% of all human genes. The authors tested four chimeric transcripts experimentally and showed that they are detectable in normal human tissues [44]. Promoters for pol III-driven transcription may also be derived from TEs. Multiple Alu repeats were shown to serve as promoters for miRNA transcription in the C19MC miRNA cluster [204].

TEs are an abundant source of tissue-specific promoters. Faulkner et al. estimated that up to 30% of human and mouse capped transcripts initiate within repetitive elements. Transcripts that arise from these sites are generally tissue-specific [46]. Mätlik et al. screened GenBank data and characterized 49 chimeric mRNAs expressed from the L1 antisense promoter and overlapping with mRNAs derived from annotated genes. For three of these, tissue-specific expression in 16 different human tissues was confirmed experimentally [205]. More recently, Roller et al. used ChIP-seq to map histone modifications associated with regulatory activity in different tissues of 10 diverse mammalian species. The authors identified multiple TE-derived promoters and showed that tissue-specific active promoters are enriched with L1 transposons. In contrast, tissue-shared active promoters were enriched with L2 LINE elements [206].

TE-derived promoters may also orchestrate stage-dependent transcription [24,207,208]. For example, mouse endogenous retrovirus type L (MERVL) controls transcription at the

two-cell stage of the embryo. More than 50 genes transcribed at this stage generate chimeric transcripts linked to MERVL elements. MERVL is sharply induced at this stage and gets rapidly repressed at the next stage [115,208].

TEs contributed to the formation of new enhancers. Computational methods showed that MIR SINEs are highly enriched within putative enhancers genome-wide, that the chromatin environment of MIRs providing core enhancer sequences is similar to that of canonical enhancers, and that there are strong associations of MIR-derived enhancers with gene expression levels and tissue-specific gene expression [209]. It is also speculated that Alu elements may become new enhancers [210–212]. Alus upstream of transcription starting sites are enriched with epigenetic enhancer/enhancer-like marks (H3K4me1) in a tissue-specific manner, and the gain of these marks and TF binding motifs is positively correlated with the evolutionary age of Alus [212]. Canonical SVA sequence includes a known enhancer element [103].

Insulators are regulatory elements necessary for organizing eukaryotic chromatin into functionally distinct domains. Enhancer-blocking insulators prevent interactions between enhancers and promoters, and chromatin barrier insulators (boundary elements) separate active chromatin and heterochromatin. A common method to test the insulator activity of some sequences in vivo is an enhancer-blocking assay, where the tested sequence is cloned in a construct between an enhancer and a promoter that drives expression of a reporter gene. Many TE-derived insulators bind the CCCTC-binding factor (CTCF). Another insulation mechanism attributed to TEs involves binding pol III, or pol III-specific TF TFIIC [213]. CTCF ChIP-seq data for immune cells from human and mouse showed that 70 different types of TEs have contributed strongly to CTCF binding in both species, constituting ~35% of all CTCF binding sites. TE-derived sites are overrepresented among species-specific CTCF binding sites (regardless of whether the TE insertion occurred before or after speciation) [214]. A complex mechanism of insulation involving transcription of insulating sequence has been shown for rodent B1 SINEs. While basal insulator activity is maintained by pol III transcription, induced insulation involves release of pol III and engagement of pol II transcription on the same strand. CTCF also participates in this mechanism [215]. An insulation mechanism involving TE transcription was also reported for B2 SINE elements. In the murine growth hormone (GH) locus, a B2 repeat serves as a boundary, blocking the influence of repressive chromatin modifications. *Gh* gene is initially silenced in GH-producing cells, and TF PIT-1 activates it at a specific stage of embryonic development. Both pol II- and pol III-driven transcripts of the B2 element are required to restructure the regulated locus to facilitate this activation [216]. MIR SINEs share sequence characteristics with known pol III-dependent insulators. Wang et al. predicted 1178 MIR insulators in the human genome. Only six of them overlap with previously characterized CTCF barrier elements, suggesting that the mechanism of insulation for MIR elements is CTCF-independent. Three of the predicted MIR insulators were tested in an enhancer-blocking assay, and their insulator properties were validated [213]. Another ancient transposon family with insulator activity is MER20. Targeted ChIP assays demonstrated that MER20s bind CTCF and other known insulator proteins. It was shown that insulator-type MER20s are located between differentially expressed genes in human endometrial stromal cells [217].

TE copies distributed across the genome and containing similar regulatory elements are able to establish or rewire gene regulatory networks, coordinating the expression rates of multiple genes [218]. An impressive example of this is gene regulation in mammalian pregnancy [198,217,219,220]. Some TE-derived components of transcription regulation in placenta show signs of convergent evolution in different species [221]. Another example of a transcriptional network shaped by TE-derived enhancers is interferon response; Chuong et al. showed that ERVs have dispersed numerous IFN-inducible enhancers independently in diverse mammalian genomes. Deletion of a subset of these ERV elements in HeLa cells impaired expression of adjacent IFN-induced genes [222].

Many *cis*-regulatory elements in TEs evolved to have similar activity in multiple diverse species, making TEs a promising source for the development of new expression vectors in biotechnology [223].

### 1.8.3. Transcription Factors Binding Sites within TEs

Despite ubiquitous transcriptional repression of TEs, host TFs have been shown to directly activate expression of specific TE subfamilies, mostly relatively young retrotransposons. TEs appear to extensively contribute to mammalian transcription regulation, providing TF binding sites for a number of tissue-specific TFs [12].

Sun et al. performed a computational analysis of L1 transcriptional regulators in human cells using ENCODE ChIP-seq datasets and showed that 178 TFs are bound to L1 in at least one biological condition, with 138 TFs localized to the promoter [224]. Multiple TFs were experimentally confirmed to bind with human L1 elements and/or regulate transcription of nearby sequence: YY1, RUNX3, SOX factors [111,225], p53 [226], MYC, CTCF [224] and TCF/LEF [227]. Not all TF binding sites in L1 are located within its 5'-UTR; for example, a few overlapping SOX/LEF sites were found in ORF2 coding sequence, and these sites may control transcription of nearby genes [227]. Importantly, most of the TFs listed above are involved in neurogenesis [228]. TF binding sites may facilitate either activation or inhibition of L1 transcription; for some sites, both effects have been shown. The highly conserved YY1 binding site facilitates full-length L1 transcription, and, at the same time, is used by the host cell for L1 repression via DNA methylation. It is possible that YY1 recruits DNA methyltransferases directly [229]. Mutations in this site disrupt the accurate position of the transcription initiation complex [230]. One RUNX binding site in the human L1 5'-UTR has been proven to facilitate L1 transcription and retrotransposition [231]. There are two functional SOX binding sites in the human L1 5'-UTR, and experiments with a reporter gene system showed that different members of the SRY TF family have different effects upon binding these sites. SOX11 and SOX3 increased the reporter gene expression with different efficiencies, while SRY reduced the expression [232]. *Sox2* expression is inversely correlated with L1 expression during neuronal differentiation; a temporary decrease in *Sox2* expression is correlated with chromatin remodelling, allowing increased L1 transcription [225] (see also Section 2.4). There are controversial data about the role of p53 in L1 expression. In 2009, Harris et al. showed that the L1 promoter directly binds p53 and is positively regulated by it [226]. More recent studies brought the opposite results: p53 was shown to repress L1 transcription in various models. Apparently, p53 regulates the deposition of repressive histone marks to the L1 promoter [233,234]. One of L1 p53 binding sites responsible for L1 transcriptional repression is the same sequence that was earlier [226] reported to mediate p53-dependent transcriptional activation, suggesting that some additional factors modify the action of this site in vivo [234]. Kuwabara et al. demonstrated that some SOX binding sites within L1 overlap with TCF/LEF-binding sites involved in the canonical Wnt signaling pathway [235]. The regulation of L1 transcription by Wnt3a was confirmed experimentally [227]. Sun et al. demonstrated the involvement of Myc and CTCF in regulating L1 transcription [224].

Alu elements have dozens of predicted TF binding sites, some of these specific to certain Alu subfamilies. Most of these putative sites correspond to TFs involved in stress response or the development of specific tissues [91,236]. TFs that have been experimentally confirmed to increase transcription of Alu-containing sequences include SP1 [237], LXRE $\alpha$  [238], estrogen receptor [239], retinoic acid receptors [240] and NF- $\kappa$ B [241]. As was already mentioned, SINEs are transcribed by pol III. Interestingly, binding sites of pol III-associated TFs in SINEs were shown to participate in pol II transcription upregulation. Alu sequence contains A- and B-boxes recognized by TFIIC. Ferrari et al., studying cell cycle regulation in breast cancer cells, demonstrated that Alus located nearby pol II promoters may recruit TFIIC without pol III or other components of the pol III transcription machinery. The authors suggested a complex mechanism where specific Alus pre-marked by the ADNP protein recruit TFIIC; TFIIC then acetylates histone 3 lysine-18 (H3K18ac)

which increases chromatin accessibility, while also reorganizing chromatin loops by interaction with CTCF [242]. TFIIC-dependent regulation of pol II transcripts was also shown for murine SINEs in neurons, but in this case TFIIC is involved in transcription downregulation [243].

SVA elements contain many predicted TF binding sites: hormone response element (HRE) half-sites, a glucocorticoid response element, binding sites for SP1 [103], YY1/2 and OCT4 [244]. The ability of different parts of the SVA sequence to act as transcriptional regulators was confirmed experimentally [20,109]; however, no specific TFs were identified in these studies.

Ito et al., using ChIP-Seq datasets, identified putative TF binding sites in HERV sequences and predicted the binding of 84 TFs. Clustering analysis showed that HERV insertions can be divided into four groups: HERVs bounded to pluripotent TFs, to embryonic endoderm/mesendoderm TFs, to hematopoietic TFs, and to CTCF. It was also predicted that HERV regulatory elements tend to locate nearby and/or interact three-dimensionally with genes involved in immune responses [245]. HERV-K transcription was experimentally confirmed to be mediated by SP1, SP3 [246], YY1 and MITF-M [247], IRF1 and NF- $\kappa$ B [248].

#### 1.8.4. TE Insertions Modifying Coding Regions and Causing Formation of New Genes

In the overwhelming majority of cases, TE insertions affecting protein-coding regions are deleterious, since the most common results of this are frameshift or introduction of a premature stop codon. TE-encoded sequences were found only in 0.1% of functional proteins in the Protein Data Bank [249].

Nevertheless, there is a chance that TE-encoded sequence insertion within a coding region may create a new functional protein. There are so-called “orphan” genes that do not show homology to sequences in other species. The origin of such genes is often unclear, but apparently it often involves TE activity. It was shown that 53% of all orphan primate genes contain TEs. In primates, the most common mechanism of TE-mediated new gene formation is Alu exonization (see the next section), but other transposition events may also contribute to the formation of new genes [250].

Both L1 and SVA elements may bypass their own polyadenylation sites and use a downstream host site, taking some adjacent genomic sequence with them and generating readthrough transcripts that are then retrotransposed. This mechanism is called 3' transduction. Transduced host regions may reach a few kilobases in size, and they may contain coding sequences [104]. Moran et al., using artificial transposition reporter constructs in HeLa cells, confirmed that L1 retrotransposition potentially can mediate mobilization and duplication (shuffling) of exons [251]. Pickeral et al. screened 129 full-length L1 elements and showed that at least 10% of them have a putative 3'-transduced sequence. The authors extrapolated that the total amount of DNA transduced by L1 may constitute about 1% of the human genome [252]. Notably, because most L1 retrotranspositions are 5'-truncated, it is possible that some transduction events are not accompanied by L1 sequences at all (if a truncation happened during an insertion of a hybrid sequence before its L1 part was processed), and their L1-related origin would likely be unrecognizable [251].

At least 10% of SVA elements were found to have 3' transductions [104]. SVAs are also capable of 5' transduction: these elements recruit external upstream promoters to be mobilized, leading to mobilization of an upstream genomic region together with the SVA sequence. SVAs may also mobilize coding sequences by “exon trapping”, if an upstream exon is spliced to cryptic acceptor sites within the SVA sequence [107,253].

#### 1.8.5. Exonization of TEs Leading to the Formation of Novel Transcripts

Exonization is a TE insertion into an intron followed by the recognition of some part of the TE sequence as a new exon. This happens relatively often because many TEs carry potential splice sites [11,254]. An indirect proof that exonization is the most common mechanism of TEs entering coding sequences is the fact that the majority of TEs found in protein-coding regions are fragmented. TE-derived novel exons are usually “cassette

exons" that may be skipped as a result of splicing. Most TE-derived exons are included only in minor splice isoforms. In these cases, exonization does not destroy the function of the gene [2,255]. Analysis of recently formed coding cassette exons showed that about 80% of them create frameshifts or introduce premature stop codons [256]. Single-base mutations may change an alternative TE splice site into a constitutive one. A few hereditary diseases (Alport syndrome, mucopolysaccharidosis type VII, CCFDN syndrome) have been associated with constitutive inclusions of TE-derived exons [93]. Nevertheless, a substantial number of recently formed constitutive exons overlap with repetitive elements in the human genome [257].

Exonization mechanism is well-known for Alu elements. Their sequence contains multiple cryptic splicing sites, requiring a few mutations to become functional [2,11,93,256]. However, only two of these sites are usually selected in Alu exonizations. Most exonizations involve the right arm of Alu in antisense orientation [254,258]. Interestingly, Alu insertion into an exon may cause formation of a new intron (intronization) [259]. Using multiple genome comparisons, Zhang and Chasin found that 62% of recently evolved (primate-specific) cassette exons and 28% of recently evolved constitutive exons overlap with Alu elements [257]. Mandal et al. analyzed the complete human transcriptome and showed that 12.7% of all transcripts have Alu exonizations. Of these, 80% of transcripts have Alu exonizations within 3'-UTR [260]. A notable example of a gene harboring an Alu exonization is *ADAR2* encoding an enzyme performing RNA A-I editing (this mechanism is discussed in the Section 1.8.7). In human *ADAR2*, exon 8 is a new Alu-derived exon that is alternatively spliced in high inclusion levels in a tissue-specific manner. The new exon is inserted in the catalytic domain of *ADAR2* and alters its catalytic activity. Interestingly, a mouse-specific exonization occurred at the same position in the mouse *Adar2* gene [256].

The L1 sequence contains a few functional splicing sites [261], and L1-related exonization events were reported in mouse [262] and human [263] transcriptomes.

Comparative genomic analysis showed that an ancient LF-SINE element provided sequence for 19 different coding exons in tetrapod genomes (in the human genome, an LF-SINE-derived exon is found within *PCBP2* gene). All affected proteins have different functions and belong to different protein families. A total of 16 out of 19 LF-SINE-related exons are alternatively spliced; 11 of them introduce an early stop codon. For 3 of these exons, a regulatory function was suggested [264].

#### 1.8.6. TEs Introducing Alternative Polyadenylation Sites

Many retroelements have an A-rich tail, and a single point mutation in it may produce the polyadenylation signal, AATAAA. Thus, retroposition downstream of the coding region of a host gene could be a source of an alternative poly(A) signal. However, poly(A) signal is not sufficient to direct strong polyadenylation by itself; the poly(A) site region contains other *cis*-elements involved in the recognition of poly(A) signal by the polyadenylation complex [265]. Such elements may also be possibly introduced by a TE insertion. Over half of all human genes have multiple poly(A) sites. A total of 3188 human poly(A) sites from 2565 genes (corresponding to ~8% of all poly(A) sites and ~16% of all genes surveyed) were found to be associated with TEs. Notably, ~94% of human TE-associated poly(A) sites are nonconserved in mouse. Intragenic poly(A) sites associated with TEs are less frequently utilized compared with conserved sites at 3' ends of transcripts. Different TE classes are associated with poly(A) sites differently: most DNA transposons and LTRs contain a whole poly(A) site region, while many LINEs and SINEs are located either upstream or downstream of poly(A) sites, suggesting the contribution of *cis* elements [266]. The role of Alu elements in generating new poly(A) sites seems to be especially important. There are 107 Alu-derived poly(A) sites in the human genome; importantly, almost half of them constitute major or even unique poly(A) sites in a gene [267].

### 1.8.7. RNA A-I Editing

A very important RNA modification mechanism is adenosine-to-inosine RNA editing (A-I editing). ADARs (adenosine deaminases that act on RNA) target double-stranded RNA without any apparent sequence specificity and deaminate adenosines to produce inosines (recognized by most RNA-interacting molecules as guanosines). A-I editing may alter the sequence of a protein translated from an affected mRNA and/or create a new splicing site. Moreover, “hyperedited” mRNAs that contain many inosines are retained in the nucleus, providing a specific mechanism of gene silencing [11,268–270]. Alu sequence was identified as the primary target of A-I editing. By recent estimations, Alus account for >99% of A-I editing events found in humans [271]. Two nearby Alu repeats in opposite orientations easily anneal with each other, forming an extended region of dsRNA recognized by ADARs [11]. Additionally, it was found that editing of non-Alu sites is dependent on nearby edited Alu sites in the human transcriptome [272,273]. In mammals that do not have Alus in their genomes, other TEs serve as A-I editing targets (in mice, most A-I editing sites are distributed between B1 and B2 SINEs, L1 LINE, and MaLR LTR elements) [274]. Most mRNAs that undergo A-I editing are expressed in the nervous system [275] and are discussed in detail below, in the Section 2.3. Interestingly, ADAR2 mRNA is also subject to A-I editing, providing an example of auto-regulation [276,277].

### 1.8.8. Other TE-Related Mechanisms Altering the Fate of Transcripts

Back-splicing (circular splicing) is an unusual splicing mechanism that generates a circular RNA (circRNA) and a linear RNA with exon exclusion. It was recently discovered that circRNAs can bind miRNAs and regulate mRNA splicing and transcription [278,279]. Alu elements may facilitate circularization of pre-mRNA and back-splicing because two Alus in an opposite orientation on different ends of a pre-mRNA may easily form an imperfect duplex with each other [2]. Indeed, bioinformatic analysis revealed that circularized exons usually have long bordering introns containing complementary Alu repeats [280].

*Trans*-splicing happens when the splicing machinery joins splice donor and acceptor sites from two different transcripts. Clayton et al. proposed a potential mechanism of TE sequences facilitating *trans*-splicing events by providing complementary sequences for hybridization between pre-mRNAs from different loci. It has not yet been confirmed experimentally [281].

TE sequences within an mRNA may affect its degradation rate by different mechanisms. The poly(T) sequence in antisense Alu elements is the source of ~40% of identified 3'-UTR AU-rich elements (AREs), which regulate mRNA half-life through the competitive binding of proteins that stabilize or destabilize the transcript. Alu elements can also mediate intermolecular base pairing between two RNA molecules. The resulted dsRNA structures may be subject to Staufen-mediated mRNA decay [2]. Another mechanism by which TEs regulate mRNA degradation rate is miRNA-mediated decay and/or translational repression of mRNAs. The most common miRNA target site coincides with the most evolutionary conserved part of Alu [282].

TE-derived RNAs may compete for regulatory molecules with host mRNAs. At least one TE-rich transcript was confirmed to function as an “miRNA sponge” participating in complex gene network regulation [283].

TE insertions may provide alternative translation starts. For example, the insertion of an LTR element at the 5' end of the *MTH1* (*NUDT1*) gene generated three additional upstream in-frame start codons [284].

### 1.8.9. TE-Encoded Proteins and Non-Coding RNAs Used by the Host Cell

Proteins encoded by TEs are able to bind, cut, ligate or degrade nucleic acids, as well as interact with other proteins, and all these functions may be used by the host cell as well [3,34,198]. It was estimated that the human genome includes over 40 genes derived from DNA transposases [285] and at least 85 genes derived from the *gag* gene of LTR elements [198]. We will list some specific TE-related genes below.

A famous example of TE-derived protein-coding genes are the genes encoding the RAG1 and RAG2 recombinases, which catalyse the V(D)J somatic recombination in lymphocytes, generating a highly diverse repertoire of antibodies, immunoglobulins and T-cell receptors in vertebrates. The RAG1 gene has been derived from a *Transib* DNA transposon [286], and then its recombination activity was enhanced by the capture of a pro-RAG2 sequence [287,288]. In the context of the present review, it is interesting that RAG1 and RAG2 were shown to be expressed in neurons [289–291], even though the exact function of these recombinases in neurons is not clear.

The human *CENPB* gene encoding centromere-associated protein B is related to the *pogo* superfamily of DNA transposons [292,293]. CENP-B recognizes and binds a specific DNA sequence (CENP-B box). CENP-B lost its capacity to act as a transposase, although it apparently retains nickase activity, and this might contribute to the high-order structure of the centromeres by promoting recombination hotspots [294].

Syncytins involved in the development of placenta are encoded by ERV *env*-derived genes. Human syncytins-1 and -2 are essential for the formation and function of syncytiotrophoblast, the boundary layer between maternal and fetal tissue [295,296]. Syncytiotrophoblast is a fused cell layer, and Env proteins have membrane fusogenic capacity [296]. A conserved *env* immunosuppressive domain in syncytins is important for maternal immune tolerance during pregnancy [297]. Notably, domestication of *env* genes from ERVs of different families generating syncytin-like genes has occurred independently in different mammalian lineages [296,298].

Other TE-derived protein-coding genes expressed in the placenta and necessary for its normal development belong to the *MART* family (mammalian retrotransposon transcripts; another name of this family is *SIRH*, *sushi-ichi*-related retrotransposon-homologs). *MART* genes (at least 11 in placental mammals) are derived from the vertebrate Ty3/*gypsy* LTR family called *Sushi* (more specifically, they show similarity to the *sushi-ichi* element). *MART* genes are incapable of mobilization, having lost some structural features such as LTRs; the *pol* region is also completely or partially deleted in these genes. However, all *MART* genes carry a protein-coding region derived from the *gag* region of their ancestral retrotransposon [198,299,300]. It must be noted that genes of this family have many synonymous names used in different publications, which may be confusing. Some important examples of *MART* genes with placental function are *PEG10* (*RTL2*, *MART2*, *SIRH1*), *RTL1* (*PEG11*, *MART1*, *SIRH2*) [301], and *SIRH7* (*LDOC1*, *MART7*, *RTL7*) [302]. One gene of this family, *RTL4* (*MART4*, *SIRH11*, *ZCCHC16*) is expressed in the brain [303]; its possible function is described below in the Section 2.7.2.

New protein-coding genes can also be generated by a fusion between host and TE coding sequences. An example of this is the *SETMAR* gene encoding a protein involved in nonhomologous end-joining (NHEJ) DNA repair [304]. This chimeric gene is a result of fusion of the pre-existing *SET* gene with the *Hsmar1* DNA transposon [305].

Long non-coding RNAs (lncRNAs) are structurally similar to mRNAs but are not translated, instead participating in gene expression regulation. In striking contrast with mRNAs, most lncRNAs include TE-derived sequences [306,307]. HERV-H is a TE family particularly enriched in long intergenic non-coding RNAs (lincRNAs). A total of 127 HERV-H-lincRNAs exhibit dramatic stem cell-specific expression [306].

TEs have also provided coding sequences for multiple small silencing RNAs. piRNAs and siRNAs derived from TEs have been discussed above in the Section 1.7.4. miRNAs may originate from TEs as well. Piriyaopongsa et al. identified 55 experimentally characterized human miRNA genes that are derived from TEs. L2 LINEs, MIR SINEs and DNA transposons were enriched among TE-derived miRNA genes. One notable TE-derived miRNA is hsa-mir-566, which is Alu-related and has 1184 predicted target sites [308].

In a recent work by Percharde et al., an important role for the host cell gene expression regulation during embryonic development was confirmed for L1 RNA. This RNA is highly expressed and located in the nucleus in mouse ESCs. Surprisingly, KD of L1 RNA results in a dramatic decrease in ESC self-renewal, reduction in total RNA levels, and reduction

of rRNA synthesis. L1 KD ESCs showed upregulation of many genes normally expressed specifically at the two-cell stage (2C); the most important of these genes is *Dux* encoding a master TF. It was shown that L1 RNA interacts with *Dux* and rDNA loci. Experiments in vivo confirmed the importance of L1 RNA for normal embryonic development: the development of mouse embryos with L1 KD was arrested at the 2C. Thus, L1 expression is necessary for the exit from the 2C by repressing the transcription of genes controlled by *Dux* and activating rRNA synthesis to support rapid proliferation [115].

Even though ERV interactions with host immunity are commonly associated with pathologies [309–311], ERV-expressed products may also be used by host organisms to tune immune response mechanisms. While ERV proteins are presented to developing lymphocytes among other “self” peptides, their low expression in healthy tissues due to silencing leads to incomplete immunological tolerance to ERV-derived antigens. By mimicking viral infection, endogenous retroelements may thus provide “intrinsic adjuvant” for the immune response to poorly immunogenic targets [26]. In specific cases of cancer, overexpressed ERVs may serve as tumor-specific antigens recognized by the immune system, thus helping to regress tumors [295]. In addition, antisense ERV transcripts have been proposed to hybridize homologous RNA sequences of exogenous retroviruses, forming dsRNA molecules recognized by innate immunity machinery [297].

#### 1.8.10. TEs Participating in DNA Repair and Chromosome Maintenance

The interplay between TE activity and DNA-damage repair systems may be very complicated. While a TE insertion is a kind of DNA damage itself, so the host DNA repair systems must actively prevent it, it also seems logical that during the long co-evolution of TEs and their host genomes, mobile elements that are able to insert themselves more or less anywhere in the host DNA should eventually be used by the host cell to “patch” some damaged genome regions.

Noncoding parts of chromosomes that only have a structural function also get damaged, and restoring these using TE machinery would be a much simpler task for TEs than the repair of a coding sequence, since the exact sequence of inserted DNA would not be as important. Indeed, TE machinery is used to maintain telomeres. It remains debatable whether telomerases are domesticated retrotransposons; however, their phylogenetic and functional relation to TEs is clear [312–314]. The catalytic core of telomerase consists of the telomerase reverse transcriptase (TERT) and the integral telomerase RNA (TR). TR serves as a template for RT. Therefore, TERT, just like L1 ORF2p, is a reverse transcriptase that carries a specific RNA [315–318]. The amino-acid sequence of telomerase catalytic subunits is similar to that of viral reverse transcriptases [318]. Curiously, there are reports about telomerase being capable to reverse transcribe telomere RNA to internal chromosomal sites, a process quite similar to retrotransposition. Telomerase was also shown to bind RNAs different from TR [79,319]. The TERT-dependent telomere maintenance mechanism is conserved in most eukaryotes; however, dipteran genomes lack the *TERT* gene [320]. *Drosophila melanogaster* cells rely on three retrotransposons (HetA, TART, and TAHRE) to maintain telomeres. Therefore, in this species, not only TE-derived enzymes but also TE DNA are used for telomere elongation [79,314]. TE function in maintaining telomeres has also been suggested in some other eukaryotes [321,322]; on the other hand, even other species of the genus *Drosophila* have different TERT-independent telomere elongation mechanisms that rely on TE insertions to various degrees [323].

There are multiple reports about TE insertions at DNA break sites in coding regions, and in some cases these insertions contain retrotransposed host mRNA sequence. However, the possibility of TEs providing means for the accurate repair of genes is still unclear. RNA-templated DNA repair is a recently discovered repair mechanism [316,324]. It is apparently coupled with transcription [325,326], but the key process in this mechanism must be RT, and currently it is not well understood how it happens. RT may possibly be performed by DNA-dependent DNA polymerases [327], and the role of such a mechanism in DNA damage repair was proposed [328]. However, mammalian cells express three specialized

reverse transcriptases: TERT [315], L1 ORF2p [34] and the enzyme encoded by ERV *pol* gene [3]; all these enzymes have a TE-related origin.

Probably the earliest study of TE-mediated DNA repair was performed in 1996 on yeast transformed with plasmids encoding various TE-encoded reverse transcriptases, including the L1 enzyme [329]. A more recent study showed that an endogenous mechanism of RNA-dependent DNA repair exists in yeast, that it normally happens using a cDNA intermediate, and that the reverse transcriptase of the yeast transposon Ty performs the RT [330]. Ono et al. introduced DSBs into mouse zygotes and NIH-3T3 cells by the CRISPR/Cas system and identified long de-novo insertions at DSB-targeted sites. It was shown that DSBs are repaired by sequences deriving from retrotransposons, genomic DNA, mRNA and sgRNA (single guide RNA used in the CRISPR/Cas method). Some insertions derived from intronless mRNA sequences, indicating the involvement of RT. RT inhibitors prevented mRNA- or retrotransposon-derived insertions into the DSB sites [331].

Canonical homologous recombination (HR) DNA repair mechanism uses a sister chromatid as a matrix for repair, and so HR only works during S or G2 cell cycle phases [332]. Wei et al., using U2OS osteosarcoma cells in G0 or G1 phases, showed that HR factors, including RAD52, are recruited to transcriptionally active damage sites. The authors suggested that RNA may be used as an HR template instead of DNA, and experiments with RNase H treatment confirmed that RNA is required for this mechanism [325]. This suggests that one of the TE-encoded reverse transcriptases may be involved in such repair. The same research group then discovered a similar repair mechanism in neurons [333] (discussed in the Section 2.6).

Human L1s were shown to be able to integrate at endogenous DNA lesions in cells defective in the NHEJ repair pathway; experiments with mutant L1 forms confirmed that such integration is endonuclease-independent (ENi) but needs L1 reverse transcriptase function to be intact. ENi L1 retrotransposition was demonstrated to generate unusual insertions: L1 copies inserted at preexisting DNA lesions were truncated predominantly at their 3' ends, lacked target site duplications, and two out of nine examined insertions contained host cDNA fragments at their 3' ends [64]. The same research group later showed that ENi L1 retrotransposition can use dysfunctional telomeres as integration sites [79,317]. To check whether ENi L1 integration is possible in vivo, Sen et al. scanned the human genome and identified 21 young non-classical L1 insertions (NCLI) distinguishable from TPRT-mediated ones. A total of 15 out of 21 NCLI loci involved additions of non-L1 DNA segments (up to 312 bp) to the L1 DNA. At two NCLI loci, fragments of cellular RNAs appeared to have been reverse transcribed along with the L1 RNA [61]. Similar observations were made later for Alus (by the same research group). A total of 23 non-classical Alu insertions (NCAI) that have signs of ENi integration were found. Out of these, 13 NCAI loci included non-Alu DNA sequences (the largest one being about 2 kb), and 2 NCAI loci had signs of cellular RNAs apparently being transcribed along with the Alu fragment. In 5 NCAI loci, capture of another TE RNAs along with the main Alu sequence was detected [334]. It must be noted that the methods used in the two latter papers only allowed to reveal germline non-classical TE insertions. We may speculate that in somatic cells ENi insertions are more common.

## 2. The Role of TEs in the Normal Function of Neuronal Tissue

Both germline and somatic TE insertions were found to have functional roles in neurons and neural precursor cells (NPCs). These mechanisms are summarized in Table 2.

**Table 2.** Functions of exapted TEs in neuronal tissue.

Function	Cell Type	Type of TE Insertions	Section Number
A source of mutagenesis contributing to neuronal mosaicism	Neurons and NPCs	Somatic	2.2
RNA A-I editing of neuronal transcripts	Neurons, possibly NPCs	Germline	2.3
Facilitating neuronal differentiation via host gene transcription regulation	NPCs	Germline	2.4
Possible TE participation in neuronal plasticity	Neurons, possibly NPCs	Somatic	2.5
Possible TE participation in DNA repair	Neurons	Somatic	2.6
TE-derived <i>cis</i> -regulatory sequences in neuronal gene expression	Neurons and NPCs	Germline	2.7.1
TE-encoded neuronal proteins	Neurons, possibly NPCs	Germline	2.7.2
TE-encoded neuronal ncRNAs	Neurons and NPCs	Germline	2.7.3

### 2.1. Somatic Mosaicism in Neurons: Its Sources and Possible Functions

Somatic mosaicism is a feature of a normal brain [72,275,335–339]. Neurons of the human brain are extremely diverse. Even compared with the immune system, cellular diversity in the nervous system is estimated to be twice as high. Some classifications suggest that there might be up to 10,000 different types of neurons. Neurons with similar morphologies may differ in gene expression and function. Neuronal diversity is generated by a combination of mechanisms acting on different levels: DNA (mutations and different activity of promoters), RNA (alternative splicing and polyadenylation, RNA editing) and proteins (post-translational modifications). At least some of these mechanisms, such as RNA editing and alternative polyadenylation, seem to be more common in the CNS than in other tissues [275]. In this section, we will discuss neuron-specific genomic mutations.

Neurons in general have the highest quantity of “acquired” DNA compared to other cells. Frontal cortex neurons were found to have about 4% more DNA in their genomes compared to lymphocytes. In the same study, it was shown that neurons in the frontal cortex have more DNA compared to the cerebellum of the same brain [340]. A single neuron may contain in its genome at least 800 single-nucleotide variants (SNVs) [341] and 1–2 mutations caused by random L1 integrations [342]. Single-cell sequencing of human frontal cortex neurons revealed that 13–41% of them have megabase-scale de-novo copy number variations (CNV), with deletions being twice as common as duplications. A subset of neurons have highly aberrant genomes with multiple alterations [343]. A noticeable fraction of neurons in the normal brain even have aneuploidy. When specific chromosomes were labeled in adult brain samples by in-situ hybridization, it was found that about 1% of murine cortical neurons have loss or gain of X or Y chromosomes [344], while, in the human brain, chromosome 21 aneuploid cells (including non-neuronal cells) constitute ~4% of all cells [336]. Importantly, experiments with mice demonstrated that aneuploid neurons are not dying cells, but are actually functionally active and integrated into the normal brain circuitry [339].

Neuronal genomic diversity is caused by multiple factors. Neurons are the most long-living cells in the human body [345]; therefore, during their life, multiple mutations inevitably accumulate in their genomes. However, it is important to specify that many neuron-specific mutations actually happen very early, even before the full differentiation of the cell. NPCs undergo multiple double-strand breaks and genomic recombinations (with some long genes being particularly affected) [346,347]. Spectral karyotype analysis demonstrated that a striking 33% of proliferating neural precursors in the brain of a normal mouse embryo have aneuploidy (most of these cells lacked one chromosome). Apparently, most of these aneuploid cells undergo apoptosis, since the level of aneuploidy in adult mouse cortex was much lower [344]. Indeed, it was shown that during embryonic

neurogenesis only 15–40% of post-migratory cells survive, suggesting that some kind of selection takes place [275]. In mature neurons, one of the most important causes of mutations is activity-induced DSBs. DSBs happen in neurons as part of their normal function [348]. Such DSBs are not random; they form at specific locations in the genome. DSBs in the promoters of a subset of immediate early genes are necessary to trigger their fast expression by resolving topological constraints [349].

## 2.2. Transposable Elements Are an Important Source of Neuronal Genetic Mosaicism

The brain is the only known somatic tissue where retroelements are de-repressed throughout the life of a healthy human [350]. There are multiple reports [7,39,47,72,174,225,351–355] about NPCs and neurons being more permissive to retrotransposition than other types of cells in mammals (only embryonal [351] and tumor cells [356] have retrotransposition rates comparable to that in neural tissue). These reports are described at length below.

It must be considered that even if transposition were equally frequent in all cell types, neurons would likely still be the most affected (in the sense of TEs having an effect on cell function). Firstly, neurons express a larger fraction of their genome compared to most other cell types. The initial estimates were that up to 50% of all protein-coding genes in the human genome are expressed in the brain (with only kidney and testes having comparable transcriptome complexity) [357]. However, more recently, it was confirmed that up to 95% of protein-coding genes in the genome are expressed in at least one human brain region during at least one period of development or adulthood [358,359]. Secondly, many long genes are expressed in the brain or involved in neuronal-specific processes [7,346]. Recent studies have demonstrated that transcripts of genes longer than 100 kb are highly enriched in neurons [360–362]. Long neuronal genes have an increased susceptibility to random TE insertions simply because of their size [7]. Interestingly, *NF1*, a very rare example of a human gene containing specific hotspots for L1-mediated insertions, is quite long, about 280 kb [363]. While *NF1* is not strictly neuron-specific, it has distinct functions in neurons [364]. TE insertion hotspots within *NF1* were associated with a disease [363] described in the Section 3.2.18.

TE activity (L1 transposition in particular) is increased in neural tissue. Experiments with the L1-EGFP retrotransposition reporter demonstrated that engineered human L1s can retrotranspose in adult rat hippocampus NPCs *in vitro* and in the mouse brain *in vivo* [225]. Later, L1 mobilization was similarly confirmed in NPCs isolated from human fetal brain and NPCs derived from hESCs [352]. Macia et al. found that L1 expression and engineered retrotransposition is much higher in NPCs than in mesenchymal (MSCs) and hematopoietic (HSCs) somatic stem cells. They have also demonstrated for the first time that engineered L1s can mobilize in mature, non-dividing neurons [354].

Similar observations were made for endogenous L1 activity *in vivo*. Coufal et al. detected an increased copy number of endogenous L1s in several regions of adult human brain compared with heart and liver samples [352]. Similar results were obtained by Jacob-Hirsch et al.: whole-genome sequencing of human samples (including different brain regions, peripheral blood, and lymphoblastoid cell lines) showed that the number of retrotranspositions in brain tissues is higher than in non-brain samples. Most identified somatic brain transposition events are L1 insertions. Interestingly, the majority of these insertions were shown to be nested in pre-existing L1 elements. These nested insertions usually had signs of ENi transposition [39]. On the protein level, ORF1p expression in different brain regions was confirmed by immunohistochemistry of postmortem brain samples. Notably, kidney, liver, lung and heart samples showed little to no ORF1p expression in this study [365].

A number of studies using single-cell sequencing approaches have confirmed that among all somatic transposition events in the human brain, L1 insertions and L1-mediated deletions are the most frequent. However, due to methodological variation, the estimated rate of L1 insertions in these studies varies considerably: from one insertion per 25 neurons to 13.7 insertions per a single neuron [70,353,366]. Importantly, experiments that calculated

the lower bound [366] have higher rates of validation. Even with the lowest estimate, considering the enormous total number of neurons constituting the human brain, there must be tens of billions somatic retrotransposition events within an individual brain. In addition, taking into account the highly networked structure of the brain, a mutation in a single neuron may potentially alter the function of many other neurons, causing far-reaching effects on neuronal circuitry [7,116]. Indeed, it was found that severe neuropathologies may be caused by some somatic mutations with low levels of mosaicism (mutation detected in 10% of blood cells, or as low as 8% cells in the brain samples). This may be explained by cell migration during the brain development. For example, cerebral cortical progenitors generate cells that migrate to the outer layers of the cortex, so cortical neurons with the same somatic mutation would likely be intermingled with neurons without this mutation. Importantly, glial cells also have complex migration patterns [338]. It must be noted that the reports about endogenous L1 activity in glial cells are somewhat controversial [116]—while some studies showed that glial cells harbor less L1 insertions compared to neurons [225,353], other authors found no significant difference between the number of somatic L1-associated variants (SLAVs) in these cell types. SLAVs include both L1 insertions and retrotransposition-independent L1 endonuclease-mediated mutations; surprisingly, the latter happens about as often as the former [70].

It was found that TE insertions in the brain are not only common, but also prone to target neuron-specific genes. Baillie et al. developed a high-throughput protocol called retrotransposon capture sequencing (RC-Seq) to map individual retrotransposition events. The authors applied this method to identify TE insertions in human hippocampus and caudate nucleus samples. It was demonstrated that TEs mobilize to protein-coding genes that are differentially expressed and active in the brain. Protein-coding loci were disproportionately affected by L1 and Alu insertions compared to random expectation and compared to prior germline frequencies. Genes containing somatic intronic L1 insertions were twice as likely to be differentially overexpressed in the brain compared to random expectation. Loci with such insertions were associated with neurogenesis and synaptic function: among them were tumor suppressor genes deleted in neuroblastoma and glioma, genes encoding dopamine receptors and neurotransmitter transporters. Somatic Alu insertions were not overrepresented in introns but were even more common in exons than somatic L1 insertions [175].

New approaches also allowed researchers to pinpoint specific TE variants in the brain. Sanchez-Luque et al. performed a single-cell genomic analysis of human hippocampal neurons, identified and validated one specific somatically active hot donor L1 element. This particular L1 lacked the YY1 binding site and was highly mobile when tested in vitro [229].

Not only mammalian brains have been shown to contain genetically diverse neurons with relatively high TE activity. Perrat et al. performed cell-type specific gene-expression profiling of *Drosophila* neurons and demonstrated that TE expression is more abundant in mushroom body  $\alpha\beta$  neurons than in neighboring neurons. Deep sequencing identified over 200 de-novo TE insertions in  $\alpha\beta$  neurons [367]. However, a more recent study found no evidence for TEs being mobile in the fly brain. Even though increased TE expression in the  $\alpha\beta$  neurons was confirmed, TE sequences were not found to accumulate in neuronal genomic DNA with age [368], making the subject of TE-related neuronal mosaicism in the fly brain controversial.

