Assessing flavanone 3 beta-hydroxylase (F3H) expression using scaled RNA-Seq counts per million (cpm) measurements

Ann Loraine

June 12, 2023

Introduction

This Markdown uses scaled RNA-Seq counts data to assess expression of Solyc02g083860.3 and PRAM_26184, encoding flavanone 3-dioxygenase (F3H) from tomato genome assemblies SL4 and SL5, respectively, in tomato pollen tubes germinated *in vitro* and undergoing a heat stress treatment lasting 15, 30, 45, and 75 minutes.

A genome-wide, scaled counts data containing scaled expression data for these genes and all other genes in the SL4 and SL5 genome assemblies will be saved to files named `results/muday-144-SL4_counts-salmon_scaled.txt` and `results/muday-144-SL5_counts-salmon_scaled.txt`.

For more information about this gene, see its NCBI sequence record:

- https://www.ncbi.nlm.nih.gov/protein/NP_001316412.1

Background

Previously, we found that F3H expression was higher in mutant genotype `are,` in which the F3H gene (encoded by `Solyc02g083860.3` in assembly SL4 and by `Solyc02g083860.3` in assembly SL5) contains a point mutation and produces an inactive protein. Also, we found that two over-expression lines exhibited no elevated expression compared to their wild-type, non-transgenic progenitor line `VF36`. The previous experiment tested gene expression after 2 hours of heat treatment. The data set analyzed here comes from heat-treated and control samples collected at 15, 30, 45, and 75 minutes of heat treatment. The analysis shown here will investigate if the previously observed trend with respect of F3H expression re-occurred in this new data set.

Results

Load functions:

```
source("Common.R")
```

```
## Loading required package: limma
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'

## The following object is masked from 'package:limma':
##
##     plotMA

## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which.max, which.min

##
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:utils':
##
##     findMatches

## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname

## Loading required package: IRanges

## Loading required package: GenomicRanges

## Loading required package: GenomeInfoDb

## Loading required package: SummarizedExperiment

## Loading required package: MatrixGenerics

## Loading required package: matrixStats

##
## Attaching package: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':
##
##     colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##     colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##     colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##     colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##     colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##     colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##     colWeightedMeans, colWeightedMedians, colWeightedSds,
##     colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##     rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##     rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##     rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##     rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##     rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
```

```
##     rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##     rowWeightedSds, rowWeightedVars

## Loading required package: Biobase

## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.

##
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
##
##     rowMedians

## The following objects are masked from 'package:matrixStats':
##
##     anyMissing, rowMedians

## Loading required package: ggplot2

## Loading required package: ggrepel
```

Read `results/muday-144-SL4_counts-salmon.txt` and ‘results/muday-144-SL5_counts-salmon.txt:

```
sl4_unscaled=getCounts(assembly="SL4",keep_description = T)
sl5_unscaled=getCounts(assembly="SL5",keep_description = T)
```

Load edgeR library:

```
library(edgeR)
```

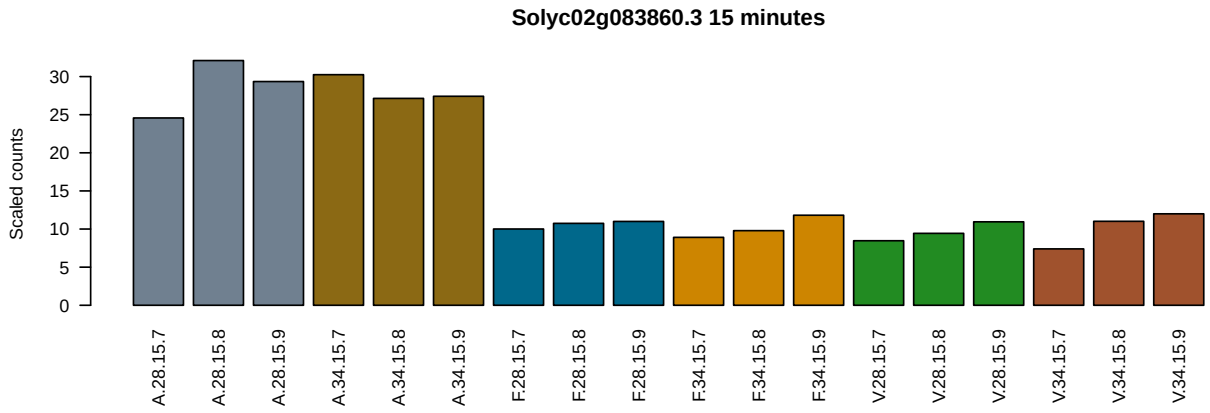
Scale the data using edgeR function `cpm`:

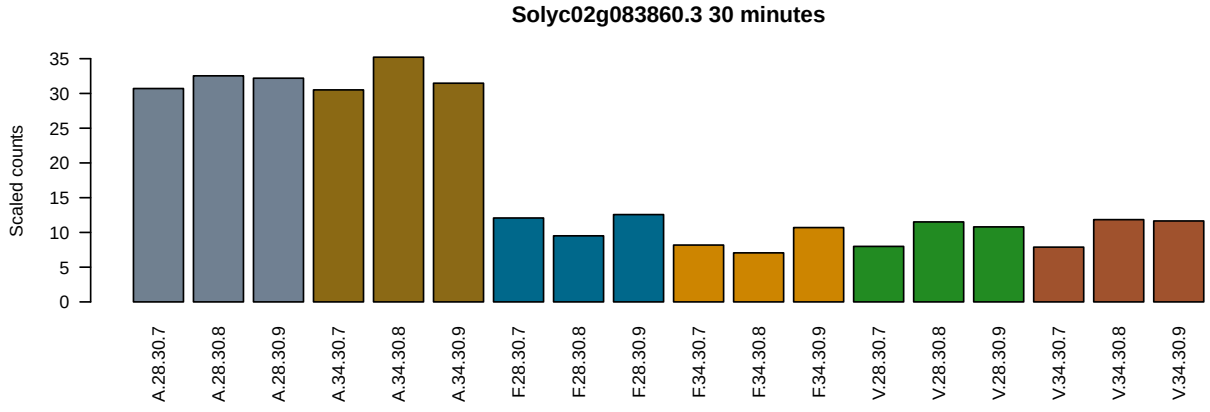
```
description_index = which(names(sl4_unscaled)=="description")
sl4_scaled = data.frame(cpm(sl4_unscaled[,-description_index]))
sl4_scaled$description=sl4_unscaled$description
description_index = which(names(sl5_unscaled)=="description")
sl5_scaled = data.frame(cpm(sl5_unscaled[,-description_index]))
sl5_scaled$description=sl5_unscaled$description
```

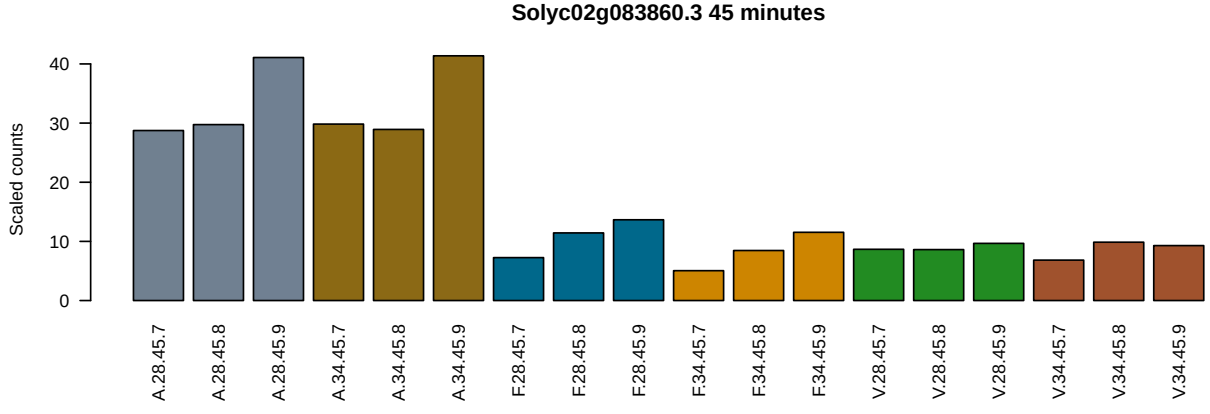
Create plot that shows scaled gene expression for SL4 gene `Solyc02g083860.3`:

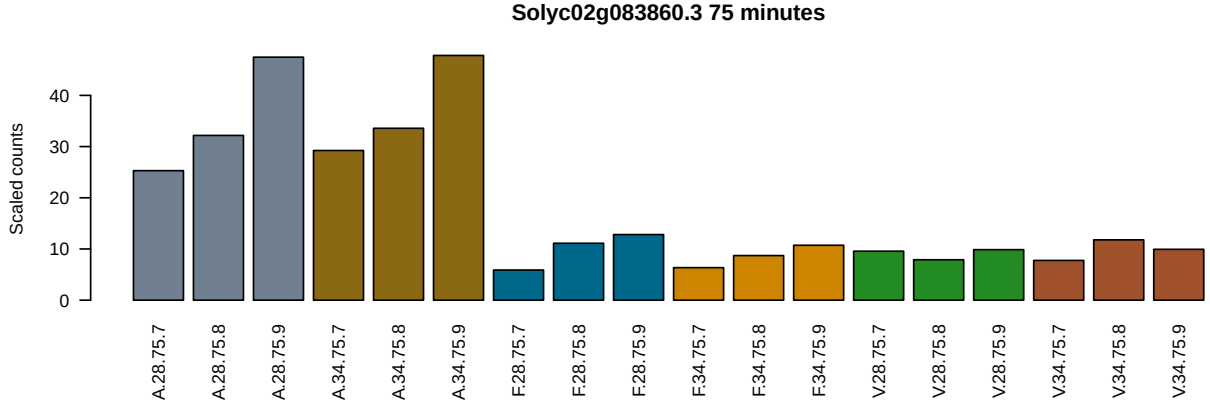
```
oldpar=par()
par(mfrow=c(4,1))
scaled=sl4_scaled
g = sl4_g
v = colnames(scaled)
indexes = c(grep("A.28.15",v),
            grep("A.34.15",v),
            grep("F.28.15",v),
            grep("F.34.15",v),
            grep("V.28.15",v),
            grep("V.34.15",v))
all = indexes
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"15 minutes"))
```

```
indexes = c(grep("A.28.30",v),
            grep("A.34.30",v),
            grep("F.28.30",v),
            grep("F.34.30",v),
            grep("V.28.30",v),
            grep("V.34.30",v))
all = c(all,indexes)
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"30 minutes"))
indexes = c(grep("A.28.45",v),
            grep("A.34.45",v),
            grep("F.28.45",v),
            grep("F.34.45",v),
            grep("V.28.45",v),
            grep("V.34.45",v))
all = c(all,indexes)
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"45 minutes"))
indexes = c(grep("A.28.75",v),
            grep("A.34.75",v),
            grep("F.28.75",v),
            grep("F.34.75",v),
            grep("V.28.75",v),
            grep("V.34.75",v))
all = c(all,indexes)
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"75 minutes"))
```


Solyc02g083860.3 15 minutes
Scaled counts
0
5
10
15
20
25
30
A.28.15.7
A.28.15.8
A.28.15.9
A.34.15.7
A.34.15.8
A.34.15.9
F.28.15.7
F.28.15.8
F.28.15.9
F.34.15.7
F.34.15.8
F.34.15.9
V.28.15.7
V.28.15.8
V.28.15.9
V.34.15.7
V.34.15.8
V.34.15.9


Solyc02g083860.3 30 minutes
Scaled counts
0
5
10
15
20
25
30
35
A.28.30.7
A.28.30.8
A.28.30.9
A.34.30.7
A.34.30.8
A.34.30.9
F.28.30.7
F.28.30.8
F.28.30.9
F.34.30.7
F.34.30.8
F.34.30.9
V.28.30.7
V.28.30.8
V.28.30.9
V.34.30.7
V.34.30.8
V.34.30.9


Solyc02g083860.3 45 minutes
Scaled counts
0
10
20
30
40
A.28.45.7
A.28.45.8
A.28.45.9
A.34.45.7
A.34.45.8
A.34.45.9
F.28.45.7
F.28.45.8
F.28.45.9
F.34.45.7
F.34.45.8
F.34.45.9
V.28.45.7
V.28.45.8
V.28.45.9
V.34.45.7
V.34.45.8
V.34.45.9


