Published in final edited form as:

Curr Opin Genet Dev. 2013 June ; 23(3): 264–270. doi:10.1016/j.gde.2013.02.007.

Functional Impact of the Human Mobilome

Timothy D. Babatz^{1,2} and Kathleen H. Burns^{1,2,3,4,*}

¹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

²High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD

³Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD

⁴The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD

Abstract

The human genome is replete with interspersed repetitive sequences derived from the propagation of mobile DNA elements. Three families of human retrotransposons remain active today: LINE1, *Alu*, and SVA elements. Since 1988, *de novo* insertions at previously recognized disease loci have been shown to generate highly penetrant alleles in Mendelian disorders. Only recently has the extent of germline-transmitted retrotransposon insertion polymorphism (RIP) in human populations been fully realized. Also exciting are recent studies of somatic retrotransposition in human tissues and reports of tumor-specific insertions, suggesting roles in tissue heterogeneity and tumorigenesis. Here we discuss mobile elements in human disease with an emphasis on exciting developments from the last several years.

Introduction

Envisioning the RNA origin of DNA may evoke thoughts of a world teeming with self-replicating ribonucleic acids with functions that blur the boundary between the inanimate and simple life. Though sequence evidence for this world is obscure, the RNA origins of our modern human genome are readily evident; even the casual genome gazer can take in a vast genomic fossil record. Our 3 billion base pairs are overwhelmingly non-coding, and 50% or more is recognizable as repetitive sequences. Most repeats are interspersed, meaning they occur discontinuously as singularly scattered copies in the genome. These include Long Interspersed Elements (LINEs) and Short Interspersed Elements (SINEs). Most abundant in mammals are the retrotransposons, which self-replicate through the reverse transcription of RNA intermediates.

Genomes are not static strings of bases, but instead are mutable, evolving, and structurally dynamic. Three families of human retrotransposons have been active in relatively recent genome evolution and continue to generate human genomic variation today. First, the autonomous LINE-1 element (L1) is 6 kb in its full-length form and contains an internal

© 2013 Elsevier Ltd. All rights reserved.

*Corresponding author: Kathleen H. Burns, M.D., Ph.D., Johns Hopkins University School of Medicine, 720 Rutland Avenue, Ross Building, Room 524, Baltimore, MD 21205, phone (410) 502-7214, pager (410) 283-6213, kburns@jhmi.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

promoter, two open reading frames (ORFs), and a polyadenylation signal. After RNA polymerase II transcription, mRNA processing, and export to the cytoplasm, ORF1 and ORF2 encode two proteins (ORF1p and ORF2p) that preferentially associate with their encoding RNA [1,2]. The resulting ribonucleoprotein (RNP) particles return the RNA to the nucleus, where ORF2p initiates target-primed reverse transcription (TPRT) to insert a copied DNA sequence at a new site in the host cell genome [3–5]. There are approximately 500,000 copies of L1-derived sequence in the human genome, most of which have been deactivated by 5′-truncation on integration or accumulated mutations. The youngest human-specific L1 subfamily, L1Hs, accounts for about 1% of these copies, with a much smaller subset of “hot L1” elements driving most retrotransposition events.

The remaining families of active retrotransposons are non-autonomous, relying on L1-encoded machinery for mobilization [6,7]. *Alu* elements are ~300 bp long and there are approximately 1,000,000 *Alu*-derived sequences in the human genome. Nucleotide variations within the *Alu* consensus sequence distinguish several subfamilies including the most active *AluY* elements (reviewed in [8]). SVA elements comprise the third family of active retrotransposons; these are a relatively heterogeneous group of sequences with several subfamilies recognized. SVAs are composite elements, named for the sources of their component parts, SINE-R, VNTR, and *Alu* sequences. The SVA consensus sequence is approximately 2 kb and there are roughly 3,700 SVA-derived sequences in the human genome (reviewed in [9]).

Functional Impact

Host genomes are vulnerable to disruption by retrotransposons. How a gene is affected is dependent on both where the mobile DNA insertion occurs and the internal sequence of the element. Exonic insertions can interrupt the coding region of a gene. In addition, *in vitro* cell-based assays have revealed that features of L1 sequence such as the promoter, 5′-UTR antisense promoter, and internal splice sites can affect the transcription of target regions and transcript processing [10,11]. Concordantly, intronic L1 insertions can induce aberrant transcript initiation, premature truncation, or attenuate transcript elongation, affecting the structure and abundance of target gene transcripts [12–14]. Similarly, intronic and gene proximal *Alu* insertions can act as enhancers or promoters, introduce cryptic splice donors or acceptors, or abrogate endogenous exon splicing [15–19]. In clinical genetics, retrotransposon insertions have been implicated in rare cases of highly penetrant Mendelian disorders including hemophilia, cystic fibrosis, neurofibromatosis, and breast and colon cancers [20,21]. In each respective example the novel insertion creates a highly damaging allele affecting the coding capacity of a known disease gene and thus is fairly definitively implicated as the cause of the disorder.

