

WES GWAS by PLINK

Xulong Wang

February 22, 2016

```
library(ggplot2)

chrlen <- read.delim("~/Dropbox/GitHub/X/genomes/human.hg19.genome", header = F)
chrlen <- chrlen[match(paste0("chr", 1:22), chrlen$V1), ]
chrlen <- cumsum(as.numeric(chrlen$V2)) * 1e-6; names(chrlen) <- c(1:22)
chrmid <- diff(c(0, chrlen)) * 0.5 + c(0, chrlen[-length(chrlen)])
```

GitHub/wes/gwas/gwas.sh

Covariates: age, sex, and sequencing centers (Broad, Baylor, WashU). Broad and Baylor were coded as dummy variables (0/1).

```
options(stringsAsFactors = F)
plink <- read.table("~/Dropbox/GitHub/wes/gwas/plink_greg.assoc.logistic", header = T)
head(plink)
```

```
##   CHR     SNP     BP    A1 TEST NMISS      OR      STAT      P
## 1   1 1-13417 13417 CGAGA  ADD  9133 0.9714 -0.2662 7.901e-01
## 2   1 1-13417 13417 CGAGA COV1  9133 0.7532 -47.3300 0.000e+00
## 3   1 1-13417 13417 CGAGA COV2  9133 1.6840  8.5030 1.852e-17
## 4   1 1-13417 13417 CGAGA COV3  9133 1.4670  5.2700 1.366e-07
## 5   1 1-13417 13417 CGAGA COV4  9133 1.1060  1.2670 2.052e-01
## 6   1 1-69270 69270      G ADD  9133 1.0140  0.2458 8.058e-01
```

COV1/Age, COV2/Sex, COV3/Broad, were significant. COV4/Baylor was not significant. Sex was not significant by running Fisher's test on the AD status and Sex table. Sex also was also not significant if sex was the only covariate.

```
plink <- plink[plink$TEST == "ADD", ]
plink$POS <- c(0, chrlen)[plink$CHR] + plink$BP * 1e-6
```

Remove variants that were called only by Broad or the WashU/Baylor.

```
load("~/Dropbox/GitHub/wes/genotypes/suspicious_variants.rdt")

plink = plink[! plink$SNP %in% c(broad.na, other.na), ]
nrow(plink)

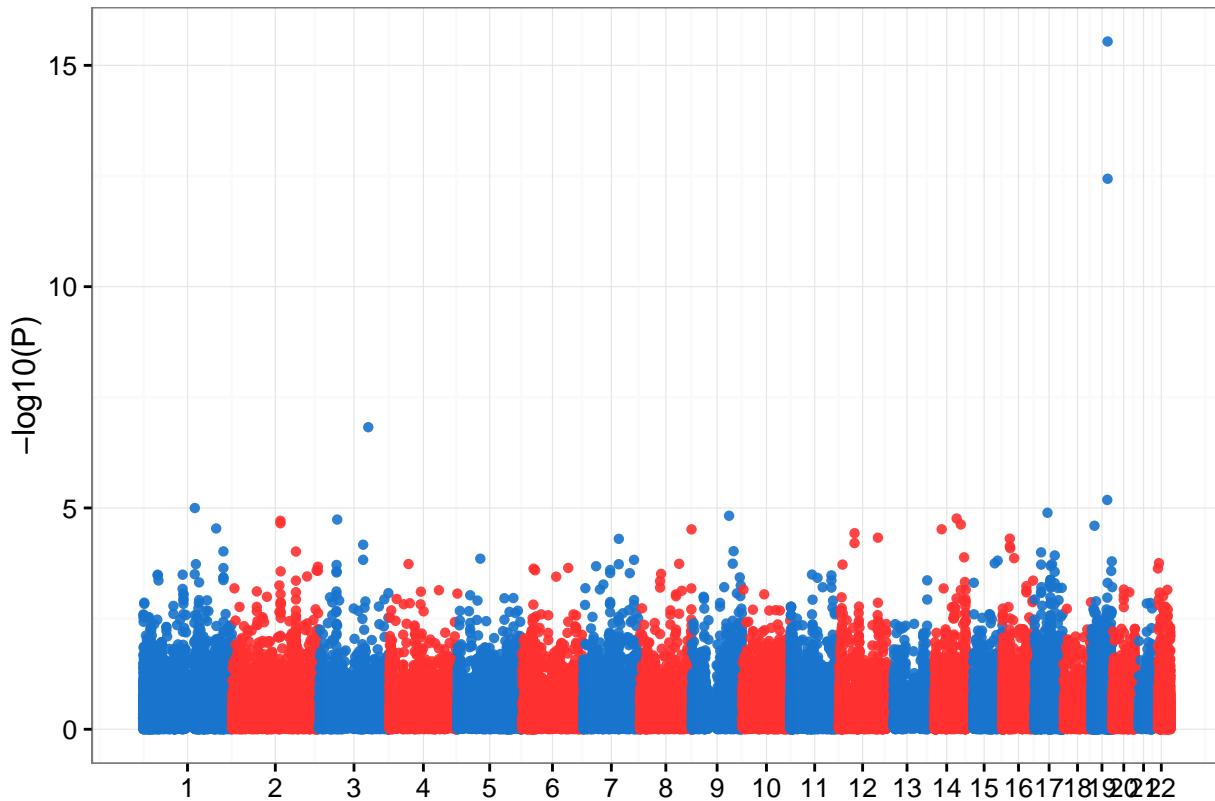
## [1] 59797
```

Manhattan

```
vId <- plink$SNP
chr <- as.numeric(gsub("-.*", "", vId))
pos <- as.numeric(gsub(".*-", "", vId)) * 1e-6 # Mb
pos <- c(0, chrlen)[chr] + pos

manhattan <- data.frame(uid = vId, chr = chr, pos = pos, pval = -log10(plink$P))
manhattan$col <- rep("o", nrow(manhattan))
manhattan$col[chr %% 2 == 1] <- "e"
manhattan$peak = "N"

ggplot(manhattan, aes(x = pos, y = pval, color = col, shape = peak)) +
  geom_point(alpha = 0.9) + # geom_hline(yintercept = 7.3, color = "black") +
  scale_x_continuous(breaks = chrmid, labels = names(chrlen), limits = c(0, max(pos))) +
  scale_color_manual(values = c("dodgerblue3", "firebrick1")) +
  theme_bw() + xlab("") + ylab("-log10(P)") + guides(shape = F, color = F) +
  theme(legend.key = element_blank())
```



ADSP WES results do not support the IGAP results.

```
load("~/Dropbox/GitHub/glmm/igap/igap.rdt")
```

```
table(plink$SNP %in% igap_s1$UID)
```

```
##  
## FALSE TRUE  
## 16447 43350
```

```
plink = plink[plink$SNP %in% igap_s1$UID, ]
igap = igap_s1[match(plink$SNP, igap_s1$UID), ]

cor(plink$P, igap$Pvalue)

## [1] 0.03229777

plot(-log10(plink$P), -log10(igap$Pvalue))
```

