ExpressionAnalysis

Alvaro Barbeira

June 28, 2016

Preliminaries

This report will compare predicted gene expressions (built using different transcriptome models) to a measured (reference) transcriptome.

Reference expression is loaded from:

```
REFERENCE <- load_file(WB_PATH)
```

Predicted data is loaded from:

```
PRS_BA <- load_file(PRS_BA_PATH)
PRS_BA$tag = "PRS_PB8K_BETA_ALL"

PRS_BB <- load_file(PRS_BB_PATH)
PRS_BB$tag = "PRS_PB8K_BETA_BEST"

PRS_ZA <- load_file(PRS_ZA_PATH)
PRS_ZA$tag = "PRS_PB8K_ZSCORE_ALL"

PRS_ZB <- load_file(PRS_ZB_PATH)
PRS_ZB$tag = "PRS_PB8K_ZSCORE_BEST"

DGN <- load_file(DGN_PATH)
DGN$tag = "EN_DGN"

predicted <- list(DGN, PRS_BA, PRS_BB, PRS_ZA, PRS_ZB)
```

We will build a linear regression of predicted expression over reference expression with the next function.

```
comparison_results <- function(predicted, expected) {
  d1 <- predicted
  d2 <- expected

  d1 <- d1[rownames(d1) %in% rownames(d2), colnames(d1) %in% colnames(d2)]
  d1 <- d1[order(rownames(d1)),]

  d2 <- d2[rownames(d2) %in% rownames(d1), colnames(d2) %in% colnames(d1)]
  d2 <- d2[order(rownames(d2)),]

  results <- build_results(d1, d2)
  results$tag <- paste0(d1$tag, "_vs_", d2$tag)
  results
}
```

And we'll be using some standard issue R tools:

```
library(ggplot2)
```

Expression analysis

We'll gather all predicted expressions and run their comparison to thye reference expression.

```
WB_DGN_COMPARISON <- comparison_results(DGN, REFERENCE)
WB_PRS_BA_COMPARISON <- comparison_results(PRS_BA, REFERENCE)
WB_PRS_BB_COMPARISON <- comparison_results(PRS_BB, REFERENCE)
WB_PRS_ZA_COMPARISON <- comparison_results(PRS_ZA, REFERENCE)
WB_PRS_ZB_COMPARISON <- comparison_results(PRS_ZB, REFERENCE)
```

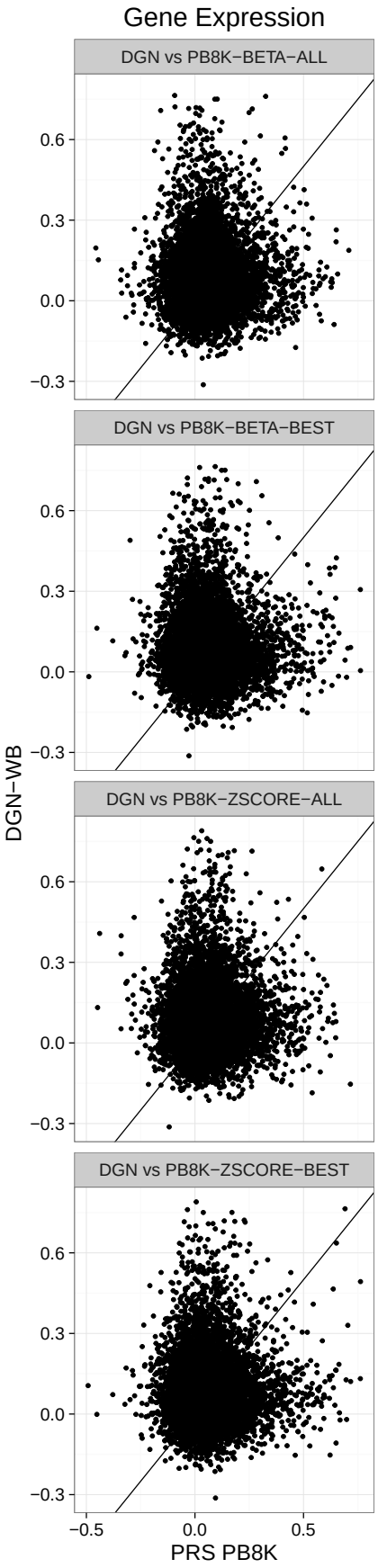
And then, we'll build a plot comparing the pearson coefficients among them

```
build_comparison_plot_data <- function(d1, d2, facet) {
  d1 <- d1[rownames(d1) %in% rownames(d2), ]
  d1 <- d1[order(rownames(d1)),]

  d2 <- d2[rownames(d2) %in% rownames(d1), ]
  d2 <- d2[order(rownames(d2)),]

  result <-  data.frame(x = d1$R, y = d2$R, the_facet=facet)
  result
}
```

