

Arabidopsis thaliana

Alice Gee

12/11/2021

```
library(BiocManager)
source("rnaseq_utils.R")
library(DESeq2)
library(BiocParallel)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(tidyr)
library(sys)
library(knitr)
library(pheatmap)
```

```
nrow(rnaCounts)
```

```
## [1] 19110
```

```
nrow(riboCounts)
```

```
## [1] 19110
```

rna-seq

```
start_time <- Sys.time()

DESeqDataSet = DESeqDataSetFromMatrix(
  countData = rnaCounts,
  colData = sampleAnnotation,
  design = ~ time + genotype + time:genotype
)
```

```
## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

```
DESeqDataSet = DESeq(
  DESeqDataSet,
  parallel=FALSE,
  test = "LRT",
  reduced = ~ time + genotype
)
```

```
## estimating size factors
```

```
## estimating dispersions
```

```

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

DESeq_Results <- results(DESeqDataSet)
# indexes that all have a value
clean_DESeq_padj <- which(!is.na(DESeq_Results$padj))
sum(DESeq_Results[clean_DESeq_padj, "padj"] <= 0.1)

## [1] 1060

sum(DESeq_Results[clean_DESeq_padj, "padj"] <= 0.1) * 0.1

## [1] 106

end_time <- Sys.time()
print(end_time - start_time)

## Time difference of 25.33242 secs

```

cut out -> not needed

```

temp = DESeq_Results[clean_DESeq_padj,]
temp = as.data.frame(temp)
rna_sig_genes = c()
rna_sig_values = c()
for (i in 1:nrow(temp)){
  if (temp$padj[i] <= 0.1){
    rna_sig_genes = c(rna_sig_genes, rownames(temp)[i])
    rna_sig_values = c(rna_sig_values, temp$baseMean[i])
  }
}

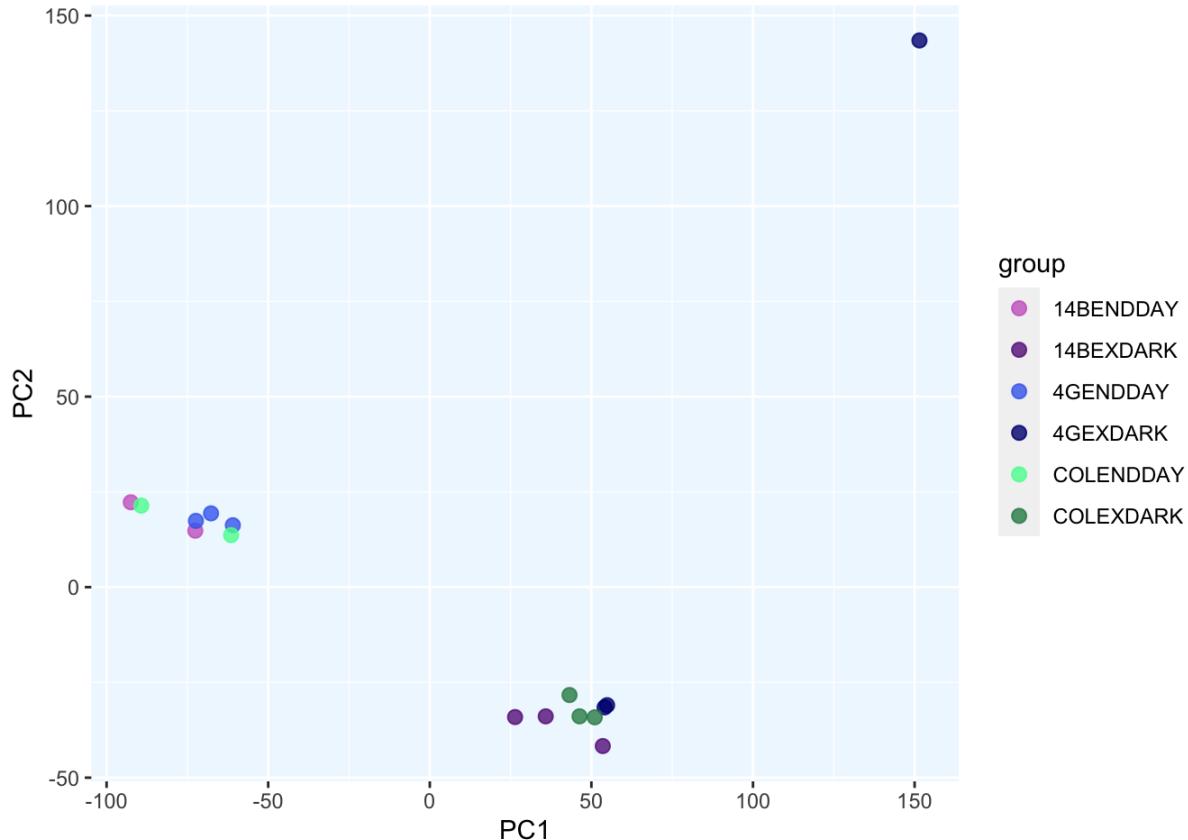
```

log transformed the data

```
lgNorm = log2(counts(DESeqDataSet, normalized=TRUE) + 1)
```

Overall PCA Plot

```
pca = prcomp(t(lgNorm))
pcaData = data.frame(pca$x[ , 1:2])
pcaData$group = sampleAnnotation[rownames(pcaData), "group"]
pcaData$sample = rownames(pcaData)
gg = ggplot(pcaData, aes(x=PC1, y=PC2, color=group, label=sample))
gg = gg + geom_point(size=2.5, alpha=0.8)
gg = gg + scale_color_manual(values=groupColors)
gg = gg + theme(panel.background = element_rect(fill = 'aliceblue'))
print(gg)
```



```
# remove the outlier
rnaCounts <- rnaCounts %>% select(-`4GEXDARK4`)
sampleAnnotation <- sampleAnnotation[colnames(rnaCounts),]
```

```
start_time <- Sys.time()

DESeqDataSet = DESeqDataSetFromMatrix(
  countData = rnaCounts,
  colData = sampleAnnotation,
  design = ~ time + genotype + time:genotype
)
```

```
## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

```

DESeqDataSet = DESeq(
  DESeqDataSet,
  parallel=FALSE,
  test = "LRT",
  reduced = ~ time + genotype
)

## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

DESeq_Results <- results(DESeqDataSet)
clean_DESeq_padj <- which(!is.na(DESeq_Results$padj))
RNA_Sig <- sum(DESeq_Results[clean_DESeq_padj, "padj"] <= 0.1)
RNA_Sig

## [1] 1225

sum(DESeq_Results[clean_DESeq_padj, "padj"] <= 0.1) * 0.1

## [1] 122.5

end_time <- Sys.time()
print(end_time - start_time)

## Time difference of 17.38915 secs

temp <- as.data.frame(DESeq_Results[clean_DESeq_padj,])
head(temp %>% filter(padj <= 0.1) %>% arrange(padj))

##          baseMean log2FoldChange      lfcSE      stat     pvalue      padj
## AT5G36910    378.8466     -2.6201284  0.4081485  78.54299 8.802504e-18 1.328210e-13
## AT4G15110    407.7008     -0.3427123  0.2261603  76.57551 2.354178e-17 1.776110e-13
## AT3G18080    1444.6423     -1.7381051  0.2019068  73.17519 1.288870e-16 4.633429e-13
## AT3G24190    789.7563      1.1735899  0.1468745  73.59185 1.046484e-16 4.633429e-13
## AT4G17090    3058.4734      2.7247848  0.3413467  72.82518 1.535366e-16 4.633429e-13
## AT3G48560    5818.1473      1.2314859  0.1711858  64.48119 9.956070e-15 2.503786e-11

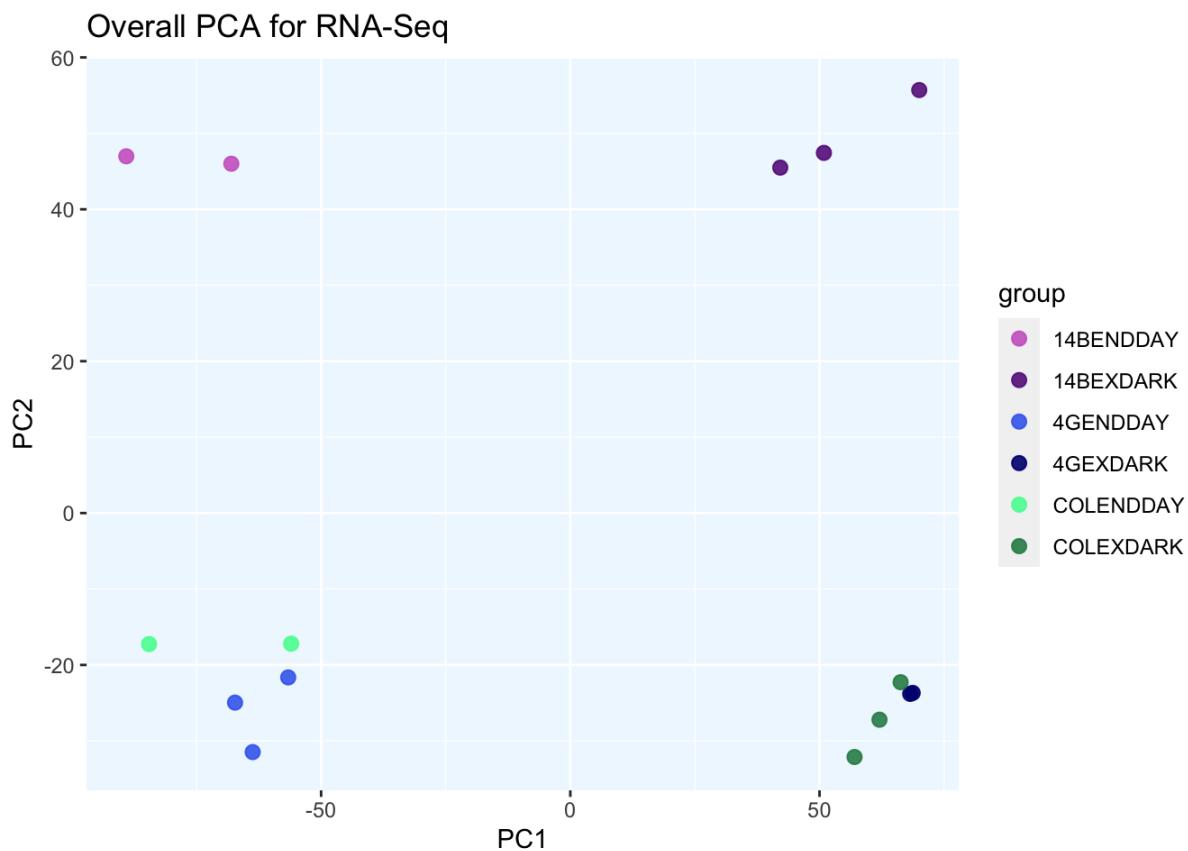
```

log transformed the data

```
lgNorm = log2(counts(DESeqDataSet, normalized=TRUE) + 1)
```

Overall PCA Plot

```
library(ggplot2)
pca = prcomp(t(lgNorm))
pcaData = data.frame(pca$x[, 1:2])
pcaData$group = sampleAnnotation[rownames(pcaData), "group"]
pcaData$sample = rownames(pcaData)
gg = ggplot(pcaData, aes(x=PC1, y=PC2, color=group, label=sample))
gg = gg + geom_point(size=2.5, alpha = 0.9)
gg = gg + scale_color_manual(values=groupColors)
gg = gg + theme(panel.background = element_rect(fill = 'aliceblue')) + ggtitle("Overall PCA for RNA-Seq")
print(gg)
```



```
joined_set = inner_join(goAssociations, geneNamesAndDescriptions, by = "gene")
ontology_names <- joined_set %>% distinct(gene_ontology_name)
ontology_names <- ontology_names[["gene_ontology_name"]]
ontology_names_all <- goAssociations2 %>% distinct(gene_ontology_name)
ontology_names_all <- ontology_names_all[["gene_ontology_name"]]
```

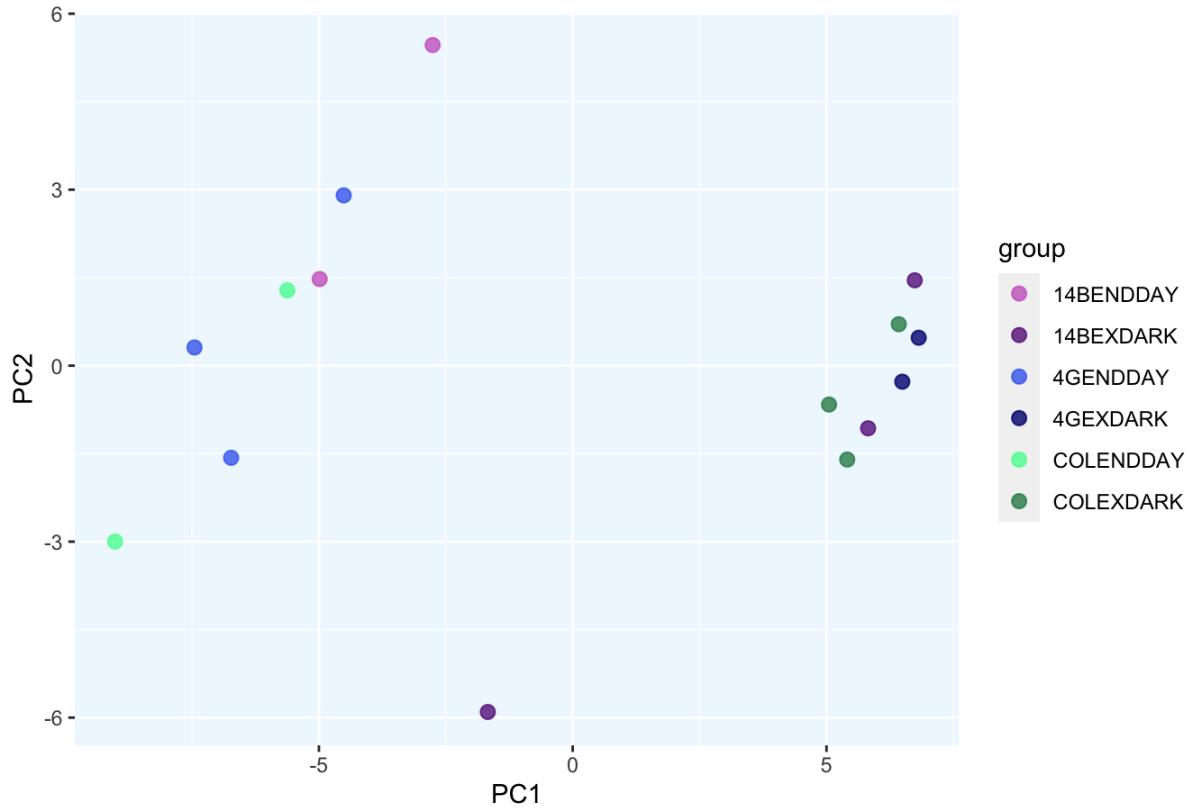
PCA based on Gene Groupings

```
for (i in 1:length(ontology_names)){
  temp_set <- joined_set %>% filter(gene_ontology_name == ontology_names[i])
  lgGo <- lgNorm[temp_set$gene, ]

  pca1 <- prcomp(t(lgGo))
  pcaData1 = data.frame(pca1$x[ , 1:2])
  pcaData1$group = sampleAnnotation[rownames(pcaData1), "group"]
  pcaData1$sample = rownames(pcaData1)

  gg1 = ggplot(pcaData1, aes(x=PC1, y=PC2, color=group, label=sample))
  gg1 = gg1 + geom_point(size=2.5, alpha=0.8) + ggtitle(ontology_names[i])
  gg1 = gg1 + scale_color_manual(values=groupColors)
  gg1 = gg1 + theme(panel.background = element_rect(fill = 'aliceblue'))
  print(gg1)
}
```

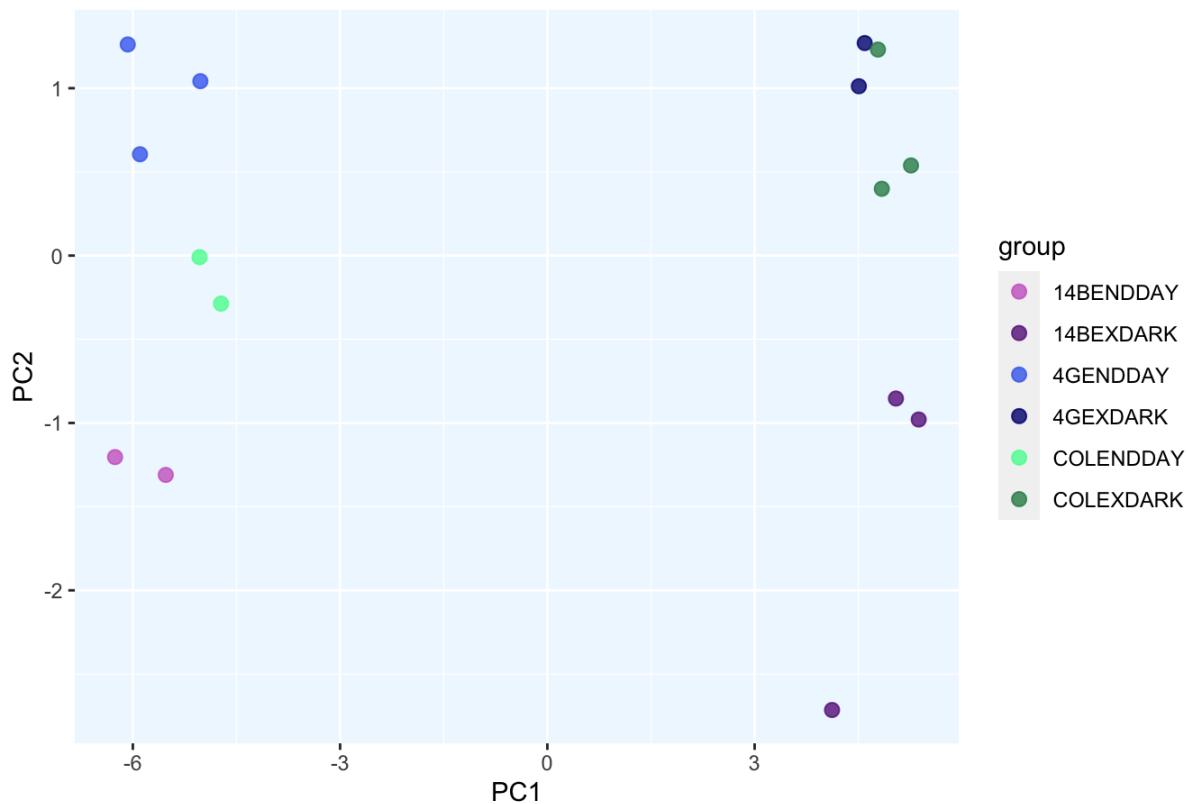
cold acclimation



group

- 14BENDDAY
- 14BEXDARK
- 4GENDDAY
- 4GEXDARK
- COLENDAY
- COLEXDARK

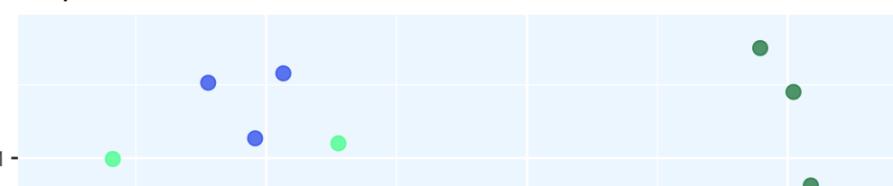
protein-chromophore linkage

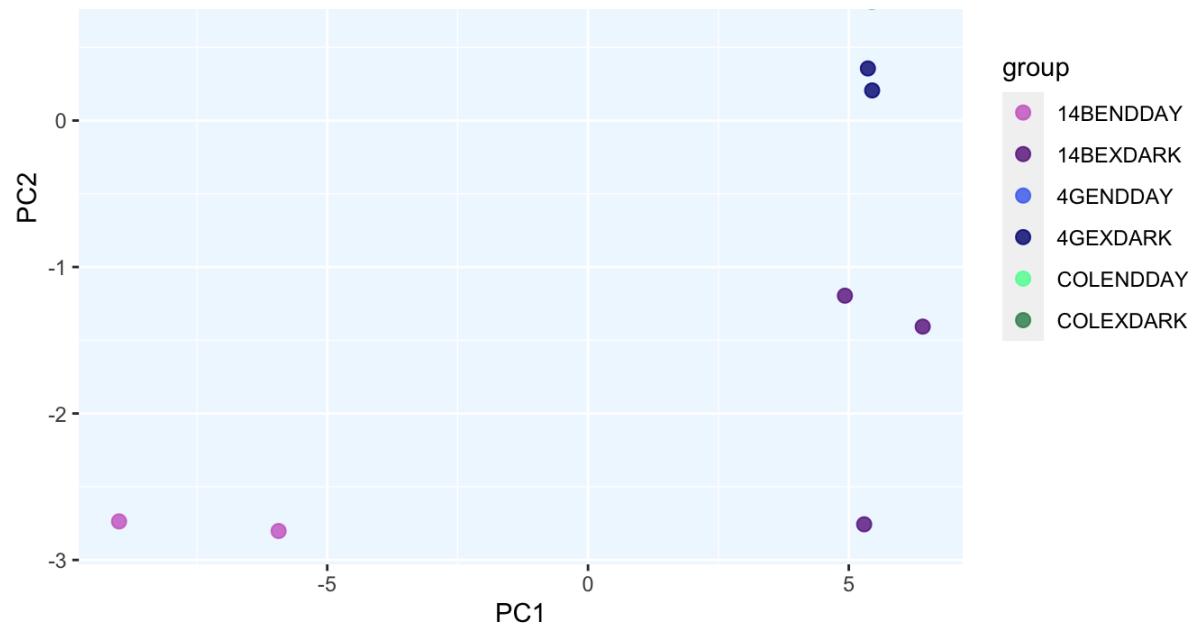


group

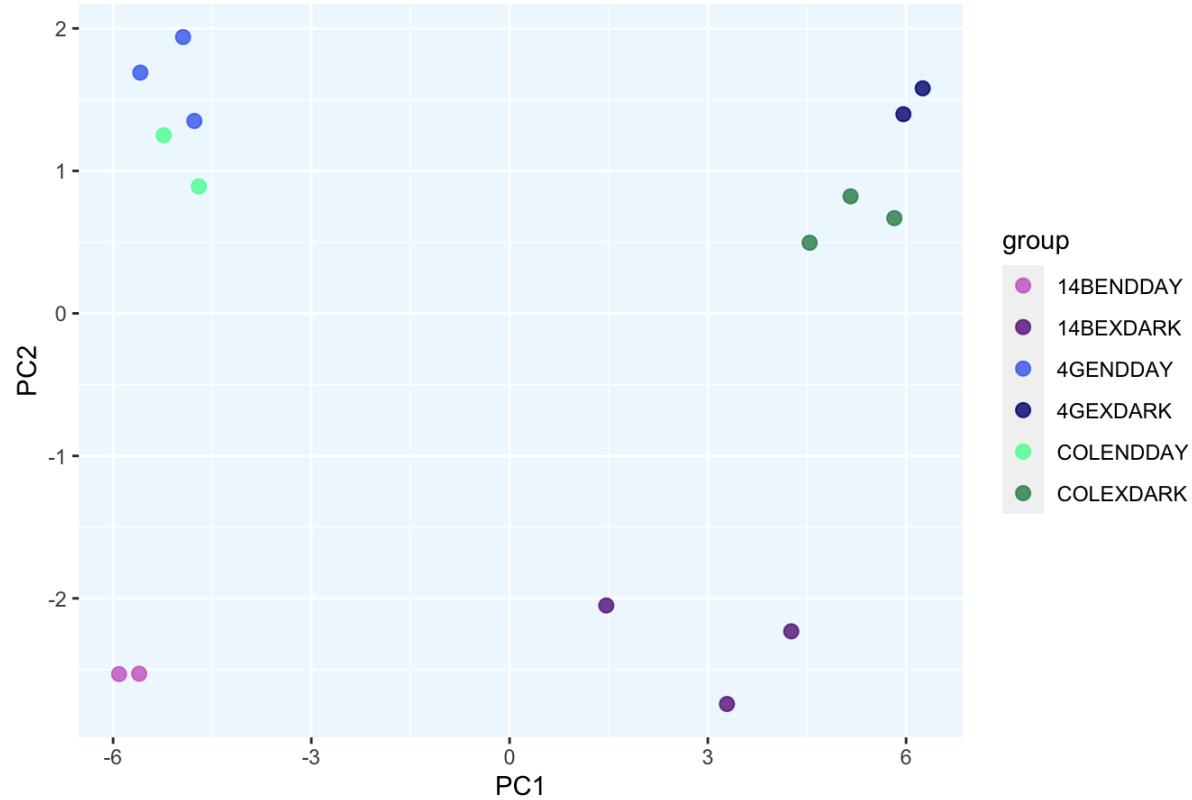
- 14BENDDAY
- 14BEXDARK
- 4GENDDAY
- 4GEXDARK
- COLENDAY
- COLEXDARK

