main

August 23, 2021

1 Examine sample make-up

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

1.1 Samples after quality control

[2]:
```
save_ggplots <- function(p, fn, w, h){
    for(ext in c('.pdf', '.png', '.svg')){
        ggsave(paste0(fn, ext), plot=p, width=w, height=h)
    }
}
```

1.1.1 Load Caudate data

[3]:
```
# Load counts and phenotype R variable
load("../../input/counts/_m/caudate_brainseq_phase3_hg38_rseGene_merged_n464.
 ↪rda")
### Subset and recode
keepIndex = which((rse_gene$Dx %in% c('Control', "Schizo")) &
                  rse_gene$Race %in% c('CAUC', 'AA'))
rse_gene = rse_gene[, keepIndex]
### Extract phenotypes
pheno_C <- colData(rse_gene) %>% as.data.frame
```

1.1.2 Load DLPFC data

[4]:
```
# Load counts and phenotype R variable
load("../../input/counts/_m/
 ↪dlpfc_ribozero_brainseq_phase2_hg38_rseGene_merged_n453.rda")
### Subset and recode
keepIndex = which((rse_gene$Dx %in% c('Control', "Schizo")) &
                  rse_gene$Race %in% c('CAUC', 'AA'))
rse_gene = rse_gene[, keepIndex]
### Extract phenotypes
pheno_D <- colData(rse_gene) %>% as.data.frame
```

1.1.3 Load Hippocampus data

[5]:
```
# Load counts and phenotype R variable
load("../../input/counts/_m/hippo_brainseq_phase2_hg38_rseGene_merged_n447.rda")
### Subset and recode
keepIndex = which((rse_gene$Dx %in% c('Control', "Schizo")) &
                  rse_gene$Race %in% c('CAUC', 'AA'))
rse_gene = rse_gene[, keepIndex]
### Extract phenotypes
pheno_H <- colData(rse_gene) %>% as.data.frame
```

1.1.4 Load DG data

[6]:
```
# Load counts and phenotype R variable
load("../../input/counts/_m/astellas_dg_hg38_rseGene_n263.rda")
### Subset and recode
keepIndex = which((rse_gene$Dx %in% c('Control', "Schizo")) &
                  rse_gene$Race %in% c('CAUC', 'AA'))
rse_gene = rse_gene[, keepIndex]
### Extract phenotypes
pheno_dg <- colData(rse_gene) %>% as.data.frame
```

1.1.5 Merge data

[7]:
```
allCols <- intersect(intersect(intersect(colnames(pheno_C), colnames(pheno_D)),
                               colnames(pheno_H)),
                     colnames(pheno_dg))
pheno = rbind(pheno_C[, allCols], pheno_D[, allCols],
              pheno_H[, allCols], pheno_dg[, allCols]) %>%
    filter(Age > 13) %>% mutate(Race=gsub("CAUC", "EA", Race))
```

1.2 STRUCTURE analysis

[8]:
```
ancestry = data.table::fread("../../input/ancestry_structure/structure.
↪out_ancestry_proportion_raceDemo_compare")
ancestry %>% head()
```

A data.table: 6 × 4

id <chr>	Afr <dbl>	Eur <dbl>	group <chr>
Br2374	0.007	0.993	CAUC
Br1857	0.001	0.999	CAUC
Br1306	0.759	0.241	AA
Br2605	0.644	0.356	AA
Br1802	0.840	0.160	AA
Br2565	0.005	0.995	CAUC

[9]:
```
ancestry %>% mutate_if(is.character, as.factor) %>%
    group_by(group) %>% summarize(AA=mean(Afr), EA=mean(Eur))
```

