Supporting Information for the article:

Transduction-Specific ATLAS (TS-ATLAS) reveals a cohort of highly active L1 retrotransposons in human populations

Catriona M. Macfarlane, Pamela Collier, Raheleh Rahbari, Christine R. Beck , John F. Wagstaff, Samantha Igoe, John V. Moran, and Richard M. Badge

Supp. Methods

L1 elements

For convenience previously described L1s are named as listed in Supp. Table S2. L1 elements discussed in this study are named according to their corresponding insertion site accession number from the HGR.

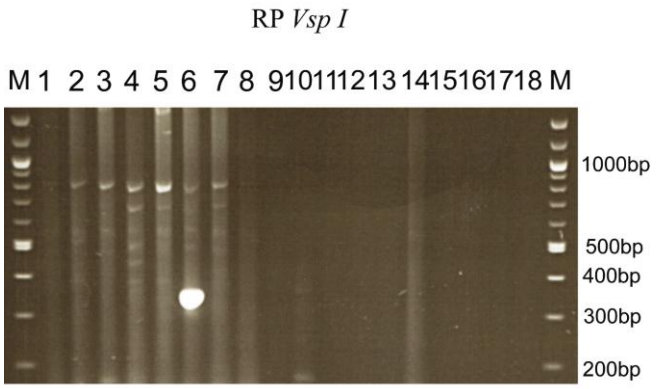
DNA samples

The 9 blood samples were obtained from individuals of Northern European origin. TS-ATLAS for the LRE3 transduction was performed on the CEPH/FRENCH pedigree 12 and CEPH/UTAH pedigrees 1333, 1340, 1424, and 1347. TS-ATLAS for the 2980 transduction was carried out using the CEPH/UTAH pedigrees 1333, 1340, 1424, and 13291. Each family was selected on the basis that one copy of either LRE3 or AC002980 was segregating in the pedigree, with one grandparent being heterozygous and the remaining being homozygous for the selected transduction locus. These assays were used to verify that TS-ATLAS amplification patterns recapitulated the segregation of AC002980 and LRE3 in the pedigree, as determined by genotyping. The results of applying this procedure to 9 unrelated individuals carrying the AC002980 full-length L1 insertion are illustrated in Figure 3B in the main text.

Library Construction and Amplification

Protocol 1. TS-ATLAS RP-specific VspI Library Construction and Amplification

The method is as described within the main text. Supp. Figure S1 shows a representative RP-specific TS-ATLAS display gel of 8 unrelated individuals, who necessarily lack the private disease causing $L1_{RP}$ insertion. Faint bands in lanes 1-8 may be $L1_{RP}$ related loci but present insufficient DNA for characterisation. The bright band in lane 6 corresponds to the putative $L1_{RP}$ progenitor element (AL050308) reported in this study.

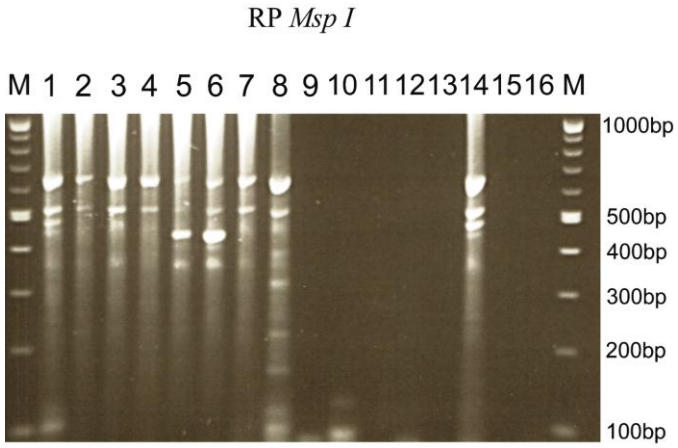


Supp. Figure S1. Representative TS-ATLAS Display Gel Showing the Results of Applying the RP-Specific (*Vsp*I) Assay. Lanes 1-8 show display patterns of 8 unrelated individuals none of whom carries the $L1_{RP}$ insertion. Lane 9 - Reaction in the absence of genomic DNA. Lane 10 - Reaction in the absence of restriction enzyme. Lane 11 - Reaction in the absence of genomic DNA. Lane 12 - Reaction in the absence of T4 Ligase. Lane 13 - Reactions in the absence of linker. Lane 14 – Replicate of Lane 1. Lanes 15 to 18 - Controls for PCR with DNA omitted. MW - molecular weight marker (100bp ladder (NEB)).

Protocol 2. TS-ATLAS RP-specific MspI Library Construction and Amplification

An aliquot (600ng) of genomic DNA was digested to completion with 20 units of *Msp*I (NEB) in the manufacturer's recommended buffer at 37°C for 3 hours. After incubation reactions were heated to 65°C for 20 minutes to inactivate the restriction enzyme. Prior to setting up the ligation reaction, linker oligonucleotides were freshly annealed by mixing equal volumes of 20 μM RBMSL2 and RBD5, heating to 65°C for 10 minutes, and then slowly cooling to room temperature. An aliquot (100ng) of the digested DNA was ligated to a molar excess of the annealed suppression linker (2.7μl of 10uM annealed linker for *Msp*I libraries) with 4 Weiss units T4 DNA ligase (Promega) in 1X Ligase Buffer (Invitrogen) overnight (~16hrs) at 15°C, in a final volume of 20μl. After ligation the reaction was heated to 70°C for 10 minutes to inactivate the ligase. Excess linkers and short DNA fragments (i.e., < 100 bp) were removed with the Qiaquick PCR purification system (Qiagen), following the manufacturer's protocol, but eluting the DNA in 30μl 5mM Tris HCl pH7.5. An aliquot (1μl) of ligated genomic DNA was amplified in 10μl PCR reactions containing 1 X PCR buffer, 1.25μM RBX4, 1.25μM RB3PA1, and 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30 X

