Three sample pairwise sample analysis output

Tomas Bjorklund

Mon Nov 2 23:55:54 2020

This is the final script presenting top candidates and overview plots.

```
suppressPackageStartupMessages(library(knitr))
```

Generation of infective library

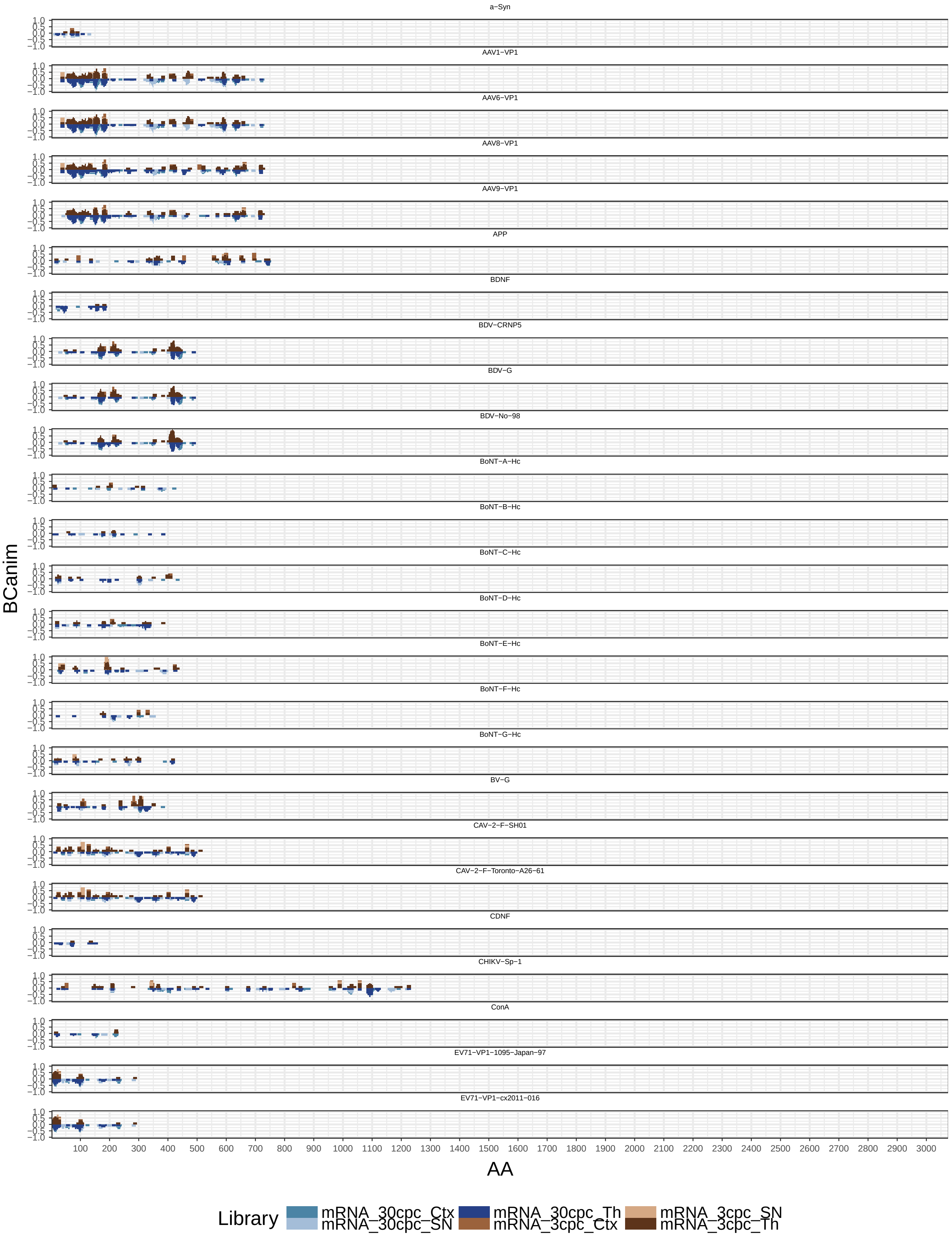
```
all.samples <- readRDS("data/allSamplesDataTable.RDS")

all.samples$Group[all.samples$Group == "mRNA_3cpc_HEK293T"] <- "mRNA_3cpc_HEK293T"
all.samples$Group[all.samples$Group == "mRNA_30cpc_HEK293T"] <- "mRNA_30cpc_HEK293T"
```

