

# main

August 20, 2021

## 1 Visualize results, local splicing

### 1.1 Load Libraries

```
[1]: library(tidyverse)
library(ggplot2)
library(DT)
library(leafcutter)
library(reshape2)
library(gridExtra)
library(intervals) # needed for pretty strand arrow placement
library(foreach)
library(grid)
library(gtable)
library(ggrepel)
```

```
Attaching packages
1.3.1 tidyverse
```

ggplot2	3.3.5	purrr	0.3.4
tibble	3.1.2	dplyr	1.0.7
tidyr	1.1.3	stringr	1.4.0
readr	1.4.0	forcats	0.5.1

#### Conflicts

```
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag() masks stats::lag()
```

Loading required package: Rcpp

Attaching package: 'reshape2'

The following object is masked from 'package:tidyr':

smiths

Attaching package: 'gridExtra'

The following object is masked from 'package:dplyr':

combine

Attaching package: 'intervals'

The following object is masked from 'package:purrr':

reduce

The following object is masked from 'package:tidyr':

expand

Attaching package: 'foreach'

The following objects are masked from 'package:purrr':

accumulate, when

## 1.2 Summary of results

```
[2]: lname = load('../_m/leafviz.RData')  
      lname
```

1. 'introns' 2. 'clusters' 3. 'counts' 4. 'meta' 5. 'exons\_table' 6. 'pca' 7. 'intron\_summary' 8. 'cluster\_summary' 9. 'introns\_to\_plot' 10. 'cluster\_ids' 11. 'sample\_table' 12. 'annotation\_code' 13. 'code'

```
[3]: sample_table
```

A data.frame: 2 × 2	group	count
	<chr>	<int>
	AA	133
	EA	109

```
[4]: cluster_summary
```

A data.frame: 3 × 2	Results	n
	<chr>	<int>
	Number of differentially spliced clusters at FDR = 0.05	1332
	Fully annotated	454
	Contain unannotated junctions	878

```
[5]: intron_summary
```

A data.frame: 5 × 2	Results	n
	<chr>	<int>
	Number of fully annotated junctions	4149
	Number of junctions with cryptic 5' splice site	821
	Number of junctions with cryptic 3' splice site	941
	Number of junctions with two cryptic splice sites	382
	Number of novel junctions that connect two annotated splice sites	538

```
[6]: clusters['gene'] <- gsub("</i>", "", gsub("<i>", "", clusters$gene))
head(clusters)
```

A data.frame: 6 × 6		clusterID	N	coord	gene	annotation	FL
		<chr>	<dbl>	<chr>	<chr>	<chr>	<chr>
	880	clu_64405_+	16	chr22:45323427-45330603	FAM118A	cryptic	8.0
	272	clu_111929_-	11	chr12:124911899-124913724	UBC	cryptic	1.3
	273	clu_111930_-	26	chr12:124911952-124913549	UBC	cryptic	5.0
	1233	clu_14450_-	11	chr8:101719648-101915809	NCALD	cryptic	5.0
	312	clu_92553_?	14	chr12:124911899-124913725	UBC	cryptic	4.0
	555	clu_146837_+	20	chr17:55774859-55965579	PCTP	cryptic	1.3

```
[7]: write.table(clusters, file="cluster_ds_results_annotated.txt", sep="\t",
quote=FALSE, row.names=FALSE)
```

## 1.3 Generate plots

### 1.3.1 Define functions

```
[8]: filter_intron_table <- function(introns, clu){
  d <- introns %>% filter(clusterID == clu) %>%
    select(chr, start, end, verdict, deltapsi) %>%
    arrange(desc(abs(deltapsi))) %>%
    rename("ΔPSI" = deltapsi)
  row.names(d) <- letters[1:nrow(d)] # letters is just a:z
  return(d)
```

```

}

getGeneLength <- function(gene_name, exons_table){
  exons      <- exons_table[ exons_table$gene_name == gene_name, ]
  geneStart  <- min(exons$start)
  geneEnd    <- max(exons$end)
  geneLength <- geneEnd - geneStart
  if( geneLength > 1e6){
    pixels <- 5000
  } else if ( geneLength > 5e5 & geneLength < 1e6){
    pixels <- 3000
  } else if ( geneLength > 1.5e5 & geneLength <= 5e5){
    pixels <- 2000
  } else {
    stopifnot(geneLength <= 1.5e5)
    pixels <- "auto"
  }
  return(pixels)
}

select_data <- function(sel, clusters, exons_table){
  gene <- clusters[ sel, ]$gene
  width <- getGeneLength(gene, exons_table)
  clusterID <- clusters[ sel, ]$clusterID
  coord <- clusters[ sel, ]$coord
  return(list(gene = gene, width = width, cluster = clusterID, coord = coord))
}

