Supplemental Text

ME-Scan protocol

ME-Scan was performed on blood-derived DNA of 460/599 CEPH individuals prior to sequencing on the Illumina 2000 platform (Supplemental Table S1). ME-Scan is a targeted sequencing protocol that amplifies specific regions of the genome prior to Illumina sequencing (Witherspoon et al. 2010; Ha et al. 2016; Witherspoon et al. 2013; Feusier et al. 2017; Ha et al. 2017). This protocol targets the 7bp insertion that is diagnostic of Yb8/9 elements (Witherspoon et al. 2013, 2010; Feusier et al. 2017). Two individuals (grandparent and offspring) failed sequencing and were dropped from the analysis. Data were mapped to hg19 using bwa align (bwa version 0.7.9a) (Li and Durbin 2009) and uploaded to SQL developer for analysis. Read set processing was the same as described in (Witherspoon et al. 2013). A detailed report of the ME-Scan protocol including primers is reported in (Feusier et al. 2017). *De novo* elements were detected by finding loci that were present in at least one offspring and absent in the parents/grandparents. Loci with less than eight reads of support were removed, based on (Feusier et al. 2017). We identified and validated one *de novo Alu* Yb8 insertion, *Alu* #1 (Table 1 and Supplemental Tables S2, S12).

ME-Scan sequencing library construction for the 16 grandparent-parent trios

Genomic DNA samples from 48 individuals were obtained from Coriell Cell Repositories (https://coriell.org/). The samples contain 16 parent-offspring trios with northern and western European ancestry from the CEPH collection. Information including population, family and individual relationships is shown in Table S1.

The ME-Scan libraries were prepared following the ME-Scan protocol described previously (Ha et al. 2017). All the adaptor and primer sequences are described previously

(Feusier et al. 2017). For each sample, five µg genomic DNA was fragmented to about 1 kb in size using Covaris system (Covaris, Woburn, MA, USA) using the following protocol: duty cycle: 5%; intensity: 3; cycles/burst: 200; time: 15 seconds. The fragmented samples were concentrated using AMPure XP beads (cat. no. A63881, Beckman Coulter, Brea, CA, USA), following the manufacturer's protocol. The concentrated DNA fragments were then used to construct the sequencing library using KAPA Library Preparation Kits with SPRI solution for Illumina (KAPA Biosystems, Wilmington, MA, USA, cat. no KK8201). For each sample, after the DNA fragments were end-repaired, A-tailed on both ends, and ligated with adaptors, the concentration of ligated DNA was quantified using Nanodrop. The 48 individual libraries were then pooled into one single library with equal concentration. All of the following steps were performed using the pooled library.

*Alu*Yb-specific first amplification was conducted for 10 cycles with 360 ng of pooled template DNA and 2.5 µl of primer, following the KAPA kit amplification protocol (initial denaturation at 98 ºC for 45 seconds, followed by the thermocycling conditions of 98 ºC for 15 seconds, 65 ºC for 30 seconds, and 72 ºC for 30 seconds, and a final extension at 72 ºC for 1 minute). The amplified PCR product was electrophoresed at 120 volts for 90 minutes on a 2 % NuSieve^{R} GTG^{R} Agarose gel (cat. no. 50080, Lonza, Rockland, Maine, USA). Fragments around 600 bp were size selected and purified using Wizard SV Gel and PCR Clean-up system (cat. no. A9281, Promega, Madison, WI, USA). After size selection, biotinylated *Alu*-enriched DNA fragments were magnetically separated from other genomic DNA fragments using 5 µl Dynabeads^{R} M-270 Streptavidin (cat. no. 65305, Invitrogen, Life Technologies, Oslo, Norway) following the manufacturer's protocol. Second amplification was conducted for 20 cycles under the same condition as first amplification, with 3 µl of biotinylated *Alu*-enriched DNA as template

using the P7 primer and a mix of six *Alu*_head primers with the Illumina P5 sequence in a 50 µl reaction. The amplified PCR product was electrophoresed at 120 volts for 90 minutes on a 2 % NuSieve^{R} GTG^{R} Agarose gel (cat. no. 50080, Lonza, Rockland, Maine, USA). Fragments around 400 bp were size selected and purified using Wizard SV Gel and PCR Clean-up system (cat. no. A9281, Promega, Madison, WI, USA). Before the library was sequenced, its fragment size and concentration was determined using Bioanalyzer and quantitative PCR by the RUCDR Infinite Biologics (Piscataway, NJ, USA). The library was sequenced using the Illumina Hiseq 2000 with 100PE format at RUCDR Infinite Biologics.

Whole Genome Sequencing

Whole-genome DNA sequencing libraries of 350bp inserts were generated from 500ng of blood-derived genomic DNA using a KAPTA HTP Library Prep Kit (KAPA Biosystems, Boston, MA) on the SciClone NGS instrument (Perkin Elmer, Waltham, MA). The genomic DNA was fragmented and size-selected with AMPure XP beads using a 0.6x/0.8x ratio. Each library was PCR amplified 4-6 cycles using KAPA HiFi and purified with two 0.7x AMPureXP bead cleanups. qPCR (KAPA Biosystems, Boston, MA) was utilized to determine the concentration of each library. The libraries were pooled and sequenced on the Illumina HiSeqX instrument (Illumina, San Diego, CA) to generate 2x150bp paired-end sequencing data of ~30X coverage for each sample.

Feusier J, Witherspoon DJ, Scott Watkins W, Goubert C, Sasani TA, Jorde LB. 2017. Discovery of rare, diagnostic AluYb8/9 elements in diverse human populations. *Mob DNA* **8**: 9.

Ha H, Loh JW, Xing J. 2016. Identification of polymorphic SVA retrotransposons using a mobile element scanning method for SVA (ME-Scan-SVA). *Mob DNA* **7**: 15.

Ha H, Wang N, Xing J. 2017. Library construction for high-throughput mobile element identification and genotyping. *Methods Mol Biol* **1589**: 1–15.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.

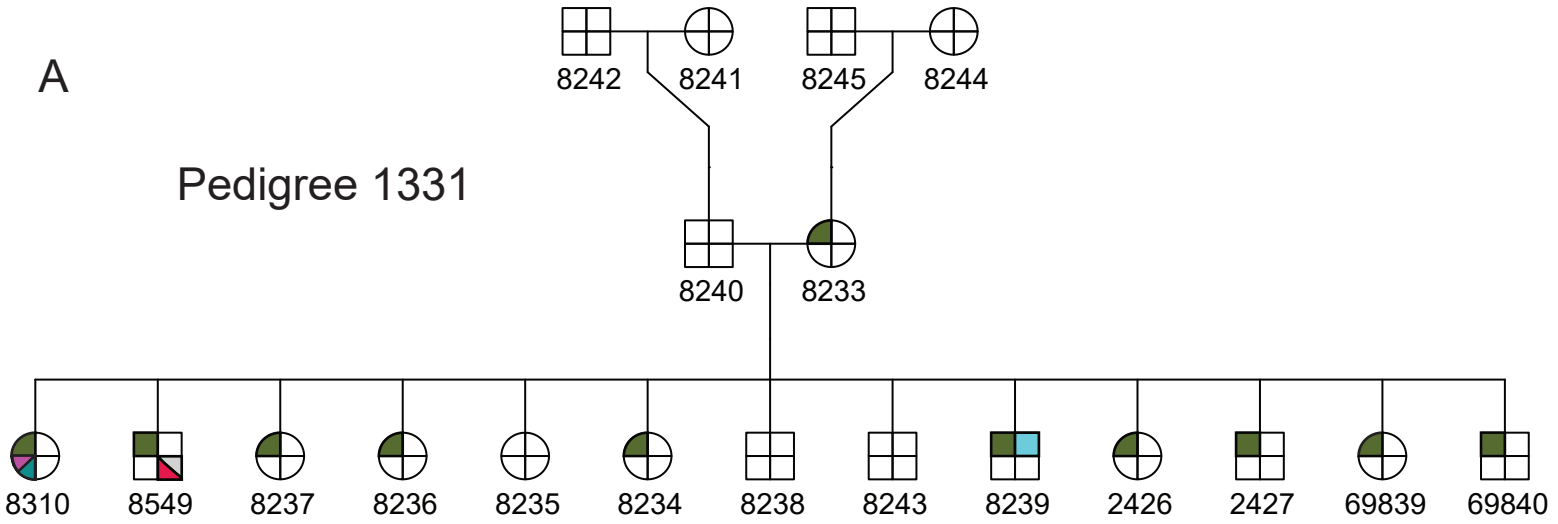
Bioinformatics **25**: 1754–1760.

Witherspoon DJ, Xing J, Zhang Y, Watkins WS, Batzer M a, Jorde LB. 2010. Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. *BMC Genomics* **11**: 410.

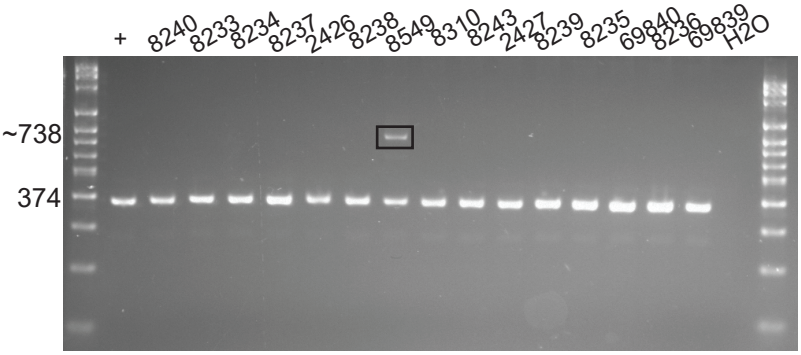
Witherspoon DJ, Zhang YH, Xing JC, Watkins WS, Ha H, Batzer MA, Jorde LB. 2013. Mobile element scanning (ME-Scan) identifies thousands of novel Alu insertions in diverse human populations. *Genome Res* **23**: 1170–1181.

A

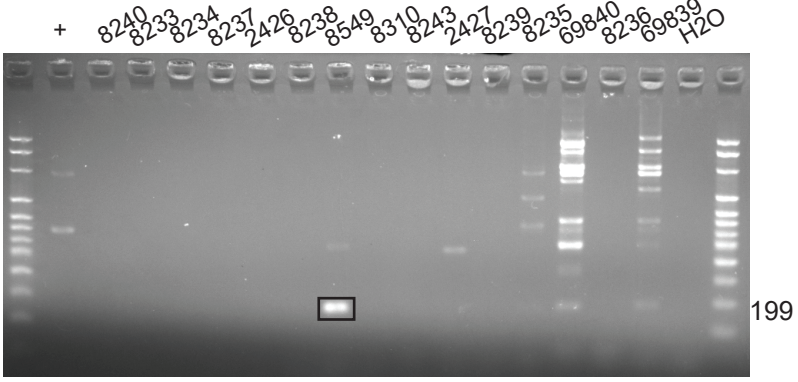
Pedigree 1331



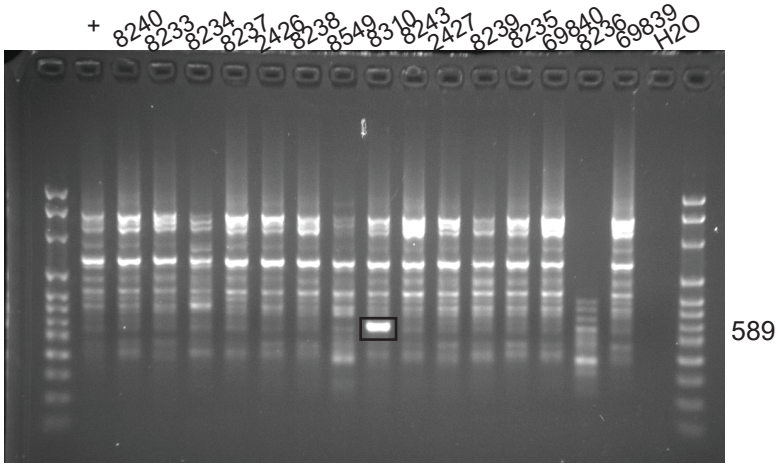
Alu #1 empty/fill



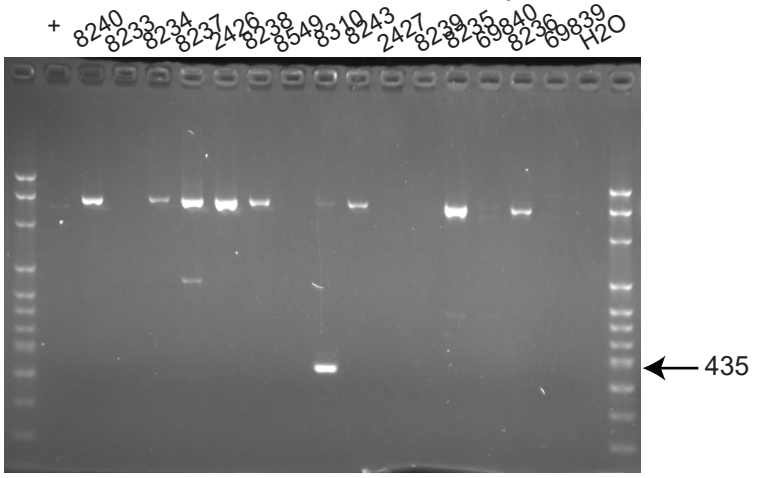
SVA #4 breakpoint



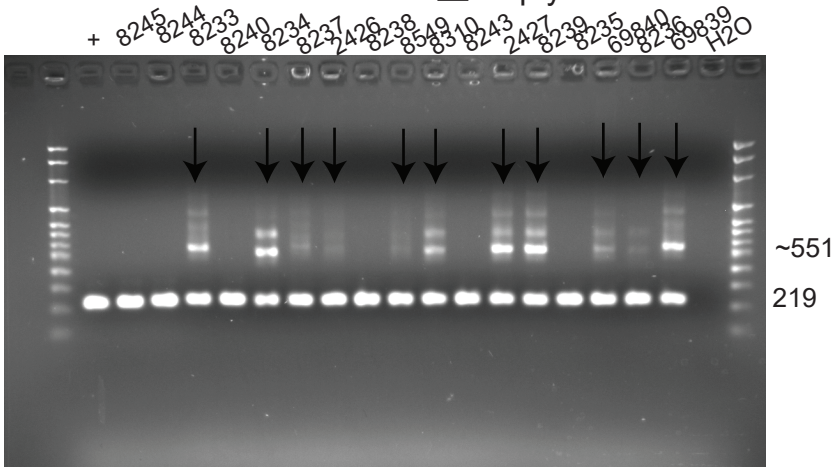
SVA #1 breakpoint



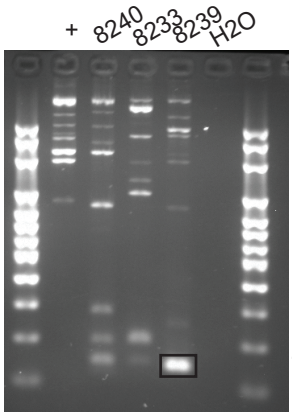
L1 #5 breakpoint



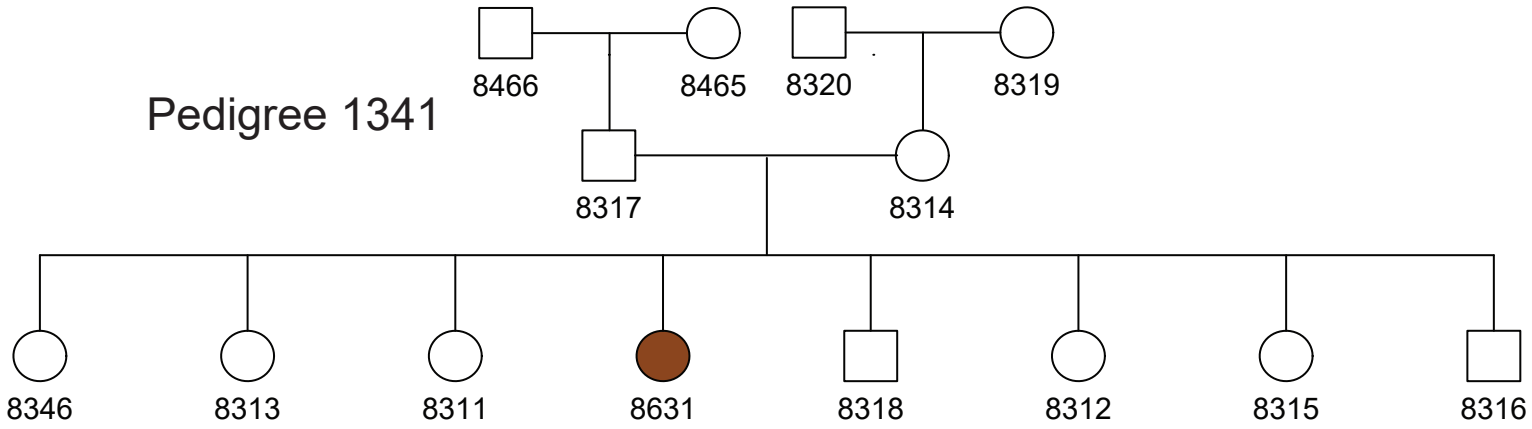
Alu #2 empty/fill

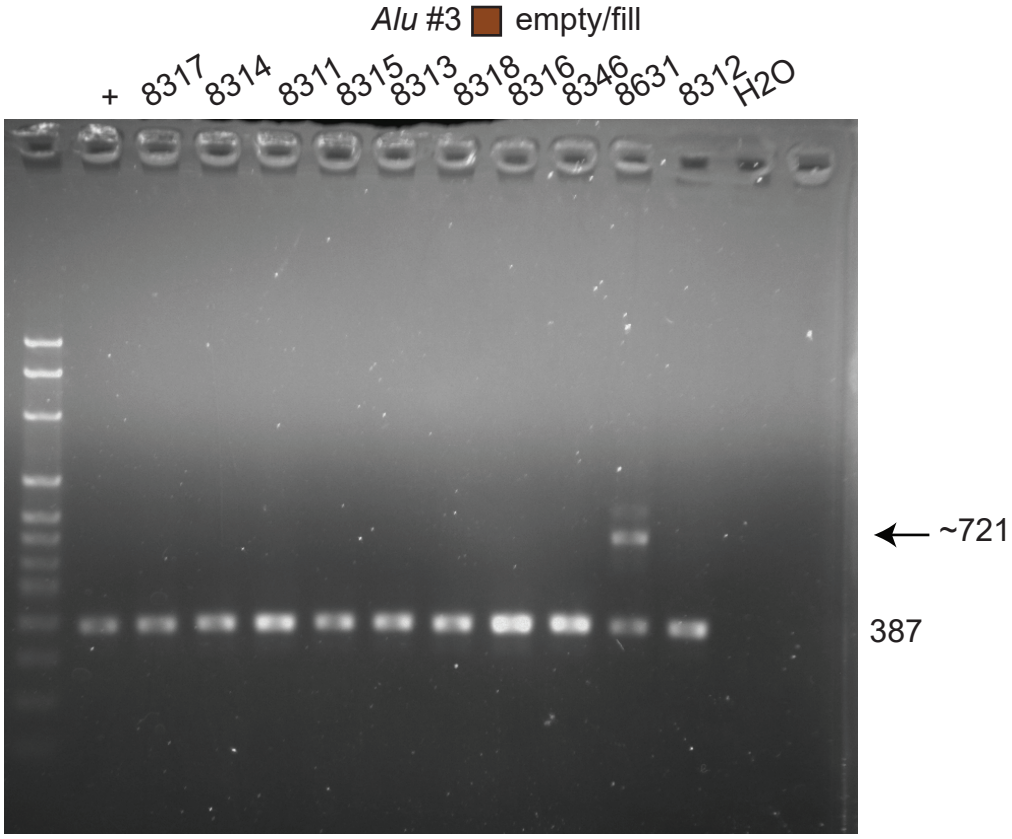


L1 #7 breakpoint



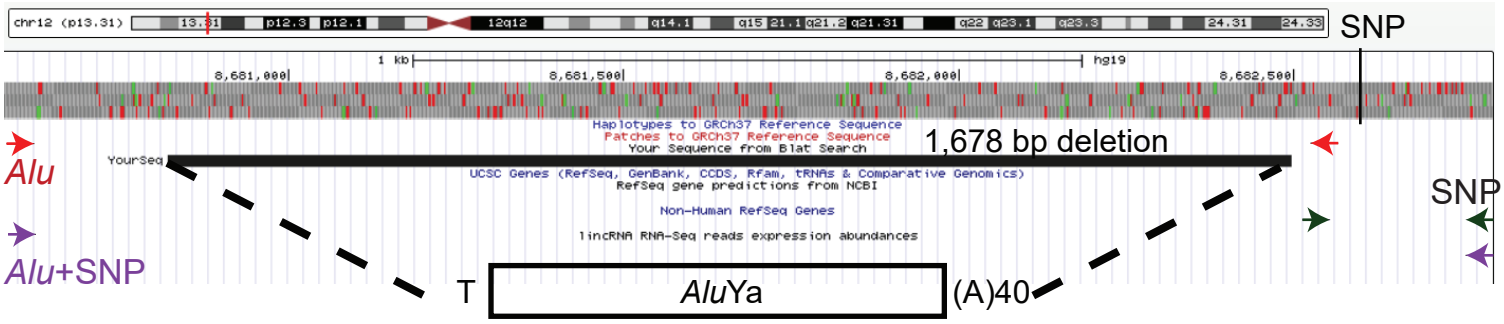
B


Pedigree 1341
8466
8465
8320
8319
8317
8314
8346
8313
8311
8631
8318
8312
8315
8316

