Alternative-based splicing analysis of *Apis mellifera* queens and drones - Part 3: Overlap

Joe Colgan

Contents

Introduction:

The purpose of this script is to examine overlap in genes that are differentially expressed between queens and drones in the gonads and the brains.

The script loads RDS objects generated from previous scripts:
- deseq_object_results_sig_gonad.rds
- deseq_object_results_sig_brain.rds

These files were generated from a DESeq2-based analysis using gene-level counts directly generated by STAR.

In addition, we load output from our maser-based analysis:
- splice_events_gonad_jc.rds
- splice_events_brain_jc.rds

The scripts examines overlap between differentially expressed genes (both in amplitude and splicing) in both tissues and generates a Euler plot to visualise genes that differ in expression between queens and drones.

1. Load libraries:

```
library(tidyverse)
library(DESeq2)
library(eulerr)
```

2. Load datasets:

We load our outputs for each tissue from DESeq2, which contain genes (rownames), baseMean (mean expression of a gene across all samples), log2FoldChange (expression difference between nurses and foragers), as well as adjusted p value (padj).

```
deseq_object_results_sig_brain <-  readRDS(
        file = "input/deseq_object_results_sig_brain.rds")
```

```
deseq_object_results_sig_gonad <-  readRDS(
        file = "input/deseq_object_results_sig_gonad.rds")
```

We next load the output files from maser and store each as an object:

```
maser_object_results_sig_brain <-  readRDS(
        file = "results/splice_events_brain_jc.rds")

maser_object_results_sig_gonad <-  readRDS(
        file = "results/splice_events_gonad_jc.rds")
```

3. Extract significantly differentially expressed genes:

As mentioned above, the row names of our DESeq2 output contain the gene name, which we can extract using the function 'rownames()' and store as an object for each individual tissue.

```
brain_degs <- rownames(deseq_object_results_sig_brain)
gonad_degs <- rownames(deseq_object_results_sig_gonad)
```

We can extract gene names from maser using information stores in the 'GeneID' column. However, as we can have many entries for the same gene in our dataset (as one gene may have more than one splicing event), we run the base R function 'unique()', which returns a unique (non-redundant) list of IDs for genes that are alternatively spliced.

```
brain_asgs <- unique(maser_object_results_sig_brain$GeneID)
gonad_asgs <- unique(maser_object_results_sig_gonad$GeneID)
```

4. Combine input genes lists to examine overlap:

Next, we use the 'euler()' function from the eulerr R package, which creates a list using the DEGs found in each tissue.

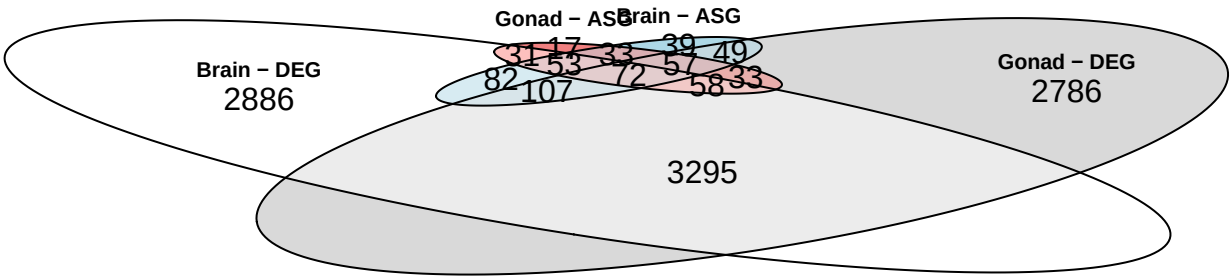
```
euler_lists <- euler(combinations = list("Brain - DEG" = brain_degs,
                                         "Gonad - DEG" = gonad_degs,
                                         "Brain - ASG" = brain_asgs,
                                         "Gonad - ASG" = gonad_asgs),
                     shape = "ellipse")
```

5. Generate plot:

We can visualise the overlap between differentially expressed genes using the 'plot()' function.

```
euler_all_plot <- plot(euler_lists,
                       quantities = TRUE,
                       edges = TRUE,
                       labels = list(fontsize = 8))

## Plot to console:
euler_all_plot
```



The last thing to do is save the image to file:

```
ggsave("results/euler_all_plot.png",
       plot = euler_all_plot,
       height = 5,
       width = 5)
```

6. Tissue-based comparison:

Lastly, we can examine the overlap of differentially expressed both in terms of amplitude and splicing across tissues.

First, for the **brain**, we can again use Eulerr to examine overlap.