```

### 1.3.2 Plotting functions

```

[9]: plot_cluster <- function(num, clusters, dir='./'){
  mydata = select_data(num, clusters, exons_table)
  while(mydata$gene == '.'){
    num = num+1
    mydata = select_data(num, clusters, exons_table)
  }
  gene_name = mydata$gene
  plotTitle <- paste0(gene_name, '_', mydata$cluster, '_top_', num, '.pdf')
  pdf(file=paste0(dir, plotTitle), width = 10, height = 5)
  print(make_cluster_plot(mydata$cluster,
    main_title = plotTitle,
    meta = meta,
    cluster_ids = cluster_ids,
    exons_table = exons_table,
    counts = counts,

```

```

                                introns = introns))

dev.off()

if (is.numeric(mydata$width)) {
  new_width = mydata$width / 100
} else {
  new_width = mydata$width
}

pdf(file=paste0(dir, gene_name, '_allClusters_top_',num,'.pdf'),
    width=new_width, height=6)
print(make_gene_plot(mydata$gene,
                     counts = counts,
                     introns = introns,
                     exons_table = exons_table,
                     cluster_list = clusters,
                     clusterID = mydata$clusterID,
                     introns_to_plot = introns_to_plot, debug=F))

dev.off()
}

```

### 1.3.3 Plot splicing

```

[10]: dir.create("top10")
      for(num in 1:10){
        plot_cluster(num, clusters, "top10/")
      }

```

Warning message:

"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = \"none\")` instead."

TableGrob (2 x 1) "arrange": 2 grobs

	z	cells	name	grob
1	1	(1-1,1-1)	arrange	gtable[layout]
2	2	(2-2,1-1)	arrange	gtable[layout]

Warning message:

"`mode(width)` differs between new and previous  
==> NOT changing 'width'"

Warning message:

"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = \"none\")` instead."

TableGrob (2 x 1) "arrange": 2 grobs

	z	cells	name	grob
1	1	(1-1,1-1)	arrange	gtable[layout]
2	2	(2-2,1-1)	arrange	gtable[layout]

```

Warning message:
"‘mode(width)’ differs between new and previous
  ==> NOT changing ‘width’"
Warning message:
"ggrepel: 20 unlabeled data points (too many overlaps). Consider increasing
max.overlaps"
Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")`
instead."
Warning message:
"Removed 1 row(s) containing missing values (geom_path)."
```

```

Warning message:
"Removed 1 row(s) containing missing values (geom_path)."
```

```

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]
```

```

Warning message:
"‘mode(width)’ differs between new and previous
  ==> NOT changing ‘width’"
Warning message:
"ggrepel: 20 unlabeled data points (too many overlaps). Consider increasing
max.overlaps"
Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")`
instead."
```

```

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]
```

```

Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")`
instead."
```

```

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]
```

```

Warning message:
"‘mode(width)’ differs between new and previous
  ==> NOT changing ‘width’"
Warning message:
"ggrepel: 20 unlabeled data points (too many overlaps). Consider increasing
max.overlaps"
Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")`
```

```

instead."

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]

Warning message:
"'mode(width)' differs between new and previous
==> NOT changing 'width'"

Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = \"none\")`
instead."

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]

Warning message:
"'mode(width)' differs between new and previous
==> NOT changing 'width'"

Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = \"none\")`
instead."

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]

Warning message:
"'mode(width)' differs between new and previous
==> NOT changing 'width'"

Warning message in min(exons$start):
"no non-missing arguments to min; returning Inf"
Warning message in max(exons$end):
"no non-missing arguments to max; returning -Inf"
Warning message:
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = \"none\")`
instead."

TableGrob (2 x 1) "arrange": 2 grobs
  z      cells      name      grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (2-2,1-1) arrange gtable[layout]

Warning message:
"'mode(width)' differs between new and previous
==> NOT changing 'width'"

Warning message:

```

```
"`guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> = "none")` instead."
```

```
TableGrob (2 x 1) "arrange": 2 grobs  
  z      cells      name      grob  
1 1 (1-1,1-1) arrange gtable[layout]  
2 2 (2-2,1-1) arrange gtable[layout]
```

```
Warning message:
```

```
"'mode(width)' differs between new and previous  
==> NOT changing 'width'"
```

```
[ ]:
```