Zoom-in gene sets

Jamie Scheper

2024-01-26

Overview

In this log, we visualize the three final gene sets through boxplots per cluster (ER and ICU) and severity level.

We briefly discuss all genes from the eleven-gene set and their relationship to oxidative stress and mitochondria dysfunction.

Setup

We first import ER and ICU metadata and the counts dataset. Then, we extract counts for each gene set and merging with the metadata. The function `visualize_group_boxplots` colors boxplots cohort dependently via the `type` parameter. We have the option to plot the boxplots separetly (each has its own box via `face_wrap`), or we plot them together. Note: the count data is VST normalized.

```r
library(dplyr)
library(tidyverse)
library(reshape2)
library(ggvenn)
library(glue)

meta_er <- read.csv("data/clustering/er_clusters.csv")
meta_icu <- read.csv("data/clustering/icu_clusters.csv")
counts <- read.csv("data/degs/all_degs.csv")
five_genes <- c('SORD', 'UCP3', 'SLC44A1', 'POLRMT', 'ACADSB')
nine_genes <- c('P2RY1', 'OAT', 'RAB5IF', 'TSPO', 'SARM1',
                'CLYBL', 'TMEM70', 'CYP1B1', 'DEPP1')
eleven_genes <- c('SORD', 'UCP3', 'KIF28P', 'THEM4', 'SLC44A1',
                  'POLRMT', 'ANKRD37', 'UCP2', 'CYP2U1', 'DEPP1', 'ARG1')

visualize_group_boxplots <- function(df, meta, x, y, group,
                                     separate = F, gene_set, type) {
  size <- length(gene_set)

  long_format <- df %>% as.data.frame() %>%
    pivot_longer(cols = -gene, names_to = "sample_identifier",
                 values_to = "counts") %>%
    merge(., meta, by = "sample_identifier") %>%
    filter(gene %in% gene_set)
```

```r
  if (type == "icu" && group != 'sepsis_severity') {
    group_colors <- c("blue", "orange")
    names <- c("ICU-mild", "ICU-severe")
  } else if ((type == "er" && group != 'sepsis_severity')) {
    group_colors <- c("purple", "green")
    names <- c("ER-severe", "ER-mild")
  } else{
    group_colors <- c("red", "blue", "green")
    names <- c("High", "Intermediate", "Low")
  }

  legend_title <- ifelse(length(group_colors) == 3, 'Severity level', 'Cluster')

  if (separate) {
    p1 <- ggplot(long_format, aes(x = !!sym(x), y = !!sym(y))) +
      geom_boxplot(aes(fill = as.factor(!!sym(group)))) +
      facet_wrap(~ gene, scales = "free") +
      scale_fill_manual(values = group_colors, labels=names) +
      theme_gray() +
      labs(fill = legend_title) +
    ggtitle(glue("{str_to_title(size)} sized gene set based on endotype ({type} clusters)"))
  } else {
    p1 <- ggplot(long_format, aes(x = !!sym(x), y = !!sym(y))) +
    geom_boxplot(aes(fill = as.factor(!!sym(group)))) +
    scale_fill_manual(values = group_colors, labels=names) +
    theme_gray() +
    labs(fill = legend_title) +
    theme(axis.text.x = element_text(angle = 90)) +
    ggtitle(glue("{str_to_title(size)} sized gene set based on cluster ({type} clusters)"))
  }

  gridExtra::grid.arrange(p1)
}
```

five-gene set

Below we highlight the differences between gene expression patterns for the five-gene set per cluster and severity level for both ER and ICU cohorts separately.

POLRMT has the largest distribution between clusters and severity, whereas other do not exhibit that much change in expression, such as SORD This could be attributed to a tightly controlled gene expression.

```
visualize_group_boxplots(counts, meta_er, 'gene', 'counts', group='cluster',
                         gene_set=five_genes, separate=F, type="er")
```
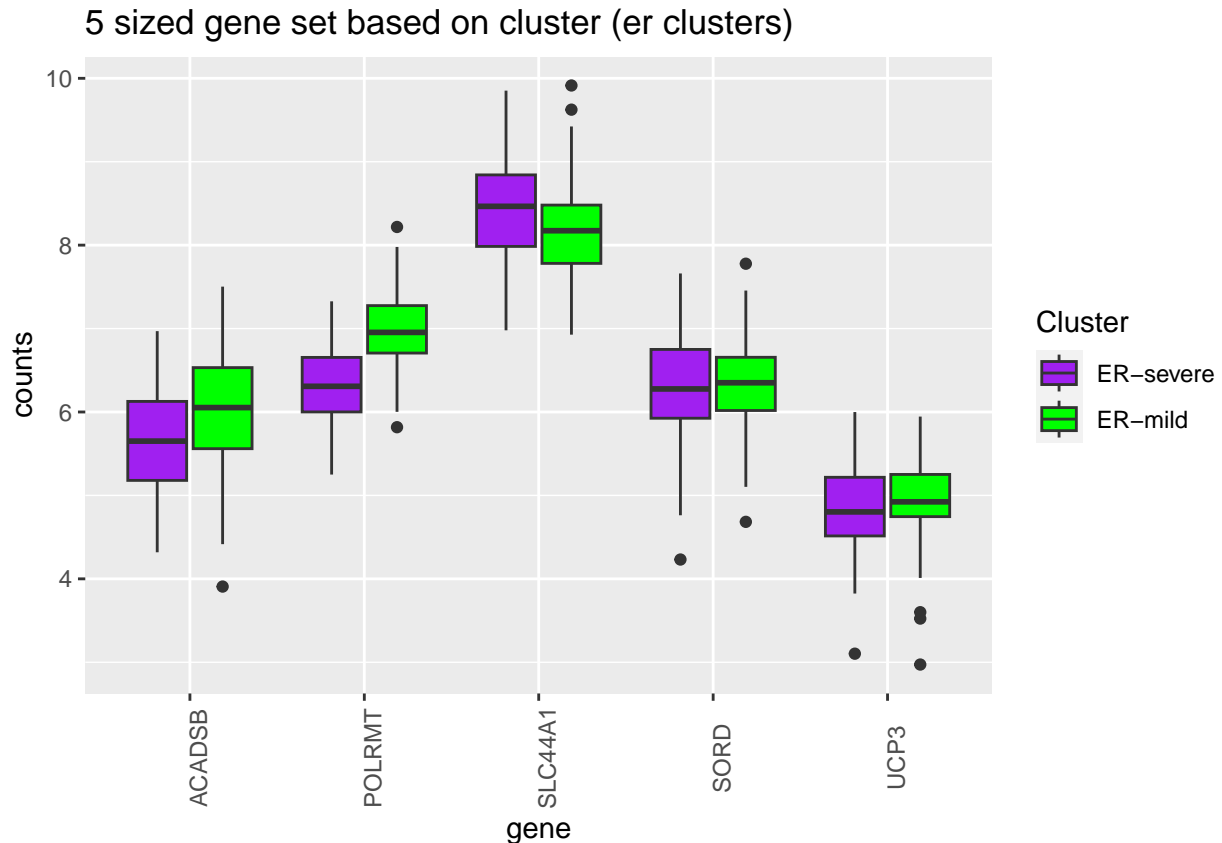

