01_plot_HM3_obs_v_pred

Heather Wheeler

2016-10-13 13:24:08

```r
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
```

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

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

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

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

```r
pops <- c('CHB','GIH','JPT','MEX','MKK','YRI')
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")
}
```
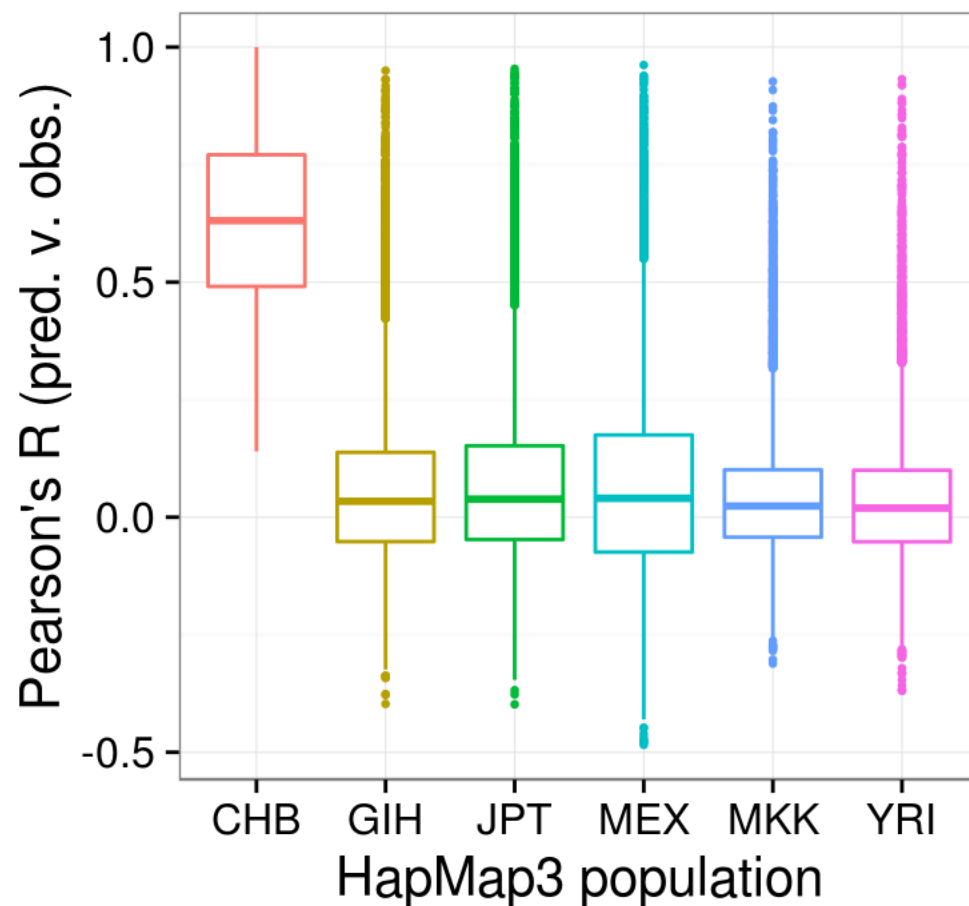
```
##      CHB                 GIH                  JPT                  MEX
## Min.   :0.1400    Min.   :-0.3970    Min.   :-0.3980    Min.   :-0.4840
## 1st Qu.:0.4910    1st Qu.:-0.0519    1st Qu.:-0.0474    1st Qu.:-0.0742
## Median :0.6310    Median : 0.0340    Median : 0.0384    Median : 0.0403
## Mean   :0.6364    Mean   : 0.0638    Mean   : 0.0779    Mean   : 0.0661
## 3rd Qu.:0.7710    3rd Qu.: 0.1380    3rd Qu.: 0.1520    3rd Qu.: 0.1750
## Max.   :1.0000    Max.   : 0.9500    Max.   : 0.9540    Max.   : 0.9620
##                   NA's   :498        NA's   :717        NA's   :427
##      MKK                 YRI
## Min.   :-0.3110    Min.   :-0.3690
## 1st Qu.:-0.0425    1st Qu.:-0.0523
## Median : 0.0238    Median : 0.0195
## Mean   : 0.0482    Mean   : 0.0403
## 3rd Qu.: 0.1010    3rd Qu.: 0.0999
## Max.   : 0.9270    Max.   : 0.9320
## NA's   :687        NA's   :496
```
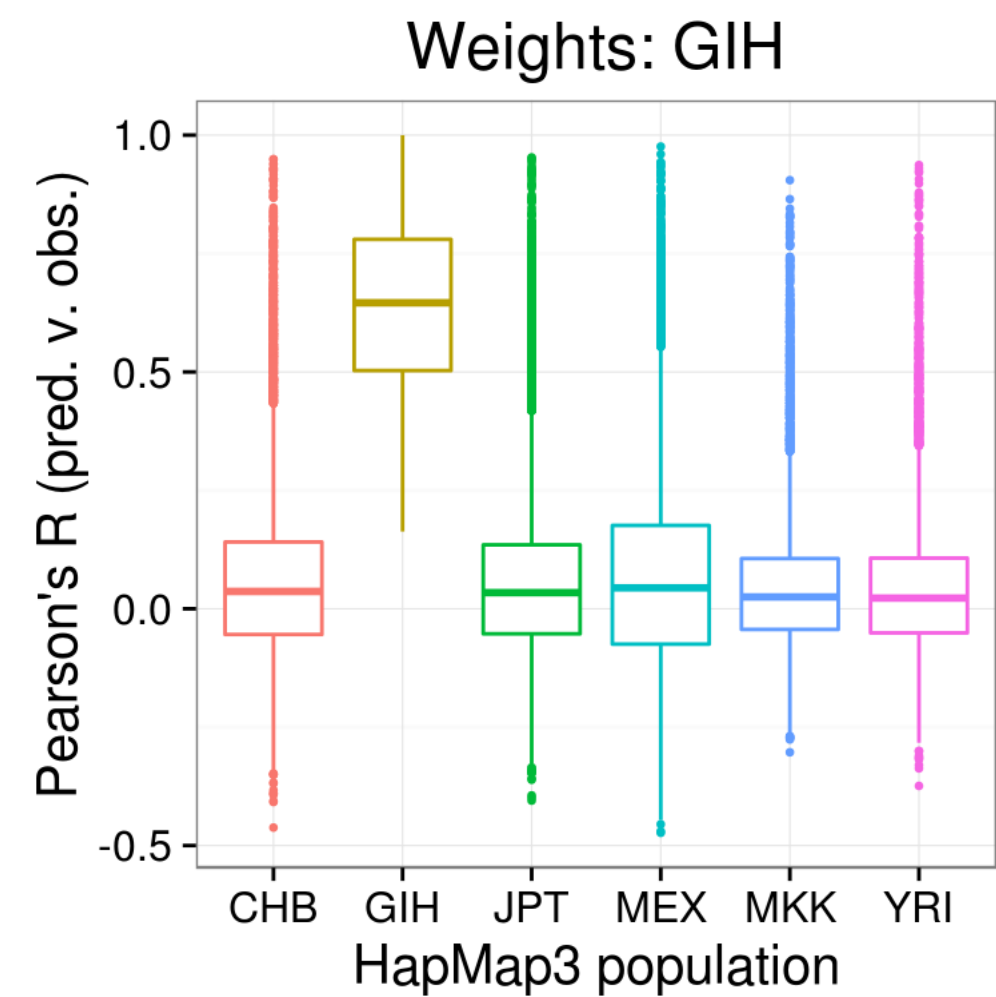

