Supplementary Materials and Methods

Identification of human-specific mobile elements (HS-MEs): BLAT-based method (Fig. 1): In this method, DNA sequences covering the MEs and their flanking regions were retrieved from the human genome sequences based on the MEs’ genomic coordinates from the input list. For each ME, three sequences were used to detect its presence or absence in the corresponding regions of the out-group genomes. Sequence 1 (S1) consists of 200 bp sequences from joining the 100 bp sequences flanking from each side of the ME as one to represent the pre-integration sequence (in most cases contains two copies of the TSD); Sequence 2 (S2) consists of 200 bp sequences with 100 bp into the 5’ flanking region and 100 bp into 5’ end of the ME to represent the 5’ junction area of the ME; Sequence 3 (S3) is similar to sequence 2, but is for the 3’ junction area of the ME. Each of these 3 sequences was aligned against each of the 9 out-group genome sequences using BLAT with a species-specific set of empirically determined blat parameters (e.g., minScore and minIdentity) based on the use of a training dataset of known polymorphic MEs documented in dbRIP (see details in later section of this document). The BLAT results of the three sequences for each ME were grouped together based on their genomic positions. For a sequence with multiple mapping locations, the one that provided the best blat score and locates into the same approximate locations with the other two sequences in the out-group genomes was selected. Those with two junction sequences mapped to different locations were discarded as they likely represent random matches to non-orthologous regions. If the blat result for a sequence meet the minimal blat score and percentage of identity, coverage of the query sequences, and span length in the genome, which were determined and optimized based on the training dataset, a “+” sign was given for this sequence to indicate the presence of an orthologous sequence in the out-group genomes. Otherwise a “-” sign was assigned to indicate the absence of an orthologous
sequence in each out-group genome. A typical pattern for a HS-ME would have a pattern of “+/−/−” for the 3 sequences in the order of sequence 1, 2, and 3, representing the pre-integration, the 5’- and 3’-ME junction sequences, respectively, while the pattern for a non HS-ME (i.e., shared) would be “+/+/+”. A HS-ME pattern is provided only if all out-group genomes show support a HS-ME.

LiftOver-based method (Fig. 1): LiftOver is a tool originally developed by the UCSC genome team (http://genome.ucsc.edu/cgi-bin/hgLiftOver) to allow conversion of genome coordinates between different assemblies of the same genome. But its use has also been extended for finding corresponding/orthologous regions between genomes of closely related species. In this method, the genomic coordinates for the three regions of a ME, corresponding to 100 bp of the 5’ flanking (R1), the ME (R2), and the 100 bp 3’ flanking regions (R3), were liftOver’ed onto each of the out-group genomes using a command line version of the liftOver tool with the related chain files between hg19 and each of the out-group genomes. The best hit, which is the one with an identifiable orthologous region closest in length to the human query sequence, for each of the three sequences among the four out-group genomes was used. A query sequence is marked as “+” for having a qualified hit in any out-group genome or as “−” otherwise. The pattern for a typical HS-ME is “+/−/+” for the three regions corresponding to 5’ flanking, ME, and 3’ flanking, respectively, while the pattern for a typical non-HS-REs should be “+/+/+”.

The final list of HS-MEs was generated by combining the results of the blat-based and liftOver-based methods. HS-MEs were then divided into four categories to reflect their types and confidence levels. Category I HS-MEs are those supported by both methods to be HS-ME with at least one of the two flanking sequences present and the ME absent in all outgroup genomes. HS-
MEs with orthologous counterpart for both flanking sequences has the highest confidence. Entries with one only one flanking sequences missing in the out-group genomes can be caused by the presence of transduced sequences or the presence of another HS-ME, and they are candidates for identifying transduction as described later. Category II HS-MEs are those with the MEs and both flanking sequences missing from all out-group genomes (i.e., a pattern of “-/-/-” by both methods). These mostly represent HS-MEs in HS-MEs or with their orthologous regions in the out-group genomes falling into sequence gaps. To exclude the latter cases, we removed those with the closest identifiable chimp orthologous regions located next to a sequence gap.

**Determination of optimal BLAT and liftOver criteria using dbRIP data as test datasets:** We used a training dataset consisting of half of the known polymorphic retrotransposon insertions (RIPs) documented in dbRIP to identify the optimal parameters for the blat and liftOver methods described above for using with each of the outgroup genomes. RIPs are results of recent retrotransposition events and show as polymorphic for its presence or absence among the human populations. Therefore, they are considered human specific, and were used to train our algorithms to identify an optimal set of criteria that gave an error (only false negative in this case) rate no high than 5% (i.e. sensitivity >=95%). A total of four criteria were used for the BLAT based method: blat score, identity, coverage and span length. The score equals the number of matches minus numbers of mis-matches and gaps in the target sequences (Score = match – mis-match – T gap count). A higher score means a better match. The identity is a percentage of the similarity between the query sequence and the target sequence. It is calculated as $100 \times \frac{(\text{match} + \text{repeat match} (\text{Rep. match}))}{(\text{match} + \text{mismatch} + \text{Rep.match})}$. Span and coverage are used to provide additional constrains about the match. Span is the total length of the target
sequence (Span = Target end – Target start). Since BLAT tolerates gaps in the target sequence, it is able to report matches consisting of fragments with gaps in between. So there is possibility that an entry could have both high score and identity while having a long gap in the target sequence. Similarly, coverage is used to monitor the quality of match for the query sequence: Coverage = (Q end – Q start). The liftOver utilizes a criterion called minMatch, which represents the minimum ratio of bases that must map on to the new genome. For example, if the minMatch ratio is set to be 0.5, liftOver would report a positive hit when 50% of the region can be mapped on to the test genome.

