Endogenous retroviral promoter exaptation in human cancer

Artem Babaian¹,² and Dixie L. Mager¹,²*

Abstract
Cancer arises from a series of genetic and epigenetic changes, which result in abnormal expression or mutational activation of oncogenes, as well as suppression/inactivation of tumor suppressor genes. Aberrant expression of coding genes or long non-coding RNAs (lncRNAs) with oncogenic properties can be caused by translocations, gene amplifications, point mutations or other less characterized mechanisms. One such mechanism is the inappropriate usage of normally dormant, tissue-restricted or cryptic enhancers or promoters that serve to drive oncogenic gene expression. Dispersed across the human genome, endogenous retroviruses (ERVs) provide an enormous reservoir of autonomous gene regulatory modules, some of which have been co-opted by the host during evolution to play important roles in normal regulation of genes and gene networks. This review focuses on the “dark side” of such ERV regulatory capacity. Specifically, we discuss a growing number of examples of normally dormant or epigenetically repressed ERVs that have been harnessed to drive oncogenes in human cancer, a process we term onco-exaptation, and we propose potential mechanisms that may underlie this phenomenon.

Keywords: Gene regulation, Endogenous retrovirus, Long terminal repeat, Retrotransposon, Epigenetics, Cancer, Alternative promoter, Exaptation, Transcription

Background
Sequences derived from transposable elements (TEs) occupy at least half the human genome [1, 2]. TEs are generally classified into two categories; DNA transposons, which comprise 3.2% of the human genome; and the retroelements, short interspersed repeats (SINEs, 12.8% of the genome), long interspersed repeats (LINEs, 20.7%) and long terminal repeat (LTR) elements, derived from endogenous retroviruses (ERVs, 8.6%). Over evolutionary time, TE sequences in the genome can become functional units that confer a fitness advantage, a process called “exaptation” [3, 4]. Exaptation includes protein coding, non-coding and regulatory effects of TEs. This is in contrast to the designation of “nonaptations” for genetic units that perform some function (such as initiate transcription) but don’t impact host fitness [4]. Besides their roles in shaping genomes during evolution, TEs continue to have impact in humans through insertional mutagenesis, inducing rearrangements and affecting gene regulation, as discussed in recent reviews [5–12].

Efforts to explore the role of TEs in human cancer have focused primarily on LINEs and ERVs. While nearly all L1s, the major human LINE family, are defective, a few hundred retain the ability to retrotranspose [13] and these active elements occasionally cause germ line mutations [9, 14, 15]. Several recent studies have also documented somatic, cancer-specific L1 insertions [16–23], and a few such insertions were shown to contribute to malignancy [9]. For example, two L1 insertions were documented to disrupt the tumor suppressor gene APC in colon cancer [16, 23]. However, it is probable that most insertions are non-consequential “passenger mutations”, as recently discussed by Hancks and Kazazian [9]. Thus, the overall biological effect size of LINE retrotransposition on the process of oncogenesis may be limited.

No evidence for retrotranspositionally active ERVs in humans has been reported [24–26], so it is unlikely that human ERVs activate oncogenes or inactivate tumor suppressor genes by somatic retrotransposition. This is in contrast to the frequent oncogene activation by insertions of exogenous and endogenous retroviruses in
chickens or mice, where retrotranspositional activity of ERVs is very high [27–29]. Therefore, to date, most studies into potential roles for ERVs in human cancer have focused on their protein products. Indeed, there is strong evidence that the accessory proteins Np9 and Rec, encoded by members of the relatively young HERV-K (HML-2) group, have oncogenic properties, particularly in germ cell tumors [30–33].

Regardless of their retrotranspositional or coding capacity, ERVs may play a broader role in oncogenesis involving their intrinsic regulatory capacity. De-repression/activation of cryptic (or normally dormant) promoters to drive ectopic expression is one mechanism that can lead to oncogenic effects [34–40]. Because TEs, and especially ERV LTRs, are an abundant reservoir of natural promoters in the human genome [6, 41, 42], inappropriate transcriptional activation of typically repressed LTRs may contribute to oncogenesis. Here we review examples of such phenomena, which we term “onco-exaptation”, and propose two explanatory models to understand the role of LTRs in oncogenesis.

**Promoter potential of ERVs**

Hundreds of ERV “families” or groups, which is the more proper designation [43], are remnants of ancient retroviral infections of the germ line and occupy at least 8.67% of the human genome [1, 24, 44]. These range from groups that integrated before the divergence of rodents and primates, such as older members of the large MaLR/ERV-L class, to the youngest HERV-K (HML-2) group, a few members of which are insertionally polymorphic in humans [45, 46]. While it has been postulated that rare “active” HERV-K elements exist at very low allele frequencies [45], there is currently no evidence for new somatic or germ line insertions of ERVs in humans and nearly all have lost coding potential [24–26]. The situation is starkly different in inbred mice, where at least 10% of documented, phenotype-producing germ line mutations and numerous somatic, cancer-associated insertions are due to ongoing retrotranspositions of ERVs [28, 29, 47]. Table 1 lists select major ERV groups found in humans, members of which are mentioned in this review.

Approximately 90% of the “ERV-related” human genomic DNA is in the form of solitary LTRs, which are created over evolutionary time via recombination between the 5’ and 3’ LTRs of an integrated provirus [48, 49]. LTRs naturally contain transcriptional promoters and enhancers, and often splice donor sites, required for autonomous expression of the integrated LTR element. Furthermore, unlike for LINEs (see below), the integration process nearly always retains the primary transcriptional regulatory motifs, i.e. the LTR, even after recombination between the LTRs of a full-length proviral form. Mutations will degrade LTR promoter/enhancer motifs over time, but many of the >470,000 ERV/LTR loci in the genome [50] likely still retain some degree of their ancestral promoter/enhancer function, and hence a gene regulatory capacity.

LTR-mediated regulation of single genes and gene networks has been increasingly documented in the literature. For example, studies have implicated ERV LTRs in species-specific regulatory networks in ES cells [51], in the interferon response [52], in p53-mediated regulation [53], as tissue-specific enhancers [54, 55] and in regulating pluripotency by promoting genes and lncRNAs in stem cells [56–60]. LTR regulatory capacity arises from both their “ready-to-use” ancestral transcriptional factor (TF) binding sites and by mutation/evolution of novel sites, possibly maintained through epistatic capture [61] (recently reviewed in [42]). For more in depth discussion of the evolutionary exaptation of enhancers/promoters of LTRs and other TEs in mammals, we refer the reader to a rapidly growing number of reviews on this subject.

### Table 1 ERV/LTR groups mentioned in this review

| ERV Class | ERV Group | Associated LTRs (Repbase names) | ~Copies of internal regions | ~Copies of solitary LTRs |
|-----------|-----------|---------------------------------|----------------------------|-------------------------|
| I (ERV1)  | HERV-H    | LTR7, 7B, 7C, 7Y                 | 1060                       | 1270                    |
|           | HERV-9    | LTR12, 12B-12 F                 | 450                        | 6500                    |
|           | HERV-E    | LTR2, 2B, 2C                   | 250                        | 720                     |
|           | HUERS-P2  | LTR1, 1A-1 F                   | 120                        | 3000                    |
|           | LOR1      | LOR1a, 1b                     | 175                        | 1080                    |
|           | MER41     | MER41A-41G                    | 275                        | 4110                    |
| II (ERV/K)| HERV-K (HML-2) | LTRs, 5A, 5B, 5Hs | 80                     | 1200                    |
| III (ERV/L)| HERV16 | LTR16A-16E                    | 860                        | 18100                   |
| III (MaLR)| THE | THE1A-1D                      | 7900                       | 21260                   |

*Copy numbers of internal regions estimated from Dfam (dfam.org) [50]

*Solitary LTR numbers estimated from Dfam coverage minus 2x internal region numbers, assuming all internal regions are associated with two LTRs*
Promoter potential of LINEs and other non-LTR TEs
Besides via new retrotransposition events, existing L1 elements can also impact genes through promoter donation. Full-length L1 elements harbor two internal promoters at their 5' end, a sense promoter that drives expression of the element and an antisense promoter that has been shown to control expression of nearby genes through formation of chimeric transcripts [66–69]. Recently, this antisense promoter was also shown to promote expression of a small protein ORF0, which plays a regulatory role in retrotransposition [70]. While there are approximately 500,000 L1 loci in the human genome [1], the vast majority of them are 5' truncated due to incomplete reverse transcription during the retrotransposition process. Only ~3500-7000 are full length, retaining their promoters and hence, the potential ability to lend these promoters to nearby genes [71, 72]. Therefore, irrespective of differences in promoter strength, epigenetic regulation or mutational degradation, the vast copy number difference (~500,000 LTRs versus ~5000 promoter-containing L1s), is likely a major reason why the great majority of TE-initiated transcripts involve LTRs rather than L1s. In genome-wide screens of TE-initiated transcripts, small fragments of old L2 elements, which do not span the canonical L2 promoter, can be found as TSSs of lowly expressed transcripts [73] (unpublished data). Such instances likely represent “de novo” promoters, those arising naturally from genomic DNA which happens to be derived from a TE fragment, (possibly because L2 fragments have a GC rich base composition), rather than an “ancestral” or “ready-made” promoter, one which utilizes a TE’s original regulatory sequence.

