

Genome Wide Association Studies (GWAS)

Part 2

BCB 504: Applied Bioinformatics

Matt Settles

University of Idaho
Bioinformatics and Computational Biology Program

April 25, 2012

1 GWAS Case/Control studies

- Population Stratification

Case/Control studies

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

Considering analysis with plink. Some summary statistics can be performed using the following commands

- `-missing`
- `-test-missing`
- `-cluster-missing`
- `-hardy`
- `-freq`
- `-check-sex`
- `-impute-sex -make-bed`
- `-homozyg`

Look for trends, sample/marker outliers, etc.

Population Stratification

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

Population stratification is the presence of a systematic difference in allele frequencies between subpopulations in a population possibly due to different ancestry, especially in the context of association studies. Population stratification is also referred to as population structure, in this context.

- Prune snps
 - `-indep-pairwise 50 5 0.2 -out pair-indep`
 - window of 50kb, step of 5kb and R^2 of 0.2
- Cluster on pruned set
 - `-extract pair-indep.prune.in -cluster -matrix -mds-plot 4 -out keep.mds-plot`
 - produce mds-plot data, 4 dimensions

MDS-plot

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

```
## R mds-plot
pheno <- read.table("bse.ped")
pheno <- pheno[,1:6]
colnames(pheno) <- c("FID", "IID", "S", "D", "S", "T")
kmbs <- read.table("keep.mds-plot.mds", header=TRUE)
pheno <- pheno[match(kmbs[,2], pheno[,2]),]

col=c("black", "red")[as.numeric(as.factor(pheno[,6]))]

pdf("Figures/mdsplotKeep.pdf", width=5, height=5, pointsize=8)
plot(kmbs$C1, kmbs$C2, col=col, pch=1, xlab="C1", ylab="C2")
legend("topleft", legend=c("Control", "BSE-positive"), col=c("black", "red"), pch=c(1))
dev.off()

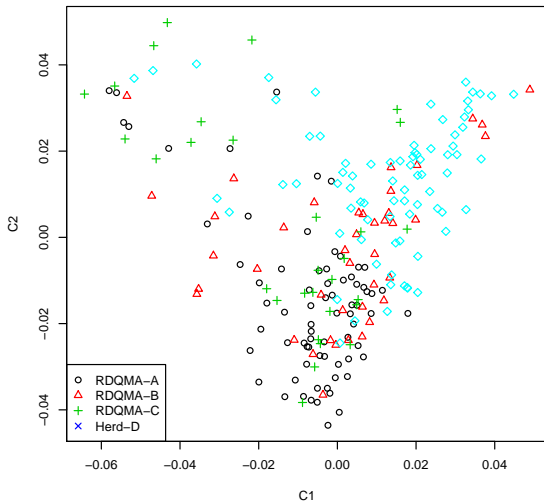
png("Figures/mdsplotpaper.png", width=7, height=7, pointsize=12, res=150, units="in")
pdf("Figures/mdsplotpaper.pdf", width=5, height=5, pointsize=12)
par(mar=c(0.5, 0.5, 0.5, 0.5))
plot(kmbs$C1, kmbs$C2, col=col, pch=c('A', 'B', 'C', 'E', 'D')[pch],
      , xlab="", ylab="", cex=0.7, axes=FALSE)
box()
legend("topleft", legend=c("Herd A", "Herd B", "Herd C", "Herd D", "Removed Animals"),
      col=c("black", "red", "green", "blue", "orange"), pch=c('A', 'B', 'C', 'D', "X"), cex=0.7)
dev.off()
```

