Supplemental Material

Supplemental Results

Sensitivity and Specificity

Based on the validation results, specificity of our assay to detect non-reference insertions is > 80% when considering peaks with more than 20 reads that are at least 100bp in width. ‘Peaks’ refers to the sequence coverage of the L1 flanking sequence proximal to the 3’ UTR (see Methods). By this definition, a ‘peak’ consists of all reads mapping within the 500bp window 3’ of the predicted L1 insertion site. This specificity estimate does not account for the failure rate of PCR validation that, based on the frequency of a validation attempt failing to amplify both filled and empty sites for a given locus, appears to be 11.5%. Since these failures are considered false positives, the actual specificity is likely to be higher than 80%. Increasing stringency with respect to the peak characteristics increases the specificity (Figure S1).

The sensitivity of our technique was determined by considering biological replicates and samples from monozygotic twins. The average sensitivity of three pairs of replicates and three pairs of monozygotic twins with respect to validated non-reference L1Hs insertions is 93% +/- 4%, corresponding to a false negative rate of ~ 7%. As with specificity, sensitivity varies with peak characteristics, especially the width of the peaks, (Figure S2).

Supplemental Discussion

The ability to catalog RIPs on a large scale has important applications to the study of the genetic basis of human disease. We can divide disease-associated L1 insertions into two categories: *de novo* and inherited. Retrotransposition events that occur in the proband, but are absent from the genomes of both parents, are *de novo* insertions. The majority of disease-causing insertions falls into this category and are generally found by thorough investigation of loci suspected to be associated with the disease (for example Kazazian et al. 1988, reviewed in Belancio et al. 2008b). In these cases, disruption of gene products occurs if the L1 interrupts an exon or otherwise disrupts the normal regulation of transcription, for example, through the disruption of splice sites (Martinez-Garay et al. 2003).

Heritable single-base substitutions that have reached some threshold allele frequency are referred to as single-nucleotide polymorphisms (SNPs). Most SNPs that are associated with disease are not implicated directly, but rather through association studies in which many loci are tested for statistical association with the disease phenotype. We expect that this will also be the case for RIPs: many more human RIPs may be associated with disease than the ~65 known *de novo* insertions for L1, *Alu*, and SVA associated with disease, but it has not been possible to examine RIPs on the scale necessary for genome-wide association studies until now. Since a large number of samples must be used for GWA studies, it should be feasible to bring the cost of performing our assay down to where it is competitive with existing large-scale SNP genotyping assays through multiplexing. In addition to simply being associated with disease, it is possible that the effect sizes of RIP-based associations will be larger than those for SNP-based associations because of the nature of the markers. Furthermore, full-length L1 insertions are relatively large pieces of DNA (6 kb), which encode multiple promoters (Speek 2001), could potentially affect splicing (Belancio et al. 2006, Belancio et al. 2008a), and can mediate ectopic recombination (Song and Boissinot 2007). As a result, the association between a polymorphic L1 and a genetic disease might be causal rather than simply correlative.

Supplemental Methods

Acquisition of human genomic DNA

EBV-transformed lymphoblastoid cell lines (LBLs) corresponding to HapMap individuals (GM11993, GM12878, GM12891, GM12892, GM19238, GM19239, GM19240) were obtained either from the laboratory of Vivian Cheung at the Children's Hospital of Philadelphia or from the Coriell Institute for Medical Research. Cells were grown in suspension culture in RPMI 1640 media supplemented with 15% FBS (Hyclone), 1% Penicillin/Streptomycin, and 1% L-glutamine. Genomic DNA was obtained from cultured cells with a Qiagen DNA Mini kit using the manufacturer's protocol. Genomic DNAs corresponding to Japanese individuals were obtained by Hiroki Kano from the laboratory of Tatsushi Toda at Osaka University. Genomic DNAs corresponding to monozygotic twin pairs were obtained from Robert Yolken of the Johns Hopkins University School of Medicine. Additional genomic DNAs corresponding to trios (SB3Mo, SB3Fa, SB3Ch, SB4Mo, SB4Fa, SB4Ch) were obtained from the laboratory of Nancy Spinner of the Children's Hospital of Philadelphia.

Estimating the Rate of Retrotransposition

The parameter θ is widely used to assess the extent of genetic variation in a population. The Watterson estimator of θ is $\hat{\theta}_W = K / a_n$ (Watterson 1975), where K is the number of segregating sites among n unrelated individuals and $a_n = \sum_{i=1}^{n-1} \frac{1}{i}$. In our dataset of 15 unrelated individuals, there are 770 segregating sites, defined in this case as insertions present in less than all 15 individuals. Thus, $\hat{\theta}_W = 770/3.25 \approx 241$, which compares favorably to the empirically determined mean value of 285 (95% c.i. 148-422) based on the average number of differences in L1Hs content between all pair-wise combinations of 2 individuals. This estimate of pair-wise variation is equivalent to Tajima's parameter $\hat{k}$ (Tajima 1989). While both these estimates have wide variances (the variance of $\hat{\theta}_W$ is calculated to be ~12241 corresponding to a standard deviation of ~110), their means are comparable. Indeed, a Tajima's D of ~0.44 confirms the neutrality of the L1 markers used in this estimation. Tajima's D test was carried out as described in Tajima (1989) where $D = \frac{\pi - \hat{\theta}_W}{\sqrt{\hat{V}}}$. We can then calculate the per sequence per generation mutation rate (or in this case, retrotransposition rate) μ using the formula $\theta = 4N_e\mu$, assuming a value of 10,000 for the effective population size for humans (Harpending 1998). Using $\theta = 285$, $\mu \approx 1/140$ L1Hs retrotransposition events per meiosis. Substituting the upper and lower confidence bounds for θ, we obtain values for μ of 1/95 and 1/270 L1 retrotransposition events per meiosis, respectively.

Protocol for retrotransposon insertion site discovery via Illumina sequencing

1. Primers

```
L1HsTAILSP1A2
GGGAGATATACCTAATGCTAGATGACAC

Adap1L1HsG
CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTGCACATGTACCCTAAAACTTAG

Adap2Seq1
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

DEGSeq1N5TCTGT
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTCTGT
```

DEGSeq1N5CTTCT
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTTCT

DEGSeq1N5CTGCA
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTGCA

DEGSeq1N5TGCCT
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTGCCT

DEGSeq1N5TCTCA
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTCTCA

DEGSeq1N5CAGAG
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCAGAG

DEGSeq1N5TTGAA
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNTTGAA

DEGSeq1N5CTTTG
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNCTTTG

2. PCR Recipes

a. Round 1: Linear amplification of L1Hs flanks followed by hemi-specific PCR (incorporates Illumina sequencing primer)

Mix: (Using Promega GoTaq Flexi)

10 uL buffer 5x
6 uL MgCl$_2$ 25mM
2 uL L1SP1A2 Primer 20uM (or SVAstarter for detection of SVA insertions)
(4 uL degenerate primer 5uM), added when linear amplification is finished
0.5 uL DMSO 100%
0.5 uL FlexiTaq
1uL dNTPs 10mM
2uL gDNA (at 100ng/uL for ~200ng per reaction)
24uL DEPC ddH$_2$O

=50uL, make mix for 9 reactions and aliquiot into 8 tubes, adding 1 of each of the 8 degenerate primers after the linear reaction is complete.

Cycling Program:

	Temperature (˚C)	Time	Notes
1	95	2:30	
2	95	0:30	
3	58	1:00	
4	72	2:00	

			Goto 2 5x
6	60	Pause	Add Degenerate Primers
7	95	0:30	
8	55	0:30	
9	72	1:30	
10			Goto 7 14x
11	72	10:00	
12	4	Hold	

- Purify all 8 reactions on Qiagen PCR Cleanup columns following Qiagen protocol, eluting in 50 uL Buffer EB. Proceed to Round 2.

b. Round 2: Amplification of library, Incorporation of Illumina sequencing adapters.

Mix (using Promega GoTaq Green Mastermix)

12.5 uL GoTaq MMix
1.5 uL Primer Adap1L1HsG (or Seq1SVA if detecting SVA insertions)
1,5 uL Primer Adap2Seq1
2.5 uL Round 1 product (1 degenerate primer per rxn)
7 uL DEPC ddH$_2$O

=25uL, make mix for 9 reactions, aliquot into 8 tubes corresponding to the 8 reactions from round 1.

Cycling Program:

	Temperature (˚C)	Time	Notes
1	95	2:00	
2	95	0:30	
3	62	0:30	
4	72	1:00	
			Goto 2 19x
5	72	5:00	
6	4	Hold	

Run products on a 1% TAE gel. Excise the constellation of bands and smears between 200 and 500 bp with a sterile scalpel and purify DNA using the Qiagen Gel Purification protocol. Elute library in 50 uL Buffer EB. Measure DNA concentrations on UV-Spectrophotometer and try to mix in equimolar ratios. After mixing, purify products again on a Qiagen column, elute in 55uL DEPC ddH$_2$O and proceed to end-polishing step.

c. End-polishing

- Taq leaves adenine overhangs that could cause problems when the library is annealed to the Illumina GA flow cell. It was necessary to use Taq because proof-reading polymerases tend to remove the 3' most base(s) of the primers due to the 3' → 5' exonuclease activity.

Mix:

6 uL 10x Pfu Buffer
2.5 uL Pfu polymerase
2.5 uL dNTPs
49 uL library

=60 uL, incubate at 72˚C for 1hr

Purify reaction on a Qiagen MinElute column and elute in 10uL EB. Minelute columns will provide the necessary concentration for accurate dilution for the Illumina GA.

References for Supplemental Material

Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C. 2009. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* **19:** 1622-1629.

Belancio VP, Hedges DJ, and Deininger P. 2006. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* **34**: 1512-1521.