Human SINE elements, namely ALUs and the older MIRs, can also promote transcription of nearby genes but these instances are relatively rare [68] given their extremely high copy numbers (~1.85 million fragments) [50]. This likely partly reflects the fact that SINEs, being derived from small functional RNAs, inherently possess PolIII promoters, rather than PolII, and their autonomous promoter strength is weak [74, 75]. Old MIR elements, as well as other ancient SINEs and DNA TEs, have been more prominent as enhancers, rather than genic promoters, as shown in several studies [76–81].

TEs and the cancer transcriptome
While some TE components have assumed cellular functions over evolutionary time, such as the syncytin genes in mammalian placenta, derived from independent ERV env genes in multiple mammals [6, 44, 82–84], the vast majority of TE/ERV insertions will be neutral or detrimental to the host. Given the potential for harm, multiple host mechanisms to repress these sequences have evolved. In mammals, ERV and L1 transcription is suppressed in normal cells by DNA methylation and/or histone modifications as well as many other host factors [9, 85–92]. The epigenetic regulation of TEs is relevant in cancer because epigenetic changes are common in malignancy and frequently associated with mutations in “epigenome-modifying” genes [93–97]. While the ultimate effects of many such mutations are not yet clear, their prominence indicates a central role for epigenomic dysregulation in oncogenesis [94, 98]. The most well established epigenetic changes are promoter hypermethylation and associated silencing of tumor suppressor genes [95, 99, 100] as well as genome-wide DNA hypomethylation [101–103]. Hypomethylation of ERVs and L1s in many tumors has been documented [104–106] and general transcriptional up-regulation of ERVs and L1s is often observed in cancers [33, 107–109]. However, other studies have shown no significant changes in ERV expression in selected human cancers compared to corresponding normal tissues [110, 111].

General conclusions about overall TE transcriptional deregulation in malignancy, or in any other biological state, are not always well founded and can depend on the type and sensitivity of the assay. For example, expression studies that use consensus probes for internal L1 or ERV regions to assay expression by custom microarrays or RT-PCR don’t resolve individual loci, so high expression signals could reflect dispersed transcriptional activation of many elements or the high expression of only one or a few loci. Such assays typically also cannot distinguish between expression due to TE promoter derepression or due to increased transcription of transcripts harboring TEs. RNA-Seq has the potential to give information on expression of individual TE loci, but interpretations of expression levels can be confounded by mapping difficulties, length of read and sequencing depth [112]. In any event, in most cases where transcriptional up-regulation of TE groups or individual TEs has been detected in cancer, the biological relevance of such aberrant expression is poorly understood.

Onco-exaptation of ERV/TE promoters
We propose that transcriptional up-regulation of LTR (and to a lesser extent L1) promoters is widespread in epigenetically perturbed cells such as cancer cells. Here we present specific published examples of onco-exaptation of TE-derived promoters affecting protein-coding genes (Table 2, Fig. 1). Although many other TE-initiated transcripts have been identified in cancer
cells (see below), in this section we restrict the discussion to those cases where some role of the TE-driven gene in cancer or cell growth has been demonstrated.

Ectopic and overexpression of protein-coding genes

The most straightforward interaction between a TE promoter and a gene is when a TE promoter is activated, initiates transcription, and transcribes a downstream gene without altering the open reading frame (ORF), thus serving as an alternative promoter. Since the TE promoter may be regulated differently than the native promoter, this can result in ectopic and/or overexpression of the gene, with oncogenic consequences.

The first case of such a phenomenon was discovered in the investigation of a potent oncogene colony stimulating factor one receptor (CSFIR) in Hodgkin Lymphoma (HL). Normally, CSFIR expression is restricted to macrophages in the myeloid lineage. To understand how this gene is expressed in HL, a B-cell derived cancer, Lamprecht et al. [113] performed 5’ RACE which revealed that the native, myeloid-restricted promoter is silent in HL cell lines, with CSFIR expression instead being driven by a solitary THE1B LTR, of the MalR-ERV1 class (Fig. 1a). THE1B LTRs are ancient, found in both Old and New World primates, and are highly abundant in the human genome, with a copy number of \(\sim 17,000\) [50, 114] (Table 1). The THE1B-CSFIR transcript produces a full-length protein in HL, which is required for growth/survival of HL cell lines [113] and is clinically prognostic for poorer patient survival [115]. Ectopic CSFIR expression in HL appears to be completely dependent on the THE1B LTR, and CSFIR protein or mRNA is detected in 39–48% of HL patient samples [115, 116].

To detect additional cases of onco-exaptation, we screened whole transcriptomes (RNA-Seq libraries) from a set of HL cell lines as well as from normal human B cells for TE-initiated transcripts, specifically transcripts that were recurrent in HL and not present in normal B cells [117]. We identified the Interferon Regulatory Factor 5 gene (IRF5) as a recurrently up-regulated gene being promoted by a LOR1a LTR located upstream of the native/canonical TSS (Fig. 1b), LOR1a LTRs are much less abundant compared to THE1 LTRs (Table 1) but are of similar age, with the IRF5 copy having inserted prior to New World-Old World primate divergence. IRF5 has multiple promoters/TSSs and complex transcription [118] and, contrary to the CSFIR case, the native promoters are not completely silent in HL. However, LTR activity correlates with strong overexpression of the IRF5 protein and transcript, above normal physiological levels [117]. While our study was ongoing, Kreher et al. reported that IRF5 is upregulated in HL and is a central regulator of the HL transcriptome [119]. Moreover, they found that IRF5 is crucial for HL cell survival. Intriguingly, we noted that insertion of the LOR1a LTR created an interferon regulatory factor-binding element (IRFE) that overlaps the 5’ end of the LTR. This IRFE was previously identified to be critical for promoter activity as a positive feedback loop through binding of various IRFs, including IRF5 itself [120]. Hence, the inherent promoter motifs of the LTR, coupled with the creation of the IRFE upon insertion, combined to provide an avenue for ectopic expression of IRF5 in HL.

Expression of truncated proteins

In these cases, a TE-initiated transcript results in the expression of a truncated open reading frame of the affected gene, typically because the TE is located in an intron, downstream of the canonical translational start site. The TE initiates transcription, but the final transcript structure depends on the position of downstream splice sites, and protein expression requires usage of a downstream ATG. Protein truncations can result in oncogenic effects due to loss of regulatory domains or through other mechanisms, with a classic example being...
Fig. 1 (See legend on next page.)
v-myb, a truncated form of myb carried by acutely transforming animal retroviruses [121, 122]. The first such case reported involving a TE was identified in a screen of human ESTs to detect transcripts driven by the antisense promoter within L1 elements. Määtlik et al. identified an L1PA2 within the second intron of the proto-oncogene MET (a THE1 LTR initiates transcription and contains a splice donor site which joins to an exon within a LINE L1MB5 element and then into the first exon of CSF1R. The TE-initiated transcript has a different, longer 5′ UTR than the canonical transcript but the same full-length protein coding sequence. An LOR1a1a LTR initiates transcription and splices into the canonical second exon of MET that contains the standard translational initiation site (TIS) to produce a full-length protein. There also is a novel second exon which is non-TE derived which is incorporated into a minor isoform of LOR1a-IRF5. Within the canonical intron 2 of the proto-oncogene MET, a full length LINE L1PA2 element initiates transcription (anti-sense to itself), splicing through a short exon in a SINE MIR element and into the third exon of MET. The first TIS of the canonical MET transcript is 14 bp into exon 2, although an alternative TIS exists in exon 3, which is believed to also be used by the L1-promoter isoform. An LTR6B2 element in intron 19 of the ALK gene initiates transcription and transcribes into the canonical exon 20 of ALK. An in-frame TIS within the 20th exon results in translation of a shortened oncogenic protein containing only the intra-cellular tyrosine kinase domain, but lacking the transmembrane and extracellular receptor domains of ALK. There are two TE-promoted isoforms of ERBB4, the minor variant initiates in an MLT1C LTR in the 12th intron and the major variant initiates in a MLT1H LTR in the 20th intron. Both isoforms produce a truncated protein, although the exact translation start sites are not defined. f In the third exon of SLCO1B3, two adjacent partly full-length HERV elements conspire to create a novel first exon. Transcription initiates in the anti-sense orientation from an LTR7 and transcribes to a sense-oriented splice donor in an adjacent MER4 LTR, which then splices into the fourth exon of SLCO1B3, creating a smaller protein. g An LTR2 element initiates anti-sense transcription (relative to its own orientation) and splices into the native second exon of FABP7. The LTR-derived isoform has a non-TE 5′ UTR and splice donor which creates a different N-terminal protein sequence of FABP7.

found that ALKATI stimulates several oncogenic signaling pathways, drives cell proliferation in vitro, and promotes tumor formation in mice [38]. The ALKATI promoter is a sense-oriented solitary LTR (termed LTR16B2) derived from the ancient ERVL family (Fig. 1d). LTR16B2 elements are found in several hundred copies in both primates and rodents [50, 114] and this particular element is present in the orthologous position in mouse. Therefore, the promoter potential of this LTR has been retained for at least 70 million years. Although not the first such case, the authors state that their findings “suggest a novel mechanism of oncogene activation in cancer through de novo alternative transcript initiation”. Evidence that this LTR is at least occasionally active in normal human cells comes from Capped Analysis of Gene Expression (CAGE) analysis through the FANTOM5 project [128]. A peak of CAGE tags from monocyte-derived macrophages and endothelial progenitor cells occurs within this LTR, 60 bp downstream of the TSS region identified by Wiesner et al. [38] (Fig. 2a), although a biological function, if any, of this isoform in normal cells is unknown.