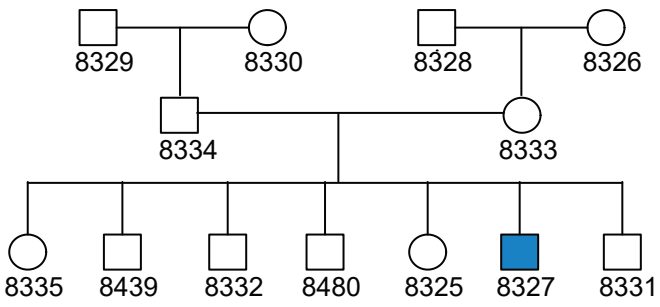

Alu #3 empty/fill
+
8317
8314
8311
8315
8313
8318
8316
8346
8631
8312
H2O
~721
387

C

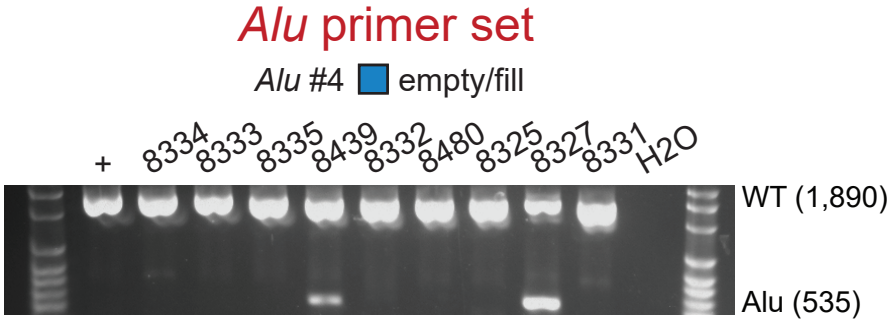
1



2



3



4

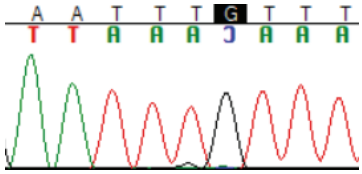
SNP primer set

Alu + SNP primer set

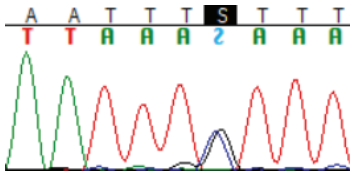
Wild Type Band

Alu Band

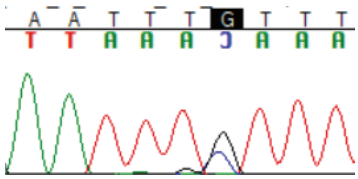
8334 (dad)
G/G

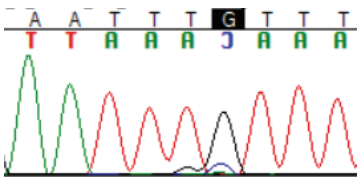


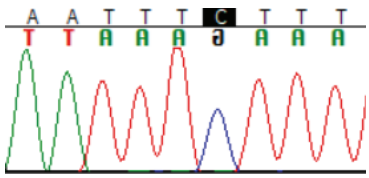
8333 (mom)
G/C



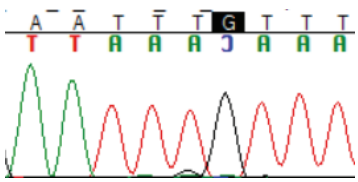
8327
G/C

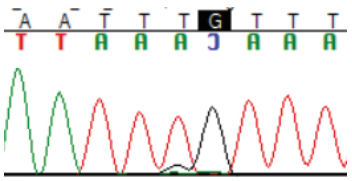


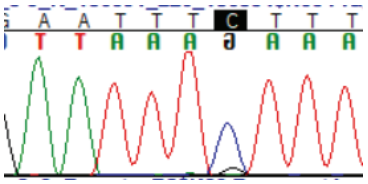




8439
G/G

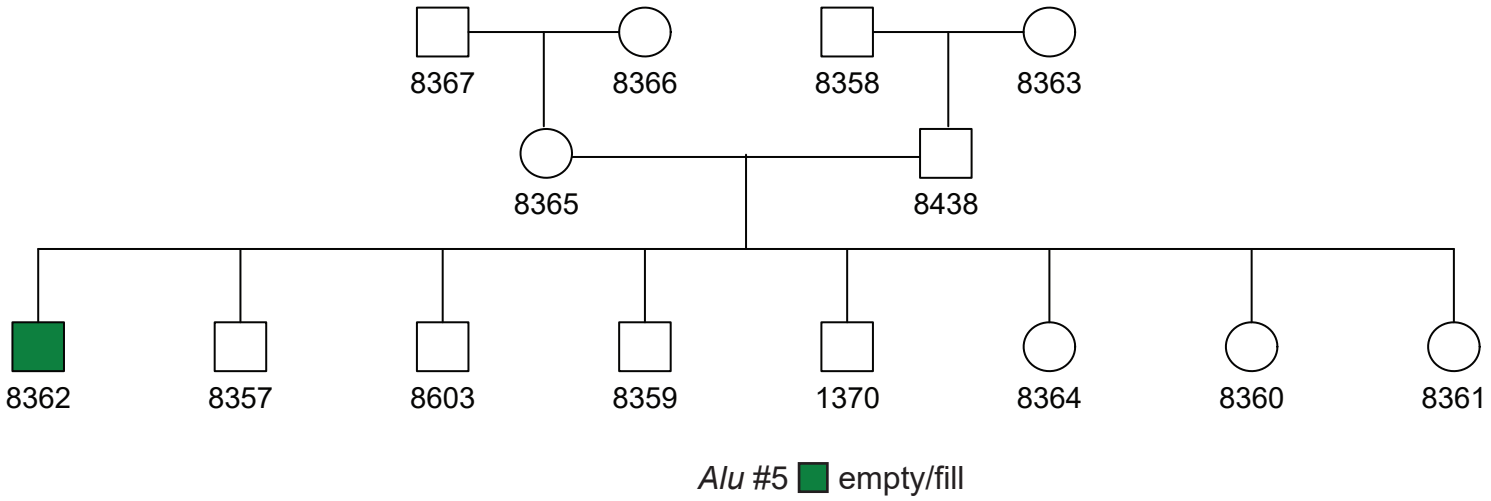


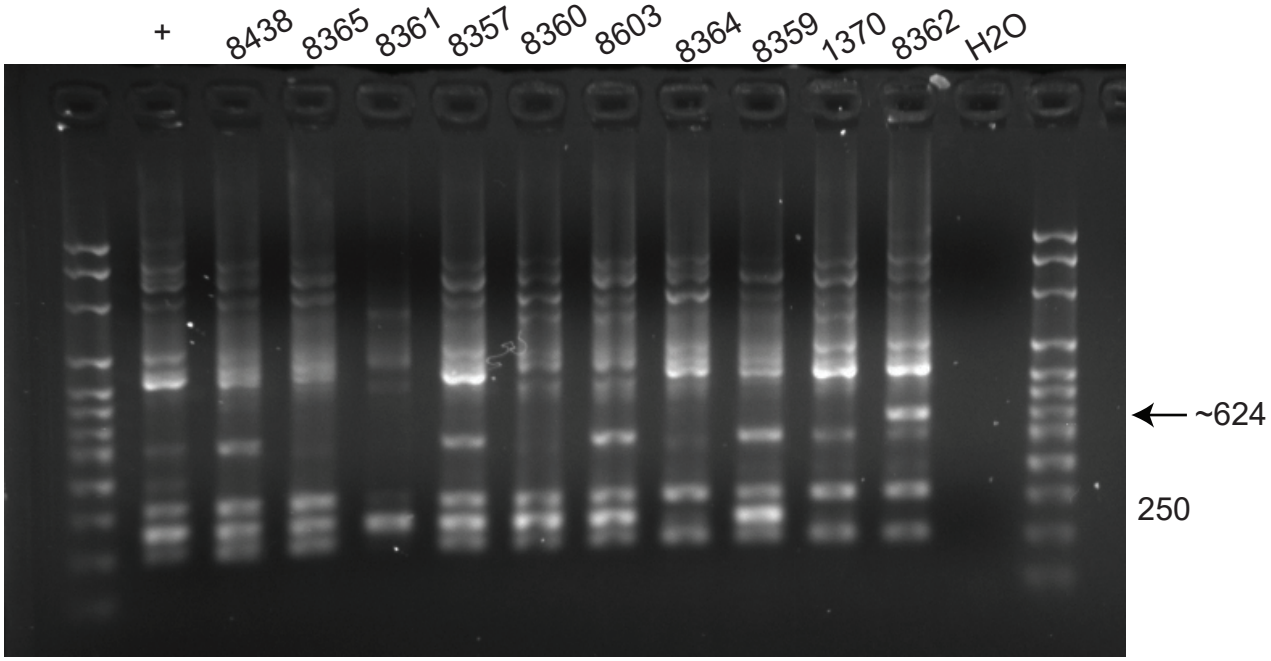




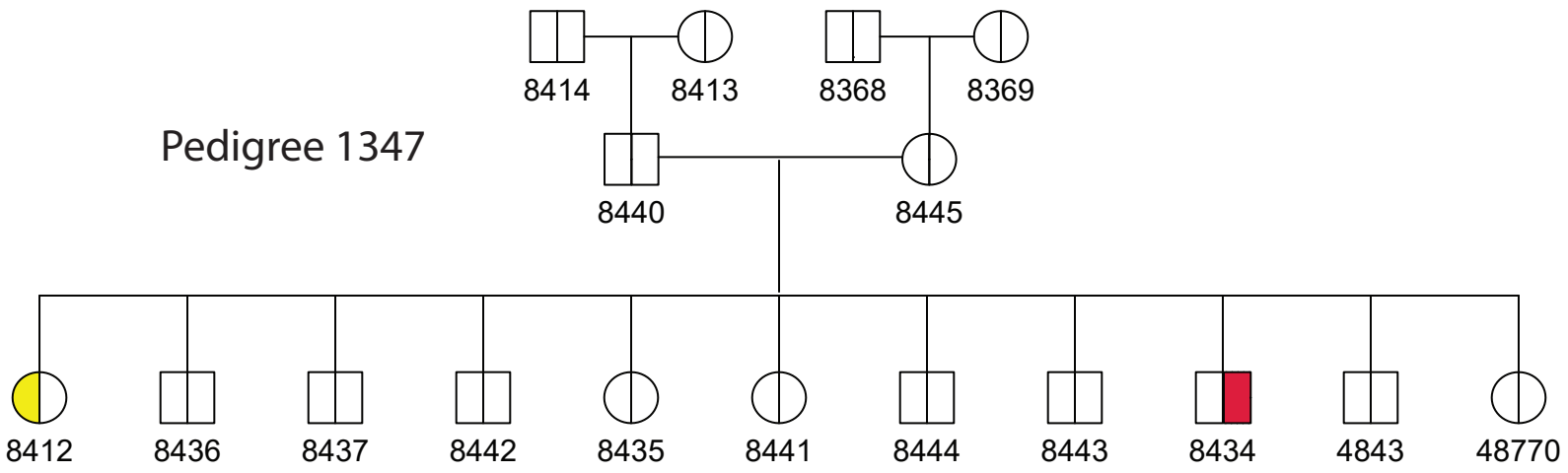
D

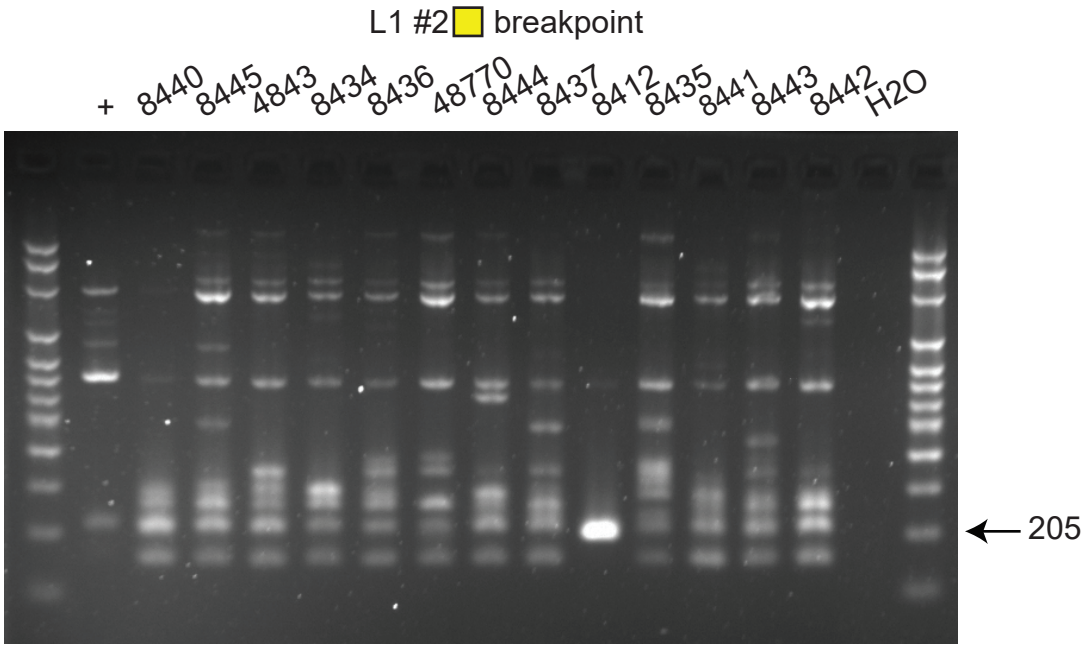
Pedigree 1346

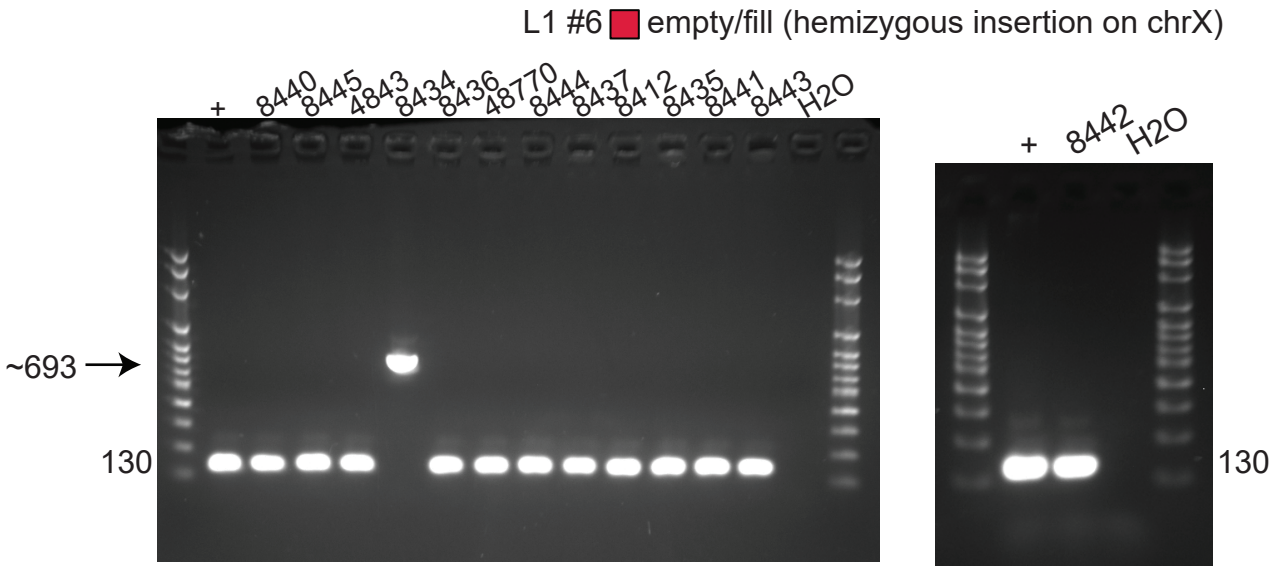




E

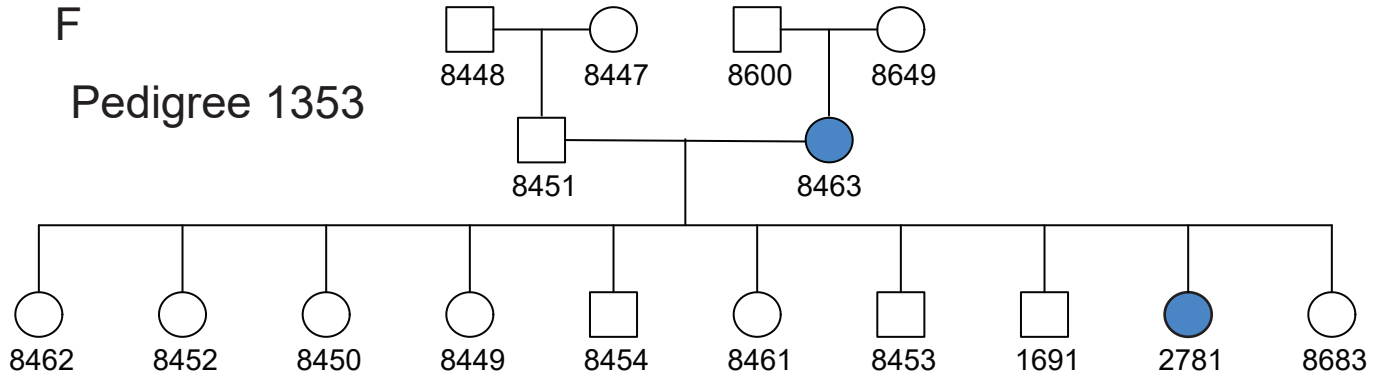

Pedigree 1347
8414
8413
8368
8369
8440
8445
8412
8436
8437
8442
8435
8441
8444
8443
8434
4843
48770


L1 #2 breakpoint
+
8440
8445
4843
8434
8436
48770
8444
8437
8412
8435
8441
8443
8442
H2O
205

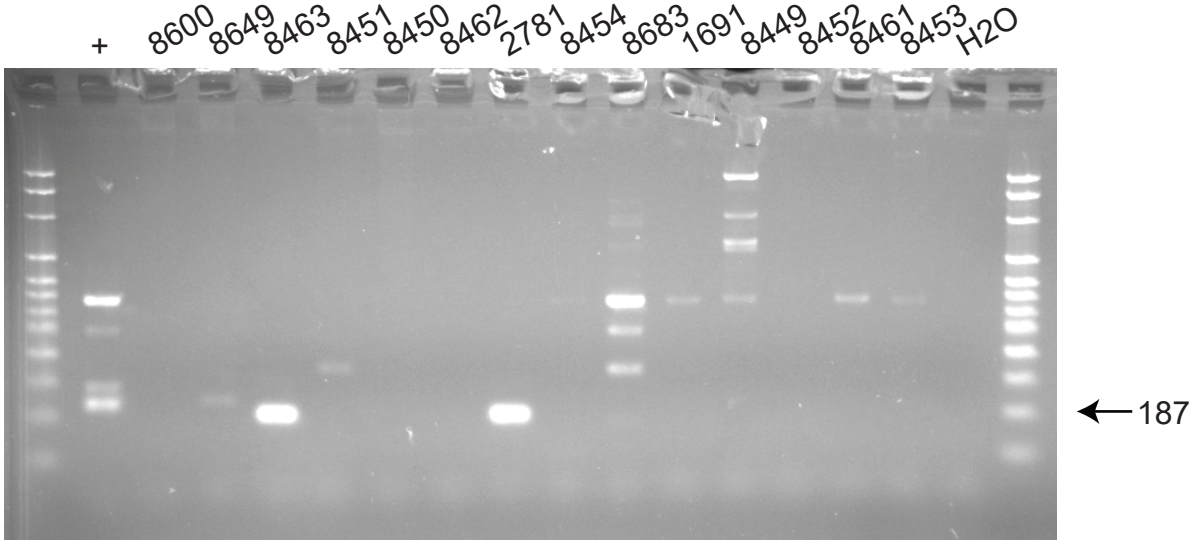

L1 #6 empty/fill (hemizygous insertion on chrX)
+
8440
8445
4843
8434
8436
48770
8444
8437
8412
8435
8441
8443
H2O
~693
130
+
8442
H2O
130

