ALU REPEATS AND HUMAN GENOMIC DIVERSITY

Mark A. Batzer and Prescott L. Deininger‡§*

During the past 65 million years, Alu elements have propagated to more than one million copies in primate genomes, which has resulted in the generation of a series of Alu subfamilies of different ages. Alu elements affect the genome in several ways, causing insertion mutations, recombination between elements, gene conversion and alterations in gene expression. Alu-insertion polymorphisms are a boon for the study of human population genetics and primate comparative genomics because they are neutral genetic markers of identical descent with known ancestral states.

MICROSATELLITE
A class of repetitive DNA that is made up of repeats that are 2–8 nucleotides in length. They can be highly polymorphic and are frequently used as molecular markers in population genetics studies.

**Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA. ‡Tulane Cancer Center, SL-66, Department of Environmental Health Sciences, Tulane University Health Sciences Center, 1430 Tulane Avenue, New Orleans, Louisiana 70112, USA. §Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, 1516 Jefferson Highway, New Orleans, Louisiana 70121, USA. Correspondence to M.A.B. e-mail: mbatzer@lsu.edu*
DOI: 10.1038/nrg798

The role of mobile elements in the shaping of eukaryotic genomes is becoming more and more recognized. Mobile elements make up over 45% of the human genome. These elements continue to amplify and, as a result of negative effects of their transposition, they contribute to a notable number of human diseases. All eukaryotic genomes contain mobile elements, although the proportion and activity of the classes of elements varies widely between genomes. Mobile elements are important in insertional mutagenesis and unequal homologous recombination events. They use extensive cellular resources in their replication, expression and amplification. There is considerable debate as to whether they are primarily an intracellular plague that attacks the host genome and exploits cellular resources, or whether they are tolerated because of their occasional positive influences in genome evolution. The recent completion of the draft sequence of the human genome provides an unprecedented opportunity to assess the biological properties of Alu repeats and the influence that they have had on the architechture of the human genome. Here, we present an overview of the biology and the impact of Alu repeats — the largest family of mobile elements in the human genome.

Discovery and origin of Alu elements

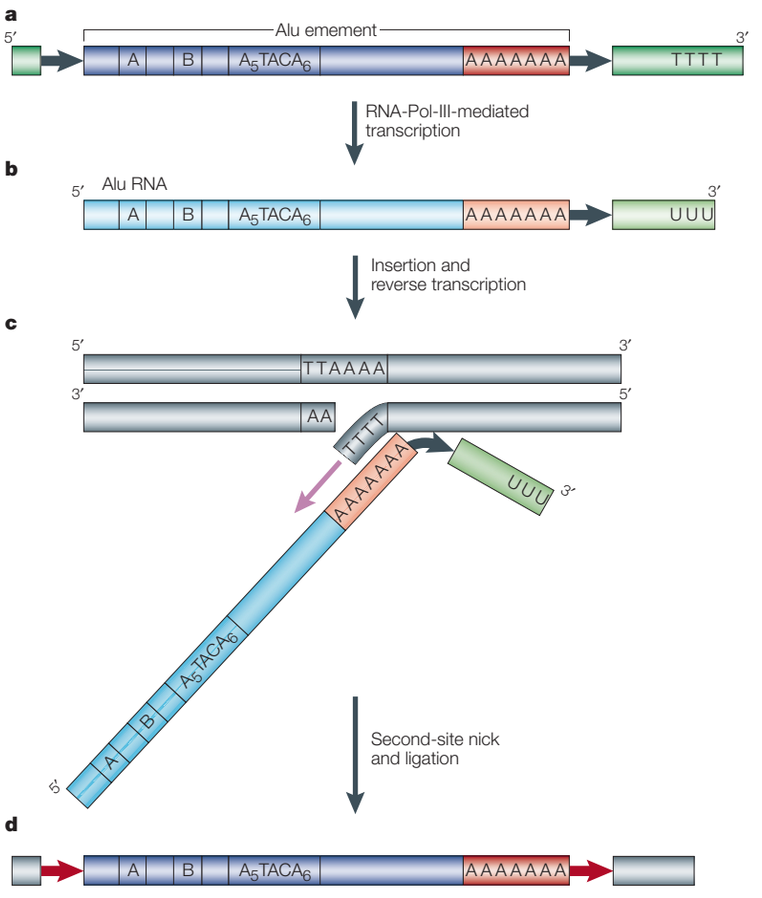
The term 'repetitive element' describes various DNA sequences that are present in multiple copies in the genomes in which they reside. Repetitive elements can be subdivided into those that are tandemly arrayed (for example, MICROSATELLITES, MINISATELLITES and telomeres) or interspersed (for example, mobile elements and processed PSEUDOGENES). Interspersed elements can be subdivided on the basis of size, with short interspersed elements (SINEs) being less than 500 bp long[1–4]. Alu SINEs were identified originally almost 30 years ago as a component in human DNA RENATURATION CURVES[5,6]. The name 'Alu elements' was given to these repeated sequences as members of this family of repeats contain a recognition site for the restriction enzyme *Alu*I (REF. 5). Subsequent detailed analyses of this portion of the renaturation curves led to sequence analysis of individual Alu elements. They were initially cloned using linkers with *Bam*HI restriction endonuclease sites that resulted in the generation of Bam-linked ubiquitous repeat (BLUR) clones[7,8]. Full-length Alu elements are ~300 bp long and are commonly found in introns, 3′ untranslated regions of genes and intergenic genomic regions (BOX 1). Initial estimates indicated that these mobile elements were present in the human genome at an extremely high copy number (~500,000 copies)[7]. Recently, a detailed analysis of the draft sequence of the human genome has shown that, out of more than one million copies, Alu elements are the most abundant SINEs, which makes them the most abundant of all mobile elements in the human genome[9]. Because of

their high copy number, the Alu gene family comprises more than 10% of the mass of the human genome[9] and as Alu sequences accumulate preferentially in gene-rich regions, they are not uniformly distributed in the human genome[9–11].

Box 1 | **A typical human Alu element and its retroposition**

The structure of each Alu element is bi-partite, with the 3′ half containing an additional 31-bp insertion (not shown) relative to the 5′ half. The total length of each Alu sequence is ~300 bp, depending on the length of the 3′ oligo(dA)-rich tail. The elements also contain a central A-rich region and are flanked by short intact direct repeats that are derived from the site of insertion (black arrows). The 5′ half of each sequence contains an RNA-polymerase-III promoter (A and B boxes). The 3′ terminus of the Alu element almost always consists of a run of As that is only occasionally interspersed with other bases (a).

Alu elements increase in number by retrotransposition — a process that involves reverse transcription of an Alu-derived RNA polymerase III transcript. As the Alu element does not code for an RNA-polymerase-III termination signal, its transcript will therefore extend into the flanking unique sequence (b). The typical RNA-polymerase-III terminator signal is a run of four or more Ts on the sense strand, which results in three Us at the 3′ terminus of most transcripts. It has been proposed that the run of As at the 3′ end of the Alu might anneal directly at the site of integration in the genome for target-primed reverse transcription (mauve arrow indicates reverse transcription) (c). It seems likely that the first nick at the site of insertion is often made by the L1 endonuclease at the TTAAAA consensus site. The mechanism for making the second-site nick on the other strand and integrating the other end of the Alu element remains unclear. A new set of direct repeats (red arrows) is created during the insertion of the new Alu element (d).



The origin and amplification of Alu elements are evolutionarily recent events that coincided with the radiation of primates in the past 65 million years[12]. Detailed sequence analysis of the structure of Alu element RNAs has indicated that Alu elements were ancestrally derived from the 7SL RNA gene, which forms part of the ribosome complex[13]. Therefore, the origins of more than 1.1 million Alu elements that are dispersed throughout the human genome can be traced to an initial gene duplication early in primate evolution, and to the subsequent and continuing amplification of these elements. This type of duplication, followed by the expansion of a SINE family, has occurred sporadically throughout evolutionary history in mammalian and non-mammalian genomes (for reviews, see REFS 1,14). The origins of a variety of SINEs can be traced to the genes of various small, highly structured RNAs, such as transfer RNA genes, the transcription of which depends on RNA polymerase III (REFS 1,15–18). The expansion of SINEs of different origins has occurred simultaneously in several diverse genomes, and although the reasons for this simultaneous expansion are unknown, there have been many interesting discussions about the factors that might have contributed to it[1].