**Identification of insertion-mediated transductions and deletions:** The target site duplications (TSDs), as well as transduction and RIMD for all HS-MEs were identified using in-house perl scripts incorporating the utility of the NCBI bl2seq and UCSC liftOver as described. Pre-integration site sequences retrieved from outgroup genomes were aligned against sequences consisting of the ME + both flanking sequences from human genome using blast. An in-house perl script is used to identify candidates for transduction and insertion mediated deletion. For identifying transduction, MEs were considered as candidates if they have (1) presence of TSDs and (2) extra sequences between the TSD and the ME. The extra sequence was then searched against the human reference genome using blat (with the similarity cutoff set as >=95%). The qualified candidates were then manually verified before they were included in the transduction list. For identifying insertion mediated deletion, a ME was considered as candidates if it had (1) absence of TSDs and (2) extra sequence at the pre-integration site in the outgroup genomes. The pre-integration sequence retrieved from the closest available outgroup genome was then used for BLAT searches (similarity cutoff set as >=90%) against other outgroup genomes. If the same
extra sequence was seen in at least 2 out-group genomes, then the ME was included in the insertion mediated deletion list after manually verification.

**PCR verification of HS-MEs:** We performed PCR analysis to verify the human-specific MEs. Genomic DNA from *Homo sapiens* (human; NA10851, Coriell Cell Repository, Camden, NJ), *Pan troglodytes* (common chimpanzee), *Gorilla gorilla* (gorilla), and *Pongo pygmaeus* (Bornean orangutan) were used as a DNA template for each PCR reaction. Genomic DNA for three apes was kindly provided by Dr. Takenaka (Primate Research Institute, Kyoto University). PCR oligonucleotide primers were designed using the software Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). Several ME loci were failed to amplify due to variable sequences in flanking sequences. Thus, we manually designed new primers for amplification of these MEs. The primers are listed in Table S3. Each PCR amplification performed in 25 µl reaction using 10 µl of a DNA polymerase kit (SolGent 2X EF-Taq, SolGent, South Korea), 2.5 µl oligonucleotide primer (4pmol/µl), 5 µl of distilled water, and 2 µl of template DNA (10-20 ng/µl). The PCR conditions were an initial denaturation at 95°C for 5 min, followed by 35 cycles of 30 sec of denaturation at 95°C, 40 sec of annealing temperature, and 1 to 5 min of extension (depending on the expected size of PCR product) at 72°C followed by a final extension at 72°C for 10 min. Five microliters of the resulting PCR products were loaded on 1% agarose gel, stained with ethidium bromide, and visualized using UV fluorescence. In a few cases, in which the expected size of PCR product is longer than 3 kb, we alternatively used Ex TaqTM polymerase (TaKaRa Japan), KOD FX (Toyobo, Japan) in order.
Genome plots of HS-MEs: The HS-MEs were plotted onto the human chromosomes with the cyto-band ideogram (based on GRCh38/hg38) based on the cyto-band data downloaded from the UCSC genome browser with the use of an in-house perl script.
Table S1: The number of MEs in different versions of the human reference genomes

| ME type*  | GRCh 35/hg17 (2,851,331,196 bp#) | GRCh 37/hg19 (2,861,327,216 bp#) | GRCh 38/hg38 (2,937,641,526 bp#) |
|-----------|-------------------------------|-------------------------------|-------------------------------|
|           | raw counts | integrated counts | total size | % of genome | raw counts | integrated counts | total size | % of genome | raw counts | integrated counts | total size | % of genome |
| DNA       | 384,847     | 305,949           | 85,545,893 | 3.0%      | 450,267     | 368,746           | 97,553,533 | 3.4%      | 479,941     | 395,965           | 101,978,857 | 3.5%       |
| L1        | 912,769     | 522,014           | 500,573,594 | 17.6%    | 938,484     | 535,248           | 513,383,945 | 17.9%    | 962,085     | 564,195           | 522,364,326 | 17.8%      |
| Alu       | 1,173,780   | 1,120,316         | 307,699,723 | 10.8%   | 1,175,329   | 1,122,711         | 307,920,501 | 10.8%   | 1,181,072   | 1,132,541         | 308,374,836 | 10.5%      |
| SVA       | 3,347       | 2,997             | 4,015,581  | 0.1%     | 3,608       | 3,028             | 4,015,157  | 0.1%     | 5,397       | 4,933             | 4,228,693  | 0.1%       |
| LTR       | 654,898     | 453,887           | 249,382,220 | 8.7%    | 698,594     | 479,372           | 262,989,620 | 9.2%    | 709,475     | 488,208           | 265,865,981 | 9.1%       |
| Others    | 1,858,138   | 1,625,340         | 243,598,757 | 8.5%    | 1,939,555   | 1,725,401         | 262,620,207 | 9.2%    | 1,949,763   | 1,733,490         | 328,324,214 | 11.2%      |
| Total     | 4,987,779   | 4,030,503         | 1,390,815,768 | 48.8%   | 5,205,837   | 4,234,506         | 1,448,482,963 | 50.6%   | 5,287,733   | 4,319,332         | 1,531,136,907 | 52.1%      |