### 2.3. RNA A-I Editing in TE Sequences Contributes to Neuronal Somatic Mosaicism on RNA Level

The A-I editing described in the Section 1.8.7 has been confirmed to supply an additional level of functional modulation to neuronal genes. A-I editing introduces codon changes in mRNAs and may generate specific protein isoforms that do not correspond to genomic DNA sequences. Disruption of this mechanism was associated with multiple disorders, primarily neurological or psychiatric [275,369–372]. Mammals have three A-I editing enzymes; ADAR3 is restricted to the brain, while ADAR1 and ADAR2 are preferentially expressed in the nervous system [373]. A-I editing is subject to complex

spatiotemporal regulation within CNS: different regions of the brain show up to three-fold differences in A-I editing levels [11], and A-I editing efficiency for both coding [374] and non-coding RNAs [375] increases during murine brain development. Brain-related genes in the human genome are enriched with intronic Alus and have an increased propensity for mRNA editing [7]. Most targets for A-I editing are found in the nervous system, and these mRNAs encode such crucial neuronal proteins as voltage-dependent potassium channels [370,376], various glutamate receptor subunits [268,275,370,376], GABA receptors [7], and serotonin receptors [275,369,370]. The lack of A-I editing of mRNA encoding the AMPA receptor subunit GluA2 in mice causes drastic phenotype changes, with severe epilepsy and premature death within 3 weeks after birth. Mutated receptors have increased rate of tetramerization and provide increased  $\text{Ca}^{2+}$  influx [268,370,376]. Remarkably, the defects of the ADAR2 KO animals are completely rescued by replacing both WT GluA2 alleles with ones encoding the edited codon, even though A-I editing of many other sites is not restored [268]. In humans, deficient GluA2 RNA A-I editing was linked to spinal motor neuron degeneration in amyotrophic lateral sclerosis [377].

#### 2.4. TEs and Cell Differentiation in Neurogenesis

In different works, the L1-EGFP reporter system has been shown to readily mobilize in NSCs, NPCs, and post-mitotic neurons [116]. However, L1 derepression seems to be specifically associated with neuronal differentiation.

Kurnosov et al. applied next-generation sequencing to directly compare the rate of L1 and Alu somatic insertions in neurogenic and non-neurogenic regions of human adult brain. The percentage of somatic L1 insertions from the total number of reads in the corresponding library was approximately equal for all studied regions except the neurogenic dentate gyrus (DG), where it was about 1.5 times higher. The highest percentage of somatic Alu insertions was also observed in the DG. In this region, Alu insertions in genes happened more often than predicted, while Alu insertions in promoters had lower than the expected rate. Another neurogenic region, the subventricular zone (SVZ), was not found to have unusual TE insertion distribution like the DG; however, there is some data that in humans, unlike rodents, the neurogenesis in the adult SVZ is negligible [174].

L1 disinhibition happens during neuronal differentiation in vitro. Erwin et al. performed artificial differentiation of hESC-derived hippocampal NPCs and showed highly upregulated L1 ORF2 expression during initial stages of neuronal differentiation [70]. Muotri et al. transfected rat hippocampus neural stem (HCN) cells with an L1-EGFP retrotransposition indicator construct with a puromycin resistance gene. Selected puromycin-resistant clones did not express EGFP at the onset of the differentiation experiment. The authors selectively stimulated HCN differentiation into neurons, astrocytes or oligodendrocytes. EGFP expression was registered only during neuronal, but not glial differentiation. EGFP could be detected as early as 2 h after the induction of neuronal differentiation, indicating that it might have resulted from the derepression of already integrated but silenced *EGFP* rather than from additional retrotransposition events. The authors then performed a mixed differentiation experiment and discovered that most HCN clones harboring detectable L1 retrotransposition events had a tendency to differentiate into neurons rather than glial cells. L1 activation in HCN cells correlated with decreased expression of *Sox2*. Moreover, recruitment of SOX2 and HDAC1 on the endogenous rat L1 promoter region in undifferentiated HCNs was confirmed by ChIP-PCR. The authors suggested that SOX2 represses L1 prior to neuronal differentiation, and release of SOX2 leads to restructuring of chromatin, allowing L1 expression. It was also noted that some of the L1 insertions in the selected HCN clones occurred within neuron-specific genes [225]. In their consequent work, Muotri et al. showed that the L1 promoter undergoes demethylation during neuronal differentiation. ChIP-qPCR showed high levels of MeCP2 in association with endogenous L1 promoter regions in rat NSCs compared to neurons generated from these cells. Specific CpG sites within the L1 promoter were found to demethylate during

neuronal differentiation. MeCP2 was also associated with ORF2 but this association did not change during differentiation [130].

Some data suggest that L1 derepression during neuronal differentiation reflects L1 *cis*-regulatory sites being used to promote transcription of nearby genes. Jönsson et al. performed a KO of *DNMT1* gene in fetal-derived human NPCs to cause global DNA demethylation. This resulted in transcriptional activation of evolutionarily young, hominoid-specific L1 elements. The authors also established that protein-coding genes located within 50 kb of activated L1 elements were upregulated; most of these genes are normally upregulated upon neuronal differentiation of NPCs. Gene Ontology analysis confirmed that these genes were enriched for functions such as synaptic transmission and cell communication. Many of the activated L1 elements gave rise to fusion transcripts, confirming their role as promoters [378]. Kuwabara et al. discovered that L1 elements in rat adult hippocampal NSCs are subject to regulation by the same TFs as *NeuroD1* (a proneural TF essential for the CNS development and adult neurogenesis). There are a few SOX2 and TCF/LEF binding sites within the *NeuroD1* promoter, including overlapping SOX/LEF-binding sites that may bind both of these factors. SOX2 and SOX/LEF binding sites are necessary for expression downregulation by FGF2, while TCF/LEF and SOX/LEF binding sites are necessary for expression upregulation by Wnt3a. Importantly, the L1 sequence also contains multiple SOX/LEF-binding sites, not only in its 5'-UTR but also in the ORF2 coding region. In reporter gene assays, the ORF2 sequence demonstrated promoter activity in both forward and reverse orientations in cultured NSCs undergoing neuronal differentiation; this activity was the highest 1 d after neuronal induction. L1 promoter activity in differentiating neurons was then confirmed *in vivo* in an experiment with a lentiviral vector encoding ORF2 fragment fused with *EGFP* gene. This virus was injected into the DG of young adult rats. Notably, EGFP expression observed after the infection was restricted to neurogenic areas, and EGFP-positive cells contained markers of NPCs and newborn neurons. These data indicate that both *NeuroD1* and L1 expression are specifically induced during NSCs transition to newborn neurons. The authors scanned human, mouse and rat genomes for L1 elements near genes and identified a few genes expressed in neurons that may possibly be subject to transcription regulation by Sox/LEF sites within proximal L1 sequences [227].

Importantly, some specific TEs need to be downregulated during neuronal differentiation [32]. Primate-specific transcription repressors ZNF417 and ZNF587 silence HERV-K and SVA elements in hESC. Neurons derived from ZNF417/587-depleted iPSC were characterized by the aberrant expression of nonneuronal genes and upregulation of transcripts related to potassium channel activity or to GABAergic neurotransmission. Another important observation in this study is that six of HERV-K elements predicted to encode neurotoxic Env proteins were upregulated in iPSC-derived neurons as a result of ZNF417/587 KD [379]. Accordingly, it was shown that downregulation of one specific *env* gene within an evolutionary young HERV-K subtype HML-2 in iPSC resulted in the dissociation of stem cell colonies and enhanced differentiation along neuronal pathways [380]. The same HERV-K subtype was studied by Padmanabhan Nair et al. They used a CRISPR-based approach to directly upregulate several distinct HML-2 promoters in hESCs. The authors differentiated transgenic and control hESCs into cortical neurons and estimated their properties at day 60. Neurons produced from cells with activated HML-2 transcription exhibited altered morphology and a drastic reduction in levels of MAP2a/b (neuron-specific cytoskeletal proteins). Forebrain organoids generated from transgenic hESCs showed a disrupted layer structure and were smaller in size compared with organoids produced from control hESCs. RNA-seq showed that HML-2 overexpression upregulates neuronal development-related genes in cortical neurons. Interestingly, dopaminergic neurons generated from control or transgenic hESCs had no difference in morphology or MAP2 expression at day 50 [381]. Skariah et al. confirmed the importance of *Mov10* RNA helicase in brain development. They found that *Mov10* level is elevated in the postnatal mouse brain and that *Mov10* KO has an embryonic lethal phenotype, while heterozygous *Mov10*<sup>+/-</sup> mice have disrupted dendritic arborization, increased anxiety and hyperactivity. Sequencing experiments showed that

Mov10 binds multiple RNAs involved in neuronal development and neurite outgrowth. However, Mov10 is also a strong suppressor of L1 mobilization. In brain samples from *Mov10*<sup>+/-</sup> mice (postnatal day 2), ORF2 genomic copy number was two times higher than in WT samples [172].

Even though most studies of TE derepression in neuronal tissue cited above either focused specifically on NPCs or did not discriminate between neurons and NPCs, a few recent works showed that postmitotic neurons permit L1 expression and/or mobilization. Macia et al. reported mobilization of engineered L1s in neurons [354]. Salvador-Palomeque et al. measured L1 promoter methylation at multiple points of neurodifferentiation in cells generated from a cultured hiPSC line. Surprisingly, a reduction in methylation was observed in day 112 neurons compared to day 72 neurons. (However, the methylation level was increased again in day 156 neurons [355].) Erwin et al. differentiated hESCs to hippocampal neurons and measured endogenous ORF2 expression at different stages. ORF2 level was low in NPCs, then it increased at days 1–2 of in vitro differentiation, diminished by day 5 and peaked in post-mitotic neurons (day 28) [70].

### 2.5. Specific TE Regulation in the Hippocampus: Is There a Role for TEs in Learning and Memory?

The hippocampus is the key brain region participating in learning and memory that also contains a site of adult neurogenesis [382]. This region also plays an important role in brain reaction to stress and early-life experiences [383,384]. The hippocampus seems to stand out when somatic retrotransposition rate is considered. While L1 retrotranspositions are more frequent in the mouse brain compared with non-neuronal tissues, in the hippocampus L1 activity is the highest among brain regions [385]. Single-cell RC-Seq showed that genes preferentially expressed in the hippocampus are significantly enriched for intragenic somatic L1 insertions in both hippocampal neurons and hippocampal glia. Curiously, no similar enrichment was found for cortical neurons. In the hippocampus samples, transcribed enhancers active in NSCs were also enriched for L1 insertions [353]. These observations suggest some functional significance of TE insertions specifically in the hippocampus.

Importantly, in many studies examining the hippocampus in the context of learning retrotransposition may be overlooked, since de-novo DNA synthesis registered in this structure is usually attributed solely to replication necessary for neurogenesis [386]. There is currently only a few published works in which the role of retrotransposition in learning and memory was specifically addressed, and while some correlations between TE activity and memory formation were confirmed, molecular mechanisms underlying these correlations are still poorly understood. Changes in TE activity specifically in the hippocampus were also reported to take place after stress or voluntary exercise.

A link between learning and reverse transcriptase activity in the hippocampus was first proposed quite a long time ago, in 1983, even though the methods used in this work were not accurate enough by modern-day criteria [387]. In the study performed by Kokaeva et al., a mixture of two antisense probes, one binding with the reverse-transcriptase-encoding region of L1 and the other binding with a unique L1 region, was used. Intrahippocampal injection of this mixture to rats was performed 6 h before training in a Morris water maze. Short-term memory (STM) and long-term memory (LTM) formation were assessed 30–40 min and 48 h after the training, respectively. While no difference in performance between the groups was registered during the training session or 30–40 min later, animals treated with anti-L1 oligos showed impaired results in the memory retention test 48 h after the training [388]. Bachiller et al. obtained similar results in mice using a different learning paradigm. Mice were systemically injected with lamivudine (RT inhibitor) just after passive avoidance (PA) training. Lamivudine did not affect the STM formation estimated 1 h after the training; however, at 72 h after the training, lamivudine-treated mice showed impaired LTM formation. The functional role of L1 expression in LTM formation was confirmed by intrahippocampal injection of antisense probes binding with ORF1 or ORF2; both probes impaired LTM in the PA paradigm and in the novel object recognition test. In

another series of experiments, these authors demonstrated a neuronal-activity-dependent activation of L1 elements in murine hippocampus after exposure to a novel environment. A transient increase in both ORF1 and ORF2 mRNA levels was observed at 1 h after the novel exploratory session. In the genomic DNA extracted from the same hippocampi, a permanent increase in the number of ORF2 insertions, but not ORF1 insertions, was registered 1 h after the exposure. Systemic administration of a lamivudine (immediately after the exposure) or glutamatergic antagonist cocktail (before the exposure) both blocked de-novo ORF2 insertions. No exact mechanism of L1 participation in memory-formation processes was found in this study, but it was speculated that L1 insertions may contribute to LTM formation by generating neurons with unique genomes and, therefore, long-lasting distinct transcriptomes by altering expression of nearby genes and/or the fate of their transcripts. However, it was shown that lamivudine administration had no effect on the exploratory-dependent expression of immediate early genes *Fos* and *Egr1*; therefore, the genes that are possibly targeted by neuronal activity-dependent L1 insertions are yet to be discovered [389]. Bedrosian et al. reported a correlation between early life experience and L1 retrotransposition in murine hippocampus. It is known that L1 activity increases during neuronal differentiation, and during the first week of life the hippocampus still undergoes extensive cell division and differentiation in mice. Thus, the sensitivity of L1 transposition to experience may be detected at this stage. The authors did not use artificial training paradigms, but compared animals who received high or low maternal care. These two “maternal styles” were identified by monitoring the behavior of animals in their home cage. It was found that pups reared with low maternal care had more L1 copies in their hippocampi. Copy numbers of other murine mobile elements did not correlate with maternal care. There was no such correlation for L1 copy number in the frontal cortex or heart. Cross-fostering experiments showed that L1 3'-UTR copy number in the hippocampi of rats at the age of 21 days correlated more with the maternal care level of their foster mother than that of their biological mother. This work shows that there is neuronal plasticity at the level of the DNA sequence in response to environmental changes at early age; however, physiological or pathological mechanisms in which such plasticity may be involved are not exactly clear [385].

Low maternal care may be considered a chronic stress exposure. In that context, the increase in hippocampal TE activity reported in [385] is in line with the general tendency of TE activation as a result of stress on cellular or organismal levels [11,33,36,176–179,390]. However, another research revealed a retrotransposon silencing response in the hippocampus following a stress exposure (specifically, acute restraint stress). Hunter et al. first observed an overall increase in the repressive histone H3 lysine 9 trimethylation (H3K9me3) level in the hippocampus of rats subjected to restraint. Notably, this effect was hippocampus-specific; there were no changes in H3K9me3 level in cerebellum, frontal cortex, heart, liver, and skeletal muscle. More detailed screening with ChIP-Seq to determine the genomic localization of the H3K9me3 response showed that, after the stress exposure, H3 methylation level in repetitive regions increased more prominently compared with its increase in genes or intergenic regions. RT-qPCR was then performed to estimate the expression of different TEs; remarkably, while hippocampal RNA levels of B2 and ERV elements were reduced as a result of stress, L1 RNA level was not; therefore, it seems like TE repression by methylation spared L1 elements. The authors point out that, previously, H3K9me3 was associated with ERV but not LINE elements [196]. In another study, the changes in L1 expression after a stress exposure were only observed in the hippocampus, but not in prefrontal cortex or amygdala [391] (this paper is also discussed in Section 3.1.3).

Muotri et al. generated L1-EGFP transgenic mice with an L1 indicator cassette that activates the expression of the EGFP reporter only after somatic retrotransposition, and provided them unlimited access to running wheels. The authors found that EGFP expression in granular cells in the DG of the hippocampus is increased in mice exposed to voluntary exercise when compared with sedentary animals (with no access to running wheels). The authors suggested that NPCs in adult brain may support de-novo L1 mobilization upon

exposure to a new environment. Interestingly, it was noted that EGFP expression in non-neurogenic brain regions also increased as a result of exercise [392]. This is in agreement with the results reported by Ivashkina et al., who observed increased DNA synthesis in DG and in non-neurogenic brain regions in mice 3 days after PA training. This study did not identify the source of the new DNA, but retrotransposition was one of the supposed mechanisms [386].

### 2.6. *Is It Possible That Neurons Use TEs for RNA-Templated DNA Damage Repair?*

Neurons and NPCs are especially prone to DSBs compared to other cell types. Two kinds of genes are particularly affected by this: immediate early genes, the expression of which was shown to be regulated by promoter breaks [349,393], and genes with extremely long introns and multiple small exons [346,347]. Neurons have to live for many years, and the integrity of their genome needs to be constantly maintained to make this possible. Accumulation of unrepairable or incorrectly repaired DNA lesions was linked with neuropsychiatric diseases [394] and neurodegeneration [395]. Repair mechanisms such as NHEJ, DNA mismatch repair (MMR), base excision repair (BER), and nucleotide excision repair (NER) (including transcription-coupled repair (TCR, or TC-NER)), were confirmed to be necessary for normal neuronal function [396–398]. However, HR, the most accurate mechanism of DSB repair, is considered to be impossible in neurons, since HR may happen only during the S or G2 phases [332]. This still leaves the opportunity of RNA-templated recombination that was discovered not long ago. This mechanism in yeast [329,330] and the U2OS cell line [325] was described above, in Section 1.8.10. The same research group that studied this mechanism in U2OS cells then continued investigating DSB repair in neurons. Welty et al. observed that in neurons, just like in U2OS cells, RAD52 is recruited to the sites of DNA damage and that this process is dependent on the presence of nascent mRNA generated during active transcription; both RNase H treatment or transcription inhibition prevented RAD52 recruitment. Electrophoretic mobility shift assay demonstrated that RAD52 binding affinity to ssRNA is even higher than its affinity to ssDNA. Furthermore, RAD52 had a high binding affinity to R-loops, the damage-prone structures at transcriptionally active sites in the genome. The authors also showed that high concentrations of amyloid  $\beta$  ( $A\beta$ ) inhibit both the expression of RAD52 and its association with DNA damage sites. Importantly,  $A\beta$  had a similar effect on RAD52-dependent repair in neurons and U2OS cells, suggesting that the same repair pathway functions in dividing and non-dividing cells [333]. All these results, taken together, provide conclusive evidence that RAD52 participates in RNA-templated DNA repair in neurons. Nevertheless, in this study, the reverse transcriptase responsible for RNA-dependent DNA synthesis was not identified. We may speculate that, since neurons were shown to have increased L1 activity [353,354], L1 ORF2p is a likely candidate. Importantly, experiments with non-neuronal cells demonstrated that L1 retrotransposition happens mainly during the S phase, while cells artificially blocked in the G1 phase show very low levels of retrotransposition [51]. However, mature neurons are naturally in the G0/G1 phase and still show increased L1 activity, and this may be considered an indirect argument for the unusual functions of L1 elements in neurons. Further experiments to investigate these functions are necessary.

### 2.7. *Specific Examples of Domesticated Mobile Elements in Neurons*

#### 2.7.1. Neuron-Specific Transcription Regulation Provided by Exapted TEs

TEs from different families were exapted to provide regulatory sequences important for mammalian brain development and function. Emera et al. calculated how many neocortex-specific enhancers overlap with repetitive elements in the human genome. They found that neocortical enhancers in general do not exhibit strong evidence of TE exaptation. However, when recently evolved and more ancient enhancers were assessed separately, it was found that less than 6% of bases in mammalian-specific and older neocortical enhancer cores intersect a repeat element, while for eutherian-specific neocortical enhancer cores this figure was about 30% [399].

Sasaki et al. identified a new SINE family in the genomes of *Amniota* and named it AmnSINE1. A total of 32 of AmnSINE1 loci are located near genes involved in brain development. The authors focused on locus AS071 located 178 kb downstream from *Fgf8* (fibroblast growth factor 8) gene. They used *lacZ* transgenic mouse system to test whether AS071 may regulate the expression of adjacent genes. It was found that AS071 is expressed in the lateral diencephalon and hypothalamus of mouse embryos, and that the expression pattern of *Fgf8* in these regions is identical to that of AS071, suggesting that AS071 functions as a *Fgf8* enhancer [400]. Further experiments with constructs encoding AS071 with various deletions showed that AS071 contains at least three distinct sub-elements directing enhancer activity in different diencephalic domains. AS071 includes three putative TF binding sites, but deletions of these sites did not significantly affect *lacZ* expression pattern in enhancer assays [401]. The same research group studied another conserved AmnSINE1 locus, AS021. Tashiro et al. used the *lacZ* reporter transgene to confirm that AS021 has enhancer activity in developing mouse brain. The expression pattern of AS021 was compared with those of genes located within 2 Mb of AS021. AS021 expression recapitulated the expression of the *Satb2* gene (encoding the TF Special AT-rich sequence-binding protein 2) at later embryonic and postnatal stages in the deep-layer cortical neurons that project axons into *corpus callosum*. The conserved region of AS021 contains 27 predicted TF binding sites; modified AS021 sequence with mutations in 15 of these sites had no enhancer activity. However, the exact mechanism of AS021 enhancer function was not established in this work [402].

MER130 (a very ancient transposon family that cannot be easily classified [19]) also displayed enhancer properties in the brain. Notwell et al. used ChIP-seq for p300 (a transcriptional co-activator recruited by active enhancers) to identify putative enhancers. MER130 elements were highly enriched among p300-binding sequences in the mouse dorsal cerebral wall (which gives rise to the neocortex) at embryonic day 14.5, but not in other tissues, including the forebrain at an earlier time point. Six of these MER130s are located next to genes important for telencephalon formation. Almost all MER130 instances contain a conserved core with five putative TF binding sites: a NeuroD/Neurog motif, an NFI dimer and two additional NFI sites. These TFs are implicated in brain development. In luciferase reporter assays in primary neuronal cultures, 22 of 23 MER130s strongly enriched for p300 signal induced more than two-fold expression upregulation. Mutations in most of the TF binding sites caused a significant reduction in reporter activity [403].

Another important example is the hypothalamus-specific enhancer that regulates neuronal expression of the proopiomelanocortin (*POMC*) gene in mammals. Santangelo et al. compared multiple mammalian genomes and demonstrated that the neuronal *POMC* enhancer nPE2 originated from the exaptation of a CORE-SINE (MIR) element in the lineage leading to mammals and remained under purifying selection during the mammalian evolution. Analysis of *POMC* expression in the brains of transgenic mice featuring deletions in different parts of ectopically expressed nPE2 showed that two regions of nPE2, both derived from CORE-SINE, are essential for enhancer activity in *POMC* hypothalamic neurons [404]. The same research group, using the same in-silico paleogenomics strategy, then showed that another *POMC* enhancer, nPE1, also has a TE-related origin. nPE1 evolved from an LTR of the MaLR family, and it happened after the emergence of nPE2; nPE1 is a placental mammalian novelty. Transgenic mice expressing different fluorescent markers driven by nPE1 or nPE2 demonstrated coexpression of both reporter genes along the entire arcuate nucleus. The onset of reporter gene expression guided by nPE1 and nPE2 was also identical and coincidental with the onset of *Pomc* expression in the presumptive mouse diencephalon [405].

Bejerano et al. identified an ancient TE called LF-SINE and confirmed experimentally that at least one copy of this element, conserved between mammals, chicken and frog, is a functional enhancer. Expression pattern of a representative reporter gene construct driven by the human *ISL1*-proximal LF-SINE in whole-mount mouse embryos closely resembled

the expression pattern of the murine *Isl1* gene. *ISL1* encodes a TF necessary for motor neuron differentiation [264].

Crepaldi et al. showed that SINEs binding pol III-specific TF TFIIC play an important role in pol II-dependent transcription of genes induced by neuronal activation. The authors performed ChIP-seq for H3K9K14ac (used to identify active neuronal genes) in the somatosensory cortex of mice exposed to novel enriched environmental conditions (NEE). Increased H3K9K14ac level in response to NEE was detected in SINEs located within 100 kb of transcription start sites (TSS) of NEE-induced genes. These SINEs contain TFIIC binding sites. ChIP-PCR confirmed that TFIIC binds to specific SINEs located near immediate early genes: an RSINE1 downstream of *Fos* and a B1F element downstream of *Gadd45b*. An immuno-DNA FISH experiment used to visualize transcription factories in the nuclei of neurons showed that co-localization of both *Fos* and *Gadd45b* loci with transcription factories increased in response to depolarization. KD of a TFIIC subunit increased the co-localization of *Fos* and *Gadd45b* with transcription factories in resting conditions, while depolarization has no further effect in these neurons. Thus, TFIIC apparently functions as a transcriptional “brake” in non-stimulated neurons. Indeed, on the morphological level, neurons with TFIIC KD resembled neurons subject to chronic depolarization [243]. Later, the same research group showed that pol III also participates in pol II transcription regulation by SINEs. Policarpi et al. identified more than 1000 enhancer SINEs (eSINEs) that are enriched within 100 kb of TSS of depolarization-inducible genes and have H3K27ac and H3K4me1 marks (associated with putative enhancers). eSINEs bind TFIIC upon depolarization and are transcribed in activated neurons. One specific eSINE,  $Fos^{RSINE1}$ , had been previously identified as a regulatory element [243]. RT-qPCR performed at different time points after depolarization showed that transcription of the  $Fos^{RSINE1}$  enhancer RNA (eRNA) starts earlier than transcription of the corresponding *Fos* mRNA. Experiments with mouse neuroblastoma cells Neuro-2a showed that  $Fos^{RSINE1}$  eRNA is needed for *Fos* relocation to transcription factories, and that a pol III binding site is crucial for the enhancer activity of  $Fos^{RSINE1}$ . Importantly, RNA immunoprecipitation confirmed that  $Fos^{RSINE1}$  eRNA directly binds with pol II. In Neuro-2a cells with  $Fos^{RSINE1}$  eRNA KD, recruitment of initiating and elongating forms of pol II at the *Fos* gene was hindered. Study of primary cortical neurons in vitro and in utero electroporation experiments in vivo showed that that  $Fos^{RSINE1}$  eRNA regulates migration and neuronal differentiation of NPCs and is required for activity-dependent dendritogenesis in maturing neurons [406].

### 2.7.2. Neuron-Specific Proteins Encoded by TE-Derived Genes

The most important example of a TE-derived gene participating in normal neuronal function is *ARC*, which encodes Arc (activity-regulated cytoskeleton-associated) protein. The *ARC* gene in tetrapods and its analogs in flies (*darc1*, *darc2*) independently originated from distinct lineages of Ty3/*gypsy* retrotransposons [407]. *ARC* has homology with retroviral *gag* genes. Some aspects of Arc mRNA regulation also resemble those of viral RNAs, as Arc mRNA contains an internal ribosomal entry site (IRES) that allows cap-independent translation [408]. Arc protein is crucial for multiple stages of long-term plasticity in neurons. Arc plays a critical role in AMPA receptor trafficking and has been implicated in actin polymerization at synapses. Most likely, Arc is involved in other plasticity-related processes that are not yet identified [408,409]. Arc-deficient mice exhibit profound deficits in memory consolidation, despite intact STM and learning acquisition. Synapses in the visual cortex of Arc-deficient animals were shown to be insensitive to the effects of both experience and deprivation [408]. Arc function in neurons is surprisingly similar to the behavior of the viral Gag protein. Experiments with Arc solutions prepared by protein expression in bacteria showed that Arc self-assembles into oligomers resembling viral capsids, detectable by electron microscopy, and that these capsids contain RNA and require RNA for successful assembly. Arc protein and Arc mRNA were then detected in extracellular vehicles (EVs) from cultured neurons. Stimulation of neuronal activity increased the release of Arc-containing EVs, and Arc-KO neurons were able to uptake Arc

protein and Arc mRNA from EVs purified from WT neuron culture medium. Since Arc capsids did not show specificity in RNA binding in vitro, the authors suggested that Arc transfers abundant mRNAs between neurons, providing a novel mechanism in intercellular signaling [407]. A similar mechanism was shown for dArc1 protein in *Drosophila*. dArc1 was confirmed to bind its own mRNA and enter EVs transferred from motor neurons to muscle cells. Artificial disruption of this transfer affected both synapse maturation and synaptic plasticity [410].

*POGZ* (*pogo* transposable element with zinc finger domain) encodes a protein containing a CENP-B-like DNA-binding domain. Mutations in this gene were associated with intellectual disability (ID) with comorbid autism spectrum disorders (ASD). In the human brain, *POGZ* is co-expressed with ASD- and ID-associated genes involved in chromatin remodeling and transcription regulation. *POGZ* was suggested to serve as a neuron-specific transcriptional regulator. *POGZ* was also shown to be involved in mitosis [411,412]. White et al. identified five patients with gene-disrupting mutations in *POGZ* and variable neurodevelopmental disorders. All patients had ID, global developmental delay, behavioral abnormalities and vision abnormalities. Furthermore, 3 out of 5 subjects had microcephaly [412]. Stessman et al., after screening sequencing data of thousands of patients, collected a cohort of 24 individuals diagnosed with various neurodevelopmental disorders and possessing likely gene-disrupting de-novo mutations in *POGZ*. Patients with these mutations all had some degree of developmental delay; most of them had mild ID. In a *Drosophila* model, KD of *row* (an ortholog of *POGZ*) in neurons led to deficits in habituation, a form of learning that is highly relevant for both ID and ASD [411].

The *JRK* gene is also related to *pogo* DNA transposons and shows similarity with *CENPB* [293,413]. It was originally named *Jerky* because mice lacking this gene are prone to epileptic seizures. *JRK* was shown to bind RNA with high affinity. In mice, *JRK* mRNA is expressed in multiple tissues, but on the protein level *JRK* showed brain-specific expression. In primary rat hippocampal cultures, *JRK* was labeled mostly in neurons. It was shown that *JRK* may participate in mRNA conservation in neurons: in brain extracts, *JRK* was identified in translationally inactive mRNP particles but did not associate with ribosomes. *JRK* is also supposed to have a CENP-B-like function in the nucleus based on similarity between these proteins [413].

The *RTL4* gene (also called *MART4*, *SIRH11* or *ZCCHC16*) encodes a putative protein that has a homology with Gag retroviral protein and includes a conserved RNA-binding domain. It must be noted that human *RTL4* is located in the region of X-chromosome where several X-linked ID genes have been identified. *Rtl4* KO mice exhibit a number of behavioral abnormalities—hyperactivity, poor working memory, enhanced impulsivity and/or lack of attention. Mutant mice also have abnormal regulation of the monoamine metabolism in their prefrontal cortex. A negative correlation between expression levels of *Rtl4* (more precisely, its 3'-UTR, left intact in the *Rtl4*-KO mice) and *Dbh*, the gene encoding dopamine  $\beta$ -monooxygenase, was observed in the brain, suggesting that the activity of these two genes may be regulated by the same environmental factors. Surprisingly, while *Rtl4* expression in various brain regions was confirmed by RT-PCR, numerous experiments provided no direct evidence of *Rtl4* mRNA or protein localization in the brain. This leaves the possibility that *Rtl4* in neurons functions as a non-coding RNA. However, the conservation of the amino acid sequence of the *Rtl4* provides indirect evidence that *Rtl4* must be a protein-coding gene [303].

The *NRIF* (*ZFP369*) gene includes the conserved SCAN domain, a protein–protein interaction motif derived from the sequence encoding the C-terminal portion of the Gag retroviral protein [199,414]. *NRIF* (neurotrophin receptor-interacting factor) encoded by this gene is a zinc finger protein that binds the neurotrophin receptor p75NTR and serves as a mediator of neuronal apoptosis. The retinae of mice with a targeted mutation in the *NRIF* gene show reduced cell death, similar to nerve growth factor (NGF)-null mice [414]. Another important function of the *NRIF* protein is regulation of neuronal cholesterol biosyn-

thesis in the hippocampus. Mice with *NRIF* deletion showed a reduction in cholesterol biosynthesis-related gene expression in their hippocampi [415].

The *PGBD5* gene is related to *piggyBac* DNA transposons. It encodes a neuron-specific domesticated transposase with a yet unknown function. This gene is highly conserved in vertebrates, and it was shown that, in mammals, *PGBD5* is almost exclusively expressed in the brain, with preferential expression in certain granule cell lineages. The putative *PGBD5* promoter has a binding site for NRSF/REST, a factor that works as a scaffold for proteins that repress neural genes in non-neural cells. The *PGBD5* protein in mouse brain is detected in both nuclear and cytoplasmic fractions, but it was shown that nuclear *PGBD5* is not bound to chromatin, since *PGBD5* was not released from the crude nuclear pellet either by complete DNase I digestion or by elevated NaCl concentrations. These results suggested that *PGBD5* is probably a structural component of the nucleus [416]. More recently, it was confirmed that ectopically expressed *PGBD5* is capable to induce stereotypical cut-and-paste DNA transposition in HeLa cells [285].

### 2.7.3. Neuronal Non-Coding RNAs Encoded by TE-Derived Genes

It was estimated that about two thirds of all known lncRNAs in zebrafish, mouse, and human contain at least one TE-derived sequence [417]. In the human genome, about 40% of all identified lncRNAs are specifically expressed in the brain [418]. lncRNAs are abundant in the brain and show spatially limited and time-varying expression. They are suggested to play a key role in the hippocampal neurogenesis [419]. Most lncRNAs expressed in developing mouse brain overlap with genes involved in neurogenesis and are coexpressed with these genes [420].

*BCYRN1* is a gene derived from a monomeric Alu element. *BCYRN1* encodes brain cytoplasmic RNA BC200, a primate-specific non-coding RNA [7,421,422]. Interestingly, a functional BC200 analog in rodents, BC1 RNA, is also related to a family of SINEs (ID elements); apparently *Bc1* gene is not a domesticated ID repeat, but rather a “master gene” that gave rise to a whole subfamily of IDs [423]. Both BC1 RNA and BC200 RNA are expressed in neurons, providing rare examples of tissue-specific pol III transcription [95,424], and have the same function, playing an important role in synaptic plasticity. BC1/BC200 participates in local translation regulation in neuronal dendrites. In the absence of synaptic activation, BC1/BC200 binds translation initiation factors, preventing the “premature” translation of dendritic mRNAs. Once the synapse is stimulated, BC RNA releases these factors and the local translation may start [425,426]. *BCYRN1* kept its transpositional ability: more than 200 retrocopies of this gene were identified [422].

Carrieri et al. identified a lncRNA antisense to mouse *Uchl1* gene (encoding ubiquitin carboxy-terminal hydrolase L1). UCHL1 is a neuron-restricted protein; loss of UCHL1 function has been associated with different neurodegenerative disorders. The authors showed that antisense *Uchl1* transcript (AS *Uchl1*) increases *Uchl1* mRNA translation in the MN9D dopaminergic cell line. AS *Uchl1* includes a B2 SINE, which is necessary for the lncRNA activity. In control cells, rapamycin treatment caused increased UCHL1 translation; this did not occur in cells with AS *Uchl1* KD or cells overexpressing the mutant form of AS *Uchl1* lacking the B2 element [427]. Later, more lncRNAs like this were discovered; this new functional class of lncRNAs was named SINEUPs (SINE element-containing translation up-regulators). Schein et al. screened human brain transcriptome and identified 129 putative SINEUPs. The authors selected two of these transcripts (not neuron-specific) and confirmed their SINEUP properties experimentally [428].

lncRNAs may also perform regulatory functions as “miRNA sponges”, competing with mRNAs for binding of corresponding miRNAs. There is an example of TEs providing multiple miRNA binding sites within one RNA sequence expressed in neurons. Bhat-tacharya et al. characterized CYP20A1\_Aluc-LT RNA (a non-coding transcript of the orphan human gene *CYP20A1*) with an unusually long (8.93 kb) 3'-UTR, almost two thirds of which is derived from exonizations of 23 Alus of different subfamilies. CYP20A1\_Aluc-LT is widely expressed; it is present in published data on the single nucleus RNA-seq of human cortical

neurons. CYP20A1\_Alu-LT 3'-UTR harbors thousands of predicted miRNA recognition sites. There are 140 specific miRNAs having 10 or more recognition sites each within this sequence. The miRNA sponge function of CYP20A1\_Alu-LT was indirectly confirmed in experiments with primary human neurons. Heat shock downregulated CYP20A1\_Alu-LT transcription in these cultures, while the application of the HIV1-Tat neurotoxic protein upregulated it. RNA-seq identified 380 genes, expression of which positively correlates with that of CYP20A1\_Alu-LT after heat shock or HIV1-Tat treatment. All these genes contain at least one recognition site for neuron-expressed miRNAs predicted to bind to CYP20A1\_Alu-LT. These genes and corresponding miRNAs are known to be involved in hemostasis and neuron development [283].

Two other TE-related non-coding RNAs that apparently play an important role in neurons are SLC7A2-IT1A and SLC7A2-IT1B isoforms. These two RNAs have different transcription starts within an intron of *SLC7A2* gene, and both contain the sequence from an L1 insertion nested in an Alu insertion. SLC7A2-IT1A/B expression was confirmed in several human brain structures but the function of these transcripts is not known. They have no homology with any known protein or regulatory motif. It was demonstrated that a mutation increasing their piRNA-mediated silencing leads to lethal encephalopathy [429].

TE-derived small RNA species were also found to be expressed in neurons. piRNAs, many of which have a TE-related origin, used to be considered germline-specific. However, recent studies suggest that this class of small RNAs may also have some functions in neurons in both vertebrates and invertebrates [146–148,156,430]. There are also reports about endogenous siRNAs being expressed in neurons [156,430]. Small RNAs of these two classes may arise not only from TEs but from other genomic regions as well. In some papers, the origin of neuron-specific piRNAs is unclear [146] (Rajasethupathy, personal communication), while other works explicitly state that such piRNAs are mapped to unique sites in the mouse genome [147]. However, a large-scale sequencing of small noncoding RNAs in mouse primary hippocampal neurons performed by Nandi et al. discovered numerous TE-related piRNAs and siRNAs. RNA-seq analysis of mouse brain regions revealed the expression of transcripts encoding proteins involved in piRNA and endo-siRNA biogenesis. Mice with KO of the *Mili* (*Piwi2*) gene (encoding a murine analog of PIWI) displayed hyperactivity and reduced anxiety. Analysis of brain samples from these mice allowed the authors to identify specific L1 promoter regions that are putative targets for methylation by a MILI/piRNA-related mechanism in neurons [156].

## 2.8. TEs in the Human Brain Evolution

TE activity is one of the most notable sources of mutagenesis, and it was linked to major genome rearrangements [431] and the formation of new species [432]. Notably, the human genome is particularly enriched for large, highly homologous segmental duplications. Bailey et al. examined sequence features at the junctions of duplications and observed a highly significant enrichment of Alu sequences near or within junctions [433]. Analysis of UTRs of human and mouse mRNAs showed that, in both species, recently expanded gene classes have transcripts enriched in TEs [207]. The key part of brain evolution is changes in the brain development, and there is significant evidence that TE-related mechanisms participated in shaping mammalian embryonic development, including neuronal differentiation [9]. Here, we will focus on primate-specific and human-specific TEs, and, in particular, on neuron-specific genes that could be affected by TE insertions during human evolution.

### 2.8.1. An overview of TE Evolution in Primate Lineage

During primate evolution, there have been multiple waves of LINE retrotransposition as well as the birth of completely new mobile elements such as Alu and SVA [7]. Comparative genomic analysis between human and chimpanzee genomes allowed the identification of about 11,000 TEs that are differentially present. It was determined that, during the past several million years (My), the overall level of transposition was higher in human than in chimpanzee. L1, Alu, and SVA elements together comprised more than 95% of recent

insertions in both species. For all three types of TEs, the number of recent TE insertions in the human genome was at least 1.5 times higher than in the chimpanzee genome (more than three times higher for Alus). Alu insertions constituted 71% of all human-specific TE insertions and 56.1% of chimpanzee-specific insertions [434]. More recent analysis of 83 fully sequenced human, chimpanzee, bonobo, orangutan and gorilla genomes focused on L1 and Alu elements showed that Alu retrotransposition rate underwent the most dramatic changes over the shortest time intervals, with rates of accumulation differing by 15 fold and varying significantly among all branches. Differences in the L1 insertion rate were much more modest and gradual. The human lineage showed the most notable decline in L1 accumulation and increase in Alu accumulation, even though humans, chimpanzees, and bonobos all experienced increased Alu retrotransposition [435]. Although L1 elements were not extensively mobilized, their retrotransposition machinery was used by other elements to propagate. About 40–50 My ago, coincident with the onset of the radiation of the higher primates, ancestral primate genomes underwent intensive amplification of Alus, pseudogenes generated by retrocopying, and specific L1 subfamilies (L1PA6, L1PA7 and L1PA8). This suggests that these L1 subfamilies must have mobilized RNAs in *trans* at accelerated rates [436]. Hominid-specific SVA elements first originated ~25 My ago; this coincided with the increase in the frontal cortex size. The largest increase in *Homo* brain size occurred within the past ~5 My. During this period, L1Hs, AluYa5, AluYb8, and SVA flourished, and these elements are still the most active TEs in the human genome [7].