Figure 1: Gen expression of five-gene set based on ER per cluster.

```
visualize_group_boxplots(counts, meta_er, 'gene', 'counts',
                         group='sepsis_severity', gene_set=five_genes,
                         separate=F, type="er")
```

```
visualize_group_boxplots(counts, meta_icu, 'gene', 'counts', group='cluster',
                         gene_set=five_genes, separate=F, type="icu")
```

```
visualize_group_boxplots(counts, meta_icu, 'gene', 'counts',
                         group='sepsis_severity', gene_set=five_genes,
                         separate=F, type="icu")
```
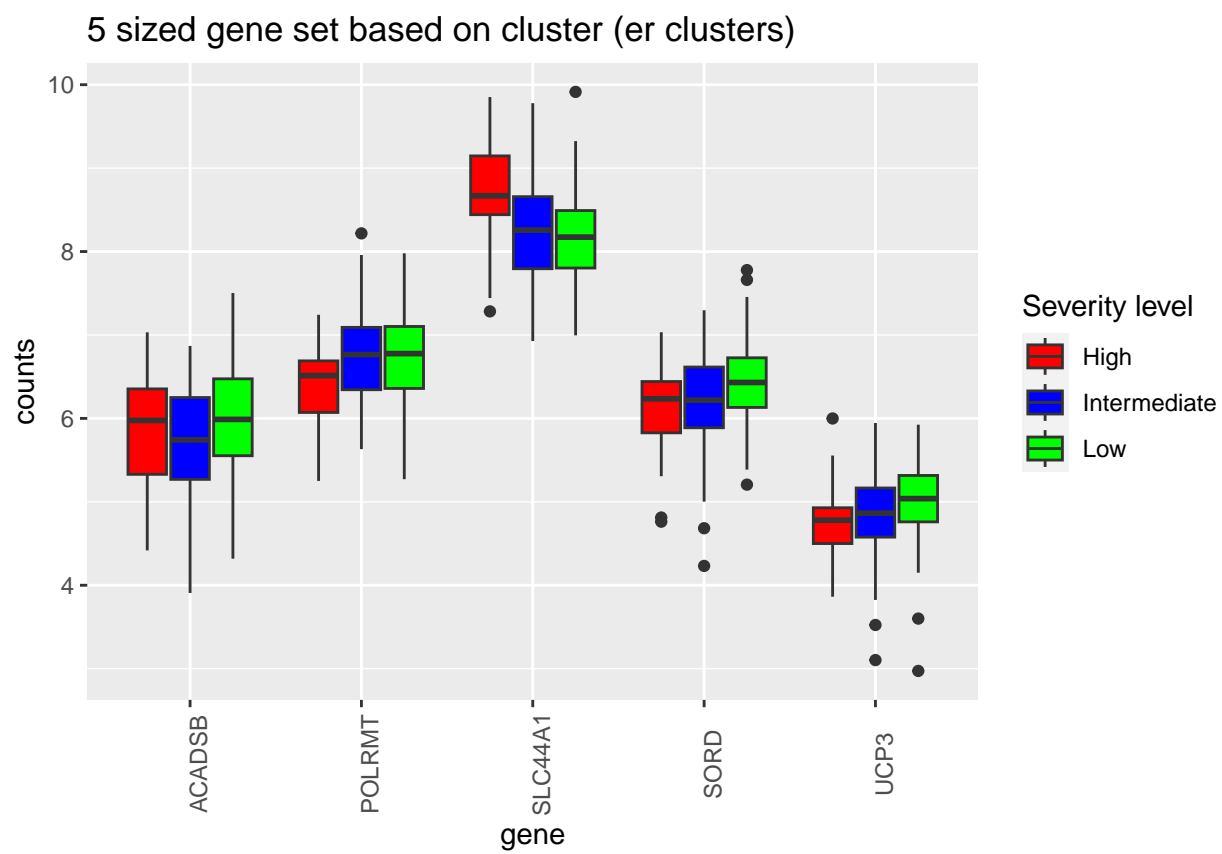
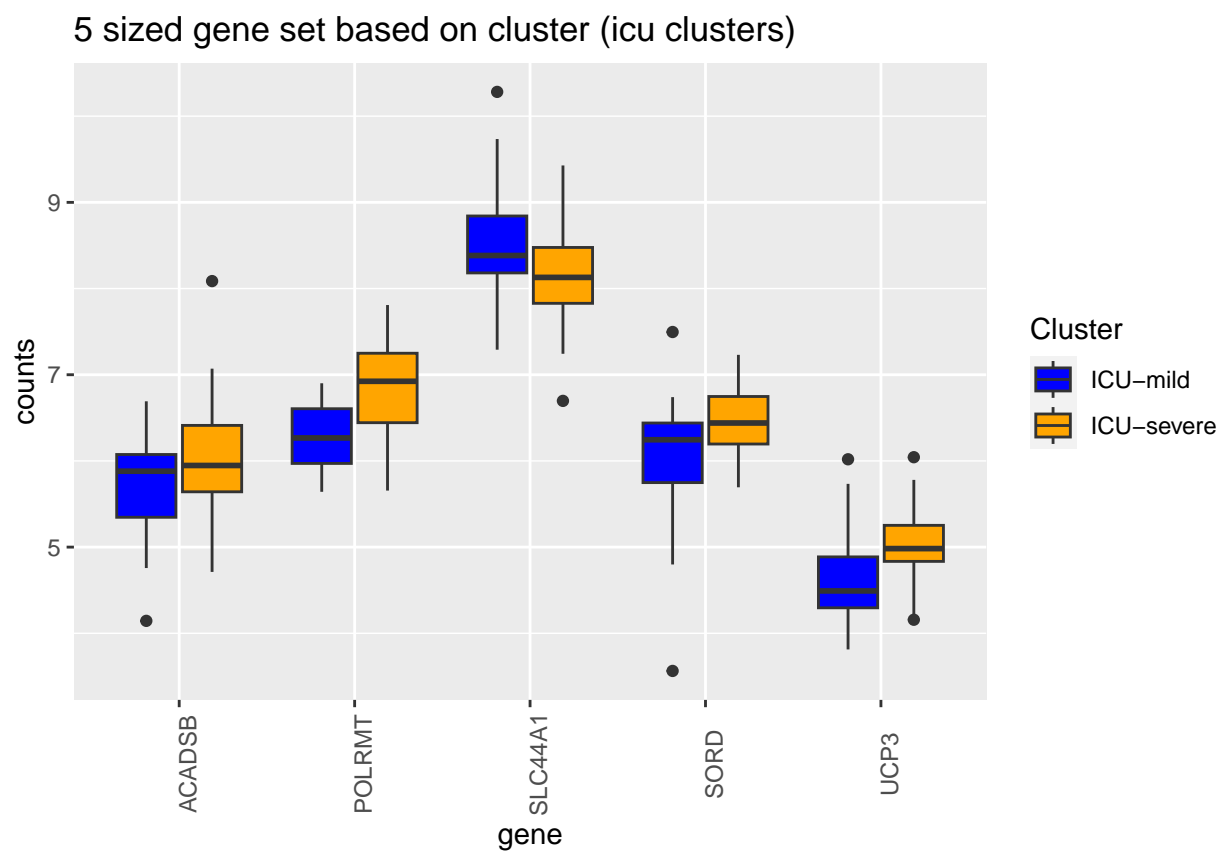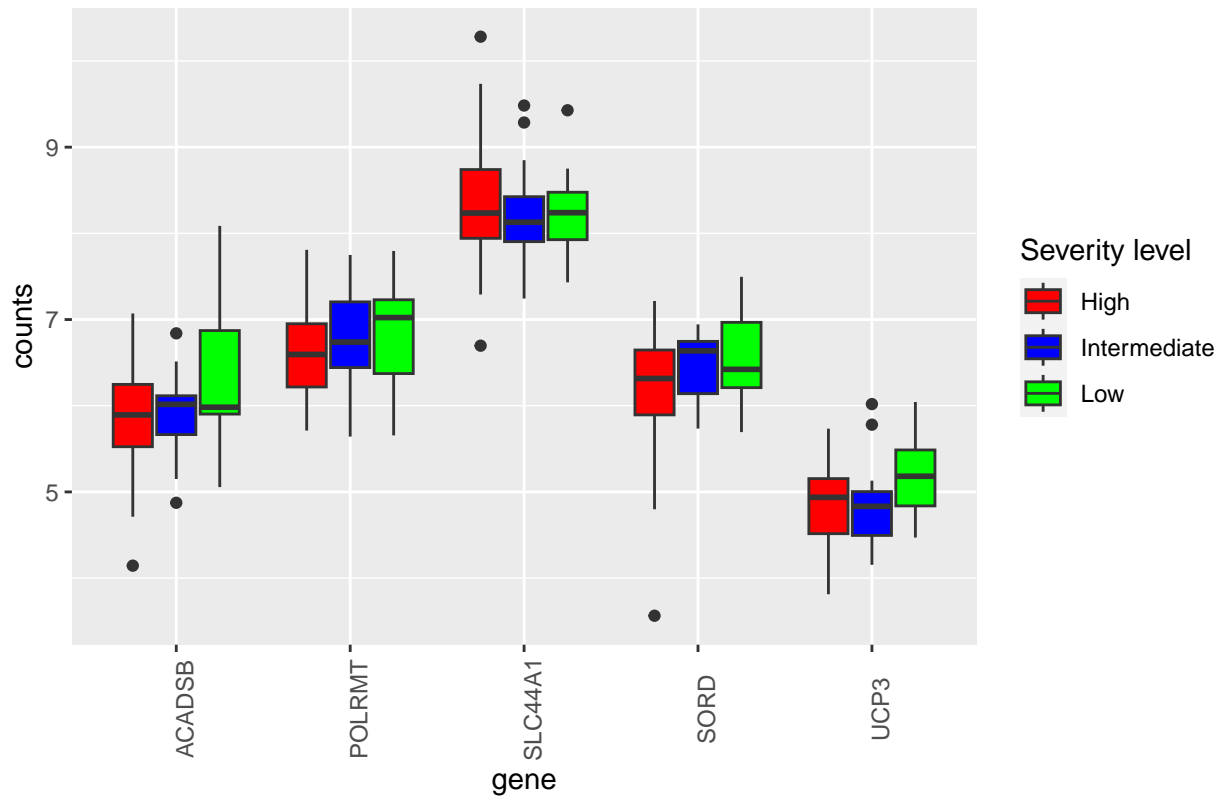Figure 2: Gen expression of five-gene set based on ER per severity level.