[96°C -30s; 58°C -30s; 72°C -1min]; 72°C -10min. Primary suppression PCR reactions were diluted 1:50 in SMDD and 1µl diluted PCR reaction was added to 9µl secondary PCR reactions containing 1 X PCR buffer, 0.625mM RBY1, 0.625mM RB011TD1, and 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30[96°C -30s; 58°C -30s; 72°C -1min]; 72°C -10min. An aliquot (10µl) of secondary PCR products were fractionated on 2% Seakem LE (Cambrex) 0.5X TBE agarose gels against the 100bp ladder (NEB) size marker and visualised by ethidium bromide (0.5 µg/ml) staining. Novel PCR products (i.e. amplicons not corresponding in size to the suppressed known transduction locus) were excised from the gel and purified using the Qiagen Minelute system (Qiagen) following the manufacturer's protocol, but eluting the DNA in 10µl of 5mM Tris HCl pH7.5. Purified PCR products were directly sequenced with ABI BigDye Ver. 3.0 ReadyReaction, using 3.3uM RBY1 as the primer. Sequencing reactions were purified using Performa DTR spin columns (Edge BioSystems) and the sequencing data collected using an ABI 3730 capillary sequencer by the PNACL DNA sequencing service (University of Leicester). Supp. Figure S2 illustrates a representative TS-ATLAS display gel for the RP-Specific (*Msp*I) assay. Strongly amplifying bands showing variable presence absence between individuals were characterised by excision and direct sequencing. Faint bands in Lanes 1-8 may represent other $L1_{RP}$ related loci, but were not characterised due to insufficient amplification.



Supp. Figure S2. Representative TS-ATLAS Display Gel Showing the Results of Applying the RP-Specific (*Msp*I) Assay. Lanes 1-8 show display patterns of 8 unrelated individuals none of whom carries the $L1_{RP}$ insertion. Lane 9 - Reaction in the absence of genomic DNA. Lane 10 - Reaction in the absence of restriction enzyme. Lane 11 - Reaction in the absence of genomic DNA. Lane 12 - Reaction in the absence of T4 Ligase. Lane 13 - Reactions in the absence of linker. Lane 14 – Replicate of Lane 1. Lanes 15 and 16 - Controls for PCR with DNA omitted.MW - molecular weight marker (100bp ladder (NEB)).

Protocol 3. TS-ATLAS AC002980-specific Nla*III *Library Construction and Amplification

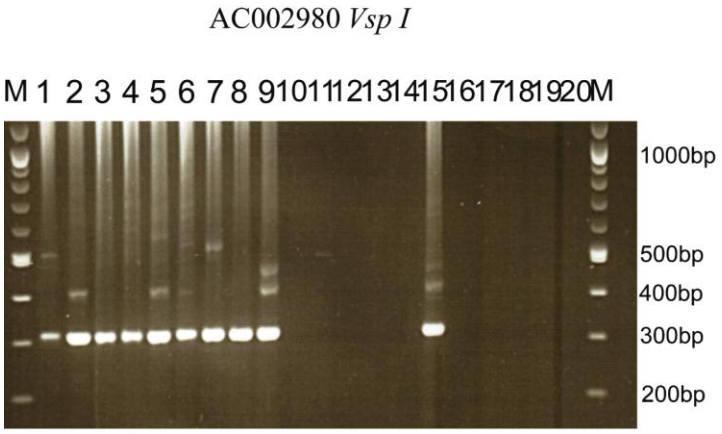
An aliquot (600ng) of genomic DNA was digested to completion with 20 units of *Nla*III (NEB) in the manufacturer's recommended buffer at 37°C for 3 hours. Incubation reactions were heated to 65°C for 20 minutes to inactivate the restriction enzyme. Prior to setting up the ligation reaction, linker oligonucleotides were freshly annealed by mixing equal volumes of 20 μM RBMSL3 and RBD4, heating to 65°C for 10 minutes, and then slowly cooling to room temperature. An aliquot (100ng) of the digested DNA was ligated to a molar excess of the annealed suppression linker (2.7μl of 10μM annealed linker for *Vsp*I libraries) with 4 Weiss units T4 DNA ligase (Promega) in 1X Ligase Buffer (Invitrogen) overnight (~16hrs) at 15°C, in a final volume of 20μl. After ligation the reaction was heated to 70°C for 10 minutes to inactivate the ligase. Excess linkers and short DNA fragments (i.e., < 100 bp) were removed with the Qiaquick PCR purification system (Qiagen), following the manufacturer's protocol, but eluting the DNA in 30μl 5mM Tris HCl pH7.5. To suppress amplification of the L1 AC002980, an aliquot (10μl) of the ligation reaction was incubated with 10 units *Mun* I (Roche) for 3 hours at 37°C**,** in a final reaction volume of 20μl. Reactions were heated to 65^{o}C for 20 minutes to inactivate the enzyme, cooled on ice, and centrifuged briefly. An aliquot (1μl) of ligated and

