This folder contains two scripts for linear regression analyses. The first R scrip investigates the relationship between the frequency of RTEs per chromosome and chromosome size. The second R scrip investigates the relationship between RTEs density and gene density per chromosome. Both R scrips process a tab delimited text file in table format and produces regression plots that presents the relationship between each of the independent variables (i.e. chromosome size and gene density) with each type of reference and non-reference RTEs consisting of L1s, Alus, and SVAs.

1. Linear regression\_RTE per chromosome and chromosome size.R: R script for the regression analysis between the frequency of L1s, Alus, and SVAs per chromosome and chromosome size. Each script is executable in RStudio by pressing Ctrl+Shift+S or by clicking the Source button on the top right corner. There are 3 similar scripts for each RTE type:

Linear regression\_L1s per chromosome and chromosome size.R

Linear regression\_Alus per chromosome and chromosome size.R

Linear regression\_SVAs per chromosome and chromosome size.R

1. Linear regression\_RTE and gene density per chromosome.R: R script for the regression analysis between density of L1s, Alus, and SVAs per chromosome and gene density. Each script is executable in RStudio by pressing Ctrl+Shift+S or by clicking the Source button on the top right corner. There are 3 similar scripts for each RTE type:

Linear regression\_L1s and gene density per chromosome.R

Linear regression\_Alus and gene density per chromosome.R

Linear regression\_SVAs and gene density per chromosome.R

**Input file**: RegressionAnalysis\_inputFile.txt

The input file is a tab delimited file consisting of 23 columns in the following order:

Column 1: Chromosome name

Columns 2 & 3: Chromosome length in base pairs (bp) and in megabases (Mb)

Columns 4 & 5: Total number of genes per chromosome and gene density per Mb.

Columns 6-23 contain the chromosomal distribution information for reference and non-reference L1s, Alus, and SVAs including count of insertions per chromosome, frequency of insertions per chromosome, and insertions density per Mb.

Gene density was calculated as:

Frequency of insertions per chromosome was calculated as:

Density of insertions per chromosome was calculated as: