

R Notebook

Expression QTL analysis in expression genetic data

```
#####
# 1. install some packages
#####
setwd("/Users/shiluzhang/Box/UWResearch/Rotations/Rotation2/")
# install.packages("qtl")
# install.packages("lineup")

#####
# 2. Download data
#####
## The data are at https://phenome.jax.org/projects/Attie1
# zipurl <- "https://phenomedoc.jax.org/QTL_Archive/attie_2015/Attie_2015_eqtl_clean.zip"
# dir_for_data <- "Attie_data"
# zipfile <- file.path(dir_for_data, "Attie_2015_eqtl_clean.zip")
#
# # check if directory exists; if not, create it
# if(!dir.exists(dir_for_data))
#   dir.create(dir_for_data)
# download.file(zipurl, zipfile) # about 913 MB
#
# unzipped <- unzip(zipfile, exdir=dir_for_data) # about 2.6 GB expanded
#
# ## data gets placed in "Clean" subdirectory
# data_dir <- file.path(dir_for_data, "Clean")

#####
# 3. load data
#####
## annotation file
library(data.table)
data_dir=".~/Attie_data/Clean"
annot <- fread(file.path(data_dir, "microarray_annot.csv"), data.table=FALSE)
# "a_gene_id" is the main probe identifier
# "chr", "pos.cM", and "pos.Mb" are the genomic positions

## QTL cross
library(qtl)
f2g <- read.cross("csv", data_dir, "genotypes_clean.csv",
                   genotypes=c("BB", "BR", "RR"), alleles=c("B", "R"))

## --Read the following data:
## 544 individuals
## 2060 markers
## 3 phenotypes
```

```

## Warning in summary.cross(cross): Some markers at the same position on chr
## 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,X; use jittermap().
## --Cross type: f2
f2g <- jittermap(f2g) # avoid having markers at exactly the same location

## load the islet expression data
islet <- fread(file.path(data_dir, "islet_mlratio_clean.csv"), header=TRUE, data.table=FALSE)
# make first column (mouse IDs) the row names
rownames(islet) <- islet[,1]
islet <- islet[,-1]
# 491 rows (the mice) and 40572 columns (the microarray probes)

#####
# 4. keep only probes that have genomic positions
# and are on an autosome (1-19)
#####
probeindex=which(!is.na(annot$pos.cM) & annot$chr!="X")
probes2keep=as.character(annot$a_gene_id[probeindex])
# 36364 probes kept

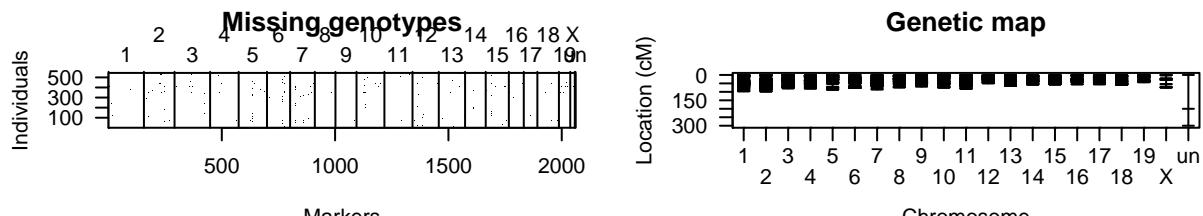
# subset the islet data to just these probes
islet1 <- islet[,probes2keep]

# probe location in cM
probeloc <- data.frame(chr=annot$chr[probeindex] ,
                         pos=annot$pos.cM[probeindex])
rownames(probeloc) <- probes2keep

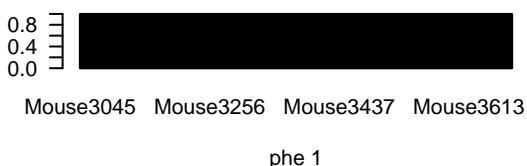
#####
# 5. calculate conditional QTL genotype probabilities
#####

f2g <- calc.genoprob(f2g, step=0.5, error.prob=0.002, map.function="c-f")
# probabilities are now embedded inside f2g
# f2g$geno[[6]]$prob is a 3d array for chr 6, mouse x position x genotype
#pdf("datasummary_islet.pdf")
plot(f2g)

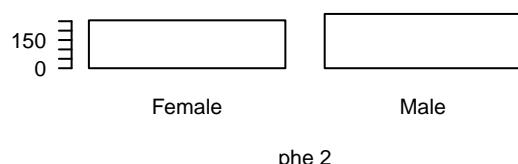
```