In contrast, less deleterious insertions that are inherited and constitutive in an individual are major sources of structural variation in the human genome, genetic variation in populations, and occasional causes of genetic disease. Several studies in 2010 demonstrated that retrotransposon insertion polymorphisms (RIPs) are significantly more polymorphic among individual human genomes than previously recognized.

Germline Insertions

Because *de novo* insertions are templated from the relatively few active subfamilies of repeats, each new insertion bears high internal sequence homology to the subfamily consensus sequence. This feature can be leveraged in targeted approaches for finding mobile DNA variants. These methods can be broadly categorized as either using laboratory-based genomic assays to map repeats or using computational means to recognize insertions in unselected reads in sequencing studies. The first category has included variations on hemi-

specific PCRs using one primer within the repeat family of interest and a reaction schema that amplifies adjacent DNA [22–25]. The latter has had the advantage of using data from more large-scale collaborative efforts, such as the 1000 Genomes project, to identify large numbers of new insertions and phase these with neighboring SNPs [26,27] (Figure 1). At the same time, Beck *et al.* reported polymorphism of full-length “hot” L1Hs elements among populations, suggesting possible inter-individual variation in endogenous L1 activity [28]. Collectively, new perspectives resulting from these discovery efforts have upheld earlier assertions that mobile DNAs are a major source of human genetic diversity. We currently estimate that between 1/100 and 1/200 individuals harbor a *de novo* LINE-1 insertion, and that about 1/20 individuals have a *de novo* *Alu* element insertion. This degree of activity is expected to result in approximately 11,000 L1Hs and 65,000 *Alu*Y common variable insertion alleles (MAF > 0.05) segregating in human populations and additional increasingly rare alleles.

Recently, two studies elucidated the mutational mechanism of inherited insertions in recessive diseases, one of which was a novel insertion discovery. A third study reported a novel mechanism of accumulated *Alu* RNA cytotoxicity.

Taniguchi-Ikeda *et al.* functionally characterized the polymorphic SVA insertion in the 3' UTR of the *FUKUTIN* gene, which causes Fukuyama muscular dystrophy (FCMD). FCMD is a rare autosomal recessive disease mainly described in the Japanese population [29] and currently has no effective treatments. This ancient SVA is on a haplotype shared by >80% of FCMD chromosomes and likely arose as a single ancestral allele [30]. Long-range RT-PCR and sequence analysis in FCMD lymphoblast mRNA and controls revealed a novel splicing event between a weak alternative donor site in the final *FUKUTIN* coding exon and an acceptor site within the SVA insertion. This aberrant splicing excises the normal stop codon, deletes coding sequences for the 38 C-terminal amino acids of FUKUTIN, and adds 129 amino acids encoded by the SVA sequence. Normal FUKUTIN localizes to the Golgi, while FUKUTIN with SVA-derived C-terminus mislocalizes to the ER. Remarkably, the authors were able to restore normal mRNA splicing of FCMD alleles to 40% of control levels using a combination of antisense oligonucleotides (AONs) targeted to the alternative splice donor, acceptor, and an exonic splicing enhancer. This strategy increased normal *FUKUTIN* mRNA expression significantly and rescued functional Fukutin protein *in vivo* in knock-in mice homozygous for humanized SVA-containing *FUKUTIN*. This study elucidated the mutational mechanism of a polymorphic disease-causing retrotransposon and demonstrated the possibility of splicing modulation therapy as a treatment for FCMD.

