Mobile DNA

CrossMark

Endogenous retroviral promoter exaptation in human cancer

Artem Babaian[1,2] and Dixie L. Mager[1,2]*

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

* Correspondence: dmager@bccrc.ca
[1]Terry Fox Laboratory, British Columbia Cancer Agency, 675 West 10th Avenue, Vancouver, BC V5Z1L3, Canada
[2]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

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 accessary 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[a]	~Copies of solitary LTRs[b]
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 (ERVK)	HERV-K (HML-2)	LTR5, 5A, 5B, 5Hs	80	1200
III (ERVL)	HERV16	LTR16A-16E	860	18100
III (MaLR)	THE	THE1A-1D	7900	21260
	MLT1	MLT1A-1O	3820	146,550

[a]Copy numbers of internal regions estimated from Dfam (dfam.org) [50]
[b]Solitary LTR numbers estimated from Dfam coverage minus 2x internal region numbers, assuming all internal regions are associated with two LTRs

[6, 10, 42, 62–65]. Suffice it to say that, retrotranspositionally incompetent ERV LTRs, long considered the "poor cousin" of active L1 elements, have emerged from the shadowy realm of junk DNA and are now recognized as a major source of gene regulatory evolution through exaptation of their promoters and enhancers.

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

Table 2 Activation of oncogenes by Onco-exaptation of TE-derived promoters

Gene[a]	Gene function	Primary result of TE-driven expression	TE type	TE promoter coordinates (hg38)	Cancer type	references
CSF1R	Tyrosine kinase receptor	Ectopic expression of normal protein	(ERVL-MaLR) THE1B LTR	chr5:150092453–150092809	HL	[113]
IRF5	Transcription factor	Ectopic expression of normal protein	(ERV1) LOR1a LTR	chr7:128936859–128937097	HL	[117, 119]
MET	Tyrosine kinase receptor	Protein truncation	(L1) L1PA2	chr7:116718498–116724489	CML, others?	[124, 125]
ALK	Tyrosine kinase receptor	Protein truncation	(ERVL) LTR16B2	chr2:29223783–29224196	melanoma	[38]
ERRB4	Tyrosine kinase receptor	Protein truncation	(ERVL-MaLR) MLT1C LTR (ERVL-MaLR) MLT1H2 LTR	chr2:211693702–211694209 chr2:211465146–211465419	ALCL	[129]
SLCO1B3	Anion transporter	Chimeric protein	(ERV1) LTR7[b]	chr12:20822187–20822617	colon, others	[133, 136]
FABP7	Fatty acid binding	Chimeric protein	(ERV1) LTR2	chr6:122748805–122749262	DLBCL	[138]

[a]Only those cases with supporting evidence of a role in the cancer are listed
[b]The fact that the promoter for these isoforms is an LTR was not noted in the cited papers

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 (*CSF1R*) in Hodgkin Lymphoma (HL). Normally, *CSF1R* 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 *CSF1R* expression instead being driven by a solitary THE1B LTR, of the MaLR-ERVL 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 ~17,000 [50, 114] (Table 1). The *THE1B-CSF1R* 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 *CSF1R* expression in HL appears to be completely dependent on the THE1B LTR, and CSF1R 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 *CSF1R* 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
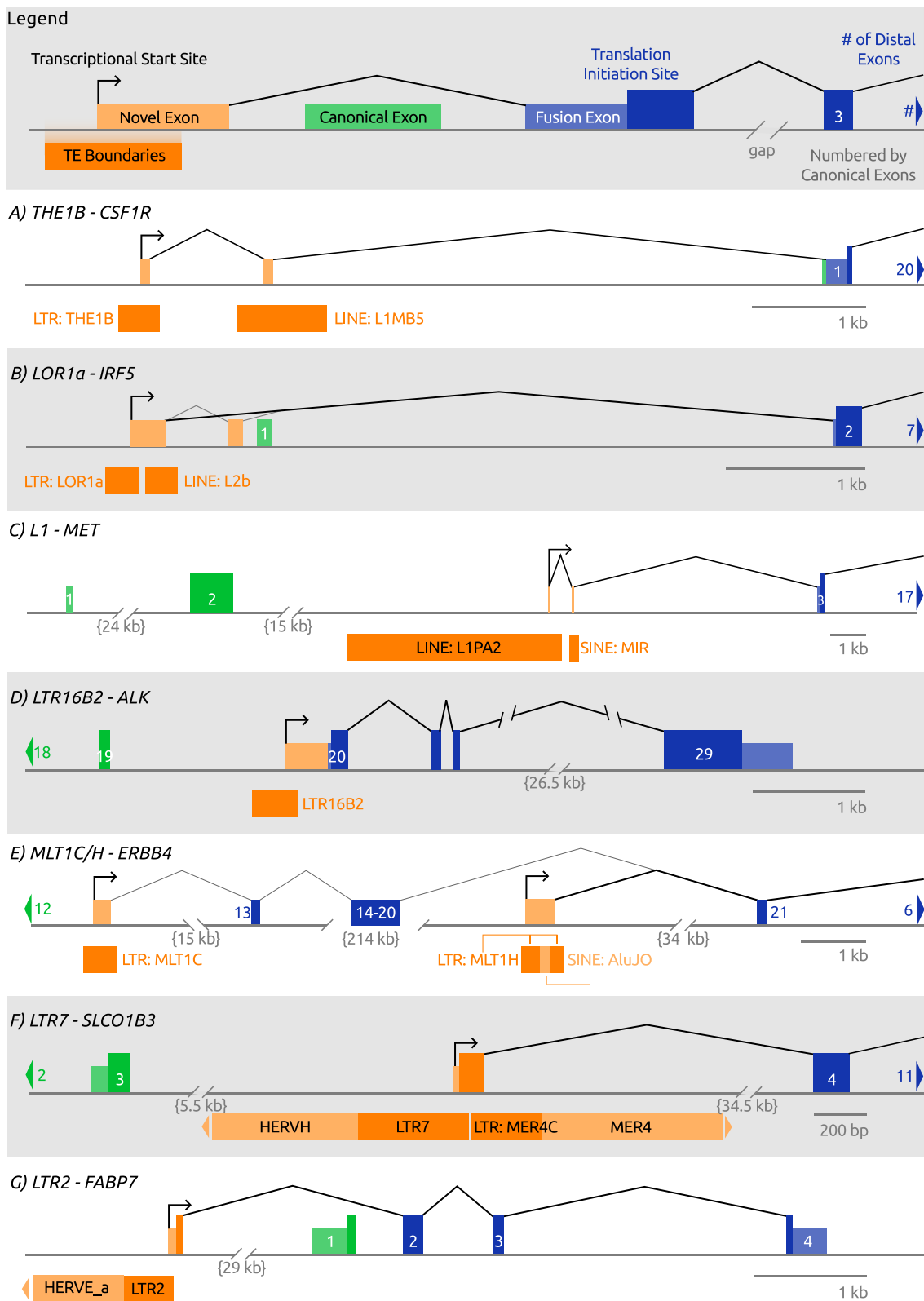
Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Examples of Onco-exaptation. Gene models of known TE-derived promoters expressing downstream oncogenes and listed in Table 2. Legend is shown at the top. **a** 6 kb upstream of *CSF1R*, a THE1B 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. **b** An LOR1a LTR initiates transcription and splices into the canonical second exon of *IRF5* 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. **c** 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-promoterd isoform. **d** An LTR16B2 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*. **e** 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 MER4C 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 TIS and splice donor which creates a different N-terminal protein sequence of FABP7

