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

**Bioinformatics pipeline to guide post-GWAS studies in Alzheimer’s: A new catalogue of disease candidate short structural variants**

By Michael W. Lutz and Ornit Chiba-Falek

R version 4.0.5 (2021-03-31) -- "Shake and Throw"

Copyright (C) 2021 The R Foundation for Statistical Computing

Platform: x86\_64-w64-mingw32/x64 (64-bit)

> library(motifbreakR)

> library(BSgenome)

> library(BSgenome.Hsapiens.UCSC.hg38)

> library(MotifDb)

>data(“hocomoco”)

> tfn <- system.file("extdata","Bell\_p1.txt",package="motifbreakR")

> tvars <- variants.from.file(file=tfn,search.genome=BSgenome.Hsapiens.UCSC.hg38,format="bed")

> tvars

GRanges object with 200 ranges and 3 metadata columns:

seqnames ranges strand |

<Rle> <IRanges> <Rle> |

chr1:108876393:T:TTTTG chr1 108876393 \* |

chr1:108877450-108877453:TTGC:T chr1 108877450-108877453 \* |

chr1:108877514-108877515:CT:C chr1 108877514-108877515 \* |

chr1:108962693:A:AC chr1 108962693 \* |

chr1:108962870:C:CA chr1 108962870 \* |

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

chr7:12210391-12210394:CCTA:C chr7 12210391-12210394 \* |

chr7:12210476-12210478:AAT:A chr7 12210476-12210478 \* |

chr7:12210501-12210507:CCTGTGA:C chr7 12210501-12210507 \* |

chr7:12211745-12211747:TAC:T chr7 12211745-12211747 \* |

chr7:12211989:C:CT chr7 12211989 \* |

SNP\_id REF ALT

<character> <DNAStringSet> <DNAStringSet>

chr1:108876393:T:TTTTG chr1:108876393:T:TTTTG T TTTTG

chr1:108877450-108877453:TTGC:T chr1:108877450-10887.. TTGC T

chr1:108877514-108877515:CT:C chr1:108877514-10887.. CT C

chr1:108962693:A:AC chr1:108962693:A:AC A AC

chr1:108962870:C:CA chr1:108962870:C:CA C CA

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

chr7:12210391-12210394:CCTA:C chr7:12210391-122103.. CCTA C

chr7:12210476-12210478:AAT:A chr7:12210476-122104.. AAT A

chr7:12210501-12210507:CCTGTGA:C chr7:12210501-122105.. CCTGTGA C

chr7:12211745-12211747:TAC:T chr7:12211745-122117.. TAC T

chr7:12211989:C:CT chr7:12211989:C:CT C CT

-------

seqinfo: 7 sequences from hg38 genome

> rtout <- motifbreakR(snpList=tvars,filterp=TRUE,pwmList=hocomoco,method="ic",threshold=1E-4, bkg=c(A=0.25, C=0.25, G=0.25, T=0.25),legacy.score=FALSE)

> write.csv(rtout,"Bell\_p1\_output\_hocomoco.csv")

\*\*\*\* repeat for Encode motifs

> data("encodemotif")

> rtout <- motifbreakR(snpList=tvars,filterp=TRUE,pwmList=encodemotif,method="ic",threshold=1E-4, bkg=c(A=0.25, C=0.25, G=0.25, T=0.25),legacy.score=FALSE)

> write.csv(rtout,"Bell\_p1\_output\_encode.csv")

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

> rs4647710 <- results[names(results) %in% "rs4647710"]

> rs4647710

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

> rs4647710 <- calculatePvalue(rs4647710, granularity = 1e-4)

> rs4647710

\*\*\*\*\* calculate p values for a set of variants

> rtout\_pvalues <- calculatePvalue(rtout, granularity = 1e-4)