

# elt2\_promoter\_regulation

Robert Williams

6/3/2022

The goal of this analysis is to evaluate if there is a higher abundance of RNA-seq reads aligning to the elt-2 5'UTR or 3'UTR in elt-2(-) compared to wildtype.

## ELT-2 UTR RNA-seq analysis

Load libraries

```
library(DESeq2)
```

```
## Loading required package: S4Vectors
## Warning: package 'S4Vectors' was built under R version 4.1.1
## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:stats':
##
##   IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##   anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##   dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##   grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##   order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##   rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##   union, unique, unsplit, which.max, which.min
##
## Attaching package: 'S4Vectors'
## The following objects are masked from 'package:base':
##
##   expand.grid, I, unname
```

```

## Loading required package: IRanges
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Warning: package 'GenomeInfoDb' was built under R version 4.1.1
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Warning: package 'MatrixGenerics' was built under R version 4.1.1
## Loading required package: matrixStats
##
## Attaching package: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
##   colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##   colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##   colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##   colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##   colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##   colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##   colWeightedMeans, colWeightedMedians, colWeightedSds,
##   colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##   rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##   rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##   rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##   rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##   rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##   rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##   rowWeightedSds, rowWeightedVars
## Loading required package: Biobase
## Welcome to Bioconductor
##
##   Vignettes contain introductory material; view with
##   'browseVignettes()'. To cite Bioconductor, see
##   'citation("Biobase")', and for packages 'citation("pkgname)".
##
## Attaching package: 'Biobase'
## The following object is masked from 'package:MatrixGenerics':
##
##   rowMedians
## The following objects are masked from 'package:matrixStats':
##
##   anyMissing, rowMedians
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr  1.0.8

```

```

## v tidyr 1.2.0 v stringr 1.4.0
## v readr 2.1.2 v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.1.2
## Warning: package 'readr' was built under R version 4.1.2
## Warning: package 'dplyr' was built under R version 4.1.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::collapse() masks IRanges::collapse()
## x dplyr::combine() masks Biobase::combine(), BiocGenerics::combine()
## x dplyr::count() masks matrixStats::count()
## x dplyr::desc() masks IRanges::desc()
## x tidyr::expand() masks S4Vectors::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::first() masks S4Vectors::first()
## x dplyr::lag() masks stats::lag()
## x ggplot2::Position() masks BiocGenerics::Position(), base::Position()
## x purrr::reduce() masks GenomicRanges::reduce(), IRanges::reduce()
## x dplyr::rename() masks S4Vectors::rename()
## x dplyr::slice() masks IRanges::slice()

#Load and format data

utr_analysis <- function(infile){
countsData <- read.delim(file = infile, header = TRUE, sep = "\t") %>% column_to_rownames(var = "Geneid")
head(countsData)
metadata1 <- c("elt2D_sorted_1",
"elt2D_sorted_2",
"elt2D_sorted_3",
"elt2D_sorted_4",
"elt2Delt7D_sorted_1",
"elt2Delt7D_sorted_2",
"elt2Delt7D_sorted_3",
"wt_sorted_1",
"wt_sorted_2",
"wt_sorted_3",
"wt_sorted_4",
"elt7D_sorted_1",
"elt7D_sorted_2",
"elt7D_sorted_3")

metadata1 <- data.frame(names = metadata1) %>%
  separate(names, sep = "_", into = c("genotype", "sorted", "rep"), remove = FALSE)

metadata1<- metadata1 %>%
  mutate(genotype = fct_relevel(genotype, c("wt", "elt7D", "elt2D", "elt2Delt7D"))) %>%
  arrange(genotype)

countsData <- countsData %>% select(Chr:Length, metadata1$names)

cts <- as.matrix(countsData %>% select(metadata1$names))
rownames(metadata1)<- metadata1$names
coldata <- metadata1[,c("names", "genotype", "rep")]
rownames(coldata) <- as.vector(metadata1$names)