#All genome sizes are for non-gap sequences; *All uncertain entries are excluded(e.g. those designated as "DNA?", "Unknown")
Table S2: Comparison of HS-MEs with results from prior studies

| ME type | Mills et al 2006 | Tang et al 2018 | shared MEs (original*) | unique to Mills et al 2009 | unique to Tang et al 2018 |
|---------|------------------|-----------------|------------------------|---------------------------|--------------------------|
| Alu     | 5530             | 8817            | 4748(4695)             | 782                       | 4069                     |
| L1      | 1174             | 3912            | 1040(1033)             | 134                       | 2872                     |
| SVA     | 865              | 1571            | 833(823)               | 32                        | 738                      |
| LTR     | 170              | 530             | 117(116)               | 53                        | 413                      |
| Other   | 47               | 0               | 0(0)                   | 47                        | 0                        |
| Total   | 7786             | 14830           | 6738(6667)             | 1048                      | 8092                     |

*number of shared MEs with Mills et al 2009 before adding the false negatives
### Table S3: PCR primers and results for experimental validation of HS-MEs

| Position          | ID       | Strand | Family | Type | ME   | Category | Experiment | Forward Primer                      |
|-------------------|----------|--------|--------|------|------|----------|------------|-------------------------------------|
| chr5:33122370-33128395 | 03099614 | +      | L1PA2  | LINE | L1   | category I | HS         | ATTAAGCAAAGTGTTAGGTGC                 |
| chr4:87347103-87353146 | 02915560 | -      | L1HS   | LINE | L1   | category II | HS         | GCTGGTACTAAAGTAGACCC                 |
| chr3:58814308-58816278 | 02585297 | +      | L1PA2  | LINE | L1   | category II | HS(internal primer fail)             | AGCCTGACTACCTGTTATGC                 |
| chr1:84052342-84058406 | 00151933 | -      | L1HS   | LINE | L1   | category II | Fail       | AGTCTGCTTCACATATACAGCC               |
| chr12:88708857-88714885 | 00907324 | -      | L1PA2  | LINE | L1   | category II | Fail       | CCTGTTCATAATGCCCCAATACA              |
| chrY:22475174-22476068 | 04449852 | +      | MER11D | LTR  | ERVK | category II | HS         | CAACACACTGGACTAGATTC                |
| chr7:23039855-23040823 | 03595003 | +      | LTR5_Hs| LTR  | ERVK | category II | HS         | CACTTAAGACCCCTGTCTCCCC              |
| chr3:195927524-195928492 | 02783330 | -      | LTR5_Hs| LTR  | ERVK | category II | HS         | GAAGAAGAACAAGCATAGAGC               |
| chr7:23637166-23638734 | 03596147 | +      | SVA_E  | Other | Other | category II | HS         | CAGTGACAGACGGTTATGGG                |
| chr2:191397192-191399262 | 02184897 | -      | SVA_E  | Other | Other | category I  | HS         | GGAAAGTGTCAGAACAGGC                |
| chr17:19627119-19628872 | 01579849 | -      | SVA_F  | Other | Other | category I  | HS         | ACGGATAGAAGAACAGCCGCTG              |
| chr11:77489410-77491145 | 00690834 | +      | SVA_D  | Other | Other | category I  | HS         | CTAATATACCTGACAGCCAGGC             |
| chr1:112897624-112899310 | 00194215 | -      | SVA_F  | Other | Other | category I  | HS         | GTCAAGTTTGGCTTCTTAAACC             |
| chr6:61183414-61183725 | 03410386 | -      | Alu_Ya5| SINE | Alu  | category II | HS         | TCAATGCTGGTTTCAAGG                 |
| chr19:47055494-47055803 | 01884351 | -      | Alu_Y  | SINE | Alu  | category I  | No HS      | GGCCTGACAATTGAACCTG                |
| Reverse Primer | Internal Primer | Annealing Temperature | Product sizes(+/−) |
|---------------|----------------|-----------------------|-------------------|
| AGTAAGGCCCTGCTGAATGGG | Int2R:GCGTCCGTCACCCCTTTCTT | 55°C/55°C | 6731bp(661bp)/438bp |
| TGTGCTAAGCTGGGTGTGCC | Int3F:CAAAGACTTGGAAACCAAACC | 55°C/59°C | 6730bp(788bp)/685bp |
| CTATGTTCTGTCAGCAG | N/A | 53°C | 8972bp/689bp |
| CTCATTTTGAAAGCTACAAGGG | IntF:ATTGTGGAAGTCAGTGTGC/IntR:GTTIACCTAAGCAAGCCTGGG | 51°C | N/A |
| TTGGGATTCTTTGAACCTAGC | N/A | N/A | N/A |
| GAGCTGGGAATGTTTAATG | N/A | 52°C | ~3500bp/~1000bp |
| CACCAAAACAAATCCACTGCC | Int2F:ATACTAAGGGAACTCAGGAGC | 60°C/58°C | 9753bp(2695bp)/2715bp |
| GAATGGGTGTACCTGGAC | N/A | 52°C | 1425bp/451bp |
| TTCCCGTTACACCTGCC | N/A | 57°C | 3574bp/543bp |
| AACCATCTTGAGCGATAACC | N/A | 56°C | 2706bp/506bp |
| CCTCATATTGGAACATTCCTG | N/A | 55°C | 2671bp/744bp |
| TCTCACCTGATCCTGAGGCC | N/A | 50°C | 2436bp/593bp |
| CTITTCTGACCTTGCACCC | N/A | 55°C | 2766bp/871bp |
| ATVAAAGTCCATATGGCTCTTC | N/A | 55°C | 871bp/549bp |
| CATCTGGTAAATGGCTCC | N/A | 57°C | 1198bp |
### Table S4a: Sources for novel HS-MEs (each source considered independently)

