

# main

July 10, 2021

## 1 Generate phenotypes

```
[1]: suppressMessages({library(SummarizedExperiment)
                        library(dplyr)
                        library(sva)})
```

### 1.1 Function

```
[2]: # Function from jaffelab github
merge_rse_metrics <- function(rse) {
  stopifnot(is(rse, 'RangedSummarizedExperiment'))

  rse$overallMapRate = mapply(function(r, n) {
    sum(r*n)/sum(n)
  }, rse$overallMapRate, rse$numReads)
  rse$mitoRate = mapply(function(r, n) {
    sum(r*n)/sum(n)
  }, rse$mitoRate, rse$numMapped)
  rse$rRNA_rate = mapply(function(r, n) {
    sum(r*n)/sum(n)
  }, rse$rRNA_rate, rse$numMapped)
  rse$totalAssignedGene = mapply(function(r, n) {
    sum(r*n)/sum(n)
  }, rse$totalAssignedGene, rse$numMapped)

  rse$numMapped = sapply(rse$numMapped, sum)
  rse$numReads = sapply(rse$numReads, sum)
  rse$numUnmapped = sapply(rse$numUnmapped, sum)
  rse$mitoMapped = sapply(rse$mitoMapped, sum)
  rse$totalMapped = sapply(rse$totalMapped, sum)
  return(rse)
}
```

## 1.2 Main

```
[3]: fields = c('BrNum', 'RNum', 'Region', 'RIN', 'Age', 'Sex', 'Race',
               'Dx', 'mitoRate', 'rRNA_rate', 'overallMapRate')

[4]: load("../..//counts/_m/degradation_rse_phase3_caudate.rda")
dm <- assays(cov_rse_caudate)$counts
qSV <- qsva(dm)

if("TRUE" %in% grepl("_", rownames(qSV))){# Remove underscore if needed
  ↪
  rownames(qSV) <- sapply(strsplit(rownames(qSV), "_"), "[", 1)
}
qSV %>% as.data.frame %>% write.csv("qSV_caudate.csv")

load("../..//counts/_m/caudate_brainseq_phase3_hg38_rseGene_merged_n464.rda')

rse_gene <- merge_rse_metrics(rse_gene)
colData(rse_gene)$RIN = sapply(colData(rse_gene)$RIN, "[", 1)

colData(rse_gene)[,fields] %>%
  as.data.frame %>% write.csv(file = 'caudate_phenotypes.csv')

[5]: load("../..//counts/_m/degradation_rse_phase2_dlpfc.rda")
dm <- assays(cov_rse_dlpfc)$counts
qSV <- qsva(dm)

if("TRUE" %in% grepl("_", rownames(qSV))){# Remove underscore if needed
  ↪
  rownames(qSV) <- sapply(strsplit(rownames(qSV), "_"), "[", 1)
}
qSV %>% as.data.frame %>% write.csv("qSV_dlpfc.csv")

load('../..//counts/_m/dlpfc_ribozero_brainseq_phase2_hg38_rseGene_merged_n453.
  ↪rda')

rse_gene <- merge_rse_metrics(rse_gene)
colData(rse_gene)$RIN = sapply(colData(rse_gene)$RIN, "[", 1)

colData(rse_gene)[,fields] %>%
  as.data.frame %>% write.csv(file = 'dlpfc_phenotypes.csv')

[6]: load("../..//counts/_m/degradation_rse_phase2_hippo.rda")
dm <- assays(cov_rse_hippo)$counts
qSV <- qsva(dm)
```

```

if("TRUE" %in% grepl("_", rownames(qSV))){# Remove underscore if needed
  ↪
  rownames(qSV) <- sapply(strsplit(rownames(qSV), "_"), "[", 1)
}
qSV %>% as.data.frame %>% write.csv("qSV_hippo.csv")

load("../../counts/_m/hippo_brainseq_phase2_hg38_rseGene_merged_n447.rda")

rse_gene <- merge_rse_metrics(rse_gene)
colData(rse_gene)$RIN = sapply(colData(rse_gene)$RIN, "[", 1)

colData(rse_gene)[,fields] %>% as.data.frame %>%
  write.csv(file = 'hippo_phenotypes.csv')

