SKAT-O Gene-based Association Analysis

```
## load phenotype data
if(!file.access(PHENO.FN)==0) {
  ## phenotype data has not yet been generated -- do that now
  source(purl('phenotype_preprocessing.Rmd'))
} else {
  ## load already processed phenotype file
  load(PHENO.FN)
}
## subset to european-like individuals
eur.pheno <- pheno.df[which(pheno.df$iseuro == TRUE),]
```

SKAT-O Association

Load and annotate VCF file of PGRNseq variants.

```
vcfFile <- '../data/consensus_seq_variants/consensus.geno.vcf.gz'
annot <- annotate.genotypes(vcfFile)
```

```
## Warning: trimmed start values to be positive
## Warning: trimmed end values to be <= seqlengths
```

```
sparseX <- load.snp.mat(vcfFile, eur.pheno)
```

```
## Loading required package: splines
##
## Attaching package: 'Matrix'
##
## The following object is masked from 'package:VariantAnnotation':
##
##     expand
##
## The following object is masked from 'package:IRanges':
##
##     expand
```

```
## Warning: non-diploid variants are set to NA
```

```
genes <- unique(annot$GENEID[which(!is.na(annot$GENEID))])
```

Fit a SKAT-O model for each unique gene in our annotation. This scan is farmed out across multiple cores.

```
## Warning: 14  samples have either missing phenotype or missing covariates. They are exclud
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
```

```
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: invalid factor level, NA generated
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
```

Check the QQ-plot of your result.

Take a look at the top hits, their p-values, and the genes they correspond to.

```
top.hits <- clean.res[order(clean.res$p),]
dim(clean.res)
```

[1] 365 6

```
kable(head(top.hits), format='markdown')
```

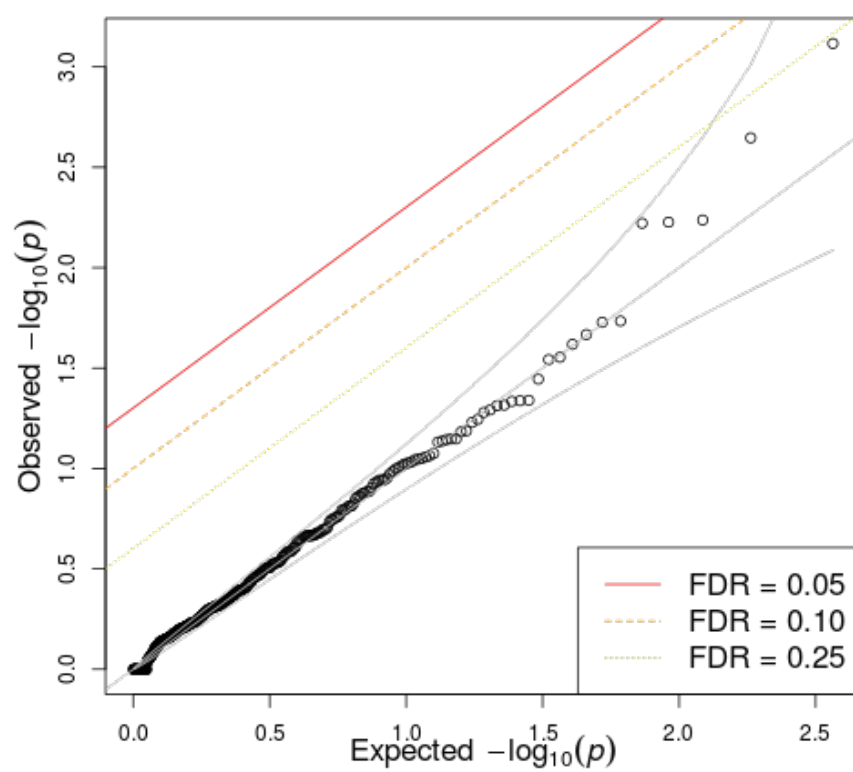|	gene	p	param.p	rho	nsnps	symbol
237	100286922	0.0008	0.0008	1	9	LOC100286922
45	2052	0.0023	0.0012	0	27	EPHX1
350	2938	0.0058	0.0031	1	24	GSTA1
83	3177	0.0059	0.0035	0	8	SLC29A2
130	1588	0.0060	0.0035	1	22	CYP19A1
140	6799	0.0184	0.0114	0	14	SULT1A2

Figure 1: plot of chunk qqplot