Plotting function

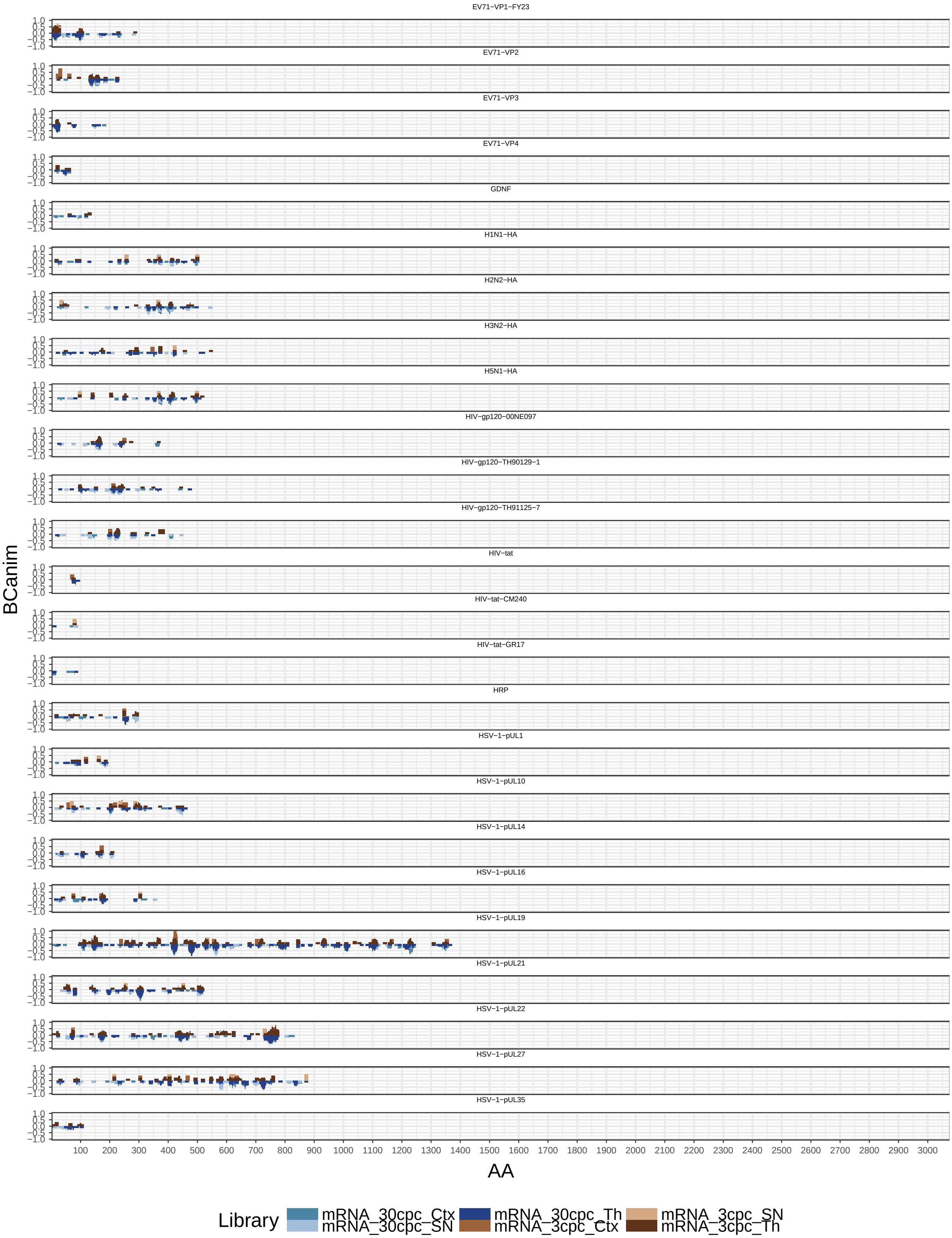
```
# Select samples ===================

topSampleOne <- "mRNA_3cpc_Th"
topSampleTwo <- "mRNA_3cpc_Ctx"
topSampleThree <- "mRNA_3cpc_SN"
bottomSampleOne <- "mRNA_30cpc_Th"
bottomSampleTwo <- "mRNA_30cpc_Ctx"
bottomSampleThree <- "mRNA_30cpc_SN"
filterBC <- FALSE
filterAnimal <- FALSE
AnimaladjustPlot <- FALSE
NormalizePlot <- TRUE
size.bin <- 1
winWidth = 1
PlotBC = TRUE

fill.values <- eval(parse(text = paste("c(", topSampleOne, "= rgb(93,52,27, maxColorValue = 255), ",
    topSampleTwo, "= rgb(155,98,60, maxColorValue = 255), ", topSampleThree,
    "= rgb(213,168,132, maxColorValue = 255), ", bottomSampleOne, "= rgb(38,64,135, maxColorValue = 255),",
    bottomSampleTwo, "= rgb(75,132,165, maxColorValue = 255),", bottomSampleThree,
    "= rgb(164,189,216, maxColorValue = 255))", sep = "")))
setkey(all.samples, Group)
select.samples <- all.samples[J(names(fill.values))]  #Select the six compared groups
select.samples[, `:=`(RNAcount, log2(RNAcount + 1))]

setorder(select.samples, Group, GeneName, start, width)

windowTable <- select.samples[, c("GeneName", "start", "width"), with = FALSE]
windowTable <- unique(windowTable, by = c("GeneName", "start", "width"))
windowTable <- windowTable[, (seq(start, start + width - winWidth)), by = c("GeneName",
    "start", "width")]
setnames(windowTable, "V1", "winStart")
windowTable[, `:=`(winEnd, winStart + winWidth - 1)]
```