Belancio VP, Roy-Engel AM, Deininger P. 2008a. The impact of multiple splice sites in human L1 elements. *Gene* **411:** 38-45.

Belancio VP, Hedges DJ, Deininger P. 2008b. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res* **18:** 343-358.

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. 1998. Genetic traces of ancient demography. *Proc Natl Acad Sci* **95:** 1961-1967.

Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332:** 164-166.

Martinez-Garay I, Ballesta MJ, Oltra S, Orellana C, Palomeque A, Molto MD, Prieto F, Martinez F. 2003. Intronic L1 insertion and F268S, novel mutations in RPS6KA3 (RSK2) causing Coffin-Lowry syndrome. *Clin Genet* **64:** 491-496.

Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, et al. 2010. Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463:** 943-947.

Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* **21:** 1973-1985.

Song M, Boissinot S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **390:** 206-213.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7:** 256-276.

Supplemental Tables

Table S1: Cataloged non-reference insertion sites. FS and ES primers correspond to primer positions in Figure 2b. The ‘Indiv.’ column indicates the number of unrelated individuals a given insertion was detected out of a possible 13. If an insertion falls within an intron the gene symbol is noted in the ‘Gene’ column. Rows marked by an asterisk (*) indicate insertions identified through analysis of mate-paired resequencing data from three personal genomes (Ahn et al. 2009, Schuster et al. 2010).

Chr	Peak Center	Strand	Indiv.	FS Primer	ES Primer	Gene
*1	12798245	-	3	TGAAGTTGCCTATCAGGAGATT	CCATGACTAAACCAGGAAGAAG	
1	24648099	+	3	ACAGGCCCAAGATACAGTACAC	TTGCCCATAAAGTGATTTTAGC	
*1	42256618	+	6	GGGAAGGTAAATATATGGGAGG	TTGTCAACATTTGGTATTGTCC	
*1	74965343	-	9	TTTCTCCTCAGAAAATGGGTT	GGAGTGTAGTGGCATCTCTCA	CRYZ
1	78795733	+	5	TATTTGGGCTCTTTTTGGTTC	AAACTACAGGCCAATACCACC	
1	83381713	-	6	AGACTGGCCTCTTCTCTTTAGG	AACAAACAGTCAGATGGCTTTC	
1	102317644	-	5	TTCAGGCTTTTTGTTTGTTTG	CTCTGCATTCTTGCCAATATC	
1	102341702	+	5	ATCCTTCTTGCTTCTAACCTCC	TGCTATTTGATTTTCACCTGAG	
*1	113690572	+	13	CAAATGCACATGTTGATTCTTC	CTGACAAAGGGCTAATATCCAG	
1	114778405	-	6	GCTGGCCACTCATACCTTATC	CAGTATGGTATCATTTTGTATTTGTG	TRIM33
*1	118631050	+	15	CTGTCACAACAAAAGGAACAGA	GACTGGATTAAGAAAATGTGGC	
*1	119355099	+	7	AATTCCGGGGTTTTAAAAGAT	TTGTTCAGACTCCTCCTTTGA	
*1	147575280	+	15	TGAATGAATTTTTGGAGGACA	GCCATTCCTCTAAATAGCTGC	
1	150710322	-	2	TGAACACCTCAGGCTAATCTTT	CTCTCTGAAACATGTGCTGTGT	
*1	157841030	+	14	CATGTAGTTCTCGAGCATTGG	GTCTGAGATCAAACTGCAAGG	
1	163798330	-	1	TTTTCACTAGTCAGGCAGTTCC	GAGACGAGATTCTCCTGAAATG	
1	177841820	-	7	TACTGAATAGTTGCACTCCCCT	CAAAGGCACAATACATGAAAGA	TDRD5
1	190528785	-	5	TATCATAGAAAATTGGGCATCC	CCCCATCTCTACTACAAATCCA	
*1	221408094	+	8	GATGATGGTGATGTACAGATGG	TTCAAATTACGTATCGTTTGGG	
1	240563219	+	9	AGGATATGCAGTAGATTGGTGG	CTTGACGTTCACTCGTAGATACC	FLJ40773
*2	13100917	+	5	ATAAGTCATGGGAGGGTGTATG	TTGGATCAAGCAGACATGATAG	
2	22560190	+	2	CTTATCTACATGCACCTCCCAT	TGTGCAAAAATGTGTTTGATTT	
*2	23044293	-	12	TGGTTAAGCATTTTATCTATGTGG	TATTGTGGTAGGAACCTGCTG	
*2	23813798	+	8	TTTAAGGACTTCTCTGCATTGG	GCATTGCTGAGTTAGTTGTTTG	
2	35732947	+	9	TTCAGGATTTTGATCTTTCAGG	TGTAAGTATTCCATGGTCTGTGTC	

*2	41905115	+	12	TGTGAACACAGCTCTTCATTTC	ATTTGCAAACCATAATCAAACC	
2	53313665	+	10	TATTGGACCACAGCCATATGTA	TTAGCACCCTAAGATTTCTCCA	
*2	56637626	-	7	TCTGAAGCCTTCTTCTCTCAAC	CCCATGGAGTCTCACTGATTAC	
*2	66713391	-	7	TCTGTTGGAATTTGCTAGAGGT	ACACCATATTTGAGGCAGTAGC	
*2	69022148	-	1	ATTGATTTGCCTATGTTGAACC	TTATCTGCACATGGAACACATT	
*2	76431546	-	1	GTTGGAGTTTGCTAGAGGTCC	AACATGGCCAAATACAAACAG	
*2	79783143	+	2	GCATGGAGTGTTTTCTCTAAGG	GACTGCTGTCTGCTTTTGTATG	CTNNA2
*2	81951698	+	6	AAACTGTTCACTTGTTCTTGGG	TTTTGCCTTTTCTTCTGTTGTT	
*2	91004083	+	11	GCCTTACTAGATTGGGGAAGTC	CTCTGAGACCAAACTTCCAGAG	
2	91670522	+	8	CCAAATGTCCACTTCCAGATAC	GATAGCTTTGAGGATTTCGTTG	
2	91683159	-	5	GGAAATATCTTCCCCTGAAAAC	GTTTCCAATGAAATCCTCAAAG	
2	91688210	-	5	AATATCCACTTGCAGATTCCAC	GACCTCAGGAATACGGTGATAA	
*2	102862296	-	14	CCTTTGGTTTGAATTTCCTTC	TAATACTGCGCTTTTCTGACG	
2	105517882	-	12	GCAGTGAGCCAAAATCATAATA	AGGAAGGGGAATATCACACTCT	
2	105529652	+	3	AAGATAGCTAAAAGCTGGGACC	CAAAGCAAAGCAGAAAGGTTAG	
*2	112983902	-	13	TTGTTTACCTAATCAAGCCTGG	GGTGTATACCTGTAGTCCCAGC	TTL
*2	134385211	+	15	TAGTTCTCGAGCCTTGGTTTT	TTCTGAGTCAAAGAAAGGGGT	
*2	143001810	+	5	GGCAACAGAGAGAGACTCCAT	AAAGTCCACTCTGTGAAGCCT	
2	143087173	-	14	AGCATTCTTCACGTAGTTCTCG	AGTCTGAGATCAAACTGCAAGG	
2	143727317	+	4	GTACCAATTAGGCCAAAGAGTG	GAAGGATTTGCCAGAGAGC	ARHGAP15
*2	146737646	+	13	AACAACAACAACAGCAACCTT	TCCCTGCTGTGTAATTGTAGC	
2	168443545	-	2	GTTTCTCTGCTTCAGAGCTCAC	ATACAAAATAATTAGCTGGGCG	
*2	180709139	+	3	ACCAAATCTAAGTGGGAAGGAT	TCTGATCCGAAGGAATTAGAAA	
*2	183260037	+	13	TCCTTTTAACAGTCAGGACCC	CAGTCTACAGCTCCCAGTGTG	
*2	195645783	-	5	GAGACTTCTGCATTCTTCACG	GAATATTGCGCTTTTCAGACC	
*2	199003965	-	12	CAAAGGAGTAATAGGCAGATGG	TCACCAAAGTTGAAATGAAGG	
2	212415966	+	2	CACCCAATATGAAAGCTGACTT	TGAGATTTCAAATACGTTTCCA	ERBB4
*2	221211555	-	4	CATCCAGGTAGAGGTTTATGGA	AGTTTCAAGGGGAAAAAGAGAC	
2	230050212	+	5	CACAGAATTAACACAGGTGGG	TCAAAAACCAAACAAAATAGAAATC	DNER
2	235184583	+	8	CCTTGAAGAATCAGGAGTTACG	CCATACTGGTCTTGTGACAGTG	
*3	631199	-	15	GTCTGTTAGGTCTGCTTGGCT	GCAAACGCTTTGACTAAATGA	
3	3383404	+	1	ATTACAGGCATGTGCTATGTTG	AAACAGATGATTGAGGAACCAC	
3	21210010	+	1	CCAATCTCTGCCTCTGTAATTC	AGTTGCACTGGAAACAGAGACT	
3	22415117	+	5	AACCATCAGAAAAATGCAAATC	ATTGAGCAGCTATCACCACAG	
*3	35471977	+	14	CTAACTTGGTTCCATTCTCCC	GCTTAGGTAAACAAAGCAGCC	
3	38601133	+	10	TTCATCCATCAACCATCTGTC	CTTCAGCATTGCCTGTTTATC	SCN5A
3	44275495	-	1	ATCCTGTGGATATTCAACCAAG	AGCAAGGTAAAGAACAGTACGC	
3	58758987	+	2	CTGCTGCTTTATCAACTCAGTG	CTTTAAGTTGAAGCCAGTGGTC	FLJ42117
*3	59828056	+	12	TGGTCCTGAATTTTAGAGCAGT	TTAGGAAGGTTTTATGTCACCG	FHIT
3	65660615	+	4	GAGATGGAGTCTCGCTCTGTAG	AAACCAAACACCTCATATGCTC	
*3	70582150	-	12	TCAGCTCCTTTAAGCACTTCTC	TTACCTACCGGTCATTCAAGAG	
*3	75184134	+	5	CAAAACAGTTGCAAAAATAGCA	CTGGTAGGTAATTGGTCCTTTG	
3	75752701	-	5	TTTTTGTAGAGGCAGGGTCTC	ATGGGTCTGTCTCCCTTTATG	
3	75773828	-	12	GCAAGGAGAGTGAGTACAGAGG	AAAATGAACAAAGGTCAGCTTG	
3	79299110	+	4	CTTAATTACCAAGGCAGAATGC	CGTTGATCCGCTATTTTTCTTA	ROBO1
3	82073421	-	4	CTTTTATGCCTGTCGTAGTTCC	CATACAAAAACAACATGCAAATG	
*3	84675703	-	9	TTAGATCCCACTTGCCAATTAC	GGATGATAGAATCTGTACCCCA	
3	85659321	+	9	TTGACAGAAAAGAAGATAGACTGAAC	CATATCCTTCACAGACACTTGC	
3	90539210	-	3	ATGTGAAGATATTCCCGTTTCC	GATGTTTGCATTCAACTCACAG	