Alu-element mobilization

The amplification of Alu elements is thought to occur by the reverse transcription of an Alu-derived RNA polymerase III transcript in a process called retrotransposition[19]. A schematic diagram of the generally accepted mechanism for Alu-element mobilization is shown in BOX 1. The Alu-derived transcript is thought to use a nick at its genomic integration site to allow target-primed reverse transcription (TPRT) to occur[20–22]. However, there is limited direct evidence for the TPRT mechanism, and it is possible that other mechanisms, such as self-priming of reverse transcription by the Alu RNA[23], might also contribute to the amplification process. Because Alu elements have no open reading frames, they are thought to 'borrow' the factors that are required for their amplification from long interspersed elements (LINEs)[24]. These elements have been shown to encode a functional reverse transcriptase[24,25] that also has an endonuclease domain[20,26], which makes them putative providers of the exogenous enzymatic functions that are thought to be crucial for Alu-element amplification. Furthermore, the poly(A) tails of LINEs and Alu elements are thought to be the common structural features that are involved in the competition of these mobile elements for the same enzymatic machinery for mobilization[27]. In support of this connection between LINE and Alu mobilization, it is interesting to note that the number of LINEs that is present in mammalian genomes has increased during the past 150 million years of evolution[28,29] — a period that also encompasses Alu-amplification activity. Therefore, LINEs seem to have supplied the crucial reverse transcriptase activity that resulted in the subsequent generation of various SINE families in different mammalian genomes that have amplified to extremely high copy numbers in a relatively short evolutionary time frame.

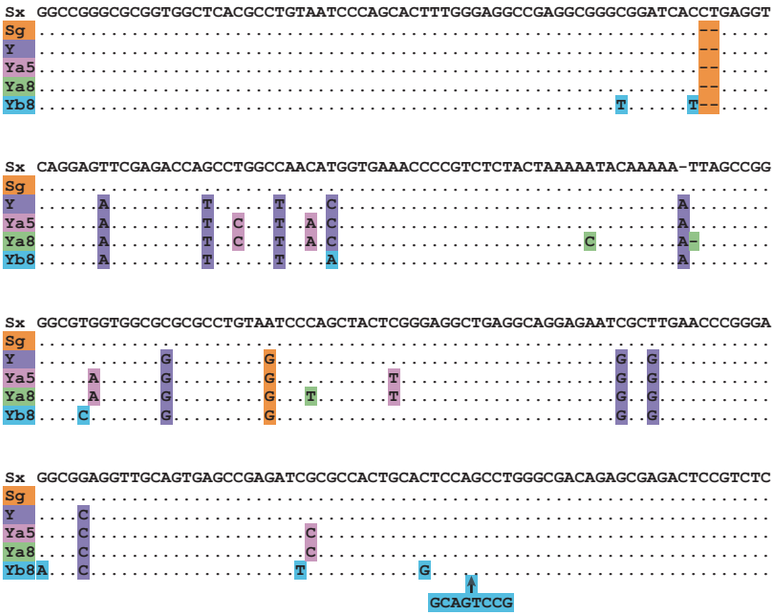


Figure 1 | **Alignment of Alu-subfamily consensus sequences.** The consensus sequence for the Alu Sx subfamily is shown at the top, with the sequences of progressively younger Alu subfamilies underneath. The dots represent the same nucleotides as the consensus sequence. Deletions are shown as dashes, and mutations are shown in coloured boxes; all are colour-coded according to the family in which the ancestral mutation arose. Each of the newer subfamilies, such as Ya5 or Yb8, has all the mutations of the ancestral Alu elements, as well as five or eight extra mutations, respectively, that are diagnostic for the particular Alu subfamily. This figure primarily illustrates the newer subfamilies and does not attempt to show many of the older Alu subfamilies.

MINISATELLITE
A class of repetitive sequences, 7–100 nucleotides each, that span 500–20,000 bp, and are especially located throughout the genome, towards chromosome ends.

PSEUDOGENE
A DNA sequence that was derived originally from a functional protein-coding gene that has lost its function owing to the presence of one or more inactivating mutations.

RENATURATION CURVE
A plot of DNA annealing as a function of DNA concentration and time. The amount of DNA (as a percentage) that has renatured (reassociated/reannealed) plotted against 'C_0t', where 'C_0' refers to the initial DNA concentration and 't' is the time of renaturation.

PAIRWISE DIVERGENCE
The number of nucleotide differences between two aligned DNA sequences.

Alu source genes and subfamily structure

Only a few human Alu elements, the so-called 'master' or source genes, seem to be retrotransposition competent[30]. Individual Alu copies contain an internal RNA-polymerase-III promoter, but this promoter is not sufficient for active transcription *in vivo*[31], as appropriate flanking sequences are required for its activation[32]. So, most new Alu copies in the human genome are, by definition, non-functional fossil relics with respect to retrotransposition unless they fortuitously land in a region of the genome that confers activity to the incomplete RNA-polymerase-III promoter. Transposition of elements that are fortuitously activated might be short lived, because individual Alu elements carry 24 or more CpG dinucleotides[33] that are prone to mutation as a result of the deamination of 5-methylcytosine residues[34,35]. Mutations in the CpG dinucleotides of a newly integrated Alu element could therefore minimize or eliminate the retrotransposition capability of a newly integrated Alu repeat. In addition, the homopolymeric-A-rich tails of individual Alu repeats are thought to be important in the amplification process[27] and might rapidly mutate into simple sequence repeats after the integration of a new Alu element[36–41]. The decay of A-rich Alu tails provides a second potential mechanism for the retrotranspositional quiescence of individual Alu repeats. Therefore, individual Alu repeats seem to have very little chance of acting as long-lived amplification drivers for the expansion of Alu-element copy number[30]. Although the essential features that define an Alu element as a retrotransposition-competent source gene are not fully understood, several factors have been suggested to influence the amplification process. These include transcriptional capacity of individual elements, ability of the specific transcript to associate with the retrotransposition mechanism, and possibly the length and homogeneity of the A-tail to allow effective priming[3,23,30,42,43].

Mutations that accumulate in the source genes are subsequently inherited by their copies. Therefore, the human Alu family is composed of several distinct subfamilies of different genetic ages that are characterized by a hierarchical series of mutations. Several laboratories have identified a number of human Alu elements that share common diagnostic sequence features and comprise subfamilies or clades that have expanded in different evolutionary time frames, as reviewed in REF. 1. FIGURE 1 compares the consensus sequences of several Alu subfamilies. Older Alu subfamilies are characterized by the smallest number of diagnostic subfamily-specific mutations. These older elements have also accumulated the largest number of random mutations (up to 20% PAIRWISE DIVERGENCE), which confirms their ancient origin[8]. By contrast, the younger families of Alu elements are characterized by an increasing number of subfamily-specific mutations, together with a smaller number of random mutations (as little as 0.1% pairwise divergence) that accumulate after the individual Alu elements integrate into the genome[35,44–46].

Alu amplification rate

The rate of amplification of human Alu elements has not been uniform[47]. FIGURE 2 illustrates the pattern of expansion of the Alu family in primate genomes in relation to the approximate subfamily size. Most of the Alu repeats duplicated more than 40 million years ago. Early in primate evolution, there was approximately one new Alu insertion in every primate birth. By contrast, the current rate of Alu amplification is estimated to be of the order of one Alu insertion in every 200 births[48]. So, the rate of Alu amplification has decreased by at least two orders of magnitude throughout the expansion of the family. Although the underlying reasons behind the decrease in the amplification rate are unknown, changes in the retrotransposition potential of mobilization-competent Alu elements that result from altered transcription or reverse transcription might be to blame[47]. It might also be a consequence of a decreased availability of empty insertion sites for the integration of new Alu copies — most of these sites are already occupied by older Alu elements. Furthermore, one might speculate that the human genome has evolved towards restricting the amplification of these elements, similar to the way that genomes of model organisms, such as *Drosophila melanogaster*, restrict amplification of other types of mobile elements[49].