*Mun*I digested genomic DNA was amplified in 10µl PCR reactions containing 1 X PCR buffer (45mM Tris HCl pH 8.8, 11mM NH_4SO_4, 0.9mM $MgCl_2$, 6.7mM β-mercaptoethanol, 113 µg/ml BSA, 1mM dNTPs), 1.25µM RBX4, 1.25µM RB980TD2 and 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30X [96°C -30s; 59.6°C -30s; 72°C -1min]; 72°C -10min. Primary suppression PCR reactions were diluted 1:50 in Single Molecule Dilution Diluent (SMDD: 5mM Tris HCl pH7.5, 5ng/µl sonicated *E.coli* genomic DNA) and 1µl diluted PCR reaction was added to 9µl secondary PCR reactions containing 1 X PCR buffer, 0.625mM RBY1, 0.625mM RB980TD3, 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30 [96°C -30s; 64°C -30s; 72°C -1min]; 72°C -10min. An aliquot (10µl) of secondary PCR products were fractionated on 2% Seakem LE (Cambrex) 0.5X TBE agarose gels against the 100bp ladder (NEB) size marker and visualised by ethidium bromide (0.5 µg/ml) staining. Novel PCR products (i.e. amplicons not corresponding in size to the suppressed known transduction locus) were excised from the gel and purified using the Qiagen Minelute system (Qiagen) following the manufacturer's protocol, but eluting the DNA in 10µl of 5mM TrisHCl pH7.5. Purified PCR products were directly sequenced with ABI BigDye Ver. 3.0 ReadyReaction, using 3.3µM RBY1 as the primer. Sequencing reactions were purified using Performa DTR spin columns (Edge BioSystems) and the sequencing data collected using an ABI 3730 capillary sequencer by the PNACL DNA sequencing service (University of Leicester). The results of applying this protocol to *Nla*III libraries with *Mun*I suppression to 9 unrelated individuals carrying the AC002980 full-length L1 insertion are illustrated in Figure 3B of the main text.

Protocol 4. TS-ATLAS AC002980-specific VspI Library Construction and Amplification

An aliquot (600ng) of genomic DNA was digested to completion with 20 units of *Vsp*I (Promega) in the manufacturer's recommended buffer at 37°C for 3 hours. After incubation reactions were heated to 65°C for 20 minutes to inactivate the restriction enzyme. Prior to setting up the ligation reaction, linker oligonucleotides were freshly annealed by mixing equal volumes of 20 µM RBMSL2 and RBD3, heating to 65°C for 10 minutes, and then slowly cooling to room temperature. An aliquot (100ng) of the digested DNA was ligated to a molar excess of the

annealed suppression linker (2.7µl of 10uM annealed linker for *Vsp*I libraries) with 4 Weiss units T4 DNA ligase (Promega) in 1X Ligase Buffer (Invitrogen) overnight (~16hrs) at 15°C, in a final volume of 20µl. After ligation the reaction was heated to 70°C for 10 minutes to inactivate the ligase. Excess linkers and short DNA fragments (i.e., < 100 bp) were removed with the Qiaquick PCR purification system (Qiagen), following the manufacturer's protocol, but eluting the DNA in 30µl 5mM TrisHCl pH7.5. An aliquot (1µl) of ligated genomic DNA was amplified in 10µl PCR reactions containing 1 X PCR buffer, 1.25µM RBX4, 1.25µM RB3PA1, and 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30X [96°C -30s; 59.6°C -30s; 72°C -1min]; 72°C -10min. Primary suppression PCR reactions were diluted 1:50 in SMDD and 1µl diluted PCR reaction was added to 9µl secondary PCR reactions containing 1 X PCR buffer, 0.625mM RBY1, 0.625mM RB980TD3, 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30X [96°C -30s; 64°C -30s; 72°C -1min]; 72°C -10min. An aliquot (10µl) of secondary PCR products were fractionated on 2% Seakem LE (Cambrex) 0.5X TBE agarose gels against the 100bp ladder (NEB) size marker and visualised by ethidium bromide (0.5 µg/ml) staining. Novel PCR products (i.e. amplicons not corresponding in size to the suppressed known transduction locus) were excised from the gel and purified using the Qiagen Minelute system (Qiagen) following the manufacturer's protocol, but eluting the DNA in 10µl of 5mM Tris HCl pH7.5. Purified PCR products were directly sequenced with ABI BigDye Ver. 3.0 ReadyReaction, using 3.3uM RBY1 as the primer. Sequencing reactions were purified using Performa DTR spin columns (Edge BioSystems) and the sequencing data collected using an ABI 3730 capillary sequencer by the PNACL DNA sequencing service (University of Leicester). Supp. Figure S3 illustrates how the presence of the AC002980 amplicon (strong band ~300bp) inhibits amplification of related transduction carrying amplicons (weak bands), necessitating the use of enzymatic treatment to suppress this amplicon, as used in protocol 1.