F

Pedigree 1353

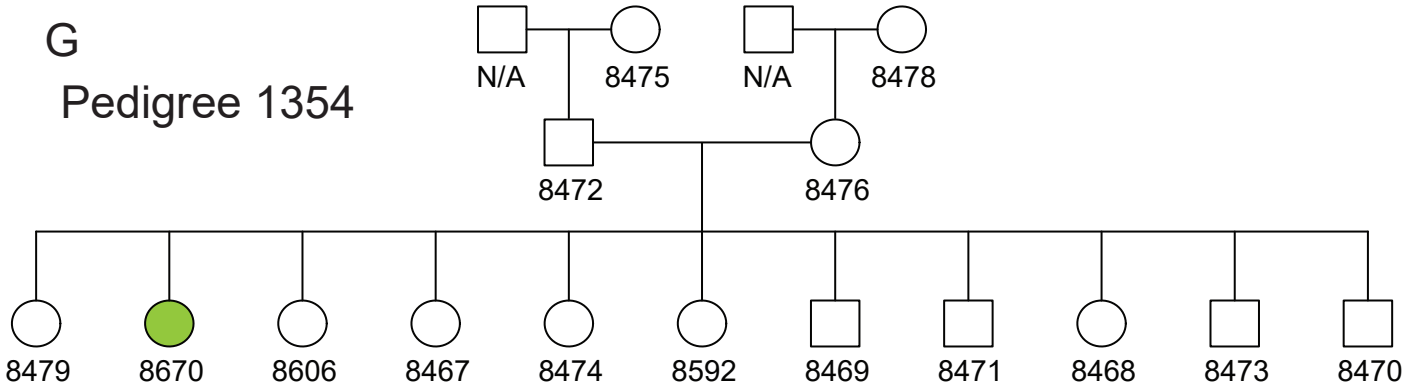


SVA #2 breakpoint

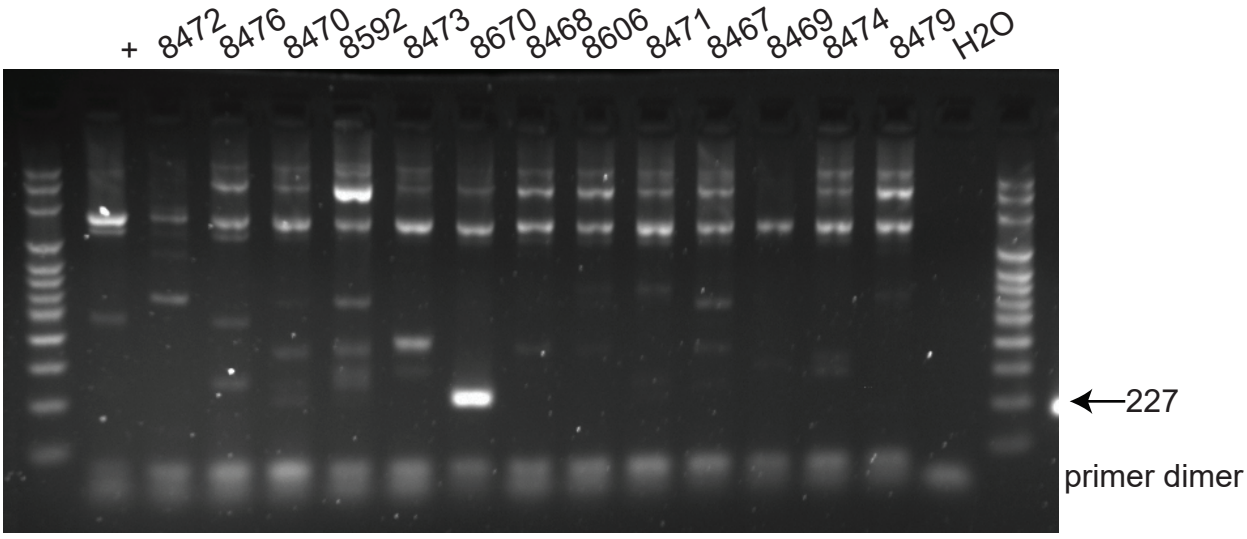


G

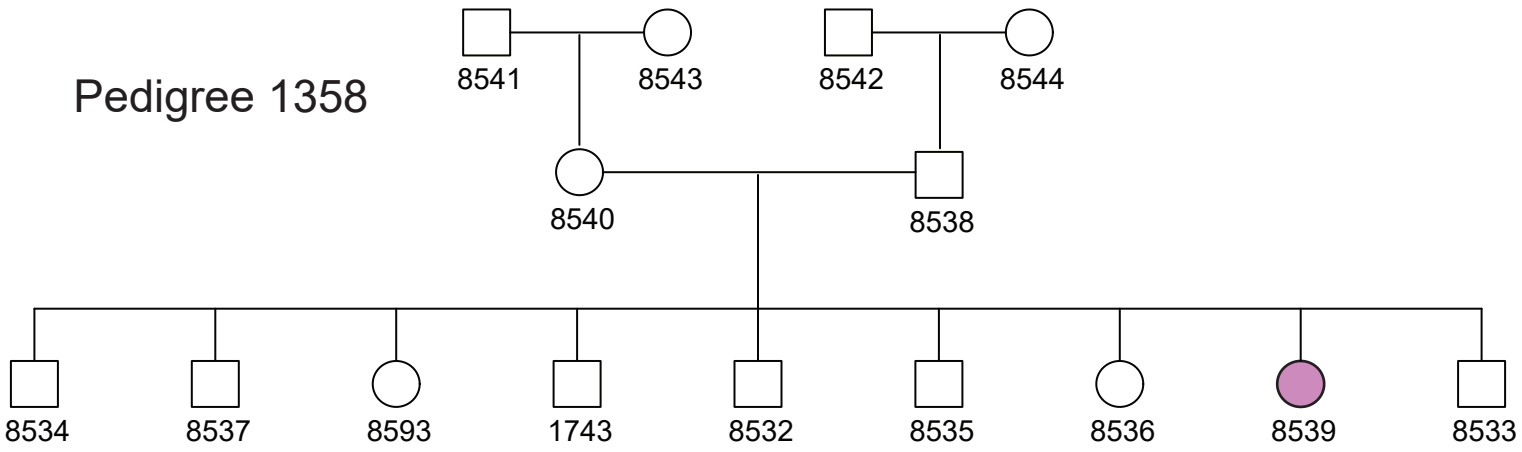
Pedigree 1354

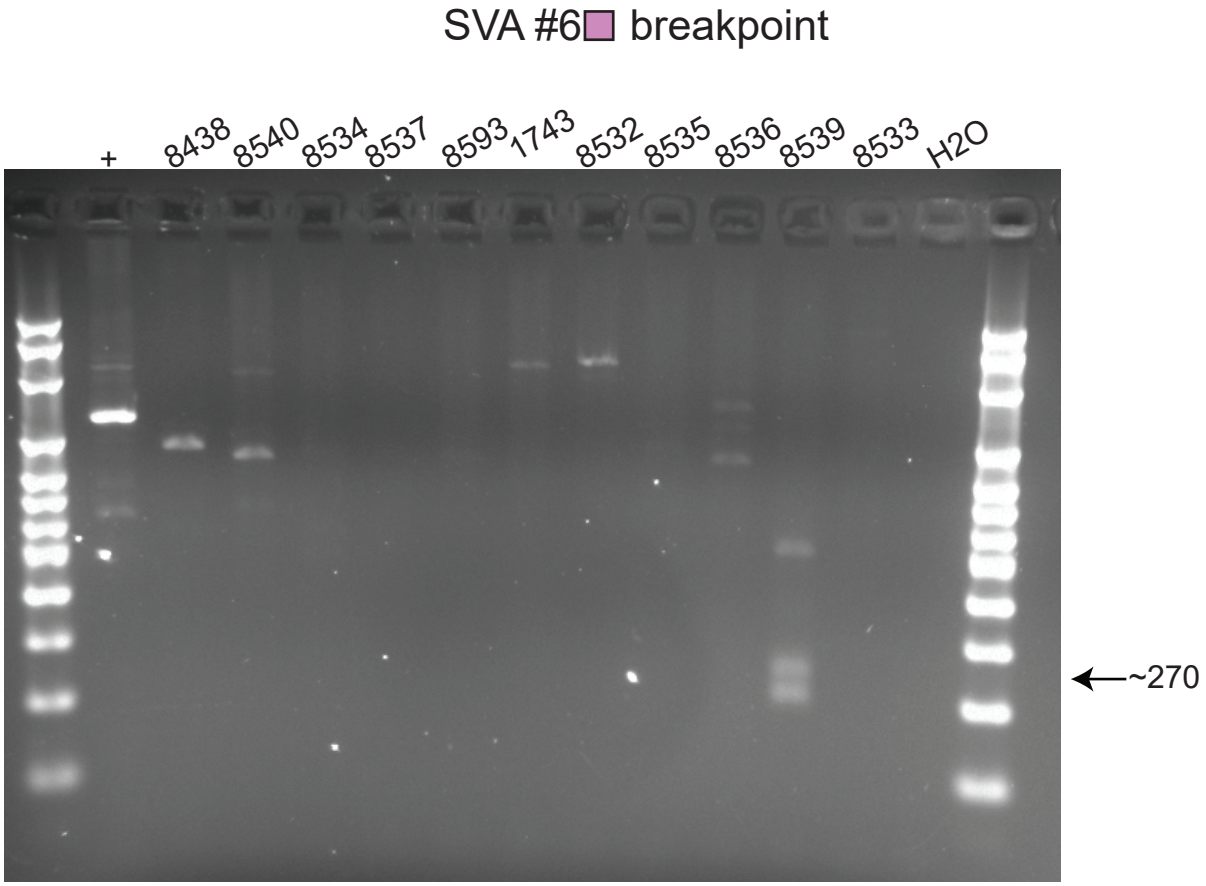


L1 #3 breakpoint

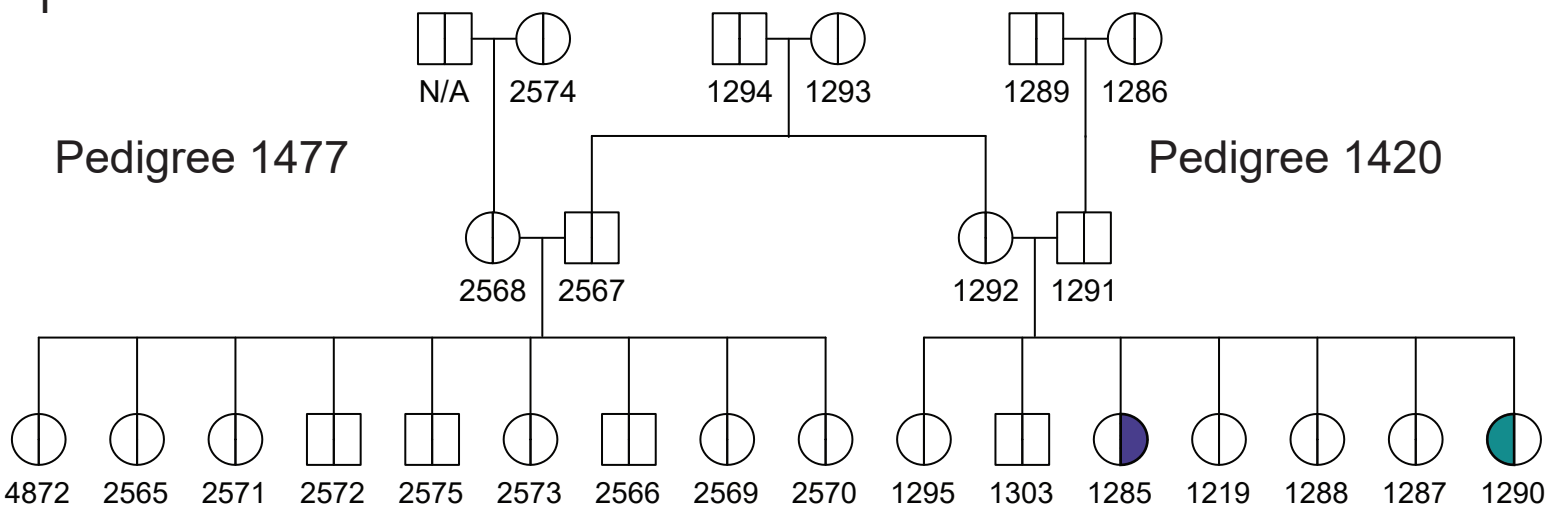


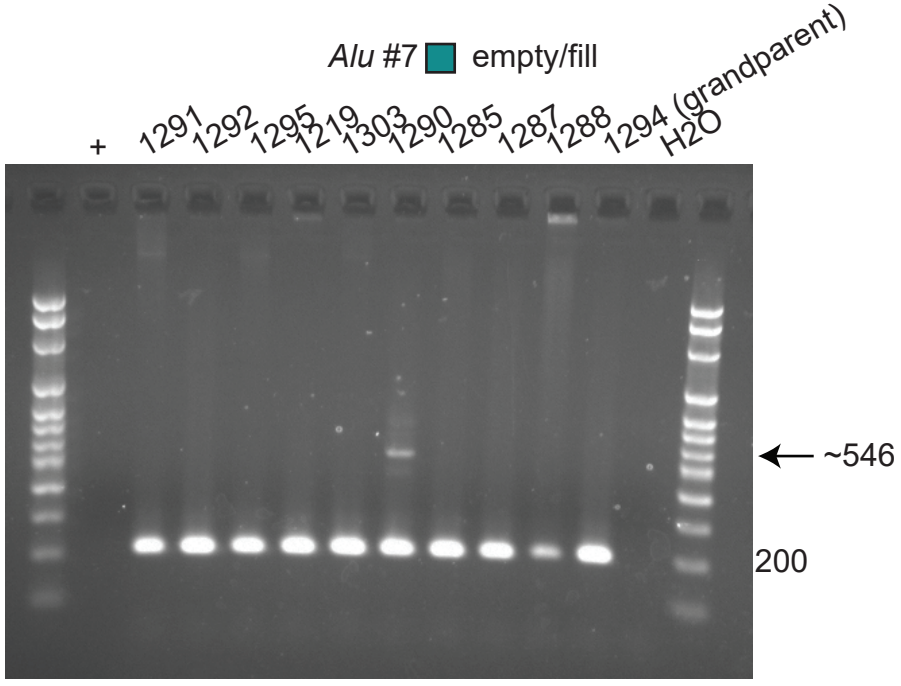
H

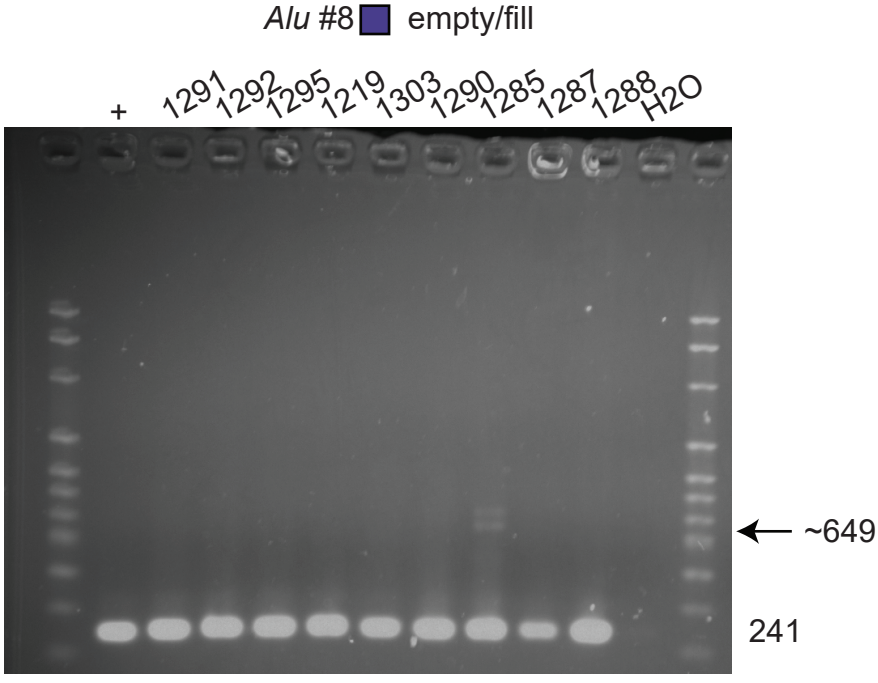

Pedigree 1358
8541
8543
8542
8544
8540
8538
8534
8537
8593
1743
8532
8535
8536
8539
8533


SVA #6 breakpoint
+
8438
8540
8534
8537
8593
1743
8532
8535
8536
8539
8533
H2O
~270

I

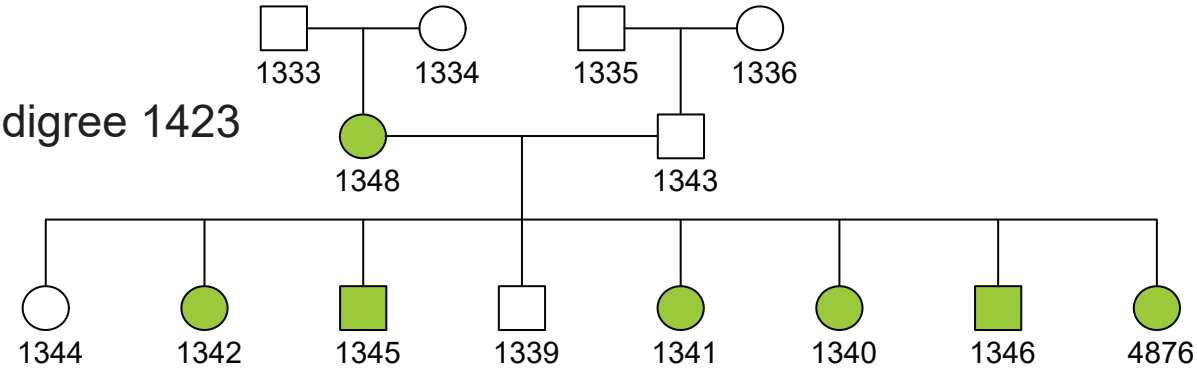

Pedigree 1477
Pedigree 1420
N/A
2574
1294
1293
1289
1286
2568
2567
1292
1291
4872
2565
2571
2572
2575
2573
2566
2569
2570
1295
1303
1285
1219
1288
1287
1290


Alu #7 empty/fill
+
1291
1292
1295
1219
1303
1290
1285
1287
1288
1294 (grandparent)
H2O
~546
200

