Genome Dynamics in Aging Mice

Martijn E.T. Dollé[1,3] and Jan Vijg[1,2]

[1]Sam and Ann Barshop Center for Longevity and Aging Studies, University of Texas Health Science Center, San Antonio, Texas 78245, USA; [2]Geriatric Research Education and Clinical Center, South Texas Veterans Health Care System, San Antonio, Texas 78229, USA

Random spontaneous genome rearrangements are difficult to detect in vivo, especially in postmitotic tissues. Using a lacZ-plasmid reporter mouse model, we have previously presented evidence for the accumulation of large genome rearrangements in various tissues, including postmitotic tissues, during aging. These rearrangements, which were found to be organ-specific and to increase with age, have one breakpoint in the lacZ-reporter locus and the second elsewhere in the mouse genome. In this present work, we have used a mouse genome sequence database to physically characterize a total of 49 genome rearrangements in the brain, heart, and liver from young and old mice at two lacZ-plasmid reporter loci. Half of all breakpoints in the mouse genome occurred in chromosomes 3 and 4, each carrying a lacZ-reporter cluster, at distances varying from <100 kb to 66 Mb, indicating intrachromosomal deletions or inversions. The other half of the breakpoints in the mouse genome was found randomly on any of the other chromosomes, indicating translocations. Alternatively, part of the intra- and extrachromosomal events could involve transpositions. Regions of extended homology were not found at the breakpoints. These results lead us to postulate potential mechanisms for the origin of large genome rearrangements in mouse tissues and to predict their possible impact as a potential cause of aging.

Somatic mutations are thought to play a major role in cancer and possibly aging (DePinho 2000; Vijg 2000). To monitor tissue-specific patterns of somatic mutation accumulation during aging, a plasmid transgenic mouse model sensitive to a broad range of mutational events has been developed (Boerrigter et al. 1995). These mice harbor chromosomally integrated plasmids that can be efficiently recovered from genomic DNA and transferred into a suitable *Escherichia coli* host for mutant selection, quantitation, and characterization. Using this model, we have previously reported organ-specific differences in mutation accumulation with age (Dollé et al. 1997, 2000). A major advantage of this system is its sensitivity for genome rearrangements. Indeed, we have demonstrated that in addition to point mutations and small deletions, large rearrangements with one breakpoint in the lacZ-reporter gene and one in the mouse genome contribute significantly to the spontaneous mutant spectra (Dollé et al. 1997, 2000). Such large mutations are more likely to contribute to age-related cellular degeneration and death than randomly induced point mutations, for example, by haploidization of entire genomic regions or disruption of higher order genome organization (Vijg 2000; Vijg and Dollé 2002).

In view of the potentially high functional impact of genome rearrangements, characterization of such mutations could provide a better understanding of genome dynamics and its possible role in the aging process and in age-related diseases. Until recently, characterization of the rearrangements detected at the lacZ-reporter loci was limited to cumbersome genetic mapping, yielding little or no sequence information regarding the breakpoints in the mouse genome (Dollé et al. 2000). With the availability of a near complete mouse genome sequence database (Marshall 2001), physical locations and sequence information surrounding the breakpoints can now be readily obtained.

Here we present complete physical characterization of 49 spontaneous genome rearrangements recovered from somatic tissues of the mouse. Intrachromosomal mutations, up to 66 Mb in size, comprised 53% of all genome rearrangements analyzed. The remaining 47% may have resulted from translocations. No regions of extended homology were found at the breakpoints.

RESULTS

The lacZ-plasmid transgenic mice used in this study were homozygous for two integration sites, one on chromosome 3 and one on chromosome 4 (Dollé et al. 2000). The transgenes are integrated as arrays of multiple plasmids arranged in a head to tail fashion (Fig. 1). Previously, mutant frequencies and spectra at the lacZ-plasmid reporter loci were determined in the brain, liver, heart, and small intestine of mice of different ages (Dollé et al. 1997, 2000). Individual plasmids were obtained by digesting genomic DNA with HindIII, cutting only once in the wild-type plasmid sequence (Fig. 1). Subsequently, the plasmids were purified using lacI-coated magnetic beads, circularized through self-ligation and electro-transferred into a ΔlacZ, galE⁻ E. coli host for mutant selection. The recovered mutant plasmids were characterized by restriction enzyme analysis on agarose gels. Mutant plasmids with similar restriction patterns as wild-type plasmids, such as point mutations and small deletions, were classified as no-change mutants, whereas plasmids with altered restriction patterns were termed size-change mutants. About 83% of the size-change mutants have been found to represent genome rearrangements with one breakpoint in the transgene and one in the mouse genome (Dollé et al. 2000). These genome rearrangements were recovered as truncated plasmids attached to a piece of mouse genomic sequence (Fig. 1).

As indicated in Table 1, the frequency of genome rearrangements was found to vary from organ to organ, being the

[3]Corresponding author.
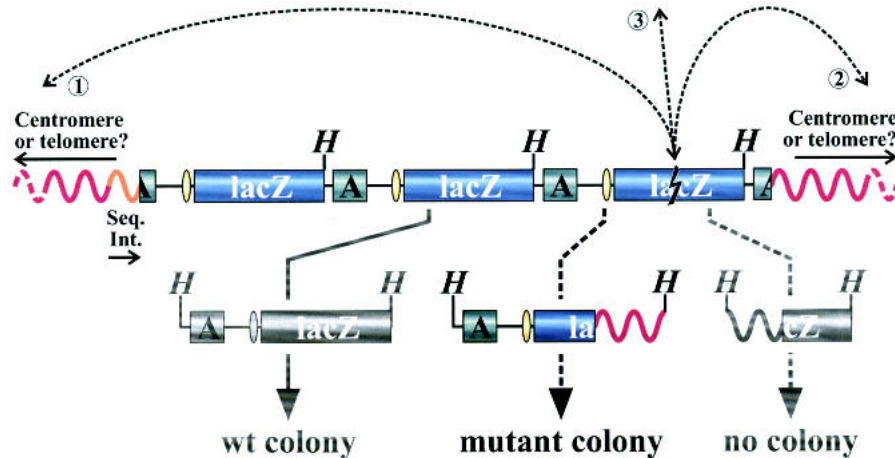E-MAIL dolle@uthscsa.edu; FAX (210) 562-5028.

