SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Correspondence in repeat number between human reference (hg18) alleles and modal alleles among 48 humans resequenced as part of the HapMap-ENCODE project. A total of 181 mono- and dinucleotide loci were examined. Population-modal allele and the reference allele were identical at all examined repeat numbers (3 to 6) of dinucleotide loci. For mononucleotide loci, such an identity was observed for most loci with less than 12 repeats; note that in our birth/death study we only examined mononucleotide loci with less than 10 repeats.

Figure S2. Special cases of microsatellite births and deaths. (**A**) Microsatellite resurrection. Ancient microsatellite was eliminated due to substitution, but was resurrected in *H* lineage by removal of the interrupting nucleotide, by substitution. (**B**) Multi-step event including a birth followed by subsequent death. (**C**) Illustration of microsatellite birth by partial motif insertion. (**D**) Examples of microsatellite loci whose mutational steps towards births/deaths could not be deciphered because the states of the two outgroup species (marmoset and macaque) differed, or because of large inter-species variation in locus lengths and/or due to close clustering of multiple mutations separating the loci's microsatellite/non-microsatellite states. For instance, in the first and the third examples, the outgroups have different birth/death status, whereas for the second example, the exact ancestry of the interrupting $[C]_n$ cannot be clearly determined due to lack of ancestral information at these locations.

Figure S3. Phylogenetic tree constructed from orthologous (A) *Alu*Y elements and (B) L1PA3/4/5/6 extracted from hg18, panTro2 and ponAbe2 genomes. The blue, red and green colored branches lead to orangutan, chimpanzee and human lineages, respectively. Grey circles mark branch points that have greater than 80% bootstrap support. We investigated whether gene conversion among members of individual *Alu* subfamilies (Aleshin and Zhi 2010; Batzer, et al. 1995; Zhi 2007) may influence the observed microsatellite birth/death trends. For this, we considered *AluY* elements only, as gene conversion is more likely to occur among elements of younger subfamilies due to lower sequence divergence (Roy, et al. 2000). Sequences of all *AluY*s that were inserted prior to the human-chimpanzee-orangutan divergence were obtained from all three species, and were aligned using CLUSTALW with default parameters. CLUSTALW was used for obtaining a neighbor-joining tree with 1000 bootstraps. Figures were generated using iTol (Letunic and Bork 2007). Of the 173 orthologous *AluY* elements that harbored microsatellite births, only 23 clustered together with other *Alu* elements within the same species and not with orthologous elements in related species (Fig. S2A). Similar results were obtained for deaths (data not shown). In a similar analysis for investigating gene conversion among young L1PA subfamilies, we examined tree topology of 300 bp 3' end of L1PA3/4/5/6 elements that harbored microsatellites births and deaths. Of the 218 L1PA3/4/5/6 elements that harbored microsatellite births, only 11 clustered together with other L1PA3/4/5/6 elements of the same species, and including these 11, only 51 elements did not cluster with their orthologs in other species (Fig. S2B).

Figure S4. Numbers of microsatellite birth and death events and stationary microsatellites in *H, C, HC* and *O* lineages mapping to (**A**) *Alu* elements and NCNR; (**B**) L1 elements and NCNR; and (**C**) coding exons and NCNR. Microsatellite thresholds used are [9,5,4,3]. See legend of Figure 4 for explanation of the color scheme.

Figure S5. Numbers of microsatellite birth and death events and stationary microsatellites in *H, C, HC* and *O* lineages mapping to 3' terminal 300 bp of different L1PA subfamilies, pooled by age. Microsatellite thresholds used are [9,5,4,3]. See legend of Figure 4 for the explanation of the color scheme. The majority of the biases in enrichment of births/deaths and stationary loci were statistically not significant (due to overall lower number of events considered for terminal 300 bps

than for full length elements), apart from the significant enrichment of oldest L1PA elements for deaths.

Figure S6. Frequencies per alignment representation of (**A**) microsatellite births, (**B**) microsatellite deaths, and for (**C**) stationary microsatellite loci along L1 consensus sequence divided into 60-bp bins. In each plot, the frequency bars are decomposed (indicated by different colors) into frequencies of loci mapping to different L1 families (plotted above the horizontal midline), and into GC-richness of the respective microsatellite loci (plotted below the horizontal midline).

Figure S7. Births/deaths in the disease-associated loci. Partial multiple alignment blocks are shown. Microsatellite sequences are indicated in blue color (and also in red color wherever more than one microsatellites were present) and human (hg18) sequences are underlined.

Figure S8. The regional variation in birth/death density (i.e., number of births/deaths per nucleotide) at the scale of 5-Mb windows along hg18 chromosome 1. Here, for births and deaths taking place along *C, HC* and *O* lineages, hg18 positions were marked.

Figure S9. Multiple logistic regressions predicting microsatellite births and deaths (pooled) as a function of regional genomic features. For each window size, performance (last row) is measured as percent reduction in null deviance, and significant predictors (P<0.05; body of the Table) are evaluated by their contribution to this reduction (see Methods), Also noted, in parentheses, are the significance values ($\log_{10}$ P value with Bonferroni correction for multiple tests applied) of each predictor in the respective regression models. Blue and brown cells represent negative and positive predictors, respectively. *This fit employed event windows with ≥4 births/deaths.

Figure S10. Results of multiple logistic regression for predicting microsatellite births and deaths, using (A) *Alu*Y content (and no other SINE elements) (B) *Alu*S content (and no other SINE elements); and (C) *Alu*J content (and no other SINE elements) as individual predictor. For each regression model, % Null Deviance reduced (the relative reduction of Null deviance by implementing all the model predictors) is written in bold. In each cell, details for models constructed without/with proto-microsatellites as predictors are mentioned. For each predictor, the "% contribution towards reduction of null deviance" is written when the predictor is significant (P value < 0.5) in the model. Red cells highlight positive, and blue cells highlight negative predictors.

Figure S11. Two illustrations depict modifications of multiZ alignments that were carried out wherever an identical nucleotide not belonging to a microsatellite motif was found in more than one species, on the condition that the immediately flanking motifs were identical (i.e., in the same phase) among all relevant species. In such cases, we re-aligned the interrupting nucleotide to the same alignment positions, to minimize over-estimation of substitutions and non-motif indels.

