This folder contains the scripts needed for investigating the enrichment of retrotransposable element (RTE) insertions at genomic loci associated with disease risk.

A list of SNPs associated with disease risk by genome-wide association studies (GWAS) was downloaded from the NHGRI-EBI Catalog of published GWAS ([www.ebi.ac.uk/gwas](http://www.ebi.ac.uk/gwas)) (Buniello et al., 2019). GWS SNPs (Genome-wide significant with a P-value ≤ 5x10-8) that have been identified in cohorts of European descents and were not in the human leukocyte antigen (HLA) region were extracted.

The genotypes of SNPs identified the European super population from the 1,000 genome project phase 3 release were obtained via the steps included in 1kGPEuropeanGenotypes.txt file.

**Note**: A copy of PLINK binary files “allchromosome\_european\_1kPGph3” is available in OneDrive via the following links:

allchromosome\_european\_1kPGph3.bed link:

<https://liveplymouthac-my.sharepoint.com/:u:/g/personal/randa_ali_plymouth_ac_uk/EXfl_LlolGlFhZjcTSNFAdkB2LMVYFbVqLlxY1qwrQ6j_A?e=v4Ji8m>

allchromosome\_european\_1kPGph3.bim link:

<https://liveplymouthac-my.sharepoint.com/:u:/g/personal/randa_ali_plymouth_ac_uk/EW_qCUOcx25Gn4Y0Yju74AcBqPPIqHV1yK9IgDLz4HnURA?e=eaxL61>

allchromosome\_european\_1kPGph3.fam link:

<https://liveplymouthac-my.sharepoint.com/:u:/g/personal/randa_ali_plymouth_ac_uk/EddfsZsma2lBrC25PC_-t6UBr4ARTzWYDrVUHLfLLUA_VA?e=a6foVq>

Download the .bed .bim .fam files, and add them to the following directory:

Path\_to\_ RTE\_enrichment\_in\_GWAS\_loci/DataFile/

A thousand set of random SNPs with genomic properties that match the properties of GWAS SNPs were generated using SNPsnap default settings (<https://data.broadinstitute.org/mpg/snpsnap/>) (Pers et al., 2015).

An LD-block was generated around each SNP within the 1,000 random sets using the first 5-’ and last 3’- tagging SNPs (r2≥ 0.8) via the GenerateLDblocks.R code.

Each set of random LD-blocks generated from the previous step were intersected with Polymorphic RTEs using the EnrichmentInGWASloci.R code.

The R codes are executable in the command line from the “Path\_To\_ RTE\_enrichment\_in\_GWAS\_loci/Codes” directory.

**Required software**:

PLINK, (Purcell et al., 2007).

BEDtools, (Quinlan, 2014).

**List of references**:

Buniello, A. *et al.* (2019) ‘The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019’, *Nucleic Acids Research*. Oxford University Press, 47(D1), pp. D1005–D1012. doi: 10.1093/nar/gky1120.

Pers, T. H., Timshel, P. and Hirschhorn, J. N. (2015) ‘SNPsnap: A Web-based tool for identification and annotation of matched SNPs’, *Bioinformatics*, 31(3), pp. 418–420. doi: 10.1093/bioinformatics/btu655.

Purcell, S. *et al.* (2007) ‘PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses’, *The American Journal of Human Genetics*, 81(September), pp. 559–575. doi: 10.1086/519795.

Quinlan, A. R. (2014) BEDTools: The Swiss-Army tool for genome feature analysis, Current Protocols in Bioinformatics. doi: 10.1002/0471250953.bi1112s47.