A tibble: 2 × 3

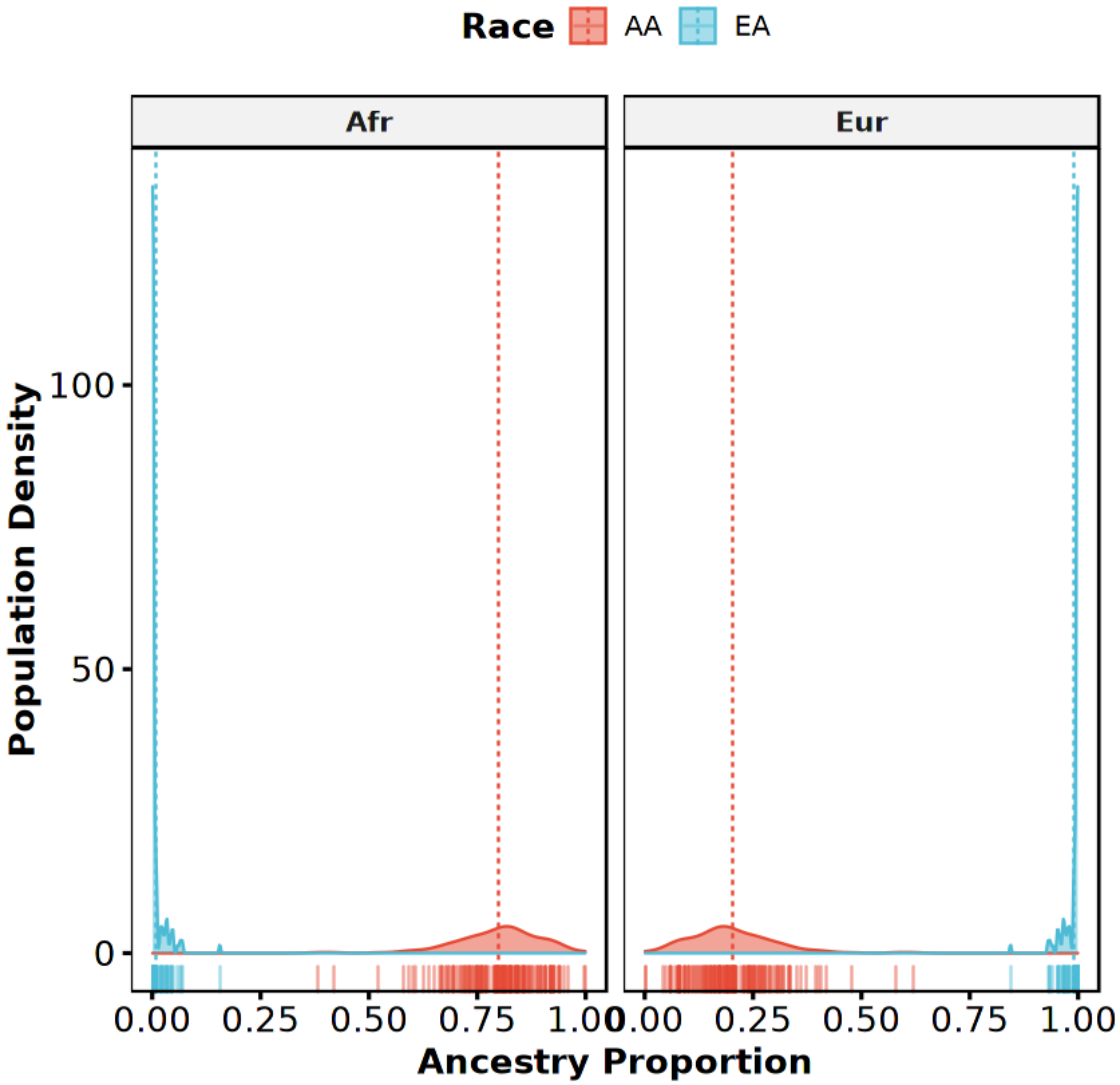
group	AA	EA
<fct>	<dbl>	<dbl>
AA	0.782219451	0.2177805
CAUC	0.007510536	0.9924895

[10]:
```
ancestry %>% inner_join(pheno, by=c("id"="BrNum")) %>%
    filter(Age > 17, Dx == "Control") %>% select(group, Afr, Eur) %>%
    mutate_if(is.character, as.factor) %>% distinct %>%
    group_by(group) %>%
    summarize(AA_mean=mean(Afr), AA_sd=sd(Afr), AA_max=max(Afr),␣
 ↪AA_min=min(Afr),
              EA_mean=mean(Eur), EA_sd=sd(Eur), EA_max=max(Eur),␣
 ↪EA_min=min(Eur))
```

A tibble: 2 × 9

group	AA_mean	AA_sd	AA_max	AA_min	EA_mean	EA_sd	EA_max
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
AA	0.78962609	0.10611682	0.999	0.381	0.2103739	0.10611682	0.619
CAUC	0.03087879	0.02997578	0.156	0.001	0.9691212	0.02997578	0.999

[11]:
```
bxp = ancestry %>% inner_join(pheno, by=c("id"="BrNum")) %>%
    filter(Age > 17, Dx == "Control") %>% select(id, Race, Afr, Eur) %>%
    mutate_if(is.character, as.factor) %>% distinct %>%
    pivot_longer(-c("Race", "id"), names_to="Ancestry", values_to="Proportion")␣
 ↪%>%
    ggdensity(x="Proportion", color="Race", fill="Race", facet.by="Ancestry",
              ncol=2, rug=TRUE, add="mean", palette="npg", ylab="Population␣
 ↪Density",
              xlab="Ancestry Proportion", panel.labs.font=list(face='bold'),
              ggtheme=theme_pubr(base_size=15, border=TRUE)) +
    font("xy.title", face="bold") + font("legend.title", face="bold")
bxp
```