To gain a molecular understanding of ALK-negative anaplastic large-cell lymphoma (ALCL) cases, Scarfo et al. conducted gene expression outlier analysis and identified high ectopic co-expression of ERBB4 and COL29A1 in 24% of such cases [129]. Erb-b2 receptor tyrosine kinase 4 (ERBB4), also termed HER4, is a member of the ERBB family of RTKs, which includes EGFR and HER2, and mutations in this gene have been implicated in some cancers [130]. Analysis of the ERBB4 transcripts expressed in these ALCL samples revealed two isoforms initiated from alternative promoters, one within intron 12 (I12-ERBB4) and one within intron 20 (I20-ERBB4), with little or no expression from the native/canonical promoter. Both isoforms produce truncated proteins that show oncogenic potential, either alone
(I12 isoform) or in combination. Remarkably, both promoters are LTR elements of the ancient MaLR-ERVL class (Fig. 1e). Of note, Scarfo et al. reported that two thirds of ERBB4 positive cases showed a "Hodgkin-like" morphology, which is normally found in only 3% of ALCLs [129]. We therefore examined our previously published RNA-Seq data from 12 HL cell lines [117] and found evidence for transcription from the intron 20 MLTH2 LTR in two of these lines (unpublished observations), suggesting that truncated ERBB4 may play a role in some HLs.

**TE-promoted expression of chimeric proteins**

Perhaps the most fascinating examples of onco-exaptation involve generation of a novel "chimeric" ORF via usage of a TE promoter that fuses otherwise non-coding DNA to downstream gene exons. These cases involve both protein and transcriptional innovation and the resulting product can acquire de novo oncogenic potential.

The solute carrier organic transporter family member 1B3, encodes organic anion transporting polypeptide 1B3 (OATP1B3, or SLCO1B3), is a 12-transmembrane transporter with normal expression and function restricted to the liver [131]. Several studies have shown that this gene is ectopically expressed in solid tumors of non-hepatic origin, particularly colon cancer [131–134]. Investigations into the cause of this ectopic expression revealed that the normal liver-restricted promoter is silent in these cancers, with expression of “cancer-type” (Ct)-OATP1B3 being driven from an alternative promoter in the second canonical intron [133, 134]. While not previously reported as being within a TE, we noted that this alternative promoter maps within the 5’LTR (LTR7) of a partly full-length antisense HERV-H element that is missing the 3’LTR. Expression of HERV-H itself and LTR7-driven chimeric long non-coding RNAs is a noted feature of embryonic stem cells and normal early embryogenesis, where several studies indicate an intriguing role for this ERV group in pluripotency (for recent reviews see [8, 10, 60]). A few studies have also noted higher general levels of HERV-H transcription in colon cancer [109, 135]. The LTR7-driven isoform of SLCO1B3 makes a truncated protein lacking the first 28 amino acids but also includes protein sequence from the LTR7 and an adjacent MER4C LTR (Fig. 1f). The novel protein is believed to be intracellular and its role in cancer remains unclear. However, one study...
showed that high expression of this isoform is correlated with reduced progression-free survival in colon cancer [136].

In another study designed specifically to look for TE-initiated chimeric transcripts, we screened RNA-seq libraries from 101 patients with diffuse large B-cell lymphoma (DLBCL) of different subtypes [137] and compared transcriptomes from normal B-cells. This screen resulted in the detection of 98 such transcripts that were found in at least two DLBCL cases and no normals [138]. One of these involved the gene for fatty acid binding protein 7 (FABP7). FABP7, normally expressed in brain, is a member of the FABP family of lipid chaperones involved in fatty acid uptake and trafficking [139]. Overexpression of FABP7 has been reported in several solid tumor types and is associated with poorer prognosis in aggressive breast cancer [139, 140]. In 5% of the DLBCL cases screened, we found that FABP7 is expressed from an antisense LTR2 (the 5′LTR of a HERV-E element) (Fig. 1g). Since the canonical ATG is in the first exon of FABP7, the LTR driven transcript encodes a chimeric protein with a different N-terminus (see accession NM_001319042.1) [138]. Functional analysis in DLBCL cell lines revealed that the LTR-FABP7 protein isoform is required for optimal cell growth and also has subcellular localization properties distinct from the native form [138].

Overall, among all TE types giving rise to chimeric transcripts detected in DLBCL, LTRs were over represented compared to their genomic abundance and, among LTR groups, we found that LTR2 elements and THE1 LTRs were over represented [138]. As discussed above, this predominance of LTRs over other TE types is expected.

**Table 3** LTR-driven LncRNAs with oncogenic role

| LncRNAa | TE type | TE promoter coordinates (hg38) | Cancer type | references |
|---------|---------|-----------------------------|-----------|-----------|
| SchLAP1 | (ERV1) LTR12C | chr2:180691205-180692425 | prostate | [143] |
| ROSE2 | (ERV1) LTR7 | chr1857072052-57072502 | breast, others | [147, 150] |
| HOST2 | (ERV1) LTR2B | chr1084171987-81472465 | ovarian | [154] |
| AFAP1-AS1 | (ERV1-MaLR) THE1A LTR6 | chr4:77533884-7754236 | several | [156, 158] |
| SAMMSON | (ERV1) LTR1A2D | chr3:69999501-70003359 | melanoma | [161] |
| HULC | (ERV1-MaLR) MLT1A LTR | chr6:8662095-8662454 | liver | [163] |
| UCA1 | (ERV1) LTR7C | chr19:15828738-15829200 | several | [165, 167] |
| BANCR | (ERV1) MER41B LTR | chr9:69306939-69307567 | melanoma, others | [169] |

*Only those cases with supporting evidence of a role in the cancer are listed.

The fact that the promoter is an LTR was not previously noted.

**TE-initiated LncRNAs with oncogenic properties**

In an extensive study, Prensner et al. reported that the lncRNA *SchLAP1* (SWI/SNF complex antagonist associated with prostate cancer 1) is overexpressed in ~25% of prostate cancers, is an independent predictor of poor clinical outcomes and is critical for invasiveness and metastasis [143]. Intriguingly, they found that *SchLAP1* inhibits the function of the SWI/SNF complex, which is known to have a tumor suppressor roles [144]. While not mentioned in the main text, the authors report in supplementary data that the promoter for this lncRNA is an LTR (Fig. 3a). Indeed, this LTR is a sense-oriented solitary LTR12C (of the ERV9 group).

**Linc-ROR** is a non-coding RNA (long intergenic non-protein coding RNA, regulator of reprogramming) promoted by the 5′ LTR (LTR7) of a full length HERV-H element [56] (Fig. 3b) and has been shown to play a role in human pluripotency [145]. Evidence suggests it acts as a microRNA sponge of miR-145, which is a repressor of the core pluripotency transcription factors Oct4, Nanog and Sox2 [146]. Several recent studies have reported an oncogenic role for **Linc-ROR** in different cancers by sponging miR-145 [147–149] or through other mechanisms [150, 151].

Using Serial Analysis of Gene Expression (SAGE), Rangel et al. identified five Human Ovarian cancer Specific Transcripts (HOSTs) that were expressed in ovarian cancer but not in other normal cells or cancer types. These cases can be broadly divided into those with direct, measurable oncogenic properties (Table 3) and those with expression correlated with a cancer. It should be noted that we have likely missed some examples if the nature of the promoter was not highlighted or mentioned in the original publications.
examined [152]. One of these, HOST2, is annotated as a spliced lncRNA entirely contained within a full length HERV-E and promoted by an LTR2B element (Fig. 3c). Perusal of RNA-Seq from the 9 core ENCODE cell lines shows robust expression of HOST2 in GM12878, a B-lymphoblastoid cell line, which extends beyond the HERV-E. As with Linc-ROR, HOST2 appears to play an oncogenic role by functioning as a miRNA sponge of miRNA let-7b, an established tumor suppressor [153], in epithelial ovarian cancer [154].