*3	90567924	-	14	GCATGGAGTAATTTGATCGTCT	GTCAAAGAAAGGCATGACAGAT	
*3	96864297	-	13	GTTTTCAATTTCTTTGCCTTTG	AATACTGTGCTTTTCTGATGGG	
*3	122180500	-	5	AACTTCTTTGCCTTTGGTTTG	ATTATATCCCACACCTGGCTC	
3	124516484	+	1	ATTCTCTAGCCACCTCCTCTCT	GATTCTGGAACCGTAAGGGTAT	ADCY5
3	125073586	+	10	AATTATCAAACCCAAGAAGGG	ATTTGCTGTTGAAAATCATGC	
*3	128003373	+	12	TGAAGCCTTCTTCTCTCAACTC	GATACTGCACTTTTCTGATGGG	CHCHD6
3	132552671	+	2	AATACCCCACATTTGATAGCTG	CAGAATTTTCAACCATGGAACT	
*3	137426639	+	11	TGGTTTGAATTTCCTCCTGTAG	AGTCTGAGATCAAACTGCAAGG	
*3	140527178	+	4	ACACAATTCAGTCCAGAGCAG	CTGGGACCAATTACACTGATG	
3	144604357	-	3	AATGTGGCATTTCTCAAACTCT	GTGGCATGAGACAGTAACTCAA	SLC9A9
*3	146092566	-	2	GAGAGATTCCCAGAGTGTGTGT	TGTATGTTGTGAGCACTGTGAA	
3	151693251	-	11	CATCAAGCAGGTTGTAACTGG	AATGCTAGATGACACGTCAGTG	
3	162628541	+	7	TCCACTACACACACACACACAC	CTAACACTGGATCTCTGGCTTC	
*3	169212206	+	11	TCATGTAGTTCTCGAGCCTTG	GTCACTCCCACCCAAATACTG	GOLPH4
3	170753054	+	7	AGGAATGATTTCTTGAGTCTAAGG	TTGAGAAACCTCTTGGTGAAG	MDS1
*3	177160510	-	9	TGCTACGGGGTTTATTCTAGC	CCTTCTACTGGAAAAAGGAGC	
*3	177574260	+	13	ATGTGTTTTGGTATTGGTTTCC	TCAAAACCACAATGAGGTATCA	
*3	182676002	+	4	TCATCCAGGAGAACTTTCCTAA	TTGTTGAATTGAACCCTTTACC	
*3	187854979	+	11	AGATGAGCCAGAGAAGAGGAG	CTGAGATAATGCCACTGTTCG	
*3	189317736	-	15	TGAGTGGCTGGAACTACAGG	ACTCCAGTCTGGGAGACTTTG	
3	194836724	-	4	GCCAGTGAATGATTAGACCACT	AACCAGGAAGACTGTTAAACCA	
*4	13823193	-	7	TGAAGCCTTCTTCTCTCAACTC	CAAATACTGTGCTTTTCCAATG	
*4	14458442	+	2	ACTGACCTTCACAAAAAGGAAA	GAAAGTTGGAGGTGTGGTGTAT	
*4	17771395	-	13	GCCATAAAGGCTAGATAGGGAT	GTTATGAATAAAAGGCATGGGA	
4	28931312	+	5	CAAGCTGAATGTGAAATGAAAA	CTTGATGGTTTCCATGATTTCT	
*4	46702487	-	15	GGTGAAACCCCGTCTCTACTA	TGTCGGAAATGATCTAATGGA	
*4	57191579	-	7	GATGCTGTCTTCTCATTGTATCC	TGAATCAACCTAATCCAAAAGG	
4	57821248	-	8	GTCTGCTCAAGGATTCAATTTC	GAGACTTTAACAGCCCACTGAC	
4	58612005	-	6	TTTTTCTGCATCTAGTCCGAAT	CGGATTTTTAGTAGAGAGGGGT	
4	71127982	+	5	GTTGAGTCCCATTTTGAAAGAA	CTGAACTTGACCGGAATTTTAC	C4orf7
4	80987674	+	8	CTCCTGACAACACTGCAAAATA	GAGGTCAGAAGTTCAAGACCAG	
*4	103562913	+	6	AAGACTTTCTAGCCAGGTGACA	CAAATGTTCTGGATTCCATCTT	
4	104434223	+	5	TCAGAATCAGAGGGACAGATG	GCGCAGAAAACTGTACAAAGA	
4	115352276	+	9	GCAAATTTTTAGAATCAGGTAAAAAC	GCCAGAATTATAAAAGGAAATACTG	
*4	123715205	-	5	TCGAACTTCCTCCTTTAGCTC	TTCTGACGGTCTTAGCAAATG	
4	129184647	-	2	CTCCTGAGTAGCTGGGATTACA	TAGAAACGGGTTTTTACCATGT	
4	132400781	-	5	AAGAGATGGTTGTGGTGATGTT	TCCCATTTTCTATTGGCATAAC	
*4	134816193	+	7	CCTAAAAGAAAGCTGCTCAAAA	AAGATTATCATAGTGAAAGTTGGACA	
*4	137986550	-	3	TACCTCTGGTAGAATTCGGCT	CAAGCAGACCTAATCGATGTC	
*4	138397092	+	8	TCCTCTCTCATCTGTGGGTAGT	AATATGTCACAGTGCTGCAGAG	
*4	147444897	+	7	TTGGAAGTAAAGCTCTCCTCAG	TTTACAGGCATGCATATTTCAG	
*4	150497831	+	4	TAAACATTGGTGCAAAAATCC	AAAGTCATTGGTAGCTTGATGG	
4	151802069	+	2	CACCTGAGGTCAGGAGTTCAT	TGCTGTATACAAACATCCCAGA	LRBA
4	152952051	-	3	CTGCTGTGAAAATAGCTCACTG	GCATGAGTCCAAAGCTAAAAAT	
*4	170518322	-	13	TTTACACACCACAATTACAGGG	AAATTCTGGAGTTGAGAAATGC	
*4	189020300	-	8	GGATGTCCTTTCTGCTTGTTAG	TGAGGTACCAGGTTCATCTCAC	
4	190896936	+	8	GATGAGAGTGCAGAGTGCTGT	AGCACATGCCATGAGAGTTAG	
*5	8802690	+	10	GGAAGATTTGTCTCTTCTTCCA	ATATGTAGAATGGCACAGGCTT	
5	9108982	+	2	TAAGGGCTCCAATTTCTTCAC	AAGAGGATGCCAAAATCATTC	SEMA5A

