Transcription factors

Tae Kim

3/18/2019

scale the data and regress out the mean

```
X$A[X$A > 0] = 1
X$A = scale(X$A) * sqrt(length(X$A)) / sqrt(length(X$A)-1)
Y = resid(lm(Y~X$A))
Y = scale(Y) * sqrt(length(X$A)) / sqrt(length(X$A)-1)
```

read TF data

```
TF = fread("../data/TFcheckpoint.txt", header=TRUE)
TF = setDF(TF)
TF = TF[,1]
results = getBM(attributes = c('hgnc_symbol','ensembl_gene_id'),
                filters = "hgnc_symbol",
                values = TF,
                mart = ensembl)

genes = colnames(Y)[which(colnames(Y) %in% results$ensembl_gene_id)]
results = results[match(genes, results$ensembl_gene_id),]

scores = newscores = matrix(0, length(genes), ncol(Y)-1)
for (i in 1:length(genes)){
  ind = which(colnames(Y) == results$ensembl_gene_id[i])
  W = store_W_c(Y[,ind], Y[, -ind])
  scores[i,] = get_score_W_c(X$A, W)
}

ref_chisq = qchisq(seq(0,1, length=ncol(Y)-1), 1)
par(mfrow = c(1,2))
for (i in 1:length(genes)){
    if(results$hgnc_symbol[i] %in% c("FOXM1", "POLR2G")){
      qqplot(ref_chisq, scores[i,],
            main = paste(results$hgnc_symbol[i], 'unadjusted'),
            cex = 0.5)
      abline(0,1,col = 'red')
  }
}
```
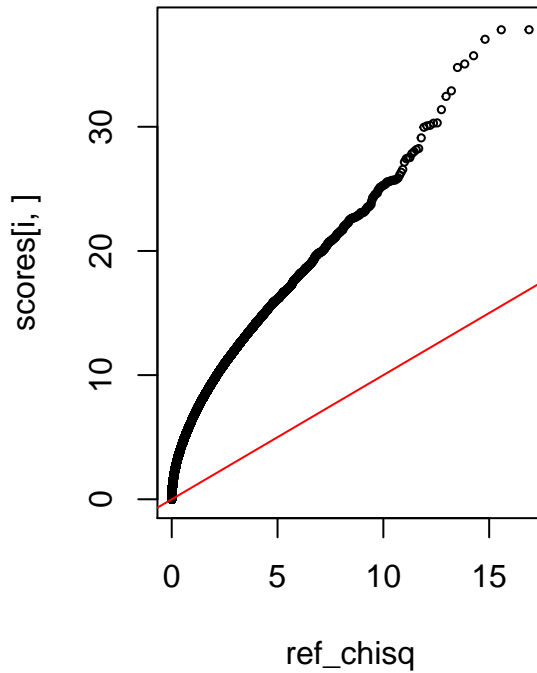
## POLR2G unadjusted
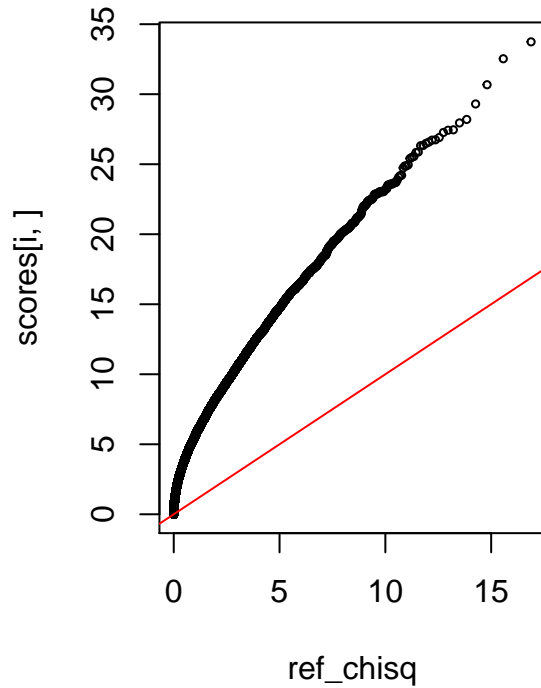
## FOXM1 unadjusted



```
par(mfrow = c(1,1))
d = rowSums(scores)
df = data.frame(d = d, genes = results$hgnc_symbol, xaxis = 1:length(d))
plot(d,
     ylab = 'd',
     main = 'Transcription Factors')
abline(h=25190, col='red')
ind = which(d > ncol(Y) + 200 * sqrt(2*ncol(Y)))
df = df[ind, ]
with(df, text(d ~ xaxis, labels = genes, pos = 4))
```

Transcription Factors