Figure S1

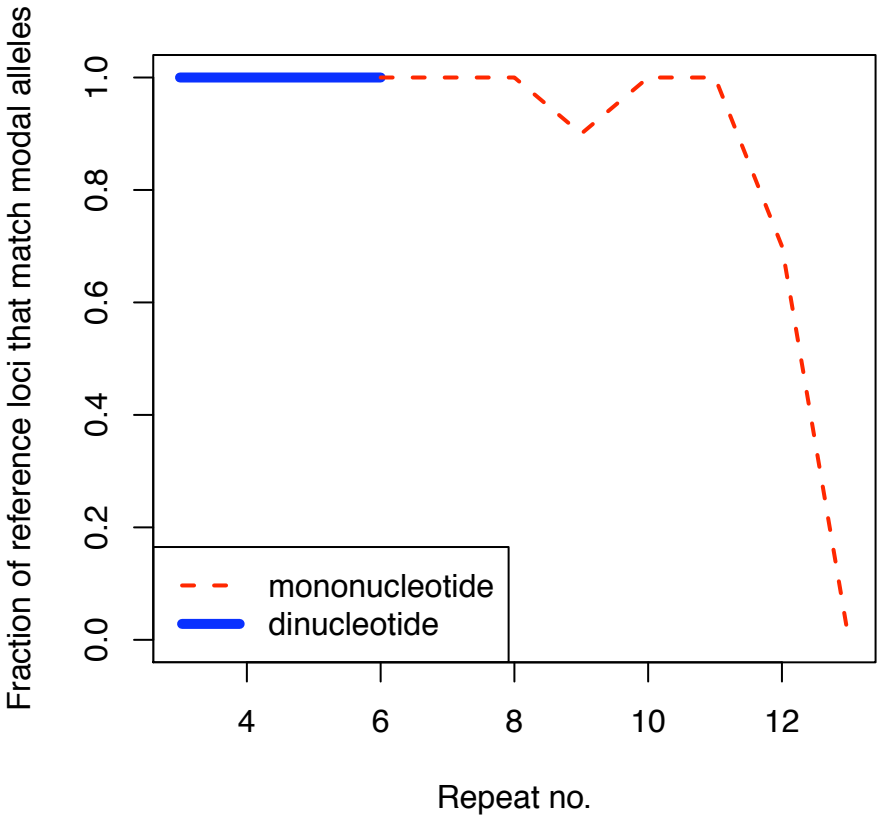
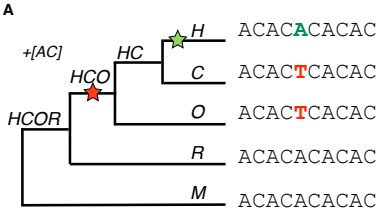
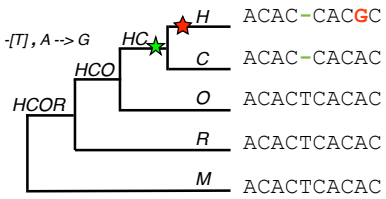

Fraction of reference loci that match modal alleles
0.0
0.2
0.4
0.6
0.8
1.0
4
6
8
10
12
Repeat no.
mononucleotide
dinucleotide

Figure S2

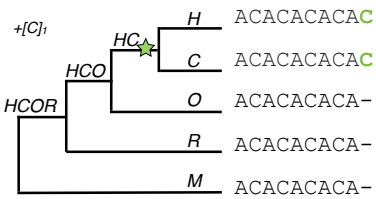
A



B



C



D

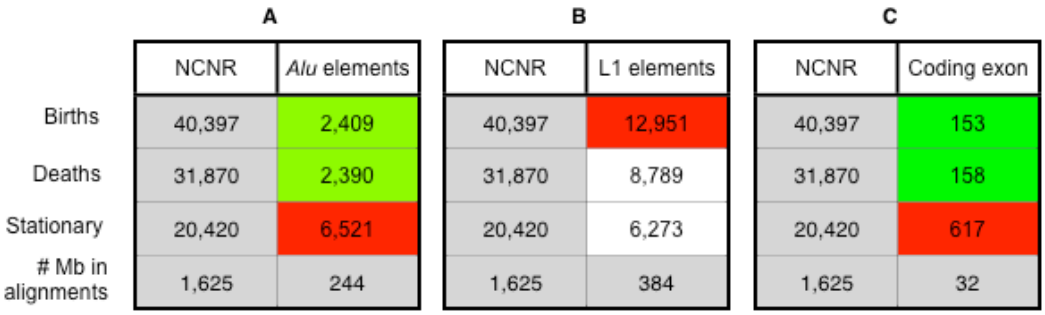
```
      Human  agttatgttaa---------------------------------------tttttttttttttttt----gagacagagtctcc
      Chimp  agttatgttaa-----------------------------------attttttttttttttttttt----gagacagagtctcc
  Orangutan  agttatgttaa-----------------------------------ttttttttttttttttttttt--------ttgagtctcc
     Rhesus  agttatgttaa---------------------------------------ttttttttctttttttgggggggacagagtctcc
   Marmoset  agttatgctaatataaatgggggttttttgtttgattttgttttttgttttttgttttct----gagacggagtctca


  Orangutan  AACA--CTATC---AAA----------AAAAAAAATGGCACA----
      Human  AACA--CTATTAAAAAA----------AAAACCAATGACACA----
      Chimp  AACA--CTATT---AAA----------AAAACCAATGGCACA----
     Rhesus  AACA--CTATC---AAAAGCAAAAACGAAAAACAATGGTACACACT
   Marmoset  AACAGTCTATT---TAA----------AA----AATGGCTCACATT


  Orangutan  GGGCTTTCA----gtttgtttgtttgtttgt---ttTTAATAGCTT
      Human  GGGCTTTCA----GTTTGTGTGTTTGTTTGT---TTTTAATAGCTT
      Chimp  GGGCTTTCAgtttgtttgtttgtttgtttgt---ttTTAATAGCTT
     Rhesus  GGGCCTTCA----GTTTG--TGTTTGTTTGT---TTTTAATAGCTT
   Marmoset  GGGCTTTCA----gtttgtttgtttgtttgtttatttTAATAGCTT
```

Figure S3. See a separate file.

Figure S4.



	A		B		C	
	NCNR	*Alu* elements	NCNR	L1 elements	NCNR	Coding exon
Births	40,397	2,409	40,397	12,951	40,397	153
Deaths	31,870	2,390	31,870	8,789	31,870	158
Stationary	20,420	6,521	20,420	6,273	20,420	617
# Mb in alignments	1,625	244	1,625	384	1,625	32

Figure S5.

	L1PA (all)	L1PA3/4/5/6	L1PA7/8/8a/10	L1PA11/12/13/14
Births	103	20	57	26
Deaths	136	14	65	57
Stationary	105	31	46	28
Total	344	65	168	111
# Mb in alignments	0.2	0.03	0.09	0.08

Figure S6.