Tucker *et al.* [31] performed exome sequencing on a single proband with retinitis pigmentosa (RP), a typically autosomal recessive ciliopathy characterized by loss of photoreceptor cells and a significant amount of genetic heterogeneity. After stringent filtering of sequence data from both ABI and Illumina sequencing instruments and prioritization of identified variants, there was a surprising lack of plausible disease-causing mutations verified by Sanger sequencing and present in both datasets. The ABI data showed a compound heterozygous pair of mutations in exon 9 of male germ cell-associated kinase (*MAK*) that were not in the Illumina data, and Sanger sequencing of the exon surprisingly revealed a homozygous 353-bp *Alu* insertion. The authors investigated why this important variant was missed by one method and misidentified by the other. The ABI protocol completely removed the *Alu*-*MAK* junction sequences and therefore sequenced chimeric DNA molecules and identified two heterozygous variants, and the Illumina protocol sequenced the junction sequences but trimmed the *Alu* sequence from the ends of the junction reads during analysis, making it invisible. The authors went on to functionally characterize the homozygous *Alu* variant in induced pluripotent stem cells (iPSCs) derived from the RP patient's dermal fibroblasts and controls, showing that the *Alu* insertion

abrogates an essential developmental switch to alternatively include exon 9 and a previously unknown retina-specific exon 12 in retinal lineage differentiated cells. Screening of 1,798 additional RP patients yielded 21 individuals homozygous for the *Alu*, all of whom reported Jewish ancestry, and the *Alu* was absent from controls. This study highlights current strengths and weaknesses of relatively new mutation discovery tools, and suggests that there may be significantly more disease causing insertions than previously appreciated.

Kaneko *et al.* [32] reported a pathogenic role for accumulating *Alu* RNA in retinal pigment epithelial (RPE) degeneration in geographic atrophy (GA), an untreatable advanced form of age-related macular degeneration. This is the first known example of *Alu* RNA cytotoxicity, rather than *Alu*-mediated genomic rearrangement or insertional mutagenesis, causing disease. The authors observed a significant deficit of *DICER1* mRNA and protein and an excess of dsRNA in RPE of GA donor eyes compared to controls. Subsequent cloning and sequencing of the RNA species showed homology to the human *AluSq* subfamily. *DICER1* knockdown increased *Alu* RNA accumulation, and overexpression of exogenous *Alu* in human RPE cells overwhelmed endogenous Dicer and reduced cell viability. Finally, RPE cell viability was rescued by antisense oligonucleotides (AONs) targeting *AluSq*. The authors conclude that *DICER1* is critical for RPE cell survival by silencing cytotoxic *Alu* transcripts, and highlight a new role for *DICER1* in protecting against dsRNA transcribed from the most abundant element in the human genome. The study raises the intriguing questions of whether upregulation of *Alu*-containing RNA polIII-transcribed genes could be pathogenic, or whether *Alu* RNA toxicity plays a role in other tissues and diseases.

Somatic Insertions

Perhaps the most exciting developments in the study of human mobile elements relate to somatic retrotransposition, both in terms of normal variability and disease (Figure 2). A series of three papers from the laboratory of Dr. Fred Gage reported the first evidence of somatic retrotransposition in mammalian tissues and suggested the tantalizing hypothesis that ongoing retrotransposition contributes to somatic mosaicism and phenotypic variability in humans. Muotri *et al.* [33] reported in 2005 that an engineered human L1 can retrotranspose in differentiating rat neural progenitor cells (NPCs) *in vitro* and in the brains of transgenic mice *in vivo*, however this study did not address insertions from endogenous mobile elements. In 2009 Coufal *et al.* [34] went on to show that human fetal brain stem cells and derived neural progenitor cells support retrotransposition of the engineered L1, and used a quantitative multiplexing PCR assay targeting a 3' region of L1 ORF2 to examine endogenous L1 copy number in several tissues. Notably, they observed a subtle yet statistically significant increase of genomic L1 copy number in brain tissues compared to liver and heart. There was variability among individuals and among brain regions, suggesting ongoing somatic retrotransposition during development. Furthermore, in 2010 Muotri *et al.* [35] investigated the relationship between L1 retrotransposition and MeCP2, a methylation-sensitive DNA binding protein frequently mutated in Rett Syndrome (RTT). The authors examined L1 retrotransposition in NPCs derived from induced pluripotent stem cells (iPSCs) from Rett Syndrome patients and controls. The engineered L1 retrotransposed twice as frequently in RTT NPCs than in controls. Genomic qPCR revealed an increase in L1 ORF2 copy number in RTT brain relative to controls, and in brain relative to heart tissue from the same individual. Taken collectively, this series of papers supported the model of L1 driving genetic heterogeneity in neuronal lineages. The missing piece of evidence was then mapping individual retrotransposition events in the resulting heterogeneous cell populations.