| ME type | Extra in hg38 (vs. hg17) | ME integration | MEs in MEs | non-canonical MEs* | Multiple genomes |
|---------|--------------------------|----------------|------------|-------------------|------------------|
| L1      | 54                       | 670            | 1088       | 1598              | 347              |
| Alu     | 86                       | 49             | 2138       | 953               | 326              |
| SVA     | 13                       | 113            | 326        | 444               | 34               |
| LTR     | 8                        | 166            | 192        | 261               | 115              |
| Total   | 161                      | 998            | 3744       | 3256              | 822              |

### Table S4b: Sources for novel HS-MEs (non-redundant when considered step-wise in the given order)

| ME type | Extra in hg38 (vs. hg17) | ME integration | MEs in MEs | non-canonical MEs* | More primate genomes | Others | Total | % HS-MEs |
|---------|--------------------------|----------------|------------|-------------------|-----------------------|--------|-------|---------|
| L1      | 54                       | 658            | 836        | 793               | 37                    | 494    | 2872  | 73%     |
| Alu     | 86                       | 40             | 2091       | 479               | 55                    | 1318   | 4069  | 46%     |
| SVA     | 13                       | 112            | 267        | 186               | 5                     | 112    | 695   | 44%     |
| LTR     | 8                        | 162            | 119        | 82                | 4                     | 38     | 413   | 78%     |
| Total   | 161                      | 972            | 3313       | 1540              | 101                   | 1962   | 8049  | 54%     |

*Non-canonical MEs include MEs with transductions, RMID, and no TSDs
| Family | Subfamily* | Total copies | HS-MEs | HS%  |
|--------|------------|--------------|--------|------|
| Alu    | AluYa5     | 3,861        | 3,007  | 77.9%|
|        | AluYb8     | 2,828        | 2,108  | 74.5%|
|        | AluYb9     | 327          | 240    | 73.4%|
|        | AluYd8     | 237          | 154    | 65.0%|
|        | AluYg6     | 835          | 363    | 43.5%|
|        | AluYi6     | 455          | 164    | 36.0%|
|        | AluYk12    | 201          | 68     | 33.8%|
|        | AluYa8     | 343          | 103    | 30.0%|
|        | AluYe5     | 1,318        | 269    | 20.4%|
|        | AluYi6 4d  | 149          | 20     | 13.4%|
|        | AluYh7     | 153          | 16     | 10.5%|
|        | AluYc3     | 543          | 32     | 5.9% |
|        | AluYk11    | 1,256        | 68     | 5.4% |
|        | AluYk4     | 1,010        | 29     | 2.9% |
|        | AluYh3     | 2,627        | 74     | 2.8% |
|        | AluYe6     | 194          | 5      | 2.6% |
|        | AluYc      | 4,521        | 107    | 2.4% |
|        | AluYh9     | 142          | 3      | 2.1% |
|        | AluYh3a3   | 313          | 5      | 1.6% |
|        | Alu        | 4,280        | 65     | 1.5% |
|        | AluYk3     | 1,152        | 17     | 1.5% |
|        | AluY      | 102,844      | 1,442  | 1.4% |
|        | AluYj4     | 3,487        | 38     | 1.1% |
| ERV    | ERVK       | 7,369        | 217    | 2.9% |
|        | ERV1       | 103,982      | 204    | 0.2% |
| L1     | L1HS       | 1,346        | 991    | 73.6%|
|        | L1PA2      | 4,096        | 1,739  | 42.5%|
|        | L1P1       | 2,851        | 211    | 7.4% |
|        | L1PA3      | 8,780        | 269    | 3.1% |
|        | L1PA8      | 6,541        | 131    | 2.0% |
|        | L1P2       | 1,231        | 22     | 1.8% |
|        | L1P        | 140          | 2      | 1.4% |
| SVA    | SVA_D      | 1,325        | 895    | 67.5%|
|        | SVA_F      | 821          | 418    | 50.9%|
|        | SVA_E      | 595          | 156    | 26.2%|
|        | SVA_C      | 418          | 17     | 17.0%|
|        | SVA_A      | 1,001        | 19     | 1.9% |
|        | SVA_B      | 768          | 12     | 1.6% |
| All    | All        | 2,194,815    | 14,870 | 0.7% |