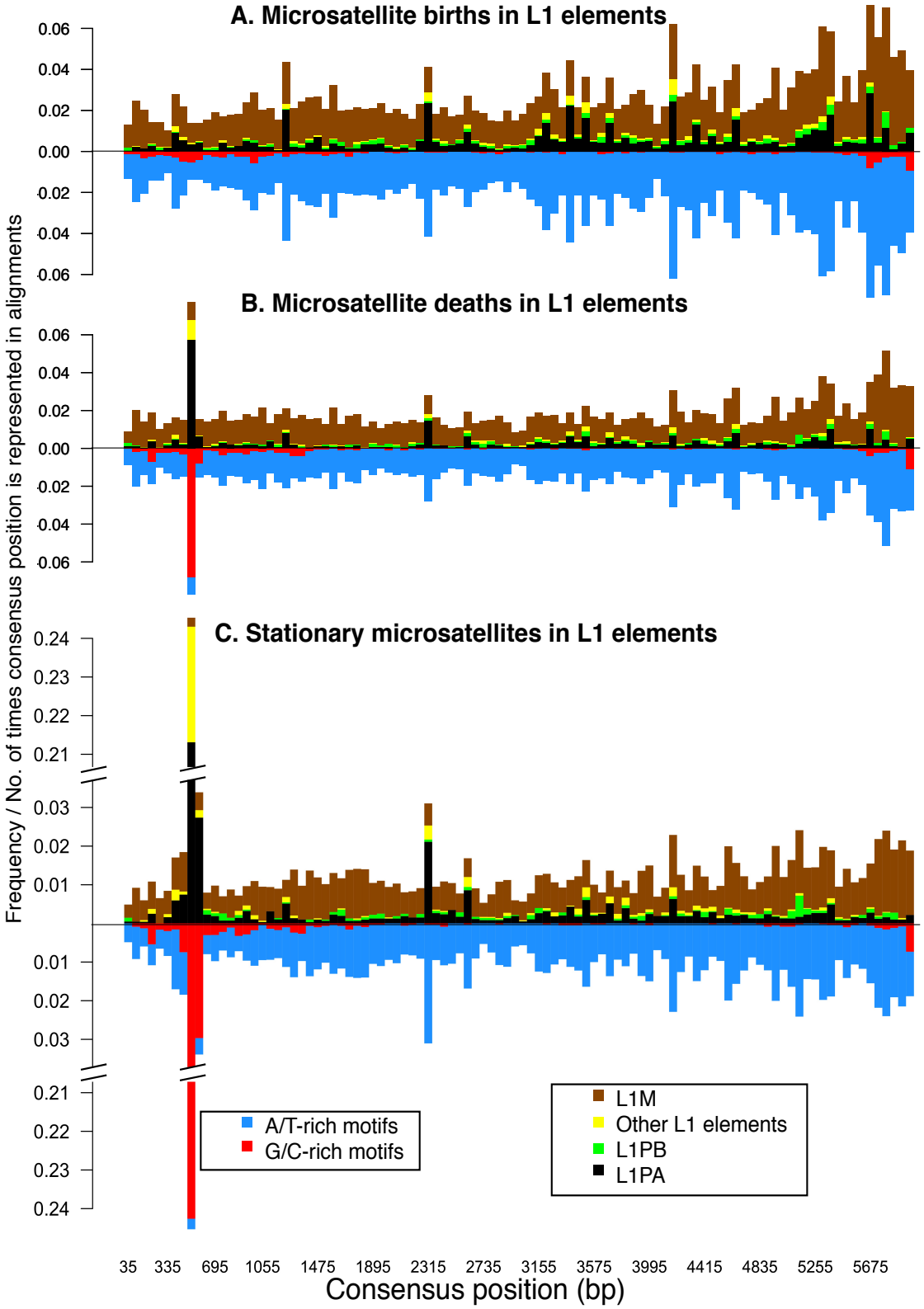

A. Microsatellite births in L1 elements
B. Microsatellite deaths in L1 elements
C. Stationary microsatellites in L1 elements
Frequency / No. of times consensus position is represented in alignments
0.06
0.04
0.02
0.00
0.02
0.04
0.06
0.24
0.23
0.22
0.21
0.03
0.02
0.01
A/T-rich motifs
G/C-rich motifs
L1M
Other L1 elements
L1PB
L1PA
35
335
695
1055
1475
1895
2315
2735
3155
3575
3995
4415
4835
5255
5675
Consensus position (bp)

Figure S7.

Spinocerebellar ataxia 7 (Ataxin-7)

```
##maf version=1
a score=521191.0
s hg18.chr3          63873382 69 + 199501827 agcggccgcggccgcccggcagcagcagcagcagcagcagcagcagccgccgcctccgcagcccca
s panTro2.chr3       65356498 57 + 203962478 agcggtcgcggccgcccggcagcagcagcagcagcagcagc------------ctccgcagcccca
s ponAbe2.chr3      120045492 54 - 202140232 agcagctgcggccgcccggcagcagcagcagcagcaacagc--------------ctccgcagcccca
s rheMac2.chr2      117258002 48 - 189746636 agcggccgcggccgcccggcagcagcagcagcagc--------------------ctccgcagcccca
s calJac1.Contig92     620222 57 +    730829 agcggctgcggccgctcggcggcagcagcagcagcagcaacagc------------cttcgcagcccca
```

OPMD: Oculopharyngeal Muscular Dystrophy (poly(A)-binding protein-2)

```
##maf version=1
a score=790599.0
s hg18.chr14         22860502 69 + 106368585 CCCAGTCTGAgcggcgatggcggcggcggcggcggcggcagcagcagcgggggctgcggGCGGTCGGGG
s panTro2.chr14      22278478 69 + 107349158 CCCAGTCTGAgcggcgatggcggcggcggcggcggcggcagcagcagcgggggctgcggGCGGTCGGGG
s ponAbe2.chr14      22805678 69 + 108868599 CCCAGTCTGATCAgcgatggcggcggcggcggcggcggcagcagcaacgggggctgcgAGGCGTCGGGG
s rheMac2.chr7       86161173 69 + 169801366 CCCAGTCTGAgcggcgatggcggcggcggcggcggcggcagcagcagcgggggctgcggGCGGTCGGGG
s calJac1.Contig824     79684 69 +    643256 CCCAGTCTGAgcggcgatggcggcggcggcggcggcggcagcagcagcgggggctgcggGCGGTCGGGG
```

COMP: Multiple Skeletal dysplasias (cartilage oligomeric matrix protein)

```
##maf version=1
a score=389405.0
s hg18.chr19          18757819 54 + 63811651 CCCGACTGTCAGGGACTCCGTCATTGTCGTCGTCGTCGTCGCAGGCATCACCCT
s panTro2.chr19       19256500 54 + 64473437 CCCGACTGTCAGGGACTCCGTCATTGTCGTCGTCGTCGTCGCAGGCATCACCCT
s ponAbe2.chr19       19184941 54 + 60714840 CCCGACTGTCAGGGACTCcgtcattgtcgtcgtcgtcgtcgcaggcatcaccct
s rheMac2.chr19       18409947 54 + 64391591 CCCGACTGTCAGGGACTCCGTCATTGTCGTCGTCGTTGTCGCAGGCATCACCCT
s calJac1.Contig13901    15472 54 +    31514 CTCGACTGTCAGGGACTCTgtcgttgtcgtcgtcgtcgtcgcaagcatcaccct
```

FRDA: Friedreich's ataxia (frataxin)

```
##maf version=1
a score=677349.0
s hg18.chr9          70842003  32 + 140273252 tacaaaaaaaaaaaaaaagaagaagaagaagaagaagaaaataaagaaaagttagccgg
s panTro2.chr9       67807593  32 + 138509991 tacaaaaaaaaaaaaaaaaaaaaagaagaagaagaaaat----aaaagttagccgg
s ponAbe2.chr9       63730253  30 + 135191526 tac--aaaaaaaaaaaaagaagaagaagaagaagaaag-----aaaagttagccgg
```

SCA8: Spinocerebellar ataxia 8 (SCA8 gene)

```
##maf version=1
a score=1278093.0
s hg18.chr10          102745016 45 + 135374737 TGAAACATCCGGAGCAAGGCTGCTGGAAAGCTGCTCTCTGTGAGG
s panTro2.chr10       101345612 45 + 135001995 TGAAACATCCGGAGCAAGGCTGCTGGAAAGCTGCTCTCTGTGAGG
s ponAbe2.chr10       100018529 45 + 133410057 TGAAACATCCGGAGCAAAGCTGCTGGAAAGCTGCTCTCTGTGAGG
s rheMac2.chr9        100606938 45 + 133323859 TGAAACATCCAGAGCAAGGCTGCTGGAAAGCTGCTCTCTGTGAGG
s calJac1.Contig13802     14656 45 +     18126 TGAAACATGGGGAGCAAGGCTGCTGGAAAGCTGCTGTCTGTGAGG
```

