

Looking and visualizing individual genes: PIN1

date: June 25, 2014

The goal of this report is to look at PIN1 genes in the LCM data.

```
library(rmarkdown)
render("PIN1.Rmd", "pdf_document")
```

Specify gene of interest

```
gene1 <- "Solyc10g078370"
gene2 <- "Solyc10g080880"
gene3 <- "Solyc03g118740"
```

```
library(reshape2)
library(ggplot2)
library(plyr)
```

```
countData <- read.csv("../data/normalized_read_count.csv")
```

Set up the dataframe for analysis and visualization

```
head(countData)
```

```
##           X tf2ambr1 tf2ambr3 tf2ambr4 tf2ambr6 tf2aother1
## 1 Solyc00g005040.2.1  6.0449  1.1849  0.08117  7.0721  2.136e-01
## 2 Solyc00g005050.2.1 17.5503  8.3433  3.81482  8.3945  3.561e-01
## 3 Solyc00g005060.1.1  0.1465 10.7562  0.06630  0.2796  1.388e-17
## 4 Solyc00g005070.1.1  4.8208 49.9373  9.35780  0.5832  1.134e+01
## 5 Solyc00g005080.1.1  7.0960 12.0352  9.50432 18.5486  1.780e+00
## 6 Solyc00g005150.1.1  1.4389  0.3072  0.00000  2.4538  2.136e-01
##   tf2aother2 tf2aother4 tf2aother7 tf2bmr2 tf2bmr5 tf2bmr6 tf2bmr7
## 1 8.628e-01  0.000e+00  4.134e+00  0.0797  0.0000  0.4807  0.910
## 2 1.756e+00  2.666e+00  4.100e-01  1.5143  2.0529  8.3241  9.698
## 3 1.388e-17  1.388e-17  1.388e-17 39.2198  0.0000  0.0000  0.000
## 4 4.761e+00  2.980e+01  1.092e+01 475.8553  0.5787  0.0000  3.306
## 5 1.070e+01  8.702e+00  9.964e+00 21.9756  8.9378 15.5335 11.108
## 6 0.000e+00  1.781e+00  2.499e+00  0.1594  0.7986  0.3704  3.741
##   tf2bmr3 tf2bmr4 tf2bmr6 tf2cmbr1.4 tf2cmbr3 tf2cmbr6 tf2cmbr7
## 1 2.6718  2.9960  1.0400  0.19425  1.9319  6.003  2.287
## 2 7.6530 12.8115 13.9365  3.35700 15.6790 12.014  6.780
## 3 0.8983  0.5569  0.6358  0.03642  0.3887  0.000  0.000
## 4 3.2208  5.1564  6.2361 30.68191  4.0895  5.622  5.776
## 5 10.3549  6.6472  5.9642 30.86637  1.4849  5.497  8.007
## 6 0.0000  0.0000  0.3964  1.69136  1.5024  0.000  0.000
```

```
##      tf2cother2 tf2cother5 tf2cother6 tf2cother7 wtambr2 wtambr4 wtambr5
## 1      2.7089      3.9771      2.4497      0.7023  0.5210  1.5839  7.1866
## 2      3.3915      3.0412     10.6643     10.2733  1.5230  4.8497  4.9325
## 3      0.9326      0.2604      1.3326      0.8252  0.1603  0.0000  0.8881
## 4      3.5883      3.1124      1.9268      3.8031 41.3735  2.3180  7.8886
## 5      8.3383      9.2101      8.5928      2.0241 19.0512 12.5689 16.4443
## 6      0.9177      3.2909      0.4408      0.0000  0.2805  0.7957  1.7274
##      wtaother1 wtaother5 wtaother6 wtaother7 wtaother8 wtbmbr2 wtbmbr3
## 1      0.143      0.1762 1.0073042  0.019732      2.004  0.5940  2.1243
## 2     11.057      6.4862 9.0547131  2.392538      3.011  9.7030  1.0309
## 3      0.000      0.0000 0.0006911  0.009866      2.003  0.2529  0.9928
## 4      4.696      1.3633 5.0348399  5.510508      6.013  5.5179  1.4796
## 5     10.132     16.2462 6.0630040 11.174775      7.023 11.0368  6.0589
## 6      1.146      2.2655 0.0024189  0.034531      2.004  0.0000  0.0000
##      wtbmbr6 wtbmbr8 wtbother1.4 wtbother3 wtbother5 wtbother8 wtcnbr10
## 1  0.9167  0.4284  1.462e+00 4.177e+00 8.084e-01 2.847e+00 0.03860
## 2  6.7745  2.0470  5.182e+00 1.256e+01 2.866e+00 1.339e+01 5.78363
## 3  0.1486  3.7447  1.388e-17 1.388e-17 1.388e-17 1.388e-17 1.09692
## 4  3.0671 40.5179  1.595e+00 2.660e+00 5.775e+00 3.816e+00 5.49736
## 5  9.9635  3.1895  5.315e+00 1.528e+01 2.940e+00 1.045e+01 0.24350
## 6  1.4997  0.1428  5.315e-01 8.585e-01 2.940e-01 1.920e+00 0.02969
##      wtcnbr1.4.6 wtcnbr2 wtcnbr3 wtcnbr7 wtcnbr9 wtcother1.3.4 wtcother2
## 1      5.6209      1.503  0.0000  0.1193 0.10210      0.7675  0.3561
## 2     21.6291     11.056  1.8104  6.3556 9.53545      9.5583 10.3781
## 3      1.7497      0.000  0.0000  1.3238 0.03927     22.8475  0.3858
## 4      6.2600      3.372  0.4806  7.7951 1.67951     300.4438 17.4571
## 5      1.0269      9.804 23.6396  6.2916 9.47282     29.6349 37.1138
## 6      0.1882      1.639  1.1578  2.5947 1.35316      4.7879  1.7939
##      wtcother6
## 1      4.962
## 2     13.879
## 3      0.000
## 4      0.000
## 5      8.338
## 6      1.491
```

