Allan Award Lecture: On Jumping Fields and "Jumping Genes"

Haig H. Kazazian, Jr.[1,*]

Haig H. Kazazian, Jr.

First, let me thank the American Society of Human Genetics for this great honor. For the 40 years that I have attended the ASHG meeting, I have been impressed with the significance of the Allan Award in honoring research accomplishment in human genetics. This is a very important day for me, and I greatly appreciate your recognition.

Before going further, I'd like to say that there is someone we all miss very much at this meeting, Victor McKusick. He did so much for the field of human genetics and for our society. Victor meant a lot to me during my 25 years at Johns Hopkins, both personally and professionally. His intellect was extraordinary, but he never flaunted it. What a wonderful man he was!

Now, I'd like to tell you about my journey in human genetics, hitting the high points and providing a few anecdotes and philosophical comments. I'll also stress the role of teamwork and interaction with colleagues in the research process. I'll end by discussing what has gotten me excited in the lab presently. In doing research, I've tried to work on important problems, yet ones that were solvable using available methods. I've also tried not to follow the crowd, but in the words of Robert Frost to take "the road less traveled by, And that has made all the difference." Every researcher is a gambler, but the smart gambler likes odds that are not too long. In research, problems with odds of success about 5 or 10:1 are good problems. If the odds are 2:1, the problem is probably not so important. If the odds are 100:1, the problem is too risky. I'm going to tell you about my research course starting from globin regulation in the thalassemias, then turning to "jumping genes," and how even though those subjects seem quite disparate, the course followed a logical route. Because of time constraints, I won't speak about other projects, namely, cystic fibrosis work by Garry Cutting, schizophrenia research with Ann Pulver, characterization of hemophilia A mutations with Stelios Antonarakis, and gene therapy of hemophilia A in the mouse and dog models by Rita Sarkar and Denise Sabatino.

My first research experience was at Dartmouth Medical School in the summers of 1958 to 1960. Before starting medical school after my junior year in 1958, I asked the Dean whether I should work as a hospital orderly. He told me to spend the summer doing biochemical research with Lafayette Noda on rabbit creatine kinase. The following two summers I worked with Lucille Smith on bacterial electron transport. However, after transferring to Hopkins for the last 2 years of medical school, I remained unconvinced that research was my career path. I enjoyed an elective in child psychiatry, and decided to train in pediatrics.

As a fourth year medical student in 1962, I took an elective taught by Barton Childs in genetics. Parenthetically, this was the only time that 8 week course was given. Each of six students and two faculty gave a seminar on a literature topic. I remember telling Barton in his office that I knew nothing about genetics, and he said, "Don't worry, you'll learn." My seminar was on the cytogenetic abnormalities that were being published almost weekly at that time, and I got hooked on human genetics. At that time, I also got hooked on Lilli, my loving wife of many years. After an internship in pediatrics at the University of Minnesota, I decided to do a genetics fellowship after only one year of residency. I wrote Barton asking him to suggest places to train. He mentioned a few places including Hopkins, where he had recently obtained an NIH training grant. After reading the ground-breaking Davidson, Nitowsky, and Childs paper[1] in PNAS demonstrating the validity of the Lyon hypothesis in humans, I decided to train with Dr. Childs.

[1]Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
*Correspondence: kazazian@mail.med.upenn.edu
DOI 10.1016/j.ajhg.2009.01.003.

At Hopkins, Childs suggested that I study genetic variation in fruit flies in Bill Young's lab, and we were successful in demonstrating that X chromosome dosage compensation in flies does not proceed by the same X inactivation mechanism used by mammals.[2] That was the gist of my first paper in 1965.

After 20 months at Hopkins and with the U.S. Army breathing down my neck, I joined the U.S. Public Health Service at the NIH. Luckily, I found a position with Harvey Itano largely due to a recommendation from my Dartmouth mentor, Lafayette Noda. Note the Japanese-American connection. With Itano, I did research on the regulation of globin synthesis in humans. Then Bob Cooke, Chair of Pediatrics at Hopkins, offered me a faculty position, but I asked Dr. Cooke to let me finish a third year of pediatric training at Hopkins first. During elective time in that residency year, I wrote my first successful RO1 application on globin gene regulation. So I had funding when I joined the faculty in 1969.

I started working on separating the alpha and beta globin messenger RNAs in a small operation, a technician and me. However, I occupied the lab designated for Childs, so the space was ample. I knew that in order to succeed competing against larger, smarter labs I would need to spend most of my time on research. My goal was to emulate the high-quality research in the basic sciences at Hopkins.

Shortly thereafter, Mike Kaback started his Tay-Sachs screening program in the Baltimore Jewish community.[3] He suggested that if adult hemoglobin were produced in the fetus, one might similarly be able to do carrier screening and prenatal diagnosis for sickle cell anemia. I was keen to look into this, and Morley Hollenberg, an MD student with an Oxford DPhil, quickly found hemoglobin A synthesis in the midtrimester fetus using techniques that I had learned at NIH with Itano.[4] Prenatal detection of hemoglobin A was my entree into the prenatal diagnosis field.

In the early 1970s I continued to work on the mRNA imbalance in the alpha- and beta-thalassemias. In 1975, John Phillips, a former chief resident at Boston Children's, joined the lab. John did lovely studies of normal globin mRNA ratios,[5] but his breakthrough came through a fortuitous chain of events.

In early 1978, Ned Boyer, a senior faculty in McKusick's Division of Medical Genetics, sent Alan Scott, a postdoc, to Alec Jeffreys's lab in England to learn Southern blotting. After Alan returned, Phillips suggested and I agreed that he learn Southern blotting from Alan. Y.W. Kan had recently found a HpaI polymorphic site 3′ of the beta-globin gene, and demonstrated its usefulness in prenatal diagnosis of sickle cell anemia.[6] Soon thereafter, Jeffreys found two polymorphic HindIII sites in the gamma-globin genes.[7] Phillips used Southern blotting to find extended linkage disequilibrium involving the Beta S mutation and the HpaI and HindIII sites.[8] Out of eight possibilities for the three sites, there were only four varieties of Beta S-bearing chromosomes. The 60% with the HpaI site lacked both HindIII sites, whereas the 40% without the HpaI site usually contained one HindIII site. This setup allowed an increase in precise prenatal detection of sickle cell anemia by linkage analysis from 60% to 85%.[8] This paper was crucial to the thinking that went into our later work characterizing mutations in beta-thalassemia. In 1979, we hired Corinne Boehm to run our molecular diagnosis program, which consisted only of prenatal diagnosis of sickle cell anemia by linkage analysis.