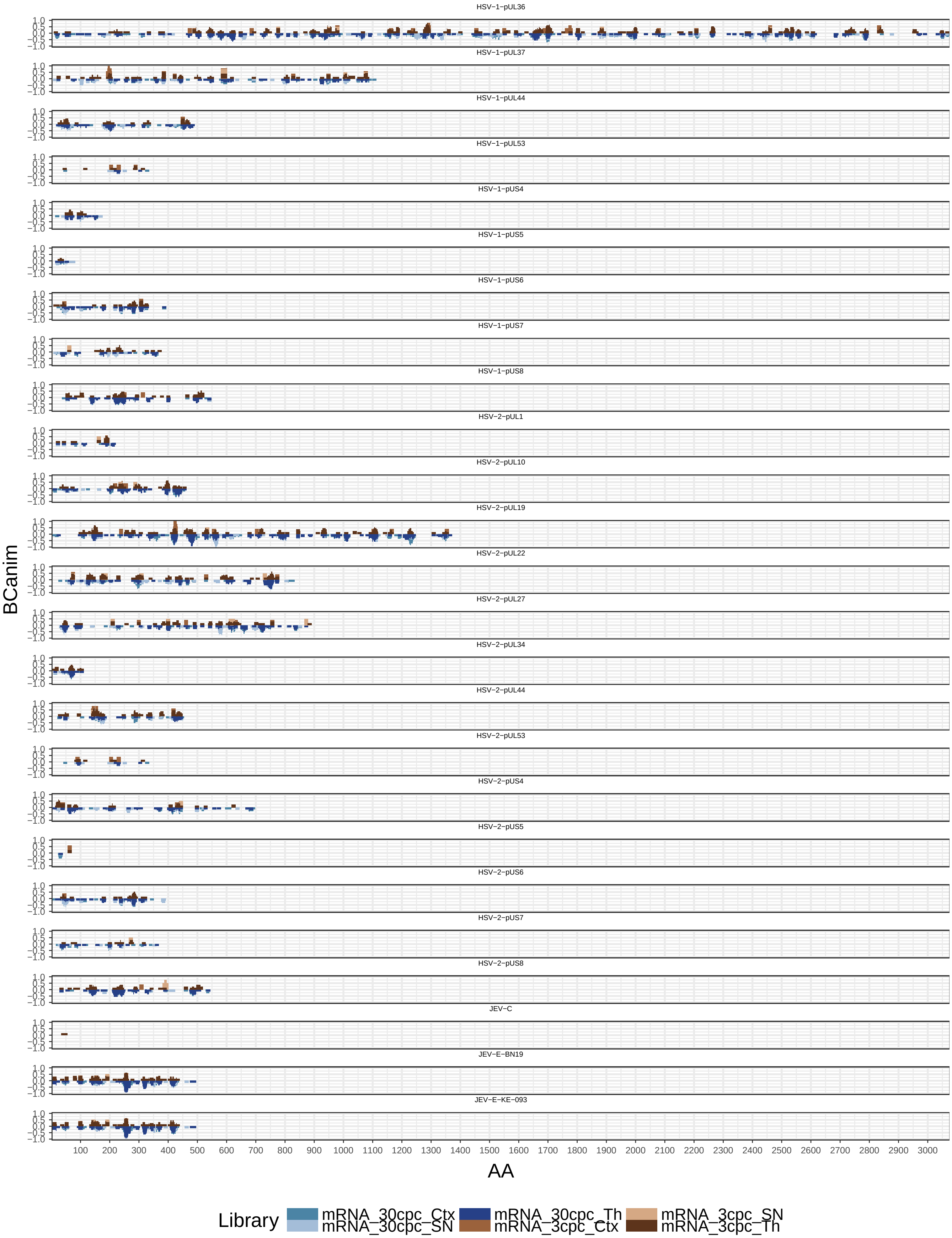
```r
setkeyv(windowTable, c("GeneName", "start", "width"))
setkeyv(select.samples, c("GeneName", "start", "width"))
select.samples.windowBin <- select.samples[windowTable, allow.cartesian = TRUE]

setkeyv(select.samples.windowBin, c("Group", "GeneName", "winStart", "winEnd"))
select.samples.windowBin <- select.samples.windowBin[, list(Overlaps = .N, BC = paste(t(BC),
    collapse = ","), Animals = paste(t(Animals), collapse = ","), LUTnrs = paste(t(LUTnrs),
    collapse = ","), RNAcount = sum(RNAcount)), by = c("Group", "GeneName",
    "winStart", "winEnd", "seqlength")]

plot.data.dt <- unique(select.samples.windowBin, by = c("Group", "GeneName",
    "winStart", "winEnd"))

# =================== Binning of data ===================
FullLength <- max(plot.data.dt$winStart)
position <- seq(0, FullLength, size.bin)
plot.data.dt[, `:=`(bin, findInterval(winStart, position))]

plot.data.bin <- plot.data.dt[, list(.N, seqlength = min(seqlength), BCsum = length(table(strsplit(paste(t(
    collapse = ","), ","))), AA = position[findInterval(mean(winStart), position)],
    AnimalCount = length(table(strsplit(paste(t(Animals), collapse = ","), ","))),
    LUTnrs = paste(unique(names(table(strsplit(paste(t(LUTnrs), collapse = ","),
        ",")))), collapse = ","), NormCount = sum(RNAcount)), by = c("Group",
    "GeneName", "bin")]
plot.data.bin <- unique(plot.data.bin, by = c("Group", "GeneName", "bin"))

plot.data.bin[, `:=`(BCanim, as.double(BCsum + AnimalCount))]



# =================== Filtration parameters ===================

if (NormalizePlot) {
    for (this.name in names(fill.values)) {
        plot.data.bin[plot.data.bin$Group == this.name]$NormCount <- plot.data.bin[plot.data.bin$Group ==
            this.name]$NormCount/max(plot.data.bin[plot.data.bin$Group == this.name]$NormCount)
    }
}