SCA12: Spinocerebellar ataxia 12 (protein phosphatase 2, regulatory subunit B)

```
##maf version=1
a score=1212654.0
s hg18.chr5           146238462 75 + 180857866 ACCCACACGCGCGCACTCGCA
------gctgctgctgctgctgctgctgctgctgcAGGAGGCTGGAGGCGGCTGGTG
s panTro2.chr5        148788110 75 + 183994906 ACCCACACGCGCGCACTCGCA
------gctgctgctgctgctgctgctgctgctgcAGGAGGCTGGAGGCGGCTGGTG
s ponAbe2.chr5        148555516 69 + 183952662 ACCCACACGCGCGCACTCGCA
------------gctgctgctgctgctgctgctgcAGGAGGCTGGAGGCGGCTGGTG
s rheMac2.chr6        143346047 78 + 178205221 ACCCACACGCGCGCACTCGCA
---gctgctgctgctgctgctgctgctgctgctgcAGGAGGCTGGAGGCGGCTGGTG
s calJac1.Contig7915     48874 81 +     90797 ACCCACACGCGCGCACTCGCA
GCAgctgctgctgctgctgctgctgctgctgctgcAGAAGGCTGGAGGCGGCTGGTG
```

HDL2: Huntington's disease-like 2 (junctophilin-3)

```
##maf version=1
a score=354523.0
s hg18.chr16          86195375 81 + 88827254 CCGGAAGCCAGGGAgctgcctgctgctg-ctgctgctgctgctgctgctgctgctgctgctg
TAAGATGGTTTCTGTGCAGG
s panTro2.chr16       87952686 72 + 90682376 CCGGAAGCCAGGGAgctgc----------ctgctgctgctgctgctgctgctgctgctgctg
TAAGATGGTTTCTGTGCAGG
s ponAbe2.chr16       75233263 69 + 77800216 CCGGAAGCCAGGGAgctgc-------------ctgctgctgctgctgttgctgctgctgctg
TAAGATGGTTTCTGTGCAGG
s rheMac2.chr20       85728194 69 + 88221753 CCGGAAGCCAGGGAgctgc-------------ctgctgctgctgctgttgctgctgctgctg
TAAGATGGTTTCTGTGCAGG
s calJac1.Contig5110      5446 75 +   272301 CCGGAAGCCAGGGAgctgc-------tgcctgctcctgctcctgctgctgccgctgccgctg
TAAGATGGTTTCTGTGCAGG
```

FRA10A: No confirmed disease association (FRA10A candidate gene 1)

```
##maf version=1
a score=360051.0
s hg18.chr10          95452250 63 + 135374737 AGCCTCGCCACAAccaccaccgccgccgccgccgccgccgccgccCGCAACCCGCCTCTCCCT
s panTro2.chr10       94050162 57 + 135001995 AGCCTCGCCACAAcca---ccgccgccgccgccgccgccgcc---CGCAACCCGCCTCTCCCT
s ponAbe2.chr10       92466184 60 + 133410057 AGCCTcgccaccaccaccaccgccgccgccgccgccg---tcgccCGCAACCCGCCTCTCCCT
s rheMac2.chr9        93161312 48 + 133323859 AGCCTCCCCACAA------CCGCCGCCGCCG---------TCCCCCGCTACCCGCCTCTCCCT
s calJac1.Contig164     873528 48 +    998040 AGCCTCGCCACAGT------CGCCGCCGC---------AGTCGCCTGCTACCCGCCTCTCCCT
```

DRPLA: Dentatorubral-pallidouysian atrophy (Dentatorubral-pallidouysian atrophy)

```
##maf version=1
a score=416031.0
s hg18.chr12           6916133 84 + 132349534 TCACCACcag---caacagca
acagcagcagcagcagcagcagcagcagcagcagcagcagcagcaTCACGGAAACTCTGGGCC
s ponAbe2.chr12        7169766 72 + 136387465 TCACCACcag---caacagca
------------gcagcagcagcaacagcagcagcagcagcagcaTCACGGAAGCTCTGGGCC
s rheMac2.chr11        7118124 57 + 134511895 TCACCACcag---caacagca
--------------------------gcagcagcagcagcagcaTCACGGAAGCTCTGGGCC
s calJac1.Contig3751     84135 57 +    254053 TCACCACcagcagcaacagca
-----------------------------gcagcaacagcagcaCCATCATGGAAGCTCTGGGCC
```

SCA1: Spinocerebellar ataxia 1 (Ataxin-1)

```
##maf version=1
a score=1297277.0
s hg18.chr6           16435826 123 + 170899992 CCGGAGCCCTGCT---GAGG
tgctgctgctgctgctgctgctgctgctgctgctgctgctgctgatgctgatgctgctgctgctgctgctgctgctgctgctgctgctCAGCCTTGTGTCCCG
s panTro2.chr6        16652979 105 + 173908612 CCGGAGCCCTGCT---GAGG
tgctgctgctgctgctgctgctgctgc------------tgctgatgctgatgctgctgctgctgc------tgctgctgctgctgctgctCAGCCTTGTGTCCCG s
ponAbe2.chr6       16956324 111 + 174210431 CCGGAGCCCTGCTGAGGAGG
tgctgctgctgctgctgctgctgctgc------------tgctgatgctgctgctgctgctgctgatgctgctgctgctgctgct---CAGCCTTGTGTCCCG
s rheMac2.chr4        16344561  84 + 167655696 CTGGAGCCCTGCT---GAGG
TGGtgctgctgctgctgctgctgctgctgc----------------------------------------tgctgctgctgctgctgctCAGCCTTGTGTCCCG
s calJac1.Contig850     354739  69 -    593183 CCGGAGCCCTGCT---GAGG
TGCTGCTGCGGCGGTGGCGGCTGCTGCGGCTGCTGCT----------------------------------------------------CAGCCTTGTGTCCCG
```

SCA2: Spinocerebellar ataxia 2 (Ataxin-2)

```
##maf version=1
a score=525198.0
s hg18.chr12          110521118 108 + 132349534 TGGCAGCCGCGGGCGGCG
gctgctgctgctgctgctgctgctgct---gttgct----------------gctgctgctgctgctgctgctgctgctgctgctgctgGGGCTTCAGCGACATGGTGA
s panTro2.chr12       112918382 117 + 135371336 TGGCAGCCGCGGGCGGCG
gctgctgctgctgctgctgct---------gttgctgctgctgttgttgttgttgttgttgctgctgctgctgctgctgctgctgctgGGGCTTCAGCGACATGGTGA
s ponAbe2.chr12       113416023  84 + 136387465 TGGCAGCCGCGGGCGgct
gctgctgct---------------gct---gttgctgctgctgttgctgct----------------------gctgctgctgctgGGGCTTCACCGACTTGGTGA
s rheMac2.chr11       112542371  88 + 134511895 TGGCAGCCGCGGGCGgct
gctgctgctgctgct---gctgctgct---gttgctgctgct------------------------------gctgctgc-ctgctgGGGCTTCAGCGACATGGTGA
s calJac1.Contig1525     16582  84 -    445262 TGGCGGCCGCAGGCGgct
gcggctgctgctgctgctgctgctgctgttgctgctgct----------------------------------------gGGGCTTCAGCGACATGGTGA
```