### 2.8.2. TE-Mediated Recombinations in Human Brain Evolution

Sen et al. compared human and chimpanzee genomes and identified 492 human-specific deletions attributable to recombinations between Alus. The majority of the deletions happened within known or predicted genes (including three deletions that removed functional exons). Of particular interest is the deletion of the fourth exon in the predicted chimpanzee gene *LOC471177*, which is orthologous to human *CHRNA9* gene encoding neuronal acetylcholine receptor subunit  $\alpha 9$ , associated with cochlea hair cell development. Mutations in this gene may affect the dynamic range of hearing [437].

*RHOXF2* is a homeobox gene expressed primarily in the testis, but a growing body of evidence suggests that this gene also has some important function in the brain. There are two copies of this gene in the human genome; in striking contrast to non-human primates, humans appear to have homogenized their two *RHOXF2* copies by the ERV-mediated non-allelic recombination mechanism [438].

### 2.8.3. Multiple New Regulatory Elements Evolved from TEs in the Human Lineage

In mammalian brains, TE-derived sequences have been implicated in multiple layers of gene regulation, from transcriptional and post-transcriptional control to three-dimensional chromatin organization. There are a few examples of convergent evolution of such regulatory elements in primates and rodents (reviewed in [9]).

It is now considered that the speciation and evolution of human lineage happened due to many incremental independent genomic regulatory changes over extended evolutionary periods, not due to a singular phenotype-defining event. Guffanti et al. analyzed human dorsolateral prefrontal cortex transcriptomes and noted that a large set of TEs expressed in this region are highly conserved during 8 My of primate evolution. Although TEs cannot be considered the only mechanisms driving the evolution of the human brain, these conserved elements are likely conveying important primate-specific regulatory functions [439]. As was already mentioned, analysis of chromatin accessibility maps in multiple non-neuronal human cell types showed that almost half of open chromatin sites (DHS) overlapped with TEs; notably, this overlap was higher for primate-specific sequences [202]. In a study by del Rosario et al., anthropoid-specific constrained (ASC) regions were identified in human, orangutan, rhesus macaque, and common marmoset genomes. It was shown that 99.7% of the ASC base pairs are noncoding, suggesting their *cis*-regulatory functions. ASCs

were highly enriched in loci associated with fetal brain development, motor coordination, neurotransmission, and vision. Notably, 56% of ASCs overlapped with TEs [440].

Luo et al. compared 3D genomes [441] of human, macaque, and mouse brains and identified hundreds of human-specific chromatin structure changes. The authors showed that human-specific chromatin loops are enriched in enhancer–enhancer interactions, and the regulated genes show human-specific expression in the subplate, a transient zone of the developing brain critical for neural circuit formation and plasticity. Importantly, human-specific topologically associating domains (TAD) boundaries are enriched with TEs. For Alu and L1, it was shown that only evolutionarily younger subfamilies of these elements (AluY and L1PA) are enriched in human-specific TAD boundaries. The conserved TAD boundaries are more enriched in housekeeping genes, while the human-specific boundaries are more enriched in brain-development-related genes. One example is the human gained TAD boundary around *CNTN5*, a gene involved in neuron circuit formation and autism spectrum disorders; this boundary contains AluY elements and is correlated with an increased *CNTN5* expression in human compared with macaque [442].

Nataf et al. screened the EnHERV database to determine whether candidate lists of cognition-/behavior-related genes are enriched in genes harboring promoter-localized HERV LTRs. The authors found that promoter regions of human genes associated with ID (but not with autism or schizophrenia) are uniquely enriched in LTR sequences of primate-specific MER41 elements. MER41 LTRs in these promoters harbor TF binding sites for YY1 and immunity-associated TFs such as STAT1, STAT3 and NFκB1. Human and chimpanzee genomes were found to have different MER41 LTR insertion sites. A survey of the human proteome allowed the authors to map a protein–protein network which links the suggested immune/MER41/cognition pathway to FOXP2, a key TF involved in the emergence of human speech [443].

SVA elements, despite being the youngest and smallest active human TE family, are supposed to play an important role in human evolution, providing new *cis*-regulatory elements. Some genes important for neuronal function have SVA insertions in their upstream flanking sequence. For two specific SVA D elements upstream of *PARK7* (involved in pathogenesis of Parkinson’s disease) and *FUS* (linked to multiple neuropathologies including amyotrophic lateral sclerosis), the ability of SVA sequence to regulate gene expression rate was confirmed experimentally [20,109]. Of particular interest is possible SVA-related regulation of neuropeptide signaling. VNTRs that are a part of the SVA sequence have been demonstrated to be key neuropeptide regulatory domains. The arginine vasopressin 1a receptor gene (*AVPR1A*) contains two upstream VNTRs, and their regulatory role was confirmed experimentally. SVA insertions were also found in the upstream regions of *OXT* (the gene encoding oxytocin), *TACR3* (tachykinin receptor 3 gene), and *TRPV3* (the gene encoding a vanilloid receptor involved in the modulation of neuropeptide release) [108].

#### 2.8.4. Alu Elements Participated in Our Brain Evolution in Diverse Ways

A recent computational analysis suggested that Alu elements have the largest overall regulatory impact compared to other types of TEs in the human genome: Alu elements showed the highest probability of contributing to regulatory regions of any type (defined by chromatin state). Interestingly, this study showed that most TE types are primarily associated with reduced gene expression, while Alus are associated with upregulated gene expression [444]. Considering Alu expansion in the human lineage [434,435], the effect of these elements on our gene regulation is very important.

An important example of regulatory sequences that have been widely distributed across the genome by Alu expansion is retinoic acid response elements [7]. Retinoic acid is a crucial regulator of both embryonal and adult neurogenesis [445,446]. In humans, 90% of DR2 DNA motifs recognized by retinoic acid receptors are present within Alu repeats. Alu-DR2 motifs are located adjacent to numerous known retinoic acid target genes, and some of these elements were confirmed to recruit retinoic acid receptors in cell culture

experiments. Alus also contain other retinoic acid response elements, including 50% of consensus DR4 motifs [447].

The Alu sequence is the primary target of A-I editing in human RNAs. The intensity of A-I editing in human transcriptome is much higher compared with other animals [274,369,448]. In mice, for example, A-I editing sites are distributed between TEs of different families, and two elements with a low level of homology are less likely to form a stable dsRNA structure recognized by ADAR enzymes (while two human Alus form such a structure easily) [274]. However, the A-I editing level in human brain was found to be higher even when compared with other primates that have similar Alu repeats in their genomes [449]. The number of Alu RNA editing sites in the human genome is more than a million [450], and by some estimations may even be more than 100 million [271]. A-I editing is necessary for normal neuronal function, since it affects genes encoding key proteins involved in neurotransmission, neurogenesis, gliogenesis, and synaptogenesis [451]. The examples of neuron-specific transcripts that are the most important targets for A-I editing have been already provided above in the Section 2.3. In addition, Mattick and Mehler analyzed the human RNA editing datasets and found many transcripts involved in nervous system development, formation of the neural tube, and NSCs regulation [371]. Importantly, Paz-Yaacov et al. found that new editable species-specific Alu insertions that happened after the human–chimpanzee split are significantly enriched in genes related to neuronal functions and neurological diseases [449]. It must be noted that recent studies using full sequencing of genomes and transcriptomes of diverse species revealed that recoding mRNA A-I editing sites affecting the structure of proteins are actually rare; in most cases, editing happens in noncoding mRNA regions, and in human transcriptome the percentage of recoding A-I editing sites is negligible compared to some invertebrate species [372,452]. However, non-coding regulatory RNAs may also be A-I editing targets [371,453]. A-I editing may also influence splicing and miRNA/mRNA interaction [453]. The transcript of the *NARF* gene provides an interesting example of alternative splicing made possible by a few A-I editing events within Alu elements. The resulting new exon is alternatively spliced in a tissue-dependent manner, with a high level of inclusion in the brain [454].

Alu exonizations in neuron-specific genes seem to have distinct properties compared to Alu exonizations in other genes. While analysis of the whole human transcriptome showed that most Alu exons are present in 3'-UTRs [260], in the cerebellum transcriptome most Alu exons are located in 5'-UTR regions. It was also found that, among the cerebellum-expressed genes, Alu exonization events are strongly enriched in genes encoding zinc finger TFs (ZNF); notably, ZNF genes underwent rapid expansion during human evolution [455]. An important example of a novel Alu exon is noncoding exon 2 of the *GPR56* gene encoding a G-protein-coupled receptor required for cortical development. This gene in humans has at least 17 alternative first exons, while only 5 such exons were identified in the mouse *Gpr56* gene. Interestingly, this gene also has an upstream *cis*-regulatory element showing a homology to a 4/RTE mammalian LINE [456].

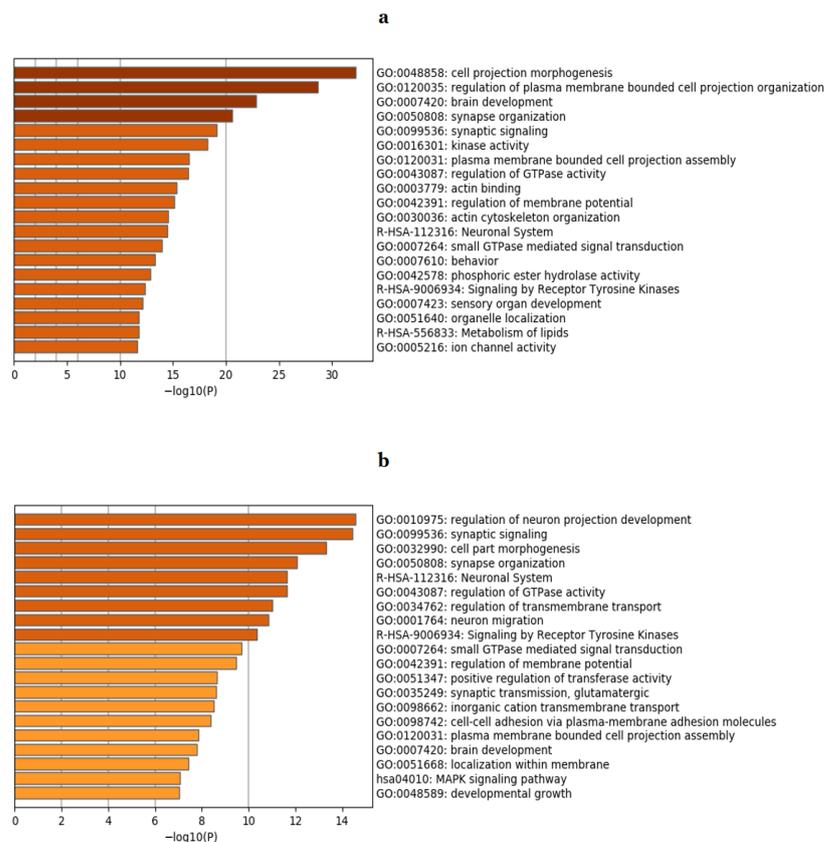
Recombinations between Alu copies caused deletions [437] and duplications [433] of genome sequences; the papers cited here have been already discussed above.

There are two examples of specific Alu insertions that generated new genes active in human neurons. A monomeric Alu insertion that appeared after the split of prosimians and anthropoids generated the *BCYRN1* gene encoding the functional non-coding RNA—brain cytoplasmic RNA BC200 [7,421,422]. However, this RNA is not exclusively primate-specific; it has an analog in rodents [424,426]. Its structure and function are described in Section 2.7.3. Parrott et al. identified a small non-coding RNA family named *snaR* that also evolved from an Alu monomer. *snaR* RNAs are transcribed by pol III and associate with the nuclear factor 90 (NF90/ILF3) protein; they are suggested to participate in cell growth and translation regulation. *SNAR* genes were only found in African great apes, and some are unique to humans; however, other primates have *snaR*-related elements. Human *SNAR* genes are expressed in testis and brain, with differential expression of *snaR* subsets in different regions of the brain, but their function in neurons is not clear yet [457].

A deleterious Alu insertion was traced to happen shortly before the time when brain expansion began in human ancestors. This insertion inactivated the *CMAH* gene encoding the key enzyme for the synthesis of N-glycolylneuraminic acid (Neu5Gc). In mammals with a functional *CMAH* gene, its expression seems to be selectively downregulated in the brain. These facts led the authors to propose an intriguing hypothesis about Neu5Gc somehow restricting brain growth [458]. However, *Cmah* inactivation in mice had no effect on general appearance, sensorimotor reflexes, or learning and memory [459]. On the other hand, brain-specific *Cmah* overexpression in mice resulted in abnormal locomotor activity, impaired object recognition memory, and abnormal axon myelination, confirming that Neu5Gc has negative effects on neural functions, even though mutant mice showed no gross defects in brain size [460].

### 2.8.5. Human-Specific TE Insertions within Neuron-Associated Genes

In the study by Mills et al. where human and chimpanzee TEs were compared, a third of all new insertions were located within known genes in both species. Further analysis of insertion patterns revealed that they were far from random. Only ~14% of all human genes were affected by TEs, and about a third of these genes contained multiple insertions. Similar results were observed in the chimpanzee genome. It appears that a large fraction of the new transposon insertions in both humans and chimpanzees were somehow targeted preferentially to specific genes [434]. We further analyzed data sets published in this paper using Gene Ontology [461–463]. In the human dataset, TEs were significantly overrepresented in 1866 genes; in the chimpanzee dataset, the analysis yielded significant over-representation of TEs in 820 genes. Most interestingly, many of these genes are associated with neuronal function (see GO terms in Figure 2).



**Figure 2.** Analysis of datasets from [434]. (a). Gene Ontology enrichment analysis for the list of human genes targeted by recently mobilized transposons. (b). Gene Ontology enrichment analysis for the list of chimpanzee genes targeted by recently mobilized transposons.

### 3. The Role of Neuronal TEs in Pathology

We have found reports about neuronal TE activation in multiple diverse neuropathologies. This information is summarized in Table 3.

**Table 3.** Neuropathologies associated with TE dysregulation.

Disease	Type of Disease	Associated TE Activity	Suggested Mechanisms of TE Participating in Pathogenesis	Section Number
Alzheimer's disease	Neurodegenerative	Transcriptional activation of different TE types	Neuroinflammation; generation of defective <i>APP</i> retrocopies; speculated Alu-mediated alterations of <i>MAPT</i> transcripts	3.2.1
Parkinson's disease	Neurodegenerative	Increased L1 transcription; germline TE insertions	Aggravation of oxidative stress; germline <i>PRKN</i> mutations	3.2.2
Huntington's disease	Neurodegenerative	Increased L1 mobilization and transcription	Dysregulation of cell survival signaling pathways	3.2.3
Ataxia telangiectasia	Neurodegenerative	Increased L1 mobilization; L1 insertions with increased average length	No specific mechanisms proposed	3.2.4
Spinal muscular atrophy	Neurodegenerative, neuromuscular	Alu-mediated recombinations	Germline <i>SMN1</i> mutations	3.2.5
Amyotrophic lateral sclerosis	Neurodegenerative	Increased HERV-K transcription; increased L1 mobilization	Env protein toxicity; ORF1p aggregation with mutant host proteins	3.2.6
Fronto-temporal lobar degeneration	Neurodegenerative	Decreased binding between TDP-43 and its multiple TE RNA targets; increased L1 mobilization and transcription	No specific mechanisms proposed	3.2.6
Fragile X-associated tremor/ataxia syndrome	Neurodegenerative	Increased LTR transcription	The causal role of TEs in pathogenesis was confirmed, but the exact mechanisms of damage are unknown	3.2.7
Multiple sclerosis	Autoimmune, inflammatory	Increased HERV-W transcription	Env protein toxicity; Env-induced autoimmunity to myelin proteins	3.2.8
Aicardi–Goutières syndrome	Inflammatory, neurodevelopmental	Increased L1 mobilization and transcription	Accumulation of L1 cDNA in the cytoplasm leading to IFN-I response and neuroinflammation	3.2.9
Glioblastoma	Malignant tumor	Increased HERV transcription; selective packaging of HERV RNA in microvesicles	No specific mechanisms proposed	3.2.10
Autism spectrum disorders	Neurodevelopmental	Increased L1 transcription; germline L1 and Alu insertions in autism-associated genes; increased HERV-H transcription	Increased L1 RNA level correlated with oxidative stress; germline TE insertions likely affect gene expression rate	3.2.11

Table 3. Cont.

Disease	Type of Disease	Associated TE Activity	Suggested Mechanisms of TE Participating in Pathogenesis	Section Number
Rett syndrome	Neurodevelopmental	Increased L1 mobilization; less frequent L1 insertions in exons	No specific mechanisms proposed	3.2.12
Schizophrenia	Mental (psychotic)	Increased L1 mobilization; germline and somatic L1 insertions in genes associated with SZ and similar pathologies; increased HERV-K10 and MSRV <i>env</i> transcription	L1 insertions affect gene expression rate (confirmed in vitro for somatic insertions). A link between ERV expression and myelin inflammation was proposed	3.2.13
Bipolar disorder	Mental (mood disorder)	Increased HERV-K10 and MSRV <i>env</i> transcription	A link between ERV expression and myelin inflammation was proposed	3.2.13
Major depressive disorder	Mental (mood disorder)	Increased L1 mobilization in blood cells; decreased L1 mobilization in prefrontal cortex	No specific mechanisms proposed	3.2.14
Post-traumatic stress disorder	Mental (anxiety)	Altered L1 and Alu methylation levels in blood cells; increased L1 transcription in the brain after stress exposure	No specific mechanisms proposed	3.2.15
Drug addiction	Substance use disorder	Increased L1 mobilization and transcription caused by metamphetamine, cocaine, morphine administration; L1 insertions in genes associated with cocaine addiction in addict brain samples; increased LTR and L1 transcription in alcoholic brain samples	Germline or early developmental stage somatic insertions in genes influencing predisposition to cocaine addiction	3.2.16
Creutzfeldt–Jakob disease	Neurodegenerative, prion	Increased HERV-W and HERV-L transcription	Neuroinflammation	3.2.17
Neurofibromatosis type I	Neurocutaneous	Germline TE insertions within <i>NF1</i> with unexpectedly high rate	Germline <i>NF1</i> mutations	3.2.18
Age-related macular degeneration	Neurodegenerative	Increased Alu transcription	Alu RNA accumulation in the cytoplasm causing innate immune response and apoptosis	3.2.19
X-linked dystonia-parkinsonism	Neurodegenerative	SVA insertion	Germline <i>TAF1</i> mutation	3.2.20
Ravine encephalopathy	Neurodegenerative	Point mutation within a TE-derived region	Germline <i>SLC7A2</i> mutation	3.2.21

### 3.1. TE-Related Mechanisms of Neuropathology

#### 3.1.1. Known Mechanisms of TE-Associated Disorders in General

TEs have long been associated with different pathologies. It was demonstrated that full-size L1 insertions in human genome are subject to negative selection [71]. In 1988, Kazazian et al. were the first to demonstrate that a germline L1 retrotransposition event is a cause of a human disease (hemophilia A) [464]. Since then, more than a hundred de-novo germline TE insertions disrupting normal gene function have been implicated in multiple inherited diseases, most notably various cancer types, hemophilia and neurofibromatosis type I [465]. Somatic mutations caused by transposition or TE-mediated chromosomal rearrangements have been causally linked to several types of cancer. Derepression of a TE may also activate transcription of an adjacent proto-oncogene via the TE regulatory elements [24]. The vast majority of TEs that are transcriptionally activated in pathology are no longer replication competent. However, there are other ways in which their activity may be harmful for the host cell. Some of these transcripts encode toxic proteins or RNAs. Exotic nucleic acid structures derived from activated TEs trigger immune responses [24,33,297,310].

#### 3.1.2. Neuropathologies Associated with TE Activation

In this review, we will focus on diseases that are thought to be caused by either de-novo germline TE insertions in neuron-specific genes, or somatic transposition events in neurons. Most of TE-associated neuropathologies involve neurodegeneration and/or neuroinflammation. To date, multiple authors have provided evidence that TE activity may be involved in pathogenesis of Alzheimer's disease, autism, schizophrenia, Rett syndrome, ataxia-telangiectasia, amyotrophic lateral sclerosis, macular degeneration and some other neuropathologies [39,56,112,113,350,466]. Physiological conditions such as stress and aging are not formally considered disorders, but since they are associated with some detrimental effects as well as with TE derepression [384,467–469], we discuss these conditions as well.

Multiple studies described in the previous chapter confirmed that neuronal tissue is normally more permissive to TE activity compared to other somatic tissues. Even higher TE expression and/or transposition levels were found in a number of neurological disorders. In a recent review, Terry and Devine discussed that mutations in genes that normally play a role in regulating TE activity were only confirmed in some neuropathologies. Other neuropathologies with known overall TE activation are disorders caused by a combination of genetic predisposition and environmental stress factors or disorders associated with aging. It is likely that for some diseases TE activation may be merely a side effect of cell stress [350]. However, examples of specific mechanisms directly linking TEs with detrimental processes in neurons are also known [470,471]. There are also case reports about very specific TE insertions affecting certain genes relevant for neuropathology [363,472–474].

Abrusán used flux balance analysis, a mathematical method used for simulation of genome-scale metabolic networks, to estimate the influence of somatic retrotransposition on human brain metabolism. The simulation showed that TE mobilization may influence the biosynthesis of more than 250 molecules, including dopamine, serotonin and glutamate. When metabolic changes associated with various neuropathologies were compared with metabolic changes caused by TE insertions, it was found that the number of metabolites influenced by TEs is significantly higher than expected by chance in the case of Parkinson's disease, schizophrenia, and autism [475]. Larsen et al. listed 37 neurological and neurodegenerative disorders, including Alzheimer's disease, Parkinson's disease, epilepsy and amyotrophic lateral sclerosis, in which deleterious Alu activity has been implicated as a contributing factor [451]. Jacob-Hirsch et al. used whole-genome sequencing to characterize the retrotransposition landscape of patients affected by various neurodevelopmental disorders. To estimate the role of somatic retrotranspositions, they studied diseases characterized by extensive variability of clinical features, including intra-family variability. The authors showed that the number of retrotransposition events is higher in diseased brains compared to healthy ones. They also found out that somatic L1 insertions in neurons mostly target

genes associated with neural functions and diseases; this was especially prominent for brain samples from patients with subependymal giant cell astrocytoma, Rett syndrome, ataxia-telangiectasia and non-syndromic autism [39].

### 3.1.3. TEs in Neurons and Stress

“Stress” is a very broad term for any intrinsic or extrinsic stimulus that evokes a biological response. Stress can induce both beneficial and harmful consequences. It has many effects on human CNS, and may even cause structural changes in the brain. Stress exposure activates autonomous nervous system and the hypothalamus–pituitary–adrenal axis and triggers the release of glucocorticosteroids and catecholamines. The hippocampus is the brain region with the highest density of glucocorticosteroid receptors and the highest level of response to stress. Other key brain structures involved in stress response are the amygdala and temporal lobe [476]. Many complex human diseases, including neurological ones, are affected by chronic and/or severe types of stress, and the control of TE activity is partly lost in many of these diseases. Many TFs involved in the stress response (such as the glucocorticoid receptor) regulate expression of epigenetic modifying enzymes (DNA-methyltransferases or histone methyltransferases) that participate in TE silencing [384,467]. There are most likely other mechanisms of TE regulation in stress conditions. Understanding of such mechanisms is hindered by the fact that, during stress, preprogrammed molecular mechanisms work in the background of multiple random, unspecific events that only happen because normal cell homeostatic mechanisms are strained. There are numerous reports of TE activity in the brain or other tissues being affected by stress, and the observed patterns seem to be quite complex.

Hunter et al. reported repressive histone H3 lysine 9 trimethylation of TEs and decrease in SINE and ERV transcription levels in the hippocampus of rats subjected to acute restraint stress. Interestingly, while there were no changes in H3K9me3 level in the cerebellum, RT-qPCR showed L1 and ERV transcription upregulation by acute stress [196]. In the study of mechanisms underlying stress-enhanced fear learning in rats (a model of PTSD) gene expression profiles in the amygdala were examined 3 weeks after the exposure. One surprising finding was a statistically distinct module containing tightly coregulated transcripts that mapped to L1 elements. All these transcripts were highly upregulated in the basolateral amygdala of rats subjected to stress-enhanced fear learning [390]. A study by Cappucci et al. showed that neuronal L1 expression in stress conditions depends on the genetic background. BALB mice subjected to five daily restraint sessions showed increased L1 transcription in the hippocampus, while in the hippocampi of C57 mice L1 expression was downregulated after the same stress experience [391].

There are some data about early-life stress influencing TE activity. The paper by Bedrosian et al. about low maternal care causing L1 mobilization in the hippocampus [385] has been discussed above, in the Section 2.5. Cuarenta et al. used the neonatal predator odor exposure paradigm to model early-life stress in rats. On the last day of exposure (postnatal day 3), the authors measured L1 ORF1 and ORF2 DNA copy number and mRNA expression in the amygdala, hippocampus and prefrontal cortex of the experimental animals by qPCR. It was found that stress causes an increase in ORF1 copy number in the hippocampus, but only in males. The differences in mRNA expression after the stress exposure were also observed only in male animals: ORF1 mRNA level increased in both amygdala and prefrontal cortex, while ORF2 mRNA level increased only in amygdala [477]. The same group then studied delayed effects of early-life stress. L1 mobilization and expression after the neonatal predator odor exposure were assessed like in the previous work, with the difference that the animals were sacrificed at postnatal day 33, a month after the exposure. In this case, the only significant effect of stress on L1 activity was the increase in ORF1 copy number in the amygdala. Behavioral tests showed that ORF1 copy number in the amygdala negatively correlated with juvenile social-play levels. The authors also observed sex differences in L1 copy number and transcription levels in all three studied regions [478]. Nätt et al. showed that high cortisol level in hair samples (a common

biomarker for chronic stress) correlates with a genome-wide decrease in DNA methylation (measured in whole blood DNA) in preschool children. The authors found that cortisol-associated differential methylation generally targets SINEs, but not other types of repeats. SINEs were associated with both hypo- and hypermethylated regions in this study, more so with the latter. Hypomethylated, but not hypermethylated differentially methylated regions (DMRs) were strongly associated with genes. Many of these genes are involved in neurodevelopment and calcium transport. More than a half of the DMR-associated genes had previously been linked to disease, mostly with neuropsychiatric disorders, especially those related to aging [479].

#### 3.1.4. TEs in Neurons during Normal Aging

Aging is a result of gradual accumulation of damage in cells and tissues [469]. A phylogenetically conserved feature of aging is induction of stress response pathways that was confirmed in all systems studied, including whole organism analysis of invertebrates and analysis of the brain of different mammalian species [480]. There is also evidence that human aging is associated with mild chronic inflammation in the absence of infection, revealed by elevated levels of inflammatory biomarkers [481]. Old age is a major risk factor for neurodegenerative disorders. Moreover, even normal aging in healthy people includes decrements in learning and memory, attention, decision-making speed, sensory perception and motor coordination, and even reduction in brain size [476]. Many pathological features of neurodegeneration are aggravated hallmarks of normal brain aging (including genomic instability and epigenetic alterations), suggesting that neurodegeneration may be actually seen as accelerated aging [468,469]. One of the most important mechanisms in neuronal aging and age-associated cognitive impairment is oxidative damage [482]. Just like in the case of stress, studying the role of TEs in aging is complicated by preprogrammed mechanisms coinciding with unspecific cell damage consequences.

TEs, and L1 elements in particular, become unsilenced in different organs and cell types during aging. De Cecco et al. discovered that, during replicative senescence in human fibroblasts, chromatin undergoes extensive changes, and expression levels of L1 and Alu elements and L1 genomic copy number are increased [483]. These observations were confirmed in vivo—expression levels of murine LTRs, LINEs and SINEs are elevated in liver and skeletal muscle samples of older mice [484]. SIRT6 (sirtuin 6) is a multifunctional protein participating in cell stress response, transcription regulation and DNA repair, and SIRT6 depletion is strongly associated with premature aging [485]. SIRT6 is involved in silencing of L1 elements. SIRT6-deficient cells show excess genomic instability, DNA damage, accumulation of cytoplasmic L1 cDNA and strong IFN-I response; this is prevented by RT inhibitors or L1 RNA KD. SIRT6 KO mice show a strong activation of L1 in multiple tissues (including the brain) and develop a severe premature aging phenotype; treatment with RT inhibitors improves the health and lifespan of these mice [486].

Elevated TE expression was reported in aging brain. Van Meter et al. examined L1 expression in brain, liver and heart tissue from young (4 months old) and old (24 months old) mice. The brain exhibited the most significant deregulation of L1 expression with age: in old brain tissue L1 expression increased more than seven fold compared to the young brain. In contrast, in the old liver, L1 expression level was 65% higher than in the young liver, and no significant difference in L1 expression between old and young heart was found [487]. Ramirez et al. analyzed TE RNA levels in forebrain lysates from B6C3HF1 mice aged 6, 12 and 20 months from publicly available RNA-seq data. Among all TEs differentially expressed between age groups, 92.7% were upregulated in 20-month-old mice; most of these TEs were ERVs [488]. In *Drosophila* brain, expression of R1, R2 and *gypsy* TEs increases with age. Mutations of *dAgo2* (encoding an Argonaute protein) cause overexpression of TEs in brains of young flies to the levels comparable with those in older flies. This is accompanied by rapid age-dependent memory impairment and shortened life span [489].

### 3.1.5. TEs, Autoimmunity and Neuroinflammation

It is now well-known that autoimmunity and inflammation are among underlying pathological mechanisms in multiple neurological disorders [56,490–492].

The most obvious connection between TEs and inflammation exists for endogenous retroviruses. dsRNA and RNA:DNA hybrids produced during ERV transcription and mobilization are known to trigger various innate immunity cascades in different cell types [297,310]. HERV proteins are also capable to trigger antiviral responses in the absence of exogenous viral infection [32,309]. Experiments with non-neuronal cells, neurons and astrocytes showed that type-I interferon (IFN-I) upregulates transcription of HERV-K via interferon-stimulated response elements in the viral promoter, meaning there is a positive feedback loop increasing inflammatory response [248]. HERVs were implicated in various neuropathologies [56], including multiple sclerosis [493], amyotrophic lateral sclerosis [248], schizophrenia and bipolar disorder [494], and Alzheimer's disease [495].

L1 elements are also able to cause immune responses. As mentioned in the previous section, there is a link between cellular senescence, increased L1 expression and activation of the IFN-I response demonstrated in various cell types of human and mice [486,496]. Elevated L1 DNA content has been reported in several neurological disorders associated with autoimmunity, including Rett syndrome, ataxia telangiectasia, schizophrenia and Aicardi–Goutières syndrome. The abundance of L1 DNA in these disorders is most likely a result of accumulation of L1 DNA molecules not integrated into the nuclear genome. Accumulation of DNA in the cytoplasm may trigger immune responses mediated by various sensor pathways [116].

### 3.1.6. TEs and Mitochondrial Dysfunction in the Context of Neurodegeneration

The human brain has the highest energy requests among all organs. Neurons utilize highly oxidative glucose metabolism, making neuronal mitochondrial function especially important and vulnerable. Beyond their function in energy metabolism, mitochondria are also involved in cellular  $\text{Ca}^{2+}$  homeostasis, and calcium signaling plays a fundamental role in neuronal function. Mutations in mitochondrial DNA are associated with aging-related pathological processes in neurons [482]. Notably, in mammals, mitochondrial DNA only encodes 13 proteins, and more than 1000 other proteins necessary for mitochondrial function are encoded by nuclear DNA (mitonuclear genes) [497]. Thus, TE-related alterations of the nuclear genome can affect mitochondrial function.

Larsen et al. suggested that cell stress at early stages of neurodegeneration may be caused by somatic mutations in mitonuclear genes. These genes were found to be enriched with Alu elements (compared with protein-coding genes not related to mitochondria). Alu-mediated mutations were documented in a number of mitonuclear genes implicated in various neuropathologies, such as *ABCD1* being associated with adrenoleukodystrophy, *OPA1* being linked to autosomal dominant optic atrophy, or *NDUFS2* and *PDHA1* being involved in neurodegeneration [498]. There are also 57 mitonuclear genes with A-I editing occurring within putative Alu exons; in 52 cases, the editing changes protein sequences. Many of these genes encode proteins involved in essential neuronal processes [451].

Baeken et al. confirmed the involvement of L1 activation in mitochondrial dysfunction and related neurodegeneration using in vitro and in vivo models of Parkinson's disease [499]. This paper is described below in the Section 3.2.2.

TE RNA accumulation may activate innate immunity machinery, which has been discussed in the previous section; it must be noted that these processes also increase ROS production in mitochondria, which contributes to cell death [471].

### 3.2. Specific Neurological Diseases with Reported Changes in TE Activity

Here, we compile some reports about changed TE mobilization and/or expression in various neuropathologies. While some of these studies suggest specific molecular mechanisms of TE involvement in disease, many papers cited below merely provide observations that are yet to be explained.

### 3.2.1. Alzheimer's Disease and Tauopathy

Alzheimer's disease (AD) is the leading cause of dementia. At the tissue level, the main features of AD pathology are extracellular neuritic plaques and intracellular neurofibrillary tangles, comprised of aggregated, misfolded A $\beta$  peptide and Tau protein, respectively. Other abnormalities such as astrogliosis, microglial activation and cerebral amyloid angiopathy are often observed; the downstream consequences include neurodegeneration leading to macroscopic atrophy. The vast majority of AD cases are sporadic, but it is now thought that ~70% of AD risk is attributable to genetic factors. The  $\epsilon$ 4 allele of the *APOE* gene is the single biggest risk for sporadic AD. Having one or two copies of this allele increases the risk 3- or 12-fold, respectively [500].

Guo et al. analysed cortical transcriptomes of elderly people with and without AD and found evidence of global TE transcriptional activation in AD. Activation of three ERV clades correlated with global cognitive performance in the year proximate to death. A total of 9 individual TEs that were associated with Tau tangles were identified: a single L1, 5 SINEs, and 3 ERVs. The authors used two *Drosophila* genetic models to establish the link between TE expression levels and tauopathy. Several TEs showed transcriptional activation in neurons in one or both Tau transgenic strains. For some elements, activation was further enhanced with aging. Tau pathology was previously linked to chromatin reorganization, and the authors supposed that it may derepress silenced TEs, which, in turn, may cause genome instability and neuroinflammation [501]. The role of chromatin reorganization by Tau was confirmed by Sun et al.: they found that most TEs that were found to be upregulated in Tau transgenic *Drosophila* are similarly upregulated in flies with mutations that cause depletion of constitutive heterochromatin. Tau<sup>R406W</sup> fly brains also showed downregulation of multiple piRNAs, as well as decreased PIWI protein level. Pan-neuronal *piwi* overexpression reduced neuronal death in Tau<sup>R406W</sup> flies, suggesting a causal contribution of PIWI-piRNA mechanism dysfunction in Tau-induced neurotoxicity. Accordingly, piRNA system disruption caused by mutations of *flamenco* (a locus encoding a major piRNA cluster) enhanced neuronal death and exacerbated Tau-induced locomotor deficits in Tau<sup>R406W</sup> flies (but not in flies without the Tau-encoding gene) [502]. More recently, the same research group made similar observations in murine models of tauopathy, rTg4510 and PS19 transgenic mouse strains. L1 and ERV genomic copy numbers were increased in the cortex of 12-month-old rTg4510 mice compared to control animals. TE (predominantly ERV) transcription levels were increased in CNS samples from 9-month-old rTg4510 and PS19 mice. Publicly available RNA-seq data from JNPL3 tauopathy strain (but not from APdE9, an A $\beta$ -based mouse model of AD) also showed TE upregulation. Gag protein level was found to be elevated in frontal cortex of 6- and 12-month-old rTg4510 mice [488]. Interestingly, *MAPT* gene encoding Tau protein incorporates multiple Alu elements and Alu-related A-I editing sites, which may be involved in the formation of pathogenic Tau isoforms [451].

The role of TE demethylation in AD pathogenesis was suggested. Bollati et al. evaluated methylation of repetitive elements in peripheral blood leukocytes from AD patients and matched healthy control subjects. Alu methylation was decreased in AD samples, while L1 methylation level was increased in this group. However, the observed changes were small in size. Methylation level of L1 or Alu did not correlate with age, APOE status, or CSF levels of A $\beta$  and Tau [503].

The role of retrocopying in AD was also demonstrated. Lee et al. studied small populations of nuclei from AD cerebral cortex samples to detect mosaic alterations of *APP* gene encoding A $\beta$ . RT-PCR detected unusual APP mRNA isoforms with loss of central exons, and some of these isoforms were also detected in genomic DNA samples. Sequencing experiments showed that *APP* retrocopies with exon loss and single-nucleotide variations were more prevalent in AD than in control samples. Two *APP* retrocopy variants with exon loss were also identified in neurons of J20 mice (a genetic model of AD); importantly, one of these variants accumulated with age. When the cytotoxicity of three proteins encoded by atypical APP mRNA variants was examined, two of them could induce cell death. Ex-

periments with cell cultures showed that incorporation of *APP* retrocopies into the genome only happens if DNA damage is induced; application of RT inhibitors prevented *APP* retrocopy formation. The reverse transcriptase responsible for *APP* retrocopy formation was not identified, but the authors do not exclude the possibility that it is ORF2p [504].

L1 mobilization is likely not involved in AD pathogenesis, since no significant difference in L1 genomic copy number in brain and blood samples between AD patients and age-matched controls without dementia was detected by qPCR [505].

### 3.2.2. Parkinson's Disease

Parkinson's disease (PD, or parkinsonism) is the second most common neurodegenerative disorder after AD. The main histopathology associated with PD is degeneration of mesencephalic dopaminergic (mDa) neurons in the *pars compacta* of the *substantia nigra* (SNpc). The most prominent symptoms are motor abnormalities—resting tremor, bradykinesia, rigidity and postural reflex impairment. As the disease progresses, decreased cognitive ability may appear as well. The etiology of PD is complex, involving both genetic and environmental factors. Molecular mechanisms of PD pathogenesis are still largely unknown, even though at least 11 genes have been associated with PD. The most important of these genes is *PRKN* (*PARK2*) [506].

Blaudin de Thé et al. studied L1 expression in mDa neurons using diverse models of PD. For in-vitro experiments, the authors used primary cultures of ventral midbrain neurons subjected to oxidative stress by H<sub>2</sub>O<sub>2</sub> application. For experiments in vivo, either *En1*<sup>+/-</sup> mutant mice or mice injected with 6-OHDA were used. *En1* encodes Engrailed-1, a homeobox TF specifically expressed in adult mDa neurons; in heterozygous *En1*<sup>+/-</sup> mice, mDa neurons show a progressive degeneration that starts at 6 weeks of age. Used in this study 6-OHDA is a drug that induces oxidative stress specifically in mDa neurons. The authors observed increased L1 expression in SNpc mDa neurons of 6-week-old *En1*<sup>+/-</sup> mice compared to WT siblings. In experiments with acute oxidative stress in both cultured midbrain neurons and adult mDa neurons in vivo, DSB formation correlated with enhanced L1 expression. The causal role of L1 activity in oxidative stress in neurons was confirmed in further experiments. The authors inhibited L1 expression in vitro and in vivo by four independent methods: transcriptional repression by Engrailed-2 application, a siRNA targeting L1 transcript, application of RT inhibitor stavudine, and ectopic overexpression of *Miwi* (leading to L1 silencing by RNA interference). In all these models, decreased L1 transcription led to less prominent oxidative stress. These results confirm that L1 overexpression in mDa neurons triggers oxidative stress-induced DNA damage. The suggested role of Engrailed proteins in these neurons is neuroprotection through the repression of L1 transcription [507]. In human dopaminergic LUHMES cells, treatment with MPP<sup>+</sup> or rotenone (respiratory chain complex I inhibitors used to model PD) caused an increase in ORF1p expression. Similar ORF1p expression increase was registered in midbrain lysates from mice treated with a metabolic precursor of MPP<sup>+</sup> that selectively targets dopaminergic cells. Experiments with cell cultures also showed that respiratory chain complex I inhibitors cause a decrease in overall DNA methylation level, suggesting a potential mechanism of TE mobilization [499]. Pfaff et al. utilized whole-genome sequencing data to investigate the role of retrotransposition-competent L1 elements (RC-L1) in PD. The authors identified 22 reference and 50 non-reference RC-L1 loci that are polymorphic for their presence or absence. A total of 18 of these loci were determined to be highly active. Healthy controls and PD subjects were categorized based on the total number of highly active RC-L1s. In the control group, 45% of individuals had ≥9 highly active RC-L1 in comparison to the 57.2% of PD subjects. PD patients with a higher number of highly active RC-L1s showed an increase in markers of disease progression such as the total levodopa equivalent daily dose or the degree of DA neurons degeneration. The authors hypothesize that the RC-L1s may not trigger neurodegeneration but likely contribute to the process and aggravate the cell stress [508]. In a study where the methylation of the L1 promoter in blood mononuclear cells of subjects recently diagnosed with idiopathic PD was measured,

no correlation between this parameter and the disease was found; however, the inverse correlation between PD and ever regularly smoking tobacco was strongest for subjects with the lowest L1 promoter methylation [509].

Whole-genome sequencing of three families with early onset PD revealed five structural variations in the *PRKN* gene; retrotransposon sequences were identified within 1–2 Kb of the deletions' break points, suggesting that all identified variations might originate through retrotransposition events [472].

