# Annual Review of Pathology: Mechanisms of Disease Our Conflict with Transposable Elements and Its Implications for Human Disease 

Kathleen H. Burns<br>Department of Pathology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; email: kburns@jhmi.edu

## Annu. Rev. Pathol. Mech. Dis. 2020. 15:51-70

The Annual Review of Pathology: Mechanisms of Disease is online at pathol.annualreviews.org
https://doi.org/10.1146/annurev-pathmechdis-012419-032633

Copyright © 2020 by Annual Reviews. All rights reserved

## Anul

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media


## Keywords

retrotransposon, endogenous retrovirus, LINE-1, Alu, SVA, interferon


#### Abstract

Our genome is a historic record of successive invasions of mobile genetic elements. Like other eukaryotes, we have evolved mechanisms to limit their propagation and minimize the functional impact of new insertions. Although these mechanisms are vitally important, they are imperfect, and a handful of retroelement families remain active in modern humans. This review introduces the intrinsic functions of transposons, the tactics employed in their restraint, and the relevance of this conflict to human pathology. The most straightforward examples of disease-causing transposable elements are germline insertions that disrupt a gene and result in a monogenic disease allele. More enigmatic are the abnormal patterns of transposable element expression in disease states. Changes in transposon regulation and cellular responses to their expression have implicated these sequences in diseases as diverse as cancer, autoimmunity, and neurodegeneration. Distinguishing their epiphenomenal from their pathogenic effects may provide wholly new perspectives on our understanding of disease.


## 1. OUR JUNK-FILLED GENOMES

### 1.1. Introduction to Mobile Genetic Elements

Like other complex eukaryotic genomes, ours is replete with junk DNA: a fossil record of waves of invasions of mobile genetic elements (1, 2). It appears that our ancestral species contended with these endogenous viruses and other transposons as far back in time as we can perceive by sequence homology. Each successful element type left behind copies of its sequence interspersed in our genome. While many of these insertions were lost, others became fixed through genetic drift and, thus, trapped in the amber of our DNA for millions of years after the loss of their ability to propagate in the genome $(3,4)$.

In humans and other mammals, this fossil record is dominated by various types of retroelements (also known as class I mobile elements, retrotransposons, or retroposons). This is in contrast to the class II DNA transposons discovered by Barbara McClintock, which excise themselves and reinsert in the genome and are colloquially known as jumping genes or cut-and-paste elements. Retroelements propagate through the reverse transcription of an RNA intermediate. These copy-and-paste elements make up nearly half of the human genome (5) (Figure 1).

The most commonly recurrent elements are the so-called short interspersed elements (SINEs), which rely on other retrotransposons to encode the protein functions for their reverse transcription and integration in genomes. $A l u(6)$ and mammalian-wide interspersed repeat $(7,8)$ elements are the most prevalent. Together, they make up $20 \%$ of our DNA; Alu sequences number nearly 1 million in the human reference genome assembly. Essentially, all ongoing retrotransposition of these sequences stems from a few AluY subfamilies $(9,10)$, the last letter signifying that these are the young remnants in humans of a burst of $A l u$ retrotransposition that reached a crescendo 35-40 million years ago in ancestral primates (11-13). Alu sequences depend on the endonuclease and reverse transcriptase functions of the long interspersed element-1 (LINE-1, L1) for their propagation in genomes (14).

L1 activity has accompanied the evolution of our species from our eutherian mammalian ancestors ( $15-17$ ). L1 elements are carried in human genomes as $\sim 500,000$ mostly fragmented copies that make up about $17 \%$ of our $\operatorname{DNA}(5,18)$. Today, L1 activity in any individual resides in a small subset of these sequences: 100 or so members of the L1PA1 and L1PA2 subfamilies, or Homo sapiens-specific (L1Hs) elements (19, 20). The most active are termed the hot elements. L1 also transposes SVA [SINE/variable number tandem repeat (VNTR)/Alu] sequences (21, 22), composite retroelements with SINE, VNTR, and Alu sequence homologies. Ongoing germline L1 activity makes L1, Alu, and SVA integrations commonplace, and insertions of these sequences represent common structural variants segregating in populations (20,23-27). We each inherit a unique complement of these mobile element insertions, distinguishing us from other individuals at many hundreds of sites.

Collectively, L1 and the sequences it copies are known as the non-long terminal repeat (LTR) retroelements. This distinguishes them from the endogenous retroviruses, which are flanked by LTRs and have a different mechanism of retrotransposition. In humans, LTR element activity has been extinguished (or all but extinguished) within the past few million years (28), although their sequences persist as about $8 \%$ of our genome. Endogenous retroviruses (ERVs) are named for the transfer RNA that primes their reverse transcription reaction, and the most recently active subfamily is HERV-K (HML2, primed by lysine transfer RNA) (Figure 2). There remain insertion variants of this type in modern humans-that is, genomic loci where both insertion and preinsertion (or empty) alleles can be found (29). The LTRs that flank full-length proviral insertions have a tendency to recombine, deleting the intervening sequences and reducing an insertion to a solo LTR. Such recombinations have led to structural variants of both ERV-K and older ERVs


Figure 1
Transposons in the human genome. The schematic shows the structure of different transposable elements and retrotransposition products. Protein-coding ORFs are shown in dark arrows. The full-length LINE-1 (L1) (6 kb) includes an ORF (ORF0) in its $5^{\prime}$ untranslated region in a $3^{\prime}$ to $5^{\prime}$ orientation (132). ORF2p encodes endonuclease and reverse transcriptase activities. Recurrent types of $5^{\prime}$ truncated L1 are shown next; these vary in size. Alu sequences are about 300 bp . The prototypical SVA insertion is 2 kb , although individual genomic insertions are variable in size. ERV proviral insertions are approximately 10 kb in size. The Pol ORF encodes reverse transcriptase and integrase activities. Abbreviations: Env, envelope; ERV, endogenous retrovirus; FLAM, free left Alu monomer; FRAM, free right Alu monomer; Gag, group antigen; LINE-1, long interspersed element-1; LTR, long terminal repeat; ORF, open reading frame; ORF2p, ORF2 protein; pA, polyA tail; Pol, polymerase; SVA, short interspersed element (SINE)/variable number tandem repeat (VNTR)/Alu. Figure adapted from images created with BioRender.
in modern humans-that is, genomic loci where both the proviral sequence and the collapsed (or solo) LTR-containing alleles can be found (30).