Recently integrated human Alu repeats

Alu elements that are unique to the human genome were initially identified on the basis that they share a higher number of diagnostic point mutations, and that

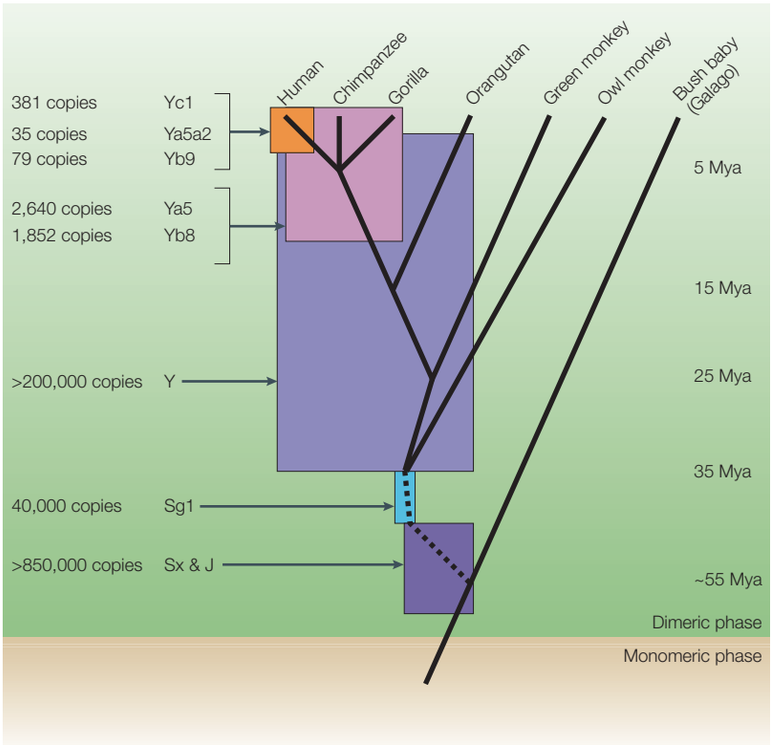


Figure 2 | **The expansion of Alu elements in primates.** The expansion of Alu subfamilies (Yc1, Ya5a2, Yb9, Yb8, Y, Sg1, Sx and J) is superimposed on a tree of primate evolution. The expansion of the various Alu subfamilies is colour coded to denote the times of peak amplification. The approximate copy numbers of each Alu subfamily are also noted. Mya, million years ago.

they were polymorphic with respect to their presence or absence in diverse human genomes[35,50–52]. Almost all of the recently integrated human Alu elements belong to one of several small and closely related 'young' Alu subfamilies, known as Y, Yc1, Yc2, Ya5, Ya5a2, Ya8, Yb8 and Yb9 (REFS 35,44–46,52–55). With the exception of the Alu Y-family elements, and of a small number of elements from the other 'young' subfamilies[43,56–58], individual members of these young Alu subfamilies that are present in the human genome are not found at ORTHOLOGOUS positions in the genomes of other great apes. These largely human-specific Alu subfamilies represent only ~0.5% of all the Alu repeats in the human genome and have amplified in the human genome in an overlapping time frame, as shown in FIG. 3.

Although some newly integrated Alu elements result in detrimental mutation events in the human genome (see below), the vast majority of recently integrated Alu elements have had no apparent negative impact on the genome and represent new, essentially neutral, mutation events. After a new, neutral Alu insertion integrates into the genome, it is subjected to GENETIC DRIFT. So, the probability that it will be lost from the population is initially quite high, depending on the size of the population (the greater the population size, the more likely it is to be lost). But, over a short period of time, the Alu element will increase in frequency in the population. Because the amplification of Alu repeats is a continuing process, a series of Alu elements must have integrated into the human genome at different times. Therefore, the time of origin of a new Alu insertion directly affects the spread of this insertion through the species or the population. Depending on when, in primate evolution, an Alu element has integrated into a primate genome, it will be shared by one or more species. But even the elements that are only found in a single species might have arisen at different times. Some members of the 'young' Alu subfamilies have inserted into the human genome so recently that they are polymorphic with respect to the presence or absence of insertion in different human genomes[51]. Those relatively few elements that are present in the genomes of some individuals and absent from others are referred to as Alu-insertion polymorphisms[51,53,59,60]. Individual Alu elements might be found in a single population, a single family or, in the case of the *de novo* Alu insertions, in a single individual, depending on the genetic drift that occurs after the initial integration of that element into the human genome (FIG. 4).

The 'young' Alu subfamilies are composed of ~5,000 Alu elements that have integrated into the human genome in the past 4–6 million years after the divergence of humans and African apes[45,46,51,52,54], but most of them integrated before the African radiation of humans[44–46,51,54,61]. So, these Alu repeats are monomorphic for their insertion sites among diverse human genomes. However, ~25% of the young Alu repeats (~1,200 elements) have inserted into the human genome so recently that they are dimorphic for the presence or absence of the insertion, which makes them a useful source of genomic polymorphism[44–46,51,54].

Alu-insertion polymorphisms

The analysis of human Alu-insertion polymorphisms has been used to address several questions about human origins and demography[59,60,62–71]. In several instances, many types of genetic variation (such as mitochondrial DNA sequences or restriction-fragment length polymorphisms (RFLPs)) have been examined in overlapping, diverse human populations and have provided largely congruent results with respect to the history of the human population[62,65,70]. Alu-insertion polymorphisms have several characteristics that make them unique reagents for the study of human population genetics[51,59–61]. Individuals that share Alu-insertion polymorphisms have inherited the Alu elements from a common ancestor, which makes the Alu-insertion alleles identical by descent. The identical-by-descent nature of SINE insertions that are used in phylogenetic studies[72–75] has previously been questioned[76], and several examples of SINE insertions that have occurred at or near the same genomic region have recently been reported[77,78]. However, variation in the presence or absence of SINE insertions seems to be quite rare, and is a function of both evolutionary time and retrotransposition rate. This is particularly true with respect to Alu-insertion polymorphisms, as the probability of two independent Alu insertions occurring in the same genomic region in the human population, given the current rate of Alu retrotransposition and the relatively short evolutionary time frame that is involved, is essentially zero[59,78]. Therefore,

ORTHOLOGOUS GENES
Loci in two species that are derived from a common ancestral locus by a speciation event. This is different from paralogous members of a gene family that are derived from duplication events.

GENETIC DRIFT
Random changes in allele frequency that result from the sampling of gametes from generation to generation.

Alu-insertion polymorphisms are essentially HOMOPLASY-free characteristics that can be used to study human population genetics[59,78]. In addition, there is no evidence for any type of process that specifically removes Alu elements from the genome; even when a rare deletion occurs, it leaves behind a molecular signature[79]. By contrast, other types of genetic polymorphism, such as variable numbers of tandem repeats[80], RFLPs[81] and single-nucleotide polymorphisms (SNPs)[82–84], are merely identical by state; that is, they have arisen as the result of several independent parallel mutations at different times and have not been inherited from a common ancestor. Alleles that are identical by descent have been directly inherited from a common ancestor. Alleles that are identical by state have the same character state, but have not been inherited from a common ancestor. The ancestral state of Alu-insertion polymorphisms is known to be the absence of the Alu element at a particular genomic location[51,59,60]. Precise knowledge of the ancestral state of a genomic polymorphism allows us to draw trees of population relationships without making too many assumptions[59,60,63,69].

Alu elements as insertion mutations

The diversity created by a new Alu insertion can have a rare positive impact on the genome; for example, through the advantageous alteration of gene expression or the occasional incorporation of the Alu element into the protein-encoding portion of a gene[85–87]. More commonly, the insertion of a new Alu repeat results in one of several negative effects (for a review, see REF. 48). Genetic disorders can result from different types of mutation that arise following the insertion of an Alu repeat (FIG. 5a).

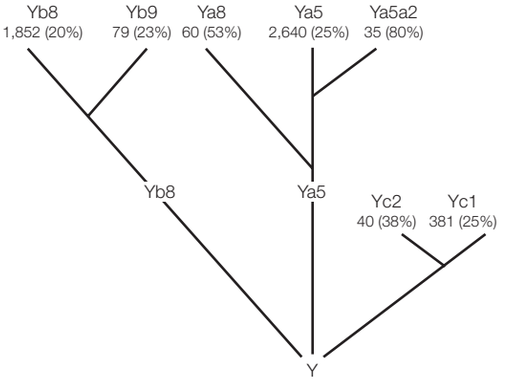


Figure 3 | **Expansion of recently integrated human Alu subfamilies.** Several subfamilies of Alu elements have expanded simultaneously in the human genome primarily from three Y-subfamily lineages, termed 'Ya', 'Yb' and 'Yc' in accordance with standard Alu nomenclature on the basis of commonly shared mutations. The approximate copy number of each subfamily is given as estimated from computational analysis of the draft sequence of the human genome[9]. The percentage of insertion polymorphisms in each family is given in brackets. Alu subfamilies with smaller copy numbers and higher levels of insertion polymorphism are generally thought to be more recent in origin in the human genome. The tree is based in the mutations that define each Alu subfamily. The time-frame for dispersal of these Alu subfamilies is shown in FIG. 2.