*5	11444312	+	7	GACAAAATCAAACTTTTCTCGC	TTTCCTGGGATTACAATCAGAC	CTNND2
5	20494373	+	5	AAGAACTATGTGGAATATTCAGGTG	ATCTTGAAATTGGTGTGGTCA	
5	23106036	+	1	TGAAAACTGGCAATTTGTAATTC	GTTGCTGTCTATGATTTTTAGGG	
*5	26129040	+	9	GCAAGACTCTGTCTCAACAACA	AGACTGTTGGAAAATCTGATGC	
5	33833098	-	4	TAGGGTCTATGAAGAGCAGGAA	TCCATTAATCCTTACGACAACC	ADAMTS12
*5	40076824	-	2	AGTTCATGGATTTCCATAGGTG	GATGCTGAATTAAGGCTTTCTG	
5	41637878	+	3	TCCCGAGTAGCTAGGACTACAG	AAAAGATAATGAATCTCAGCGG	
*5	44050173	+	13	CTTTGATGATGGTGACGTACAG	AGGTACCAGGTTCATCTCACTG	
*5	45481408	+	9	TGCTCTGCTTTTTAGAGTTTCC	TAGTCAAAGAAAGGGGTGACAG	HCN1
*5	49505921	-	6	AGTTTCCAGTTTTTCTGCTCTG	CACACTTGGAAATTCATGTCAC	
*5	57315355	+	3	GGTTTATCAGGATGTAACCCC	ACCAGCGTTGATCTCTCTTTC	
5	57463865	+	1	CCAACTACTCAGGAGGTTGAAG	AATGAAAGGGCTGAAGTGTCTA	
5	59626798	+	1	CATCTAGTTGGTAGAAGCCAGG	AGAGTCAGGTAGAATTGCTGTTG	
*5	62913890	+	11	GGCTGTAGAACAGTGGTTTTTC	AAAACAATTTGTGTGAAGGGAG	
5	67757507	+	5	TATAGCTGAGCGAGTGACCC	CCACACATTTGCTCTCTTCTG	
5	68653761	-	5	GGCATGAATATCCTAGCAATG	AGAATTGCTTGAATCCAGGAG	
*5	79234390	+	9	GATGATGGTGATGTACAGATGG	TCTGAGGTACCAGGTTCATCTC	
*5	84275156	-	7	GCTCTCTGTGTATCCCATTAGG	CTGGAAGACCAAAACACTTTTC	
*5	89486677	+	13	AACTGCATACAAACATTCAGCA	CCAATCTGCAACAAATCTACCT	
*5	98708608	-	6	TTTTTAACTTCTTTGCCTTTGG	GGAGATTATATCCTGCACATGG	
*5	98926938	-	14	TTCCCATCTTTGTGGTTTTATC	TAGTCAAAGAAAGGGGTGACAG	
*5	109595533	+	4	GCCTGTGCTTATGAGGTCTTAG	TGGCATTTGATACAATTCAAAA	
5	112477435	-	2	AGCATTCCTCTGTACTCTGCAT	CAAACCTCCATAAACCACTCAT	MCC
5	124183163	-	2	TGACTTGATTCCCCTACCATAC	GTGCCTCCCTTATCTTAAGGAC	
5	125068078	+	6	TTCCAAGATTCTCTTTGTGCTT	TACACACATGGCATCAGAAAAT	
5	132962016	+	3	GCATTTTCTTAAGCCTCACCTA	TAAGTTCATCTTTCCCATCACC	
5	144213566	-	1	TTCAATTTTGCCTATGGTTTTC	TGATTTCCCCTAAGATTAGAAGC	
5	144310932	-	2	ACACACACACGCATTTTCTTT	ACACACACACGCATTTTCTTT	
5	155266284	-	3	TTGTATGTGTTCTGACTGCTCC	CTCCAAAAAGAGAATCAACCTG	
*5	155552622	+	4	TGTTAGGACATATCAGCTGTGC	ATAAATTCTCAGGGTGATGGTG	
5	157799748	-	2	TTTTCCCATTTTGTAGGTTGTC	TCACAAGCATTCTTATACACCG	
5	159283584	-	13	TGGAGGTACCATTTGTTGAGA	CCATGGTGTCTTCTGTGATTC	ADRA1B
*5	161395800	-	13	TAGTTCTCGAGCCTTGGTTTT	ACTTGGAGAATCGGGTCACT	
6	13243006	+	1	TTCAGTGTCTAATGAATGGCAG	AATCAGATTGTTGCTGTGTCTG	
6	13611232	+	12	AATTGGCATTGCTCTTGTTACT	CAATATGCATTGGTCTCAGAAA	
6	18335716	+	1	TACAGAAAAAGTTGCTGACCCT	CCGATTGTACGTTCCATCTACT	DEK
6	36606018	+	1	CTGTACTTTTTGAGAGGCCAAG	ACTTACAATTAGCCCCCTTTTC	STK38
6	46042230	+	1	GTGGAGATGCAACCATAAAAAT	AGACATCCCAAAATCTGTCATC	CLIC5
*6	46418330	+	13	TTCAAATGTCTCTGCATCATTC	AGAAACAAGAACCTGGCTATGA	
6	53392986	+	10	CCCTTATCCTCTGCTAAAACCT	CCCTACGTCAGGTTGATTACAT	
6	56866318	-	7	TGAGTCCTCATAGCTTAGCTCC	CTGGGTCTGTGTGGATAGAAGT	
*6	72037524	+	9	GTCTGTTGGAGTTTGCTAGAGG	AGGTACTGGGTTCATCTCACTG	
6	73773301	+	3	AACCTGGTTGTATAGGCTGAAA	TTGAGGTCAGGAGTTTGATACC	KCNQ5
*6	74992108	-	2	TTTTAATTTCGGTGTTTTCTGC	GAAAGTGAAAAGATGGAAAAGG	
6	77436574	-	2	TTCATCATCATTTTTGTTCTGC	ACAGTGATTATATGACCCCCAG	
*6	78037300	-	14	TGTCAAGATGAAATATGGCTCA	CTGCTTTTGTGAATGATGAGAA	
*6	78502539	-	9	CTTTGATGATGGTGACGTACAG	AGACGAATGATTTCTGCATTTC	
6	83734761	-	1	CAGTCTCCAGAATAGGTGCAAT	CATCAAACTTTAAAAACCCCAA	C6orf157
6	108893430	-	1	TCATTTTTATGCTACCCAGGTC	TACTAGGGAGGCTGAGTCAAGA	LACE1

*6	123895408	-	3	CCAATGCAGTAGCCACTAGAC	CTTCTTCACAGCATCTGTTGG	TRDN
6	125733847	-	13	TGTGATATCTGAGAGTGGGTCA	AAAACTCCTCAGGTTCAAAGTG	
*6	142333058	+	5	CAGTTTTTCTGCTCTGCTTTTT	TCCTAGTCAAATAAAGGGGTGA	
6	142491622	-	11	GTCCAATTGCTCTGGTTATGA	AGGCGAGCTCAGATTTACTACT	
*6	148163189	-	12	GCTTCTGCATTCTTCACGTAG	TTGTTACACAAATGCACATGG	
*6	148237532	+	9	AATTCAACTTGGAAGTGCAAAC	TCTTAAGGTCTTTACCCATTTCA	
*6	155709317	-	5	AATCTGTAAATTGCTTTGGGC	AATGTAAATGGGCTAAATGCC	
*6	157888337	-	9	AGTGAGTGAGGAGCTGAAATGT	GCAGAGAAGAAATGACTCCAAT	
*6	162101432	+	10	CAGTTTAGATATGTGCGGGTG	CTTTCAGTGAACGAGGCATAG	PARK2
7	22749859	+	1	TTAAAAATTAGCCAGGTGTTGC	AGGTCTTTTCAAGTCACAAAGC	
7	25047810	-	12	GTGGGTTCATTCCAGTCTTATC	ACCGCATATTCTCACTCATAGG	
7	26761499	+	1	CATTTGTTTGTTCATTCAGCA	AGGAGGCTAGGGCAATAAAAT	SCAP2
7	34407244	+	2	GTCCTTCTTCTTGATGGCATAG	ATGAGTCTTGCTTCATCAACCT	
*7	53619881	-	6	CCTGTTTTTCCCACTCAAAGT	GTTTTGGAGGGTCGTAGAAAT	
7	61816291	+	1	AAGACATTTCCTTTTTCACCG	GTGAAACCGAATTTTGATGTG	
7	66228282	-	2	TACCTCCTTATCAACATCACCC	AAGACCATAGCAGACATCTCATC	FLJ10900
*7	69826421	-	9	GGAGATTTTCCTTTTGGAACTC	AAACACAAAAATTAGCCAGGTG	
*7	84739272	-	13	AAATTGGCTTAGGGAAAAAGAG	TTATTTTGGGAAGTCATTCCAG	
7	87841033	-	5	AGCTGTTAGACAGGGACGTTTA	ATCATACCAGGGAATGCTAAGA	
*7	92984394	-	5	GATGGATGAAGTTTTCCCATAA	GCAGTGCTGTTATGGATTGTAA	CALCR
7	105755294	+	1	TCTCTGAATACTGCTGAGGGTT	TCATTTCTCCCATACTTTGTCC	
7	121911834	+	2	AAACAGTCATGATATGAGCACG	GTAAAATCTCCAGCAAGGAATG	
7	144746343	-	11	TCTGAAGCCTTCTTCTCTCAAC	CATGGAGTCTCACTGATTGCTA	
7	154851420	-	2	TTCTTTCATTACATCTGCCCTC	CCACTCAGGGTTAAATGGATTA	
*8	1110666	-	11	TGCTCTGCTTTTTAGAGTTTCC	TACCCGAATACTGTGCTTTTTC	
*8	5816535	-	12	ACATCTGAGGATTGTGATTCG	GGTCTGAAGTGATGTCACCC	
8	11768782	+	4	CTACAGCTACAGGCAGCAAAG	TCTCATGCTCAAACAATCCTC	
8	18500203	+	8	AATCTTTGAATCCTGAATCCCT	AATTGGAAACCATCATTCTCAG	
8	24982191	-	7	AAAACTCCCCTTTCAAAGAGTC	ACTGGGGAATGAAAAATCAATA	
*8	36390042	-	9	TTAAGGACTTCTCTGCATTGGT	GTTTGATTAGGAAAACAAAGCG	
8	40552049	-	2	TTTGAAATCAAAGTCACTGCTG	GGTGTCATGTTTCCTTCACTTT	
8	43903825	+	7	AAGGTGCTCCAAATATCCAAT	TGTGTGCATTGAACTCAGAGA	
8	58510740	+	1	ATGAATGAGCCTTGAAAACACT	CTTTTAAGTTCTTAGGGCCTGC	
*8	67544279	-	11	TTACTGGAGAAATGGAATTTGG	TCCAGAGGGAATAACCATGTAG	
8	72561658	-	9	CATGCTCCTACAACAGTCTTCA	TGAGATGCTTTTCTGGGTTAGT	
8	83903289	+	2	ATGATGTCATCTGGACCATGTA	TTGAACTATGCGCTTGTTTTATT	
8	85591362	+	2	TCTTTGACTCCTTGATTTTGC	TGCATAAGTGGAGAGATATACACC	
*8	98610034	-	6	AATTTTCAGCTTCTCTGCTCTG	AGGTACAGGTGCATCTCACTG	
8	112190297	-	3	AATGGTATGAAGTTTAACAAAAACAC	TGAATAACATTTCCACCTCCTC	
8	115857528	-	2	ATTGGAAATTGCAACACTCTTC	ACATGCATATAACCATCAGCAC	
*8	115950995	+	8	TAGTTCTCATAGGGCCAGAAAA	CAGCCTTTGCAATACATGATAA	
*8	131110512	+	12	CATGCCTGGCTAATTTTTGTA	CCTCCTGAGTAGCTGGGATTA	
*9	225571	-	7	GTCTGTTGGAGTTTGCTAGAGG	GACACAGAATACGGGTGATTTC	
9	4255600	+	3	CTGGATAACATGGTGAAACCTT	ACATCTGGTGAAAGGTGATAGG	
*9	6280848	+	15	AAAATGTTTTGCCTATTTGCC	CACACACACACACACACACAC	
9	7317229	-	1	TGTGAATGGAATTAGCACAGAG	CCAGATGCTTTAGCACTAGGAG	
9	7416706	-	4	GTTCTGTTCTTGAACCTCTGGT	TACAAGGAAAACACAAAGGAGG	
*9	9226751	+	8	ATCAGGAATCAGGTTTCCAAC	CCCCATCCCAAAATAAATAAA	
*9	18639955	-	15	GCTAGATTTCCACTCCAGACC	AATAAAATTCCATGGAGGAGG	ADAMTSL1