Since L1 is the engine behind modern-day mobile element activity, its mechanism of retrotransposition is considered in further detail here (Figure 3). The life cycle of L1 begins with its transcription, which is mediated by RNA polymerase II (Pol II) at an internal promoter sequence. The RNA intermediate for retrotransposition must encompass the full 6 kb sequence, the socalled unit or authentic L1 transcript (31, 32), which begins with the L1 promoter and extends up to (or slightly past) the $3^{\prime}$ polyA sequence. The overwhelming majority of L1 RNAs in cells


Figure 2
Endogenous retroviral (ERV) retrotransposition. Genomic elements code for RNA (red) that is translated to produce viral proteins (purple). Functional envelope (Env) protein is produced by the rough endoplasmic reticulum and expressed on the surface of the cell, where it is available to coat viral particles that bud from the cell. ERVs with functional Env can reinfect cells, whereas those without are obligate retrotransposons. Group antigen (Gag) forms the viral capsid. Polymerase (Pol) is responsible for transfer RNA-primed reverse transcription and integration of the double-stranded DNA (dsDNA) into the genome. Reverse transcribed DNA is shown in blue. Figure adapted from images created with BioRender.
do not have this structure (32), but rather are read-through transcripts that variously consist of a short-lived, prespliced mRNA; a long noncoding RNA; or a persistent structural RNA (33). So investigators should not equate the presence of L1 RNA or increases in its level with the potential for retrotransposition.

The unit L1 transcript is bicistronic, coding for two proteins in its regular (sense) orientation, ORF1p and ORF2p. ORF1p is produced in greater quantity, and its expression can be readily detected in human tissues by Western blotting or immunostaining (34-36), where it is a marker of L1 expression and a surrogate for retrotransposition potential. ORF1p encodes an RNA binding protein that forms a homotrimeric complex through interactions of its long coiled-coil domain $(37,38)$. A hinge in the ORF1p coiled-coil domain gives it the potential to form higher-order oligomers and lattices as well $(39,40)$. The precise role of ORF1p in retrotransposition is not known, although its presence, RNA binding capacity, and oligomerization all are required for L1 retrotransposition. Dominant negative forms with premature stop codons exist; these retain their coiled-coil domains and presumably suppress retrotransposition in trans by interacting with intact ORF1p molecules (41). ORF1p binds to L1 RNA, but its affinity for RNA is not sequence


Figure 3
Long interspersed element-1 (LINE-1, L1) retrotransposition. L1 RNA is transcribed by RNA polymerase II and translated in the cytosol by host ribosomes. The ORF1 protein (ORF1p) is an RNA binding protein and forms ribonucleoprotein (RNP) complexes with ORF2p and L1 RNA. ORF2p encodes endonuclease and reverse transcriptase activities. Upon nuclear entry, ORF2p nicks target DNA and reverse transcribes the L1 RNA. It uses a $3^{\prime} \mathrm{OH}$ from DNA at the target site to prime the reaction, a mechanism termed target-primed reverse transcription (TPRT). RNA is shown in red; L1-encoded proteins are in blue; and complementary DNA is in black. Figure adapted from images created with BioRender.
specific (42). ORF2p is notoriously difficult to detect. It encodes the enzymatic activities required for retrotransposition, including an apurinic/apyrimidinic endonuclease-like nuclease domain (43, $44)$ and a telomerase-like reverse transcriptase $(45,46)$. These effect retrotransposition through a process known as target-primed reverse transcription (47).

L1 insertions bear several signature sequences of retrotransposition by ORF2p.

1. There is an insertion site preference for the junction between short stretches of thymine (T) and adenine (A) residues (e.g., $5^{\prime}$ TTTT-AA $3^{\prime}$ ), reflecting the preferences of the endonuclease and the need for the L1 RNA to base-pair with the priming DNA strand.
2. Many L1 insertions fail to incorporate the full-length element, but rather are $5^{\prime}$ truncated; a significant proportion is also $5^{\prime}$ inverted (48).
3. Occasionally, insertions will incorporate $3^{\prime}$ transductions (49)—that is, a reverse transcribed sequence matching the genome immediately downstream of the source element.
4. The entire insert is flanked by short ( $<20 \mathrm{bp}$ ) target site duplications; unlike the LTRs of endogenous retroviruses, these do not have a strong tendency to recombine.
5. Newly acquired L1 insertions are uniformly members of the L1PA1 and L1PA2 subfamilies, and the originating element encodes intact proteins. ORF2p shows a strong cis preferencethat is, the tendency of the protein to copy the RNA that templated it $(50,51)$.
$A l u$ is highly effective at breaking this cis preference in the germline. These noncoding RNAs are autonomously transcribed by RNA Pol III and form a highly structured signal recognition particle (52) made up of 7SL-derived Alu RNA and SRP9-14 proteins that together stall ORF2p translation on the ribosome (53) and provide an opportunity for Alu RNA to interact with ORF2p. While ORF2p is necessary for $A l u$ retrotransposition (14), ORF1p appears dispensable. SVA elements use a similar mechanism for retrotransposition via L1-encoded ORF2p (22). Like L1 genomic integration, both $A l u$ and SVA elements bear the insertion site preference and target site duplication features described above. However, unlike L1, they are more likely to insert as intact units as opposed to being $5^{\prime}$ truncated and rearranged. There are limited numbers of subfamilies of retrotransposition-competent $A l u$ and SVA in modern humans (e.g., $A l u \mathrm{Ya5}, A l u \mathrm{Yb} 8, A l u \mathrm{Yb} 9$, and SVA_E and SVA_F), although these are more diverse than the active human-specific L1Hs elements, which are restricted to the L1PA1 and L1PA2 subfamilies.

Estimates are that 1 person in 100-200 people carries a de novo germline integration of an L1 or SVA element-that is, an insertion that is not inherited $(23,24)$. The frequency is predicted to be higher for $A l u$ insertions: About 1 in 20 individuals is estimated to carry a new insertion (54). Less commonly, ORF2p can interact with mRNAs and generate spliced pseudogene integrations (55) or with U6 RNAs to form integrations consisting of U6- $3^{\prime}$ L1 chimeras $(56,57)$.

As is considered in Section 2.1, L1 activity in the germline can be mutagenic, and there is measurable negative selection against full-length, active L1 loci (58), which are relegated to allelic variants in many cases (20). As more complete collections of these active elements and their inheritance patterns become available, it will be interesting to develop combinatorial models to test whether there are limits on the coinheritance of multiple hot L 1 loci.

### 1.2. Mobile Element Silencing and Escape

Retrotransposons contain regions that act as transcriptional activators (i.e., promoters and enhancers), and they rely on these sequences for their expression and, ultimately, their retrotransposition. Because of the threat that this poses to genome integrity, host species have developed epigenetic silencing mechanisms as a counter tactic. These elaborate pathways are most thoroughly studied in experimental models of gametogenesis and early cleavage stage embryogenesis. In these systems, DNA methylation and repressive histone marks are produced by piwi-interacting RNA pathways $(59,60)$ and various protein complexes, such as the Krüppel-associated box (KRAB) domain-containing zinc-finger proteins with KRAB-associated protein-1 (KAP1, TRIM28) complexes (61-63); human silencing hub (HUSH) complex (64, 65); and Polycomb repressor complex 2 (PRC2) (66, 67). Their combined effects tolerize the genome to the accumulation of retroelements by minimizing transposon transcription.