Figure 1 Recovery of integrated *LacZ*-plasmids. Each integration site comprises a concatamer of ~10 head-to-tail organized plasmid copies (only three plasmid copies are depicted). Individual plasmid copies are released from mouse genomic DNA (red curving lines) by *Hin*dIII digestion (*H* indicates *Hin*dIII site). After purification, self-ligation, and transformation into *Escherichia coli* host cells, the plasmids are recovered in the form of ampicillin-resistant colonies. The dashed arrows represent the occurrence and detection of a hypothetical genome rearrangement with one breakpoint in a *lacZ*-reporter gene (lightning-sign) and one in the mouse genome 5′ (①) or 3′ (②) of the integration site or on another chromosome (③). When not deleted, only the upstream truncated plasmid sequence will result in a mutant colony after *Hin*dIII digestion, as the origin of replication (yellow ellipse) and the ampicillin-resistance gene (A) are needed for selective propagation of the transformed *E. coli* host. The orientation of the cloned 5′ sequences of the integration sites (Seq. Int.; orange curve), indicated by the arrow, corresponds with the direction of the blue arrowheads in Fig. 3 and is arbitrary within this scheme. The physical location and orientation of the cloned 5′ integration site sequence and the recovered mouse genomic sequence in the mutant plasmid allow for characterization of the recovered genome rearrangement. See text for more details.

lowest in the brain and the highest in heart. In addition, a significant increase in the frequency of genome rearrangements was found with age in the heart and liver but not in the brain or small intestine. Assuming that the *lacZ*-plasmid reporter locus is representative for the genome overall, we extrapolated the genome rearrangement frequencies from the 3-kb *lacZ* transgene to the entire 6×10^9-bp diploid genome. The outcome was subsequently divided by a factor of two, because only one of the two breakpoints needs to occur in a *lacZ*-reporter gene. The results indicate surprisingly high numbers of five to 37 genome rearrangements per cell, depending on the tissue of origin and donor age (Table 1), which prompted us to characterize a representative portion of these mutants in more detail.

Mouse sequences present in 49 mutant *lacZ*-plasmids, recovered from the brain, liver, and heart of untreated mice of different ages, were sequenced. Subsequent homology searches with these sequence fragments against a mouse genome database led to the characterization of the breakpoints in the mouse genome (Table 2). For a total of 11 breakpoints, a physical location within the genome could not be established: Five mutant plasmids comprised L1-repeat sequences; five recovered mouse genomic fragments were homologous to sequence fragments in the database of unknown location; and one genomic mouse fragment was too short (11 bases) for unambiguous identification. Of the remaining 38 analyzed mutants, a total of 20 (53%) recovered mouse sequences were situated on the chromosomes containing the *lacZ*-plasmid integration sites of this reporter mouse strain: nine and 11 mouse sequences on chromosomes 3 and 4, respectively (Fig. 2). These mutant plasmids presumably represent intrachro-

mosomal rearrangements. Mouse sequence fragments present in the remaining 18 (47%) mutant plasmids localized seemingly random to chromosomes other than 3 and 4 (Fig. 2) and were classified as translocations.

None of the characterized rearrangements showed extended regions of homology between the breakpoints in the plasmid and the mouse genome, perhaps with the exception of liver mutant Li133 sharing a 50% homology with the first 38 bases directly 5′ of the breakpoints (Table 2). Occasionally, direct homology at the breakpoints was found for 1 to 6 bp (Table 2). Furthermore, we did not detect any obvious recurring sequence patterns between the breakpoints of different rearrangements.

For further characterization, we took into account that in this mouse model, genome rearrangements can only be detected if the upstream region of the plasmid, containing the ampicillin resistance gene (*Amp^r*) and the origin of replication, remains intact (Fig. 1). The downstream region of a breakpoint in *lacZ* will not provide ampicillin resistance to the *E. coli* host cell during selection for transformed colonies (Fig. 1). The orientation of the integrated plasmids in combination with the location and orientation of the recovered mouse genomic sequence allowed us to deduce the simplest mutagenic event that could have resulted in the genome rearrangement detected.

We first determined the physical location and orientation of the integrated *lacZ*-plasmid concatamers. Previously, the 5′ sequences of both integration sites (Fig. 1) had been cloned and localized to chromosomes 3 and 4 by both fluorescence in situ hybridization (FISH) and genetic mapping (Dollé et al. 2000). Part of these flanking sequences and the first 20 bases of the first plasmid copy of each integrated transgene cluster are given in Table 2. Homology searches with the 5′ flanking sequences of the integration sites against the mouse genomic database confirmed localization on chromosomes 3 and 4 (Table 2). The plasmid concatamers are positioned 59.5 and 87.2 Mb from the centromere on chromosome 3 and 4, respectively (Fig. 3). The orientation of the integrated plasmids could now be inferred from the orientation of the 5′ flanking sequence of the integration sites (Fig. 1). The blue arrowheads in Figure 3 indicate the order of *Amp^r*, origin of replication, and *lacZ*: proximal (toward the centromere) on chromosome 3 and distal (toward the telomere) on chromosome 4.

