**Sharma et al., Supplemental Information.**

**Extraction of *LINE-1* (*L1*) repeats in *POTE* genes, and presentation Illumina 450K coordinates in *POTE* genes:** To extract *L1* repeats in *POTEs*, we determined the start/end coordinates of each *POTE* using the genome annotation file (hg38 assembly)[[1]](#footnote-1). We then wrote a Python script to find *L1* elements within -/+ 10K of each *POTE* [[2]](#footnote-2). We generated a .BED file containing all *L1* coordinates that span *POTE* genes, which allowed visualization in the genome browser[[3]](#footnote-3) along with CpG islands (CGI). We also visualized 450K Illumina Infinium probe coordinates.

**Table:** *POTE* family gene coordinates (hg38 assembly).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| gene\_name | ensembl\_id | gene\_biotype | chromosome | start | end | strand |
| *POTEA* | ENSG00000188877 | polymorphic\_pseudogene | chr8 | 43292483 | 43363518 | + |
| *POTEB* | ENSG00000233917 | protein\_coding | chr15 | 21846329 | 21877703 | - |
| *POTEB2* | ENSG00000230031 | protein\_coding | chr15 | 20835372 | 20866314 | - |
| *POTEB3* | ENSG00000278522 | protein\_coding | chr15 | 21408243 | 21440451 | - |
| *POTEC* | ENSG00000183206 | protein\_coding | chr18 | 14507339 | 14543585 | - |
| *POTED* | ENSG00000166351 | protein\_coding | chr21 | 13609858 | 13641585 | + |
| *POTEE* | ENSG00000188219 | protein\_coding | chr2 | 131218067 | 131265278 | + |
| *POTEF* | ENSG00000196604 | protein\_coding | chr2 | 130074030 | 130129222 | - |
| *POTEG* | ENSG00000187537 | protein\_coding | chr14 | 19402486 | 19434341 | - |
| *POTEH* | ENSG00000198062 | protein\_coding | chr22 | 15690026 | 15721631 | + |
| *POTEI* | ENSG00000196834 | protein\_coding | chr2 | 130459455 | 130509666 | - |
| *POTEJ* | ENSG00000222038 | protein\_coding | chr2 | 130611481 | 130658091 | + |
| *POTEKP* | ENSG00000204434 | unprocessed\_pseudogene | chr2 | 131591752 | 131627266 | + |
| *POTEM* | ENSG00000222036 | protein\_coding | chr14 | 18967434 | 18999012 | + |

**Figure:** Steps performed for extracting *L1* elements in *POTE* genes:



**Figure. Visualization of *POTEs* in the UCSC genome browser, including *L1* sequences and Illumina 450K CpG sites within *POTE* genes or 30kb flanking regions.**





**Table. Ilumina 450K CpG probes within *POTE* genes or 10kb flanking regions.**

|  |  |  |
| --- | --- | --- |
|  |  | **Ilumina 450K CpG sites** |
|  | **Ensembl ID** | **−10kb 🡪transcription start site (TSS)** | **TSS🡪 transcription end site (TES)** | **10kb downstream TES** |
| ***POTEA*** | **ENSG00000188877** | 29 | 2 | N/A |
| ***POTEB*** | **ENSG00000233917** | N/A | N/A | N/A |
| ***POTEB2*** | **ENSG00000230031** | N/A | N/A | N/A |
| ***POTEB3*** | **ENSG00000278522** | N/A | N/A | N/A |
| ***POTEC*** | **ENSG00000183206** | 5 | 3 | N/A |
| ***POTED*** | **ENSG00000166351** | 2 | 4 | N/A |
| ***POTEE*** | **ENSG00000188219** | 3 | 4 | 2 |
| ***POTEF*** | **ENSG00000196604** | 20 | 7 | N/A |
| ***POTEG*** | **ENSG00000187537** | N/A | 1 | N/A |
| ***POTEH*** | **ENSG00000198062** | 3 | 4 | N/A |
| ***POTEI*** | **ENSG00000196834** | 3 | N/A | N/A |
| ***POTEJ*** | **ENSG00000222038** | 2 | N/A | 1 |
| ***POTEKP*** | **ENSG00000204434** | 6 | 3 | N/A |
| ***POTEM*** | **ENSG00000222036** | 2 | 2 | N/A |

1. The file was downloaded from ensembl and currently stored in the following path: ~/POTE\_project/data/ensembl\_downloads/hg38/genes/raw/Homo\_sapiens.GRCh38.92.chr.gtf [↑](#footnote-ref-1)
2. For more detail about the script and the data used, go back to the script file: ~/POTE\_project/[Active]LINE\_coverage\_with\_POTEs.py [↑](#footnote-ref-2)
3. The path to the .BED file is: ~/POTE\_project/data/outputs/hg38\_LINEs\_within\_POTEs\_10K\_flankings.bed [↑](#footnote-ref-3)