However, these silencing mechanisms are imperfect, and some elements escape and are transcribed. For older and noncoding mobile elements, the consequences of this may be unimportant. However, at retrotransposition-competent L1 loci, the consequence may be significant, and an example of such an escape was recently reported (68) (Figure $4 a$ ). It occurred in a patient with a high-grade mucinous adenocarcinoma of the colon. In this case, the specific L1Hs locus at issue-located on chromosome 17-had an unmethylated promoter sequence and produced RNA in normal tissue samples from the patient. This active element caused a de novo retrotransposon insertion that is inferred to have occurred in the individual's preneoplastic colonic epithelium. The new, somatically acquired L1 insertion disrupted an exon of the adenomatous polyposis coli (APC) tumor suppressor gene (chromosome 5) during its integration, producing an early driver mutation that likely contributed to the subsequent colon cancer. The source element on chromosome 17 is a polymorphic structural variant in human populations, and methylation


Figure 4
Retroelement insertions and rearrangements implicated in disease. (a) A long interspersed element-1 (LINE-1, L1) insertion interrupting the last protein-coding codon in the adenomatous polyposis coli (APC) tumor suppressor gene. The adjacent diagram of the colon illustrates that this insertion was somatically acquired, and it contributed a tumor-driving mutation in colon cancer. (b) An L1 insertion interrupting a protein-coding sequence in the coagulation factor VIII (F8) gene on chromosome X. This and all subsequent loci drawn here are germline (constitutional) alleles. (c) A common Alu insertion variant promotes exon skipping in the CD58 molecule (CD58) gene and appears to contribute to multiple sclerosis risk through this mechanism. (d) An intronic short interspersed element/variable number tandem repeat/Alu (SVA) insertion in the TATA-box binding protein associated factor 1 (TAF1) gene causes X-linked dystonia-parkinsonism. Expansion of the hexanucleotide repeat domain $(\text { CCCTCT })_{n}$ within the SVA is associated with earlier disease onset. This is predicted to promote G-quadruplex formation (shown). (e) An Alu-Alu recombination at the ubiquitin conjugating enzyme E2 T (UBE2T) gene causing Fanconi anemia (complementation group T). The rearrangement between two $A l u \mathrm{Ya5}$ repeats caused deletion of exons 2-6 of the paternal allele and a corresponding duplication of the maternal allele. Figure adapted from images created with BioRender.
studies in other individuals indicate that this specific locus can be effectively silenced. In the patient, the locus was one of $\sim 10$ expressed L 1 Hs ; all other loci appear to have been silenced.

This case brings into focus many significant knowledge gaps. We do not currently have catalogs of epialleles of active L1 in humans, nor have these loci been systematically evaluated as markers of cancer risk. More fundamentally, we have little insight into the mechanisms that might be responsible for the variation in silencing versus escape among the retrotransposition-competent

L1. Why do some individuals silence this same insertion effectively while it was expressed and caused cancer in this patient? Why was this patient able to silence hundreds of nearly identical L1 loci elsewhere in their genome and yet miss this one? Possible scenarios include a stochastic quality to heterochromatin formation in early development, environmental exposures or other changes that affect the maintenance of silencing over time in a locus-specific manner, as well as genetic mechanisms, such as sequence variants within the insertion allele that allow L1 to evade silencing or variations in silencing genes that confer to different individuals distinct competencies for recognizing and silencing different L1. We also have few insights into the frequency of somatic retrotransposition events outside of malignancy. Given that a retrotransposition-competent L1 was left unchecked in this individual, it seems plausible that the person would be prone to acquire other L1 insertions.

### 1.3. Seeing the Greater Good

Retrotransposons that escape from silencing are not always deleterious to the host. Indeed, transcriptional activity intrinsic to older elements-that is, those that are no longer retrotransposition competent-appears to sometimes be repurposed in the service of gene regulation. Transposition of interspersed repeats widely distributes similarly functioning regulatory sequences in the genome, and this provides an avenue for the evolution of gene networks with coordinated expression. For example, LTR elements have strong promoters, and MER41 and other endogenous retroviral sequences have been co-opted to control interferon-responsive genes $(69,70)$.

Host genes that have evolved for the purpose of transposon silencing can similarly be used as a substrate for evolutionary innovation. KRAB zinc-finger protein (ZFP)-coding genes are the most abundant type of gene in our species. They have evolved to silence transposable elements in early development (61-63, 71, 72): our stockpiles in the so-called arms race with transposons. With each win, the target retroelements are kept transcriptionally quiescent and eventually become incompetent for retrotransposition, slowly relaxing evolutionary constraints for their cognate ZFPs. Meanwhile, the repertoire of ZFPs that has emerged from this conflict represents a rich collection of DNA sequence-specific regulators in various stages of repurposing for gene regulation (73, 74).

Sometimes, transposable element protein-coding functions are co-opted by a host species, such that the transposon becomes a so-called domesticated host gene. Recombination activating genes (RAG1 and RAG2) were taken from an ancient Transib DNA transposable element, and its transposition mechanism now mediates $\mathrm{V}(\mathrm{D}) \mathrm{J}$ recombination in B and T cell precursors (75-79) (Figure 5). Mutations in RAG1 and RAG2 cause immunodeficiencies. Other ancient cut-and-paste transposons have been co-opted as centromere protein B (CENPB) and other pogo-derived genes, and sets of piggyBac- and tigger-derived genes (80-82). Of these, mutations in pogo transposable element derived with ZNF domain (POGZ) cause a neurodevelopmental syndrome ( 83,84 ).

In summary, while transposable elements pose real risks as endogenous mutagens and may play roles more broadly in disease, their existence also provides benefits to a species over time. Much of the most exciting work in the field today is leveraging advances in genomics and CRISPR-Cas9enabled functional studies $(65,85)$ to understand transposable element regulation and how this intersects with critical genome functions.

## 2. THE DARK SIDE OF MOBILE ELEMENTS

### 2.1. Disease-Causing Insertion Alleles

The most clear-cut cases of how transposable elements cause disease are those instances in which an insertion in the germline critically affects gene function and creates a characteristic monogenic disease phenotype. The inciting event either can be a de novo insertion (i.e., acquired in the patient


Figure 5
Parallels between $V(D) J$ recombination and DNA transposition. Recombination activating genes RAG1 and RAG2 are derived from a Transib DNA transposon. (Left) Canonical V(D)J recombination. Juxtaposition of the two recombination signal sequences (RSSs) leads to excision of the intervening sequence. RSSs are tripartite modules made up of partially conserved heptamer ( 7 bp ) and nonamer ( 9 bp ) sequences with an intervening spacer of 12 bp or 23 bp . The ends of the excised fragment are ligated as the signal joint. The rearranged strand is repaired as the coding joint and encodes an immunoglobulin or T cell receptor chain. (Right) The alternative to signal joint ligation and loss of the circularized sequence: genomic integration. Figure adapted from images created with BioRender.
or a mosaic parent) or can come from an inherited insertion allele segregating in a population. Highly penetrant insertion alleles are established through genetic drift and are often limited to a founder population (see, e.g., 86-88).