### 3.2.3. Huntington's Disease

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder, characterized by progressive motor impairment, cognitive decline and psychiatric disturbances. HD patients show selective loss of medium-sized spiny neurons of the striatum; at later stages, pathologic alterations happen in other brain regions. HD is caused by mutations in *HTT*, the gene encoding huntingtin. Pathological alleles of *HTT* have an increased number of CAG repeats, and as a result mutant protein forms have an expanded polyglutamine tract (polyQ) and the tendency to aggregate [510]. A link between huntingtin polyQ-expansion and chromosomal instability was also suggested [511].

In a murine model of HD, it was shown that increased L1 mobilization may be involved in HD pathogenesis. L1 DNA copy number in HD mouse brain caudate regions was increased compared to WT control. Expression of ORF1 and ORF2 RNAs and proteins was detected in brain caudate of HD but not WT mice. Pyrosequencing showed that, in HD mice, L1 regions are less methylated. Increased L1 content in HD mouse neurons correlated with increased cell death and dysregulation of cell survival signaling pathways (indicated by decreased mTORC1 activity and AMPK $\alpha$  level). Overexpression of L1 ORF2, but not ORF1, in 293T cells caused similar dysregulation of cell survival signaling and cell death [512]. Conversely, in another study, L1 expression was found to be decreased in HD. Analysis of published RNA-seq data of two HD mouse models showed decreased expression of some specific L1 RNAs in the striatum, but not the cortex, of transgenic mice expressing pathological alleles of human *HTT* (compared with mice expressing its normal allele). L1 downregulation in striatum was more prominent in mice expressing the *HTT* allele with more CAG repeats. In human HD post-mortem brain samples (prefrontal cortex Brodmann area 9 that is involved in HD pathogenesis but suffers less neuronal death than striatum), L1 expression was downregulated compared with normal controls, but the difference was not significant [510]. TE expression and mobilization were found to be increased in a *Drosophila* HD model; moreover, inhibition of TE mobilization with RT inhibitors rescued polyQ-dependent eye neurodegeneration and genome instability and increased fly lifespan [513].

### 3.2.4. Ataxia Telangiectasia

Ataxia telangiectasia (AT) patients exhibit progressive neurodegeneration and eventual death in the second or third decade of life. The disease is caused by mutations in the *ATM* gene, which plays an important role in DNA damage signaling. Coufal et al. performed KD of *ATM* in hESCs that were then differentiated to NPCs and transfected with L1-EGFP retrotransposition reporter. L1 retrotransposition level in *ATM*-deficient NPCs was increased, but these cells were still able to differentiate into functional neurons. L1 expression was not increased in *ATM*-deficient cells; insertion of L1 copies in unusual sites of the genome was not registered. However, L1 insertions in *ATM*-deficient cells were found to be longer (with truncation being less frequent). The authors suppose that *ATM* may recognize intermediates generated during the process of L1 integration as DNA damage and initiate its repair, reducing the length of the resultant retrotransposition events. The correlation between *ATM* deficiency and elevated retrotransposition level discovered in cultured NPCs was confirmed in vivo using *ATM* KO mice with human L1 and EGFP retrotransposition reporter in their genome. Brain samples of 3-month-old *ATM* KO mice had more EGFP-positive cells compared with WT animals. The most noticeable increase

was observed in the hippocampus, and EGFP-positive cells were not apparent in non-neuronal tissues. In the same study, L1 DNA copy number was found to be increased in AT postmortem brain samples compared with matched controls. It is not clear whether L1 retrotransposition contributes to AT pathogenesis [58].

### 3.2.5. Spinal Muscular Atrophy

Spinal muscular atrophy (SMA) is a devastating childhood motor neuron disease that often leads to death. The major pathological characteristic of SMA is selective degeneration of lower  $\alpha$ -motor neurons in the ventral horn of the spinal cord, resulting in progressive muscle denervation, skeletal muscular atrophy and eventual paralysis. SMA is a rare example of a monogenetic neurodegenerative disease: the causative factor of SMA is the loss of SMN, a protein encoded by the *SMN1* gene. The loss of SMN affects not only motor neurons but also some other neurons, muscle cells and some other cells and tissues. SMN is a multifunctional protein, it is expressed ubiquitously, and homozygous deletion of *Smn* in mice is lethal. Humans have two *SMN* genes, and the loss of SMN in SMA is partially compensated by expression from *SMN2*. The *SMN2* gene does not fully substitute for *SMN1* because almost 90% of *SMN2* transcripts lack exon 7, producing a truncated SMN $\Delta$ 7 protein that is quickly degraded. Disease severity varies according to the number of *SMN2* copies carried by the patient. In the most severe form (one *SMN2* copy), disease onset occurs before 6 months of age with death of respiratory distress usually within 2 years. With multiple *SMN2* copies, the disease is characterized by an early-adulthood clinical onset and typically a normal life expectancy [514,515].

Each *SMN* gene contains more than 60 Alu-like sequences. Some *SMN1* mutations in SMA patients were found to be caused by Alu-mediated recombinations, and it was suggested that more yet-unidentified Alu-related mutations in this gene may lead to SMA. However, SMA is a rare example of a neuropathology in which TE-related mechanisms may not only cause the disease but also compensate the damage. Ottesen et al. identified a novel exon within *SMN* gene sequence, 6B, generated by exonization of an Alu element located within intron 6. Transcripts containing exon 6B are not abundant because they are prone to nonsense-mediated decay and Staufen-mediated mRNA decay. However, these transcripts were found in all tissues. The protein produced from these transcripts, SMN6B, has the same length as normal SMN but has different 16 amino acids on the C-terminus. SMN6B is less stable than SMN but more stable than SMN $\Delta$ 7. SMN6B was shown to have similar intracellular localization as SMN and is able to interact with Gemin2, a protein partner of SMN. It is very likely that SMN6B may substitute for SMN loss better than SMN $\Delta$ 7. However, levels of SMN6B produced in SMA patients remain unknown [516].

### 3.2.6. Amyotrophic Lateral Sclerosis and Fronto-Temporal Lobar Degeneration

Two other neurodegenerative disorders—amyotrophic lateral sclerosis (ALS) and fronto-temporal lobar degeneration (FTLD)—were linked to TE activity in the brain and share mechanisms of pathogenesis. Both diseases typically start after the age of 50. ALS is a progressive motor neuron degeneration, and increasing muscle weakness leads to respiratory failure and death in 3–5 years. ALS is usually sporadic—only 5–10% of cases are inherited [517]. At least 25 genes have been linked to ALS, the most important of them are *SOD1* (superoxide dismutase) and *C9ORF72* [518]. FTLD is a heterogeneous syndrome characterized by progressive decline in behavior or language. FTLD patients show gross atrophy of the frontal and anterior temporal lobes: loss of pyramidal neurons, cortical gliosis, loss of myelin and axons. FTLD is a common cause of early-onset dementia. Survival of patients with FTLD is shorter, and their cognitive decline is more rapid compared with AD. FTLD patients may suffer from ALS at the same time. Up to 40% of FTLD patients have a history that suggests familial transmission. Most inherited FTLD cases are caused by mutations in the genes encoding Tau protein and progranulin, leading to abnormal protein aggregation in neurons [519]. Frontotemporal dementia (FTD) is a subgroup of FTLD-associated disorders [520].

One of the key proteins involved in both diseases is TDP-43 (trans-activation response DNA-binding protein of 43 kDa), a multifunctional RNA-binding protein [120,519,521]. Accumulation of TDP-43-containing cytoplasmic inclusions is a shared pathological hallmark in a broad spectrum of neurodegenerative disorders [112,521]. TDP-43 has thousands of RNA targets [120] and multiple functions, including alternative splicing, regulation of mRNA stability and miRNA biogenesis [112]. Some data indicate that TDP-43 must be a stress-responsive RNA-associated factor [521]. In mouse models of TDP-43 proteinopathy, transcription and splicing of hundreds of RNAs were misregulated [522,523].

Li et al. re-analyzed the accessible data of deep sequencing of RNA targets co-purifying with immunoprecipitated TDP-43 from a few studies performed on mice [522–524], rat [524] and human [525] samples. An extensive binding of TDP-43 to RNA derived from all major classes of TEs (LINE, SINE, LTR and DNA transposons) was revealed. It was found that, in mice with disrupted TDP-43 function, multiple LINE, SINE and LTR transcripts are upregulated. In cortical tissue of FTLD patients, binding between TDP-43 and its TE targets is selectively and dramatically reduced compared to control subjects. When similar analysis was performed for FUS, another RNA-binding protein implicated in neurodegenerative disorders, no preferential interaction of FUS with TE-derived RNAs was found. All these findings suppose that TDP-43 normally plays a protective role by specifically binding and silencing multiple transposon transcripts [112]. Liu et al. used subcellular fractionation and FACS to enrich for diseased neuronal nuclei without TDP-43 from post-mortem FTD-ALS human brain samples. The authors showed that loss of nuclear TDP-43 is associated with chromatin decondensation around LINE elements and increased L1 DNA content. The authors noted some specificity toward decondensation of LINEs compared with other repeat elements, but the exact mechanism of this is unclear [526].

Conversely, some other studies suggest that TDP-43 causes activation of TEs (ERVs in particular) and contributes to neuropathology. Krug et al. expressed human TDP-43 in *Drosophila* neurons or glia and reported that it disrupted siRNA silencing, caused derepression of LINE and LTR transposons and led to age-dependent neurological deterioration and reduced lifespan. In fly glial cells (but not in neurons), TDP-43 expression caused unsilencing of *gypsy* (an ERV that is related to human HERV-K). It was found to causally contribute to degeneration, because inhibiting *gypsy* genetically or pharmacologically was sufficient to rescue the phenotypic effects [527]. Combined ectopic expression of TDP-43 and HERV-K in cultured human cells increased HERV-K transcription and reverse transcriptase activity (as compared with cells cotransfected with HERV-K and a control plasmid); KD of endogenous TDP-43 decreased HERV-K transcription. A total of 5 binding sites for TDP-43 in HERV-K promoter (LTR) were identified. In the same study, it was found that HERV-K is activated in a subpopulation of patients with sporadic ALS. Env protein was detected in some cortical and spinal neurons of these patients. Transgenic mice expressing the *env* gene under a neuron-specific promoter showed an expression pattern of the Env protein similar to that of ALS patient tissues. Mutant mice after 2–3 months of age displayed progressive motor dysfunction, selective atrophy of the motor cortex, abnormalities in neuron shape and function, and DNA damage; however, the exact mechanism of HERV-K Env neurotoxicity is not well understood. The authors suggest that virus particles may be transmitted between neurons, explaining the spread of the disease along anatomical pathways in ALS [30].

Pereira et al. suggested that ORF1p protein aggregation may be involved in ALS pathogenesis. They showed that ALS-associated mutant forms of FUS, TDP-43, and SOD1 proteins, but not WT proteins, strongly colocalize with ORF1p in granules in cell cultures. Experiments with cell culture transfections showed that two ALS-associated mutant forms of TDP-43 fail to inhibit L1 retrotransposition with the same efficiency as the WT TDP-43. A similar observation was made for one ALS-associated mutant form of angiogenin. Conversely, two different ALS-associated mutant forms of FUS reduced L1 retrotransposition, while the WT FUS had no such effect [518].

Analyses of transcriptomes of postmortem cortex samples with machine-learning algorithms showed that ALS is most likely a heterogeneous condition, and TE activation

was only observed in one of three distinct molecular subtypes of ALS, comprising 20% of all examined samples. In this subtype, altered expression of genes in multiple pathways linked to TDP-43 was also noted [528].

### 3.2.7. Fragile X-Associated Tremor/Ataxia Syndrome

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset neurodegenerative disorder characterized by a progressive action tremor with ataxia. Patients have degeneration in the cerebellum (including Purkinje cell loss and Bergman gliosis) and intranuclear inclusions in neurons and glia throughout the brain. FXTAS is distinct from fragile X syndrome, but is also caused by an increased number of CGG trinucleotide repeats in the 5'-UTR of the *FMR1* gene. Tan et al. studied a *Drosophila* FXTAS model (rCGG-repeat transgenic flies) that shows a distinct neurodegenerative eye phenotype. The authors measured mRNA levels of different TEs in the brain of mutant flies and registered increased expression of *gypsy* and *copia* LTR elements. Derepression of *gypsy* expression achieved by crossing rCGG-repeat transgenic flies with another mutant line was shown to enhance neurodegenerative eye changes. Accordingly, reduction in *gypsy* expression by RNAi rescued neurodegenerative eye changes. The authors suggest that rCGG-repeat RNA and TE RNA share the same binding proteins, and accumulation of rCGG-repeat RNA leads to sequestration of these proteins and TE unsilencing that contributes to neurodegeneration [470].

### 3.2.8. Multiple Sclerosis

Multiple sclerosis (MS) is a complex multifactorial disease of the CNS. Pathogenesis of MS involves neuroinflammation, death of oligodendrocytes, demyelination, axonal damage and defects in myelin repair, which result in physical and cognitive disabilities [311,529]. In MS, infiltrating lymphocytes and macrophages are observed in the perivascular space of the CNS [530]. Environmental factors, viral infections in particular, are able to trigger MS pathogenesis. Multiple studies suggested a correlation between expression of HERVs (mostly HERV-W) and onset and progression of MS (reviewed in [493]).

Two specific ERVs of the HERV-W family are activated in MS: MSR<sub>V</sub> (MS-associated retrovirus), and ERVWE1 [531]. Env proteins encoded by these two ERVs are called, respectively, MSR<sub>V</sub>env and syncytin-1 (the latter also has a function in placenta formation [296], discussed in the Section 1.8.9). MSR<sub>V</sub>env and syncytin-1 are highly homologous (>94% identity at the RNA level) and very difficult to discriminate. A major difference between MSR<sub>V</sub> and ERVWE1 is that only MSR<sub>V</sub> is found as an extracellular virus, visualized by electron microscopy, while syncytin-1 is detected only intracellularly or on the plasma membrane [531]. MSR<sub>V</sub>env is detected on macrophages or microglial cells in active plaques of MS brains [311]. Similar observations were reported for syncytin [529].

Experiments in cell cultures showed that ectopic expression of syncytin in human fetal astrocytes (HFAs) and monocyte-derived macrophages (MDMs) increased production of proinflammatory cytokine IL-1. Conditioned medium from HFAs (but not MDMs) ectopically expressing syncytin was highly cytotoxic to human or rat oligodendrocytes [529]. Application of MSR<sub>V</sub>env stimulated both adhesion of HL-60 cells (promyeoloblasts) to the monolayer of HCMEC/D3 cells (endothelium) and transmigration of HL-60 cells through the endothelial monolayer. MSR<sub>V</sub>env induced production of pro-inflammatory cytokines in HCMEC/D3 cells and primary endothelial cultures. In the same study, syncytin had a similar but less prominent effect on primary endothelial cultures. KD of toll-like receptor 4 (TLR4) abolished HCMEC/D3 response to MSR<sub>V</sub>env, suggesting that TLR4 is implicated in MSR<sub>V</sub>env recognition [530].

In vivo, MSR<sub>V</sub>env was shown to induce autoimmunity to myelin proteins and cause experimental allergic encephalomyelitis in mice (an animal model for MS) [311]. Mice infected with viral particles carrying the syncytin-expressing vector developed neuroinflammation, oligodendrocyte and myelin damage. These animals also demonstrated decreased coordination, muscle strength and seeking behavior [529].

It was shown that some exogenous viruses may cause reactivation of HERV-W copies. *Herpesviridae* are the most common triggers known to upregulate MSR expression [532].

### 3.2.9. Aicardi-Goutières Syndrome

Aicardi–Goutières syndrome (AGS) is a very rare hereditary interferonopathy with a neurological phenotype. Neurological symptoms include severe encephalitis, demyelination of motor neurons, and psychomotor retardation. Patients have elevated IFN-I levels in CSF. Many AGS symptoms resemble those caused by congenital viral infection. Mutations in seven genes have been confirmed to produce AGS phenotypes; all these genes are involved in regulating cytosolic nucleic acids [168,169,533]. One of the genes linked to AGS is *TREX1* encoding an anti-viral DNase. Stetson et al. purified cytoplasmic DNA fragments from hearts of WT and *Trex1* KO mice. In mutant mice, this DNA was more abundant and more complex. DNA fragments derived from retrotransposons were enriched in the samples from mutant mice. More specifically, L1-, LTR- and SINE-derived cDNAs were identified. Experiments on cell cultures cotransfected with vectors expressing retroelements and *TREX1* confirmed that *TREX1* metabolizes DNA derived from these elements and prevents retrotransposition [168]. Thomas et al. generated *TREX1*-deficient pluripotent stem cells (two cell lines generated from hESCs with introduced mutations in *TREX1*, and an iPSC line from an AGS patient with a naturally occurring *TREX1* mutation). The authors differentiated these cells into NPCs, neurons, and astrocytes. All these cell types exhibited increased levels of ssDNA, as was shown by immunostaining. Application of RT inhibitors to *TREX1*-deficient NPCs reduced ssDNA amount to near-control level. Sequencing of extrachromosomal DNA in *TREX1*-deficient NPCs showed increased level of L1 cDNA, and the relative presence of L1s compared to other transposable elements far exceeded its genomic representation. L1 mRNA levels were elevated in *TREX1*-deficient astrocytes (but not NPCs and neurons) compared to their WT counterparts. Neurons were found to be more vulnerable to the loss of *TREX1* compared to NPCs and astrocytes. *TREX1*-deficient neurons had downregulation of canonical neuronal markers, altered shape, increased levels of cleaved caspase-3, and excessive DNA breakage. These effects were alleviated by RT inhibitors [169]. Another gene mutated in AGS patients and involved in TE silencing is *SAMHD1* encoding a dNTPase. Zhao et al. demonstrated that mutant *SAMHD1* proteins from AGS patients are defective in L1 inhibition. The authors also showed that dNTPase activity may not be critical for *SAMHD1*-mediated L1 inhibition, since a dNTPase-defective *SAMHD1* version maintained the ability to inhibit L1 retrotransposition [170]. Yet another mechanism supposedly linking TE activation with neuroinflammation in AGS is disruption of *ADAR1* A-I editing. Rice et al. performed whole-exome sequencing of patients diagnosed with AGS but screened negative for mutations in *TREX1* and a few other genes known to be relevant to the AGS phenotype. The authors found *ADAR1* mutations in some of the studied patients. All tested individuals with mutations in *ADAR1* showed upregulation of interferon-stimulated genes (assessed by RT-qPCR of 15 targets in whole blood samples). The authors speculated that, since most of the known *ADAR1* editing substrates are TE-derived RNAs, *ADAR1* dysfunction may cause TE dsRNA accumulation which prompts IFN-I response [534].

### 3.2.10. Glioblastoma

Glioblastoma is the most common of all primary malignant CNS tumors, and one of the most lethal of all human cancers. Glioblastomas display striking cellular heterogeneity and plasticity between different cellular states of the tumor [535].

Quite paradoxically, increased L1 mobilization is not associated with brain cancers, even though one might expect this based on L1 unsilencing in both normal neuronal tissue and various cancer types. A few studies showed very little L1 mobilization in glioblastoma [536,537].

Microarray analysis of cellular and exosomal RNA from a low-passage glioblastoma cell line indicated high transcription levels of retrotransposons. The levels of Alu, L1 and, especially, HERV transcripts were frequently higher in microvesicles than in cells (while RNA from DNA transposons was similar in content in cells and microvesicles). This suggests selective packaging of retrotransposon RNA sequences, especially HERV, in tumor microvesicles. The most abundant HERV species in microvesicles from glioblastoma cells was HERV-H. This finding is consistent with HERV-H being the most active HERV in fetal brain cells; glioblastomas likely derive from dedifferentiated glia cells or neuroprecursor cells [538], even though other versions of glioblastoma origin were suggested [535].

Interestingly, it was also found that expression of L2-derived miRNAs and their targets is disturbed in glioblastoma [539].

### 3.2.11. Autism Spectrum Disorders

Autism is a set of heterogeneous neurodevelopmental conditions, characterised by early-onset difficulties in social communication, unusually restricted interests and repetitive behaviors. More than 70% of patients have comorbid conditions, about 45% of individuals with autism have ID, and 32% show loss of previously acquired skills. Electrophysiology, neuroimaging and molecular data suggest that patients with autism have atypical neural connectivity. Autism has high heritability, but its genetics is very complex. Multiple loci were linked to this disorder (up to 1000 by some estimations), and many genetic variants contributing to the autism phenotype have a high degree of pleiotropy [540]. The term “autism spectrum disorders” (ASD) includes autism, Asperger’s syndrome and pervasive developmental disorder-not otherwise specified [541]. Heterogeneity in the phenotypic presentation of ASD makes it difficult to identify specific genes involved in pathogenesis of this condition. Hu et al., using multiple clustering algorithms, suggested dividing ASD population into four phenotypic subgroups based on similarity of symptom severity [542].

Shpyleva et al. studied postmortem brain samples of patients diagnosed with autism and matched controls. Four brain regions were considered: three neocortex areas and the cerebellum. L1 expression was found to be increased in the cerebellum but not in the cortex of autism patients. The binding of MeCP2 and histone H3K9me3 to L1 sequences was lower in autism cerebellum. The link between L1 overexpression and oxidative stress was suggested: the increase in L1 RNA level inversely correlated with glutathione redox status, and the expression of FOXO3 (a TF inducible by oxidative stress) was also increased [176]. Tangsuwansri et al. investigated the association between L1 elements and altered gene expression profiles in ASD using published sequencing data. Genes containing L1 insertions were overrepresented among differentially expressed genes (both up- and downregulated) in ASD blood cells. A total of 351 L1-containing genes were differentially expressed in ASD in at least two independent transcriptomic studies. Ingenuity Pathway Analysis showed that 25 of these genes are associated with autism or ID. Neurodevelopmental disorders comorbid with ASD (mental retardation, cognitive impairment, and neuromuscular disease) were also associated with this gene set. It was found that the percentage of overall L1 methylation is decreased in the subgroup of ASD individuals with severe language deficits, but not in other ASD subgroups [543]. The same research group performed similar analysis for Alu elements. It was found that genes containing Alu insertions are differentially expressed (mostly downregulated) in ASD blood cells. A total of 320 Alu-containing genes were shown to be downregulated in more than one study. Ingenuity Pathway Analysis showed association of these genes with neurological diseases, nervous system development and function. A total of 21 genes were associated with autism or ID. There was no significant difference in AluS methylation between ASD and matched control groups when all ASD individuals were combined; however, significant AluS methylation patterns were associated with specific ASD subgroups [544]. Balestrieri et al. studied expression of four HERV families in peripheral blood mononuclear cells (PBMCs) from ASD patients. RT-qPCR showed that HERV-H expression is upregulated, while HERV-W is downregulated in PBMCs from ASD patients. Cultured PMBCs from ASD individuals

showed an increased potential to upregulate HERV-H expression upon stimulation with a T-lymphocyte-specific mitogen. It was also found that younger ASD patients have higher levels of HERV-H, and that high HERV-H expression is associated with severe impairment of developmental level in ASD [545].

### 3.2.12. Rett Syndrome

Rett syndrome (RTT) is a major neurodevelopmental disorder, characterized by mental retardation and autistic behavior and caused by mutations in the *MeCP2* gene. *MeCP2* encodes methyl-CpG-binding protein 2, an epigenetic regulator highly expressed in the nervous system. This gene is located in the X chromosome, and almost all patients with RTT are female (in males with one *MeCP2* copy its mutation is usually lethal). Approximately 99% of RTT cases are sporadic [546,547]). RTT is associated with comorbid autism [176].

Zhao et al. profiled genome-wide somatic L1 insertions in prefrontal cortex samples obtained from RTT patients. Non-brain tissues (heart, eye, and fibroblasts) were also studied. L1 activity in RTT and control subjects was found to be higher in prefrontal cortex compared to other tissues. In cortical neurons of RTT patients, the pattern of insertions was different compared to controls: unexpectedly, genomes of RTT neurons had fewer insertions within exons. L1 depletion in the exons was significant for long (>100 kb) genes (many of which encode important neuronal proteins [360]). The authors suggest that this observation might reflect the negative selection against cells with multiple exonic insertions (especially within important genes). In cells that already have defective MeCP2, such retrotransposition events would presumably compromise the genome integrity so badly that these cells would die [342]. Muotri et al. used fibroblasts taken from a patient with RTT and from a control subject to generate iPSCs. These cells were then artificially differentiated into NPCs. An L1-EGFP reporter construct was used to measure L1 retrotransposition frequency. Retrotransposition was increased about twofold in NPCs generated from cells of the RTT patient. In the same study, qPCR confirmed that L1 genomic copy number in RTT post-mortem brain samples is higher than in control samples [130].

### 3.2.13. Schizophrenia and Bipolar Disorder

Schizophrenia (SZ) is a devastating chronic illness characterized by psychosis (delusions and hallucinations), alterations in drive and volition, impairment of memory and attention, and affective dysregulation. SZ typically debuts in adolescence and early adulthood. Symptoms and course of the disease vary greatly across individuals [548]. Dorsolateral prefrontal cortex (DLPFC) is one of the brain regions particularly affected in SZ [549,550]. Clinical heterogeneity of SZ is underpinned by considerable genetic heterogeneity. Thousands of common polymorphisms, as well as multiple individually rare CNVs, were suggested to contribute to SZ risk [551]. The inherited component to risk for SZ is estimated at 64%. Genome-wide association studies reported more than 100 genetic loci containing common alleles conveying only minor influences on SZ risk. Rare CNVs increase disease risk substantially, but only in a small fraction of patients diagnosed with SZ [552]. Bipolar disorder (BD) is characterized by mood instability with periods of mania or hypomania and periods of depression. During manic episodes, patients experience psychotic symptoms similar to signs of SZ. Genome-wide association studies suggest genetic overlap between SZ and BD. Many of the risk alleles identified in these studies are implicated in both disorders [553]. Family, twin, and adoption studies show that genetic correlation between SZ and BD is about 60% [554]. Considering this, in the present review, we discuss TE involvement in SZ and BD in the same section. Importantly, a genetic overlap between SZ and neurodevelopmental disorders including autism has also been reported [548,551]. On the other hand, involvement of HERV-W elements and myelin inflammation in pathogenesis of SZ and BD links these diseases with multiple sclerosis [494,555].

Guffanti et al. identified 1689 TEs differentially expressed in the DLPFC of SZ patients using a novel bioinformatics approach. Most of these TEs are located within introns of 1137 protein-coding genes [439].

Bundo et al. reported a high L1 copy number in neurons from the prefrontal cortex of SZ patients. Whole-genome sequencing revealed brain-specific L1 insertions localized preferentially to synapse- and SZ-related genes. In animal models of perinatal environmental risk factors for SZ, brain L1 content was also increased [556]. Doyle et al. identified intragenic L1 insertions absent in the reference genome in postmortem DLPFC samples from SZ patients. At least 23 of 63 genes with such insertions identified in the SZ samples from this and previous [556] studies were known to have an association with SZ. Notably, out of 210 insertions identified by Doyle et al., only one is likely identical by its position to an insertion detected earlier by Bundo et al. It is located within an intron of the *NPAS1* gene [552] encoding a TF that regulates multiple genes involved in SZ and ID [557]. Guffanti et al. performed a deep whole-genome sequencing of blood samples from six members of the same family, four of which were diagnosed with SZ, to identify non-reference L1 insertions. Using two different computational algorithms, the authors registered 110 such insertions in this family (about 40 insertions per subject). A total of 52 non-reference L1 insertion loci mapped to protein-coding or RNA genes. It was found that L1 polymorphic insertions in SZ subjects have an impact on genes previously implicated in SZ, such as *GABRB1*, *FHIT*, and *RYR3* [473]. In a recent study, Zhu et al. analyzed L1 and Alu retrotransposition in sorted cells using deep whole-genome sequencing and a machine-learning method. Two brain-specific L1 insertions detected in the same SZ donor were thoroughly investigated. Both insertions were within introns of genes previously associated with neuropsychiatric disorders (*CNNM2* and *FRMD4A*). *CNNM2* is located within a locus that has been associated with SZ, and the L1 insertion within *CNNM2* affected a putative transcriptional regulatory element. For insertions in *CNNM2* and *FRMD4A*, respectively, mosaicism levels quantified using ddPCR were 0.72% and 1.2% in neurons and 0.54% and 0.53% in glia. Both insertions were present in neurons and glia from different brain regions across both hemispheres, indicating that these two retrotransposition events occurred during early embryogenesis. Cell culture experiments with reporter gene system confirmed that these insertions could reduce *CNNM2* and *FRMD4A* expression in the subject's brain cells [474].

HERV-W RNA was identified in plasma and CSF from patients with SZ and related pathologies [558]. Frank et al. performed a microarray-based analysis of HERV *pol* transcriptional activity in DLPFC samples from individuals with SZ or BD and matched controls. HERV *pol* expression profiles of patient groups were found to be generally similar to those of the control group. The most active HERV elements were equally represented in all three groups. Among less active elements, HERV-IP, NMWV7, and HERV-L were overrepresented in SZ compared to BD specimens. Seq63 and NMWV7 elements were underrepresented in BD samples compared to control samples. Most notably, HERV-K10 was significantly overrepresented in both BD and SZ samples compared to controls [31]. Perron et al. addressed expression levels and mutations in the MSR *env* gene in peripheral mononuclear cells collected from patients with SZ and BD. MSR *env* transcription was elevated in both patient groups compared with healthy controls. MSR *env* overexpression was more prominent in the BD group. Paradoxically, MSR *env* DNA copy number was lower in the genomes of patients with BD or SZ compared with the control group. Differences in nucleotide sequence of MSR *env* were found between all three groups. The authors also noted that, in patients with BD or SZ molecular characteristics of MSR, *env* are different from those reported previously in patients with MS [494].

### 3.2.14. Major Depressive Disorder

Major depressive disorder (MDD) is a psychiatric disorder defined by episodes of depressed mood lasting for more than 2 weeks with disturbed sleep and appetite, excessive guilt, reduced concentration, slowed movements and suicidal thoughts. MDD affects approximately 17% of the population. Both genetic and environmental factors are thought to be involved in the etiology of MDD. Liu et al. examined DNA samples from peripheral blood cells and showed that MDD patients have increased L1 genomic copy number and decreased L1 5'-UTR methylation level compared with healthy control subjects. In the same

study, mice were subjected to chronic unpredictable mild stress as a model of depression. L1 copy number in peripheral blood cells of stressed mice was increased, just like in MDD patients. However, in the prefrontal cortex of stressed mice, L1 copy number was decreased compared with the control group (while no change was found in other studied brain regions) [559].

### 3.2.15. Post-Traumatic Stress Disorder

Severe stress or trauma may cause permanent changes in brain circuitry, leading to dysregulation of fear responses and the development of post-traumatic stress disorder (PTSD). PTSD is considered an anxiety disorder. Symptoms include re-experiencing the trauma, avoiding stimuli associated with the trauma, numbing of general responsiveness, and increased arousal. Many PTSD patients show an exaggerated reaction to mild stressors. The amygdala is the key brain structure involved in the consolidation and recall of traumatic memories [390]. PTSD is quite common in combat troops after deployment. Rusiecki et al. examined LINE and Alu methylation in serum DNA from US military soldiers with diagnosed PTSD and matched controls; in all cases, serum samples taken before and after deployment were available. It was found that pre-deployment Alu methylation level was higher in PTSD subjects versus controls; these groups did not differ post-deployment. In the control group, but not in the PTSD group, L1 methylation increased post-deployment [560]. An animal model of PTSD is stress-enhanced fear learning (SEFL) where animals are pre-exposed to a stressor before classic fear conditioning. Ponomarev et al. used the SEFL paradigm with and without exposure to isoflurane in the concentration blocking the behavioral effects of stress and performed gene expression profiling in lateral/basolateral amygdala 3 weeks after the exposure. It was found that L1 transcripts were upregulated by shock and downregulated by isoflurane. The authors suggested that L1 expression in SEFL may simply reflect general chromatin remodeling during unsilencing of multiple genes [390].

### 3.2.16. Drug Addiction and Alcoholism

Drug addiction is a chronic relapsing disorder characterized by maladaptive learning about drugs and predictive cues that leads to compulsive drug seeking. Genetic factors contribute to drug addiction, but it is, primarily, a behavioral learning disorder that only develops after repeated exposure to a drug of abuse. The dopamine signaling pathway involved in reward learning is dysregulated in drug addiction; the most important brain region in this pathway is *nucleus accumbens* (NAc). Recently, the role of epigenetic changes in drug-induced plasticity was discovered [561,562]. Some drugs of abuse were shown to be able to induce L1 retrotransposition.

In experiments with neuronal cell lines, Okudaira et al. demonstrated that methamphetamine and cocaine in non-cytotoxic concentrations trigger L1 mobilization. Other drugs tested in these experiments, ethanol and sodium barbital, did not have the same effect. It was confirmed that methamphetamine and cocaine do not induce DSBs. Both drugs had no effects on the expression or splicing of L1 RNA. A mutation in the reverse transcriptase domain of ORF2 prevented the drug-induced retrotransposition. It was also shown that methamphetamine or cocaine recruited ORF1p to chromatin in a CREB-dependent manner [187]. A few other studies showed a link between cocaine administration and TE activation in vivo. Voskresenskiy et al. administered cocaine chronically to pregnant rats and registered increased L1 ORF1 expression in blood cells of the mothers and in brain, spinal cord and heart of the offspring at the first day of life [563]. Maze et al. reported that, in mice, both acute and repeated cocaine administration dynamically alters H3K9me3 distribution and heterochromatin formation in NAc (but not in other brain regions implicated in cocaine action). ChIP-Seq experiments showed that repeated cocaine administration significantly changed H3K9me3 level at 16 repetitive element types, and for 11 of these H3K9me3 level was reduced. The most prominent H3K9 demethylation happened in L1 elements, and L1 unsilencing was confirmed by RT-qPCR [564]. Doyle et al. studied postmortem samples of

medial prefrontal cortex from individuals with cocaine addiction. The authors used L1-seq, a specialised method to identify specific L1 insertions by their flanking sequences, in sorted neuronal and non-neuronal nuclei. Novel L1 insertions were identified in neurons from both addiction and control brain samples. Notably, in the addiction group, these insertions were enriched in GO terms and pathways previously associated with cocaine addiction. Comparison between different samples suggested that most L1 insertions identified in the addiction group are either germline or occurred de novo at an early developmental stage. This means that these insertions could influence predisposition to the addiction, rather than be a consequence of cocaine intake [565].

Another drug of abuse associated with TE activation is morphine. Trivedi et al. characterized the influence of morphine on DNA methylation in cultured neuronal SH-SY5Y cells. L1 methylation (assessed by the MethylCap-Seq method) was increased after 4 h of morphine treatment and decreased after 24 h of treatment. Correspondingly, L1 RNA expression was decreased after short-term morphine treatment and increased after 24 h of treatment. Changes in L1 methylation were in agreement with changes in global DNA methylation induced by morphine, even though methylation level of individual CpG regions in L1 was not predictive of global methylation level [566]. Sun et al. demonstrated that H3K9me2 repressive mark is downregulated in mouse NAc (but not in dorsal striatum) by repeated morphine administration. ChIP-seq was then used to map the distribution of H3K9me2 binding, and 24 types of repetitive elements were found to have significant morphine regulation, most of them showing decreased H3K9me2 enrichment and de-repression. Among unsilenced TEs were five L1 subfamilies [567].

Ponomarev et al. performed a gene expression profiling in postmortem human brain samples of alcoholics. It was shown that alcohol abuse resulted in upregulation of LTR retrotransposons. Overrepresentation analysis of coexpression modules identified several modules that showed significant enrichment with TEs in all brain regions. Many of these TEs have retained functional promoters, and DNA methylation in the LTR region of these transposons was decreased. The authors suggested that these TE promoters may regulate the expression of adjacent genes [568]. Only one Illumina probe corresponding to a retrotransposition-competent L1HS family element was used in this study, and this element was found to be highly upregulated in alcoholic brain [120] (Ponomarev, personal communication).

### 3.2.17. Creutzfeldt-Jakob Disease

Creutzfeldt–Jakob disease (CJD) is a rare, fatal neurodegenerative disorder caused by prion proteins. The most common form of CJD is sporadic and is diagnosed in one person per million per year worldwide. The key molecular event in the pathogenesis of CJD is the conformational change of cellular prion protein PrP<sup>C</sup> to its abnormal isoform PrP<sup>Sc</sup>. Accumulation of PrP<sup>Sc</sup> multimers induces vacuolation of gray matter by microglial activation and neuronal loss, leading to progressive neurodegeneration and astrogliosis. Usually, it manifests as a rapidly progressing dementia with ataxia and myoclonus, leading to death within one year [569,570].

Jeong et al. investigated the prevalence of different HERV families in CSF samples from healthy controls, patients with sporadic CJD, and patients with other (not specified in the paper) neurological diseases. RT-PCR showed that levels of HERV-W, HERV-L, FRD and ERV-9 are increased in CSF of individuals with sporadic CJD compared to controls. Concentrations of HERV-W and HERV-L were higher in CJD samples than in samples from the “other diseases” group. Importantly, the HERV-L family has not previously been shown to be associated with any other disease. The exact role of HERV expression in the pathogenesis of sporadic CJD is not clear, but the authors suggested it may be involved in neuroinflammatory processes [571].

### 3.2.18. Neurofibromatosis Type I

Neurofibromatosis type I (NF1) is the most common autosomal dominant disorder affecting the nervous system [572], as well as the most common neurocutaneous disease [573]. Birth incidence of NF1 is estimated to be between 1:2000 and 1:3500 [364,572,574–576]. NF1 is a multisystem genetic disorder; it has a highly variable phenotype; and its symptoms include dermatological, ocular, neurological, musculoskeletal, vascular and cardiac manifestations, and tumors [575,576]. While most patients with NF1 have normal intelligence, their motor development is frequently delayed. Some patients show learning disabilities, ID, ASDs or epileptic seizures [573,574,576]. Learning or attention deficits or behavioral problems are evident in 50–80% of patients [576]. NF1 patients are prone to the development of benign and malignant tumors [572]; they have a five-fold risk for cancer in general and more than 2000-fold risk for neurogenic malignancies compared to the general population [575].

The only gene responsible for this disease is *NF1* [576] encoding neurofibromin 1, a protein participating in Ras and cAMP signaling pathways; in neurons it is involved in synaptic signaling and neural development [364,575,577]. No mutational hotspots have been identified in *NF1* [575]. However, it seems that this gene is particularly vulnerable for TE insertions. Wimmer et al. described 18 novel retrotransposon insertions within *NF1* in patients with neurofibromatosis type I. Strikingly, three of the identified TE integration sites were each used twice in non-related patients. Moreover, 6 out of 18 insertions were located close to each other, within a specific 1.5 kb region that constitutes less than 1% of this long gene. All this suggests that L1-mediated retrotransposition is not as random as previously thought, and some genomic sites are particularly vulnerable to it. Retrotransposon insertions account for ~0.4% of all *NF1* mutations, which is 2–4 times higher than expected considering the average insertion rate. Splicing defects were shown to be the prevalent effect of L1-mediated insertions identified in this study; there were 13 exonic Alu, L1 and poly(T) insertions that led to exon skipping or use of a cryptic splice site [363].

### 3.2.19. Age-Related Macular Degeneration

Age-related macular degeneration (AMD) results from retinal pigmented epithelium (RPE) degeneration and is a leading cause of blindness. One of molecular traits specific for AMD is downregulation of *DICER1* expression; other miRNA-processing enzymes are not affected. Kaneko et al. sequenced dsRNA extracted from human RPE with AMD and identified increased level of Alu RNAs compared to control RPE samples. Experiments with pol III inhibitor applied to cultured RPE cells confirmed that Alu RNA molecules accumulating as a result of *DICER1* deficiency are primary Alu transcripts and not sequences embedded in pol II-transcribed RNAs. Transfection of human or mouse RPE cells with a plasmid encoding Alu reduced cell viability. In vivo, subretinal transfection of plasmids encoding two different Alu RNAs induced RPE degeneration in WT mice; similar results were obtained with plasmids encoding B1 or B2 RNAs. Alu RNA accumulation in AMD and its experimental models correlated with increased caspase-3 cleavage, indicating a role for excess Alu RNA in apoptosis [578]. A following study by the same research group uncovered that the mechanism of Alu RNA cytotoxicity involves inflammatory response. Despite the immune privilege of the retina, multiple innate immune receptors are expressed in the RPE. Tarallo et al. tested a number of dsRNA sensor proteins for their interaction with Alu RNA in RPE cells using genetically modified mice strains. MyD88, a protein previously considered to be specific for immune cells, was identified as a dsRNA sensor responsible for Alu RNA cytotoxicity in RPE monoculture. Conditional KO of *MyD88* in the RPE protected against Alu RNA-induced RPE degeneration in vivo. Further experiments identified other participants of Alu RNA-mediated inflammatory response in RPE cells: IL-18, NLRP3 inflammasome and caspase-1 [471].