The Ref-Seq annotated lncRNA AFAP1 antisense RNA 1 (AFAP1-AS1) runs antisense to the actin filament associated protein 1 (AFAP1) gene and several publications report its up-regulation and association with poor survival in a number of solid tumor types [155–158]. While the oncogenic mechanism of AFAP1-AS1 has not been extensively studied, one report presented evidence that it promotes cell proliferation by upregulating RhoA/Rac2 signaling [159] and its expression inversely correlates with AFAP1. Although clearly annotated as initiating within a solitary THE1A LTR (Fig. 3d), this fact has not been mentioned in previous publications. In screens for TE-initiated transcripts using RNA-seq data from HL cell lines, we noted recurrent and cancer-specific up-regulation of AFAP1-AS1 (unpublished observations), suggesting that it is not restricted to solid tumors. The inverse correlation of expression between AFAP1 and AFAP1-AS1 suggests an interesting potential mechanism by which TE-initiated transcription may suppress a gene; where an anti-sense TE-initiated transcript disrupts the transcription, translation or stability of a tumor suppressor gene transcript through RNA interference [160].
The SAMMSON IncRNA (survival associated mitochondrial melanoma specific oncogenic non-coding RNA), which is promoted by a solitary LTR1A2 element, was recently reported as playing an oncogenic role in melanoma [161]. This IncRNA is located near the melanoma-specific oncogene MITF and is always included in genomic amplifications involving MITF. Even in melanomas with no genomic amplification of this locus, SAMMSON is expressed in most cases, increases growth and invasiveness and is a target for SOX10 [161], a key TF in melanocyte development which is deregulated in melanoma [162]. Interestingly, the two SOX10 binding sites near the SAMMSON TSS lie just upstream and downstream of the LTR (Fig. 2b), suggesting that both the core promoter motifs provided by the LTR and adjacent enhancer sites combine to regulate SAMMSON.

Other examples of LTR-promoted oncogenic IncRNAs include HULC for Highly Upregulated in Liver Cancer [163, 164], UCA1 (urothelial cancer associated 1) [165–168] and BANCR (BRAF-regulated IncRNA 1) [169–171]. Although not mentioned in the original paper, three of the four exons of BANCR were shown to be derived from a partly full length MER41 ERV, with the promoter within the 5’LTR of this element annotated MER41B [141]. Intriguingly, MER41 LTRs were recently shown to harbor enhancers responsive to interferon, indicating a role for this ERV group in shaping the innate immune response in primates [52]. It would be interesting to investigate roles for BANCR with this in mind.

**TE-initiated IncRNAs as cancer-specific markers**

There are many examples of TE-initiated RNAs with potential roles in cancer or which are preferentially expressed in malignant cells but for which a direct oncogenic function has not yet been demonstrated. Still, such transcripts may underlie a predisposition for transcription of specific groups of LTRs/TEs in particular malignancies and therefore function as a marker for a cancer or cancer subtype. Since these events potentially do not confer a fitness advantage for the cancer cell, they are not “exaptations” but “nonaptations” [4].

One of these is a very long RNA initiated by the antisense promoter of an L1PA2 element as reported by Tufarelli’s group and termed LCT13 [172, 173]. EST evidence indicates splicing from the L1 promoter to the GNTG1 gene, located over 300 kb away. The tumor suppressor gene, tissue factor pathway inhibitor 2, (TFPI-2), which is often epigenetically silenced in cancers [174], is antisense to LCT13 and it was shown that LCT13 transcript levels are correlated with down regulation of TFPI-2 and associated with repressive chromatin marks at the TFPI-2 promoter [172].

Gibb et al. analyzed RNA-Seq from colon cancers and matched normal colon to find cancer-associated IncRNAs and identified an RNA promoted by a solitary MER48 LTR, which they termed EVADR, for Endogenous retroviral-associated ADenocarcinoma RNA [175]. Screening of data from The Cancer Genome Atlas (TCGA) [176] showed that EVADR is highly expressed in several types of adenocarcinomas, it is not associated with global activation of MER48 LTRs across the genome and its expression correlated with poorer survival [175]. In another study, Gosencsa et al. used a custom microarray to measure overall expression of several HERV groups in urothelial carcinoma compared to normal urothelial tissue and generally found no difference [111]. However, they found one full-length HERV-E element, located in the antisense direction in an intron of the PLA2G4A gene that is transcribed in urothelial carcinoma and appears to modulate PLA2G4A expression, thereby possibly contributing to carcinogenesis, although the mechanism is not clear.

By mining long nuclear RNA datasets from ENCODE cell lines, normal blood and Ewing sarcomas, one group identified over 2000 very long (~50–700 kb) non coding transcripts termed vlincRNAs [142]. They found the promoters for these vlincRNAs to be enriched in LTRs, particularly for cell type-specific vlincRNAs, and the most common transcribed LTR types varied in different cell types. Moreover, among the datasets examined, they reported that the number of LTR-promoted vlincRNAs correlated with degree of malignant transformation, prompting the conclusion that LTR-controlled vlincRNAs are a “hallmark” of cancer [142].

In a genome-wide CAGE analysis of 50 hepatocellular carcinoma (HCC) primary samples and matched non-tumor tissue, Hashimoto et al. found that many LTR-promoted transcripts are upregulated in HCC, most of these apparently associated with non-coding RNAs as the CAGE peaks in the LTRs are far from annotated protein coding genes [177]. Similar results were found in mouse HCC. Among the hundreds of human LTR groups, they found the LTR-associated CAGE peaks to be significantly enriched in LTR12C (HERV9) LTRs and mapped the common TSS site within these elements, which agrees with older studies on TSS mapping of this ERV group [178]. Moreover, this group reported that HCCs with highest LTR activity mostly had a viral (Hepatitis B) etiology, were less differentiated and had higher risk of recurrence [177]. This study suggests widespread tissue-inappropriate transcriptional activity of LTRs in HCC.

**LTR12s as flexible promoters in cancer and normal tissues**

Most recent human ERV LTR research has been focused on HERV-H (LTR7/7Y/7B/7C) due to roles for
HERV-H/LTR7-driven RNAs in pluripotency [56–58, 60, 179, 180] or on the youngest HERV group, HERV-K (LTR5/5Hs), due to its expression in early embryogenesis [181–183], coding capacity of some members [30, 184] and potential roles for its proteins in cancer and other diseases [30–33, 185]. LTR12s (including LTR12B,C,D,E and F subtypes), which are the LTRs associated with the HERV-9 group [186], are generally of similar age to HERV-H [187] but are much more numerous than HERV-H or HERV-K, with solitary LTRs numbering over 6000 (Table 1). There are several examples of LTR12s providing promoters for coding genes or IncRNAs in various normal tissues [63, 188–191]. LTR12s, particularly LTR12C, are longer and more CpG rich than most other ERV LTRs, possibly facilitating development of diverse inherent tissue-specificities and flexible combinations of TF binding sites, which may be less probable for other LTR types. For example, the consensus LTR7 (HERV-H) is 450 bp whereas LTR12C (of similar age) is 1577 bp [114], which is usually long for retroviral LTRs. As noted above, LTR12 elements are among the most enriched LTR types activated as promoters in HCC [177] and appear to be the most active LTR type in K562 cells [142]. It is important to point out, however, that only a very small fraction of genomic LTR12 copies are transcriptionally active in any of these contexts, so general conclusions about activity of ‘a family of LTRs’ should be made with caution.

A number of other recent investigations on LTR12-driven chimeric transcription have been published. One study specifically screened for and detected numerous LTR12-initiated transcripts in ENCODE cell lines, some of which extend over long genomic regions and emanate from bidirectional promoters within these LTRs [192]. The group of Dobbelstein discovered that a male germ line-specific form of the tumor suppressor TP63 gene is driven by an LTR12C [190]. Interestingly, they found that this LTR is silenced in testicular cancer but reactivated upon treatment with histone deacetylase inhibitors (HDACi), which also induces apoptosis [190]. In follow-up studies, this group used 3’ RACE to detect more genes controlled by LTR12s in primary human testis and in the GH testicular cancer cell line and reported hundreds of transcripts, including an isoform of TNFRSF10B which encodes the death receptor DR5 [193]. As with TP63, treating GH or other cancer cell lines with HDAC inhibitors such as trichostatin A activated expression of the LTR12-driven TNFRSF10B and some other LTR12-chimeric transcripts and induced apoptosis [193, 194]. Therefore, in some cases, LTR-driven genes can have a proapoptotic role. In accord with this notion is a study reporting that LTR12 antisense U3 RNAs were expressed at higher levels in non-malignant versus malignant cells [195]. It was proposed that the antisense U3 RNA may act as a trap for the transcription factor NF-Y, known to bind LTR12s [196], and hence participate in cell cycle arrest [195].

### Chromosomal translocations involving TEs in cancer

Activation or creation of oncogenes via chromosomal translocations most commonly involves either the fusion of two coding genes or juxtaposition of new regulatory sequences next to a gene, resulting in oncogenic effects due to ectopic expression [197]. One might expect some of the latter cases to involve TE-derived promoters/enhancers but, to date, there are very few well-documented examples of this mechanism in oncogenesis. The ETS family member ETV1 (ETS variant 1) is a transcription factor frequently involved in oncogenic translocations, particularly in prostate cancer [198]. Although not a common translocation, Tomlins et al. identified a prostate tumor with the 5’ end of a HERV-K (HML-2) element on chromosome 22q11.23 fused to ETV1 [199]. This particular HERV-K element is a complex locus with two 5’ LTRs and is quite highly expressed in prostate cancer [200]. Indeed, while a possible function is unknown, this HERV-K locus produces a IncRNA annotated as PCAT-14, for prostate cancer–associated ncRNA transcript-14 [201]. In the HERV-K-ETV1 fusion case, the resultant transcript (Genbank Accession Ef632111) initiates in the upstream 5’ LTR, providing evidence that the LTR controls expression of ETV1.