SCA17: Spinocerebellar ataxia 17 (TATA-box binding protein)

```
##maf version=1
a score=802930.0
s hg18.chr6           170712901 153 + 170899992 TTTGGAAGAGCAACAAAG
gcagcagcagcaacaacaacagcagcagcagcagcagcagcagcaacagcaacagca
gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaacaggcagtggcagctgcagcCGT
s panTro2.chr6        173817680 137 + 173908612 TTTGGAAGAGCAACAAAG
gcagcagcagcaacaacaacagcagcagcagcagcagcagca---------------
s ponAbe2.chr6        174151460 117 + 174210431 TTTGGAAGagcagcaaag
gcagcagcagcaacaacagcag-----------------------------------
cagcagcagcagcagcagcaacagcaacagcagcagcagcagcagcagcagcaacaggcagtggcagctgcagcCGT
-cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaacaggcagtggcagctgcagcCGT
s rheMac2.chr4        167567085 114 + 167655696 TTTGGAAGAGCAACAAAG
gcagcagcagcaacaacag--------------------------------------
cagcagcagcagcaacaacagcaacaacagcagcagcagcagcagcagcagcagcaacagGCAGTGGCCGCTGCAGCCGT
s calJac1.Contig3625     209475 120 -    242075 TTTGGAAGAGCAACAAAG
gcagcagcagcagcaacagcagcag--------------------------------
cagcagcagcagcagcagcagcagcagcagcagcaacaacagcagcaacagcagcaggcagtggcagccgcagcCGT
```

AIB-I: Increased prostate cancer risk (Nuclear receptor coactivator 3)

```
##maf version=1
a score=555048.0
s hg18.chr20          45713203 126 + 62435964 GAGGGTGGCTATGATGATgcagcagcagcagcagcagcaacagcag
cagcagcagcagcagcagcaacagcaacagcaacagcaacagcagcaacagcagcaAACCCAGGCCTTCAGCCCACC
s panTro2.chr20       45073718 122 + 62293572 GAGGGTGGCTATGATGATgcagcagcagcagcaacagcagca----
cagcagcagcagcagcagcagcagcaacagcaacagcaacagcagcaacagcagcaAACCCAGGCCTTCAGCCCACC
s ponAbe2.chr20       45263245 105 + 62736349 GAGGGTGGCTATGATGATgcagcagcaa----------------
cagcagcagcagcagcagcagcagcaacagcaacagcaa---cagcaacagcagcaAACCCAGGCTTTCAGCCCACC
s calJac1.Contig1469   162963 105 -   486727 GAGGGTGGCTATGATGATgcagcagcag----------------
cagcagcagcagcaacaacaacagcagcaacagcagcagcag---cagcagcagcagcaAACCCAGGCCTTCAGCCCACC
s rheMac2.chr10       78041815 123 - 94855758 GAGGGTGGCTATGATGATgcagcagcagcagcagcaacagcag---
caacagcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaAACCCAGGCCTTCAGCCCACC
```

HOXD13: Synpolydactyly (Homeobox D13)

```
##maf version=1
a score=1108042.0
s hg18.chr2          176666013 84 + 242951149
CGGGACGCATTCGGGGCGggcggcggcggcggcagcggcggctgcggcggcggcggcgg---------cagcCTCCGGCTTTGCGTACCCCGG
s panTro2.chr2b      181095886 84 + 248603653
CGGGACGCATTCGGGGCGggcggcggcggcggcggcggcggctgcggcggcgg---------cagcggcagcCTCCGGCTTTGCGTACCCCGG
s ponAbe2.chr2b       66248520 84 + 135000294
CGGGACGCATTCGGGGCGggctgcggcggctgcggctgcggctgcggctgcggcggcgg---------cagcCTCCGGCTTTGCGTACCCCGG
s calJac1.Contig349     461839 84 +    777646
CGGGACACATTCGGGACGggcggcggcggcggcggctgcggctgcggcggcgg---------cagcagcagcCTCTGGCTTTGCGTACCCCGG
s rheMac2.chr12       39733878 80 + 106505843
----ACACATTCCGGGCGggcggcggcggcaga---------tgcggcggcggcggcggcggcagcggcagcCTCCGGCTTTGCATACCCCGG
```

CBFA1:cleidocranial dysplasia (runt-related transcription factor 2)

```
##maf version=1
a score=214737.0
s hg18.chr6    45498446 90 + 170899992
gcagcagcagcagcaggaggcggcggcggcggctgcggcggcggcggcggctgcggcggcggcagctgcagTGCCCCGGTTGCGGCCGCC
s panTro2.chr6 46497743 90 + 173908612
gcagcagcagcagcaggaggcggcggcggcggctgcggcggcggcggcggctgcggcggcggcagctgcagTGCCCCGGTTGCGGCCGCC
s ponAbe2.chr6 45402644 90 + 174210431
acagcagcagcagcaggaggcggcggcggcggctgcggcggcggcggcggctgcggcggcggcagctgcagTGCCCCGGTTGCGGCCGCC
s rheMac2.chr4 45311954 90 + 167655696
acagcagcagcagcaggaggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcagctgcagTGCCCCGGTTGCGGCCGCC
```

MAB21L1:no confirmed disease association (MAB21, C. elegans, homologue-like 1)

```
##maf version=1
a score=111560.0
s hg18.chr13          34948414 96 + 114142980 TGCTGCAGCGTTGGTTTCC
ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctTTTCCCTTCCTTTTATCT
s panTro2.chr13       35251355 78 + 115868456 TGCTGCAGCGTTGGTTTCC
------------------ctgctgctgctgctgctgctgctgctgccgccgctgcCTTTCCCTTCCTTTTATCT
s ponAbe2.chr13       35179473 63 + 117095149 TGCTGCAGCGTTGGTTTCC
---------------------------------ctgctgctgctgctgctgctgctgctTTTCCCTTCCTTTTATCT
s calJac1.Contig3565    136453 57 +    251835 TACTGCCGCGTTGGTTTCC
---------------------------------------CTGCTGCTGCTCCTGCTGCTTTTCCCTTCCTTTAATCT
s rheMac2.chr17       14743868 54 +  94452569 TGCTGCAGCGTTGGTTTCC
-----------------------------------------CTGCTGCTGCTGCTGCTTTTCCCTTCCTTTTATCT
```

FRA11B: Jacobsen syndrome (CAS-BR-M murine ectopic retroviral transforming sequence homology)

```
##maf version=1
a score=584691.0
s hg18.chr11          118582190 72 + 134452384
CTCCCTCCGCCCGGATAGCcggcggcggcggcggcggcggcggcggcggcggcCGGGAGAGGCCCCTCCTTC
s panTro2.chr11       118088877 60 + 134204764
CTCCCTCCGCCCGGATAGC------------cggcggcggcggcggcggcggcCGGGAGAGGCCCCTCCTTC
s ponAbe2.chr11       116154296 72 + 132107971
CTCCCTCCGCCCCGATAGCcggcggcggcggcggcggcggcggcggcggcggcCGGGAGAGGCCCCTCCTTC
s rheMac2.chr14       117615281 63 + 133002572
CTCCCTCAGCCCGGATAGC---------cggcggcggcggcggcggcggcggcCGGGAGAGGCCCCTCCTTC
s calJac1.Contig373     44546 54 -    996610
CTCCCTCCGCCCGGATAgc------------------cggcggcggcggcggcgggAGAGGCTCCTCCTTC
```

Figure S8.