### 3.2.20. X-Linked Dystonia-Parkinsonism

X-linked dystonia-parkinsonism (XDP) is a Mendelian neurodegenerative disease endemic to one island in Philippines. The clinical phenotype combines features of dystonia and parkinsonism, beginning with hyperkinetic symptoms that shift to hypokinetic movements over time; the mean disease onset age is 39.7 years. Aneichyk et al. integrated multiple genome and transcriptome assembly technologies to identify the mutation causing this disease. They found out that this is a SVA insertion within an intron of the *TAF1* gene [579].

### 3.2.21. Ravine Encephalopathy

A very rare inherited condition was identified in a small population in the Ravine region of Reunion Island. This disease is characterized by extreme infantile anorexia and lethal encephalopathy with brainstem atrophy. It was associated with a point mutation in a degenerated TE (L1PA8 embedded in AluSz) in an intron of the *SLC7A2* gene. Two lncRNAs with unknown function, *SLC7A2-IT1A* and *SLC7A2-IT1B*, originate from this intron. The mutation greatly increases the probability for the mutant transcripts to be targeted by piRNA-mediated silencing. More than eight-fold decrease in *SLC7A2-IT1A/B* RNA level in patient brain tissue was registered. KD of *SLC7A2-IT1A/B* in human neuroblastoma cells caused their apoptosis, suggesting the causal role of *SLC7A2-IT1A/B* silencing in Ravine encephalopathy [429].

## 4. Conclusions

TEs have fascinatingly diverse and complex interactions with host genomes. There are multiple examples of TE exaptation in different tissues and organs, and the introduction of whole-genome approaches in recent years allowed us to realize that the extent of TE involvement in multiple gene-regulation processes is actually much higher than had been previously considered.

Some germline TE insertions caused formation of new genes. Such TE domestication events played a role in the evolution of different cell types and tissues. It seems like domestications of a number of genes encoding neuron-specific proteins and non-coding RNAs happened by random chance rather than by some predictable pattern; the resulting genes are diverse in their function and derived from different TE families.

Apparently, the most prominent role of TE-related germline mutations in the human brain evolution was providing new regulatory elements. Of particular note is the impact of rapidly expanding Alu insertions in human evolution: it seems like these elements disproportionately contributed to the evolution of neuron-specific genes and their transcriptional regulation.

There is extensive evidence that in the brain, and specifically in the hippocampus, somatic TE insertions are much more common compared to non-neuronal somatic tissues. Most of the studies demonstrating increased TE mobilization in the brain did not discriminate between TE insertions that happened during and after neuronal differentiation; it should be noted that hippocampus includes an adult neurogenesis niche that contains NPCs. A number of in-vitro studies focused on NPCs showed that notable TE derepression (and activation of L1 elements in particular) happens during neuronal differentiation. This process is not random; some specific TE types were shown to be silenced in neurogenesis. It was suggested that at this stage L1-derived regulatory sequences are used to promote expression of specific genes. Another proposed role of TE mobilization in the context of neurogenesis is increasing heterogeneity in newly generated neurons. However, somatic retrotransposition was also confirmed to happen in adult neurons, but its function remains obscure.

TE activity was linked to multiple diseases with diverse mechanisms. TE mobilization or transcription was registered in various neuropathologies. In many cases, it is not obvious if there is a causal link between TE activity and pathological changes in neurons. HERV and L1 elements are associated with most of the diseases described in this review, since

activation of these elements causes mutations and triggers innate immunity cascades. It was suggested that, at least in some cases, TE activation in neuropathology may simply reflect genome destabilization caused by some other factors.

TE involvement in myriad of normal and pathological processes emphasizes the importance of further studies in this field. However, TE abundance and repetitive nature make pinpointing specific transposition events particularly hard, especially in somatic cells. It is also important to highlight that functional assays are necessary to establish the causal role of transposition events in specific processes.

**Author Contributions:** Conceptualization, E.C.; Formal analysis, A.B.; Funding acquisition, P.K.; Investigation, E.C. and A.B.; Project administration, P.K.; Supervision, P.K.; Visualization, E.C. and A.B.; Writing—original draft, E.C. and A.B.; Writing—review and editing, E.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was supported by Russian Ministry of Science and High Education (agreement 075-15-2020-801).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets provided by third parties were analyzed in this study. These data can be found at [https://www.cell.com/ajhg/fulltext/S0002-9297\(07\)63704-5#supplementaryMaterial](https://www.cell.com/ajhg/fulltext/S0002-9297(07)63704-5#supplementaryMaterial) (accessed on 20 May 2020). This publication is also included in the References list.

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

## Abbreviations

6-OHDA	6-hydroxydopamine
7SL RNA	signal recognition particle 7S RNA
ABCD1	ATP binding cassette subfamily D member 1
AD	Alzheimer's disease
ADAR	adenosine deaminase that acts on RNA
ADNP	activity-dependent neuroprotector homeobox protein
AGS	Aicardi–Goutières syndrome
ALS	amyotrophic lateral sclerosis
AluI	<i>Arthrobacter luteus</i> endonuclease
AMD	age-related macular degeneration
AmnSINE1	<i>Amniota</i> SINE1
AMPA	$\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
APOBEC3	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3
APOE	apolipoprotein E
APP	amyloid $\beta$ precursor protein
ARC	activity-regulated cytoskeleton-associated protein
ARE	adenylate-uridylate-rich elements
ASC	anthropoid-specific constrained region
ASD	autism spectrum disorders
AT	ataxia telangiectasia
ATM	ataxia telangiectasia mutated
AVPR1A	arginine vasopressin receptor 1A
A $\beta$	amyloid $\beta$
BCYRN1	brain cytoplasmic RNA 1
BD	bipolar disorder
BDNF	brain derived neurotrophic factor
BER	base excision repair

bp	base pair
C19MC	chromosome 19 microRNA cluster
C9ORF72	guanine nucleotide exchange factor C9orf72
CCFDN	congenital cataracts facial dysmorphism neuropathy
CENP-B	centromere-associated protein B
ChIP-PCR	chromatin immunoprecipitation-polymerase chain reaction
ChIP-Seq	chromatin immunoprecipitation sequencing
CHRNA9	cholinergic receptor nicotinic $\alpha$ 9 subunit
CJD	Creutzfeldt–Jakob disease
CMAH	cytidine monophospho-N-acetylneuraminic acid hydroxylase
CNNM2	cyclin and CBS domain divalent metal cation transport mediator 2
CNS	central nervous system
CNTN5	contactin 5
CNV	copy number variation
CREB	cAMP responsive element binding protein 1
CRISPR/Cas	clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins
CSF	cerebrospinal fluid
CTCF	CCCTC-binding factor
CYP20A1	cytochrome P450 family 20 subfamily A member 1
DBH	dopamine $\beta$ -hydroxylase
DDP	DNA-dependent DNA polymerase
ddPCR	droplet digital polymerase chain reaction
DG	dentate gyrus
DHS	DNase I-hypersensitive site
DLPFC	dorsolateral prefrontal cortex
DNMT	DNA methyltransferase
DNMT3L	DNA methyltransferase 3-like
DR2	direct repeats of RGKTCA motifs separated by 2 bp
DR4	direct repeats of RGKTCA motifs separated by 4 bp
DSB	DNA double-strand break
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
EGFP	enhanced green fluorescent protein
EGR1	early growth response 1
EN	engrailed homeobox
ENi	endonuclease-independent
Env	envelope protein
ERCC2	ERCC excision repair 2, TFIIH core complex helicase subunit
ERV	endogenous retrovirus
ESCs	embryonic stem cells
eSINE	enhancer SINEs
EV	extracellular vehicle
FACS	fluorescent-activated cell sorting
FGF	fibroblast growth factor
FHIT	fragile histidine triad diadenosine triphosphatase
FMR1	fragile X messenger ribonucleoprotein 1
FOS	FBJ osteosarcoma oncogene
FOX	forkhead box
FOXO3	forkhead box O3
FOXP2	forkhead box P2
FRMD4A	FERM domain containing 4A
FTD	fronto-temporal dementia
FTLD	fronto-temporal lobar degeneration
FUS	fused in sarcoma

---

FXTAS	fragile X-associated tremor/ataxia syndrome
GABA	$\gamma$ -aminobutyric acid
GABRB1	$\gamma$ -aminobutyric acid type A receptor subunit $\beta$ 1
GADD45B	growth arrest and DNA damage inducible $\beta$
Gag	group-specific antigen
GFP	green fluorescent protein
GH	growth hormone
GluA2	glutamate ionotropic receptor AMPA type subunit 2
GO	gene ontology
GPR56	G-protein-coupled receptor 56
HCN	hippocampus neural stem cells
HD	Huntington's disease
HDAC1	histone deacetylase 1
HERV	human endogenous retrovirus
hESCs	human embryonic stem cells
HetA	healing transposon
HFA	human fetal astrocyte
HIV1-Tat	human immunodeficiency virus type 1 trans-activator of transcription
hnRNPA1	heterogeneous nuclear ribonucleoprotein A1
HOX	homeobox
HR	homologous recombination
HSCs	hematopoietic stem cells
HTT	huntingtin
hYRNA	human RNA of the Y family
ID	intellectual disability
IFN	interferon
IL	interleukin
ILF3	interleukin enhancer-binding factor 3
iPSCs	induced pluripotent stem cells
IRES	internal ribosome entry site
IRF1	interferon regulatory factor 1
ISL1	ISL LIM homeobox 1
KAP1	KRAB-associated protein 1
kb	thousand base pairs
KD	knock-down
KDM4B	lysine demethylase 4B
KO	knock-out
KRAB-ZFP	Krüppel-associated box domain zinc finger protein
KZNF	Krüppel-associated box domain zinc finger protein
L1Hs	L1 human specific
lacZ	$\beta$ -galactosidase
LDOC1	LDOC1 regulator of NF $\kappa$ B signaling
LF-SINE	'living fossil' SINE
lincRNA	long intergenic non-coding RNA
LINE	long interspersed nucleotide element
lncRNA	long non-coding RNA
LTM	long-term memory
LTR	long terminal repeat
LXRE $\alpha$	liver X receptor $\alpha$
MAP2	microtubule-associated protein 2
MAPT	microtubule-associated protein tau
MART	mammalian retrotransposon transcript
Mb	million base pairs
mDa	mesencephalic dopaminergic neurons

MDD	major depressive disorder
MDM	monocyte-derived macrophage
MeCP2	methyl-CpG-binding protein 2
MER	medium reiterated repeat
MERVL	mouse endogenous retrovirus type L
MethylCap-Seq	Methyl-CpG binding domain-based capture and sequencing
MILI	mouse PIWI-like
MIR	mammalian-wide interspersed repeat
miRNA	microRNA
MITF-M	melanoma specific microphthalmia-associated transcription factor
MMR	DNA mismatch repair
Mov10	Mov10 RISC complex RNA helicase
MPP+	1-methyl-4-phenylpyridinium
mRNP	messenger RNA
mRNP	messenger ribonucleoprotein
MS	multiple sclerosis
MSCs	mesenchymal stem cells
MTH1	methylated purine nucleoside triphosphate hydrolase
My	million years
MYC	MYC proto-oncogene, bHLH transcription factor
MyD88	myeloid differentiation primary response 88
NAc	<i>nucleus accumbens</i>
NCAI	non-classical Alu insertion
NCLI	non-classical L1 insertions
NDUFS2	NADH:ubiquinone oxidoreductase core subunit S2
NEE	novel enriched environmental conditions
NER	nucleotide excision repair
Neu5Gc	<i>N</i> -glycolylneuraminic acid
NeuroD	neuronal differentiation
Neurog	neurogenin
NF1	neurofibromatosis type I
NF90	nuclear factor of activated T-cells, 90 kD
NFI	nuclear factor I
NF-κB	nuclear factor kappa B
NF-κB1	nuclear factor kappa B subunit 1
NHEJ	non-homologous end joining
NLS	nuclear localization signal
NPAS1	neuronal PAS domain protein 1
NPCs	neural precursor cells
NRIF	neurotrophin receptor-interacting factor
NRSF	neuron-restrictive silencer factor
NSCs	neural stem cells
NUDT1	nudix hydrolase 1
OPA1	optic atrophy protein 1
ORF	open reading frame
OXT	oxytocin/neurophysin I prepropeptide
p75NTR	nerve growth factor receptor
PA	passive avoidance
PARK2	Parkinson juvenile disease protein 2
PARK7	Parkinsonism-associated deglycase
PBMCs	peripheral blood mononuclear cells
PCBP2	poly(RC) binding protein 2
PD	Parkinson's disease
PDHA1	pyruvate dehydrogenase E1 subunit α1

PEG	paternally expressed
piRNA	PIWI-interacting RNA
PIT-1	pituitary-specific positive transcription factor 1
PIWI	P-element induced wimpy testis
PKR	protein kinase R
POGZ	<i>pogo</i> transposable element with zinc finger domain
pol II	RNA polymerase II
pol III	RNA polymerase III
POMC	proopiomelanocortin
PRC2	Polycomb repressive complex 2
PRKN	parkin RBR E3 ubiquitin protein ligase
PrP	prion protein
PTSD	post-traumatic stress disorder
qPCR	quantitative polymerase chain reaction
RAD52	RAD52 homolog, DNA repair protein
RC-L1	retrotransposition-competent L1 elements
RC-Seq	retrotransposon capture sequencing
rDNA	ribosomal DNA
RDP	RNA-dependent DNA polymerase
REST	RE1-silencing transcription factor
RHOXF2	Rhox homeobox family member 2
RNP	ribonucleoprotein particle
ROS	reactive oxygen species
RPA1	replication protein A1
RPE	retinal pigmented epithelium
rRNA	ribosomal RNA
RT	reverse transcription
RTL	retrotransposon Gaglike
RT-qPCR	quantitative reverse transcription polymerase chain reaction
RTT	Rett syndrome
RUNX3	RUNX family transcription factor 3
RYR3	ryanodine receptor 3
SAMHD1	SAM domain and HD domain-containing protein 1
SCAN	SRE-ZBP, CTfin51, AW-1 and number 18 cDNA
SEFL	stress-enhanced fear learning
SETDB1	SET domain bifurcated histone lysine methyltransferase 1
SETMAR	SET domain and mariner transposase fusion gene
sgRNA	single guide RNA
SINE	short interspersed nucleotide element
SINEUP	SINE element-containing translation up-regulator
SIRH	<i>sushi-ichi</i> -related retrotransposon-homolog
siRNA	small interfering RNA
SIRT6	sirtuin 6
SLAV	somatic L1-associated variants
SLC7A2	solute carrier family 7 member 2
SMA	spinal muscular atrophy
SMN1	survival of motor neuron 1, telomeric
SMN2	survival of motor neuron 2, centromeric
SNAR	small NF90 (ILF3) associated RNA
SNpc	<i>substantia nigra pars compacta</i>
SNV	single-nucleotide variant
SOD1	superoxide dismutase
SOX	SRY-associated high mobility group box
SRY	sex-determining region Y
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA

STAT	signal transducer and activator of transcription
STG	superior temporal gyrus
STM	short-term memory
SUV39H1	suppressor of variegation 3–9 homolog 1
SVA	SINE-R, VNTR, and Alu
SVZ	subventricular zone
SZ	schizophrenia
Ta	transcriptionally active
TACR3	tachykinin receptor 3
TAD	topologically associating domain
TAF1	TATA-box binding protein associated factor 1
TAHRE	telomere associated and HeTA-related
TART	telomere associated retrotransposon
TBX	T-box
TCF/LEF	T-cell factor/lymphoid enhancer factor
TC-NER	transcription-coupled nucleotide excision repair
TDP-43	TAR DNA-binding protein 43
TE	transposable element
TERT	telomerase reverse transcriptase
TF	transcription factor
TFIIIC	general transcription factor IIIC
TFR	transposon-free region
TLR4	toll-like receptor 4
TNPO1	transportin 1
TPRT	target-site primed reverse transcription
TR	telomerase RNA
TREX1	three-prime repair exonuclease 1
TRIM28	tripartite motif containing 28
tRNA	transfer RNA
TRPV3	transient receptor potential cation channel subfamily V member 3
Ty	transposons of yeast
U snRNA	uridine-rich small nuclear RNA
UCHL1	ubiquitin carboxy-terminal hydrolase L1
UTR	untranslated region
UV	ultraviolet
VNTR	variable number of tandem repeats
Wnt3a	wingless-type MMTV integration site family member 3a
WT	wild type
XPD	X-linked dystonia-parkinsonism
YY1	Yin Yang-1
ZCCHC16	zinc finger CCHC domain-containing protein 16
ZNF	zinc finger protein

## References

1. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 344. [[CrossRef](#)]
2. Elbarbary, R.A.; Lucas, B.A.; Maquat, L.E. Retrotransposons as regulators of gene expression. *Science* **2016**, *351*, aac7247. [[CrossRef](#)] [[PubMed](#)]
3. 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](#)] [[PubMed](#)]
4. van't Hof, A.E.; Campagne, P.; Rigden, D.J.; Yung, C.J.; Lingley, J.; Quail, M.A.; Hall, N.; Darby, A.C.; Saccheri, I.J. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **2016**, *534*, 102–105. [[CrossRef](#)] [[PubMed](#)]
5. Cordaux, R.; Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703. [[CrossRef](#)] [[PubMed](#)]
6. Bourque, G.; Burns, K.H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H.L.; Macfarlan, T.S.; et al. Ten things you should know about transposable elements. *Genome Biol.* **2018**, *19*, 199. [[CrossRef](#)] [[PubMed](#)]
7. Linker, S.B.; Marchetto, M.C.; Narvaiza, I.; Denli, A.M.; Gage, F.H. Examining non-LTR retrotransposons in the context of the evolving primate brain. *BMC Biol.* **2017**, *15*, 68. [[CrossRef](#)]
8. Kramerov, D.A.; Vassetzky, N.S. SINEs. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 772–786. [[CrossRef](#)]

9. Ferrari, R.; Grandi, N.; Tramontano, E.; Dieci, G. Retrotransposons as Drivers of Mammalian Brain Evolution. *Life* **2021**, *11*, 376. [[CrossRef](#)]
10. Dieci, G.; Conti, A.; Pagano, A.; Carnevali, D. Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochim. Biophys. Acta* **2013**, *1829*, 296–305. [[CrossRef](#)]
11. Walters, R.D.; Kugel, J.F.; Goodrich, J.A. Invaluable junk: The cellular impact and function of Alu and B2 RNAs. *IUBMB Life* **2009**, *61*, 831–837. [[CrossRef](#)] [[PubMed](#)]
12. Hermant, C.; Torres-Padilla, M.-E. TFs for TEs: The transcription factor repertoire of mammalian transposable elements. *Genes Dev.* **2021**, *35*, 22–39. [[CrossRef](#)] [[PubMed](#)]
13. Platt, R.N.; Vandeweghe, M.W.; Ray, D.A. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* **2018**, *26*, 25–43. [[CrossRef](#)] [[PubMed](#)]
14. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
15. Eickbush, T.H.; Jamburuthugoda, V.K. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* **2008**, *134*, 221–234. [[CrossRef](#)] [[PubMed](#)]
16. 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](#)] [[PubMed](#)]
17. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.; Smit, A.F.A.; Wheeler, T.J. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [[CrossRef](#)]
18. de Koning, A.P.J.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet.* **2011**, *7*, e1002384. [[CrossRef](#)]
19. Kojima, K.K. Human transposable elements in Repbase: Genomic footprints from fish to humans. *Mob. DNA* **2018**, *9*, 2. [[CrossRef](#)]
20. Savage, A.L.; Bubb, V.J.; Breen, G.; Quinn, J.P. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol. Biol.* **2013**, *13*, 101. [[CrossRef](#)]
21. Schaack, S.; Gilbert, C.; Feschotte, C. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **2010**, *25*, 537–546. [[CrossRef](#)] [[PubMed](#)]
22. Platt, R.N.; Mangum, S.F.; Ray, D.A. Pinpointing the vesper bat transposon revolution using the *Miniopterus natalensis* genome. *Mob. DNA* **2016**, *7*, 12. [[CrossRef](#)] [[PubMed](#)]
23. Palazzo, A.; Caizzi, R.; Moschetti, R.; Marsano, R.M. What Have We Learned in 30 Years of Investigations on Bari Transposons? *Cells* **2022**, *11*, 583. [[CrossRef](#)] [[PubMed](#)]
24. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)]
25. Jern, P.; Coffin, J.M. Effects of Retroviruses on Host Genome Function. *Annu. Rev. Genet.* **2008**, *42*, 709–732. [[CrossRef](#)]
26. Kassiotis, G.; Stoye, J.P. Immune responses to endogenous retroelements: Taking the bad with the good. *Nat. Rev. Immunol.* **2016**, *16*, 207–219. [[CrossRef](#)]
27. 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](#)]
28. Löwer, R.; Boller, K.; Hasenmaier, B.; Korbmayer, C.; Müller-Lantzsch, N.; Löwer, J.; Kurth, R. Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4480–4484. [[CrossRef](#)]
29. Boller, K.; König, H.; Sauter, M.; Mueller-Lantzsch, N.; Löwer, R.; Löwer, J.; Kurth, R. Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology* **1993**, *196*, 349–353. [[CrossRef](#)]
30. Li, W.; Lee, M.-H.; Henderson, L.; Tyagi, R.; Bachani, M.; Steiner, J.; Campanac, E.; Hoffman, D.A.; von Geldern, G.; Johnson, K.; et al. Human endogenous retrovirus-K contributes to motor neuron disease. *Sci. Transl. Med.* **2015**, *7*, 307ra153. [[CrossRef](#)]
31. Frank, O.; Giehl, M.; Zheng, C.; Hehlmann, R.; Leib-Mösch, C.; Seifarth, W. Human Endogenous Retrovirus Expression Profiles in Samples from Brains of Patients with Schizophrenia and Bipolar Disorders. *J. Virol.* **2005**, *79*, 10890–10901. [[CrossRef](#)] [[PubMed](#)]
32. Gröger, V.; Emmer, A.; Staeger, M.S.; Cynis, H. Endogenous Retroviruses in Nervous System Disorders. *Pharmaceuticals* **2021**, *14*, 70. [[CrossRef](#)] [[PubMed](#)]
33. Cardelli, M.; Giacconi, R.; Malavolta, M.; Provinciali, M. Endogenous Retroelements in Cellular Senescence and Related Pathogenic Processes: Promising Drug Targets in Age-Related Diseases. *Curr. Drug Targets* **2016**, *17*, 416–427. [[CrossRef](#)] [[PubMed](#)]
34. Beck, C.R.; Garcia-Perez, J.L.; Badge, R.M.; Moran, J.V. LINE-1 elements in structural variation and disease. *Annu. Rev. Genom. Hum. Genet.* **2011**, *12*, 187–215. [[CrossRef](#)] [[PubMed](#)]
35. Stoye, J.P. Endogenous retroviruses: Still active after all these years? *Curr. Biol.* **2001**, *11*, R914–R916. [[CrossRef](#)]
36. Levin, H.L.; Moran, J.V. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* **2011**, *12*, 615–627. [[CrossRef](#)]
37. Wilson, A.S.; Power, B.E.; Molloy, P.L. DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* **2007**, *1775*, 138–162. [[CrossRef](#)]
38. Gilbert, N.; Labuda, D. CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 2869–2874. [[CrossRef](#)]

39. Jacob-Hirsch, J.; Eyal, E.; Knisbacher, B.A.; Roth, J.; Cesarkas, K.; Dor, C.; Farage-Barhom, S.; Kunik, V.; Simon, A.J.; Gal, M.; et al. Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. *Cell Res.* **2018**, *28*, 187–203. [[CrossRef](#)]
40. Han, J.S.; Szak, S.T.; Boeke, J.D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **2004**, *429*, 268–274. [[CrossRef](#)]
41. Brouha, B.; Schustak, J.; Badge, R.M.; Lutz-Prigge, S.; Farley, A.H.; Moran, J.V.; Kazazian, H.H., Jr. Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5280–5285. [[CrossRef](#)] [[PubMed](#)]
42. Beck, C.R.; Collier, P.; Macfarlane, C.; Malig, M.; Kidd, J.M.; Eichler, E.E.; Badge, R.M.; Moran, J.V. LINE-1 retrotransposition activity in human genomes. *Cell* **2010**, *141*, 1159–1170. [[CrossRef](#)] [[PubMed](#)]
43. Finnegan, D.J. Retrotransposons. *Curr. Biol.* **2012**, *22*, R432–R437. [[CrossRef](#)] [[PubMed](#)]
44. Criscione, S.W.; Theodosakis, N.; Micevic, G.; Cornish, T.C.; Burns, K.H.; Neretti, N.; Rodić, N. Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genom.* **2016**, *17*, 463. [[CrossRef](#)]
45. Speek, M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* **2001**, *21*, 1973–1985. [[CrossRef](#)]
46. Faulkner, G.J.; Kimura, Y.; Daub, C.O.; Wani, S.; Plessy, C.; Irvine, K.M.; Schroder, K.; Cloonan, N.; Steptoe, A.L.; Lassmann, T.; et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **2009**, *41*, 563–571. [[CrossRef](#)]
47. Singer, T.; McConnell, M.J.; Marchetto, M.C.; Coufal, N.G.; Gage, F.H. LINE-1 retrotransposons: Mediators of somatic variation in neuronal genomes? *Trends Neurosci.* **2010**, *33*, 345–354. [[CrossRef](#)]
48. Bodak, M.; Yu, J.; Ciaudo, C. Regulation of LINE-1 in mammals. *Biomol. Concepts* **2014**, *5*, 409–428. [[CrossRef](#)]
49. Denli, A.M.; Narvaiza, I.; Kerman, B.E.; Pena, M.; Benner, C.; Marchetto, M.C.; Diedrich, J.K.; Aslanian, A.; Ma, J.; Moresco, J.J.; et al. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* **2015**, *163*, 583–593. [[CrossRef](#)]
50. Lavie, L.; Maldener, E.; Brouha, B.; Meese, E.U.; Mayer, J. The human L1 promoter: Variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.* **2004**, *14*, 2253–2260. [[CrossRef](#)]
51. Mita, P.; Wudzinska, A.; Sun, X.; Andrade, J.; Nayak, S.; Kahler, D.J.; Badri, S.; LaCava, J.; Ueberheide, B.; Yun, C.Y.; et al. LINE-1 protein localization and functional dynamics during the cell cycle. *Elife* **2018**, *7*, e30058. [[CrossRef](#)] [[PubMed](#)]
52. Freeman, B.T.; Sokolowski, M.; Roy-Engel, A.M.; Smither, M.E.; Belancio, V.P. Identification of charged amino acids required for nuclear localization of human L1 ORF1 protein. *Mob. DNA* **2019**, *10*, 20. [[CrossRef](#)] [[PubMed](#)]
53. Idica, A.; Sevrioukov, E.A.; Zisoulis, D.G.; Hamdorf, M.; Daugaard, I.; Kadandale, P.; Pedersen, I.M. MicroRNA miR-128 represses LINE-1 (L1) retrotransposition by down-regulating the nuclear import factor TNPO1. *J. Biol. Chem.* **2017**, *292*, 20494–20508. [[CrossRef](#)] [[PubMed](#)]
54. Fung, L.; Guzman, H.; Sevrioukov, E.; Idica, A.; Park, E.; Bochnakian, A.; Daugaard, I.; Jury, D.; Mortazavi, A.; Zisoulis, D.G.; et al. miR-128 Restriction of LINE-1 (L1) Retrotransposition Is Dependent on Targeting hnRNPA1 mRNA. *Int. J. Mol. Sci.* **2019**, *20*, 1955. [[CrossRef](#)] [[PubMed](#)]
55. Cost, G.J.; Feng, Q.; Jacquier, A.; Boeke, J.D. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* **2002**, *21*, 5899–5910. [[CrossRef](#)]
56. Saleh, A.; Macia, A.; Muotri, A.R. Transposable Elements, Inflammation, and Neurological Disease. *Front. Neurol.* **2019**, *10*, 894. [[CrossRef](#)]
57. Gasior, S.L.; Wakeman, T.P.; Xu, B.; Deininger, P.L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **2006**, *357*, 1383–1393. [[CrossRef](#)]
58. Coufal, N.G.; Garcia-Perez, J.L.; Peng, G.E.; Marchetto, M.C.N.; Muotri, A.R.; Mu, Y.; Carson, C.T.; Macia, A.; Moran, J.V.; Gage, F.H. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20382–20387. [[CrossRef](#)]
59. Suzuki, J.; Yamaguchi, K.; Kajikawa, M.; Ichiyangi, K.; Adachi, N.; Koyama, H.; Takeda, S.; Okada, N. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.* **2009**, *5*, e1000461. [[CrossRef](#)]
60. Gasior, S.L.; Roy-Engel, A.M.; Deininger, P.L. ERCC1/XPF limits L1 retrotransposition. *DNA Repair* **2008**, *7*, 983–989. [[CrossRef](#)]
61. Sen, S.K.; Huang, C.T.; Han, K.; Batzer, M.A. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res.* **2007**, *35*, 3741–3751. [[CrossRef](#)] [[PubMed](#)]
62. Szak, S.T.; Pickeral, O.K.; Makalowski, W.; Boguski, M.S.; Landsman, D.; Boeke, J.D. Molecular archeology of L1 insertions in the human genome. *Genome Biol.* **2002**, *3*, research0052.0051. [[CrossRef](#)] [[PubMed](#)]
63. Ostertag, E.M.; Kazazian, H.H., Jr. Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **2001**, *11*, 2059–2065. [[CrossRef](#)] [[PubMed](#)]
64. Morrish, T.A.; Gilbert, N.; Myers, J.S.; Vincent, B.J.; Stamato, T.D.; Taccioli, G.E.; Batzer, M.A.; Moran, J.V. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **2002**, *31*, 159–165. [[CrossRef](#)] [[PubMed](#)]
65. Esnault, C.; Maestre, J.; Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **2000**, *24*, 363–367. [[CrossRef](#)]
66. Sultana, T.; van Essen, D.; Siol, O.; Bailly-Bechet, M.; Philippe, C.; Zine El Aabidine, A.; Pioger, L.; Nigumann, P.; Saccani, S.; Andrau, J.-C.; et al. The Landscape of L1 Retrotransposons in the Human Genome is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Mol. Cell* **2019**, *74*, 555–570. [[CrossRef](#)]

67. Gilbert, N.; Lutz-Prigge, S.; Moran, J.V. Genomic Deletions Created upon LINE-1 Retrotransposition. *Cell* **2002**, *110*, 315–325. [[CrossRef](#)]
68. Gilbert, N.; Lutz, S.; Morrish, T.A.; Moran, J.V. Multiple Fates of L1 Retrotransposition Intermediates in Cultured Human Cells. *Mol. Cell. Biol.* **2005**, *25*, 7780. [[CrossRef](#)]
69. Graham, T.; Boissinot, S. The genomic distribution of L1 elements: The role of insertion bias and natural selection. *J. Biomed. Biotechnol.* **2006**, *2006*, 75327. [[CrossRef](#)]
70. Erwin, J.A.; Paquola, A.C.M.; Singer, T.; Gallina, I.; Novotny, M.; Quayle, C.; Bedrosian, T.A.; Alves, F.I.A.; Butcher, C.R.; Herdy, J.R.; et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* **2016**, *19*, 1583–1591. [[CrossRef](#)]
71. Boissinot, S.; Davis, J.; Entezam, A.; Petrov, D.; Furano, A.V. Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9590. [[CrossRef](#)] [[PubMed](#)]
72. Bodea, G.O.; McKelvey, E.G.Z.; Faulkner, G.J. Retrotransposon-induced mosaicism in the neural genome. *Open Biol.* **2018**, *8*. [[CrossRef](#)] [[PubMed](#)]
73. Ewing, A.D.; Ballinger, T.J.; Earl, D.; Harris, C.C.; Ding, L.; Wilson, R.K.; Haussler, D. Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. *Genome Biol.* **2013**, *14*, R22. [[CrossRef](#)] [[PubMed](#)]
74. Kaessmann, H.; Vinckenbosch, N.; Long, M. RNA-based gene duplication: Mechanistic and evolutionary insights. *Nat. Rev. Genet.* **2009**, *10*, 19–31. [[CrossRef](#)] [[PubMed](#)]
75. Pei, B.; Sisu, C.; Frankish, A.; Howald, C.; Habegger, L.; Mu, X.J.; Harte, R.; Balasubramanian, S.; Tanzer, A.; Diekhans, M.; et al. The GENCODE pseudogene resource. *Genome Biol.* **2012**, *13*, R51. [[CrossRef](#)] [[PubMed](#)]
76. Cheetham, S.W.; Faulkner, G.J.; Dinger, M.E. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat. Rev. Genet.* **2020**, *21*, 191–201. [[CrossRef](#)]
77. Ding, W.; Lin, L.; Chen, B.; Dai, J. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB Life* **2006**, *58*, 677–685. [[CrossRef](#)]
78. Wei, W.; Gilbert, N.; Ooi, S.L.; Lawler, J.F.; Ostertag, E.M.; Kazazian, H.H.; Boeke, J.D.; Moran, J.V. Human L1 retrotransposition: Cis preference versus trans complementation. *Mol. Cell. Biol.* **2001**, *21*, 1429–1439. [[CrossRef](#)]
79. Kopera, H.C.; Moldovan, J.B.; Morrish, T.A.; Garcia-Perez, J.L.; Moran, J.V. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20345. [[CrossRef](#)]
80. Mandal, P.K.; Ewing, A.D.; Hancks, D.C.; Kazazian, H.H., Jr. Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. *Hum. Mol. Genet.* **2013**, *22*, 3730–3748. [[CrossRef](#)]
81. Briggs, E.M.; McKerrow, W.; Mita, P.; Boeke, J.D.; Logan, S.K.; Fenyö, D. RIP-seq reveals LINE-1 ORF1p association with p-body enriched mRNAs. *Mob. DNA* **2021**, *12*, 5. [[CrossRef](#)] [[PubMed](#)]
82. Grechishnikova, D.; Poptsova, M. Conserved 3' UTR stem-loop structure in L1 and Alu transposons in human genome: Possible role in retrotransposition. *BMC Genom.* **2016**, *17*, 992. [[CrossRef](#)] [[PubMed](#)]
83. Garcia-Perez, J.L.; Doucet, A.J.; Bucheton, A.; Moran, J.V.; Gilbert, N. Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res.* **2007**, *17*, 602–611. [[CrossRef](#)] [[PubMed](#)]
84. Richardson, S.R.; Salvador-Palomeque, C.; Faulkner, G.J. Diversity through duplication: Whole-genome sequencing reveals novel gene retrocopies in the human population. *Bioessays* **2014**, *36*, 475–481. [[CrossRef](#)]
85. Yang, L.; Emerman, M.; Malik, H.S.; McLaughlin, R.N.J. Retrocopying expands the functional repertoire of APOBEC3 antiviral proteins in primates. *Elife* **2020**, *9*, e58436. [[CrossRef](#)]
86. Zhang, L.; Richards, A.; Barrasa, M.I.; Stephen, H.H.; Richard, A.Y.; Jaenisch, R. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2105968118. [[CrossRef](#)]
87. Ostertag, E.M.; Prak, E.T.; DeBerardinis, R.J.; Moran, J.V.; Kazazian, H.H., Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **2000**, *28*, 1418–1423. [[CrossRef](#)]
88. Grover, D.; Mukerji, M.; Bhatnagar, P.; Kannan, K.; Brahmachari, S.K. Alu repeat analysis in the complete human genome: Trends and variations with respect to genomic composition. *Bioinformatics* **2004**, *20*, 813–817. [[CrossRef](#)]
89. Houck, C.M.; Rinehart, F.P.; Schmid, C.W. A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* **1979**, *132*, 289–306. [[CrossRef](#)]
90. Batzer, M.A.; Deininger, P.L. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **2002**, *3*, 370–379. [[CrossRef](#)]
91. Deininger, P. Alu elements: Know the SINEs. *Genome Biol.* **2011**, *12*, 236. [[CrossRef](#)] [[PubMed](#)]
92. Ullu, E.; Tschudi, C. Alu sequences are processed 7SL RNA genes. *Nature* **1984**, *312*, 171–172. [[CrossRef](#)] [[PubMed](#)]
93. Häslser, J.; Samuelsson, T.; Strub, K. Useful 'junk': Alu RNAs in the human transcriptome. *Cell. Mol. Life Sci.* **2007**, *64*, 1793–1800. [[CrossRef](#)] [[PubMed](#)]
94. Lütcke, H. Signal Recognition Particle (SRP), a Ubiquitous Initiator of Protein Translocation. *Eur. J. Biochem.* **1995**, *228*, 531–550. [[CrossRef](#)]
95. Martignetti, J.A.; Brosius, J. BC200 RNA: A neural RNA polymerase III product encoded by a monomeric Alu element. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11563–11567. [[CrossRef](#)]
96. Li, T.H.; Schmid, C.W. Differential stress induction of individual Alu loci: Implications for transcription and retrotransposition. *Gene* **2001**, *276*, 135–141. [[CrossRef](#)]

97. Weiner, A.M. SINEs and LINEs: The art of biting the hand that feeds you. *Curr. Opin. Cell Biol.* **2002**, *14*, 343–350. [[CrossRef](#)]
98. Ohshima, K. RNA-Mediated Gene Duplication and Retroposons: Retrogenes, LINEs, SINEs, and Sequence Specificity. *Int. J. Evol. Biol.* **2013**, *2013*, 424726. [[CrossRef](#)]
99. Kriegs, J.O.; Churakov, G.; Jurka, J.; Brosius, J.; Schmitz, J. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.* **2007**, *23*, 158–161. [[CrossRef](#)]
100. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker. Available online: <https://github.com/rmhubble/RepeatMasker/blob/master/repeatmasker.help> (accessed on 15 April 2022).
101. Hanks, D.C.; Kazazian, H.H., Jr. Active human retrotransposons: Variation and disease. *Curr. Opin. Genet. Dev.* **2012**, *22*, 191–203. [[CrossRef](#)]
102. Ostertag, E.M.; Goodier, J.L.; Zhang, Y.; Kazazian, H.H., Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* **2003**, *73*, 1444–1451. [[CrossRef](#)] [[PubMed](#)]
103. Hanks, D.C.; Kazazian, H.H., Jr. SVA retrotransposons: Evolution and genetic instability. *Semin. Cancer Biol.* **2010**, *20*, 234–245. [[CrossRef](#)] [[PubMed](#)]
104. Wang, H.; Xing, J.; Grover, D.; Hedges, D.; Han, K.; Walker, J.A.; Batzer, M. SVA elements: A hominid-specific retroposon family. *J. Mol. Biol.* **2005**, *354*, 994–1007. [[CrossRef](#)] [[PubMed](#)]
105. Raiz, J.; Damert, A.; Chira, S.; Held, U.; Klawitter, S.; Hamdorf, M.; Löwer, J.; Strätling, W.H.; Löwer, R.; Schumann, G.G. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res.* **2012**, *40*, 1666–1683. [[CrossRef](#)] [[PubMed](#)]
106. Goodier, J.L.; Mandal, P.K.; Zhang, L.; Kazazian, H.H., Jr. Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum. Mol. Genet.* **2010**, *19*, 1712–1725. [[CrossRef](#)] [[PubMed](#)]
107. Damert, A.; Raiz, J.; Horn, A.V.; Löwer, J.; Wang, H.; Xing, J.; Batzer, M.A.; Löwer, R.; Schumann, G.G. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res.* **2009**, *19*, 1992–2008. [[CrossRef](#)]
108. Gianfrancesco, O.; Bubb, V.J.; Quinn, J.P. SVA retrotransposons as potential modulators of neuropeptide gene expression. *Neuropeptides* **2017**, *64*, 3–7. [[CrossRef](#)] [[PubMed](#)]
109. Savage, A.L.; Wilm, T.P.; Khursheed, K.; Shatunov, A.; Morrison, K.E.; Shaw, P.J.; Shaw, C.E.; Smith, B.; Breen, G.; Al-Chalabi, A.; et al. An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS ONE* **2014**, *9*, e90833. [[CrossRef](#)]
110. Miskey, C.; Izsvák, Z.; Kawakami, K.; Ivics, Z. DNA transposons in vertebrate functional genomics. *Cell. Mol. Life Sci.* **2005**, *62*, 629–641. [[CrossRef](#)]
111. Rosser, J.M.; An, W. L1 expression and regulation in humans and rodents. *Front. Biosci.* **2012**, *4*, 2203–2225. [[CrossRef](#)]
112. Li, W.; Jin, Y.; Prazak, L.; Hammell, M.; Dubnau, J. Transposable Elements in TDP-43-Mediated Neurodegenerative Disorders. *PLoS ONE* **2012**, *7*, e44099. [[CrossRef](#)] [[PubMed](#)]
113. Lapp, H.E.; Hunter, R.G. Early life exposures, neurodevelopmental disorders, and transposable elements. *Neurobiol. Stress* **2019**, *11*, 100174. [[CrossRef](#)] [[PubMed](#)]
114. Rodriguez-Terrones, D.; Torres-Padilla, M.E. Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet.* **2018**, *34*, 806–820. [[CrossRef](#)] [[PubMed](#)]
115. Percharde, M.; Lin, C.-J.; Yin, Y.; Guan, J.; Peixoto, G.A.; Bulut-Karslioglu, A.; Biechele, S.; Huang, B.; Shen, X.; Ramalho-Santos, M. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **2018**, *174*, 391–405.e319. [[CrossRef](#)]
116. Faulkner, G.J.; Billon, V. L1 retrotransposition in the soma: A field jumping ahead. *Mob. DNA* **2018**, *9*, 22. [[CrossRef](#)]
117. Simons, C.; Pheasant, M.; Makunin, I.V.; Mattick, J.S. Transposon-free regions in mammalian genomes. *Genome Res.* **2006**, *16*, 164–172. [[CrossRef](#)]
118. Simons, C.; Makunin, I.V.; Pheasant, M.; Mattick, J.S. Maintenance of transposon-free regions throughout vertebrate evolution. *BMC Genom.* **2007**, *8*, 470. [[CrossRef](#)]
119. Yang, A.S.; Estécio, M.R.H.; Doshi, K.; Kondo, Y.; Tajara, E.H.; Issa, J.-P.J. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res.* **2004**, *32*, e38. [[CrossRef](#)]
120. Reilly, M.T.; Faulkner, G.J.; Dubnau, J.; Ponomarev, I.; Gage, F.H. The role of transposable elements in health and diseases of the central nervous system. *J. Neurosci.* **2013**, *33*, 17577–17586. [[CrossRef](#)]
121. Haggerty, C.; Kretzmer, H.; Riemenschneider, C.; Kumar, A.S.; Mattei, A.L.; Bailly, N.; Gottfreund, J.; Giesselmann, P.; Weigert, R.; Brändl, B.; et al. Dnmt1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol.* **2021**, *28*, 594–603. [[CrossRef](#)]
122. Bourc'his, D.; Bestor, T.H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **2004**, *431*, 96–99. [[CrossRef](#)] [[PubMed](#)]
123. Schulz, W.A.; Steinhoff, C.; Florl, A.R. Methylation of endogenous human retroelements in health and disease. *Curr. Top. Microbiol. Immunol.* **2006**, *310*, 211–250. [[CrossRef](#)] [[PubMed](#)]
124. Byun, H.M.; Motta, V.; Panni, T.; Bertazzi, P.A.; Apostoli, P.; Hou, L.; Baccarelli, A.A. Evolutionary age of repetitive element subfamilies and sensitivity of DNA methylation to airborne pollutants. *Part. Fibre Toxicol.* **2013**, *10*, 28. [[CrossRef](#)] [[PubMed](#)]
125. Miousse, I.R.; Chalbot, M.-C.G.; Lumen, A.; Ferguson, A.; Kavouras, I.G.; Koturbash, I. Response of transposable elements to environmental stressors. *Mutat. Res. Rev. Mut. Res.* **2015**, *765*, 19–39. [[CrossRef](#)] [[PubMed](#)]