The fibroblast growth factor receptor 1 (FGFR1) gene on chromosome 8 is involved in translocations with at least 14 partner genes in stem cell myeloproliferative disorder and other myeloid and lymphoid cancers [202]. One of these involves a HERV-K3 element on chromosome 19 and this event creates a chimeric ORF with HERV-K3 gag sequences [203]. While it was reported that the LTR promoter may contribute to expression of the fusion gene [203], no supporting evidence was presented. Indeed, perusal of public expression data (Expressed sequence tags) from a variety of tissues indicates that the HERV-K3 element on chromosome 19 is highly expressed, but from a non-ERV promoter just upstream (see chr19:58,305,253–58,315,303 in human hg38 assembly). Therefore, there is little current evidence for LTR/TE promoters playing a role in oncogene activation via chromosomal translocations or rearrangements.

### Models for onco-exaptation

The aforementioned cases of onco-exaptation are a distinct mechanism by which proto-oncogenes become oncogenic. Classical activating mutations within TEs may also lead to transcription of downstream oncogenes but we are unaware of any evidence for DNA mutations resulting in LTR/TE transcriptional activation, including cases where local DNA was sequenced [38] (unpublished
results). Thus, it is important to consider the etiology through which LTRs/TEs become incorporated into new regulatory units in cancer. The mechanism could possibly be therapeutically or diagnostically important and perhaps even model how TEs influence genome regulation in evolutionary time.

In some of the above examples, there is no or very little detectable transcription from the LTR/TE in any cell type other than the cancer type in which it was reported, suggesting the activity is specific to a particular TE in a particular cancer. In other cases, CAGE or EST data show that the LTR/TE can be expressed in other normal or cancer cell types, perhaps to a lower degree. Hence the term “cancer-specific” should be considered a relative one. Indeed, the idea that the same TE-promoted gene transcripts occur recurrently in tumors from independent individuals is central to understanding how these transcripts arise. Below we present two models that may explain the phenomenon of onco-exaptation.

**The De-repression model**

Lamprecht and co-workers proposed a ‘De-repression model’ for the LTR driven transcription of CSF1R [204]. The distinguishing feature of this model is that onco-exaptations arise deterministically, as a consequence of molecular changes that occur during oncogenesis, changes which act to de-repress LTRs or other TEs (Fig. 4). It follows that ‘activation’ of normally dormant TEs/LTRs could lead to robust oncogene expression. In the CSF1R case, the THE1B LTR, which promotes CSF1R in HL, contains binding sites for the transcription factors Sp1, AP-1 and NF-kB, each of which contributes to promoter activity in a luciferase reporter experiment [113]. High NF-kB activity, which is known to be up-regulated in HL, loss of the epigenetic corepressor CBFA2T3 as well as LTR hypomethylation all correlated with CSF1R-positive HL driven by the LTR [113]. Under the de-repression model, the THE1B LTR is repressed by default in the cell but under a particular set of

![De-repression model for onco-exaptation](image_url)

**Fig. 4** De-repression model for onco-exaptation. In the normal or pre-malignant state TEs (grey triangles) are largely silenced across the genome. There is low transcriptional activity to produce long non-coding RNA (orange box), or express coding genes in the case of evolutionary exaptations (not shown). The example proto-oncogene (green box) is under the regulatory control of its native, restrictive promoter. During the process of transformation and/or oncogenesis, a change in the molecular state of the cell occurs leading to loss of TE repressors (black circles), i.e. DNA hypomethylation, loss of transcriptional or epigenetic repressive factors. The change could also be accompanied by a change/gain in activating factor activities (red and purple shapes). Together these de-repression events result in higher TE promoter activity (orange triangles) and more TE-derived transcripts based on the factors that become deregulated. Oncogenic activation of proto-oncogenes is a consequence of a particular molecular milieu that arises in the cancerous cells.
conditions (gain of NF-kB, loss of CBFA2T3, loss of DNA methylation) the LTR promoter is remodeled into an active state [204]. More generally, the model proposes that a particular LTR activation is a consequence of the pathogenic or disrupted molecular state of the cancer cell. In a similar vein, Weber et al. proposed that the L1-driven transcription of MET arose as a consequence of global DNA hypomethylation and loss of repression of TEs in cancer [124].

The LOR1a-IRF5 on-co-exaptation in HL [117] can be interpreted using a de-repression model. An interferon regulatory factor binding element site was created at the intersection of the LOR1a LTR and genomic DNA. In normal and HL cells negative for LOR1a-IRF5, the LTR is methylated and protected from DNase digestion, a state that is lost in de-repressed HL cells. This transcription factor-binding motif is responsive to IRF5 itself and creates a positive feedback loop between the IRF5 and the chimeric LOR1a-IRF5 transcript. Thus epigenetic de-repression of this element may reveal an oncogenic exploitation, resulting in high recurrence of LOR1a LTR-driven IRF5 in HL [117].

A de-repression model explains several experimental observations, such as the necessity for a given set of factors to be present (or absent) for a certain promoter to be active, especially when those factors differ between cell states. Indeed, experiments probing the mechanism of TE/LTR activation have used this line of reasoning, often focusing on DNA methylation [113, 117, 125, 129]. The limitation of these studies is that they fail to determine if a given condition is sufficient for onco-exaptation to arise. For instance, the human genome contains ~37,000 THE1 LTR loci (Table 1), and indeed this set of LTRs is generally more active in HL cells compared to B-cells as would be predicted [113] (unpublished results). The critical question is why particular THE1 LTR loci, such as THE1b-CSF1R, are recurrently de-repressed in HL, yet thousands of homologous LTRs are not.

**The Epigenetic Evolution model**

A central premise in the TE field states that TEs can be beneficial to a host genome since they increase genetic variation in a population and thus increase the rate at which evolution (by natural selection) occurs [62, 205, 206]. The epigenetic evolution model for onco-exaptation (Fig. 5) draws a parallel to this premise within the context of tumor evolution.

Key to the epigenetic evolution model is that there is high epigenetic variance, both between LTR loci and at the same LTR locus between cells in a population. This epigenetic variance fosters regulatory innovation, and increases during oncogenesis. In accord with this idea are several studies showing that DNA methylation variation, or heterogeneity, increases in tumor cell populations and this isn’t simply a global hypomethylation relative to normal cells [207–209] (reviewed in [210]). In contrast to the de-repression model, a particular pathogenic molecular state is not sufficient or necessary for TE-driven transcripts to arise; instead the given state only dictates which sets of TEs in the genome are permissive for transcription. Likewise, global de-repression events, such as DNA hypomethylation or mutation of epigenetic regulators, are not necessary, but would increase the rate at which novel transcriptional regulation evolves.

Underpinning this model is the idea that LTRs are highly abundant and self-contained promoters dispersed across the genome that can stochastically initiate low or noisy transcription. This transcriptional noise is a kind of epigenetic variation and thus contributes to cell-cell variation in a population. Indeed, by re-analyzing CAGE datasets of retrotransposon-derived TSSs published by Faulkner et al. [73], we observed that TE-derived TSSs have lower expression levels and are less reproducible between biological replicates, compared to non-TE promoters (unpublished observations). During malignant transformation, TFs can become deregulated and genomewide epigenetic perturbations occur [94, 98, 211] which would change the set of LTRs that are potentially active as well as possibly increasing the total level of LTR-driven transcriptional noise. Up-regulation of specific LTR-driven transcripts would initially be weak and stochastic, from the set of permissive LTRs. Those cells gaining an LTR-driven transcript which confers a growth advantage would then be selected for, and the resultant oncogene expression would increase in the tumor population as that epiallele increases in frequency, in a similar fashion as proposed for the epigenetic silencing of tumor suppressor genes [95, 99, 100]. Notably, this scenario also means that within a tumor, LTR-driven transcription would be subject to epigenetic bottleneck effects as well, and that transcriptional LTR noise can become “passenger” expression signals as the cancer cells undergo somatic, clonal evolution.

It may be counter-intuitive to think of evolution and selection as occurring outside the context of genetic variation, but the fact that both genetic mutations and non-genetic/epigenetic variants can contribute to somatic evolution of a cancer is becoming clear [209, 212–215]. Epigenetic information or variation by definition is transmitted from mother to daughter cells. Thus, in the specific context of a somatic/asexual cell population such as a tumor, this information, which is both variable between cells in the population and heritable, will be subject to evolutionary changes in frequency. DNA methylation in particular has a well-established mechanism by which information (mainly gene repression) is transmitted epigenetically from mother to daughter cells [216] and DNA hypomethylation at LTRs often correlates with their
Fig. 5 Epigenetic evolution model for onco-exaptation. In the starting cell population there is a dispersed and low/noisy promoter activity at TEs (colored triangles) from a set of transcriptionally permissive TEs (grey triangles). TE-derived transcript expression is low and variable between cells. Some transcripts are more reliably measurable (orange box). Clonal tumor evolutionary forces change the frequency and expression of TE-derived transcripts by homogenizing epialleles and use of TE promoters (highlighted haplotype). A higher frequency of ‘active’ TE epialleles at a locus results in increased measurable transcripts initiating from that position. TE epialleles that promote oncogenesis, namely onco-exaptations, can be selected for and arise multiple times independently as driver epialleles, in contrast to the more dispersed passenger epialleles, or “nonaptations”
expression [113, 117, 217]. Thus, this model suggests that one important type of “epigenetic variant” or epiallele is the transcriptional status of the LTR itself, since the phenotypic impact of LTR transcription may be high in onco-exaptation. Especially in light of the fact that large numbers of these highly homologous sequences are spread across the genome, epigenetic variation, and possibly selection, at LTRs creates a fascinating system by which epigenetic evolution in cancer may occur.