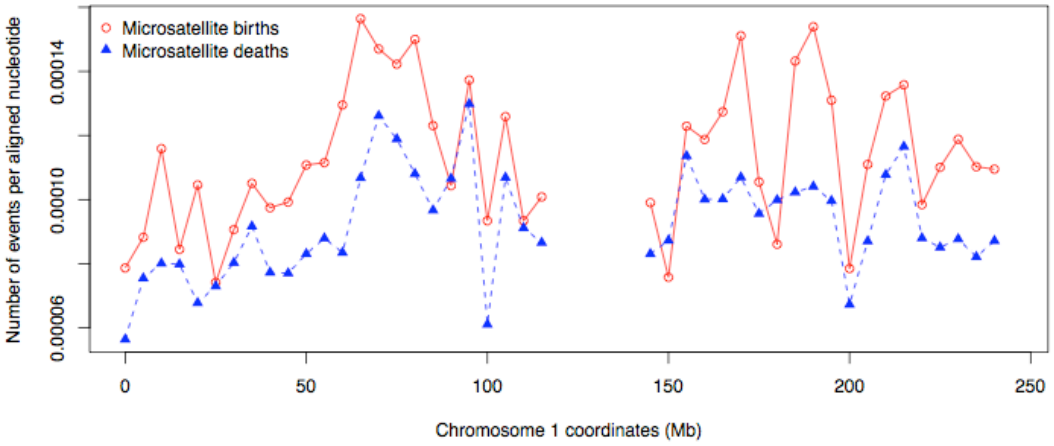

Microsatellite births
Microsatellite deaths
Number of events per aligned nucleotide
0.00006
0.00010
0.00014
0
50
100
150
200
250
Chromosome 1 coordinates (Mb)

Figure S9.

Feature	Window size					
	0.1 Kb	1 Kb	10 Kb	50 Kb	100 Kb	100 Kb*
GC content	0.31 (-15)	0.31 (-15)	0.37 (-15)	0.27 (-15)	0.19 (-15)	0.12 (-15)
L1 content	0.00	0.10 (-15)	0.08 (-15)	0.00	0.00	0.00
Alu content	0.19 (-15)	0.37 (-13)	0.46 (-15)	0.63 (-15)	0.72 (-15)	0.83 (-15)
Substitution rate	0.05 (-15)	0.06 (-15)	0.09 (-15)	0.06 (-15)	0.08 (-15)	0.04 (-15)
Protomicrosatellite	0.44 (-15)	0.15 (-2)	0.00	0.02 (-6)	0.01 (-2)	0.00
% deviance reduced	**0.07**	**0.02**	**0.01**	**0.03**	**0.05**	**0.13**

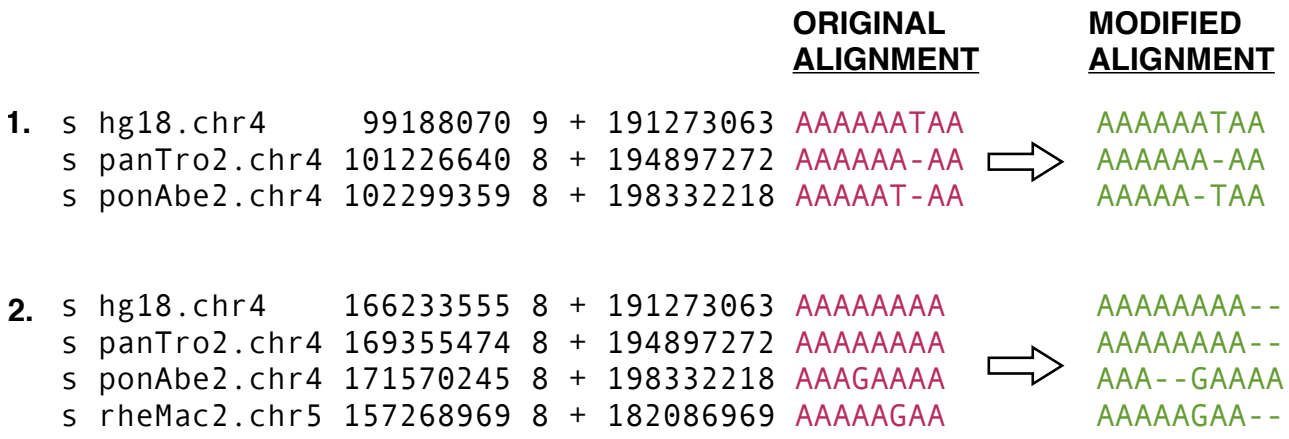
Figure S10.

A: AluY	Size of event and control windows				
	0.1 Kb	1 Kb	10 Kb	50 Kb	100 Kb
Births	**6.09/8.43**	**1.54/1.89**	**1.09/1.09**	**1.93/1.95**	**3.38/3.40**
G+C content	73.54/37.47	51.43/60.46	50.97/51.13	51.68/52.07	39.42/39.25
LINE content	1.23/1.24	14.75/11.07	9.55/9.84	0/0	0/0.57
AluY content	0.59/1.28	1.61/3.92	2.59/1.94	6.04/2.87	12.04/6.42
Subst. rate	10.01/5.81	8.55/5.94	10.52/10.32	6.39/5.63	5.65/4.90
Protomicrosat.	/27.72	/18.8	0	/1.03	0
Deaths	**4.47/5.67**	**1.02/1.16**	**0.71/0.71**	**1.02/1.04**	**1.93/1.96**
G+C content	81.35/52.06	65.16/60.42	60.96/60.39	61.81/62.30	51.45/52.02
LINE content	0.23/0	7.74/4.73	3.65/3.65	0.24/0.23	0.46/0
SINE content	0.00/0	1.72/3.22	2.25/2.25	2.39/0	3.81/0
Subst. rate	6.98/4.60	9.68/7.77	12.36/12.36	5.25/4.22	4.87/3.75
	/21.13	/8.71	0	/1.87	0

B: AluS	Size of event and control windows				
	0.1 Kb	1 Kb	10 Kb	50 Kb	100 Kb
Births	**6.17 / 8.75**	**1.64/2.15**	**1.13 / 1.13**	**2.03 / 2.03**	**3.56 / 3.59**
G+C content	61.14 / 26.00	38.95 / 51.77	33.59 / 29.19	33.30 / 25.30	25.52 / 15.72
LINE content	0.82 / 0.72	15.00 / 10.02	9.02 / 8.23	0.00 / 0.00	0.00 / 0.00
AluS content	1.89 / 4.85	7.67 / 15.34	6.38 / 5.59	10.51 / 6.63	16.55 / 11.47
Subst. rate	10.17 / 6.01	8.84 / 6.74	11.04 / 11.18	6.42 / 6.52	5.44 / 6.15
Protomicrosats.	29.45	23.14	0	0	0
Deaths	**4.52 / 5.87**	**1.02 / 1.19**	**0.71 / 0.71**	**1.06 / 1.06**	**2.06 / 2.06**
G+C content	69.65 / 39.73	58.19 / 47.32	46.78 / 41.06	43.09 / 37.33	34.96 / 25.57
LINE content	0.00 / 0.00	7.33 / 3.70	3.64 / 3.07	0.00 / 0.00	0.00 / 0.00
SINE content	1.14 / 3.39	1.51 / 5.55	2.52 / 2.79	5.76 / 2.07	9.46 / 0.00
Subst. rate	7.21 / 4.80	10.34 / 8.50	12.61 / 12.85	5.30 / 4.84	4.73 / 0.00
Protomicrosats.	19.72	13.68	0	0	0

C: AluJ	Size of event and control windows				
	0.1 Kb	1 Kb	10 Kb	50 Kb	100 Kb
Births	**6.17/8.34**	**1.52/1.82**	**1.08/1.09**	**1.92/1.93**	**3.25/3.27**
G+C content	75.21/38.91	54.94/62.34	44.16/43.60	41.87/40.30	33.50/31.36
LINE content	1.60/1.40	15.02/11.30	9.58/9.72	0.00/0.00	0.00/0.00
AluJ content	1.94/0.00	0.62/0.00	2.27/1.46	5.38/1.86	8.56/2.75
Subst. rate	9.49/5.77	8.39/6.07	10.39/10.21	5.61/5.11	4.45/4.00
Protomicrosat.	/25.04	/14.85	0	0	0
Deaths	**4.71/5.78**	**1.06/1.15**	**0.69/0.69**	**1.01/1.03**	**1.97/1.98**
G+C content	81.21/53.33	70.93/67.43	61.21/60.46	53.49/53.65	40.96/37.65
LINE content	0.58/0.53	7.84/5.33	3.74/4.01	0.00/0.00	0.30/0.00
SINE content	5.10/1.96	5.77/2.67	0.00/0.00	0.00/0.00	0.00/1.64
Subst. rate	6.41/4.45	8.87/7.62	12.07/12.03	4.82/3.76	4.04/3.57
	/13.36	/6.29	0	1.41	0

Figure S11.