if (PlotBC && NormalizePlot) {
    for (this.name in names(fill.values)) {
        plot.data.bin[plot.data.bin$Group == this.name]$BCanim <- plot.data.bin[plot.data.bin$Group ==
            this.name]$BCanim/max(plot.data.bin[plot.data.bin$Group == this.name]$BCanim)
    }
}

for (this.name in names(fill.values)[seq((length(fill.values)/2) + 1, length(fill.values))]) {
    plot.data.bin[plot.data.bin$Group == this.name]$NormCount <- plot.data.bin[plot.data.bin$Group ==
        this.name]$NormCount * -1  #This line flips the values for the second half of the groups
    plot.data.bin[plot.data.bin$Group == this.name]$BCanim <- plot.data.bin[plot.data.bin$Group ==
        this.name]$BCanim * -1
}
# =================== Output plot ===================

if (PlotBC) {
    outVar <- "BCanim"
} else {
    outVar <- "NormCount"
}

plot.out <- eval(parse(text = paste("ggplot(plot.data.bin,aes(x=AA,y=", outVar,
    ", fill = Group))", sep = "")))
plot.out <- plot.out + geom_bar(stat = "identity", position = "identity") +
    theme_bw() + scale_fill_manual(name = "Library", values = fill.values) +
    scale_colour_manual(name = "Library", values = fill.values) + scale_x_continuous(breaks = c(seq(0,
    3000, 100)), expand = c(0, 0)) + theme(plot.margin = unit(x = c(0, 0, 0,
    0), units = "mm"), legend.position = "bottom", legend.spacing = unit(0,
    "cm"), legend.key.height = unit(0, "cm"), plot.background = element_rect(fill = "white"),
    axis.text = element_text(size = rel(0.45)), axis.ticks = element_line(size = rel(0.5)),
    axis.ticks.length = unit(0.05, "cm"), strip.text.x = element_text(size = rel(0.5),
        colour = "black", angle = 0, lineheight = 0.1, vjust = 0.1), strip.background = element_blank(),
    panel.spacing.y = unit(-0.15, "cm"), panel.spacing.x = unit(0, "cm"))

facet_multiple(plot = plot.out, facets = "GeneName", ncol = 1, nrow = 25)
```