```

```

[7]: load("../../counts/_m/degradation_rse_dg_hippo_n263.rda")

dm <- assays(cov_rse)$counts
qSV <- qsva(dm)

if("TRUE" %in% grepl("_", rownames(qSV))){# Remove underscore if needed
  ↪
  rownames(qSV) <- sapply(strsplit(rownames(qSV), "_"), "[", 1)
}
qSV %>% as.data.frame %>% write.csv("qSV_dg.csv")

load("../../counts/_m/astellas_dg_hg38_rseGene_n263.rda")

rse_gene <- merge_rse_metrics(rse_gene)
colData(rse_gene)$RIN = sapply(colData(rse_gene)$RIN, "[", 1)
rownames(colData(rse_gene)) <- sapply(strsplit(rownames(colData(rse_gene)),
  ↪ "_"), "[", 1)

dg_hippo = colData(rse_gene)[,fields] %>% as.data.frame %>%
  write.csv(file = 'dg_phenotypes.csv')

```

### 1.3 Reproducibility Information

```

[8]: Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()

```

```
[1] "2021-07-10 11:06:34 EDT"
```

```

      user  system elapsed
8797.312 4130.982  846.486

```

```
Session info
```

```

setting  value
version  R version 4.0.3 (2020-10-10)
os       Arch Linux
system   x86_64, linux-gnu
ui       X11
language (EN)
collate  en_US.UTF-8
ctype    en_US.UTF-8
tz       America/New_York
date     2021-07-10

```

#### Packages

package	* version	date	lib	source
annotate	1.68.0	2020-10-27	[1]	Bioconductor
AnnotationDbi	1.52.0	2020-10-27	[1]	Bioconductor
assertthat	0.2.1	2019-03-21	[1]	CRAN (R 4.0.2)
base64enc	0.1-3	2015-07-28	[1]	CRAN (R 4.0.2)
Biobase	* 2.50.0	2020-10-27	[1]	Bioconductor
BiocGenerics	* 0.36.1	2021-04-16	[1]	Bioconductor
BiocParallel	* 1.24.1	2020-11-06	[1]	Bioconductor
bit	4.0.4	2020-08-04	[1]	CRAN (R 4.0.2)
bit64	4.0.5	2020-08-30	[1]	CRAN (R 4.0.2)
bitops	1.0-7	2021-04-24	[1]	CRAN (R 4.0.3)
blob	1.2.1	2020-01-20	[1]	CRAN (R 4.0.2)
cachem	1.0.5	2021-05-15	[1]	CRAN (R 4.0.3)
cli	3.0.0	2021-06-30	[1]	CRAN (R 4.0.3)
crayon	1.4.1	2021-02-08	[1]	CRAN (R 4.0.3)
DBI	1.1.1	2021-01-15	[1]	CRAN (R 4.0.2)
DelayedArray	0.16.3	2021-03-24	[1]	Bioconductor
digest	0.6.27	2020-10-24	[1]	CRAN (R 4.0.2)
dplyr	* 1.0.7	2021-06-18	[1]	CRAN (R 4.0.3)
edgeR	3.32.1	2021-01-14	[1]	Bioconductor
ellipsis	0.3.2	2021-04-29	[1]	CRAN (R 4.0.3)
evaluate	0.14	2019-05-28	[1]	CRAN (R 4.0.2)
fansi	0.5.0	2021-05-25	[1]	CRAN (R 4.0.3)
fastmap	1.1.0	2021-01-25	[1]	CRAN (R 4.0.2)
genefilter	* 1.72.1	2021-01-21	[1]	Bioconductor
generics	0.1.0	2020-10-31	[1]	CRAN (R 4.0.2)
GenomeInfoDb	* 1.26.7	2021-04-08	[1]	Bioconductor
GenomeInfoDbData	1.2.4	2021-02-02	[1]	Bioconductor
GenomicRanges	* 1.42.0	2020-10-27	[1]	Bioconductor
glue	1.4.2	2020-08-27	[1]	CRAN (R 4.0.2)
htmltools	0.5.1.1	2021-01-22	[1]	CRAN (R 4.0.2)
httr	1.4.2	2020-07-20	[1]	CRAN (R 4.0.2)
IRanges	* 2.24.1	2020-12-12	[1]	Bioconductor
IRdisplay	1.0	2021-01-20	[1]	CRAN (R 4.0.2)
IRkernel	1.2	2021-05-11	[1]	CRAN (R 4.0.3)
jsonlite	1.7.2	2020-12-09	[1]	CRAN (R 4.0.2)