The physical locations and orientations of recovered mouse genomic sequences localized on chromosomes 3 and 4 relative to the integrated plasmids on the respective chromosome are also indicated in Figure 3. For both chromosomes, mouse sequences were recovered proximal and distal to the integration sites, with an average distance of ~20 Mb and up

Table 1. Mutant Frequencies and Genome Rearrangements in Organs of Young and Old Mice

Organ	Age (mo)	LacZ mutant frequency ($\times 10^{-5}$)[a]		Genome rearrangements per cell[b]	P[c]
		total	genome rearrangements		
Brain	4	3.1 ± 0.3	0.5 ± 0.1	5	0.7333
	32	5.0 ± 1.3	0.6 ± 0.4	6	
Liver	4	3.1 ± 0.3	0.9 ± 0.2	9	0.0140
	32	10.5 ± 2.4	2.7 ± 1.0	27	
Heart	3	4.1 ± 1.4	1.9 ± 0.8	19	0.0262
	32	10.1 ± 2.7	3.7 ± 2.0	37	
Small intestine	3	6.1 ± 1.8	1.5 ± 0.5	15	0.9999
	32	19.5 ± 1.0	2.4 ± 2.1	24	

[a]Mean ± SD.
[b]Average number of genome rearrangements per diploid genome, based on a 3-kb target locus, a 6×10^6 kb diploid genome, and a twofold chance of a breakpoint falling within a reporter gene.
[c]Significance between age groups within organs for genome rearrangements using the Wilcoxon rank sum test.

to 66 Mb. The orientation of the recovered mouse sequences did not correlate with a proximal or distal location relative to the integrated plasmid cluster and its orientation on either chromosome (Fig. 3).

As mentioned above, by taking into account that for a genome rearrangement to be detected, the 5′ plasmid sequence of the breakpoint in *lacZ* must remain intact and end immediately in front of the recovered mouse sequence, the simplest intrachromosomal mutation that could have taken place was inferred (Fig. 4). Rearrangements with breakpoints in the mouse genome on either site of the integrated plasmid concatamer, but with reversely oriented sequences, could be inversions (Fig. 4). Rearrangements in the direction of the integrated plasmids, proximal for chromosome 3 and distal for chromosome 4 (Fig. 3), with similarly oriented breakpoints in the mouse genome, could be deletions (Fig. 4). Rearrangements in the reverse direction of the integrated plasmids, with reversely oriented mouse sequences, are more complicated and might be owing to transpositions (Fig. 4). According to these schemes, half of the intrachromosomal rearrangements would have been inversions, whereas deletions and transpositions each made up one fourth (Fig. 3). Alternatively, these rearrangements could be explained by translocations involving the transgene clusters integrated on either the homolog or the other chromosome.

The location of breakpoints on chromosomes other than 3 and 4 (Fig. 2, blue bars) were found throughout the chromosomes. There was no correlation between chromosomal location and orientation; that is, both distal and proximal oriented fragments were found on distal and proximal locations of the chromosomes (data not shown). Because of the contrasting orientation between integrated transgene clusters on chromosomes 3 and 4 (Fig. 3), and the inability to identify which transgene cluster was involved in a particular recovered translocation, we cannot infer the nature of the underlying mechanisms, such as chromosome arm exchanges or chromosome fusions.

DISCUSSION

The detection and characterization of somatic genome rearrangements occurring at random in vivo at low frequencies has been difficult, especially in postmitotic tissues. For example, previous attempts to demonstrate site-specific recombination in the mouse brain using *lacZ*-based reporter constructs yielded conflicting results and left the question open of whether such processes occur in postmitotic tissues (Matsuoka et al. 1991; Schatz and Chun 1992). In this present work, using a different type of transgenic reporter mouse in combination with the mouse genome sequence database, which has recently become available (Marshall 2001), we now provide sequence information on the breakpoints of rearrangements detected as part of mutant *lacZ* genes in a chromosomally integrated plasmid. The use of this system could help explain the mechanisms by which these events occur in different organs and tissues, why they are subject to organ-specificity, and why they increase with age in some but not in all organs. Surprisingly, the total number of rearrangements in the heart and liver at old age was found to be very high, that is, up to almost 40 events per cell in old heart (Table 1). Based on cytogenetic assays applied to metaphases in cultured cells, mostly lymphocytes, estimates of the frequency of chromosomal aberrations are generally 0.01 to 1 per cell, depending on cell type, age, and the laboratory where the analyses had been carried out. There are at least two explanations for this apparent discrepancy.

One possible explanation for the high number of genome rearrangements observed in this present study is that some or even most of the events scored by us as genome rearrangements are artifacts of the procedure applied to recover the mutant plasmids (Fig. 1). Although it is impossible to completely rule this out, we have addressed the possibility of artifacts extensively in a previous paper in which various control experiments had been performed on plasmids grown in *E. coli*, mixed with nontransgenic mouse genomic DNA, and mock-rescued into *E. coli*. Such experiments generally indicated significantly lower mutation frequencies in *E. coli* than in the mouse and no evidence for genome rearrangements as indicated by a mouse sequence at a *lacZ* breakpoint (Dollé et al. 1999b). Also, enhanced instability caused by the artificial nature of the *lacZ*-plasmid cluster in the mouse genome is unlikely to be responsible for the observed mutations. Indeed, neither the mutation frequencies nor their spectra are dramatically different from results reported with endogenous reporter genes such as HPRT, APRT, or HLA. Mutation frequencies at these loci were generally in the same range as our own values and also indicated a significant fraction of all mutations caused by genome rearrangements (Grist

Table 2. Breakpoints in the pUR288 Transgene and the Mouse Genome of the Two Integration Sites and a Representative Selection of Genome Rearrangements