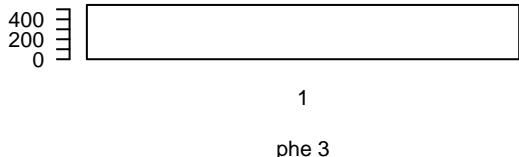
MouseNum



Sex



pgm



```
#dev.off()
```

```
#####
# 6. find pseudomarker nearest each gene
#####

library(lineup)
pmar <- find.gene.pseudomarker(f2g, pull.map(f2g), probeloc)

## Warning in find.gene.pseudomarker(f2g, pull.map(f2g), probeloc): 2055 genes
## differ from pseudomarker pos by > 2 Mbp, with gaps as big as 20.1 Mbp
# doing all this with cM rather than Mbp
# some genes quite far from any marker, but we can ignore this for now

#####
# 7. calculate a and d for each sex in islet
#####
n=length(probes2keep)
result=matrix(0,nrow=n,ncol=6)
adjrsq=NULL
rsq=NULL
fit=NULL
for(i in 1:10) #36364
{
  probe=probes2keep[i]
  chr <- as.character(pmar[probe, "chr"])
  this_pmar <- pmar[probe, "pmark"]
```

```

# probabilities are embedded in
pr <- f2g$geno[[chr]]$prob[,this_pmar,] # 544 x 3 matrix

# put IDs as row names
rownames(pr) <- f2g$pheno$MouseNum

# sex of the mice ("Male" and "Female")
sex <- f2g$pheno$Sex

# lineup the mice in the genotype data and the islet data
# (function in R/lineup package)
id <- findCommonID(rownames(pr), rownames(islet))

# subset the two; also subset sex
pr <- pr[id$first,]
islet <- islet[id$second,]
sex <- sex[id$first]

# the expression data for this particular probe
y <- islet[,probe]

# calculate X matrix; can leave out the intercept
x <- cbind(a = (pr[,3] - pr[,1])/2,
            d = pr[,2] - (pr[,1] + pr[,3])/2)

# estimate a and d in females and males separately
lm_fem <- lm(y ~ x, subset=(sex=="Female"))
lm_mal <- lm(y ~ x, subset=(sex=="Male"))

# results in a vector
result[i,] <- c(a_fem=lm_fem$coef[2],
                  d_fem=lm_fem$coef[3],
                  sig_fem=summary(lm_fem)$sigma,
                  a_mal=lm_mal$coef[2],
                  d_mal=lm_mal$coef[3],
                  sig_mal=summary(lm_mal)$sigma)
adjrsq[i]=summary(lm_fem)$adj.r.squared
rsq[i]=summary(lm_mal)$r.squared
fit[[2*i-1]]=lm_fem
fit[[2*i]]=lm_mal
}

#save(result,file="islet/Result_islet.Rdata") #i=12964
load("islet/Result_islet.Rdata")
#result=result.islet
rownames(result)=probes2keep
colnames(result)=c("a_fem","d_fem","sig_fem","a_mal","d_mal","sig_mal")

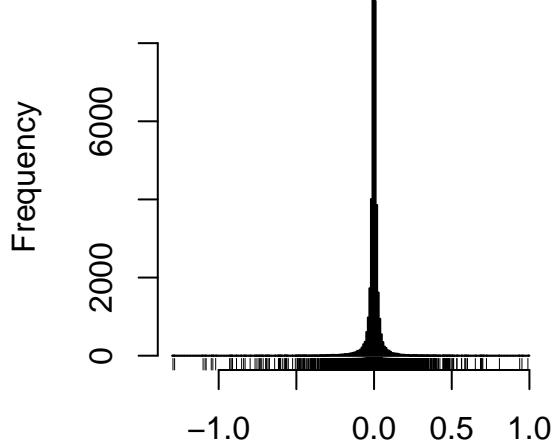
#####
# 7. some histograms plots to explore a vs d for each sex in islet
#####
par(mfrow=c(1,2),pty = "s")
#pdf("Histogram_islet.pdf")

