

# main

August 5, 2021

## 1 Sample Summary

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

### 1.1 Functions

```
[2]: get_mds <- function(){
      mds_file = "/ceph/projects/v4_phase3_paper/inputs/genotypes/mds/_m/
      ↪LIBD_Brain_TopMed.mds"
      mds = data.table::fread(mds_file) %>%
        rename_at(.vars = vars(starts_with("C")),
                  function(x){sub("C", "snpc", x)}) %>%
        mutate_if(is.character, as.factor)
      return(mds)
    }

memMDS <- memoise::memoise(get_mds)

get_pheno <- function(){
  counts_lt = "/ceph/projects/v4_phase3_paper/inputs/counts/_m/
  ↪caudate_brainseq_phase3_hg38_rseGene_merged_n464.rda"
  load(counts_lt)
  rse_df = rse_gene
  keepIndex = which(rse_df$Age > 13 & rse_df$Race %in% c("AA", "CAUC"))
  rse_df = rse_df[, keepIndex]
  rse_df$Sex <- factor(rse_df$Sex)
  rse_df <- jaffelab::merge_rse_metrics(rse_df)
  colData(rse_df)$RIN = sapply(colData(rse_df)$RIN, "[", 1)
  rownames(colData(rse_df)) <- sapply(strsplit(rownames(colData(rse_df)), "_",
  ↪"_"), "[", 1)
  pheno = colData(rse_df) %>% as.data.frame %>%
    inner_join(memMDS(), by=c("BrNum"="FID")) %>%
    distinct(RNum, .keep_all = TRUE)
  return(pheno)
}
```

```
memPHENO <- memoise::memoise(get_pheno)
```

## 1.2 Sample breakdown

### 1.2.1 eQTL analysis

```
[3]: cols = c('BrNum', 'RNum', 'Region', 'RIN', 'Age', 'Sex', 'Race', 'Dx',
              'ERCCsumLogErr', 'overallMapRate', 'totalAssignedGene',
              'mitoRate', 'rRNA_rate', 'snpc1', 'snpc2', 'snpc3')

df = memPHENO() %>% select(all_of(cols))
df %>% dim
```

```
1. 443 2. 16
```

```
[4]: print(paste("There are", df$BrNum %>% unique %>% length, "unique BrNum."))
```

```
[1] "There are 443 unique BrNum."
```

```
[5]: table(df$Dx)
```

```
Bipolar Control Schizo
      44      245      154
```

```
[6]: table(df$Dx, df$Sex)
```

```
      F  M
Bipolar 16 28
Control 76 169
Schizo  50 104
```

```
[7]: table(df$Dx, df$Race)
```

```
      AA CAUC
Bipolar  4  40
Control 123 122
Schizo  83  71
```

```
[8]: table(df$Sex)
```

```
  F  M
142 301
```

```
[9]: table(df$Race)
```

```
  AA CAUC
```

210 233

### Mean

```
[10]: df %>% group_by(Dx) %>%  
      summarise(across(c("RIN", "Age"), ~ mean(.x, na.rm = TRUE)))
```

	Dx <chr>	RIN <dbl>	Age <dbl>
A tibble: 3 × 3	Bipolar	7.822727	43.21182
	Control	7.857143	47.46478
	Schizo	7.875974	51.74896

### Median

```
[11]: df %>% group_by(Dx) %>%  
      summarise(across(c("RIN", "Age"), ~ median(.x, na.rm = TRUE)))
```

	Dx <chr>	RIN <dbl>	Age <dbl>
A tibble: 3 × 3	Bipolar	7.65	44.785
	Control	7.80	48.690
	Schizo	7.90	51.970

### Standard deviation

```
[12]: df %>% group_by(Dx) %>%  
      summarise(across(c("RIN", "Age"), ~ sd(.x, na.rm = TRUE)))
```

	Dx <chr>	RIN <dbl>	Age <dbl>
A tibble: 3 × 3	Bipolar	0.8066190	12.51277
	Control	0.7914502	16.54702
	Schizo	0.9692700	14.96908

## 1.2.2 Expression analysis

```
[13]: df2 = df %>% filter(Age > 17)  
df2 %>% dim
```

1. 437 2. 16

```
[14]: print(paste("There are", df2$BrNum %>% unique %>% length, "unique BrNum."))
```

```
[1] "There are 437 unique BrNum."
```

```
[15]: table(df2$Dx)
```

Bipolar	Control	Schizo
44	239	154

```
[16]: table(df2$Dx, df2$Sex)
```

	F	M
Bipolar	16	28
Control	71	168
Schizo	50	104

```
[17]: table(df2$Sex, df2$Race)
```

	AA	CAUC
F	78	59
M	131	169

```
[18]: table(df2$Race)
```

	AA	CAUC
	209	228

```
[19]: table(df2$Sex)
```

	F	M
	137	300

### Mean

```
[20]: df2 %>% group_by(Dx) %>%
      summarise(across(c("RIN", "Age"), ~ mean(.x, na.rm = TRUE)))
```

	Dx	RIN	Age
	<chr>	<dbl>	<dbl>
A tibble: 3 × 3	Bipolar	7.822727	43.21182
	Control	7.855649	48.29879
	Schizo	7.875974	51.74896

### Median

```
[21]: df2 %>% group_by(Dx) %>%
      summarise(across(c("RIN", "Age"), ~ median(.x, na.rm = TRUE)))
```

	Dx	RIN	Age
	<chr>	<dbl>	<dbl>
A tibble: 3 × 3	Bipolar	7.65	44.785
	Control	7.80	48.770
	Schizo	7.90	51.970

### Standard deviation

```
[22]: df2 %>% group_by(Dx) %>%
      summarise(across(c("RIN", "Age"), ~ sd(.x, na.rm = TRUE)))
```

	Dx <chr>	RIN <dbl>	Age <dbl>
A tibble: 3 × 3	Bipolar	0.8066190	12.51277
	Control	0.7925331	15.87895
	Schizo	0.9692700	14.96908

### 1.3 Reproducibility Information

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

```
[1] "2021-08-05 12:10:10 EDT"
```

```
   user  system elapsed
12.346   0.617  13.288
```

```
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     2021-08-05
```

```
Packages
package      * version  date      lib source
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
bitops        1.0-7    2021-04-24 [1] CRAN (R 4.0.3)
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)
data.table    1.14.0   2021-02-21 [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)
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)
fansib        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)
```

fs	1.5.0	2020-07-31	[1]	CRAN (R 4.0.2)
gargle	1.2.0	2021-07-02	[1]	CRAN (R 4.0.3)
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)
googledrive	2.0.0	2021-07-08	[1]	CRAN (R 4.0.3)
htmltools	0.5.1.1	2021-01-22	[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)
jaffelab	0.99.30	2021-02-02	[1]	Github (LieberInstitute/
↪jaffelab@42637ff)				
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
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)
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)
rafalib	1.0.2	2021-04-26	[1]	Github (ririzarr/rafalib@2580666)
RColorBrewer	1.1-2	2014-12-07	[1]	CRAN (R 4.0.2)
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)
S4Vectors	* 0.28.1	2020-12-09	[1]	Bioconductor
segmented	1.3-4	2021-04-22	[1]	CRAN (R 4.0.3)
sessioninfo	1.1.1	2018-11-05	[1]	CRAN (R 4.0.2)
SummarizedExperiment	* 1.20.0	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)
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
```