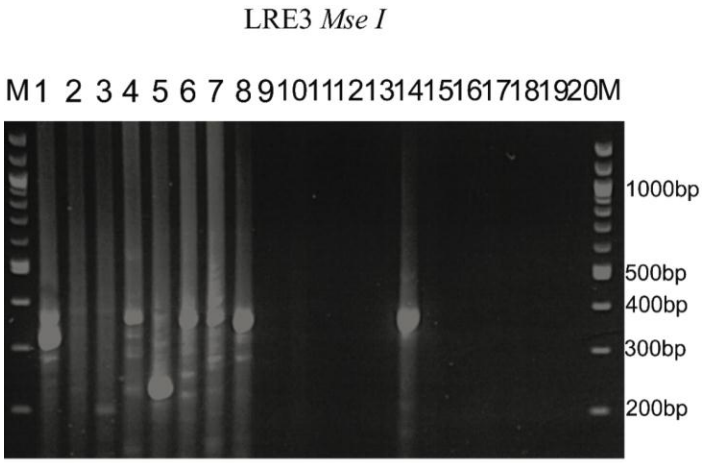


Supp. Figure S3. Representative TS-ATLAS Display Gel Showing the Results of Applying the AC002980-Specific (*Vsp*I) Assay. Lanes 1-9 show display patterns of 9 unrelated individuals. Lane 10 - Reaction in the absence of genomic DNA. Lane 11 - Reaction in the absence of restriction enzyme. Lane 12 - Reaction in the absence of genomic DNA. Lane 13 - Reaction in the absence of T4 Ligase. Lane 14 - Reaction in the absence of linker. Lane 15 – Replicate of Lane 9. Lanes 16 to 20 - Controls for PCR with DNA omitted.MW - molecular weight marker (100bp ladder (NEB)).

Protocol 5. TS-ATLAS LRE3-specific MseI Library Construction and Amplification

An aliquot (600ng) of genomic DNA was digested to completion with 15 units of *Mse*I (NEB) in the manufacturer's recommended buffer at 37°C for 3 hours. After incubation reactions were heated to 65°C for 20 minutes to inactivate the restriction enzyme. Prior to setting up the ligation reaction, linker oligonucleotides were freshly annealed by mixing equal volumes of 20 μM RBMSL2 and RBD3, heating to 65°C for 10 minutes, and then slowly cooling to room temperature. An aliquot (100ng) of the digested DNA was ligated to a molar excess of the annealed suppression linker (2.7μl of 10μM annealed linker for *MseI* libraries) with 4 Weiss units T4 DNA ligase (Promega) in 1 X Ligase Buffer (Invitrogen) overnight (~16hrs) at 15°C, in a final volume of 20μl. After ligation the reaction was heated to 70°C for 10 minutes to inactivate the ligase. Excess linkers and short DNA fragments (*i.e.*, < 100 bp) were removed with the Qiaquick PCR purification system (Qiagen), following the manufacturer's protocol, but eluting the DNA in 30μl 5mM Tris HCl pH7.5. To suppress amplification of known transduction locus (loci?) 10μl of the ligation reaction was incubated with 10 units *Bbs* I (NEB) for 3 hours at 37°C**,** in a final reaction volume of 20ul. Reactions were heated to 65^{o}C for 20 minutes to

inactivate the enzyme, cooled on ice, and centrifuged briefly. An aliquot (1µl) of ligated and *Bbs* I digested genomic DNA was amplified in 10µl PCR reactions containing 1 X PCR buffer, 1.25µM RBX4, 1.25µM RB3PA1, and 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30 X [96°C -30s; 59.6°C -30s; 72°C -1min]; 72°C -2min. Primary suppression PCR reactions were diluted 1:50 in SMDD and 1µl diluted PCR reaction was added to 9µl secondary PCR reactions containing 1 X PCR buffer, 0.625mM RBY1, 0.625mM CM958TD1, and 0.4 units of *Taq* DNA polymerase (ABgene). Reactions were cycled in a Tetrad 2 Thermal Cycler (MJ Research / Biorad, Hercules, CA) using the following conditions: 96°C -1min; 30X [96°C -30s; 59.6°C -30s; 72°C -1min]; 72°C -2min. An aliquot (10µl) of secondary PCR products were fractionated on 2% Seakem LE (Cambrex) 0.5X TBE agarose gels against the 100bp ladder (NEB) size marker and visualised by ethidium bromide (0.5 µg/ml) staining. Novel PCR products (i.e. amplicons not corresponding in size to the suppressed known transduction locus) were excised from the gel and purified using the Qiagen Minelute system (Qiagen) following the manufacturer's protocol, but eluting the DNA in 10µl of 5mM TrisHCl pH7.5. Purified PCR products were directly sequenced with ABI BigDye Ver. 3.0 ReadyReaction, using 3.3uM RBY1 as the primer. Sequencing reactions were purified using Performa DTR spin columns (Edge BioSystems) and the sequencing data collected using an ABI 3730 capillary sequencer by the PNACL DNA sequencing service (University of Leicester). Supp. Figure S2 illustrates a representative TS-ATLAS display gel for the LRE3-Specific (*Mse*I) Assay. In the absence of the LRE3 supressing restriction enzyme *Bbs*I only the LRE3 amplicon is generated (Supp. Figure S2, lane 14, ~350bp band). With *Bbs*I digestion (Supp. Figure S4, Lanes 1-8) a range of differently sized fragments can amplify to levels consistent with excision and cloning. Faint bands in Lanes 1-8 may represent other LRE3 related loci, but were not characterised due to insufficient amplification.