v-myb, a truncated form of *myb* carried by acutely transforming animal retroviruses [121, 122].
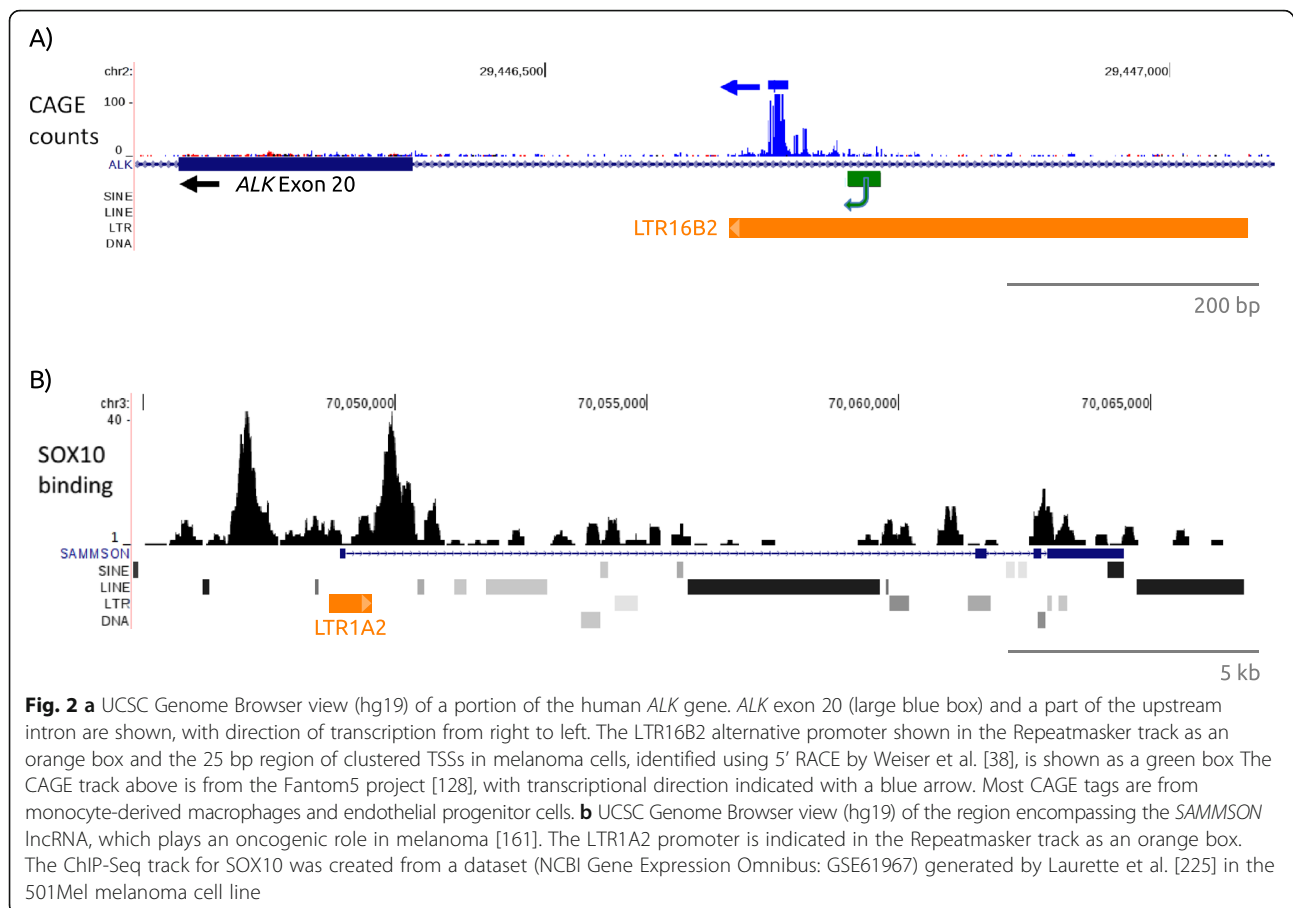
The first such reported case involving a TE was identified in a screen of human ESTs to detect transcripts driven by the antisense promoter within L1 elements. Mätlik et al. identified an L1PA2 within the second intron of the proto-oncogene *MET* (*MET* proto-oncogene, receptor tyrosine kinase) that initiates a transcript by splicing into downstream *MET* exons (Fig. 1c) [67]. Not surprisingly, transcriptional activity of the CpG rich promoter of this L1 in bladder and colon cancer cell lines is inversely correlated to its degree of methylation [123, 124]. A slightly truncated MET protein is produced by the TE-initiated transcript and one study reported that L1-driven transcription of MET reduces overall MET protein levels and signaling, although by what mechanism is not clear [124]. Analyses of normal colon tissues and matched primary colon cancers and liver metastasis samples showed this L1 is progressively demethylated in the metastasis samples, which strongly correlates with increased L1-MET transcripts and protein levels [125]. Since MET levels are a negative prognostic indicator for colon cancer [126], these findings suggest an oncogenic role for L1-MET.

More recently, Wiesner et al. identified a novel isoform of the receptor tyrosine kinase (RTK), anaplastic lymphoma kinase (*ALK*), initiating from an alternative promoter in its 19th intron [38]. This alternative transcription initiation (ATI) isoform or *ALK*[ATI] was reported to be specific to cancer samples and found in ~11% of skin cutaneous melanomas. *ALK*[ATI] transcripts produce three protein isoforms encoded by exons 20 to 29. These smaller isoforms exclude the extracellular domain of the protein but contain the catalytic intracellular tyrosine kinase domain. This same region of *ALK* is commonly found fused with a range of other genes via chromosomal translocations in lymphomas and a variety of solid tumors [127]. In the Wiesner et al. study it was

found that ALK[ATI] stimulates several oncogenic signaling pathways, drives cell proliferation in vitro, and promotes tumor formation in mice [38].

The *ALK*[ATI] 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 ERRB4 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

Fig. 2 a UCSC Genome Browser view (hg19) of a portion of the human *ALK* gene. *ALK* exon 20 (large blue box) and a part of the upstream intron are shown, with direction of transcription from right to left. The LTR16B2 alternative promoter shown in the Repeatmasker track as an orange box and the 25 bp region of clustered TSSs in melanoma cells, identified using 5' RACE by Weiser et al. [38], is shown as a green box The CAGE track above is from the Fantom5 project [128], with transcriptional direction indicated with a blue arrow. Most CAGE tags are from monocyte-derived macrophages and endothelial progenitor cells. **b** UCSC Genome Browser view (hg19) of the region encompassing the *SAMMSON* lncRNA, which plays an oncogenic role in melanoma [161]. The LTR1A2 promoter is indicated in the Repeatmasker track as an orange box. The ChIP-Seq track for SOX10 was created from a dataset (NCBI Gene Expression Omnibus: GSE61967) generated by Laurette et al. [225] in the 501Mel melanoma cell line

(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 anion 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 to 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.

TE-initiated non-coding RNAs in cancer
Since TEs, particularly ERV LTRs, provide a major class of promoters for long non-coding RNAs [56, 141, 142], it is not surprising that multiple LTR-driven lncRNAs

have been shown to be involved in cancer. 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. Unlike the coding genes discussed above which have non-TE or native promoters in normal tissues, the lncRNAs described here typically have LTRs as their only promoter in normal or malignant cells.