[12]:
```
save_ggplots(bxp, "ancestry_structure_distribution", 10, 5)
```

1.3 eQTL analysis

[13]:
```
pheno %>% dim
```

1. 1334 2. 21

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

```
[1] "There are 509 unique BrNum."
```

[15]:
```
pheno %>% select(BrNum, Region) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Region) %>% count()
```

A grouped_df: 4 × 2

Region <fct>	n <int>
Caudate	400
DentateGyrus	161
DLPFC	378
HIPPO	395

[16]:
```
pheno %>% select(BrNum, Race) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Race) %>% count()
```

A grouped_df: 2 × 2

Race <fct>	n <int>
AA	256
EA	253

[17]:
```
pheno %>% select(BrNum, Race, Region) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Region, Race) %>% count()
```

A grouped_df: 8 × 3

Region <fct>	Race <fct>	n <int>
Caudate	AA	206
Caudate	EA	194
DentateGyrus	AA	78
DentateGyrus	EA	83
DLPFC	AA	204
DLPFC	EA	174
HIPPO	AA	213
HIPPO	EA	182

[18]:
```
pheno %>% select(BrNum, Sex, Region) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Region, Sex) %>% count()
```

A grouped_df: 8 × 3

Region <fct>	Sex <fct>	n <int>
Caudate	F	126
Caudate	M	274
DentateGyrus	F	48
DentateGyrus	M	113
DLPFC	F	121
DLPFC	M	257
HIPPO	F	126
HIPPO	M	269

[19]:
```
pheno %>% group_by(Region) %>%
  summarise_at(vars(c("Age")), list(mean = mean, sd = sd))
```

A tibble: 4 × 3

Region	mean	sd
<chr>	<dbl>	<dbl>
Caudate	49.12390	16.05379
DentateGyrus	50.06770	15.43849
DLPFC	45.83574	16.49445
HIPPO	45.49527	16.41527

[20]:
```
pheno %>% group_by(Region, Race) %>%
  summarise_at(vars(c("Age")), list(mean = mean, sd = sd))
```

A grouped_df: 8 × 4

Region	Race	mean	sd
<chr>	<chr>	<dbl>	<dbl>
Caudate	AA	48.81325	14.49676
Caudate	EA	49.45376	17.58900
DentateGyrus	AA	50.18423	15.53374
DentateGyrus	EA	49.95819	15.44210
DLPFC	AA	46.97896	15.34261
DLPFC	EA	44.49542	17.70090
HIPPO	AA	46.34080	15.61922
HIPPO	EA	44.50571	17.29140

[21]:
```
pheno %>% filter(RIN != "NA") %>% mutate("RIN"=as.numeric(unlist(RIN))) %>%
    group_by(Region) %>% summarise_at(vars(c("RIN")), list(mean = mean, sd =␣
 ↪sd))
```

A tibble: 4 × 3

Region	mean	sd
<chr>	<dbl>	<dbl>
Caudate	7.861000	0.8648983
DentateGyrus	5.208403	1.1871187
DLPFC	7.699471	0.9340876
HIPPO	7.616962	1.0311104

[22]:
```
pheno %>% filter(RIN != "NA") %>% mutate("RIN"=as.numeric(unlist(RIN))) %>%
    group_by(Region, Race) %>% summarise_at(vars(c("RIN")), list(mean = mean,␣
 ↪sd = sd))
```

A grouped_df: 8 × 4

Region	Race	mean	sd
<chr>	<chr>	<dbl>	<dbl>
Caudate	AA	7.859709	0.8416464
Caudate	EA	7.862371	0.8911055
DentateGyrus	AA	5.206349	1.2062837
DentateGyrus	EA	5.210714	1.1760765
DLPFC	AA	7.678922	0.9445184
DLPFC	EA	7.723563	0.9238440
HIPPO	AA	7.604225	1.0509344
HIPPO	EA	7.631868	1.0101014

1.4 Adult individuals for expression related analysis

[23]:
```
pheno = pheno %>% filter(Age > 17, Dx == "Control")
pheno %>% dim
```

1. 785 2. 21

[24]:
```
print(paste("There are", unique(pheno$BrNum) %>% length, "unique BrNum."))
```

```
[1] "There are 292 unique BrNum."
```

[25]:
```
pheno %>% select(BrNum, Region) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Region) %>% count()
```

A grouped_df: 4 × 2

Region <fct>	n <int>
Caudate	240
DentateGyrus	90
DLPFC	212
HIPPO	243

[26]:
```
pheno %>% select(BrNum, Race) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Race) %>% count()
```

A grouped_df: 2 × 2

Race <fct>	n <int>
AA	151
EA	141

[27]:
```
pheno %>% select(BrNum, Race, Region) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Region, Race) %>% count()
```

A grouped_df: 8 × 3

Region <fct>	Race <fct>	n <int>
Caudate	AA	122
Caudate	EA	118
DentateGyrus	AA	47
DentateGyrus	EA	43
DLPFC	AA	123
DLPFC	EA	89
HIPPO	AA	133
HIPPO	EA	110