An insertion of an Alu element might alter the transcription of a gene by changing the methylation status of its promoter, by disrupting its promoter or by introducing additional regulatory sequences, such as the binding sites for steroid-hormone receptors, that are contained in some Alu-family members[88]. Alternatively, an Alu repeat might integrate directly into the coding region of a gene and disrupt the open reading frame, which generates a nonsense or frameshift mutation, or disrupt the splicing of a gene. Alu insertions account for ~0.1% of all human genetic disorders, such as neurofibromatosis, haemophilia, breast cancer, Apert syndrome, cholinesterase deficiency and complement deficiency[48].

Alu elements and recombination

Because Alu repeats are the largest multigene family in the human genome they might also act as nucleation points for homologous recombination[48]. Homologous recombination between dispersed Alu elements might result in various genetic exchanges, including duplications, deletions and translocations (FIG. 5b). Across longer evolutionary time frames, these types of event are probably a mechanism for the creation of genetic diversity in the human genome, and they have been suggested as a putative mechanism for the diversification of the TROPOELASTIN genes during primate evolution[89].

Alu-mediated recombination events might occur in the soma or in the germ line. Some regions of the genome, such as the low-density lipoprotein locus, seem to be more susceptible to Alu-mediated recombination events than others. Although a high density of Alu elements is likely to contribute to a high level of unequal homologous recombination, it does not seem to be sufficient, because several genes with very high Alu content are not particularly prone to this type of recombination damage — for example, thymidine kinase or β-tubulin[90]. Levels of intrachromosomal recombination have previously been shown to be directly related to the length of uninterrupted regions of nucleotide identity, with higher rates of recombination being associated with longer stretches of nucleotide identity[91]. Therefore, the level of recombination between Alu elements from different subfamilies should vary as a function of pairwise sequence divergence between elements, with older Alu elements that have higher pairwise divergence (~15–20%) being much less likely to recombine than younger Alu insertions that have lower pairwise divergence (<1%). It is also interesting to note that the rapid mutation of methylated CpG dinucleotides in newly integrated Alu repeats[34,35] would tend to increase the pairwise divergence between Alu elements and provide one potential mechanism for the establishment of a barrier against subsequent Alu-mediated homologous recombination events in the genome. Various inherited disorders have been caused by Alu-mediated recombination, including insulin-resistant diabetes type II, Lesch–Nyhan syndrome, Tay–Sachs disease, complement component C3 deficiency, familial hypercholesterolaemia and α-thalassaemia[48]. Several types of cancer, including Ewing sarcoma, breast cancer

HOMOPLASY
Similarity due to independent evolutionary change; an allelic variant (such as a nucleotide variant or a mobile-element insertion at a particular location) that is present in two or more genes, but absent in their common ancestor.

TROPOELASTIN
The soluble precursor of elastin (one of the most hydrophobic proteins known). Mammalian tropoelastin is a moderately conserved protein.

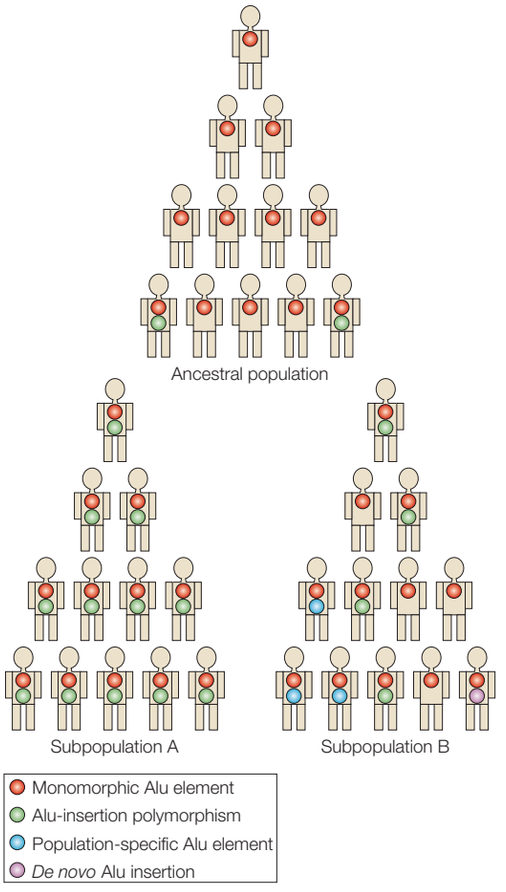


Figure 4 | **Spread of an Alu insertion.** The ancestral human population is shown at the top, and two separate subpopulations are shown below. A monomorphic Alu insertion (red) is shared by all members of the population. Several Alu insertion polymorphisms are also shown, including an intermediate-frequency Alu insertion polymorphism in the ancestral and subpopulations (green), a population-specific element (blue) and a *de novo* insert in subpopulation B (mauve).

and acute myelogenous leukaemia have also been associated with Alu-mediated recombination[48]. Overall, ~0.3% of all human genetic diseases seem to have resulted from Alu-mediated unequal homologous recombination[48].

There is also some evidence that Alu elements that insert into an inverted orientation are more prone than others to illegitimate recombination[92,93]. It has been suggested that these types of recombination events might have resulted in a genomic depletion of Alu elements with inverted orientation. However, identifying Alu elements that are responsible for such events has proven much more difficult than detecting Alu-mediated homologous recombination events for two reasons — the inverted elements result in illegitimate recombination events, and it is difficult to determine, in individual recombination events, whether the Alu elements contributed to the event or were merely located fortuitously nearby. So, the total contribution of Alu elements to recombination-mediated damage to the human genome might be much higher than the estimates quoted above.

LOSS OF HETEROZYGOSITY (LOH). A loss of one of the alleles at a given locus as a result of a genomic change, such as mitotic deletion, gene conversion or chromosome missegregration.

Apart from the propensity of certain genes to have highly increased Alu-mediated recombination, it is probable that the extent to which this process actually occurs varies between individuals. For example, model studies show that *TP53* (which encodes p53) mutants are much more prone to both homologous and inverted Alu-mediated recombination events[94]. So, individuals with defects or polymorphisms in *TP53* might be more prone to these types of event as a result of increased levels of homologous recombination, as well as possibly decreased sensitivity to base-pairing fidelity that would presumably allow recombination between more poorly matched homologues. Furthermore, as genes such as *TP53* become inactivated in tumorigenesis, Alu-mediated recombination events are likely to be a principal factor in progression of the tumour through LOSS OF HETEROZYGOSITY and genomic rearrangements.

It is also possible that Alu insertions have more subtle consequences for genomic structure and function — for example, chromosomal recombination rates are influenced by non-homologous regions. Previous studies have indicated that a mobile-element insertion might be responsible for a marked decrease of recombination in the vicinity of insertion[95,96]. Such a decrease of recombination might influence the reassociation of haplotypes in the vicinity of a polymorphic Alu insertion. Early in primate evolution, this type of local disruption of chromosomal recombination might have contributed to chromosome incompatibilities that accelerated speciation.

Alu elements are distributed in the genome with a strong bias towards the more gene-rich chromosomal regions[9–11]. It seems unlikely that this bias is due to insertional preferences, because L1 elements have almost the opposite chromosomal distribution, and the younger Alu elements do not show this chromosomal bias[97]. Therefore, it has been suggested that Alu elements might have a function that imposes post-insertional selection pressures that change the distribution of the older Alu elements[9], although some recombination-based process that can alter their distribution cannot be ruled out. However, even the younger Alu elements in that study were old enough to be fixed in the human genome. Once elements are fixed in the genome, there is no longer enough diversity for natural selection to act on, and therefore natural selection is unlikely to be important[98]. Therefore, we believe that the relatively high Alu–Alu recombination rate is likely to be responsible for the gradual depletion of Alu elements in the gene-poor regions. Recombination events in the more gene-rich regions are more likely to provide a selective disadvantage, resulting in the gradual loss of Alu elements from the gene-poor regions.