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].
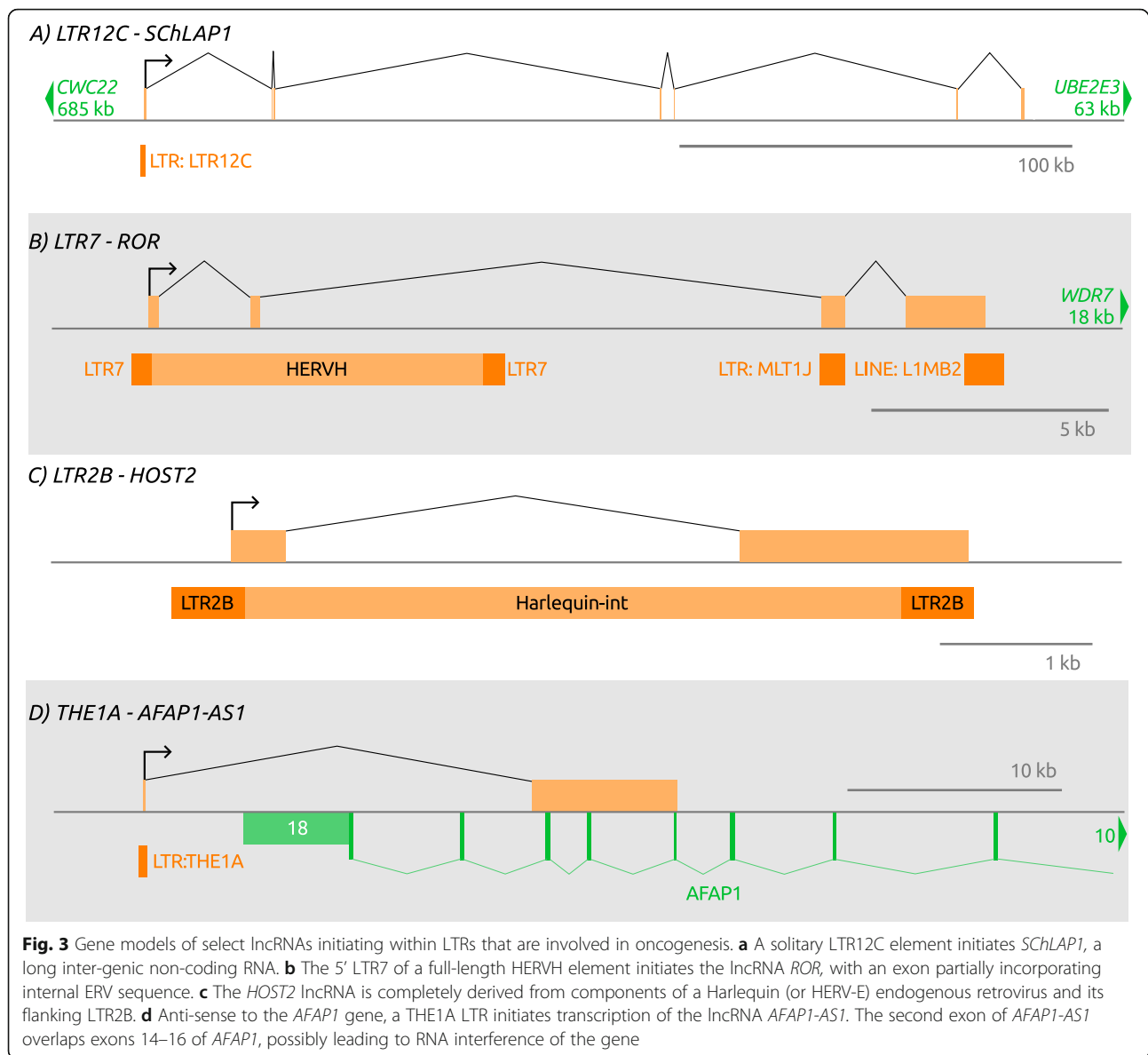
Using Serial Analysis of Gene Expression (SAGE), Rangel et al. identified five Human Ovarian cancer Specific Transcripts (HOSTs) that were expressed in ovarian cancer but not in other normal cells or cancer types

Table 3 LTR-driven LncRNAs with oncogenic role

lncRNA[a]	TE type	TE promoter coordinates (hg38)	Cancer type	references
SchLAP1	(ERV1) LTR12C	chr2:180691205–180692425	prostate	[143]
ROR	(ERV1) LTR7	chr18:57072052–57072502	breast, others	[147, 150]
HOST2	(ERV1) LTR2B	chr10:84171987–84172465	ovarian	[154]
AFAP1-AS1	(ERVL-MaLR) THE1A LTR[b]	chr4:7753884–7754236	several	[156, 158]
SAMMSON	(ERV1) LTR1A2[b]	chr3:69999501–70000359	melanoma	[161]
HULC	(ERVL-MaLR) MLT1A LTR	chr6:8652095–8652454	liver	[163]
UCA1	(ERV1) LTR7C	chr19:15828738–15829200	several	[165, 167]
BANCR	(ERV1) MER41B LTR	chr9:69306939–69307567	melanoma, others	[169]

[a]Only those cases with supporting evidence of a role in the cancer are listed
[b]The fact that the promoter is an LTR was not previously noted

Fig. 3 Gene models of select lncRNAs initiating within LTRs that are involved in oncogenesis. **a** A solitary LTR12C element initiates *SChLAP1*, a long inter-genic non-coding RNA. **b** The 5′ LTR7 of a full-length HERVH element initiates the lncRNA *ROR,* with an exon partially incorporating internal ERV sequence. **c** The *HOST2* lncRNA is completely derived from components of a Harlequin (or HERV-E) endogenous retrovirus and its flanking LTR2B. **d** Anti-sense to the *AFAP1* gene, a THE1A LTR initiates transcription of the lncRNA *AFAP1-AS1*. The second exon of *AFAP1-AS1* overlaps exons 14–16 of *AFAP1*, possibly leading to RNA interference of the gene