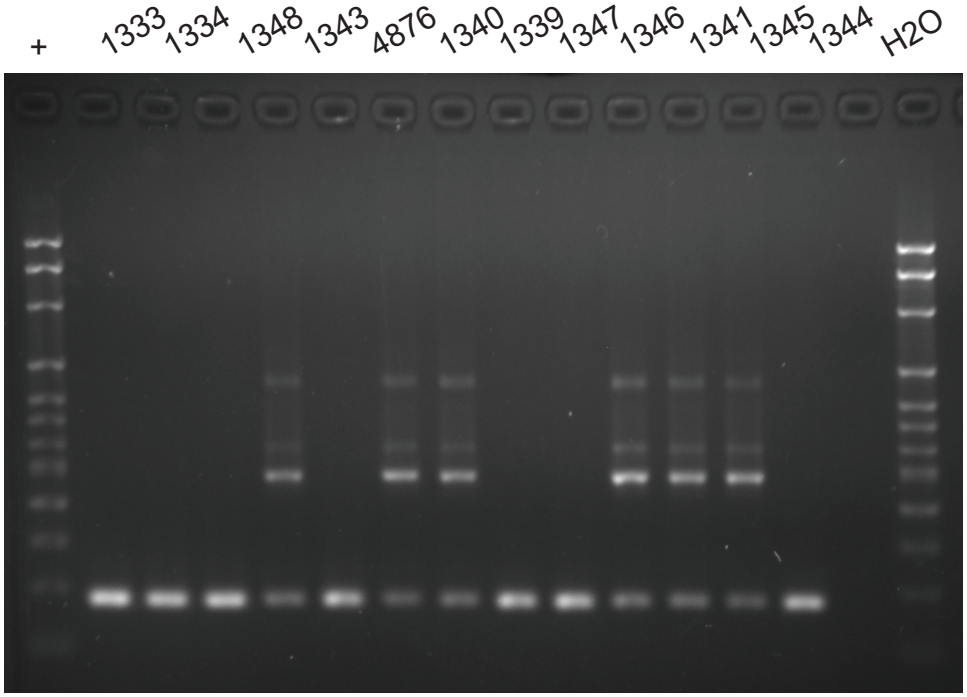

Alu #8 empty/fill
+
1291
1292
1295
1219
1303
1290
1285
1287
1288
H2O
~649
241

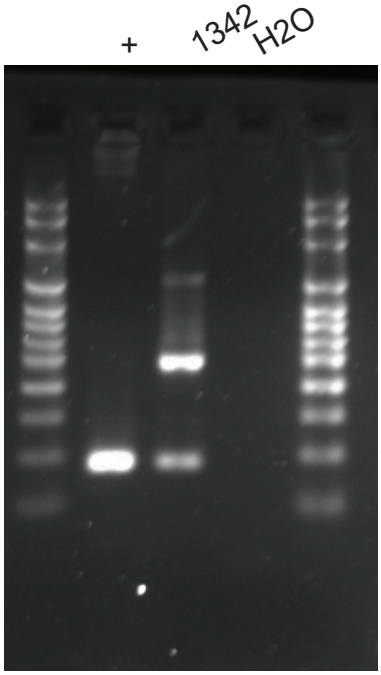
J

Pedigree 1423



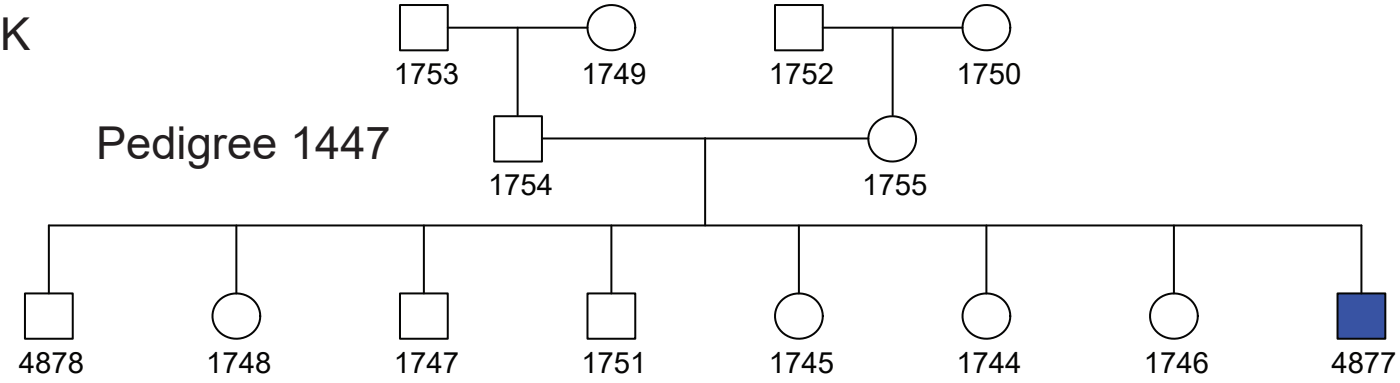
Alu #9 empty/fill



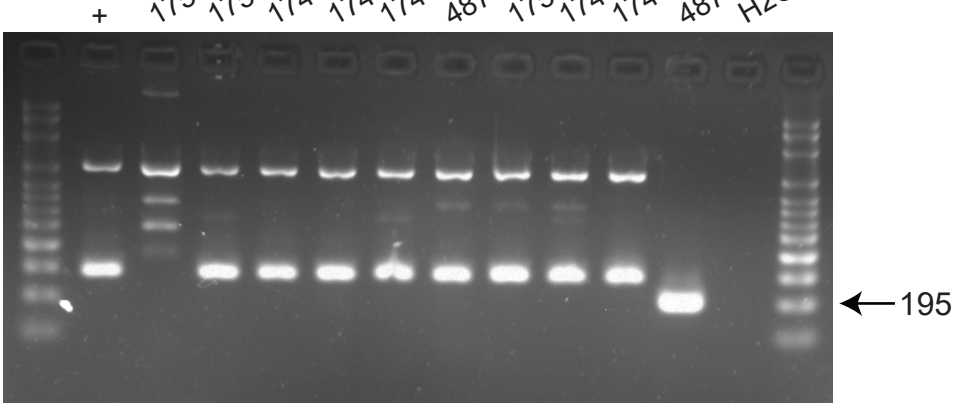


K

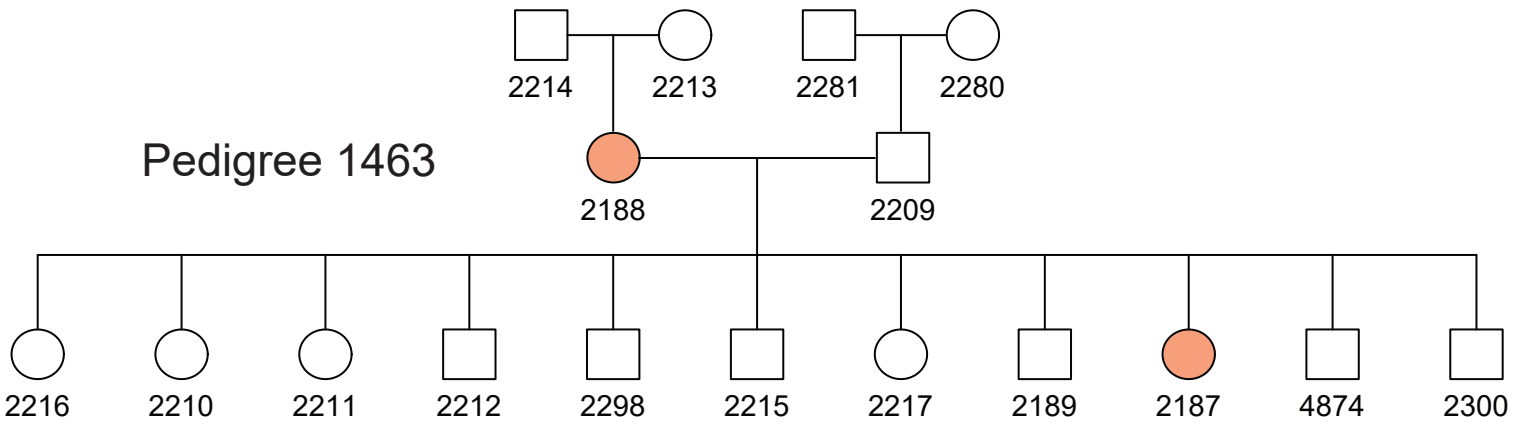
Pedigree 1447

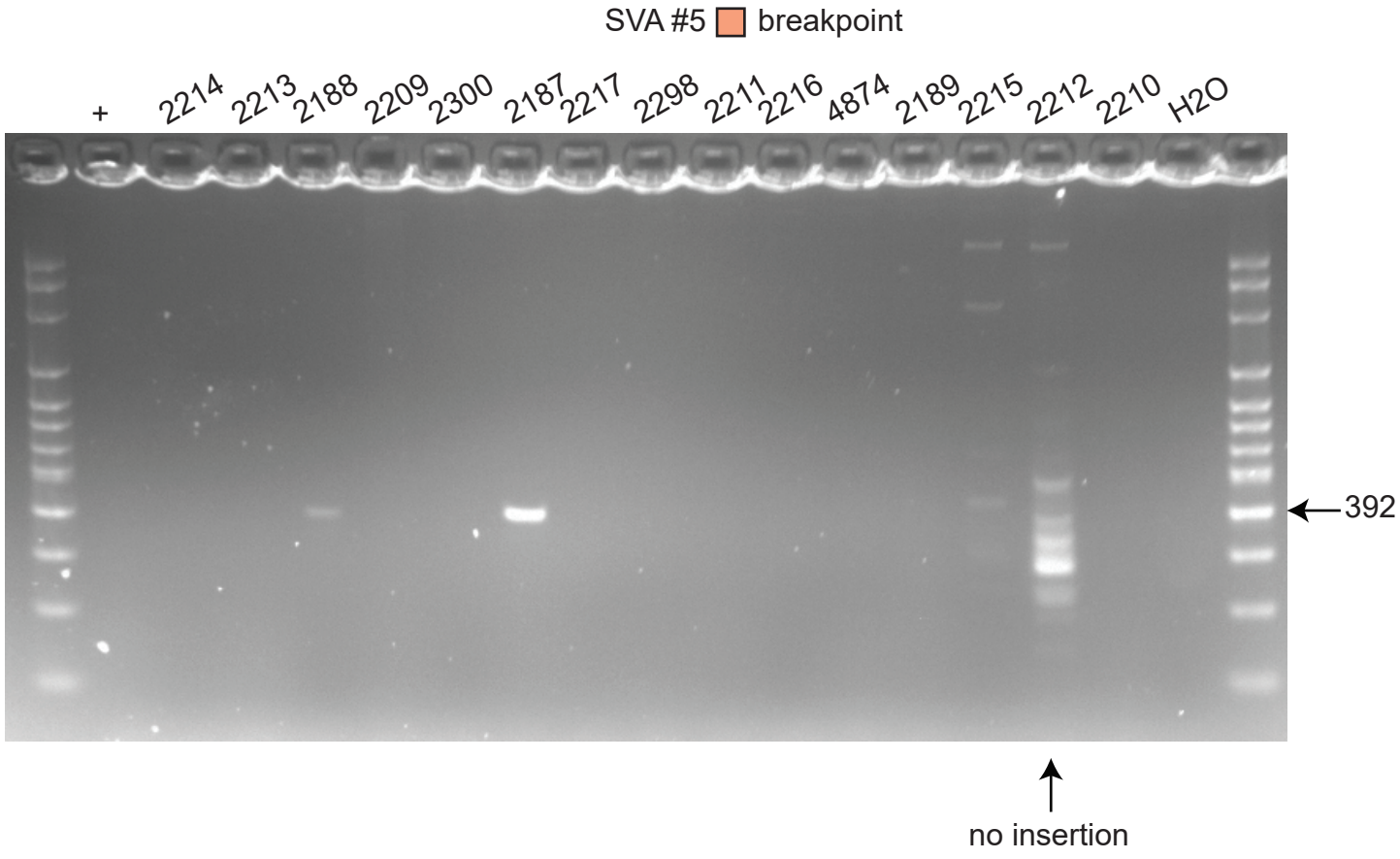


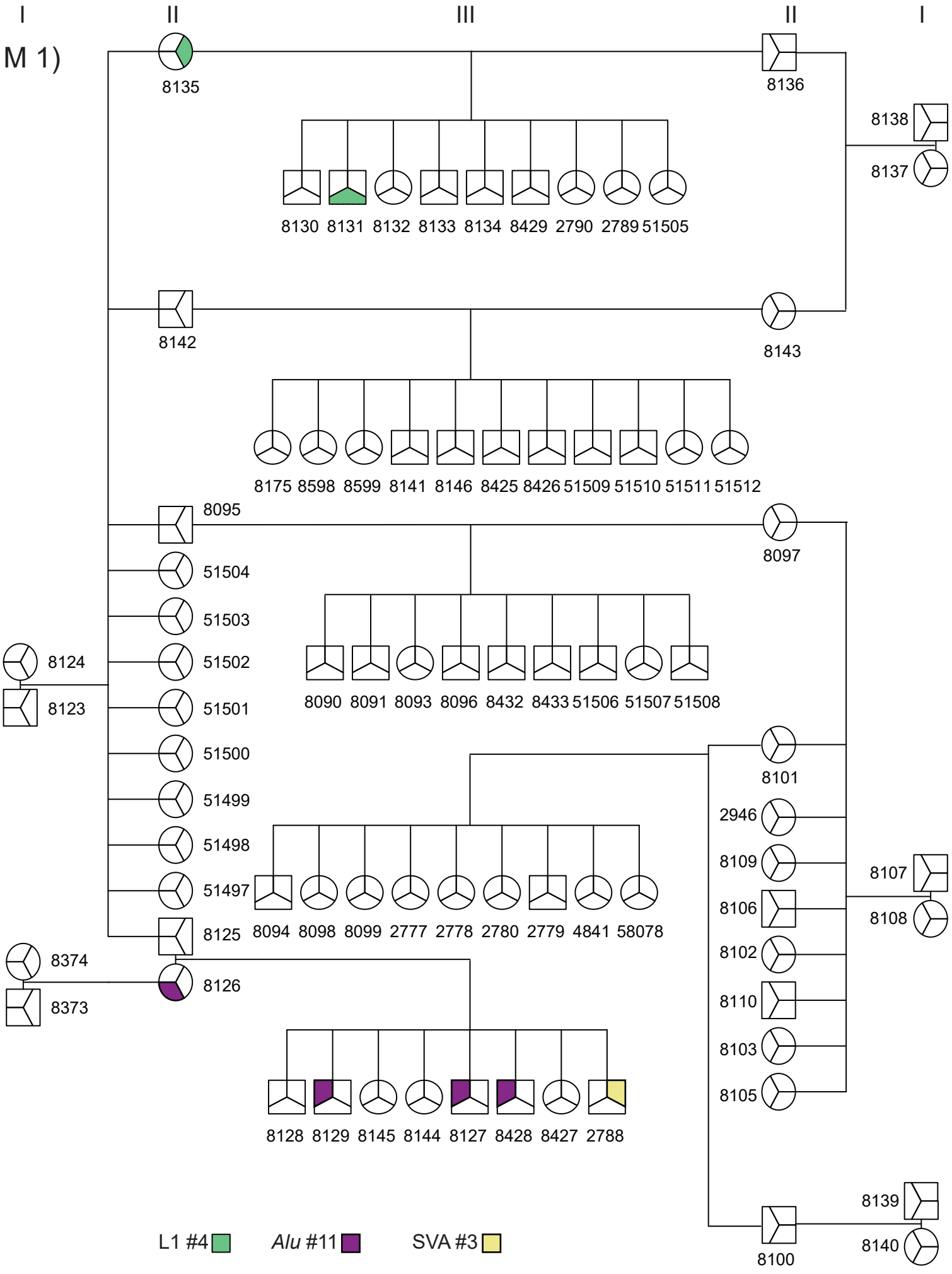
L1 #8 breakpoint

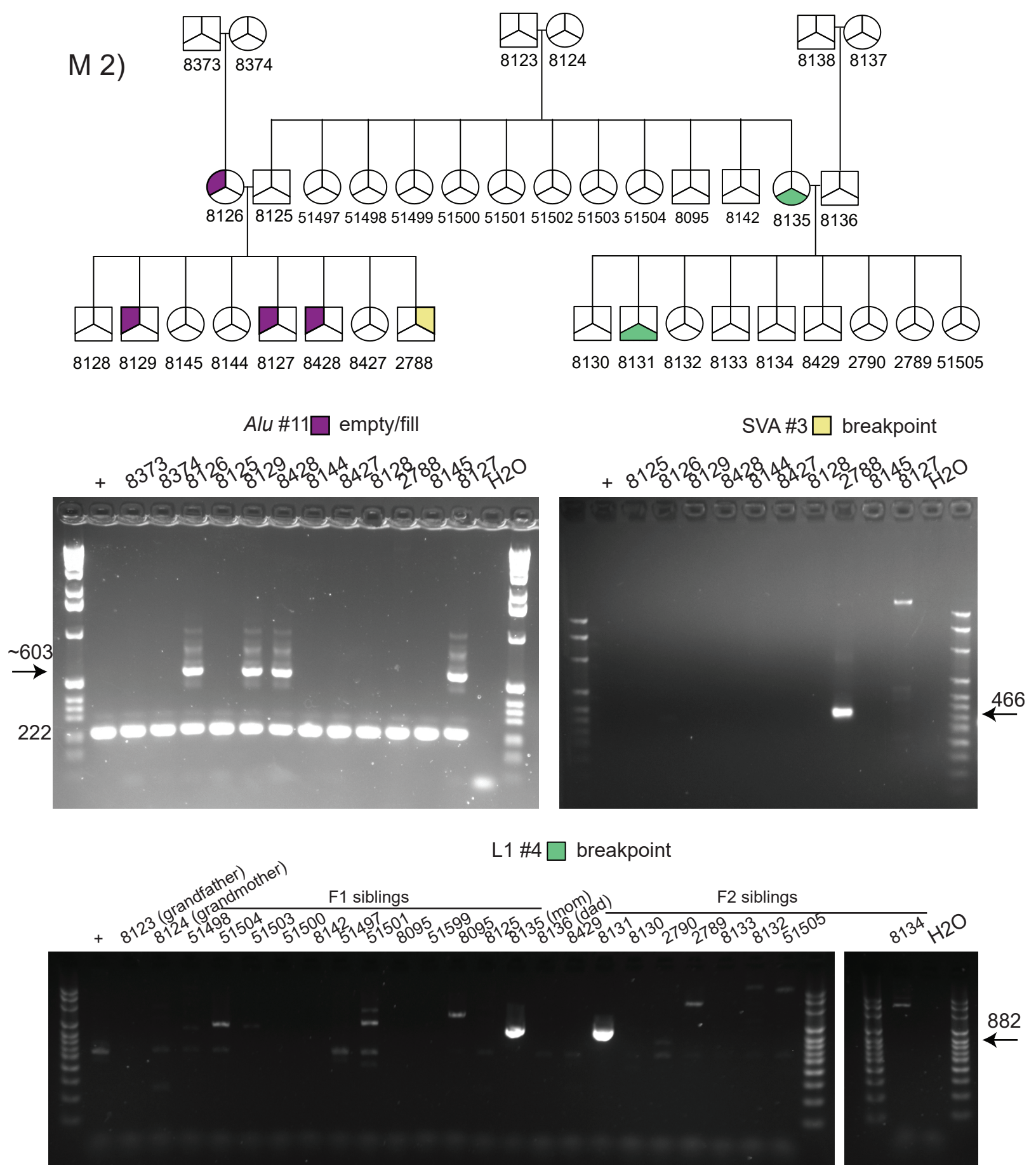


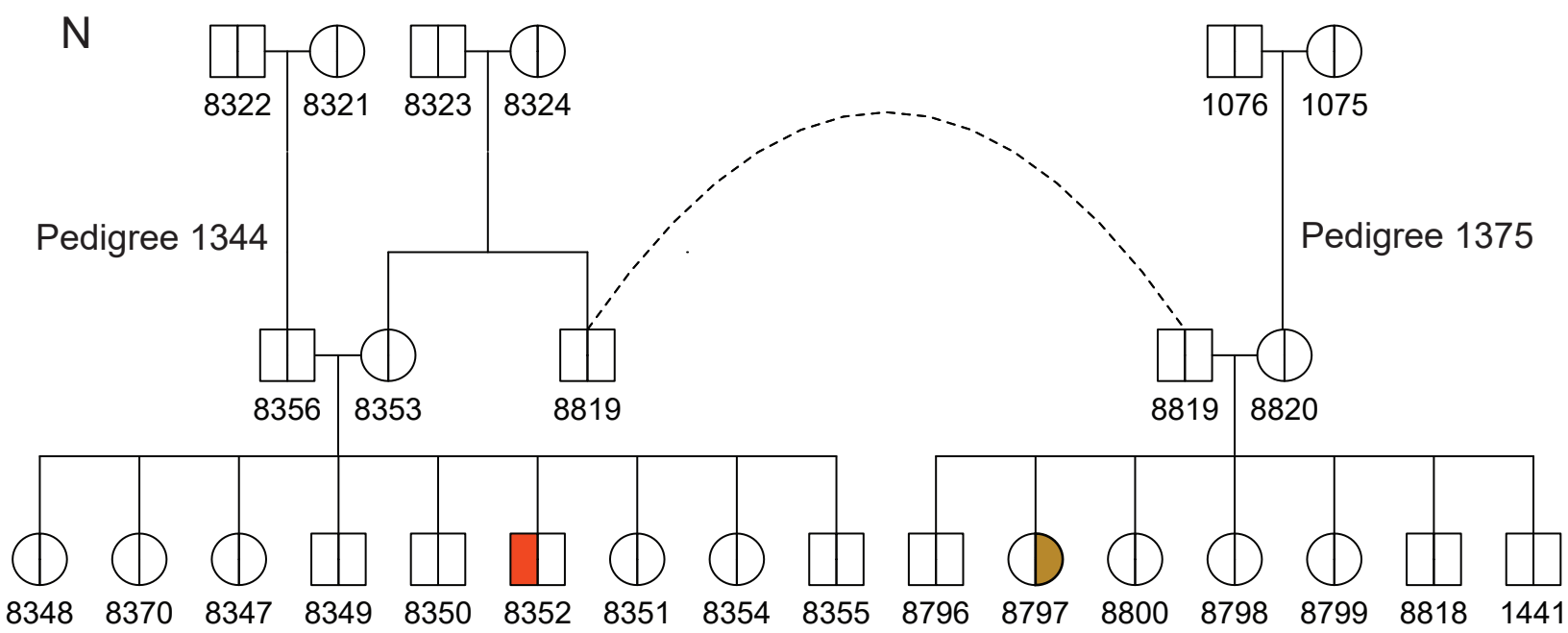
L

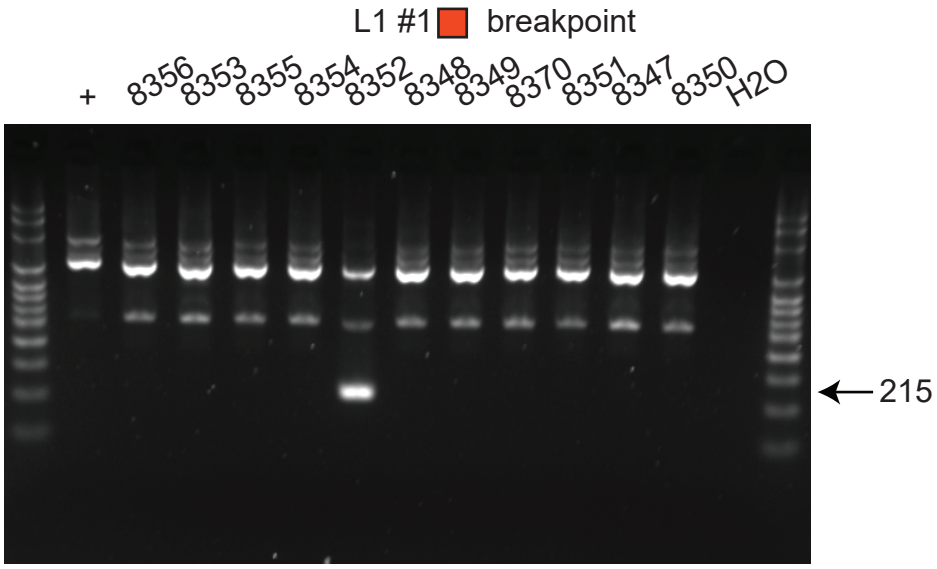

Pedigree 1463
2214
2213
2281
2280
2188
2209
2216
2210
2211
2212
2298
2215
2217
2189
2187
4874
2300