response to disaccharide

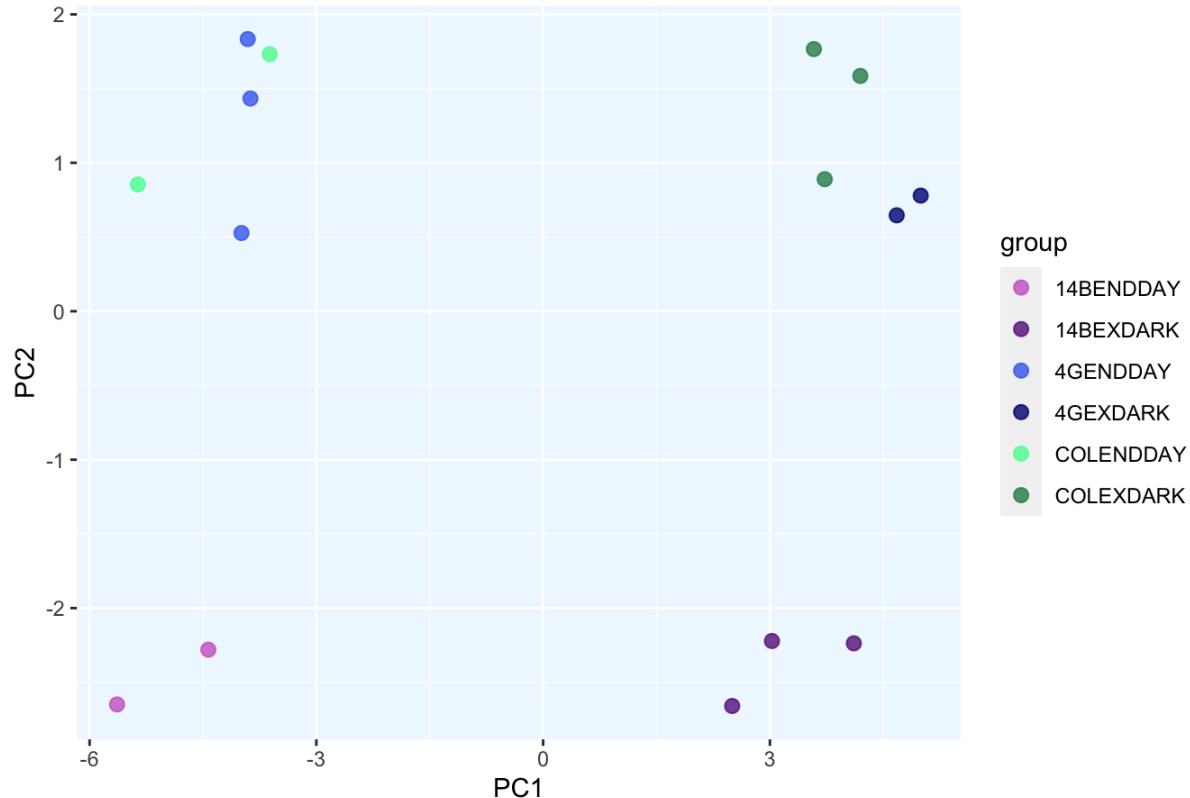




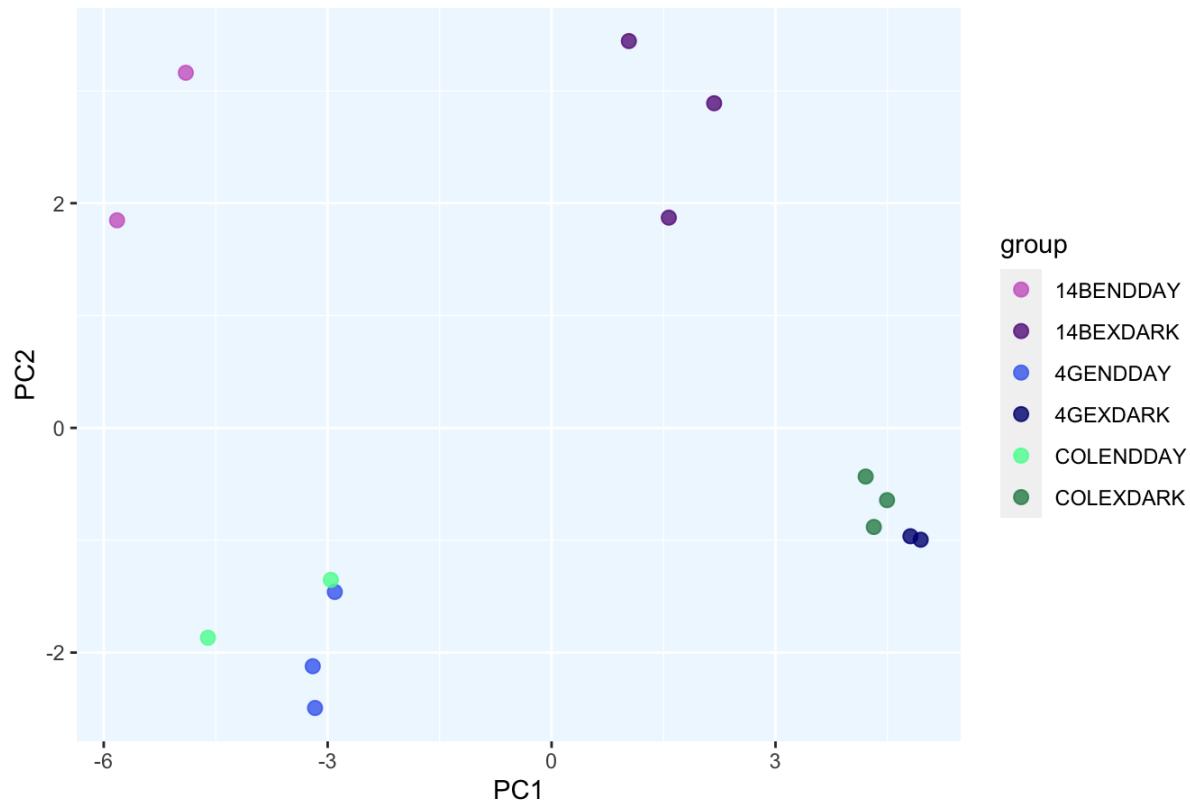
tetrapyrrole metabolic process



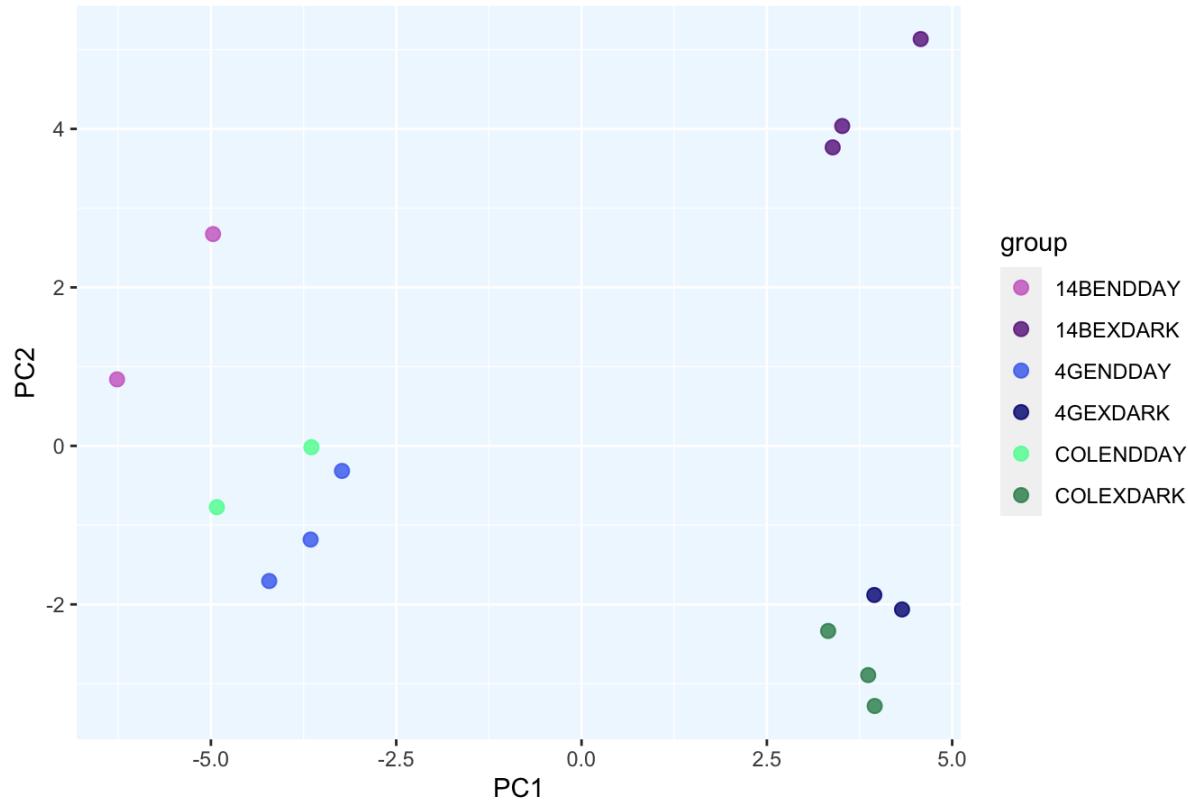
cellular aldehyde metabolic process



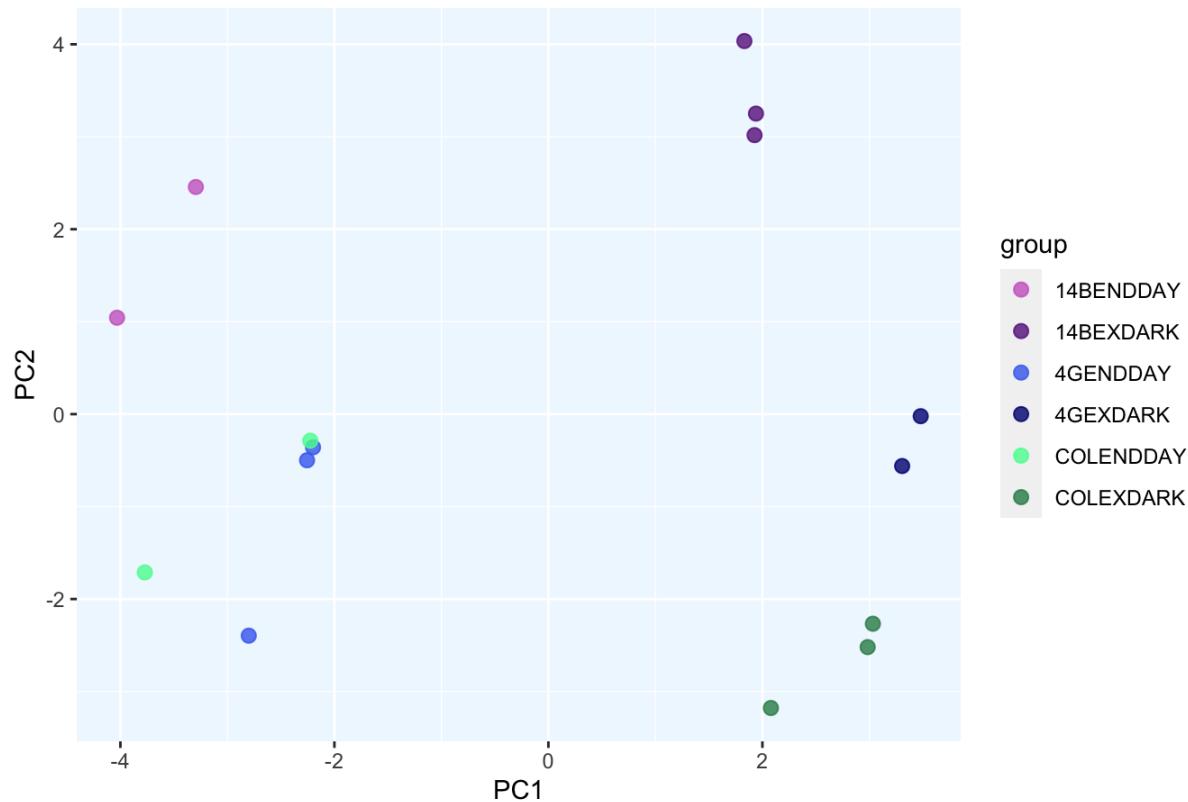
pyruvate metabolic process



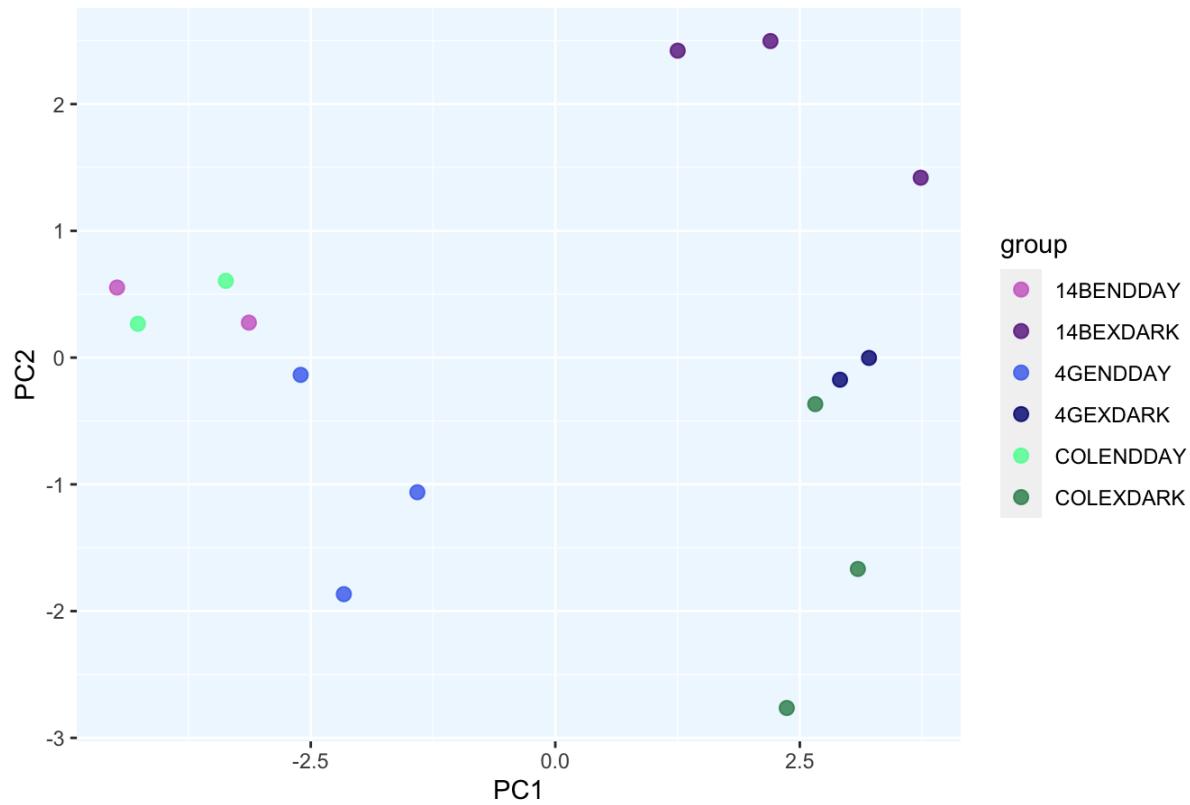
amino acid transport



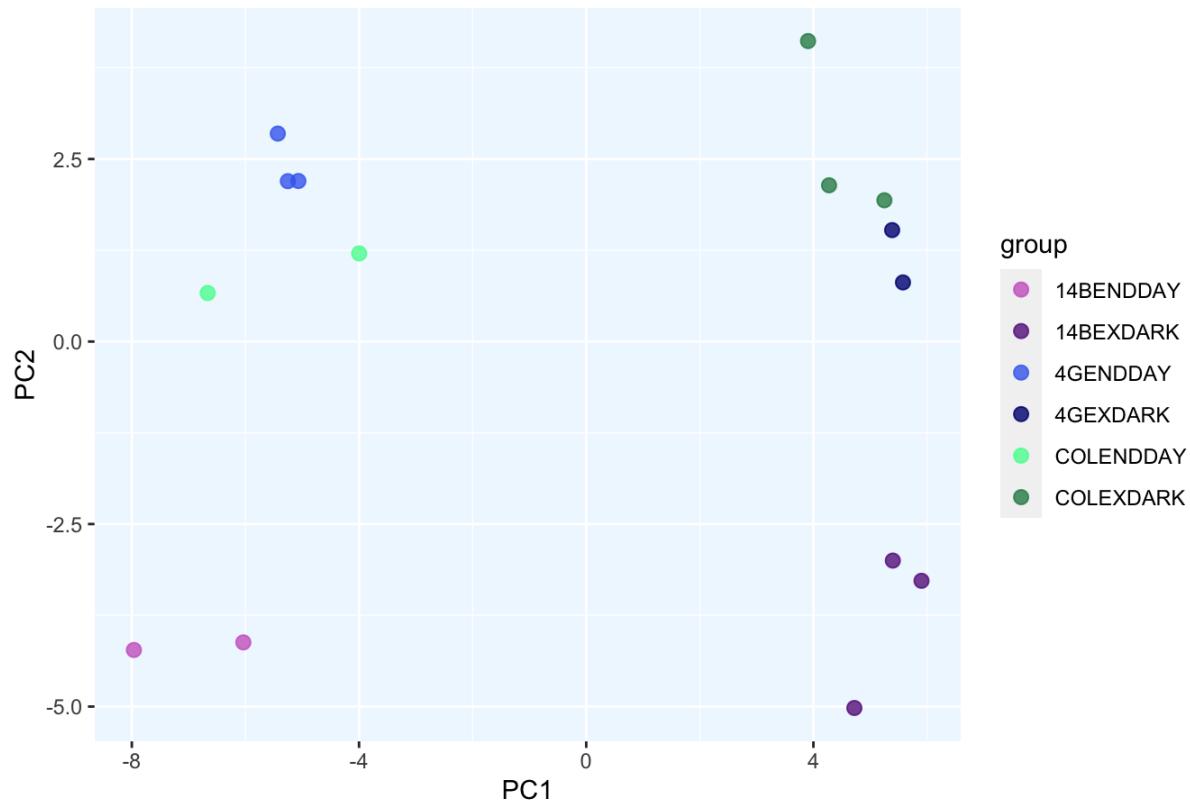
anion homeostasis



response to ozone



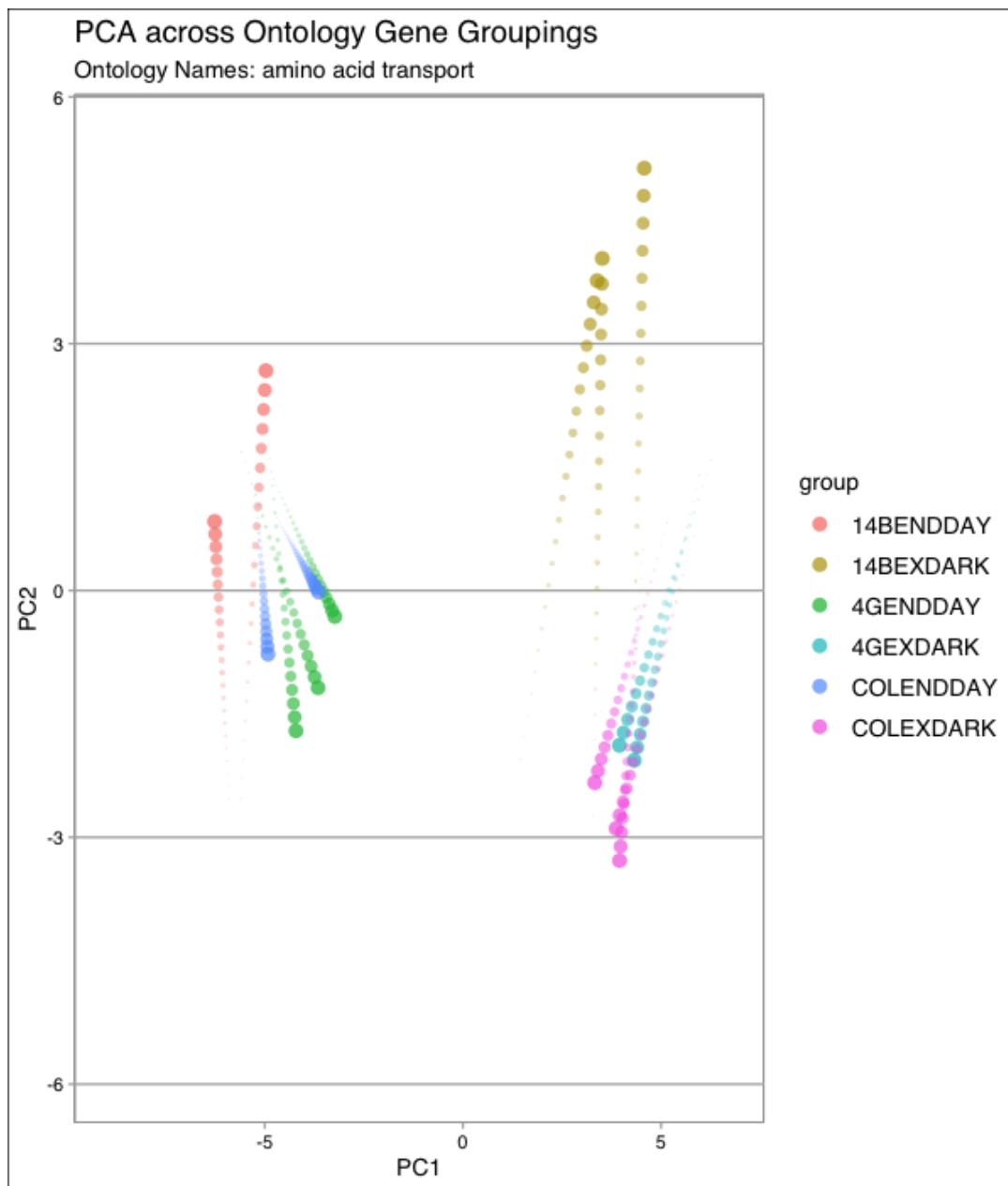
cellular response to endogenous stimulus



```

all_pca = data.frame()
for (i in 1:length(ontology_names)){
  temp_set <- joined_set %>% filter(gene_ontology_name == ontology_names[i])
  lgGo <- lgNorm[temp_set$gene, ]
  pca1 <- prcomp(t(lgGo))
  pcaData1 = data.frame(pca1$x[, 1:2])
  pcaData1$group = sampleAnnotation[rownames(pcaData1), "group"]
  pcaData1$sample = rownames(pcaData1)
  pcaData1$ontology_name = ontology_names[i]
  all_pca = rbind(all_pca, pcaData1)
}
write.table(all_pca,
            "all_pca.csv",