In July 1980, Antonarakis entered the picture. He had written me asking for a genetics training position. I had no extra money to pay him, but he said he'd come without pay. His letter languished on my desk until another colleague, George Stamatoyannopoulos, called from Seattle to push for Antonarakis. George was very persuasive, so I hired Stelios without pay. Once he arrived, he got $5000 from the Greek church in Baltimore. On Christmas day 1980, at a family get-together in Detroit, I received a call from the Armenian industrialist and family friend I had solicited. He said he'd provide another $5000 for Antonarakis. So Stelios was paid something.

Antonarakis used Southern blotting to find more RFLPs in the beta globin cluster. He found three new HincII site polymorphisms in the epsilon and pseudo-beta regions. Meanwhile, Kan had found a polymorphic BamHI site 3′ of the beta gene.[9] Then Stelios found a polymorphic AvaII site in the beta gene itself. Parenthetically, these RFLPs became SNPs 20 years later. In studying these polymorphic sites in families, there emerged an interesting pattern. There was clear linkage disequilibrium of the five sites 5′ to the delta gene (only three common polymorphic patterns were present out of a potential 32). There was also linkage disequilibrium for the AvaII and BamHI sites in and 3′ of the beta gene (three patterns out of a predicted four).[10] However, the two regions of disequilibrium combined together randomly, i.e., the three 5′ patterns randomly associated with the three 3′ patterns. This suggested that within the intervening ~12 kb was a recombination hot spot.[10] Indeed, a number of recombination events have been seen in this hot-spot region (Figure 1).

One day in 1981, I asked the lab what to call the pattern of polymorphic restriction sites on a chromosome. Corinne Boehm suggested the term "haplotype" after the usage for HLA haplotypes coined a few years earlier by Ceppelini. "Haplotype" in the DNA context appeared for the first time in January 1982.[10] Later, we found a polymorphic TaqI site upstream of the epsilon gene that was not in linkage disequilibrium with the sites in the beta globin cluster, and thus the region extending from the epsilon gene to the delta gene constituted the first "haplotype block" as analyzed by Chakravarti.[11] This was my introduction to genomic analysis to which I later returned in studies of mobile DNA.

In 1980, we used haplotypes to carry out prenatal diagnosis of beta-thalassemia in families with a previously affected child.[12] However, we wondered whether the lesson of restriction-site polymorphisms and their linkage disequilibrium with the Beta S mutation would carry over to beta-thalassemia. By this time, three labs had

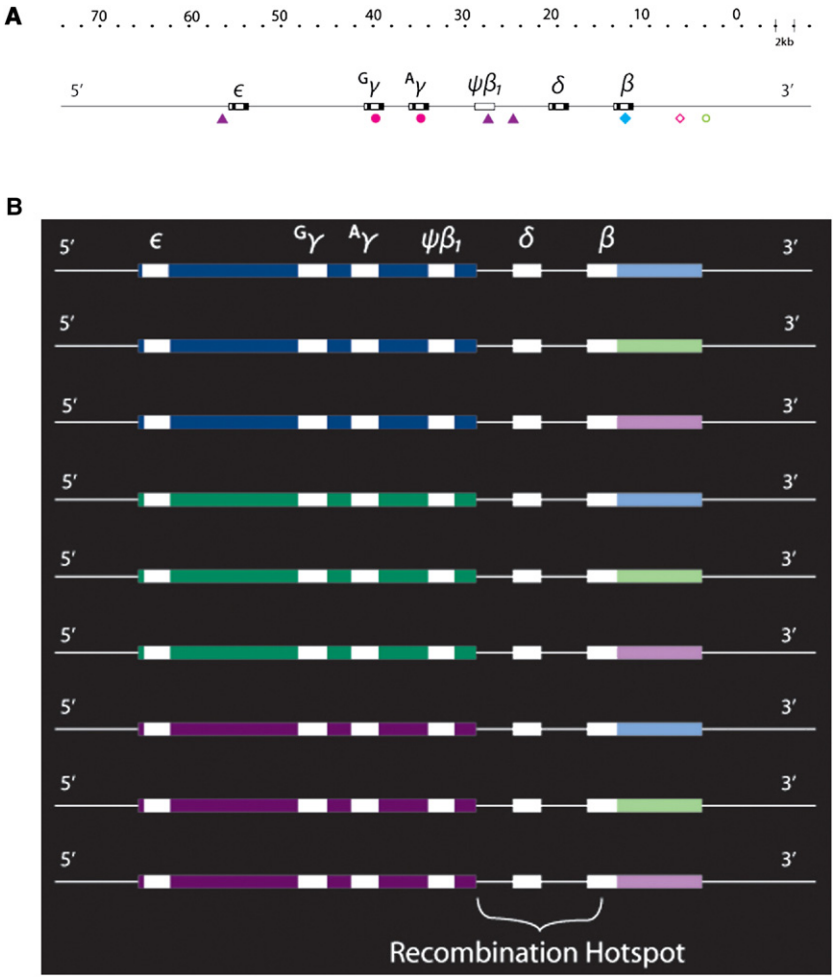

Figure 1. Linkage Disequilibrium in the β-globin Gene Cluster with a Recombination Hot Spot

(A) The β-globin gene cluster with seven polymorphic restriction sites is shown: HincII (▲), HindIII (⬟), AvaII (◆), HpaI (◇), and BamHI (▢).

(B) The five 5′ polymorphic sites from the epsilon gene to the pseudo-beta gene (HincII and HindIII) are in linkzage disequilibrium because of 32 expected haplotypes (2^5), only three are commonly present (shown in dark blue, green, and purple). Likewise, the two 3′ polymorphic sites from the beta gene and 3′ to it (AvaII and BamHI) are also in linkage disequilibrium because of four expected haplotypes; only three are found in human populations (shown in lighter shades of blue, green, and purple). However, because the 5′ and 3′ haplotypes recombine randomly, there is a ~12 kb region between them that is a "recombination hot spot."

sequenced six beta-thalassemia genes, but only two different mutations had been found, each one three times. Since we had few beta-thalassemia patients at Hopkins, I called Pat Giardina, Director of the Thalassemia Clinic at New York Hospital. She had a large patient population, and could get blood samples on families. To do haplotype analysis, we needed blood on trios, the patient and both parents. Amazingly, within 1 month Giardina had sent blood on some 30 trios which the lab quickly haplotyped.