Overall, about 120 examples of transposable element insertions causing genetic disease have been reported (89). There is a widely held expectation that these mutant alleles are being underascertained, as clinical sequencing efforts by and large do not employ tailored pipelines to identify retroelements.

Disease-causing insertions are mostly L1 or Alu elements that interrupt a protein-coding exon or land within a few base pairs of an exon so as to directly disrupt mRNA splicing ( 87,90 ). SVA and pseudogene insertions have been described as well, as have insertions that occur deeper within intronic sequences that disrupt normal splicing through an alternative mechanism referred to as
gene trapping, in which the exon $5^{\prime}$ of the insertion site is spliced to RNA encoded by the insert $(91,92)$.

The first-and still a prototypical-example of a disease-causing insertion was discovered in a patient with hemophilia A by investigators at the Johns Hopkins Hospital (90). An L1 was inserted into an exon of the coagulation factor VIII (F8) gene on chromosome X (Figure 4b) and interrupted the $F 8$ open reading frame. The discovery was impactful since it was the first demonstration of an active mobile genetic element in humans. Subsequently, researchers were able to clone elements and develop in vitro assays for retrotransposition, which have been powerful tools to probe transposon and host factor requirements for transposition $(20,65,93)$.

### 2.2. Mobile Element Polymorphisms and the Common Disease Paradigm

Transposable element insertion variants with lesser impact on gene function may play roles in common diseases. Genome-wide association studies (GWASs) have reinforced the observation that these diseases (i.e., diseases traditionally viewed as sporadic) have a genetic basis, and studies have attributed a portion of that heritability to specific haplotypes. We recently identified hundreds of Alu element insertion variants occurring at the most well-defined disease risk loci (94). Moreover, in many instances, we see that the mobile element insertion is in linkage disequilibrium (LD) with markers of the disease haplotype (i.e., the trait-associated single nucleotide polymorphisms, or SNPs). In these cases, this is genetic evidence suggesting that the GWAS is identifying functional effects of the mobile element insertion. Although the molecular mechanism at these loci is unproven, our study shows that inherited mobile insertion alleles that are common in all of our genomes may be modifying the risk for developing many types of diseases.

One example is an Alu insertion at the CD58 molecule (CD58) locus on a haplotype that affects multiple sclerosis (MS) susceptibility (Figure 4c). GWASs have identified numerous risk alleles for the disease $(95,96)$, including one at $C D 58$. The $C D 58$ genotype is one of the strongest genetic indicators of MS risk (odds ratio $\left.=1.3, p=3 \times 10^{-10}\right)(96,97)$. The risk allele reduces CD58 expression in lymphoblastoid cell lines (97), and MS patients in relapse have lower expression of CD58 on peripheral blood mononuclear cells than do MS patients in remission (97). Thus, decreased expression of CD58 is associated with the risk for developing MS and for relapse. Fine-mapping studies, including resequencing of a 76 kb interval at $C D 58$, and targeted SNP genotyping in 1,278 trio families with MS and 3,341 MS cases and controls indicate a single risk allele in LD with SNP rs2300747 is responsible (97). We found perfect LD between rs2300747 and an Alu insertion variant in an intron of CD58. This is a 302 bp insertion that is present in some individuals and absent in others (allele frequency $=0.66$ ). Further, we found that the Alu insertion alters splicing of CD58 mRNA, promoting skipping of exon 3 in splice reporter assays, and associates with this splice event in lymphoblastoid cell lines (98). The resulting transcript is frame-shifted and presumed to be nonfunctional. Together, these data suggest a mechanism by which a commonly occurring Alu insertion compromises CD58 expression, thereby creating an allele that increases susceptibility to MS.

### 2.3. Rearrangements of Mobile Element Insertions

Some mobile element sequences are intrinsically unstable, and sequence changes within a repeat or rearrangements between different repeats can have phenotypic consequences. A striking example of this was recently reported by investigators studying a disease-causing SVA insertion in an intron of the TATA-box binding protein associated factor 1 (TAF1) gene (Figure 4d). The SVA insertion allele causes X-linked dystonia-parkinsonism, a prevalent movement disorder where the
allele was discovered on Panay Island in the Philippines (99). It is now clear that an allelic series exists in the population there, with variably sized expansions of the hexanucleotide repeat domain $(\text { CCCTCT })_{n}$ within the SVA. Repeat length correlates strongly with disease onset, with longer repeat stretches being associated with earlier manifestations of symptoms (100). The presence of the SVA compromises normal TAF1 mRNA splicing and results in retention of the intron (101). The precise role of hexanucleotide length as a modifying factor is unclear, although this affects the transcriptional potential of the retroelement and may promote G-quadruplex formation (100). These findings underscore the idea that transposable element insertion polymorphisms should not be considered only in simple biallelic terms-that is, an as an insertion allele versus an empty allele-and that the dynamics of SVA sequences deserve special attention.

Rearrangements between two nonallelic insertions can cause genetic disease. For Alu elements, this recombination causes disease alleles as commonly as do de novo retrotransposition events (102). An illustrative example is shown in Figure $4 e$. In that case, an $A l u-A l u$ recombination occurred between tandem AluYa5 insertions at the ubiquitin conjugating enzyme E2 T (UBE2T) gene, causing Fanconi anemia (complementation group T). The rearrangement caused an interstitial deletion encompassing exons 2-6 of the paternal allele, with a corresponding duplication of the maternal allele in the patient (103).

## 3. GOING SLOW TO GO FAST: MOBILE ELEMENT EXPRESSION IN MALIGNANCY

### 3.1. L1 Expression in Cancer

The previous section focused on how germline insertions of transposable elements can affect surrounding genes and cause genetic disease. For the remainder of this review, the focus is on diseases characterized by the dysregulation of transposable elements and hypotheses about how this may affect pathogenesis.