            sep = ",",
            row.names = FALSE,
            quote = FALSE)

```

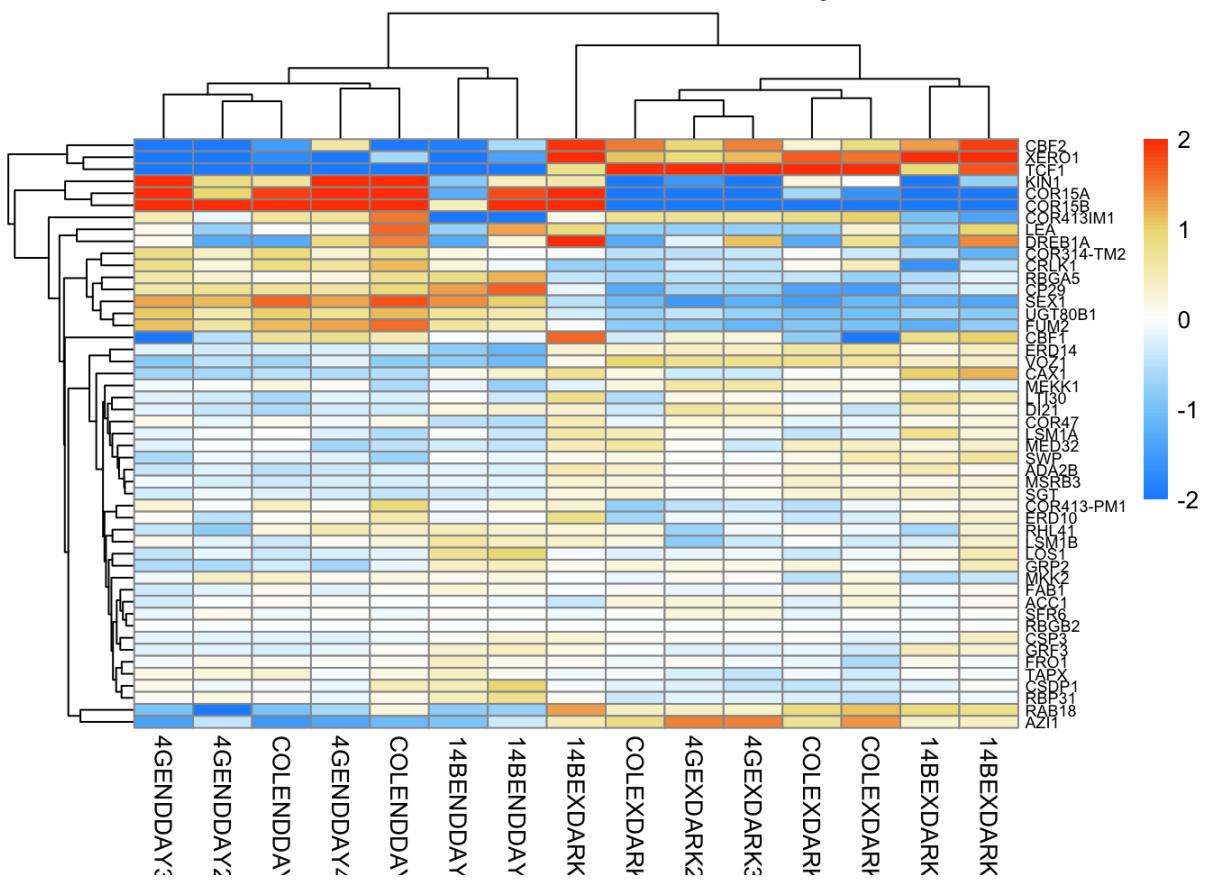


Heatmap per Gene Grouping

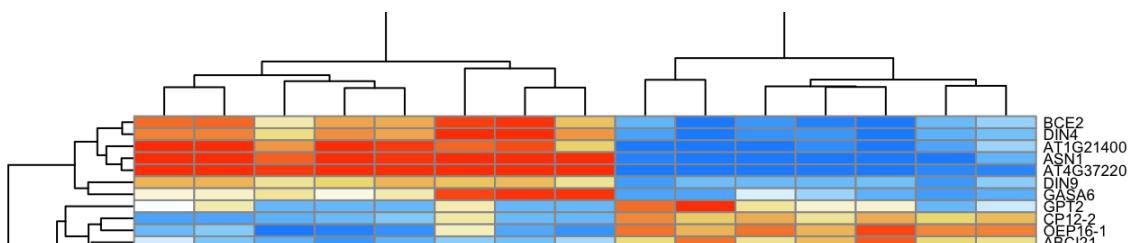
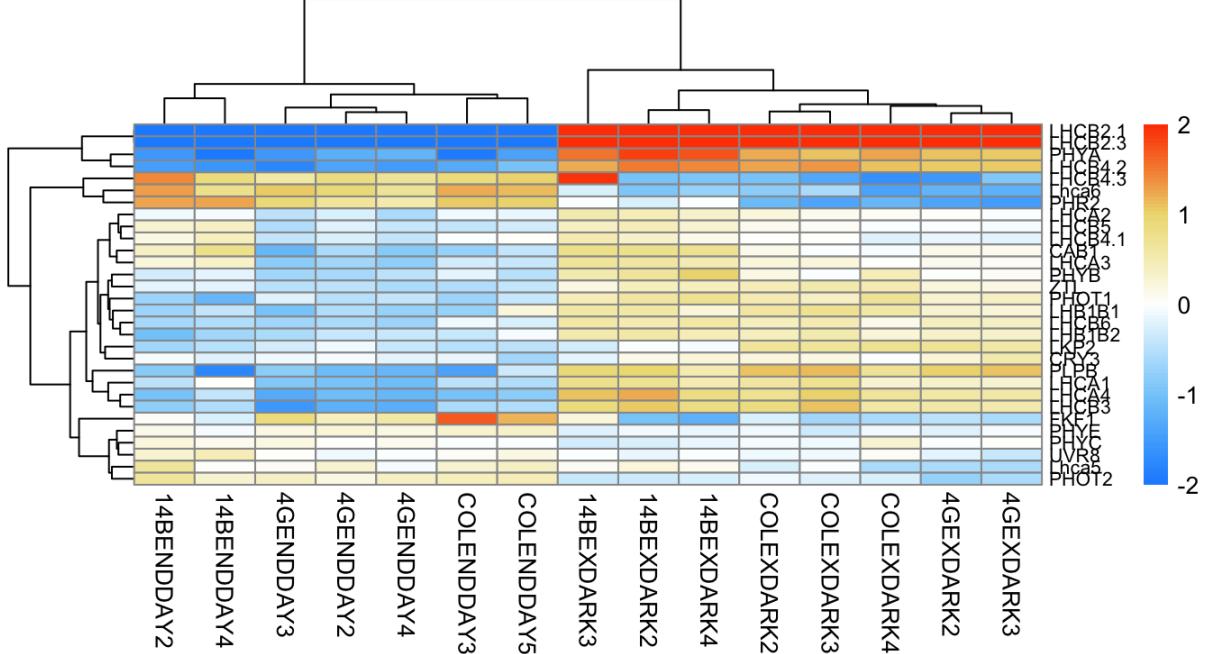
```
for (i in 1:length(ontology_names)){
  temp_set <- joined_set %>% filter(gene_ontology_name == ontology_names[i])
  lgGo <- lgNorm[temp_set$gene, ]

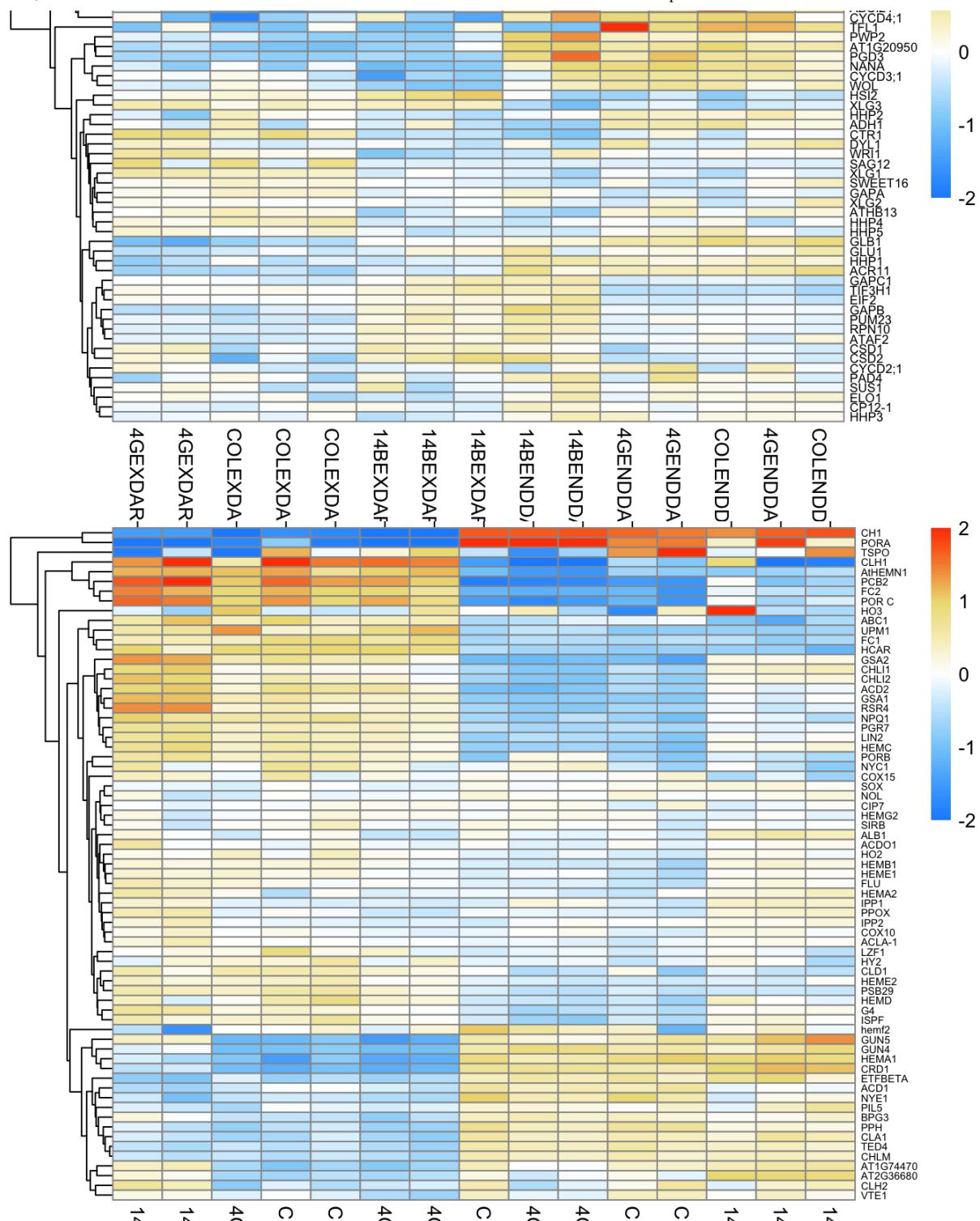
  heatData = lgGo - rowMeans(lgGo)
  heatData = as.data.frame(heatData)
  heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
  fontsize_row = 10 - nrow(heatData) / 15
  pheatmap(
    heatData,
    color = heatPalette,
    clustering_method = "average",
    labels_row=geneNamesAndDescriptions[rownames(heatData), "symbol"],
    main = ontology_names[i],
    fontsize_row = fontsize_row,
    cellheight = 5,
    margins = c(10, 10)
  )
}
```

Arabidopsis thaliana

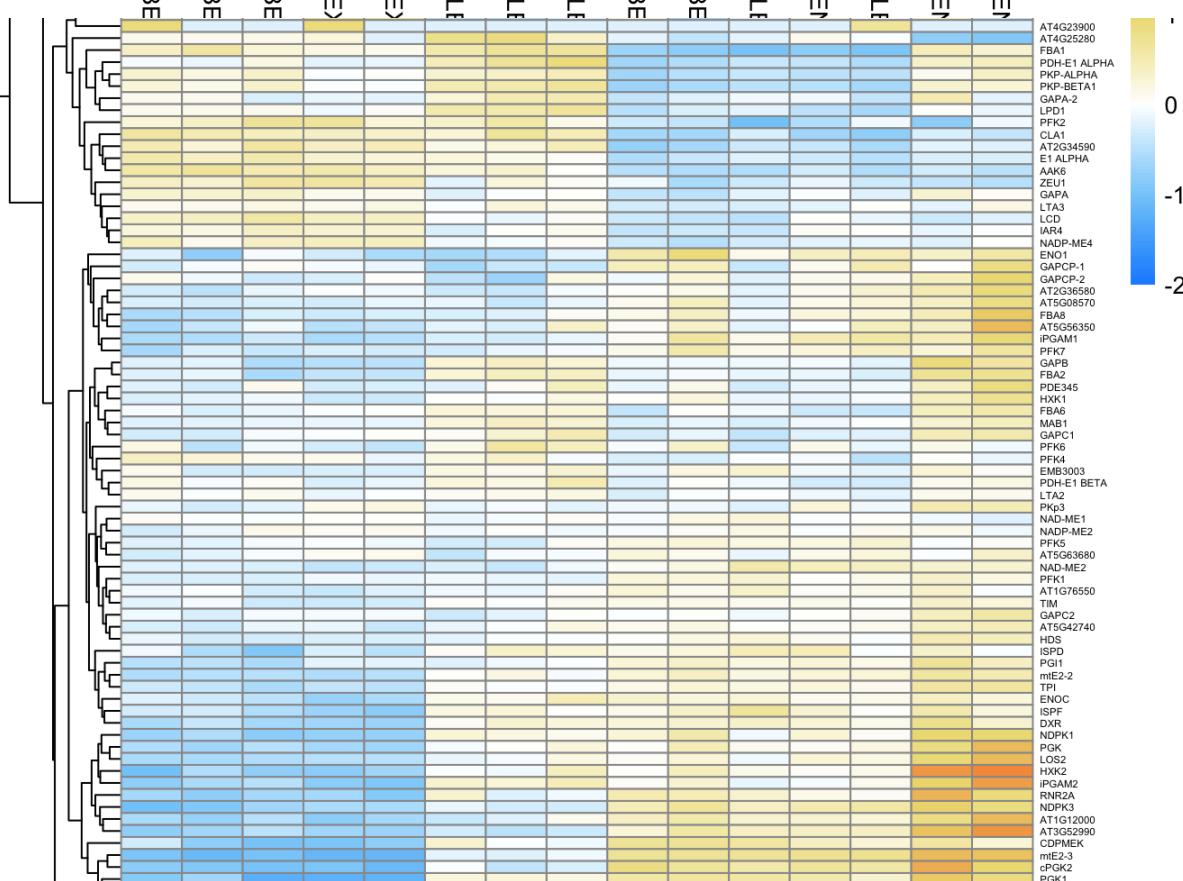
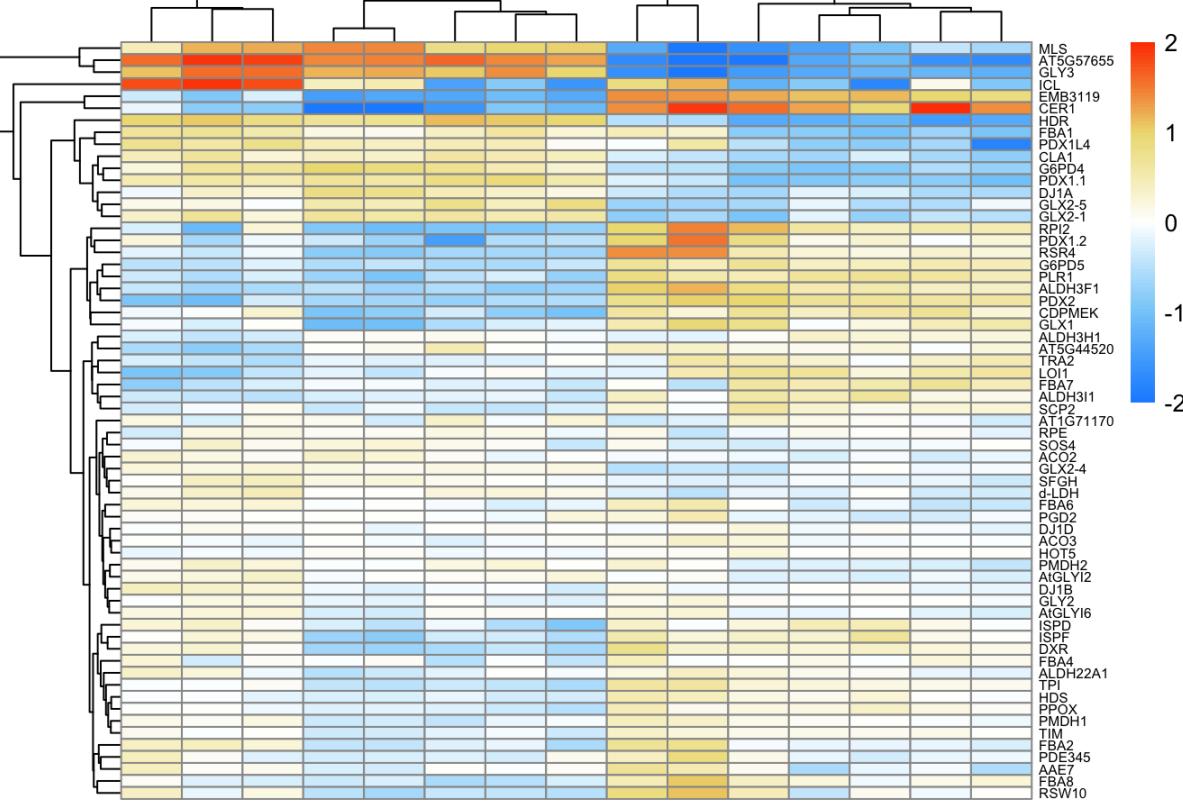