Around this time I contacted Stuart Orkin, a friend and globin colleague, at Boston Children's. Stu was interested in sequencing beta-thalassemia genes. I thought that perhaps different normal chromosome backgrounds were hit by different mutations (Figure 2), and because the beta-globin cluster was so small there had been insufficient time in generations for the mutations to move to different haplotypes by recombination. Thus, we might have a neat analytical way to characterize all the common mutations causing beta-thalassemia. I suggested to Stu that I send him specific beta-thalassemia DNAs each containing beta-thalassemia mutations on different haplotypes. After sending the first three DNAs to Boston, I remember Stu's call a few weeks later. He had sequenced three beta genes of interest, and they all had different mutations that had not been seen previously. I nearly fainted from the excitement! We quickly finished cloning and sequencing the beta-thalassemia gene from nine different haplotypes in Mediterranean patients, finding eight different mutations.[13] At this time, Orkin and I were communicating new information daily by phone. Stu came to Baltimore and stayed at our home as we finished the paper. Later, we used haplotyping followed by cloning and sequencing at Hopkins to characterize the common beta-thalassemia alleles in Asian Indians, Chinese, and African-American patients.[14–16] The procedure of sequencing beta-thalassemia genes from different haplotypes worked well. A new allele was discovered 80% of the time. We characterized roughly 40 alleles that acted at a variety of steps in gene action: transcription, RNA splicing, RNA processing, translation, and protein stability (Figure 3). In 1984, Orkin and I reviewed the topic.[17] Beta-thalassemia was the first disorder caused by multiple mutations that was essentially completely characterized at the molecular level.

However, it looked like the beta-thalassemia project would soon become humdrum as thalassemia alleles were characterized in different ethnic groups. Orkin and I discussed this problem at the 1983 ASHG meeting in Newport Beach, Virginia. He had found his future niche in transcription factors in hematopoiesis beginning with GATA-1. I was still looking for mine, and it would come, but not until the summer of 1987.

In 1985, there was another lucky occurrence. Henry Erlich and Norm Arnheim wanted me to consult for Cetus in the East Bay, but I declined because the travel distance was too great and I foolishly thought I wouldn't learn

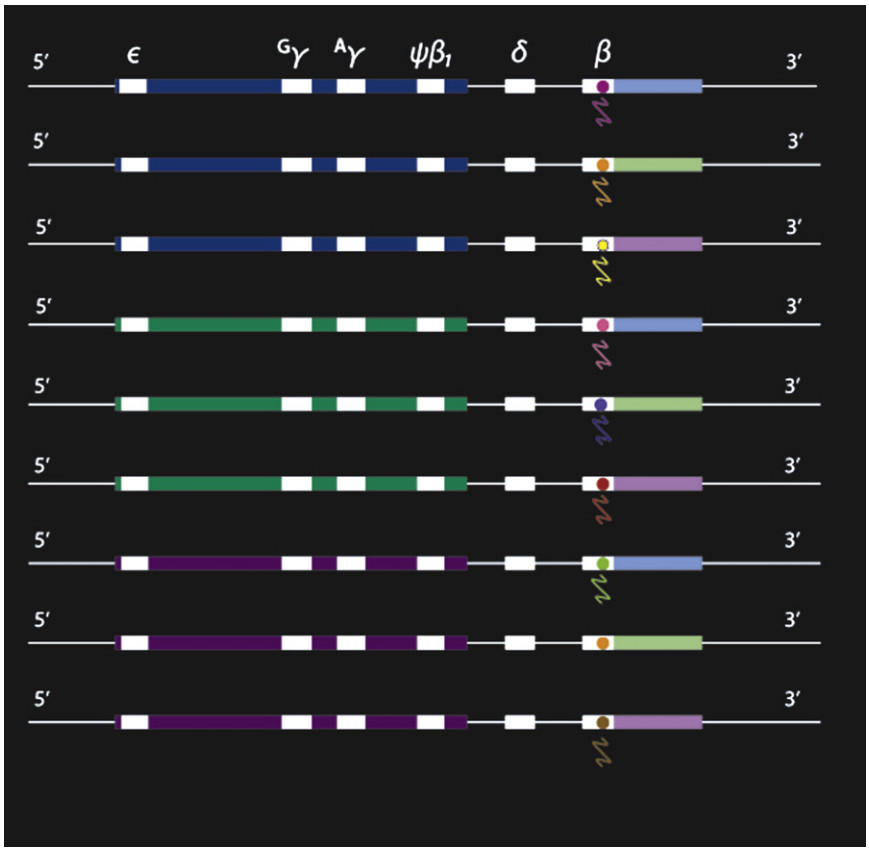

Figure 2. Different β-Thalassemia Mutations on Different β-Globin Haplotypes

The nine different 5′ plus 3′ combined haplotypes have been hit by eight different β-thalassemia mutations. The occurrence of different mutations on the different haplotypes is shown by different colored lightning bolts striking the β-globin genes. A recombination event between haplotypes II and VIII in the recombination hotspot has led to the same mutation on both of these haplotypes.

anything new at the company. Boy, was I wrong! In 1985, our daughter, Sonya, decided to attend Stanford, and it became convenient to consult at Cetus when visiting Palo Alto. In September 1985, Lilli and I took Sonya to Stanford, and while at Cetus Henry and Norm ushered me into a small conference room to hear a wild idea from Kerry Mullis. This was my introduction to PCR, but although Randy Saiki had made some progress, the idea hadn't proceeded very far. Over the following year, we had little success in the lab with PCR. But Lilli talked me into returning to Stanford for sophomore parent's weekend in October, 1986. Now upon entering Erlich's office, Henry's excitement was palpable. David Gelfand had cloned Taq polymerase and the resultant PCR bands were incredible. We then used PCR with Taq polymerase to sequence beta-thalassemia genes directly from leukocyte DNA to discover mutations.[18] From 1982 to 1987 haplotype analysis had aided mutation discovery in beta globin, alpha globin, phenyalanine hydroxylase,