Weights: CHB

```
##       CHB                GIH                JPT                MEX
## Min.   :-0.4620    Min.   :0.1630     Min.   :-0.4050    Min.   :-0.4730
## 1st Qu.:-0.0543    1st Qu.:0.5030     1st Qu.:-0.0529    1st Qu.:-0.0747
## Median : 0.0365    Median :0.6460     Median : 0.0340    Median : 0.0441
## Mean   : 0.0665    Mean   :0.6464     Mean   : 0.0652    Mean   : 0.0711
## 3rd Qu.: 0.1410    3rd Qu.:0.7800     3rd Qu.: 0.1350    3rd Qu.: 0.1760
## Max.   : 0.9490    Max.   :1.0000     Max.   : 0.9530    Max.   : 0.9760
## NA's   :927                           NA's   :1087       NA's   :468
##       MKK                YRI
## Min.   :-0.3030    Min.   :-0.3740
## 1st Qu.:-0.0437    1st Qu.:-0.0510
## Median : 0.0250    Median : 0.0225
## Mean   : 0.0515    Mean   : 0.0450
## 3rd Qu.: 0.1060    3rd Qu.: 0.1070
## Max.   : 0.9050    Max.   : 0.9370
## NA's   :811        NA's   :590
```
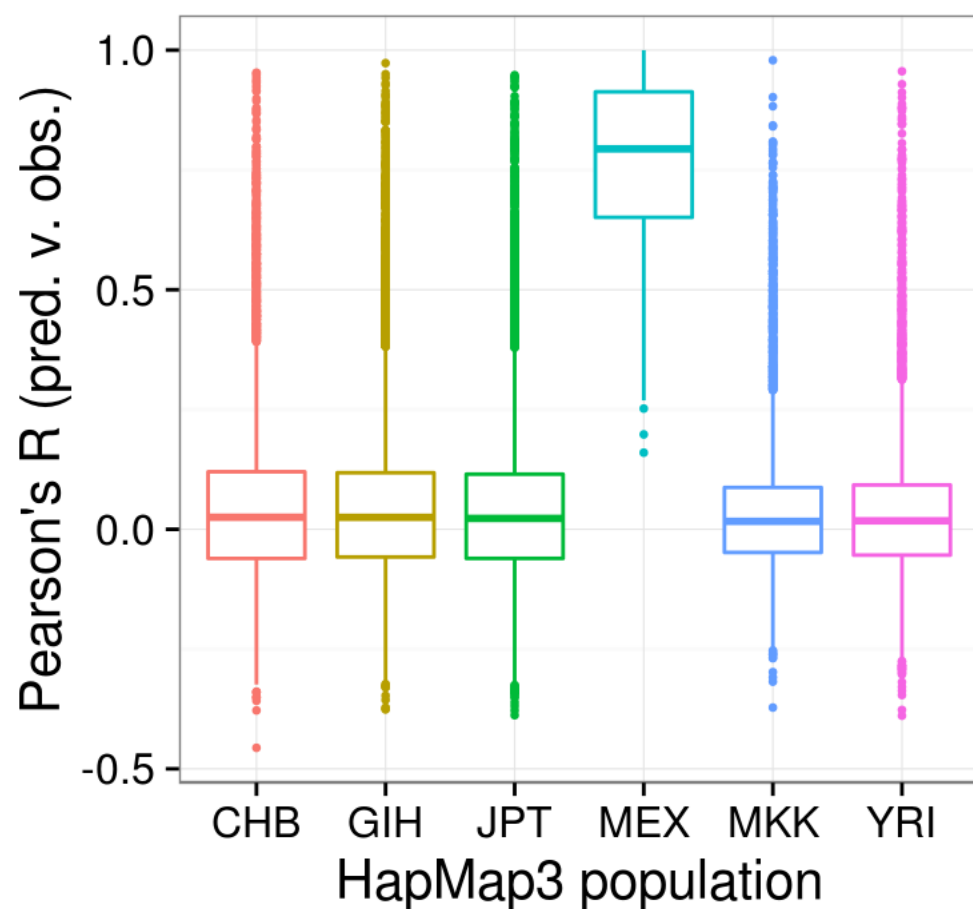

Weights: GIH

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


Weights: MEX

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


Weights: YRI