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("curatedGenes.Rmd", "pdf_document")
```

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

Read in Data

```
countData <- read.csv("../../data/normalized_read_count.csv")
geneList1 <- read.csv("pnas.1402835111.sd06.csv")
geneList2 <- read.csv("pnas.1402835111.sd07.csv")
```

Set up the counts dataframe for analysis and visualization

```
#Melt Data
countData <- melt(countData)
```

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

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


#set genotype

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

#set type

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

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

#Set tissueRegion

countData$type <- paste(countData$region, countData$tissue,  sep = "")
```

Subset for each gene

First I take the genes and make them into a characters, so that I can find them and loop through. Takes ~ 20 minutes.

```
genes2 <- geneList1[,1]
genes1 <- geneList2[,1]

genesOfInterest1 <- levels(genes1)
genesOfInterest2 <- levels(genes2)

genesOfInterest <- c(genesOfInterest1, genesOfInterest2)
```

Intitialize datafram to fit in all the visualization information.

```
largeGeneList <- data.frame(t(rep(NA,7)))
colnames(largeGeneList) <- c("type", "genotype", "N", "mean", "sd", "se", "gene")
head(largeGeneList)
```

```
##   type genotype  N mean sd se gene
## 1   NA       NA NA   NA NA NA   NA
```

Making the data table for Visualization

This loop (Takes about 20 minutes)

```
for(GENE in genesOfInterest) {

  if(length(grep(GENE, countData$gene)) < 1){ #this is just making sure that the curated
    next;
    }

  geneData <- subset(countData, grepl(GENE, countData$gene))

  sumGraph <- ddply(geneData, c("type", "genotype"), summarise,
               N    = length(count),
               mean = mean(count),
               sd   = sd(count),
               se   = sd / sqrt(N))

  sumGraph$gene <- GENE

largeGeneList  <- rbind(largeGeneList, sumGraph) #bind together all the new rows per loop.

}

#make for use
finalList <- largeGeneList
finalList <- finalList[-1,] #remove the first row

#get the log2 of mean.
```

```
finalList$log2Mean <- log2(finalList$mean) # why are there negative numbers now?
finalList <- subset(finalList, log2Mean > 0 ) #fix that problem, but must go back to figure out why

#Is order for the lines to correct properly, they must be grouped by both genotype and gene.  So I am g

finalList <- within(finalList, lineGroup <- paste(genotype, gene,sep='.'))
head(finalList)
```

```
##     type genotype N  mean     sd    se           gene log2Mean
## 2   Ambr      tf2 4 34.02 12.718 6.359 Solyc00g009100    5.088
## 3   Ambr       wt 3 39.99 14.698 8.486 Solyc00g009100    5.322
## 4 Aother      tf2 4 39.31 17.291 8.646 Solyc00g009100    5.297
## 5 Aother       wt 5 30.72  7.405 3.311 Solyc00g009100    4.941
## 6   Bmbr      tf2 3 27.89 15.936 9.201 Solyc00g009100    4.802
## 7   Bmbr       wt 4 36.28  5.319 2.659 Solyc00g009100    5.181
##            lineGroup
## 2 tf2.Solyc00g009100
## 3  wt.Solyc00g009100
## 4 tf2.Solyc00g009100
## 5  wt.Solyc00g009100
## 6 tf2.Solyc00g009100
## 7  wt.Solyc00g009100
```

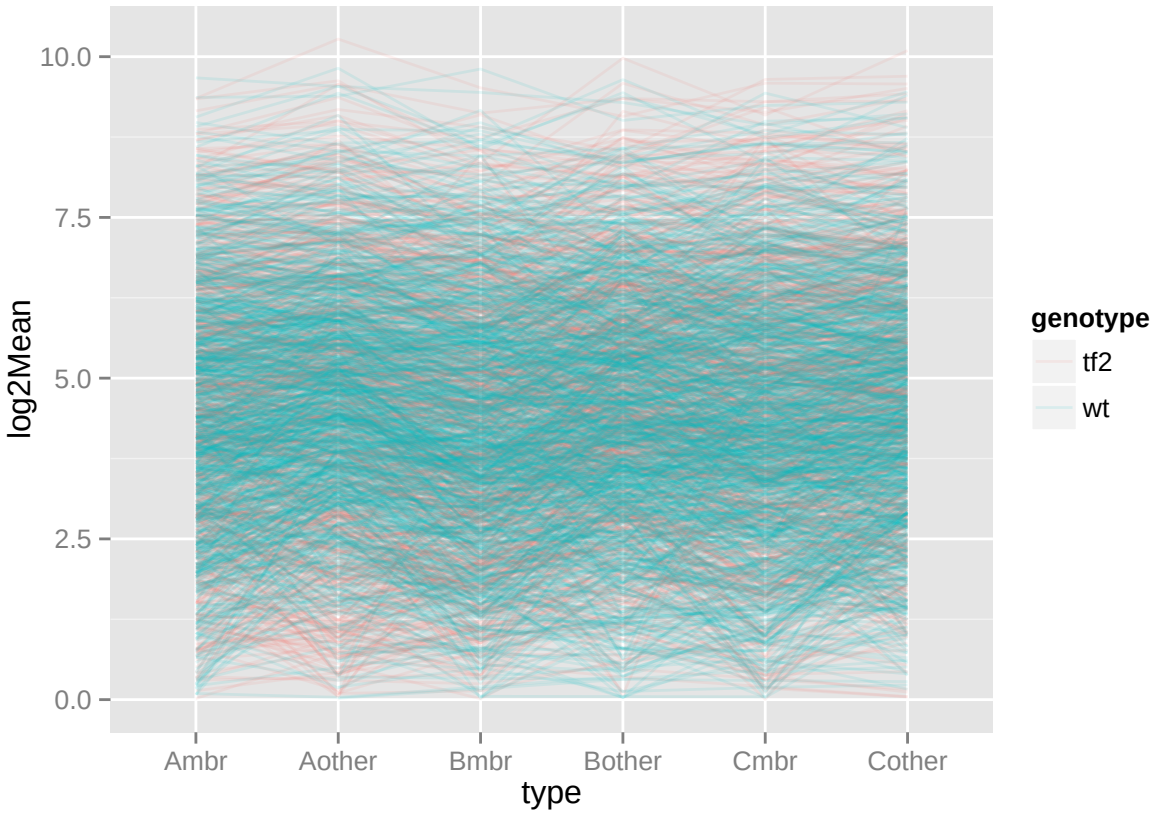
```
#for subsetting and optimiztion
#finalListSub <- finalList[1:100,]
```

Visualization

Now that I have the dataset, I can begin visualization. the main ways I want to visualize are:

1. All genes, all tissue types, colored by genotype.

```
ggplot(finalList, aes(type, log2Mean, group = lineGroup, color =  genotype )) +
  geom_line(alpha = .1, (aes(color = factor(genotype)))) +
  geom_point(alpha = .0)
```


10.0
7.5
5.0
2.5
0.0
log2Mean
Ambr
Aother
Bmbr
Bother
Cmbr
Cother
type
genotype
tf2
wt