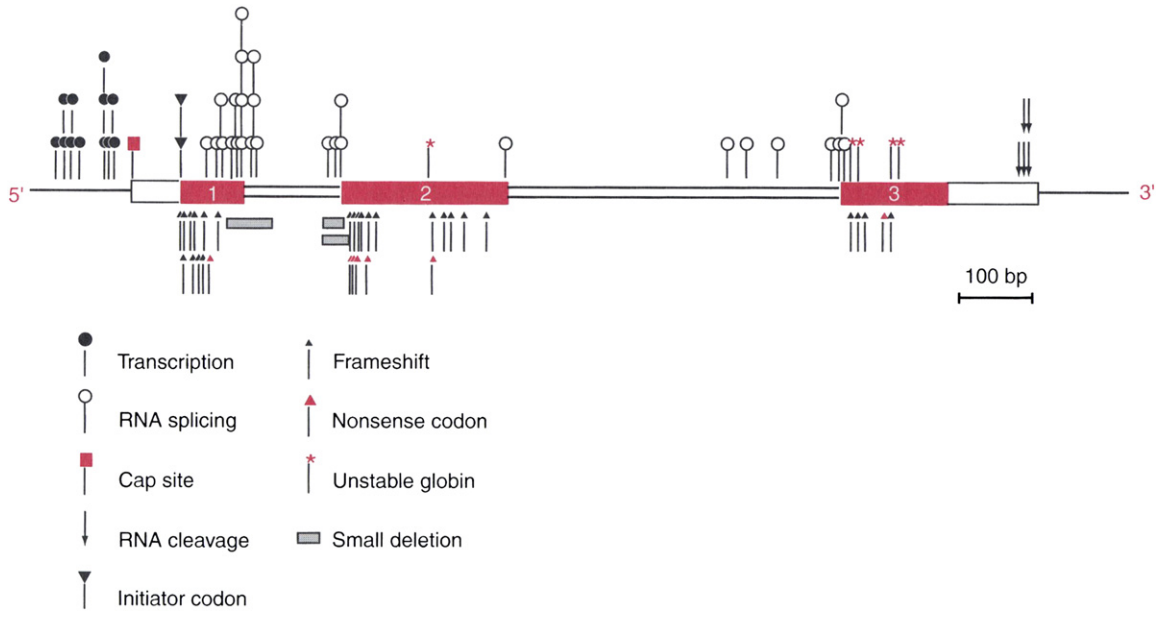

Figure 3. β-Thalassemia Mutations in Different Steps in Gene Action

β-thalassemia mutations in different steps in gene action, including transcription, RNA processing (Cap and polyadenylation), RNA splicing, translation (initiation codon, nonsense codons, and frameshifts), and protein stability.

and LDL receptor genes. Now it was passé! PCR could be used to sequence mutations directly from leukocyte DNA in any gene whose sequence was known.

Meanwhile, in 1984, to explore the full spectrum of mutations causing disease, we wanted to work on a large gene, whose mutations unlike those in beta-globin were not under selection, and where nearly every unrelated individual probably had a different mutation. The ideal gene appeared when Gitscher and colleagues at Genentech and Wozney, Toole, and colleagues at GI cloned the factor VIII gene that is mutated in hemophilia A.[19,20] Here was the perfect gene, very large (nearly 200 kb) and X-linked, and Haldane had predicted in 1938 that every unrelated affected male would have a different mutation (Of course, the recurrent FVIII inversion found later by Gitscher made that prediction incorrect.[21]). In 1984, Antonarakis and I traveled to GI in Boston, and obtained the FVIII cDNA for mutation analysis.

In May 1987, Hagop Youssoufian, a genetics fellow, made a key discovery. He had been characterizing FVIII mutations in 240 hemophilia A patients by Southern blotting using FVIII cDNA probes. Two patients had abnormal restriction fragments, suggesting that their mutations were insertions. Although Hagop was leaving in July, he persuaded me to let him clone a portion of one insertion. One Monday morning, Hagop told me that over the weekend he had succeeded, blotted the clone with L1 and Alu probes, and found the insertion was part of an L1 repetitive element. Immediately, I thought this is the problem that I've been looking for. L1 is an insertional mutagen that can jump into genes, disrupt them, and cause disease. "I'm going to work on human transposable elements."

At the time, I knew something about L1s. At Hopkins and Cold Spring Harbor, I had heard Maxine Singer speak on repeat sequences in human DNA. Moreover, Allan Scott had joined the faculty in Pediatric Genetics, and after cloning and sequencing a number of human L1s, he had built an important consensus sequence for the full-length 6 kb L1.[22] Scott's consensus sequence had two open reading frames, called ORF1 and ORF2. ORF2 contained a sequence that when translated had similarity to the reverse transcriptases of retroviruses. It was thought that these sequences were retrotransposons, elements that could duplicate themselves through reverse transcription of an RNA intermediate.

After Youssoufian left, Corinne Wong took up the project of characterizing the two L1 insertions, both of which were 5′ truncated. Meanwhile, at Singer's suggestion we found that the inserted L1s came from a special subfamily of transcribed L1 elements. Shown in Figure 4 are the parents and the young man (patient JH-27) with hemophilia A who had an L1 insertion. In Figure 5 is the 3.8 kb insertion of a 5′ truncated L1 into exon 14 of JH-27's factor VIII gene. Note the insertion was not present in his mother, meaning that it occurred either in her germ cells or early in his development. We published this paper in early 1988[23] and I entered the "black hole" of transposable elements.

Figure 4. Hemophilia A Patient JH-27 and His Parents at His Bar Mitzvah
Shown with parental permission.

In early 1988, Beth Dombroski joined the lab as a postdoc, and we decided to try to isolate the full-length 6 kb L1 precursor of the JH-27 insertion. Our premise was that the precursor would have the identical sequence as the insertion over the 3.8 kb of the insertion, and would be an active mobile element. Of course, if I had known more about mobile elements, I would have realized that this premise depended on the element acting in *cis*, i.e., mobilizing a copy of itself into a new genomic site, not in *trans*, i.e., mobilizing a copy of another element as is the case for nearly every other transposable element. Thus, with more experience in the field I may not have done the experiment. I compared Alan Scott's L1 consensus sequence with the sequence of the 3.8 kb JH-27 insertion, and found a region of 20 nucleotides (nts) at roughly nt. 4200 in the 6 kb L1 consensus sequence that differed from the insertion by 3 nts. We made an oligonucleotide (JH-27) containing the 20 nts from the insertion sequence, and Beth carried out the blot against DNA digested with BamHI from JH-27, his parents, and two controls. The result was amazing! We thought then that the human genome contained about 100,000 L1 sequences, but now we know that the number is greater than 500,000. So it was a real surprise when Beth found that instead of a smear of thousands of L1 sequences, each individual had only a handful of bands and the patient had a new band not present in his parents (Figure 6). This