126. Bollati, V.; Schwartz, J.; Wright, R.; Litonjua, A.; Tarantini, L.; Suh, H.; Sparrow, D.; Vokonas, P.; Baccarelli, A. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech. Ageing Dev.* **2009**, *130*, 234–239. [[CrossRef](#)] [[PubMed](#)]
127. Jintaridith, P.; Mutirangura, A. Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol. Genom.* **2010**, *41*, 194–200. [[CrossRef](#)]
128. Xie, M.; Hong, C.; Zhang, B.; Lowdon, R.F.; Xing, X.; Li, D.; Zhou, X.; Lee, H.J.; Maire, C.L.; Ligon, K.L.; et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* **2013**, *45*, 836–841. [[CrossRef](#)]
129. Philippe, C.; Vargas-Landin, D.B.; Doucet, A.J.; van Essen, D.; Vera-Otarola, J.; Kuciak, M.; Corbin, A.; Nigumann, P.; Cristofari, G. Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife* **2016**, *5*, e13926. [[CrossRef](#)]
130. Muotri, A.R.; Marchetto, M.C.N.; Coufal, N.G.; Oefner, R.; Yeo, G.; Nakashima, K.; Gage, F.H. L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **2010**, *468*, 443–446. [[CrossRef](#)]
131. Slotkin, R.K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **2007**, *8*, 272–285. [[CrossRef](#)]
132. Almeida, M.V.; Vernaz, G.; Putman, A.L.K.; Miska, E.A. Taming transposable elements in vertebrates: From epigenetic silencing to domestication. *Trends Genet.* **2022**, *38*, 529–553. [[CrossRef](#)] [[PubMed](#)]
133. Délérís, A.; Berger, F.; Duhaucourt, S. Role of Polycomb in the control of transposable elements. *Trends Genet.* **2021**, *37*, 882–889. [[CrossRef](#)] [[PubMed](#)]
134. Xiang, Y.; Yan, K.; Zheng, Q.; Ke, H.; Cheng, J.; Xiong, W.; Shi, X.; Wei, L.; Zhao, M.; Yang, F.; et al. Histone Demethylase KDM4B Promotes DNA Damage by Activating Long Interspersed Nuclear Element-1. *Cancer Res.* **2019**, *79*, 86. [[CrossRef](#)] [[PubMed](#)]
135. Varshney, D.; Vavrova-Anderson, J.; Oler, A.J.; Cowling, V.H.; Cairns, B.R.; White, R.J. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat. Commun.* **2015**, *6*, 6569. [[CrossRef](#)]
136. Bruno, M.; Mahgoub, M.; Macfarlan, T.S. The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu. Rev. Genet.* **2019**, *53*, 393–416. [[CrossRef](#)]
137. Wolf, G.; Greenberg, D.; Macfarlan, T.S. Spotting the enemy within: Targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. *Mob. DNA* **2015**, *6*, 17. [[CrossRef](#)]
138. Rowe, H.M.; Jakobsson, J.; Mesnard, D.; Rougemont, J.; Reynard, S.; Aktas, T.; Maillard, P.V.; Layard-Liesching, H.; Verp, S.; Marquis, J.; et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **2010**, *463*, 237–240. [[CrossRef](#)]
139. Jacobs, F.M.J.; Greenberg, D.; Nguyen, N.; Haussler, M.; Ewing, A.D.; Katzman, S.; Paten, B.; Salama, S.R.; Haussler, D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **2014**, *516*, 242–245. [[CrossRef](#)]
140. Ghildiyal, M.; Zamore, P.D. Small silencing RNAs: An expanding universe. *Nat. Rev. Genet.* **2009**, *10*, 94–108. [[CrossRef](#)]
141. Meister, G. Argonaute proteins: Functional insights and emerging roles. *Nat. Rev. Genet.* **2013**, *14*, 447–459. [[CrossRef](#)]
142. Li, L.; Liu, Y. Diverse small non-coding RNAs in RNA interference pathways. *Methods Mol. Biol.* **2011**, *764*, 169–182. [[CrossRef](#)] [[PubMed](#)]
143. Sasidharan, R.; Gerstein, M. Protein fossils live on as RNA. *Nature* **2008**, *453*, 729–731. [[CrossRef](#)] [[PubMed](#)]
144. Ghildiyal, M.; Seitz, H.; Horwich, M.D.; Li, C.; Du, T.; Lee, S.; Xu, J.; Kittler, E.L.; Zapp, M.L.; Weng, Z.; et al. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **2008**, *320*, 1077–1081. [[CrossRef](#)] [[PubMed](#)]
145. Saito, K.; Siomi, M.C. Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Dev. Cell* **2010**, *19*, 687–697. [[CrossRef](#)]
146. Rajasethupathy, P.; Antonov, I.; Sheridan, R.; Frey, S.; Sander, C.; Tuschl, T.; Kandel, E.R. A Role for Neuronal piRNAs in the Epigenetic Control of Memory-Related Synaptic Plasticity. *Cell* **2012**, *149*, 693–707. [[CrossRef](#)]
147. Lee, E.J.; Banerjee, S.; Zhou, H.; Jammalamadaka, A.; Arcila, M.; Manjunath, B.S.; Kosik, K.S. Identification of piRNAs in the central nervous system. *RNA* **2011**, *17*, 1090–1099. [[CrossRef](#)]
148. Kim, K.W. PIWI Proteins and piRNAs in the Nervous System. *Mol. Cells* **2019**, *42*, 828–835. [[CrossRef](#)]
149. Watanabe, T.; Totoki, Y.; Toyoda, A.; Kaneda, M.; Kuramochi-Miyagawa, S.; Obata, Y.; Chiba, H.; Kohara, Y.; Kono, T.; Nakano, T.; et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **2008**, *453*, 539–543. [[CrossRef](#)]
150. Tam, O.H.; Aravin, A.A.; Stein, P.; Girard, A.; Murchison, E.P.; Cheloufi, S.; Hodges, E.; Anger, M.; Sachidanandam, R.; Schultz, R.M.; et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **2008**, *453*, 534–538. [[CrossRef](#)]
151. Aravin, A.A.; Hannon, G.J.; Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **2007**, *318*, 761–764. [[CrossRef](#)]
152. Huang, S.; Yoshitake, K.; Asakawa, S. A Review of Discovery Profiling of PIWI-Interacting RNAs and Their Diverse Functions in Metazoans. *Int. J. Mol. Sci.* **2021**, *22*, 11166. [[CrossRef](#)] [[PubMed](#)]
153. Sato, K.; Siomi, M.C. The piRNA pathway in *Drosophila* ovarian germ and somatic cells. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **2020**, *96*, 32–42. [[CrossRef](#)] [[PubMed](#)]

154. Kuramochi-Miyagawa, S.; Watanabe, T.; Gotoh, K.; Totoki, Y.; Toyoda, A.; Ikawa, M.; Asada, N.; Kojima, K.; Yamaguchi, Y.; Ijiri, T.W.; et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **2008**, *22*, 908–917. [[CrossRef](#)] [[PubMed](#)]
155. Czech, B.; Munafo, M.; Ciabrelli, F.; Eastwood, E.L.; Fabry, M.H.; Kneuss, E.; Hannon, G.J. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu. Rev. Genet.* **2018**, *52*, 131–157. [[CrossRef](#)] [[PubMed](#)]
156. Nandi, S.; Chandramohan, D.; Fioriti, L.; Melnick, A.M.; Hébert, J.M.; Mason, C.E.; Rajasethupathy, P.; Kandel, E.R. Roles for small noncoding RNAs in silencing of retrotransposons in the mammalian brain. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12697–12702. [[CrossRef](#)] [[PubMed](#)]
157. Kawamura, Y.; Saito, K.; Kin, T.; Ono, Y.; Asai, K.; Sunohara, T.; Okada, T.N.; Siomi, M.C.; Siomi, H. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **2008**, *453*, 793–797. [[CrossRef](#)] [[PubMed](#)]
158. Soares, Z.G.; Gonçalves, A.N.A.; de Oliveira, K.P.V.; Marques, J.T. Viral RNA recognition by the Drosophila small interfering RNA pathway. *Microb. Infect.* **2014**, *16*, 1013–1021. [[CrossRef](#)]
159. Okamura, K. Diversity of animal small RNA pathways and their biological utility. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 351–368. [[CrossRef](#)]
160. Aliyari, R.; Wu, Q.; Li, H.-W.; Wang, X.-H.; Li, F.; Green, L.D.; Han, C.S.; Li, W.-X.; Ding, S.-W. Mechanism of Induction and Suppression of Antiviral Immunity Directed by Virus-Derived Small RNAs in Drosophila. *Cell Host Microbe* **2008**, *4*, 387–397. [[CrossRef](#)]
161. Yang, N.; Kazazian, H.H., Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* **2006**, *13*, 763–771. [[CrossRef](#)]
162. Berrens, R.V.; Andrews, S.; Spensberger, D.; Santos, F.; Dean, W.; Gould, P.; Sharif, J.; Olova, N.; Chandra, T.; Koseki, H.; et al. An endosRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell* **2017**, *21*, 694–703.e697. [[CrossRef](#)] [[PubMed](#)]
163. Goodier, J.L. Restricting retrotransposons: A review. *Mob. DNA* **2016**, *7*, 16. [[CrossRef](#)] [[PubMed](#)]
164. Gao, J.; Choudhry, H.; Cao, W. Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like family genes activation and regulation during tumorigenesis. *Cancer Sci.* **2018**, *109*, 2375–2382. [[CrossRef](#)] [[PubMed](#)]
165. Feng, Y.; Goubran, M.H.; Follack, T.B.; Chelico, L. Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. *Sci. Rep.* **2017**, *7*, 10881. [[CrossRef](#)]
166. Kinomoto, M.; Kanno, T.; Shimura, M.; Ishizaka, Y.; Kojima, A.; Kurata, T.; Sata, T.; Tokunaga, K. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* **2007**, *35*, 2955–2964. [[CrossRef](#)]
167. Koito, A.; Ikeda, T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Front. Microbiol.* **2013**, *4*, 28. [[CrossRef](#)]
168. Stetson, D.B.; Ko, J.S.; Heidmann, T.; Medzhitov, R. Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell* **2008**, *134*, 587–598. [[CrossRef](#)]
169. Thomas, C.A.; Tejwani, L.; Trujillo, C.A.; Negraes, P.D.; Herai, R.H.; Mesci, P.; Macia, A.; Crow, Y.J.; Muotri, A.R. Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. *Cell Stem Cell* **2017**, *21*, 319–331.e318. [[CrossRef](#)]
170. Zhao, K.; Du, J.; Han, X.; Goodier, J.L.; Li, P.; Zhou, X.; Wei, W.; Evans, S.L.; Li, L.; Zhang, W.; et al. Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutières syndrome-related SAMHD1. *Cell Rep.* **2013**, *4*, 1108–1115. [[CrossRef](#)]
171. Aoto, S.; Katagiri, S.; Wang, Y.; Pagnamenta, A.T.; Sakamoto-Abutani, R.; Toyoda, M.; Umezawa, A.; Okamura, K. Frequent retrotransposition of endogenous genes in ERCC2-deficient cells derived from a patient with xeroderma pigmentosum. *Stem Cell Res. Ther.* **2019**, *10*, 273. [[CrossRef](#)]
172. Skariah, G.; Seimetz, J.; Norsworthy, M.; Lannom, M.C.; Kenny, P.J.; Elrakhawy, M.; Forsthoefel, C.; Drnevich, J.; Kalsotra, A.; Ceman, S. Mov10 suppresses retroelements and regulates neuronal development and function in the developing brain. *BMC Biol.* **2017**, *15*, 54. [[CrossRef](#)] [[PubMed](#)]
173. Boissinot, S.; Entezam, A.; Furano, A.V. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **2001**, *18*, 926–935. [[CrossRef](#)] [[PubMed](#)]
174. Kurnosov, A.A.; Ustyugova, S.V.; Nazarov, V.I.; Minervina, A.A.; Komkov, A.Y.; Shugay, M.; Pogorelyy, M.V.; Khodosevich, K.V.; Mamedov, I.Z.; Lebedev, Y.B. The Evidence for Increased L1 Activity in the Site of Human Adult Brain Neurogenesis. *PLoS ONE* **2015**, *10*, e0117854. [[CrossRef](#)] [[PubMed](#)]
175. Baillie, J.K.; Barnett, M.W.; Upton, K.R.; Gerhardt, D.J.; Richmond, T.A.; De Sapio, F.; Brennan, P.M.; Rizzu, P.; Smith, S.; Fell, M.; et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **2011**, *479*, 534–537. [[CrossRef](#)]
176. Shpyleva, S.; Melnyk, S.; Pavliv, O.; Pogribny, I.; James, S.J. Overexpression of LINE-1 Retrotransposons in Autism Brain. *Mol. Neurobiol.* **2018**, *55*, 1740–1749. [[CrossRef](#)]
177. Kimura, R.H.; Choudary, P.V.; Schmid, C.W. Silk worm Bm1 SINE RNA increases following cellular insults. *Nucleic Acids Res.* **1999**, *27*, 3380–3387. [[CrossRef](#)]
178. Capomaccio, S.; Verini-Supplizi, A.; Galla, G.; Vitulo, N.; Barcaccia, G.; Felicetti, M.; Silvestrelli, M.; Cappelli, K. Transcription of LINE-derived sequences in exercise-induced stress in horses. *Anim. Genet.* **2010**, *41*, 23–27. [[CrossRef](#)]
179. Li, T.; Spearow, J.; Rubin, C.M.; Schmid, C.W. Physiological stresses increase mouse short interspersed element (SINE) RNA expression in vivo. *Gene* **1999**, *239*, 367–372. [[CrossRef](#)]

180. Dubin, M.J.; Scheid, O.M.; Becker, C. Transposons: A blessing curse. *Curr. Opin. Plant Biol.* **2018**, *42*, 23–29. [[CrossRef](#)]
181. Lanciano, S.; Mirouze, M. Transposable elements: All mobile, all different, some stress responsive, some adaptive? *Curr. Opin. Genet. Dev.* **2018**, *49*, 106–114. [[CrossRef](#)]
182. Dai, J.; Xie, W.; Brady, T.L.; Gao, J.; Voytas, D.F. Phosphorylation Regulates Integration of the Yeast Ty5 Retrotransposon into Heterochromatin. *Mol. Cell* **2007**, *27*, 289–299. [[CrossRef](#)] [[PubMed](#)]
183. Giorgi, G.; Marcantonio, P.; Del Re, B. LINE-1 retrotransposition in human neuroblastoma cells is affected by oxidative stress. *Cell Tissue Res.* **2011**, *346*, 383–391. [[CrossRef](#)] [[PubMed](#)]
184. Kale, S.P.; Moore, L.; Deininger, P.L.; Roy-Engel, A.M. Heavy metals stimulate human LINE-1 retrotransposition. *Int. J. Environ. Res. Public Health* **2005**, *2*, 14–23. [[CrossRef](#)] [[PubMed](#)]
185. El-Sawy, M.; Kale, S.P.; Dugan, C.; Nguyen, T.Q.; Belancio, V.; Bruch, H.; Roy-Engel, A.M.; Deininger, P.L. Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. *J. Mol. Biol.* **2005**, *354*, 246–257. [[CrossRef](#)]
186. Farkash, E.A.; Kao, G.D.; Horman, S.R.; Prak, E.T.L. Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res.* **2006**, *34*, 1196–1204. [[CrossRef](#)]
187. Okudaira, N.; Ishizaka, Y.; Nishio, H. Retrotransposition of long interspersed element 1 induced by methamphetamine or cocaine. *J. Biol. Chem.* **2014**, *289*, 25476–25485. [[CrossRef](#)]
188. Liu, W.M.; Chu, W.M.; Choudary, P.V.; Schmid, C.W. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.* **1995**, *23*, 1758–1765. [[CrossRef](#)]
189. Rudin, C.M.; Thompson, C.B. Transcriptional activation of short interspersed elements by DNA-damaging agents. *Genes Chromosomes Cancer* **2001**, *30*, 64–71. [[CrossRef](#)]
190. Mariner, P.D.; Walters, R.D.; Espinoza, C.A.; Drullinger, L.F.; Wagner, S.D.; Kugel, J.F.; Goodrich, J.A. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* **2008**, *29*, 499–509. [[CrossRef](#)]
191. Espinoza, C.A.; Goodrich, J.A.; Kugel, J.F. Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA* **2007**, *13*, 583–596. [[CrossRef](#)]
192. Chu, W.M.; Ballard, R.; Carpick, B.W.; Williams, B.R.; Schmid, C.W. Potential Alu function: Regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell. Biol.* **1998**, *18*, 58–68. [[CrossRef](#)] [[PubMed](#)]
193. Horváth, V.; Merenciano, M.; González, J. Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. *Trends Genet.* **2017**, *33*, 832–841. [[CrossRef](#)] [[PubMed](#)]
194. Goodier, J.L.; Zhang, L.; Vetter, M.R.; Kazazian, H.H., Jr. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.* **2007**, *27*, 6469–6483. [[CrossRef](#)] [[PubMed](#)]
195. Hu, S.; Li, J.; Xu, F.; Mei, S.; Le Duff, Y.; Yin, L.; Pang, X.; Cen, S.; Jin, Q.; Liang, C.; et al. SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. *PLoS Genet.* **2015**, *11*, e1005367. [[CrossRef](#)] [[PubMed](#)]
196. Hunter, R.G.; Murakami, G.; Dewell, S.; Seligsohn, M.; Baker, M.E.R.; Datson, N.A.; McEwen, B.S.; Pfaff, D.W. Acute stress and hippocampal histone H3 lysine 9 trimethylation, a retrotransposon silencing response. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17657. [[CrossRef](#)] [[PubMed](#)]
197. Ali, A.; Han, K.; Liang, P. Role of Transposable Elements in Gene Regulation in the Human Genome. *Life* **2021**, *11*, 118. [[CrossRef](#)] [[PubMed](#)]
198. Naville, M.; Warren, I.A.; Haftek-Terreau, Z.; Chalopin, D.; Brunet, F.; Levin, P.; Galiana, D.; Volff, J.N. Not so bad after all: Retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **2016**, *22*, 312–323. [[CrossRef](#)]
199. Kaneko-Ishino, T. The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front. Microbiol.* **2012**, *3*, 262. [[CrossRef](#)]
200. Volff, J.N. Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **2006**, *28*, 913–922. [[CrossRef](#)]
201. Estécio, M.R.H.; Gallegos, J.; Dekmezian, M.; Lu, Y.; Liang, S.; Issa, J.-P.J. SINE Retrotransposons Cause Epigenetic Reprogramming of Adjacent Gene Promoters. *Mol. Cancer Res.* **2012**, *10*, 1332–1342. [[CrossRef](#)]
202. Jacques, P.-É.; Jeyakani, J.; Bourque, G. The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. *PLoS Genet.* **2013**, *9*, e1003504. [[CrossRef](#)] [[PubMed](#)]
203. Kellner, M.; Makołowski, W. Transposable elements significantly contributed to the core promoters in the human genome. *Sci. China Life Sci.* **2019**, *62*, 489–497. [[CrossRef](#)] [[PubMed](#)]
204. Borchert, G.M.; Lanier, W.; Davidson, B.L. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1097–1101. [[CrossRef](#)] [[PubMed](#)]
205. Mätlik, K.; Redik, K.; Speek, M. L1 antisense promoter drives tissue-specific transcription of human genes. *J. Biomed. Biotechnol.* **2006**, *2006*, 71753. [[CrossRef](#)] [[PubMed](#)]
206. Roller, M.; Stamper, E.; Villar, D.; Izuogu, O.; Martin, F.; Redmond, A.M.; Ramachandran, R.; Harewood, L.; Odom, D.T.; Flicek, P. LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biol.* **2021**, *22*, 62. [[CrossRef](#)]
207. van de Lagemaat, L.N.; Landry, J.R.; Mager, D.L.; Medstrand, P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* **2003**, *19*, 530–536. [[CrossRef](#)]

208. Schoorlemmer, J.; Pérez-Palacios, R.; Climent, M.; Guallar, D.; Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. *Front. Oncol.* **2014**, *4*, 14. [[CrossRef](#)]
209. Jjingo, D.; Conley, A.B.; Wang, J.; Mariño-Ramírez, L.; Lunyak, V.V.; Jordan, I.K. Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mob. DNA* **2014**, *5*, 14. [[CrossRef](#)]
210. Su, M.; Han, D.; Boyd-Kirkup, J.; Yu, X.; Han, J.J. Evolution of Alu elements toward enhancers. *Cell Rep.* **2014**, *7*, 376–385. [[CrossRef](#)]
211. Zhang, X.O.; Gingeras, T.R.; Weng, Z. Genome-wide analysis of polymerase III-transcribed Alu elements suggests cell-type-specific enhancer function. *Genome Res.* **2019**, *29*, 1402–1414. [[CrossRef](#)]
212. Chen, L.L.; Yang, L. ALU Alternative Regulation for Gene Expression. *Trends Cell Biol.* **2017**, *27*, 480–490. [[CrossRef](#)] [[PubMed](#)]
213. Wang, J.; Vicente-García, C.; Seruggia, D.; Moltó, E.; Fernandez-Miñán, A.; Neto, A.; Lee, E.; Gómez-Skarmeta, J.L.; Montoliu, L.; Lunyak, V.V.; et al. MIR retrotransposon sequences provide insulators to the human genome. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4428–E4437. [[CrossRef](#)] [[PubMed](#)]
214. 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](#)] [[PubMed](#)]
215. Román, A.C.; González-Rico, F.J.; Moltó, E.; Hernando, H.; Neto, A.; Vicente-García, C.; Ballestar, E.; Gómez-Skarmeta, J.L.; Vavrova-Anderson, J.; White, R.J.; et al. Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch. *Genome Res.* **2011**, *21*, 422–432. [[CrossRef](#)] [[PubMed](#)]
216. Lunyak, V.V.; Prefontaine, G.G.; Núñez, E.; Cramer, T.; Ju, B.G.; Ohgi, K.A.; Hutt, K.; Roy, R.; García-Díaz, A.; Zhu, X.; et al. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **2007**, *317*, 248–251. [[CrossRef](#)] [[PubMed](#)]
217. Lynch, V.J.; Leclerc, R.D.; May, G.; Wagner, G.P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **2011**, *43*, 1154–1159. [[CrossRef](#)]
218. 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](#)]
219. Lynch, V.J.; Nnamani, M.C.; Kapusta, A.; Brayer, K.; Plaza, S.L.; Mazur, E.C.; Emera, D.; Sheikh, S.Z.; Grützner, F.; Bauersachs, S.; et al. Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy. *Cell Rep.* **2015**, *10*, 551–561. [[CrossRef](#)]
220. Chuong, E.B. Retroviruses facilitate the rapid evolution of the mammalian placenta. *Bioessays* **2013**, *35*, 853–861. [[CrossRef](#)]
221. Chuong, E.B.; Rumi, M.A.K.; Soares, M.J.; Baker, J.C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **2013**, *45*, 325–329. [[CrossRef](#)]
222. 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](#)] [[PubMed](#)]
223. Palazzo, A.; Marsano, R.M. Transposable elements: A jump toward the future of expression vectors. *Crit. Rev. Biotechnol.* **2021**, *41*, 792–808. [[CrossRef](#)] [[PubMed](#)]
224. Sun, X.; Wang, X.; Tang, Z.; Grivainis, M.; Kahler, D.; Yun, C.; Mita, P.; Fenyö, D.; Boeke, J.D. Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5526–E5535. [[CrossRef](#)] [[PubMed](#)]
225. Muotri, A.R.; Chu, V.T.; Marchetto, M.C.; Deng, W.; Moran, J.V.; Gage, F.H. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **2005**, *435*, 903–910. [[CrossRef](#)]
226. Harris, C.R.; Dewan, A.; Zupnick, A.; Normart, R.; Gabriel, A.; Prives, C.; Levine, A.J.; Hoh, J. p53 responsive elements in human retrotransposons. *Oncogene* **2009**, *28*, 3857–3865. [[CrossRef](#)]
227. Kuwabara, T.; Hsieh, J.; Muotri, A.; Yeo, G.; Warashina, M.; Lie, D.C.; Moore, L.; Nakashima, K.; Asashima, M.; Gage, F.H. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat. Neurosci.* **2009**, *12*, 1097–1105. [[CrossRef](#)]
228. Erwin, J.A.; Marchetto, M.C.; Gage, F.H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **2014**, *15*, 497–506. [[CrossRef](#)]
229. Sanchez-Luque, F.J.; Kempen, M.H.C.; Gerdes, P.; Vargas-Landin, D.B.; Richardson, S.R.; Troskie, R.L.; Jesuadian, J.S.; Cheetham, S.W.; Carreira, P.E.; Salvador-Palomeque, C.; et al. LINE-1 Evasion of Epigenetic Repression in Humans. *Mol. Cell* **2019**, *75*, 590–604.e512. [[CrossRef](#)]
230. Athanikar, J.N.; Badge, R.M.; Moran, J.V. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **2004**, *32*, 3846–3855. [[CrossRef](#)]
231. Yang, N.; Zhang, L.; Zhang, Y.; Kazazian, H.H., Jr. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* **2003**, *31*, 4929–4940. [[CrossRef](#)]
232. Tchénio, T.; Casella, J.F.; Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **2000**, *28*, 411–415. [[CrossRef](#)] [[PubMed](#)]
233. Wylie, A.; Jones, A.E.; D’Brot, A.; Lu, W.-J.; Kurtz, P.; Moran, J.V.; Rakheja, D.; Chen, K.S.; Hammer, R.E.; Comerford, S.A.; et al. p53 genes function to restrain mobile elements. *Genes Dev.* **2016**, *30*, 64–77. [[CrossRef](#)] [[PubMed](#)]
234. Tiwari, B.; Jones, A.E.; Caillet, C.J.; Das, S.; Royer, S.K.; Abrams, J.M. p53 directly represses human LINE1 transposons. *Genes Dev.* **2020**, *34*, 1439–1451. [[CrossRef](#)] [[PubMed](#)]
235. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **2006**, *127*, 469–480. [[CrossRef](#)]

236. Polak, P.; Domany, E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genom.* **2006**, *7*, 133. [[CrossRef](#)] [[PubMed](#)]
237. Piedrafita, F.J.; Molander, R.B.; Vansant, G.; Orlova, E.A.; Pfahl, M.; Reynolds, W.F. An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. *J. Biol. Chem.* **1996**, *271*, 14412–14420. [[CrossRef](#)]
238. Li, Y.; Bolten, C.; Bhat, B.G.; Woodring-Dietz, J.; Li, S.; Prayaga, S.K.; Xia, C.; Lala, D.S. Induction of Human Liver X Receptor  $\alpha$  Gene Expression Via an Autoregulatory Loop Mechanism. *Mol. Endocrinol.* **2002**, *16*, 506–514. [[CrossRef](#)]
239. Norris, J.; Fan, D.; Aleman, C.; Marks, J.R.; Futreal, P.A.; Wiseman, R.W.; Iglehart, J.D.; Deininger, P.L.; McDonnell, D.P. Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* **1995**, *270*, 22777–22782. [[CrossRef](#)]
240. Vansant, G.; Reynolds, W.F. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8229–8233. [[CrossRef](#)]
241. Antonaki, A.; Demetriades, C.; Polyzos, A.; Banos, A.; Vatsellas, G.; Lavigne, M.D.; Apostolou, E.; Mantouvalou, E.; Papadopoulou, D.; Mosialos, G.; et al. Genomic analysis reveals a novel nuclear factor- $\kappa$ B (NF- $\kappa$ B)-binding site in Alu-repetitive elements. *J. Biol. Chem.* **2011**, *286*, 38768–38782. [[CrossRef](#)]
242. Ferrari, R.; de Llobet Cucional, L.I.; Di Vona, C.; Le Dilly, F.; Vidal, E.; Lioutas, A.; Oliete, J.Q.; Jochem, L.; Cutts, E.; Dieci, G.; et al. TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Mol. Cell* **2020**, *77*, 475–487.e411. [[CrossRef](#)] [[PubMed](#)]
243. Crepaldi, L.; Policarpi, C.; Coatti, A.; Sherlock, W.T.; Jongbloets, B.C.; Down, T.A.; Riccio, A. Binding of TFIIIC to sine elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet.* **2013**, *9*, e1003699. [[CrossRef](#)] [[PubMed](#)]
244. Barnada, S.M.; Isopi, A.; Tejada-Martinez, D.; Goubert, C.; Patoori, S.; Pagliaroli, L.; Tracewell, M.; Trizzino, M. Genomic features underlie the co-option of SVA transposons as cis-regulatory elements in human pluripotent stem cells. *bioRxiv* **2022**, 2022.2001.2010.475682. [[CrossRef](#)]
245. 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*, e1006883. [[CrossRef](#)] [[PubMed](#)]
246. Fuchs, N.V.; Kraft, M.; Tondera, C.; Hanschmann, K.-M.; Löwer, J.; Löwer, R. Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3. *J. Virol.* **2011**, *85*, 3436–3448. [[CrossRef](#)] [[PubMed](#)]
247. Manghera, M.; Douville, R.N. Endogenous retrovirus-K promoter: A landing strip for inflammatory transcription factors? *Retrovirology* **2013**, *10*, 16. [[CrossRef](#)] [[PubMed](#)]
248. Manghera, M.; Ferguson-Parry, J.; Lin, R.; Douville, R.N. NF- $\kappa$ B and IRF1 Induce Endogenous Retrovirus K Expression via Interferon-Stimulated Response Elements in Its 5' Long Terminal Repeat. *J. Virol.* **2016**, *90*, 9338–9349. [[CrossRef](#)] [[PubMed](#)]
249. Gotea, V.; Makołowski, W. Do transposable elements really contribute to proteomes? *Trends Genet.* **2006**, *22*, 260–267. [[CrossRef](#)]
250. Toll-Riera, M.; Bosch, N.; Bellora, N.; Castelo, R.; Armengol, L.; Estivill, X.; Mar Albà, M. Origin of Primate Orphan Genes: A Comparative Genomics Approach. *Mol. Biol. Evol.* **2009**, *26*, 603–612. [[CrossRef](#)]
251. Moran, J.V.; DeBerardinis, R.J.; Kazazian, H.H. Exon Shuffling by L1 Retrotransposition. *Science* **1999**, *283*, 1530. [[CrossRef](#)]
252. Pickeral, O.K.; Makołowski, W.; Boguski, M.S.; Boeke, J.D. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* **2000**, *10*, 411–415. [[CrossRef](#)] [[PubMed](#)]
253. Hanks, D.C.; Ewing, A.D.; Chen, J.E.; Tokunaga, K.; Kazazian, H.H., Jr. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res.* **2009**, *19*, 1983–1991. [[CrossRef](#)] [[PubMed](#)]
254. Lev-Maor, G.; Sorek, R.; Shomron, N.; Ast, G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* **2003**, *300*, 1288–1291. [[CrossRef](#)] [[PubMed](#)]
255. Sorek, R.; Ast, G.; Graur, D. Alu-containing exons are alternatively spliced. *Genome Res.* **2002**, *12*, 1060–1067. [[CrossRef](#)] [[PubMed](#)]
256. Sorek, R. The birth of new exons: Mechanisms and evolutionary consequences. *RNA* **2007**, *13*, 1603–1608. [[CrossRef](#)]
257. Zhang, X.H.F.; Chasin, L.A. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13427–13432. [[CrossRef](#)] [[PubMed](#)]
258. Gal-Mark, N.; Schwartz, S.; Ast, G. Alternative splicing of Alu exons—Two arms are better than one. *Nucleic Acids Res.* **2008**, *36*, 2012–2023. [[CrossRef](#)]
259. Sela, N.; Mersch, B.; Gal-Mark, N.; Lev-Maor, G.; Hotz-Wagenblatt, A.; Ast, G. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol.* **2007**, *8*, R127. [[CrossRef](#)]
260. Mandal, A.K.; Pandey, R.; Jha, V.; Mukerji, M. Transcriptome-wide expansion of non-coding regulatory switches: Evidence from co-occurrence of Alu exonization, antisense and editing. *Nucleic Acids Res.* **2013**, *41*, 2121–2137. [[CrossRef](#)]
261. Belancio, V.P.; Roy-Engel, A.M.; Deininger, P. The impact of multiple splice sites in human L1 elements. *Gene* **2008**, *411*, 38–45. [[CrossRef](#)]
262. Zemojtel, T.; Penzkofer, T.; Schultz, J.; Dandekar, T.; Badge, R.; Vingron, M. Exonization of active mouse L1s: A driver of transcriptome evolution? *BMC Genom.* **2007**, *8*, 392. [[CrossRef](#)] [[PubMed](#)]

263. Kaer, K.; Branovets, J.; Hallikma, A.; Nigumann, P.; Speek, M. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. *PLoS ONE* **2011**, *6*, e26099. [[CrossRef](#)] [[PubMed](#)]
264. Bejerano, G.; Lowe, C.B.; Ahituv, N.; King, B.; Siepel, A.; Salama, S.R.; Rubin, E.M.; Kent, W.J.; Haussler, D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **2006**, *441*, 87–90. [[CrossRef](#)] [[PubMed](#)]
265. Roy-Engel, A.M.; El-Sawy, M.; Farooq, L.; Odom, G.L.; Perepelitsa-Belancio, V.; Bruch, H.; Oyeniran, O.O.; Deininger, P.L. Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* **2005**, *110*, 365–371. [[CrossRef](#)]
266. Lee, J.Y.; Ji, Z.; Tian, B. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. *Nucleic Acids Res.* **2008**, *36*, 5581–5590. [[CrossRef](#)]
267. Chen, C.; Ara, T.; Gautheret, D. Using Alu elements as polyadenylation sites: A case of retroposon exaptation. *Mol. Biol. Evol.* **2009**, *26*, 327–334. [[CrossRef](#)]
268. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **2002**, *71*, 817–846. [[CrossRef](#)]
269. Chen, L.-L.; DeCerbo, J.N.; Carmichael, G.G. Alu element-mediated gene silencing. *EMBO J.* **2008**, *27*, 1694–1705. [[CrossRef](#)]
270. Chen, L.-L.; Carmichael, G.G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: Functional role of a nuclear noncoding RNA. *Mol. Cell* **2009**, *35*, 467–478. [[CrossRef](#)] [[PubMed](#)]
271. Bazak, L.; Haviv, A.; Barak, M.; Jacob-Hirsch, J.; Deng, P.; Zhang, R.; Isaacs, F.J.; Rechavi, G.; Li, J.B.; Eisenberg, E.; et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **2014**, *24*, 365–376. [[CrossRef](#)]
272. Daniel, C.; Silberberg, G.; Behm, M.; Öhman, M. Alu elements shape the primate transcriptome by cis-regulation of RNA editing. *Genome Biol.* **2014**, *15*, R28. [[CrossRef](#)] [[PubMed](#)]
273. Ramaswami, G.; Lin, W.; Piskol, R.; Tan, M.H.; Davis, C.; Li, J.B. Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods* **2012**, *9*, 579–581. [[CrossRef](#)] [[PubMed](#)]
274. Neeman, Y.; Levanon, E.Y.; Jantsch, M.F.; Eisenberg, E. RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA* **2006**, *12*, 1802–1809. [[CrossRef](#)]
275. Muotri, A.R.; Gage, F.H. Generation of neuronal variability and complexity. *Nature* **2006**, *441*, 1087–1093. [[CrossRef](#)] [[PubMed](#)]
276. Rueter, S.M.; Dawson, T.R.; Emeson, R.B. Regulation of alternative splicing by RNA editing. *Nature* **1999**, *399*, 75–80. [[CrossRef](#)]
277. Feng, Y.; Sansam, C.L.; Singh, M.; Emeson, R.B. Altered RNA editing in mice lacking ADAR2 autoregulation. *Mol. Cell. Biol.* **2006**, *26*, 480–488. [[CrossRef](#)]
278. Starke, S.; Jost, I.; Rossbach, O.; Schneider, T.; Schreiner, S.; Hung, L.-H.; Bindereif, A. Exon Circularization Requires Canonical Splice Signals. *Cell Rep.* **2015**, *10*, 103–111. [[CrossRef](#)]
279. Chen, L.-L.; Yang, L. Regulation of circRNA biogenesis. *RNA Biol.* **2015**, *12*, 381–388. [[CrossRef](#)]
280. Jeck, W.R.; Sorrentino, J.A.; Wang, K.; Slevin, M.K.; Burd, C.E.; Liu, J.; Marzluff, W.F.; Sharpless, N.E. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **2013**, *19*, 141–157. [[CrossRef](#)]
281. Clayton, E.A.; Rishishwar, L.; Huang, T.-C.; Gulati, S.; Ban, D.; McDonald, J.F.; Jordan, I.K. An atlas of transposable element-derived alternative splicing in cancer. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2020**, *375*, 20190342. [[CrossRef](#)]
282. Lehnert, S.; Van Loo, P.; Thilakarathne, P.J.; Marynen, P.; Verbeke, G.; Schuit, F.C. Evidence for co-evolution between human microRNAs and Alu-repeats. *PLoS ONE* **2009**, *4*, e4456. [[CrossRef](#)] [[PubMed](#)]
283. Bhattacharya, A.; Jha, V.; Singhal, K.; Fatima, M.; Singh, D.; Chaturvedi, G.; Dholakia, D.; Kutum, R.; Pandey, R.; Bakken, T.E.; et al. Multiple Alu Exonization in 3'UTR of a Primate-Specific Isoform of CYP20A1 Creates a Potential miRNA Sponge. *Genome Biol. Evol.* **2021**, *13*, evaa233. [[CrossRef](#)] [[PubMed](#)]
284. Oda, H.; Nakabeppu, Y.; Furuichi, M.; Sekiguchi, M. Regulation of expression of the human MTH1 gene encoding 8-oxo-dGTPase. Alternative splicing of transcription products. *J. Biol. Chem.* **1997**, *272*, 17843–17850. [[CrossRef](#)] [[PubMed](#)]
285. Henssen, A.G.; Henaff, E.; Jiang, E.; Eisenberg, A.R.; Carson, J.R.; Villasante, C.M.; Ray, M.; Still, E.; Burns, M.; Gandara, J.; et al. Genomic DNA transposition induced by human PGBD5. *Elife* **2015**, *4*, e10565. [[CrossRef](#)]
286. Kapitonov, V.V.; Jurka, J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **2005**, *3*, e181. [[CrossRef](#)]
287. Carmona, L.M.; Fugmann, S.D.; Schatz, D.G. Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. *Genes Dev.* **2016**, *30*, 909–917. [[CrossRef](#)]
288. Carmona, L.M.; Schatz, D.G. New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J.* **2017**, *284*, 1590–1605. [[CrossRef](#)]
289. Chun, J.J.; Schatz, D.G.; Oettinger, M.A.; Jaenisch, R.; Baltimore, D. The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. *Cell* **1991**, *64*, 189–200. [[CrossRef](#)]
290. Jessen, J.R.; Jessen, T.N.; Vogel, S.S.; Lin, S. Concurrent expression of recombination activating genes 1 and 2 in zebrafish olfactory sensory neurons. *Genesis* **2001**, *29*, 156–162. [[CrossRef](#)]
291. Álvarez-Lindo, N.; Baleriola, J.; de Los Ríos, V.; Suárez, T.; de la Rosa, E.J. RAG-2 deficiency results in fewer phosphorylated histone H2AX foci, but increased retinal ganglion cell death and altered axonal growth. *Sci. Rep.* **2019**, *9*, 18486. [[CrossRef](#)]
292. Mateo, L.; González, J. Pogo-like transposases have been repeatedly domesticated into CENP-B-related proteins. *Genome Biol. Evol.* **2014**, *6*, 2008–2016. [[CrossRef](#)] [[PubMed](#)]