*Including subfamilies with top two activities in the family or activities at 1% or more
Table S6a: Ratio of all MEs in MEs by ME class

| TE type | MEs in MEs | All MEs | MEs-in-MEs/All MEs |
|---------|------------|---------|--------------------|
| LINE    | 172773     | 969873  | 17.8%              |
| DNA     | 88614      | 399590  | 22.2%              |
| LTR     | 139906     | 496946  | 28.2%              |
| SINE    | 526647     | 1689416 | 31.2%              |
| SVA     | 1799       | 4933    | 36.5%              |

Table S6b. Ratios of HS-MEs in MEs by ME class

| ME type | HS-MEs-in-MEs | Total HS-MEs | MEs-in-MEs% |
|---------|---------------|--------------|-------------|
| L1      | 1388          | 3912         | 35.5%       |
| LTR     | 187           | 530          | 35.3%       |
| Alu     | 4086          | 8817         | 46.3%       |
| SVA     | 640           | 1571         | 40.7%       |

Table S6c. The densities of HS-MEs in MEs by ME type (#HS-MEs/Mbp host MEs)

| HS-ME | LINE | DNA | LTR | SINE | SVA | All MEs |
|-------|------|-----|-----|------|-----|---------|
| L1    | 1.7  | 0.9 | 0.9 | 0.3  | 0.0 | 1.0     |
| LTR   | 0.2  | 0.1 | 0.4 | 0.0  | 0.0 | 0.2     |
| Alu   | 4.7  | 3.6 | 2.5 | 1.4  | 0.2 | 3.1     |
| SVA   | 0.7  | 0.4 | 0.3 | 0.1  | 12.7| 0.5     |
| Total | 7.3  | 5.0 | 4.0 | 1.8  | 13.0| 4.8     |
Table S7: Chromosome distributions of HS-MEs by ME type