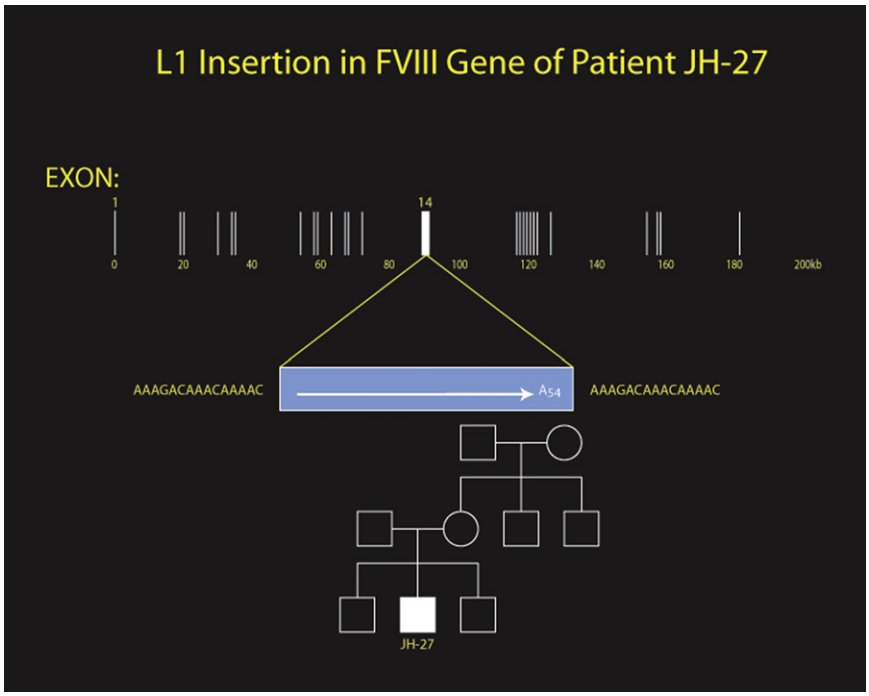

Figure 5. L1 Insertion into Exon 14 of the Factor VIII Gene in Patient JH-27
The 3.8 kb 5′ truncated insertion ends in a poly A tract and is bracketed by short target-site duplications. In the pedigree, the insertion is not seen in the mother so it occurred either in one of her germ cells or in early embryogenesis of JH-27.

band represented the insertion! Dombroski then obtained a commercial lambda genomic library and cloned four full-length L1s that hybridized with the JH-27 oligonucleotide and corresponded to four of the bands on the gel. Upon checking sequence of these four L1s, the second looked promising.

With 5500 nts out of 6000 in hand, we had a perfect match with the insertion sequence (Figure 7). However, in those last 500 nts, Beth found two changes from the JH-27 insertion. When I saw those changes on the sequencing run, my first thought (optimistically) was that perhaps we had isolated an allele of the real precursor from the commercial library. To check that possibility, Beth got genomic sequence flanking the L1, did a PCR from the JH-27 oligonucleotide to the flank in both parents, and sequenced the region of the two changes from her PCR product. Voila! The parental sequences matched the insertion—the changes were not present in either parent and the allele hypothesis was probably correct. We then got blood from the mother, and Beth cloned the same L1 from the mother's genomic library. The mother's L1 was indeed identical in sequence over its last 3.8 kb to the insertion. It was also the first L1 isolated that contained two intact ORFs. It had originated on chromosome 22 and a portion of it had been copied and inserted into the factor VIII gene in JH-27. We had successfully obtained the precursor from among hundreds of thousands of L1s, and had strong evidence for *cis* preference of L1 elements.

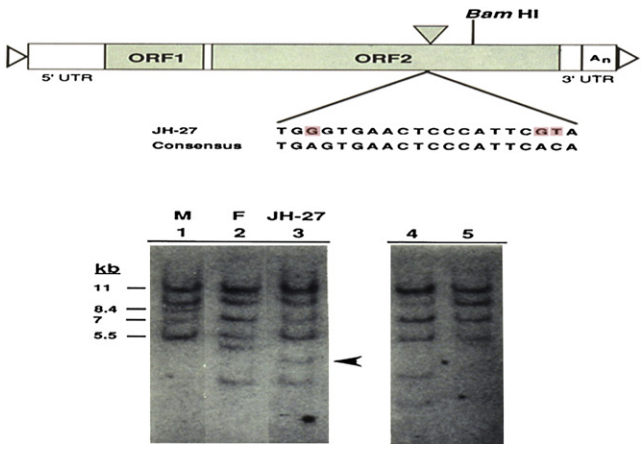

Figure 6. Hybridization of JH-27 Oligonucleotide to BamHI Digests of DNA from JH-27, His Parents, and Two Controls
The JH-27 oligonucleotide contains three sequence changes (pink boxes) from the L1 consensus sequence over 20 nts. Note band (arrow) in JH-27 not seen in either parent, representing the L1 insertion. From Dombroski et al., 1991.[24] Reprinted with permission from AAAS.

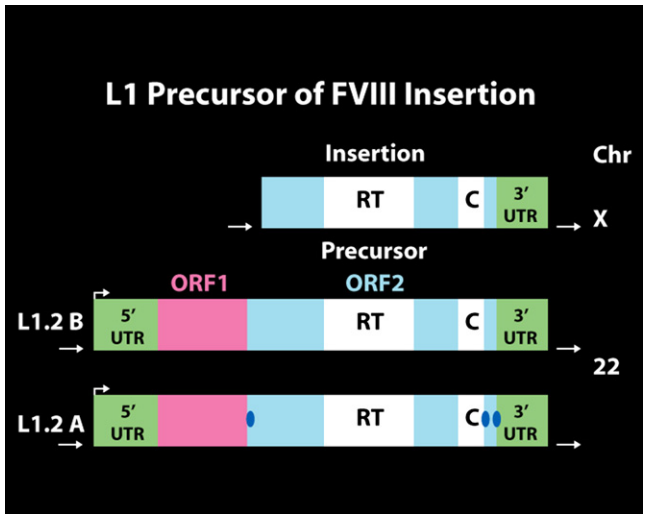

Figure 7. Isolation of the Full-Length L1 Precursor to the JH-27 Insertion
L1.2A has two nucleotide changes from the insertion near its 3′ end (blue ovals) and was isolated from a commercial genomic library. However, L1.2B was isolated from a library made with the mother's DNA and matches the insertion sequences over its entire 3.8 kb. It is the precursor to the insertion in JH-27 and is located on chromosome 22q.