```

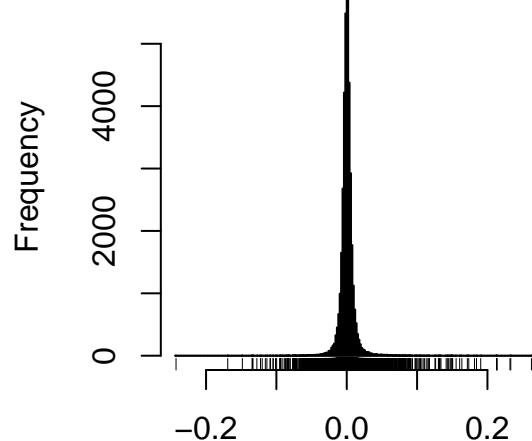
```
hist(result[,1],breaks=300,main="Histogram of Additive effect in female",xlab="Additive effect in female",rug(result[,1]))
```

```
hist(result[,2],breaks=300,main="Histogram of Dominance effect in female",xlab="Dominance effect in female",rug(result[,2]))
```

Histogram of Additive effect in female



Histogram of Dominance effect in female



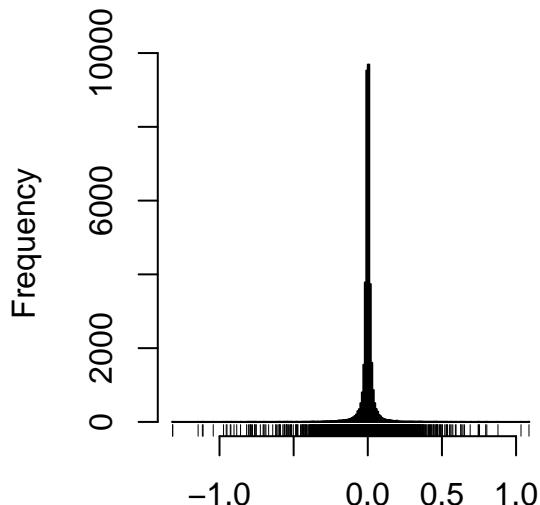
Additive effect in female

Dominance effect in female

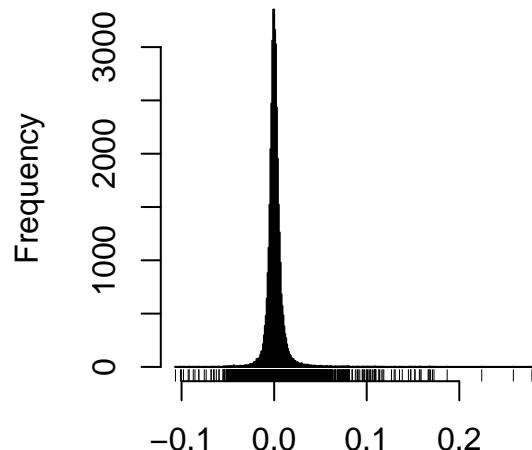
```
hist(result[,4],breaks=300,main="Histogram of Additive effect in male",xlab="Additive effect in male",rug(result[,4]))
```

```
hist(result[,5],breaks=300,main="Histogram of Dominance effect in male",xlab="Dominance effect in male",rug(result[,5]))
```