Figure 3: Gen expression of five-gene set based on ICU per cluster.

5 sized gene set based on cluster (icu clusters)

Nine-gene set Below we highlight the differences between gene expression patterns for the nine-gene set per cluster and severity level for both ER and ICU cohorts separately.

CYP1B1 and TSPO has the largest distribution between clusters and severity, whereas other do not exhibit that much change in expression, such as TMEM70 This could be attributed to a tightly controlled gene expression.

```
visualize_group_boxplots(counts, meta_er, 'gene', 'counts', group='cluster',
                         gene_set=nine_genes, separate=F, type="er")
```
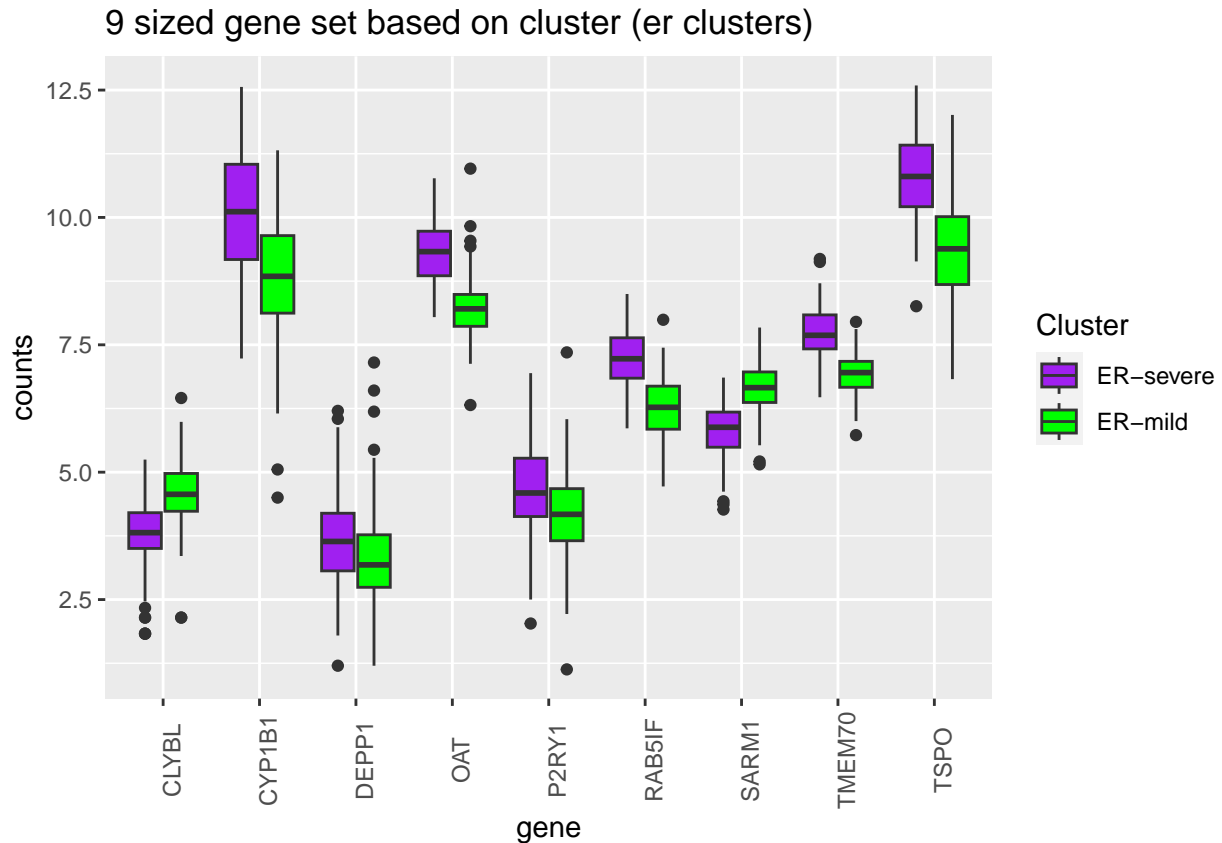