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* lncRNA (survival associated mito-chondrial 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 lncRNA 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 lncRNAs include *HULC* for Highly Upregulated in Liver Cancer [163, 164], *UCA1* (urothelial cancer associated 1) [165–168] and *BANCR* (BRAF-regulated lncRNA 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 lncRNAs 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 lncRNAs 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, Gosenca 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 lncRNAs 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 lncRNA 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 HERVK3 element on chromosome 19 and this event creates a chimeric ORF with HERVK3 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 HERVK3 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.
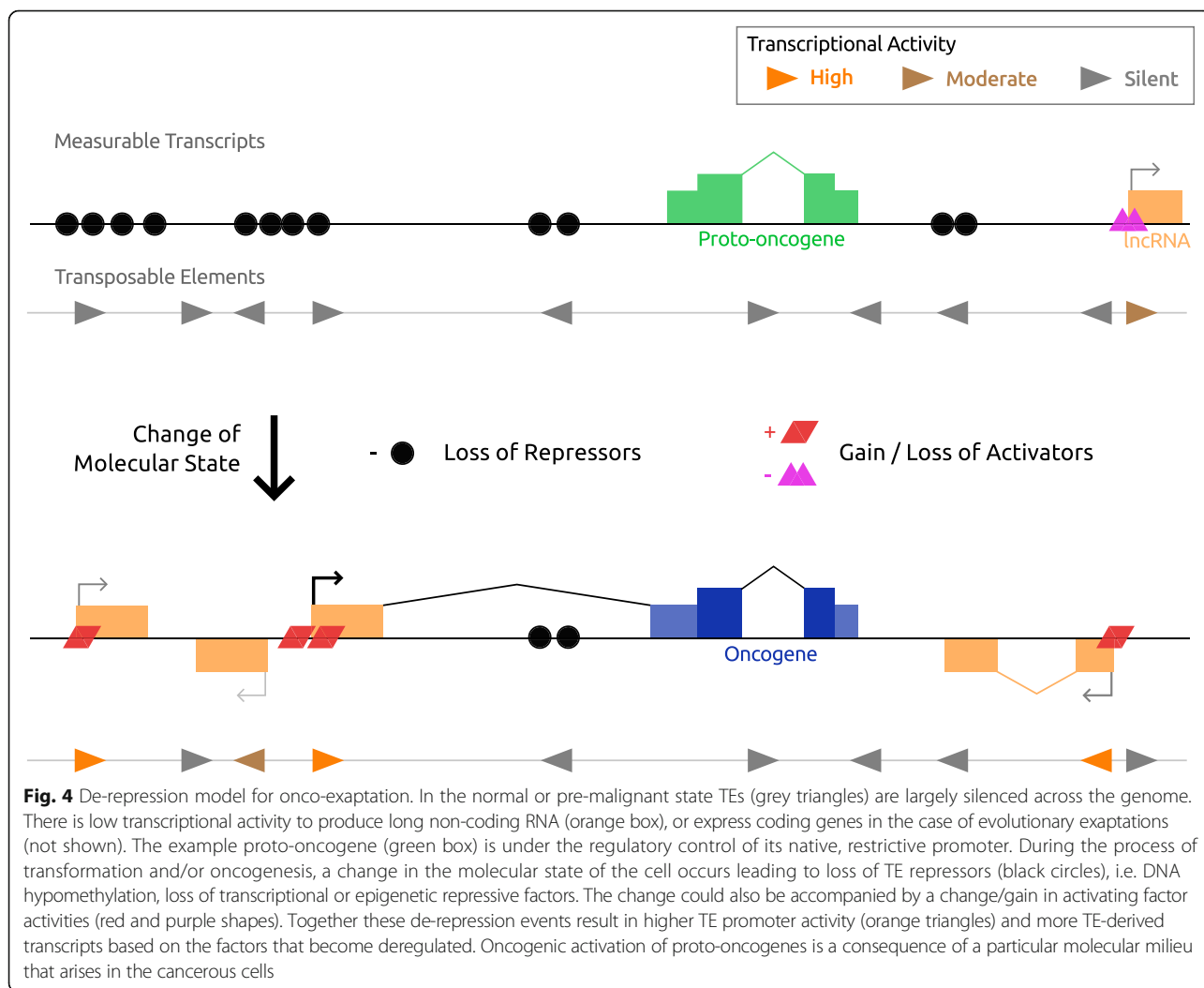
Models for onco-exaptation

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

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* onco-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 genome-wide 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
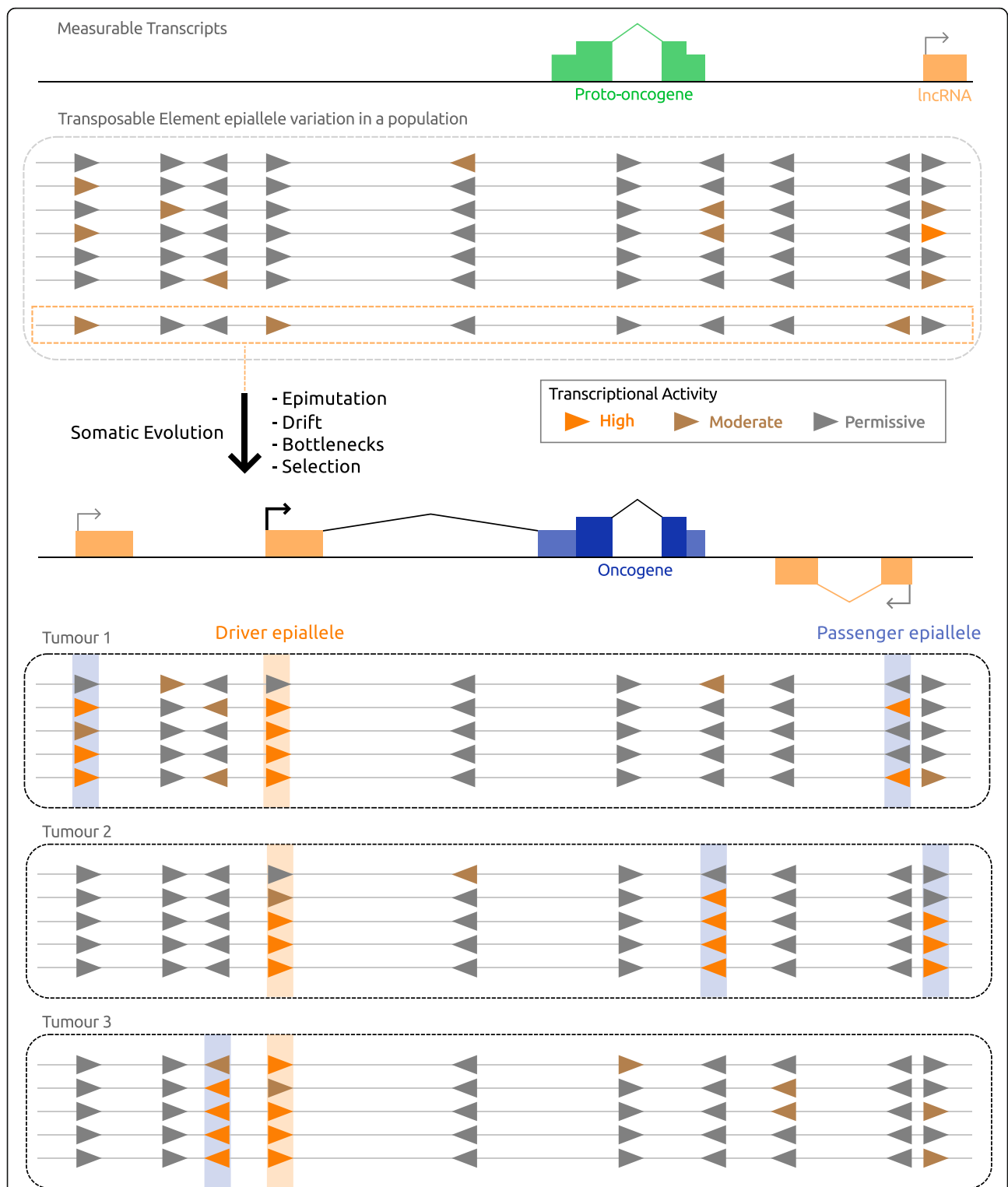
Fig. 5 Epigenetic evolution model for onco-exaptation. In the starting cell population there is a dispersed and low/noisy promoter activity at TEs (colored triangles) from a set of transcriptionally permissive TEs (grey triangles). TE-derived transcript expression is low and variable between cells. Some 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-kB 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 cryptic 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.

SAMMSON: 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; TFPI-2: Tissue factor pathway inhibitor 2; TIS: Translation initiation site; TSS: Transcriptional start site; UCA1: Urothelial cancer associated 1

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.