Page 1 *of* 6

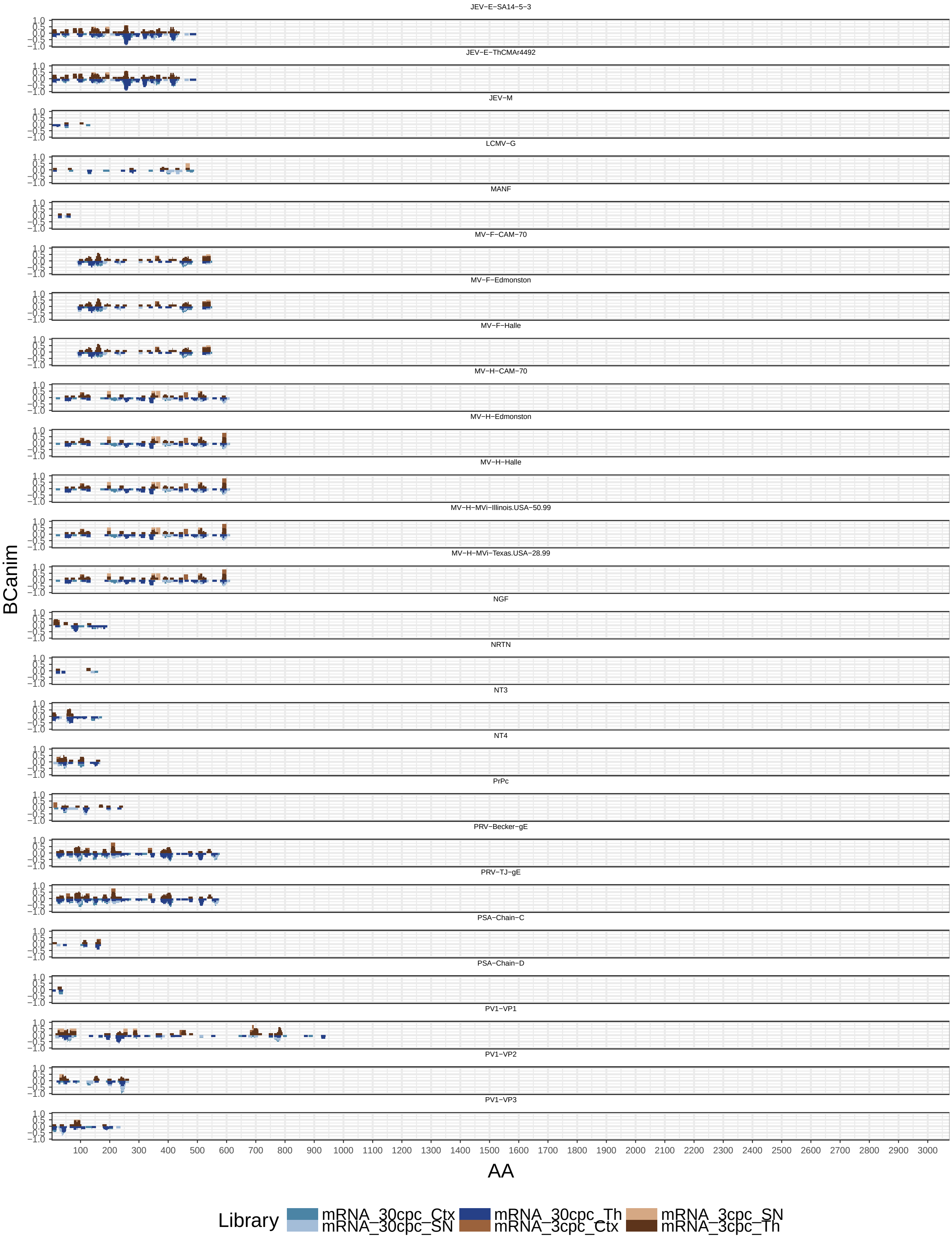

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
1.0
0.5
0.0
−0.5
−1.0
BCanim
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th