Conclusions
Here we have reviewed the growing number of examples of LTR/TE onco-exaptation. Although such TEs have the potential to be deleterious by contributing to oncogenesis if transcriptionally activated, their fixation in the genome and ancient origin suggests that their presence is not subject to significant negative selection. This could be due to the low frequency of onco-exaptation at a particular TE locus and/or to the fact that cancer is generally a disease that occurs after the reproductive years. However, it is generally assumed that negative selection is the reason why TEs are under-represented near or within genes encoding developmental regulators [218–220]. Similarly we hypothesize that LTR/TE insertions predisposed to causing potent onco-exaptations at a high frequency would also be depleted by selective forces.

In this review we have also presented two models that may explain such onco-exaptation events. These two models are not mutually exclusive but they do provide alternative hypotheses by which TE-driven transcription may be interpreted. This dichotomy is possibly best exemplified by the ERBB4 case (Fig. 1e) [129]. There are two LTR-derived promoters which result in aberrant ERBB4 expression in ALCL. From the de-repression model viewpoint, both LTR elements are grouped MLT1 (MLT1C and MLT1H) and thus this group can be interpreted as de-repressed. From the epigenetic evolution model viewpoint, this is convergent evolution/selection for onco-exaptations involving ERBB4.

Through application of the de-repression model, TE-derived transcripts could be used as a diagnostic marker in cancer. If the set of TE/LTR derived transcripts are a deterministic consequence of a given molecular state, by understanding which set of TEs correspond to which molecular state, it might be possible to assay cancer samples for functional molecular phenotypes. In HL for example, CSF1R status is prognostically important [115] and this is dependent on the transcriptional state of a single THE1B. HL also has a specific increase in THE1 LTR transcription genome-wide (unpublished observations). Thus, it’s reasonable to hypothesize that the prognostic power can be increased if the transcriptional status of all THE1 LTRs is considered. A set of LTRs can then be interpreted as an in situ ‘molecular sensor’ for aberrant NF-κB function in HL/B-cells for instance.

The epigenetic evolution model proposes that LTR-driven transcripts can be interpreted as a set of epimutations in cancer, similar to how oncogenic mutations are analyzed. Genes that are recurrently (and independently) onco-exapted in multiple different tumors of the same cancer type may be a mark of selective pressure for acquiring that transcript. This is distinct from the more diverse/noisy “passenger LTR” transcription occurring across the genome. These active but “passenger LTRs” may be expressed to a high level within a single tumor population due to epigenetic drift and population bottlenecks but would be more variable across different tumors. Thus analysis of recurrent and cancer-specific TE-derived transcripts may enrich for genes of significance to tumor biology.

While we focused in this review on TE-initiated transcription in cancer, many of the concepts presented here can be applied to other regulatory functions of TEs such as enhancers, insulators, or repressors of transcription. Although less straightforward to measure, it is probable that perturbations to such TE regulatory functions contribute to some malignancies. Furthermore, several studies have shown that TEs play substantial roles in splicing in humans [221–223] and thus may be a further substrate of transcriptional innovation in cancer, particularly since DNA methylation state can affect splicing [224].

Regardless of the underlying mechanism, onco-exaptation offers a tantalizing opportunity to model evolutionary exaptation. Specifically, questions such as “How do TEs influence the rate of transcriptional/ regulatory change?” can be tested in cell culture experiments. As more studies that focus on regulatory aberrations in cancer are performed in the coming years, we predict that this phenomenon will become increasingly recognized as a significant force shaping transcriptional innovation in cancer. Moreover, we propose that studying such events will provide insight into how TEs have contributed to reshaping transcriptional patterns during species evolution.

Abbreviations
AFAP1-AS1: AFAP1 antisense RNA 1; ALCCL: Anaplastic large-cell lymphoma; ALK: Anaplastic lymphoma kinase; BANCR: BRAF-regulated IncRNA 1; CAGE: Capped analysis of gene expression; CSFIR: Colony stimulating factor one receptor; DLBCL: Diffuse large B-cell lymphoma; ERBB4: Erb-b2 receptor tyrosine kinase 4; ERV: Endogenous retrovirus; EST: Expressed sequence tag; ETV1: ETS variant 1; EVADR: Endogenous retroviral-associated Adenocarcinoma RNA; FABP7: Fatty acid binding protein 7; HCC: Hepatocellular carcinoma; HL: Hodgkin lymphoma; HOST2: Human ovarian cancer specific transcript-2; HULC: Highly upregulated in liver cancer; IRF5: Interferon regulatory factor 5; IRFE: Interferon regulatory factor-binding element; Linc-ROR: Long intergenic non-protein coding RNA, regulator of reprogramming; LINE-1: L1: Long interspersed repeat-1; LncRNA: Long non-coding RNA; LTR: Long terminal repeat; MET: MET proto-oncogene, receptor tyrosine kinase; OAT1B3: Organic anion transporting polypeptide 1B3;
SAMWSON: Survival associated mitochondrial melanoma specific oncogenic non-coding RNA; SChLAP1: SWI/SNF complex antagonist associated with prostate cancer 1; SINE: Short interspersed element; SLCO1B3: Solute carrier organic anion transporter family member 1B3; TCGA: The cancer genome atlas; TE: Transposable element; TF: Transcription factor; TFP1-2: Tissue factor pathway inhibitor 2; TIS: Translation initiation site; TSS: Transcriptional start site; UCA1: Urothelial cancer associated 1

Acknowledgments
We thank Matt Lorincz and the anonymous reviewers for comments and helpful suggestions on this manuscript. We apologize to colleagues and other researchers if we failed to cite relevant work on this subject.

Funding
Work on this topic in our laboratory has been funded by grants from the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Cancer Society and the Leukemia and Lymphoma Society of Canada, with core support provided by the BC Cancer Agency. AB is supported by a studentship award from NSERC.

Availability of data and materials
Data sharing not applicable as no datasets were generated or analyzed during the current study.

Authors’ contributions
AB and DLM wrote the manuscript and both authors approved the final version.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Received: 1 September 2016 Accepted: 11 November 2016
Published online: 01 December 2016

