main

September 2, 2021

1 Generate covariates matrix

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

[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)
}


prep_covs <- function(tissue){
    ancestry = "../../../ancestry_structure/structure.
 ↪out_ancestry_proportion_raceDemo_compare"
```

```
    counts_lt = list("caudate"="../../../counts/_m/
↪caudate_brainseq_phase3_hg38_rseGene_merged_n464.rda",
                     "dg"="../../../counts/_m/astellas_dg_hg38_rseGene_n263.
↪rda",
                     'dlpfc'="../../../counts/_m/
↪dlpfc_ribozero_brainseq_phase2_hg38_rseGene_merged_n453.rda",
                     "hippocampus"="../../../counts/_m/
↪hippo_brainseq_phase2_hg38_rseGene_merged_n447.rda")
    qsv_lt = list("caudate"="../../_m/qSV_caudate.csv", "dg"="../../_m/qSV_dg.
↪csv",
                  "dlpfc"="../../_m/qSV_dlpfc.csv", "hippocampus"="../../_m/
↪qSV_hippo.csv")
    load(counts_lt[[tissue]])
    rse_df = rse_gene
    keepIndex = which((rse_df$Dx %in% c("Control", "Schizo")) &
                      rse_df$Race %in% c("AA", "CAUC"))
    rse_df = rse_df[, keepIndex]
    rse_df <- 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(data.table::fread(ancestry), by=c("BrNum"="id",␣
↪"Race"="group")) %>%
        select(RNum, Eur, Dx, Sex, Age, mitoRate, rRNA_rate, totalAssignedGene,␣
↪overallMapRate) %>%
        column_to_rownames("RNum")
    mod = model.matrix(~Eur + Dx + Sex + Age + mitoRate + rRNA_rate +
                       totalAssignedGene + overallMapRate, data = pheno)
    colnames(mod) <- gsub("Dx", "", colnames(mod))
    colnames(mod) <- gsub("SexM", "Male", colnames(mod))
    colnames(mod) <- gsub("Eur", "EA", colnames(mod))
    colnames(mod) <- gsub("\\(Intercept\\)", "Intercept", colnames(mod))
    mod %>% as.data.frame %>% rownames_to_column("RNum") %>%
        inner_join(data.table::fread(qsv_lt[[tissue]]), by=c("RNum"="V1")) %>%
        rename_all(list(~str_replace_all(., 'PC', 'qPC'))) %>%
        data.table::fwrite(paste0(tissue, "_covariates.csv"), sep=',')
}
```

1.1 Main loop

[3]:
```
for(tissue in c("caudate", "dg", "dlpfc", "hippocampus")){
    prep_covs(tissue)
}
```

1.2 Reproducibility Information

[4]:
```
Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-09-02 16:38:16 EDT"
```

```
   user  system elapsed
 20.484   1.677  22.975
```

```
 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     2021-09-02

 Packages
 package              * version  date       lib source
 assertthat             0.2.1    2019-03-21 [1] CRAN (R 4.0.2)
 backports              1.2.1    2020-12-09 [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)
 broom                  0.7.8    2021-06-24 [1] CRAN (R 4.0.3)
 cellranger             1.1.0    2016-07-27 [1] CRAN (R 4.0.2)
 cli                    3.0.0    2021-06-30 [1] CRAN (R 4.0.3)
 colorspace             2.0-2    2021-06-24 [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)
 dbplyr                 2.1.1    2021-04-06 [1] CRAN (R 4.0.3)
 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)
 forcats              * 0.5.1    2021-01-27 [1] CRAN (R 4.0.2)
 fs                     1.5.0    2020-07-31 [1] CRAN (R 4.0.2)
 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
ggplot2              * 3.3.5    2021-06-25 [1] CRAN (R 4.0.3)
glue                   1.4.2    2020-08-27 [1] CRAN (R 4.0.2)
gtable                 0.3.0    2019-03-25 [1] CRAN (R 4.0.2)
haven                  2.4.1    2021-04-23 [1] CRAN (R 4.0.3)
hms                    1.1.0    2021-05-17 [1] CRAN (R 4.0.3)
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)
lubridate              1.7.10   2021-02-26 [1] CRAN (R 4.0.3)
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)
modelr                 0.1.8    2020-05-19 [1] CRAN (R 4.0.2)
munsell                0.5.0    2018-06-12 [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)
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)
readr                * 1.4.0    2020-10-05 [1] CRAN (R 4.0.2)
readxl                 1.3.1    2019-03-13 [1] CRAN (R 4.0.2)
repr                   1.1.3    2021-01-21 [1] CRAN (R 4.0.2)
reprex                 2.0.0    2021-04-02 [1] CRAN (R 4.0.3)
rlang                  0.4.11   2021-04-30 [1] CRAN (R 4.0.3)
rstudioapi             0.13     2020-11-12 [1] CRAN (R 4.0.2)
rvest                  1.0.0    2021-03-09 [1] CRAN (R 4.0.3)
S4Vectors            * 0.28.1   2020-12-09 [1] Bioconductor
scales                 1.1.1    2020-05-11 [1] CRAN (R 4.0.2)
sessioninfo            1.1.1    2018-11-05 [1] CRAN (R 4.0.2)
stringi                1.7.3    2021-07-16 [1] CRAN (R 4.0.3)
stringr              * 1.4.0    2019-02-10 [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)
tidyr                * 1.1.3    2021-03-03 [1] CRAN (R 4.0.3)
tidyselect             1.1.1    2021-04-30 [1] CRAN (R 4.0.3)
tidyverse            * 1.3.1    2021-04-15 [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)
 xml2                  1.3.2    2020-04-23 [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
```