# Custom array sequence generation

*Tomas Bjorklund*

*Wed Nov 1 10:28:10 2017*

This script generates all unique AA sequences for the CustomArray production

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

## Loading source files

```
source(file.path("functions", "AAtoDNA.R"))
source(file.path("functions", "GeneCodon.R"))
# Override the GeneCodon function with local version containing human codons
unlockBinding("GeneCodon", as.environment("package:GeneGA"))
assign("GeneCodon", GeneCodon, as.environment("package:GeneGA"))

allSequences <- readFasta("input/DNA-lib_RetrogradeTransport.fasta")
AAlist <- data.frame(Class = character(), Family = character(), Strain = character(),
    Note = character(), Number = character(), Name = character(), AAfragment = character(),
    stringsAsFactors = FALSE)

# allSequences <- allSequences[124:129] #Debug row
```

## Generation of AA table for the selected proteins

```
strt <- Sys.time()
for (i in 1:length(allSequences)) {
    thisID <- as.character(ShortRead::id(allSequences[i]))
    thisSeq <- sread(allSequences[i])
    thisAA <- Biostrings::translate(thisSeq, genetic.code = GENETIC_CODE, if.fuzzy.codon = "solve")
    AAlist[i, c("Class", "Family", "Strain", "Note", "Number", "Name", "AAfragment")] <- c(BBmisc::explode(t
        sep = ","), as.character(thisAA))
}
```

## The generateFragments function

```
generateFragments <- function(minLength, maxLength, frequency = 1) {
    # Generate a table to store all AA sequences
    fragList <- data.frame(Class = character(), Family = character(), Strain = character(),
        Note = character(), Number = character(), Name = character(), AAstart = integer(),
        AAstop = integer(), AAfragment = character(), stringsAsFactors = FALSE)


    makeAllFrags <- function(k) {
        thisFullAA <- AAlist[k, "AAfragment"]
        count = length(fragList[, 1]) + 1
        for (l in seq(1, (width(thisFullAA) - minLength), frequency)) {
            for (m in (l + minLength - 1):(min(l + maxLength - 1, width(thisFullAA)))) {
```

```
                # Truncate string to the relevant fragment:
                thisFragment <- substr(thisFullAA, l, m)
                # Take away any sequence that starts with a start codon ATG:
                if (substr(thisFragment, 1, 1) != "M") {
                  fragList[count, c("Class", "Family", "Strain", "Note", "Number",
                    "Name", "AAstart", "AAstop", "AAfragment")] <- c(AAlist[k,
                    c("Class", "Family", "Strain", "Note", "Number", "Name")],
                    l, m, thisFragment)
                  ## Inserts the fragment with information into the new data frame
                  count <- count + 1
                }
            }
        }
        return(fragList)
    }

    fragList <- do.call(rbind, mclapply(1:length(AAlist[, 1]), makeAllFrags,
        mc.preschedule = TRUE, mc.cores = detectCores()))

    # Control if any sequences contain non-AA characters and save them into a
    # separate list
    discardList <- fragList[grep("[[:punct:]|X]", fragList[, "AAfragment"]),
        ]

    # Remove any sequence containing non AA characters
    fragList <- fragList[grep("[[:punct:]|X]", fragList[, "AAfragment"], invert = TRUE),
        ]


    # Sort the fragments, find unique strings and count number of duplicates
    sortedFragments <- rev(sort(table(fragList[, "AAfragment"])))

    # Run the AAtoDNA function to convert all AA sequences to human
    # codon-optimized DNA sequences
    row.names(sortedFragments) <- mclapply(row.names(sortedFragments), fullOPT = FALSE,
        species = "hsa", AAtoDNA, mc.preschedule = TRUE, mc.set.seed = TRUE,
        mc.silent = FALSE, mc.cores = detectCores(), mc.cleanup = TRUE)  #

    sortedFragments <- sortedFragments[order(row.names(sortedFragments))]

    return(sortedFragments)
}
```

## Execution of the function

```
sortedFragments.14aa <- generateFragments(14, 14, 1)
```

Add the overhangs for amplication PCR and Gibson assembly into the AAV plasmid

```
fivePrime <- tolower("AACCTCCAGAGAGGCAACGCT")
threePrime <- tolower("GCCAGACAAGCAGCTACCGCA")
row.names(sortedFragments.14aa) <- paste(fivePrime, row.names(sortedFragments.14aa),
    threePrime, sep = "")

sortedFragments.14aa.G4S <- generateFragments(14, 14, 3)
sortedFragments.14aa.A5 <- sortedFragments.14aa.G4S
```

Add the overhangs including G4S spacers for amplication PCR and Gibson assembly into the AAV plasmid

```
fivePrime <- tolower("AACCTCCAGAGAGGCAACGGAGGCGGAGGAAGT")
threePrime <- tolower("GGAGGCGGCGGAAGCAGACAAGCAGCTACCGCA")
row.names(sortedFragments.14aa.G4S) <- paste(fivePrime, row.names(sortedFragments.14aa.G4S),
    threePrime, sep = "")
```

Add the overhangs including A5 spacers for amplication PCR and Gibson assembly into the AAV plasmid

```
fivePrime <- tolower("AACCTCCAGAGAGGCAACGCTGCTGCAGCAGCC")
threePrime <- tolower("GCAGCTGCAGCTGCCAGACAAGCAGCTACCGCA")
row.names(sortedFragments.14aa.A5) <- paste(fivePrime, row.names(sortedFragments.14aa.A5),
    threePrime, sep = "")
```

Generate 22aa fragments

```
sortedFragments.22aa <- generateFragments(22, 22, 3)
```

Add the overhangs for amplication PCR and Gibson assembly into the AAV plasmid

```
fivePrime <- tolower("AACCTCCAGAGAGGCAACGCT")
threePrime <- tolower("GCCAGACAAGCAGCTACCGCA")
row.names(sortedFragments.22aa) <- paste(fivePrime, row.names(sortedFragments.22aa),
    threePrime, sep = "")
```

Merge all separate fragment lists into one complete list

```
sortedFragments <- c(sortedFragments.22aa, sortedFragments.14aa, sortedFragments.14aa.A5,
    sortedFragments.14aa.G4S)

print(paste("Number of unique fragments:", length(unique(names(sortedFragments))),
    sep = " "))
```

```
[1] "Number of unique fragments: 92343"
```

```
write.table(c("Sequence", unique(names(sortedFragments))), "data/SortedFragments_all.txt",
    row.names = F, col.names = F, quote = F, sep = "\t")

print(Sys.time() - strt)
```

```
Time difference of 2.421197 mins
```

```
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     2017-11-01
```

```
Packages -----------------------------------------------------------------
```

```
 package             * version  date       source
 acepack               1.4.1    2016-10-29 CRAN (R 3.4.2)
 ade4                  1.7-8    2017-08-09 CRAN (R 3.4.2)
 backports             1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                * 3.4.2    2017-10-06 local
 base64enc             0.1-3    2015-07-28 CRAN (R 3.4.2)
```

```
BBmisc                1.11     2017-03-10 CRAN (R 3.4.2)
Biobase             * 2.36.2   2017-10-13 Bioconductor
BiocGenerics        * 0.22.1   2017-10-13 Bioconductor
BiocParallel        * 1.10.1   2017-10-13 Bioconductor
Biostrings          * 2.44.2   2017-10-13 Bioconductor
bitops                1.0-6    2013-08-17 CRAN (R 3.4.2)
checkmate             1.8.4    2017-09-25 CRAN (R 3.4.2)
cluster               2.0.6    2017-03-16 CRAN (R 3.4.2)
colorspace            1.3-2    2016-12-14 CRAN (R 3.4.2)
compiler              3.4.2    2017-10-06 local
data.table            1.10.4-2 2017-10-12 CRAN (R 3.4.2)
datasets            * 3.4.2    2017-10-06 local
DelayedArray        * 0.2.7    2017-10-13 Bioconductor
devtools            * 1.13.3   2017-08-02 CRAN (R 3.4.2)
digest                0.6.12   2017-01-27 CRAN (R 3.4.2)
evaluate              0.10.1   2017-06-24 CRAN (R 3.4.2)
foreign               0.8-69   2017-06-21 CRAN (R 3.4.2)
formatR               1.5      2017-04-25 CRAN (R 3.4.2)
Formula             * 1.2-2    2017-07-10 CRAN (R 3.4.2)
GeneGA              * 1.26.0   2017-10-13 Bioconductor
GenomeInfoDb        * 1.12.3   2017-10-13 Bioconductor
GenomeInfoDbData      0.99.0   2017-10-13 Bioconductor
GenomicAlignments   * 1.12.2   2017-10-13 Bioconductor
GenomicRanges       * 1.28.6   2017-10-13 Bioconductor
ggplot2             * 2.2.1    2016-12-30 CRAN (R 3.4.2)
graphics            * 3.4.2    2017-10-06 local
grDevices           * 3.4.2    2017-10-06 local
grid                  3.4.2    2017-10-06 local
gridExtra             2.3      2017-09-09 CRAN (R 3.4.2)
gtable                0.2.0    2016-02-26 CRAN (R 3.4.2)
hash                * 2.2.6    2013-02-21 CRAN (R 3.4.2)
Hmisc               * 4.0-3    2017-05-02 CRAN (R 3.4.2)
hms                   0.3      2016-11-22 CRAN (R 3.4.2)
htmlTable             1.9      2017-01-26 CRAN (R 3.4.2)
htmltools             0.3.6    2017-04-28 CRAN (R 3.4.2)
htmlwidgets           0.9      2017-07-10 CRAN (R 3.4.2)
httr                  1.3.1    2017-08-20 CRAN (R 3.4.2)
hwriter               1.3.2    2014-09-10 CRAN (R 3.4.2)
IRanges             * 2.10.5   2017-10-13 Bioconductor
kableExtra          * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr               * 1.17     2017-08-10 CRAN (R 3.4.2)
lattice             * 0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra          0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval              0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr              1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats         * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise               1.1.0    2017-04-21 CRAN (R 3.4.2)
methods             * 3.4.2    2017-10-06 local
munsell               0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                  7.3-12   2016-02-02 CRAN (R 3.4.2)
parallel            * 3.4.2    2017-10-06 local
plyr                  1.8.4    2016-06-08 CRAN (R 3.4.2)
R6                    2.2.2    2017-06-17 CRAN (R 3.4.2)
RColorBrewer          1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                  0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                 1.95-4.8 2016-03-01 CRAN (R 3.4.2)
```

```
readr                   1.1.1    2017-05-16 CRAN (R 3.4.2)
rlang                   0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown               1.6      2017-06-15 CRAN (R 3.4.2)
rpart                   4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot               1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools             * 1.28.0   2017-10-13 Bioconductor
rvest                   0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors             * 0.14.7   2017-10-13 Bioconductor
scales                  0.5.0    2017-08-24 CRAN (R 3.4.2)
seqinr                * 3.4-5    2017-08-01 CRAN (R 3.4.2)
ShortRead             * 1.34.2   2017-10-13 Bioconductor
splines                 3.4.2    2017-10-06 local
stats                 * 3.4.2    2017-10-06 local
stats4                * 3.4.2    2017-10-06 local
stringi                 1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                 1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment * 1.6.5     2017-10-13 Bioconductor
survival              * 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                  1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                   3.4.2    2017-10-06 local
utils                 * 3.4.2    2017-10-06 local
withr                   2.0.0    2017-07-28 CRAN (R 3.4.2)
xml2                    1.1.1    2017-01-24 CRAN (R 3.4.2)
XVector               * 0.16.0   2017-10-13 Bioconductor
yaml                    2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc                1.22.0   2017-10-13 Bioconductor
```