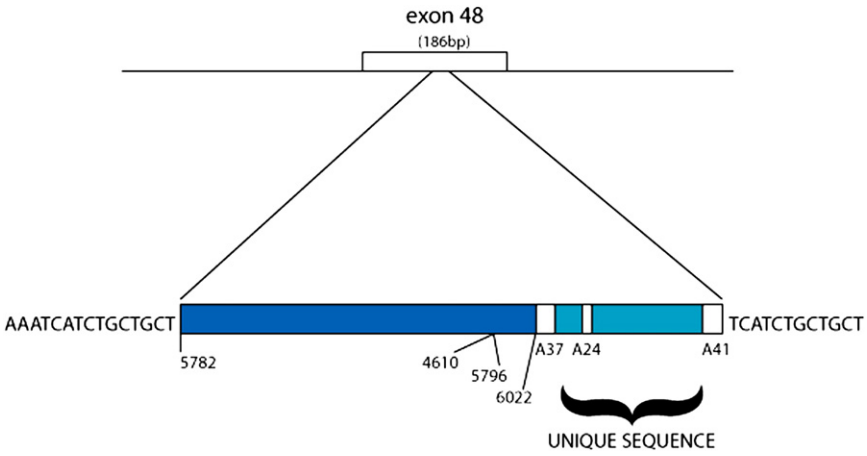

Figure 8. L1 Insertion into an Exon of the Dystrophin Gene Containing an Extra 500 Nucleotides 3′ to the L1
This insertion disrupted expression of the dystrophin gene, leading to Duchenne muscular dystrophy. The extra 3′ nts resulted from transcription read-through of the L1 and cleavage of the transcript only after a downstream poly A signal. The extra unique sequence was then used to clone the precursor of this L1 insertion.

Then Abram Gabriel, an MD postdoc at Hopkins, suggested an assay for reverse transcriptase activity using a hybrid yeast Ty1-human L1 combination. Steve Mathias, a PhD student with Alan Scott, working with Abram and Jef Boeke, the yeast retrotransposon expert at Hopkins, demonstrated reverse transcriptase activity encoded by ORF2 and the two papers were published together.[24,25] All this time, I was learning more about transposable elements through quarterly lab meetings with Maxine Singer.

In 1993, Susan Holmes, a grad student, worked up another interesting patient sample from the DNA diagnostic lab, a Duchenne patient with an L1 insertion knocking out his dystrophin gene. The insertion into an exon was unusual because it contained 500 nts of single-copy sequence 3′ to the L1 (Figure 8), which Susan used to clone the full-length precursor on chromosome 1q.[26] We know now that 20% of L1 insertions contain extra 3′ sequence because the L1 poly A signal is weak and cleavage of the L1 RNA often occurs only after a downstream poly A signal. Later, John Moran showed that this mechanism, called 3′ transduction, can be used by L1s to shuffle exons.[27] This is just one of a host of mechanisms by which L1 retrotransposons drive genome evolution.[28]

In 1994, I moved to Penn, and Moran joined the lab as a postdoc with one goal—to develop a tissue culture assay for L1 retrotransposition. In 6 months, he had put a retrotransposition cassette containing an antisense *neo* gene disrupted by a forward intron into the 3′ UTR of our insertion precursor and had the assay working in HeLa cells beyond our wildest dreams (Figure 9). In this assay, the only way to get G418 resistant colonies is by a retrotransposition event. I remember telling Moran that all those G418-resistant colonies (up to 1 colony per 10–20 plated cells) couldn't possibly represent retrotransposition events. But they did! Donna Sassaman found that many other potentially active L1s we isolated also retrotransposed in culture. Moran showed that both ORFs were critical for retrotransposition, and we helped Jef Boeke demonstrate that L1 ORF2 encodes an endonuclease activity critical for L1 mobility.[29–31]

Then, Thierry Nass and Ralph DeBerardinis demonstrated that disease-causing, full-length mouse L1s are also active in the assay.[32] John Goodier found that many mouse L1s from three different L1 subfamilies are active, suggesting that there are about 3000 active L1s in the mouse.[33] In 2003, Brook Brouha isolated nearly all the potential active L1s from the reference human genome and found that approximately half are active, indicating that the average human diploid genome has 80–100 active L1s, a small number compared to 3000 in the mouse.[34] He also found that only a handful of these L1s in any individual are very active or "hot," but that "hot" L1s account for nearly all of the disease-causing insertions in humans. Five out of six disease-causing, full-length L1s were "hot." In 2006, Seleme showed that there is extensive individual variation in retrotransposition capability among human beings.[35]

Meanwhile, in 2002 Eric Ostertag made a number of transgenic mouse models of L1 retrotransposition using "hot" human L1s containing heterologous promoters as transgenes.[36] Daria Babushok characterized a large number of L1 insertions in transgenic mice, showing that they are very similar to the endogenous insertions found in humans.[37]

This has led to recent exciting work by Hiroki Kano on the timing of retrotransposition. The field has believed that nearly all retrotransposition occurs in germ cells or sufficiently early in the embryo so that the insertions are heritable. After all, the human genome contains over 500,000 L1s. But just because so many insertions are present in the genome does not mean that present-day insertions are all heritable. Kano used endogenous human L1s and endogenous mouse L1s both with only their own promoters as transgenes in mice and rats. The readout for retrotransposition was removal of an intron from the 3′ UTR of the L1. He made two very surprising findings that were observed for all transgenic lines. First, he found L1 RNA can be carried over from the germ cell, either sperm or ovum, into the embryo. Moreover, insertion can occur from carried over RNA in the morula or blastocyst. The second surprising finding was that most L1 retrotransposition occurs not in germ cells even though the L1 RNA is present, but in early embryogenesis. L1 RNA is abundant in spermatogenic fractions, but very little