Alu elements and simple sequence repeats

Several laboratories have done computational and empirical studies of Alu insertions and of simple sequence repeats in the human genome, and noted an association in the distribution of these two classes of repeated sequences[37,39–41,99]. When a new Alu element

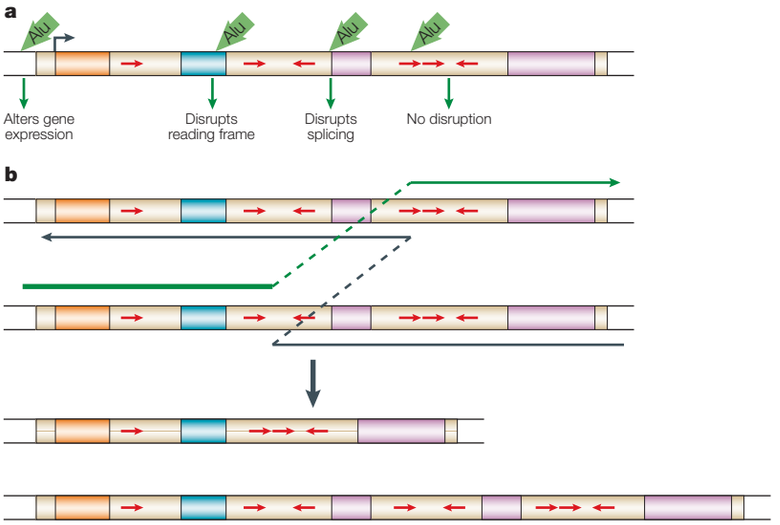


Figure 5 | **Schematic of Alu-induced damage to the human genome. a** | Potential consequences of insertion of a new element in the vicinity of a gene. The coloured boxes represent exons. The red arrows show existing Alu elements that are orientated in different directions in the introns of the gene. The site of insertion of an Alu element influences the effect of this insertion on the genome as shown. **b** | Unequal, homologous recombination between two Alu elements that are located in two different introns. The arrows that are broken by dashed lines show the path of the recombination event. The genes below show that a deletion has occurred in one copy, whereas a duplication has occurred in the other; either is likely to be deleterious (modified with permission from figure 1 in REF. 48).

integrates into the genome, it brings along two additional sources of simple sequence repeats: in the middle, the A-rich region (that contains the sequence A_5TACA_6) and the oligo(dA)-rich tail (which can be a perfect A repeat, up to 100 bases long). In addition, individual Alu repeats are also flanked by short (A+T)-rich direct repeated sequences that form when the elements integrate into staggered chromosomal breaks, and are thought to arise as a result of the endonucleolytic activity of LINE-derived reverse transcriptase[26]. The homopolymeric simple sequences in Alu elements are the least complex simple sequence repeats in the human genome. So, the association between Alu elements and homopolymeric simple sequence repeats in general is not surprising. These simple sequences are subjected to various mutational forces, including point mutations, and inter- and intrastrand crossover events, as well as replication slippage, all of which lead to changes in both length and complexity of the repeats[100,101].

More than 25% of all the simple sequence repeats in primate genomes, including microsatellites, are associated with Alu elements[38]. In some cases, this association might result from a random integration of Alu elements near existing microsatellite sequences[36]. Alternatively, and more commonly, Alu elements themselves are the source of homopolymeric simple sequences that give rise to microsatellite sequence motifs, following additional mutational events[36]. The analysis of Alu subfamilies and of the recently integrated Alu elements has indicated that the homopolymeric adenine sequences that lie in Alu elements are a source of primate microsatellites[36–38]. In fact, because each Alu has two A-rich regions, we can estimate that together, genome-wide Alu elements provide at least 2.2 million potential sites for generating microsatellite repeats. There is also at least one example of the middle A-rich region of an Alu element in the frataxin gene that can give rise to a triplet-repeat expansion that is responsible for Friedreich ataxia[102,103]. Further computational and empirical studies are required to help us understand the mechanisms that generate these microsatellites, and how the generation of mutations in these sequences and their rates differ across the genome.

GENE CONVERSION
A non-reciprocal recombination process that results in an alteration of the sequence of a gene to that of its homologue during meiosis.

Alu elements and SNPs

Several studies have involved repeated sequence analysis of individual Alu-family members that have only recently integrated into the human genome. Because of their recent origin, these young Alu elements have low levels of SNPs[104]. Phylogenetic studies of the sequence diversity in and around Alu elements that are located in the α-fetoprotein gene cluster[105], albumin[106] and around globin genes[107,108] have indicated that, once integrated into the genome, Alu elements might mutate at a neutral rate. However, as already mentioned, the high incidence of CpG dinucleotides in new Alu inserts predisposes them to an approximately tenfold higher mutation rate[34,35]. Because there are ~24 CpG positions in a new Alu insertion[33], roughly half of the SNPs in young Alu elements fall in these CpG dinucleotides.

Several studies indicate that Alu elements, as well as other mobile elements, undergo a large amount of gene conversion[44,54,109–111]. None of the large-scale studies on Alu elements has systematically addressed the impact of GENE CONVERSION on human polymorphisms. Alu-element-mediated gene conversion has been implicated in the inactivation of the CMP-*N*-acetylneuraminic acid hydroxylase gene in the human lineage[112]. However, phylogenetic studies, as well as studies based on the very strong hierarchy of Alu subfamilies, have indicated that there might be very high levels of gene conversion among Alu elements[110]. The only gene conversions that were detectable in those studies were those that changed one or more of the diagnostic, subfamily-specific mutations. In general, the gene conversions seem to involve relatively small regions of 50–100 base pairs that alter only one or two of the diagnostic mutations, although gene conversion of a complete Alu element has also been detected[109,112]. Gene-conversion frequency varies between Alu elements. In the case of relatively recently inserted Ya5 elements that were converted to the older and much higher copy number Alu Y subfamily, a conversion frequency of ~20% was observed over a few million years[110]. In another study, of older Alu elements that were undergoing gene conversion to the low copy number Ya5 and Yb9 subfamilies, only ~1% of the elements had undergone conversion over a period of 5–10 million years[78]. Because these Alu elements are older, they would have accumulated mismatches relative to their subfamily consensus sequences. Unfortunately, we cannot determine whether these differences in the conversion rate were related to the copy number or mismatch levels between the different Alu repeats. It is important to bear in mind that, because of detection limitations, these studies might

underestimate gene-conversion events between members of the same subfamily.

The molecular mechanism behind these apparent gene-conversion events is unknown. It is possible that tandem integration of Alu elements, followed by recombination between the two adjacent Alu elements, could create Alu elements with chimeric subfamily characteristics[109,110]. Although this might explain a small proportion of Alu gene conversions, most would require double crossover events that are much more likely to occur by more traditional gene-conversion events. So far, the relative influences of copy number, mismatch and sequence polymorphism on Alu-related gene conversion have not been determined.

Irrespective of the molecular mechanism that underlies Alu-mediated gene conversion, this type of sequence exchange is of great biological importance to Alu repeats and to the alteration of the sequence architecture of the human genome[42,78,109,110]. Other types of mobile element, such as Tf2 in yeast, have been shown to mobilize by gene-converting pre-existing elements[113]. Therefore, this process of gene conversion between Alu elements might, in fact, represent a second pathway for their mobilization in the human genome[109,110]. Although such a process will not influence the copy number of Alu elements, it can alter the copy number of specific subfamilies and also potentially result in activation or silencing of an Alu element at a specific locus by altering its promoter sequences. Because Alu elements make up more than 10% of the human genome, and because they are associated with such a high level of gene conversion, it is probable that this type of gene conversion contributes considerably to the overall frequency of SNPs in the human genome[110]. Similar types of gene conversion seem to contribute to SNP diversity throughout the human genome[114,115]. Depending on their frequency, gene-conversion events could have an important impact on the use of these SNPs as identical-by-descent markers, because gene conversion would effectively generate parallel forward or backward SNP mutations.

Alu elements and gene expression

An estimated one-third of all human CpG dinucleotides are found in Alu sequences[116,117], and those that lie in mobilization-competent Alu elements have been retained throughout 65 million years of evolution[33,35]. Because the remainder of Alu sequence (non-CpG bases) has mutated at a neutral rate throughout subfamily evolution, these CpG dinucleotides might have some function, either in the Alu sequences themselves or in the genome. In eukaryotes, cytosines can become methylated to form 5-methylcytosine — an important genomic modification that frequently leads to a methylcytosine-to-thymidine conversion on replication and, potentially, to large-scale changes in gene expression, as a result of alterations in DNA methylation patterns[118]. Therefore, the conservation of these dinucleotides in Alu subfamilies might indicate a form of selection for their retention in active Alu elements. This is in contrast with the rest of the Alu sequences, which seem to have evolved at a neutral rate throughout primate evolution[35].