References

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
2. Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive Sequences in Complex Genomes: Structure and Evolution. Annu Rev Genomics Hum Genet. 2007;8(1):241–9.
3. Brosius J, Gould SJ. On “genomenculture”: a comprehensive (and respectful) taxonomy for pseudogenes and other “Junk DNA”. Proc Natl Acad Sci. 1992;89(22):10706–10.
4. Gould SJ, Vija ES. Exaptation-A Missing Term in the Science of Form. Paleobiology. 1982(8):14–15.
5. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and implications for pseudogenes and other “junk DNA”. Proc Natl Acad Sci. 1992;89(22):10706–10.
6. Rebollo R, Romanius MT, Mager DL. Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. Annu Rev Genet. 2012;46:21–42.
7. Richardson SR, Doucet AJ, Kopeca HC, Moldovan JB, Garcia-Pérez JL, Moran JV. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr. 2015;3(2):1112.18(microbiolspec/MN12:0061-2014
8. Robbez-Masson L, Rowe H. Retrotransposons shape species-specific embryonic stem cell gene expression. Retrovirology. 2015;12(1):45.
9. Hancks DC, Kazanjan HH. Roles for retrotransposon insertions in human disease. Mob DNA. 2016;7(1):1–28.
10. Gerdes P, Richardson SR, Mager DL, Faulkner GJ. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. Genome Biol. 2016;17(1):1–17.
11. Weiss RA. Human endogenous retroviruses: friend or foe? PAMIS. 2016;124(1–2):24–10.
12. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. 2016;351(6274):ac2747.
13. Brouha B, Schustaj J, Badge RM, Lutz-Pirgrte S, Farley AH, Moran JV, Kazanjan JH. Hot L1s account for the bulk of retrotransposition in the human population. PNAS. 2003;100(9):5280–5.
14. Chen J, Stenson P, Cooper D, Furec A. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet. 2005;117(5):411–27.
15. Kaer K, Speek M. Retroelements in human disease. Gene. 2013;518(2):231–41.
16. Miyi Y, Nishioh I, Honi A, Miyoshi Y, Utsunomiya J, Kinder KW, Vogelstein B, Nakamura Y. Disruption of the APC gene by a retrotranspositional insertion of L1 sequence in a colon cancer. Cancer Res. 1992;52(3):643–5.
17. Lee E, Isikow R, Yang L, Gokcumen O, Haslepy P, Luquette LJ, Lohr JG, Harris CC, Ding L, Wilson RK, et al. Landscape of Somatic Retrotransposition in Human Cancers. Science. 2012;337(6097):967–71.
18. Sol Smy S, Ewing AD, Rahimman EP, Doucet TT, Nelson HH, Burns MB, Harris RS, Sigmon DF, Casella A, Erlanger B, et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res. 2012;22(12):2328–38.
19. Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baille JK, Collino A, Giandetti S, et al. Endogenous retrotransposition activates oncogenic pathways in hematopoietic cell carcinogenesis. Cell. 2013;153(1):101–11.
20. Tubio JMC, Li Y, Ju YS, Martincorea I, Cokele SL, Tojo M, Gundem G, Pipnikas CP, Zamora J, Raine K, et al. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science. 2014;345(6196):1251343.
21. Rodic N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, Kohutek ZA, Huang CR, Ahn D, Mita P, et al. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nat Med. 2015; 21(9):1060–4.
22. Ewing AD, Gacita A, Wood LD, Ma F, Xing D, Kim M-S, Manda SS, Abril G, Pereira G, Makohon-Moore A, et al. Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. Genome Res. 2015;25(10):1536–45.
23. Scott EC, Gardner EJ, Masood A, Chuang NT, Devine SE. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res. 2016;26(6):745–55.
24. Bannett N, Kurth R. The Evolutionary Dynamics of Human Endogenous Retroviral Families. Annu Rev Genomics Hum Genet. 2006;7(1):149–73.
25. Jern P, Coffin JM. Effects of Retroviruses on Host Genome Function. Annu Rev. Genet. 2008;42(1):709–32.
26. Magiorkinis G, Blanco-Melo D, Belshaw R. The decline of human endogenous retroviruses: extinction and survival. Retrovirology. 2015;12(1):1–12.
27. Rosenberg N, Jolicoeur P. Retroviral pathogenesis: In: Coffin JM, Hughes SH, Varmus H, editors. Retroviruses. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1997. p. 475–586.
28. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. Activation and transposition of endogenous retroviral elements in hypomethylated induced tumors in mice. Oncogene. 2008;27(3):404–8.
29. Fan H, Johnson C. Insertional Oncogenesis by Non-Acute Retroviruses: Implications for Gene Therapy. Viruses. 2011;3(6):398–422.
30. Hahn W, Kenke K, Bannett N. HERV-K(HML-2), the best preserved family of endogenous retroviruses: the possible source of retroviral insertions in cancer genomes. Hum Genet. 2004;115(6):822–3.
31. Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baille JK, Collino A, Giandetti S, et al. Endogenous retrotransposition activates oncogenic pathways in hematopoietic cell carcinogenesis. Cell. 2013;153(1):101–11.
32. Tubio JMC, Li Y, Ju YS, Martincorea I, Cokele SL, Tojo M, Gundem G, Pipnikas CP, Zamora J, Raine K, et al. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science. 2014;345(6196):1251343.
33. Kassiotis G. Endogenous Retroviruses and the Development of Cancer. J Immunol. 2014;192(4):1343–9.
34. Gomez-del Arco P, Kashiwagi M, Jackson AF, Naito T, Zhang J, Liu F, Kee B, Voois J, Radtke F, Redondo JM, et al. Alternative promoter usage at the Tcf7l2 locus supports ligand-independent signaling in T cell development and leukemogenesis. Immunity. 2010;33(5):648–59.
35. Thorsen K, Schepeler T, Ouster B, Rasmussen NH, Wang J, Mosher S, Hansen HH, Diwald S, et al. Extensive L1 retrotransposition in colorectal tumors. Genome Res. 2012;22(12):2328–38.
36. Muratani M, Deng N, Ooi WF, Lin SJ, Xing M, Xu C, Quamra A, Tay ST, Malik S, Wu J, et al. Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements. Nat Commun. 2014;5:4361.
37. Nagarajan RP, Zhang B, Bell RJA, Johnson BE, Otshen AB, Sundaram V, Li D, Graham AE, Diaz A, Fouse SD, et al. Recurrent epimutations activate gene body promoters in primary glioblastoma. Genome Res. 2014;24(5):761–74.

38. Wiesner T, Lee W, Obenauf AC, Ren L, Murali R, Zhang QF, Wong EWP, Hu W, Scott SN, Shah RH, et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. Nature. 2015;526(7573):453–7.

39. O’Connell MR, Sarkar S, Luthra GK, Okugawa Y, Toiyama Y, Gajjar AH, Qiu S, Goel A, Singh P. Epigenetic changes and alternate promoter usage by human colon cancers for expressing DCLK1-Isomorphs: Clinical Implications. Sci Rep. 2015;5:14983.

40. Gratelli E, Pisanó F, Ciealdiella A, Bonomo S, Missaglia C, Cerutti MG, Masiero L, Ianzano L, Giordano F, Cisselli V, et al. A novel oncogenic BTC isoform is overexpressed in colon cancers and required for RAS-mediated transformation. Oncogene. 2016;35:4368–78.

41. Maeso I, Tena JJ. Favorable genomic environments for cis-regulatory evolution: A novel theoretical framework. Semin Cell Dev Biol. 2015;26:17–20.

42. Thompson Peter J, Macfarlan Todd S, Lorincz Matthew C. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Mol Cell. 2016;62(5):766–76.

43. Blomberg J, Benachenhou F, Bilskard V, Sperber G, Mayer J. Classification and nomenclature of endogenous retroviral sequences (ERVs): Problems and recommendations. Gene. 2009;448(2):115–23.

44. Mager DL, Stoeyp J. Mammalian Endogenous Retroviruses. Microbiol Spectr. 2015, 3(1). doi: 10.1128/microbiolspec.AMDNA3-0009-2014.

45. Belshaw R, Watson J, Katzourakis A, Howe A, Woolven-Allen J, Burt A, Tristem M. Rate of Recombinational Deletion among Human Endogenous Retroviruses. J Virol. 2005;79(19):12507–14.

46. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JP. Retroviruses. J Virol. 2007;81(17):9437–47.

47. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Prabhakar S, Kimura-ćN. Genome-wide characterization of human L1 antisense promoter-driven transcripts. BMC Genomics. 2016;17(1):26.

48. Belshaw R, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083–8.

49. Jimenez-Capdevila JM, Siles-Menorca MC, Ortego A, Miquel J. The majority of primate-specific ORF0 contributes to retrotransposon-mediated diversity. J Biomed Biotechnol. 2006;2006:71753.

50. Rebolledo R, Fairvar S, Mager DL. Transposable Elements Transformed the Uterine Regulatory Landscape and preimplantation development and nuclear reprogramming. Nat Genet. 2016;48:44–52.

51. Izvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERV-K: Conflict or serendipity? BioEssays. 2016;38(1):109–17.

52. Emers D, Wagner GP. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. Brief Funct Genomics. 2012;11(4):267–76.

53. Feschtotte C. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 2008;9(5):397–405.

54. Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Genome. 2009;448(2):105–14.

55. Rayan NA, del Rosario RCH, Prabhakar S. Massive contribution of transposable elements to mammalian regulatory sequences. Semin Cell Dev Biol. 2015;57:51–6.
132. Lee W, Belkhiria A, Lockhart AC, Merchant N, Glaser H, Harris EL, Washington MK, Brunet EM, Zalika A, Kim RB, et al. Overexpression of OATP1B1 Confers Apoptotic Resistance in Colon Cancer. Cancer Res. 2008;68(24):10315–23.

133. Nagar M, Furfata T, Matsumoto S, Sini S, Motoshashi S, Yoshihito I, Uigajin M, Miyajima A, Matsumoto S, Chiba K. Identification of a new organic anion transporting polypeptide 1B3 mRNA isoform primarily expressed in human cancerous tissues and cells. Biochem Biophys Res Commun. 2012;418(4):818–23.

134. Imai S, Kikuchi R, Tsuruya Y, Nao S, Nishida S, Kusuhara S, Hugiyama Y. Epigenetic Regulation of Organic Anion Transporting Polypeptide 1B3 in Cancer Cell Lines. Pharm Res. 2013;30(11):2880–90.

135. Li C, Xu Z, Wu L, Zheng S. Characterization of novel human ovarian cancer-specific transcripts (HOSTs) in human embryonic stem cell self-renewal. Dev Cell. 2013;25(1):69–72.

136. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H. Linc-RoR promotes c-Myc expression through hnRNP I and AUF1. Nucleic Acids Res. 2016;44(15):5858–70.

137. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Curran M, Onder T, Agarwal S, et al. Large intergenic non-coding RNA-RoR decoys gene-specific histone methylation to associate with miR-145. Pathol Oncol Res. 2016;22(4):733–42.

138. Long non-coding RNA ROR decoys gene-specific histone methylation to predict non-coding RNA upregulated in lung cancer and predicts invasion and metastasis. Tumor Biol. 2016;37(1):239–37.

139. Yung F, Lyu S, Dong S, Liu Y, Zhang X, Wang O. Expression profile analysis of long non-coding RNA in HER-2-enriched subtype breast cancer by next-generation sequencing and bioinformatics. OncoTargets Ther. 2016;9:761–72.

140. Wang X-S, Zhang Z, Wang H-C, Cai J-L, Xu Q-W, Li M-Q, Chen Y-C, Qian X-P, Wu H-Z, Gong P-H, Lu S. Epigenetic Regulation of Organic Anion Transporting Polypeptide 1B3 mRNA isoform primarily expressed in human cancerous tissues and cells. Biochem Biophys Res Commun. 2012;418(4):818–23.