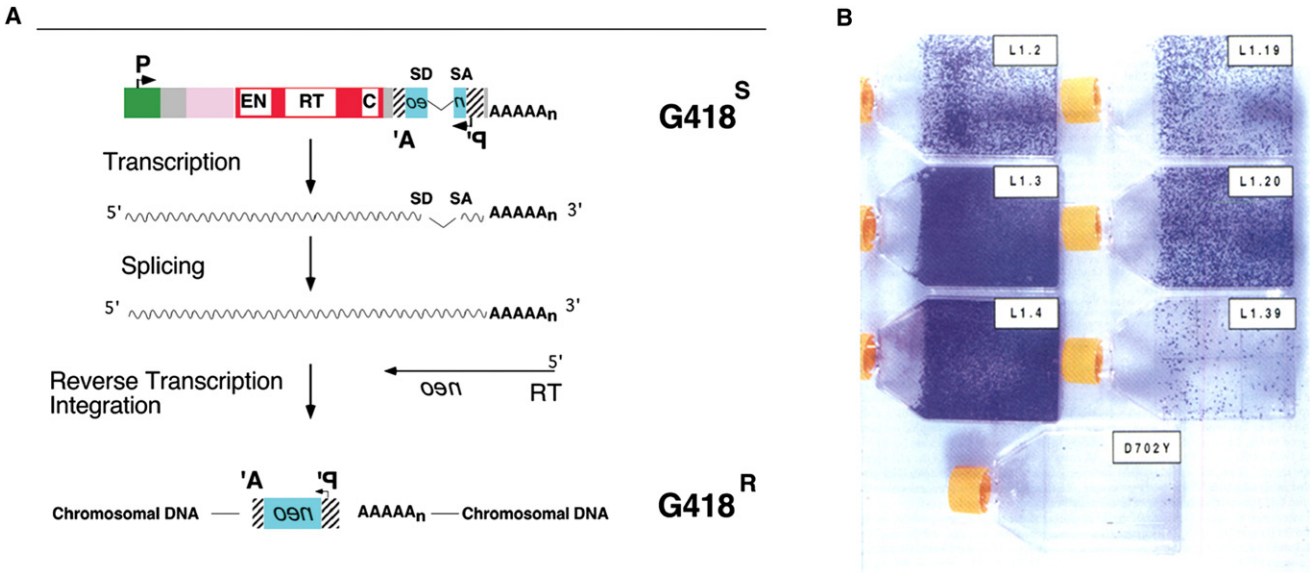

Figure 9. Assay for Retrotransposition in Cultured Cells

(A) A retrotransposition cassette containing a backward *neo* gene disrupted by a forward γ-globin intron is placed into the 3′ UTR of an active L1. This marked L1 is then cloned into a pCEP4 expression vector that is transfected into transformed human cells, e.g., HeLa cells. The only way to get $G418^R$ colonies is by transcription from the L1 promoter, splicing out of the intron, reverse transcription, and integration of the now-intact *neo* gene.

(B) Results of retrotransposition assays for different human L1s. Each blue-stained colony represents a clone of cells with a retrotransposition event. The D702Y mutation affects reverse transcriptase activity of L1.2 and causes a major reduction in retrotransposition activity. Note the variability in the retrotransposition activity of various active human L1s.

L1 insertion is observed in those fractions. However, there is considerable retrotransposition in the preimplantation embryo even though there is much less total DNA in the sample (H. Kano, I. Godoy, C. Courtney, M.R. Vetter, G.L. Gerton, E.M. Ostertag, and H.H.K, unpublished data).

These results suggest a posttranscriptional block for insertion in germ cells and the very early embryo. Although there is abundant L1 RNA in germ cells, very little actually retrotransposes, but this block is lifted by the morula stage (around the 100-cell embryo). The results suggest that we are mosaic for L1 insertions in our tissues, and that we may contain a very large number of somatic insertions at low frequencies per cell. Insertions in important genes could lead to susceptibility to common diseases, including oncogenic and neurologic conditions. We are exploring these possibilities using new high-throughput sequencing methods. There are still a number of unanswered questions in the retrotransposition field, some of which include:

What is the frequency of both heritable and nonheritable L1 retrotransposition from one generation to the next?

What are the genetic and epigenetic factors that lead to individual differences in retrotransposition capability?

What host factors are critical for retrotransposition?

We know that retrotransposition events have caused ~65 single gene disorders. How extensive is their role in common, multifactorial disease?

As you can tell, progress in human genetics takes an army. I have been blessed with many outstanding fellows, graduate students, postdocs, and research assistants over the years at Hopkins and Penn. I have had outstanding colleagues at these institutions and elsewhere. I thank them all for their many contributions. It's been a wonderful journey that isn't over yet! I particularly thank Barton Childs, my mentor for many years. He always said, "You've got to burn to do research." You need very high motivation! I especially thank my wife, Lilli, and children, Haig and Sonya, for putting up with me all these years. Thanks again to the American Society of Human Genetics for this wonderful honor and to all of you for listening.

Acknowledgments

I thank Sarah Robinson for assistance in preparing the manuscript and figures. The work was supported by grants from the NIH, a grant from the Penn Genomics Frontiers Insititute and a SRA from Transposagen Biopharmaceuticals, Inc.

References

1. Davidson, R.G., Nitowski, H.M., and Childs, B. (1963). Demonstration of two populations of cells in the human female heterozygous for glucose=6 phosphate dehydrogenase variants. Proc. Natl. Acad. Sci. USA *50*, 481–485.
2. Kazazian, H.H., Jr., Young, W.J., and Childs, B. (1965). X-linked 6-phosphogluconate dehydrogenase in Drosophila: Subunit associations. Science *150*, 1601–1602.