# Extraction of Barcodes and gene fragments

*Tomas Bjorklund*

*Wed Nov 1 10:30:45 2017*

This workflow extracts barcodes and the gene fragments synthesized with the CustomArray using bbmap2. The fragments are then suitable for alignment to reference sequences using blastn.

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

## Sequencing files

```
dataDir <- config$Value[1]
in.name.P5 <- file.path(dataDir, config$Value[2])
in.name.P7 <- file.path(dataDir, config$Value[3])
name.out <- config$Value[4]
paired.alignment <- as.logical(config$Value[5])
```

## Analysis parameters

```
knitr::kable(config, format = "latex", booktabs = T) %>% kable_styling(latex_options = "striped")
```

| Parameter | Value |
|---|---|
| dataDir | seqFiles |
| in.name.P5 | DNA_pscAAVlib_R1.fastq.gz |
| in.name.P7 | DNA_pscAAVlib_R2.fastq.gz |
| name.out | AAVlibrary_complete |
| paired.alignment | TRUE |
| run.subset | FALSE |
| max.cores | 32 |
| subset.count | 250000 |

```
run.subset <- as.logical(config$Value[6])
max.cores <- as.integer(config$Value[7])
subset.count <- as.integer(config$Value[8])

strt <- Sys.time()
```

## Selection of real amplicons

```
# This section searches the sequencing file and only select the files with
# valid amplicons
out.name.P5 <- tempfile(pattern = "P5_", tmpdir = tempdir(), fileext = ".fastq.gz")
out.name.P7 <- tempfile(pattern = "P7_", tmpdir = tempdir(), fileext = ".fastq.gz")
command.args <- paste("overwrite=true k=15 rcomp=f skipr2=t qhdist=0 maskmiddle=f",
    " hammingdistance=2 findbestmatch=f ordered=t threads=", detectCores(),
    " in=", in.name.P5, " in2=", in.name.P7, " outm=", out.name.P5, " outm2=",
```

```
    out.name.P7, " fliteral=", "GTATGTTGTTCTGGAGCGGGAGGGTGCTATTTTGCCTAGCGATAA",
    sep = "")

sys.out <- system2(path.expand("~/bbmap/bbduk2.sh"), args = command.args, stdout = TRUE,
    stderr = TRUE)

sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("bbduk2 Identification of real amplicons")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[3:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

| | bbduk2 Identification of real amplicons |
|---|---|
| 3 | |
| 4 | BBDuk2 version 37.02 |
| 5 | Set ORDERED to true |
| 6 | Set threads to 32 |
| 7 | k=15 |
| 8 | hamming distance=2 |
| 9 | kfiltering using 1 literal. |
| 10 | |
| 11 | Initial: |
| 12 | Memory: max=77002m, free=74592m, used=2410m |
| 13 | |
| 14 | Added 30721 kmers; time: 0.072 seconds. |
| 15 | Memory: max=77002m, free=71779m, used=5223m |
| 16 | |
| 17 | Input is being processed as paired |
| 18 | Started output streams: 0.130 seconds. |
| 19 | Processing time: 108.946 seconds. |
| 20 | |
| 21 | Input: 23191088 reads 3490215687 bases. |
| 22 | Contaminants: 23095890 reads (99.59%) 3475908371 bases (99.59%) |
| 23 | Total Removed: 23095890 reads (99.59%) 3475908371 bases (99.59%) |
| 24 | Result: 95198 reads (0.41%) 14307316 bases (0.41%) |
| 25 | |
| 26 | Time: 109.162 seconds. |
| 27 | Reads Processed: 23191k 212.45k reads/sec |
| 28 | Bases Processed: 3490m 31.97m bases/sec |

```
in.name.P5 <- out.name.P5
in.name.P7 <- out.name.P7
```

## Extraction of a subset

```
if (run.subset) {
    suppressWarnings(sampler <- FastqSampler(gsub("([\\])", "", in.name.P5),
        subset.count, readerBlockSize = 1e+09, ordered = TRUE))
    set.seed(123)
    tmp.P5 <- yield(sampler)
    in.name.P5 <- tempfile(pattern = "P5_", tmpdir = tempdir(), fileext = ".fastq.gz")
```

```
    writeFastq(tmp.P5, in.name.P5, compress = TRUE)
    rm(tmp.P5)
    suppressWarnings(sampler <- FastqSampler(gsub("([\\])", "", in.name.P7),
        subset.count, readerBlockSize = 1e+09, ordered = TRUE))
    set.seed(123)
    tmp.P7 <- yield(sampler)
    in.name.P7 <- tempfile(pattern = "P7_", tmpdir = tempdir(), fileext = ".fastq.gz")
    writeFastq(tmp.P7, in.name.P7, compress = TRUE)
    rm(tmp.P7)
}

output.Reads <- as.integer(system(paste("gunzip -c ", shQuote(gsub("([\\])",
    "", in.name.P5)), " | echo $((`wc -l`/4)) 2>&1", sep = ""), intern = TRUE,
    ignore.stdout = FALSE))  #Stores the read count utilized
print(paste("Utilized sequences:", output.Reads))
```

```
[1] "Utilized sequences: 11547945"
```

## Extraction of barcodes

```
out.name.P5 <- tempfile(pattern = "BC_", tmpdir = tempdir(), fileext = ".fastq.gz")

sys.out <- system(paste("~/bbmap/bbduk2.sh overwrite=true k=18 mink=18 hammingdistance=2 findbestmatch=t ",
    "rcomp=f findbestmatch=f qhdist=1 minavgquality=0 maxns=0 minlength=18 ",
    "maxlength=22 threads=", detectCores(), " in=", shQuote(in.name.P5), " out=",
    out.name.P5, " lliteral=", "GGCCTAGCGGCCGCTTTACTT", " rliteral=", "ATAACTTCGTATAATGTATGC",
    " 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)
sys.out <- as.data.frame(sys.out)

in.name.P5 <- out.name.P5


colnames(sys.out) <- c("bbduk2 Extraction of barcodes")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[3:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

```
rm(sys.out)

reads.BC <- readFastq(in.name.P5)
sread(reads.BC)
```

```
  A DNAStringSet instance of length 11374106
            width seq
       [1]    20 AAGTAATTGTGGCAGGAAGT
       [2]    20 GTTGGCGTGTGTCCTTATCC
       [3]    20 GCGCCAATGCGTGTAGCTGT
       [4]    20 CTTGATTGCTTGATGGCCTC
       [5]    20 GAATATTGGTGCATTCCCTT
       ...   ... ...
[11374102]    20 GCGCGTTGATGGGTTGGCTC
[11374103]    20 CAGGCCTGAATGGCGGGCGG
[11374104]    20 GTTTGCTAGTTCCCGGGTAC
[11374105]    20 CAGGGCGGGTGTGCAGGCGG
[11374106]    20 GTACGTTTCCGTCCATATTG
```

| | bbduk2 Extraction of barcodes |
|---|---|
| 3 | |
| 4 | BBDuk2 version 37.02 |
| 5 | Set threads to 32 |
| 6 | k=18 |
| 7 | maskMiddle=true |
| 8 | hamming distance=2 |
| 9 | right-ktrimming using 1 literal. |
| 10 | left-ktrimming using 1 literal. |
| 11 | |
| 12 | Initial: |
| 13 | Memory: max=76982m, free=73767m, used=3215m |
| 14 | |
| 15 | Added 5104 kmers; time: 0.044 seconds. |
| 16 | Memory: max=76982m, free=70955m, used=6027m |
| 17 | |
| 18 | Added 5104 kmers; time: 0.010 seconds. |
| 19 | Memory: max=76982m, free=70955m, used=6027m |
| 20 | |
| 21 | Input is being processed as unpaired |
| 22 | Started output streams: 0.082 seconds. |
| 23 | Processing time: 212.211 seconds. |
| 24 | |
| 25 | Input: 11547945 reads 1738273591 bases. |
| 26 | KTrimmed: 23030031 reads (199.43%) 1505373909 bases (86.60%) |
| 27 | Low quality discards: 6 reads (0.00%) 122 bases (0.00%) |
| 28 | Total Removed: 173839 reads (1.51%) 1511015923 bases (86.93%) |
| 29 | Result: 11374106 reads (98.49%) 227257668 bases (13.07%) |
| 30 | |
| 31 | Time: 212.365 seconds. |
| 32 | Reads Processed: 11547k 54.38k reads/sec |
| 33 | Bases Processed: 1738m 8.19m bases/sec |

```
(unique.BCs <- unique(sread(reads.BC)))
```

```
  A DNAStringSet instance of length 3934570
          width seq
      [1]    20 AAGTAATTGTGGCAGGAAGT
      [2]    20 GTTGGCGTGTGTCCTTATCC
      [3]    20 GCGCCAATGCGTGTAGCTGT
      [4]    20 CTTGATTGCTTGATGGCCTC
      [5]    20 GAATATTGGTGCATTCCCTT
      ...   ... ...
[3934566]    20 GCTCCCGGGAAGCTTCCCGT
[3934567]    20 AAATACTGGCTGATAACCTG
[3934568]    20 GCATCCTTATTTCATGCTTT
[3934569]    20 GCGCGCTGATGTGTTCGCGG
[3934570]    20 GTTTGCTAGTTCCCGGGTAC
```

```
output.BCs <- length(unique.BCs)
print(paste("Utilized barcodes:", output.BCs))
```

```
[1] "Utilized barcodes: 3934570"
```

```
barcodeTable <- data.table(ID = as.character(ShortRead::id(reads.BC)), BC = as.character(sread(reads.BC)))
```