141. Li C, Xu Z, Wu L, Zheng S. Expression Patterns of Non-Coding Spliced Transcripts from Human Endogenous Retrovirus HER-H Elements in Colon Cancer. PLoS One. 2012;7(1):e29552.

142. Teft WA, Welch S, Lenehan J, Parfitt J, Chai YH, Winquist E, Kim RB. OATP1B1 and tumour OATP1B3 modulate exposure, toxicity, and survival after irinotecan-based chemotheraphy. Br J Cancer. 2015;112(5):857–65.

143. Moor RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Curran M, Onder T, Agarwal S, et al. Large intergenic non-coding RNA-RoR decoys gene-specific histone methylation to associate with miR-145. Pathol Oncol Res. 2016;22(4):733–42.

144. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H. Linc-RoR promotes c-Myc expression through hnRNP I and AUF1. Nucleic Acids Res. 2016;44(15):5858–70.

145. Loewer S, Cabili MN, Guttman M, Loh Y-H, Thomas K, Park IH, Garber M, Rangel LBA, Sherman-Baust CA, Wernyj RP, Schwartz DR, Cho KR, Morin PJ. Genet. 2010;42(12):1113–26.

146. Imai S, Kikuchi R, Tsuruya Y, Nao S, Nishida S, Kusuhara S, Hugiyama Y. Epigenetic Regulation of Organic Anion Transporting Polypeptide 1B3 in Cancer Cell Lines. Pharm Res. 2013;30(11):2880–90.

147. Li C, Xu Z, Wu L, Zheng S. Characterization of novel human ovarian cancer-specific transcripts (HOSTs) in human embryonic stem cell self-renewal. Dev Cell. 2013;25(1):69–72.

148. Eades G, Wolfson B, Zhang Y, Li Q, Yao Y, Zhou Q, Imai S, Kikuchi R, Tsuruya Y, Nao S, Nishida S, Kusuhara S, Hugiyama Y. Epigenetic Regulation of Organic Anion Transporting Polypeptide 1B3 in Cancer Cell Lines. Pharm Res. 2013;30(11):2880–90.

149. Li C, Xu Z, Wu L, Zheng S. Characterization of novel human ovarian cancer-specific transcripts (HOSTs) in human embryonic stem cell self-renewal. Dev Cell. 2013;25(1):69–72.

150. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell. 2013;25(1):69–72.

151. Eades G, Wolfson B, Zhang Y, Li Q, Yao Y, Zhou Q, lincRNA-RoR and miR-145-3p Regulate Invasion in Triple-Negative Breast Cancer via Targeting ARF6. Mol Cancer Res. 2015;13(2):330–8.

152. Gao S, Wang P, Hua Y, Xi H, Meng Z, Liu T, Chen Z, Liu L. ROR functions as a ceRNA to regulate Nanog expression by sponging miR-145 and predicts poor prognosis in pancreatic cancer. Oncotarget. 2016;7(24):16088–16097.

153. Huang J, Zhang A, Ho T-T, Zhang Z, Zhu N, Ding X, Zhang X, Xu M, Mo Y-Y. linc-RoR promotes c-Myc expression through interaction with hnRNP I and AUF1. Nucleic Acids Res. 2016;44(7):3059–69.

154. Rangel LBA, Sherman-Baust CA, Wernyj RP, Schwartz DR, Cho KR, Morin PJ. Characterization of novel human ovarian cancer-specific transcripts (HOSTs) identified by serial analysis of gene expression. Oncogene. 2003;22(46):7225–32.
175. Gibb E, Warren R, Wilson G, Brown S, Robertson G, Morin G, Holt R. Activation of an endogenous retrovirus-associated long non-coding RNA in human adenocarcinoma. Genome Med. 2015;7(1):22.

176. Tomczak K, Czerwi

177. Hashimoto K, Suzuki AM, Dos Santos A, Desterke C, Collino A, Ghisletti S, Braun E, Bonetti A, Fort A, Qin XY, et al. CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors. Genome Res. 2015;25(12):1812–24.

178. Lania L, Di Cristofano A, Stazzullo M, Pengue G, Majello B, La Mantia G. Structural and functional organization of the human endogenous retroviral ERV sequences. Virology. 1992;191(1):94–6.

179. Santoni F, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology. 2012;9(1):111.

180. Ohnuki M, Tanabe K, Sutoh K, Teramoto I, Sawamura Y, Naito M, Nakamura M, Tokunaga Y, Nakamura M, Watanabe A, et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proc Natl Acad Sci. 2014;111(19):4216–31.

181. Fuchs N, Loever S, Daley G, Izsvak Z, Lower J, Lower R. Human endogenous retrovirus K (HERV-K) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology. 2013;10(1):115.

182. Grow EF, Flynn RA, Chavez SL, Baleses NL, Wossidlo M, Wescott DJ, Martin L, Ware CB, Blish CA, Chang HY, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522(7553):211–5.

183. Göke J, Lu X, Chan YS, Ng H-H, Ly L-H, Sachs F, Szazerbinska I. Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. Cell Stem Cell. 2015; 16(2):135–43.

184. Haeusler K, Ruther UM, Dobbelstein M. Endogenous retroviruses in cancer. J Cell Biochem. 2005;94(2):245–53.

185. Li W, Lee M-H, Henderson L, Tyagi R, Bachani M, Steiner J, Campanac E, Rades MJ, Calvert V, Juhl H, Mills F, Mason K, et al. A Novel Function of RNAs Arising from the 8p12 retrovirus-K contributes to motor neuron disease. Proc Natl Acad Sci U S A. 2011;108(9):3624–9.

186. Costas J, Naveira H. Evolutionary History of the Human Endogenous Retrovirus Family ERV9. Mol Biol Evol. 2000;17(2):320–7.

187. Di Cristofano A, Stazzullo M, Longo L, La Mantia G. Characterization and genomic mapping of the ZNF80 locus. Expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. Nucleic Acids Res. 1995;23:2823.

188. Chen H-J, Carr K, Jerome RE, Edengberg HJ. A Retroviral Repetitive Element Confers Tissue-Specificity to the Human Alcohol Dehydrogenase 1C (ADH1C) Gene. DNA Cell Biol. 2002;21(11):793–801.

189. Beyer U, Moll-Rocek J, Moll UM, Dobbelstein M. Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. Proc Natl Acad Sci U S A. 2011;108(9):3624–9.

190. Li W, Zhu X, Wu M, Wang Y, Fulzele S, Eroluo A, Liu J, Tuan D. Long-range function of an intergenic retrotransposon. Proc Natl Acad Sci U S A. 2010;107(29):12992–7.

191. Sokol M, Jensen KM, Pedersen FS. Human endogenous retroviruses sustain complex and cooperative regulation of gene-containing loci and unannotated megabase-sized regions. Retrovirology. 2015;12:32.

192. Beyer U, Kronung SK, Leh a, Walter L, Dobbelstein M. Comprehensive identification of genes driven by ERV9-LTRs reveals TNRSSF10A as a re-activatable mediator of testicular cancer cell death. Cell Death Differ. 2016;23:674–86.

193. Kloning SK, Beyer U, Chiaramonte ML, Dolfini D, Mantovani R, Dobbelstein M, LTR12 promoters activation in a broad range of human tumor cells by HDAC inhibition. Oncotarget. 2016;7(23):33484–497.

194. Xu L, Elkahouli AG, Candotti F, Grakowski A, Beaucage SL, Petricoin EF, Calvert V, Juhl H, Mills F, Mason K, et al. A Novel Function of RNAs Arising From the Long Terminal Repeat of Human Endogenous Retrovirus 9 in Cell Cycle Arrest. J Virol. 2013;87(1):25–36.

195. Xu L, Elkahouli AG, Candotti F, Grakowski A, Beaucage SL, Petricoin EF, Calvert V, Juhl H, Mills F, Mason K, et al. A Novel Function of RNAs Arising From the Long Terminal Repeat (LTR) of ERV9 Human Endogenous Retrovirus Ends to NF-Y in the Assembly of an Active LTR Enhancer Complex NF-Y/MZF1/GATA-2. J Biol Chem. 2005;280(42):35184–94.

196. Gasparini P, Sozzi G, Pierotti MA. The role of chromosomal alterations in human cancer development. J Cell Biochem. 2007;102(2):320–31.
222. Vorechovsky I. Transposable elements in disease-associated cryptic exons. Hum Genet. 2010;127(2):135–54.

223. Darby MM, Leek JT, Langmead B, Yolken RH, Sabunciyan S. Widespread splicing of repetitive element loci into coding regions of gene transcripts. Hum Mol Gen 2016, in press.

224. Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. Trends Genet. 2015;31(5):274–80.

225. Laurette P, Strub T, Koludrovic D, Keime C, Le Gras S, Seberg H, Van Otterloo E, Imrichova H, Siddaway R, Aerts S, et al. Transcription factor MITF and remodeler BRG1 define chromatin organisation at regulatory elements in melanoma cells. eLife. 2015;4:e06857.