#Melt Data

```
countData <- melt(countData)
```

```
## Using X as id variables
```

```
head(countData)
```

```
##           X variable  value
## 1 Solyc00g005040.2.1 tf2ambr1 6.0449
## 2 Solyc00g005050.2.1 tf2ambr1 17.5503
## 3 Solyc00g005060.1.1 tf2ambr1  0.1465
## 4 Solyc00g005070.1.1 tf2ambr1  4.8208
## 5 Solyc00g005080.1.1 tf2ambr1  7.0960
## 6 Solyc00g005150.1.1 tf2ambr1  1.4389
```

```
colnames(countData) <- c("gene", "sample", "count")
head(countData)
```

```
##           gene      sample      count
## 1 Solyc00g005040.2.1 tf2ambr1  6.0449
## 2 Solyc00g005050.2.1 tf2ambr1 17.5503
## 3 Solyc00g005060.1.1 tf2ambr1  0.1465
## 4 Solyc00g005070.1.1 tf2ambr1  4.8208
## 5 Solyc00g005080.1.1 tf2ambr1  7.0960
## 6 Solyc00g005150.1.1 tf2ambr1  1.4389
```

```
#set genotype
```

```
countData$genotype <- ifelse(grepl("wt", countData$sample, ignore.case = T), "wt",
                             ifelse(grepl("tf2", countData$sample, ignore.case = T), "tf2", "unknown"))
```

```
#countData$genotype <- if(grepl("wt", countData$sample)) {
#  "wt"
# } else {
  "tf2"
```

```
## [1] "tf2"
```

```
# }
```

```
#set type
```

```
countData$tissue <- ifelse(grepl("other", countData$sample, ignore.case = T), "other",
                           ifelse(grepl("mbr", countData$sample, ignore.case = T), "mbr", "unknown"))
```

```
head(countData)
```

```
##           gene      sample      count genotype tissue
## 1 Solyc00g005040.2.1 tf2ambr1  6.0449      tf2     mbr
## 2 Solyc00g005050.2.1 tf2ambr1 17.5503      tf2     mbr
## 3 Solyc00g005060.1.1 tf2ambr1  0.1465      tf2     mbr
## 4 Solyc00g005070.1.1 tf2ambr1  4.8208      tf2     mbr
## 5 Solyc00g005080.1.1 tf2ambr1  7.0960      tf2     mbr
## 6 Solyc00g005150.1.1 tf2ambr1  1.4389      tf2     mbr
```

```
#Set Region
```

```
countData$region <- ifelse(grepl("a", countData$sample, ignore.case = T), "A",
                           ifelse(grepl("c", countData$sample, ignore.case = T), "C", "B"))
```

Plot Data

Subsetting for genes of interest. This is specified at the start of the script.

```

gene1Data <- subset(countData, grepl(gene1, countData$gene))
gene2Data <- subset(countData, grepl(gene2, countData$gene))
gene3Data <- subset(countData, grepl(gene3, countData$gene))

graphData <- rbind(gene1Data, gene2Data, gene3Data)
graphData$sample2 <- paste(graphData$region, graphData$tissue, sep= "")