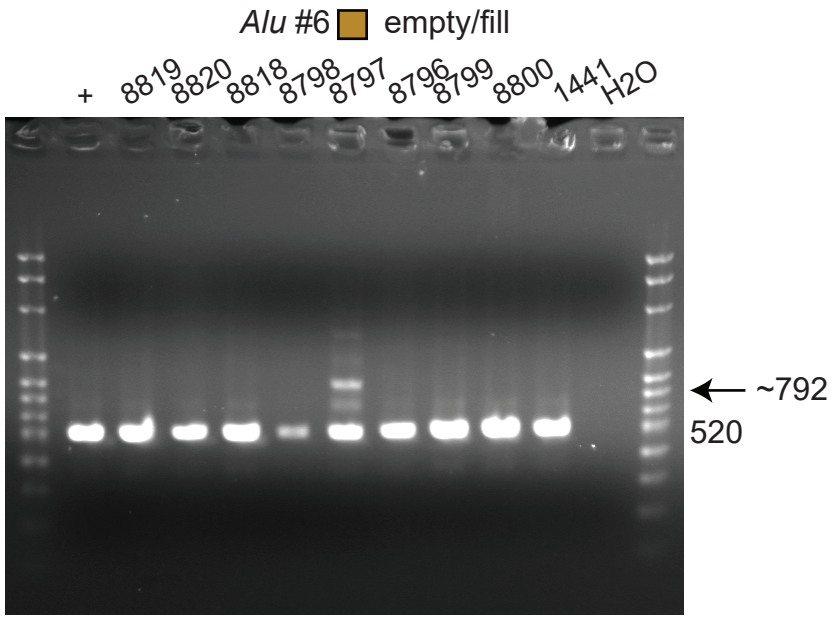

SVA #5 breakpoint
+
2214
2213
2188
2209
2300
2187
2217
2298
2211
2216
4874
2189
2215
2212
2210
H2O
392
no insertion


I
II
III
II
I
M 1)
8135
8136
8138
8137
8130 8131 8132 8133 8134 8429 2790 2789 51505
8142
8143
8175 8598 8599 8141 8146 8425 8426 51509 51510 51511 51512
8095
8097
51504
51503
51502
51501
51500
51499
51498
51497
8125
8124
8123
8090 8091 8093 8096 8432 8433 51506 51507 51508
8101
2946
8109
8106
8102
8110
8103
8105
8107
8108
8094 8098 8099 2777 2778 2780 2779 4841 58078
8374
8373
8126
8128 8129 8145 8144 8127 8428 8427 2788
8139
8140
8100
L1 #4
Alu #11
SVA #3

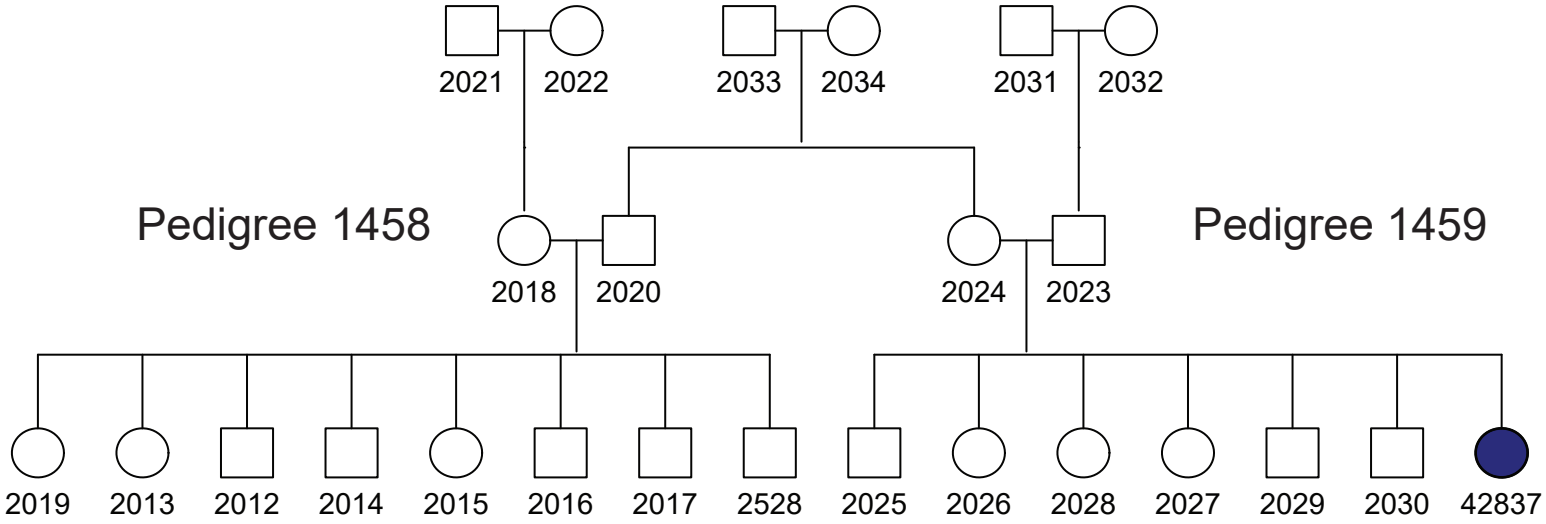

M 2)
8373
8374
8123
8124
8138
8137
8126
8125
51497
51498
51499
51500
51501
51502
51503
51504
8095
8142
8135
8136
8128
8129
8145
8144
8127
8428
8427
2788
8130
8131
8132
8133
8134
8429
2790
2789
51505
Alu #11 empty/fill
SVA #3 breakpoint
+
8373
8374
8126
8125
8129
8428
8144
8427
8128
2788
8145
8127
H2O
~603
222
+
8125
8126
8129
8428
8144
8427
8128
2788
8145
8127
H2O
466
L1 #4 breakpoint
+
8123 (grandfather)
8124 (grandmother)
F1 siblings
51498
51504
51503
51500
8142
51497
51501
8095
51599
8095
8125
8135 (mom)
8136 (dad)
F2 siblings
8429
8131
8130
2790
2789
8133
8132
51505
8134
H2O
882

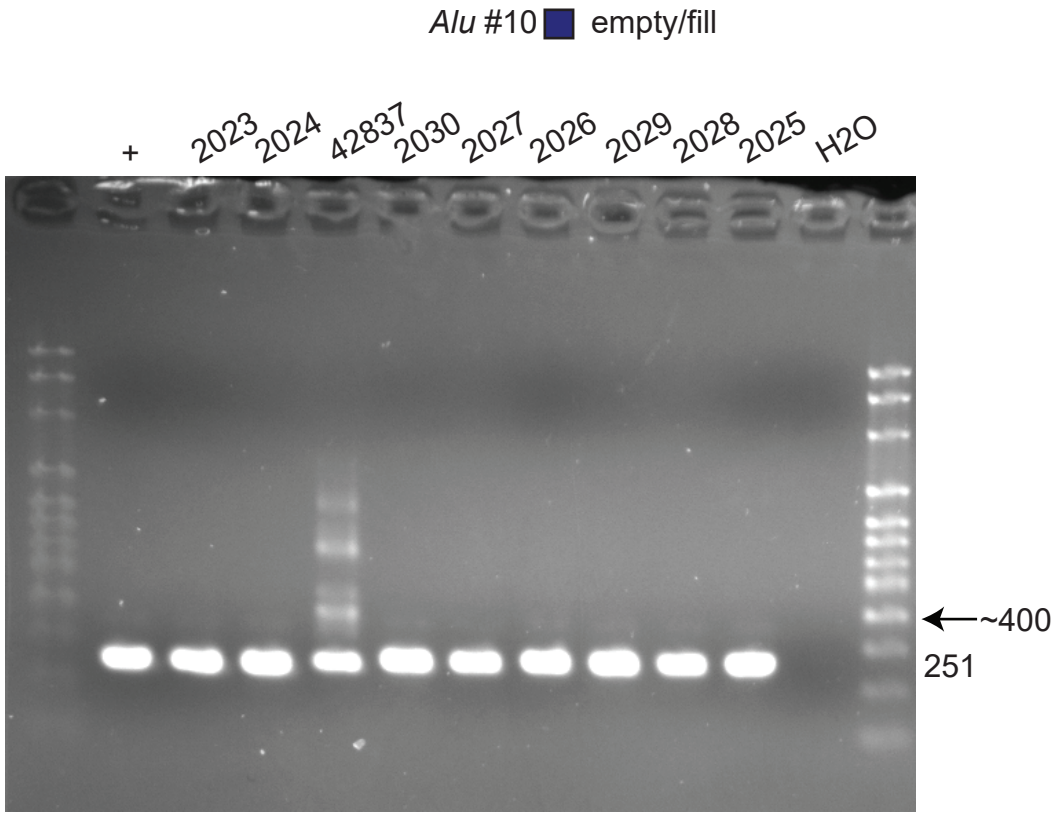

N
8322
8321
8323
8324
1076
1075
Pedigree 1344
Pedigree 1375
8356
8353
8819
8819
8820
8348
8370
8347
8349
8350
8352
8351
8354
8355
8796
8797
8800
8798
8799
8818
1441


L1 #1 breakpoint
+
8356
8353
8355
8354
8352
8348
8349
8370
8351
8347
8350
H2O
215

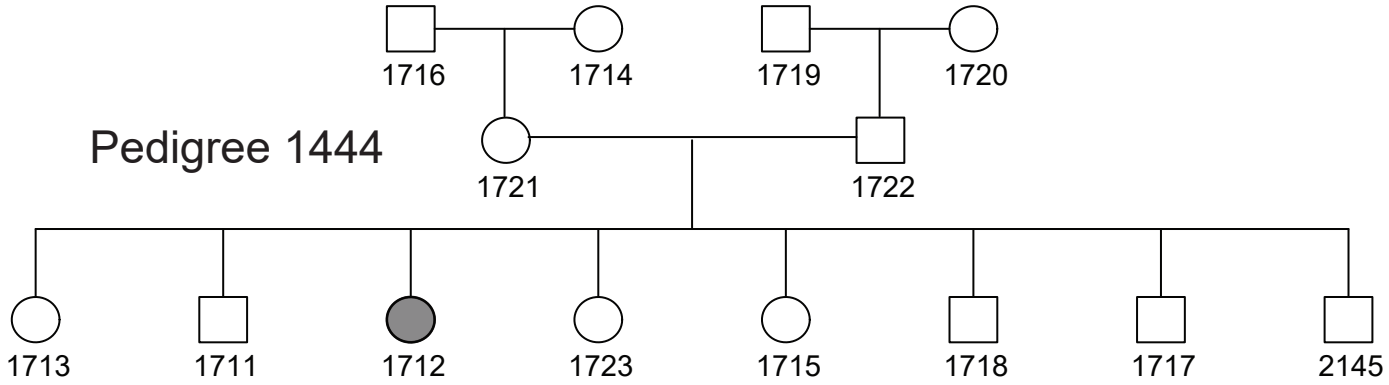

Alu #6 empty/fill
+
8819
8820
8818
8798
8797
8796
8799
8800
1441
H2O
~792
520

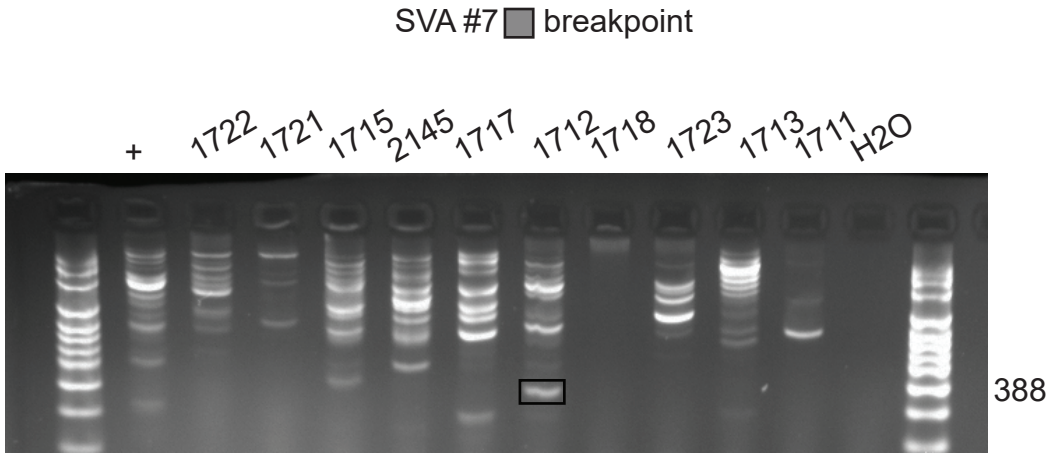
O


2021
2022
2033
2034
2031
2032
Pedigree 1458
Pedigree 1459
2018
2020
2024
2023
2019
2013
2012
2014
2015
2016
2017
2528
2025
2026
2028
2027
2029
2030
42837


Alu #10 empty/fill
+
2023
2024
42837
2030
2027
2026
2029
2028
2025
H2O
~400
251

P


Pedigree 1444
1716
1714
1719
1720
1721
1722
1713
1711
1712
1723
1715
1718
1717
2145


SVA #7 breakpoint
+
1722
1721
1715
2145
1717
1712
1718
1723
1713
1711
H2O
388

Supplemental Figure S1. Pedigree structure and PCR of the 26 *de novo* MEI. Each PCR reaction was run with a non-pedigree individual control DNA ("+") and H_2O. Empty/fill refers to amplifying around the MEI, including amplification of the WT chromosome. Breakpoint refers to amplification using one internal primer and one primer outside of the MEI. The black arrow indicates the MEI band. *Alu* #11 is shown with the NEB Quick-load 1kb DNA ladder. The rest of the MEIs are shown with the GeneMate Quanti-Marker 100bp ladder.

A) Pedigree 1331 and *Alu* #1-2, L1 #5, #7, and SVA #1, #4. Amplification of L1 #7 leads to non-specific bands of the same size as the amplicon; therefore, a full pedigree gel is not provided.
B) Pedigree 1341 and *Alu* #3.
C) Pedigree 1345 and *Alu* #4. 1) UCSC genome browser screenshot of the MEI insertion. This *Alu* caused a 1.7kb 5' deletion. The three primer sets are labeled in red (*Alu* primer set), black (SNP primer set) and purple (*Alu* + SNP primer set). 2) Pedigree diagram. 3) Gel of amplification of *Alu* #4 using the *Alu* primer set. The *Alu* band amplifies faintly in individual 8439. 4) Sanger sequence of 8327, 8439, and their parents at this locus. When the locus is amplified around a nearby SNP, 8327 is heterozygous G/C and 8439 is homozygous G/G. When the MEI insertion and SNP are amplified together (*Alu* + SNP primer set), the C allele is on the same band as the *Alu* band, indicating that the *Alu* landed on the maternal chromosome. Individual 8439 has a C in the *Alu* band. Given the haplotype discrepancy and lack of recombination evidence, we suspect that the 8439 DNA has low-level contamination of individual 8327.
D) Pedigree 1346 and *Alu* #5.
E) Pedigree 1347 and L1 #2 and #6.
F) Pedigree 1353 and SVA #2.
G) Pedigree 1354 and L1 #3.
H) Pedigree 1358 and SVA #6.
I) Pedigree 1420/1477 and *Alu* #7-8.
J) Pedigree 1423 and *Alu* #9.
K) Pedigree 1447 and L1 #8.
L) Pedigree 1463 and SVA #5.
M) Pedigree 1328/1329 and *Alu* #11, L1 #4, and SVA #3. 1) Full diagram of the pedigree. 2) Subset of pedigree and gel of the three MEIs.
N) Pedigree of 1344/1375 and *Alu* #6 and L1 #1.
O) Pedigree of 1458/1459 and *Alu* #10.
P) Pedigree of 1444 and SVA #7

Alu #1
Chr4q24

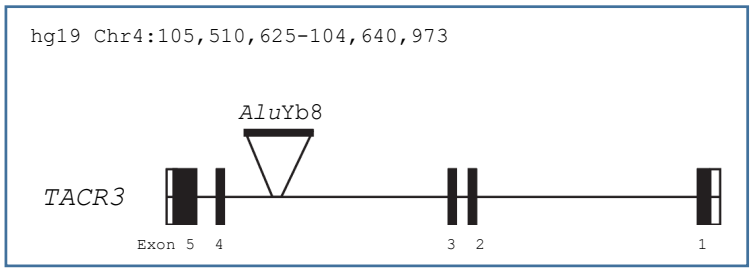
Empty site:

1st strand cleavage

5′-ATGCAAAAAGAAAGAAACTGCTTTTTATTATAAGAAATAACTCAA-3′
3′-TAGCTTTTTCTTTCTTTGACGAAAAATAATATTCTTTATTGAGTT-5′

Filled site:

5′-ATGCAAAAAGAAAGAAACTGCTTTTTT$_{63}$ *Alu*Yb8 GAAACTGCTTTTTATTATAAGAAATAACTCAA-3′
3′-TAGCTTTTTCTTTCTTTGACGAAAAA$_{63}$ CTTTGACGAAAAATAATATTCTTTATTGAGTT-5′



- strand
TSD: 13bp
Chr4:104,530,552-104,530,567

Alu #2
Chr1p36.22

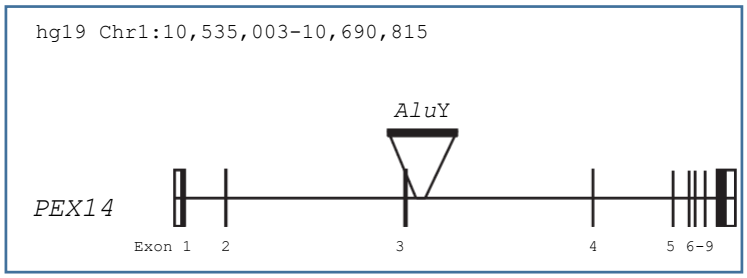
Empty site:

5′-AAGACTCCATCTCAAAAATAGAAACAAAAACAAACAAA-3′
3′-TTCTGAGGTAGAGTTTTTATCTTTGTTTTTGTTTGTTT-5′

1st strand cleavage

Filled site:

5′-AAGACTCCATCTCAAAAATAGAAACA *Alu*Y A$_{44}$AAAAAATAGAAACAAAAACAAACAAA-3′
3′-TTCTGAGGTAGAGTTTTTATCTTTGT T$_{44}$TTTTTTATCTTTGTTTTTGTTTGTTT-5′



+ strand
TSD: 13
Chr1:10,602,818-10,602,830

Alu #3
Chr3p24.2

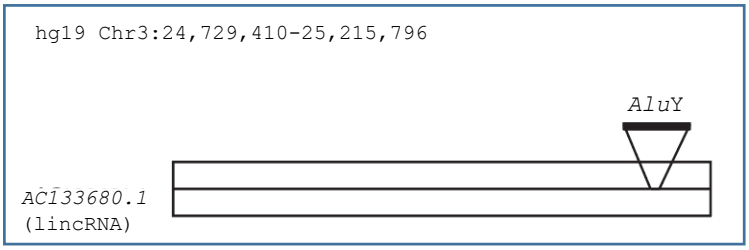
Empty site:

5′-TTCTACACATTTATTAAGATTTATCTATTGAACAGTCAACTCCT-3′
3′-AAGATGTGTAAATAATTCTAAATAGATAACTTGTCAGTTGAGGA-5′

1st strand cleavage

Filled site:

5′-TTCTACACATTTATTAAGATTTATCT *AluY* A_{80}AAGATTTATCTATTGAACAGTCAACTCCT-3′
3′-AAGATGTGTAAATAATTCTAAATAGA T_{80}TTCTAAATAGATAACTTGTCAGTTGAGGA-5′



+ strand
TSD: 11
Chr3:25,163,270-25,163,280

Alu #4
Chr12p13.31

Empty site:

deleted

5′-GATTAATGGGTTGATGTATTA...1.7kb...TTTTGATAGAAATAGAAACATGGAACTCAATAGAAA-3′
3′-CTAATTACCCAACTACATAAT...1.7kb...AAAACTATCTTTATCTTTGTACCTTGAGTTATCTTT-5′

1st strand cleavage

Filled site:

5′-GATTAATGGGTTGATGTATT**T** *AluYa* A_{41}AGAAACATGGAACTCAATAGAAA-3′
3′-CTAATTACCCAACTACATAA**A** T_{41}TCTTTGTACCTTGAGTTATCTTT-5′

+ strand
TSD: 5′ deletion
5′ end: Chr12:8,680,801-8,680,820
3′ end: Chr12:8,682,497-8,682,519

Intergenic

Alu #5
Chr6p22.2

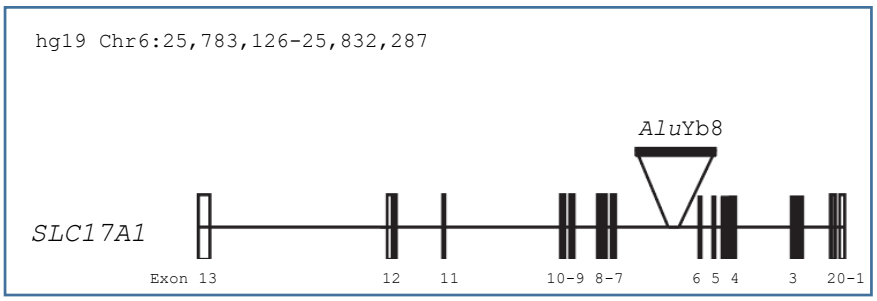
Empty site:

