Scripts

Best view with “Navigation Pane” in Word turn on.

# gwasplot.R

P.plot function for plotting colorful Manhattan plot

USAGE:

maledata2 <- maledata[,c("chr","SNPID","position","p","HWE\_pval")]

names(maledata2) <- c("CHR","SNP","BP","P","hwe")

png(filename="maleCRP.png",width=11,height=8.5,res=300,units="in")

P.plot(d=maledata2,flag=TRUE,hwe.cut=1e-6,suggestiveline=5,title="Male CRP")

dev.off()

Input : plink.qassoc (plink association result) file structure with at least

3 columns below [HWE is optional]

CHR :: chromosome

BP :: positions

P :: P value of GWAS results

hwe :: HWE pvalue (required for flag=T option)

# gwasGEE.R

Analyze family data for association using GEE with exchangeable correlation structure and robust variance estimators

USAGE:

gwasGEE(genodata=genotype,phendata=phenotype,k=1000,genostart=7, genoend=ncol(genodata),outcomes=names(phenotype)[-2:-1], famid="FID",covar="",suffix="",outdir="./results",mincount=0 ,...)

--------------------------------------------------------------------------------------------------------------------

genodata : a data frame containing genotype data from PLINK obtained

using plink --bfile [plink\_file] --recodeA --out [output\_file]

phendata : a data frame containing phenotype data with "ID" column for combining with genodata

k : an integer used to opimize the size of data to be handled at a time. Default = 1000

genostart: an integer denoting the first column of the genotype in genodata

genoend : an integer denoting the column number of the last marker to be analyzed,

by default this is the last column of genodata

outcomes : a character string or array containing the name of phenotypes in "phendata" to be analyzed

famid : a character string denoting the variable name in genodata that is corresponding to the cluster ID

to be used with GEE model

covar : a character string containing covariates in the form

"+X1+X2+X3"

suffix : a character string to be added to the result filename in addition to the phenotype name

outdir : a character string denoting the output directory

mincount : an integer denoting the minimum genotype count of SNPs to be analyzed e.g.

specifying mincount=5 will only analyze SNP with at least 5 counts for any genotype

... : additional options to be passed onto gee() function

# combinedResult.sh

Combine a series of files with the same prefix name following by an index number. Each file also has a header. This script will use the first file's header and add additional file to create one bigger file.

USAGE: sh combinedResult.sh [prefix] [ext] [from] [to]