## Extraction of fragments

```
out.name.P7 <- tempfile(pattern = "P7_", tmpdir = tempdir(), fileext = ".fastq.gz")
command.args <- paste("overwrite=true k=18 mink=18 rcomp=f qhdist=1 maskmiddle=t",
    " hammingdistance=2 findbestmatch=t minlength=38 maxlength=78 ordered=t ",
    "threads=", detectCores(), " in=", in.name.P7, " out=", out.name.P7, " lliteral=",
    "AGCAACCTCCAGAGAGGCAACG", " rliteral=", "CAGACAAGCAGCTACCGCAGAT", sep = "")

sys.out <- system2(path.expand("~/bbmap/bbduk2.sh"), args = command.args, stdout = TRUE,
    stderr = TRUE)  #

sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("bbduk2 extraction of fragments")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[3:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

```
in.name.P7 <- out.name.P7

out.name.P5 <- tempfile(pattern = "P5_", tmpdir = tempdir(), fileext = ".fastq.gz")
out.name.P7 <- tempfile(pattern = "P7_", tmpdir = tempdir(), fileext = ".fastq.gz")
out.name.P5_singlet <- tempfile(pattern = "P5_singlet_", tmpdir = tempdir(),
    fileext = ".fastq.gz")
out.name.P7_singlet <- tempfile(pattern = "P7_singlet_", tmpdir = tempdir(),
    fileext = ".fastq.gz")

command.args <- paste("makepairs -c 'gzip' -f ", in.name.P5, " -r ", in.name.P7,
    " -fp ", out.name.P5, " -rp ", out.name.P7, " -fs ", out.name.P5_singlet,
    " -rs ", out.name.P7_singlet, " --stats 2>&1", sep = "")
sys.out <- system2("/usr/local/bin/pairfq", args = command.args, stdout = TRUE,
    stderr = TRUE)
sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("pairfq pair matching")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[1:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

```
rm(sys.out)


system(paste("mv ", out.name.P5, " ./data/barcodes_", name.out, ".fastq.gz",
    sep = ""))
system(paste("mv ", out.name.P7, " ./data/fragments_", name.out, ".fastq.gz",
    sep = ""))

unlink(paste(tempdir(), "/*", sep = ""), recursive = FALSE, force = FALSE)  #Cleanup of temp files

print("Total execution time:")
```

```
[1] "Total execution time:"
```

| | bbduk2 extraction of fragments |
|---|---|
| 3 | |
| 4 | BBDuk2 version 37.02 |
| 5 | Set ORDERED to true |
| 6 | Set threads to 32 |
| 7 | k=18 |
| 8 | maskMiddle=true |
| 9 | hamming distance=2 |
| 10 | right-ktrimming using 1 literal. |
| 11 | left-ktrimming using 1 literal. |
| 12 | |
| 13 | Initial: |
| 14 | Memory: max=75573m, free=72417m, used=3156m |
| 15 | |
| 16 | Added 6380 kmers; time: 0.070 seconds. |
| 17 | Memory: max=75573m, free=69657m, used=5916m |
| 18 | |
| 19 | Added 6380 kmers; time: 0.002 seconds. |
| 20 | Memory: max=75573m, free=69657m, used=5916m |
| 21 | |
| 22 | Input is being processed as unpaired |
| 23 | Started output streams: 0.090 seconds. |
| 24 | Processing time: 232.650 seconds. |
| 25 | |
| 26 | Input: 11547945 reads 1737634780 bases. |
| 27 | KTrimmed: 22321617 reads (193.30%) 1078844314 bases (62.09%) |
| 28 | Total Removed: 698793 reads (6.05%) 1146258928 bases (65.97%) |
| 29 | Result: 10849152 reads (93.95%) 591375852 bases (34.03%) |
| 30 | |
| 31 | Time: 232.830 seconds. |
| 32 | Reads Processed: 11547k 49.60k reads/sec |
| 33 | Bases Processed: 1737m 7.46m bases/sec |

| pairfq pair matching |
|---|
| ========== pairfq version : 0.17.0 (completion time: Wed Nov 1 10:47:44 UTC 2017) |
| Total forward reads (/tmp/RtmpmemuBF/BC_2162781c3c.fastq.gz) : 11374106 |
| Total reverse reads (/tmp/RtmpmemuBF/P7_2161be9256f.fastq.gz) : 10849152 |
| Total forward paired reads (/tmp/RtmpmemuBF/P5_21663064762.fastq.gz) : 10698072 |
| Total reverse paired reads (/tmp/RtmpmemuBF/P7_2162b7ef9d1.fastq.gz) : 10698072 |
| Total forward unpaired reads (/tmp/RtmpmemuBF/P5_singlet_2161f485577.fastq.gz) : 676034 |
| Total reverse unpaired reads (/tmp/RtmpmemuBF/P7_singlet_216961813b.fastq.gz) : 151080 |
| |
| Total paired reads : 21396144 |
| Total unpaired reads : 827114 |

```
print(Sys.time() - strt)
```

```
Time difference of 16.88165 mins
```

```
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     2017-11-01
```

Packages ----------

```
 package              * version  date       source
 acepack                1.4.1    2016-10-29 CRAN (R 3.4.2)
 backports              1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                 * 3.4.2    2017-10-06 local
 base64enc              0.1-3    2015-07-28 CRAN (R 3.4.2)
 Biobase              * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics         * 0.22.1   2017-10-13 Bioconductor
 BiocParallel         * 1.10.1   2017-10-13 Bioconductor
 Biostrings           * 2.44.2   2017-10-13 Bioconductor
 bitops                 1.0-6    2013-08-17 CRAN (R 3.4.2)
 checkmate              1.8.4    2017-09-25 CRAN (R 3.4.2)
 cluster                2.0.6    2017-03-16 CRAN (R 3.4.2)
 codetools              0.2-15   2016-10-05 CRAN (R 3.4.2)
 colorspace             1.3-2    2016-12-14 CRAN (R 3.4.2)
 compiler               3.4.2    2017-10-06 local
 data.table           * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets             * 3.4.2    2017-10-06 local
 DelayedArray         * 0.2.7    2017-10-13 Bioconductor
 devtools             * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 digest                 0.6.12   2017-01-27 CRAN (R 3.4.2)
 doParallel           * 1.0.11   2017-09-28 CRAN (R 3.4.2)
 evaluate               0.10.1   2017-06-24 CRAN (R 3.4.2)
 foreach              * 1.4.3    2015-10-13 CRAN (R 3.4.2)
 foreign                0.8-69   2017-06-21 CRAN (R 3.4.2)
 formatR                1.5      2017-04-25 CRAN (R 3.4.2)
 Formula              * 1.2-2    2017-07-10 CRAN (R 3.4.2)
 GenomeInfoDb         * 1.12.3   2017-10-13 Bioconductor
 GenomeInfoDbData       0.99.0   2017-10-13 Bioconductor
 GenomicAlignments    * 1.12.2   2017-10-13 Bioconductor
 GenomicRanges        * 1.28.6   2017-10-13 Bioconductor
 ggplot2              * 2.2.1    2016-12-30 CRAN (R 3.4.2)
 graphics             * 3.4.2    2017-10-06 local
 grDevices            * 3.4.2    2017-10-06 local
 grid                   3.4.2    2017-10-06 local
 gridExtra              2.3      2017-09-09 CRAN (R 3.4.2)
 gtable                 0.2.0    2016-02-26 CRAN (R 3.4.2)
 Hmisc                * 4.0-3    2017-05-02 CRAN (R 3.4.2)
 hms                    0.3      2016-11-22 CRAN (R 3.4.2)
 htmlTable              1.9      2017-01-26 CRAN (R 3.4.2)
 htmltools              0.3.6    2017-04-28 CRAN (R 3.4.2)
 htmlwidgets            0.9      2017-07-10 CRAN (R 3.4.2)
 httr                   1.3.1    2017-08-20 CRAN (R 3.4.2)
 hwriter                1.3.2    2014-09-10 CRAN (R 3.4.2)
 IRanges              * 2.10.5   2017-10-13 Bioconductor
 iterators            * 1.0.8    2015-10-13 CRAN (R 3.4.2)
 kableExtra           * 0.5.2    2017-09-15 CRAN (R 3.4.2)
 knitr                * 1.17     2017-08-10 CRAN (R 3.4.2)
```

```
lattice             * 0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra          0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval              0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr              1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats         * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise               1.1.0    2017-04-21 CRAN (R 3.4.2)
methods             * 3.4.2    2017-10-06 local
munsell               0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                  7.3-12   2016-02-02 CRAN (R 3.4.2)
parallel            * 3.4.2    2017-10-06 local
plyr                  1.8.4    2016-06-08 CRAN (R 3.4.2)
R6                    2.2.2    2017-06-17 CRAN (R 3.4.2)
RColorBrewer          1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                  0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                 1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                 1.1.1    2017-05-16 CRAN (R 3.4.2)
rlang                 0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown             1.6      2017-06-15 CRAN (R 3.4.2)
rpart                 4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot             1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools           * 1.28.0   2017-10-13 Bioconductor
rvest                 0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors           * 0.14.7   2017-10-13 Bioconductor
scales                0.5.0    2017-08-24 CRAN (R 3.4.2)
ShortRead           * 1.34.2   2017-10-13 Bioconductor
splines               3.4.2    2017-10-06 local
stats               * 3.4.2    2017-10-06 local
stats4              * 3.4.2    2017-10-06 local
stringi               1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr               1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment * 1.6.5   2017-10-13 Bioconductor
survival            * 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                 3.4.2    2017-10-06 local
utils               * 3.4.2    2017-10-06 local
withr                 2.0.0    2017-07-28 CRAN (R 3.4.2)
xml2                  1.1.1    2017-01-24 CRAN (R 3.4.2)
XVector             * 0.16.0   2017-10-13 Bioconductor
yaml                  2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc              1.22.0   2017-10-13 Bioconductor
```

# Reverse mapping of CustumArray oligos to original proteins

*Tomas Bjorklund*

*Wed Nov 1 10:47:48 2017*

This workflow aligns the short oligos from the CustomArray order to the full reference sequences using Bowtie2. This enables the mapping to all genes sharing the same sequence.