*9	28980487	+	7	TGCTCTGCTTTTTAGAGTTTCC	CCTAGTCAAAGAAAGGGGTGAC	
*9	32082619	-	7	AATGGTATCCCAGATCTGCTC	GCTCACACCTGTAATCTCAGC	
*9	32453715	-	8	TTTCTAGAATTGAAATCCCTGG	AGAGTACTGGCAGACTGCTTTC	RIG-I
9	70189008	+	11	TGATATCGAGTTTGGGAAGTG	CGAGGAAATGCATCACATATT	PGM5
9	81272431	-	2	TCGTAACTTTCCCTGAAACTGT	CAGTGCTAACCAGAAATGATGA	
*9	94707443	-	9	TGGTTTGAATTTCCTCCTGTAG	AGTCTGAGATCAAACTGCAAGG	
*9	104113602	+	7	TGGATTGAGATTACACTGACCA	TCACACAGTGGAGATTTGACAT	
9	109058128	-	8	GGAGGCTGAGACATGAGAATAG	ACATGGGAGACTTTTCATTTTG	
9	110605000	+	2	CCAGCTACATGACTCCTAAACC	AACCCTCTATGATTTCCTCTCC	
*9	124459441	+	1	AGCTAGTAAGTGGATGAGCCAG	GTTTAAATCAACTGAGGGGTCC	
10	19105556	-	2	TGCTAGGAGATAAAACACCAGAC	CATGCCTGACTAGCTACCTACC	
*10	31777639	-	13	CTGAGGCTTCTGCATTCTTC	AGACTATATCCCACACCTGGC	
*10	43927619	-	4	TGGTTTGAATTTCCTCCTGTAG	GTCTGAGATCAAACTGCTAGGC	
*10	57877492	-	8	GTCTGTTGGAGTTTGCTAGAGG	TCTGAGGTACTGGGTTCATCTC	
10	91707512	-	2	CCCAAAAAGTTTTGCACTTTAG	CAGATTGAACACCAACTGACAC	
*10	108604142	-	1	AATTTTACCCTGTTGGTTGATG	AATTGGGAAGGAGAATTAGGAA	SORCS1
10	110443921	+	13	TCTCGTACTGTGGTTTTCAGC	TAGGACAGCAGTCTGAGTTCG	
*10	124445297	+	7	TCAAATTCAAAGCACACATAGC	GGACTCAGGGTGAGATTGTTAG	
10	132906631	-	3	AGACGAGGTTTCACCATGTTAG	AAAGGGAAGATGATTTGAGATG	TCERG1L
10	133378471	-	1	CACAGAGGTCTGAGGAGACAG	TAAACACACTCACCTGTGAAGC	
11	15336940	+	9	AAAGAAAAGCCACCAAATACTAA	GAATGGCTGTTACCAAAAAGAG	
11	19608869	-	8	TTTCTGCTCCTGCCTTTACTAC	TTGTGCACATGTACCCTAAAAC	
11	26116503	-	2	AAAATTACTGGTGTTTGCTGCT	GACCTTGAGGAAGTCGTTAGTG	
11	30137279	+	1	TGAAATACATATTTTCCAGGGG	TTAGTGCAGACCCCTAGTTTGT	
11	37767945	+	2	AAACCCTAAAAACACCTCCAGT	AATCTGTGGACTGTTCCTTTTG	
*11	48712714	+	14	CTTTGGTTTTTGATGATGGTG	AAGGCAGAAGACAGGTGATTT	
11	67356067	+	12	AGGCAGATCTGGTTCTCAAAT	TGCTTTTGAGTGAATTTTTCC	
11	67533650	-	1	CTCTGGAGCAAGAAGCTCTG	CTGAGGAGGAGGAGTGGG	
11	83329941	-	1	TTACCTCTATGCCTCAGTTTCC	CCTGAGGTCAAGAGTTCAAAAC	DLG2
11	90842710	-	2	ATGGAGGCAGTTCAGTCATTAT	GGAAGGAAATGACAACTACAGC	
*11	94324169	+	13	ACACAATCTGTTTTCCTCAACC	GCCTGTAATTCCAGTACTTTGG	
*11	104061193	-	6	GTTTCAATATCCAGCTTATGCC	ACACACACAGACACACACACAC	
11	104984392	+	1	CGTTACATTTCACCACAGATTG	AGCACAATACCTTGCACATTAG	
*11	109882864	-	6	TGCAGAACACAAGTCAATGTG	AATGAGCCTTTCCTCAGAGTC	
11	114400944	+	2	GATTTTGTCTTGGTAGGTTTCG	AGGTGAACTATGAGGATGCAAG	
11	122747116	+	4	GCCTAGAAAATTCACTTCATGG	GAGTTCAGAAAGAATGGAAACG	
*11	133810552	-	15	AGGTCAGGAGTTTGAAACCAG	GAGGGGCATCACTCAGATTAC	
*12	12824829	-	15	GACAATGTCTGGCTCTATTGC	TGGAATTATAGGTGTGAGCCA	
12	25667033	-	1	TGATTCTATTAGCTCCTCAGGG	GCTTTAACATAGCAACACCCTC	
*12	28117528	-	8	GGCAAATATGTGGAGAAAAGAG	GTCTGATCCATAGAATGCCATC	
12	33908428	-	6	TAAGCCTAGATTCTGTGAACCC	TTCAACAAAGGGATAGGAAATG	
*12	37877494	-	8	CAACTTCTTTTCCTTTGGTTTG	TACAGGAGCGGATATTTCAATC	
12	40923329	+	5	TGTTTATACCACCCAGTCTGC	ATGCCTGTGAATAGACACTGC	
*12	41543046	-	13	AGTAAGCAAGGTGTGCAGGG	TAGCAGTGGGAACTATCCATC	
*12	47005368	-	15	TGGTTTGAATTTCCTCCTGTAG	AGTCTGAGATCAAACTGCAAGG	
12	47215424	+	5	TTCCCTTTTCTCTGTAACCTGA	AAAACCAAAAATTGACAAATGG	
*12	53891078	+	9	AGTTTCCAGTTTTTCTGCTCTG	TAGTCAAAGAAAGGGGTGACAG	
*12	54068500	+	3	GAGCTAATTACAAGCTGGATGC	GCATGATTTATAGTCCTTTGGG	
*12	72075545	+	9	TCAGCTCCTTTAAGCACTTCTC	ACCACGAGATTATATCCTGCAC	