ORIGINAL ALIGNMENT
MODIFIED ALIGNMENT
1.
s hg18.chr4 99188070 9 + 191273063 AAAAAATAA
s panTro2.chr4 101226640 8 + 194897272 AAAAAA-AA
s ponAbe2.chr4 102299359 8 + 198332218 AAAAAT-AA
AAAAAATAA
AAAAAA-AA
AAAAA-TAA
2.
s hg18.chr4 166233555 8 + 191273063 AAAAAAAA
s panTro2.chr4 169355474 8 + 194897272 AAAAAAAA
s ponAbe2.chr4 171570245 8 + 198332218 AAAGAAAA
s rheMac2.chr5 157268969 8 + 182086969 AAAAAGAA
AAAAAAAA--
AAAAAAAA--
AAA--GAAAA
AAAAAGAA--

SUPPLEMENTAL TABLES

Table S1. The numbers of microsatellites remaining after each filtering step implemented.

Motif size	Threshold	Orthologous loci identified	No. of loci remaining after each filtering step		
			Interrupted, compound, & loci closer than n bp	Loci with low sequence quality (PHRED <20)	Loci with low complexity or flanking sequence similarity
1	5	9654415	3520671	439424	222386
	6	6725552	1837882	1349607	343147
	7	3445015	250922	225829	157905
	8	2370977	250866	225779	44451
	9*	1503526	1433786	1306667	127490
	9	1503526	1407654	1296474	21027
	10	92764	45835	39876	33226
2	3	675360	5366270	260297	123069
	4	588474	1234764	1311287	87135
	5*	478955	402763	387654	29290
	5	478955	394756	370989	4497
	6	200741	23456	21110	3283
	7	213983	119826	43348	2627
	8	87265	19746	16191	1133
3	2	683236	6376363	475796	233236
	3	332798	389622	350659	158934
	4*	201854	129342	193654	19663
	4	201854	128033	193030	2951
	5	86945	93611	81441	2756
	6	60828	16027	13462	928
4	2	875878	1308482	618438	312905
	3*	509904	484293	475435	60635
	3	509904	466401	472795	9770
	4	72654	35154	30583	4882
	5	61741	16363	13417	2058

*n=10, otherwise n=25

Table S2. Counts of filtered microsatellite loci for each examined threshold value, and counts of filtered loci for which the mutational pathway was deciphered.

Motif size	Threshold (no. of repeats)	Total number of loci	No. of deciphered loci*	Percent deciphered loci	No. of deciphered stationary loci	No. of deciphered births	No. of deciphered deaths
1	5	222386	176153	79.21	114703	37194	24256
	6	343147	248394	72.38	142003	60179	46212
	7	157905	108434	68.67	50274	35078	23082
	8	44451	27616	62.12	9922	8130	9564
	9	21027	19195	91.28	8601	6070	4524
	10	33226	24612	74.07	9103	5565	9944
2	3	123069	92091	74.82	46429	27659	18003
	4	87135	58194	66.78	21946	19133	17115
	5	4497	4057	90.21	1644	1313	1100
	6	3283	2257	68.74	681	594	982
	7	2627	1587	60.41	284	869	434
	8	1133	807	71.22	191	280	336
3	2 (7 bp)	233236	177373	76.04	88181	56545	32647
	3	158934	110684	69.64	46938	33030	30716
	4	2951	2605	88.27	1015	749	841
	5	2756	1693	61.42	383	681	629
	6	928	651	70.15	98	203	350
4	2	312905	240292	76.79	114320	76246	49726
	3	9770	8804	90.11	4138	2342	2324
	4	4882	3164	64.80	825	1192	1147
	5	2058	1492	72.49	249	494	749

* Undeciphered loci had large between species locus-size variation leading to problematic alignments, or because deciphering involved large number of closely spaced mutations (see Fig. 2D for examples), or because the two out-group species differed in their birth/death states.

Table S3. Proportions of birth events along individual branches. Expected proportions based on species divergence times are given in parentheses.

Motif size	Threshold (no. of repeats)	H (1.00)	C (1.00)	HC (1.33)	O (2.33)
1	5	1.00	1.01	1.37	2.71
	6	1.00	1.01	1.38	2.49
	7	1.00	1.00	1.25	2.26
	8	1.05	1.00	1.08	2.03
	9	1.06	1.00	1.02	1.84
	10	1.02	1.17	1.87	1.00
2	3	1.00	1.02	1.35	2.59
	4	1.00	1.04	1.35	2.49
	5	1.08	1.00	1.33	2.12
	6	1.04	1.00	4.11	1.49
	7	1.32	1.00	1.57	2.22
	8	1.26	1.00	3.81	0.88
3	2 (7 bp)	1.04	1.00	1.36	2.39
	3	1.05	1.00	1.46	2.49
	4	1.13	1.00	1.53	2.36
	5	1.00	1.01	1.53	1.91
	6	1.00	1.64	4.48	1.40
4	2	1.05	1.00	1.46	2.56
	3	1.11	1.00	1.49	2.38
	4	1.01	1.00	1.46	2.17
	5	1.00	1.49	5.68	1.28

Table S4. Proportions of death events along individual branches. The expected proportions based on species divergence time are in parentheses.

Motif size	Threshold (no. of repeats)	H (1.00)	C (1.00)	HC (1.33)	O (2.33)
1	5	1.04	1.00	1.58	2.54
	6	1.00	1.04	1.41	2.21
	7	1.00	1.13	1.38	1.94
	8	1.00	1.05	1.53	1.69
	9	1.00	1.07	1.39	1.51
	10	1.06	1.06	1.00	1.30
2	3	1.03	1.00	1.86	2.66
	4	1.00	1.09	1.51	2.33
	5	1.00	1.09	1.30	2.11
	6	1.00	1.16	1.08	2.22
	7	1.12	1.00	1.58	1.74
	8	1.00	1.30	1.22	1.69
3	2 (7 bp)	1.00	1.10	1.77	2.56
	3	1.00	1.03	1.61	2.44
	4	1.00	1.03	1.55	2.33
	5	1.00	1.10	1.45	2.08
	6	1.00	1.42	1.11	1.98
4	2	1.00	1.00	1.71	2.59
	3	1.03	1.00	1.60	2.42
	4	1.25	1.00	1.38	1.82
	5	1.10	1.00	1.53	1.96

Table S5. Substitution matrix for Births and Deaths (summed across *H*, *C*, *O*, and *HC* branches). Final transition:transversion ratios for births and deaths are 2.17 and 2.01 respectively. No correction for multiple hits was applied.

Births	**A**	**T**	**G**	**C**
A		1472	2127	1132
T	1481		1252	2129
G	913	341		336
C	358	887	358	
Deaths	**A**	**T**	**G**	**C**
A		375	763	527
T	375		460	840
G	919	226		263
C	919	924	269	

Table S6. Number of different types of mutations leading to microsatellite births/deaths and the total number of stationary microsatellites differentiated by motif sizes (the last column), in coding exons.