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):241–59.
3. Brosius J, Gould SJ. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc Natl Acad Sci. 1992;89(22):10706–10.
4. Gould SJ, Vrba ES. Exaptation-A Missing Term in the Science of Form. Paleobiology. 1982;8(1):4–15.
5. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 2012;13(4):283–96.
6. Rebollo R, Romanish 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, Kopera HC, Moldovan JB, Garcia-Pérez JL, Moran JV: The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr 2015, 3(2):10.1128/microbiolspec.MDNA1123-0061-2014
8. Robbez-Masson L, Rowe H. Retrotransposons shape species-specific embryonic stem cell gene expression. Retrovirology. 2015;12(1):45.
9. Hancks DC, Kazazian 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? APMIS. 2016; 124(1–2):4–10.
12. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. 2016;351(6274):aac7247.
13. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian Jr HH. 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, Ferec C. 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. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res. 1992;52(3):643–5.
17. Lee E, Iskow R, Yang L, Gokcumen O, Haseley 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. Solyom S, Ewing AD, Rahrmann 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, Baillie JK, Collino A, Ghisletti S, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell. 2013;153(1):101–11.
20. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas 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, Vertino PM, Devine SE. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res. 2016;26(6):745–55.
24. Bannert 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 hypomethylation 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(4):398–422.
30. Hohn O, Hanke K, Bannert N. HERV-K(HML-2), the best preserved family of HERVs: endogenisation, expression and implications in health and disease. Front Oncol. 2013;3:246.
31. Chen T, Meng Z, Gan Y, Wang X, Xu F, Gu Y, Xu X, Tang J, Zhou H, Zhang X, et al. The viral oncogene Np9 acts as a critical molecular switch for co-activating beta-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. Leukemia. 2013;27(7):1469–78.
32. Downey RF, Sullivan FJ, Wang-Johanning F, Ambs S, Giles FJ, Glynn SA. Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice? Int J Cancer. 2015;137(6):1249–57.
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, Vooijs M, Radtke F, Redondo JM, et al. Alternative promoter usage at the Notch1 locus supports ligand-independent signaling in T cell development and leukemogenesis. Immunity. 2010;33(5):685–98.
35. Thorsen K, Schepeler T, Øster B, Rasmussen MH, Vang S, Wang K, Hansen KQ, Lamy P, Pedersen JS, Eller A, et al. Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis. BMC Genomics. 2011;12(1):1–14.
36. Muratani M, Deng N, Ooi WF, Lin SJ, Xing M, Xu C, Qamra 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, Olshen 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, Ran 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-isoforms: Clinical Implications. Scie Rep. 2015;5:14983.

40. Grassilli E, Pisano F, Cialdella A, Bonomo S, Missaglia C, Cerrito MG, Masiero L, Ianzano L, Giordano F, Cicirelli V, et al. A novel oncogenic BTK 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. 2016;57:2–10.

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, Blikstad 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, Stoye JP: Mammalian Endogenous Retroviruses. Microbiol Spectr 2015, 3(1). doi: 10.1128/microbiolspec.MDNA3-0009-2014.

45. Belshaw R, Dawson ALA, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K(HML2): Implications for Present-Day Activity. J Virol. 2005;79(19):12507–14.

46. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc Natl Acad Sci. 2016;113(16):E2326–34.

47. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. PLoS Genet. 2006;2(1):e2.

48. 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. 2007;81(17):9437–42.

49. Gemmell P, Hein J, Katzourakis A. Phylogenetic Analysis Reveals That ERVs "Die Young" but HERV-H Is Unusually Conserved. PLoS Comput Biol. 2016; 12(6):e1004964.

50. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AFA, Wheeler TJ. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016;44(D1):D81–9.

51. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010;42(7):631–4.

52. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083–7.

53. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci. 2007;104(47):18613–8.

54. Jacques P-E, Jeyakani J, Bourque G. The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. PLoS Genet. 2013;9(5):e1003504.

55. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat Genet. 2013; 45(7):836–41.

56. Kelley D, Rinn J. Transposable elements reveal a stem cell specific class of long noncoding RNAs. Genome Biol. 2012;13(11):R107.

57. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat Struct Mol Biol. 2014;21(4):423–5.

58. Wang J, Xie G, Singh M, Ghanbarian AT, Raško T, Szvetnik A, Cai H, Besser D, Prigione A, Fuchs NV, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014;516(7531):405–9.

59. Durruthy-Durruthy J, Sebastiano V, Wossidlo M, Cepeda D, Cui J, Grow EJ, Davila J, Mall M, Wong WH, Wysocka J, et al. The primate-specific noncoding RNA HPAT5 regulates pluripotency during human

preimplantation development and nuclear reprogramming. Nat Genet. 2016;48:44–52.

60. Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? BioEssays. 2016;38(1): 109–17.

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

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

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

64. Rayan NA, del Rosario RCH, Prabhakar S. Massive contribution of transposable elements to mammalian regulatory sequences. Semin Cell Dev Biol. 2016;57:51–6.

65. Friedli M, Trono D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. Annu Rev Cell Dev Biol. 2015;31(1):429–51.

66. Nigumann P, Redik K, Matlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics. 2002;79(5):628–34.

67. Mätlik K, Redik K, Speek M. L1 Antisense Promoter Drives Tissue-Specific Transcription of Human Genes. J Biomed Biotechnol. 2006;2006:71753.

68. Rebollo R, Farivar S, Mager DL. C-GATE - catalogue of genes affected by transposable elements. Mob DNA. 2012;3(1):9.

69. Criscione SW, Theodosakis N, Micevic G, Cornish TC, Burns KH, Neretti N, Rodić N. Genome-wide characterization of human L1 antisense promoter-driven transcripts. BMC Genomics. 2016;17(1):1–15.

70. Denli Ahmet M, Narvaiza I, Kerman Bilal E, Pena M, Benner C, Marchetto Maria CN, Diedrich Jolene K, Aslanian A, Ma J, Moresco James J, et al. Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. Cell. 2015;163(3):583–93.

71. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. Molecular archeology of L1 insertions in the human genome. Genome Biol. 2002;3(10):research0052. 0051–research0052.0018.

72. Khan H, Smit A, Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 2006;16(1):78–87.

73. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009;41(5):563–71.

74. Roy AM, West NC, Rao A, Adhikari P, Alemán C, Barnes AP, Deininger PL. Upstream flanking sequences and transcription of SINEs1. J Mol Biol. 2000; 302(1):17–25.

75. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011;12(12):1–12.

76. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, James Kent W, Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature. 2006;441(7089):87–90.

77. Lowe CB, Bejerano G, Haussler D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc Natl Acad Sci. 2007;104(19):8005–10.

78. 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. 2008; 105(11):4220–1.

79. Jjingo D, Conley AB, Wang J, Mariño-Ramírez L, Lunyak VV, Jordan IK: Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. Mobile DNA 2014, 5:14–14

80. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet. 2011;43(11):1154–9.

81. Lynch Vincent J, Nnamani Mauris C, Kapusta A, Brayer K, Plaza Silvia L, Mazur Erik C, Emera D, Sheikh Shehzad 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(4):551–61.

82. Sinzelle L, Izsvak Z, Ivics Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. Cell Mol Life Sci. 2009;66(6):1073–93.

83. Bowen NJ, Jordan IK. Transposable elements and the evolution of eukaryotic complexity. Curr Issues Mol Biol. 2002;4(3):65–76.

84. 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(9):663–71.

85. Wolf D, Goff SP. Host Restriction Factors Blocking Retroviral Replication. Annu Rev Genet. 2008;42(1):143–63.

86. Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Diaz N, Rowe HM, Ecco G, Unzu C, Planet E, Lombardo A, et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome Res. 2014;24(8):1251–9.

87. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature. 2014;516(7530):242–5.

88. Wolf G, Greenberg D, Macfarlan T. Spotting the enemy within: Targeted silencing of foreign DNA in mammalian genomes by the Kruppel-associated box zinc finger protein family. Mob DNA. 2015;6(1):17.

89. Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virology. 2011;411(2):273–87.

