This folder contains the information for the distribution of retrotransposable elements (RTEs) in different recombination regions and an R script for plotting the data. The R script process tab delimited text file in table format and creates a bar chart figure that presents the frequency distribution of RTEs in recombination regions to facilitate visualising the data.

BarplotForRecombinationDis.R: R code for plotting the distribution of reference and non-reference RTEs in hot, cold, and intermediate recombination regions. The script is executable in RStudio from the “Path\_To\_MyAnalysis/MyAnalysis/RStudio\_Scripts” directory:

**Input file:** RecombinationFrequencies.txt

The input file is a tab delimited text file in table format consisting of 4 columns in the following order:

Column 1: Recombination region. Either cold, intermediate, or hot.

Column 2: Frequency distribution of reference RTEs in each recombination region.

Column 3: Frequency distribution of non-reference RTEs in each recombination region.

Column 4: RTE type. Either L1, Alu, or SVA.

The frequency distribution of each RTE type in each genomic region was determined by overlapping the position of RTEs with the standardized sex-averaged Decode recombination map (https:www.decode.com/; Kong et al., 2010). The Decode recombination map was downloaded from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTables; Karolchik, 2004); accessing the ‘decodeSexAveraged’ table under the group ‘All Tables’ under the ‘hg19’ database. The recombination map file is available in “MyAnalysis/ Recombination\_rate\_analysis/ Recombination\_map” folder.

The position of autosomal RTEs were overlapped with the recombination map using the closest tool of BEDtools suite version 2.25.0 (Quinlan, 2014) with the following parameters:

-a: RTE features in bed file format.

-b: Recombination map in bed file format.

-d: Reports the distance between the two features. The reported distance for overlapping features is 0.

Command:

$ awk '{ if ($1!="X" && $1!="Y") print $0} BEGIN {OFS="\t"}' RTE.bed\* | sort -k1,1V -k2,2n > RTE\_autosomal.bed

$ bedtools closest -a RTE\_autosomal.bed -b decodeSexAverage\_hg19\_noChr\_19032020.txt -d > output.bed

The output file was sorted in ascending order based on the position of RTE and distance to closest recombination region. Duplicate lines of insertions closet to multiple recombination regions are then removed using the following command: awk '!seen[$0]++'

This leaves one entry for each RTE insertions with its closest recombination region. The RTE insertions were then grouped based on the standardized recombination rate (SRR) of the region. SRR of 0 represent recombination cold-spots, SRR ≥ 10 represent recombination hotspots, and SRR between 0 and 10 represent intermediate recombination regions.

Frequency of insertions per recombination region was then calculated as:

\* RTE.bed are 3 reference L1s, Alus, and SVA files and 3 non-reference L1s, Alus, and SVAs files available in “MyAnalysis/RTE\_files” folder. The three files prefixed with RTEdb\_\* are non-reference RTEs while the remaining three are the reference RTE files.