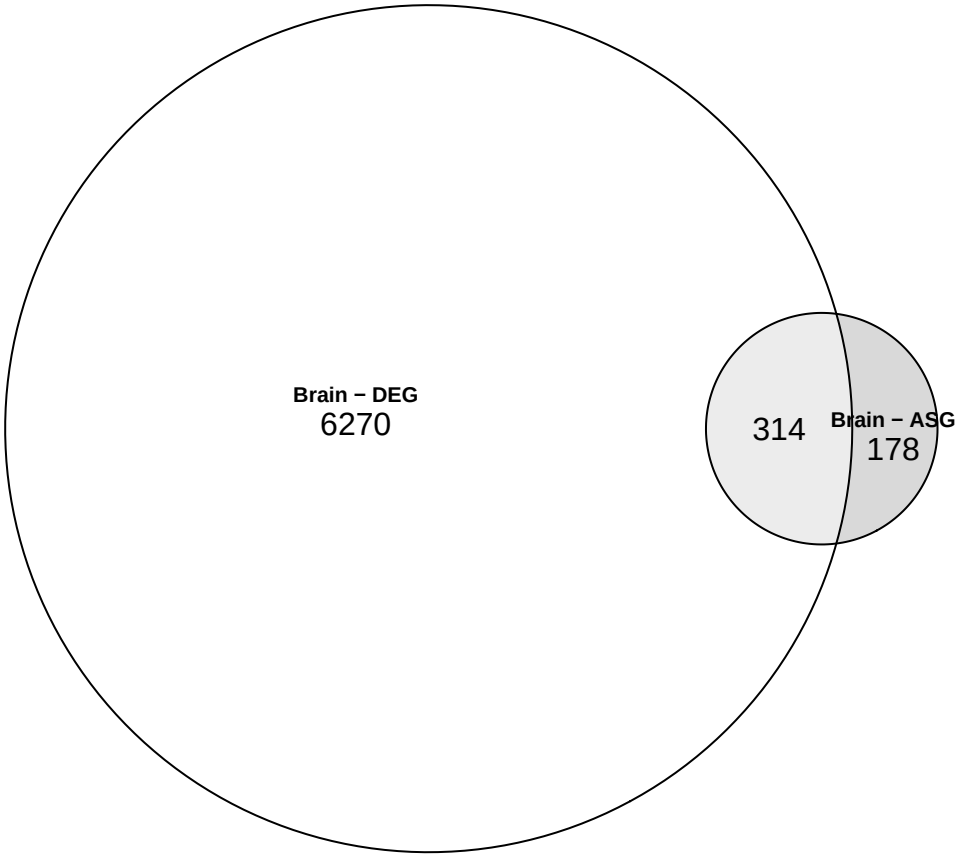
```
euler_brain_lists <- euler(combinations = list("Brain - DEG" = brain_degs,
                                               "Brain - ASG" = brain_asgs))
```

Visualise overlap for genes with differential expression in the brain:

```
euler_brain_plot <- plot(euler_brain_lists,
                         quantities = TRUE,
                         edges = TRUE,
                         labels = list(fontsize = 8))
```

You can print this to console by just typing and running the following command.

```
euler_brain_plot
```



Lastly, we can save the plot to file.

```
ggsave("results/euler_brain_plot.png",
       plot = euler_brain_plot,
       height = 5,
       width = 5)
```

We can also the same for the data genrated from the **gonads**.

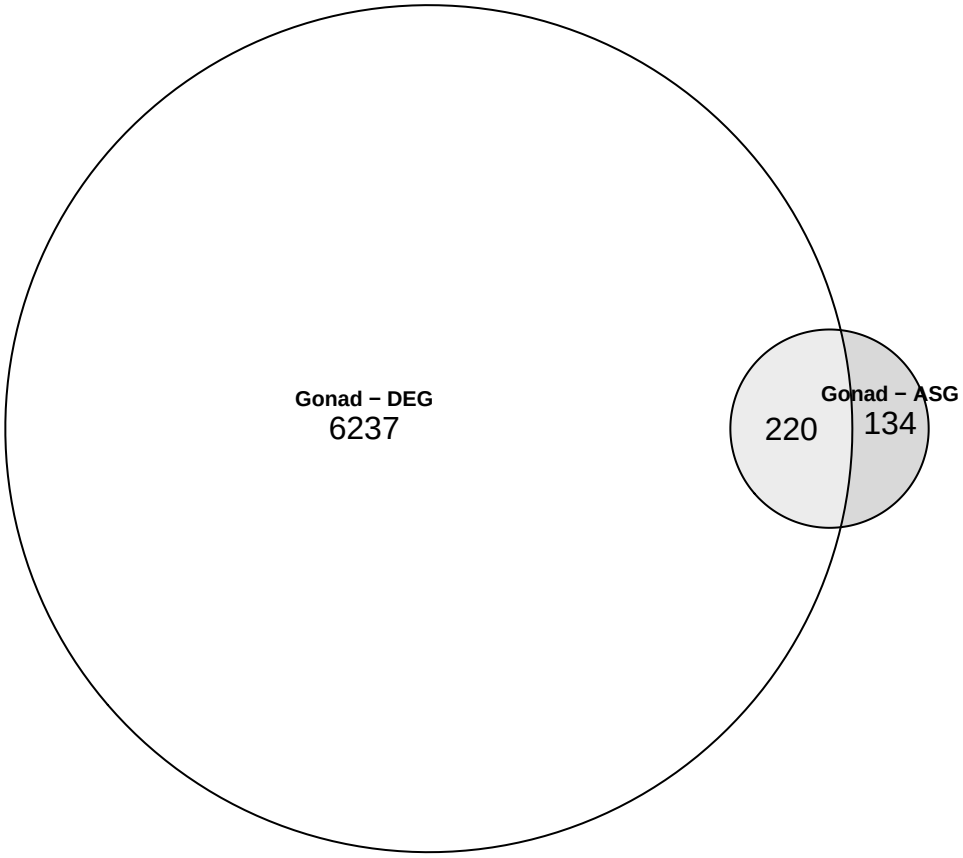
```
euler_gonad_lists <- euler(combinations = list("Gonad - DEG" = gonad_degs,
                                               "Gonad - ASG" = gonad_asgs))
```

Visualise overlap for genes with differential expression in the gonads:

```
euler_gonad_plot <- plot(euler_gonad_lists,
                         quantities = TRUE,
                         edges = TRUE,
                         labels = list(fontsize = 8))
```

Similar, we can print the plot to the console:

```
euler_gonad_plot
```



Lastly, we can save the plot to file:

```
ggsave("results/euler_gonad_plot.png",
       plot = euler_gonad_plot,
       height = 5,
       width = 5)
```