protein-chromophore linkage

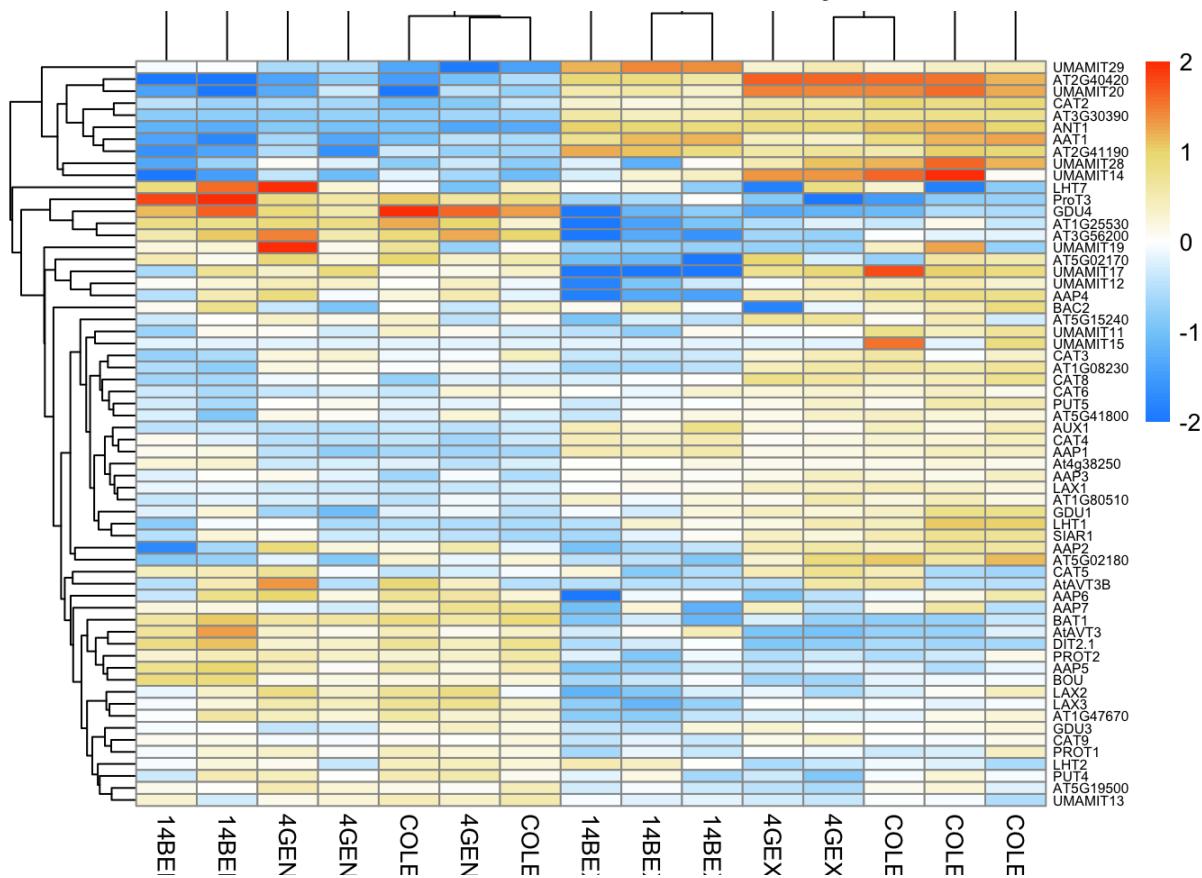




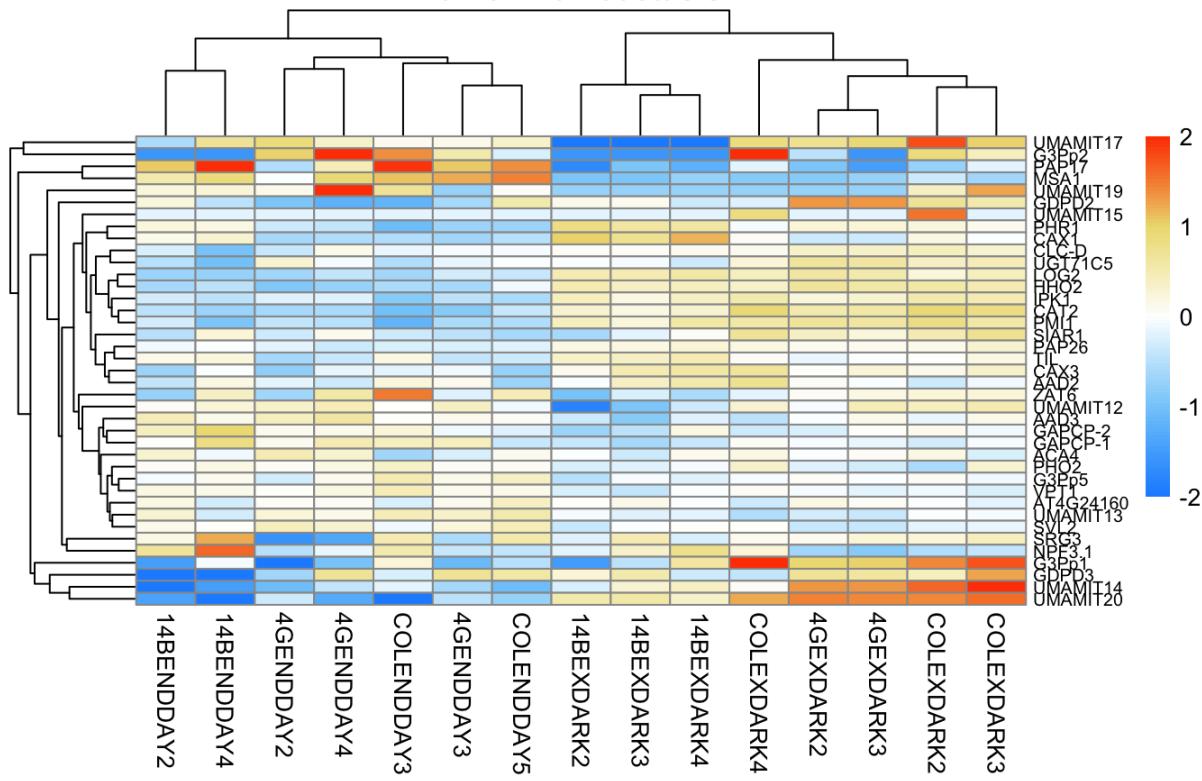
Arabidopsis thaliana

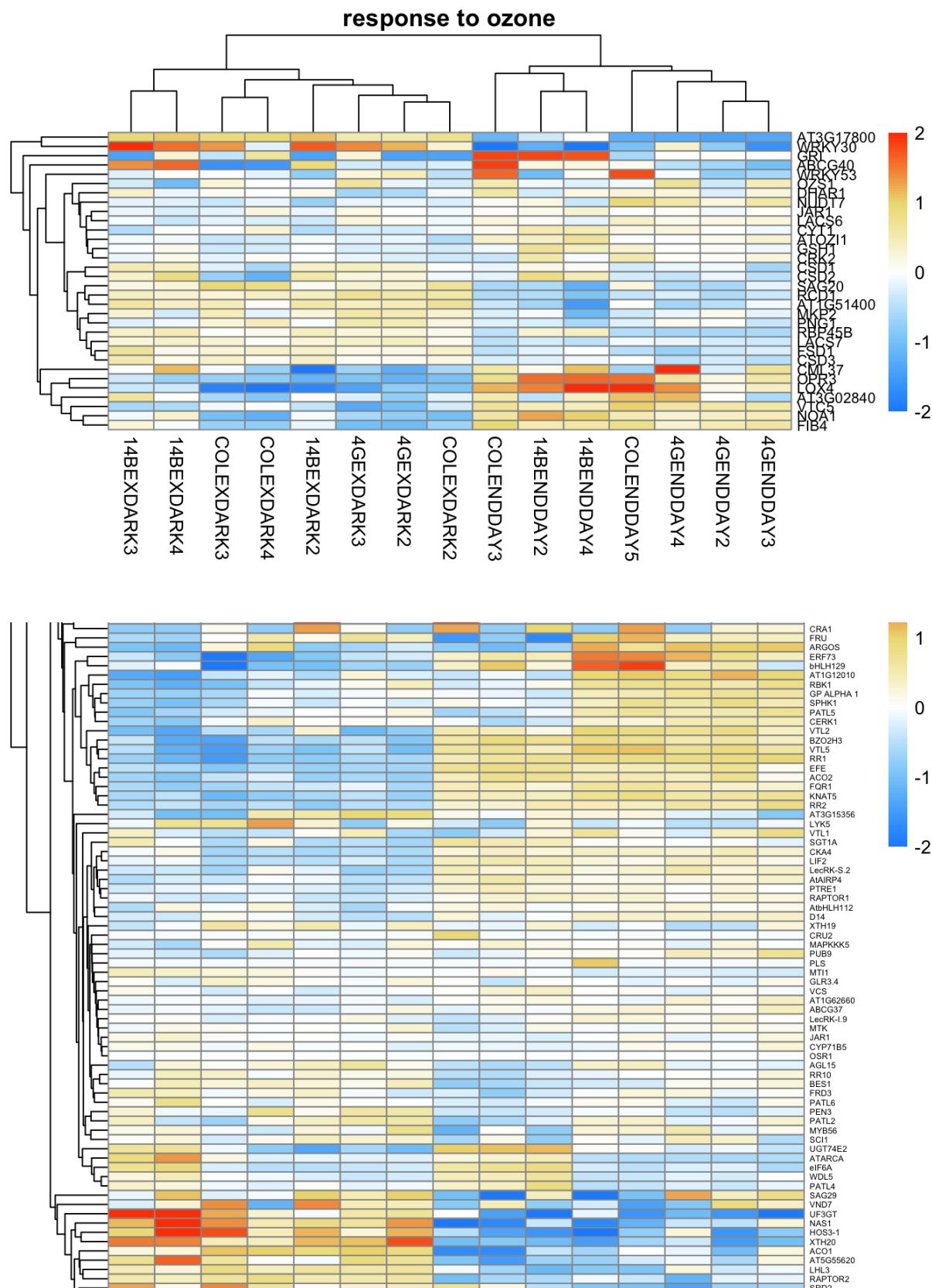


Arabidopsis thaliana



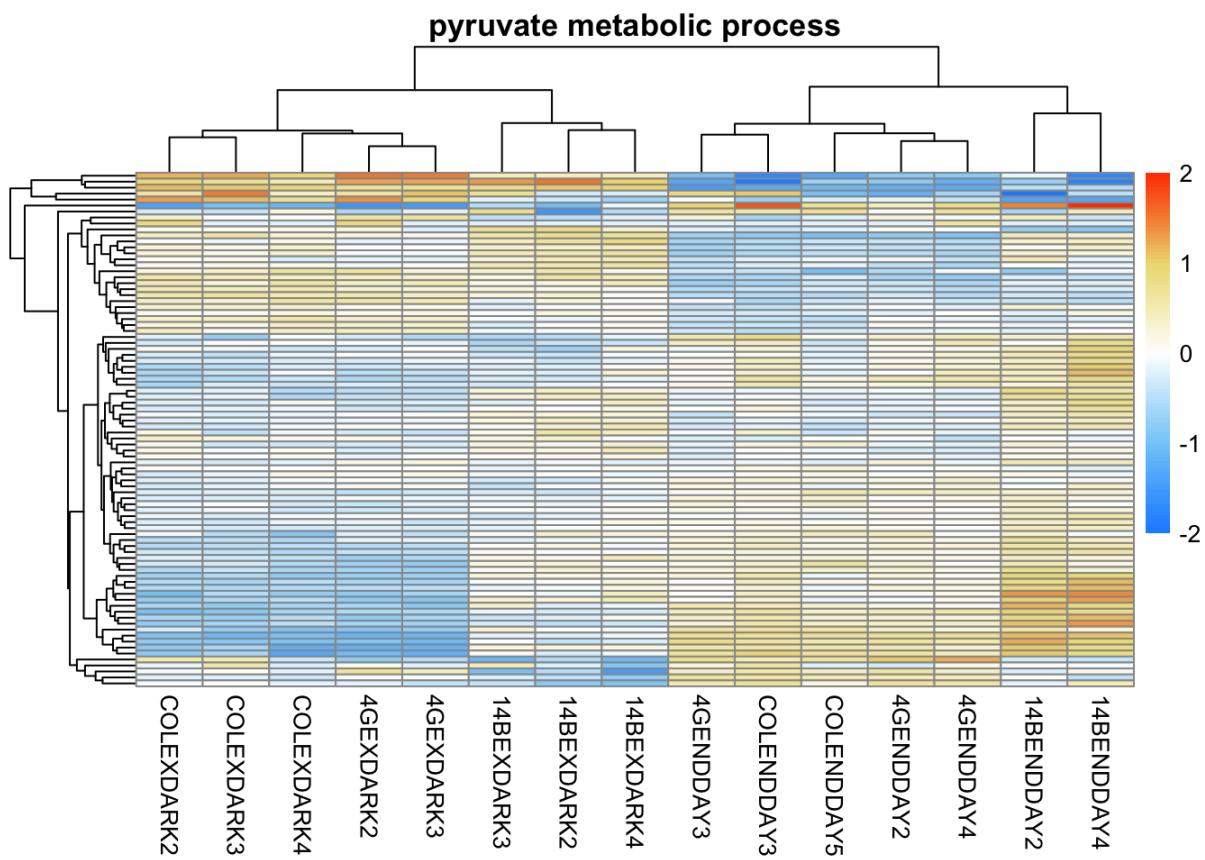
anion homeostasis



**Specific pHatmap per Gene Grouping**

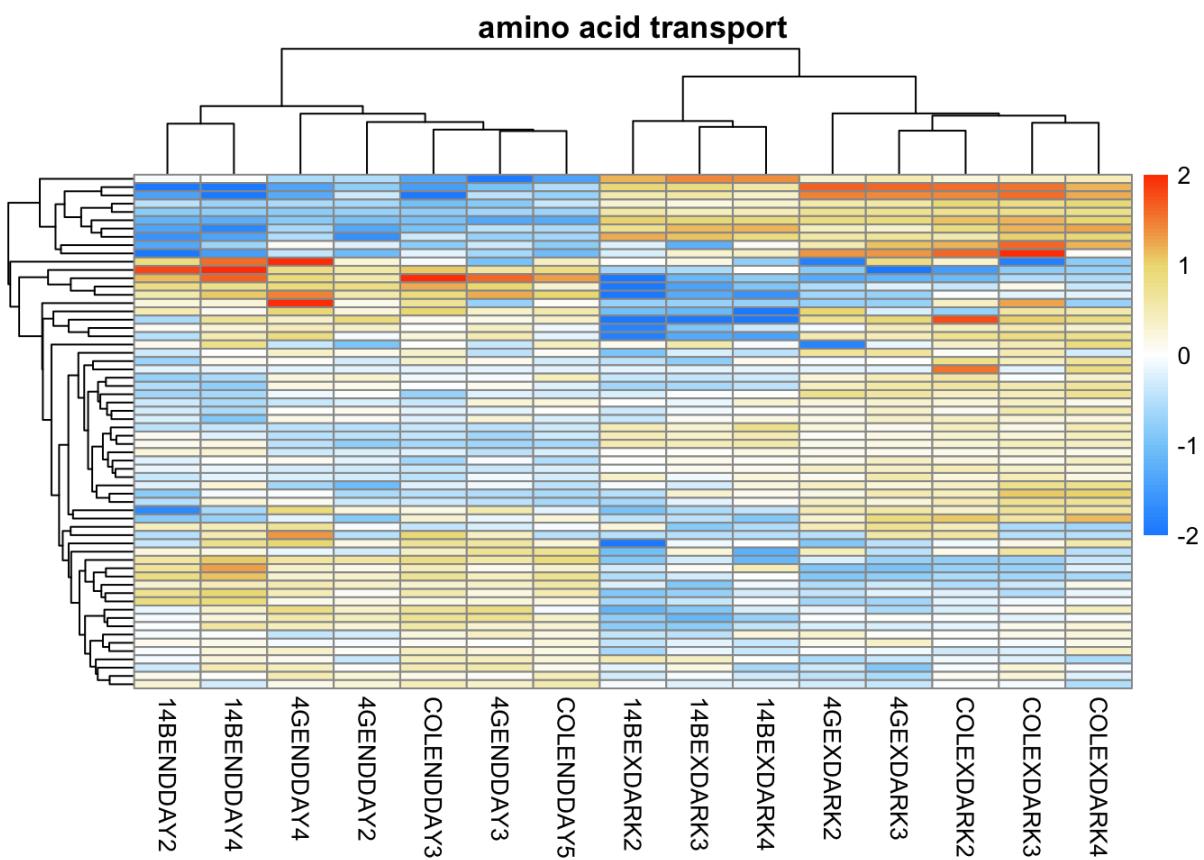
```
# "pyruvate metabolic process"
temp_set <- joined_set %>% filter(gene_ontology_name == "pyruvate metabolic process")
lgGo <- lgNorm[temp_set$gene, ]

heatData = lgGo - rowMeans(lgGo)
heatData = as.data.frame(heatData)
heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
fontsize_row = 10 - nrow(heatData) / 15
pheatmap(
  heatData,
  color = heatPalette,
  clustering_method = "average",
  labels_row= geneNamesAndDescriptions[rownames(heatData), "symbol"],
  main = "pyruvate metabolic process",
  show_rownames = FALSE
)
```



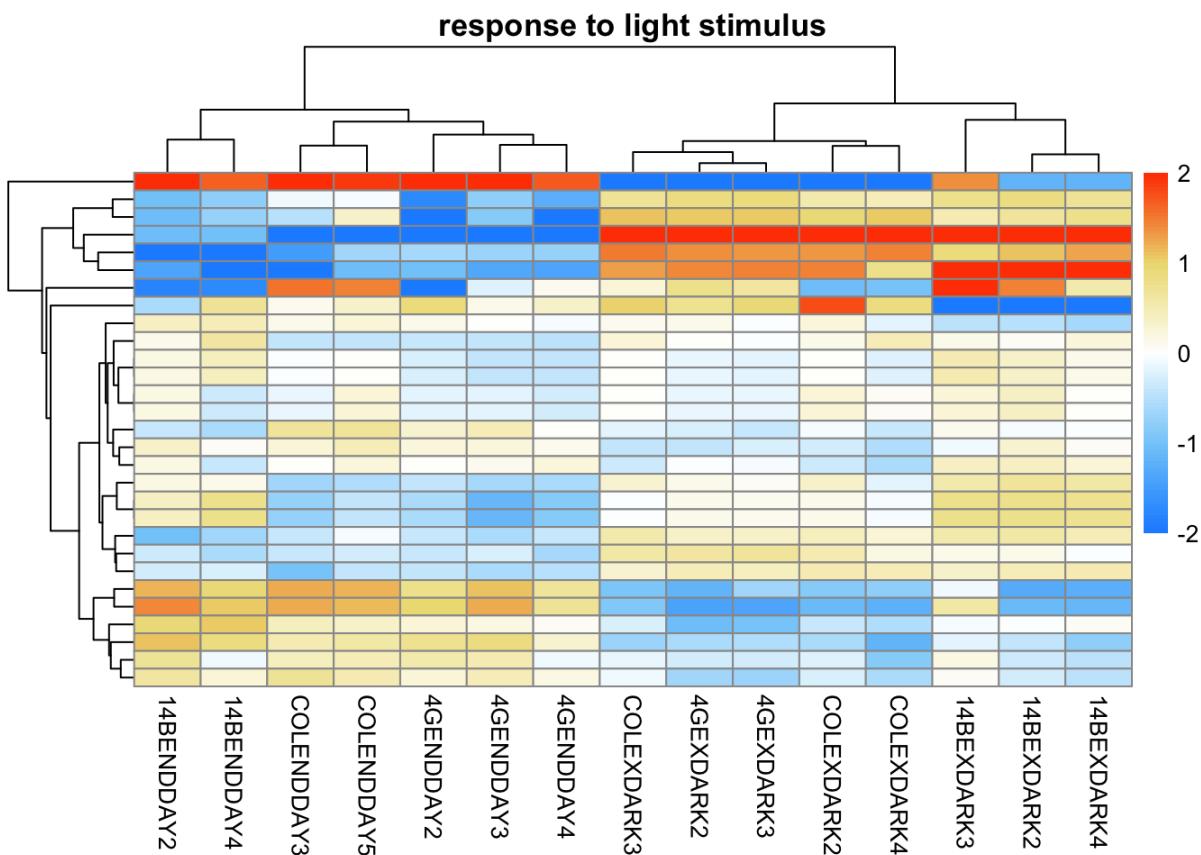
```
# "amino acid transport"
temp_set <- joined_set %>% filter(gene_ontology_name == "amino acid transport")
lgGo <- lgNorm[temp_set$gene, ]

heatData = lgGo - rowMeans(lgGo)
heatData = as.data.frame(heatData)
heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
fontsize_row = 10 - nrow(heatData) / 15
pheatmap(
  heatData,
  color = heatPalette,
  clustering_method = "average",
  labels_row= geneNamesAndDescriptions[rownames(heatData), "symbol"],
  main = "amino acid transport",
  show_rownames = FALSE
)
```



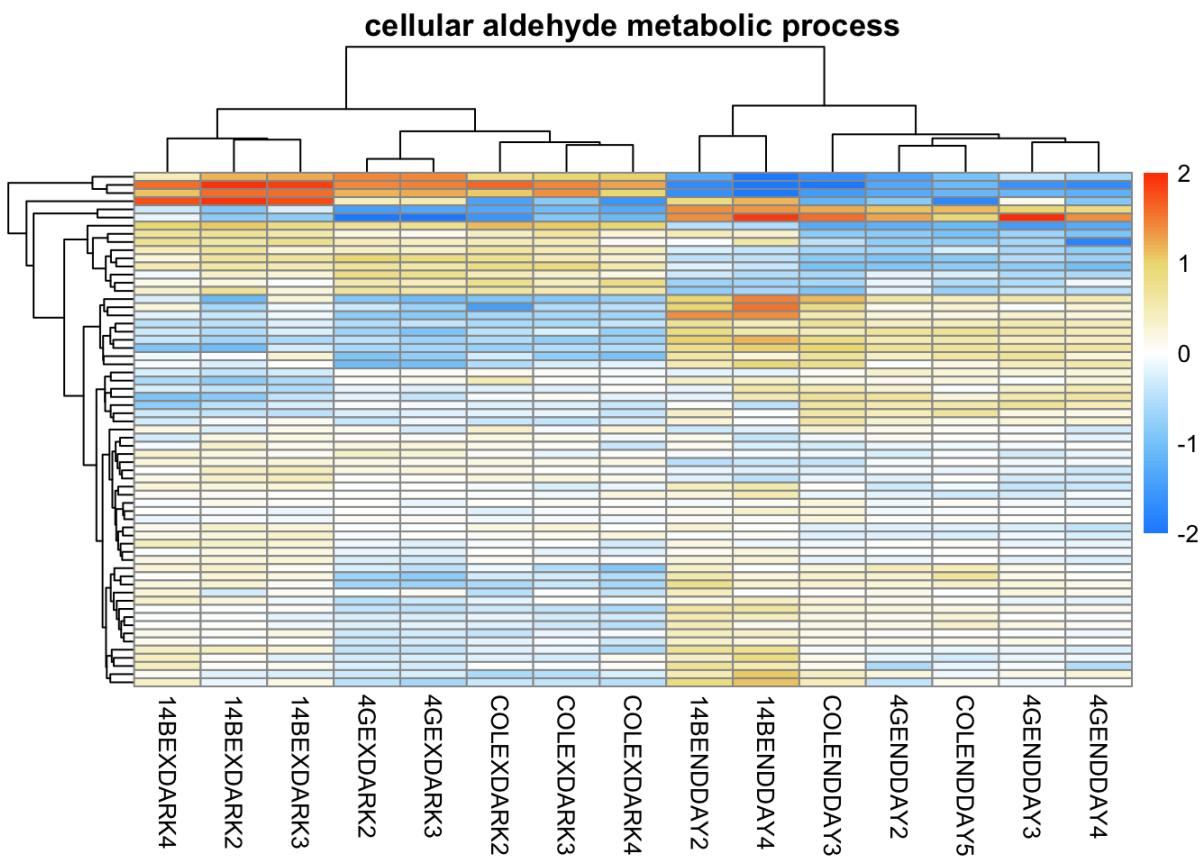
```
# "response to light stimulus"
temp_set <- goAssociations2 %>% filter(gene_ontology_name == "response to light stimulus")
lgGo <- lgNorm[temp_set$gene, ]

heatData = lgGo - rowMeans(lgGo)
heatData = as.data.frame(heatData)
heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
fontsize_row = 10 - nrow(heatData) / 15
pheatmap(
  heatData,
  color = heatPalette,
  clustering_method = "average",
  main = "response to light stimulus",
  show_rownames = FALSE
)
```