lattice	0.20-41	2020-04-02	[2]	CRAN (R 4.0.3)
lifecycle	1.0.0	2021-02-15	[1]	CRAN (R 4.0.3)
limma	3.46.0	2020-10-27	[1]	Bioconductor
locfit	1.5-9.4	2020-03-25	[1]	CRAN (R 4.0.2)
magrittr	2.0.1	2020-11-17	[1]	CRAN (R 4.0.2)
Matrix	1.3-4	2021-06-01	[1]	CRAN (R 4.0.3)
MatrixGenerics	* 1.2.1	2021-01-30	[1]	Bioconductor
matrixStats	* 0.59.0	2021-06-01	[1]	CRAN (R 4.0.3)
memoise	2.0.0	2021-01-26	[1]	CRAN (R 4.0.2)
mgcv	* 1.8-33	2020-08-27	[2]	CRAN (R 4.0.3)
nlme	* 3.1-152	2021-02-04	[1]	CRAN (R 4.0.3)
pbdZMQ	0.3-5	2021-02-10	[1]	CRAN (R 4.0.3)
pillar	1.6.1	2021-05-16	[1]	CRAN (R 4.0.3)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN (R 4.0.2)
purrr	0.3.4	2020-04-17	[1]	CRAN (R 4.0.2)
R6	2.5.0	2020-10-28	[1]	CRAN (R 4.0.2)
Rcpp	1.0.7	2021-07-07	[1]	CRAN (R 4.0.3)
RCurl	1.98-1.3	2021-03-16	[1]	CRAN (R 4.0.3)
repr	1.1.3	2021-01-21	[1]	CRAN (R 4.0.2)
rlang	0.4.11	2021-04-30	[1]	CRAN (R 4.0.3)
RSQLite	2.2.7	2021-04-22	[1]	CRAN (R 4.0.3)
S4Vectors	* 0.28.1	2020-12-09	[1]	Bioconductor
sessioninfo	1.1.1	2018-11-05	[1]	CRAN (R 4.0.2)
SummarizedExperiment	* 1.20.0	2020-10-27	[1]	Bioconductor
survival	3.2-7	2020-09-28	[2]	CRAN (R 4.0.3)
sva	* 3.38.0	2020-10-27	[1]	Bioconductor
tibble	3.1.2	2021-05-16	[1]	CRAN (R 4.0.3)
tidyselect	1.1.1	2021-04-30	[1]	CRAN (R 4.0.3)
utf8	1.2.1	2021-03-12	[1]	CRAN (R 4.0.3)
uuid	0.1-4	2020-02-26	[1]	CRAN (R 4.0.2)
vctrs	0.3.8	2021-04-29	[1]	CRAN (R 4.0.3)
withr	2.4.2	2021-04-18	[1]	CRAN (R 4.0.3)
XML	3.99-0.6	2021-03-16	[1]	CRAN (R 4.0.3)
xtable	1.8-4	2019-04-21	[1]	CRAN (R 4.0.2)
XVector	0.30.0	2020-10-27	[1]	Bioconductor
zlibbioc	1.36.0	2020-10-27	[1]	Bioconductor

[1] /home/jbenja13/R/x86\_64-pc-linux-gnu-library/4.0

[2] /usr/lib/R/library