01_plot_HM3_obs_v_pred

Heather Wheeler

2016-09-27 15:41:01

```
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'
```

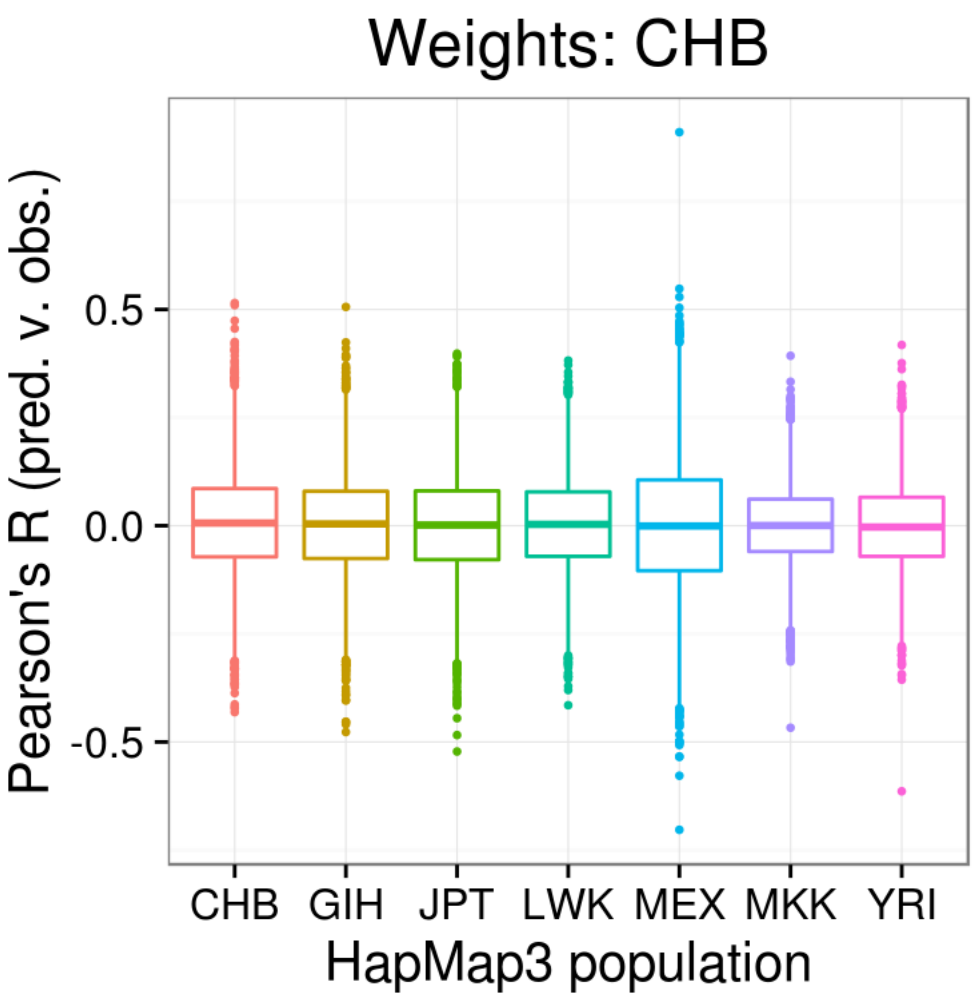
```
## The following objects are masked from 'package:dplyr':
##
##     between, last
```

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

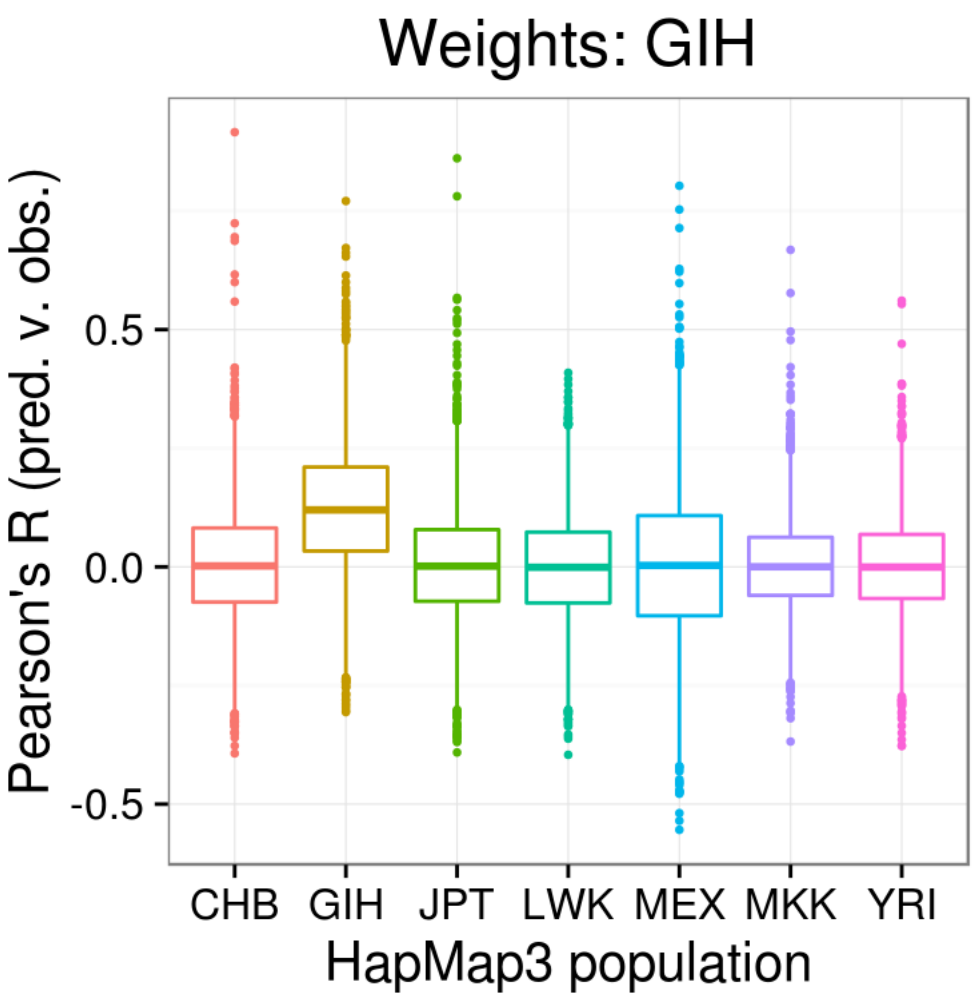
```
pops <- c('CHB','GIH','JPT','LWK','MEX','MKK','YRI')
##add other pops in here
dbs <- c('CHB','GIH','MEX','YRI')

for(d in dbs){
  for(pop in pops){
    predexp1 <- data.frame(fread(px.dir %&% d %&% "_db_" %&% pop %&% "_predicted_0.5/
predicted_expression.txt"))
    rownames(predexp1) <- predexp1[,1]
```

```
obsexp <- data.frame(fread(obs.dir %&% pop %&% "_Expression.txt"))
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[,colnames(tobsexp) %in% colnames(predexp1)])
obs <- obs2[order(rownames(obs2)),order(colnames(obs2))]

pred2 <- predexp1[,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")
}
```

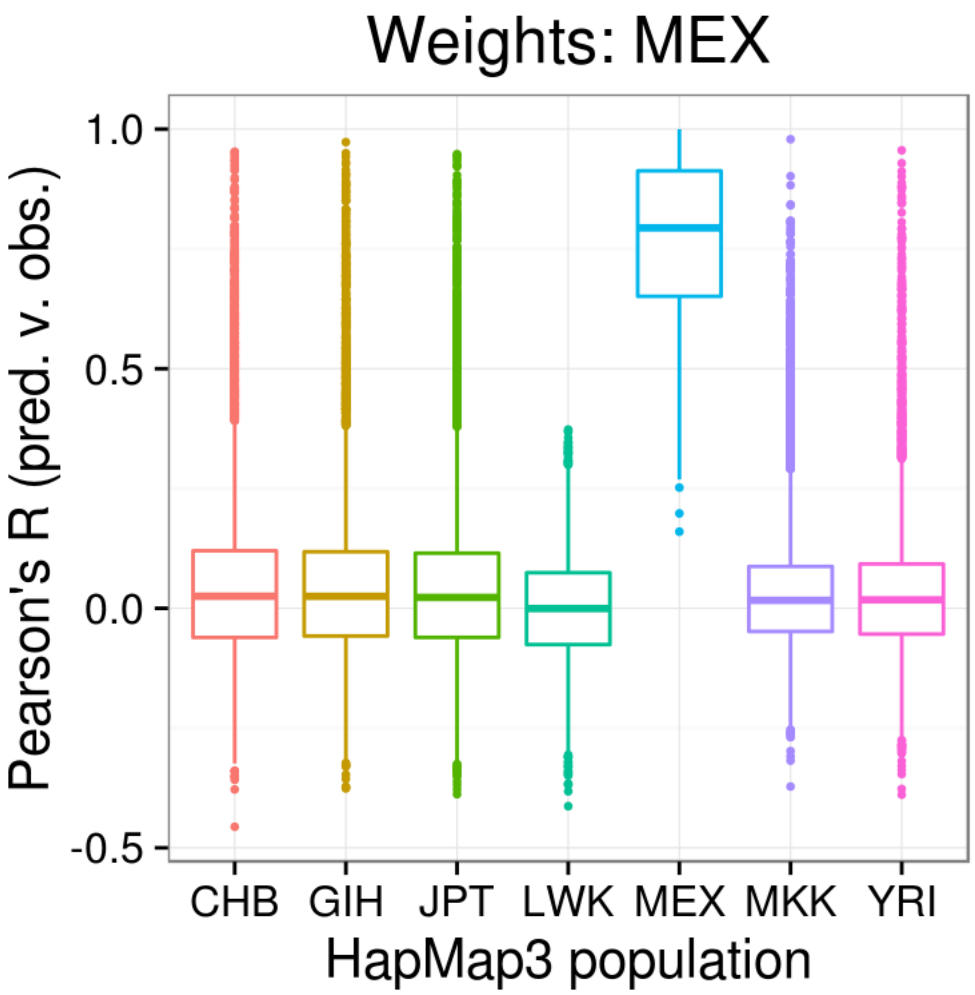
```
##       CHB                 GIH                 JPT                 LWK
##  Min.   :-0.431000   Min.   :-0.4770   Min.   :-0.5220   Min.   :-0.4150
##  1st Qu.:-0.072000   1st Qu.:-0.0758   1st Qu.:-0.0785   1st Qu.:-0.0709
##  Median : 0.006270   Median : 0.0042   Median : 0.0016   Median : 0.0031
##  Mean   : 0.007379   Mean   : 0.0019   Mean   : 0.0010   Mean   : 0.0029
##  3rd Qu.: 0.085900   3rd Qu.: 0.0798   3rd Qu.: 0.0805   3rd Qu.: 0.0783
##  Max.   : 0.515000   Max.   : 0.5060   Max.   : 0.3980   Max.   : 0.3820
##                      NA's   :462       NA's   :718       NA's   :711
##       MEX                 MKK                 YRI
##  Min.   :-0.7030   Min.   :-0.4670   Min.   :-0.6140
##  1st Qu.:-0.1040   1st Qu.:-0.0596   1st Qu.:-0.0708
##  Median :-0.0007   Median : 0.0005   Median :-0.0031
##  Mean   : 0.0002   Mean   : 0.0001   Mean   :-0.0029
##  3rd Qu.: 0.1060   3rd Qu.: 0.0614   3rd Qu.: 0.0659
##  Max.   : 0.9100   Max.   : 0.3930   Max.   : 0.4180
##  NA's   :410       NA's   :755       NA's   :534
```



```
##       CHB                 GIH                 JPT                 LWK
##  Min.   :-0.3930   Min.   :-0.30600   Min.   :-0.3910   Min.   :-0.3960
##  1st Qu.:-0.0743   1st Qu.: 0.03338   1st Qu.:-0.0726   1st Qu.:-0.0760
##  Median : 0.0020   Median : 0.12000   Median : 0.0016   Median :-0.0009
##  Mean   : 0.0041   Mean   : 0.12270   Mean   : 0.0040   Mean   :-0.0009
##  3rd Qu.: 0.0819   3rd Qu.: 0.21025   3rd Qu.: 0.0784   3rd Qu.: 0.0732
##  Max.   : 0.9160   Max.   : 0.77100   Max.   : 0.8610   Max.   : 0.4090
##  NA's   :1110                         NA's   :1226      NA's   :906
##       MEX                 MKK                 YRI
##  Min.   :-0.5540   Min.   :-0.3680   Min.   :-0.3770
##  1st Qu.:-0.1030   1st Qu.:-0.0601   1st Qu.:-0.0667
##  Median : 0.0030   Median : 0.0001   Median :-0.0005
##  Mean   : 0.0029   Mean   : 0.0028   Mean   : 0.0012
##  3rd Qu.: 0.1080   3rd Qu.: 0.0623   3rd Qu.: 0.0683
##  Max.   : 0.8030   Max.   : 0.6680   Max.   : 0.5610
##  NA's   :595       NA's   :965       NA's   :717
```



```
##       CHB                 GIH                 JPT                 LWK
##  Min.   :-0.4560   Min.   :-0.3760   Min.   :-0.3880   Min.   :-0.4130
##  1st Qu.:-0.0609   1st Qu.:-0.0579   1st Qu.:-0.0608   1st Qu.:-0.0758
##  Median : 0.0250   Median : 0.0249   Median : 0.0226   Median :-0.0006
##  Mean   : 0.0476   Mean   : 0.0495   Mean   : 0.0465   Mean   :-0.0008
##  3rd Qu.: 0.1200   3rd Qu.: 0.1180   3rd Qu.: 0.1150   3rd Qu.: 0.0743
##  Max.   : 0.9530   Max.   : 0.9730   Max.   : 0.9480   Max.   : 0.3730
##  NA's   :793       NA's   :443       NA's   :929       NA's   :664
##       MEX                 MKK                 YRI
##  Min.   :0.1600   Min.   :-0.3720   Min.   :-0.3890
##  1st Qu.:0.6510   1st Qu.:-0.0484   1st Qu.:-0.0539
##  Median :0.7940   Median : 0.0165   Median : 0.0176
##  Mean   :0.7753   Mean   : 0.0358   Mean   : 0.0328
##  3rd Qu.:0.9130   3rd Qu.: 0.0872   3rd Qu.: 0.0925
##  Max.   :1.0000   Max.   : 0.9790   Max.   : 0.9560
##                   NA's   :756       NA's   :464
```



```
##       CHB                 GIH                 JPT                 LWK
##  Min.   :-0.4540   Min.   :-0.6510   Min.   :-0.5190   Min.   :-0.4090
##  1st Qu.:-0.0597   1st Qu.:-0.0550   1st Qu.:-0.0600   1st Qu.:-0.0774
##  Median : 0.0284   Median : 0.0296   Median : 0.0254   Median : 0.0011
##  Mean   : 0.0569   Mean   : 0.0595   Mean   : 0.0530   Mean   :-0.0002
##  3rd Qu.: 0.1280   3rd Qu.: 0.1300   3rd Qu.: 0.1230   3rd Qu.: 0.0755
##  Max.   : 0.9460   Max.   : 0.9700   Max.   : 0.9480   Max.   : 0.4060
##  NA's   :1550      NA's   :1165      NA's   :1675      NA's   :619
##       MEX                 MKK                 YRI
##  Min.   :-0.5810   Min.   :-0.3080   Min.   :0.1560
##  1st Qu.:-0.0775   1st Qu.:-0.0376   1st Qu.:0.4960
##  Median : 0.0328   Median : 0.0340   Median :0.6430
##  Mean   : 0.0586   Mean   : 0.0699   Mean   :0.6418
##  3rd Qu.: 0.1660   3rd Qu.: 0.1240   3rd Qu.:0.7810
##  Max.   : 0.9590   Max.   : 0.9280   Max.   :1.0000
##  NA's   :1079      NA's   :642
```