Methylation levels that are associated with several Alu sequences have been shown to vary in different tissues at different times throughout development[116]. Such spatial and temporal variation in Alu methylation is important in silencing their expression[119,120]. In addition, the decay of methylated CpG dinucleotides into TpG dinucleotides would also tend to increase the pairwise divergence between Alu repeats over time, thereby decreasing the recombination between elements. Methylation of these CpG dinucleotides has been shown to influence gene expression in a subtle, gene-specific manner, as well as in genome-wide imprinting. These data indicate that Alu elements might act as global modifiers of gene expression through changes in their own methylation status.

The expression of Alu RNAs has been shown to increase in response to cellular stress, and to viral and translational inhibition[121,122]. In addition, Alu RNA has been shown to stimulate the translation of a reporter gene, which indicates that Alu RNAs might have a role in maintaining or regulating translation (C. M. Rubin *et al.*, unpublished data). Even though Alu elements form a large multigene family, the expression of individual genomic Alu elements in response to stress induction seems to vary from locus to locus, and to depend on the local genomic environment. So, aspects of both local and global changes in response to stress can be attributed to Alu elements.

Conclusion and future directions

Alu repeats continue to generate genomic diversity in several ways. Their amplification has resulted in the generation of the largest family of mobile elements in the human genome. Several thousand Alu elements have integrated into the human genome since the divergence of humans and African apes[44,46,78,110]; some of them have caused new detrimental mutations[48]. Additionally, recombination between Alu elements has contributed to the generation of human genetic diversity and is responsible for several human genetic disorders[48]. Many Alu sequences affect gene expression through changes in their own methylation status, whereas the expression of Alu RNA might influence translation levels[121,123,124]. Alu repeats that have undergone extensive gene conversion influenced the accumulation of SNPs in the genome[44,78,110] — a phenomenon that has a significant impact on genetic-linkage studies and on population genetics. Detailed knowledge of the levels of temporal and spatial variation in Alu-related gene conversion will provide further insight into the magnitude of this process.

But, most of the newly integrated Alu insertions are an innocuous source of genetic variation with a subset of homoplasy-free Alu-insertion polymorphisms that are useful for studying the relationships between populations, and the evolution and organization of tandemly arrayed gene families[44,46,78,110]. These elements will also be useful as genomic anchors for comparative genomic studies of the organization of non-human primate genomes[44,46,78]. Future studies of the expansion of recently integrated Alu elements and

LINEs in non-human primate genomes will allow a detailed analysis of the interplay between the amplification dynamics of these elements, using whole primate genomes as 'test-tubes'. These types of studies will facilitate an evolutionary examination of the current working hypothesis that Alu elements and LINEs use a common pathway for amplification[27]. In addition, they will result in the generation of new genetic markers for primate conservation biology and studies of non-human primate demographics, as well as providing an insight into the genetic differences between humans and non-human primates. These studies should also shed new light on the biology of these interesting mobile elements and provide a comparative assessment of their role in shaping various non-human primate genomes.

1. Deininger, P. L. & Batzer, M. A. Evolution of retroposons. *Evol. Biol.* **27**, 157–196 (1993).
2. Okada, N. SINEs. *Curr. Opin. Genet. Dev.* **1**, 498–504 (1991).
3. Schmid, C. W. Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. *Prog. Nucleic Acids Res. Mol. Biol.* **53**, 283–319 (1996).
4. Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
5. Houck, C. M., Rinehart, F. P. & Schmid, C. W. A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* **132**, 289–306 (1979).
6. Schmid, C. W. & Deininger, P. L. Sequence organization of the human genome. *Cell* **6**, 345–358 (1975).
7. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T. & Schmid, C. W. Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. *Nature* **284**, 372–374 (1980).
8. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. & Schmid, C. W. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151**, 17–33 (1981).
9. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
An assembly and annotation of the first draft sequence of the entire human genome that includes a comprehensive analysis of repeated DNA sequences.
10. Korenberg, J. R. & Rykowski, M. C. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**, 391–400 (1988).
11. Chen, C., Gentles, A. J., Jurka, J. & Karlin, S. Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA* **99**, 2930–2935 (2002).
12. Deininger, P. L. & Daniels, G. R. The recent evolution of mammalian repetitive DNA elements. *Trends Genet.* **2**, 76–80 (1986).
13. Ullu, E. & Tschudi, C. Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171–172 (1984).
14. Shedlock, A. M. & Okada, N. SINE insertions: powerful tools for molecular systematics. *Bioessays* **22**, 148–160 (2000).
15. Ohshima, K., Hamada, M., Terai, Y. & Okada, N. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol. Cell. Biol.* **16**, 3756–3764 (1996).
16. Ohshima, K. & Okada, N. Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus. *J. Mol. Biol.* **243**, 25–37 (1994).
17. Okada, N. & Hamada, M. The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. *J. Mol. Evol.* **44**, S52–S56 (1997).
18. Okada, N. & Ohshima, K. A model for the mechanism of initial generation of short interspersed elements (SINEs). *J. Mol. Evol.* **37**, 167–170 (1993).
19. Rogers, J. Retroposons defined. *Nature* **301**, 460 (1983).
20. Feng, Q., Moran, J. V., Kazazian, H. H. Jr & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
21. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
This manuscript presents the development and characterization of an *in vitro* assay to measure retrotransposition in mammalian cells.
22. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. 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 (1993).
The authors provide strong experimental evidence for the role of target-primed reverse transcription in retroelement mobilization.
23. Shen, M. R., Brosius, J. & Deininger, P. L. *BC1* RNA, the transcript from a master gene for ID element amplification, is able to prime its own reverse transcription. *Nucleic Acids Res.* **25**, 1641–1648 (1997).
24. Mathias, S. L., Scott, A. F., Kazazian, H. H. Jr, Boeke, J. D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808–1810 (1991).
25. Deragon, J. M., Sinnett, D. & Labuda, D. Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. *EMBO J.* **9**, 3363–3368 (1990).
26. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl Acad. Sci. USA* **94**, 1872–1877 (1997).
This paper provides the first computational evidence for the involvement of enzymatic activity in the integration of retroposons in the genome.
27. Boeke, J. D. LINEs and Alus — the polyA connection. *Nature Genet.* **16**, 6–7 (1997).
28. Fanning, T. G. & Singer, M. F. LINE-1: a mammalian transposable element. *Biochim. Biophys. Acta* **910**, 203–212 (1987).
29. Skowronski, J. & Singer, M. F. The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 457–464 (1986).
30. Deininger, P. L., Batzer, M. A., Hutchison, C. A. & Edgell, M. H. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**, 307–311 (1992).
A comparison of amplification models for mobile elements that are proposed as a result of the initial discovery of mobile-element subfamily structure.
31. Paulson, K. E. & Schmid, C. W. Transcriptional inactivity of Alu repeats in HeLa cells. *Nucleic Acids Res.* **14**, 6145–6158 (1986).
32. Ullu, E. & Weiner, A. M. Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature* **318**, 371–374 (1985).
33. Batzer, M. A. *et al.* Standardized nomenclature for Alu repeats. *J. Mol. Evol.* **42**, 3–6 (1996).
34. Labuda, D. & Striker, G. Sequence conservation in Alu evolution. *Nucleic Acids Res.* **17**, 2477–2491 (1989).
35. Batzer, M. A. *et al.* Structure and variability of recently inserted Alu family members. *Nucleic Acids Res.* **18**, 6793–6798 (1990).
36. Arcot, S. S., Wang, Z., Weber, J. L., Deininger, P. L. & Batzer, M. A. Alu repeats: a source for the genesis of primate microsatellites. *Genomics* **29**, 136–144 (1995).
37. Economou, E. P., Bergen, A. W., Warren, A. C. & Antonarakis, S. E. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl Acad. Sci. USA* **87**, 2951–2954 (1990).
38. Jurka, J. & Pethiyagoda, C. Simple repetitive DNA sequences from primates: compilation and analysis. *J. Mol. Evol.* **40**, 120–126 (1995).
39. Zuliani, G. & Hobbs, H. H. A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. *Am. J. Hum. Genet.* **46**, 963–969 (1990).
40. Toth, G., Gaspari, Z. & Jurka, J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967–981 (2000).
41. Beckman, J. S. & Weber, J. L. Survey of human and rat microsatellites. *Genomics* **12**, 627–631 (1992).
42. Aleman, C., Roy-Engel, A. M., Shaikh, T. H. & Deininger, P. L. *Cis*-acting influences on Alu RNA levels. *Nucleic Acids Res.* **28**, 4755–4761 (2000).
43. Shaikh, T. H., Roy, A. M., Kim, J., Batzer, M. A. & Deininger, P. L. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* **271**, 222–234 (1997).
44. Carroll, M. L. *et al.* Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* **311**, 17–40 (2001).
45. Roy, A. M. *et al.* Recently integrated human Alu repeats: finding needles in the haystack. *Genetica* **107**, 149–161 (1999).
46. Roy-Engel, A. M. *et al.* Alu insertion polymorphisms for the study of human genomic diversity. *Genetics* **159**, 279–290 (2001).
47. Shen, M. R., Batzer, M. A. & Deininger, P. L. Evolution of the master Alu gene(s). *J. Mol. Evol.* **33**, 311–320 (1991).
48. Deininger, P. L. & Batzer, M. A. Alu repeats and human disease. *Mol. Genet. Metab.* **67**, 183–193 (1999).
This article provides an overview of the data that show a role for Alu elements in human genetic instability and disease.
49. Misra, S. & Rio, D. C. Cytotype control of *Drosophila P* element transposition: the 66 kd protein is a repressor of transposase activity. *Cell* **62**, 269–284 (1990).
50. Deininger, P. L. & Slagel, V. K. Recently amplified Alu family members share a common parental Alu sequence. *Mol. Cell. Biol.* **8**, 4566–4569 (1988).
51. Batzer, M. A. & Deininger, P. L. A human-specific subfamily of Alu sequences. *Genomics* **9**, 481–487 (1991).
52. Matera, A. G., Hellmann, U. & Schmid, C. W. A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10**, 5424–5432 (1990).
53. Batzer, M. A. *et al.* Amplification dynamics of human-specific (HS) Alu family members. *Nucleic Acids Res.* **19**, 3619–3623 (1991).
54. Batzer, M. A. *et al.* Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *J. Mol. Biol.* **247**, 418–427 (1995).
55. Jurka, J. A new subfamily of recently retroposed human Alu repeats. *Nucleic Acids Res.* **21**, 2252 (1993).
56. Leeflang, E. P., Chesnokov, I. N. & Schmid, C. W. Mobility of short interspersed repeats within the chimpanzee lineage. *J. Mol. Evol.* **37**, 566–572 (1993).
57. Leeflang, E. P., Liu, W. M., Chesnokov, I. N. & Schmid, C. W. Phylogenetic isolation of a human Alu founder gene: drift to new subfamily identity. *J. Mol. Evol.* **37**, 559–565 (1993).
58. Leeflang, E. P., Liu, W. M., Hashimoto, C., Choudary, P. V. & Schmid, C. W. Phylogenetic evidence for multiple Alu source genes. *J. Mol. Evol.* **35**, 7–16 (1992).
59. Batzer, M. A. *et al.* African origin of human-specific polymorphic Alu insertions. *Proc. Natl Acad. Sci. USA* **91**, 12288–12292 (1994).
This paper shows the use of Alu elements for the study of human population genetics and includes the first comprehensive survey of Alu-insertion-polymorphism-related human variation.
60. Perna, N. T., Batzer, M. A., Deininger, P. L. & Stoneking, M. Alu insertion polymorphism: a new type of marker for human population studies. *Hum. Biol.* **64**, 641–648 (1992).
61. Arcot, S. S. *et al.* Alu fossil relics — distribution and insertion polymorphism. *Genome Res.* **6**, 1084–1092 (1996).
62. Bamshad, M. *et al.* Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**, 994–1004 (2001).
63. Batzer, M. A. *et al.* Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.* **42**, 22–29 (1996).
64. Comas, D. *et al.* Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum. Genet.* **107**, 312–319 (2000).
65. Jorde, L. B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).
66. Nasidze, I. *et al.* Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur. J. Hum. Genet.* **9**, 267–272 (2001).
67. Novick, G. E. *et al.* Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum. Biol.* **70**, 23–39 (1998).
68. Sherry, S. T., Harpending, H. C., Batzer, M. A. & Stoneking, M. Alu evolution in human populations: using the coalescent