12	83489465	+	6	CACTGCATCCAGTCTAACAAAA	TTCTATGTAAGTGTCCATGCCA	
12	84030526	+	3	TCTGAGTAAACAGAATGCCAAC	ATCCTTGGATAGGACATTCTTC	
12	116298750	-	11	ATTTTCGTAGGCTCAGCATTT	GCTCCAAGCTCTTTCTAGTGG	
*12	116671475	+	6	ACTATCTTTTGGTGTTTGCCC	TCTTAATTTAGGTTCCCCCAA	KSR2
*12	125368458	-	9	TGATTAATATTCTCAGGGGCTC	TCTTGCCTGGAATTTTATTTTG	
*13	17976303	+	12	TCCAATTTTTCTGCTCTGTTTT	TTTCCTAGTCAAAGAAAGGGGT	
*13	26386225	+	7	GGAATTTGCCAACTCTGATAAA	CAAGTCATAAAGGTAGAACCCG	
13	41669428	+	5	ACCTACAAGAACAAGGGAAACA	GGATAACCCCTCTGGAATAAAG	DGKH
13	60119024	-	12	AGCACCTTACTTTCTTCATTGG	GTATGATTCCTAGCACTGCCTC	
*13	60360128	-	11	TAAGATAGACTTGAGGGGAGCA	TGCATTTCTCCTTTCTGAGATT	
*13	60632771	+	8	TCTCCTCAATTGTTTTTGCAG	TGGGTGAGAGTTTAATGAGCA	
*13	64446933	-	3	TTTTCCTGGTTGTTAGGTTTTC	CTGCCAGTGGAACATACTGTAA	
*13	76953231	+	1	CTATCTCCCAGCTGCTCTTTTA	AGTTTCTCCCAGGGTATTGATT	
*13	84769525	-	8	TTAAAAGGAAATTTTATTCACAAGC	CATCATTGATTAAAGCAGGAAAG	
14	30220276	-	2	GCCTATTTTGGATTTAACTTGC	TGTGATGCTAAGAAATGGTTTG	
*14	33209695	-	4	CGGATTCGTGTTCTTATTTGAT	CAGATGAGGAGTAGGCAGTGAT	NPAS3
*14	35744836	+	14	AAAAATCACAGACACATTTTCTTC	CCTCACTAGGGGCTGAGATAG	
14	50381232	-	8	TACGTAAGCACAAATGAAGGC	GTGGCTTAAATAATTGAGGGG	
14	58290253	+	4	TCTTCAAAGCCTAGCTGAAGTC	TCCCAATCTTGTGATTTTCTTC	
*14	65568174	-	11	CTTCCATCACTGATACCCTTTC	CAGACAAACAAAAAGACAGCAG	
14	66489363	-	7	GTACTCCTTTTTATGGCTGCAC	TAATAAAAAGACCAACAACCCG	GPHN
14	67854103	+	1	CATCACCTTGGTGTAAAGCAT	TGCTTACCAAATGGTATGACCT	RAD51L1
14	70267559	+	2	TGGATGAGCAGTAAATGAACAG	GTAGAGTCCTGCTGAGTGTTCC	
14	71403274	-	3	CCATAAAAGAGAATGATCTGGG	AAACAATGTGGAAAACAACTCC	
14	85377323	+	13	AATGTTGTGTTTCTCAAACTGAAG	CCAAGTTCTGAGGACAGAGTTC	
14	85451415	-	9	CACCTATTCATCCTTCCTTCC	ACTCAAATTAATGACCCCCAC	
14	89566936	-	12	GTTCTCACGCTTGGTATTCTTC	TTGTGCACATGTACCCTAAAAC	
15	22805045	-	5	TAAACTGATGAGAGCACAGCC	GGAGACAAGGGACTAGGGATT	
15	25549420	+	1	TATTTGCTCCCTTTTCCTGTAA	ACACACACACACACACACAGAG	
15	33441295	-	1	TGTATCTAAGTGCAGGCCTTTT	AGCACTGCAGACAGACACTCTA	
15	35495023	-	5	GTAAAAGGGGCACTTTGAGAG	TGTTTATTTCCTTTGTGCCAG	
15	43354873	-	1	CTTATGTTCTTGTGGCATTTTG	TAGTAGAGACGGGGTTTCACTG	SLC28A2
*15	45972163	-	7	TACCAGCTCCTCTTTGTACCTC	AGAAGCAAGAGCAGACAAATTC	
15	48976862	+	5	TTTACTGAAAATCTCTGCCCAC	ATGGTGAAACCCCATCTTTAC	
15	49631202	+	1	CTACGACACTGATGTCCTTTCA	TAGGATCCTTACCAGAAAAGCA	RC3
15	51743660	+	2	CATGCCTCAAAGCAGTAGAGTT	CCGGTTCTCAGCTAATACTCTG	FLJ38736
15	54038943	+	7	ATTCATTATGGATGCTGAGGC	TTCTCTTTCACTGATTTTGGC	NEDD4
15	59116100	-	5	CCAGAAGCATAAGCCTCTTAAA	CATCAAAATCCCCTGTAGTGTT	RORA
*15	81348528	-	10	TGACCTCATCATCCACTCATT	CCTGATGTCCTCATTGTGTTC	
15	83450506	+	1	GCCTTTCATCTCTCACTGTTTC	TTCAGGAGCTGGTAGTCTCATT	
*15	98748598	-	7	TCTTTAAAGGCAGAAGCAAGAC	TTTTCTTTTCCTTTGTGCTTTG	
16	5278880	+	1	AGAATGACTCGACCAGACAGAT	TCCCTTCATAATAACATTAGCCA	
16	16841748	-	11	ATGATCTCAGCTCACAAAGGTT	AATTGGAAACCATCATTCTCAG	
*16	17662466	-	7	TCTTTCTCTGACATTCCAGGTT	CTTCCAGGATGACAATCAAAAT	
16	22853202	-	5	TGAGTTCAAATTTTGGCTCTG	TCCTTAGTCAGCAACTTCGAC	
16	32432877	-	15	GTGTTTCCAAACTGCTGAATC	TGATTTGTGATGTGTGCATTC	
16	53849733	+	1	AAACGTGTATTTTATCGGTTGG	ATCAAGTGTTGAGAAGGCTCAT	
16	88505727	+	5	TCTGTGCACCATCAGTCAAC	AATGTAAATGGGTCACGCTCT	
17	25422607	+	4	ATCTACGTTGTGGCAAATGAC	CTGAAGAAAACATGAGGGAAAC	FLJ46247

17	29838023	+	7	GCTATGACAACTTGAATAGGGC	GAAACCCCATCTCTAGAAAACA	
*17	31296358	+	2	TTCATGATCTTTGACATTTTGC	TGAGCCTCAGTTTAATCCTGTT	
17	61981578	-	11	ATTACCTACAGCTCTTCTCCCC	TGTGTGTTGTAGCCAAGAAGTC	PRKCA
17	65869157	+	9	TGGTAGATATTACAAGCCCAGG	AAGTGGAAAGGTAGGGTTGTTT	
*18	43418386	+	12	TGCTCTGCTTTTTAGAGTTTCC	CAGGGAGTTCCCTTACCTAGTC	
18	47470081	-	5	AGACAGAATCTCTGCTTTCAGG	GCATCCTGACACATGGTAAATA	
*18	48351662	-	2	TGAATTCCTTTCCGTTATTCTG	CTGATAGCAAGACCTCACCAG	DCC
*18	50016501	-	12	TTAAATGAGCCTTTTGTGACC	AAGCTTCTTCAGTTCCCTTTG	
18	57730962	+	8	TGGAGTTTCACTCTTGTTACCC	CCAATCTTAACCTCTCTGGCTA	
18	67768379	+	3	GCAGCCATGAGAGATAACAGA	TGTGTCTCTTTGTCCAATCCT	
19	24393994	-	12	GCAGTTTTGAAACTCTGTTTTTG	AACGCACACATTACAAGGAAG	
20	6053791	-	2	CTGGTTTTAAGCGTTCTTCTTG	TTTTTAACCCACTTCTAACCCA	
*20	9425795	-	1	GAGTAGTTTGACAAGGAGCAGG	GTGGGAGGATCTGGAAGATTAT	
20	16427373	-	5	AAGCAGACAGAGACAAGAGAGG	GGACATCTCTGAGGACTACTGC	C20orf23
20	28205337	+	14	AACTCTTTCCAATCATCTCTGG	TCTGAGAATAAAATCAATGGGG	
*20	29384816	+	4	CTTTGTTCTGTTGCTGGTGAG	AGGGAGTTCCCTTTCCTAGTC	
20	41385386	-	13	TTTTGGAAGAATATGGCTCTG	CAGACTGCAAATCCCACTTTA	
21	18439085	-	10	GCATGTCACTTGTGATGATTTC	TGTATTTTTAGTAGAGACGGGGG	
*21	27991000	-	12	TTTTCACCTTCACTTATTGCCT	GGTACATGTGCAGGTGTGTTAC	
22	15626312	+	7	TCTGGGATTATAGGCACACAG	TTACATGCTTCTGGAACACTTG	
*22	15627723	+	8	AGAACAGTGGATTTTCGTGAAC	AATAGGAACAGCTCCGGTCTAC	
22	17590951	+	4	AACTTCCTGAAGGTGTGTTCC	TGCTTTGGGAAATAAGTGAGA	CLTCL1
*22	21045839	+	8	TTCCATTAGTTCTTTCCCTCTG	CAGCTATTTGTTAAAGGGGTTG	
22	42655713	-	11	ATCCTCAGGGTCTAGCAGAGA	ATGTGTGAGCACACTTCAGAG	C22orf20
X	3888925	-	1	GGTCTTGATCTGTCAATTAGGC	ATTTGCTAAGGATAATGGCCTC	
X	11863204	-	6	CTCTATTCCCAAGGCCTAGAAC	AATTGGAAACCATCATTCTCAG	
X	14061589	-	5	AGTCAAGCCCTCTCAAAATGTA	ACTAGGGAGCTGGAATCTTAGG	
X	21210715	+	13	TTTTGCTTAGGATTGTTTTGG	GAAAACTTCAGGCCAATATCC	
*X	46996130	-	4	GTAGAGACAGGGTTTCACCATC	CATCATTCTCAGCAAACTATCG	
X	55601659	-	7	TTTCATATTGCTTGCTGACTTC	TCACGTTTATTTGCAAACCTC	
X	61787710	+	6	TTTCCTTTTCCACCATAGACC	TTTTTGTGGTACTTGCAAACG	
*X	62975936	+	14	CCTGTAGCTCGGAGTAATTTG	AATACTTCGCTTTTCTGACGG	
*X	69776322	+	5	CTGATACCCTTTCTTCCAGTTG	GCTTAGGTAAACAAAGCAGCC	
*X	75866302	+	14	TTTTCTTTCATTTGTCTTGTGG	TGACACATTAGTGGGTGCAG	
*X	95739669	+	2	ACCTGTTTGCCTGGGTATTAG	GGAAAAATTCCTGAAAGTGGA	
X	98807470	-	8	TTTGATTTGCATTTCTTTGATG	AAAAGACAAAACTGGGAGCATA	
X	102249963	+	2	ATCTCAGGTCACTACAACCCC	ACTTGTCAGCTACTGCAAACC	
X	111444323	-	5	GGAAATTTTGATGCCTCTCAT	GGTCAGAGATAGTGTGCATGG	
X	119776563	-	4	CAAGACAGCCTCTTTTCCTAGA	TCCCATCACAATAATTTTCCAT	
*X	122492207	-	12	CAGATGATTGTAGGTGTGTGGT	AGCTCAATGGACACAGAAAAG	
X	124464713	-	1	ATTATTGCTCATTAATGTCCCC	TGTGTTGAGGAAACTCAAAGAA	
X	129352743	-	5	GACATTTTTCTACCTACCACGC	ATATATGACAAGTTATCCGGCG	
X	136008636	-	5	TCAAGGTCTACAAGCCAAGTC	GCACGTTTATCAGGTCATTTG	
*X	143512470	+	2	AATCTTTCTGGCCACTGTGTAG	ACACACACACACACACACACAC	
Y	11950753	+	12	CACCTGGGATCCAAAACTACT	AAATCAAATGTTCCAGAACCAC	
Y	12388782	+	7	TACCAGGGTTTGTGTTATAGGG	AGGAAGGGGAATATCACACTCT	
*Y	17607004	+	7	TCTGAAGCCTTCTTCTCTCAAC	GTCAAAGAAAGAGGTGATGGAC	
Y	57390031	-	7	GTATCTCACTCTGTCACCCAGC	ATGGAGTGGAATGGAATGGTAT	

Table S2: Raw statistics for sequence data appearing in this paper. Indicated are the number of reads generated for each library, the number of peaks into which the reads cluster, and the number of uniquely aligned reads after filtering. The column entitled "Signal" refers to the number of reads for a given lane that were incorporated into peaks with represent bona fide insertions. The percentage of uniquely aligning reads which represent bona fide human-specific L1 insertions for each lane is listed under "% Signal". The read length used for the run, and the specific Illumina Genome Analyzer (GA) platform used for each lane are also listed.

