# main

August 30, 2021

```
[1]: suppressMessages(library(SummarizedExperiment))
     suppressMessages(library(limma))
     suppressMessages(library(jaffelab))
```

```
[2]: load("/ceph/projects/v4_phase3_paper/inputs/counts/_m/
     ↪caudate_brainseq_phase3_hg38_rseGene_merged_n464.rda")
```

```
[3]: load("/ceph/projects/v4_phase3_paper/inputs/counts/_m/
     ↪caudate_brainseq_phase3_hg38_rseJxn_merged_n464.rda")
```

```
[4]: ls()
```

1. 'rse_gene' 2. 'rse_jxn'

```
[5]: all(colnames(rse_jxn) == colnames(rse_gene))
```

TRUE

```
[6]: rowData(rse_jxn)$Length <- 100
```

```
[7]: rse_genes_and_drd2junctions = rbind(rse_jxn[c('chr11:113412884-113415420(-)',␣
     ↪'chr11:113414462-113415420(-)', 'chr11:113412884-113414374(-)'),],
         rse_gene)
```

```
[8]: rowData(rse_genes_and_drd2junctions)
```

```
DataFrame with 58040 rows and 21 columns
                              inGencode inGencodeStart inGencodeEnd
                              <logical>      <logical>    <logical>
chr11:113412884-113415420(-)       TRUE           TRUE         TRUE
chr11:113414462-113415420(-)       TRUE           TRUE         TRUE
chr11:113412884-113414374(-)       TRUE           TRUE         TRUE
ENSG00000223972.5                    NA             NA           NA
ENSG00000227232.5                    NA             NA           NA
...                                 ...            ...          ...
ENSG00000198695.2                    NA             NA           NA
ENSG00000210194.1                    NA             NA           NA
ENSG00000198727.2                    NA             NA           NA
ENSG00000210195.2                    NA             NA           NA
```

```
ENSG00000210196.2                                  NA              NA          NA
                                          gencodeGeneID        ensemblID      Symbol
                                            <character>      <character> <character>
chr11:113412884-113415420(-) ENSG00000149295.13 ENSG00000149295         DRD2
chr11:113414462-113415420(-) ENSG00000149295.13 ENSG00000149295         DRD2
chr11:113412884-113414374(-) ENSG00000149295.13 ENSG00000149295         DRD2
ENSG00000223972.5                                  NA ENSG00000223972     DDX11L1
ENSG00000227232.5                                  NA ENSG00000227232      WASH7P
…                                                   …               …           …
ENSG00000198695.2                                  NA ENSG00000198695      MT-ND6
ENSG00000210194.1                                  NA ENSG00000210194       MT-TE
ENSG00000198727.2                                  NA ENSG00000198727      MT-CYB
ENSG00000210195.2                                  NA ENSG00000210195       MT-TT
ENSG00000210196.2                                  NA ENSG00000210196       MT-TP
                               gencodeStrand
                                 <character>
chr11:113412884-113415420(-)               -
chr11:113414462-113415420(-)               -
chr11:113412884-113414374(-)               -
ENSG00000223972.5                         NA
ENSG00000227232.5                         NA
…                                          …
ENSG00000198695.2                         NA
ENSG00000210194.1                         NA
ENSG00000198727.2                         NA
ENSG00000210195.2                         NA
ENSG00000210196.2                         NA
                                                                                   ␣

 →gencodeTx
                                                                               ␣

 →<CharacterList>
chr11:113412884-113415420(-)                                              ␣
 →ENST00000346454.7
chr11:113414462-113415420(-) ENST00000362072.7,ENST00000544518.5,ENST00000542968.
 →5,…
chr11:113412884-113414374(-)        ENST00000362072.7,ENST00000544518.
 →5,ENST00000542968.5
ENSG00000223972.5                                    ENST00000456328.
 →2,ENST00000450305.2
ENSG00000227232.5                                                         ␣
 →ENST00000488147.1
…                                                                             ␣
 →…
ENSG00000198695.2                                                         ␣
 →ENST00000361681.2
ENSG00000210194.1                                                         ␣
 →ENST00000387459.1
```