5′-AATATACATGTAATAAAAAAAATCACAAGGTGTGATATACATGTA-3′
3′-TTATATGTACATTATTTTTTTTTAGTGTTCCACACTATATGTACAT-5′

1st strand cleavage

Filled site:

5′-AATATACATGTAATAAAAAAAATCAC *Alu*Yb8 A_{66}AAAAAAAATCACAAGGTGTGATATACATGTA-3′
3′-TTATATGTACATTATTTTTTTTTAGTG T_{66}AAAAAAAATCACAAGGTGTGATATACATGTA-5′



+ strand
TSD: 12
Chr6:25,815,318-25,815,329

Alu #6
Chr2p13.1

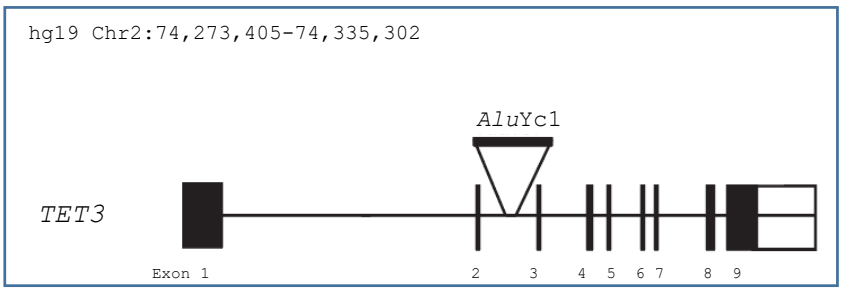
Empty site:

5′-AAGAGCTTGTTTTAAAAAATTAATTCGTTAATTAAAAAAAAAT-3′
3′-TTCTCGAACAAAATTTTTTAATTAAGCAATTAATTTTTTTTTA-5′

1st strand cleavage

Filled site:

5′-AAGAGCTTGTTTTAAAAAATTAATTCGTT *Alu*Yc1 A_{24}AAAAAATTAATTCGTTAATTAAAAAAAAAT-3′
3′-TTCTCGAACAAAATTTTTTAATTAAGCAA T_{24}TTTTTTAATTAAGCAATTAATTTTTTTTTA-5′



+ strand
TSD: 16
Chr2:74,305,312-74,305,327

Alu #7
Chr1p13.3

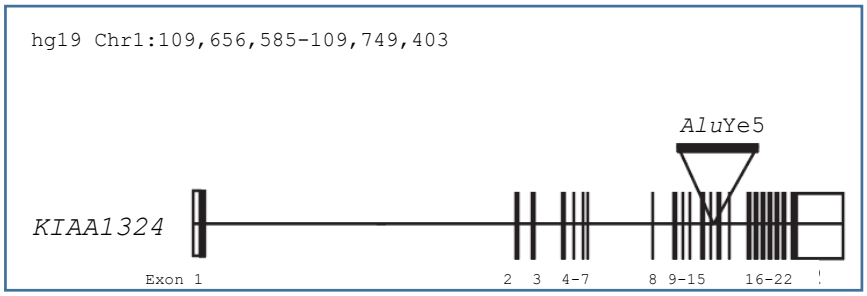
Empty site:

5′-TAAAAATAAACAAAAATAAGAAATAAAATCAGCCATAATCCTACT-3′
3′-ATTTTTATTTGTTTTTATTCTTTATTTTAGTCGGTATTAGGATGA-5′

1st strand cleavage

Filled site:

5′-TAAAAATAAACAAAAATAAGAAATAAAATCAG [*Alu*Ye5] A_{47}AAGAAATAAAATCAGCCATAATCCTACT-3′
3′-ATTTTTATTTGTTTTTATTCTTTATTTTAGTC T_{47}TTCTTTATTTTAGTCGGTATTAGGATGA-5′



+ strand
TSD: 15
Chr1:109,734,843-109,734,857

Alu #8
Chr2q22.2

Empty site:

1st strand cleavage

5′-TTGGAACACCCAGATGTATAAAACAAATATTATTAGATCTAAAG-3′
3′-AACCTTGTGGGTCTACATATTTTGTTTATAATAATCTAGATTTC-5′

Filled site:

5′-TTGGAACACCCAGATGTATAAAACAAATATTTT_{113} [*Alu*Ya5] ATAAAACAAATATTATTAGATCTAAAG-3′
3′-AACCTTGTGGGTCTACATATTTTGTTTATAAA_{113} TATTTTGTTTATAATAATCTAGATTTC-5′

- strand
TSD: 14
Chr2:143,161,204-143,161,217

Intergenic

Alu #9
Chr7q31.1

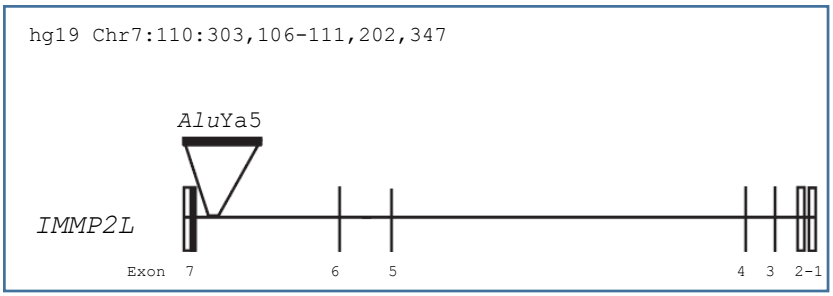
Empty site:

5′-TAGAGTCAAGTTCCATCAAATCAGGAGGCACAAAATATCGGTGC-3′
3′-ATCTCAGTTCAAGGTAGTTTAGTCCTCCGTGTTTTATAGCCACG-5′

1st strand cleavage

Filled site:

5′-TAGAGTCAAGTTCCATCAAATCAGGAGGCAC *AluYa5* A_{107}AAATCAGGAGGCACAAAATATCGGTGC-3′
3′-ATCTCAGTTCAAGGTAGTTTAGTCCTCCGTG T_{107}TTTAGTCCTCCGTGTTTTATAGCCACG-5′



\+ strand
TSD: 14
Chr7:110,305,748-110,305,761

Alu #10
Chr4q22.3

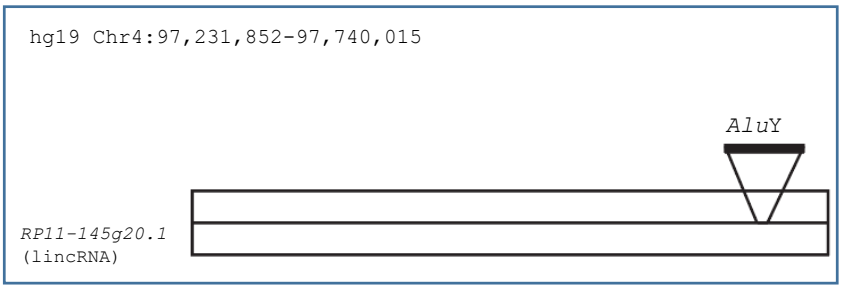
Empty site:

1st strand cleavage

5′-TTTTCCTGCTTTATTCTAAGTCATTTTAACTATATTTATTACAT-3′
3′-AAAAGGACGAAATAAGATTCAGTAAAATTGATATAAATAATGTC-5′

Filled site:

5′-TTTTCCTGCTTTATTCTAAGTCATTTTT_{113} *AluY* TAAGTCATTTTAACTATATTTATTACAT-3′
3′-AAAAGGACGAAATAAGATTCAGTAAAAA_{113} ATTCAGTAAAATTGATATAAATAATGTC-5′



\- strand
TSD: 11
Chr4:97,701,273-97,701,283

Alu #11
Chr17p11.2

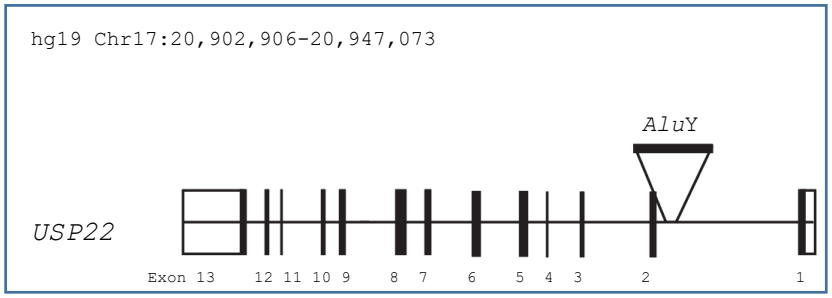
Empty site:

1st strand cleavage

5′-ACCTACTATGTACCCATCAGATTTTTTAAAAAAGAAATTAA-3′
3′-TGGATGATACATGGGTAGTCTAAAAAATTTTTTCTTTAATT-5′

Filled site:

5′-ACCTACTATGTACCCATCAGATTTTTTT$_{113}$ *AluY* CCATCAGATTTTTTAAAAAAGAAATTAA-3′
3′-TGGATGATACATGGGTAGTCTAAAAAAA$_{113}$ GGTAGTCTAAAAAATTTTTTCTTTAATT-5′



- strand
TSD: 14
Chr17:20,932,724-20,932,737

L1 #1
Chr5q15

Empty site:

double-strand break

5′-ATTTAAAATTAGTTTGAATCCTCTGCTTGTTCCTGATTAAA-3′
3′-TAAATTTTAATCAAACTTAGGAGACGAACAAGGACTAATTT-5′

Filled site:

L1 microhomology

5′-ATTTAAAATTAGTTTGA partial ORF2 TCCTCTGCTTGTTCCTGATTAAA-3′
3′-TAAATTTTAATCAAACT AGGAGACGAACAAGGACTAATTT-5′

- strand
1bp deletion
Chr5:95,081,487

L1 #2
Chr4q28.2

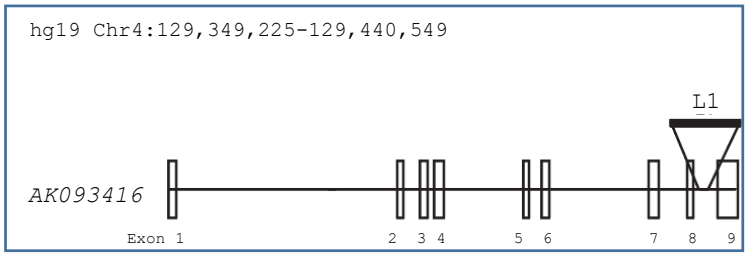
Empty site:

5′-TGCACTTAACAGTATCATGTAAGTATCAGTTTCT-3′
3′-ACGTGAATTGTCATAGTACATTCATAGTCAAAGA-5′

1st strand cleavage

Filled site:

5′-TGCACTTAACAGTATCATGTAA UTR ORF1 ORF2/UTR A_{29} 82 bp A_{96}AACAGTATCATGTAAGTATCAGTTTCT-3′
3′-ACGTGAATTGTCATAGTACATT T_{29} T_{96}TTGTCATAGTACATTCATAGTCAAAGA-5′



+ strand
TSD: 15
Chr4:129,438,004-129,438,018
Transduced sequence:
Chr4:112,628,960-112,629,041

L1 #3
Chr3p14.3

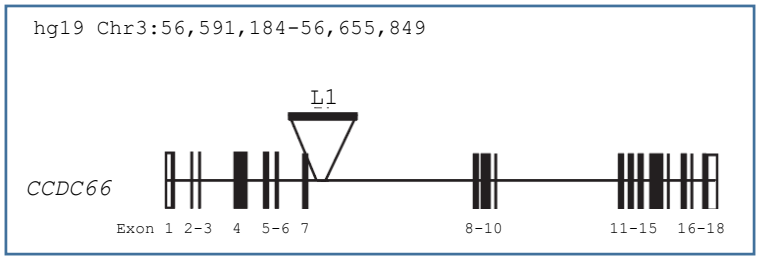
Empty site:

1st strand cleavage

5′TTTTGGGTTTTTTGTATGTGTGTTTTTTTTAAAAAAAGAAACAG-3′
3′AAAACCCAAAAAACATACACACAAAAAAAATTTTTTTCTTTGTC-5′

Filled site:

5′-TTTTGGGTTTTTTGTATGTGTGTTTTTTTTT$_{??}$ ORF2/UTR GTATGTGTGTTTTTTTTAAAAAAAGAAACAG-3′
3′-AAAACCCAAAAAACATACACACAAAAAAAAA$_{??}$ CATACACACAAAAAAAATTTTTTTCTTTGTC-5′



- strand
TSD: 17
Chr3:56,605,741-56,605,757

L1 #4
Chr1p31.2

Empty site:

5′-TCTCGTGTAGTTGTAAGAGGAGCAGCTTTGATAATTGTTC-3′
3′-AGAGCACATCAACATTCTCCTCGTCGAAACTATTAACAAG-5′

1st strand cleavage

Filled site:

5′-TCTCGTGTAGTTGTAAGAGGAGCAGCT [ORF2/UTR] A_{32} [846 bp] A_{94}AAGAGGAGCAGCTTTGATAATTGTTC-3′
3′-AGAGCACATCAACATTCTCCTCGTCGA T_{32} T_{94}TTCTCCTCGTCGAAACTATTAACAAG-5′

+ strand
TSD: 13
Chr1:69,238,322-69,238,334
Transduced sequence:
Chr5:112,703,068-112,703,913

Intergenic

L1 #5
Chr4q28.2

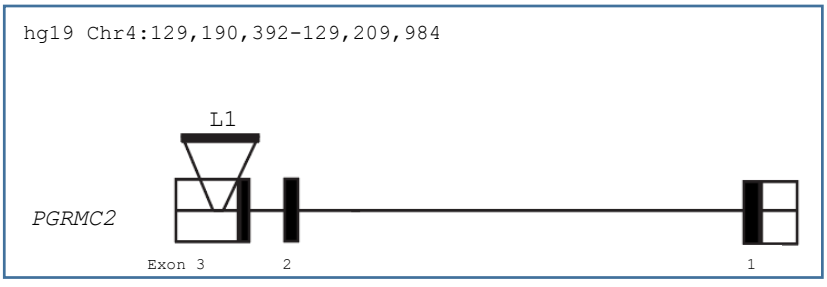
Empty site:

1st strand cleavage

5′-TTTGAAAACAGAAAAGCA...CAATATATTTTAAAAAGAAAGA-3′
3′-AAACTTTTGTCTTTTCGT...GTTATATAAAATTTTTCTTTCT-5′

Filled site:

5′-TTTGAAAACAGAAAAGCA...CAATATATTTTAT_{83} [UTR ORF2 | ORF1] AACAGAAAAGCA...CAATATATTTTAAAAAGAAAGA-3′
3′-AAACTTTTGTCTTTTCGT...GTTATATAAAATA_{83} TTGTCTTTTCGT...GTTATATAAAATTTTTCTTTCT-5′



- strand
TSD: 628
Chr4:129,191,510-129,192,137

L1 #6
ChrXq13.1

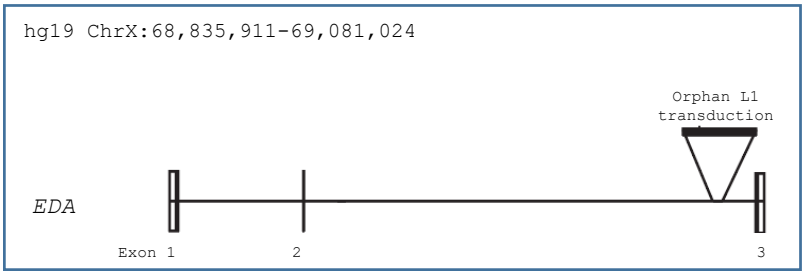
Empty site:

5′-GAAGACGAAGTTGAAACTTAATAACTTGCCTACGGGCACAGGCTGTTATA -3′
3′-CTTCTGCTTCAACTTTGAATTATTGAACGGATGCCCGTGTCCGACAATAT-5′

1st strand cleavage

Filled site:

5′-GAAGACGAAGTTGAAACTTAATAACTT [497bp] A_{59}AATAACTTGCCTACGGGCACAGGCTGTTATA-3′
3′-CTTCTGCTTCAACTTTGAATTATTGAA [497bp] T_{59}TTATTGAACGGATGCCCGTGTCCGACAATAT-5′



+ strand
TSD: 8
ChrX:68,964,693-68,964,700
Orphan Transduction
Chr13:61,460,432-61,460,928

L1 #7
Chr6q15

Empty site:

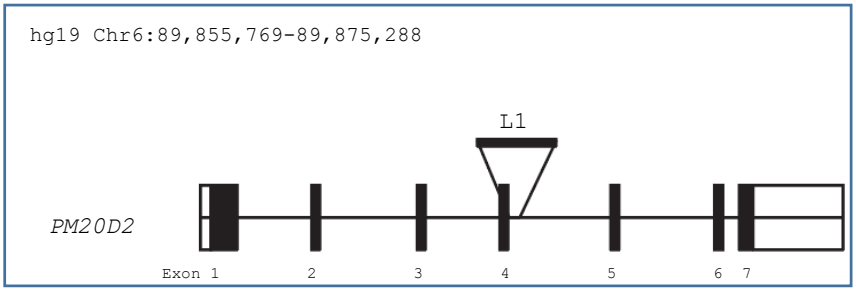
exonic

5′-**GCACA**GTAAGAACTTTTAAAAGTTAATCTAAGTTACAAT-3′
3′-**CGTGT**CATTCTTGAAAATTTTCAATTAGATTCAATGTTA-5′

1st strand cleavage

Filled site:

5′-**GCACA**GTAAGAACTTTTAAAAGTTAATCTA [UTR | ORF1 | ORF2/UTR] A_{XX}AAAAGTTAATCTAAGTTACAAT-3′
3′-**CGTGT**CATTCTTGAAAATTTTCAATTAGAT [UTR | ORF1 | ORF2/UTR] T_{XX}TTTTCAATTAGATTCAATGTTA-5′



+ strand
TSD: 13
Chr6:89,864,634-89,864,646

L1 #8
ChrYq11.21

Empty site:

1st strand cleavage

5′-ATGTTTATTACCCACTTGTGTTAGTTTCTTAAGGCTGCTGTGACAAATCA-3′
3′-TACAAATAATGGGTGAACACAATCAAAGAATTCCGACGACACTGTTTAGT-5′

Filled site:

5′-ATGTTTATTACCCACTTGTGTTAGTTTCTTTT$_{??}$ ORF2/UTR TGTTAGTTTCTTAAGGCTGCTGTGACAAATCA-3′
3′-TACAAATAATGGGTGAACACAATCAAAGAAAT$_{??}$ ACAATCAAAGAATTCCGACGACACTGTTTAGT-5′

Intergenic

- strand
TSD: 12
ChrY:13,871,221-13,871,232

SVA #1
Chr6p22.1

Empty site:

1st strand cleavage

5′-GCAATCTGATTTGGCAAATATTTTTAAAGATGATATTTGAAT-3′
3′-CGTTAGACTAAACCGTTTATAAAAATTTCTACTATAAACTTA-5′

Filled site:

19bp upstream

5′-GCAATCTGATTTGGCAAATATTTTTT$_{74}$ SINE_R VNTR *Alu* GGCAAATATTTTTAAAGATGATATTTGAAT-3′
3′-CGTTAGACTAAACCGTTTATAAAAA$_{74}$ GCGTTTATAAAAATTTCTACTATAAACTTA-5′

SVA_D

spliced out

725bp

Chr17 + strand: GT AG

*Alu*Sq | VNTR | SINE_R

Intergenic

- strand
TSD: 13
Chr6:29,684,013-29,684,026
Source SVA:
Chr17: 42,314,401-42,316,970

SVA #2
Chr4q31.1

Empty site:

5′-GCTATGCTTAAGAAAACAAAGCTGCATTTGGGAGGCTG-3′
3′-CGATACGAATTCTTTTGTTTCGACGTAAACCCTCCGAC-5′

1st strand cleavage

Filled site:

hexamer repeats

5′-GCTATGCTTAAGAAAACAAAGCTGCAT | *AluSq* | VNTR | SINE_R | A_{45}AAGAAAACAAAGCTGCATTTGGGAGGCTG-3′
3′-CGATACGAATTCTTTTGTTTCGACGTA | | | | T_{45}TTCTTTTGTTTCGACGTAAACCCTCCGAC-5′

SVA_E

+ strand
TSD: 18
Chr4:140408752-140408769

Intergenic

SVA #3
Chr11p15.4

Empty site:

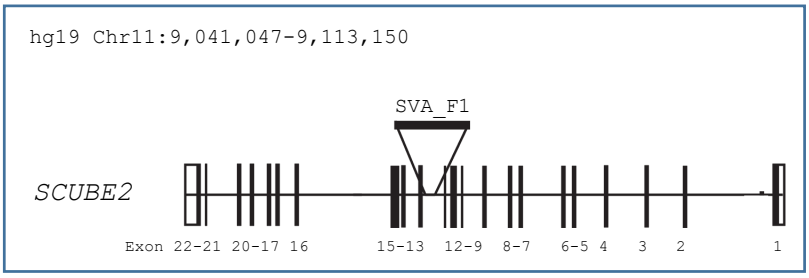
5′-CCTATCTATCTATCTAAAACTGAGGGTCAAAAATAGCCTCCA-3′
3′-GGATAGATAGATAGATTTTGACTCCCAGTTTTTATCGGAGGT-5′