Solyc02g083860.3 75 minutes
Scaled counts
0
10
20
30
40
A.28.75.7
A.28.75.8
A.28.75.9
A.34.75.7
A.34.75.8
A.34.75.9
F.28.75.7
F.28.75.8
F.28.75.9
F.34.75.7
F.34.75.8
F.34.75.9
V.28.75.7
V.28.75.8
V.28.75.9
V.34.75.7
V.34.75.8
V.34.75.9

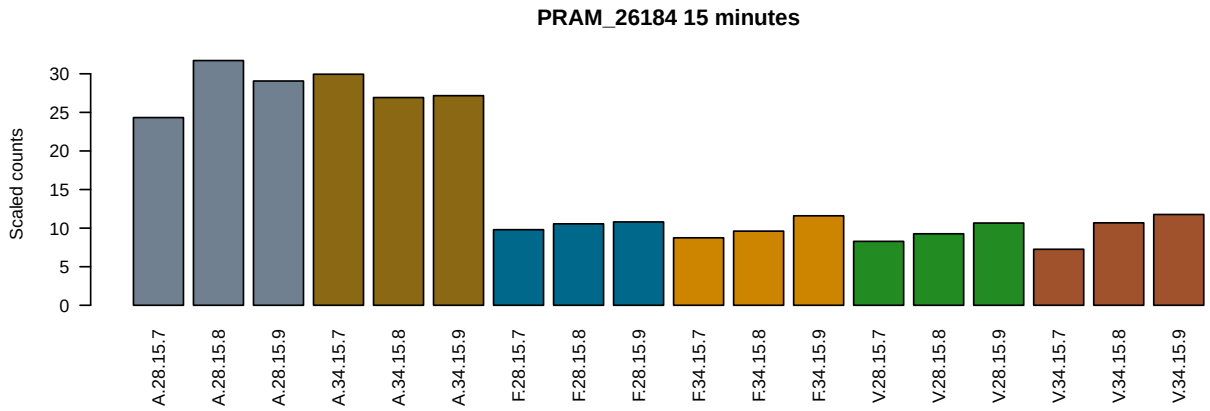
```
par(oldpar)
```

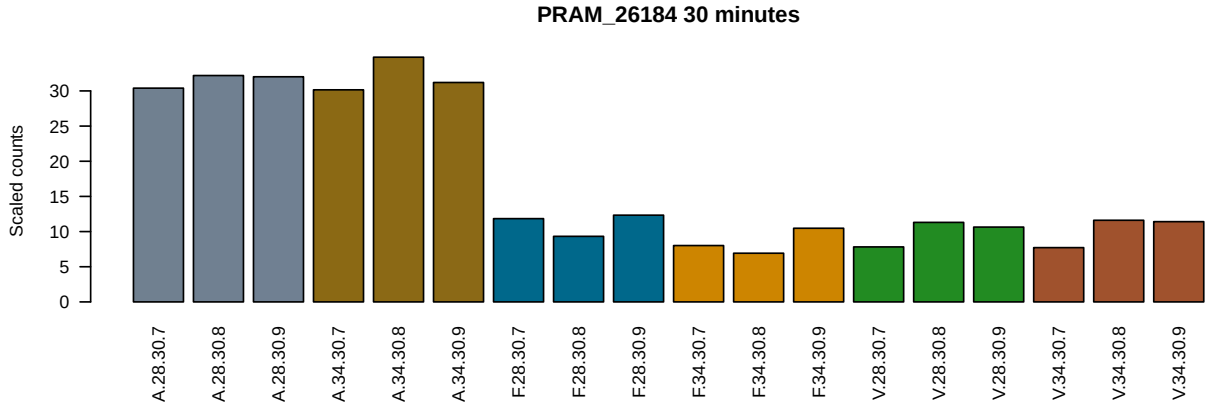
The preceding plots show the following aspects of SL4 `Solyc02g083860.3` gene expression:

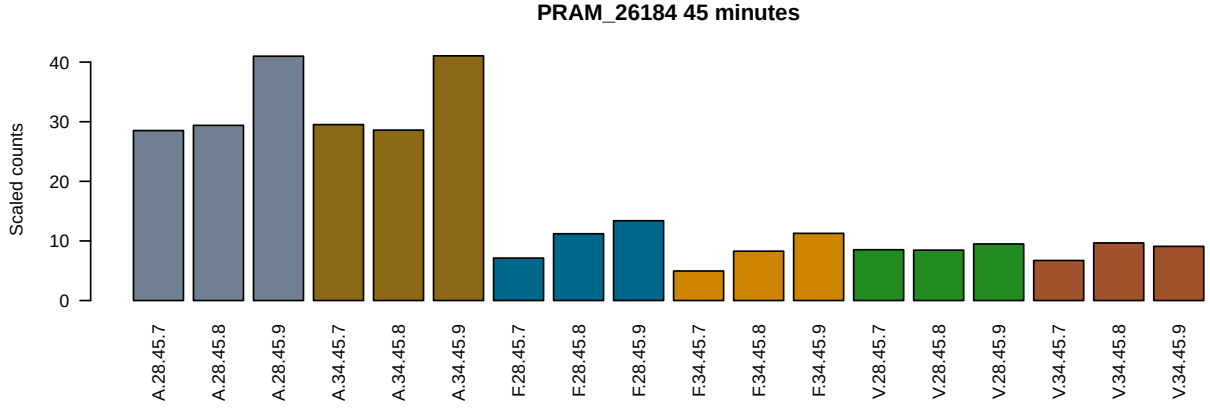
- At each treatment duration time point, expression was higher in the `are` mutant than in the wildtype and over-expression genotypes, as was observed previously.
- At each time point, expression of the over-expression genotype was about the same as in the `VF36` genotype.
- Heat stress did not appear to affect expression in any of the genotypes, at any of the treatment duration time points.

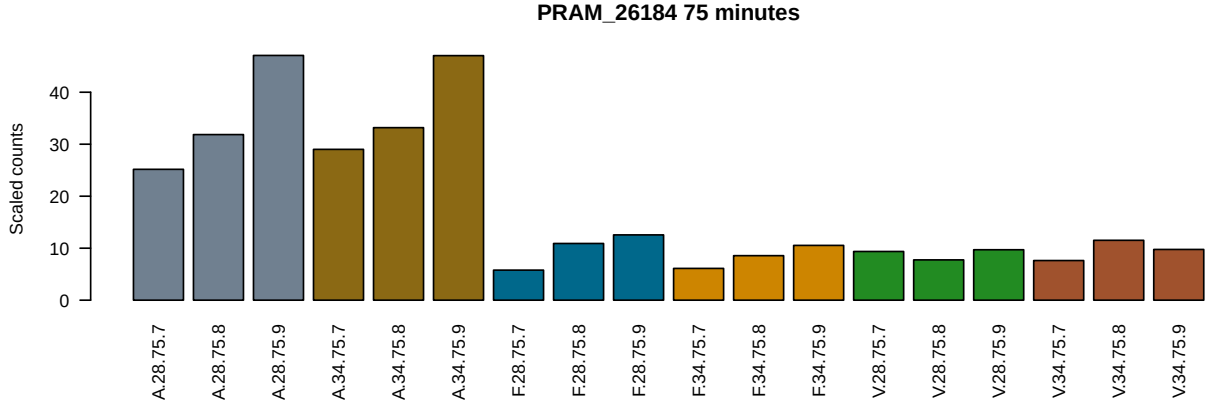
Create plot that shows scaled gene expression for SL5 gene `PRAM_26184`:

```
oldpar=par()
par(mfrow=c(4,1))
scaled=sl5_scaled
g = sl5_g
v = colnames(scaled)
indexes = c(grep("A.28.15",v),
            grep("A.34.15",v),
            grep("F.28.15",v),
            grep("F.34.15",v),
            grep("V.28.15",v),
            grep("V.34.15",v))
all = indexes
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"15 minutes"))
indexes = c(grep("A.28.30",v),
            grep("A.34.30",v),
            grep("F.28.30",v),
            grep("F.34.30",v),
            grep("V.28.30",v),
            grep("V.34.30",v))
all = c(all,indexes)
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"30 minutes"))
indexes = c(grep("A.28.45",v),
            grep("A.34.45",v),
            grep("F.28.45",v),
            grep("F.34.45",v),
            grep("V.28.45",v),
            grep("V.34.45",v))
all = c(all,indexes)
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"45 minutes"))
indexes = c(grep("A.28.75",v),
            grep("A.34.75",v),
            grep("F.28.75",v),
            grep("F.34.75",v),
            grep("V.28.75",v),
            grep("V.34.75",v))
all = c(all,indexes)
makeBarPlot(g,scaled[,indexes],"Scaled counts",
            beside=T,main=paste(g,"75 minutes"))
```


