01_plot_HM3_obs_v_pred

Heather Wheeler

2016-10-13 11:58:18

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(ggplot2)
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

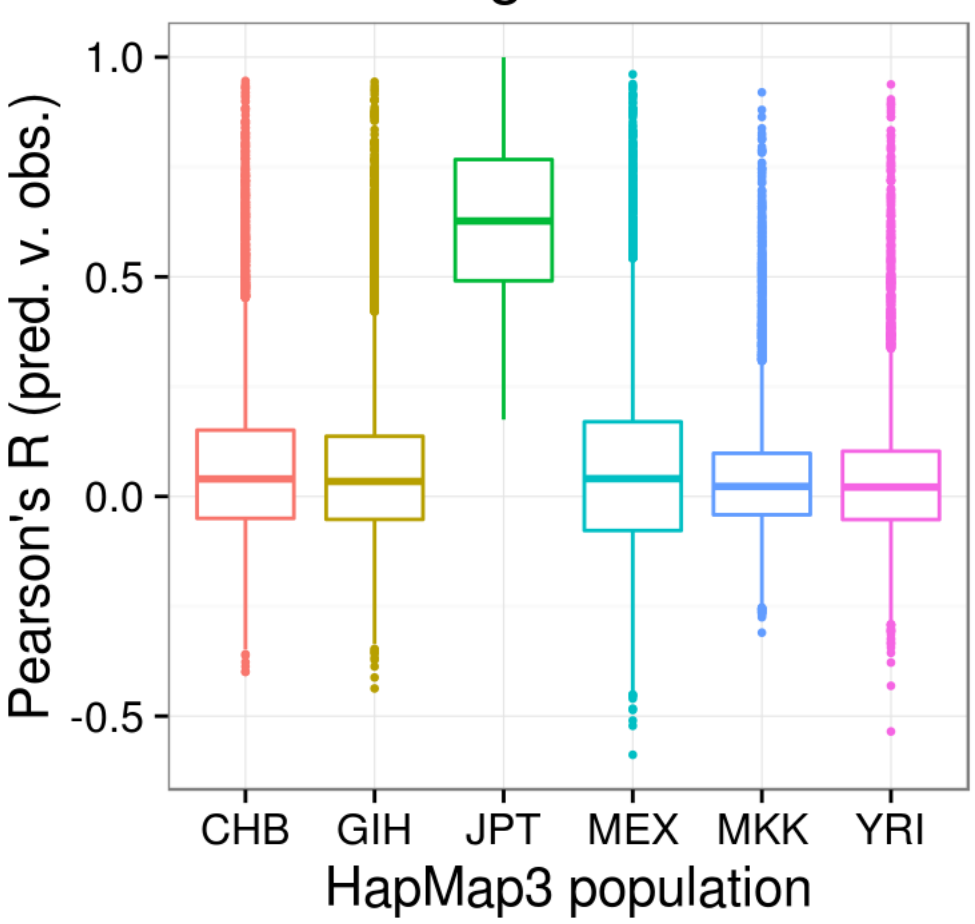
```
## The following objects are masked from 'package:dplyr':
##
##     between, last
```

```
"%&%" = function(a,b) paste(a,b,sep="")
px.dir = "/home/natalie/Desktop/PrediXcan/"
obs.dir = "/home/natalie/Desktop/Matrix_eQTL/Expression/Expression_for_ElasticNet/"
```

```
pops <- c('CHB','GIH','JPT','MEX','MKK','YRI')
dbs <- c('JPT','MKK')


for(d in dbs){
  for(pop in pops){
    cat(d,pop,'\n')
    predexp1 <- data.frame(fread(px.dir %&% d %&% "_db_" %&% pop %&% "_predicted/pred
icted_expression.txt"))
```

```
    rownames(predexp1) <- predexp1[,1]
    obsexp <- read.table(obs.dir %&% pop %&% "_Expression.txt",header=T)
    rownames(obsexp)<-obsexp[,1]
    tobsexp <- t(obsexp[,-1]) #transpose the observed exp matrix

    #get the same genes in obs & pred and sort by ID and gene
    obs2 <- data.frame(tobsexp[rownames(tobsexp) %in% rownames(predexp1),colnames(tob
sexp) %in% colnames(predexp1)])
    obs <- obs2[order(rownames(obs2)),order(colnames(obs2))]

    pred2 <- predexp1[rownames(predexp1) %in% rownames(obs2),colnames(predexp1) %in%
colnames(obs2)]
    pred <- pred2[order(rownames(pred2)),order(colnames(pred2))]

    #convert to matrix and transpose
    predexp <- as.matrix(pred)
    obsexp <- as.matrix(obs)

    popres <- matrix(NA,ncol=1,nrow=dim(obsexp)[2])

    for(i in 1:dim(obsexp)[2]){
      corres <- cor.test(predexp[,i] , obsexp[,i])
      r <- signif(corres$estimate,3)
      popres[i,] <- r
    }
    if(exists("allres") == FALSE){
      allres = popres
    }else{
      allres<- cbind(allres,popres)
    }
  }

  colnames(allres) <- pops
  #print(ggpairs(allres,diag=list(continuous='blank'),title="Weights: GEUVADIS " %&%
geu %&% ", HapMap3 pred v obs R"))
  print(summary(allres))
  gres <- gather(data.frame(allres),key=pop,value=R)
  print(ggplot(gres,aes(x=pop,y=R,color=pop)) + geom_boxplot(outlier.size = 0.5) + th
eme_bw(15) + guides(color=FALSE) + ggtitle("Weights: " %&% d) + xlab("HapMap3 populat
ion")+ylab("Pearson's R (pred. v. obs.)"))
  rownames(allres) <- colnames(obs)
  write.table(allres,px.dir %&% "R_pred_v_obs_" %&% d %&% "_db.txt",quote=F)
  rm("allres")
}
```

```
## JPT CHB
## JPT GIH
## JPT JPT
## JPT MEX
## JPT MKK
## JPT YRI
##       CHB                GIH                JPT               MEX
##  Min.   :-0.3990   Min.   :-0.4370   Min.   :0.1750   Min.   :-0.5880
##  1st Qu.:-0.0500   1st Qu.:-0.0523   1st Qu.:0.4910   1st Qu.:-0.0776
##  Median : 0.0399   Median : 0.0342   Median :0.6270   Median : 0.0404
##  Mean   : 0.0765   Mean   : 0.0611   Mean   :0.6335   Mean   : 0.0619
##  3rd Qu.: 0.1510   3rd Qu.: 0.1370   3rd Qu.:0.7670   3rd Qu.: 0.1700
##  Max.   : 0.9460   Max.   : 0.9440   Max.   :1.0000   Max.   : 0.9610
##  NA's   :509       NA's   :419                        NA's   :341
##       MKK                YRI
##  Min.   :-0.3100   Min.   :-0.5350
##  1st Qu.:-0.0418   1st Qu.:-0.0526
##  Median : 0.0228   Median : 0.0213
##  Mean   : 0.0440   Mean   : 0.0384
##  3rd Qu.: 0.0983   3rd Qu.: 0.1030
##  Max.   : 0.9200   Max.   : 0.9380
##  NA's   :743       NA's   :455
```



```
## MKK CHB
## MKK GIH
## MKK JPT
## MKK MEX
## MKK MKK
## MKK YRI
##       CHB                GIH                JPT                MEX
##  Min.   :-0.4070   Min.   :-0.4010   Min.   :-0.4920   Min.   :-0.5960
##  1st Qu.:-0.0539   1st Qu.:-0.0546   1st Qu.:-0.0533   1st Qu.:-0.0684
##  Median : 0.0367   Median : 0.0362   Median : 0.0343   Median : 0.0468
##  Mean   : 0.0693   Mean   : 0.0695   Mean   : 0.0665   Mean   : 0.0740
##  3rd Qu.: 0.1440   3rd Qu.: 0.1470   3rd Qu.: 0.1400   3rd Qu.: 0.1830
##  Max.   : 0.9530   Max.   : 0.9690   Max.   : 0.9390   Max.   : 0.9750
##  NA's   :1468      NA's   :1102      NA's   :1751      NA's   :1087
##       MKK                YRI
##  Min.   :0.1150   Min.   :-0.3600
##  1st Qu.:0.4330   1st Qu.:-0.0402
##  Median :0.5550   Median : 0.0426
##  Mean   :0.5608   Mean   : 0.0780
##  3rd Qu.:0.6780   3rd Qu.: 0.1430
##  Max.   :0.9990   Max.   : 0.9490
##                   NA's   :394
```