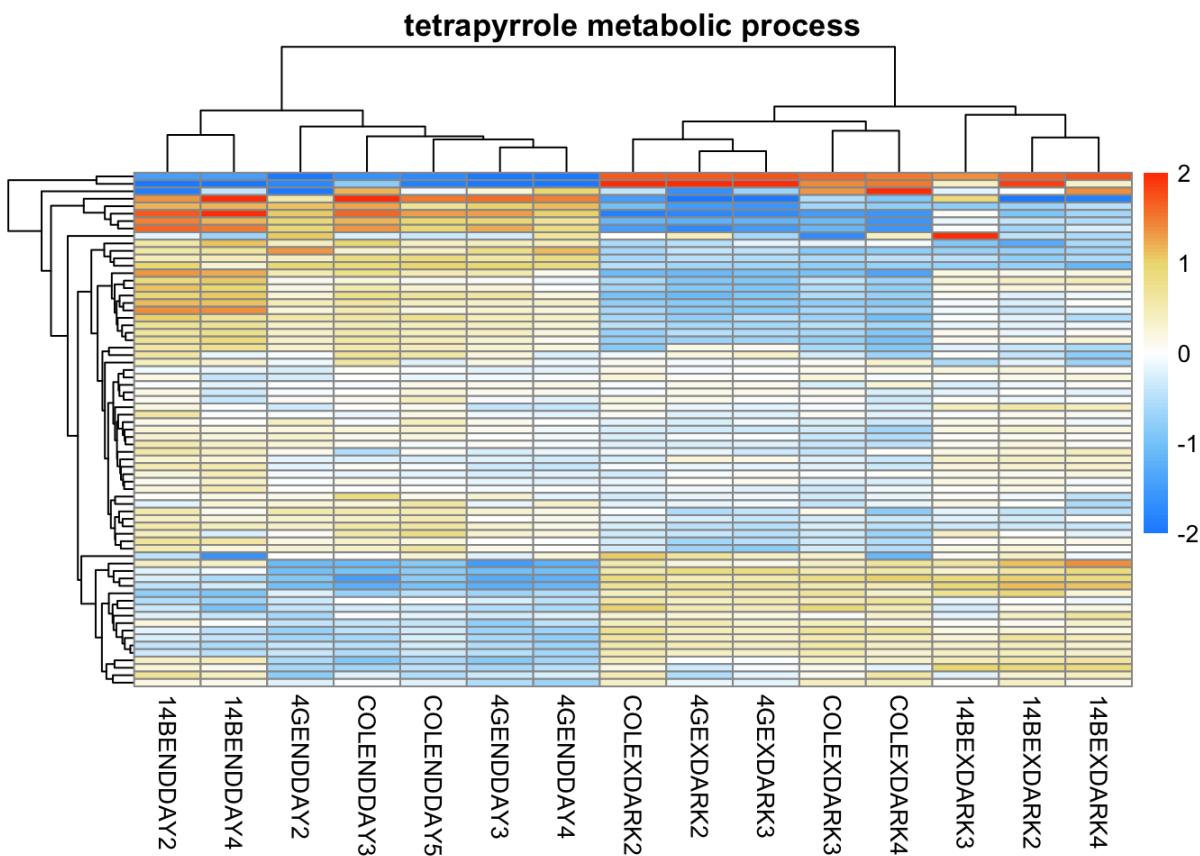
```
# "cellular aldehyde metabolic process"
temp_set <- joined_set %>% filter(gene_ontology_name == "cellular aldehyde metabolic process")
lgGo <- lgNorm[temp_set$gene, ]

heatData = lgGo - rowMeans(lgGo)
heatData = as.data.frame(heatData)
heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
fontsize_row = 10 - nrow(heatData) / 15
pheatmap(
  heatData,
  color = heatPalette,
  clustering_method = "average",
  labels_row= geneNamesAndDescriptions[rownames(heatData), "symbol"],
  main = "cellular aldehyde metabolic process",
  show_rownames = FALSE
)
```



```
# "tetrapyrrole metabolic process"
temp_set <- joined_set %>% filter(gene_ontology_name == "tetrapyrrole metabolic process")
lgGo <- lgNorm[temp_set$gene, ]

heatData = lgGo - rowMeans(lgGo)
heatData = as.data.frame(heatData)
heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
fontsize_row = 10 - nrow(heatData) / 15
pheatmap(
  heatData,
  color = heatPalette,
  clustering_method = "average",
  labels_row= geneNamesAndDescriptions[rownames(heatData), "symbol"],
  main = "tetrapyrrole metabolic process",
  show_rownames = FALSE
)
```



```
clean_DESeq_padj <- which(!is.na(DESeq_Results$padj))
clean_DESeq_Results <- DESeq_Results[clean_DESeq_padj, ]
significant_genes <- rownames(clean_DESeq_Results[which(clean_DESeq_Results[, "padj"] <= 0.1), ])
clean_DESeq_Results <- data.frame(clean_DESeq_Results[which(clean_DESeq_Results[, "padj"] <= 0.1),
])
clean_DESeq_Results$gene <- significant_genes
RNA_Result_groupings <- clean_DESeq_Results %>% inner_join(goAssociations2, by = "gene")
RNA_Result_groupings %>% group_by(gene_ontology_name) %>% summarize(count = n()) %>% arrange(desc(count)) %>%
  filter(count >= 10) %>%
knitr::kable(caption = "Distribution of Significant Genes Groupings for RNA-Seq")
```

Distribution of Significant Genes Groupings for RNA-Seq

gene_ontology_name	count
pyruvate metabolic process	18
amino acid transport	16
response to light stimulus	14
cellular aldehyde metabolic process	13
tetrapyrrole metabolic process	13
cold acclimation	11
protein-chromophore linkage	10

```

clean_DESeq_padj <- which(!is.na(DESeq_Results$padj))
clean_DESeq_Results <- DESeq_Results[clean_DESeq_padj, ]
significant_genes <- rownames(clean_DESeq_Results[which(clean_DESeq_Results[, "padj"] <= 0.1), ])
clean_DESeq_Results <- data.frame(clean_DESeq_Results[which(clean_DESeq_Results[, "padj"] <= 0.1),
])
clean_DESeq_Results$gene <- significant_genes
RNA_Result_groupings <- clean_DESeq_Results %>% inner_join(goAssociations2, by = "gene")
RNA_Result_groupings %>% group_by(gene_ontology_name) %>% summarize(count = n()) %>% arrange(desc(count)) %>%
knitr:::kable(caption = "Distribution of Significant Genes Groupings for RNA-Seq")

```

Distribution of Significant Genes Groupings for RNA-Seq

gene_ontology_name	count
pyruvate metabolic process	18
amino acid transport	16
response to light stimulus	14
cellular aldehyde metabolic process	13
tetrapyrrole metabolic process	13
cold acclimation	11
protein-chromophore linkage	10
response to cold	9
cellular response to endogenous stimulus	8
response to abscisic acid	7
response to disaccharide	7
response to ozone	7
translation	7
anion homeostasis	6
circadian rhythm	5
photosynthesis	5
response to red or far red light	5

gene_ontology_name	count
regulation of transcription	4
response to far red light	4
carbohydrate metabolic process	3
cellular amino acid metabolic process	3
protein transport	3
proteolysis	3
response to light intensity	3
alpha-amino acid metabolic process	2
chlorophyll biosynthetic process	2
defense response	2
developmental growth	2
gluconeogenesis	2
intracellular protein transport	2
nitrate assimilation	2
photosynthetic electron transport chain	2
protein autoubiquitination	2
protein folding	2
response to oxidative stress	2
response to salicylic acid	2
translational elongation	2
abscisic acid-activated signaling pathway	1
actin filament bundle assembly	1
ammonium homeostasis	1
auxin metabolic process	1
branched-chain amino acid biosynthetic process	1
carbohydrate transport	1
cell differentiation	1
cell division	1
cellular amino acid biosynthetic process	1
cellular catabolic process	1
cellular lipid metabolic process	1
cellular response to abscisic acid stimulus	1
cellular response to hypoxia	1
cellular response to light stimulus	1

gene_ontology_name	count
cellular response to salicylic acid stimulus	1
chloroplast thylakoid membrane	1
chloroplast-nucleus signaling pathway	1
cysteine biosynthetic process	1
cytokinin-activated signaling pathway	1
cytoplasmic translation	1
de novo' NAD biosynthetic process from aspartate	1
exocytosis	1
gibberellic acid mediated signaling pathway	1
glycerophospholipid catabolic process	1
hormone-mediated signaling pathway	1
inositol catabolic process	1
jasmonic acid metabolic process	1
maturation of LSU-rRNA from tricistronic rRNA transcript	1
mitochondrial translational elongation	1
molecular_function	1
mRNA export from nucleus	1
negative regulation of gene expression	1
negative regulation of transcription	1
negative regulation of translation in response to stress	1
nucleosome assembly	1
oxoacid metabolic process	1
peptidyl-cysteine S-nitrosylation	1
phospholipid metabolic process	1
plant-type secondary cell wall biogenesis	1
positive regulation of ubiquitin-protein transferase activity	1
protein quality control for misfolded or incompletely synthesized proteins	1
proton transmembrane transport	1
regulation of defense response	1
regulation of gene expression	1
regulation of nitrogen utilization	1
response to wounding	1
starch biosynthetic process	1
sterol biosynthetic process	1

gene_ontology_name	count
translational initiation	1
transmembrane transport	1
valine biosynthetic process	1

Strip Charts per Gene Groupings

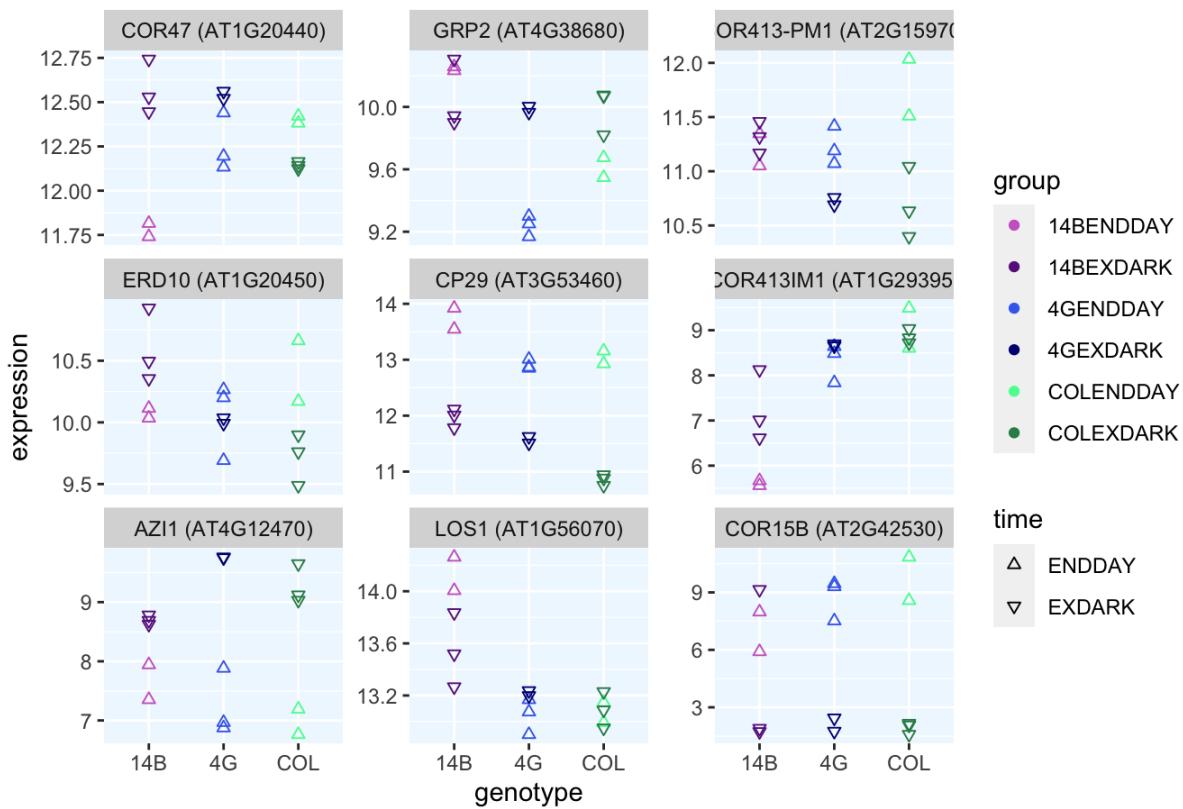
```

for (i in 1:length(ontology_names)){
  temp_set <- joined_set %>% filter(gene_ontology_name == ontology_names[i])
  lgGo <- lgNorm[temp_set$gene, ]

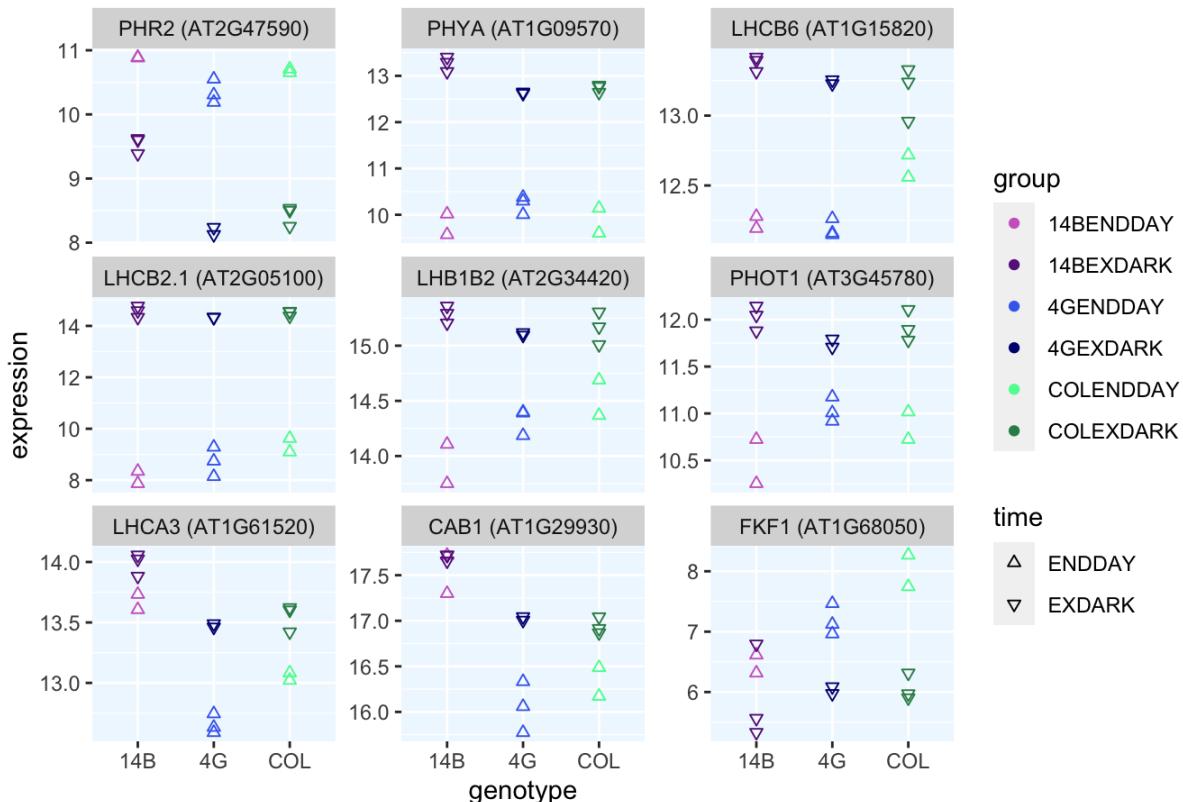
  DESeq_Results1 = DESeq_Results[temp_set$gene, ]
  DESeq_Results1 = DESeq_Results1[order(DESeq_Results1$pvalue), ]
  DESeq_Results1 = DESeq_Results1[1:9, ]
  lgGo1 <- lgNorm[rownames(DESeq_Results1), ]
  stripChart <- stripchart321g(data = lgGo1,
                                sampleAnnotation = sampleAnnotation,
                                geneNames = geneNamesAndDescriptions,
                                colorValues = groupColors
                                )
  stripChart = stripChart + ggtitle(ontology_names[i])
  stripChart = stripChart + theme(panel.background = element_rect(fill = 'aliceblue'))
  print(stripChart)
}