1st strand cleavage

Filled site:

5′-CCTATCTATCTATCTAAAACTGAGGGTC | MAST2 | VNTR | SINE_R | A_{115}TAAAACTGAGGGTCAAAAATAGCCTCCA-3′
3′-GGATAGATAGATAGATTTTGACTCCCAG | | | | T_{115}ATTTTGACTCCCAGTTTTTATCGGAGGT-5′

SVA_F1



+ strand
TSD: 14
Chr11:9,072,822-9,072,835

SVA #4
Chr8p12

Empty site:

5′-GAAGATAAACTTAGAAAATTTTGTGTCACAAGGTATAT-3′
3′-CTTCTATTTGAATCTTTTAAAACACAGTGTTCCATATA-5′

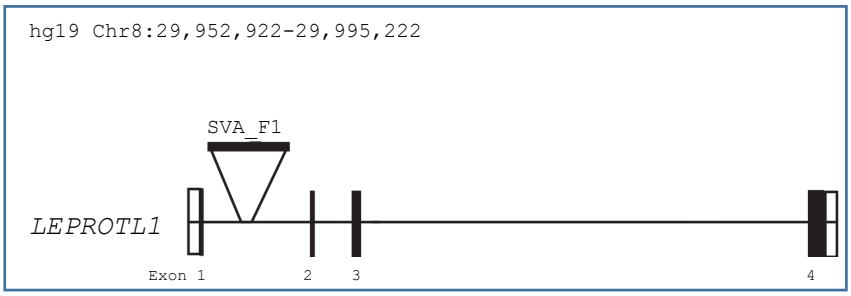
1st strand cleavage

Filled site:

479bp

5′-GAAGATAAACTTAGAAAATTTTGTGTCA MAST2 VNTR SINE_R AluSp A_{62}AGAAAATTTTGTGTCACAAGGTATAT-3′
3′-CTTCTATTTGAATCTTTTAAAACACAGT T_{62}TCTTTTAAAACACAGTGTTCCATATA-5′

SVA_F1



+ strand
TSD: 16
Chr8:29,955,718-29,955,733

SVA #5
Chr3p13

Empty site:

1st strand cleavage

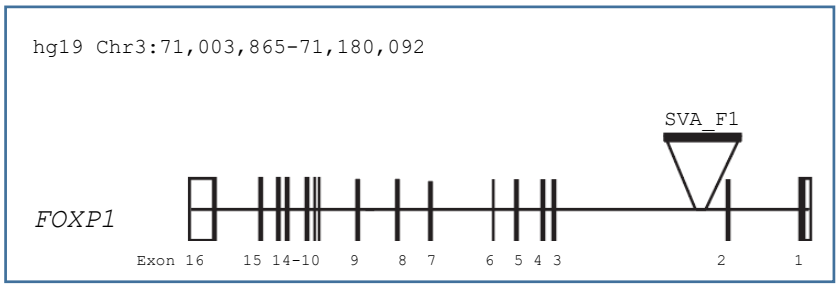
5′-TGGAAAAAGAAGAAAATAGATTATGTTTCAATAAACCCATAAGGA-3′
3′-ACCTTTTTCTTCTTTTATCTAATACAAAGTTATTTGGGTATTCCT-5′

Filled site:

22bp deletion

5′-GAAAAAGAAGAAAATAGATTATGTTTCT$_{40}$ SINE_R VNTR MAST2 AluSc GATTATGTTTCAATAAACCCATAAGGA-3′
3′-CTTTTTCTTCTTTTATCTAATACAAAGA$_{40}$ CTAATACAAAGTTATTTGGGTATTCCT-5′

SVA_F1 122bp



- strand
TSD: 11
Chr3:71,599,786-71,599,796
Source element:
Chr3:48,251,893-48,254,907

SVA #6
Chr12p13.2

Empty site:

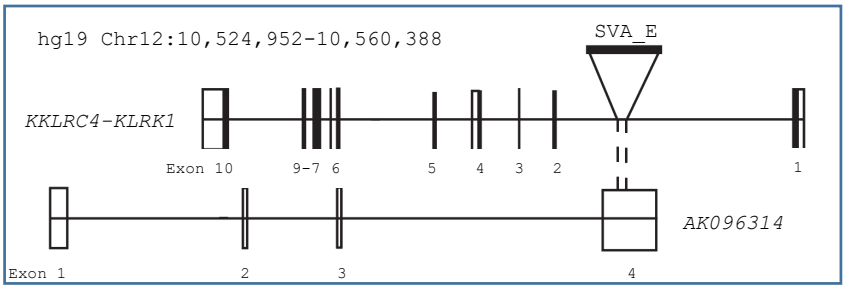
5′-GAGCCTCTGTCTCAAAGAAAGAAAGAGTATGTACCTGGATATGTATTTTT -3′
3′-CTCGGAGACAGAGTTTCTTTCTTTCTCATACATGGACCTATACATAAAAA-5′

1st strand cleavage

Filled site:

5′-GAGCCTCTGTCTCAAAGAAAGAAAGAGTATG [VNTR|SINE_R] A_{126}AAAGAAAGAGTATGTACCTGGATATGTATTTTT-3′
3′-CTCGGAGACAGAGTTTCTTTCTTTCTCATAC [VNTR|SINE_R] T_{126}TTTCTTTCTCATACATGGACCTATACATAAAAA-5′

SVA_E



+ strand
TSD: 14
Chr12:10,549,100-10,549,113

SVA #7
Chr1q42.2

Empty site:

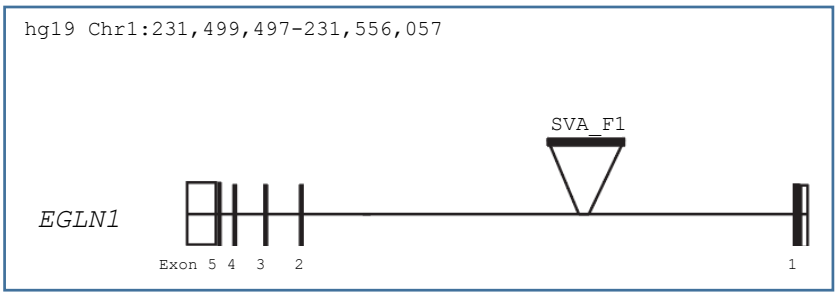
5′-CATATGTTATGATTTACATGAAAAAACTAACCTGCTCAAAAGATT-3′
3′-GTATACAATACTAAATGTACTTTTTTGATTGGACGAGTTTTCTAA-5′

1st strand cleavage

Filled site:

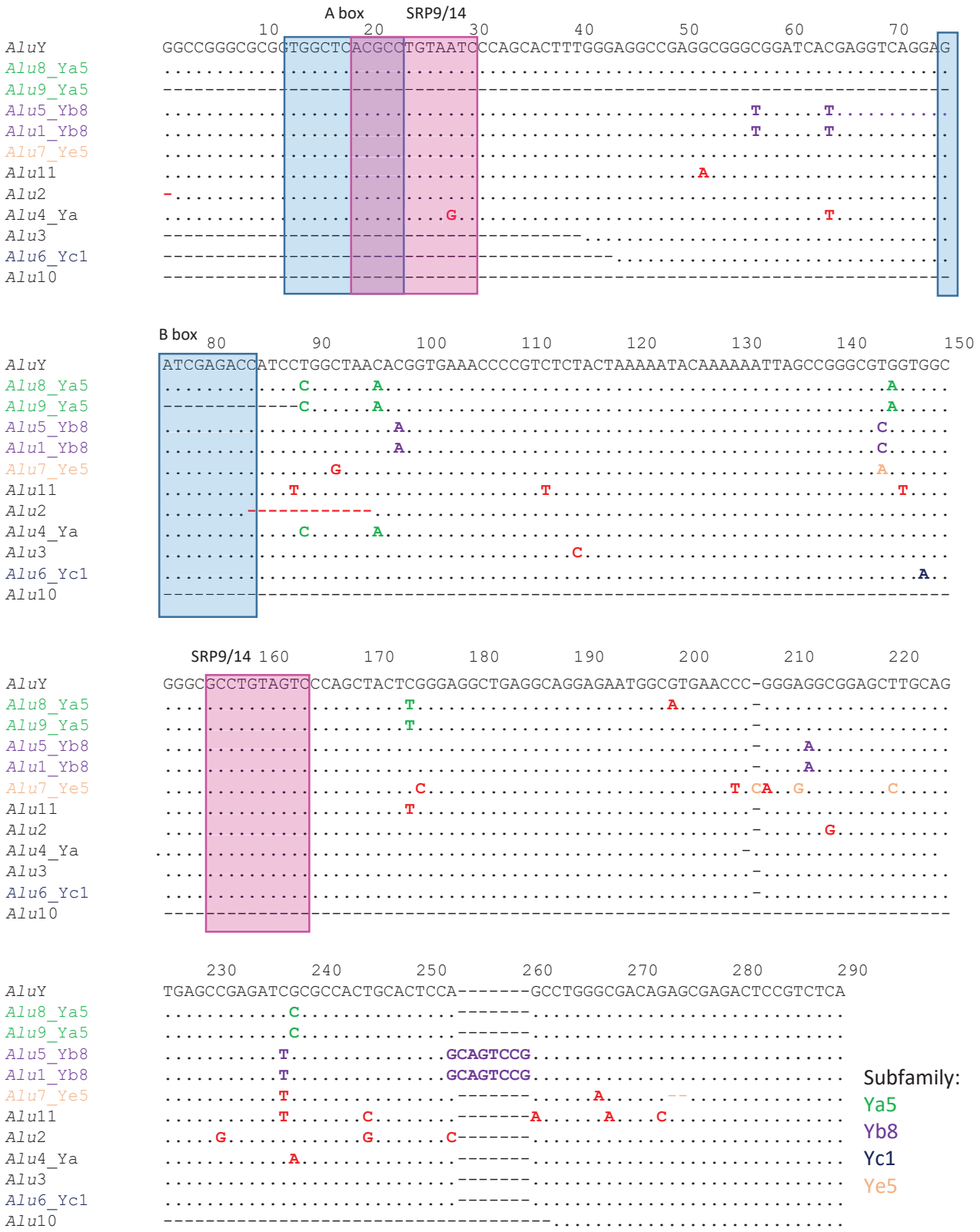
5′-CATATGTTATGATTTACATGAAAAAACTA [MAST2|VNTR|SINE_R] A_{58}GAAAAAACTAACCTGCTCAAAAGATT-3′
3′-GTATACAATACTAAATGTACTTTTTTGAT [MAST2|VNTR|SINE_R] T_{58}CTTTTTTGATTGGACGAGTTTTCTAA-5′

SVA_F1



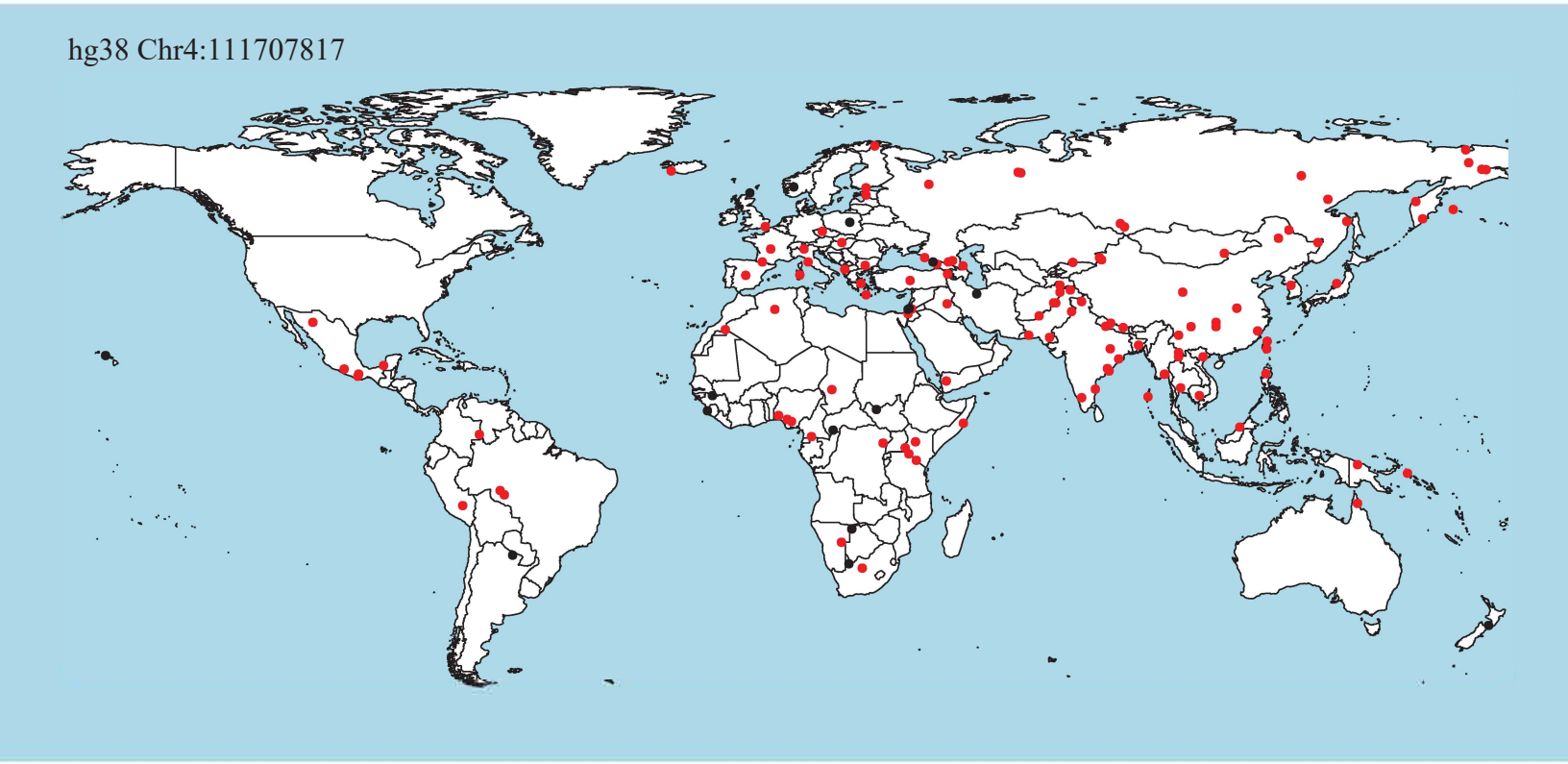
+ strand
TSD: 10
Chr1:231,536,301-231,536,310

Supplemental Figure S2. Detailed insertion information for the 26 *de novo* MEI. The figures are listed from *Alu* #1 – SVA #7. Empty site refers to the non-MEI chromosome, and filled site refers to the MEI chromosome. The TSD is labeled in red. All chromosome regions are in hg19. The MEI are depicted in relation to a gene where applicable (not drawn to scale). For L1 #1, the microhomology site is labeled in green, and the deleted region is in purple. For SVA #1, there is a schematic that shows that SVA #1 likely has internal splicing from its source element on Chr17.

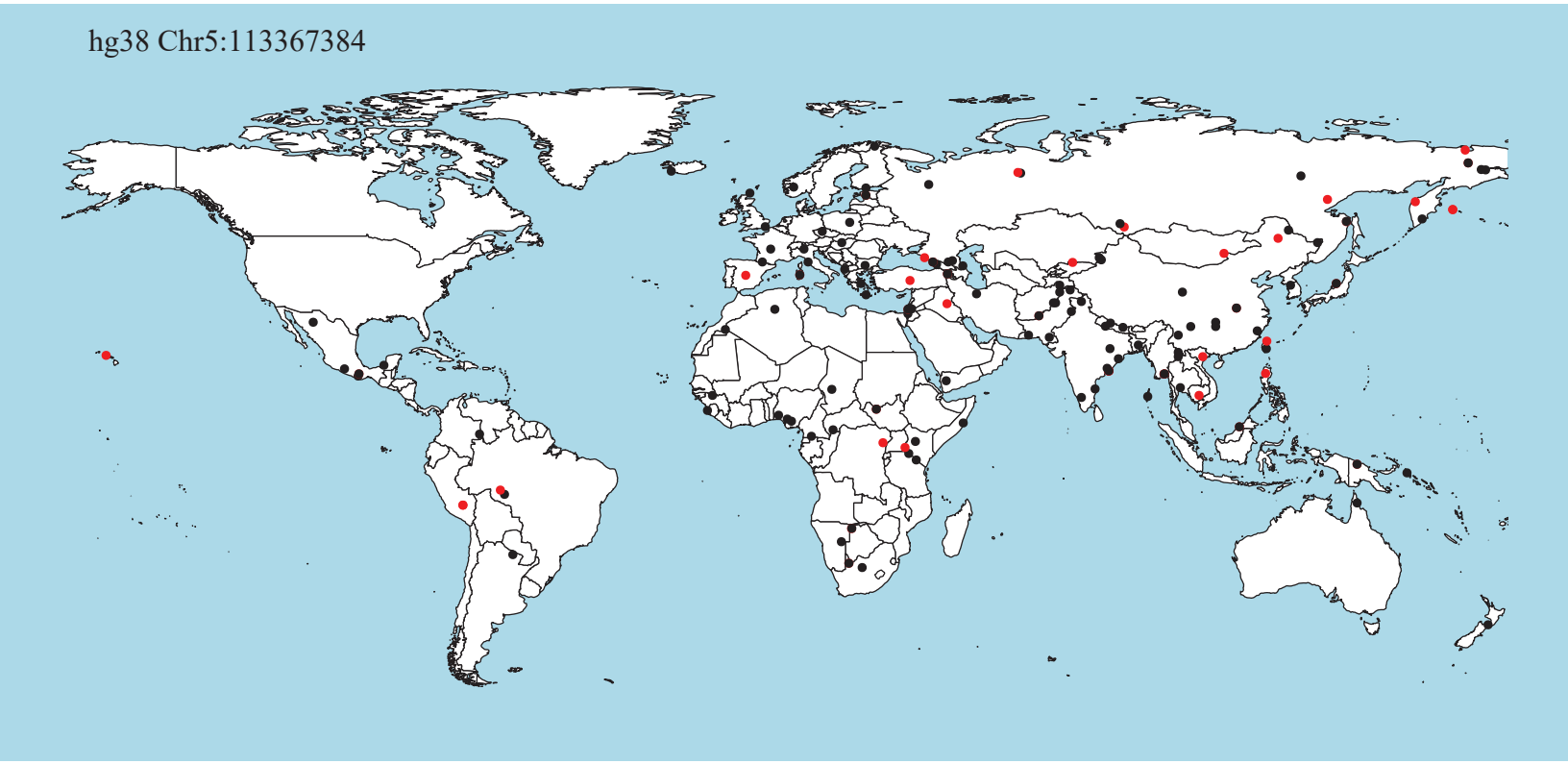


Supplemental Figure S3. Alignment of 11 *de novo Alu* elements to the *Alu*Y consensus.