3. Kaback, M., Lim-Steele, J., Dabholkar, D., Brown, D., Levy, N., and Zeiger, K. (1993). Tay-Sachs disease–carrier screening, prenatal diagnosis, and the molecular era. An international perspective, 1970 to 1993. The International TSD Data Collection Network. JAMA *270*, 2307–2315.
4. Hollenberg, M.D., Kaback, M.M., and Kazazian, H.H., Jr. (1971). Adult hemoglobin synthesis by reticulocytes from the human fetus at midtrimester. Science *174*, 698–702.
5. Phillips, J.A., III, Snyder, P.G., and Kazazian, H.H., Jr. (1977). Ratios of α to β globin mRNA and regulation of globin synthesis in reticulocytes. Nature *269*, 442–445.
6. Kan, Y.W., and Dozy, A.M. (1978). Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic-fluid cells. Lancet *2*, 910–912.
7. Jeffreys, A.J. (1979). DNA sequence variants in the G gamma-, A gamma-, delta- and beta-globin genes of man. Cell *18*, 1–10.
8. Phillips, J.A., III, Panny, S.R., Kazazian, H.H., Jr., Boehm, C.D., Scott, A.F., and Smith, K.D. (1980). Prenatal diagnosis of sickle cell anemia by restriction endonuclease analysis: Hind III polymorphisms in γ-globin genes extend test applicability. Proc. Natl. Acad. Sci. USA *77*, 2853–2856.
9. Kan, Y.W., Lee, K.Y., Furbetta, M., Angius, A., and Cao, A. (1980). Polymorphism of DNA sequence in the beta-globin gene region. Application to prenatal diagnosis of beta 0 thalassemia in Sardinia. N. Engl. J. Med. *302*, 185–188.
10. Antonarakis, S.E., Boehm, C.D., Giardina, P.J.V., and Kazazian, H.H., Jr. (1982). Non-random association of polymorphic restriction sites in the β-globin gene cluster. Proc. Natl. Acad. Sci. USA *79*, 137–141.
11. Chakravarti, A., Buetow, K.H., Antonarakis, S.E., Waber, P.G., Boehm, C.D., and Kazazian, H.H., Jr. (1984). Non-uniform recombination within the human β-globin gene cluster. Am. J. Hum. Genet. *36*, 1239–1284.
12. Kazazian, H.H., Jr., Phillips, J.A., III, Boehm, C.D., Vik, T., Mahoney, M.J., and Ritchey, A.K. (1980). Prenatal diagnosis of β-thalassemia by amniocentesis: Linkage analysis of multiple polymorphic restriction endonuclease sites. Blood *56*, 926–930.
13. Orkin, S.H., Kazazian, H.H., Jr., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G., and Giardina, P.J.V. (1982). Linkage of β-thalassemia mutations and β-globin gene polymorphisms with DNA polymorphisms in the human β-globin gene cluster. Nature *296*, 627–631.
14. Kazazian, H.H., Jr., Orkin, S.H., Antonarakis, S.E., Sexton, J.P., Boehm, C.D., Goff, S.C., and Waber, P.G. (1984). Molecular characterization of seven β-thalassemia mutations in Asian Indians. EMBO J. *3*, 593–596.
15. Cheng, T.C., Orkin, S.H., Antonarakis, S.E., Potter, M.J., Sexton, J.P., Giardina, P.J.V., Li, A., and Kazazian, H.H., Jr. (1984). β-thalassemia in Chinese: Use of in vivo RNA analysis and oligonucleotide hybridization in systematic characterization of molecular defects. Proc. Natl. Acad. Sci. USA *81*, 2821–2825.
16. Antonarakis, S.E., Orkin, S.H., Cheng, T.-c., Scott, A.F., Sexton, J.P., Trusko, S., Charache, S., and Kazazian, H.H., Jr. (1984). β-thalassemia in American Blacks: Novel mutations in the TATA box and IVS-2 acceptor splice site. Proc. Natl. Acad. Sci. USA *81*, 1154–1158.
17. Orkin, S.H., and Kazazian, H.H., Jr. (1984). Mutation and polymorphism of the human β-globin gene and its surrounding DNA. Annu. Rev. Genet. *18*, 131–171.
18. Wong, C., Dowling, C.E., Saiki, R.K., Higuchi, R.G., Erlich, H.A., and Kazazian, H.H., Jr. (1987). Characterization of β-thalassemia mutations using direct genomic sequencing of amplified single copy DNA. Nature *330*, 384–386.
19. Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Vehar, G.A., Capon, D.J., and Lawn, R.M. (1984). Characterization of the human factor VIII gene. Nature *312*, 326–330.
20. Toole, J.J., Knopf, J.L., Wozney, J.M., Sultzman, L.A., Buecker, J.L., Pittman, D.D., Kaufman, R.J., Brown, E., Shoemaker, C., Orr, E.C., et al. (1984). Molecular cloning of a cDNA encoding human antihaemophilic factor. Nature *312*, 342–347.
21. Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E., and Gitschier, J. (1993). Inversions disrupting the factor VIII gene as a common cause of severe hemophilia A. Nat. Genet. *5*, 236–241.
22. Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., and Margolet, L. (1987). Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics *1*, 113–125.
23. Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D., and Antonarakis, S.E. (1988). A novel mechanism of mutation in man: Hemophilia A due to de novo insertion of L1 sequences. Nature *332*, 164–166.
24. Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., and Kazazian, H.H., Jr. (1991). Isolation of an active human transposable element. Science *254*, 1805–1808.
25. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., and Gabriel, A. (1991). Reverse transciptase encoded by a human transposable element. Science *254*, 1808–1810.
26. Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the LRE 2 locus on chromosome 1q produces a chimeric insertion. Nat. Genet. *7*, 143–148.
27. Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retro-transposition. Science *283*, 1530–1534.
28. Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposition revisited: The restraint and rehabilitation of parasites. Cell *135*, 23–35.
29. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, 917–927.
30. Feng, Q., Moran, J.V., Kazazian, H.H., Jr., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, 905–916.
31. Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., and Kazazian, H.H., Jr. (1997). Many human L1 elements are capable of retrotransposition. Nat. Genet. *16*, 37–43.
32. Naas, T.P., DeBerardinis, R.J., Moran, J.V., Kingsmore, S.F., Seldin, M.F., Hayashizaki, Y., Martin, S.L., and Kazazian, H.H., Jr. (1998). An actively-retrotransposing, novel subfamily of mouse L1 elements. EMBO J. *17*, 590–597.
33. Goodier, J.L., Ostertag, E.M., Du, K., and Kazazian, H.H., Jr. (2001). Characterization of a novel active L1 retrotransposon subfamily in the mouse. Genome Res. *11*, 1677–1685.
34. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s

account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. USA *100*, 5280–5285.
35. Seleme MdC. (2006). Vetter MR, Cordaux R, Bastone L, Batzer MA, and Kazazian, HH, Jr. Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. Proc. Natl. Acad. Sci. USA *103*, 6611–6616.
36. Ostertag, E.M., Deberardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G., and Kazazian, H.H., Jr. (2002). A mouse model of human L1 retrotransposition. Nat. Genet. *32*, 655–660.
37. Babushok, D.V., Ostertag, E.M., Courtney, C.E., Choi, J.M., and Kazazian, H.H., Jr. (2006). L1 integration in a transgenic mouse model. Genome Res. *16*, 240–250.