MDS-plot

GWAS

Matt Settles

GWAS
Case/Control
studies
Population
Stratification



Hierarchical cluster

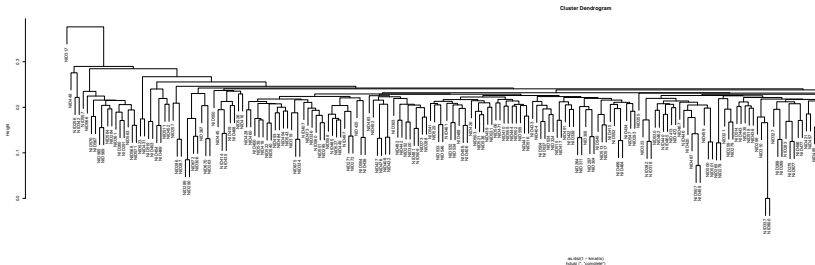
GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

```
kmatrix <- read.table("keep.mds-plot.mibs")  
rownames(kmatrix) <- kmids[,1]  
colnames(kmatrix) <- kmids[,1]  
hc <- hclust(as.dist(1-kmatrix))  
pdf("Figures/hclustkeep.pdf",width=30,height=7,pointsize=8)  
plot(hc)  
dev.off()
```



Association Analysis

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

- Random data (no stratification)
 - `-assoc -adjust -qq-plot`
 - perform case/control association (Chi-square), apply multiple testing adjustment and produce data for a qq-plot
- Stratified data
 - `-mh -adjust -qq-plot -within sourcecluster.dat`
 - perform a Cochran-Mantel-Haenszel test, apply multiple testing adjustment and produce data for a qq-plot

```
## qq-plot
res1 <- read.table("default4.assoc.assoc.adjusted",header=TRUE)
png (file="Figures/default.assoc.png",width=5,height=5,units="in",pointsize=8,res=600)
plot(c(0,6), c(0,6), col="red", lwd=3, type="l", xlab="Expected (-logP)",
     ylab="Observed (-logP)", xlim=c(0,8), ylim=c(0,8), las=1, xaxs="i", yaxs="i", bty="l")
points(-log10(res1$QQ), -log10(res1$UNADJ), pch=23, cex=.4, bg="black")
dev.off()
```

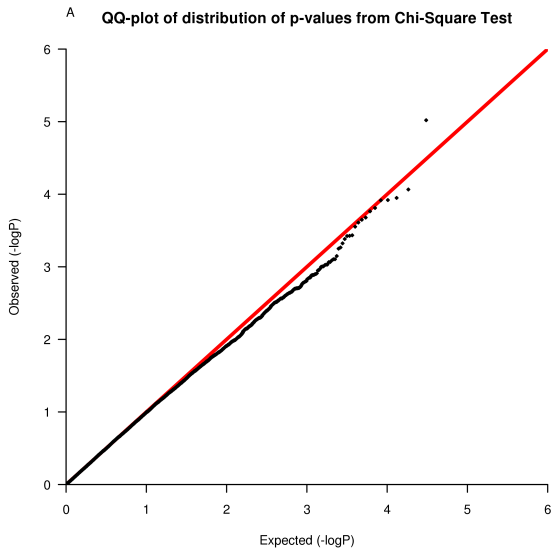

QQ-plot

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification



adjusted data

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

CHR		SNP	UNADJ	GC	QQ	BONF	HC
3		ARS-BFGL-NGS-113303	3.062e-07	3.062e-07	1.094e-05	0.01399	0.01
1		Hapmap57114-rs29012843	3.264e-05	3.264e-05	3.283e-05	1	
1	ARS-USMARC-Parent-DQ381153-rs29012842		3.264e-05	3.264e-05	5.472e-05	1	
21		Hapmap60593-rs29025761	3.453e-05	3.453e-05	7.661e-05	1	
11		BTA-93093-no-rs	6.214e-05	6.214e-05	9.85e-05	1	
27		UA-IFASA-1830	6.635e-05	6.635e-05	0.0001204	1	

Manhattan plot

GWAS

Matt Settles

GWAS
Case/Control
studies

Population
Stratification

<http://webpages.uidaho.edu/msettles/Rcode/wgplot.R>