Baillie *et al.* [36] address this problem using a novel genomic approach that complements and validates the work from the Gage group, galvanizing the idea that these elements are

mobile in somatic tissues at points beyond early embryonic development and contribute to cellular heterogeneity. The authors developed a high-throughput protocol utilizing hybridization-based capture of L1, *Alu*, and SVA sequences followed by deep sequencing of the junctions, conservative filtering, and mapping to the genome. This assay was performed on genomic DNA from hippocampus and caudate nucleus from three individuals. Putative insertions were classified as somatic if they were in one brain subregion and not the other, absent from all known polymorphism catalogs, and absent from similar data from pooled genomic DNA extracted from blood. PCR amplification and sequencing validated many somatic 5' junctions and 80% of somatic L1 insertions corresponded to the youngest subfamilies, L1-Ta and pre-Ta. This study extended somatic mobilization to *Alu* and SVA elements, and analogous to L1, 83% of somatic *Alu* insertions correspond to the youngest *AluY* subfamily. The observed somatic insertions disproportionately target protein-coding genes, raising the possibility of phenotypic variability between subpopulations of cells harboring different complements of insertions.

To investigate the frequency of somatic retrotransposition in the brain, Evonry *et al.* [40] applied the L1-seq method of Ewing *et al.* [23] to 300 individual neurons from three individuals. They used single-cell sorting followed by multiple displacement amplification (MDA) of the whole genome before L1-seq, controlling for allelic and locus dropout. They estimate <0.6 somatic L1 insertions per neuron with >80% of neurons having no somatic insertions, and conclude that somatic retrotransposition is not an essential source of neuronal diversity in the regions studied. The single somatic L1 whose full length was successfully cloned contained a 5' transduction mapping back to a full length parental L1 that is polymorphic. This study provides a perhaps sobering quantification of somatic retrotransposition in the brain, but does revisit the interesting question of polymorphic "hot" L1s contributing to variation in somatic retrotransposon activity [28].

Somatic Insertions in Cancer

Several groups have reported recently that somatic retrotransposition also occurs in various cancers. The first of these, Iskow *et al.* [24], reported high levels of somatic L1 retrotransposition in lung cancer genomes and detected a hypomethylation signature specific to L1-permissive tumors. Whether the identified insertions are a cause or symptom of neoplasia remains unclear, however there is evidence to suggest that tumor-specific insertions are not merely passenger mutations. Lee *et al.* [37] developed a computational method termed Transposable Element Analyzer (Tea) to detect TEs in whole-genome sequencing data and applied it to data from 43 tumors and matched normal blood samples. They identified somatic insertions in colorectal, prostate, and ovarian tumors with significant variation in TE activity among tumors, and no insertions were observed in multiple myeloma or glioblastoma. TE target genes were significantly enriched for genes frequently mutated in cancer and target genes showed decreased expression compared to controls. The great majority of somatic L1s were significantly truncated upon insertion. Solyom *et al.* [38] performed L1-specific amplification followed by resequencing [23] in 16 colorectal tumors and matched normal colorectal DNA and identified somatic L1 insertions with a high degree of variability among individual tumors. Many of the somatic L1s were PCR validated and 5' and 3' junctions were sequenced. Like Lee *et al.* the majority of somatic L1s were significantly truncated. Notably, this group detected an L1 insertion in one gene identified by Lee *et al.*, Roundabout homolog 2 (*ROBO2*); these findings indicate independently derived insertions in an intron of the gene. This group also observed an interesting positive correlation of number of somatic L1s with patient age. Together, these studies indicate a possible role for somatic retrotransposition in tumorigenesis, however, more specific characterization of target gene insertions sites is required.

Conclusions

The last several years have been exciting times in human transposon biology. New methods such as high throughput sequencing, hybridization-based sequence capture, and repeat-specific computational analysis have allowed for high-resolution detection of RIPs in human genomes and revealed their extensive polymorphism among diverse human populations. Recent work has validated the accumulating evidence for active retrotransposition in several tissues with the potential to affect gene expression, developmental cell fate, and contribute to tumorigenesis in cancer. Topics for further investigation include the unique properties of various tissues that foster permissive environments for retrotransposon mobilization; improved detection and awareness of insertions in exome and whole-genome sequencing studies; and determining the relative contribution of somatic insertions to functional diversity in heterogeneous cell populations. Alternatively they may be merely benign rare events with little to no functional impact.