ID	Reads	Peaks	Uniquely Aligned	Signal	% Signal	Read Length	Platform
GM11993	6785973	8296	4149010	3961252	0.954746313	36	GA I
GM12878	15673377	6806	9347568	8959785	0.958515092	76	GA II
GM12891	15472915	7392	7276548	6685826	0.918818374	76	GA II
GM12892	15271927	6916	8940513	8537167	0.954885587	76	GA II
GM19238	10527361	12487	6507150	5069265	0.77902999	76	GA II
GM19239	8505774	12181	4849718	3839593	0.791714694	36	GA I
GM19240	8997073	11628	4705241	3867141	0.821879474	36	GA I
JapnICh	9411427	9327	5095475	4310418	0.845930556	36	GA I
JapnIFa	8675808	8093	5222872	4523469	0.866088428	36	GA I
JapnIMo	9031741	8287	5433986	4801321	0.883572575	36	GA I
JapnYChR1	5322102	6598	3415112	3090557	0.904965049	36	GA I
JapnYChR2	7624670	7726	4859533	4447533	0.915218191	36	GA I
JapnYFaR1	5043944	4971	3255538	3001899	0.922089989	36	GA I
JapnYFaR2	6445758	5467	4113599	3815390	0.927506546	36	GA I
JapnYMoR1	7746479	8601	5491693	5040112	0.917770167	36	GA I
JapnYMoR2	8923068	8863	6253950	5760027	0.921022234	36	GA I
Mz10601	11135724	6647	7957533	7139302	0.897175293	72	GA II
Mz10602	10941594	7588	7725460	6890480	0.89191841	72	GA II
Mz10901	15261168	10637	7628727	6682467	0.875960957	72	GA II
Mz10902	15580191	8059	7796922	6834116	0.876514604	72	GA II
Mz12201	14558423	11256	7977118	7169008	0.898696497	72	GA II
Mz12202	18311439	7595	13569978	12688048	0.935008738	72	GA IIx
SB4Ch	13910364	10132	8093522	7578119	0.936319071	72	GA II
SB4Fa	14909304	11182	8773598	8389517	0.956223091	72	GA II
SB4Mo	15192949	11434	9007753	8537816	0.94782972	72	GA II
SB3Ch	22714563	5487	15525409	14805738	0.953645601	72	GA IIx
SB3Fa	19885341	9412	12817339	12084141	0.942796395	72	GA IIx
SB3Mo	23121059	6020	14250341	13532567	0.9496311	72	GA IIx
Average	12320768	8539	7501472	6858646	0.905195455		

Supplemental Figures

Figure S1: Specificity of the L1-seq assay in terms of peak width and reads per peak as indicated on the axes.

Figure S2: Sensitivity of the L1-seq assay in terms of peak width and reads per peak as indicated on the axes.

Figure S3: Genomic distribution of detected L1 insertions. Gray bands on schematic chromosomes indicate the positions of human-specific L1 insertions in the reference genome (hg18). Reference (pre-Ta through Ta-1) and non-reference L1 insertions are colored according to the legend.

Figure S4: The number of segregating L1 insertions increases in a logarithmic manner relative to the number of individual genomes sampled. Individuals are added in permuted order yielding a distribution for the number of segregating sites for each possible number of unrelated individuals out of 13.

Figure S5: Length distribution of 165 L1Hs shared between targeted L1-flank resequencing data and 3 personal genomes (Ahn et al. 2009, Schuster et al. 2010). Histograms of reference and non-reference insertion length show similar distributions.

Figure S1:

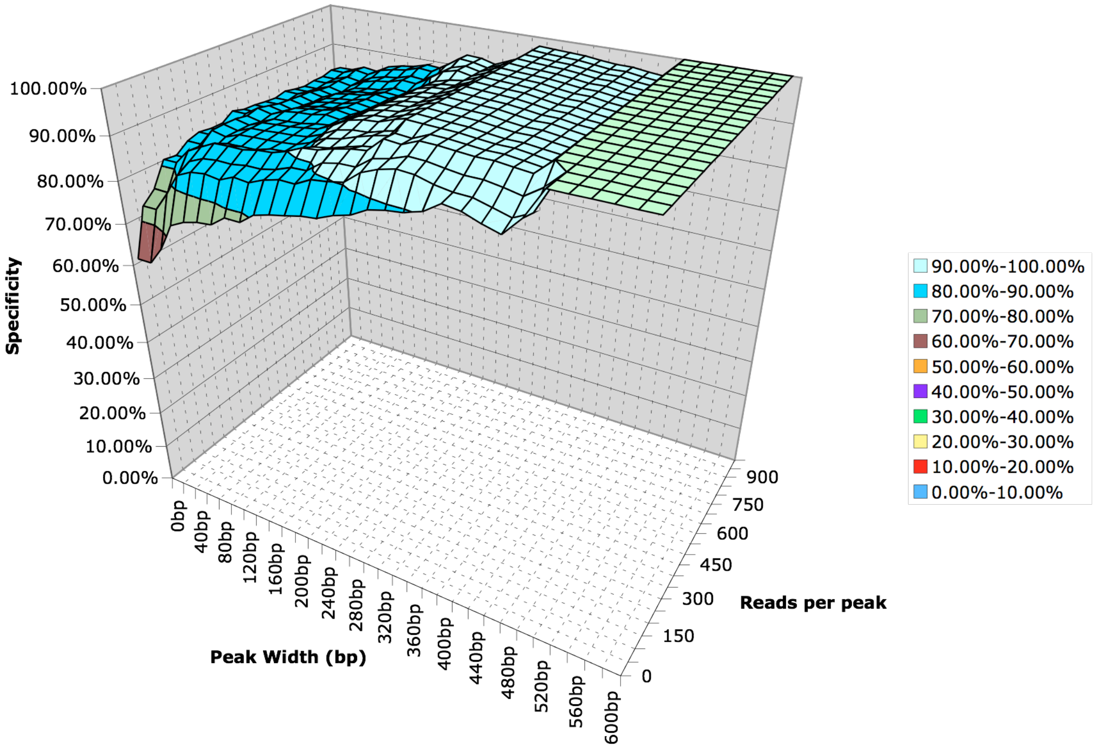
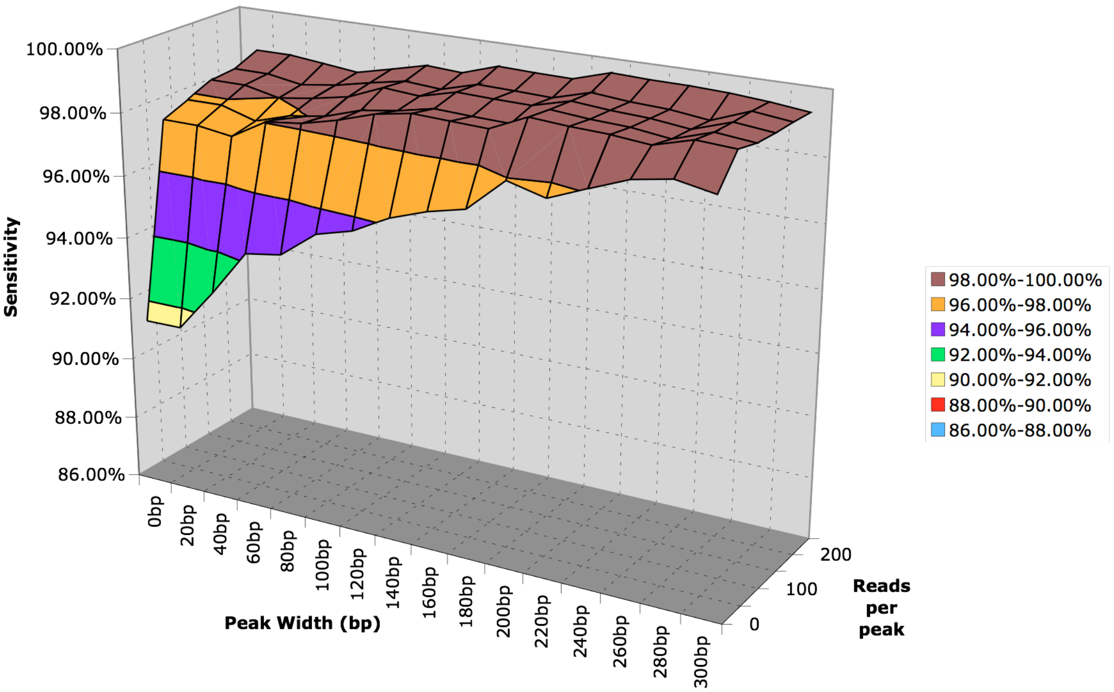