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

## Load sequences

```
LUT.dna <- read.table("data/SortedFragments_all.txt", header = TRUE, skip = 0,
    sep = "\t", stringsAsFactors = FALSE, fill = TRUE)
LUT.dna <- data.table(LUT.dna)
```

## Remove constitutive backbone sequences

```
invisible(LUT.dna[, `:=`(Sequence, gsub("aacctccagagaggcaacg", "", Sequence))])
invisible(LUT.dna[, `:=`(Sequence, gsub("cagacaagcagctaccgca", "", Sequence))])
invisible(LUT.dna[, `:=`(Sequence, toupper(Sequence))])
setkey(LUT.dna, "Sequence")
LUT.dna <- unique(LUT.dna)
LUT.dna$Names <- LUT.dna$Sequence
LUT.dna$LUTnr <- make.names(seq(nrow(LUT.dna)), unique = TRUE)
```

## Split sequences based on linker and length

```
LUT.14aaG4S <- LUT.dna[substr(LUT.dna$Sequence, 1, 14) == "GAGGCGGAGGAAGT"]
LUT.remaining <- LUT.dna[!(substr(LUT.dna$Sequence, 1, 14) == "GAGGCGGAGGAAGT")]
LUT.14aaA5 <- LUT.remaining[substr(LUT.remaining$Sequence, 1, 14) == "CTGCTGCAGCAGCC"]
LUT.remaining <- LUT.remaining[!(substr(LUT.remaining$Sequence, 1, 14) == "CTGCTGCAGCAGCC")]
LUT.22aa <- LUT.remaining[nchar(LUT.remaining$Sequence) == 70L & substr(LUT.remaining$Sequence,
    1, 2) == "CT"]
LUT.remaining <- LUT.remaining[!(nchar(LUT.remaining$Sequence) == 70L & substr(LUT.remaining$Sequence,
    1, 2) == "CT")]
LUT.14aa <- LUT.remaining[nchar(LUT.remaining$Sequence) == 46L & substr(LUT.remaining$Sequence,
    1, 2) == "CT"]
rm(LUT.remaining)
LUT.dna[LUT.dna$Sequence %in% LUT.14aaG4S$Sequence, "Structure"] <- "14aaG4S"
LUT.dna[LUT.dna$Sequence %in% LUT.14aaA5$Sequence, "Structure"] <- "14aaA5"
LUT.dna[LUT.dna$Sequence %in% LUT.22aa$Sequence, "Structure"] <- "22aa"
LUT.dna[LUT.dna$Sequence %in% LUT.14aa$Sequence, "Structure"] <- "14aa"

save(LUT.dna, file = "data/LUTdna.rda")
```

## Trim sequences

```
LUT.14aa$Sequence <- substr(LUT.14aa$Sequence, 3, 44)
LUT.14aaG4S$Sequence <- substr(LUT.14aaG4S$Sequence, 15, 56)
LUT.14aaA5$Sequence <- substr(LUT.14aaA5$Sequence, 15, 56)
LUT.22aa$Sequence <- substr(LUT.22aa$Sequence, 3, 68)
```

## Save fasta files for Bowtie alignments

```
LUT.14aa.fa <- tempfile(pattern = "LUT_14aa_", tmpdir = tempdir(), fileext = "fa")
LUT.14aa.seq = ShortRead(DNAStringSet(LUT.14aa$Sequence), BStringSet(LUT.14aa$LUTnr))
writeFasta(LUT.14aa.seq, LUT.14aa.fa)

LUT.14aaG4S.fa <- tempfile(pattern = "LUT_14aaG4s_", tmpdir = tempdir(), fileext = "fa")
LUT.14aaG4S.seq = ShortRead(DNAStringSet(LUT.14aaG4S$Sequence), BStringSet(LUT.14aaG4S$LUTnr))
writeFasta(LUT.14aaG4S.seq, LUT.14aaG4S.fa)

LUT.14aaA5.fa <- tempfile(pattern = "LUT_14aaA5_", tmpdir = tempdir(), fileext = "fa")
LUT.14aaA5.seq = ShortRead(DNAStringSet(LUT.14aaA5$Sequence), BStringSet(LUT.14aaA5$LUTnr))
writeFasta(LUT.14aaA5.seq, LUT.14aaA5.fa)

LUT.22aa.fa <- tempfile(pattern = "LUT_14aaA5_", tmpdir = tempdir(), fileext = "fa")
LUT.22aa.seq = ShortRead(DNAStringSet(LUT.22aa$Sequence), BStringSet(LUT.22aa$LUTnr))
writeFasta(LUT.22aa.seq, LUT.22aa.fa)
```

## Build Bowtie index

```
seqs.original <- readFasta("input/DNA-lib_RetrogradeTransport.fasta")

seqs.AA <- Biostrings::translate(sread(seqs.original), genetic.code = GENETIC_CODE,
    if.fuzzy.codon = "error")

source("functions/AAtoDNA.R")
seqs.optimized = ShortRead(DNAStringSet(sapply(seqs.AA, function(x) AAtoDNA(x,
    species = "hsa"))), BStringSet(gsub("([ ])", "_", ShortRead::id(seqs.original))))

bowtie.fasta <- tempfile(pattern = "bowtie_", tmpdir = tempdir(), fileext = ".fa")

writeFasta(seqs.optimized, bowtie.fasta)

bowtie.idx <- tempfile(pattern = "IDX_bowtie_", tmpdir = tempdir(), fileext = "")

sys.out <- system(paste("bowtie2-build", bowtie.fasta, bowtie.idx, "2>&1", sep = " "),
    intern = TRUE, ignore.stdout = FALSE)
```

## Align fragments to reference

Align 14aa sequences

```
name.bowtie <- tempfile(pattern = "bowtie_", tmpdir = tempdir(), fileext = "")
```

```
sys.out <- system(paste("bowtie2 --non-deterministic --threads ", detectCores(),
    " --very-sensitive -f -a", " -x ", bowtie.idx, " -U ", LUT.14aa.fa, " -S ",
    name.bowtie, ".sam 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)


sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("Bowtie 2 alignment to library")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[1:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

| Bowtie 2 alignment to library |
|---|
| 44705 reads; of these: |
| 44705 (100.00%) were unpaired; of these: |
| 0 (0.00%) aligned 0 times |
| 27420 (61.34%) aligned exactly 1 time |
| 17285 (38.66%) aligned >1 times |
| 100.00% overall alignment rate |

```
command.args <- paste("view -@ ", detectCores(), " -Sb ", name.bowtie, ".sam > ",
    name.bowtie, ".bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
command.args <- paste("sort -@ ", detectCores(), " ", name.bowtie, ".bam -o ",
    name.bowtie, "_sort.bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
frag14aa.ranges <- readGAlignments(paste(name.bowtie, "_sort.bam", sep = ""),
    use.names = TRUE)
length(names(frag14aa.ranges))
```

```
[1] 75152
```

```
length(unique(names(frag14aa.ranges)))
```

```
[1] 44705
```

```
length(unique(LUT.14aa$Sequence))
```

```
[1] 44705
```

Align 14aaG4S sequences

```
name.bowtie <- tempfile(pattern = "bowtie_", tmpdir = tempdir(), fileext = "")

sys.out <- system(paste("bowtie2 --non-deterministic --threads ", detectCores(),
    " --very-sensitive -f -a", " -x ", bowtie.idx, " -U ", LUT.14aaG4S.fa, " -S ",
    name.bowtie, ".sam 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)

sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("Bowtie 2 alignment to library")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[1:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

| Bowtie 2 alignment to library |
|---|
| 15792 reads; of these: |
| 15792 (100.00%) were unpaired; of these: |
| 0 (0.00%) aligned 0 times |
| 9150 (57.94%) aligned exactly 1 time |
| 6642 (42.06%) aligned >1 times |
| 100.00% overall alignment rate |

```
command.args <- paste("view -@ ", detectCores(), " -Sb ", name.bowtie, ".sam > ",
    name.bowtie, ".bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
command.args <- paste("sort -@ ", detectCores(), " ", name.bowtie, ".bam -o ",
    name.bowtie, "_sort.bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
frag14aaG4S.ranges <- readGAlignments(paste(name.bowtie, "_sort.bam", sep = ""),
    use.names = TRUE)
length(names(frag14aaG4S.ranges))
```

```
[1] 27778
```

```
length(unique(names(frag14aaG4S.ranges)))
```

```
[1] 15792
```

```
length(unique(LUT.14aaG4S$Sequence))
```

```
[1] 15792
```

Align 14aaA5 sequences

```
name.bowtie <- tempfile(pattern = "bowtie_", tmpdir = tempdir(), fileext = "")

sys.out <- system(paste("bowtie2 --non-deterministic --threads ", detectCores(),
    " --very-sensitive -f -a", " -x ", bowtie.idx, " -U ", LUT.14aaA5.fa, " -S ",
    name.bowtie, ".sam 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)

sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("Bowtie 2 alignment to library")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[1:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

| Bowtie 2 alignment to library |
|---|
| 15792 reads; of these: |
| 15792 (100.00%) were unpaired; of these: |
| 0 (0.00%) aligned 0 times |
| 9150 (57.94%) aligned exactly 1 time |
| 6642 (42.06%) aligned >1 times |
| 100.00% overall alignment rate |