| Chr  | Chr length (Mb) | All MEs | MEs/ Mb | ALL HS-ME | HS-ME/ 500kb | HS/1K RE | Gene | Gene/ Mb | HS- Alu | HS- Alu/ Mb | HS-L1 | HS-L1/2 MB | HS-SVA | HS-SVA/ 10Mb | HS-LTR | HS-LTR/ 2Mb |
|------|-----------------|---------|---------|-----------|--------------|---------|------|---------|---------|-----------|------|------------|--------|--------------|---------|--------------|
| chr1 | 230.5           | 175319  | 760.7   | 1147      | 5.0          | 6.5     | 5620 | 24.4    | 667     | 2.9        | 297  | 2.6        | 159   | 6.9          | 24     | 0.2          |
| chr2 | 240.5           | 170039  | 706.9   | 1190      | 4.9          | 7.0     | 4264 | 17.7    | 735     | 3.1        | 310  | 2.6        | 128   | 5.3          | 17     | 0.1          |
| chr3 | 198.1           | 138120  | 697.2   | 1036      | 5.2          | 7.5     | 3230 | 16.3    | 633     | 3.2        | 268  | 2.7        | 113   | 5.7          | 22     | 0.2          |
| chr4 | 189.8           | 126991  | 669.2   | 1025      | 5.4          | 8.1     | 2699 | 14.2    | 619     | 3.3        | 321  | 3.4        | 68    | 3.6          | 17     | 0.2          |
| chr5 | 181.3           | 123664  | 682.2   | 970       | 5.4          | 7.8     | 3086 | 17.0    | 597     | 3.3        | 256  | 2.8        | 87    | 4.8          | 30     | 0.3          |
| chr6 | 170.1           | 118657  | 697.7   | 902       | 5.3          | 7.6     | 3160 | 18.6    | 531     | 3.1        | 257  | 3.0        | 87    | 5.1          | 27     | 0.3          |
| chr7 | 159.0           | 124355  | 782.3   | 794       | 5.0          | 6.4     | 3084 | 19.4    | 505     | 3.2        | 188  | 2.4        | 84    | 5.3          | 17     | 0.2          |
| chr8 | 144.8           | 101910  | 704.0   | 692       | 4.8          | 6.8     | 2507 | 17.3    | 417     | 2.9        | 196  | 2.7        | 54    | 3.7          | 25     | 0.3          |
| chr9 | 121.8           | 91710   | 753.0   | 635       | 5.2          | 6.9     | 2443 | 20.1    | 405     | 3.3        | 144  | 2.4        | 68    | 5.6          | 18     | 0.3          |
| chr10| 133.3           | 100918  | 757.3   | 564       | 4.2          | 5.6     | 2433 | 18.3    | 350     | 2.6        | 137  | 2.1        | 67    | 5.0          | 10     | 0.2          |
| chr11| 134.5           | 92192   | 685.3   | 667       | 5.0          | 7.2     | 3434 | 25.5    | 382     | 2.8        | 184  | 2.7        | 75    | 5.6          | 26     | 0.4          |
| chr12| 133.1           | 104467  | 784.7   | 625       | 4.7          | 6.0     | 3104 | 23.3    | 382     | 2.9        | 155  | 2.3        | 74    | 5.6          | 14     | 0.2          |
| chr13| 98.0            | 66141   | 675.0   | 507       | 5.2          | 7.7     | 1464 | 14.9    | 340     | 3.5        | 113  | 2.3        | 39    | 4.0          | 15     | 0.3          |
| chr14| 90.6            | 67930   | 750.0   | 423       | 4.7          | 6.2     | 2341 | 25.8    | 243     | 2.7        | 120  | 2.6        | 48    | 5.3          | 12     | 0.3          |
| chr15| 84.6            | 66556   | 786.3   | 370       | 4.4          | 5.6     | 2310 | 27.3    | 225     | 2.7        | 92   | 2.2        | 48    | 5.7          | 5      | 0.1          |
| chr16| 81.8            | 77620   | 948.8   | 344       | 4.2          | 4.2     | 2639 | 32.3    | 210     | 2.6        | 78   | 1.9        | 46    | 5.6          | 10     | 0.2          |
| chr17| 82.9            | 79545   | 959.3   | 350       | 4.2          | 4.4     | 3153 | 38.0    | 212     | 2.6        | 49   | 1.2        | 79    | 9.5          | 10     | 0.2          |
| chr18| 80.1            | 52029   | 649.6   | 376       | 4.7          | 7.2     | 1227 | 15.3    | 238     | 3.0        | 110  | 2.7        | 22    | 2.7          | 6      | 0.1          |
| chr19| 58.4            | 72255   | 1236.4  | 280       | 4.8          | 3.9     | 3098 | 53.0    | 141     | 2.4        | 45   | 1.5        | 68    | 11.6         | 26     | 0.9          |
| chr20| 63.9            | 52602   | 822.6   | 282       | 4.4          | 5.4     | 1496 | 23.4    | 152     | 2.4        | 72   | 2.3        | 54    | 8.4          | 4      | 0.1          |
| chr21| 40.1            | 28896   | 720.8   | 162       | 4.0          | 5.6     | 903  | 22.5    | 101     | 2.5        | 17   | 1.7        | 17    | 4.2          | 9      | 0.4          |
| chr22| 39.2            | 35007   | 894.0   | 128       | 3.3          | 3.7     | 1417 | 36.2    | 68      | 1.7        | 29   | 1.5        | 27    | 6.9          | 4      | 0.2          |
| chrX | 154.9           | 112141  | 724.0   | 689       | 4.4          | 6.1     | 2550 | 16.5    | 361     | 2.3        | 257  | 3.3        | 54    | 3.5          | 17     | 0.2          |
| chrY | 26.4            | 15751   | 596.3   | 672       | 25.4         | 42.7    | 601  | 22.8    | 303     | 11.5       | 199  | 15.1       | 5     | 1.9          | 165    | 12.5         |
| Genome| 2937.6         | 2194815 | 747.1   | 14830     | 5.0          | 6.8     | 62263| 21.2    | 8817    | 3.0        | 3912 | 2.7        | 1571  | 5.3          | 530    | 0.4          |
| ME type | protein-coding genes # | # in NR | total # in reference | protein-coding genes (bp) | total bp in reference |
|---------|-------------------------|---------|----------------------|---------------------------|-----------------------|
|         | CDS | 5'-UTR | 3'-UTR | CDS | 5'-UTR | 3'-UTR | CDS | 5'-UTR | 3'-UTR | CDS | 5'-UTR | 3'-UTR | bp | bp |
| Alu     | 1   | 5      | 41     | 81  | 128   | 94     | 639 | 11,298 | 13,133 | 25,164 |
| L1      | 5   | 3      | 4      | 45  | 57    | 319    | 892 | 1,997  | 8,414  | 11,622 |
| SVA     | 32  | 3      | 7      | 58  | 100   | 10,623 | 624 | 7,580  | 21,254 | 40,081 |
| LTR     | 2   | 0      | 1      | 16  | 19    | 225    | 0   | 105    | 6,878  | 7,208  |
| Total   | 40  | 11     | 53     | 200 | 304   | 11,261 | 2,155| 20,980 | 49,679 | 84,075 |
| RepeatMasker ID for HS-MEs | ME position     | ME strand | subfamily | Class       | family | ME length (bp) | CDS position | Gene name     | Transcript ID    | ME-CDS size (bp) |
|---------------------------|-----------------|-----------|-----------|-------------|--------|----------------|--------------|---------------|----------------|------------------|
| chr9:83049539-83055571    | +               | L1HS      | LINE      | L1          | chr9:83055548-83055683 | RASEF | ENSG00000165105.9 | 24           |
| chr5:133583288-133589299  | +               | L1HS      | LINE      | L1          | chr5:133583265-133583396 | FSTL4 | ENSG00000053108.16 | 109          |
| chr18:79114907-79116935   | -               | L1HS      | LINE      | L1          | chr18:79114994-79114999 | ATP9B | ENSG00000166377.19 | 6            |
| chr3:136963693-136969736  | +               | L1HS      | LINE      | L1          | chr3:136969273-136969378 | IL20RB | ENSG00000174564.12 | 106          |
| chr15:45597902-45600342    | +               | L1PA2     | LTR       | L1          | chr15:45597923-45597996 | BLOC1S6 | ENSG00000104164.10 | 74           |
| chr5:156657705-156666885  | -               | LTR5 Hs   | LTR       | ERVK        | chr5:156660389-156660435 | AC016577.1 | ENSG00000279646.1 | 47           |
| chr22:18938673-18947848    | +               | LTR5 Hs/HE | RVK-int   | LTR         | chr22:18946791-18946968 | AC007326.1 | ENSG00000279560.1 | 178          |
| chr12:119917797-119919714  | -               | SVA C     | Retroposon| SVA         | chr12:119918001-119918372 | AC004813.1 | ENSG00000279777.1 | 372          |
| chr8:128117822-128119386   | -               | SVA D     | Retroposon| SVA         | chr8:128118011-128118382 | FKSG59 | ENSG00000280151.1 | 372          |
| chr15:47131719-47134222     | +               | SVA D     | Retroposon| SVA         | chr15:47133655-47134026 | AC066661.5 | ENSG00000259752.2 | 372          |
| chr5:60430541-60432081      | -               | SVA D     | Retroposon| SVA         | chr5:60430738-60431109 | FKSG52 | ENSG00000280447.1 | 372          |
| chr16:70224194-70225526     | -               | SVA D     | Retroposon| SVA         | chr16:70224358-70224735 | FKSG63 | ENSG00000280252.1 | 378          |
| chr20:6355588-206356116     | -               | SVA D     | Retroposon| SVA         | chr20:6355739-206356116 | AC017081.1 | ENSG00000279921.1 | 378          |
| chr14:20505334-20507630     | +               | SVA D     | Retroposon| SVA         | chr14:20505941-20506019 | RNASE10 | ENSG00000182545.6 | 79           |
| chr6:31931262-31933153      | -               | SVA D     | Retroposon| SVA         | chr6:31932475-31932493 | C2    | ENSG00000166278.14 | 19           |
| chr14:22636438-22638669     | +               | SVA D     | Retroposon| SVA         | chr14:22638102-22638473 | AC243945.1 | ENSG00000279510.1 | 372          |
| chr7:16789065-16791103      | -               | SVA D     | Retroposon| SVA         | chr7:16789242-16789619 | AC073333.1 | ENSG00000280130.1 | 378          |
| chr19:40647367-40649204     | -               | SVA D     | Retroposon| SVA         | chr19:40647563-40647934 | FKSG66 | ENSG00000279183.1 | 372          |
| chrX:42285051-42286669      | -               | SVA D     | Retroposon| SVA         | chrX:42285229-42285600 | FKSG70 | ENSG00000279849.1 | 372          |
| chr14:77416262-77416906     | -               | SVA D     | Retroposon| SVA         | chr14:77416398-77416769 | FKSG61 | ENSG00000280308.1 | 372          |
| chr21:35897707-35898625     | +               | SVA D     | Retroposon| SVA         | chr21:35898058-35898429 | FKSG68 | ENSG00000280170.1 | 372          |
| chr5:16560188-65630430      | +               | SVA D     | Retroposon| SVA         | chr5:16563020-65630402 | LEPR | ENSG00000116678.18 | 201          |
| chr19:3537871-3539763       | -               | SVA D     | Retroposon| SVA         | chr19:3538044-3538415 | FKSG51 | ENSG00000280230.1 | 372          |
| chr15:40682074-40684259     | +               | SVA D     | Retroposon| SVA         | chr15:40683697-40684068 | AC022405.1 | ENSG00000279804.1 | 372          |
| chr3:130867338-130869442    | +               | SVA D     | Retroposon| SVA         | chr3:130868875-130869246 | AC055733.1 | ENSG00000280127.1 | 372          |
| chr19:52592779-52594353     | -               | SVA D     | Retroposon| SVA         | chr19:52594212-52594304 | ZNF83 | ENSG00000167766.18 | 93           |
| chr14:64894264-64896723     | -               | SVA E     | Retroposon| SVA         | chr14:64894339-64894710 | AL135745.1 | ENSG00000279654.1 | 372          |
### Table S10. Contribution of HS-MEs to transcription factor binding sites (TFBS)