Figure 4: Gen expression of nine-gene set based on ER per cluster.

```
visualize_group_boxplots(counts, meta_er, 'gene', 'counts',
                         group='sepsis_severity', gene_set=nine_genes,
                         separate=F, type="er")
```

```
visualize_group_boxplots(counts, meta_icu, 'gene', 'counts', group='cluster',
                         gene_set=nine_genes, separate=F, type="icu")
```

```
visualize_group_boxplots(counts, meta_icu, 'gene', 'counts',
                         group='sepsis_severity', gene_set=nine_genes,
                         separate=F, type="icu")
```
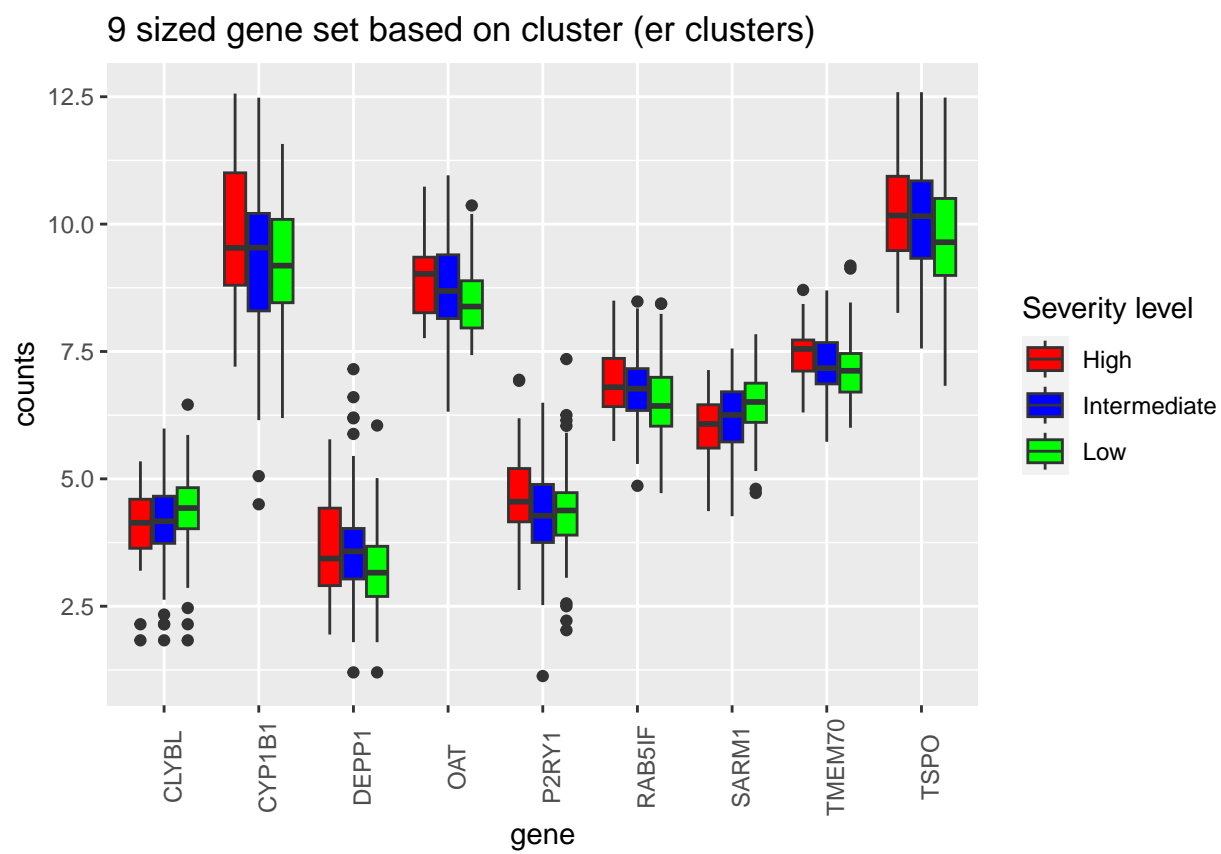
7

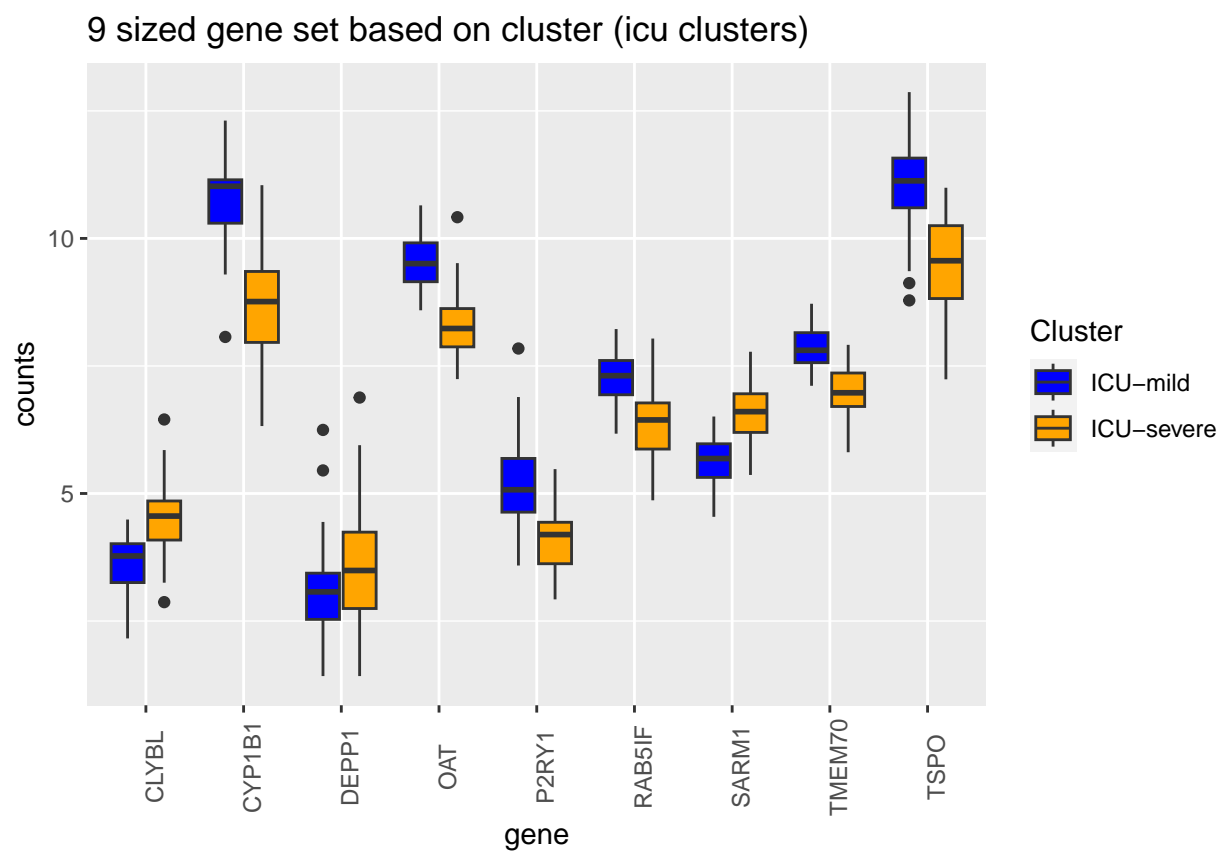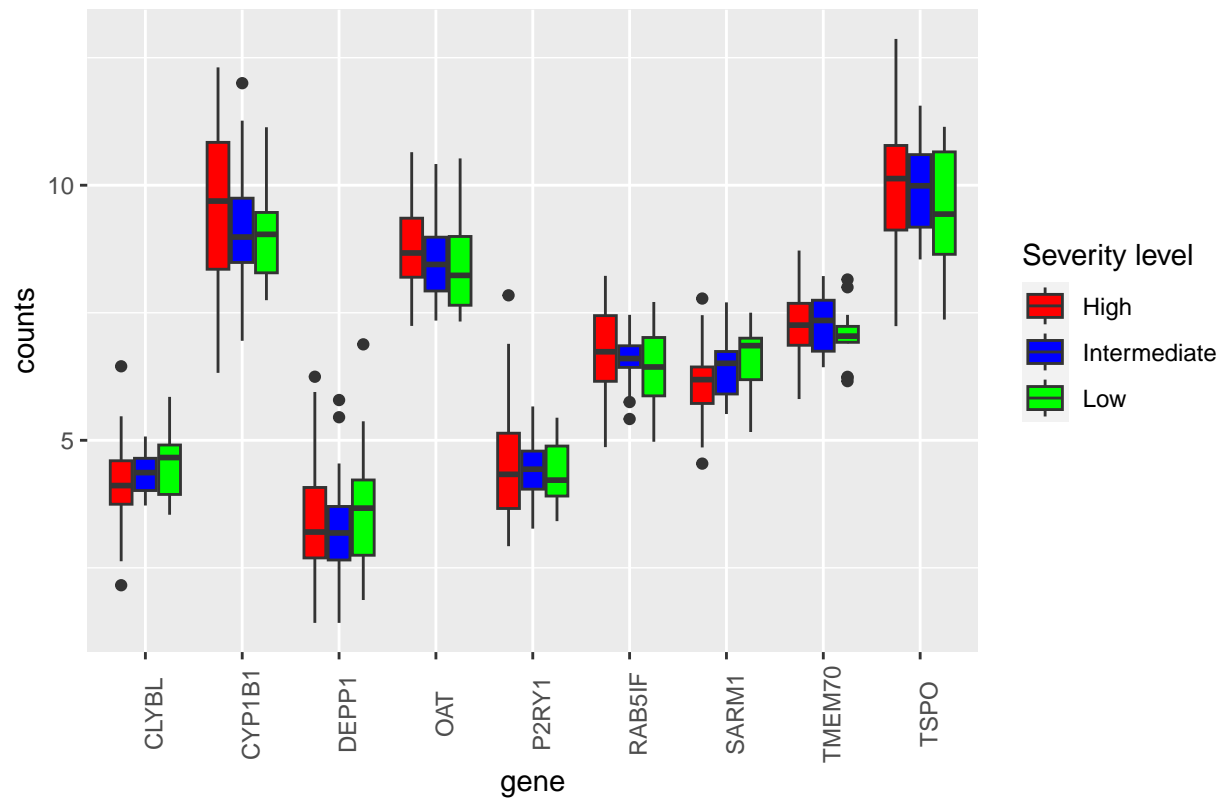Figure 5: Gen expression of nine-gene set based on ER per severity level.

