main

July 11, 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                                   tidyverse
1.3.1

  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]:
```
levels(meta$group) <- c("Female", "Male")
sample_table
```

A data.frame: 2 × 2

group	count
<chr>	<int>
F	121
M	254

[4]:
```
cluster_summary
```

A data.frame: 3 × 2

Results	n
<chr>	<int>
Number of differentially spliced clusters at FDR = 0.05	31
Fully annotated	11
Contain unannotated junctions	20

[5]:
```
intron_summary
```

A data.frame: 5 × 2

Results	n
<chr>	<int>
Number of fully annotated junctions	70
Number of junctions with cryptic 5' splice site	15
Number of junctions with cryptic 3' splice site	20
Number of junctions with two cryptic splice sites	20
Number of novel junctions that connect two annotated splice sites	7

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

A data.frame: 6 × 6

	clusterID	N	coord	gene	annotation	FDR
	<chr>	<dbl>	<chr>	<chr>	<chr>	<dbl>
31	clu_739_-	3	chrX:53217966-53220839	KDM5C	annotated	7.29e-35
29	clu_729_-	12	chrX:53176622-53193437	KDM5C	cryptic	2.62e-25
27	clu_57947_+	3	chrX:47199106-47199480	UBA1	cryptic	3.27e-11
19	clu_164860_+	3	chr7:74189918-74194741	EIF4H	annotated	1.59e-05
30	clu_736_-	3	chrX:53210576-53211497	KDM5C	cryptic	2.37e-05
3	clu_5190_-	16	chr11:62520391-62530586	.	annotated	1.02e-04

[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 Plot top 6 clusters

[9]:
```
plot_cluster <- function(num, clusters, dir='./'){
    mydata = select_data(num, clusters, exons_table)
    if(mydata$gene == '.'){
        mydata$gene = gsub("-", "_", gsub(":", "_", mydata$coord))
    }
    gene_name = mydata$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)
    try(print(make_gene_plot(mydata$gene,
                         counts = counts,
                         introns = introns,
                         exons_table = exons_table,
                         cluster_list = clusters,
                         clusterID = mydata$cluster,
                         introns_to_plot = introns_to_plot, debug=F)))
    dev.off()
}
```

1.4 Plot splicing

[10]:
```
dir.create("top10")
for(num in 1:10){
    ii = 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:
"`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 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"
```

```
Error in make_gene_plot(mydata$gene, counts = counts, introns = introns,  :
  length(unique(exons$chr)) == 1 is not TRUE
```

```
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 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"
```

```
Error in make_gene_plot(mydata$gene, counts = counts, introns = introns,  :
  length(unique(exons$chr)) == 1 is not TRUE
```

```
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 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"
```

```
Error in make_gene_plot(mydata$gene, counts = counts, introns = introns,  :
  length(unique(exons$chr)) == 1 is not TRUE
```

```
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’"
```

[11]:
```
dir.create("x_chromosome")
x_clu = clusters %>% filter(str_detect(coord, "chrX"), FDR < 0.05)
x_clu
```

A data.frame: 7 × 6

clusterID <chr>	N <dbl>	coord <chr>	gene <chr>	annotation <chr>	FDR <dbl>
clu_739_-	3	chrX:53217966-53220839	KDM5C	annotated	7.29e-35
clu_729_-	12	chrX:53176622-53193437	KDM5C	cryptic	2.62e-25
clu_57947_+	3	chrX:47199106-47199480	UBA1	cryptic	3.27e-11
clu_736_-	3	chrX:53210576-53211497	KDM5C	cryptic	2.37e-05
clu_57932_+	6	chrX:47084600-47092061	RGN	cryptic	4.17e-03
clu_1778_-	5	chrX:152989331-152991916	PNMA5	cryptic	4.33e-03
clu_59263_+	4	chrX:153768428-153769162	PLXNB3	cryptic	4.76e-02

[12]:
```
for(num in 1:dim(x_clu)[1]){
    plot_cluster(num, x_clu, "x_chromosome/")
}
```

```
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:
"`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."
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:
"`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’"
```

[13]:
```
#dir.create("ank3")
ank3 = clusters %>% filter(gene == 'ANK3')
ank3
```

A data.frame: 0 × 6

clusterID <chr>	N <dbl>	coord <chr>	gene <chr>	annotation <chr>	FDR <dbl>

[ ]:
```

```