293. Gao, B.; Wang, Y.; Diaby, M.; Zong, W.; Shen, D.; Wang, S.; Chen, C.; Wang, X.; Song, C. Evolution of pogo, a separate superfamily of IS630-Tc1-mariner transposons, revealing recurrent domestication events in vertebrates. *Mob. DNA* **2020**, *11*, 25. [[CrossRef](#)] [[PubMed](#)]
294. Dumont, M.; Fachinetti, D. DNA Sequences in Centromere Formation and Function. In *Centromeres and Kinetochores: Discovering the Molecular Mechanisms Underlying Chromosome Inheritance*; Black, B.E., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 305–336.
295. Bannert, N.; Kurth, R. Retroelements and the human genome: New perspectives on an old relation. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14572–14579. [[CrossRef](#)] [[PubMed](#)]
296. Dupressoir, A.; Lavialle, C.; Heidmann, T. From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. *Placenta* **2012**, *33*, 663–671. [[CrossRef](#)]
297. Grandi, N.; Tramontano, E. Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Front. Immunol.* **2018**, *9*, 2039. [[CrossRef](#)]
298. Nakaya, Y.; Koshi, K.; Nakagawa, S.; Hashizume, K.; Miyazawa, T. Fematrin-1 is involved in fetomaternal cell-to-cell fusion in Bovinae placenta and has contributed to diversity of ruminant placentation. *J. Virol.* **2013**, *87*, 10563–10572. [[CrossRef](#)]
299. Brandt, J.; Schrauth, S.; Veith, A.-M.; Froschauer, A.; Haneke, T.; Schultheis, C.; Gessler, M.; Leimeister, C.; Volff, J.-N. Transposable elements as a source of genetic innovation: Expression and evolution of a family of retrotransposon-derived neogenes in mammals. *Gene* **2005**, *345*, 101–111. [[CrossRef](#)]
300. Henke, C.; Strissel, P.L.; Schubert, M.-T.; Mitchell, M.; Stolt, C.C.; Faschingbauer, F.; Beckmann, M.W.; Strick, R. Selective expression of sense and antisense transcripts of the sushi-ichi-related retrotransposon-derived family during mouse placentogenesis. *Retrovirology* **2015**, *12*, 9. [[CrossRef](#)]
301. Rawn, S.M.; Cross, J.C. The Evolution, Regulation, and Function of Placenta-Specific Genes. *Annu. Rev. Cell. Dev. Biol.* **2008**, *24*, 159–181. [[CrossRef](#)]
302. Naruse, M.; Ono, R.; Irie, M.; Nakamura, K.; Furuse, T.; Hino, T.; Oda, K.; Kashimura, M.; Yamada, I.; Wakana, S.; et al. Sirh7/Ldoc1 knockout mice exhibit placental P4 overproduction and delayed parturition. *Development* **2014**, *141*, 4763–4771. [[CrossRef](#)] [[PubMed](#)]
303. Irie, M.; Yoshikawa, M.; Ono, R.; Iwafune, H.; Furuse, T.; Yamada, I.; Wakana, S.; Yamashita, Y.; Abe, T.; Ishino, F.; et al. Cognitive Function Related to the Sirh11/Zcchc16 Gene Acquired from an LTR Retrotransposon in Eutherians. *PLoS Genet.* **2015**, *11*, e1005521. [[CrossRef](#)] [[PubMed](#)]
304. Lee, S.-H.; Oshige, M.; Durant, S.T.; Rasila, K.K.; Williamson, E.A.; Ramsey, H.; Kwan, L.; Nickoloff, J.A.; Hromas, R. The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18075. [[CrossRef](#)] [[PubMed](#)]
305. Cordaux, R.; Udit, S.; Batzer, M.A.; Feschotte, C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8101–8106. [[CrossRef](#)] [[PubMed](#)]
306. Kelley, D.; Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **2012**, *13*, R107. [[CrossRef](#)]
307. Kapusta, A.; Kronenberg, Z.; Lynch, V.J.; Zhuo, X.; Ramsay, L.; Bourque, G.; Yandell, M.; Feschotte, C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **2013**, *9*, e1003470. [[CrossRef](#)]
308. Piriyaongsa, J.; Mariño-Ramírez, L.; Jordan, I.K. Origin and evolution of human microRNAs from transposable elements. *Genetics* **2007**, *176*, 1323–1337. [[CrossRef](#)]
309. Greenig, M. HERVs, immunity, and autoimmunity: Understanding the connection. *PeerJ* **2019**, *7*, e6711. [[CrossRef](#)]
310. Hurst, T.P.; Magiorkinis, G. Activation of the innate immune response by endogenous retroviruses. *J. Gen. Virol.* **2015**, *96*, 1207–1218. [[CrossRef](#)]
311. Perron, H.; Dougier-Reynaud, H.-L.; Lomparski, C.; Popa, I.; Firouzi, R.; Bertrand, J.-B.; Marusic, S.; Portoukalian, J.; Jouvin-Marche, E.; Villiers, C.L.; et al. Human endogenous retrovirus protein activates innate immunity and promotes experimental allergic encephalomyelitis in mice. *PLoS ONE* **2013**, *8*, e80128. [[CrossRef](#)]
312. Curcio, M.J.; Belfort, M. The beginning of the end: Links between ancient retroelements and modern telomerases. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9107–9108. [[CrossRef](#)]
313. Belfort, M.; Curcio, M.J.; Lue, N.F. Telomerase and retrotransposons: Reverse transcriptases that shaped genomes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20304. [[CrossRef](#)] [[PubMed](#)]
314. Casacuberta, E. Drosophila: Retrotransposons Making up Telomeres. *Viruses* **2017**, *9*, 192. [[CrossRef](#)] [[PubMed](#)]
315. Podlevsky, J.D.; Chen, J.J.L. It all comes together at the ends: Telomerase structure, function, and biogenesis. *Mutat. Res.* **2012**, *730*, 3–11. [[CrossRef](#)] [[PubMed](#)]
316. Michelini, F.; Jalihal, A.P.; Francia, S.; Meers, C.; Neeb, Z.T.; Rossiello, F.; Gioia, U.; Aguado, J.; Jones-Weinert, C.; Luke, B.; et al. From “Cellular” RNA to “Smart” RNA: Multiple Roles of RNA in Genome Stability and Beyond. *Chem. Rev.* **2018**, *118*, 4365–4403. [[CrossRef](#)] [[PubMed](#)]
317. Morrish, T.A.; Garcia-Perez, J.L.; Stamato, T.D.; Taccioli, G.E.; Sekiguchi, J.; Moran, J.V. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **2007**, *446*, 208–212. [[CrossRef](#)] [[PubMed](#)]

318. Zvereva, M.I.; Shcherbakova, D.M.; Dontsova, O.A. Telomerase: Structure, functions, and activity regulation. *Biochem. Biokhimiia* **2010**, *75*, 1563–1583. [[CrossRef](#)]
319. Maida, Y.; Yasukawa, M.; Furuuchi, M.; Lassmann, T.; Possemato, R.; Okamoto, N.; Kasim, V.; Hayashizaki, Y.; Hahn, W.C.; Masutomi, K. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* **2009**, *461*, 230–235. [[CrossRef](#)]
320. Mason, J.M.; Randall, T.A.; Frydrychova, R.C. Telomerase lost? *Chromosoma* **2016**, *125*, 65–73. [[CrossRef](#)]
321. Fujiwara, H.; Osanai, M.; Matsumoto, T.; Kojima, K.K. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. *Chromosome Res.* **2005**, *13*, 455–467. [[CrossRef](#)]
322. Gladyshev, E.A.; Arkhipova, I.R. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9352–9357. [[CrossRef](#)]
323. Saint-Leandre, B.; Nguyen, S.C.; Levine, M.T. Diversification and collapse of a telomere elongation mechanism. *Genome Res.* **2019**, *29*, 920–931. [[CrossRef](#)] [[PubMed](#)]
324. Meers, C.; Keskin, H.; Storici, F. DNA repair by RNA: Templated, or not templated, that is the question. *DNA Repair* **2016**, *44*, 17–21. [[CrossRef](#)] [[PubMed](#)]
325. Wei, L.; Nakajima, S.; Böhm, S.; Bernstein, K.A.; Shen, Z.; Tsang, M.; Levine, A.S.; Lan, L. DNA damage during the G0/G1 phase triggers RNA-templated, Cockayne syndrome B-dependent homologous recombination. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3495–E3504. [[CrossRef](#)] [[PubMed](#)]
326. Puget, N.; Miller, K.M.; Legube, G. Non-canonical DNA/RNA structures during Transcription-Coupled Double-Strand Break Repair: Roadblocks or Bona fide repair intermediates? *DNA Repair* **2019**, *81*, 102661. [[CrossRef](#)]
327. Su, Y.; Ghodke, P.P.; Egli, M.; Li, L.; Wang, Y.; Guengerich, F.P. Human DNA polymerase  $\eta$  has reverse transcriptase activity in cellular environments. *J. Biol. Chem.* **2019**, *294*, 6073–6081. [[CrossRef](#)]
328. Meers, C.; Keskin, H.; Banyai, G.; Mazina, O.; Yang, T.; Gombolay, A.L.; Mukherjee, K.; Kaparos, E.I.; Newnam, G.; Mazin, A.; et al. Genetic Characterization of Three Distinct Mechanisms Supporting RNA-Driven DNA Repair and Modification Reveals Major Role of DNA Polymerase  $\zeta$ . *Mol. Cell* **2020**, *79*, 1037–1050.e1035. [[CrossRef](#)]
329. Teng, S.C.; Kim, B.; Gabriel, A. Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* **1996**, *383*, 641–644. [[CrossRef](#)]
330. Keskin, H.; Shen, Y.; Huang, F.; Patel, M.; Yang, T.; Ashley, K.; Mazin, A.V.; Storici, F. Transcript-RNA-templated DNA recombination and repair. *Nature* **2014**, *515*, 436–439. [[CrossRef](#)]
331. Ono, R.; Ishii, M.; Fujihara, Y.; Kitazawa, M.; Usami, T.; Kaneko-Ishino, T.; Kanno, J.; Ikawa, M.; Ishino, F. Double strand break repair by capture of retrotransposon sequences and reverse-transcribed spliced mRNA sequences in mouse zygotes. *Sci. Rep.* **2015**, *5*, 12281. [[CrossRef](#)]
332. Sonoda, E.; Hohegger, H.; Saberi, A.; Taniguchi, Y.; Takeda, S. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair* **2006**, *5*, 1021–1029. [[CrossRef](#)]
333. Welty, S.; Teng, Y.; Liang, Z.; Zhao, W.; Sanders, L.H.; Greenamyre, J.T.; Rubio, M.E.; Thathiah, A.; Kodali, R.; Wetzel, R.; et al. RAD52 is required for RNA-templated recombination repair in post-mitotic neurons. *J. Biol. Chem.* **2018**, *293*, 1353–1362. [[CrossRef](#)]
334. Srikanta, D.; Sen, S.K.; Huang, C.T.; Conlin, E.M.; Rhodes, R.M.; Batzer, M.A. An alternative pathway for Alu retrotransposition suggests a role in DNA double-strand break repair. *Genomics* **2009**, *93*, 205–212. [[CrossRef](#)] [[PubMed](#)]
335. Bushman, D.M.; Chun, J. The genomically mosaic brain: Aneuploidy and more in neural diversity and disease. *Semin. Cell Dev. Biol.* **2013**, *24*, 357–369. [[CrossRef](#)] [[PubMed](#)]
336. Rehen, S.K.; Yung, Y.C.; McCreight, M.P.; Kaushal, D.; Yang, A.H.; Almeida, B.S.V.; Kingsbury, M.A.; Cabral, K.M.S.; McConnell, M.J.; Anliker, B.; et al. Constitutional aneuploidy in the normal human brain. *J. Neurosci.* **2005**, *25*, 2176–2180. [[CrossRef](#)]
337. Cai, X.; Evrony, G.D.; Lehmann, H.S.; Elhosary, P.C.; Mehta, B.K.; Poduri, A.; Walsh, C.A. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **2014**, *8*, 1280–1289. [[CrossRef](#)] [[PubMed](#)]
338. Poduri, A.; Evrony, G.D.; Cai, X.; Walsh, C.A. Somatic Mutation, Genomic Variation, and Neurological Disease. *Science* **2013**, *341*, 1237758. [[CrossRef](#)]
339. Kingsbury, M.A.; Friedman, B.; McConnell, M.J.; Rehen, S.K.; Yang, A.H.; Kaushal, D.; Chun, J. Aneuploid neurons are functionally active and integrated into brain circuitry. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6143–6147. [[CrossRef](#)]
340. Westra, J.W.; Rivera, R.R.; Bushman, D.M.; Yung, Y.C.; Peterson, S.E.; Barral, S.; Chun, J. Neuronal DNA content variation (DCV) with regional and individual differences in the human brain. *J. Comp. Neurol.* **2010**, *518*, 3981–4000. [[CrossRef](#)]
341. Lodato, M.A.; Woodworth, M.B.; Lee, S.; Evrony, G.D.; Mehta, B.K.; Karger, A.; Lee, S.; Chittenden, T.W.; D’Gama, A.M.; Cai, X.; et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **2015**, *350*, 94–98. [[CrossRef](#)]
342. Zhao, B.; Wu, Q.; Ye, A.Y.; Guo, J.; Zheng, X.; Yang, X.; Yan, L.; Liu, Q.-R.; Hyde, T.M.; Wei, L.; et al. Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. *PLoS Genet.* **2019**, *15*, e1008043. [[CrossRef](#)]
343. McConnell, M.J.; Lindberg, M.R.; Brennand, K.J.; Piper, J.C.; Voet, T.; Cowing-Zitron, C.; Shumilina, S.; Lasken, R.S.; Vermeesch, J.R.; Hall, I.M.; et al. Mosaic copy number variation in human neurons. *Science* **2013**, *342*, 632–637. [[CrossRef](#)] [[PubMed](#)]
344. Rehen, S.K.; McConnell, M.J.; Kaushal, D.; Kingsbury, M.A.; Yang, A.H.; Chun, J. Chromosomal variation in neurons of the developing and adult mammalian nervous system. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13361–13366. [[CrossRef](#)] [[PubMed](#)]

345. Kole, A.J.; Annis, R.P.; Deshmukh, M. Mature neurons: Equipped for survival. *Cell Death Dis.* **2013**, *4*, e689. [[CrossRef](#)] [[PubMed](#)]
346. Wei, P.-C.; Chang, A.N.; Kao, J.; Du, Z.; Meyers, R.M.; Alt, F.W.; Schwer, B. Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* **2016**, *164*, 644–655. [[CrossRef](#)]
347. Alt, F.W.; Schwer, B. DNA double-strand breaks as drivers of neural genomic change, function, and disease. *DNA Repair* **2018**, *71*, 158–163. [[CrossRef](#)]
348. Suberbielle, E.; Sanchez, P.E.; Kravitz, A.V.; Wang, X.; Ho, K.; Eilertson, K.; Devidze, N.; Kreitzer, A.C.; Mucke, L. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid- $\beta$ . *Nat. Neurosci.* **2013**, *16*, 613–621. [[CrossRef](#)]
349. Madabhushi, R.; Gao, F.; Pfenning, A.R.; Pan, L.; Yamakawa, S.; Seo, J.; Rueda, R.; Phan, T.X.; Yamakawa, H.; Pao, P.-C.; et al. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **2015**, *161*, 1592–1605. [[CrossRef](#)]
350. Terry, D.M.; Devine, S.E. Aberrantly High Levels of Somatic LINE-1 Expression and Retrotransposition in Human Neurological Disorders. *Front. Genet.* **2020**, *10*, 1244. [[CrossRef](#)]
351. Faulkner, G.J.; Garcia-Perez, J.L. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends Genet.* **2017**, *33*, 802–816. [[CrossRef](#)]
352. Coufal, N.G.; Garcia-Perez, J.L.; Peng, G.E.; Yeo, G.W.; Mu, Y.; Lovci, M.T.; Morell, M.; O’Shea, K.S.; Moran, J.V.; Gage, F.H. L1 retrotransposition in human neural progenitor cells. *Nature* **2009**, *460*, 1127–1131. [[CrossRef](#)]
353. Upton, K.R.; Gerhardt, D.J.; Jesuadian, J.S.; Richardson, S.R.; Sánchez-Luque, F.J.; Bodea, G.O.; Ewing, A.D.; Salvador-Palomeque, C.; van der Knaap, M.S.; Brennan, P.M.; et al. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **2015**, *161*, 228–239. [[CrossRef](#)] [[PubMed](#)]
354. Macia, Á.; Widmann, T.J.; Heras, S.R.; Ayllón, V.; Sanchez, L.P.; Benkaddour-Boumzaouad, M.; Muñoz-López, M.; Rubio, A.M.; Amador-Cubero, S.; Blanco-Jimenez, E.; et al. Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* **2017**, *27*, 335–348. [[CrossRef](#)] [[PubMed](#)]
355. Salvador-Palomeque, C.; Francisco, J.S.L.; Patrick, R.J.F.; Adam, D.E.; Ernst, J.W.; Sandra, R.R.; Geoffrey, J.F. Dynamic Methylation of an L1 Transduction Family during Reprogramming and Neurodifferentiation. *Mol. Cell. Biol.* **2019**, *39*, e00499–18. [[CrossRef](#)] [[PubMed](#)]
356. Scott, E.C.; Devine, S.E. The Role of Somatic L1 Retrotransposition in Human Cancers. *Viruses* **2017**, *9*, 131. [[CrossRef](#)]
357. Naumova, O.Y.; Lee, M.; Rychkov, S.Y.; Vlasova, N.V.; Grigorenko, E.L. Gene expression in the human brain: The current state of the study of specificity and spatiotemporal dynamics. *Child Dev.* **2013**, *84*, 76–88. [[CrossRef](#)]
358. Bae, B.I.; Jayaraman, D.; Walsh, C.A. Genetic changes shaping the human brain. *Dev. Cell* **2015**, *32*, 423–434. [[CrossRef](#)]
359. Negi, S.K.; Guda, C. Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Sci. Rep.* **2017**, *7*, 897. [[CrossRef](#)]
360. Zylka, M.J.; Simon, J.M.; Philpot, B.D. Gene length matters in neurons. *Neuron* **2015**, *86*, 353–355. [[CrossRef](#)]
361. McCoy, M.J.; Fire, A.Z. Intron and gene size expansion during nervous system evolution. *BMC Genom.* **2020**, *21*, 360. [[CrossRef](#)]
362. McCoy, M.J.; Paul, A.J.; Victor, M.B.; Richner, M.; Gabel, H.W.; Gong, H.; Yoo, A.S.; Ahn, T.H. LONGO: An R package for interactive gene length dependent analysis for neuronal identity. *Bioinformatics* **2018**, *34*, i422–i428. [[CrossRef](#)]
363. Wimmer, K.; Callens, T.; Wernstedt, A.; Messiaen, L. The NF1 gene contains hotspots for L1 endonuclease-dependent de novo insertion. *PLoS Genet.* **2011**, *7*, e1002371. [[CrossRef](#)] [[PubMed](#)]
364. Hsueh, Y.P. Neurofibromin signaling and synapses. *J. Biomed. Sci.* **2007**, *14*, 461–466. [[CrossRef](#)] [[PubMed](#)]
365. Sur, D.; Kustwar, R.K.; Budania, S.; Mahadevan, A.; Hancks, D.C.; Yadav, V.; Shankar, S.K.; Mandal, P.K. Detection of the LINE-1 retrotransposon RNA-binding protein ORF1p in different anatomical regions of the human brain. *Mob. DNA* **2017**, *8*, 17. [[CrossRef](#)] [[PubMed](#)]
366. Evrony, G.D.; Cai, X.; Lee, E.; Hills, L.B.; Elhosary, P.C.; Lehmann, H.S.; Parker, J.J.; Atabay, K.D.; Gilmore, E.C.; Poduri, A.; et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **2012**, *151*, 483–496. [[CrossRef](#)] [[PubMed](#)]
367. Perrat, P.N.; DasGupta, S.; Wang, J.; Theurkauf, W.; Weng, Z.; Rosbash, M.; Waddell, S. Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* **2013**, *340*, 91–95. [[CrossRef](#)]
368. Treiber, C.D.; Waddell, S. Resolving the prevalence of somatic transposition in Drosophila. *Elife* **2017**, *6*, e28297. [[CrossRef](#)]
369. Li, J.B.; Church, G.M. Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nat. Neurosci.* **2013**, *16*, 1518–1522. [[CrossRef](#)]
370. Maas, S.; Kawahara, Y.; Tamburro, K.M.; Nishikura, K. A-to-I RNA editing and human disease. *RNA Biol.* **2006**, *3*, 1–9. [[CrossRef](#)]
371. Mattick, J.S.; Mehler, M.F. RNA editing, DNA recoding and the evolution of human cognition. *Trends Neurosci.* **2008**, *31*, 227–233. [[CrossRef](#)]
372. Yablonovitch, A.L.; Deng, P.; Jacobson, D.; Li, J.B. The evolution and adaptation of A-to-I RNA editing. *PLoS Genet.* **2017**, *13*, e1007064. [[CrossRef](#)]
373. Mehler, M.F.; Mattick, J.S. Noncoding RNAs and RNA Editing in Brain Development, Functional Diversification, and Neurological Disease. *Physiol. Rev.* **2007**, *87*, 799–823. [[CrossRef](#)] [[PubMed](#)]
374. Wahlstedt, H.; Daniel, C.; Ensterö, M.; Ohman, M. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome Res.* **2009**, *19*, 978–986. [[CrossRef](#)] [[PubMed](#)]

375. Ekdahl, Y.; Farahani, H.S.; Behm, M.; Lagergren, J.; Öhman, M. A-to-I editing of microRNAs in the mammalian brain increases during development. *Genome Res.* **2012**, *22*, 1477–1487. [[CrossRef](#)] [[PubMed](#)]
376. Rosenthal, J.J.C.; Seeburg, P.H. A-to-I RNA editing: Effects on proteins key to neural excitability. *Neuron* **2012**, *74*, 432–439. [[CrossRef](#)] [[PubMed](#)]
377. Kwak, S.; Kawahara, Y. Deficient RNA editing of GluR2 and neuronal death in amyotrophic lateral sclerosis. *J. Mol. Med.* **2005**, *83*, 110–120. [[CrossRef](#)] [[PubMed](#)]
378. Jönsson, M.E.; Brattås, P.L.; Gustafsson, C.; Petri, R.; Yudovich, D.; Piracs, K.; Verschuere, S.; Madsen, S.; Hansson, J.; Larsson, J.; et al. Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* **2019**, *10*, 3182. [[CrossRef](#)]
379. Turelli, P.; Playfoot, C.; Grun, D.; Raclot, C.; Pontis, J.; Coudray, A.; Thorball, C.; Duc, J.; Eugenia, V.P.; Deplancke, B.; et al. Primate-restricted KRAB zinc finger proteins and target retrotransposons control gene expression in human neurons. *Sci. Adv.* **2020**, *6*, eaba3200. [[CrossRef](#)]
380. Wang, T.; Medynets, M.; Johnson, K.R.; Doucet-O’Hare, T.T.; DiSanza, B.; Li, W.; Xu, Y.; Bagnell, A.; Tyagi, R.; Sampson, K.; et al. Regulation of stem cell function and neuronal differentiation by HERV-K via mTOR pathway. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 17842–17853. [[CrossRef](#)]
381. Nair, V.P.; Liu, H.; Ciceri, G.; Jungverdorben, J.; Frishman, G.; Tchieu, J.; Cederquist, G.Y.; Rothenaigner, I.; Schorpp, K.; Klepper, L.; et al. Activation of HERV-K(HML-2) disrupts cortical patterning and neuronal differentiation by increasing NTRK3. *Cell Stem Cell* **2021**, *28*, 1566–1581.e1568. [[CrossRef](#)]
382. Vago, D.R.; Wallenstein, G.V.; Morris, L.S. Hippocampus. In *Encyclopedia of the Neurological Sciences*, 2nd ed.; Aminoff, M.J., Daroff, R.B., Eds.; Academic Press: Oxford, UK, 2014; pp. 566–570. [[CrossRef](#)]
383. McEwen, B.S.; Nasca, C.; Gray, J.D. Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex. *Neuropsychopharmacology* **2016**, *41*, 3–23. [[CrossRef](#)]
384. Hunter, R.G.; Gagnidze, K.; McEwen, B.S.; Pfaff, D.W. Stress and the dynamic genome: Steroids, epigenetics, and the transposome. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6828–6833. [[CrossRef](#)] [[PubMed](#)]
385. Bedrosian, T.A.; Quayle, C.; Novaresi, N.; Gage, F.H. Early life experience drives structural variation of neural genomes in mice. *Science* **2018**, *359*, 1395. [[CrossRef](#)] [[PubMed](#)]
386. Ivashkina, O.I.; Zots, M.A.; Bezriadnov, D.V.; Anokhin, K.V. Erratum to: Increased 5-Bromo-2’-Deoxyuridine Incorporation in Various Brain Structures Following Passive Avoidance Training in Mice. *Bull. Exp. Biol. Med.* **2012**, *154*, 171–173. [[CrossRef](#)] [[PubMed](#)]
387. Salganik, R.I.; Parvez, H.; Tomson, V.P.; Shumskaya, I.A. Probable role of reverse transcription in learning: Correlation between hippocampal RNA-dependent DNA synthesis and learning ability in rats. *Neurosci. Lett.* **1983**, *36*, 317–322. [[CrossRef](#)]
388. Kokaeva, F.F.; Den’ mukhametova, S.V.; Kanapin, A.A.; Godukhin, O.V.; Il’in, Y.; Ivanov, V.A. Antisense oligodeoxyribonucleotides for fragments of the reverse transcriptase gene of the LINE-1 element of rats disturb the formation of long-term memory. *Dokl. Biochem. Biophys.* **2002**, *383*, 93–95. [[CrossRef](#)]
389. Bachiller, S.; Del-Pozo-Martín, Y.; Carrión, Á.M. L1 retrotransposition alters the hippocampal genomic landscape enabling memory formation. *Brain Behav. Immun.* **2017**, *64*, 65–70. [[CrossRef](#)]
390. Ponomarev, I.; Rau, V.; Eger, E.I.; Harris, R.A.; Fanselow, M.S. Amygdala transcriptome and cellular mechanisms underlying stress-enhanced fear learning in a rat model of posttraumatic stress disorder. *Neuropsychopharmacology* **2010**, *35*, 1402–1411. [[CrossRef](#)]
391. Cappucci, U.; Torromino, G.; Casale, A.M.; Camon, J.; Capitano, F.; Berloco, M.; Mele, A.; Pimpinelli, S.; Rinaldi, A.; Piacentini, L. Stress-induced strain and brain region-specific activation of LINE-1 transposons in adult mice. *Stress* **2018**, *21*, 575–579. [[CrossRef](#)]
392. Muotri, A.R.; Zhao, C.; Marchetto, M.C.; Gage, F.H. Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* **2009**, *19*, 1002–1007. [[CrossRef](#)]
393. Navabpour, S.; Rogers, J.; McFadden, T.; Jarome, T.J. DNA Double-Strand Breaks Are a Critical Regulator of Fear Memory Reconsolidation. *Int. J. Mol. Sci.* **2020**, *21*, 8995. [[CrossRef](#)]
394. McConnell, M.J.; Moran, J.V.; Abyzov, A.; Akbarian, S.; Bae, T.; Cortes-Ciriano, I.; Erwin, J.A.; Fasching, L.; Flasch, D.A.; Freed, D.; et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **2017**, *356*. [[CrossRef](#)] [[PubMed](#)]
395. Leijja-Salazar, M.; Piette, C.; Proukakakis, C. Review: Somatic mutations in neurodegeneration. *Neuropathol. Appl. Neurobiol.* **2018**, *44*, 267–285. [[CrossRef](#)] [[PubMed](#)]
396. McKinnon, P.J. Genome integrity and disease prevention in the nervous system. *Genes Dev.* **2017**, *31*, 1180–1194. [[CrossRef](#)] [[PubMed](#)]
397. Ciuffardini, F.; Nicolai, S.; Caputo, M.; Canu, G.; Paccosi, E.; Costantino, M.; Frontini, M.; Balajee, A.S.; Proietti-De-Santis, L. The cockayne syndrome B protein is essential for neuronal differentiation and neurogenesis. *Cell Death Dis.* **2014**, *5*, e1268. [[CrossRef](#)] [[PubMed](#)]
398. Kitsera, N.; Gasteiger, K.; Lühnsdorf, B.; Allgayer, J.; Epe, B.; Carell, T.; Khobta, A. Cockayne syndrome: Varied requirement of transcription-coupled nucleotide excision repair for the removal of three structurally different adducts from transcribed DNA. *PLoS ONE* **2014**, *9*, e94405. [[CrossRef](#)]

399. Emera, D.; Yin, J.; Reilly, S.K.; Gockley, J.; Noonan, J.P. Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2617–E2626. [[CrossRef](#)]
400. Sasaki, T.; Nishihara, H.; Hirakawa, M.; Fujimura, K.; Tanaka, M.; Kokubo, N.; Kimura-Yoshida, C.; Matsuo, I.; Sumiyama, K.; Saitou, N.; et al. Possible involvement of SINEs in mammalian-specific brain formation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4220–4225. [[CrossRef](#)]
401. Nakanishi, A.; Kobayashi, N.; Suzuki-Hirano, A.; Nishihara, H.; Sasaki, T.; Hirakawa, M.; Sumiyama, K.; Shimogori, T.; Okada, N. A SINE-Derived Element Constitutes a Unique Modular Enhancer for Mammalian Diencephalic Fgf8. *PLoS ONE* **2012**, *7*, e43785. [[CrossRef](#)]
402. Tashiro, K.; Teissier, A.; Kobayashi, N.; Nakanishi, A.; Sasaki, T.; Yan, K.; Tarabykin, V.; Vigier, L.; Sumiyama, K.; Hirakawa, M.; et al. A Mammalian Conserved Element Derived from SINE Displays Enhancer Properties Recapitulating Satb2 Expression in Early-Born Callosal Projection Neurons. *PLoS ONE* **2011**, *6*, e28497. [[CrossRef](#)]
403. Notwell, J.H.; Chung, T.; Heavner, W.; Bejerano, G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat. Commun.* **2015**, *6*, 6644. [[CrossRef](#)]
404. Santangelo, A.M.; de Souza, F.S.J.; Franchini, L.F.; Bumashny, V.F.; Low, M.J.; Rubinstein, M. Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene. *PLoS Genet.* **2007**, *3*, e166. [[CrossRef](#)] [[PubMed](#)]
405. Franchini, L.F.; López-Leal, R.; Nasif, S.; Beati, P.; Gelman, D.M.; Low, M.J.; de Souza, F.J.; Rubinstein, M. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15270–15275. [[CrossRef](#)] [[PubMed](#)]
406. Policarpi, C.; Crepaldi, L.; Brookes, E.; Nitarska, J.; French, S.M.; Coatti, A.; Riccio, A. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. *Cell Rep.* **2017**, *21*, 2879–2894. [[CrossRef](#)] [[PubMed](#)]
407. Pastuzyn, E.D.; Day, C.E.; Kearns, R.B.; Kyrke-Smith, M.; Taibi, A.V.; McCormick, J.; Yoder, N.; Belnap, D.M.; Erlendsson, S.; Morado, D.R.; et al. The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* **2018**, *172*, 275–288.e218. [[CrossRef](#)] [[PubMed](#)]
408. Shepherd, J.D. Arc—An endogenous neuronal retrovirus? *Semin. Cell Dev. Biol.* **2018**, *77*, 73–78. [[CrossRef](#)]
409. Zhang, H.; Bramham, C.R. Arc/Arg3.1 function in long-term synaptic plasticity: Emerging mechanisms and unresolved issues. *Eur. J. Neurosci.* **2021**, *54*, 6696–6712. [[CrossRef](#)]
410. Ashley, J.; Cordy, B.; Lucia, D.; Fradkin, L.G.; Budnik, V.; Thomson, T. Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. *Cell* **2018**, *172*, 262–274.e211. [[CrossRef](#)]
411. Stessman, H.A.F.; Willemsen, M.H.; Fenckova, M.; Penn, O.; Hoischen, A.; Xiong, B.; Wang, T.; Hoekzema, K.; Vives, L.; Vogel, I.; et al. Disruption of POGZ is Associated with Intellectual Disability and Autism Spectrum Disorders. *Am. J. Hum. Genet.* **2016**, *98*, 541–552. [[CrossRef](#)]
412. White, J.; Beck, C.R.; Harel, T.; Posey, J.E.; Jhangiani, S.N.; Tang, S.; Farwell, K.D.; Powis, Z.; Mendelsohn, N.J.; Baker, J.A.; et al. POGZ truncating alleles cause syndromic intellectual disability. *Genome Med.* **2016**, *8*, 3. [[CrossRef](#)]
413. Liu, W.; Seto, J.; Donovan, G.; Toth, M. Jerky, a Protein Deficient in a Mouse Epilepsy Model, Is Associated with Translationally Inactive mRNA in Neurons. *J. Neurosci.* **2002**, *22*, 176. [[CrossRef](#)]
414. Edelstein, L.C.; Collins, T. The SCAN domain family of zinc finger transcription factors. *Gene* **2005**, *359*, 1–17. [[CrossRef](#)] [[PubMed](#)]
415. Koradi, Z.; Kenchappa, R.S.; Mirmics, K.; Carter, B.D. NRIF is a Regulator of Neuronal Cholesterol Biosynthesis Genes. *J. Mol. Neurosci.* **2009**, *38*, 152–158. [[CrossRef](#)] [[PubMed](#)]
416. Pavelitz, T.; Gray, L.T.; Padilla, S.L.; Bailey, A.D.; Weiner, A.M. PGBD5: A neural-specific intron-containing piggyBac transposase domesticated over 500 million years ago and conserved from cephalochordates to humans. *Mob. DNA* **2013**, *4*, 23. [[CrossRef](#)] [[PubMed](#)]
417. Kapusta, A.; Feschotte, C. Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. *Trends Genet.* **2014**, *30*, 439–452. [[CrossRef](#)] [[PubMed](#)]
418. Zimmer-Bensch, G. Emerging Roles of Long Non-Coding RNAs as Drivers of Brain Evolution. *Cells* **2019**, *8*, 1399. [[CrossRef](#)] [[PubMed](#)]
419. Mustafin, R.N. The Role of Transposable Elements in the Differentiation of Stem Cells. *Mol. Gen. Microbiol. Virol.* **2019**, *34*, 67–74. [[CrossRef](#)]
420. Aprea, J.; Prenninger, S.; Dori, M.; Ghosh, T.; Monasor, L.S.; Wessendorf, E.; Zocher, S.; Massalini, S.; Alexopoulou, D.; Lesche, M.; et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J.* **2013**, *32*, 3145–3160. [[CrossRef](#)] [[PubMed](#)]
421. Skryabin, B.V.; Kremerskothen, J.; Vassilacopoulou, D.; Disotell, T.R.; Kapitonov, V.V.; Jurka, J.; Brosius, J. The BC200 RNA gene and its neural expression are conserved in Anthropeidea (Primates). *J. Mol. Evol.* **1998**, *47*, 677–685. [[CrossRef](#)]
422. Kuryshev, V.Y.; Skryabin, B.V.; Kremerskothen, J.; Jurka, J.; Brosius, J. Birth of a gene: Locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropeidea lineage. *J. Mol. Biol.* **2001**, *309*, 1049–1066. [[CrossRef](#)]
423. Kim, J.; Martignetti, J.A.; Shen, M.R.; Brosius, J.; Deininger, P. Rodent BC1 RNA gene as a master gene for ID element amplification. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3607–3611. [[CrossRef](#)]

424. Martignetti, J.A.; Brosius, J. BC1 RNA: Transcriptional analysis of a neural cell-specific RNA polymerase III transcript. *Mol. Cell. Biol.* **1995**, *15*, 1642–1650. [[CrossRef](#)] [[PubMed](#)]
425. Chesnokova, E.; Kolosov, P. Local Protein Synthesis in Dendritic Terminals and Its Regulation in Normal Conditions and during Plastic Changes. *Neurosci. Behav. Physiol.* **2017**, *47*, 595–607. [[CrossRef](#)]
426. Lee, Y.; Lee, H.S.; Kim, M.; Shin, H. Brain Cytoplasmic RNAs in Neurons: From Biosynthesis to Function. *Biomolecules* **2020**, *10*, 313. [[CrossRef](#)] [[PubMed](#)]
427. Carrieri, C.; Cimatti, L.; Biagioli, M.; Beugnet, A.; Zucchelli, S.; Fedele, S.; Pesce, E.; Ferrer, I.; Collavin, L.; Santoro, C.; et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **2012**, *491*, 454–457. [[CrossRef](#)] [[PubMed](#)]
428. Schein, A.; Zucchelli, S.; Kauppinen, S.; Gustincich, S.; Carninci, P. Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. *Sci. Rep.* **2016**, *6*, 33605. [[CrossRef](#)]
429. Cartault, F.; Munier, P.; Benko, E.; Desguerre, I.; Hanein, S.; Boddaert, N.; Bandiera, S.; Vellayoudom, J.; Krejbich-Trotot, P.; Bintner, M.; et al. Mutation in a primate-conserved retrotransposon reveals a noncoding RNA as a mediator of infantile encephalopathy. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4980. [[CrossRef](#)]
430. Smalheiser, N.R. The search for endogenous siRNAs in the mammalian brain. *Exp. Neurol.* **2012**, *235*, 455–463. [[CrossRef](#)]
431. Hughes, J.F.; Coffin, J.M. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat. Genet.* **2001**, *29*, 487–489. [[CrossRef](#)]
432. Ricci, M.; Peona, V.; Guichard, E.; Taccioli, C.; Boattini, A. Transposable Elements Activity is Positively Related to Rate of Speciation in Mammals. *J. Mol. Evol.* **2018**, *86*, 303–310. [[CrossRef](#)]
433. Bailey, J.A.; Liu, G.; Eichler, E.E. An Alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **2003**, *73*, 823–834. [[CrossRef](#)]
434. Mills, R.E.; Bennett, E.A.; Iskow, R.C.; Luttig, C.T.; Tsui, C.; Pittard, W.S.; Devine, S.E. Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* **2006**, *78*, 671–679. [[CrossRef](#)]
435. Hormozdiari, F.; Miriam, K.K.; Prado-Martinez, J.; Chiatante, G.; Irene, H.H.; Jerilyn, A.W.; Nelson, B.; Alkan, C.; Peter, H.S.; Huddleston, J.; et al. Rates and patterns of great ape retrotransposition. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13457–13462. [[CrossRef](#)]
436. Ohshima, K.; Hattori, M.; Yada, T.; Gojobori, T.; Sakaki, Y.; Okada, N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol.* **2003**, *4*, R74. [[CrossRef](#)] [[PubMed](#)]
437. Sen, S.K.; Han, K.; Wang, J.; Lee, J.; Wang, H.; Callinan, P.A.; Dyer, M.; Cordaux, R.; Liang, P.; Batzer, M.A. Human Genomic Deletions Mediated by Recombination between Alu Elements. *Am. J. Hum. Genet.* **2006**, *79*, 41–53. [[CrossRef](#)] [[PubMed](#)]
438. Niu, A.-L.; Wang, Y.-Q.; Zhang, H.; Liao, C.-H.; Wang, J.-K.; Zhang, R.; Che, J.; Su, B. Rapid evolution and copy number variation of primate RHOF2, an X-linked homeobox gene involved in male reproduction and possibly brain function. *BMC Evol. Biol.* **2011**, *11*, 298. [[CrossRef](#)] [[PubMed](#)]
439. Guffanti, G.; Bartlett, A.; Klengel, T.; Klengel, C.; Hunter, R.; Glinsky, G.; Macciardi, F. Novel Bioinformatics Approach Identifies Transcriptional Profiles of Lineage-Specific Transposable Elements at Distinct Loci in the Human Dorsolateral Prefrontal Cortex. *Mol. Biol. Evol.* **2018**, *35*, 2435–2453. [[CrossRef](#)]
440. del Rosario, R.C.; Rayan, N.A.; Prabhakar, S. Noncoding origins of anthropoid traits and a new null model of transposon functionalization. *Genome Res.* **2014**, *24*, 1469–1484. [[CrossRef](#)]
441. Belaghal, H.; Dekker, J.; Gibcus, J.H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **2017**, *123*, 56–65. [[CrossRef](#)]
442. Luo, X.; Liu, Y.; Dang, D.; Hu, T.; Hou, Y.; Meng, X.; Zhang, F.; Li, T.; Wang, C.; Li, M.; et al. 3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis. *Cell* **2021**, *184*, 723–740.e721. [[CrossRef](#)]
443. Nataf, S.; Uriagereka, J.; Benitez-Burraco, A. The Promoter Regions of Intellectual Disability-Associated Genes Are Uniquely Enriched in LTR Sequences of the MER41 Primate-Specific Endogenous Retrovirus: An Evolutionary Connection Between Immunity and Cognition. *Front. Genet.* **2019**, *10*, 321. [[CrossRef](#)]
444. Zeng, L.; Pederson, S.M.; Cao, D.; Qu, Z.; Hu, Z.; Adelson, D.L.; Wei, C. Genome-Wide Analysis of the Association of Transposable Elements with Gene Regulation Suggests that Alu Elements Have the Largest Overall Regulatory Impact. *J. Comput. Biol.* **2018**, *25*, 551–562. [[CrossRef](#)] [[PubMed](#)]
445. Jacobs, S.; Lie, D.C.; DeCicco, K.L.; Shi, Y.; DeLuca, L.M.; Gage, F.H.; Evans, R.M. Retinoic acid is required early during adult neurogenesis in the dentate gyrus. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3902. [[CrossRef](#)] [[PubMed](#)]
446. Haushalter, C.; Asselin, L.; Fraulob, V.; Dollé, P.; Rhinn, M. Retinoic acid controls early neurogenesis in the developing mouse cerebral cortex. *Dev. Biol.* **2017**, *430*, 129–141. [[CrossRef](#)] [[PubMed](#)]
447. Laperriere, D.; Wang, T.T.; White, J.H.; Mader, S. Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. *BMC Genom.* **2007**, *8*, 23. [[CrossRef](#)] [[PubMed](#)]
448. Eisenberg, E.; Nemzer, S.; Kinar, Y.; Sorek, R.; Rechavi, G.; Levanon, E.Y. Is abundant A-to-I RNA editing primate-specific? *Trends Genet.* **2005**, *21*, 77–81. [[CrossRef](#)] [[PubMed](#)]