Supp. Figure S4. Representative TS-ATLAS Display Gel Showing the Results of Applying the LRE3-Specific (*Mse*I) Assay. Lanes 1-8 show display patterns of 8 unrelated individuals. Lane 9 - Reaction in the absence of genomic DNA. Lane 10 - Reaction in the absence of restriction enzyme (*Mse*I). Lane 11 - Reaction in the absence of genomic DNA. Lane 12 - Reaction in the absence of T4 Ligase. Lane 13 - Reactions in the absence of linker. Lane 14 - Reaction in the absence of supression enzyme digestion (*Bbs*I). Lanes 15 to 20 Control for PCR with DNA omitted. MW - molecular weight marker (100bp ladder (NEB)).

Results

In Silico Recovery of L1s belonging to Active Transduction Lineages within the HGR (hg18)

In total 6 L1 elements belonging to the transduction lineages AC002980, LRE3, and $L1_{RP}$ were detected within the HGR; for further detail refer to Table 1 in the main text and Supp. Table S2.

AC002980: the transduction of the L1 AC002980 produced three BLAT hits with >98% identity which were preceded by an L1. The first corresponded to the L1 AC002980 (chrX: 11863136-11863323), the second a 5′ truncated L1 AC010387 (chr5: 24406320-24406557), and the third a full length L1 AL118519 (chr6: 70782984-70783219). L1 AC010387 is a 5′ truncated and inverted L1 2563bp in length, and flanked by TSDs of 14bp. L1 AL118519 was a full length L1HS which contained a 1bp frameshifting deletion within ORF2. Comparison of the three 3′ flanking sequences and locations of each of the L1 TSDs indicated that AC002980 and AC010387 were likely derived from AL118519 (Figure 4A). An overview of the AC002980

transduction alignment and its implication for variable polyadenylation is presented in Figure 4B of the main text. A detailed, annotated alignment of the transduced sequence with primer and restriction sites is presented in Supp. Figure S5.

$L1_{RP}$*:* the poly(A) tail and 11bp transduced sequence of $L1_{RP}$ yielded two BLAT hits with >95% identity which were preceded by an L1. The first, AC093861 (chr4: 44202044 – 44202074), was a 5′ truncated L1 and did not appear to have any TSDs. The second, AC005939 (chr17: 65972606-65972764), was a full length L1 containing a stop codon within ORF2 and was flanked by 11bp TSDs (Figure 2A). A detailed, annotated alignment of the transduced sequence with primer and restriction sites is presented in Supp. Figure S6.

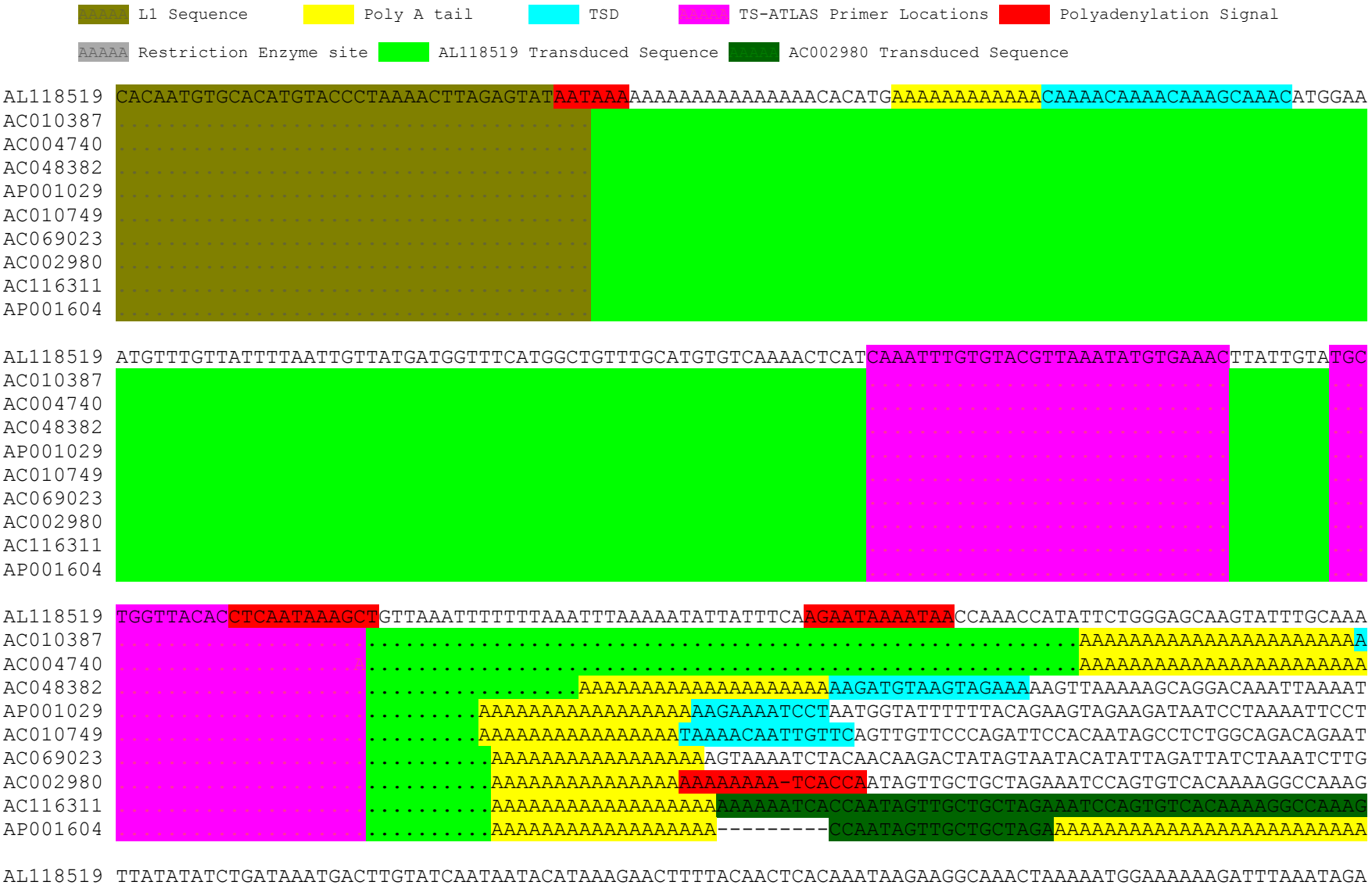
LRE3*: in silico* searches using the transduction of LRE3 produced an exact match on the long arm of chromosome 2, however no full length L1 preceded the transduced sequence indicating that the polymorphic $L1_{LRE3}$ was not present within the HGR (Brouha et al., 2002). No further L1s containing the LRE3 transduction were detected in the HGR (Figure 2B). A detailed, annotated alignment of the transduced sequence with primer and restriction sites is presented in Supp. Figure S7.