PRAM_26184 15 minutes
Scaled counts
0
5
10
15
20
25
30
A.28.15.7
A.28.15.8
A.28.15.9
A.34.15.7
A.34.15.8
A.34.15.9
F.28.15.7
F.28.15.8
F.28.15.9
F.34.15.7
F.34.15.8
F.34.15.9
V.28.15.7
V.28.15.8
V.28.15.9
V.34.15.7
V.34.15.8
V.34.15.9


PRAM_26184 30 minutes
Scaled counts
0
5
10
15
20
25
30
A.28.30.7
A.28.30.8
A.28.30.9
A.34.30.7
A.34.30.8
A.34.30.9
F.28.30.7
F.28.30.8
F.28.30.9
F.34.30.7
F.34.30.8
F.34.30.9
V.28.30.7
V.28.30.8
V.28.30.9
V.34.30.7
V.34.30.8
V.34.30.9


PRAM_26184 45 minutes
Scaled counts
0
10
20
30
40
A.28.45.7
A.28.45.8
A.28.45.9
A.34.45.7
A.34.45.8
A.34.45.9
F.28.45.7
F.28.45.8
F.28.45.9
F.34.45.7
F.34.45.8
F.34.45.9
V.28.45.7
V.28.45.8
V.28.45.9
V.34.45.7
V.34.45.8
V.34.45.9


PRAM_26184 75 minutes
Scaled counts
0
10
20
30
40
A.28.75.7
A.28.75.8
A.28.75.9
A.34.75.7
A.34.75.8
A.34.75.9
F.28.75.7
F.28.75.8
F.28.75.9
F.34.75.7
F.34.75.8
F.34.75.9
V.28.75.7
V.28.75.8
V.28.75.9
V.34.75.7
V.34.75.8
V.34.75.9

```
par(oldpar)
```

The preceding plots show that `PRAM_26184`, the SL5 counterpart of SL4 gene `Solyc02g083860.3`, exhibited the same expression profile.

Write scaled data files:

```
to_write = data.frame(gene_name=row.names(sl4_scaled))
to_write=cbind(to_write,sl4_scaled[,all])
to_write$description = sl4_scaled$description
for (i in 2:(ncol(to_write)-1)) {
  to_write[,i]=signif(to_write[,i],3)
}
write.table(to_write,file=sl4_output_fname,sep="\t",quote = F,
            row.names = F,col.names = T)
to_write = data.frame(gene_name=row.names(sl5_scaled))
to_write=cbind(to_write,sl5_scaled[,all])
to_write$description = sl5_scaled$description
for (i in 2:(ncol(to_write)-1)) {
  to_write[,i]=signif(to_write[,i],3)
}
write.table(to_write,file=sl5_output_fname,sep="\t",quote = F,
            row.names = F,col.names = T)
```

The preceding code chunk wrote files named `results/muday-144-SL4_counts-salmon_scaled.txt` and that `results/muday-144-SL5_counts-salmon_scaled.txt` contains scaled gene expression values per gene for SL4 and SL5 assemblies, expressed in counts per million.

Discussion

The expression profile for Solyc02g083860.3/PRAM_26184 resembled the previously observed patterns in the following respects.

As before, expression of Solyc02g083860.3/PRAM_26184 in the over-expression genotype relative to progenitor genotype VF36, a wildtype line, was not higher. Why not?

The *are* (anthocyanin reduced) genotype showed the highest expression of all three genotypes, as observed previously in the 120 minutes heat stress experiment. This suggests that transcriptional and/or post-transcriptional regulation of the flavonoid biosynthetic gene Solyc02g083860.3/PRAM_26184 is somehow able to respond to the levels of flavonoids produced, due to some as-yet unknown compensatory, regulatory mechanism.

Conclusion

- Expression of Solyc02g083860.3/PRAM_26184 matched what was previously observed in `../ARE-120min-analysis/R3H.Rmd` and suggested that feedback mechanisms exist that control expression of the flavonoid biosynthetic pathway genes via products of the pathway itself.
- Scaled expression files for SL4 and SL5 were written.