| ME type | # TFBS site | # TF | # HS-ME | # total HS-ME | TFBS/ total HS-ME | % all HS-TFBS |
|---------|-------------|------|---------|---------------|--------------------|---------------|
| Alu     | 1,621       | 135  | 581     | 8,817         | 6.6%               | 53.5          |
| L1      | 690         | 114  | 311     | 3,912         | 7.9%               | 22.8          |
| SVA     | 504         | 109  | 197     | 1,571         | 12.5%              | 16.6          |
| LTR     | 217         | 58   | 78      | 530           | 14.7%              | 7.2           |
| Total   | 3,032       | 142  | 1,167   | 14,830        | 7.9%               | 100.0         |
**Fig. S1: PCR validation results 4 selected HS-MEs**

**A: HS_Alu:**
- chr6:62053989-62054300 (hg19);
- chr6:61183414-61183725 (hg38);

### PCR Mixture

|   |   |   |
|---|---|---|
| 1 | EF taq polymerase | 25 ul |
| 2 | Forward primer (4 pmol) | 7.5 ul |
| 3 | Reverse primer (4 pmol) | 7.5 ul |
| 4 | D.W | 10 ul |
| 5 | DNA template (10 ng/ul) | 2 ul |
| **Total** | | **52 ul** |

### PCR Condition (35 cycles)

|   |   |   |
|---|---|---|
| 1 | Pre-denaturation | 95°C | 5 mins |
| 2 | Denaturation | 95°C | 30 s |
| 3 | Annealing | 55°C | 30 s |
| 4 | Extension | 72°C | 1 min |
| 5 | Termination | 72°C | 5 mins |
| 6 | Hold | 4°C | ∞ |