to estimate effective population size. *Genetics* **147**, 1977–1982 (1997).
69. Stoneking, M. *et al.* Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**, 1061–1071 (1997).
70. Watkins, W. S. *et al.* Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* **68**, 738–752 (2001).
71. Hammer, M. F. A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.* **11**, 749–761 (1994).
72. Shimamura, M. *et al.* Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388**, 666–670 (1997).
In this manuscript, the authors use SINE insertions to study the phylogenetic origin of whales.
73. Nikaido, M., Rooney, A. P. & Okada, N. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl Acad. Sci. USA* **96**, 10261–10266 (1999).
74. Nikaido, M. *et al.* Evolution of CHR-2 SINEs in cetartiodactyl genomes: possible evidence for the monophyletic origin of toothed whales. *Mamm. Genome* **12**, 909–915 (2001).
75. Nikaido, M. *et al.* Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl Acad. Sci. USA* **98**, 7384–7389 (2001).
76. Hillis, D. M. SINEs of the perfect character. *Proc. Natl Acad. Sci. USA* **96**, 9979–9981 (1999).
77. Cantrell, M. A. *et al.* An ancient retrovirus-like element contains hot spots for SINE insertion. *Genetics* **158**, 769–777 (2001).
78. Roy-Engel, A. M. *et al.* Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* **316**, 1033–1040 (2002).
79. Edwards, M. C. & Gibbs, R. A. A human dimorphism resulting from loss of an Alu. *Genomics* **14**, 590–597 (1992).
80. Nakamura, Y. *et al.* Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616–1622 (1987).
81. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
82. Brookes, A. J. The essence of SNPs. *Gene* **234**, 177–186 (1999).
83. Chakravarti, A. It's raining SNPs, hallelujah? *Nature Genet.* **19**, 216–217 (1998).
84. Pennisi, E. A closer look at SNPs suggests difficulties. *Science* **281**, 1787–1789 (1998).
85. Britten, R. J. DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl Acad. Sci. USA* **93**, 9374–9377 (1996).
86. Britten, R. J. Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**, 177–182 (1997).
A thorough compilation of mobile elements, which have been functionally significant in the genome.
87. Makalowski, W., Mitchell, G. A. & Labuda, D. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* **10**, 188–193 (1994).
88. Norris, J. *et al.* Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* **270**, 22777–22782 (1995).
89. Szabo, Z. *et al.* Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution. *J. Mol. Evol.* **49**, 664–671 (1999).
90. Slagel, V., Flemington, E., Traina-Dorge, V., Bradshaw, H. & Deininger, P. Clustering and subfamily relationships of the Alu family in the human genome. *Mol. Biol. Evol.* **4**, 19–29 (1987).
One of the first reports of subfamily structure in Alu elements.
91. Waldman, A. S. & Liskay, R. M. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* **8**, 5350–5357 (1988).
92. Lobachev, K. S. *et al.* Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.* **19**, 3822–3830 (2000).
93. Stenger, J. E. *et al.* Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome Res.* **11**, 12–27 (2001).
94. Gebow, D., Miselis, N. & Liber, H. L. Homologous and nonhomologous recombination resulting in deletion: effects of p53 status, microhomology, and repetitive DNA length and orientation. *Mol. Cell. Biol.* **20**, 4028–4035 (2000).
95. Hsu, S. J., Erickson, R. P., Zhang, J., Garver, W. S. & Heidenreich, R. A. Fine linkage and physical mapping suggests cross-over suppression with a retroposon insertion at the *npc1* mutation. *Mamm. Genome* **11**, 774–778 (2000).
96. Rieder, M. J., Taylor, S. L., Clark, A. G. & Nickerson, D. A. Sequence variation in the human angiotensin converting enzyme. *Nature Genet.* **22**, 59–62 (1999).
97. Arcot, S. S. *et al.* High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. *J. Mol. Biol.* **281**, 843–856 (1998).
98. Brookfield, J. F. Selection on Alu sequences? *Curr. Biol.* **11**, 900–901 (2001).
99. Iizuka, M., Jones, C., Hayashi, K. & Sekiya, T. Mapping of 28 (CA)*n* microsatellite repeats and 13 Alu markers on human chromosome 11 using a panel of somatic cell hybrids. *Genomics* **19**, 581–584 (1994).
100. Schlotterer, C. & Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**, 211–215 (1992).
101. Levinson, G. & Gutman, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203–221 (1987).
102. Justice, C. M. *et al.* Phylogenetic analysis of the Friedreich ataxia GAA trinucleotide repeat. *J. Mol. Evol.* **52**, 232–238 (2001).
103. Campuzano, V. *et al.* Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**, 1423–1427 (1996).
104. Knight, A. *et al.* DNA sequences of Alu elements indicate a recent replacement of the human autosomal genetic complement. *Proc. Natl Acad. Sci. USA* **93**, 4360–4364 (1996).
105. Ryan, S. C., Zielinski, R. & Dugaiczyk, A. Structure of the gorilla α-fetoprotein gene and the divergence of primates. *Genomics* **9**, 60–72 (1991).
106. Nishio, H., Hamdi, H. K. & Dugaiczyk, A. Genomic expansion across the albumin gene family on human chromosome 4q is directional. *Biol. Chem.* **380**, 1431–1434 (1999).
107. Bailey, W. J. *et al.* Molecular evolution of the ψε-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol. Biol. Evol.* **8**, 155–184 (1991).
108. Koop, B. F. *et al.* Tarsius δ- and β-globin genes: conversions, evolution, and systematic implications. *J. Biol. Chem.* **264**, 68–79 (1989).
109. Kass, D. H., Batzer, M. A. & Deininger, P. L. Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Mol. Cell. Biol.* **15**, 19–25 (1995).
110. Roy, A. M. *et al.* Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res.* **10**, 1485–1495 (2000).
In this paper, the authors provide an initial estimate of the impact of gene conversion on the sequence diversity of Alu elements.
111. Maeda, N., Wu, C. I., Bliska, J. & Reneke, J. Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. *Mol. Biol. Evol.* **5**, 1–20 (1988).
112. Hayakawa, T., Satta, Y., Gagneux, P., Varki, A. & Takahata, N. Alu-mediated inactivation of the human CMP-*N*-acetylneuraminic acid hydroxylase gene. *Proc. Natl Acad. Sci. USA* **98**, 11399–11404 (2001).
113. Hoff, E. F., Levin, H. L. & Boeke, J. D. *Schizosaccharomyces pombe* retrotransposon Tf2 mobilizes primarily through homologous cDNA recombination. *Mol. Cell. Biol.* **18**, 6839–6852 (1998).
114. Ardlie, K. *et al.* Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am. J. Hum. Genet.* **69**, 582–589 (2001).
115. Frisse, L. *et al.* Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831–843 (2001).
116. Rubin, C. M., VandeVoort, C. A., Teplitz, R. L. & Schmid, C. W. Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res.* **22**, 5121–5127 (1994).
117. Schmid, C. W. Human Alu subfamilies and their methylation revealed by blot hybridization. *Nucleic Acids Res.* **19**, 5613–5617 (1991).
118. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499–1504 (1980).
119. Liu, W. M., Maraia, R. J., Rubin, C. M. & Schmid, C. W. Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res.* **22**, 1087–1095 (1994).
120. Liu, W. M. & Schmid, C. W. Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. *Nucleic Acids Res.* **21**, 1351–1359 (1993).
121. Li, T. & Schmid, C. W. Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* **276**, 135–141 (2001).
122. Liu, W. M., Chu, W. M., Choudary, P. V. & Schmid, C. W. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.* **23**, 1758–1765 (1995).
123. Schmid, C. W. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26**, 4541–4550 (1998).
An interesting discussion of the evidence for potential functional roles for Alu sequences.
124. Li, T., Spearow, J., Rubin, C. M. & Schmid, C. W. Physiological stresses increase mouse short interspersed element (SINE) RNA expression *in vivo*. *Gene* **239**, 367–372 (1999).

