

# CODE for training

---

We first extracted the post clumping SNPs (using default PRSice2 clumping options) from our training data set and rerun PRSice2 with the no-clump flag selected.

```
Rscript PRSice.R --cov-file ALL_COVS_ALS.txt --out Results -t
```

# CODE for Replication

---

```
> R
library("data.table")
VanRheenenTemp <- fread("Summary_Statistics_GWAS_2016", header = F)
temp <- VanRheenenTemp[order(VanRheenenTemp$SNP, VanRheenenTemp$P), ]
temp$dupe <- duplicated(temp$SNP)
VanRheenen <- subset(temp, dupe == F)
listOfPathways <- read.table("significant.txt", header = T)
names(listOfPathways) <- c("id")
for(i in 1:length(listOfPathways$id))
{
    instrumentId <- as.character(listOfPathways$id[i])
    thisDataTemp <- fread(file = instrumentId, header = F)
```

```

thisData <- subset(thisDataTemp, V5 == "Y")
merged <- merge(thisData, VanRheenen, by.x = "V1", by.y = "V1")
merged$allele <- toupper(merged$A1)
toPlink <- merged[,c("V1", "allele", "b")]
write.table(toPlink, file = paste(instrumentId, ".signif", "plink.txt"),
}

```

```

> PLINK
cat significant.txt | while read LINE
do
echo "THIS IS LINE" $LINE
plink --bfile REPLICATION --score $LINE.bpe.txt --out $LINE.sig
done

```

```

> R
library("data.table")
listOfProfiles <- read.table("significant.txt", header = T)
names(listOfProfiles) <- c("id")
covs1 <- fread("/data/LNG/ALS_PATHWAYS/PHENO_COVARIATES_ALS_LN")
covs2 <- fread("/data/LNG/ALS_PATHWAYS/ALS_Hits_toCOVs.raw", header = T)
covs <- merge(covs1, covs2, by = "FID")
covs$index <- paste(covs$FID, covs$IID, sep = "_")
covs$CASE <- covs$PHENO - 1
outPut <- matrix(ncol = 4, nrow = length(listOfProfiles$id), data = NA)
colnames(outPut) <- c("pathway", "b", "se", "p")
for(i in 1:length(listOfProfiles$id))
{
  profileName <- as.character(listOfProfiles$id[i])
  profile <- fread(file = paste(profileName, ".significant.txt"), header = T)
  profile$index <- paste(profile$FID, profile$IID, sep = "_")
  data <- merge(covs, profile, by = "index")
}

```

```
meanControls <- mean(data$SCORE[data$CASE == 0])
sdControls <- sd(data$SCORE[data$CASE == 0])
data$zSCORE <- (data$SCORE - meanControls)/sdControls
grsTest <- glm(CASE ~ zSCORE + age_at_onset + gender +
beta <- summary(grsTest)$coefficients["zSCORE", "Estimate"]
se <- summary(grsTest)$coefficients["zSCORE", "Std. Error"]
p <- summary(grsTest)$coefficients["zSCORE", "Pr(>|z|)"]
outPut[i,1] <- profileName
outPut[i,2] <- beta
outPut[i,3] <- se
outPut[i,4] <- p
}
write.table(outPut, "REPLICATION_plink.tab", quote = F, sep =
```