```

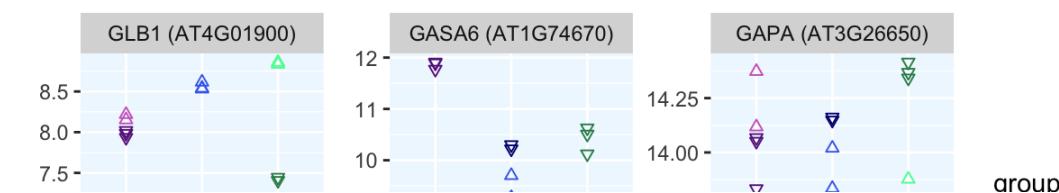
cold acclimation

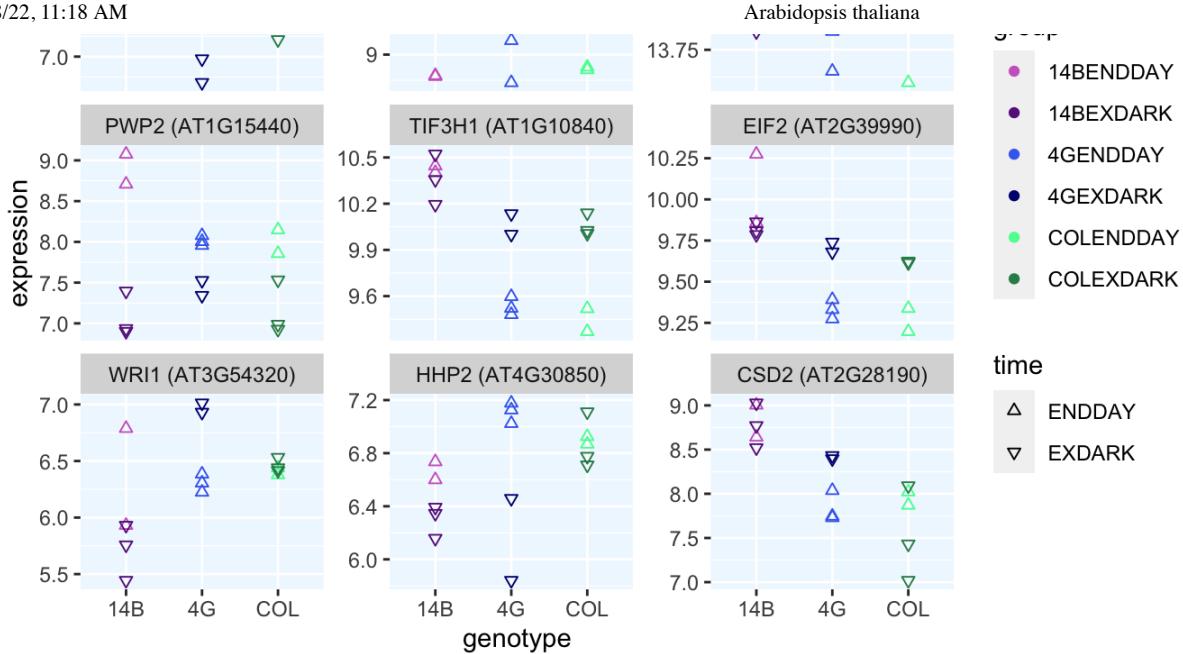


protein-chromophore linkage

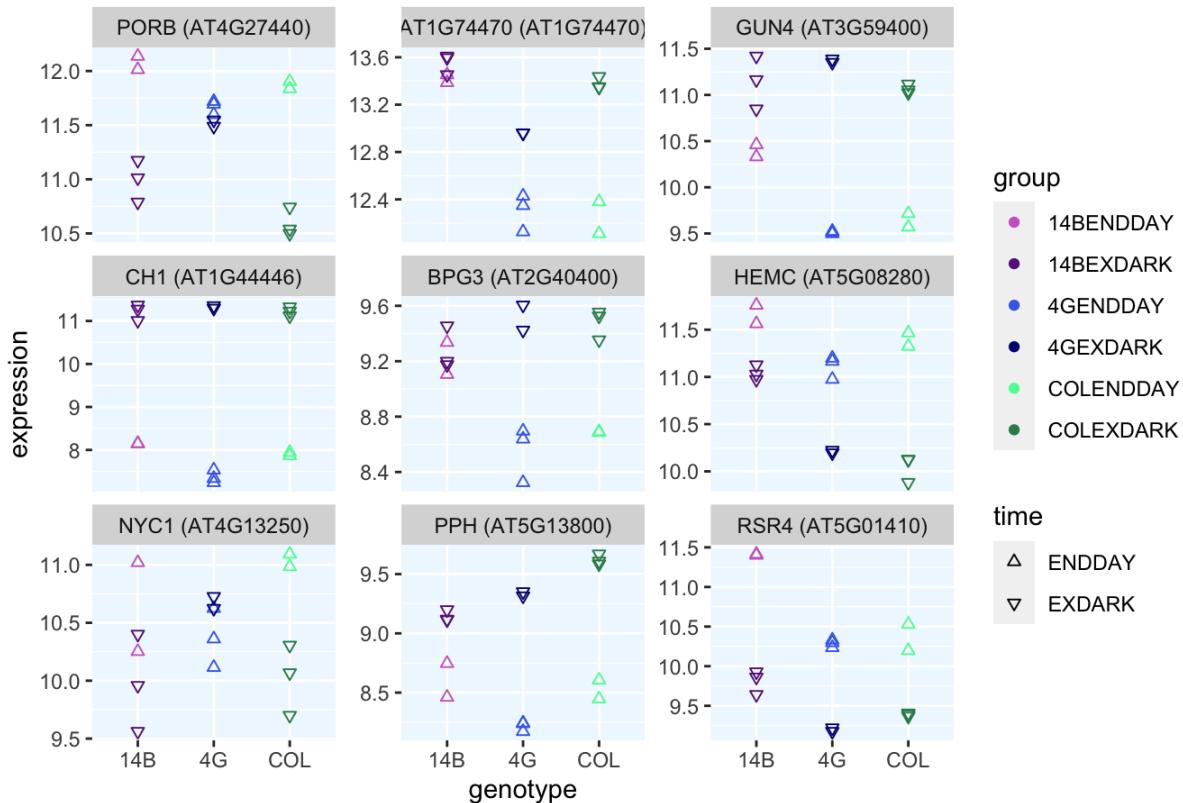


response to disaccharide

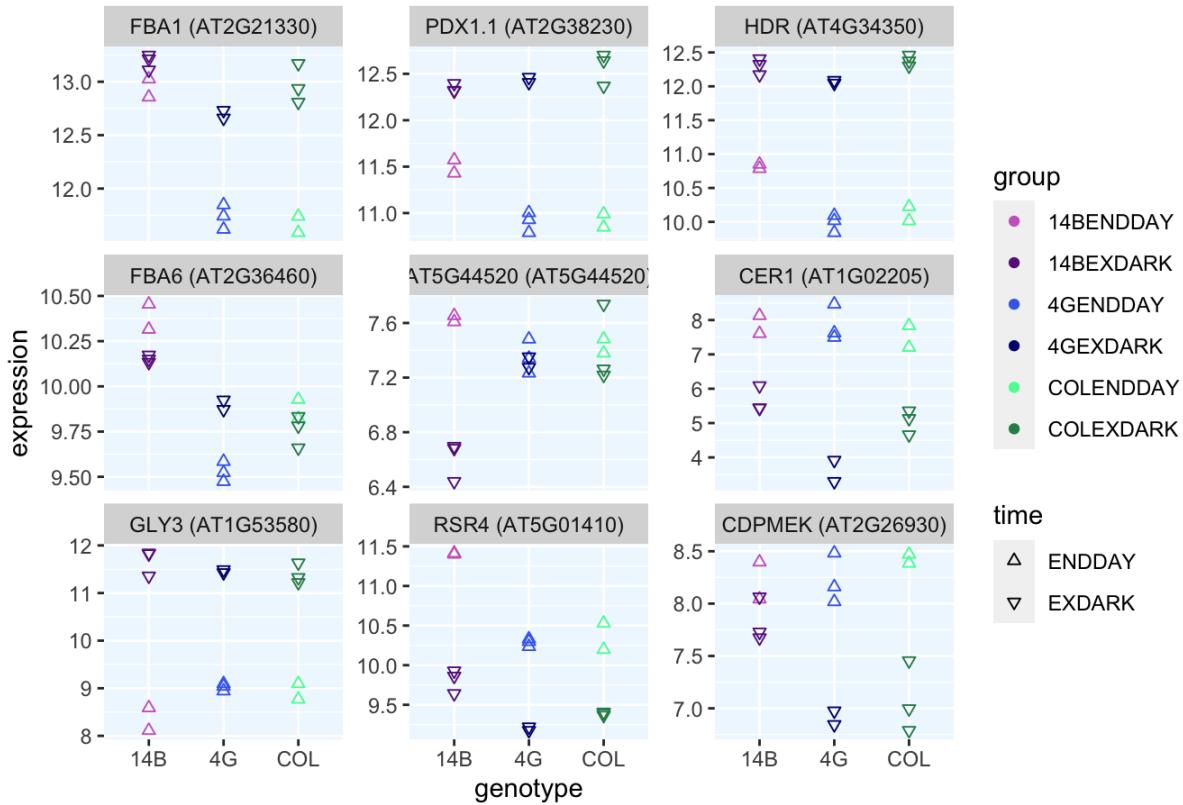




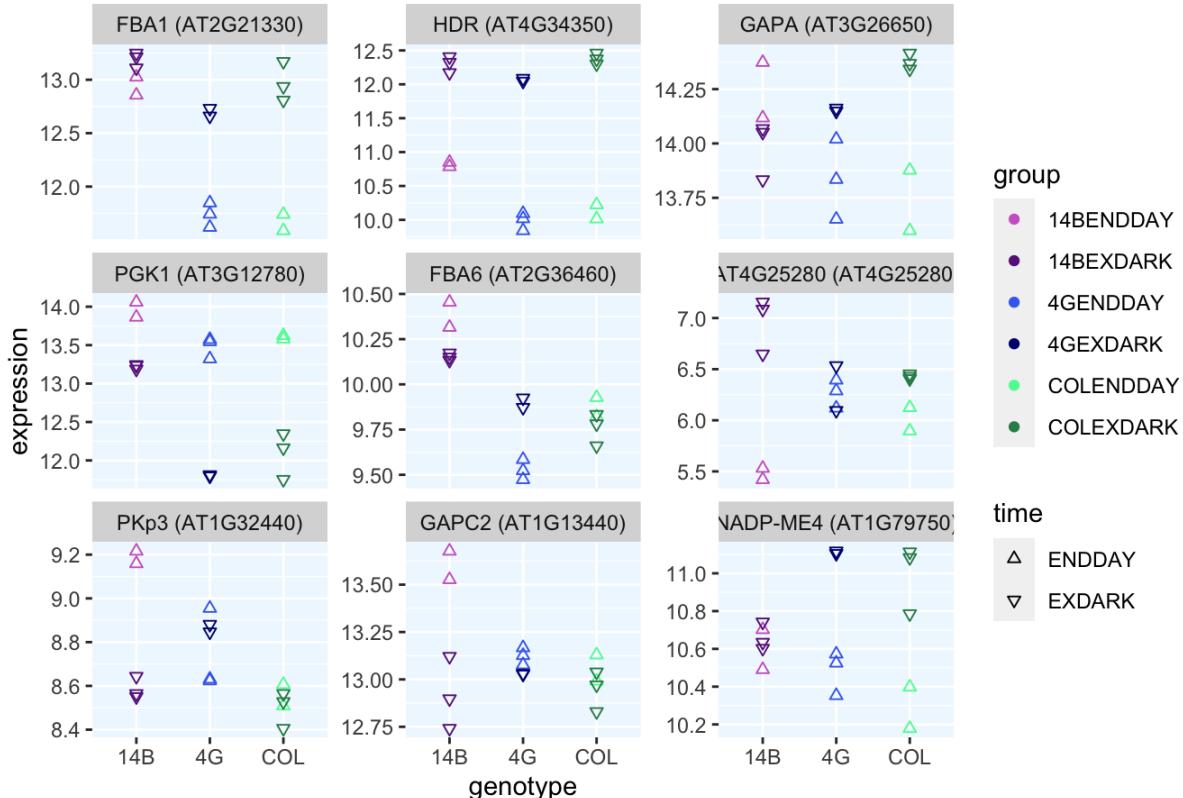
tetrapyrrole metabolic process



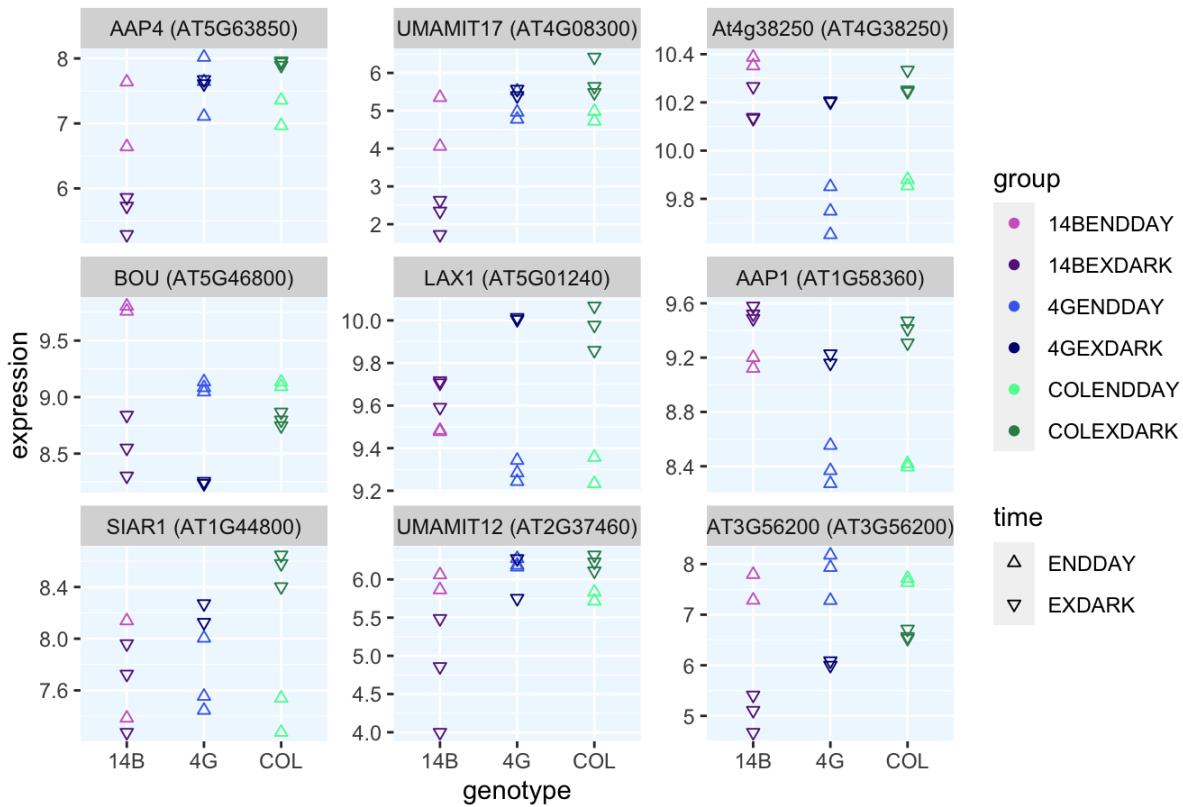
cellular aldehyde metabolic process



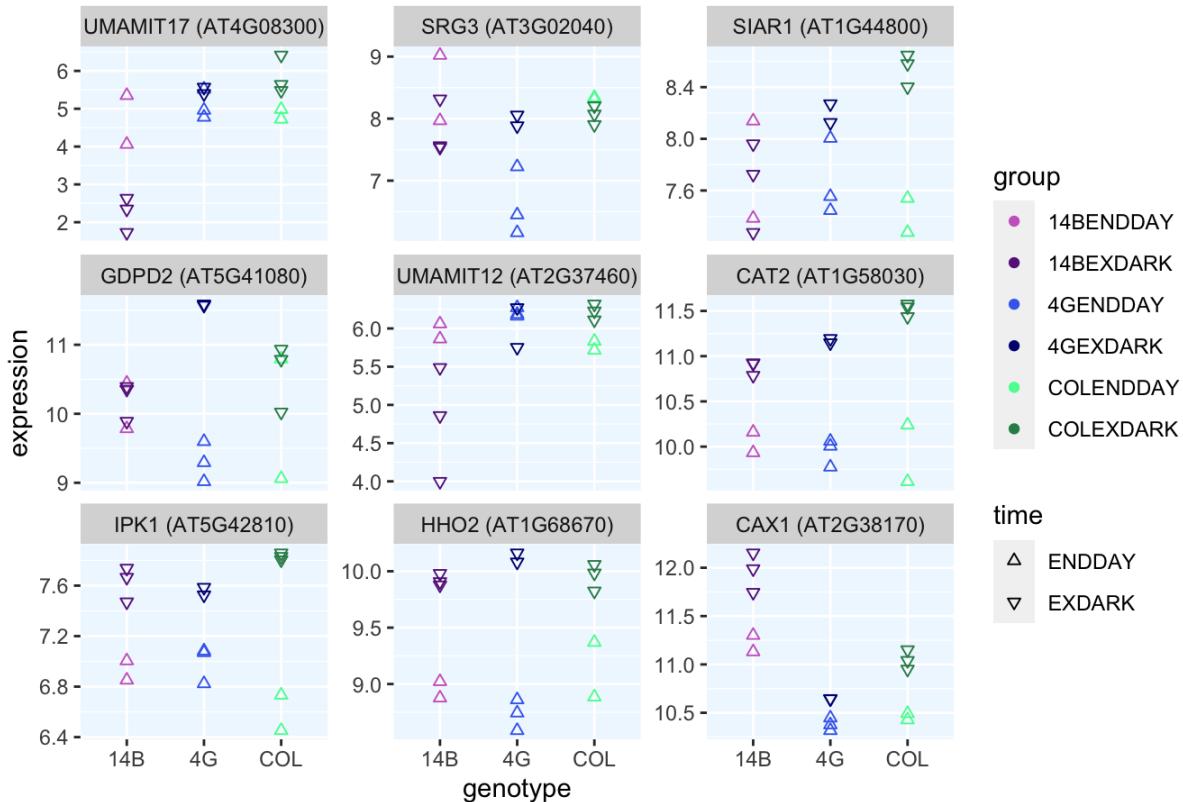
pyruvate metabolic process



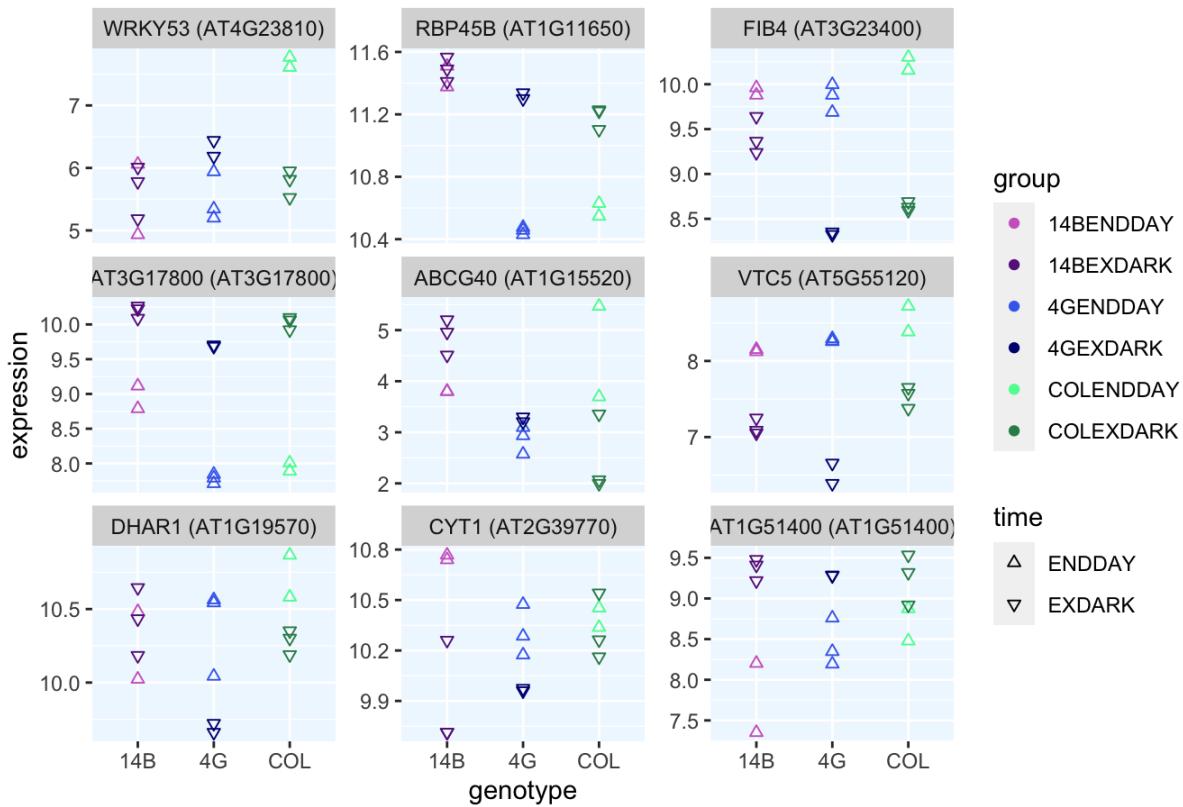
amino acid transport



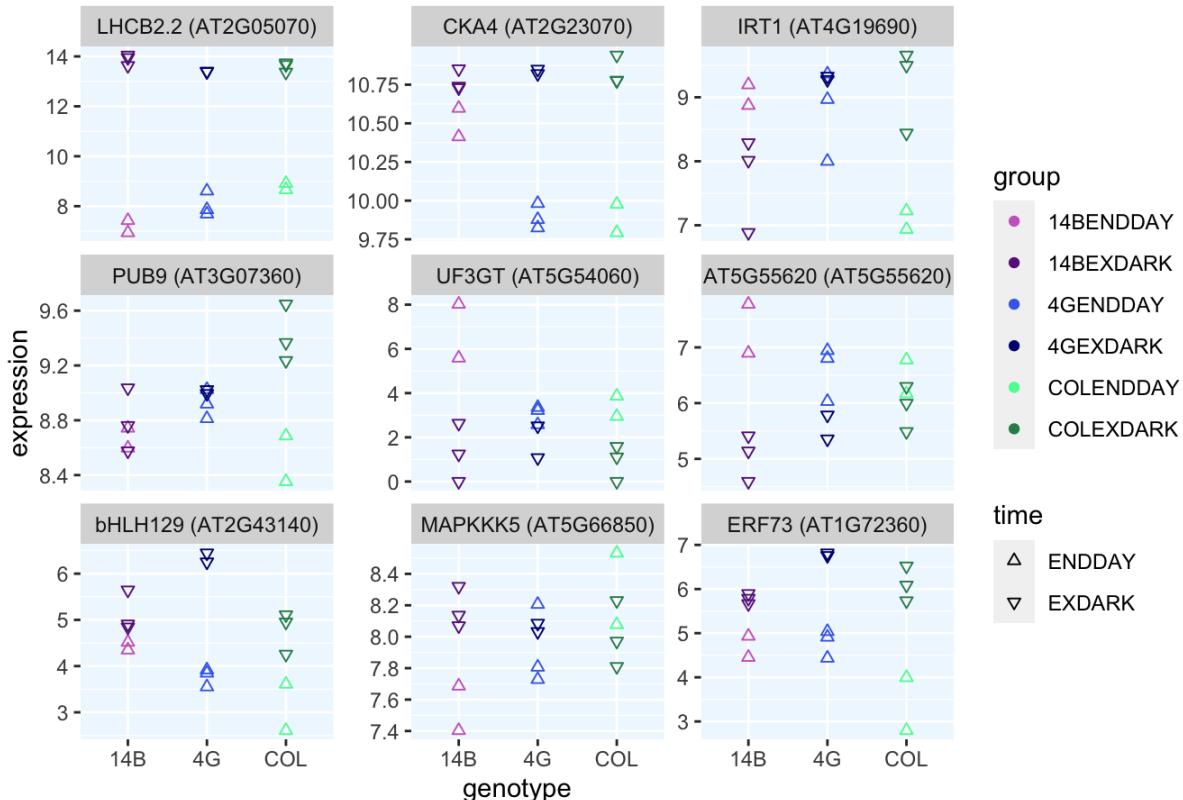
anion homeostasis



response to ozone



cellular response to endogenous stimulus



```
# remove the outlier
riboCounts <- riboCounts %>% select(-`4GEXDARK`)
sampleAnnotation2 <- sampleAnnotation2[colnames(riboCounts),]
```

DEseq with ribo-seq data

```

start_time <- Sys.time()
DESeqDataSet = DESeqDataSetFromMatrix(
  countData = riboCounts,
  colData = sampleAnnotation2,
  design = ~ time + genotype + time:genotype
)

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors

```

```

DESeqDataSet = DESeq(
  DESeqDataSet,
  parallel=FALSE,
  test = "LRT",
  reduced = ~ time + genotype
)

```

```
## estimating size factors
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
## fitting model and testing
```

```

DESeq_Results <- results(DESeqDataSet)
clean_DESeq_padj <- which(!is.na(DESeq_Results$padj))
Ribo_Sig <- sum(DESeq_Results[clean_DESeq_padj, "padj"] <= 0.1)
Ribo_Sig

```

```
## [1] 167
```

```
sum(DESeq_Results[clean_DESeq_padj, "padj"] <= 0.1) * 0.1
```

```
## [1] 16.7
```

```

end_time = Sys.time()
difference = end_time - start_time
difference

```

```
## Time difference of 16.9889 secs
```

```
temp <- as.data.frame(DESeq_Results[clean_DESeq_padj,])
head(temp %>% filter(padj <= 0.1) %>% arrange(padj))
```