Page 2 of 6

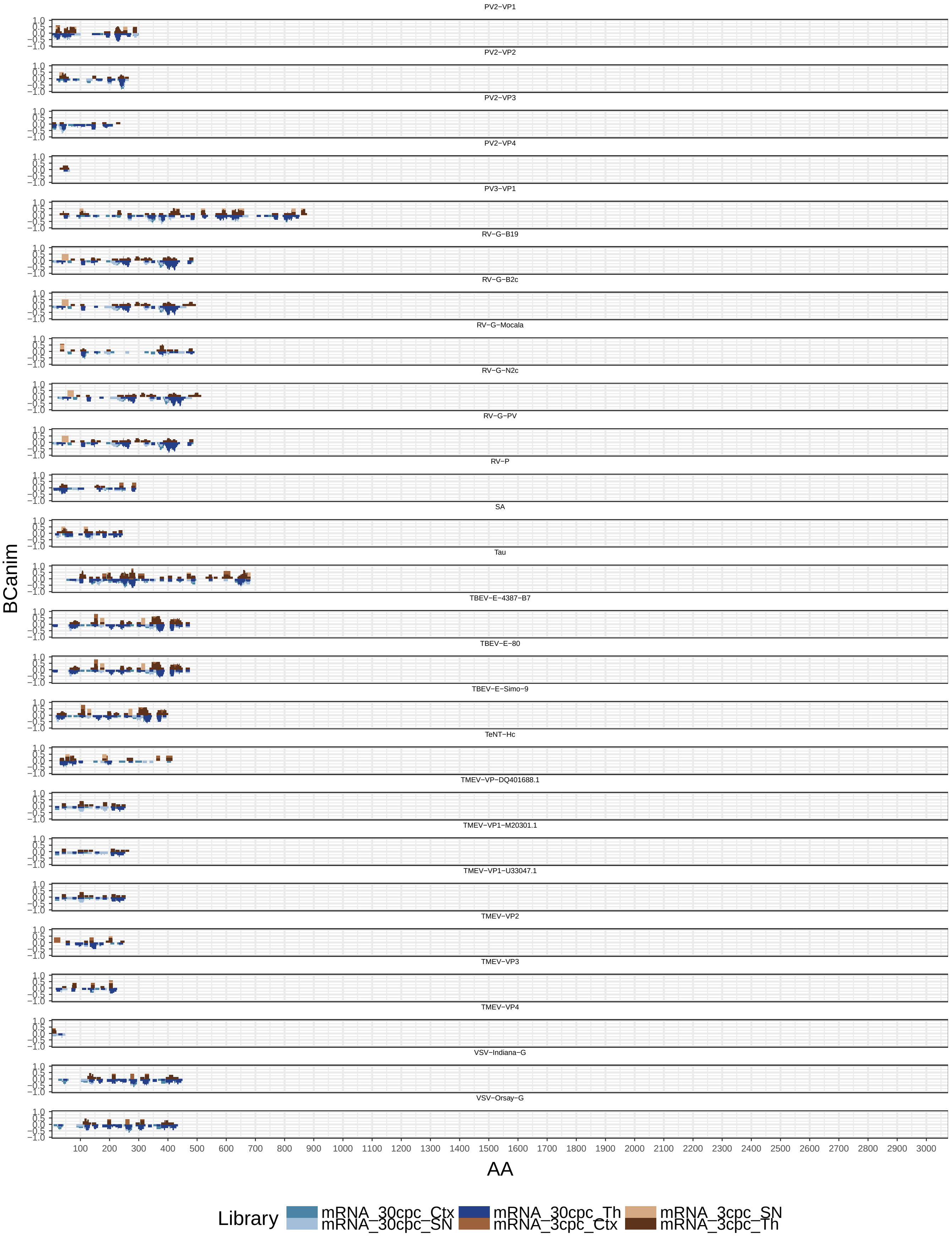

EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
HIV−tat−GR17
HRP
HSV−1−pUL1
HSV−1−pUL10
HSV−1−pUL14
HSV−1−pUL16
HSV−1−pUL19
HSV−1−pUL21
HSV−1−pUL22
HSV−1−pUL27
HSV−1−pUL35
1.0
0.5
0.0
−0.5
−1.0
BCanim
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th


HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
HSV−1−pUS4
HSV−1−pUS5
HSV−1−pUS6
HSV−1−pUS7
HSV−1−pUS8
HSV−2−pUL1
HSV−2−pUL10
HSV−2−pUL19
HSV−2−pUL22
HSV−2−pUL27
HSV−2−pUL34
HSV−2−pUL44
HSV−2−pUL53
HSV−2−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−C
JEV−E−BN19
JEV−E−KE−093
1.0
0.5
0.0
−0.5
−1.0
BCanim
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th