Figure S2:



Sensitivity
100.00%
98.00%
96.00%
94.00%
92.00%
90.00%
88.00%
86.00%
0bp
20bp
40bp
60bp
80bp
100bp
120bp
140bp
160bp
180bp
200bp
220bp
240bp
260bp
280bp
300bp
Peak Width (bp)
0
100
200
Reads per peak
🟫 98.00%-100.00%
🟧 96.00%-98.00%
🟪 94.00%-96.00%
🟩 92.00%-94.00%
🟨 90.00%-92.00%
🟥 88.00%-90.00%
🟦 86.00%-88.00%

Figure S3:

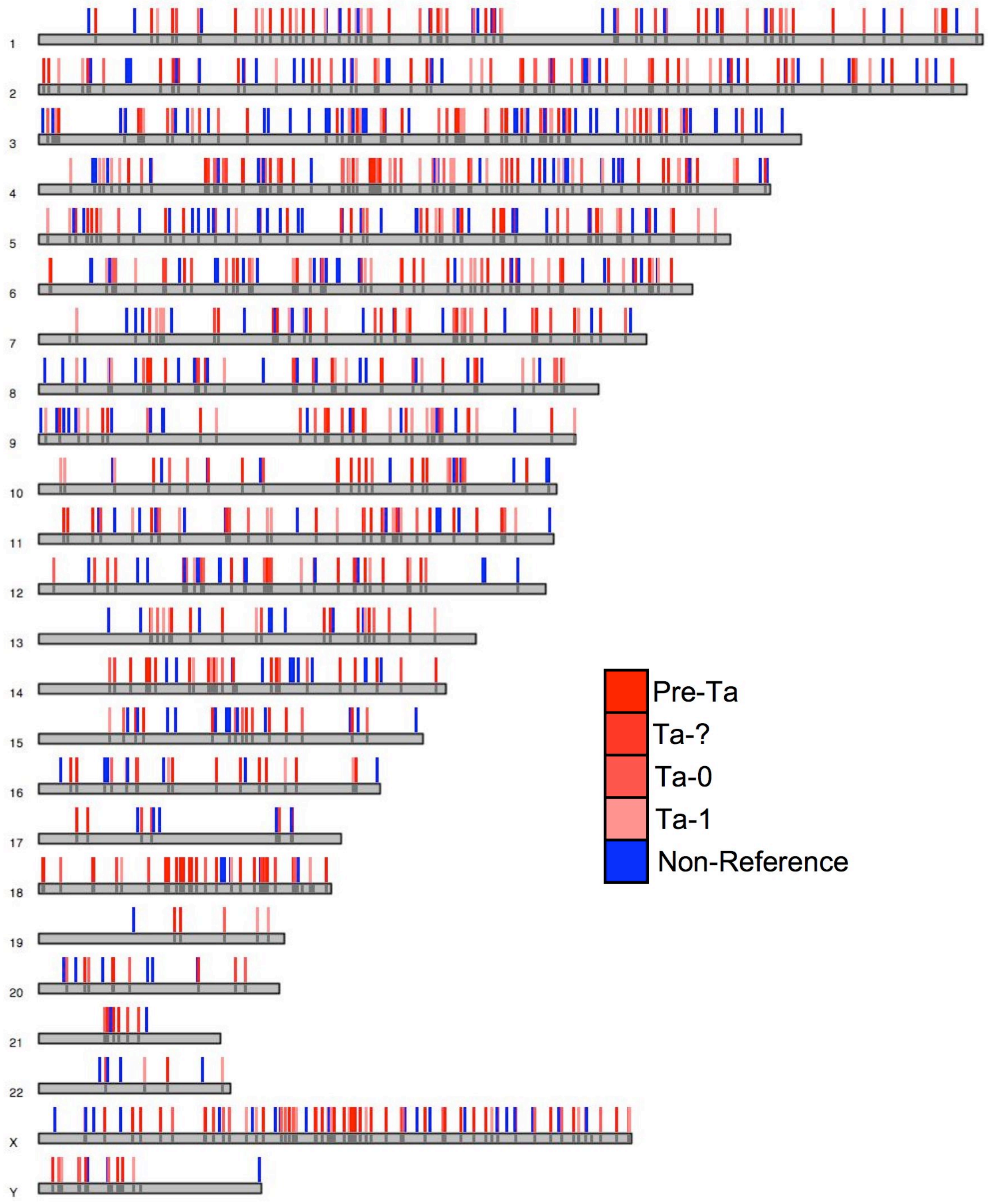

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Pre-Ta
Ta-?
Ta-0
Ta-1
Non-Reference

Figure S4:

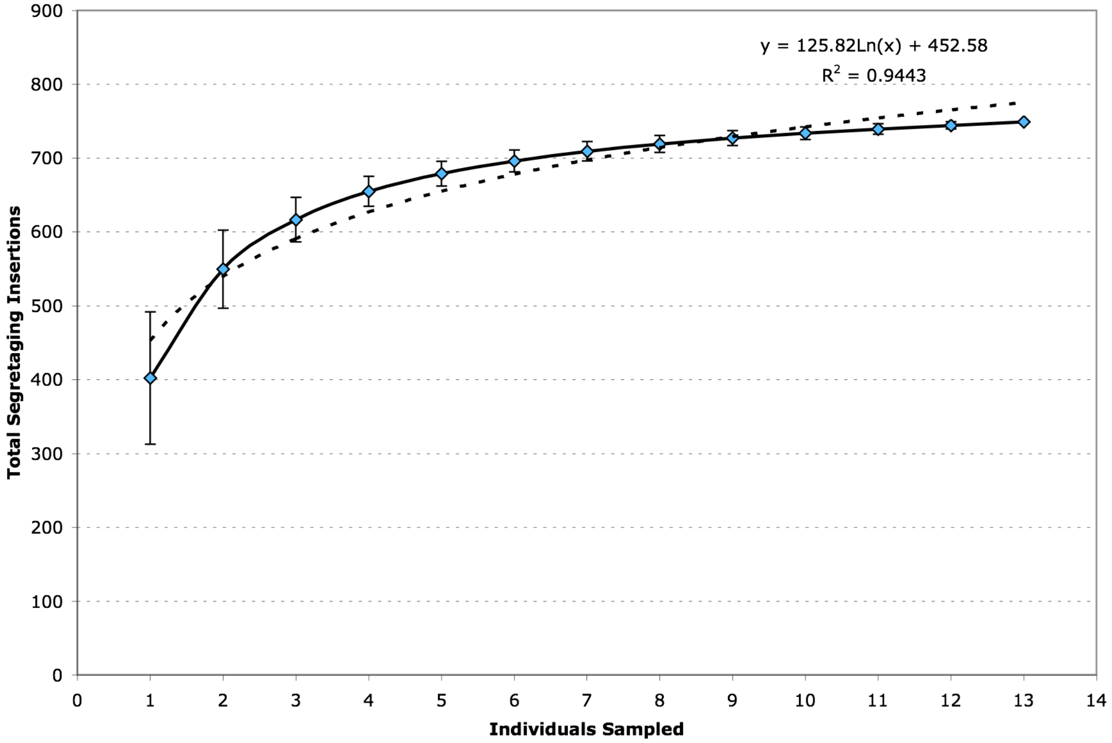
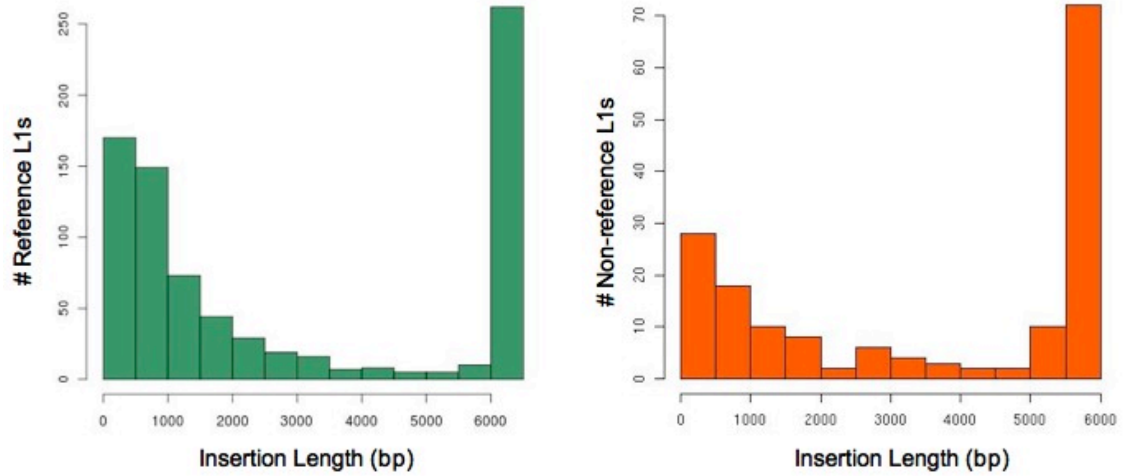

y = 125.82Ln(x) + 452.58
R^2 = 0.9443
Total Segretaging Insertions
900
800
700
600
500
400
300
200
100
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Individuals Sampled

Figure S5:


Reference L1s
0
50
100
150
200
250
0
1000
2000
3000
4000
5000
6000
Insertion Length (bp)
Non-reference L1s
0
10
20
30
40
50
60
70
0
1000
2000
3000
4000
5000
6000
Insertion Length (bp)