One can also speculate about the role of RIPs in common complex disease. We consider the “mobilome” an untapped source of human genomic variation in complex disease. We are reminded of the relatively recent realization of extensive copy number polymorphism (CNP) and subsequent association of many CNPs with risk for common disease [39]. In contrast to highly deleterious coding insertions, intronic and intergenic insertions can mildly affect transcript structure and abundance, and such mild changes of gene expression are inherently likely to have modest effect sizes and interact with multiple loci and/or environmental factors in complex disease.

Acknowledgments

We gratefully acknowledge Dr. Haig Kazazian for critical reading of the manuscript. The authors are supported by K08 CA134746 and a Career Award for Medical Scientists from the Burroughs Wellcome Foundation (to KHB).

References

1. Hohjoh H, Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* 1996; 15:630–639. [PubMed: 8599946]
2. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol.* 2006; 13:655–660. [PubMed: 16783376]
3. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell.* 1996; 87:905–916. [PubMed: 8945517]
4. Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science.* 1991; 254:1808–1810. [PubMed: 1722352]
5. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 2002; 21:5899–5910. [PubMed: 12411507]
6. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet.* 2003; 35:41–48. [PubMed: 12897783]
7. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr. Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet.* 2011
8. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet.* 2002; 3:370–379. [PubMed: 11988762]
9. Hancks DC, Kazazian HH Jr. SVA retrotransposons: Evolution and genetic instability. *Semin Cancer Biol.* 2010
10. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol.* 2001; 21:1973–1985. [PubMed: 11238933]
11. Belancio VP, Roy-Engel AM, Deininger P. The impact of multiple splice sites in human L1 elements. *Gene.* 2008; 411:38–45. [PubMed: 18261861]

12. Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* 2005; 15:1073–1078. [PubMed: 16024818]
13. Han JS, Boeke JD. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays.* 2005; 27:775–784. [PubMed: 16015595]
14. Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature.* 2004; 429:268–274. [PubMed: 15152245]
15. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature.* 2006; 441:87–90. [PubMed: 16625209]
16. Thomson SJ, Goh FG, Banks H, Krausgruber T, Kotenko SV, Foxwell BM, Udalova IA. The role of transposable elements in the regulation of IFN-lambda1 gene expression. *Proc Natl Acad Sci U S A.* 2009; 106:11564–11569. [PubMed: 19570999]
17. Romanish MT, Nakamura H, Lai CB, Wang Y, Mager DL. A novel protein isoform of the multicopy human NAIP gene derives from intragenic Alu SINE promoters. *PLoS One.* 2009; 4:e5761. [PubMed: 19488400]
18. Sorek R. The birth of new exons: mechanisms and evolutionary consequences. *Rna.* 2007; 13:1603–1608. [PubMed: 17709368]
19. Lev-Maor G, Sorek R, Shomron N, Ast G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science.* 2003; 300:1288–1291. [PubMed: 12764196]
20. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 2009; 10:691–703. [PubMed: 19763152]
21. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev.* 2012; 22:191–203. [PubMed: 22406018]
- *22. Huang CR, Schneider AM, Lu Y, Niranjana T, Shen P, Robinson MA, Steranka JP, Valle D, Civin CI, Wang T, et al. Mobile interspersed repeats are major structural variants in the human genome. *Cell.* 2010; 141:1171–1182. This study, along with [23, 24, 28], represents a significant leap in our grasp of the amount of genomic variation due to retrotransposon insertions. Microarray technology and second-generation sequencing platforms made these advances possible. [PubMed: 20602999]
- *23. Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* 2010; 20:1262–70. [PubMed: 20488934]
- *24. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell.* 2010; 141:1253–1261. [PubMed: 20603005]
25. Witherspoon DJ, Xing J, Zhang Y, Watkins WS, Batzer MA, Jorde LB. Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. *BMC Genomics.* 2010; 11:410. [PubMed: 20591181]
26. Hormozdiari F, Alkan C, Ventura M, Hajirasouliha I, Malig M, Hach F, Yorukoglu D, Dao P, Bakhshi M, Sahinalp SC, et al. Alu repeat discovery and characterization within human genomes. *Genome Res.* 2010; 21:840–849. [PubMed: 21131385]
27. Stewart C, Kural D, Stromberg MP, Walker JA, Konkel MK, Stutz AM, Urban AE, Grubert F, Lam HY, Lee WP, et al. A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* 2011; 7:e1002236. [PubMed: 21876680]
- *28. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV. LINE-1 retrotransposition activity in human genomes. *Cell.* 2010; 141:1159–1170. [PubMed: 20602998]
29. Taniguchi-Ikeda M, Kobayashi K, Kanagawa M, Yu CC, Mori K, Oda T, Kuga A, Kurahashi H, Akman HO, DiMauro S, et al. Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature.* 2011; 478:127–131. [PubMed: 21979053]
30. Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Iida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa M, et al. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature.* 1998; 394:388–392. [PubMed: 9690476]