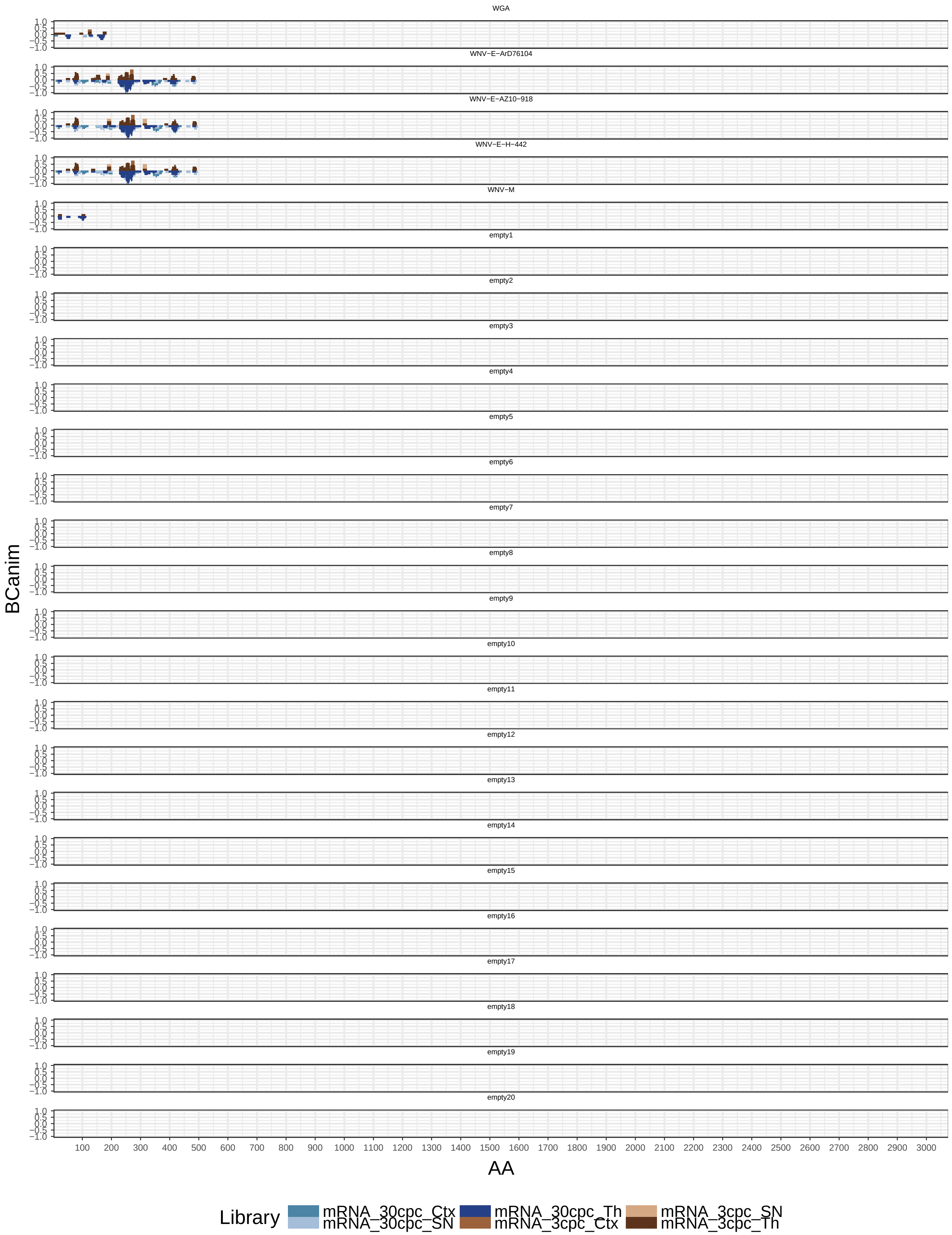
Page 4 *of* 6


JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
MANF
MV−F−CAM−70
MV−F−Edmonston
MV−F−Halle
MV−H−CAM−70
MV−H−Edmonston
MV−H−Halle
MV−H−MVi−Illinois.USA−50.99
MV−H−MVi−Texas.USA−28.99
NGF
NRTN
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PSA−Chain−C
PSA−Chain−D
PV1−VP1
PV1−VP2
PV1−VP3
1.0
0.5
0.0
−0.5
−1.0
BCanim
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th

Page 5 *of* 6


PV2−VP1
PV2−VP2
PV2−VP3
PV2−VP4
PV3−VP1
RV−G−B19
RV−G−B2c
RV−G−Mocala
RV−G−N2c
RV−G−PV
RV−P
SA
Tau
TBEV−E−4387−B7
TBEV−E−80
TBEV−E−Simo−9
TeNT−Hc
TMEV−VP−DQ401688.1
TMEV−VP1−M20301.1
TMEV−VP1−U33047.1
TMEV−VP2
TMEV−VP3
TMEV−VP4
VSV−Indiana−G
VSV−Orsay−G
1.0
0.5
0.0
−0.5
−1.0
BCanim
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th