```

Summarize Data

```

sumGraph <- dplyr::ddply(graphData, c("gene", "sample2", "genotype"), summarise,
  N      = length(count),
  mean   = mean(count),
  sd     = sd(count),
  se     = sd / sqrt(N) )

```

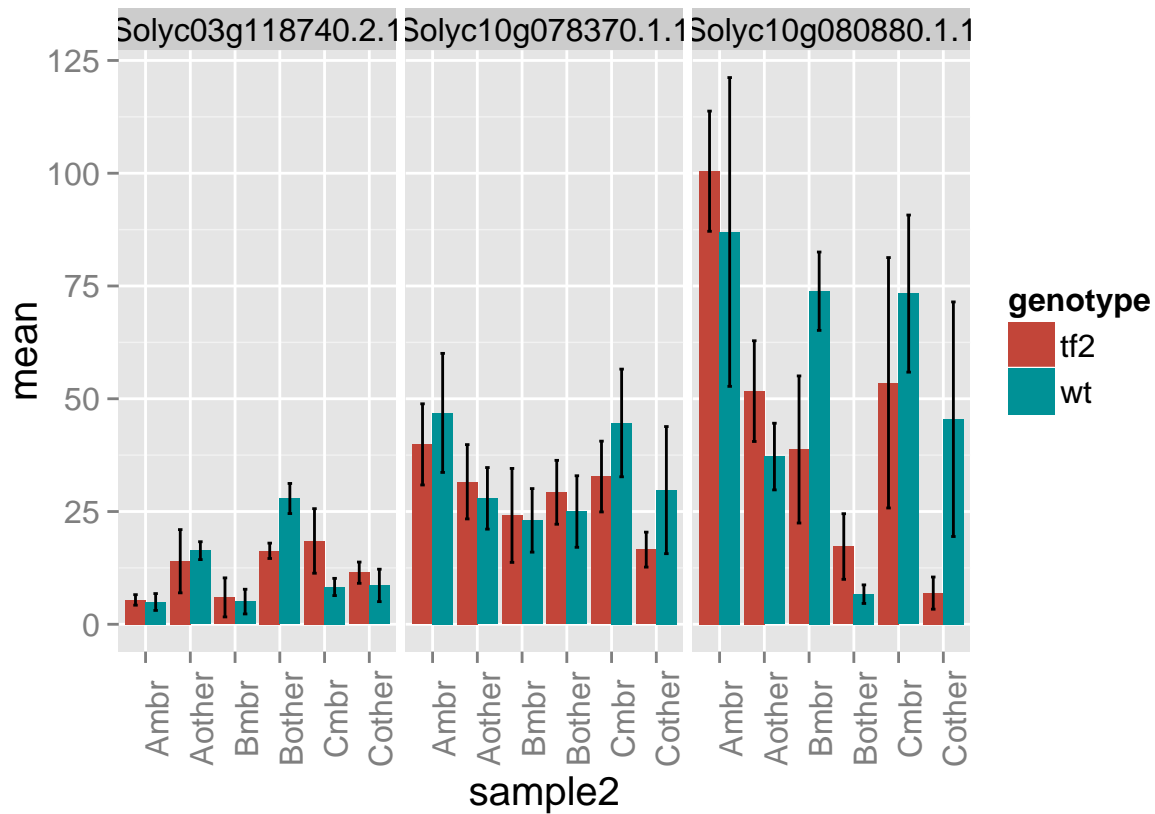
Plot 1

This organizing the data to look at differences between genotype.

```

ggplot(sumGraph, aes(sample2, mean, fill = genotype)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(.~gene) +
  geom_errorbar(aes(ymin=mean-se, ymax=mean+se),
    width=.2, # Width of the error bars
    position=position_dodge(.9)) +
  theme(text = element_text(size=15),
    axis.text.x = element_text(angle=90, vjust=1)) +
  scale_fill_hue(l=47)

```



Plot 2

This organizing the data to look at differences per region.

```
ggplot(sumGraph, aes(gene, mean, fill = sample2)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(.~genotype) +
  geom_errorbar(aes(ymin=mean-se, ymax=mean+se),
    width=.2, # Width of the error bars
    position=position_dodge(.9)) +
  theme(text = element_text(size=15),
    axis.text.x = element_text(angle=90, vjust=1)) +
  scale_fill_hue(1=47)
```