2

```
ENSG00000198727.2                                                          ␣
 ↪ENST00000361789.2
ENSG00000210195.2                                                          ␣
 ↪ENST00000387460.2
ENSG00000210196.2                                                          ␣
 ↪ENST00000387461.2
```

| | numTx | Class | startExon | endExon |
|---|---|---|---|---|
| | <integer> | <character> | <integer> | <integer> |
| chr11:113412884-113415420(-) | 1 | InGen | 320190 | 320189 |
| chr11:113414462-113415420(-) | 6 | InGen | 320193 | 320189 |
| chr11:113412884-113414374(-) | 3 | InGen | 320190 | 320193 |
| ENSG00000223972.5 | NA | InGen | NA | NA |
| ENSG00000227232.5 | NA | InGen | NA | NA |
| … | … | … | … | … |
| ENSG00000198695.2 | NA | InGen | NA | NA |
| ENSG00000210194.1 | NA | InGen | NA | NA |
| ENSG00000198727.2 | NA | InGen | NA | NA |
| ENSG00000210195.2 | NA | InGen | NA | NA |
| ENSG00000210196.2 | NA | InGen | NA | NA |

| | newGeneID | newGeneSymbol | isFusion |
|---|---|---|---|
| | <character> | <character> | <logical> |
| chr11:113412884-113415420(-) | ENSG00000149295.13 | DRD2 | FALSE |
| chr11:113414462-113415420(-) | ENSG00000149295.13 | DRD2 | FALSE |
| chr11:113412884-113414374(-) | ENSG00000149295.13 | DRD2 | FALSE |
| ENSG00000223972.5 | NA | NA | NA |
| ENSG00000227232.5 | NA | NA | NA |
| … | … | … | … |
| ENSG00000198695.2 | NA | NA | NA |
| ENSG00000210194.1 | NA | NA | NA |
| ENSG00000198727.2 | NA | NA | NA |
| ENSG00000210195.2 | NA | NA | NA |
| ENSG00000210196.2 | NA | NA | NA |

| | Length | gencodeID | gene_type |
|---|---|---|---|
| | <numeric> | <character> | <character> |
| chr11:113412884-113415420(-) | 100 | NA | NA |
| chr11:113414462-113415420(-) | 100 | NA | NA |
| chr11:113412884-113414374(-) | 100 | NA | NA |
| ENSG00000223972.5 | 1735 | ENSG00000223972.5 | transcribed_unproces.. |
| ENSG00000227232.5 | 1351 | ENSG00000227232.5 | unprocessed_pseudogene |
| … | … | … | … |
| ENSG00000198695.2 | 525 | ENSG00000198695.2 | protein_coding |
| ENSG00000210194.1 | 69 | ENSG00000210194.1 | Mt_tRNA |
| ENSG00000198727.2 | 1141 | ENSG00000198727.2 | protein_coding |
| ENSG00000210195.2 | 66 | ENSG00000210195.2 | Mt_tRNA |
| ENSG00000210196.2 | 68 | ENSG00000210196.2 | Mt_tRNA |

| EntrezID | meanExprs | NumTx |
|---|---|---|
| <integer> | <numeric> | <integer> |

```
chr11:113412884-113415420(-)           NA          NA          NA
chr11:113414462-113415420(-)           NA          NA          NA
chr11:113412884-113414374(-)           NA          NA          NA
ENSG00000223972.5                   84771  0.00186396           2
ENSG00000227232.5                      NA  1.22336500           1
…                                       …           …           …
ENSG00000198695.2                    4541   109.75770           1
ENSG00000210194.1                      NA     8.36876           1
ENSG00000198727.2                    4519   570.90682           1
ENSG00000210195.2                      NA   157.30261           1
ENSG00000210196.2                      NA   213.52484           1
```

[9]: 
```r
rpkm = recount::getRPKM(rse_genes_and_drd2junctions, 'Length')
```

Setting options('download.file.method.GEOquery'='auto')

Setting options('GEOquery.inmemory.gpl'=FALSE)