449. Paz-Yaacov, N.; Levanon, E.Y.; Nevo, E.; Kinar, Y.; Harmelin, A.; Jacob-Hirsch, J.; Amariglio, N.; Eisenberg, E.; Rechavi, G. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12174–12179. [[CrossRef](#)]
450. Ramaswami, G.; Li, J.B. RADAR: A rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* **2014**, *42*, D109–D113. [[CrossRef](#)]
451. Larsen, P.A.; Hunnicutt, K.E.; Larsen, R.J.; Yoder, A.D.; Saunders, A.M. Warning SINEs: Alu elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Res.* **2018**, *26*, 93–111. [[CrossRef](#)]
452. Porath, H.T.; Schaffer, A.A.; Kaniewska, P.; Alon, S.; Eisenberg, E.; Rosenthal, J.; Levanon, E.Y.; Levy, O. A-to-I RNA Editing in the Earliest-Diverging Eumetazoan Phyla. *Mol. Biol. Evol.* **2017**, *34*, 1890–1901. [[CrossRef](#)]
453. Behm, M.; Öhman, M. RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain. *Trends Genet.* **2016**, *32*, 165–175. [[CrossRef](#)]
454. Lev-Maor, G.; Sorek, R.; Levanon, E.Y.; Paz, N.; Eisenberg, E.; Ast, G. RNA-editing-mediated exon evolution. *Genome Biol.* **2007**, *8*, R29. [[CrossRef](#)] [[PubMed](#)]
455. Shen, S.; Lin, L.; Cai, J.J.; Jiang, P.; Kenkel, E.J.; Stroik, M.R.; Sato, S.; Davidson, B.L.; Xing, Y. Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2837. [[CrossRef](#)] [[PubMed](#)]
456. Bae, B.-I.; Tietjen, I.; Atabay, K.D.; Evrony, G.D.; Johnson, M.B.; Asare, E.; Wang, P.P.; Murayama, A.Y.; Im, K.; Lisgo, S.N.; et al. Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* **2014**, *343*, 764–768. [[CrossRef](#)] [[PubMed](#)]
457. Parrott, A.M.; Tsai, M.; Batchu, P.; Ryan, K.; Ozer, H.L.; Tian, B.; Mathews, M.B. The evolution and expression of the snaR family of small non-coding RNAs. *Nucleic Acids Res.* **2011**, *39*, 1485–1500. [[CrossRef](#)] [[PubMed](#)]
458. Chou, H.-H.; Hayakawa, T.; Diaz, S.; Krings, M.; Indriati, E.; Leakey, M.; Paabo, S.; Satta, Y.; Takahata, N.; Varki, A. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11736–11741. [[CrossRef](#)]
459. Hedlund, M.; Tangvoranuntakul, P.; Takematsu, H.; Long, J.M.; Housley, G.D.; Kozutsumi, Y.; Suzuki, A.; Wynshaw-Boris, A.; Ryan, A.F.; Gallo, R.L.; et al. N-glycolylneuraminic acid deficiency in mice: Implications for human biology and evolution. *Mol. Cell. Biol.* **2007**, *27*, 4340–4346. [[CrossRef](#)]
460. Naito-Matsui, Y.; Davies, L.R.L.; Takematsu, H.; Chou, H.-H.; Tangvoranuntakul, P.; Carlin, A.F.; Verhagen, A.; Heyser, C.J.; Yoo, S.-W.; Choudhury, B.; et al. Physiological Exploration of the Long Term Evolutionary Selection against Expression of N-Glycolylneuraminic Acid in the Brain. *J. Biol. Chem.* **2017**, *292*, 2557–2570. [[CrossRef](#)]
461. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
462. The Gene Ontology Consortium. The Gene Ontology resource: Enriching a Gold mine. *Nucleic Acids Res.* **2021**, *49*, D325–D334. [[CrossRef](#)]
463. Mi, H.; Muruganujan, A.; Ebert, D.; Huang, X.; Thomas, P.D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **2019**, *47*, D419–D426. [[CrossRef](#)]
464. Kazazian, H.H., Jr.; Wong, C.; Youssoufian, H.; Scott, A.F.; Phillips, D.G.; Antonarakis, S.E. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **1988**, *332*, 164–166. [[CrossRef](#)] [[PubMed](#)]
465. Hanks, D.C.; Kazazian, H.H. Roles for retrotransposon insertions in human disease. *Mob. DNA* **2016**, *7*, 9. [[CrossRef](#)] [[PubMed](#)]
466. Guffanti, G.; Gaudi, S.; Fallon, J.H.; Sobell, J.; Potkin, S.G.; Pato, C.; Macciardi, F. Transposable elements and psychiatric disorders. *Am. J. Med. Genet. B* **2014**, *165*, 201–216. [[CrossRef](#)] [[PubMed](#)]
467. Nätt, D.; Thorsell, A. Stress-induced transposon reactivation: A mediator or an estimator of allostatic load? *Environ. Epigenet.* **2016**, *2*, dvw015. [[CrossRef](#)] [[PubMed](#)]
468. Ravel-Godreuil, C.; Znaidi, R.; Bonnifet, T.; Joshi, R.L.; Fuchs, J. Transposable elements as new players in neurodegenerative diseases. *FEBS Lett.* **2021**, *595*, 2733–2755. [[CrossRef](#)]
469. Peze-Heidsieck, E.; Bonnifet, T.; Znaidi, R.; Ravel-Godreuil, C.; Massiani-Beaudoin, O.; Joshi, R.L.; Fuchs, J. Retrotransposons as a Source of DNA Damage in Neurodegeneration. *Front. Aging Neurosci.* **2022**, *13*, 786897. [[CrossRef](#)]
470. Tan, H.; Qurashi, A.; Poidevin, M.; Nelson, D.L.; Li, H.; Jin, P. Retrotransposon activation contributes to fragile X premutation rCGG-mediated neurodegeneration. *Hum. Mol. Genet.* **2011**, *21*, 57–65. [[CrossRef](#)]
471. Tarallo, V.; Hirano, Y.; Gelfand, B.D.; Dridi, S.; Kerur, N.; Kim, Y.; Cho, W.G.; Kaneko, H.; Fowler, B.J.; Bogdanovich, S.; et al. DICER1 loss and Alu RNA induce age-related macular degeneration via the NLRP3 inflammasome and MyD88. *Cell* **2012**, *149*, 847–859. [[CrossRef](#)]
472. Bravo, P.; Darvish, H.; Tafakhori, A.; Azcona, L.J.; Johari, A.H.; Jamali, F.; Paisán-Ruiz, C. Molecular characterization of PRKN structural variations identified through whole-genome sequencing. *Mol. Genet. Genomic Med.* **2018**, *6*, 1243–1248. [[CrossRef](#)]
473. Guffanti, G.; Gaudi, S.; Klengel, T.; Fallon, J.H.; Mangalam, H.; Madduri, R.; Rodriguez, A.; DeCrescenzo, P.; Glovienka, E.; Sobell, J.; et al. LINE1 insertions as a genomic risk factor for schizophrenia: Preliminary evidence from an affected family. *Am. J. Med. Genet. B* **2016**, *171*, 534–545. [[CrossRef](#)]
474. Zhu, X.; Zhou, B.; Pattni, R.; Gleason, K.; Tan, C.; Kalinowski, A.; Sloan, S.; Fiston-Lavier, A.-S.; Mariani, J.; Petrov, D.; et al. Machine learning reveals bilateral distribution of somatic L1 insertions in human neurons and glia. *Nat. Neurosci.* **2021**, *24*, 186–196. [[CrossRef](#)] [[PubMed](#)]

475. Abrusán, G. Somatic transposition in the brain has the potential to influence the biosynthesis of metabolites involved in Parkinson's disease and schizophrenia. *Biol. Direct.* **2012**, *7*, 41. [[CrossRef](#)] [[PubMed](#)]
476. Yaribeygi, H.; Panahi, Y.; Sahraei, H.; Johnston, T.P.; Sahebkar, A. The impact of stress on body function: A review. *EXCLI J.* **2017**, *16*, 1057–1072. [[CrossRef](#)] [[PubMed](#)]
477. Cuarenta, A.; Kigar, S.L.; Henion, I.C.; Karls, K.E.; Chang, L.; Bakshi, V.P.; Auger, A.P. Early life stress increases Line1 within the developing brain in a sex-dependent manner. *Brain Res.* **2020**, *1748*, 147123. [[CrossRef](#)] [[PubMed](#)]
478. Cuarenta, A.; Kigar, S.L.; Henion, I.C.; Chang, L.; Bakshi, V.P.; Auger, A.P. Early life stress during the neonatal period alters social play and Line1 during the juvenile stage of development. *Sci. Rep.* **2021**, *11*, 3549. [[CrossRef](#)]
479. Nätt, D.; Johansson, I.; Faresjö, T.; Ludvigsson, J.; Thorsell, A. High cortisol in 5-year-old children causes loss of DNA methylation in SINE retrotransposons: A possible role for ZNF263 in stress-related diseases. *Clin. Epigenetics* **2015**, *7*, 91. [[CrossRef](#)]
480. Haigis, M.C.; Yankner, B.A. The aging stress response. *Mol. Cell* **2010**, *40*, 333–344. [[CrossRef](#)]
481. Franceschi, C.; Campisi, J. Chronic Inflammation (Inflammaging) and Its Potential Contribution to Age-Associated Diseases. *J. Gerontol. A Biol. Sci.* **2014**, *69*, S4–S9. [[CrossRef](#)]
482. Castelli, V.; Benedetti, E.; Antonosante, A.; Catanese, M.; Pitari, G.; Ippoliti, R.; Cimini, A.; d'Angelo, M. Neuronal Cells Rearrangement During Aging and Neurodegenerative Disease: Metabolism, Oxidative Stress and Organelles Dynamic. *Front. Mol. Neurosci.* **2019**, *12*, 132. [[CrossRef](#)]
483. De Cecco, M.; Criscione, S.W.; Peckham, E.J.; Hillenmeyer, S.; Hamm, E.A.; Manivannan, J.; Peterson, A.L.; Kreiling, J.A.; Neretti, N.; Sedivy, J.M. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **2013**, *12*, 247–256. [[CrossRef](#)]
484. De Cecco, M.; Criscione, S.W.; Peterson, A.L.; Neretti, N.; Sedivy, J.M.; Kreiling, J.A. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging* **2013**, *5*, 867–883. [[CrossRef](#)] [[PubMed](#)]
485. Grabowska, W.; Sikora, E.; Bielak-Zmijewska, A. Sirtuins, a promising target in slowing down the ageing process. *Biogerontology* **2017**, *18*, 447–476. [[CrossRef](#)] [[PubMed](#)]
486. Simon, M.; Van Meter, M.; Ablaeva, J.; Ke, Z.; Gonzalez, R.S.; Taguchi, T.; De Cecco, M.; Leonova, K.I.; Kogan, V.; Helfand, S.L.; et al. LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell Metab.* **2019**, *29*, 871–885.e875. [[CrossRef](#)] [[PubMed](#)]
487. Van Meter, M.; Kashyap, M.; Rezazadeh, S.; Geneva, A.J.; Morello, T.D.; Seluanov, A.; Gorbunova, V. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat. Commun.* **2014**, *5*, 5011. [[CrossRef](#)] [[PubMed](#)]
488. Ramirez, P.; Zuniga, G.; Sun, W.; Beckmann, A.; Ochoa, E.; DeVos, S.L.; Hyman, B.; Chiu, G.; Roy, E.R.; Cao, W.; et al. Pathogenic tau accelerates aging-associated activation of transposable elements in the mouse central nervous system. *Prog. Neurobiol.* **2022**, *208*, 102181. [[CrossRef](#)]
489. Li, W.; Prazak, L.; Chatterjee, N.; Grüniger, S.; Krug, L.; Theodorou, D.; Dubnau, J. Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.* **2013**, *16*, 529–531. [[CrossRef](#)]
490. Amor, S.; Peferoen, L.A.N.; Vogel, D.Y.S.; Breur, M.; van der Valk, P.; Baker, D.; van Noort, J.M. Inflammation in neurodegenerative diseases—An update. *Immunology* **2014**, *142*, 151–166. [[CrossRef](#)]
491. Kwon, H.S.; Koh, S.-H. Neuroinflammation in neurodegenerative disorders: The roles of microglia and astrocytes. *Transl. Neurodegener.* **2020**, *9*, 42. [[CrossRef](#)]
492. Jeppesen, R.; Benros, M.E. Autoimmune Diseases and Psychotic Disorders. *Front. Psychiatry* **2019**, *10*, 131. [[CrossRef](#)]
493. Morandi, E.; Tanasescu, R.; Tarlinton, R.E.; Constantinescu, C.S.; Zhang, W.; Tench, C.; Gran, B. The association between human endogenous retroviruses and multiple sclerosis: A systematic review and meta-analysis. *PLoS ONE* **2017**, *12*, e0172415. [[CrossRef](#)]
494. Perron, H.; Hamdani, N.; Faucard, R.; Lajnef, M.; Jamain, S.; Daban-Huard, C.; Sarrazin, S.; LeGuen, E.; Houenou, J.; Delavest, M.; et al. Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder. *Transl. Psychiatry* **2012**, *2*, e201. [[CrossRef](#)] [[PubMed](#)]
495. Licastro, F.; Porcellini, E. Activation of Endogenous Retrovirus, Brain Infections and Environmental Insults in Neurodegeneration and Alzheimer's Disease. *Int. J. Mol. Sci.* **2021**, *22*, 7263. [[CrossRef](#)] [[PubMed](#)]
496. De Cecco, M.; Ito, T.; Petrashen, A.P.; Elias, A.E.; Skvir, N.J.; Criscione, S.W.; Caligiana, A.; Broccoli, G.; Adney, E.M.; Boeke, J.D.; et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **2019**, *566*, 73–78. [[CrossRef](#)] [[PubMed](#)]
497. Calvo, S.E.; Mootha, V.K. The mitochondrial proteome and human disease. *Annu. Rev. Genom. Hum. Genet.* **2010**, *11*, 25–44. [[CrossRef](#)] [[PubMed](#)]
498. Larsen, P.A.; Lutz, M.W.; Hunnicutt, K.E.; Mihovilovic, M.; Saunders, A.M.; Yoder, A.D.; Roses, A.D. The Alu neurodegeneration hypothesis: A primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. *Alzheimers. Dement.* **2017**, *13*, 828–838. [[CrossRef](#)] [[PubMed](#)]
499. Baeken, M.W.; Moosmann, B.; Hajieva, P. Retrotransposon activation by distressed mitochondria in neurons. *Biochem. Biophys. Res. Commun.* **2020**, *525*, 570–575. [[CrossRef](#)] [[PubMed](#)]
500. Lane, C.A.; Hardy, J.; Schott, J.M. Alzheimer's disease. *Eur. J. Neurol.* **2018**, *25*, 59–70. [[CrossRef](#)] [[PubMed](#)]
501. Guo, C.; Jeong, H.H.; Hsieh, Y.C.; Klein, H.U.; Bennett, D.A.; De Jager, P.L.; Liu, Z.; Shulman, J.M. Tau Activates Transposable Elements in Alzheimer's Disease. *Cell Rep.* **2018**, *23*, 2874–2880. [[CrossRef](#)] [[PubMed](#)]

502. Sun, W.; Samimi, H.; Gamez, M.; Zare, H.; Frost, B. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nat. Neurosci.* **2018**, *21*, 1038–1048. [[CrossRef](#)]
503. Bollati, V.; Galimberti, D.; Pergoli, L.; Dalla Valle, E.; Barretta, F.; Cortini, F.; Scarpini, E.; Bertazzi, P.A.; Baccarelli, A. DNA methylation in repetitive elements and Alzheimer disease. *Brain Behav. Immun.* **2011**, *25*, 1078–1083. [[CrossRef](#)]
504. Lee, M.-H.; Siddoway, B.; Kaeser, G.E.; Segota, I.; Rivera, R.; Romanow, W.J.; Liu, C.S.; Park, C.; Kennedy, G.; Long, T.; et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **2018**, *563*, 639–645. [[CrossRef](#)] [[PubMed](#)]
505. Protasova, M.S.; Gusev, F.E.; Grigorenko, A.P.; Kuznetsova, I.L.; Rogaev, E.I.; Andreeva, T.V. Quantitative Analysis of L1-Retrotransposons in Alzheimer's Disease and Aging. *Biochem. Biokhimiia* **2017**, *82*, 962–971. [[CrossRef](#)] [[PubMed](#)]
506. Wirdefeldt, K.; Adami, H.-O.; Cole, P.; Trichopoulos, D.; Mandel, J. Epidemiology and etiology of Parkinson's disease: A review of the evidence. *Eur. J. Epidemiol.* **2011**, *26*, 1. [[CrossRef](#)] [[PubMed](#)]
507. de Thé, F.X.B.; Rekaik, H.; Peze-Heidsieck, E.; Massiani-Beaudoin, O.; Joshi, R.L.; Fuchs, J.; Prochiantz, A. Engrailed homeoprotein blocks degeneration in adult dopaminergic neurons through LINE-1 repression. *EMBO J.* **2018**, *37*, e97374. [[CrossRef](#)]
508. Pfaff, A.L.; Bubb, V.J.; Quinn, J.P.; Koks, S. An Increased Burden of Highly Active Retrotransposition Competent L1s is Associated with Parkinson's Disease Risk and Progression in the PPMI Cohort. *Int. J. Mol. Sci.* **2020**, *21*, 6562. [[CrossRef](#)]
509. Nielsen, S.S.; Checkoway, H.; Butler, R.A.; Nelson, H.H.; Farin, F.M.; Longstreth, W.T., Jr.; Franklin, G.M.; Swanson, P.D.; Kelsey, K.T. LINE-1 DNA methylation, smoking and risk of Parkinson's disease. *J. Parkinsons Dis.* **2012**, *2*, 303–308. [[CrossRef](#)] [[PubMed](#)]
510. Floreani, L.; Ansaloni, F.; Mangoni, D.; Agostoni, E.; Sanges, R.; Persichetti, F.; Gustincich, S. Analysis of LINE1 Retrotransposons in Huntington's Disease. *Front. Cell. Neurosci.* **2022**, *15*, 743797. [[CrossRef](#)]
511. Ruzo, A.; Croft, G.F.; Metzger, J.J.; Galgoczi, S.; Gerber, L.J.; Pellegrini, C.; Wang, H., Jr.; Fenner, M.; Tse, S.; Marks, A.; et al. Chromosomal instability during neurogenesis in Huntington's disease. *Development* **2018**, *145*, dev156844. [[CrossRef](#)]
512. Tan, H.; Wu, C.; Jin, L. A Possible Role for Long Interspersed Nuclear Elements-1 (LINE-1) in Huntington's Disease Progression. *Med. Sci. Monit.* **2018**, *24*, 3644–3652. [[CrossRef](#)]
513. Casale, A.M.; Liguori, F.; Ansaloni, F.; Cappucci, U.; Finaurini, S.; Spirito, G.; Persichetti, F.; Sanges, R.; Gustincich, S.; Piacentini, L. Transposable element activation promotes neurodegeneration in a Drosophila model of Huntington's disease. *iScience* **2022**, *25*, 103702. [[CrossRef](#)]
514. Chaytow, H.; Faller, K.M.E.; Huang, Y.T.; Gillingwater, T.H. Spinal muscular atrophy: From approved therapies to future therapeutic targets for personalized medicine. *Cell Rep. Med.* **2021**, *2*, 100346. [[CrossRef](#)] [[PubMed](#)]
515. Hamilton, G.; Gillingwater, T.H. Spinal muscular atrophy: Going beyond the motor neuron. *Trends Mol. Med.* **2013**, *19*, 40–50. [[CrossRef](#)] [[PubMed](#)]
516. Ottesen, E.W.; Seo, J.; Singh, N.N.; Singh, R.N. A Multilayered Control of the Human Survival Motor Neuron Gene Expression by Alu Elements. *Front. Microbiol.* **2017**, *8*, 2252. [[CrossRef](#)] [[PubMed](#)]
517. Alfahad, T.; Nath, A. Retroviruses and amyotrophic lateral sclerosis. *Antiviral. Res.* **2013**, *99*, 180–187. [[CrossRef](#)]
518. Pereira, G.C.; Sanchez, L.; Schaugency, P.M.; Rubio-Roldán, A.; Choi, J.A.; Planet, E.; Batra, R.; Turelli, P.; Trono, D.; Ostrow, L.W.; et al. Properties of LINE-1 proteins and repeat element expression in the context of amyotrophic lateral sclerosis. *Mob. DNA* **2018**, *9*, 35. [[CrossRef](#)]
519. Rabinovici, G.D.; Miller, B.L. Frontotemporal lobar degeneration: Epidemiology, pathophysiology, diagnosis and management. *CNS Drugs* **2010**, *24*, 375–398. [[CrossRef](#)]
520. Mohandas, E.; Rajmohan, V. Frontotemporal dementia: An updated overview. *Indian J. Psychiatry* **2009**, *51*, S65–S69.
521. Cohen, T.; Lee, V.; Trojanowski, J. TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. *Trends Mol. Med.* **2011**, *17*, 659–667. [[CrossRef](#)]
522. Shan, X.; Chiang, P.-M.; Price, D.L.; Wong, P.C. Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16325. [[CrossRef](#)]
523. Polymenidou, M.; Lagier-Tourenne, C.; Hutt, K.R.; Huelga, S.C.; Moran, J.; Liang, T.Y.; Ling, S.-C.; Sun, E.; Wancewicz, E.; Mazur, C.; et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* **2011**, *14*, 459–468. [[CrossRef](#)]
524. Sephton, C.F.; Cenik, C.; Kucukural, A.; Dammer, E.B.; Cenik, B.; Han, Y.; Dewey, C.M.; Roth, F.P.; Herz, J.; Peng, J.; et al. Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* **2011**, *286*, 1204–1215. [[CrossRef](#)] [[PubMed](#)]
525. Tollervey, J.R.; Curk, T.; Rogelj, B.; Briese, M.; Cereda, M.; Kayikci, M.; König, J.; Hortobágyi, T.; Nishimura, A.L.; Zupunski, V.; et al. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* **2011**, *14*, 452–458. [[CrossRef](#)] [[PubMed](#)]
526. Liu, E.Y.; Russ, J.; Cali, C.P.; Phan, J.M.; Amlie-Wolf, A.; Lee, E.B. Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. *Cell Rep.* **2019**, *27*, 1409–1421.e1406. [[CrossRef](#)] [[PubMed](#)]
527. Krug, L.; Chatterjee, N.; Borges-Monroy, R.; Hearn, S.; Liao, W.-W.; Morrill, K.; Prazak, L.; Rozhkov, N.; Theodorou, D.; Hammell, M.; et al. Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLoS Genet.* **2017**, *13*, e1006635. [[CrossRef](#)]
528. Tam, O.H.; Rozhkov, N.V.; Shaw, R.; Kim, D.; Hubbard, I.; Fennessey, S.; Propp, N.; Fagegaltier, D.; Harris, B.T.; Ostrow, L.W.; et al. Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell Rep.* **2019**, *29*, 1164–1177.e1165. [[CrossRef](#)]

529. Antony, J.M.; van Marle, G.; Opii, W.; Butterfield, D.A.; Mallet, F.; Yong, V.W.; Wallace, J.L.; Deacon, R.M.; Warren, K.; Power, C. Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination. *Nat. Neurosci.* **2004**, *7*, 1088–1095. [[CrossRef](#)]
530. Duperray, A.; Barbe, D.; Raguenez, G.; Weksler, B.B.; Romero, I.A.; Couraud, P.-O.; Perron, H.; Marche, P.N. Inflammatory response of endothelial cells to a human endogenous retrovirus associated with multiple sclerosis is mediated by TLR4. *Int. Immunol.* **2015**, *27*, 545–553. [[CrossRef](#)]
531. Mameli, G.; Poddighe, L.; Astone, V.; Delogu, G.; Arru, G.; Sotgiu, S.; Serra, C.; Dolei, A. Novel reliable real-time PCR for differential detection of MSRVenv and syncytin-1 in RNA and DNA from patients with multiple sclerosis. *J. Virol. Methods* **2009**, *161*, 98–106. [[CrossRef](#)]
532. Perron, H.; Bernard, C.; Bertrand, J.B.; Lang, A.B.; Popa, I.; Sanhadji, K.; Portoukalian, J. Endogenous retroviral genes, Herpesviruses and gender in Multiple Sclerosis. *J. Neurol. Sci.* **2009**, *286*, 65–72. [[CrossRef](#)]
533. Evans, T.A.; Erwin, J.A. Retroelement-derived RNA and its role in the brain. *Semin. Cell Dev. Biol.* **2021**, *114*, 68–80. [[CrossRef](#)]
534. Rice, G.I.; Kasher, P.R.; Forte, G.M.A.; Mannion, N.M.; Greenwood, S.M.; Szykiewicz, M.; Dickerson, J.E.; Bhaskar, S.S.; Zampini, M.; Briggs, T.A.; et al. Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.* **2012**, *44*, 1243–1248. [[CrossRef](#)] [[PubMed](#)]
535. Gimple, R.C.; Bhargava, S.; Dixit, D.; Rich, J.N. Glioblastoma stem cells: Lessons from the tumor hierarchy in a lethal cancer. *Genes Dev.* **2019**, *33*, 591–609. [[CrossRef](#)] [[PubMed](#)]
536. Carreira, P.E.; Ewing, A.D.; Li, G.; Schauer, S.N.; Upton, K.R.; Fagg, A.C.; Morell, S.; Kindlova, M.; Gerdes, P.; Richardson, S.R.; et al. Evidence for L1-associated DNA rearrangements and negligible L1 retrotransposition in glioblastoma multiforme. *Mob. DNA* **2016**, *7*, 21. [[CrossRef](#)] [[PubMed](#)]
537. Goodier, J.L. Retrotransposition in tumors and brains. *Mob. DNA* **2014**, *5*, 11. [[CrossRef](#)]
538. Balaj, L.; Lessard, R.; Dai, L.; Cho, Y.J.; Pomeroy, S.L.; Breakefield, X.O.; Skog, J. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* **2011**, *2*, 180. [[CrossRef](#)]
539. Petri, R.; Brattås, P.L.; Sharma, Y.; Jönsson, M.E.; Pircs, K.; Bengzon, J.; Jakobsson, J. LINE-2 transposable elements are a source of functional human microRNAs and target sites. *PLoS Genet.* **2019**, *15*, e1008036. [[CrossRef](#)]
540. Lai, M.-C.; Lombardo, M.V.; Baron-Cohen, S. Autism. *Lancet* **2014**, *383*, 896–910. [[CrossRef](#)]
541. Sharma, S.R.; Gonda, X.; Tarazi, F.I. Autism Spectrum Disorder: Classification, diagnosis and therapy. *Pharmacol. Ther.* **2018**, *190*, 91–104. [[CrossRef](#)]
542. Hu, V.W.; Steinberg, M.E. Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism Res.* **2009**, *2*, 67–77. [[CrossRef](#)]
543. Tangsuwansri, C.; Saeli, T.; Thongkorn, S.; Chonchaiya, W.; Suphapeetiporn, K.; Mutirangura, A.; Tencomnao, T.; Hu, V.W.; Sarachana, T. Investigation of epigenetic regulatory networks associated with autism spectrum disorder (ASD) by integrated global LINE-1 methylation and gene expression profiling analyses. *PLoS ONE* **2018**, *13*, e0201071. [[CrossRef](#)]
544. Saeli, T.; Tangsuwansri, C.; Thongkorn, S.; Chonchaiya, W.; Suphapeetiporn, K.; Mutirangura, A.; Tencomnao, T.; Hu, V.W.; Sarachana, T. Integrated genome-wide Alu methylation and transcriptome profiling analyses reveal novel epigenetic regulatory networks associated with autism spectrum disorder. *Mol. Autism* **2018**, *9*, 27. [[CrossRef](#)] [[PubMed](#)]
545. Balestrieri, E.; Arpino, C.; Matteucci, C.; Sorrentino, R.; Pica, F.; Alessandrelli, R.; Coniglio, A.; Curatolo, P.; Rezza, G.; Maciardi, F.; et al. HERVs expression in Autism Spectrum Disorders. *PLoS ONE* **2012**, *7*, e48831. [[CrossRef](#)] [[PubMed](#)]
546. Pohodich, A.E.; Zoghbi, H.Y. Rett syndrome: Disruption of epigenetic control of postnatal neurological functions. *Hum. Mol. Genet.* **2015**, *24*, R10–R16. [[CrossRef](#)] [[PubMed](#)]
547. Amir, R.E.; Van den Veyver, I.B.; Wan, M.; Tran, C.Q.; Francke, U.; Zoghbi, H.Y. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **1999**, *23*, 185–188. [[CrossRef](#)] [[PubMed](#)]
548. van Os, J.; Kapur, S. Schizophrenia. *Lancet* **2009**, *374*, 635–645. [[CrossRef](#)]
549. MacDonald, A.W., 3rd; Carter, C.S. Event-related fMRI study of context processing in dorsolateral prefrontal cortex of patients with schizophrenia. *J. Abnorm. Psychol.* **2003**, *112*, 689–697. [[CrossRef](#)]
550. Guillozet-Bongaarts, A.L.; Hyde, T.M.; Dalley, R.A.; Hawrylycz, M.J.; Henry, A.; Hof, P.R.; Hohmann, J.; Jones, A.R.; Kuan, C.L.; Royall, J.; et al. Altered gene expression in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol. Psychiatry* **2014**, *19*, 478–485. [[CrossRef](#)]
551. Van Winkel, R.; Esquivel, G.; Kenis, G.; Wichers, M.; Collip, D.; Peerbooms, O.; Rutten, B.; Myin-Germeys, I.; Van Os, J. REVIEW: Genome-Wide Findings in Schizophrenia and the Role of Gene–Environment Interplay. *CNS Neurosci. Ther.* **2010**, *16*, e185–e192. [[CrossRef](#)]
552. Doyle, G.A.; Crist, R.C.; Karatas, E.T.; Hammond, M.J.; Ewing, A.D.; Ferraro, T.N.; Hahn, C.-G.; Berrettini, W.H. Analysis of LINE-1 Elements in DNA from Postmortem Brains of Individuals with Schizophrenia. *Neuropsychopharmacology* **2017**, *42*, 2602–2611. [[CrossRef](#)]
553. Doherty, J.L.; O'Donovan, M.C.; Owen, M.J. Recent genomic advances in schizophrenia. *Clin. Genet.* **2012**, *81*, 103–109. [[CrossRef](#)]
554. Cardno, A.G.; Owen, M.J. Genetic Relationships Between Schizophrenia, Bipolar Disorder, and Schizoaffective Disorder. *Schizophr. Bull.* **2014**, *40*, 504–515. [[CrossRef](#)] [[PubMed](#)]
555. Slokar, G.; Hasler, G. Human Endogenous Retroviruses as Pathogenic Factors in the Development of Schizophrenia. *Front. Psychiatry* **2016**, *6*, 183. [[CrossRef](#)] [[PubMed](#)]

556. Bundo, M.; Toyoshima, M.; Okada, Y.; Akamatsu, W.; Ueda, J.; Nemoto-Miyauchi, T.; Sunaga, F.; Toritsuka, M.; Ikawa, D.; Kakita, A.; et al. Increased L1 Retrotransposition in the Neuronal Genome in Schizophrenia. *Neuron* **2014**, *81*, 306–313. [[CrossRef](#)] [[PubMed](#)]
557. Michaelson, J.J.; Shin, M.K.; Koh, J.Y.; Brueggeman, L.; Zhang, A.; Katzman, A.; McDaniel, L.; Fang, M.; Pufall, M.; Pieper, A.A. Neuronal PAS Domain Proteins 1 and 3 Are Master Regulators of Neuropsychiatric Risk Genes. *Biol. Psychiatry* **2017**, *82*, 213–223. [[CrossRef](#)] [[PubMed](#)]
558. Karlsson, H.; Schröder, J.; Bachmann, S.; Bottmer, C.; Yolken, R.H. HERV-W-related RNA detected in plasma from individuals with recent-onset schizophrenia or schizoaffective disorder. *Mol. Psychiatry* **2004**, *9*, 12–13. [[CrossRef](#)] [[PubMed](#)]
559. Liu, S.; Du, T.; Liu, Z.; Shen, Y.; Xiu, J.; Xu, Q. Inverse changes in L1 retrotransposons between blood and brain in major depressive disorder. *Sci. Rep.* **2016**, *6*, 37530. [[CrossRef](#)] [[PubMed](#)]
560. Rusiecki, J.A.; Chen, L.; Srikantan, V.; Zhang, L.; Yan, L.; Polin, M.L.; Baccarelli, A. DNA methylation in repetitive elements and post-traumatic stress disorder: A case-control study of US military service members. *Epigenomics* **2012**, *4*, 29–40. [[CrossRef](#)] [[PubMed](#)]
561. Mews, P.; Calipari, E.S. Cross-talk between the epigenome and neural circuits in drug addiction. *Prog. Brain Res.* **2017**, *235*, 19–63. [[CrossRef](#)]
562. Volkow, N.D.; Michaelides, M.; Baler, R. The Neuroscience of Drug Reward and Addiction. *Physiol. Rev.* **2019**, *99*, 2115–2140. [[CrossRef](#)]
563. Voskresenskiy, A.M.; Sun, L.S. The housekeeping gene (GA3PDH) and the long interspersed nuclear element (LINE) in the blood and organs of rats treated with cocaine. *Ann. N. Y. Acad. Sci.* **2008**, *1137*, 309–315. [[CrossRef](#)]
564. Maze, I.; Feng, J.; Wilkinson, M.B.; Sun, H.; Shen, L.; Nestler, E.J. Cocaine dynamically regulates heterochromatin and repetitive element unsilencing in nucleus accumbens. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3035–3040. [[CrossRef](#)] [[PubMed](#)]
565. Doyle, G.A.; Doucet-O'Hare, T.T.; Hammond, M.J.; Crist, R.C.; Ewing, A.D.; Ferraro, T.N.; Mash, D.C.; Kazazian Jr, H.H.; Berrettini, W.H. Reading LINE s within the cocaine addicted brain. *Brain Behav.* **2017**, *7*, e00678. [[CrossRef](#)] [[PubMed](#)]
566. Trivedi, M.; Shah, J.; Hodgson, N.; Byun, H.M.; Deth, R. Morphine induces redox-based changes in global DNA methylation and retrotransposon transcription by inhibition of excitatory amino acid transporter type 3-mediated cysteine uptake. *Mol. Pharmacol.* **2014**, *85*, 747–757. [[CrossRef](#)] [[PubMed](#)]
567. Sun, H.; Maze, I.; Dietz, D.M.; Scobie, K.N.; Kennedy, P.J.; Damez-Werno, D.; Neve, R.L.; Zachariou, V.; Shen, L.; Nestler, E.J. Morphine epigenomically regulates behavior through alterations in histone H3 lysine 9 dimethylation in the nucleus accumbens. *J. Neurosci.* **2012**, *32*, 17454–17464. [[CrossRef](#)]
568. Ponomarev, I.; Wang, S.; Zhang, L.; Harris, R.A.; Mayfield, R.D. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. *J. Neurosci.* **2012**, *32*, 1884–1897. [[CrossRef](#)]
569. Sitammagari, K.K.; Masood, W. Creutzfeldt Jakob Disease. In *StatPearls*; StatPearls Publishing LLC.: Treasure Island, FL, USA, 2022.
570. Manix, M.; Kalakoti, P.; Henry, M.; Thakur, J.; Menger, R.; Guthikonda, B.; Nanda, A. Creutzfeldt-Jakob disease: Updated diagnostic criteria, treatment algorithm, and the utility of brain biopsy. *Neurosurg. Focus* **2015**, *39*, E2. [[CrossRef](#)]
571. Jeong, B.-H.; Lee, Y.-J.; Carp, R.I.; Kim, Y.-S. The prevalence of human endogenous retroviruses in cerebrospinal fluids from patients with sporadic Creutzfeldt-Jakob disease. *J. Clin. Virol.* **2010**, *47*, 136–142. [[CrossRef](#)]
572. Gutmann, D.H.; Giovannini, M. Mouse models of neurofibromatosis 1 and 2. *Neoplasia* **2002**, *4*, 279–290. [[CrossRef](#)]
573. Vivarelli, R.; Grosso, S.; Calabrese, F.; Farnetani, M.; Di Bartolo, R.; Morgese, G.; Balestri, P. Epilepsy in neurofibromatosis 1. *J. Child Neurol.* **2003**, *18*, 338–342. [[CrossRef](#)]
574. Garg, S.; Lehtonen, A.; Huson, S.M.; Emsley, R.; Trump, D.; Evans, D.G.; Green, J. Autism and other psychiatric comorbidity in neurofibromatosis type 1: Evidence from a population-based study. *Dev. Med. Child Neurol.* **2013**, *55*, 139–145. [[CrossRef](#)]
575. Peltonen, S.; Kallionpää, R.A.; Peltonen, J. Neurofibromatosis type 1 (NF1) gene: Beyond café au lait spots and dermal neurofibromas. *Exp. Dermatol.* **2017**, *26*, 645–648. [[CrossRef](#)] [[PubMed](#)]
576. Kang, E.; Yoon, H.M.; Lee, B.H. Neurofibromatosis type I: Points to be considered by general pediatricians. *Clin. Exp. Pediatr.* **2021**, *64*, 149–156. [[CrossRef](#)] [[PubMed](#)]
577. Costa, R.M.; Silva, A.J. Molecular and cellular mechanisms underlying the cognitive deficits associated with neurofibromatosis 1. *J. Child Neurol.* **2002**, *17*, 622–626. [[CrossRef](#)] [[PubMed](#)]
578. Kaneko, H.; Dridi, S.; Tarallo, V.; Gelfand, B.D.; Fowler, B.J.; Cho, W.G.; Kleinman, M.E.; Ponicsan, S.L.; Hauswirth, W.W.; Chiodo, V.A.; et al. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* **2011**, *471*, 325–330. [[CrossRef](#)] [[PubMed](#)]
579. Aneichyk, T.; Hendriks, W.T.; Yadav, R.; Shin, D.; Gao, D.; Vaine, C.A.; Collins, R.L.; Domingo, A.; Currall, B.; Stortchevoi, A.; et al. Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell* **2018**, *172*, 897–909.e821. [[CrossRef](#)]