[28]:
```
pheno %>% select(BrNum, Sex, Region) %>% distinct %>%
    mutate_if(is.character, as.factor) %>%
    group_by(Region, Sex) %>% count()
```

A grouped_df: 8 × 3

Region <fct>	Sex <fct>	n <int>
Caudate	F	71
Caudate	M	169
DentateGyrus	F	26
DentateGyrus	M	64
DLPFC	F	66
DLPFC	M	146
HIPPO	F	74
HIPPO	M	169

[29]:
```
pheno %>% group_by(Region) %>%
  summarise_at(vars(c("Age")), list(mean = mean, sd = sd))
```

A tibble: 4 × 3

Region <chr>	mean <dbl>	sd <dbl>
Caudate	48.31150	15.84692
DentateGyrus	47.88311	15.02380
DLPFC	45.16991	14.76717
HIPPO	44.56724	14.73045

[30]:
```
pheno %>% group_by(Region, Race) %>%
  summarise_at(vars(c("Age")), list(mean = mean, sd = sd))
```

A grouped_df: 8 × 4

Region <chr>	Race <chr>	mean <dbl>	sd <dbl>
Caudate	AA	45.63770	14.72979
Caudate	EA	51.07593	16.53588
DentateGyrus	AA	45.85043	16.32827
DentateGyrus	EA	50.10488	13.28980
DLPFC	AA	44.12511	14.97092
DLPFC	EA	46.61386	14.43996
HIPPO	AA	43.30015	14.73609
HIPPO	EA	46.09927	14.64404

[31]:
```
pheno %>% filter(RIN != "NA") %>% mutate("RIN"=as.numeric(unlist(RIN))) %>%
    group_by(Region) %>% summarise_at(vars(c("RIN")), list(mean = mean, sd =␣
 ↪sd))
```

A tibble: 4 × 3

Region <chr>	mean <dbl>	sd <dbl>
Caudate	7.850000	0.7956997
DentateGyrus	5.315152	1.2186048
DLPFC	7.699057	0.8803807
HIPPO	7.735391	0.9668378

[32]:
```
pheno %>% filter(RIN != "NA") %>% mutate("RIN"=as.numeric(unlist(RIN))) %>%
    group_by(Region, Race) %>% summarise_at(vars(c("RIN")), list(mean = mean,␣
 ↪sd = sd))
```

A grouped_df: 8 × 4

Region <chr>	Race <chr>	mean <dbl>	sd <dbl>
Caudate	AA	7.829508	0.7993477
Caudate	EA	7.871186	0.7947587
DentateGyrus	AA	5.447368	1.2173824
DentateGyrus	EA	5.135714	1.2190507
DLPFC	AA	7.696748	0.8851169
DLPFC	EA	7.702247	0.8787876
HIPPO	AA	7.715038	0.9754173
HIPPO	EA	7.760000	0.9602370

1.5 Reproducibility Information

[33]:
```
Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-08-23 09:58:16 EDT"
```

```
   user  system elapsed
 18.162   1.469  20.029
```

```
 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     2021-08-23

 Packages
 package              * version  date       lib source
 abind                  1.4-5    2016-07-21 [1] CRAN (R 4.0.2)
 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)
 Cairo                  1.5-12.2 2020-07-07 [1] CRAN (R 4.0.2)
 car                    3.0-11   2021-06-27 [1] CRAN (R 4.0.3)
 carData                3.0-4    2020-05-22 [1] CRAN (R 4.0.2)
 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)
curl                  4.3.2    2021-06-23 [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)
farver                2.1.0    2021-02-28 [1] CRAN (R 4.0.3)
forcats             * 0.5.1    2021-01-27 [1] CRAN (R 4.0.2)
foreign               0.8-80   2020-05-24 [2] CRAN (R 4.0.3)
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)
ggpubr              * 0.4.0    2020-06-27 [1] CRAN (R 4.0.2)
ggsci                 2.9      2018-05-14 [1] CRAN (R 4.0.2)
ggsignif              0.6.2    2021-06-14 [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)
labeling              0.4.2    2020-10-20 [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)
openxlsx              4.2.4    2021-06-16 [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)
 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)
 rio                    0.5.27   2021-06-21 [1] CRAN (R 4.0.3)
 rlang                  0.4.11   2021-04-30 [1] CRAN (R 4.0.3)
 rstatix                0.7.0    2021-02-13 [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
 svglite                2.0.0    2021-02-20 [1] CRAN (R 4.0.3)
 systemfonts            1.0.2    2021-05-11 [1] CRAN (R 4.0.3)
 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
 zip                    2.2.0    2021-05-31 [1] CRAN (R 4.0.3)
 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
```