Histogram of Additive effect in male



Histogram of Dominance effect in male



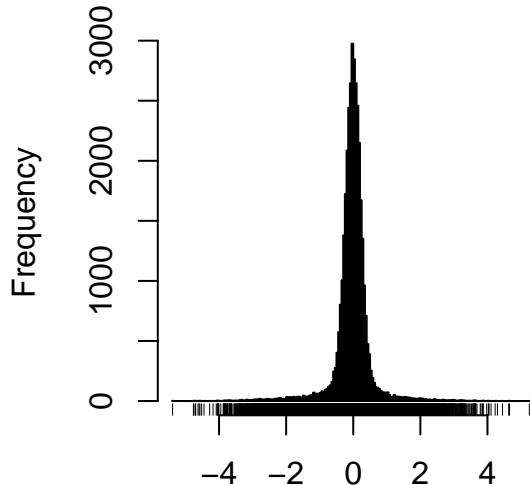
Additive effect in male

Dominance effect in male

```
hist(result[,1]/result[,3],breaks=300,main="Histogram of a/sig female")
rug(result[,1]/result[,3],cex.main=0.8)
```

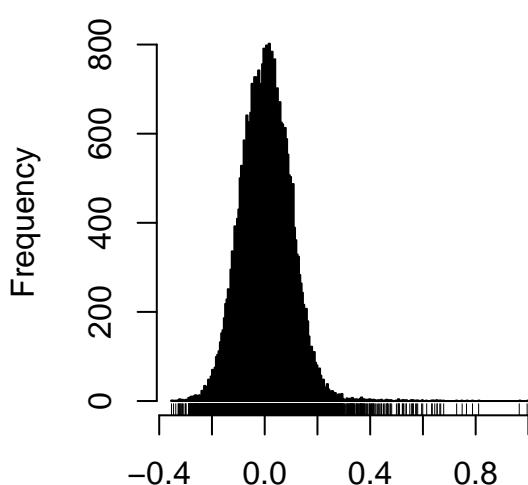
```
hist(result[,2]/result[,3],breaks=300,main="Histogram of d/sig female")
rug(result[,2]/result[,3],cex.main=0.8)
```

Histogram of a/sig female



result[, 1]/result[, 3]

Histogram of d/sig female

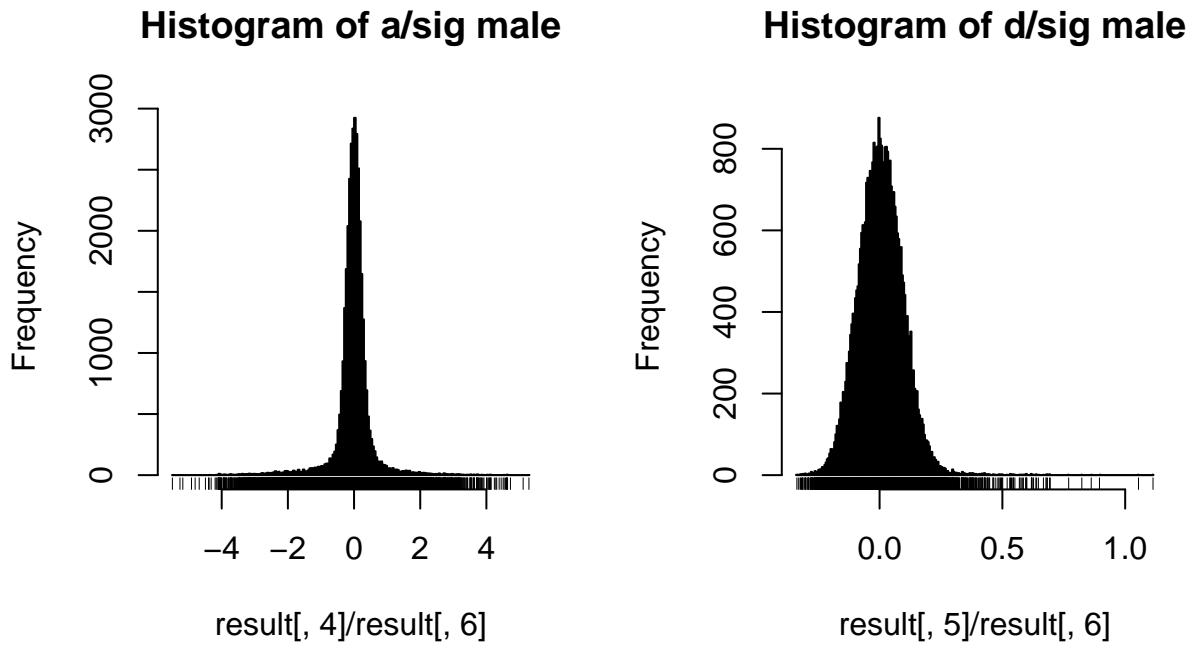


result[, 2]/result[, 3]