```
command.args <- paste("view -@ ", detectCores(), " -Sb ", name.bowtie, ".sam > ",
    name.bowtie, ".bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
command.args <- paste("sort -@ ", detectCores(), " ", name.bowtie, ".bam -o ",
    name.bowtie, "_sort.bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
frag14aaA5.ranges <- readGAlignments(paste(name.bowtie, "_sort.bam", sep = ""),
    use.names = TRUE)
length(names(frag14aaA5.ranges))
```

```
[1] 27778
```

```
length(unique(names(frag14aaA5.ranges)))
```

```
[1] 15792
```

```
length(unique(LUT.14aaA5$Sequence))
```

```
[1] 15792
```

Align 22aa sequences

```
name.bowtie <- tempfile(pattern = "bowtie_", tmpdir = tempdir(), fileext = "")

sys.out <- system(paste("bowtie2 --non-deterministic --threads ", detectCores(),
    " --very-sensitive -f -a", " -x ", bowtie.idx, " -U ", LUT.22aa.fa, " -S ",
    name.bowtie, ".sam 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)

sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("Bowtie 2 alignment to library")
invisible(sys.out[" "] <- " ")
lengthOut <- (nrow(sys.out))
knitr::kable(sys.out[1:lengthOut, ], format = "latex", booktabs = T) %>% kable_styling(latex_options = "stri
```

| Bowtie 2 alignment to library |
|---|
| 16054 reads; of these: |
| 16054 (100.00%) were unpaired; of these: |
| 0 (0.00%) aligned 0 times |
| 8730 (54.38%) aligned exactly 1 time |
| 7324 (45.62%) aligned >1 times |
| 100.00% overall alignment rate |

```
command.args <- paste("view -@ ", detectCores(), " -Sb ", name.bowtie, ".sam > ",
    name.bowtie, ".bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
command.args <- paste("sort -@ ", detectCores(), " ", name.bowtie, ".bam -o ",
    name.bowtie, "_sort.bam", sep = "")
system2("/usr/local/bin/samtools", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
frag22aa.ranges <- readGAlignments(paste(name.bowtie, "_sort.bam", sep = ""),
    use.names = TRUE)
length(names(frag22aa.ranges))
```

```
[1] 29665
```

```
length(unique(names(frag22aa.ranges)))
```

```
[1] 16054
```

```
length(unique(LUT.22aa$Sequence))
```

```
[1] 16054
```

## Merge and annotate aligned sequences

```
mcols(frag14aa.ranges)$structure <- "14aa"
mcols(frag22aa.ranges)$structure <- "22aa"
mcols(frag14aaA5.ranges)$structure <- "14aaA5"
mcols(frag14aaG4S.ranges)$structure <- "14aaG4S"
allFragments.ranges <- append(frag14aa.ranges, frag22aa.ranges)
allFragments.ranges <- append(allFragments.ranges, frag14aaA5.ranges)
allFragments.ranges <- append(allFragments.ranges, frag14aaG4S.ranges)


mcols(allFragments.ranges)$LUTnr <- names(allFragments.ranges)
setkey(LUT.dna, LUTnr)
mcols(allFragments.ranges)$Sequence <- LUT.dna[mcols(allFragments.ranges)$LUTnr]$Sequence

save(allFragments.ranges, file = "data/alignedLibraries.rda")

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     2017-11-01
```

```
Packages -----------------------------------------------------------------
```

```
 package             * version  date       source
 acepack               1.4.1    2016-10-29 CRAN (R 3.4.2)
 ade4                  1.7-8    2017-08-09 CRAN (R 3.4.2)
 backports             1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                * 3.4.2    2017-10-06 local
 base64enc             0.1-3    2015-07-28 CRAN (R 3.4.2)
 Biobase             * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics        * 0.22.1   2017-10-13 Bioconductor
 BiocParallel        * 1.10.1   2017-10-13 Bioconductor
 Biostrings          * 2.44.2   2017-10-13 Bioconductor
 bitops                1.0-6    2013-08-17 CRAN (R 3.4.2)
 checkmate             1.8.4    2017-09-25 CRAN (R 3.4.2)
```

```
cluster              2.0.6    2017-03-16 CRAN (R 3.4.2)
colorspace           1.3-2    2016-12-14 CRAN (R 3.4.2)
compiler             3.4.2    2017-10-06 local
data.table         * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
datasets           * 3.4.2    2017-10-06 local
DelayedArray       * 0.2.7    2017-10-13 Bioconductor
devtools           * 1.13.3   2017-08-02 CRAN (R 3.4.2)
digest               0.6.12   2017-01-27 CRAN (R 3.4.2)
evaluate             0.10.1   2017-06-24 CRAN (R 3.4.2)
foreign              0.8-69   2017-06-21 CRAN (R 3.4.2)
formatR              1.5      2017-04-25 CRAN (R 3.4.2)
Formula            * 1.2-2    2017-07-10 CRAN (R 3.4.2)
GeneGA             * 1.26.0   2017-10-13 Bioconductor
GenomeInfoDb       * 1.12.3   2017-10-13 Bioconductor
GenomeInfoDbData     0.99.0   2017-10-13 Bioconductor
GenomicAlignments  * 1.12.2   2017-10-13 Bioconductor
GenomicRanges      * 1.28.6   2017-10-13 Bioconductor
ggplot2            * 2.2.1    2016-12-30 CRAN (R 3.4.2)
graphics           * 3.4.2    2017-10-06 local
grDevices          * 3.4.2    2017-10-06 local
grid                 3.4.2    2017-10-06 local
gridExtra            2.3      2017-09-09 CRAN (R 3.4.2)
gtable               0.2.0    2016-02-26 CRAN (R 3.4.2)
hash               * 2.2.6    2013-02-21 CRAN (R 3.4.2)
Hmisc              * 4.0-3    2017-05-02 CRAN (R 3.4.2)
hms                  0.3      2016-11-22 CRAN (R 3.4.2)
htmlTable            1.9      2017-01-26 CRAN (R 3.4.2)
htmltools            0.3.6    2017-04-28 CRAN (R 3.4.2)
htmlwidgets          0.9      2017-07-10 CRAN (R 3.4.2)
httr                 1.3.1    2017-08-20 CRAN (R 3.4.2)
hwriter              1.3.2    2014-09-10 CRAN (R 3.4.2)
IRanges            * 2.10.5   2017-10-13 Bioconductor
kableExtra         * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr              * 1.17     2017-08-10 CRAN (R 3.4.2)
lattice            * 0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra         0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval             0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr             1.5      2014-11-22 CRAN (R 3.4.2)
Matrix               1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats        * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise              1.1.0    2017-04-21 CRAN (R 3.4.2)
methods            * 3.4.2    2017-10-06 local
munsell              0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                 7.3-12   2016-02-02 CRAN (R 3.4.2)
parallel           * 3.4.2    2017-10-06 local
plyr                 1.8.4    2016-06-08 CRAN (R 3.4.2)
R6                   2.2.2    2017-06-17 CRAN (R 3.4.2)
RColorBrewer         1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                 0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                1.1.1    2017-05-16 CRAN (R 3.4.2)
rlang                0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown            1.6      2017-06-15 CRAN (R 3.4.2)
rpart                4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot            1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools          * 1.28.0   2017-10-13 Bioconductor
rvest                0.3.2    2016-06-17 CRAN (R 3.4.2)
```

```
S4Vectors            * 0.14.7   2017-10-13 Bioconductor
scales                 0.5.0    2017-08-24 CRAN (R 3.4.2)
seqinr               * 3.4-5    2017-08-01 CRAN (R 3.4.2)
ShortRead            * 1.34.2   2017-10-13 Bioconductor
splines                3.4.2    2017-10-06 local
stats                * 3.4.2    2017-10-06 local
stats4               * 3.4.2    2017-10-06 local
stringi                1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment * 1.6.5    2017-10-13 Bioconductor
survival             * 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                 1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                  3.4.2    2017-10-06 local
utils                * 3.4.2    2017-10-06 local
withr                  2.0.0    2017-07-28 CRAN (R 3.4.2)
xml2                   1.1.1    2017-01-24 CRAN (R 3.4.2)
XVector              * 0.16.0   2017-10-13 Bioconductor
yaml                   2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc               1.22.0   2017-10-13 Bioconductor
```

# Library fragment alignment

*Tomas Bjorklund*

*Wed Nov 1 10:48:31 2017*

This workflow identifies correct fragments from the Cre-recombined AAV plasmid library and aligns them to the CustumArray ordered nucleotide fragments using Blastn. Consistant mutations in each fragment/barcode combination are also registered as is the putity of each barcode.

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

## Load the trimmed reads

```
load("data/LUTdna.rda")

fragments.file <- "data/fragments_AAVlibrary_complete.fastq.gz"
barcodes.file <- "data/barcodes_AAVlibrary_complete.fastq.gz"

reads.trim <- readFastq(fragments.file)
reads.BC <- readFastq(barcodes.file)
```

## Make CustomArray reference index for Blast

```
LUT.fa <- tempfile(pattern = "LUT_", tmpdir = tempdir(), fileext = ".fa")
LUT.seq = ShortRead(DNAStringSet(LUT.dna$Sequence), BStringSet(1:length(LUT.dna$LUTnr)))
writeFasta(LUT.seq, LUT.fa)
```

## Save unique fragments as fasta file

```
unique.reads <- unique(sread(reads.trim))
```

## Select subset

```
# unique.reads <- unique.reads[sample(length(unique.reads), 50000)]

unique.reads <- ShortRead(DNAStringSet(unique.reads), BStringSet(1:length(unique.reads)))
fragments.unique.fa <- tempfile(pattern = "FragUnique_", tmpdir = tempdir(),
    fileext = ".fa")
writeFasta(unique.reads, fragments.unique.fa)
```

## Align against the library using blast

```
blast.db <- tempfile(pattern = "blastDB_", tmpdir = tempdir(), fileext = ".db")
blast.out <- tempfile(pattern = "blastOut_", tmpdir = tempdir(), fileext = ".txt")
```