Int. site or Mutant ID	Sequence (5'→3')[a]	Origin	Homologue Celera sequence fragment ID
Int. site A[b]	AGGATAGTCTGGCATAAGCCATGTACCCTCTTGACGAGGTCCTTTGTACTGCTCTTTCTTG- *TGCAGCAATGGCAACAACGT*	chr. 3	GA_x5J8B7W3G89:4000001..4500000
Int. site B[b]	CTATTCCACTGTTTTTCTAGCAAAAGAAATATACAAGACCTTCTAATAGAAATATTTTCAA- *AGCAATGGCAACAACGTTGC*	chr. 4	GA_x5J8B7W6711:3500001..4000000
He099	GCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGTT tcatctgtggtgcaacgggcgctgggtcggttacggccag Caaccctgagattccacctcaaaccagtcagaatggctaa GATCAAAAACTCAGGTGACAGAAGATGCTGGGGAGGATGT	pUR288 chr. 3	GA_x5J8B7W4KLR:1..500000
He199	CACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCG C gaggtgcggattgaaaatggtctgctgctgctgaacggca ttgaaatgtaaataaggaaaatatccaatgcccccccccc c CAAAAAAAAAAGAAAAAAAGACAACTGGAAGAAAATACAA	pUR288 chr. 3	GA_x5J8B7W6WFA:500001..1000000
Li063	ACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTG ttcccacggagaatccgacgggttgttactcgctcacatt gataaaagtaggcctacttaggataggaatgacaattgtc ACAGACATTCTACCTTTGTCCTACACTTCCCAACCCAGGA	pUR288 chr. 3	GA_x5J8B7W86NQ:3000001..3500000
Li133	ACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGA cgcgaattattttttgatggcgttaactcggcgtttcatct gcatagtatattggcaatagtttatggaaggaacaaatga ACACACAGTATGTATTGTCCTATAGCCATTACATCTCAAG	pUR288 chr. 3	GA_x5J8B7W3G89:4000001..4500000
He216	TTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCAT ggatgagcagacgatggtgcaggatatcctgctgatgaag gatgactggtggtgttgctatatcaaactgttcaagtgtg ATTGGAGAGATCTTGATATATCAGATGATGGGATATTCAG	pUR288 chr. 4	GA_x5J8B7W5YA9:2000001..2500000
He205	ACAGGAAACAGCTATGACCATGATTACGGATTCACTGGC C gtcgttttacaacgtcgtgactggtacaaccctggcgtta tctttcaagcagacgtctattcccggatctcagagacag c ACCACAGCTGCATCTCCTACCCCAGCCACCAATGAACCCC	pUR288 chr. 4	GA_x5J8B7W7V2M:500001..964248
Li049	CAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCA gttccgtttatccgggcaaaccatcgaagtgaccagcgaa atgcataagcctgtaaatatgttgatttttatttatatac AATATATCCATGTCCATAGAGAAAGGCCAATACTAATTAT	pUR288 chr. 4	GA_x5J8B7W6711:3500001..4000000
Li012	CTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAG cgcaacgcaattaatgtgagttagctcactcattaggcac tatgcaacacatattagtagctacttaaatccagacagaa AGGTTAGCTTCAACCTGTTGTGGCCAAATATAACTAGTCT	pUR288 chr. 4	GA_x5J8B7W6711:1500001..2000000
Br032	GCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGAC ttccagttcaacatcagccgctacagtcaacagcaactga gtagtcactcaacaaactctcatgggacaggcttctgggt GCTTGGGGACCCAGTGGTGAACTAGACTAGCAAGAGGGCC	pUR288 chr. 4	GA_x5J8B7W63N7:500001..1000000
Li153	CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAA tggcgctttgcctggtttccggcaccagaagcggtgccgg tgcaactccaggtgtctggcaggaagtgactttgtcagtg ACTTCAATTTTGCTAACTGCAAACACAACTTTGATCTCTA	pUR288 chr. 4	GA_x5J8B7W6C9L:500001..1000000
Li062	TGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGC T gataccgctcgccgcaacgaccgaccgagcgcagcgagt catccgaggggtcatatggaaacttaatacagtagaagc t TGTATGAGTGTGTGAGTGTGTGTGTGTTTGTGTGTGTGTG	pUR288 chr. 4	GA_x5J8B7W3DTS:1..500000
Li152	CGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGT TTC ccgactggaaagcgggcagtgagcgcaacgcaattaatgt catgtcttgggtcttcctaaggacgaaccagtctttta ttc TGTCATCCATCAGAGTCATAGCAGCCACTCTACAGAGGGT	pUR288 chr. 1	GA_x5J8B7W6CJB:1000001..1500000
Br050	GAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGT **GACTACC** tacgggtaacagtttctttatggccagggtgaaacgcaggt tatgtctttactcgcaggagagttcttgaagca **gactacc** g AGACCAGAAGTCTGCAGGAAATCAGGAACAGCCAGACTCA	pUR288 chr. 2	GA_x5J8B7W4QPD:11500001..12000000
Li134	GAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGC A**G** cgagtcagtgagcgaggaagcggaagagcgcccaatacgc acctgggtgggggggtgtcatgctccacagtgtcacaca a**g** TCAGTGACAGTAAAGGAATGCACCAGTCCTGTTCAGTCTG	pUR288 chr. 2	GA_x5J8B7W7E7U:1..500000
He273	GCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTAC A**CG** ctgtgcgaccgctacggcctgtatgtggtggatgaagcca actcttggccccagtgctatcagaaaacagcagtagg a**gg** AGGTCGTTCTCAGGATACCCAGCTAGAGCTGAGAGCACCAGT	pUR288 chr. 5	GA_x5J8B7W6G7P:2000001..2500000
Br140	CAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCG tcgttttacaacgtcgtgactggtacaaccctggcgttac tagccatattaatatctactaaaataaactttcaaccaaa ACTAATAATTAAAAAAAAAATGACAGGGAAGGACACTTCAT	pUR288 chr. 9	GA_x5J8B7W6RL5:15000001..15500000
He256	GCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTG G cacgacaggtttcccgactggaaagcgggcagtgagcgca ctacagagttgctaaatccttttctcagtaaaatgtatca g TCGCATTGAGAGGGACATCAAGTGAGTTCATACTTCAGCC	pUR288 chr. 10	GA_x5J8B7W65AK:2000001..2500000
He271	AAGCAGAACAACTTTAACGCCGTGCCGTGTTCGCATTATC cgaaccacgcgctgtggtacacgctgtgcgaccgctacgg gcaacttcccatatatcctaatagcctttcaattctacta TATCTGGTGAGTGGTGAGCTAGAGGGGAGGTTGAACCCTT	pUR288 chr. 11	GA_x5J8B7W7UC5:1500001..2000000
He180	ACCATGATTACGGATTCACTGGCCGTCGTTTTACAACG TC gtgactggtacaaccctggcgttacccaacttaatcgcct tgtctgaagaaagaaatatttgagaagatggaaaagtc tc CCATGCTTATGGATTGGCAGGATTAATAGTAAAAATGGCT	pUR288 chr. 13	GA_x5J8B7W5BHA:1..500000
Li061	TTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGG gtgaaacgcaggtcgccagcgggcaccgcgcctttcggcgg cttatagatgtgtgtaccaccacagctctgttgtttatttc ATTACACATGAAAGTCATTGACCTCTTTCCTATTTCCTCT	pUR288 chr. 15	GA_x5J8B7W6E9R:1..500000
Br057	ACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGC T**G** attaaatatgatgaaaacggcaacccgtggtcggcttacg atttgacaaaatccaaaacccattcatgataaaagtct t**g** GAAACATCAGGAATTCAAGGCCCATACCTAAACATAATAA	pUR288 chr. 17	GA_x5J8B7W7313:1000001..1500000
Br058	GTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGACAG T cgtttgccgtctgaatttgacctgagcgcatttttacgcg attttcaaatgatacaaactgatcatggattttttttttt t AAAGAAAGCATACTACAATGTCTCACATACATAAATCTTT	pUR288 chr. 19	GA_x5J8B7W7J22:6000001..6500000
He158	TGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGAT gagcgaacgcgtaacgcgaatggtgcagcgcgatcgtaat aggacatggaatatggcaagaaaactgaaaatcatggtaa ATGAGAAACATGCAGTTGACAACTTGAAAAAAAGACGAAAT	pUR288 chr. unkn.	GA_x5J8B7TUMJ6:1..675
He288	GCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCC G gcaccagaagcggtgccggaaagctggctggagtgcgatc gccccccccccccagctactttgctaaagcccccagggtt g TGGGATAGAGGGAAGAGGGAAGAAATTCCCAACACTGACA	pUR288 chr. unkn.	GA_x5J8B7TQ14U:1..698
Br060	GAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGC cggaaagctggctggagtgcgatcttcctgaggccgatac gctatggagagaaactaagtcagtcatctctacccactaa AAGAAACAAACAAAAGAAAACAAATAAACAGAAACAGCTT	pUR288 chr. unkn.	GA_x5J8B7TRYYM:1..1484
Br137	TGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTAACG cgaatggtgcagcgcgatcgtaatcacccgagtgtgatca acctaaaccacagaaagatccaacaaagatagagaacttc AGACCAATTTCTCTTATGAATATCGATGCAAAAATCCTCA	pUR288 chr. unkn. Line 1	
Br049	GAAACTGTGGAGCGCCGAAATCCCGAATCTCATCGTGCG gtggttgaactcgcacaccgccgacggcaggcgtgattgaag accccccactcccctgcccacccactccccctttttggccc TGGTATTCCCCTGTACTGGGGCATATAAAGTTTGCAAGTC	pUR288 chr. unkn. Line 1	
Br049	ATTTCACACAGGAAACAGCTATGACCATGATTACGGAT T**C** actggccgtcgttttacaacgtcgtgactggtacaaccct gtgggttcatttctggatcttcaattctattccattgg tc TACTTGTCTGTCTCTATACCAGTACCATGCAGTTTTTATC	pUR288 chr. unkn. Line 1	
Li36H3	GTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGATCAG G atatgtggcggatgagcggcattttccgtgacgtctcgtt atttgcaaattcatctggaataacaaaaaacctaggata g CAAAAACTCTTCTCAAGGATAAAAGAACCTCTGGTGGAAT	pUR288 chr. unkn. Line 1	
Li107	GCGATCAGTTCACCCGTGCACCGCTGGATAACGACAT TGG cgtaagtgaagcgacccgcattgaccctaacgcctgggtc tgaattttttgatcttagccattctgactggtgtgagg tgg AATCTCAGGGTTGTTTTGATTTGCATTTCCCTGATGATTA	pUR288 chr. unkn. Line 1	