```

```

all(rownames(coldata) == colnames(cts))

# Make DESeqDataSet

dds <- DESeqDataSetFromMatrix(countData = cts,
                              colData = coldata,
                              design = ~ rep + genotype)

dds <- DESeq(dds)
resultsNames(dds)
plotCounts(dds, "WBGene00001250", intgroup = "genotype", returnData = TRUE)
}

utr_elt2_reads <- bind_rows(
  data.frame(utr_analysis("../01_input/five_prime_utr_counts.txt"), type = "five_prime"),
  data.frame(utr_analysis("../01_input/three_prime_utr_counts.txt"), type = "three_prime")
)

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

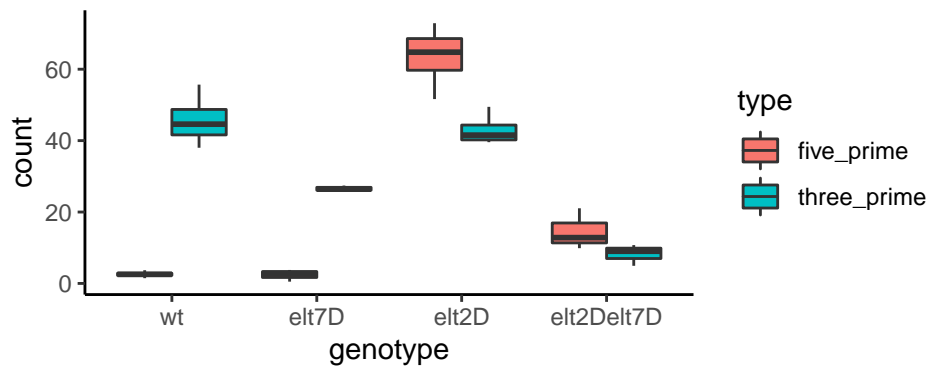
## estimating size factors
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing

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

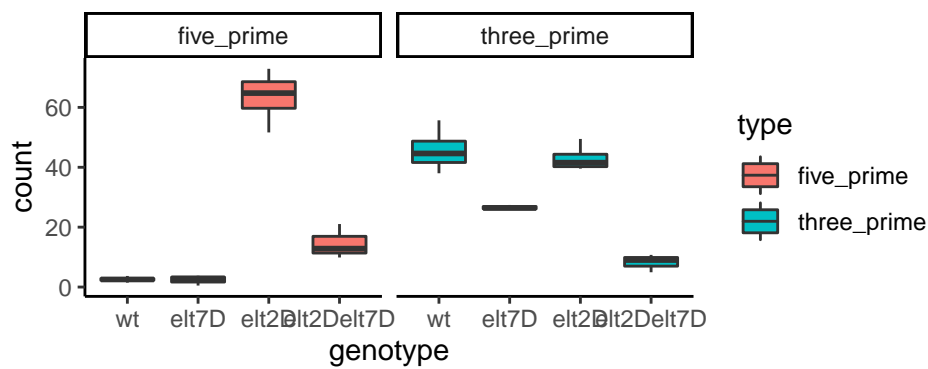
## estimating size factors
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing

ggplot(utr_elt2_reads, aes(x = genotype, y = count, fill = type)) +
  geom_boxplot(position = "dodge") +
  theme_classic()

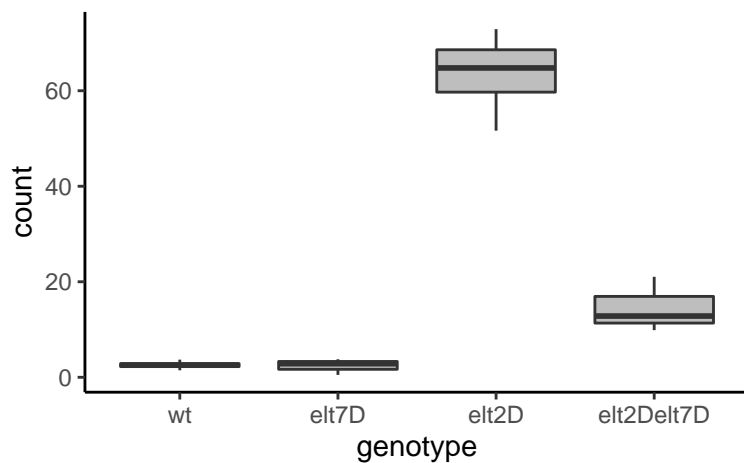
```



```
ggplot(utr_elt2_reads, aes(x = genotype, y = count, fill = type)) +
  geom_boxplot() +
  theme_classic() +
  facet_grid(.~type)
```



```
elt2_five_prime_reads <- ggplot(utr_elt2_reads %>% filter(type == "five_prime"), aes(x = genotype, y = count)) +
  geom_boxplot(fill = "grey") +
  theme_classic()
elt2_five_prime_reads
```



```
ggsave(elt2_five_prime_reads, filename = "../03_output/elt-2_five_prime_reads.pdf", width = 4, height = 4)
```