```
sys.out <- system(paste("makeblastdb -in ", LUT.fa, " -out ", blast.db, " -dbtype nucl -title LUT -parse_seq
    sep = ""), intern = TRUE, ignore.stdout = FALSE)

sys.out <- as.data.frame(sys.out)

colnames(sys.out) <- c("blastn database generation")
invisible(sys.out[" "] <- " ")
knitr::kable(sys.out[1:(nrow(sys.out)), ], format = "latex", booktabs = T) %>%
    kable_styling(latex_options = "striped")
```

| blastn database generation |
|---|
| |
| Building a new DB, current time: 11/01/2017 10:49:31 |
| New DB name: /tmp/RtmpivyakH/blastDB_66624bb2db9.db |
| New DB title: LUT |
| Sequence type: Nucleotide |
| Keep MBits: T |
| Maximum file size: 1000000000B |
| Adding sequences from FASTA; added 92343 sequences in 3.2858 seconds. |

```
sys.out <- system(paste("export SHELL=/bin/sh; cat ", fragments.unique.fa, " | parallel --block ",
    floor(length(unique.reads)/detectCores()), " --recstart '>' --pipe 'blastn -max_target_seqs 25 -word_siz
    " -num_threads 1 -outfmt 10 -db ", blast.db, " -query - '> ", blast.out,
    " 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)  #

# table.blastn <- data.table(read.table(blast.out, header = FALSE, skip = 0,
# sep=';', stringsAsFactors = FALSE, fill=FALSE) , keep.rownames=FALSE,
# key='V1')

system(paste("gzip -c ", blast.out, " > ./data/blastOutput.csv.gz", sep = " "))

table.blastn <- data.table(scan(file = "./data/blastOutput.csv.gz", what = "character",
    sep = ";"), keep.rownames = FALSE, key = "V1")


if (length(grep("Warning", table.blastn$V1)) != 0) {
    warnings.out <- unique(table.blastn[grep("Warning", table.blastn$V1), ])
    table.blastn <- table.blastn[-grep("Warning", table.blastn$V1), ]
    setnames(warnings.out, "V1", c("blastn Warnings"))
    # knitr::kable(warnings.out[1:(nrow(warnings.out)),], format = 'markdown')
}

table.blastn[, `:=`(c("Reads", "Sequence", "identity", "alignmentLength", "mismatches",
    "gapOpens", "q_start", "q_end", "s_start", "s_end", "evalue", "bitScore"),
    tstrsplit(V1, ",", fixed = TRUE)), ]



table.blastn[, `:=`(Reads, as.character(sread(unique.reads[as.integer(Reads)])))]
table.blastn[, `:=`(Sequence, as.character(sread(LUT.seq[as.integer(Sequence)])))]
setkey(table.blastn, Sequence)
setkey(LUT.dna, Sequence)
table.blastn <- table.blastn[LUT.dna, nomatch = 0]
table.blastn[, `:=`(c("V1", "identity", "alignmentLength", "gapOpens", "q_start",
```

```
    "q_end", "s_start", "s_end", "evalue", "Sequence", "Names"), NULL)]
gc()  #garbage collection to reduce memory foot print. Can be removed for speed
```

```
            used   (Mb) gc trigger    (Mb)   max used    (Mb)
Ncells   9349510  499.4  216418051 11558.0  172911949  9234.5
Vcells 964075906 7355.4 4426097682 33768.5 5509384308 42033.3
```

```
table.blastn[, `:=`(bitScore, as.numeric(bitScore))]
table.blastn[, `:=`(mismatches, as.numeric(mismatches))]

setkeyv(table.blastn, c("Reads", "LUTnr"))
setorder(table.blastn, Reads, LUTnr, -bitScore)  #This makes sure that a fragment is only aligned once to t
table.blastn <- unique(table.blastn, by = c("Reads", "LUTnr"))

gc()  #garbage collection to reduce memory foot print. Can be removed for speed
```

```
            used   (Mb) gc trigger    (Mb)   max used    (Mb)
Ncells   9350413  499.4  173134440  9246.4  172911949  9234.5
Vcells 942290072 7189.2 3540878145 27014.8 5509384308 42033.3
```

```
table.blastn.topHit <- table.blastn[table.blastn[, .I[which.max(bitScore)],
    by = "Reads"]$V1]  # Select only rows with the highest bitScore

full.table <- data.table(Reads = as.character(sread(reads.trim)), BC = as.character(sread(reads.BC)),
    key = "Reads")
all.reads <- nrow(full.table)

full.table <- full.table[table.blastn.topHit, nomatch = 0]  # Merge reads with the top hit alignment

print(paste("Alignment percentage:", percent(nrow(full.table)/all.reads)))
```

```
[1] "Alignment percentage: 99.6%"
```

## Starcode based barcode reduction

```
out.name.BC.star <- tempfile(pattern = "BCsc_", tmpdir = tempdir(), fileext = ".txt")

command.args <- paste("-c ", barcodes.file, " | starcode -t ", detectCores() -
    1, " --print-clusters -d", 1, " -r5 -q -o ", out.name.BC.star, sep = "")

system2("gunzip", args = command.args, stdout = TRUE, stderr = TRUE)
```

```
character(0)
```

```
table.BC.sc <- data.table(read.table(out.name.BC.star, header = FALSE, row.names = 1,
    skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = FALSE), keep.rownames = TRUE,
    key = "rn")  #, nrows = 1000
table.BC.sc[, `:=`(V2, NULL)]

table.BC.sc <- table.BC.sc[, strsplit(as.character(V3), ",", fixed = TRUE),
    by = rn]

SC.droppedBC <- length(unique(sread(reads.BC))) - length(unique(table.BC.sc$V1) %in%
    unique(sread(reads.BC)))
print(paste("Dropped BCs in Starcode:", SC.droppedBC))
```

```
[1] "Dropped BCs in Starcode: 16603"
```

```
# rm(reads.BC,reads.trim)

setnames(table.BC.sc, c("V1", "rn"), c("BC", "scBC"))

DNA_pscAAVlib <- table.BC.sc
save(DNA_pscAAVlib, file = "data/scBC_DNA_pscAAVlib.rda")
```

## Replacing barcodes with Starcode reduced versions

```
setkey(full.table, BC)
setkey(table.BC.sc, BC)
full.table <- full.table[table.BC.sc, nomatch = 0]
# full.table <- merge(full.table,table.BC.sc, by='BC', all = FALSE, all.x =
# FALSE) rm(table.BC.sc)

setnames(full.table, c("BC", "scBC"), c("oldBC", "BC"))

setkey(full.table, BC)

RetainedBC <- length(unique(full.table$oldBC))
scBC <- length(unique(full.table$BC))
print(paste("Original unique barcodes:", RetainedBC))
```

```
[1] "Original unique barcodes: 3579215"
```

```
print(paste("SC reduced unique barcodes:", scBC))
```

```
[1] "SC reduced unique barcodes: 3188119"
```

```
table.frag <- data.table(as.data.frame((rev(sort(table(full.table$oldBC))))[1:10],
    row.names = "Var1"), keep.rownames = TRUE)
# In R versions below 3.3 remove, row.names = 'Var1' to make this compatible
setnames(table.frag, colnames(table.frag), c("Original BC", "Count"))
knitr::kable(table.frag, format = "latex", booktabs = T) %>% kable_styling(latex_options = "striped")
```

| Original BC | Count |
|---|---|
| GTGTCCATCTGGACGCGTAG | 165 |
| CATGCATGGATGACTTATGT | 137 |
| GATGGTATATATGCGGATGG | 133 |
| GTTTAAAGCAATATGCACAC | 118 |
| AATCGCTGGATGGTTTATGG | 118 |
| GCGCAATCGTGGGAACGCAC | 108 |
| GTACATTGCTGTGAGCCTGC | 101 |
| GTGCCTTTATTGGCGGCATT | 100 |
| GCGGGCGGGTGGAATGGCGT | 99 |
| GCGTATGTGCAGGCTCATTG | 97 |

```
table.frag <- data.table(as.data.frame((rev(sort(table(full.table$BC))))[1:10],
    row.names = "Var1"), keep.rownames = TRUE)
# In R versions below 3.3 remove, row.names = 'Var1' to make this compatible
setnames(table.frag, colnames(table.frag), c("SC reduced BC", "Count"))
knitr::kable(table.frag, format = "latex", booktabs = T) %>% kable_styling(latex_options = "striped")
```

```
invisible(full.table[, `:=`(oldBC, NULL)])
```

| SC reduced BC | Count |
|---|---|
| GTGTCCATCTGGACGCGTAG | 173 |
| GATGGTATATATGCGGATGG | 151 |
| CATGCATGGATGACTTATGT | 151 |
| AATCGCTGGATGGTTTATGG | 125 |
| GTTTAAAGCAATATGCACAC | 122 |
| GCGCAATCGTGGGAACGCAC | 119 |
| GCGGGCGGGTGGAATGGCGT | 117 |
| GTACATTGCTGTGAGCCTGC | 110 |
| GCGTATGTGCAGGCTCATTG | 105 |
| CTGCGCACGCTTGTACGCGG | 103 |

## Splitting reads into single-read and multi-read barcodes

```
full.table <- full.table[order(full.table$BC), ]
full.table[, `:=`(mismatches, as.numeric(mismatches))]

temp.table.single <- full.table[full.table[, .I[.N == 1], by = "BC"]$V1]
temp.table.multi <- full.table[full.table[, .I[.N > 1], by = "BC"]$V1]

temp.table.single[, `:=`(c("mCount", "tCount"), 1)]
temp.table.single$Mode <- "Amb"
setkeyv(temp.table.multi, c("BC", "LUTnr"))


temp.table.multi[, `:=`(c("bitScore", "mismatches", "tCount"), list(mean(bitScore),
    median(mismatches), .N)), by = key(temp.table.multi)]

temp.table.multi$Mode <- "Def"
temp.table.multi <- unique(temp.table.multi)

print("Utilized reads.......")
```

```
[1] "Utilized reads......."
```

```
print(nrow(full.table))
```

```
[1] 10642223
```

```
print("Whereof single reads.......")
```

```
[1] "Whereof single reads......."
```

```
print(nrow(temp.table.single))
```

```
[1] 1353656
```

## Splitting multi-read barcodes into clean and chimeric

```
setkeyv(temp.table.multi, "BC")

temp.table.multi.clean <- temp.table.multi[temp.table.multi[, .I[.N == 1], by = "BC"]$V1]
temp.table.multi <- temp.table.multi[temp.table.multi[, .I[.N > 1], by = "BC"]$V1]
```

