Example R script to run MotifbreakR for analysis in the paper:

Bioinformatics pipeline to guide late-onset Alzheimer’s disease (LOAD) post-GWAS studies: Prioritizing **transcription regulatory variants within LOAD associated regions**

By Michael W. Lutz and Ornit Chiba-Falek

|  |
| --- |
| > setwd("~/2019 SSV for GWAS paper")  > library(motifbreakR)  > library(SNPlocs.Hsapiens.dbSNP142.GRCh37)  Loading required package: GenomeInfoDb  Loading required package: GenomicRanges  Loading required package: BSgenome  Loading required package: rtracklayer  > library(BSgenome.Hsapiens.UCSC.hg19)  > pca.snps.file <- system.file("extdata", "AD.snps", package = "motifbreakR")  > pca.snps <- as.character(read.table(pca.snps.file)[,1])  > library(MotifDb)  > snps.mb <- snps.from.rsid(rsid = pca.snps,  + dbSNP = SNPlocs.Hsapiens.dbSNP142.GRCh37,  + search.genome = BSgenome.Hsapiens.UCSC.hg19) |
|  |
| |  | | --- | | > data(motifbreakR\_motif)  > results <- motifbreakR(snpList = snps.mb, filterp = TRUE,  + pwmList = hocomoco,  + threshold = 1e-4,  + method = "ic",  + bkg = c(A=0.25, C=0.25, G=0.25, T=0.25),  + BPPARAM = BiocParallel::bpparam()) | |

\*\*\*\* show results for one variant

> rs10792832 <- results[names(results) %in% "rs10792832"]

> rs10792832

\*\*\*\* calculate p value for the motif

> rs10838725 <- calculatePvalue(rs10838725)

> rs10838725

Snps.mb output

GRanges object with 23 ranges and 4 metadata columns:

seqnames ranges strand | SNP\_id alleles\_as\_ambig REF ALT

<Rle> <IRanges> <Rle> | <character> <DNAStringSet> <DNAStringSet> <DNAStringSet>

rs10498633 chr14 92926952 \* | rs10498633 K G T

rs10792832 chr11 85867875 \* | rs10792832 R A G

rs10838725 chr11 47557871 \* | rs10838725 Y T C

rs10948363 chr6 47487762 \* | rs10948363 R A G

rs11218343 chr11 121435587 \* | rs11218343 Y T C

... ... ... ... . ... ... ... ...

rs75932628 chr6 41129252 \* | rs75932628 Y C T

rs8093731 chr18 29088958 \* | rs8093731 Y C T

rs9271192 chr6 32578530 \* | rs9271192 M C A

rs9331896 chr8 27467686 \* | rs9331896 Y C T

rs983392 chr11 59923508 \* | rs983392 R A G

output for rs10792832

> rs10792832 <- results[names(results) %in% "rs10792832"]

> rs10792832

GRanges object with 2 ranges and 18 metadata columns:

seqnames ranges strand | REF ALT snpPos motifPos geneSymbol dataSource providerName

<Rle> <IRanges> <Rle> | <DNAStringSet> <DNAStringSet> <integer> <integer> <character> <character> <character>

rs10792832 chr11 85867863-85867876 + | A G 85867875 13 HSF1 HOCOMOCO HSF1\_f2

rs10792832 chr11 85867873-85867892 + | A G 85867875 3 IRF5 HOCOMOCO IRF5\_f1

providerId seqMatch pctRef pctAlt scoreRef scoreAlt Refpvalue

<character> <character> <numeric> <numeric> <numeric> <numeric> <logical>

rs10792832 HSF1\_HUMAN ggaaaaatgtagAa 0.876730504122011 0.805890624885523 6.26652775571283 5.76019192348197 <NA>

rs10792832 IRF5\_HUMAN agAagcaaaacatacacagc 0.761913349795418 0.679592411790298 9.7051775450215 8.65658150817962 <NA>

Altpvalue alleleRef alleleAlt effect

<logical> <numeric> <numeric> <character>

rs10792832 <NA> 0.742067019991023 0.0898512406131928 weak

rs10792832 <NA> 1 0 strong

-------

seqinfo: 93 sequences (1 circular) from hg19 genome