90. Leung DC, Lorincz MC. Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem Sci. 2012;37(4):127–33.

91. Liu S, Brind'Amour J, Karimi MM, Shirane K, Bogutz A, Lefebvre L, Sasaki H, Shinkai Y, Lorincz MC. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. 2014;28(18):2041–55.

92. Yang F, Wang PJ: Multiple LINEs of retrotransposon silencing mechanisms in the mammalian germline. Semin Cell Dev Biol 2016, in press.

93. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nat Rev Cancer. 2011;11(10):726–34.

94. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nat Rev Cancer. 2013;13(7):497–510.

95. Berdasco M, Esteller M. Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry. Dev Cell. 2010;19(5):698–711.

96. Skulte KA, Phan L, Clark SJ, Taberlay PC. Chromatin remodeler mutations in human cancers: epigenetic implications. Epigenomics. 2014;6(4):397–414.

97. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012;482(7384):226–31.

98. Shen H, Laird PW. Interplay between the Cancer Genome and Epigenome. Cell. 2013;153(1):38–55.

99. Nephew KP, Huang TH-M. Epigenetic gene silencing in cancer initiation and progression. Cancer Lett. 2003;190(2):125–33.

100. Kazanets A, Shorstova T, Hilmi K, Marques M, Witcher M. Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2016;1865(2):275–88.

101. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002;21(35):5400–13.

102. Hoffmann MJ, Schulz WA. Causes and consequences of DNA hypomethylation in human cancer. Biochem Cell Biol. 2005;83(3):296–321.

103. De Smet C, Loriot A. DNA hypomethylation in cancer: Epigenetic scars of a neoplastic journey. Epigenetics. 2010;5(3):206–13.

104. Ross JP, Rand KN, Molloy PL. Hypomethylation of repeated DNA sequences in cancer. Epigenomics. 2010;2(2):245–69.

105. Szpakowski S, Sun X, Lage JM, Dyer A, Rubinstein J, Kowalski D, Sasaki C, Costa J, Lizardi PM. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. Gene. 2009;448(2):151–67.

106. Barchitta M, Quattrocchi A, Maugeri A, Vinciguerra M, Agodi A. LINE-1 Hypomethylation in Blood and Tissue Samples as an Epigenetic Marker for Cancer Risk: A Systematic Review and Meta-Analysis. PLoS One. 2014;9(10):e109478.

107. Romanish MT, Cohen CJ, Mager DL. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. Semin Cancer Biol. 2010;20(4):246–53.

108. Piskareva O, Lackington W, Lemass D, Hendrick C, Doolan P, Barron N. The human L1 element: a potential biomarker in cancer prognosis, current status and future directions. Curr Mol Med. 2011;11(4):286–303.

109. Pérot P, Mullins CS, Naville M, Bressan C, Hühns M, Gock M, Kühn F, Volff J-N, Trillet-Lenoir V, Linnebacher M, et al. Expression of young HERV-H loci in the course of colorectal carcinoma and correlation with molecular subtypes. Oncotarget. 2015;6(37):40095–111.

110. Haupt S, Tisdale M, Vincendeau M, Clements MA, Gauthier DT, Lance R, Semmes OJ, Turqueti-Neves A, Noessner E, Leib-Mösch C, et al. Human endogenous retrovirus transcription profiles of the kidney and kidney-derived cell lines. J Gen Virol. 2011;92(10):2356–66.

111. Gosenca D, Gabriel U, Steidler A, Mayer J, Diem O, Erben P, Fabarius A, Leib-Mˆsch C, Hofmann W-K, Seifarth W. HERV-E-Mediated Modulation of PLA2G4A Transcription in Urothelial Carcinoma. PLoS One. 2012;7(11):e49341.

112. Haase K, Mosch A, Frishman D. Differential expression analysis of human endogenous retroviruses based on ENCODE RNA-seq data. BMC Med Genet. 2015;8(1):71.

113. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, Kochert K, Bouhlel MA, Richter J, Soler E, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat Med. 2010;16(5):571–9.

114. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6(1):1–6.

115. Steidl C, Diepstra A, Lee T, Chan FC, Farinha P, Tan K, Telenius A, Barclay L, Shah SP, Connors JM, et al. Gene expression profiling of microdissected Hodgkin Reed-Sternberg cells correlates with treatment outcome in classical Hodgkin lymphoma. Blood. 2012;120(17):3530–40.

116. Martín-Moreno AM, Roncador G, Maestre L, Mata E, Jiménez S, Martínez-Torrecuadrada JL, Reyes-García AI, Rubio C, Tomás JF, Estévez M, et al. CSF1R Protein Expression in Reactive Lymphoid Tissues and Lymphoma: Its Relevance in Classical Hodgkin Lymphoma. PLoS One. 2015;10(6):e0125203.

117. Babaian A, Romanish MT, Gagnier L, Kuo LY, Karimi MM, Steidl C, Mager DL. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. Oncogene. 2016;35(19):2542–6.

118. Clark DN, Read RD, Mayhew V, Petersen SC, Argueta LB, Stutz LA, Till RE, Bergsten SM, Robinson BS, Baumann DG, et al. Four Promoters of IRF5 Respond Distinctly to Stimuli and are Affected by Autoimmune-Risk Polymorphisms. Front Immunol. 2013;4:360.

119. Kreher S, Bouhlel MA, Cauchy P, Lamprecht B, Li S, Grau M, Hummel F, Köchert K, Anagnostopoulos I, Jöhrens K, et al. Mapping of transcription factor motifs in active chromatin identifies IRF5 as key regulator in classical Hodgkin lymphoma. Proc Natl Acad Sci. 2014;111(42):E4513–22.

120. Mancl ME, Hu G, Sangster-Guity N, Olshalsky SL, Hoops K, Fitzgerald-Bocarsly P, Pitha PM, Pinder K, Barnes BJ. Two Discrete Promoters Regulate the Alternatively Spliced Human Interferon Regulatory Factor-5 Isoforms: Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. J Biol Chem. 2005;280(22):21078–90.

121. Introna M, Luchetti M, Castellano M, Arsura M, Golay J. The myb oncogene family of transcription factors: potent regulators of hematopoietic cell proliferation and differentiation. Semin Cancer Biol. 1994;5(2):113–24.

122. Ramsay RG, Gonda TJ. MYB function in normal and cancer cells. Nat Rev Cancer. 2008;8(7):523–34.

123. Wolff EM, Byun H-M, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer. PLoS Genet. 2010;6(4):e1000917.

124. Weber B, Kimhi S, Howard G, Eden A, Lyko F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene. 2010;29(43):5775–84.

125. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, Goel A. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut. 2014;63(4):635–46.

126. Gao H, Guan M, Sun Z, Bai C. High c-Met expression is a negative prognostic marker for colorectal cancer: a meta-analysis. Tumor Biol. 2015;36(2):515–20.

127. Mariño-Enríquez A, Dal Cin P. ALK as a paradigm of oncogenic promiscuity: different mechanisms of activation and different fusion partners drive tumors of different lineages. Cancer Genetics. 2013;206(11):357–73.

128. Fantom-Consortium. A promoter-level mammalian expression atlas. Nature. 2014;507(7493):462–70.

129. Scarfò I, Pellegrino E, Mereu E, Kwee I, Agnelli L, Bergaggio E, Garaffo G, Vitale N, Caputo M, Machiorlatti R, et al. Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts. Blood. 2016;127(2):221–32.

130. Arteaga Carlos L, Engelman Jeffrey A. ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. Cancer Cell. 2014;25(3):282–303.