Figure 6: Gen expression of nine-gene set based on ICU per cluster.

9 sized gene set based on cluster (icu clusters)

Eleven-gene set Below we highlight the differences between gene expression patterns for the eleven-gene set per cluster and severity level for both ER and ICU cohorts separately.

ARG1 has the largest distribution between clusters and severity, whereas other do not exhibit that much change in expression, such as UCP3. This could be attributed to a tightly controlled gene expression.

```
visualize_group_boxplots(counts, meta_er, 'gene', 'counts', group='cluster',
                         gene_set=eleven_genes, separate=F, type="er")
```
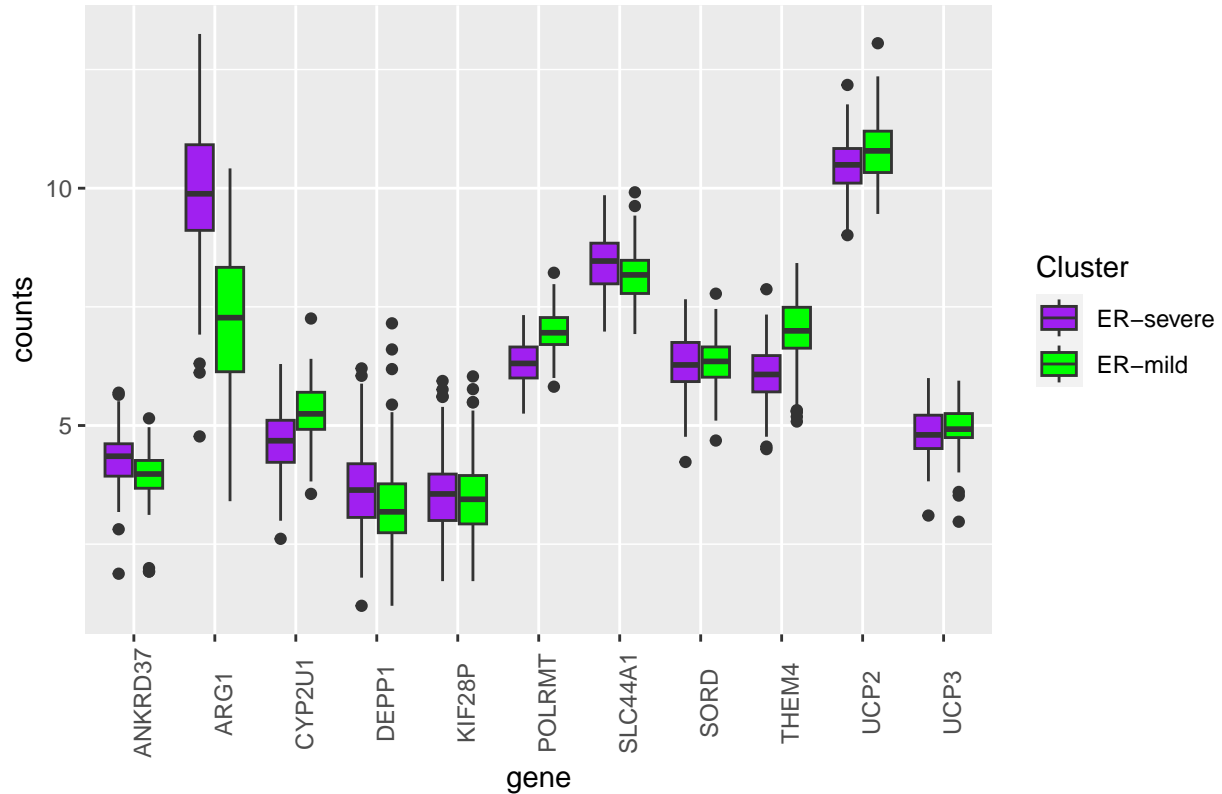