```
temp.table.multi.clean[, `:=`(mCount, tCount)]

print("Clean multi-read barcodes.......")
```

```
[1] "Clean multi-read barcodes......."
```

```
print(nrow(temp.table.multi.clean))
```

```
[1] 280333
```

```
print("Chimeric multi-read barcodes.......")
```

```
[1] "Chimeric multi-read barcodes......."
```

```
print(length(unique(temp.table.multi$BC)))
```

```
[1] 1554130
```

## Calculate consensus alignment of chimeric barcodes

```
setkey(temp.table.multi, "BC")
temp.table.multi[, `:=`("mCount", tCount)]
temp.table.multi[, `:=`("tCount", sum(tCount)), by = "BC"]

setkey(temp.table.multi, "Reads")
temp.table.multi[, `:=`(c("LUTnr", "bitScore", "mismatches", "Structure"), NULL)]
setkey(table.blastn, "Reads")
temp.table.multi <- temp.table.multi[table.blastn, nomatch = 0, allow.cartesian = TRUE]

setkeyv(temp.table.multi, c("BC", "LUTnr"))
temp.table.multi[, `:=`(c("bitScore", "mismatches", "mCount"), list(max(bitScore),
    median(mismatches), sum(mCount))), by = key(temp.table.multi)]
gc()  #garbage collection to reduce memory foot print. Can be removed for speed
```

```
            used    (Mb) gc trigger    (Mb)   max used    (Mb)
Ncells   12952728   691.8   56732692  3029.9  173134440  9246.4
Vcells 2136119104 16297.3 3540878145 27014.8 5509384308 42033.3
```

```
temp.table.multi <- unique(temp.table.multi, by = c("BC", "LUTnr"))

setkeyv(temp.table.multi, "BC")
temp.table.multi <- temp.table.multi[temp.table.multi[, .I[mCount == max(mCount)],
    by = key(temp.table.multi)]$V1]
# Select only rows with the highest mCount

temp.table.multi <- temp.table.multi[temp.table.multi[, .I[which.max(bitScore)],
    by = key(temp.table.multi)]$V1]
# Select only rows with the highest bitScore

temp.table.multi[temp.table.multi$mCount == 1]$Mode <- "Amb"

print(paste("Number of barcodes with false mCount:", nrow(temp.table.multi[mCount >
    tCount])))
```

```
[1] "Number of barcodes with false mCount: 0"
```

```
temp.table.multi.consensus <- rbind(temp.table.multi, temp.table.multi.clean)

print(paste("Total number of definitive Barcodes:", length(grep("Def", temp.table.multi.consensus$Mode))))
```

```
[1] "Total number of definitive Barcodes: 1575111"
```

```
print(paste("Total number of ambiguous Barcodes:", length(grep("Amb", temp.table.multi.consensus$Mode))))
```

```
[1] "Total number of ambiguous Barcodes: 259352"
```

```
print(paste("Total number of single-read Barcodes:", nrow(temp.table.single)))
```

```
[1] "Total number of single-read Barcodes: 1353656"
```

```
output.Table <- rbind(temp.table.multi.consensus, temp.table.single)
save(output.Table, file = "data/multipleContfragmentsComplete.rda")

print("Total analysis time:")
```

```
[1] "Total analysis time:"
```

```
print(Sys.time() - strt1)
```

```
Time difference of 4.251831 hours
```

```
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     2017-11-01
```

```
Packages -----------------------------------------------------------------
```

```
 package               * version  date       source
 acepack                 1.4.1    2016-10-29 CRAN (R 3.4.2)
 AnnotationDbi         * 1.38.2   2017-10-13 Bioconductor
 AnnotationFilter        1.0.0    2017-10-13 Bioconductor
 AnnotationHub           2.8.2    2017-10-13 Bioconductor
 backports               1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                  * 3.4.2    2017-10-06 local
 base64enc               0.1-3    2015-07-28 CRAN (R 3.4.2)
 Biobase               * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics          * 0.22.1   2017-10-13 Bioconductor
 BiocInstaller           1.26.1   2017-10-10 Bioconductor
 BiocParallel          * 1.10.1   2017-10-13 Bioconductor
 biomaRt                 2.32.1   2017-10-13 Bioconductor
 Biostrings            * 2.44.2   2017-10-13 Bioconductor
 biovizBase            * 1.24.0   2017-10-13 Bioconductor
 bit                     1.1-12   2014-04-09 CRAN (R 3.4.2)
 bit64                   0.9-7    2017-05-08 CRAN (R 3.4.2)
 bitops                  1.0-6    2013-08-17 CRAN (R 3.4.2)
 blob                    1.1.0    2017-06-17 CRAN (R 3.4.2)
 BSgenome              * 1.44.2   2017-10-13 Bioconductor
 checkmate               1.8.4    2017-09-25 CRAN (R 3.4.2)
 cluster                 2.0.6    2017-03-16 CRAN (R 3.4.2)
 codetools               0.2-15   2016-10-05 CRAN (R 3.4.2)
 colorspace              1.3-2    2016-12-14 CRAN (R 3.4.2)
 compiler                3.4.2    2017-10-06 local
 curl                    2.8.1    2017-07-21 CRAN (R 3.4.2)
 data.table            * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
```

```
datasets               * 3.4.2    2017-10-06 local
DBI                      0.7      2017-06-18 CRAN (R 3.4.2)
DelayedArray           * 0.2.7    2017-10-13 Bioconductor
devtools               * 1.13.3   2017-08-02 CRAN (R 3.4.2)
dichromat                2.0-0    2013-01-24 CRAN (R 3.4.2)
digest                   0.6.12   2017-01-27 CRAN (R 3.4.2)
doParallel             * 1.0.11   2017-09-28 CRAN (R 3.4.2)
ensembldb                2.0.4    2017-10-13 Bioconductor
evaluate                 0.10.1   2017-06-24 CRAN (R 3.4.2)
foreach                * 1.4.3    2015-10-13 CRAN (R 3.4.2)
foreign                  0.8-69   2017-06-21 CRAN (R 3.4.2)
formatR                  1.5      2017-04-25 CRAN (R 3.4.2)
Formula                * 1.2-2    2017-07-10 CRAN (R 3.4.2)
GenomeInfoDb           * 1.12.3   2017-10-13 Bioconductor
GenomeInfoDbData         0.99.0   2017-10-13 Bioconductor
GenomicAlignments      * 1.12.2   2017-10-13 Bioconductor
GenomicFeatures        * 1.28.5   2017-10-13 Bioconductor
GenomicRanges          * 1.28.6   2017-10-13 Bioconductor
ggplot2                * 2.2.1    2016-12-30 CRAN (R 3.4.2)
graphics               * 3.4.2    2017-10-06 local
grDevices              * 3.4.2    2017-10-06 local
grid                   * 3.4.2    2017-10-06 local
gridExtra                2.3      2017-09-09 CRAN (R 3.4.2)
gtable                   0.2.0    2016-02-26 CRAN (R 3.4.2)
Gviz                   * 1.20.0   2017-10-13 Bioconductor
Hmisc                  * 4.0-3    2017-05-02 CRAN (R 3.4.2)
hms                      0.3      2016-11-22 CRAN (R 3.4.2)
htmlTable                1.9      2017-01-26 CRAN (R 3.4.2)
htmltools                0.3.6    2017-04-28 CRAN (R 3.4.2)
htmlwidgets              0.9      2017-07-10 CRAN (R 3.4.2)
httpuv                   1.3.5    2017-07-04 CRAN (R 3.4.2)
httr                     1.3.1    2017-08-20 CRAN (R 3.4.2)
hwriter                  1.3.2    2014-09-10 CRAN (R 3.4.2)
interactiveDisplayBase   1.14.0   2017-10-13 Bioconductor
IRanges                * 2.10.5   2017-10-13 Bioconductor
iterators              * 1.0.8    2015-10-13 CRAN (R 3.4.2)
kableExtra             * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr                  * 1.17     2017-08-10 CRAN (R 3.4.2)
lattice                * 0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra             0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval                 0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr                 1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                   1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats            * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise                  1.1.0    2017-04-21 CRAN (R 3.4.2)
methods                * 3.4.2    2017-10-06 local
mime                     0.5      2016-07-07 CRAN (R 3.4.2)
munsell                  0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                     7.3-12   2016-02-02 CRAN (R 3.4.2)
parallel               * 3.4.2    2017-10-06 local
plyr                   * 1.8.4    2016-06-08 CRAN (R 3.4.2)
ProtGenerics             1.8.0    2017-10-13 Bioconductor
R6                       2.2.2    2017-06-17 CRAN (R 3.4.2)
RColorBrewer             1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                     0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                    1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                    1.1.1    2017-05-16 CRAN (R 3.4.2)
```

```
rlang                    0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown                1.6      2017-06-15 CRAN (R 3.4.2)
rpart                    4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot                1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools              * 1.28.0   2017-10-13 Bioconductor
RSQLite                  2.0      2017-06-19 CRAN (R 3.4.2)
rtracklayer            * 1.36.6   2017-10-13 Bioconductor
rvest                    0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors              * 0.14.7   2017-10-13 Bioconductor
scales                 * 0.5.0    2017-08-24 CRAN (R 3.4.2)
shiny                    1.0.5    2017-08-23 CRAN (R 3.4.2)
ShortRead              * 1.34.2   2017-10-13 Bioconductor
splines                  3.4.2    2017-10-06 local
stats                  * 3.4.2    2017-10-06 local
stats4                 * 3.4.2    2017-10-06 local
stringdist             * 0.9.4.6  2017-07-31 CRAN (R 3.4.2)
stringi                  1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                  1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment   * 1.6.5    2017-10-13 Bioconductor
survival               * 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                   1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                    3.4.2    2017-10-06 local
utils                  * 3.4.2    2017-10-06 local
VariantAnnotation        1.22.3   2017-10-13 Bioconductor
withr                    2.0.0    2017-07-28 CRAN (R 3.4.2)
XML                      3.98-1.9 2017-06-19 CRAN (R 3.4.2)
xml2                     1.1.1    2017-01-24 CRAN (R 3.4.2)
xtable                   1.8-2    2016-02-05 CRAN (R 3.4.2)
XVector                * 0.16.0   2017-10-13 Bioconductor
yaml                     2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc                 1.22.0   2017-10-13 Bioconductor
```

# Barcoded extraction and reduction from RNA samples

*Tomas Bjorklund*

*Wed Nov 1 15:03:54 2017*

This workflow identifies correct amplicons from in vivo & in vitro samples and extracts the barcode. Barcodes are then reduced using the starcode algorithm.

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

## Analyze tissue RNA