One of the most straightforward examples of the aberrant expression of transposons in diseased tissues is that of L1 expression in cancers. Large numbers of studies have described hypomethylation of the L1 promoter in human malignancies, and some of these have related the phenomenon to clinical features, including higher histologic grades, more advanced clinical stages, and reduced patient survival (all reviewed in Reference 36). L1 promoter hypomethylation, in turn, is associated with ORF1p expression (34). ORF1p overexpression is recognized as a hallmark of many types of cancers, including ovarian and lung, and many gastrointestinal tract tumors, including colon, pancreatic, and esophageal cancers. All are diseases with significant morbidity and mortality that are frequently diagnosed late in the evolution of the malignancy when curative treatment options are limited. Assays of circulating tumor DNA used to evaluate L1 promoter methylation and assays to detect ORF1p in clinically accessible samples may prove useful for early diagnosis, although these remain to be developed.

The proportion of tumors that become immunoreactive for ORF1p and the timing of L1 ORF1p induction depend on the tumor type. For some diseases, such as high-grade serous ovarian carcinoma, there is nearly uniform upregulation of L1 ORF1p ( $90 \%$ of cases) $(34,104)$, and aberrant expression is evident early in the disease course. In these cases, L1 induction co-occurs with fixation of tumor protein p 53 (TP53) tumor suppressor mutations in serous tubal intraepithelial carcinoma precursor lesions of the fallopian tube (104). Precursors of pancreatic carcinoma, termed pancreatic intraepithelial neoplasia, progressively accumulate immunoreactive ORF1p (34). In colon carcinoma, L1 ORF1p expression is induced in malignant cells as compared with normal colonic epithelium. However, as was considered earlier, somatic retrotransposition can antedate neoplastic transformation, and somatically acquired L 1 insertions may be the mutagenic
mechanism underlying loss of the $A P C$ tumor suppressor gene, typically ordered as an early event in the series of genetic changes that lead to colon cancer $(68,105)$. In contrast, glioblastomas express little to no L1 ORF1p and do not support significant numbers of retrotransposition events (106, 107).

### 3.2. New L1 Insertions in Cancer

All of the tumor types characterized by ORF1p expression have been shown to support retrotransposition (108-110), although there is no correspondence between the number of genomic L1 insertions and ORF1p levels. Ovarian cancers express high levels of ORF1p and accrue relatively fewer L1 insertions ( 111,112 ), while lung and colon cancers with more moderate ORF1p levels support high levels of retrotransposition ( $25,68,113$ ). Most somatically acquired insertions of transposons in cancers are $5^{\prime}$ truncated L1. Many are $5^{\prime}$ inverted (114), also a characteristic of germline insertions, and thought to reflect a second priming event during target-primed reverse transcription (48). Others carry $3^{\prime}$ transduction events that allow for the identification of source elements (110). Interestingly, unlike germline integrations, somatically acquired insertions in cancers appear to be dominated by L1 itself, with far fewer instances of Alu, SVA, and pseudogene retrotransposition $(110,115)$.

The majority of new insertions in tumors appear to be passenger events, as there is little tendency for the reoccurrence of independent insertions at the same locations in the genome (116). With exceedingly few exceptions, acquired insertions spare exons. This said, it is expected that some mutations in noncoding regions of the genome will be key events in cancer development, and as catalogs of noncoding sequence changes and somatic retrotransposition events mature, it will be important to see whether there is correspondence between the two. Such an observation would suggest that different mutagenic mechanisms disrupt the same key regulatory sequences. It may be that L1 insertions introduce new regulatory sequences to sites as well, but few functions have been ascribed to the $3^{\prime}$ end of the L1 element. In summary, there is limited evidence that L1 insertions per se have important roles in tumorigenesis.

As spatiotemporal maps of somatically acquired insertions are refined, though, we can expect that they will provide insights into the timing of L1 expression and activity throughout the evolution of a tumor, inform how well or poorly L1 activity is tolerated as cancers grow, and serve as markers of subclonal phylogenies (110, 112, 114).

### 3.3. Cellular Effects of Transposon Expression

Could L1 expression impact cancer cell biology in other ways? In experimental systems, L1 expression slows cell growth or is overtly cytotoxic and damages DNA (117-119), and it is hard to envision a scenario in which its expression confers a direct cell growth advantage. However, it is plausible that the DNA damage it exacts transforms cells or that it puts cells in a state analogous to oncogene-induced senescence $(120,121)$, creating selective advantages for preneoplastic populations of cells with critical tumor suppressor mutations.

The expression of mobile genetic elements in cancers is not limited to L1, although that has been our primary focus up to this point. Because of their relative age and lack of protein-coding capacity, respectively, SINEs and ERV elements are less likely to directly incite DNA damage. However, new studies indicate that their expression may be an important determinant of the immunologic properties of cancerous cells. Both families of retroelements can incite cellular interferon responses (122-124) (Figure 6). Excitingly, expression of ERVs can be induced by demethylating agents, an effect that mimics exogenous viral infection (123-125). The same effect


Figure 6
Demonstrated and hypothetical intersections between transposable element (TE) expression and immunity. Activation of TE transcription may directly upregulate interferon-responsive genes (ISGs), given that some long terminal repeat sequences function as promoters or enhancers at these loci. TE-derived RNAs are prevalent and can form double-stranded RNA (dsRNA) in the nucleus. Annealing of these hybrids is relaxed by adenosine (A)-to-inosine (I) editing. ADAR represents adenosine deaminases acting on RNA. Unedited hybrids are prone to come into contact with cytoplasmic dsRNA sensors, such as retinoic acid inducible gene I (RIG-I), melanoma differentiation associated protein 5 (MDA5), and others, for example, endosomal Toll-like receptors (not shown), mimicking viral infection and prompting interferon responses (red arrow). TE-encoded proteins may also play roles as well. TE-encoded endonucleases can damage DNA (red star) and generate nucleic acids that might be sensed through cyclic guanosine monophosphate-adenosine monophosphate (cGAS)/stimulator of interferon genes (STING) and related pathways. TE-encoded reverse transcriptases may generate RNA:DNA hybrids and other complementary DNA forms that feed into the same pathways. TE-encoded peptides may be presented to cytotoxic T cells by major histocompatibility complex I (MHC I) or the proteins handled by antigen presenting cells (not shown). It is unknown whether these proteins are targets of peripheral tolerization in the thymus (i.e., recognized as self or nonself). Figure adapted from images created with BioRender.
can be seen with the inhibition of the lysine-specific demethylase 1A (LSD1) gene (126). This provides a rationale for the combination of epigenetic therapies (e.g., demethylating agents and others that promote transposable element expression) with immune checkpoint inhibitors [e.g., targeting programmed cell death 1 (PD-1) and its ligand (PD-L1)].