Figure 7: Gen expression of eleven-gene set based on ER per cluster.

```
visualize_group_boxplots(counts, meta_er, 'gene', 'counts',
                         group='sepsis_severity', gene_set=eleven_genes,
                         separate=F, type="er")
```

```
visualize_group_boxplots(counts, meta_icu, 'gene', 'counts', group='cluster',
                         gene_set=eleven_genes, separate=F, type="icu")
```

```
visualize_group_boxplots(counts, meta_icu, 'gene', 'counts',
                         group='sepsis_severity', gene_set=eleven_genes,
                         separate=F, type="icu")
```
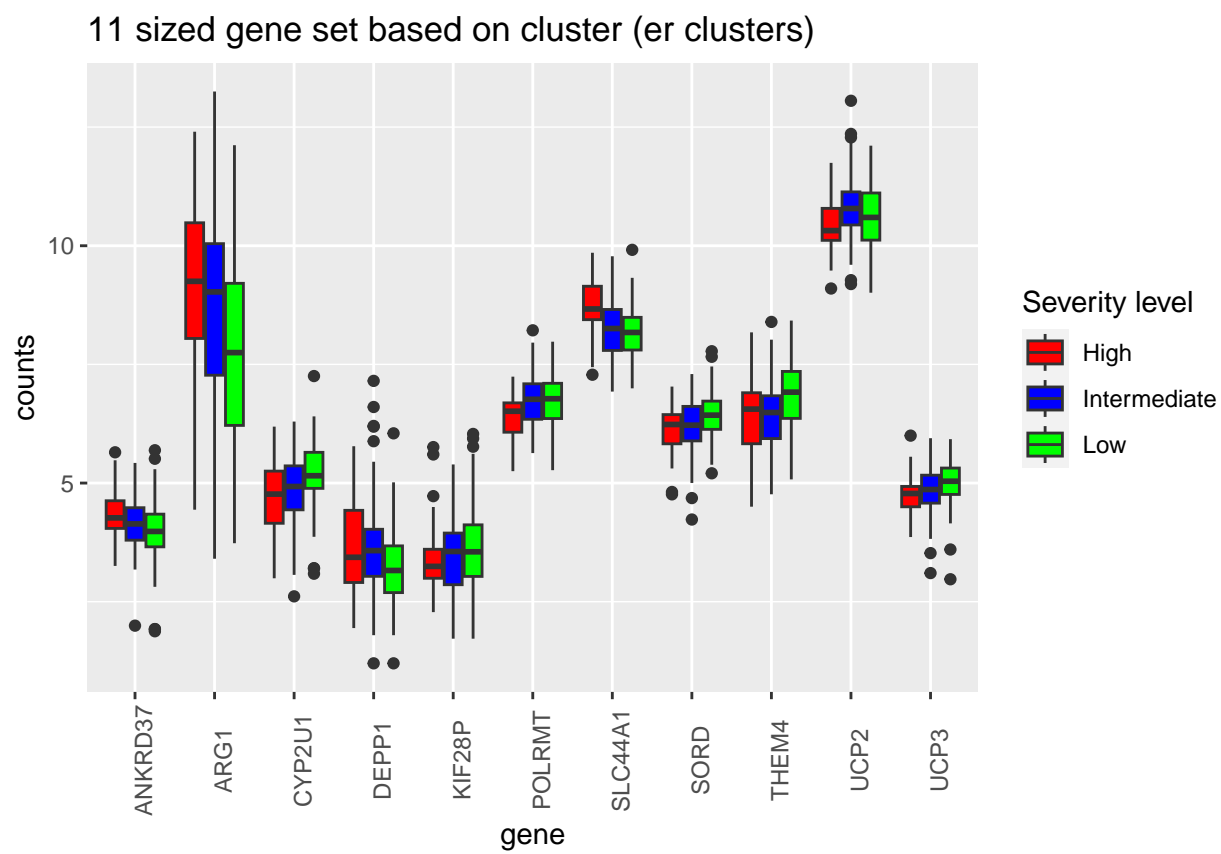
11

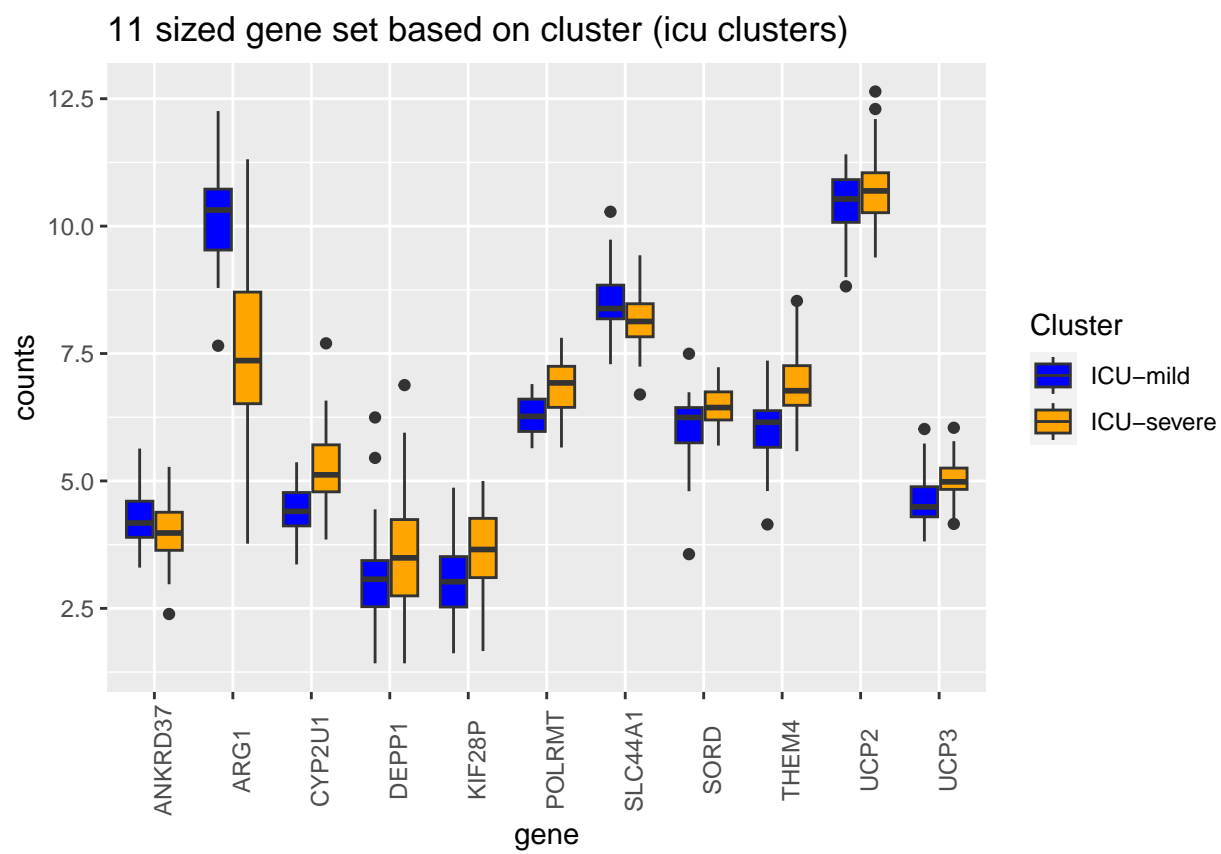Figure 8: Gen expression of eleven-gene set based on ER per severity level.