## Session info

```
sessionInfo()
```

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.7
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] parallel stats4      stats      graphics  grDevices  utils      datasets
## [8] methods    base
##
## other attached packages:
## [1] forcats_0.5.1      stringr_1.4.0
## [3] dplyr_1.0.8        purrr_0.3.4
## [5] readr_2.1.2        tidyr_1.2.0
## [7] tibble_3.1.6       ggplot2_3.3.5
## [9] tidyverse_1.3.1    DESeq2_1.32.0
## [11] SummarizedExperiment_1.22.0 Biobase_2.52.0
## [13] MatrixGenerics_1.4.3 matrixStats_0.61.0
## [15] GenomicRanges_1.44.0 GenomeInfoDb_1.28.4
## [17] IRanges_2.26.0     S4Vectors_0.30.2
## [19] BiocGenerics_0.38.0
##
## loaded via a namespace (and not attached):
## [1] fs_1.5.2           bitops_1.0-7       lubridate_1.8.0
## [4] bit64_4.0.5        RColorBrewer_1.1-3 httr_1.4.2
## [7] tools_4.1.0        backports_1.4.1    utf8_1.2.2
## [10] R6_2.5.1           DBI_1.1.2          colorspace_2.0-3
## [13] withr_2.5.0        tidysselect_1.1.2  bit_4.0.4
## [16] compiler_4.1.0     rvest_1.0.2        cli_3.2.0
## [19] xml2_1.3.3         DelayedArray_0.18.0 labeling_0.4.2
## [22] scales_1.2.0       genefilter_1.74.1  digest_0.6.29
## [25] rmarkdown_2.13     XVector_0.32.0     pkgconfig_2.0.3
## [28] htmltools_0.5.2    highr_0.9          dbplyr_2.1.1
## [31] fastmap_1.1.0      rlang_1.0.2        readxl_1.4.0
## [34] rstudioapi_0.13    RSQLite_2.2.12     farver_2.1.0
## [37] generics_0.1.2     jsonlite_1.8.0     BiocParallel_1.26.2
## [40] RCurl_1.98-1.6     magrittr_2.0.3     GenomeInfoDbData_1.2.6
## [43] Matrix_1.4-1       Rcpp_1.0.8.3       munsell_0.5.0
## [46] fansi_1.0.3        lifecycle_1.0.1    stringi_1.7.6
## [49] yaml_2.3.5         zlibbioc_1.38.0    grid_4.1.0
## [52] blob_1.2.3         crayon_1.5.1       lattice_0.20-45
## [55] Biostrings_2.60.2  haven_2.4.3        splines_4.1.0
## [58] annotate_1.70.0     hms_1.1.1          KEGGREST_1.32.0
## [61] locfit_1.5-9.5     knitr_1.38         pillar_1.7.0
## [64] geneplotter_1.70.0 reprex_2.0.1       XML_3.99-0.9
```

## [67]	glue_1.6.2	evaluate_0.15	modelr_0.1.8
## [70]	png_0.1-7	vctrs_0.4.0	tzdb_0.3.0
## [73]	cellranger_1.1.0	gtable_0.3.0	assertthat_0.2.1
## [76]	cachem_1.0.6	xfun_0.30	xtable_1.8-4
## [79]	broom_0.8.0	survival_3.3-1	AnnotationDbi_1.54.1
## [82]	memoise_2.0.1	ellipsis_0.3.2	