131. Obaidat A, Roth M, Hagenbuch B. The Expression and Function of Organic Anion Transporting Polypeptides in Normal Tissues and in Cancer. Annu Rev Pharmacol Toxicol. 2012;52(1):135–51.

132. Lee W, Belkhiri A, Lockhart AC, Merchant N, Glaeser H, Harris EI, Washington MK, Brunt EM, Zaika A, Kim RB, et al. Overexpression of OATP1B3 Confers Apoptotic Resistance in Colon Cancer. Cancer Res. 2008;68(24):10315–23.

133. Nagai M, Furihata T, Matsumoto S, Ishii S, Motohashi S, Yoshino I, Ugajin 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, Naoi S, Nishida S, Kusuhara H, Sugiyama Y. Epigenetic Regulation of Organic Anion Transporting Polypeptide 1B3 in Cancer Cell Lines. Pharm Res. 2013;30(11):2880–90.

135. Liang Q, Xu Z, Xu R, Wu L, Zheng S. Expression Patterns of Non-Coding Spliced Transcripts from Human Endogenous Retrovirus HERV-H Elements in Colon Cancer. PLoS One. 2012;7(1):e29950.

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

137. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011; 476(7360):298–303.

138. Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, Babaian A, Sistiaga-Poveda M, Lai CB, Nemirovsky O, Serrano I, et al. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. Proc Natl Acad Sci. 2014;111(34):E3534–43.

139. Thumser AE, Moore JB, Plant NJ. Fatty acid binding proteins: tissue-specific functions in health and disease. Curr Opin Clin Nutr Metab Care. 2014;17(2):124–9.

140. Liu R-Z, Graham K, Glubrecht DD, Lai R, Mackey JR, Godbout R. A fatty acid-binding protein 7/RXRβ pathway enhances survival and proliferation in triple-negative breast cancer. J Pathol. 2012;228(3):310–21.

141. Kapusta A, Kronenberg Z, Lynch VJ, 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(4):e1003470.

142. St Laurent G, Shtokalo D, Dong B, Tackett M, Fan X, Lazorthes S, Nicolas E, Sang N, Triche T, McCaffrey T, et al. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. Genome Biol. 2013; 14(7):R73.

143. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, Vergara IA, Davicioni E, Erho N, Ghadessi M, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet. 2013;45(11):1392–8.

144. Masliah-Planchon J, Bièche I, Guinebretière J-M, Bourdeaut F, Delattre O. SWI/SNF Chromatin Remodeling and Human Malignancies. Ann Rev Pathol Mech Dis. 2015;10(1):145–71.

145. Loewer S, Cabili MN, Guttman M, Loh Y-H, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010;42(12):1113–7.

146. 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):69–80.

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

148. 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(2):1608–18.

149. Zhou P, Sun L, Liu D, Liu C, Sun L. Long Non-Coding RNA lincRNA-ROR Promotes the Progression of Colon Cancer and Holds Prognostic Value by Associating with miR-145. Pathol Oncol Res. 2016;22(4):733–40.

150. Fan J, Xing Y, Wen X, Jia R, Ni H, He J, Ding X, Pan H, Qian G, Ge S, et al. Long non-coding RNA ROR decoys gene-specific histone methylation to promote tumorigenesis. Genome Biol. 2015;16(1):139.

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

152. 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.

153. Adams Brian D, Kasinski Andrea L, Slack Frank J. Aberrant Regulation and Function of MicroRNAs in Cancer. Curr Biol. 2014;24(16):R762–76.

154. Gao Y, Meng H, Liu S, Hu J, Zhang Y, Jiao T, Liu Y, Ou J, Wang D, Yao L, et al. LncRNA-HOST2 regulates cell biological behaviors in epithelial ovarian cancer through a mechanism involving microRNA let-7b. Hum Mol Genet. 2015;24(3):841–52.

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

156. Zeng Z, Bo H, Gong Z, Lian Y, Li X, Li X, Zhang W, Deng H, Zhou M, Peng S, et al. AFAP1-AS1, a long noncoding RNA upregulated in lung cancer and promotes invasion and metastasis. Tumor Biol. 2016;37(1):729–37.

157. Deng J, Liang Y, Liu C, He S, Wang S. The up-regulation of long non-coding RNA AFAP1-AS1 is associated with the poor prognosis of NSCLC patients. Biomed Pharmacother. 2015;75:8–11.

158. Wu W, Bhagat TD, Yang X, Song JH, Cheng Y, Agarwal R, Abraham JM, Ibrahim S, Bartenstein M, Hussain Z, et al. Hypomethylation of Noncoding DNA Regions and Overexpression of the Long Noncoding RNA, AFAP1-AS1, in Barrett's Esophagus and Esophageal Adenocarcinoma. Gastroenterology. 2013;144(5):956–66. e954.

159. Zhang J-Y, Weng M-Z, Song F-B, Xu Y-G, Liu Q, Wu J-Y, Qin J, Jin T, Xu J-M. Long noncoding RNA AFAP1-AS1 indicates a poor prognosis of hepatocellular carcinoma and promotes cell proliferation and invasion via upregulation of the RhoA/Rac2 signaling. Int J Oncol. 2016;48:1590.

160. Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. Nat Struct Mol Biol. 2012;19(11):1068–75.

161. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, Radaelli E, Eyckerman S, Leonelli C, Vanderheyden K, et al. Melanoma addiction to the long non-coding RNA SAMMSON. Nature. 2016;531(7595):518–22.

162. Harris ML, Baxter LL, Loftus SK, Pavan WJ. Sox proteins in melanocyte development and melanoma. Pigment Cell Melanoma Res. 2010;23(4):496–513.

163. Panzitt K, Tschernatsch MMO, Guelly C, Moustafa T, Stradner M, Strohmaier HM, Buck CR, Denk H, Schroeder R, Trauner M, et al. Characterization of HULC, a Novel Gene With Striking Up-Regulation in Hepatocellular Carcinoma, as Noncoding RNA. Gastroenterology. 2007;132(1):330–42.

164. Li C, Chen J, Zhang K, Feng B, Wang R, Chen L. Progress and Prospects of Long Noncoding RNAs (lncRNAs) in Hepatocellular Carcinoma. Cell Physiol Biochem. 2015;36(2):423–34.

165. Wang X-S, Zhang Z, Wang H-C, Cai J-L, Xu Q-W, Li M-Q, Chen Y-C, Qian X-P, Lu T-J, Yu L-Z, et al. Rapid Identification of UCA1 as a Very Sensitive and Specific Unique Marker for Human Bladder Carcinoma. Clin Cancer Res. 2006;12(16):4851–8.

166. Wang F, Li X, Xie X, Zhao L, Chen W. UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. FEBS Lett. 2008;582(13):1919–27.

167. Xue M, Chen W, Li X. Urothelial cancer associated 1: a long noncoding RNA with a crucial role in cancer. J Cancer Res Clin Oncol. 2016;142(7):1407–19.

168. Hu J-J, Song W, Zhang S-D, Shen X-H, Qiu X-M, Wu H-Z, Gong P-H, Lu S, Zhao Z-J, He M-L, et al. HBx-upregulated lncRNA UCA1 promotes cell growth and tumorigenesis by recruiting EZH2 and repressing p27Kip1/CDK2 signaling. Sci Rep. 2016;6:23521.

169. Flockhart RJ, Webster DE, Qu K, Mascarenhas N, Kovalski J, Kretz M, Khavari PA. BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. Genome Res. 2012;22(6):1006–14.