[a]The capitalized nucleotides represent the recovered mutant sequence. Bold letters indicate direct homology between the breakpoint in the transgene and the mouse genome.
[b]Mouse sequence directly 5' of the integration site. The hyphen represents the transition to the integrated pUR288 sequence 3' of the *PstI* site (in italics).
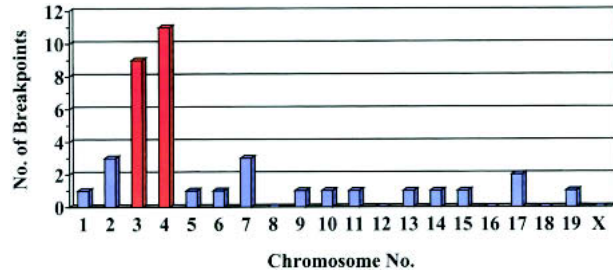
Figure 2 Chromosomal location of genomic mouse fragments recovered as parts of mutated plasmids from the brain, heart, and liver. The bars for chromosomes 3 and 4, containing the plasmid integration sites, are in orange.

et al. 1992; Dempsey et al. 1993; Stambrook et al. 1996; Albertini 2001). In general, mutation frequencies at HPRT were among the lowest, possibly because of selection against HPRT mutant lymphocytes in vivo. Interestingly, although virtually all results obtained with HPRT and other endogenous reporters involved lymphocytes, in a study using human kidney cells, significantly higher mutation frequencies were found, that is, up to ~4×10^{-4}, than in lymphocytes (Martin et al. 1996; Colgin et al. 2002). This could reflect a significantly lower selection pressure operating on kidney cells than in lymphocytes. About 15% of these *HPRT* mutations were genome rearrangements such as deletions. Based on the 44-kb target size of *HPRT*, a similar extrapolation as performed for the *lacZ*-reporter gene resulted in up to four genome rearrangements per kidney cell, which might be an underestimate owing to the lethality of such events at this X-linked locus. Preliminary data on the same *lacZ*-reporter construct, but now integrated as a single copy transgene, in *Drosophila* show a similar or even higher frequency of genome rearrangements, also indicating that the concatamer of constructs in the current mouse model is not intrinsically less stable than a single copy transgene. Finally, the observed organ specificities and age-related increase make it highly unlikely that a significant fraction of the mutants scored in our system as genome rearrangements are artifacts.