[10]: 
```r
log2rpkm = log2( rpkm + 1)
```

[11]: 
```r
summary(rowMeans(rpkm))
```
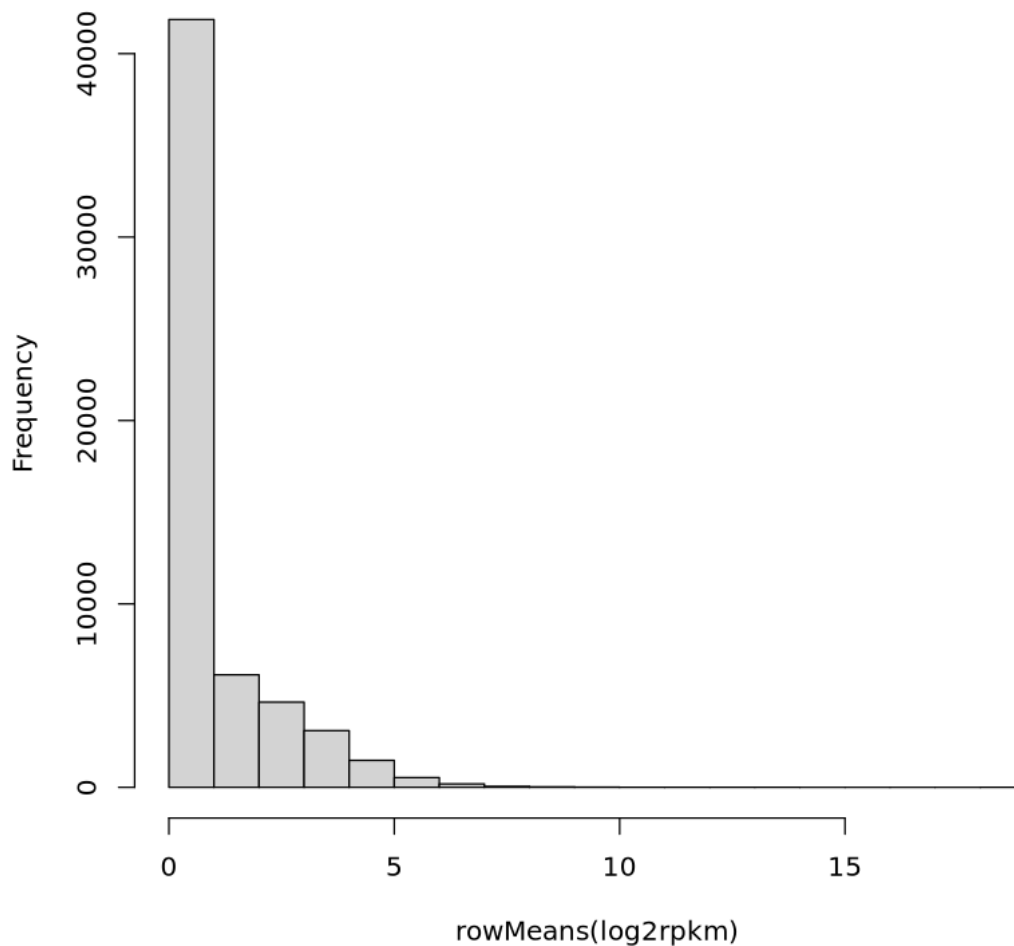
```
     Min.  1st Qu.   Median     Mean  3rd Qu.      Max.
 0.00e+00 0.00e+00 9.00e-02 9.18e+00 1.45e+00 2.85e+05
```

[12]: 
```r
summary(rowMeans(log2rpkm))
```

```
      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.000000  0.004363  0.116758  0.845061  1.221697 18.066738
```

[13]: 
```r
hist(rowMeans(log2rpkm))
```

## Histogram of rowMeans(log2rpkm)



[14]: `sum(rowMeans(log2rpkm) > 0.2)`

26513

[15]: `head(rowMeans(log2rpkm))`

**chr11:113412884-113415420(-)**            2.28619688611868 **chr11:113414462-113415420(-)**
5.43409072729181 **chr11:113412884-113414374(-)**    5.04089639513039 **ENSG00000223972.5**
0.00595840349433603 **ENSG00000227232.5**            1.75797637872393 **ENSG00000278267.1**
1.98466031706537

[16]: `filtered_log2_rpkm = log2rpkm[rowMeans(log2rpkm) > 0.2,]`

[17]: `write.csv(filtered_log2_rpkm, file='log2rpkm.csv')`