

Summary

Step 1: You will simulate some imaging measures, similar to those used in genetic association studies. Step 2: You will create polygenic scores, using summary data from a recent GWAS of schizophrenia. Step 3: You will run a simple regression to explore the association between your schizophrenia polygenic score & simulated data.

Dependences; install package built for your OS.

- Download R: <https://www.stats.bris.ac.uk/R/>
- Download plink: <https://www.cog-genomics.org/plink2> (stable, beta))

BLUE = annotation, look here for hints.

BLACK = commands, copy paste these into R (step 1 & 3) or plink (step 2).

RED = these are paths / commands that need changing to match your local pc/mac environment.

STEP 1: SIMULATE IMAGING DATA (in R, copy paste commands in)

```
###this generates 2 normally distributed values, which are similar to those extracted from  
freesurfer reconstructions (mm3)
```

```
set.seed(1234)  
hippocampal_volume <- rnorm(n=270,mean=4000,sd=3)
```

```
set.seed(1234)  
intracranial_volume <- rnorm(n=270,mean=10000,sd=3)
```

```
##simulates age, sex variables
```

```
set.seed(42)  
sex <- sample(LETTERS[c(13,6)],270,replace=T)  
set.seed(42)  
N=270  
age <- rlnbinom(N, 35, .5)
```

```
##assign ids to random data, these will be used for polygenic risk scoring  
pid <- read.table("/path/to/your/downloaded/data/genotype_ids.txt",header=F,sep="")
```

```
##merge into dataframe
```

```
phenotype.file <- data.frame(pid,hippocampal_volume,intracranial_volume,age,sex)  
colnames(phenotype.file)[1] <- "IID"
```

```
###load SCZ summary data (just to look at the schizophrenia summary data)
```

```
install.packages('qqman') ###this is a tool to look at gwas summary data  
library(qqman)
```

```
summarydata <-  
read.table("/path/to/your/downloaded/data/schizophrenia_summary_stats.score
```

```
",header=T,sep="")  
  
manhattan(summarydata, col = c("blue4", "orange3"), ymax = 12)
```

STEP 2: PROCESS GENOTYPE DATA (in cmd or terminal, copy paste command in)

```
#### open 'cmd' or 'terminal'  
###change directory to the downloaded data file  
###this includes a clumped version of the GWAS results found at  
https://www.med.unc.edu/pgc/results-and-downloads  
  
###the genotype data is just rebranded hapmap_r23a  
###first we have to merge all the chromosomes back together (storage issues)  
plink --merge-list all_chrs.txt --make-bed --out testgenotypes  
  
###then we perform the scoring using the summary stats  
plink --bfile testgenotypes --score schizophrenia_summary_stats.score 1 2 3 no-mean-  
imputation --q-score-range q.score.range schizophrenia_summary_stats.score 1 4 --out  
testgenotypes_scored_by_schizophrenia
```

STEP3: RUN POLYGENIC SCORE REGRESSION (in R, copy paste commands in)

```
####in R, load in the polygenic score and merge with phenotype.file  
  
risk.profile.score <-  
read.table("path/to/your/data/tutorial_genotypes/testgenotypes_scored_by_schizophrenia.P  
T_0.05.profile",header=T,sep="")  
  
rps <- (risk.profile.score)[c(2,6)]  
  
combined.file <- merge(risk.profile.score[,c(2,6)],phenotype.file, by = "IID")  
  
###normalise the score data  
combined.file$zSCORE.rps <- (combined.file$SCORE -  
mean(combined.file$SCORE))/sd(combined.file$SCORE)  
  
###run regression  
rps.model1 <-  
lm(hippocampal_volume~zSCORE.rps+age+sex+intracranial_volume,data=combined.file)  
  
summary(rps.model1)  
  
####plot correlation  
  
install.packages("ggplot2")  
library(ggplot2)  
  
rps.plot1 <- ggplot(combined.file, aes(x=zSCORE.rps, y= hippocampal_volume)) +  
geom_smooth(method=lm) + geom_point() + theme_classic()  
  
###you can repeat this for the other polygenic thresholds
```