A second possible explanation for the discrepancy between (1) genome rearrangements scored as *lacZ* mutants and (2) cytogenetic methods involves the nature of the different assays. Cytogenetic methods have thus far mainly been applied to cultured cells, such as lymphocytes. Although it has recently become possible to apply cytogenetic methods, such as FISH, on interphase cells (Slovak et al. 2001), no results are available for mouse organs and tissues to directly compare with our results. Indeed, the difficulty in analyzing genome rearrangements with cytogenetic methods or methods based on endogenous reporter genes was the reason to generate the *lacZ* plasmid model system in the first place (Gossen and Vijg 1993). Also the sensitivity of cytogenetic methods to detect rearrangements involving <2 Mb is low. Indeed, possibly the only types of events scored in our system that

should be readily detectable by cytogenetic means are large deletions and translocations. However, because our scoring system is based on small mouse sequence fragments, we cannot distinguish deletions and translocations from transpositions. The latter would go undetected by cytogenetic methods. Although we cannot exclude that at least a part of the recovered mutants originated from transpositions, one would then also expect to find transposons to insert into the transgene cluster. Genomic insertion mutations have thus far not been detected in the *lacZ*-plasmid reporter (Martin et al. 2001) and appeared to be extremely rare in reporter gene mouse models in general (de Boer et al. 1997). Moreover, if most of the rearrangements that were scored in our system had resulted from transpositions, it is not clear why half of them occur at chromosomes 3 and 4, where our reporter gene cluster has been integrated (Fig 2). For these reasons, we believe that although part of our rearrangements might reflect transpositions, most of them are indeed large deletions, inversions, or translocations.

With respect to the mechanisms leading to the recovered genome rearrangements, it is probable that they resulted from double-strand breaks (DSBs) in the DNA. DSBs are inflicted by a variety of processes and agents from both endogenous and exogenous sources, such as background radiation, oxygen radicals, and DNA processing enzymes. It has been shown that two simultaneously inflicted DSBs can lead to a variety of recombination processes (Richardson and Jasin 2000). Two homology-dependent and one homology-independent DSB repair pathways have been described for mammalian cells: single-strand annealing (SSA), homologous recombination (HR), and nonhomologous end joining (NHEJ), respectively (for review, see Pastink et al. 2001). The absence of (extended) homology at the breakpoints (Table 2) of the genome rearrangements characterized in the present study implies that NHEJ could have been involved. Our observation that macrohomology was mainly limited to an occasional direct homology of one to six base pairs between breakpoints (Table 2) is in agreement with a recent study by Sandoval and Labhart (2002), which shows that macrohomology is not required for the end joining process.

SSA of direct repeat sequences around a breakpoint is thought to require a minimum of 30-bp homology (Sugawara et al. 2000). Clearly, such breakpoints have not been found in this study (Table 2). It is questionable whether the plasmid transgenic mouse model is suitable to detect SSA. Although
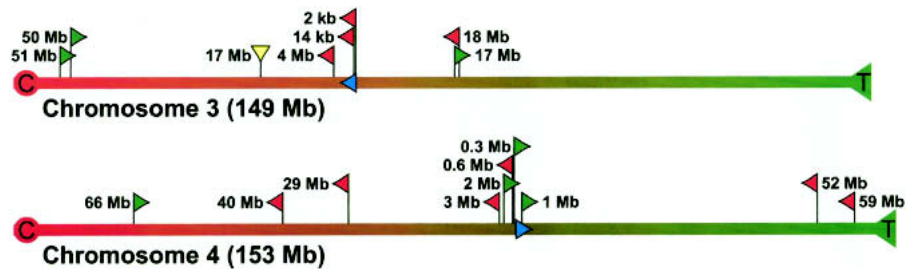

Figure 3 Schematic representation of physical locations and orientations of the integrated transgene clusters and genomic mouse fragments, recovered as parts of mutant plasmids, on chromosomes 3 and 4. The red (proximal) and green (distal) arrowheads indicate the direction of the recovered mouse fragments, such that the base and tip of the arrowhead represent the breakpoint and the *Hind*III site, respectively (yellow downward arrow indicates direction unknown). The blue arrow indicates the location and orientation of the integrated transgene cluster (see also Fig. 1). The rounded physical distances from the integration site to the localized breakpoints in the mouse genome are shown. C indicates centromere; T, telomere.
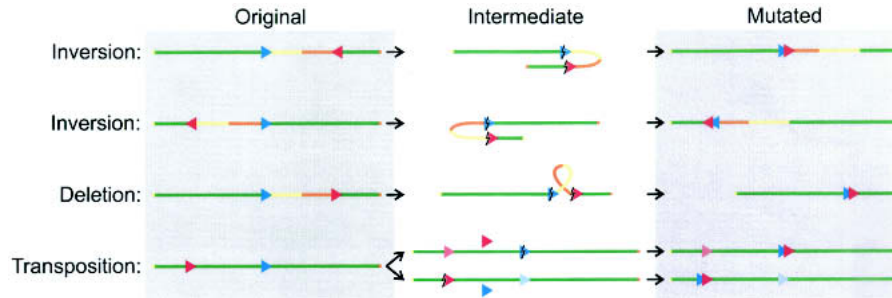
Figure 4 Schematic depiction of proposed mechanisms for observed intrachromosomal rearrangements. The blue and red arrows indicate the orientation of the integrated plasmid loci and the recovered mouse sequences, respectively, on the original non-rearranged chromosome (*left* column). All four combinations are given for an arbitrarily orientated chromosome (green line). The *middle* column shows how two breakpoints (lightning signs) could lead to the inversion or deletion of the encompassed chromosomal sequence (yellow-orange dual tone line) and result in a recoverable mutation in the *right* column. The last row indicates the two options for a transposition, in which either the transgene locus or the recovered mouse sequence is copied or excised (as indicated by the pink and light blue arrows) and integrates in the breakpoint at the other location.

