main

August 1, 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", "snpPC", x)}) %>%
             mutate_if(is.character, as.factor)
         return(mds)
     }

     memMDS <- memoise::memoise(get_mds)

     get_pheno <- function(tissue){
         counts_lt = list("caudate"="/ceph/projects/v4_phase3_paper/inputs/counts/_m/
     ↪caudate_brainseq_phase3_hg38_rseGene_merged_n464.rda",
                          "dlpfc"="/ceph/projects/v4_phase3_paper/inputs/counts/_m/
     ↪dlpfc_ribozero_brainseq_phase2_hg38_rseGene_merged_n453.rda",
                          "hippocampus"="/ceph/projects/v4_phase3_paper/inputs/
     ↪counts/_m/hippo_brainseq_phase2_hg38_rseGene_merged_n447.rda")
         load(counts_lt[[tissue]])
         rse_df = rse_gene
         keepIndex = which((rse_df$Dx %in% c("Control", "Schizo")) &
                           rse_df$Age > 13 & rse_df$Race %in% c("AA", "CAUC"))
         rse_df = rse_df[, keepIndex]
         rse_df$Dx = factor(rse_df$Dx, levels = c("Control", "Schizo"))
         rse_df$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', 'snpPC1', 'snpPC2', 'snpPC3')

    cc = memPHENO("caudate") %>% select(all_of(cols))
    dd = memPHENO("dlpfc") %>% select(all_of(cols))
    hh = memPHENO("hippocampus") %>% select(all_of(cols))
    df = bind_rows(cc, dd, hh)
    df %>% dim
```

1. 1170 2. 16

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

    [1] "There are 504 unique BrNum."

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

Caudate	DLPFC	HIPPO
399	377	394

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

	Control	Schizo
Caudate	245	154
DLPFC	229	148
HIPPO	261	133

```
[7]: table(df$Region, df$Sex)
```

	F	M
Caudate	126	273
DLPFC	121	256
HIPPO	126	268

```
[8]: table(df$Region, df$Race)
```

```
            AA  CAUC
    Caudate 206  193
    DLPFC   204  173
    HIPPO   213  181
```

```
[9]: df %>% group_by(Region, Sex, Dx) %>% summarise(N=n())
```

`summarise()` has grouped output by 'Region', 'Sex'. You can override using the
`.groups` argument.

	Region	Sex	Dx	N
	<chr>	<fct>	<fct>	<int>
	Caudate	F	Control	76
	Caudate	F	Schizo	50
	Caudate	M	Control	169
	Caudate	M	Schizo	104
A grouped_df: 12 × 4	DLPFC	F	Control	73
	DLPFC	F	Schizo	48
	DLPFC	M	Control	156
	DLPFC	M	Schizo	100
	HIPPO	F	Control	79
	HIPPO	F	Schizo	47
	HIPPO	M	Control	182
	HIPPO	M	Schizo	86

Mean

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

	Region	RIN	Age	rRNA_rate	mitoRate
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
A tibble: 3 × 5	Caudate	7.864411	49.11832	5.687892e-05	0.03651875
	DLPFC	7.698674	45.82111	5.056699e-05	0.02173725
	HIPPO	7.617513	45.48041	2.978609e-05	0.22688140

Median

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

A tibble: 3 × 5	Region	RIN	Age	rRNA_rate	mitoRate
	\<chr\>	\<dbl\>	\<dbl\>	\<dbl\>	\<dbl\>
	Caudate	7.8	49.650	4.523605e-05	0.03439737
	DLPFC	7.8	48.100	4.060600e-05	0.02063706
	HIPPO	7.8	47.515	1.991029e-05	0.23442298

Standard deviation

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

A tibble: 3 × 5	Region	RIN	Age	rRNA_rate	mitoRate
	\<chr\>	\<dbl\>	\<dbl\>	\<dbl\>	\<dbl\>
	Caudate	0.8632861	16.07355	3.630981e-05	0.01664994
	DLPFC	0.9352001	16.51391	3.180953e-05	0.00773969
	HIPPO	1.0323633	16.43348	2.769952e-05	0.09132075

1.2.2 Expression analysis

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

1. 1127 2. 16

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

```
[1] "There are 480 unique BrNum."
```

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

```
Caudate    DLPFC    HIPPO
    393      359      375
```

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

```
         Control Schizo
 Caudate     239    154
 DLPFC       211    148
 HIPPO       242    133
```

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

```
           F   M
 Caudate 121 272
 DLPFC   114 245
 HIPPO   121 254
```

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

```
           AA CAUC
   Caudate 205  188
   DLPFC   200  159
   HIPPO   207  168
```

Mean

```
[19]:  df2 %>% group_by(Region) %>%
           summarise(across(c("RIN", "Age", "rRNA_rate", "mitoRate"), ~ mean(.x, na.rm␣
       ↪= TRUE)))
```

A tibble: 3 × 5

| Region | RIN | Age | rRNA_rate | mitoRate |
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Caudate	7.863613	49.65076	5.658740e-05	0.03639159
DLPFC	7.666295	47.35662	5.149245e-05	0.02148476
HIPPO	7.598667	47.02501	3.026710e-05	0.22567167

Median

```
[20]:  df2 %>% group_by(Region) %>%
           summarise(across(c("RIN", "Age", "rRNA_rate", "mitoRate"), ~ median(.x, na.
       ↪rm = TRUE)))
```

A tibble: 3 × 5

| Region | RIN | Age | rRNA_rate | mitoRate |
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Caudate	7.8	50.00	4.447903e-05	0.03439737
DLPFC	7.8	48.66	4.175559e-05	0.02044370
HIPPO	7.8	48.34	1.997499e-05	0.23428324

Standard deviation

```
[21]:  df2 %>% group_by(Region) %>%
           summarise(across(c("RIN", "Age", "rRNA_rate", "mitoRate"), ~ sd(.x, na.rm =␣
       ↪TRUE)))
```

A tibble: 3 × 5

| Region | RIN | Age | rRNA_rate | mitoRate |
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Caudate	0.8649491	15.60086	3.634418e-05	0.016609313
DLPFC	0.9221092	15.38859	3.224609e-05	0.007513765
HIPPO	1.0321689	15.29984	2.823267e-05	0.092901235

```
[22]:  df2 %>% group_by(Region, Sex, Dx) %>% summarise(N=n())
```

`summarise()` has grouped output by 'Region', 'Sex'. You can override using the
`.groups` argument.

	Region	Sex	Dx	N
	<chr>	<fct>	<fct>	<int>
	Caudate	F	Control	71
	Caudate	F	Schizo	50
	Caudate	M	Control	168
	Caudate	M	Schizo	104
A grouped_df: 12 × 4	DLPFC	F	Control	66
	DLPFC	F	Schizo	48
	DLPFC	M	Control	145
	DLPFC	M	Schizo	100
	HIPPO	F	Control	74
	HIPPO	F	Schizo	47
	HIPPO	M	Control	168
	HIPPO	M	Schizo	86

1.3 Reproducibility Information

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

```
[1] "2021-08-01 12:10:35 EDT"

   user  system elapsed
 18.971   1.199  20.512

 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     2021-08-01

 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)
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)
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
```