Acknowledgements
Research on mobile elements in the Batzer and Deininger labs is supported by the National Institutes of Health, Department of the Army, Louisiana Board of Regents Millennium Trust Health Excellence Fund and the Office of Justice Programs, National Institute of Justice, Department of Justice. The points of view in this document are those of the authors and do not necessarily represent the official position of the US Department of Justice.

Online links

DATABASES
The following terms in this article are linked online to:
LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink
α-fetoprotein | albumin | CMP-*N*-acetylneuraminic acid hydroxylase | frataxin | *TP53*
OMIM: http://www.ncbi.nlm.nih.gov/Omim
α-thalassaemia | acute myelogenous leukaemia | Apert syndrome | breast cancer | C3 deficiency | cholinesterase deficiency | complement deficiency | Ewing sarcoma | familial hypercholesterolaemia | Friedreich ataxia | haemophilia | insulin-resistant diabetes type II | Lesch–Nyhan syndrome | neurofibromatosis | Tay–Sachs disease

FURTHER INFORMATION
Batzer laboratory: http://batzerlab.lsu.edu
Deininger laboratory: http://129.81.225.52/
Dolan DNA Learning Center, Cold Spring Harbor Laboratory — Genetic Origins and Alu Insertion Polymorphism:
http://www.geneticorigins.org/geneticorigins/pv92/aluframeset.htm
Genetic Information Research Institute:
http://www.girinst.org/index.html
Access to this interactive links box is free online.

- Alu elements are a class of short interspersed elements (SINEs) that have expanded to a copy number of more than one million elements in primate genomes.
- The expansion of Alu elements is characterized by the dispersal, in a series of subfamilies, of elements of different evolutionary age that share common nucleotide substitutions.
- Alu elements have an impact on the genome in several ways, including insertion mutations, recombination between elements, gene conversion and gene expression.
- The human diseases caused by Alu insertions include neurofibromatosis, haemophilia, familial hypercholesterolaemia, breast cancer, insulin-resistant diabetes type II and Ewing sarcoma.
- Alu elements alter the distribution of methylation and, possibly, transcription of genes throughout the genome.
- The transcription of Alu elements changes in response to cellular stress and might be involved in maintaining or regulating the cellular stress response.
- Alu elements are a primary source for the origin of simple sequence repeats in primate genomes.
- Alu-insertion polymorphisms are a boon for the study of human population genetics and primate comparative genomics because they are neutral, identical-by-descent genetic markers with known ancestral states.

Mark Batzer received his Ph.D. from the laboratory of William R. Lee at Louisiana State University (LSU), USA. He carried out postdoctoral studies with Prescott Deininger at LSU Health Sciences Center, and then with Pieter de Jong in the Human Genome Center at Lawrence Livermore National Laboratory. He became a staff scientist at Lawrence Livermore National Laboratory and then assumed a faculty position in the Department of Pathology at the LSU Health Sciences Center in 1995. He subsequently accepted a position as Professor of Biological Sciences at LSU in 2001. His laboratory focuses on comparative genomics, population genetics, human molecular genetics and the contribution of mobile elements to genomic diversity.

Prescott Deininger received his Ph.D. from the laboratory of Carl Schmid at University of California (UC), Davis, USA. He carried out postdoctoral studies with Theodore Friedmann at UC, San Diego, and then with Frederic Sanger at the Medical Research Council in Cambridge, UK. He assumed a faculty position at LSU Health Sciences Center in 1981 and moved to a position as Associate Director of the Tulane Cancer Center in 1998. He holds the Marguerite Main Zimmerman Chair in Basic Cancer Research and is Professor of Environmental Health Sciences at the Tulane University Health Sciences Center. His laboratory focuses on the mechanism and impact of mobile elements, particularly SINEs, which cause instability of the mammalian genome.

URLs

Databases

LocusLink

α-fetoprotein
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=174

albumin
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=213

CMP-*N*-acetylneuraminic acid hydroxylase
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=8418

frataxin
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=2395

TP53
http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=7157

OMIM

α-thalassaemia
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?141800

acute myelogenous leukaemia
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?601626

Apert syndrome
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?101200

breast cancer
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?114480

C3 deficiency
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?120700

cholinesterase deficiency
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=177400

complement deficiency
http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=106100

Ewing sarcoma
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?603259

familial hypercholesterolaemia
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?143890

Friedreich ataxia
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?229300

haemophilia
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?306700

insulin-resistant diabetes type II
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?125853

Lesch–Nyhan syndrome
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?300322

neurofibromatosis
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?162200

Tay–Sachs disease
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?272800