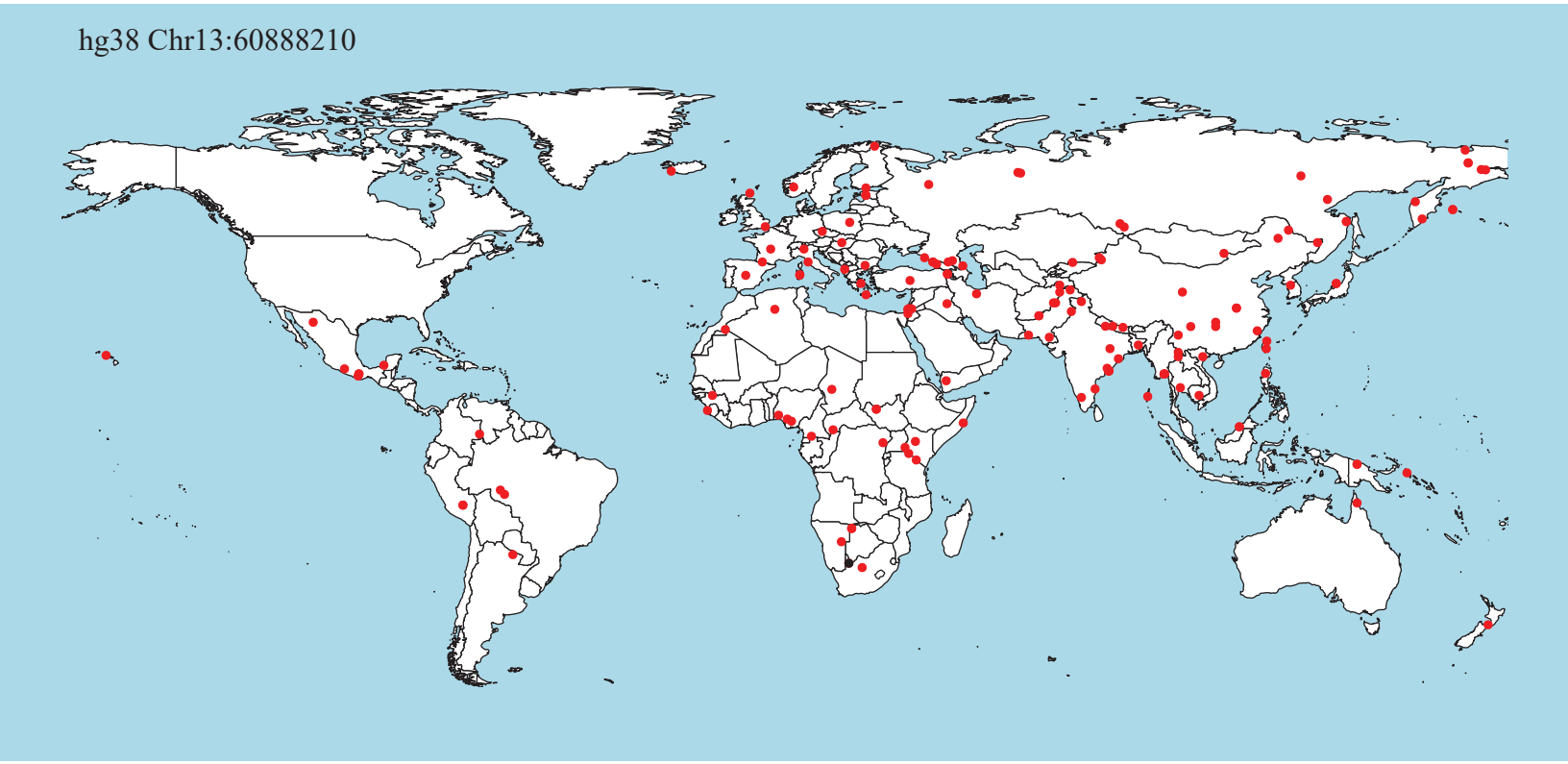
A



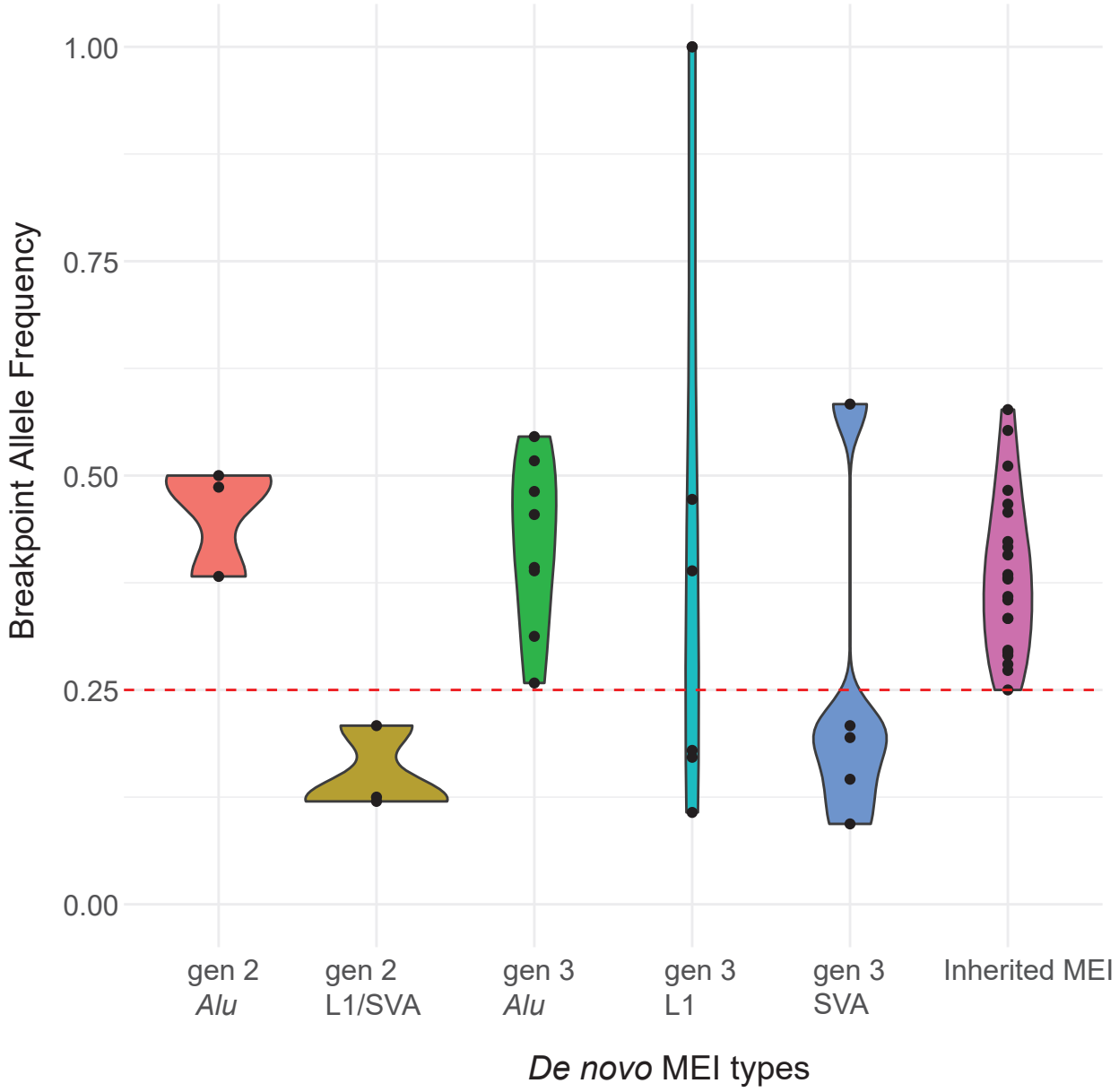
B



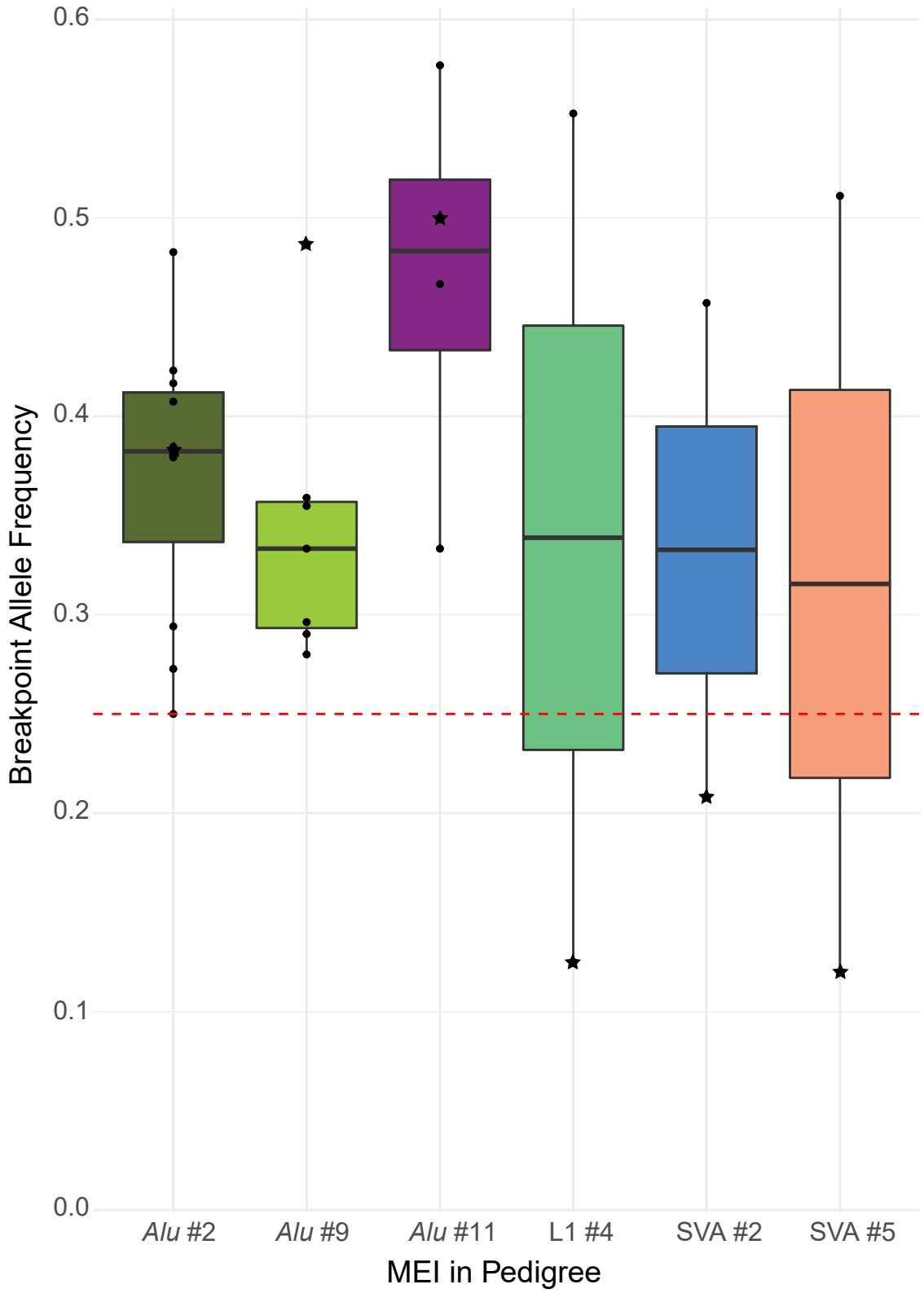
C



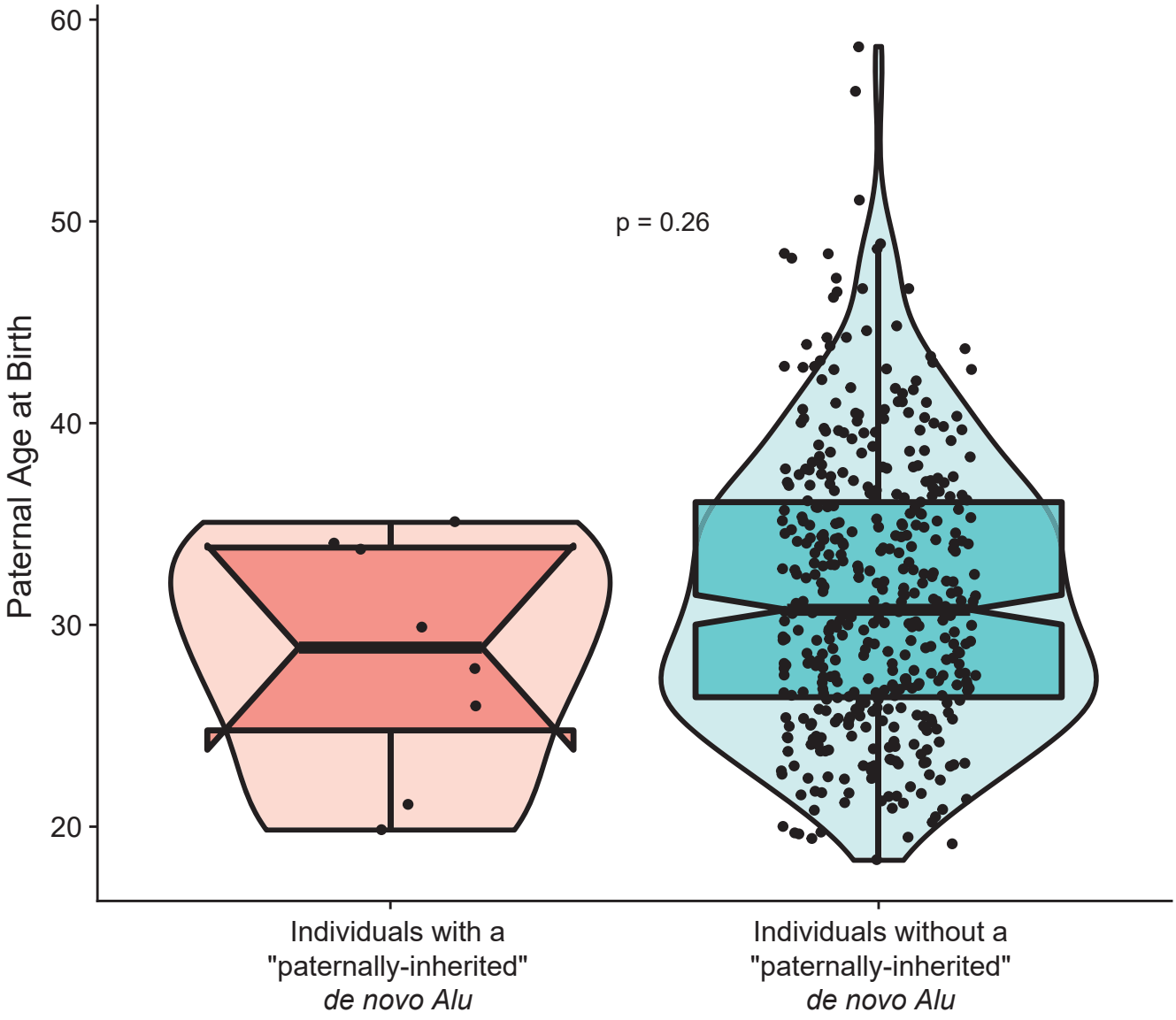
Supplemental Figure S4. Presence of Source L1 elements within SGDP populations. A red circle indicates that the L1 element was present in that population. A black circle indicates absence of the L1 element.



Supplemental Figure S5. Violin plot of BAFs in individuals with the d*e novo* MEI. Each dot represents the BAF for the *de novo* MEI in an individual. The red line indicates the lowest BAF (0.25) for an inherited *de novo* MEI in the 3rd generation.

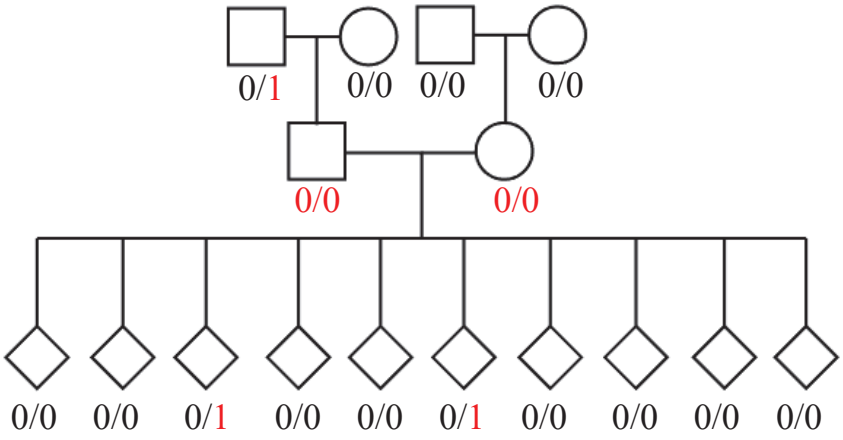


Supplemental Figure S6. Boxplot of BAFs in Pedigrees. Stars indicate the mother (generation-2) with the *de novo* MEI and dots indicate the children (generation-3) with the inherited *de novo* MEI. The red line indicates the lowest BAF (0.25) for an inherited (heterozygous) MEI in the children.

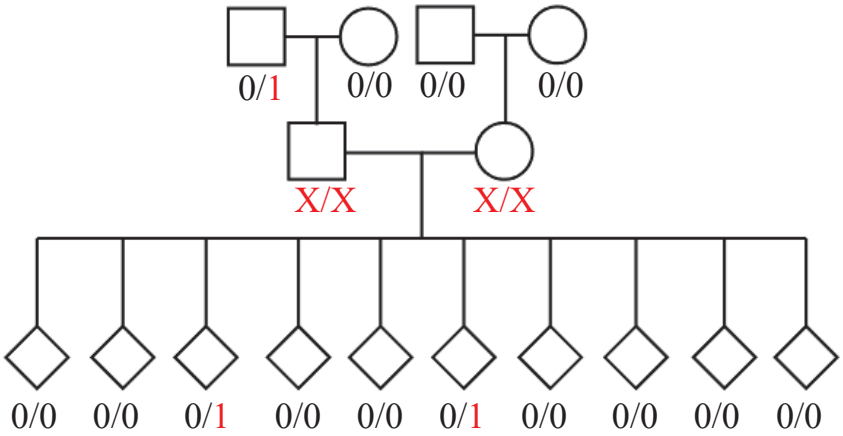


Supplemental Figure S7. Paternal age at birth comparison in the CEPH dataset. The eight CEPH individuals with a "germline" *Alu* element on the paternal chromosome are in red. The other 429 individuals with both parents in the dataset are in blue.

A False negative inheritance of MEI locus in generation 2



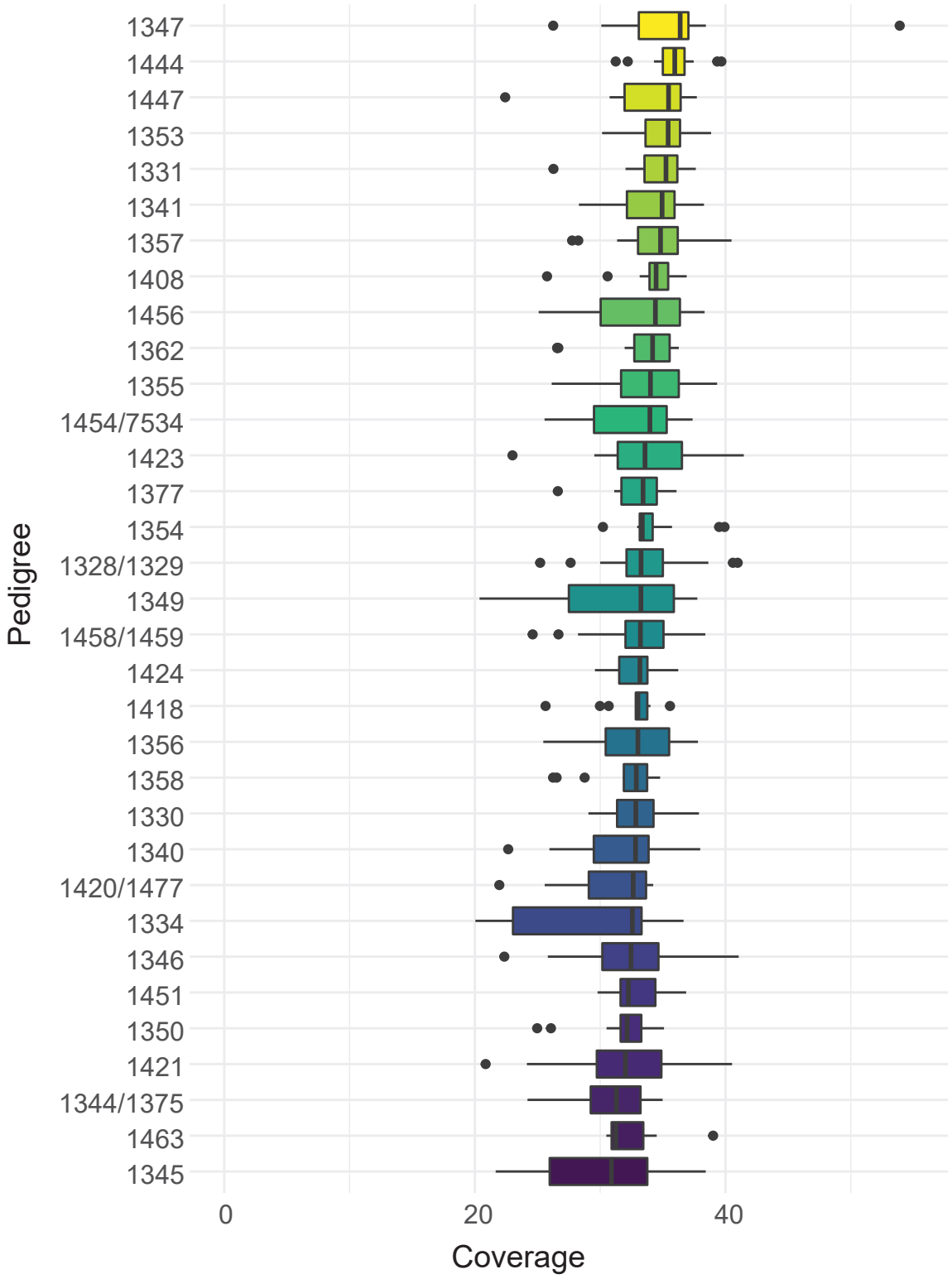
B MEI locus inherited regardless of generation 2 genotype



C Sensitivity rates in inherited MEI loci

	All loci			Filtered loci		
MEI	False Negative (#)	All inherited (#)	Sensitivity	False Negative (#)	All inherited (#)	Sensitivity
Alu	2,259	6,860	67.1%	221	3,741	94.1%
L1	284	1,022	72.2%	26	506	94.9%
SVA	128	480	73.3%	23	231	90.0%
Total	**2,671**	**8,362**	**68.1%**	**270**	**4,478**	**94.0%**

Supplemental Figure S8. False-negative inheritance rate of MELT in CEPH dataset. A) Example of a false-negative inheritance MEI insertion. The pedigree has the MEI present in at least one individual in generation 1 and at least two individuals in generation 3, but was not called in the 2nd generation. One locus has to fail in only one pedigree to count. B) Example of inherited MEI regardless of F1 genotype. C) Sensitivity rates in inherited MEI loci. Filtered loci are loci that pass MELT's filter (FILTER=PASS) Sensitivity is calculated by (1 – False Negative loci/Inherited loci).



Supplemental Figure S9. Coverage of the BAM files in each pedigree