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
## AT3G12780	966.02813	-1.973210	0.3169155	60.84003	6.148297e-14	2.073206e-10
## AT5G52310	48.02205	-3.281634	0.5017682	43.50565	3.571640e-10	6.021784e-07
## AT1G29930	587.02780	-3.330625	0.5372171	40.08236	1.977995e-09	2.223266e-06
## AT1G70700	68.55675	2.668277	0.5179129	35.02523	2.479523e-08	2.090238e-05
## AT3G27160	116.68503	-1.872538	0.4337160	33.26401	5.981533e-08	3.466640e-05
## AT5G41080	104.47642	2.634899	0.6193788	33.20248	6.168398e-08	3.466640e-05

cut out -> not needed

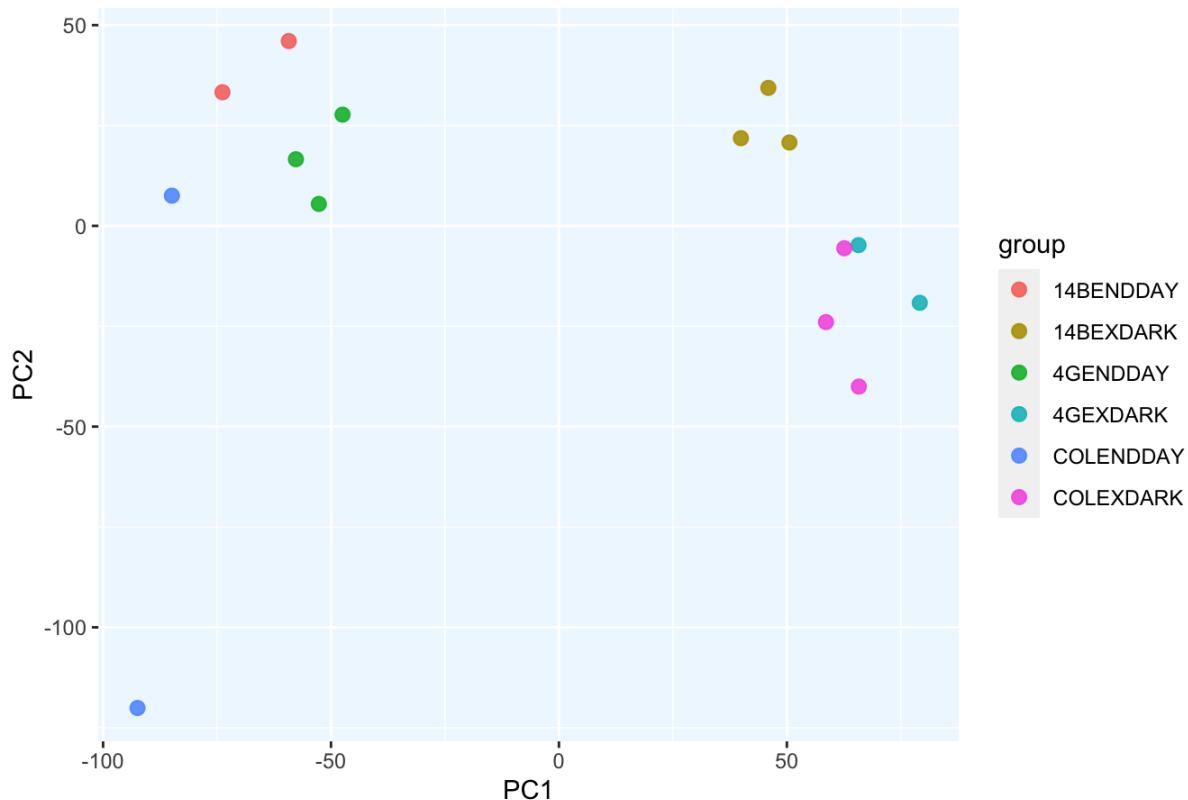
```
temp = DESeq_Results[clean_DESeq_padj,]
temp = as.data.frame(temp)
ribo_sig_genes = c()
ribo_sig_values = c()
for (i in 1:nrow(temp)){
  if (temp$padj[i] <= 0.1){
    ribo_sig_genes = c(ribo_sig_genes, rownames(temp)[i])
    ribo_sig_values = c(ribo_sig_values, temp$baseMean[i])
  }
}
```

```
lgNormRibo = log2(counts(DESeqDataSet, normalized=TRUE) + 1)
```

```
pca = prcomp(t(lgNormRibo))
pcaData = data.frame(pca$x[, 1:2])
pcaData$group = sampleAnnotation(rownames(pcaData), "group")
pcaData$sample = rownames(pcaData)

gg = ggplot(pcaData, aes(x=PC1, y=PC2, color=group, label=sample))
gg = gg + geom_point(size=2.5, alpha=0.9)
gg = gg + theme(panel.background = element_rect(fill = 'aliceblue')) + ggtitle("Overall PCA for Rib o-Seq")
print(gg)
```

Overall PCA for Ribo-Seq



```
joined_set_ribo = inner_join(goAssociations, geneNamesAndDescriptions, by = "gene")
ontology_names <- joined_set_ribo %>% distinct(gene_ontology_name)
ontology_names <- ontology_names[["gene_ontology_name"]]
```

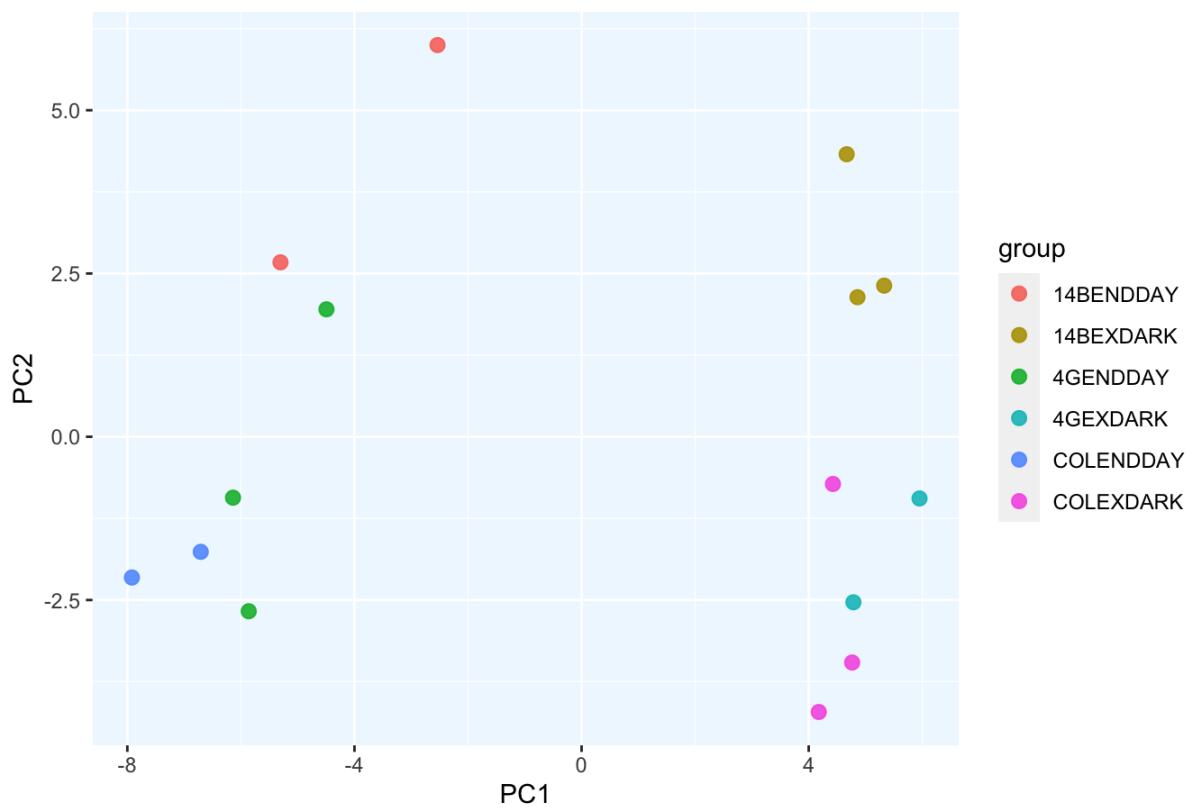
```
for (i in 1:length(ontology_names)){
  temp_set <- joined_set_ribo %>% filter(gene_ontology_name == ontology_names[i])
  lgGoRibo <- lgNormRibo[temp_set$gene, ]

  pca1 <- prcomp(t(lgGoRibo))
  pcaData1 = data.frame(pca1$x[, 1:2])
  pcaData1$group = sampleAnnotation[rownames(pcaData1), "group"]
  pcaData1$sample = rownames(pcaData1)

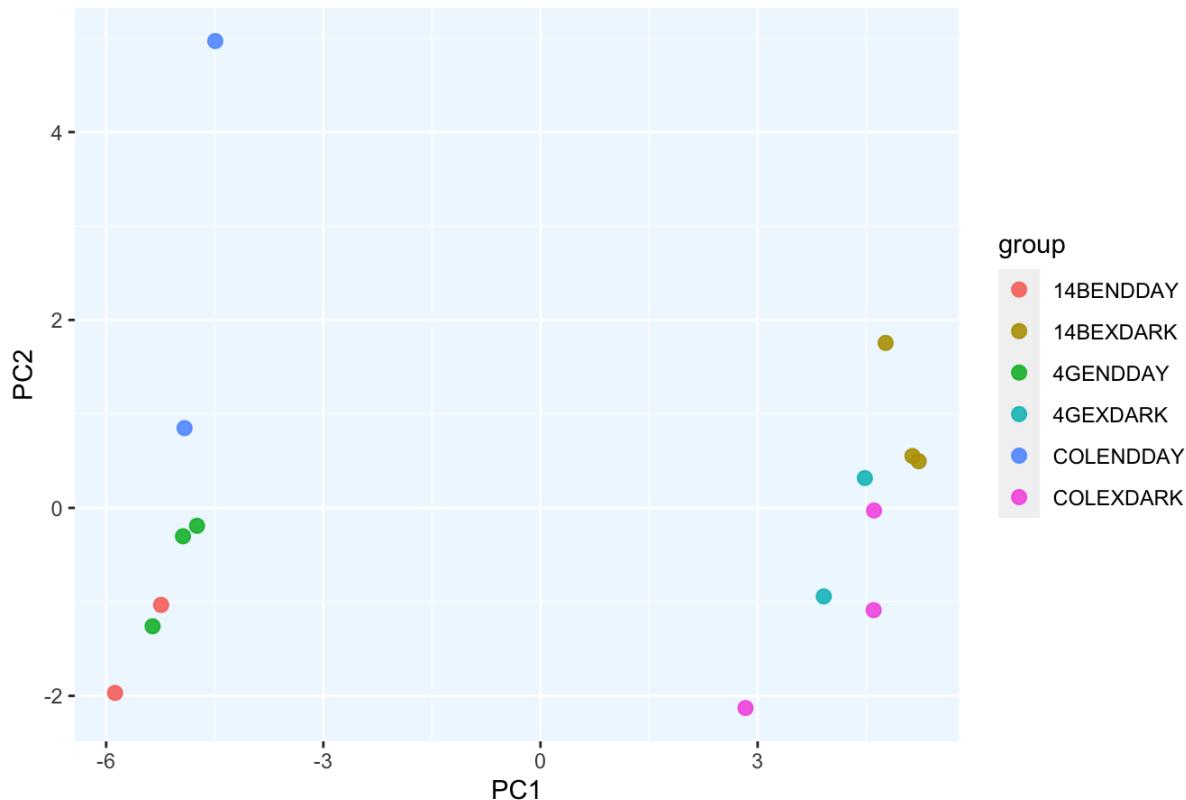
  gg = gg + geom_point(size=2.5, alpha=0.9)
  gg = gg + theme(panel.background = element_rect(fill = 'aliceblue'))

  gg1 = ggplot(pcaData1, aes(x=PC1, y=PC2, color=group, label=sample))
  gg1 = gg1 + geom_point(size=2.5, alpha=0.9) + ggtitle(ontology_names[i])
  gg1 = gg1 + theme(panel.background = element_rect(fill = 'aliceblue'))
  print(gg1)
}
```

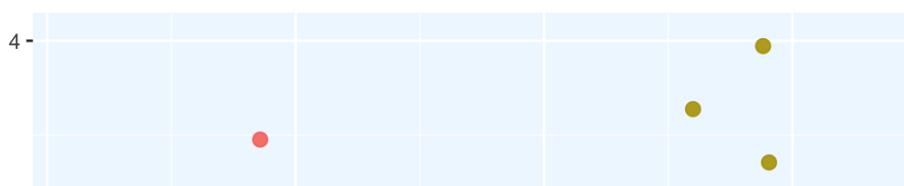

cold acclimation

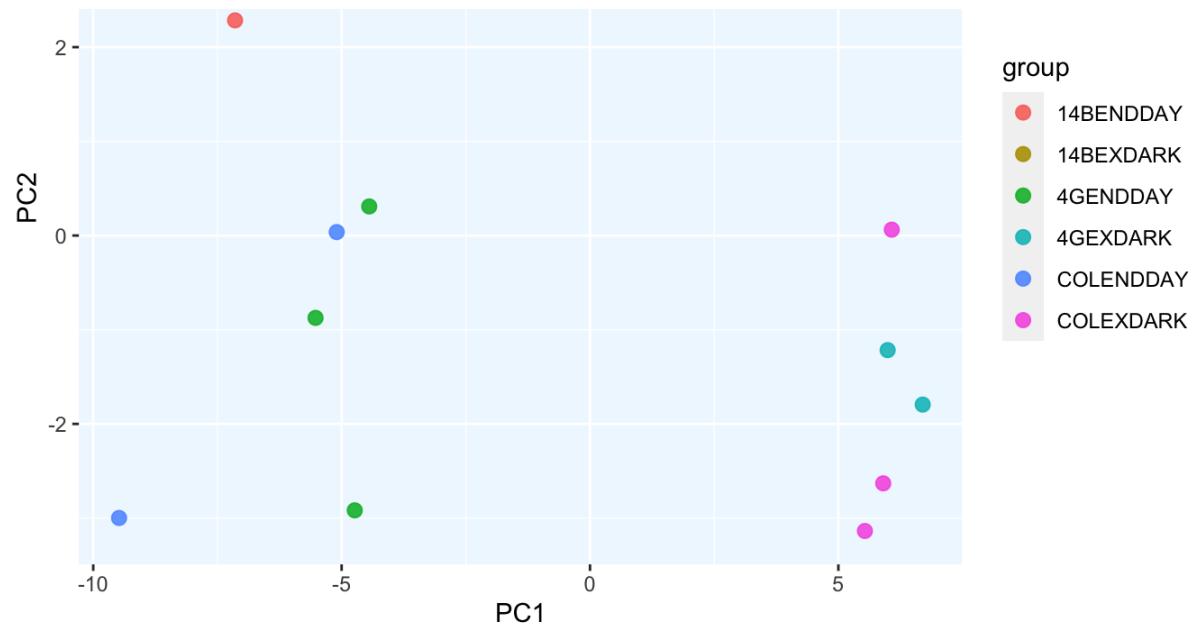
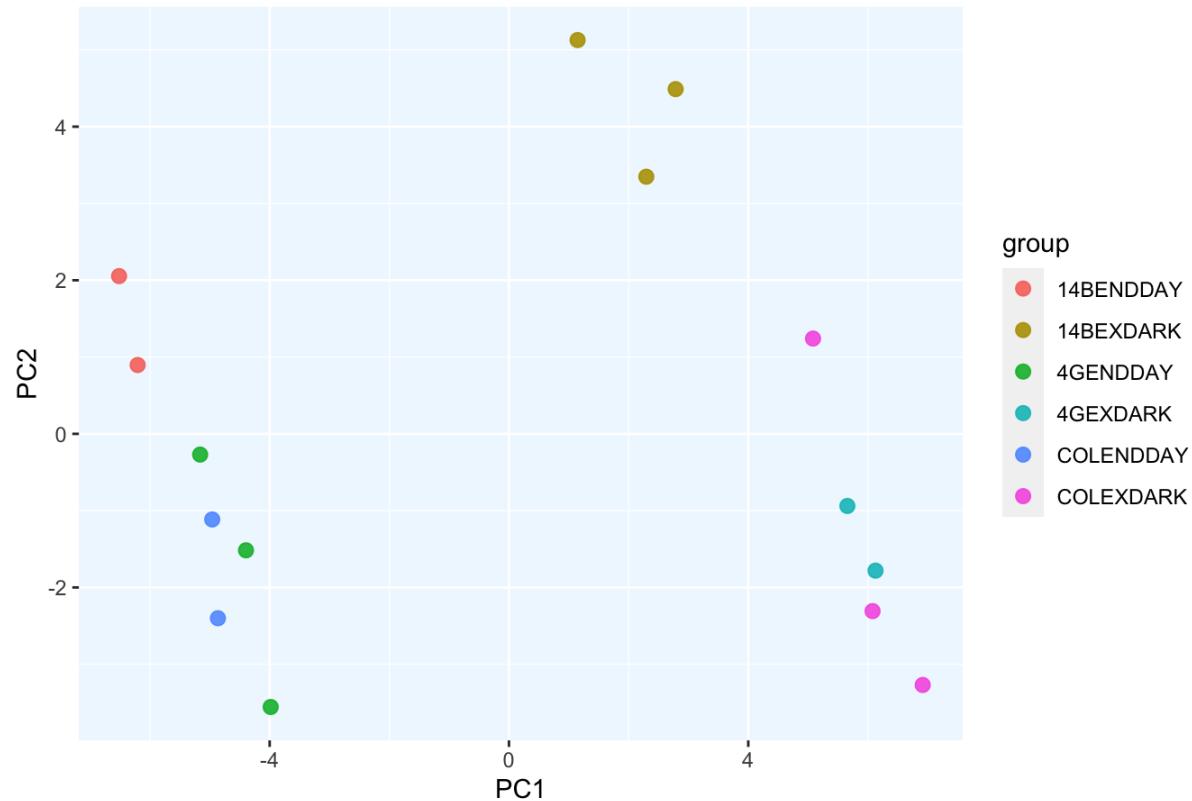