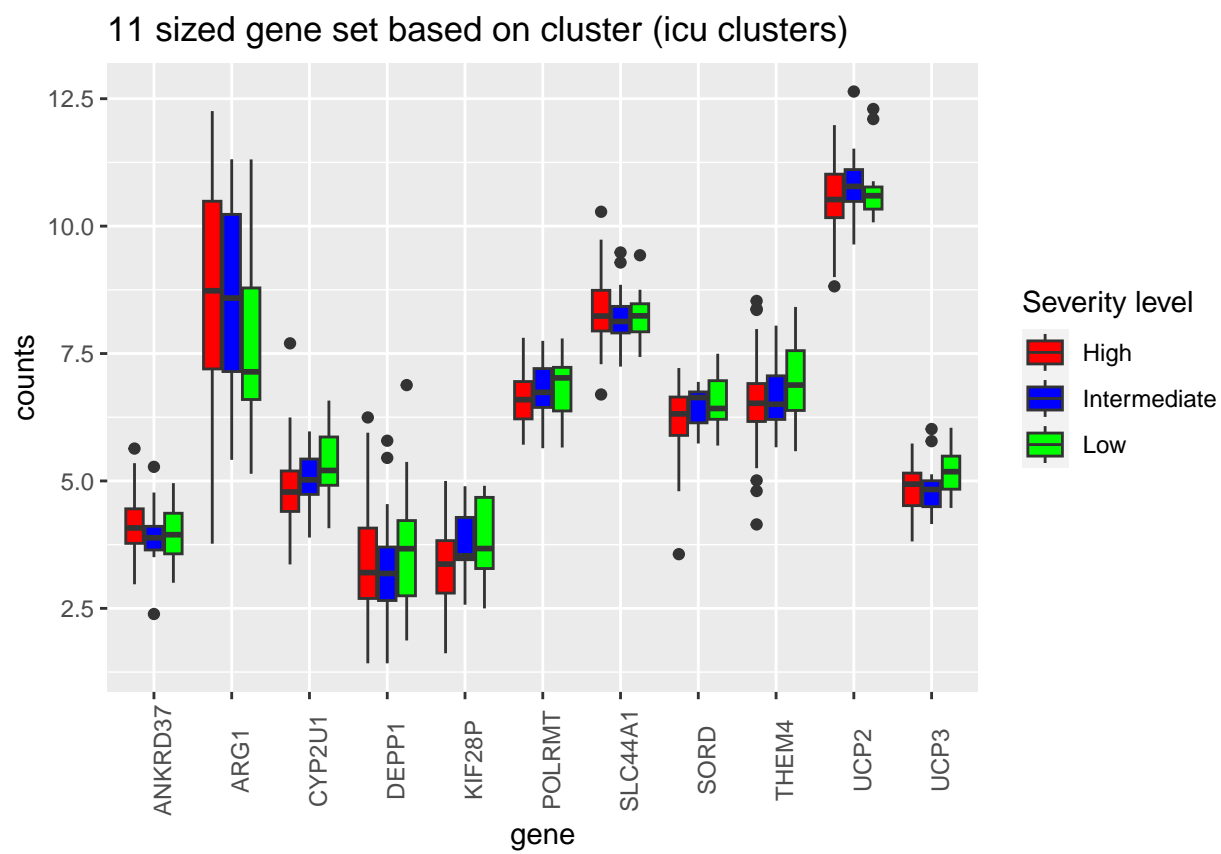Figure 9: Gen expression of eleven-gene set based on ICU per cluster.

Figure 10: Gen expression of eleven-gene set based on ICU per severity level.

Differences between sets

Differences between all gene sets is depicted below. Here, we can see that the five-gene set interely overlaps with the eleven-gene set. And there is only one shared DEG between the nine-gene and eleven-gene sets. This is probably due to the fact that the nine-gene set was extracted with mutual information, and the eleven-gene set with logistic regression with L2 regularization.

```
ggvenn(list(
        'Five' = five_genes,
        'Nine' = nine_genes,
        'Eleven' = eleven_genes)
        ) +
  ggtitle('Venn diagram of all reduced genes sets')
```
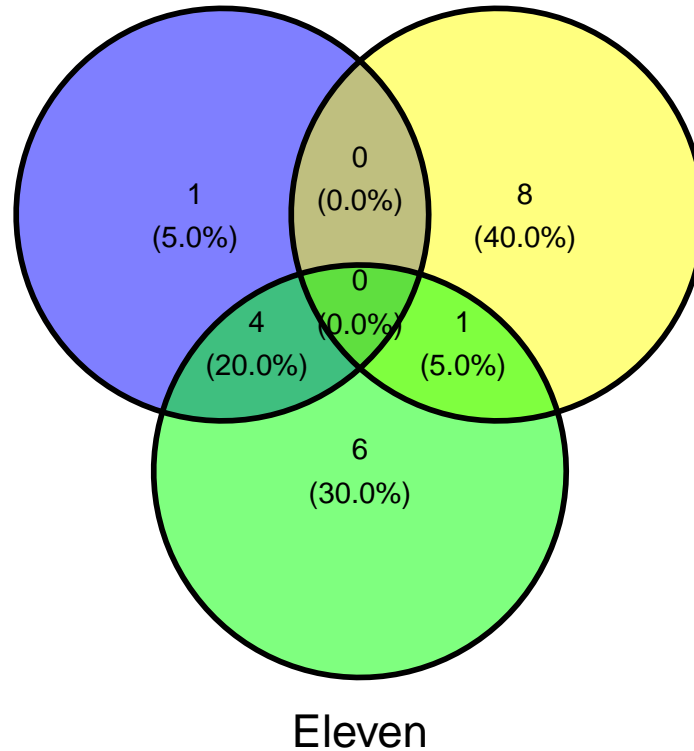

Figure 11: A venn diagram representing the differences between the genes found in all three gene sets.

Short explanation of every gene

Our best-performing gene set consisted of eleven genes, namely SORD, UCP3, KIF28P, THEM4, SLC44A1, POLRMT, ANKRD37, UCP2, CYP2U1, DEPP1, ARG1. We will explain some of them here. Some were present in multiple gene sets, underlining their importance as biomarkers related to mitochondria dysfunction. We briefly discussed the role of each gene concerning mitochondria dysfunction or oxidative stress.

ARG1 encodes the protein one of the two arginase isoenzymes and is responsible for the cleavage of L-arginine to form urea and L-ornithine. [1] Arg-1 is upregulated in many immune cells during activation, and as does the nitric oxide (NO) production, it needs L-arginine. When Arg-1 is upregulated, less L-arginine is available for NO production. [55] Reduced amounts of NO can impair mitochondria function as its role regulates mitochondria against oxidative stress through its influence on ROS production. [2] [3]

SORD is involved in the polyol pathway and encodes for sorbitol dehydrogenase, which converts sorbitol into fructose. [4] Decreased levels of SORD activity have been linked to the subsequent increased levels of sorbitol to oxidative stress and less available NADPH, disturbing the balance between free NADP+ and NADPH. [5]

PPTC7 showed up in multiple gene sets and is a mitochondrial phosphatase that prevents ROS from accumulating by regulating antioxidant Q10 through its coenzyme CoQ10. [6] In addition, the gene regulates the process of selective degradation of mitochondria by autophagy, as seen in mice. [7]

KIF28P, also known as KLP6 (Uniprot ID: B7ZC32), regulates the transport of mitochondria through mitochondrial fission and fusion. [8] It can interact with KIFIBα with KIF1-binding protein (KBP) to facilitate the localization of mitochondria in neural cells. [9] Under oxidative stress, kinesins involved in transport, especially those regulated by phosphorylation, may move damaged mitochondria in more significant numbers for reparation or degradation through mitophagy cells in axonal transportation. [9] [10] However, these findings have not been specifically linked to KIF28P/KLP6 but have been put into a broader picture regarding kinesins.