- *31. Tucker BA, Scheetz TE, Mullins RF, Deluca AP, Hoffmann JM, Johnston RM, Jacobson SG, Sheffield VC, Stone EM. Exome sequencing and analysis of induced pluripotent stem cells identify the cilia-related gene male germ cell-associated kinase (MAK) as a cause of retinitis pigmentosa. *Proc Natl Acad Sci U S A*. 2011; 108:E569–576. High throughput sequencing studies are performed for many Mendelian disorders. This study provides a cautionary lesson: do not ignore repetitive sequences during data cleaning and analysis. Here the causal mutation, an Alu element, was nearly missed because of blocking agents and sequence trimming. [PubMed: 21825139]
- *32. Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, et al. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature*. 2011; 471:325–330. Investigations of retrotransposons in human disease typically focus on insertional mutagenesis. In contrast, this study describes a cytotoxic effect of an excess of Alu RNA that causes a severe retinal degeneration phenotype. It also suggests a novel miRNA-independent role for DICER in controlling levels of dsRNA in cells by Alu transcript degradation. [PubMed: 21297615]
33. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*. 2005; 435:903–910. [PubMed: 15959507]
34. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. *Nature*. 2009; 460:1127–1131. [PubMed: 19657334]
35. Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH. L1 retrotransposition in neurons is modulated by MeCP2. *Nature*. 2010; 468:443–446. [PubMed: 21085180]
- **36. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011; 479:534–537. Before this study there was mounting evidence that brain tissues supported somatic retrotransposition, however, no endogenous somatic insertions had been mapped outside the context of disease. Using a repeat-specific sequence capture technology, these authors successfully mapped and validated somatic L1, Alu, and SVA insertions. [PubMed: 22037309]
- *37. Lee E, Iskow R, Yang L, Gokcumen O, Hanseley P, Luquette LJ, Lohr JG, Harris CC, Ding L, Wilson RK, et al. Landscape of Somatic Retrotransposition in Human Cancers. *Science*. 2012 In press. This is a landmark survey of retrotransposons in five types of cancer, analyzing whole-genome sequencing data from 43 tumors and matched normal blood DNA. Somatic insertions were found in epithelial tumors but not blood or brain, and insertions were enriched in genes frequently mutated in cancer.
38. 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
39. Wain LV, Armour JA, Tobin MD. Genomic copy number variation, human health, and disease. *Lancet*. 2009; 374:340–350. [PubMed: 19535135]
40. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell*. 2012; 151:483–496. [PubMed: 23101622]

A schematized gene containing three exons (blue) is shown with an intronic L1 insertion (orange) in the sense orientation. Zoomed in, primers for repeat detection and the resulting PCR amplicons are represented (black). These assays use a repeat-specific forward primer with a set of non-specific reverse primers or a universal primer that targets adapter sequence ligated to the genomic DNA. Resulting amplicons contain the 3' end of the repeat sequence along with flanking unique genomic sequence. This complex mixture of amplicons can be labeled and hybridized to a microarray or sequenced. Both methods provide the genomic locations of individual insertions. Intensity data from an array platform can be visualized as peaks (vertical bars), which indicate location and orientation of the insertion. In 2nd generation sequencing, short reads show the precise junction between the insertion (orange) and adjacent genomic sequence of the insertion (blue). The variable length poly-A tract (in parentheses) is not templated, but results from L1 mRNA processing.

Figure 2.

Three contexts for active retrotransposition.

This pedigree shows three examples of retrotransposition events (represented by yellow flashes) in the human genome. First, a novel retrotransposon insertion occurs in a subpopulation of germ cells in the father and is transmitted to the male child, who constitutively harbors the insertion as an inherited variant (represented in green). Second, the male child acquires a somatic insertion in his brain during neuronal development, and this leads to a subpopulation of genetically distinct neurons containing the insertion (orange). Finally, the female child acquires a tumorigenic somatic insertion in the epithelial cells of her colon that promotes clonal expansion in the developing tumor (red).