Screening for novel L1s belonging to Active Transduction Lineages within the 1000 Genomes Pilot data

The observation that none of the novel L1 insertions discovered using active lineage specific TS-ATLAS were reported in bioinformatic analyses of the 1000 Genomes Pilot study sequence data, prompted *in silico* screening. Examination of overlap between our genotyping panel and the 1000 Genomes study cohort revealed individuals carrying novel TS-ATLAS elements that were not reported by Ewing and Kazazian 2011, among them the putative $L1_{RP}$ progenitor, AL050308. To discriminate whether this incongruity arose due to stochastic coverage within the 1000 Genomes pilot dataset, or the methodological challenge of mapping insertions carrying long repetitive transductions with short read data we screened the 1000 Genomes Pilot 1, Pilot 2 and Pilot 3 sequence read archives (derived from 375 individuals) for evidence of TS-ATLAS captured insertions. Briefly, for each pilot study and individual, 454 sequence reads >45 bp in length were downloaded and converted to BLAST-formatted databases [~1.6 billion reads in total]. 300 bp of genomic DNA sequence flanking the insertion point of TS-ATLAS recovered elements, that were absent from the HGR (hg18), were used as queries for MegaBLAST

(v2.2.26) searches (Zhang *et al.*, 2000). Command line options were : megablast -F F -p 80 -b 50000 -v 50000. Alignments <45 bp and <95% identity were filtered out. The filtered BLAST output was parsed with a specifically designed perl script (available on request) to identify junction sequence reads that could be unequivocally mapped to TS-ATLAS element insertion points and that contained non-aligned sequences. Annotated results were stored in a MySQL database and accessed using custom queries. The non-aligned sequences were identified as L1 derived (5′) or poly A tail / transduction derived (3′) by re-alignment and manually verified by inspection. Elements with at least one incontrovertible junction read, with high similarity to the known insertion sequence, are indicated by the letter Y in the third column ("Detected in 1000G" of Table 1. Most insertions were supported by more than one read – the unique read identifier and positioning (5′ or 3′) of these reads is shown in Supp. Table S3.