TXNIP plays a role in redox homeostasis, and its primary function is to increase the production of ROS, which causes oxidative stress and, ultimately, apoptosis by binding to inflammasome NLRP3. [11] [12] The gene has been a target for personalized medicine in several studies regarding chronic kidney disease, cancer, and dysfunctions affected by the onset of sepsis, underlining its importance in our gene set. [12] [13]

THEM4, also known as CTMP, regulates apoptosis by reducing AKT/PKB signaling, an upregulated pathway as a protective response against oxidative stress. [14] THEM4 is upregulated in the event of mitochondrial dysfunction caused by oxidative stress. In addition, THEM4 is a known target for therapeutics in several types of cancers. [15]

DEPP-1, also known as DEPP and C10orf10, contributes to increased ROS levels and regulates autophagy in conjunction with FOXO1 or FOXO3, a mechanism to reduce damage from oxidative stress. [16] [17] Therefore, an increased expression of DEPP was expected.

CASP8 encodes for caspase-8, which plays a role as an initiator in cell apoptosis. It recruits any caspase executioner (either caspase-3, -6, or -7) via cleavage, and after that, the process of cell death starts. [18] Therefore, higher expression patterns of CASP8 are in line with mitochondria dysfunction.

UCP3, a gene encoding a mitochondrial-specific protein, is a selective counter against certain types of excessive ROS production and is expressed higher in the case of oxidative stress. [19] One study showed that overexpression of UCP3 in the presence of 4-HNE leads to a reduced increase in ROS production in older mice. [20] This highlights its importance as a potential biomarker for mitochondria dysfunction

References

[1] M. Munder, "Arginase: An emerging key player in the mammalian immune system," British Journal of Pharmacology, vol. 158, no. 3, pp. 638–651, Oct. 2009, doi: 10.1111/j.1476-5381.2009.00291.x.

[2] C. J. Darcy et al., "Increased plasma arginase activity in human sepsis: Association with increased circulating neutrophils," Clinical Chemistry and Laboratory Medicine, vol. 52, no. 4, Jan. 2014, doi: 10.1515/cclm-2013-0698.

[3] K. Wijnands, T. Castermans, M. Hommen, D. Meesters, and M. Poeze, "Arginine and citrulline and the immune response in sepsis," Nutrients, vol. 7, no. 3, pp. 1426–1463, Feb. 2015, doi: 10.3390/nu7031426.

[4] A. Y. W. LEE and S. S. M. CHUNG, "Contributions of polyol pathway to oxidative stress in diabetic cataract," The FASEB Journal, vol. 13, no. 1, pp. 23–30, Jan. 1999, doi: 10.1096/fasebj.13.1.23.

[5] P. A. Barnett, R. G. González, L. T. Chylack, and H.-M. Cheng, "The effect of oxidation on sorbitol pathway kinetics," Diabetes, vol. 35, no. 4, pp. 426–432, Apr. 1986, doi: 10.2337/diab.35.4.426.

[6] I. González-Mariscal et al., "The mitochondrial phosphatase PPTC7 orchestrates mitochondrial metabolism regulating coenzyme Q10 biosynthesis," Biochimica et Biophysica Acta (BBA) - Bioenergetics, vol. 1859, no. 11, pp. 1235–1248, Nov. 2018, doi: 10.1016/j.bbabio.2018.09.369.

[7] N. M. Niemi et al., "PPTC7 maintains mitochondrial protein content by suppressing receptor-mediated mitophagy," Nature Communications, vol. 14, no. 1, Oct. 2023, doi: 10.1038/s41467-023-42069-w.

[8] K. Tanaka, Y. Sugiura, R. Ichishita, K. Mihara, and T. Oka, "KLP6: A newly identified kinesin that regulates the morphology and transport of mitochondria in neuronal cells," Journal of Cell Science, vol. 124, no. 14, pp. 2457–2465, Jul. 2011, doi: 10.1242/jcs.086470.

[9] N. Hirokawa, Y. Noda, Y. Tanaka, and S. Niwa, "Kinesin superfamily motor proteins and intracellular transport," Nature Reviews Molecular Cell Biology, vol. 10, no. 10, pp. 682–696, Oct. 2009, doi: 10.1038/nrm2774.

[10] C. Fang, D. Bourdette, and G. Banker, "Oxidative stress inhibits axonal transport: Implications for neurodegenerative diseases," Molecular Neurodegeneration, vol. 7, no. 1, Jun. 2012, doi: 10.1186/1750-1326-7-29.

[11] R. Zhou, A. Tardivel, B. Thorens, I. Choi, and J. Tschopp, "Thioredoxin-interacting protein links oxidative stress to inflammasome activation," Nature Immunology, vol. 11, no. 2, pp. 136–140, Dec. 2009, doi: 10.1038/ni.1831.

[12] C. Yang, W. Xia, X. Liu, J. Lin, and A. Wu, "Role of TXNIP/NLRP3 in sepsis-induced myocardial dysfunction," International Journal of Molecular Medicine, Jun. 2019, doi: 10.3892/ijmm.2019.4232.

[13] J. Y. Lim, "Thioredoxin and thioredoxin-interacting protein as prognostic markers for gastric cancer recurrence," World Journal of Gastroenterology, vol. 18, no. 39, p. 5581, 2012, doi: 10.3748/wjg.v18.i39.5581.

[14] A. Parcellier et al., "Carboxy-terminal modulator protein (CTMP) is a mitochondrial protein that sensitizes cells to apoptosis," Cellular Signalling, vol. 21, no. 4, pp. 639–650, Apr. 2009, doi: 10.1016/j.cellsig.2009.01.016.

[15] M. M. Hill and B. A. Hemmings, "Inhibition of protein kinase akt," Pharmacology Therapeutics, vol. 93, no. 2–3, pp. 243–251, Feb. 2002, doi: 10.1016/s0163-7258(02)00193-6.

[16] S. Salcher, M. Hermann, U. Kiechl-Kohlendorfer, M. J. Ausserlechner, and P. Obexer, "C10ORF10/DEPP-mediated ROS accumulation is a critical modulator of FOXO3-induced autophagy," Molecular Cancer, vol. 16, no. 1, May 2017, doi: 10.1186/s12943-017-0661-4.

[17] M. W. Stepp, R. J. Folz, J. Yu, and I. N. Zelko, "The c10orf10 gene product is a new link between oxidative stress and autophagy," Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, vol. 1843, no. 6, pp. 1076–1088, Jun. 2014, doi: 10.1016/j.bbamcr.2014.02.003.

[18] D. R. McIlwain, T. Berger, and T. W. Mak, "Caspase functions in cell death and disease," Cold Spring Harbor Perspectives in Biology, vol. 7, no. 4, p. a026716, Apr. 2015, doi: 10.1101/cshperspect.a026716.

[19] E. Barreiro et al., "UCP3 overexpression neutralizes oxidative stress rather than nitrosative stress in mouse myotubes," FEBS Letters, vol. 583, no. 2, pp. 350–356, Dec. 2008, doi: 10.1016/j.febslet.2008.12.023.

[20] M. Nabben et al., "The effect of UCP3 overexpression on mitochondrial ROS production in skeletal muscle of young versus aged mice," FEBS Letters, vol. 582, no. 30, pp. 4147–4152, Nov. 2008, doi: 10.1016/j.febslet.2008.11.016.