Motif size	Branch	Synonymous substitutions	Non-synonymous substitutions	Non-frame-shift Indels	Frame-shift Indels	Total	Stationary micro-satellite count
Mono	H	0 / 0	0 / 0	1 / 0	1 / 0	2 / 0	
	C	0 / 0	0 / 0	0 / 0	1 / 1	1 / 1	
	HC	3 / 0	0 / 2	0 / 0	0 / 0	3 / 2	
	O	0 / 3	0 / 1	3 / 0	5 / 0	8 / 4	
	Total	**3 / 3**	**0 / 3**	**4 / 0**	**7 / 1**	**14 / 7**	**39**
Di	H	1 / 1	3 / 1	0 / 0	0 / 0	4 / 2	
	C	4 / 0	2 / 1	0 / 0	0 / 0	6 / 1	
	HC	3 / 5	2 / 0	0 / 0	0 / 0	5 / 5	
	O	2 / 3	5 / 2	0 / 0	1 / 0	7 / 5	
	Total	**10 / 9**	**12 / 4**	**0 / 0**	**1 / 0**	**22 / 13**	**66**
Tri	H	3 / 11	9 / 4	0 / 4	0 / 0	12 / 19	
	C	8 / 5	3 / 3	1 / 4	0 / 1	12 / 13	
	HC	10 / 26	13 / 5	3 / 1	0 / 0	12 / 32	
	O	8 / 39	11 / 9	5 / 5	0 / 0	14 / 53	
	Total	**29 / 81**	**36 / 21**	**9 / 14**	**0 / 1**	**74 / 117**	**405**
Tetra	H	4 / 5	4 / 1	0 / 0	0 / 0	8 / 6	
	C	4 / 8	2 / 4	0 / 0	0 / 0	6 / 12	
	HC	6 / 7	3 / 1	0 / 0	0 / 0	9 / 8	
	O	13 / 6	9 / 4	0 / 0	2 / 1	24 / 11	
	Total	**27 / 26**	**18 / 10**	**0 / 0**	**2 / 1**	**47 / 37**	**126**
Total		**69 / 119**	**66 / 38**	**13 / 14**	**10 / 3**	**157 / 174**	**636**

Table S7. Amino acid replacements caused by microsatellite births and deaths

	Ala	Cys	Asp	Glu	Phe	Gly	His	Lys	Leu	Asn	Pro	Gln	Arg	Ser	Stop	Thr	Val	Tyr
BIRTHS																		
Ala	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	**3**	0
Cys	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	**3**
Asp	0	0	0	2	0	1	1	0	0	0	0	0	0	0	0	0	0	1
Glu	0	0	**4**	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0
Phe	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gly	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
His	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Lys	0	0	0	0	1	0	0	0	0	**3**	0	0	1	0	0	0	1	0
Leu	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0
Pro	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0
Gln	0	0	0	1	0	0	0	0	0	0	1	0	2	0	0	0	0	0
Arg	0	0	0	0	0	0	0	2	0	0	0	**2**	0	0	0	1	0	0
Ser	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0
Thr	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0
Val	**4**	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
DEATHS																		
Ala	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0
Cys	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Asp	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Glu	1	0	**4**	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Phe	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Gly	0	0	0	**2**	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Ile	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
Lys	0	0	0	1	0	0	0	0	0	0	0	**3**	0	0	0	0	0	0
Leu	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Met	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
Asn	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
Pro	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Gln	0	0	0	0	0	0	**2**	0	0	0	0	0	0	0	0	0	0	0
Arg	0	0	0	0	0	**3**	0	0	0	0	1	0	0	0	0	0	0	0
Ser	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
Thr	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Val	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Table S8. Correlation matrix of predictors (Pearson's correlation values) used in logistic regression analysis.

Window size		Proto-micro-satellite	SINEs	GC content	LINEs	Substitution rate
0.1	Protomicrosatellite	1.00	0.11	-0.18	0.08	0.03
	SINEs	0.11	1.00	0.56	-0.22	0.06
	GC content	-0.18	0.56	1.00	-0.27	0.04
	LINEs	0.08	-0.22	-0.27	1.00	0.00
	Substitution rate	0.03	0.06	0.04	0.00	1.00
1	Protomicrosatellite	1.00	0.38	0.19	0.26	0.02
	SINEs	0.38	1.00	0.20	-0.09	0.10
	GC content	0.19	0.20	1.00	0.46	0.03
	LINEs	0.26	-0.09	0.46	1.00	0.01
	Substitution rate	0.02	0.10	0.03	0.01	1.00
10	Protomicrosatellite	1.00	0.47	-0.09	0.11	-0.03
	SINEs	0.47	1.00	0.44	-0.20	0.09
	GC content	-0.09	0.44	1.00	-0.35	0.04
	LINEs	0.11	-0.20	-0.35	1.00	0.02
	Substitution rate	-0.03	0.09	0.04	0.02	1.00
50	Protomicrosatellite	1.00	0.60	-0.03	-0.12	-0.08
	SINEs	0.60	1.00	0.50	-0.41	0.11
	GC content	-0.03	0.50	1.00	-0.53	0.08
	LINEs	-0.12	-0.41	-0.53	1.00	0.00
	Substitution rate	-0.08	0.11	0.08	0.00	1.00
100	Protomicrosatellite	1.00	0.64	-0.02	-0.19	-0.06
	SINEs	0.64	1.00	0.50	-0.49	0.12
	GC content	-0.02	0.50	1.00	-0.60	0.10
	LINEs	-0.19	-0.49	-0.60	1.00	-0.03
	Substitution rate	-0.06	0.12	0.10	-0.03	1.00

REFERENCES

Aleshin, A., and D. Zhi

2010 Recombination-associated sequence homogenization of neighboring Alu elements: signature of nonallelic gene conversion. Mol Biol Evol 27(10):2300-11.

Batzer, M. A., et al.

1995 Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. J Mol Biol 247(3):418-27.

Benson, G.

1999 Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573-80.

Boyer, J. C., et al.

2008 Sequence-dependent effect of interruptions on microsatellite mutation rate in mismatch repair-deficient human cells. Mutat Res 640(1-2):89-96.

Brinkmann, B., et al.

1998 Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62(6):1408-15.

Fujita, P. A., et al.

2010 The UCSC Genome Browser database: update 2011. Nucleic Acids Res.

Kelkar, Y. D., et al.

2010 What is a microsatellite: a computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. Genome Biol Evol 2:620-35.

Kofler, R., C. Schlotterer, and T. Lelley

2007 SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics 23(13):1683-5.

Mudunuri, S. B., and H. A. Nagarajaram

2007 IMEx: Imperfect Microsatellite Extractor. Bioinformatics 23(10):1181-7.

Myers, S., et al.

2005 A fine-scale map of recombination rates and hotspots across the human genome. Science 310(5746):321-4.

Pearson, C. E., K. Nichol Edamura, and J. D. Cleary

2005 Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6(10):729-42.

Roy, A. M., et al.

2000 Potential gene conversion and source genes for recently integrated Alu elements. Genome Res 10(10):1485-95.

Sibly, R. M., et al.

2003 The structure of interrupted human AC microsatellites. Mol Biol Evol 20(3):453-9.

Zhi, D.

2007 Sequence correlation between neighboring Alu instances suggests post-retrotransposition sequence exchange due to Alu gene conversion. Gene 390(1-2):117-21.