Supp. Figure S5



```
AC010387 AATTTTGTTTTAAAAAGTAATTAAAAATTGGAATGATTATTTATTTTGGGGATATAGATATGAAGAAAAATATTTTCAAACCATTAAACAATTTTTAATA
AC004740 AAACAATATGTAATGTTTCACTTAATATTTTGTAACCTTGTTTACCTTGAAGTCTACTTTGTCCAATGTTAAGATATTTCAGTTTTTTTATGATTACTGT
AC048382 AGAGTTTTTATTAGTTTTCTTTTTATTTGTTTCATTGTTTATTTGTTTGTTCCTTTGTGCAAGCAGTGCAAAGTTGTCATCAGTTTAAATAATGGGTTAT
AP001029 ATGGAACCACAAAAGACAATGAATAGCCAATCAATCTTGAGAAAGAATAACAAAGCCAGAGACATCACACTTTCTGATCTTACAATATATTATAAAGTTA
AC010749 GAGAACATG
AC069023 TCAGCAAACCTGAACTAAGCATCTGTTATGCACTGTTTAAATGCTGGAGGCCTGAAAGCTTTGGGGGACTAAACACATCCCCTAAAAGATTTGAATTTAC
AC002980 TTTATTGACAAATTGGTGTATATGAAAGACCTCGACCTTCAACATTCACAAGAAAAGATCAATTGCATGTAATTTGTCTTGCCCGTAATTTTGCTGTTCA
AC116311 TTTATTGACAAATTGGTGTATATGAAAGACCTCGACCTTCAACATTCACAAGAAAAGATCAATTGCATGTAATTTGTCTTGCCCGTAATTTTGCTGTTCA
AP001604 AAAAGAAAGTGCCCTGTAGTCCTTTCAACCACTGGGTGGATACACCAAAATGTGAGAAACAGTGTGGGTAAAGTAACGATGCCACCCAGAAAAGTAAAAC

AL118519 TGATTCTTCAAAGAAGATATGTGAATGTCTAACAAGCAGAGGAAATGATCATTAAT
AC010387 TTTAGTAGACACCATTTAATATAATAGCAAAACTGAGTTTTTTAATGATGTCCTCGGTTCATTGTATTAAAAGAGTTTCTCATTGGAAAAAATATGGGAA
AC004740 TTGCATG
AC048382 AAGATGGTATTTCCAAGCCTCATG
AP001029 CAATAACAAAAACAGCATG
AC010749
AC069023 TTTATTCTATGGACATTAGGGAGCCAGTCACTCTTGAGAGGCAGGTTAGTGTTGAAGTAGAGAGCTGGAATGGCTGGGGGACAAACGGTCTCCAAGGAGC
AC002980 CATAATATGGTGAAGATCCCTCTTACGCTACAAACAAATTAAATTAAACAACAACAACAAAA
AC116311 CATAATATGGTGAAGATCCCTCTTACGCTACAAACAAATTAAATTAAACAACAACAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAA
AP001604 AAGCCAAGTGCCAGATTTTGGTTAAAGTATATAAGACACAGACAGGCTTCTGGCCAAAGATACAGAGAAGCAGGAATAAAAGAAAAACAAATGTAAATAA

AL118519
AC010387 GCATTATATGAATGTAAATTTGAGCATATAAAATTGATAGATCCATTATAAGATTAAT
AC004740
AC048382
AP001029
AC010749
AC069023 CTCCGCCACCATAGCCACATG
AC002980
AC116311 TAATGTAACTAAGGAGGTGAACGATTTCTACAAGAAAAACTACAAAACACTGAAAGGAATTATAGATGACACAAATGGTAAAATATCCCATGCTCCTACAT
AP001604 AGAATGACCAGTATAGAATTCAAAAATAAAAAAGTAAAACGAGAGAAAATGTGCCACACCCAGAGCCCTTCTGGAAGTTTTCAAACCCAGAATTCTACTGA

AL118519
AC010387
AC004740
AC048382
AP001029
AC010749
AC069023
AC002980
```

```
AC116311 TAGAATAATTAAT
AP001604 GGGAAATGGAAATAGTTATAACATATTTCTAAACCCCGCTTGTCATAGTTGATT GGGCTAGGAGTGAGCTGAGCATCTAATCCAAACCTAACAAACTCTAG


AL118519
AC010387
AC004740
AC048382
AP001029
AC010749
AC069023
AC002980
AC116311
AP001604 GGTGACCAGAATTCTATGGAGTAC
```

Supp. Figure S5. Transduction Family: AC002980. Alignment showing L1 transduced sequences, polyadenylation signals and TSDs. The location of primers and restriction enzymes used in TS-ATLAS are also shown. All L1 sequences shown were recovered using TS-ATLAS.

Supp. Figure S6. Transduction Family: RP. Alignment showing L1 transduced sequences, polyadenylation signals and TSDs. The location of primers and restriction enzymes used in TS-ATLAS are also shown. AC093861 and AC005935 were present within the human genome assembly and AL590011 was recovered from fosmids. AL050308 and AC05888 were recovered using TS-ATLAS.

Supp. Figure S7

```
AC068286
AC091138
AL353685
AL592182
BX927359   GCTTGGGAGTGGGGCAATGGGGAGCGACTGTGTCATGAACACAGGACTTCTGTTTAGGGTGGTGATCATGCTTCGGAGCTCGGTGGTGGTGCTGGCTGCACAGCACTGTGAATATACCAA


AC067958
AL031548
AC068286
AC091138
AL353685
AL592182
BX927359   ATGCCACTGAATT
```

Supp. Figure S7. Transduction Family: LRE3. Alignment showing L1 transduced sequences, polyadenylation signals and TSDs. The location of primers and restriction enzymes used in TS-ATLAS are also shown. With the exception of AL353685, all L1 sequences were recovered using TS-ATLAS.

Supp. Table S1. Oligonucleotides

Library Construction	Sequence
RBMSL2	GTGGCGGCCAGTATTCGTAGGAGGGCGCGTAGCATAGAACG
RBMSL3	GTGGCGGCCAGTATTCGTAGGAGGGCGCGTAGCATAGAACGCATG
RBD3	TACGTTCTATGCTAC
RBD4	CGTTCTATGCTACG
RBD5	CGCGTTCTATGCTAC
RBX4	GTGGCGGCCAGTATTC
RBY1	GAGGGCGCGTAGCATAGAAC
RB3PA1	ATACCTAATGCTAGATGACACA
RB980TDA2	CAAATTTGTGTACGTTAAATATGTGAAAC
RB980TDA3	TGCTGGTTACACCTCAATAAAGC
CM958TD1	AGAAAAGCAAAATGTCTATTCCG
RB011TD1	AAAAAAAAAAAAAAAAAAAAAGTTTTAAATTT