Lets build some plots displaying this. **y** axis is Predicted **DGN**+ gene expression, **x** axis is predicted **PRS** gene expression, gathered by PRS model type([BETA/ZSCORE][BEST/ALL])

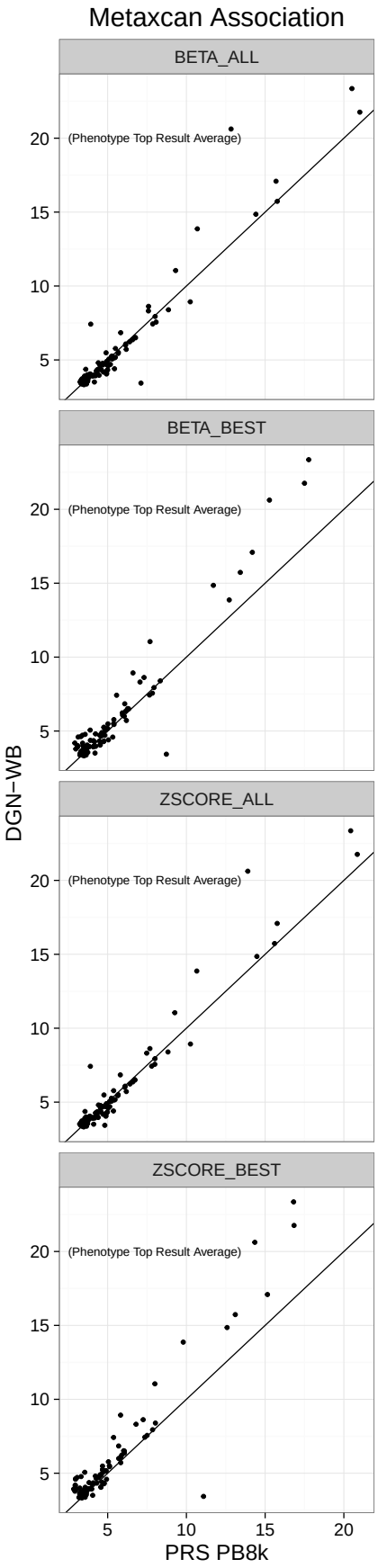

Gene Expression
DGN vs PB8K−BETA−ALL
DGN vs PB8K−BETA−BEST
DGN vs PB8K−ZSCORE−ALL
DGN vs PB8K−ZSCORE−BEST
DGN−WB
PRS PB8K
0.6
0.3
0.0
−0.3
−0.5
0.0
0.5

Metaxcan Results Analysis

Lets compare Metaxcan results when using the *DGN* model versus the *PRS* model.

```
## [1] "dgn skip CARDIoGRAM_CAD"
## [1] "dgn skip CARDIoGRAM_CAD"
## [1] "dgn skip CARDIoGRAM_CAD"
## [1] "dgn skip CARDIoGRAM_CAD"
## [1] "dgn skip pgc.adhd"
## [1] "dgn skip pgc.adhd"
## [1] "dgn skip pgc.adhd"
## [1] "dgn skip pgc.adhd"
```

```
## Warning: Removed 12 rows containing missing values (geom_point).
```


Metaxcan Association
BETA_ALL
BETA_BEST
ZSCORE_ALL
ZSCORE_BEST
(Phenotype Top Result Average)
DGN−WB
PRS PB8k
5
10
15
20