protein-chromophore linkage

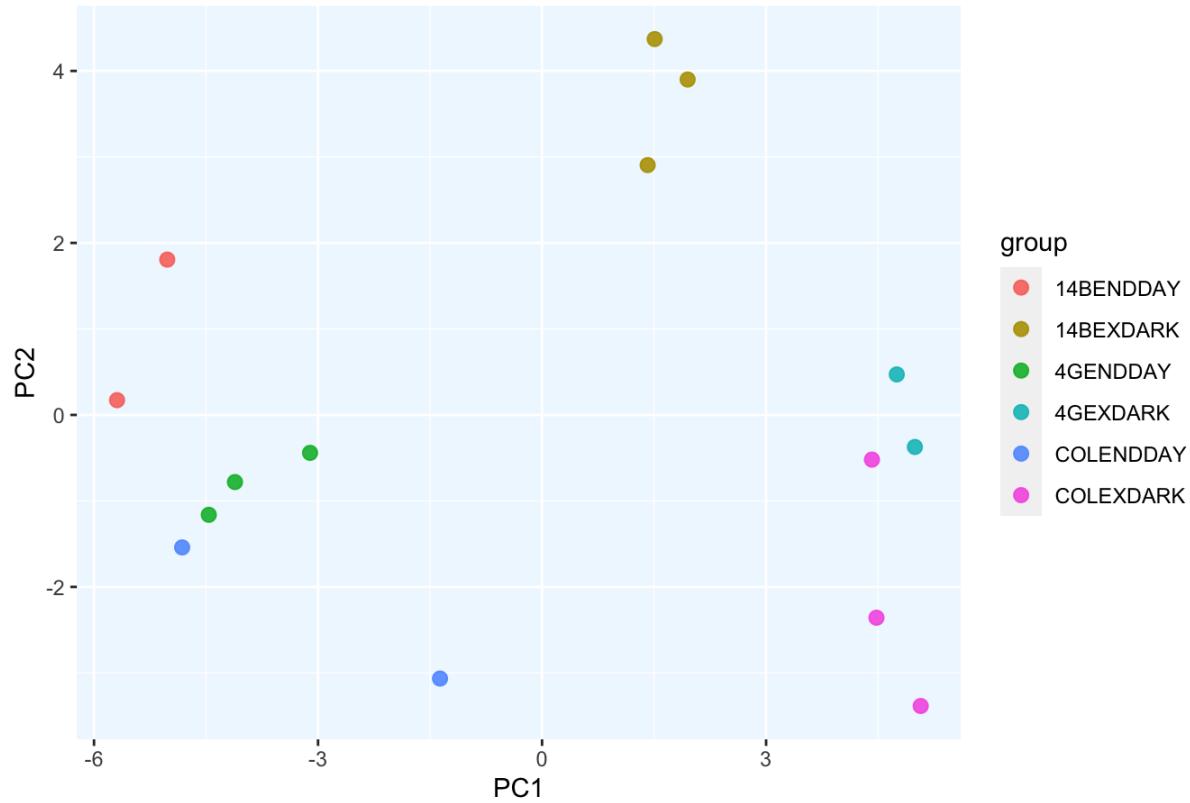


response to disaccharide

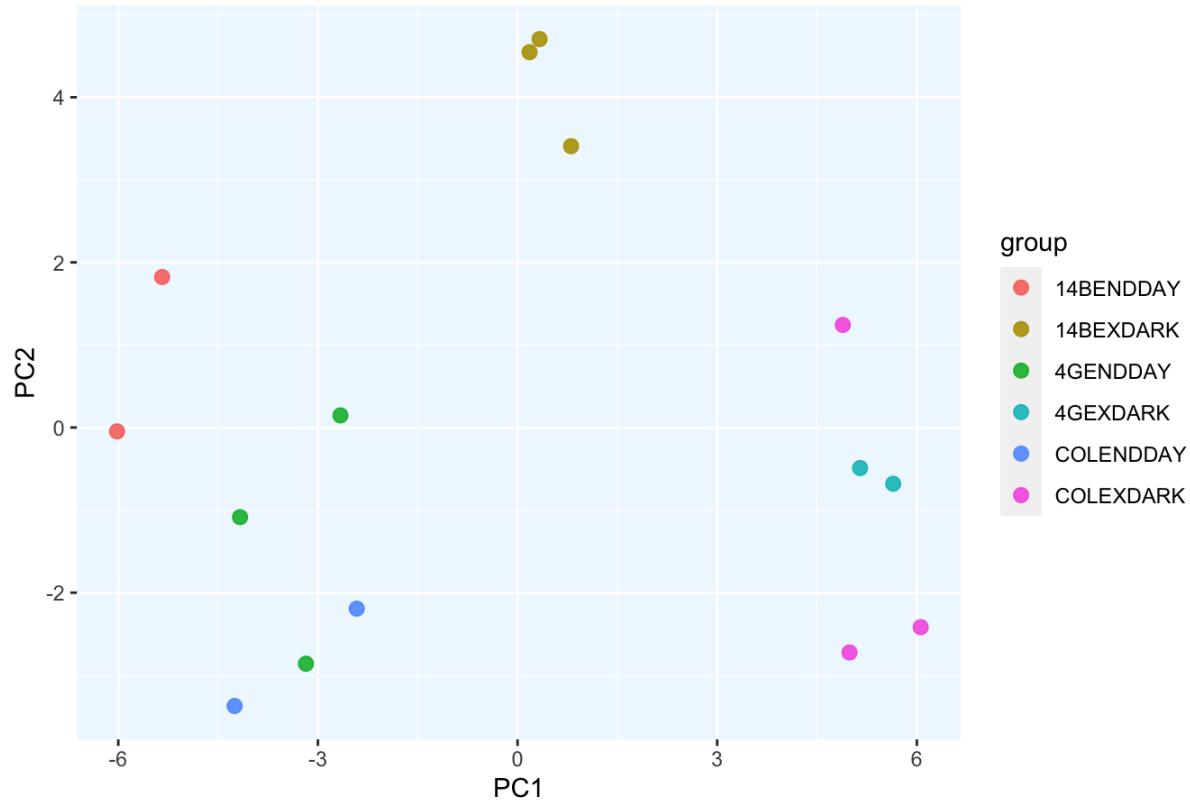


**tetrapyrrole metabolic process**

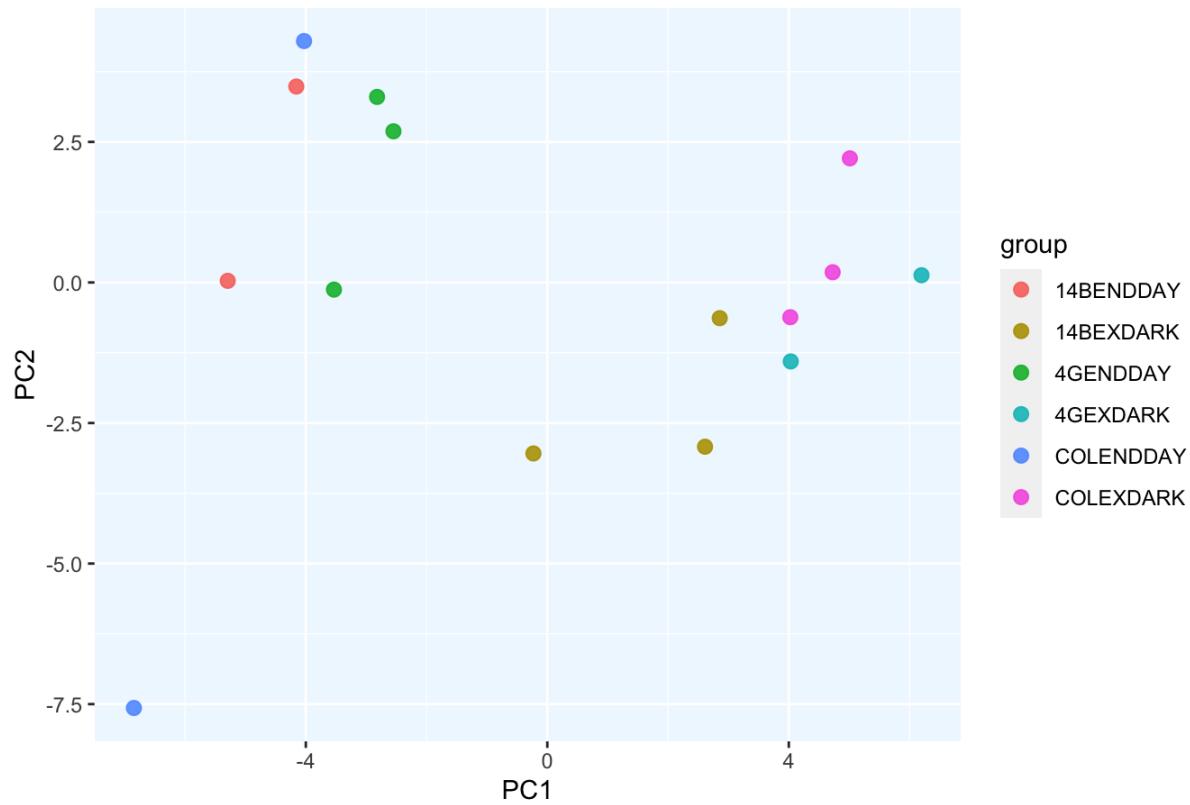
cellular aldehyde metabolic process



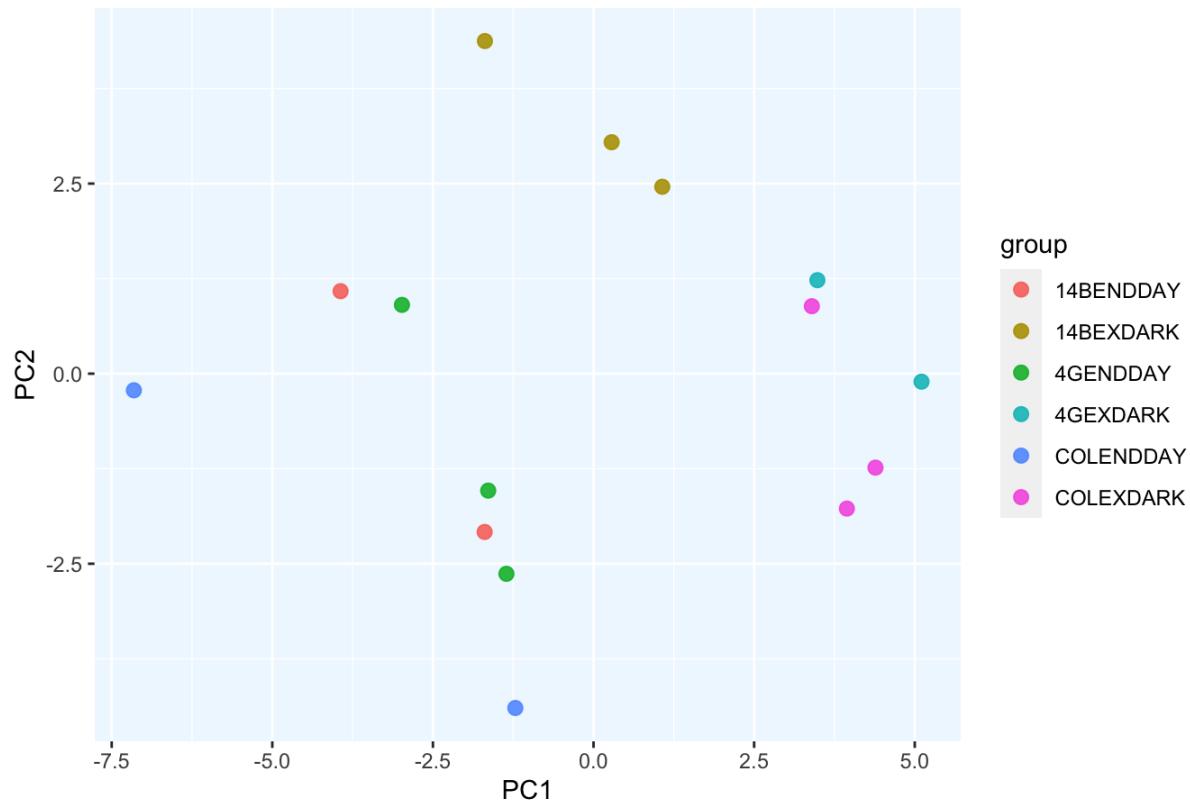
pyruvate metabolic process



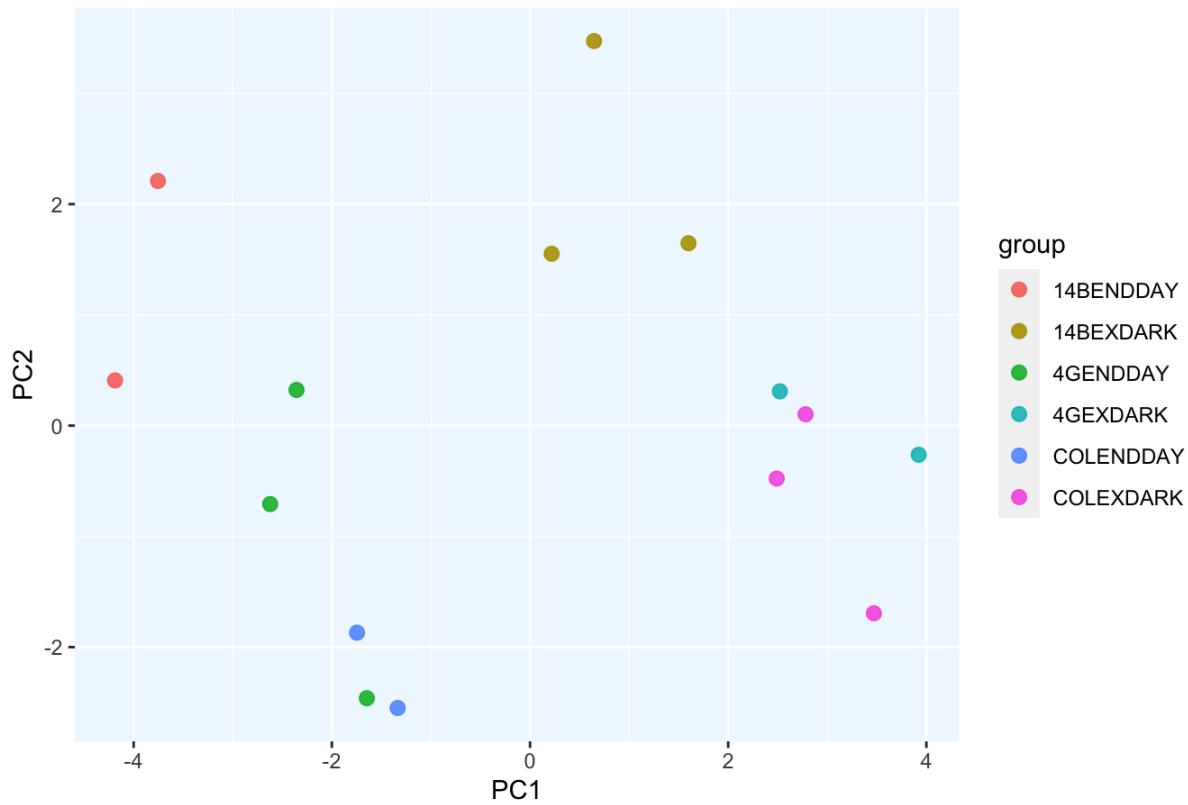
amino acid transport



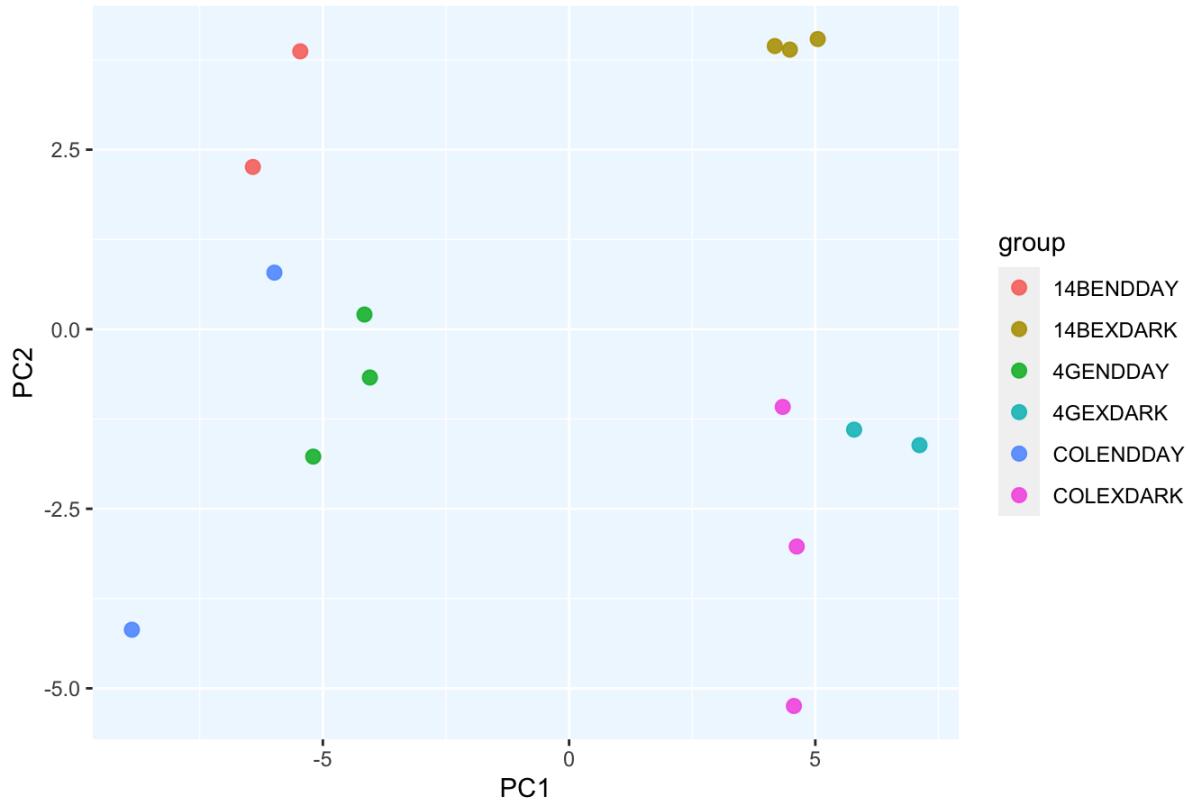
anion homeostasis



response to ozone



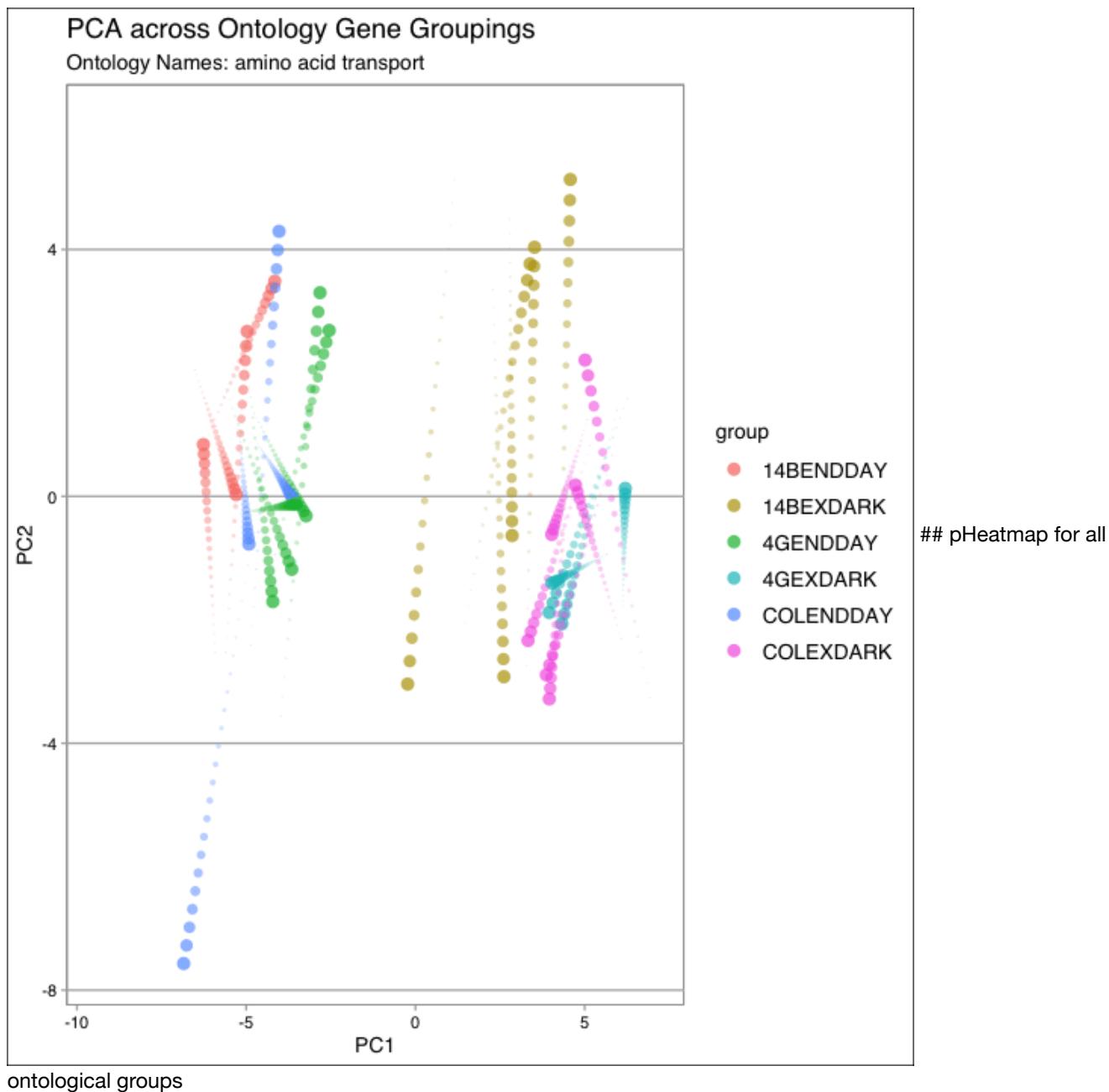
cellular response to endogenous stimulus



```

all_pca2 = data.frame()
for (i in 1:length(ontology_names)){
  temp_set <- joined_set_ribo %>% filter(gene_ontology_name == ontology_names[i])
  lgGoRibo <- lgNormRibo[temp_set$gene, ]
  pca1 <- prcomp(t(lgGoRibo))
  pcaData1 = data.frame(pca1$x[, 1:2])
  pcaData1$group = sampleAnnotation[rownames(pcaData1), "group"]
  pcaData1$sample = rownames(pcaData1)
  pcaData1$ontology_name = ontology_names[i]
  all_pca = rbind(all_pca, pcaData1)
}
write.table(all_pca,
            "all_pca2.csv",
            sep = ",",
            row.names = FALSE,
            quote = FALSE)

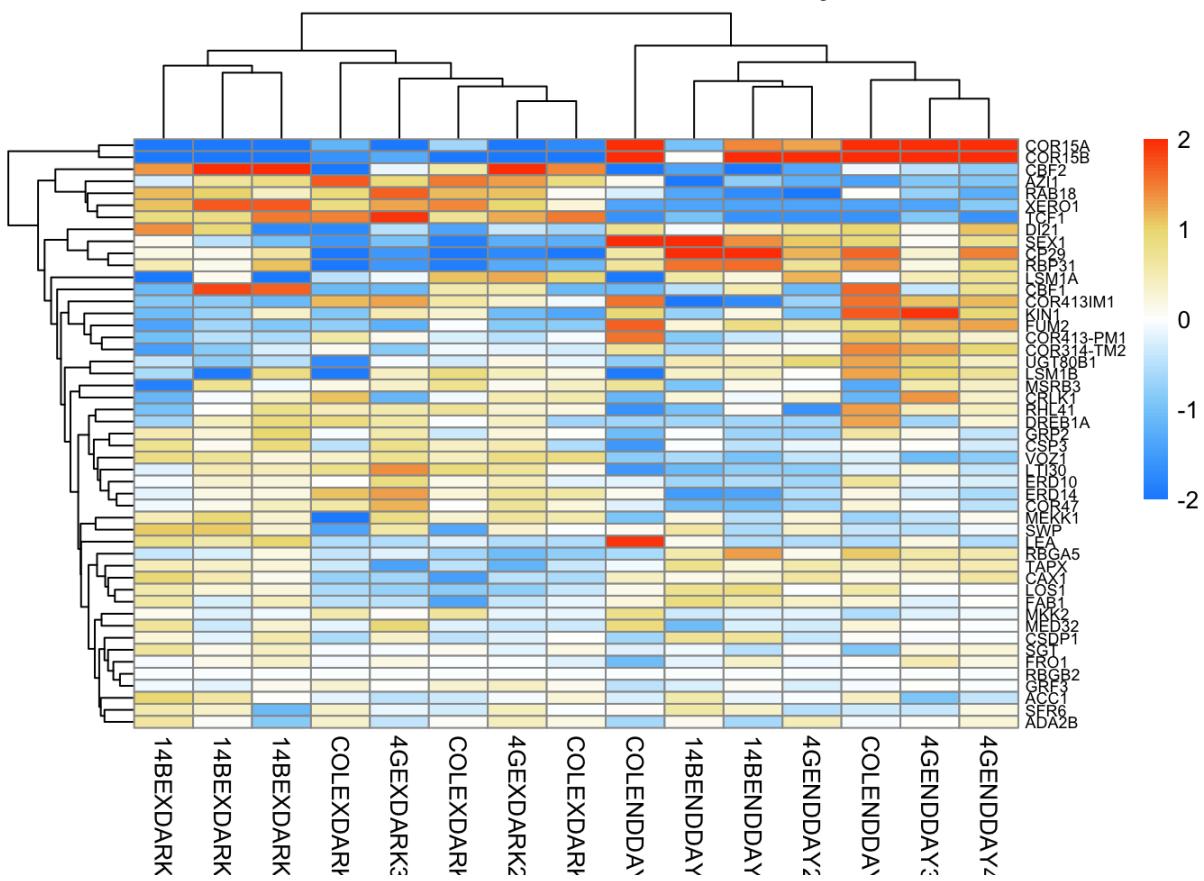
```



```
for (i in 1:length(ontology_names)){
  temp_set <- joined_set_ribo %>% filter(gene_ontology_name == ontology_names[i])
  lgGo <- lgNormRibo[temp_set$gene, ]

  heatData = lgGo - rowMeans(lgGo)
  heatData = as.data.frame(heatData)
  heatData[heatData > 2] = 2; heatData[heatData < -2] = -2
  fontsize_row = 10 - nrow(heatData) / 15
  pheatmap(
    heatData,
    color = heatPalette,
    clustering_method = "average",
    labels_row=geneNamesAndDescriptions[rownames(heatData), "symbol"],
    main = ontology_names[i],
    fontsize_row = fontsize_row,
    cellheight = 5,
    margins = c(10, 10)
  )
}
```

Arabidopsis thaliana



protein-chromophore linkage