170. Guo Q, Zhao YAN, Chen J, Hu JUN, Wang S, Zhang D, Sun Y. BRAF-activated long non-coding RNA contributes to colorectal cancer migration by inducing epithelial-mesenchymal transition. Oncol Lett. 2014;8(2):869–75.

171. Wang Y, Guo Q, Zhao YAN, Chen J, Wang S, Hu JUN, Sun Y. BRAF-activated long non-coding RNA contributes to cell proliferation and activates autophagy in papillary thyroid carcinoma. Oncol Lett. 2014;8(5):1947–52.

172. Cruickshanks HA, Vafadar-Isfahani N, Dunican DS, Lee A, Sproul D, Lund JN, Meehan RR, Tufarelli C. Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. Nucleic Acids Res. 2013;41(14):6857–69.

173. Cruickshanks HA, Tufarelli C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics. 2009;94(6):397–406.

174. Nigro CL, Wang H, McHugh A, Lattanzio L, Matin R, Harwood C, Syed N, Hatzimichael E, Briasoulis E, Merlano M, et al. Methylated Tissue Factor Pathway Inhibitor 2 (TFPI2) DNA in Serum Is a Biomarker of Metastatic Melanoma. J Invest Dermatol. 2013;133(5):1278–85.

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, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol. 2015;19(1A):A68–77.

177. Hashimoto K, Suzuki AM, Dos Santos A, Desterke C, Collino A, Ghisletti S, Braun E, Bonetti A, Fort A, Qin X-Y, 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, Strazzullo M, Pengue G, Majello B, La Mantia G. Structural and functional organization of the human endogenous retroviral ERV9 sequences. Virology. 1992;191(1):464–8.

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, Sutou K, Teramoto I, Sawamura Y, Narita 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(34):12426–31.

181. Fuchs N, Loewer S, Daley G, Izsvak Z, Lower J, Lower R. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology. 2013;10(1):115.

182. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522(7555):221–5.

183. Göke J, Lu X, Chan Y-S, Ng H-H, Ly L-H, Sachs F, Szczerbinska I. Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. Cell Stem Cell. 2015; 16(2):135–41.

184. Bannert N, Kurth R. Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci U S A. 2004;101 Suppl 2:14572–9.

185. Li W, Lee M-H, Henderson L, Tyagi R, Bachani M, Steiner J, Campanac E, Hoffman DA, von Geldern G, Johnson K, et al. Human endogenous retrovirus-K contributes to motor neuron disease. Sci Transl Med. 2015; 7(307):307ra153. 307ra153.

186. Costas J, Naveira H. Evolutionary History of the Human Endogenous Retrovirus Family ERV9. Mol Biol Evol. 2000;17(2):320–30.

187. Mager DL, Medstrand P. Retroviral Repeat Sequences. In: eLS. edn. Hoboken: Wiley; 2005.

188. Di Cristofano A, Strazzullo 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.

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

190. 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.

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

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

193. Beyer U, Kronung SK, Leha A, Walter L, Dobbelstein M. Comprehensive identification of genes driven by ERV9-LTRs reveals TNFRSF10B as a re-activatable mediator of testicular cancer cell death. Cell Death Differ. 2016; 23:64–75.

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

195. Xu L, Elkahloun AG, Candotti F, Grajkowski 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.

196. Yu X, Zhu X, Pi W, Ling J, Ko L, Takeda Y, Tuan D. The Long Terminal Repeat (LTR) of ERV-9 Human Endogenous Retrovirus Binds 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.

197. Gasparini P, Sozzi G, Pierotti MA. The role of chromosomal alterations in human cancer development. J Cell Biochem. 2007;102(2):320–31.

198. Oh S, Shin S, Janknecht R. ETV1, 4 and 5: An oncogenic subfamily of ETS transcription factors. Biochim Biophys Acta. 2012;1826(1):1–12.

199. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007;448(7153):595–9.

200. Goering W, Schmitt K, Dostert M, Schaal H, Deenen R, Mayer J, Schulz WA. Human endogenous retrovirus HERV-K(HML-2) activity in prostate cancer is dominated by a few loci. Prostate. 2015;75(16):1958–71.

201. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotech. 2011;29(8):742–9.

202. Kumar KR, Chen W, Koduru PR, Luu HS. Myeloid and Lymphoid Neoplasm With Abnormalities of FGFR1 Presenting With Trilineage Blasts and RUNX1 Rearrangement. Am J Clin Pathol. 2015;143(5):738–48.

203. Guasch G, Popovici C, Mugneret F, Chaffanet M, Pontarotti P, Birnbaum D, Pébusque M-J. Endogenous retroviral sequence is fused to FGFR1 kinase in the 8p12 stem-cell myeloproliferative disorder with t(8;19)(p12;q13.3). Blood. 2002;101(1):286–8.

204. Lamprecht B, Bonifer C, Mathas S. Repeat element-driven activation of proto-oncogenes in human malignancies. Cell Cycle. 2010;9(21):4276–81.

205. Böhne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff J-N. Transposable elements as drivers of genomic and biological diversity in vertebrates. Chromosom Res. 2008;16(1):203–15.

206. Rebollo R, Horard B, Hubert B, Vieira C. Jumping genes and epigenetics: Towards new species. Gene. 2010;454(1–2):1–7.

207. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011;43(8):768–75.

208. Landau Dan A, Clement K, Ziller Michael J, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, et al. Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. Cancer Cell. 2014;26(6):813–25.

209. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, Patel J, Dillon R, Vijay P, Brown AL, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat Med. 2016;22(7):792–9.

210. Mazor T, Pankov A, Song Jun S, Costello Joseph F. Intratumoral Heterogeneity of the Epigenome. Cancer Cell. 2016;29(4):440–51.

211. Hanahan D, Weinberg Robert A. Hallmarks of Cancer: The Next Generation. Cell. 2011;144(5):646–74.

212. Brock A, Chang H, Huang S. Non-genetic heterogeneity - a mutation-independent driving force for the somatic evolution of tumours. Nat Rev Genet. 2009;10(5):336–42.

213. Werfel J, Krause S, Bischof AG, Mannix RJ, Tobin H, Bar-Yam Y, Bellin RM, Ingber DE. How Changes in Extracellular Matrix Mechanics and Gene Expression Variability Might Combine to Drive Cancer Progression. PLoS One. 2013;8(10):e76122.

214. Pisco AO, Brock A, Zhou J, Moor A, Mojtahedi M, Jackson D, Huang S. Non-Darwinian dynamics in therapy-induced cancer drug resistance. Nat Commun. 2013;4:2467.

215. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;12(5):323–34.

216. Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska Renata Z, Ragozin S, Jeltsch A. Specificity of Dnmt1 for Methylation of Hemimethylated CpG Sites Resides in Its Catalytic Domain. Chem Biol. 2012;19(5):572–8.

217. Lavie L, Kitova M, Maldener E, Meese E, Mayer J. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). J Virol. 2005;79(2):876–83.

218. van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet. 2003;19(10):530–6.

219. Simons C, Pheasant M, Makunin IV, Mattick JS. Transposon-free regions in mammalian genomes. Genome Res. 2006;16(2):164–72.

220. Mortada H, Vieira C, Lerat E. Genes Devoid of Full-Length Transposable Element Insertions are Involved in Development and in the Regulation of Transcription in Human and Closely Related Species. J Mol Evol. 2010;71(3): 180–91.

221. Sorek R, Ast G, Graur D. Alu-Containing Exons are Alternatively Spliced. Genome Res. 2002;12(7):1060–7.

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 remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. eLife. 2015;4:e06857.