Genotyping	Sequence
RB5PA2	TGGAAATGCAGAAATCACCG
RB3PA2	ACCTAATGCTAGATGACACA
Lineage 2980	
RB980A	GGCTGTGGAGAATGCAATTGTAAG
RB980B	GCTCTATTCCCAAGGCCTAGAACA
CM1604A	GAAACCCAACCTCAACGAAA
CM1604B	ATGAACTGGTGGGAATTTGG
PC4740A	CACACCACTGGAGAGATACGCTTT
PC4740B	CACTTGACTTCTCCCAGCTTTCTG
CM0387A	TTGCATTACTTGCTTGAAATTGA
CM0387B	TGCAGAAGGCCTTACGTTTT
CM8382A	ACCTCTCACCACTCCACCAC
CM8382B	CACTGGACAGGCAGAAACAA
CM1029A	CAGCTCAATTCTGGTGGTTG
CM1029B	TTTCTGGTGACAAAGCTTCAGA
SL0749A	CACAGGTTCCTTCTTCTTTTAATCA
SL0749B	TCTCTTGGCCATCAGAAACC
CM6311A	CACAAACCAATATTCCTGACGG
CM6311B	TTTCTGTTGCATTAATTGCTCAAT
Lineage LRE3	
CM958A	GAGGCCATAAATCCCCACAT
CM958B	TGTGGAGTGTTTCTCAAACTTTTT
CM286A	TCCTGAACAACTAATGGGTCAAT
CM286B	CTTGCTCTACCTCTCAACTTTATTGAA
CM7359A	TCCTCACGCACCACACAC
CM7359B	TGCTGTCCTTCTCCTCCTTC
CM1138A	GCAGGAAGAGGGGAATAAGG
CM1138B	TTGAGCTCCCCAGATGAAAG
CM1584C	CACACACGCACAGAGGAAAC
CM1584D	TCATTTCCCGTTAAGAACTGTC
CM2182A	CAGATTGTGATAAGGGATAAGAAAAA
CM2182B	GTCAGAGGATGGGGATAGAATG
Lineage RP	
CM011A	TCTGCGGCTTCCTGATTGAG
CM011B	TGGAATGCCCCTCAAAACAA
CM0308A	GACTCTTTCAGTTGCCAGATGC
CM0308B	CCAGTGTAAAAAGATGCGGCT
CM5939A	CTGGAGAGCACGTTCAAACA
CM5939B	GTGCAGGTGTGTAGGTGTGG

CM5888A	TCTGCTGTGCTTTTGCATTC
CM5888B	TCAATGAGCCTCTCCCATTC

Supp. Table S1. ***continued***

AL050308 Amplification	Sequence
CM0308A	GACTCTTTCAGTTGCCAGATGC
JM0308D	TTTGGATTAAAAAGTTTTAAATTGGGGG

Supp. Table S2. Naming of previously described L1s

Named Here (lineage)	Named Previously	Previously Reported
AL118519 (2980)	L1HS169	(Myers et al., 2002)
AC002980 (2980)	AC002980	(Boissinot et al., 2000) (Goodier et al., 2000)
	L1HS28	(Myers et al., 2002)
AC005939 (RP)	L1HS58	(Myers et al., 2002)
AP001029 (2980)	L1HS547 (AC010966)	(Myers et al., 2002)
	3-39	(Beck et al., 2010)
AC019288 (RP)	3-31	(Beck et al., 2010)
AC004740 (2980)	2-53	(Beck et al., 2010)
AL353685 (LRE3)	1-5	(Beck et al., 2010)
AL059011 (RP)	1-3	(Beck et al., 2010)

Supp. Table S3. Transduction lineage elements identified in the 1000 Genomes (Pilot 1-3) datasets

Element (lineage)	5′ Supporting reads	3′ Supporting reads
AL050308 (RP)	SRR013212.480276	SRR003250.1197183
	SRR013225.652575	
	SRR004906.392373	
	SRR004997.233490	
AC005888 (RP)	SRR006365.408575	SRR001480.201107
AC019288 (RP)	SRR003245.794304	
AC004740 (2980)	SRR006852.240011	SRR007034.92258
AC048382 (2980)	SRR005985.3024	
AP001029 (2980)	SRR003796.581896	
	SRR003691.520238	
	SRR003691.496284	
AC010749 (2980)	SRR005998.51014	SRR005990.872145
	SRR005998.274132	SRR006505.98087
	SRR012163.14231	SRR005846.277534
	SRR006498.1079198	SRR002441.10489
	SRR006403.222421	SRR004438.356014
	SRR005847.1133907	
	SRR005847.1048525	
	SRR005847.1179564	
	SRR002559.66275	
AC067958 (LRE3)	SRR006499.541055	
	SRR006505.1078808	
	SRR003619.389189	

Supp. References

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061-73.

Beck C.R., Collier P., Macfarlane C., Malig M., Kidd J.M., Eichler E.E., Badge R.M. and Moran J.V. 2010. LINE-1 retrotransposition activity in human genomes. Cell, 141, 1159-1170.

Boissinot S., Chevret P. and Furano A.V. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol, 17, 915-28.

Brouha B., Meischl C., Ostertag E., de Boer M., Zhang Y., Neijens H., Roos D. and H. K.H.,Jr. 2002 .Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet, 71, 327-36.

Goodier J.L., Ostertag E.M. and H. K.H.,Jr. 2000 Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet, 9, 653-7.

Kimberland M.L., Divoky V., Prchal J., Schwahn U., Berger W. and H. K.H.,Jr. 1999. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum Mol Genet, 8, 1557-60.

Myers J.S., Vincent B.J., Udall H., Watkins W.S., Morrish T.A., Kilroy G.E., Swergold G.D., Henke J., Henke L., Moran J.V. et al. 2002. A comprehensive analysis of recently integrated human ta L1 elements. Am J Hum Genet, 71, 312-26.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol. 1-2:203-14.