regions of homology between *lacZ* and the mouse genome exist, the most obvious direct repeats around a breakpoint in *lacZ* are the surrounding plasmid copies of the integrated concatamer. Thus, SSA might resolve a breakpoint in one of the transgenes simply by deleting one or more plasmid copies, without creating a mutation in the remaining plasmid copies. Only when slippage or incorporation errors are made during the repair process might SSA be detected in the form of a point mutation, a small deletion, or insertion. From among the many *lacZ* point mutants that we characterized in the past (Dollé et al. 2002), evidence has been obtained that such events may indeed occur. The *lacZ*-plasmid mouse model carries multiple nonmutant single nucleotide polymorphisms (SNPs) among its wild-type integrated plasmid copies (Dollé et al. 1999b), which allow differentiation between plasmid copies. Although some of these SNPs are linked (i.e., are present within the same plasmid copy), others are not. Occasionally, point mutations flanked by SNPs that have not shown any linkage, or point mutations flanked by a single SNP that was linked to another, have been found (data not shown) and indicate that recombination between plasmid copies had occurred.

It is unlikely that the plasmid transgenic mouse model, certainly when homozygous for the transgenes, detects HR. Like SSA, only when errors occur during the repair event, might a mutated transgene result. In contrast, HR resulting in loss of heterozygosity has been detected in ~48% of all spontaneous mutations at the *Aprt* locus in mouse lymphocytes (Liang et al. 2000). However, HR is a conservative pathway, essentially restoring the original organization of the chromosome. Hence, it seems that large structural changes in the genome mainly result from nonconservative pathways such as SSA and NHEJ.

An important question is whether the observed (up to threefold) induction in spontaneous genome rearrangements, found over the adult lifespan of the mouse (Table 1), could be a causative factor in the aging process (Warner and Johnson 1997). In this respect, it is important to realize that from an evolutionary perspective, maintenance of genome integrity after the period of first reproduction offers no selective advantage (Kirkwood and Austad 2000). It is therefore likely that already at reproductive maturity, the maximum number of somatic mutations, still compatible with optimal fitness, has

been reached. Even slight increases above this threshold are likely to adversely impact on the structure and function of an organism. Although the accumulation of point mutations is unlikely to have adverse effects, even a limited number of genome rearrangements, such as the ones described in this paper, could severely affect patterns of gene regulation, namely, through gene dosage or position effects (Vijg 2000; Vijg and Dollé 2002).

METHODS

Plasmid Recovery

Aging cohorts of male C57Bl/6 pUR288-*lacZ* mice of founder-line 60 were maintained in the animal facilities of the Beth Israel Deaconess Medical Center as described previously (Dollé et al. 1997). The animals were killed by decapitation following asphyxiation by CO_2. Organs and tissues were removed, rinsed in PBS, placed in 1.5-mL microcentrifuge tubes, and frozen on dry ice. Any macroscopic lesions observed during tissue collection were excised and stored separately. The tissues were maintained at −80°C until used. DNA was extracted by routine phenol/chloroform extractions.

Complete protocols for plasmid rescue and mutant frequency determinations with this model are given elsewhere (Dollé et al. 1996). Briefly, between 10 and 20 μg genomic DNA was digested with *Hin*dIII for 1 h in the presence of magnetic beads (Dynal) precoated with lacI-lacZ fusion protein. The beads were washed 3 times to remove the unbound mouse genomic DNA. Plasmids were subsequently eluted from the beads by Isopropylthio-β-D-glastocide (IPTG). After circularization of the plasmids with T4 DNA ligase, they were ethanol-precipitated and used to electro-transform *E. coli C* (Δ*lacZ*, *galE*⁻) cells. One thousandth of the transformed cells was plated on a titer plate (with X-gal); the remainder, on a selective plate (with p-gal). The plates were incubated for 15 h at 37°C. Mutant frequencies were determined as the number of colonies on the selective plates versus the number of colonies on the titer plate (times the dilution factor of 1000).

Mutant Classification

Mutant colonies were taken from the selective plates and grown overnight in 3 mL Luria-Bertani medium. Then, 1 μL was directly plated on X-gal–supplemented LB-agar medium to screen for galactose insensitive host cells (Dollé et al. 1999a). The remainder of the cell culture was used for plasmid mini preparation (Wizard 9600, Promega). The purified plasmids were digested with *Pst*I and *Ava*I and size-separated on 1% agarose gels. Mutant plasmids with restriction patters resembling and deviating from the wild-type restriction pattern were classified as "no-change" and "size-change" mutants, respectively. Size-change mutants with a fragment of the mouse genome as part of their mutated *lacZ* gene were called genome rearrangements, as identified by sequencing and/or a polymerase chain reaction screening method as described before (Dollé et al. 2000). To exclude possible cloning artifacts as a result of the accidental inclusion of mouse sequence fragments during plasmid recovery, size-change mutants were routinely digested with *Hin*dIII; only mutant plasmids with a unique *Hin*dIII site were considered in vivo derived genome rearrangements (Dollé et al. 1999b).

Sequencing

Sequence reactions were performed with the CEQ dye terminator cycle sequencing kit (Beckman), according to the manufacturer's standard protocol, and were analyzed with a CEQ 2000 DNA analysis system (Beckman). The primers used were the same as described earlier (Dollé et al. 1999a).

Physical Characterization of Genome Rearrangements

Physical locations, directionality, and missing breakpoint sequence information of the recovered mouse sequences within these mutant plasmids were obtained through blastn searches against the entire mouse genome in the Celera Discovery System and Celera Genomics' associated databases (www.celera.com). The cloned 5′ sequences of the integration sites (Dollé et al. 2000) were used as physical locations for the breakpoints in the *lacZ* transgene to calculate sizes of intrachromosomal rearrangements.

REFERENCES

Albertini, R.J. 2001. HPRT mutations in humans: Biomarkers for mechanistic studies. *Mutat. Res.* **489:** 1–16.