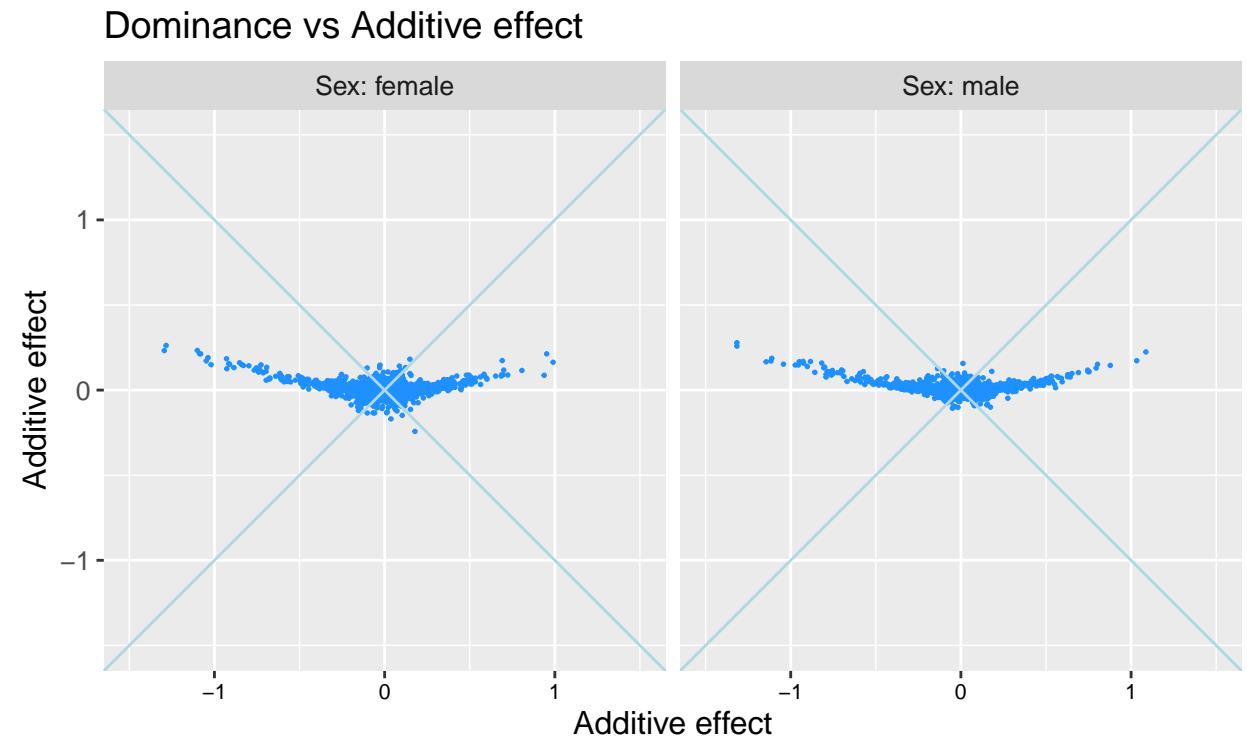
```
hist(result[,4]/result[,6],breaks=300,main="Histogram of a/sig male")
rug(result[,4]/result[,6],cex.main=0.8)
```

```
hist(result[,5]/result[,6],breaks=300,main="Histogram of d/sig male")
rug(result[,5]/result[,6],cex.main=0.8)
```

```
#####
# 8. plot a vs d for each sex in islet
#####
#plot a vs d
#pdf("Scatterplot_islet.pdf",height=6,width=10)
library(ggplot2)
```



```
df1=data.frame(result[,1:3],sex="female")
df2=data.frame(result[,4:6],sex="male")
names(df1)=c("Additive","Dominance","StDeviation","Sex")
names(df2)=c("Additive","Dominance","StDeviation","Sex")
dplot=rbind(df1,df2)
gg=ggplot(dplot, aes(x=Additive, y=Dominance))+geom_point(size=0.3,color="dodgerblue")+
  theme(text = element_text(size=12))
  ggtitle(label = paste("Dominance vs Additive effect"))+geom_abline(intercept=0,slope=1,color="lightblue")
plot(gg)
```