Page 6 *of* 6


WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
empty1
empty2
empty3
empty4
empty5
empty6
empty7
empty8
empty9
empty10
empty11
empty12
empty13
empty14
empty15
empty16
empty17
empty18
empty19
empty20
1.0
0.5
0.0
−0.5
−1.0
BCanim
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th

```r
devtools::session_info()
```

```
Session info -------------------------------------------------------------

 setting  value
 version  R version 3.4.2 (2017-09-28)
 system   x86_64, linux-gnu
 ui       X11
 language (EN)
 collate  en_US.UTF-8
 tz       UTC
 date     2020-11-03

Packages -----------------------------------------------------------------

 package                * version  date
 acepack                  1.4.1    2016-10-29
 AnnotationDbi            1.38.2   2017-11-29
 AnnotationFilter         1.0.0    2017-11-29
 AnnotationHub            2.8.3    2017-11-29
 backports                1.1.1    2017-09-25
 base                   * 3.4.2    2017-10-06
 base64enc                0.1-3    2015-07-28
 Biobase                * 2.36.2   2017-11-29
 BiocGenerics           * 0.22.1   2017-11-29
 BiocInstaller            1.26.1   2017-10-10
 BiocParallel             1.10.1   2017-11-29
 biomaRt                  2.32.1   2017-11-29
 Biostrings             * 2.44.2   2017-11-29
 biovizBase               1.24.0   2017-11-29
 bit                      1.1-12   2014-04-09
 bit64                    0.9-7    2017-05-08
 bitops                   1.0-6    2013-08-17
 blob                     1.1.0    2017-06-17
 BSgenome                 1.44.2   2017-11-29
 checkmate                1.8.4    2017-09-25
 cluster                  2.0.6    2017-03-16
 codetools                0.2-15   2016-10-05
 colorspace               1.3-2    2016-12-14
 compiler                 3.4.2    2017-10-06
 curl                     2.8.1    2017-07-21
 data.table             * 1.10.4-2 2017-10-12
 datasets               * 3.4.2    2017-10-06
 DBI                      0.7      2017-06-18
 DelayedArray           * 0.2.7    2017-11-29
 devtools               * 1.13.3   2017-08-02
 dichromat                2.0-0    2013-01-24
 digest                   0.6.12   2017-01-27
 doParallel             * 1.0.11   2017-09-28
 ensembldb                2.0.4    2017-11-29
 evaluate                 0.10.1   2017-06-24
 foreach                * 1.4.3    2015-10-13
 foreign                  0.8-69   2017-06-21
 formatR                  1.5      2017-04-25
 Formula                  1.2-2    2017-07-10
 GenomeInfoDb           * 1.12.3   2017-11-29
 GenomeInfoDbData         0.99.0   2017-11-29
 GenomicAlignments      * 1.12.2   2017-11-29
 GenomicFeatures          1.28.5   2017-11-29
```

```
GenomicRanges          * 1.28.6   2017-11-29
GGally                   1.3.2    2017-08-02
ggbio                  * 1.24.1   2017-11-29
ggplot2                * 2.2.1    2016-12-30
ggplus                 * 0.1      2017-11-29
graph                    1.54.0   2017-11-29
graphics               * 3.4.2    2017-10-06
grDevices              * 3.4.2    2017-10-06
grid                   * 3.4.2    2017-10-06
gridExtra                2.3      2017-09-09
gtable                   0.2.0    2016-02-26
Hmisc                    4.0-3    2017-05-02
hms                      0.3      2016-11-22
htmlTable                1.9      2017-01-26
htmltools                0.3.6    2017-04-28
htmlwidgets              0.9      2017-07-10
httpuv                   1.3.5    2017-07-04
httr                     1.3.1    2017-08-20
interactiveDisplayBase   1.14.0   2017-11-29
IRanges                * 2.10.5   2017-11-29
iterators              * 1.0.8    2015-10-13
kableExtra             * 0.5.2    2017-09-15
knitr                  * 1.17     2017-08-10
labeling                 0.3      2014-08-23
lattice                  0.20-35  2017-03-25
latticeExtra             0.6-28   2016-02-09
lazyeval                 0.2.0    2016-06-12
magrittr                 1.5      2014-11-22
Matrix                   1.2-11   2017-08-21
matrixStats            * 0.52.2   2017-04-14
memoise                  1.1.0    2017-04-21
methods                * 3.4.2    2017-10-06
mime                     0.5      2016-07-07
munsell                  0.4.3    2016-02-13
nnet                     7.3-12   2016-02-02
OrganismDbi              1.18.1   2017-11-29
parallel               * 3.4.2    2017-10-06
plyr                   * 1.8.4    2016-06-08
ProtGenerics             1.8.0    2017-11-29
R6                       2.2.2    2017-06-17
RBGL                     1.52.0   2017-11-29
RColorBrewer             1.1-2    2014-12-07
Rcpp                     0.12.13  2017-09-28
RCurl                    1.95-4.8 2016-03-01
readr                    1.1.1    2017-05-16
reshape                  0.8.7    2017-08-06
reshape2                 1.4.2    2016-10-22
rlang                    0.1.2    2017-08-09
rmarkdown                1.6      2017-06-15
rpart                    4.1-11   2017-04-21
rprojroot                1.2      2017-01-16
Rsamtools              * 1.28.0   2017-11-29
RSQLite                  2.0      2017-06-19
rtracklayer              1.36.6   2017-11-29
rvest                    0.3.2    2016-06-17
S4Vectors              * 0.14.7   2017-11-29
scales                   0.5.0    2017-08-24
```

```
shiny                    1.0.5    2017-08-23
splines                  3.4.2    2017-10-06
stats                  * 3.4.2    2017-10-06
stats4                 * 3.4.2    2017-10-06
stringi                  1.1.5    2017-04-07
stringr                  1.2.0    2017-02-18
SummarizedExperiment   * 1.6.5    2017-11-29
survival                 2.41-3   2017-04-04
tibble                   1.3.4    2017-08-22
tools                    3.4.2    2017-10-06
utils                  * 3.4.2    2017-10-06
VariantAnnotation        1.22.3   2017-11-29
withr                    2.0.0    2017-07-28
XML                      3.98-1.9 2017-06-19
xml2                     1.1.1    2017-01-24
xtable                   1.8-2    2016-02-05
XVector                * 0.16.0   2017-11-29
yaml                     2.1.14   2016-11-12
zlibbioc                 1.22.0   2017-11-29
source
CRAN (R 3.4.2)
Bioconductor
Bioconductor
Bioconductor
CRAN (R 3.4.2)
local
CRAN (R 3.4.2)
Bioconductor
Bioconductor
Bioconductor
Bioconductor
Bioconductor
Bioconductor
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
local
CRAN (R 3.4.2)
url
local
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
Bioconductor
Bioconductor
Bioconductor
Bioconductor
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
Github (guiastrennec/ggplus@10ca057)
Bioconductor
local
local
local
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
Bioconductor
CRAN (R 3.4.2)
url
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
url
CRAN (R 3.4.2)
CRAN (R 3.4.2)
local
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
local
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
url
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
url
CRAN (R 3.4.2)
CRAN (R 3.4.2)
```

```
Bioconductor
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
local
local
local
url
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
local
local
Bioconductor
url
CRAN (R 3.4.2)
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
Bioconductor
```