Boerrigter, M.E., Dollé, M.E., Martus, H.J., Gossen, J.A., and Vijg, J. 1995. Plasmid-based transgenic mouse model for studying in vivo mutations. *Nature* **377:** 657–659.

Colgin, L.M., Hackmann, A.F., Emond, M.J., and Monnat, Jr., R.J. 2002. The unexpected landscape of in vivo somatic mutation in a human epithelial cell lineage. *Proc. Natl. Acad. Sci.* **99:** 1437–1442.

de Boer, J.G., Suri, A., Walsh, D., and Glickman, B.W. 1997. The recovery of a B2 insertion in the lacI gene of a rat cell line containing a λ/LIZ shuttle vector. *Mutat. Res.* **375:** 79–82.

Dempsey, J.L., Pfeiffer, M., and Morley, A.A. 1993. Effect of dietary restriction on in vivo somatic mutation in mice. *Mutat. Res.* **291:** 141–145.

DePinho, R.A. 2000. The age of cancer. *Nature* **408:** 248–254.

Dollé, M.E., Martus, H.J., Gossen, J.A., Boerrigter, M.E., and Vijg, J. 1996. Evaluation of a plasmid-based transgenic mouse model for detecting in vivo mutations. *Mutagenesis* **11:** 111–118.

Dollé, M.E., Giese, H., Hopkins, C.L., Martus, H.J., Hausdorff, J.M., and Vijg, J. 1997. Rapid accumulation of genome rearrangements in liver but not in brain of old mice. *Nat. Genet.* **17:** 431–434.

Dollé, M.E., Martus, H.J., Novak, M., van Orsouw, N.J., and Vijg, J. 1999a. Characterization of color mutants in *lacZ* plasmid-based transgenic mice, as detected by positive selection. *Mutagenesis* **14:** 287–293.

Dollé, M.E., Snyder, W.K., van Orsouw, N.J., and Vijg, J. 1999b. Background mutations and polymorphisms in *lacZ*-plasmid transgenic mice. *Environ. Mol. Mutagen* **34:** 112–120.

Dollé, M.E., Snyder, W.K., Gossen, J.A., Lohman, P.H., and Vijg, J. 2000. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc. Natl. Acad. Sci.* **97:** 8403–8408.

Dollé, M.E., Snyder, W.K., Dunson, D.B., and Vijg, J. 2002. Mutational fingerprints of aging. *Nucleic Acids Res.* **30:** 545–549.

Gossen, J. and Vijg, J. 1993. Transgenic mice as model systems for studying gene mutations in vivo. *Trends Genet.* **9:** 27–31.

Grist, S.A., McCarron, M., Kutlaca, A., Turner, D.R., and Morley, A.A. 1992. In vivo human somatic mutation: Frequency and spectrum with age. *Mutat. Res.* **266:** 189–196.

Kirkwood, T.B. and Austad, S.N. 2000. Why do we age? *Nature* **408:** 233–238.

Liang, L., Deng, L., Shao, C., Stambrook, P.J., and Tischfield, JA. 2000. In vivo loss of heterozygosity in T-cells of B6C3F1 Aprt$^{+/-}$ mice. *Environ. Mol. Mutagen* **35:** 150–157.

Marshall, E. 2001. Genome sequencing: Celera assembles mouse genome; public labs plan new strategy. *Science* **292:** 822.

Martin, G.M., Ogburn, C.E., Colgin, L.M., Gown, A.M., Edland, S.D., and Monnat, Jr, R.J. 1996. Somatic mutations are frequent and increase with age in human kidney epithelial cells. *Hum. Mol. Genet.* **5:** 215–221.

Martin, S.L., Hopkins, C.L., Naumer, A., Dollé, M.E, and Vijg, J. 2001. Mutation frequency and type during ageing in mouse seminiferous tubules. *Mech. Ageing Dev.* **122:** 1321–1331.

Matsuoka, M., Nagawa, F., Okazaki, K., Kingsbury, L., Yoshida, K., Muller, U., Larue, D.T., Winer, J.A., and Sakano, H. 1991. Detection of somatic DNA recombination in the transgenic mouse brain. *Science* **254:** 81–86.

Pastink, A., Eeken, J.C., and Lohman, P.H. 2001. Genomic integrity and the repair of double-strand DNA breaks. *Mutat. Res.* **480/481:** 37–50.

Richardson, C. and Jasin, M. 2000. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* **405:** 697–700.

Sandoval, A. and Labhart, P. 2002. Joining of DNA ends bearing non-matching 3′-overhangs. *DNA Repair* **1:** 397–410.

Schatz, D.G. and Chun, J.J. 1992. V(D)J recombination and the transgenic brain blues. *New Biol.* **4:** 188–196.

Slovak, M.L., Tcheurekdjian, L., Zhang, F.F., and Murata-Collins, J.L. 2001. Simultaneous detection of multiple genetic aberrations in single cells by spectral fluorescence in situ hybridization. *Cancer Res.* **61:** 831–836.

Stambrook, P.J., Shao, C., Stockelman, M., Boivin, G., Engle, S.J., and Tischfield, J.A. 1996. APRT: A versatile in vivo resident reporter of local mutation and loss of heterozygosity. *Environ. Mol. Mutagen.* **28:** 471–482.

Sugawara, N., Ira, G., and Haber, J.E. 2000. DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae* RAD59 in double-strand break repair. *Mol. Cell. Biol.* **20:** 5300–5309.

Vijg, J. 2000. Somatic mutations and aging: A re-evaluation. *Mutat. Res.* **447:** 117–135.

Vijg, J. and Dollé, M.E. 2002. Large genome rearrangements as a primary cause of aging. *Mech. Ageing Dev.* **123:** 907–915.

Warner, H.R. and Johnson, T.E. 1997. Parsing age, mutations and time. *Nat. Genet.* **17:** 368–370.

WEB SITE REFERENCE

www.celera.com; Celera Discovery System and Celera Genomics' associated databases.