## 4. CELLULAR RESPONSES TO TRANSPOSON EXPRESSION IN INFLAMMATORY DISEASES

The paradigm that endogenous transposon expression can elicit the same interferon responses that have been associated with viral infections has implications for many other diseases. For example, $A l u, \mathrm{~L} 1$, and other repeats have been proposed as central components of the pathogenesis of Aicardi-Goutières syndrome (AGS) diseases, a spectrum of inflammatory disorders characterized by severe neurologic impairment. AGS is caused by loss-of-function mutations in genes that normally limit the feeding of nucleic acids into the viral sensor pathways, including three prime repair exonuclease 1 (TREX1), ribonuclease H 2 endonuclease complex components, adenosine deaminase acting on RNA 1 (ADAR1) double-stranded RNA (dsRNA) adenosine-to-inosine editing enzymes, and others. Gain-of-function mutations in melanoma differentiation associated protein 5 (MDA5), encoded by interferon induced with helicase C domain 1 (IFIH1), that make it hypersensitive to Alu dsRNAs (127) also cause AGS and a heritable form of systemic lupus erythematosus. Independent investigators have pointed to different types of retroelement-derived nucleic acidsincluding dsRNA, RNA:DNA hybrids, and extrachromosomal DNA—and described models for AGS that converge on sensors of these products [i.e., retinoic acid inducible gene I (RIG-I) and MDA5 or cyclic guanosine monophosphate-adenosine monophosphate (cGAS)/stimulator of interferon genes (STING)] and the triggering of interferon signaling (127-129) (Figure 6). Similarly, interferon signaling prompted by L1 reverse transcriptase activity in senescent cells has been suggested recently to be an underlying basis for aging-associated inflammation (130).

Together, these studies point toward a need to better characterize transposable element insertion variants and their expression in sporadic inflammatory diseases and in heritable epigenetic disorders (131) and to identify cellular responses to these RNAs and proteins.

## 5. CONCLUSIONS AND FUTURE PERSPECTIVES

We are at an exciting time when advances in concepts of transposable element biology will inform perspectives on disease pathogenesis. We are more capable and curious than ever to learn more about these elements: If they were previously dismissed as junk DNA and intractable to experimentation, this is less true today.

As more genomes are sequenced, comparative genomics will likely point to functional elements. Structural and sequence variants within human populations will associate with specific diseases, and new insertions responsible for clinical phenotypes will allow for a molecular accounting of more occurrences of genetic disease.

Single-cell advances will give us the chance to probe for somatic retrotransposition in noncancerous tissues in disease contexts and to build more comprehensive maps of cancer cell lineages. New sequencing technologies, including long-read single-molecule sequencing, and tailored informatics for RNA expression profiling will allow us to better map expressed transposons and distinguish read-through transcription from authentic transcription. Transcript structure will be an important determinant of, for example, protein-coding capacity or Alu-Alu dsRNA hybrid formation.

As disease-specific experimental systems are established, CRISPR-Cas9-based tools for transcriptional activation and silencing will enhance our ability to interrogate the functions of specific elements. The field will need additional reagents to detect and pharmacologically inhibit specific activities of transposon-encoded proteins.

Conclusions that a retroelement is important to disease should become less reliant on quantitative measures of transcripts or DNA copy number changes that cannot be traced unambiguously to their origins or on nonspecific nucleoside analog inhibitors of reverse transcriptase. If the field can avoid pitfalls and continue to find creative and rigorous ways to interrogate the effects of mobile elements, we are certain to discover fascinating biology and make meaningful inroads into understanding disease.

## SUMMARY POINTS

1. Much of our genome is derived from mobile genetic elements, and this reflects a rich evolutionary history of many diverse, self-propagating sequences.
2. Most elements in the genome today are no longer competent as mobile DNAs, but a subset of L1 and the sequences it mobilizes remain active.
3. New insertions can cause genetic disease by disrupting critical functional sequences, and insertions with lesser effects may contribute to common disease risk.
4. L1 is highly expressed in many human cancers, where it functions as an endogenous mutagen.
5. Transposable element expression triggers cellular interferon responses in experimental systems, an observation with implications for tumor immunotherapy, inflammatory diseases, and potentially for other pathologies.

## DISCLOSURE STATEMENT

Johns Hopkins University has licensed LINE-1 ORF1p antibodies to EMD Millipore. K.H.B. receives royalty payments from these sales.