### PCR primer information

|   |   |
|---|---|
| **Forward Primer** | **Reverse primer** |
| TCAATGCCTGGTTCAAAGG | GACAAAGTCTCAGCTTGGCTC |

### Expected PCR product size

|   |   |   |   |
|---|---|---|---|
| Human (empty size) | Chimpanzee | Gorilla | Orangutan |
| **871 bp (550 bp)** | 549 bp | 550 bp | ? |
**B: HS_L1:** chr4:87347103-87353146(hg19); chr4:87347103-87353146(hg38);

### Human Chimp Gorilla Orangutan LINE-1 HS Insertion

**PCR primer information**

|                 | Forward Primer | Reverse primer | Internal Primer |
|-----------------|----------------|----------------|-----------------|
| GCTGGTACTAAAGTAGACCC | TGTGCTAAGCTGGGTGTGGC | CAAAGACTTGGGAACCAACCC |

**Expected PCR product size**

|                  | Human (empty size) | Chimpanzee | Gorilla | Orangutan |
|------------------|--------------------|------------|---------|-----------|
| F/R              | 6730 bp (675 bp)   | 685 bp     | 687 bp  | 681 bp    |
| Int3F/R          | 788 bp             | -          | -       | -         |

### PCR Mixture

|   | EF taq polymerase | Forward primer (4 pmol) | Reverse primer (4 pmol) | D.W | DNA template (10 ng/ul) | Total |
|---|-------------------|-------------------------|-------------------------|-----|------------------------|-------|
| 1 |                   |                         |                         |     |                        | 25 ul |
| 2 |                   | 7.5 ul                  |                         |     |                        |       |
| 3 |                   |                         | 7.5 ul                  |     |                        |       |
| 4 |                   |                         |                         | 10 ul |                        |       |
| 5 |                   |                         |                         | 2 ul  |                        | 52 ul |

### PCR Condition (35 cycles) FR

|   | Annealing | Extension |
|---|-----------|-----------|
| 3 | 55°C      | 30 s      |
| 4 | 72°C      | 1 min     |

### PCR Condition (35 cycles) Int3FR

|   | Annealing | Extension |
|---|-----------|-----------|
| 3 | 59°C      | 30 s      |
| 4 | 72°C      | 1 min     |
**C: HS_HERV-K:** chr7:23079474-23080442(hg19); chr7:23039855-23040823(hg38);

- **PCR Mixture**
  1. EF taq polymerase 10 ul
  2. Forward primer (10 pmol) 1 ul
  3. Reverse primer (10 pmol) 1 ul
  4. D.W 8 ul
  5. DNA template (10 ng/ul) 2 ul
  Total 22 ul

- **PCR Condition (35 cycles) FR**
  3. Annealing 60°C 40 s
  4. Extension 72°C 3 min

- **PCR Condition (35 cycles) Int2FR**
  3. Annealing 58°C 40 s
  4. Extension 72°C 3 min

**PCR primer information**

|          | Forward Primer | Reverse primer | Internal Primer |
|----------|----------------|----------------|-----------------|
| F/R      | CACTTAAGACCTGTCTCCCC | CACCAACAAAATCCACTGGC | ATACTAAGGGAACACTCAGAGGC |

**Expected PCR product size**

|          | Human (empty size) | Chimpanzee | Gorilla | Orangutan |
|----------|--------------------|------------|---------|-----------|
| F/R      | 9753 bp (1793 bp)  | 2715 bp    | ?       | 2698 bp   |
| Int2F/R  | 2695 bp            | -          | -       | -         |
**D: HS_SVA:** chr2:192261918-192263988(hg19);
chr2:191397192-191399262(hg38);

| PCR Mixture   |         |
|---------------|---------|
| 1             | EF taq polymerase | 25 ul |
| 2             | Forward primer (4 pmol) | 5 ul |
| 3             | Reverse primer (4 pmol) | 5 ul |
| 4             | D.W     | 15 ul |
| 5             | DNA template (10 ng/ul) | 2 ul |
| **Total**     |         | 52 ul |

**PCR Condition (35 cycles)**

| Cycle | Step         | Temp (°C) | Time   |
|-------|--------------|-----------|--------|
| 1     | Pre-denaturation | 95        | 5 mins |
| 2     | Denaturation | 95        | 30 s   |
| 3     | Annealing    | 56        | 40 s   |
| 4     | Extension    | 72        | 3 min 10 s |
| 5     | Termination  | 72        | 10 mins|
| 6     | Hold         | 4         | ∞      |

**PCR primer information**

| Primer Information |         |
|--------------------|---------|
| Forward Primer     | GGAAAGTGGTCAGAACAGGC |
| Reverse primer     | AACCATCTTGCGGCTACCC |

**Expected PCR product size**

| Species          | Size     |
|------------------|----------|
| Human (empty size) | 2706 bp (507 bp) |
| Chimpanzee        | 506 bp   |
| Gorilla           | 509 bp   |
| Orangutan         | 507 bp   |
Fig. S2: Genome plots of HS-MEs. HS-MEs were plotted on the human chromosome ideograms (based on GRCh 38). The left side of the chromosome indicates the relative frequency of all HS-MEs (counts per 0.5Mb genomic region), while red track on the right represents HS-LTRs and the blue track represents HS-SVS (1 tick per entry for both tracks).