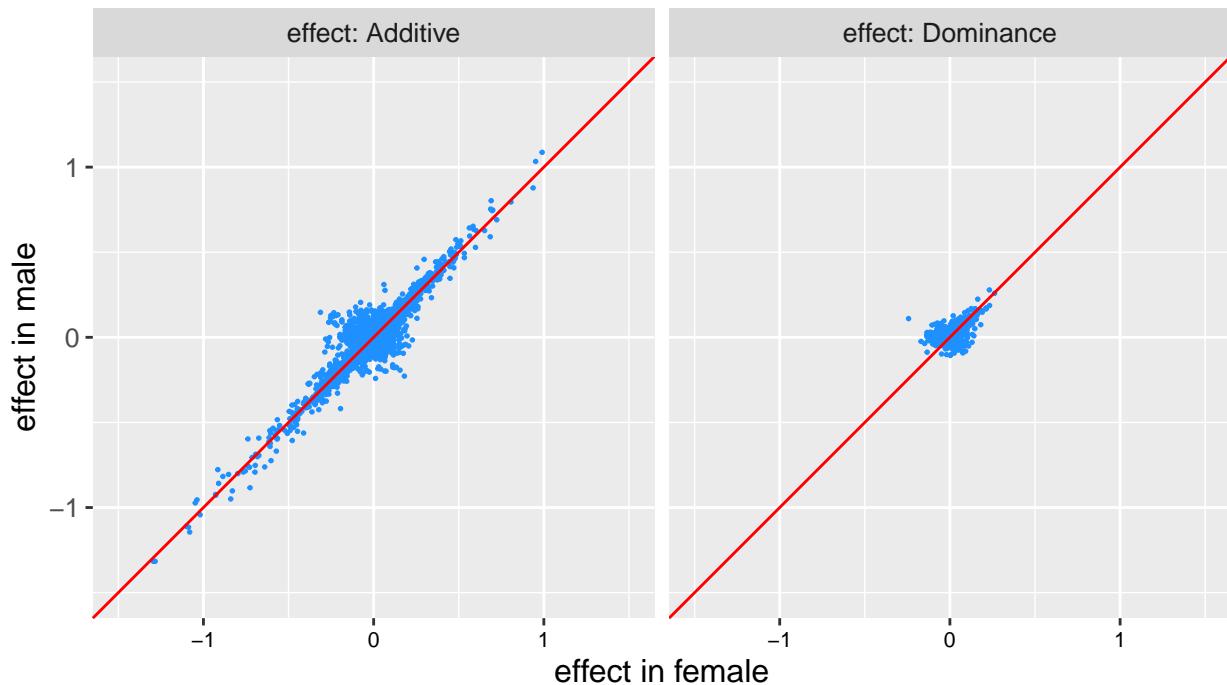


```

par(mfrow=c(1,2),pty = "s")
df1=data.frame(result[,c(1,4)],effect="Additive")
df2=data.frame(result[,c(2,5)],effect="Dominance")
names(df1)=c("fem","mal","effect")
names(df2)=c("fem","mal","effect")
dplot=rbind(df1,df2)
gg=ggplot(dplot, aes(x=fem, y=mal))+geom_point(size=0.3,color="dodgerblue")+theme(text = element_text(size=12))
  ggtitle(label = paste("Sex differences"))+ geom_abline(intercept=0,slope=1,color="red")+ylim(c(-1.5,1.5))
plot(gg)

```

Sex differences



```

# plot(result[,1]/result[,3],result[,2]/result[,1],pch=16,cex=0.5,xlab="a/sig female",ylab="d/a female")
# abline(0,1)
# abline(0,-1)
#
# plot(result[,4]/result[,6],result[,5]/result[,4],pch=16,cex=0.5,xlab="a/sig male",ylab="d/a male")
# abline(0,1)
# abline(0,-1)
#dev.off()

```