## LITERATURE CITED

1. Huang CR, Burns KH, Boeke JD. 2012. Active transposition in genomes. Annu. Rev. Genet. 46:651-75
2. Burns KH, Boeke JD. 2012. Human transposon tectonics. Cell 149:740-52
3. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110:462-67
4. Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. Institute for Systems Biology. http:// www.repeatmasker.org
5. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921
6. Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. 3:370-79
7. Jurka J, Zietkiewicz E, Labuda D. 1995. Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the Mesozoic era. Nucleic Acids Res. 23:170-75
8. Smit AFA, Riggs AD. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res. 23:98-102
9. Konkel MK, Walker JA, Hotard AB, Ranck MC, Fontenot CC, et al. 2015. Sequence analysis and characterization of active human Alu subfamilies based on the 1000 Genomes Pilot Project. Genome Biol. Evol. 7:2608-22
10. Kryatova MS, Steranka JP, Burns KH, Payer LM. 2017. Insertion and deletion polymorphisms of the ancient AluS family in the human genome. Mob. DNA 8:6
11. Kapitonov V, Jurka J. 1996. The age of Alu subfamilies. 7. Mol. Evol. 42:59-65
12. Shen MR, Batzer MA, Deininger PL. 1991. Evolution of the master Alu gene(s). 7. Mol. Evol. 33:31120
13. Konkel MK, Walker JA, Batzer MA. 2010. LINEs and SINEs of primate evolution. Evol. Anthropol. 19:236-49
14. Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. 35:41-48
15. Boissinot S, Chevret P, Furano AV. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol. Biol. Evol. 17:915-28
16. Boissinot S, Entezam A, Young L, Munson PJ, Furano AV. 2004. The insertional history of an active family of L1 retrotransposons in humans. Genome Res. 14:1221-31
17. Boissinot S, Furano AV. 2005. The recent evolution of human L1 retrotransposons. Cytogenet. Genome Res. 110:402-6
18. Beck CR, Garcia-Perez JL, Badge RM, Moran JV. 2011. LINE-1 elements in structural variation and disease. Annu. Rev. Genom. Hum. Genet. 12:187-215
19. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, et al. 2003. Hot L1s account for the bulk of retrotransposition in the human population. PNAS 100:5280-85
20. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, et al. 2010. LINE-1 retrotransposition activity in human genomes. Cell 141:1159-70
21. Raiz J, Damert A, Chira S, Held U, Klawitter S, et al. 2012. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 40:1666-83
22. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr. 2011. Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum. Mol. Genet. 20:3386-400
23. Huang CR, Schneider AM, Lu Y, Niranjan T, Shen P, et al. 2010. Mobile interspersed repeats are major structural variants in the human genome. Cell 141:1171-82
24. Ewing AD, Kazazian HH Jr. 2010. High-throughput sequencing reveals extensive variation in humanspecific L1 content in individual human genomes. Genome Res. 20:1262-70
25. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, et al. 2010. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141:1253-61
26. Stewart C, Kural D, Stromberg MP, Walker JA, Konkel MK, et al. 2011. A comprehensive map of mobile element insertion polymorphisms in humans. PLOS Genet. 7:e1002236
27. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, et al. 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526:75-81
28. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, et al. 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res. 16:1548-56
29. Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M. 2005. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERVK(HML2): implications for present-day activity. 7. Virol. 79:12507-14
30. Thomas J, Perron H, Feschotte C. 2018. Variation in proviral content among human genomes mediated by LTR recombination. Mob. DNA 9:36
31. Skowronski J, Fanning TG, Singer MF. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. Mol. Cell. Biol. 8:1385-97
32. Deininger P, Morales ME, White TB, Baddoo M, Hedges DJ, et al. 2016. A comprehensive approach to expression of L1 loci. Nucleic Acids Res. 45:e31
33. Hall LL, Carone DM, Gomez AV, Kolpa HJ, Byron M, et al. 2014. Stable C $\mathrm{C}_{0}$ T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. Cell 156:907-19
34. Rodic N, Sharma R, Sharma R, Zampella J, Dai L, et al. 2014. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am. 7. Pathol. 184:1280-86
35. Sharma R, Rodic N, Burns KH, Taylor MS. 2016. Immunodetection of human LINE-1 expression in cultured cells and human tissues. Methods Mol. Biol. 1400:261-80
36. Ardeljan D, Taylor MS, Ting DT, Burns KH. 2017. The human long interspersed element-1 retrotransposon: an emerging biomarker of neoplasia. Clin. Chem. 63:816-22
37. Martin SL, Cruceanu M, Branciforte D, Wai-Lun Li P, Kwok SC, et al. 2005. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. 7. Mol. Biol. 348:549-61
38. Khazina E, Truffault V, Buttner R, Schmidt S, Coles M, Weichenrieder O. 2011. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. Nat. Struct. Mol. Biol. 18:1006-14
39. Naufer MN, Callahan KE, Cook PR, Perez-Gonzalez CE, Williams MC, Furano AV. 2016. L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. Nucleic Acids Res. 44:281-93
40. Khazina E, Weichenrieder O. 2018. Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. eLife 7:e34960
41. Sokolowski M, Chynces M, deHaro D, Christian CM, Belancio VP. 2017. Truncated ORF1 proteins can suppress LINE-1 retrotransposition in trans. Nucleic Acids Res. 45:5294-308
42. Kolosha VO, Martin SL. 2003. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). 7. Biol. Chem. 278:811217
43. Feng Q, Moran JV, Kazazian HH Jr., Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905-16
44. Weichenrieder O, Repanas K, Perrakis A. 2004. Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. Structure 12:975-86
45. Mathias SL, Scott AF, Kazazian HH Jr., Boeke JD, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. Science 254:1808-10
46. Kopera HC, Moldovan JB, Morrish TA, Garcia-Perez JL, Moran JV. 2011. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. PNAS 108:20345-50
47. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595-605
48. Ostertag EM, Kazazian HH Jr. 2001. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res. 11:2059-65
49. Moran JV, DeBerardinis RJ, Kazazian HH. 1999. Exon shuffling by L1 retrotransposition. Science 283:1530-34
50. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, et al. 2001. Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21:1429-39
51. Kulpa DA, Moran JV. 2006. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat. Struct. Mol. Biol. 13:655-60
52. Weichenrieder O, Wild K, Strub K, Cusack S. 2000. Structure and assembly of the Alu domain of the mammalian signal recognition particle. Nature 408:167-73
53. Ahl V, Keller H, Schmidt S, Weichenrieder O. 2015. Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. Mol. Cell 60:715-27
54. Xing J, Zhang Y, Han K, Salem AH, Sen SK, et al. 2009. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 19:1516-26
55. Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. 24:363-67
56. Buzdin A, Ustyugova S, Gogvadze E, Vinogradova T, Lebedev Y, Sverdlov E. 2002. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the $3^{\prime}$ terminus of L1. Genomics 80:402-6
57. Doucet AJ, Droc G, Siol O, Audoux J, Gilbert N. 2015. U6 snRNA pseudogenes: markers of retrotransposition dynamics in mammals. Mol. Biol. Evol. 32:1815-32
58. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. 2006. Fitness cost of LINE-1 (L1) activity in humans. PNAS 103:9590-94
59. Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318:761-64
60. Newkirk SJ, Lee S, Grandi FC, Gaysinskaya V, Rosser JM, et al. 2017. Intact piRNA pathway prevents L1 mobilization in male meiosis. PNAS 114:E5635-44
61. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, et al. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature 516:242-45
62. Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, et al. 2012. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell Rep. 2:766-73
63. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, et al. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463:237-40
64. Robbez-Masson L, Tie CHC, Conde L, Tunbak H, Husovsky C, et al. 2018. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res. 28:836-45
65. Liu N, Lee CH, Swigut T, Grow E, Gu B, et al. 2018. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. Nature 553:228-32
66. Ohtani H, Liu M, Zhou W, Liang G, Jones PA. 2018. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. Genome Res. 28:1147-57
67. Walter M, Teissandier A, Perez-Palacios R, Bourc'his D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife 5:e11418
68. Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE. 2016. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res. 26:745-55
69. Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 351:1083-87
70. Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. 18:71-86
71. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, et al. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487:57-63
72. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, et al. 2014. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. 24:1260-70
73. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, et al. 2016. Transposable elements and their KRABZFP controllers regulate gene expression in adult tissues. Dev. Cell 36:611-23
74. Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543:550-54
75. Agrawal A, Eastman QM, Schatz DG. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394:744-51
76. Hencken CG, Li X, Craig NL. 2012. Functional characterization of an active Rag-like transposase. Nat. Struct. Mol. Biol. 19:834-36
77. Hiom K, Melek M, Gellert M. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell 94:463-70
78. Huang S, Tao X, Yuan S, Zhang Y, Li P, et al. 2016. Discovery of an active RAG transposon illuminates the origins of $\mathrm{V}(\mathrm{D}) \mathrm{J}$ recombination. Cell 166:102-14
79. Kapitonov VV, Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLOS Biol. 3:e181
80. Sinzelle L, Izsvak Z, Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell. Mol. Life Sci. 66:1073-93
81. Smit AF, Riggs AD. 1996. Tiggers and DNA transposon fossils in the human genome. PNAS 93:144348
82. Sarkar A, Sim C, Hong YS, Hogan JR, Fraser MJ, et al. 2003. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Mol. Genet. Genom. 270:173-80
83. Deciphering Dev. Disord. Study. 2015. Large-scale discovery of novel genetic causes of developmental disorders. Nature 519:223-28
84. Stessman HAF, Willemsen MH, Fenckova M, Penn O, Hoischen A, et al. 2016. Disruption of POGZ is associated with intellectual disability and autism spectrum disorders. Am. 7. Hum. Genet. 98:541-52
85. Fuentes DR, Swigut T, Wysocka J. 2018. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. eLife 7:e35989
86. Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, et al. 1998. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature 394:388-92
87. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, et al. 1993. Insertion of a $5^{\prime}$ truncated L1 element into the $3^{\prime}$ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. 7. Clin. Investig. 91:1862-67
88. Taniguchi-Ikeda M, Kobayashi K, Kanagawa M, Yu CC, Mori K, et al. 2011. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. Nature 478:127-31
89. Hancks DC, Kazazian HH Jr. 2016. Roles for retrotransposon insertions in human disease. Mob. DNA 7:9
90. Kazazian HH Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A resulting from de novo insertion of L 1 sequences represents a novel mechanism for mutation in man. Nature 332:164-66
91. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr. 2009. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 19:1983-91
92. de Boer M, van Leeuwen K, Geissler J, Weemaes CM, van den Berg TK, et al. 2014. Primary immunodeficiency caused by an exonized retroposed gene copy inserted in the CYBB gene. Hum. Mutat. 35:486-96
93. Kopera HC, Larson PA, Moldovan JB, Richardson SR, Liu Y, Moran JV. 2016. LINE-1 cultured cell retrotransposition assay. Methods Mol. Biol. 1400:139-56
94. Payer LM, Steranka JP, Yang WR, Kryatova M, Medabalimi S, et al. 2017. Structural variants caused by Alu insertions are associated with risks for many human diseases. PNAS 114:E3984-92
95. Int. Mult. Scler. Genet. Consort. 2013. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat. Genet. 45:1353-60
96. Int. Mult. Scler. Genet. Consort, Wellcome Trust Case Control Consort. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476:214-19
97. De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, et al. 2009. The role of the CD58 locus in multiple sclerosis. PNAS 106:5264-69
98. Payer LM, Steranka JP, Ardeljan D, Walker J, Fitzgerald KC, et al. 2019. Alu insertion variants alter mRNA splicing. Nucleic Acids Res. 47:421-31
99. Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, et al. 2007. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. Am. 7. Hum. Genet. 80:393-406
100. Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, et al. 2017. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. PNAS 114:E11020-28
101. Aneichyk T, Hendriks WT, Yadav R, Shin D, Gao D, et al. 2018. Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. Cell 172:897909
102. Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. 10:691-703
103. Rickman KA, Lach FP, Abhyankar A, Donovan FX, Sanborn EM, et al. 2015. Deficiency of UBE2T, the E2 ubiquitin ligase necessary for FANCD2 and FANCI ubiquitination, causes FA-T subtype of Fanconi anemia. Cell Rep. 12:35-41
104. Pisanic TR 2nd, S Asaka, Lin SF, Yen TT, Sun H, et al. 2019. Long interspersed nuclear element 1 retrotransposons become deregulated during the development of ovarian cancer precursor lesions. Am. 7. Pathol. 189:513-20
105. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, et al. 1992. Disruption of the $A P C$ gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res. 52:643-45
106. Achanta P, Steranka JP, Tang Z, Rodic N, Sharma R, et al. 2016. Somatic retrotransposition is infrequent in glioblastomas. Mob. DNA 7:22
107. Carreira PE, Ewing AD, Li G, Schauer SN, Upton KR, et al. 2016. Evidence for L1-associated DNA rearrangements and negligible L 1 retrotransposition in glioblastoma multiforme. Mob. DNA 7:21
108. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, et al. 2012. Landscape of somatic retrotransposition in human cancers. Science 337:967-71
109. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. 2014. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. 24:1053-63
110. Tubio JM, Li Y, Ju YS, Martincorena I, Cooke SL, et al. 2014. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science 345:1251343
111. Tang Z, Steranka JP, Ma S, Grivainis M, Rodic N, et al. 2017. Human transposon insertion profiling: analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. PNAS 114:E733-40
112. Nguyen THM, Carreira PE, Sanchez-Luque FJ, Schauer SN, Fagg AC, et al. 2018. L1 retrotransposon heterogeneity in ovarian tumor cell evolution. Cell Rep. 23:3730-40
113. Solyom S, Ewing AD, Rahrmann EP, Doucet T, Nelson HH, et al. 2012. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res. 22:2328-38
114. Rodic N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, et al. 2015. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nat. Med. 21:1060-64
115. Cooke SL, Shlien A, Marshall J, Pipinikas CP, Martincorena I, et al. 2014. Processed pseudogenes acquired somatically during cancer development. Nat. Commun. 5:3644
116. Burns KH. 2017. Transposable elements in cancer. Nat. Rev. Cancer 17:415-24
117. Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A. 2006. Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. Cancer Cell Int. 6:13
118. Haoudi A, Semmes OJ, Mason JM, Cannon RE. 2004. Retrotransposition-competent human LINE-1 induces apoptosis in cancer cells with intact p53. 7. Biomed. Biotechnol. 2004:185-94
119. Gasior SL, Wakeman TP, Xu B, Deininger PL. 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. 7. Mol. Biol. 357:1383-93
120. Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, et al. 2006. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature 444:633-37
121. Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, et al. 2006. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature 444:638-42
122. Leonova KI, Brodsky L, Lipchick B, Pal M, Novototskaya L, et al. 2013. p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. PNAS 110:E89-98
123. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, et al. 2015. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell 162:974-86
124. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, et al. 2015. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. Cell 162:961-73
125. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD. 2019. Epigenetic therapy in immune-oncology. Nat. Rev. Cancer 19:151-61
126. Sheng W, LaFleur MW, Nguyen TH, Chen S, Chakravarthy A, et al. 2018. LSD1 ablation stimulates anti-tumor immunity and enables checkpoint blockade. Cell 174:549-63
127. Ahmad S, Mu X, Yang F, Greenwald E, Park JW, et al. 2018. Breaching self-tolerance to Alu duplex RNA underlies MDA5-mediated inflammation. Cell 172:797-810
128. Lim YW, Sanz LA, Xu X, Hartono SR, Chedin F. 2015. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. eLife 4:e08007
129. Thomas CA, Tejwani L, Trujillo CA, Negraes PD, Herai RH, et al. 2017. Modeling of TREX1dependent autoimmune disease using human stem cells highlights L 1 accumulation as a source of neuroinflammation. Cell Stem Cell 21:319-31
130. De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, et al. 2019. L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature 566:73-78
131. Fahrner JA, Bjornsson HT. 2014. Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. Annu. Rev. Genom. Hum. Genet. 15:269-93
132. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, et al. 2015. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell 163:583-93