```
strt <- Sys.time()
load("data/multipleContfragmentsComplete.rda")
load("data/alignedLibraries.rda")
load("data/LUTdna.rda")

load.list <- read.table("input/loadlist.txt", header = FALSE, skip = 0, sep = "\t",
    stringsAsFactors = FALSE, fill = TRUE)

dataDir <- "seqFiles"
colnames(load.list) <- c("Name", "BaseName", "GroupName")

log.table <- data.table(Name = "Name", Reads = NA, Purity = NA, BCs = NA, SCdroppedBC = NA,
    allBCs = NA, scBCs = NA)

analyzeTissue <- function(indexNr) {
    # indexNr <- 1

    name <- unlist(strsplit(load.list$BaseName[indexNr], "/"))
    name <- name[!is.na(name)]
    if (length(name) == 2) {
        in.files <- list.files(paste(gsub("([\\])", "", dataDir), name[1], sep = "/"),
            pattern = paste(name[2], "*", sep = ""), full.names = TRUE)
    } else {
        in.files <- list.files(gsub("([\\])", "", dataDir), pattern = paste(name[1],
            "*", sep = ""), full.names = TRUE)
    }
    in.files.P5 <- in.files[grep("R1", in.files)]
    in.files.P7 <- in.files[grep("R2", in.files)]

    in.name.P5 <- tempfile(pattern = "P5_", tmpdir = tempdir(), fileext = ".fastq.gz")
    in.name.P7 <- tempfile(pattern = "P7_", tmpdir = tempdir(), fileext = ".fastq.gz")
    system(paste("cat '", paste(as.character(in.files.P5), collapse = "' '"),
        "' > ", in.name.P5, " 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)

    system(paste("cat '", paste(as.character(in.files.P7), collapse = "' '"),
        "' > ", in.name.P7, " 2>&1", sep = ""), intern = TRUE, ignore.stdout = FALSE)

    log.table$Name <- load.list$Name[indexNr]
    name.out <- log.table$Name

    # Selection of real amplicons ============================
```

```
out.name.P5 <- tempfile(pattern = "P5_", tmpdir = tempdir(), fileext = ".fastq.gz")
out.name.P7 <- tempfile(pattern = "P7_", tmpdir = tempdir(), fileext = ".fastq.gz")
command.args <- paste("overwrite=true k=10 rcomp=f skipr1=t qhdist=0 maskmiddle=t ",
    "hammingdistance=1 findbestmatch=t ordered=t threads=", detectCores(),
    " in=", in.name.P5, " in2=", in.name.P7, " outm=", out.name.P5, " outm2=",
    out.name.P7, " fliteral=", "CGCCACAACATCGAGGACGGCAGCGTG", sep = "")

sys.out <- system2(path.expand("~/bbmap/bbduk2.sh"), args = command.args,
    stdout = TRUE, stderr = TRUE)
log.table$Purity <- strsplit(sys.out[grep("Contaminants", sys.out)], split = "\t")[[1]][2]

in.name.P5 <- out.name.P5
in.name.P7 <- out.name.P7

log.table$Reads <- as.integer(system(paste("gunzip -c ", shQuote(gsub("([\\])",
    "", in.name.P5)), " | echo $((`wc -l`/4)) 2>&1", sep = ""), intern = TRUE,
    ignore.stdout = FALSE))  #Stores the read count utilized


# Extraction of barcodes ============================

out.name.BC <- tempfile(pattern = "BC_", tmpdir = tempdir(), fileext = ".fastq.gz")

sys.out <- system(paste("~/bbmap/bbduk2.sh overwrite=true k=12 mink=12 hammingdistance=2 ",
    "findbestmatch=t trd=t rcomp=f skipr2=t findbestmatch=f qhdist=0 ",
    "minavgquality=0 ordered=t maxns=0 minlength=18 maxlength=22 threads=",
    detectCores(), " in=", shQuote(in.name.P5), " out=", out.name.BC, " lliteral=",
    "GGCCTAGCGGCCGCTTTACTT", " rliteral=", "ATAACTTCGTATA", " 2>&1", sep = ""),
    intern = TRUE, ignore.stdout = FALSE)

log.table$BCs <- strsplit(sys.out[grep("Result:", sys.out)], split = "\t")[[1]][2]

reads.BC <- readFastq(out.name.BC)
barcodeTable <- data.table(ID = as.character(ShortRead::id(reads.BC)), BC = as.character(sread(reads.BC)),
    key = "BC")

# Starcode based barcode reduction ============================

out.name.BC.star <- tempfile(pattern = "BCsc_", tmpdir = tempdir(), fileext = ".txt")

system(paste("gunzip -c ", out.name.BC, " | starcode -t ", detectCores() -
    1, " --print-clusters -d", 1, " -r5 -q -o ", out.name.BC.star, " 2>&1",
    sep = ""), intern = TRUE, ignore.stdout = FALSE)

table.BC.sc <- data.table(read.table(out.name.BC.star, header = FALSE, row.names = 1,
    skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = FALSE), keep.rownames = TRUE,
    key = "rn")
table.BC.sc[, `:=`(V2, NULL)]

table.BC.sc <- table.BC.sc[, strsplit(as.character(V3), ",", fixed = TRUE),
    by = rn]

log.table$SCdroppedBC <- length(unique(sread(reads.BC))) - length(unique(table.BC.sc$V1) %in%
    unique(sread(reads.BC)))


setnames(table.BC.sc, c("V1", "rn"), c("BC", "scBC"))
```

```
    # Replacing barcodes with Starcode reduced versions
    # ============================


    setkey(table.BC.sc, BC)

    barcodeTable <- barcodeTable[table.BC.sc, nomatch = 0]

    setnames(barcodeTable, c("BC", "scBC"), c("oldBC", "BC"))

    setkey(barcodeTable, BC)

    log.table$allBCs <- length(unique(barcodeTable$oldBC))
    log.table$scBCs <- length(unique(barcodeTable$BC))

    invisible(barcodeTable[, `:=`(oldBC, NULL)])
    setkey(output.Table, "BC")

    BCcount <- data.table(as.data.frame(rev(sort(table(barcodeTable$BC))), row.names = "Var1"),
        keep.rownames = TRUE)
    # In R versions below 3.3 remove, row.names = 'Var1' to make this compatible
    setnames(BCcount, colnames(BCcount), c("BC", "RNAcount"))
    setkey(BCcount, "BC")
    foundFrags <- output.Table[BCcount, nomatch = 0]
    setkey(foundFrags, "LUTnr")
    setkey(LUT.dna, "LUTnr")
    foundFrags <- foundFrags[LUT.dna, nomatch = 0]
    setnames(foundFrags, "Sequence", "fragment")
    foundFrags[, `:=`(c("Names", "i.Structure"), NULL)]

    matchRange <- function(idxFrag) {
        # idxFrag <- 23
        matchRanges <- which(mcols(allFragments.ranges)$Sequence == foundFrags$fragment[idxFrag])
        return(cbind(matchRanges, idxFrag))
    }
    match.ranges.list <- mclapply(1:nrow(foundFrags), matchRange, mc.preschedule = TRUE,
        mc.cores = detectCores())
    match.ranges <- do.call(rbind, match.ranges.list)
    foundFragments.ranges <- allFragments.ranges[match.ranges[, 1]]
    if (ncol(match.ranges) >= 2) {
        foundFrags <- foundFrags[match.ranges[, "idxFrag"], ]
        foundFrags[, `:=`(c("Reads", "fragment", "Structure", "LUTnr"), NULL)]
        mcols(foundFragments.ranges) <- c(mcols(foundFragments.ranges), foundFrags)
        o = order(-mcols(foundFragments.ranges)$RNAcount)
        foundFragments.ranges <- foundFragments.ranges[o]
        saveRDS(foundFragments.ranges, file = paste("output/", "found.", name.out,
            ".rds", sep = ""), compress = TRUE)
    }
    return(log.table)
}
```

## Analysis summary

```
all.logs <- lapply(1:nrow(load.list), analyzeTissue)
all.logs <- rbindlist(all.logs, use.names = FALSE)
knitr::kable(all.logs, format = "latex", longtable = T, booktabs = T) %>% kable_styling(latex_options = c("s
    "scale_down", "repeat_header")) %>% landscape()
```

```
Warning in styling_latex_scale_down(out, table_info): Longtable cannot be
resized.
```

| Name | Reads | Purity | BCs | SCdroppedBC | allBCs | scBCs |
|---|---|---|---|---|---|---|
| DNA_pscAAVlib_Prep2 | 41210335 | 82420670 reads (99.65%) | 23986369 reads (58.20%) | 60046 | 4940426 | 3938187 |
| DNA_AAVlib_DNAse_3cpc | 7964874 | 15929748 reads (99.66%) | 2985701 reads (37.49%) | 154 | 341674 | 223158 |
| DNA_AAVlib_DNAse_30cpc | 17557643 | 35115286 reads (99.69%) | 8827549 reads (50.28%) | 2464 | 1157571 | 795330 |
| mRNA_30cpc_SN_RatNr7 | 1983137 | 3966274 reads (99.65%) | 1614014 reads (81.39%) | 41 | 24691 | 9962 |
| mRNA_30cpc_Ctx_RatNr7 | 1994467 | 3988934 reads (99.60%) | 1771085 reads (88.80%) | 13 | 16108 | 7547 |
| mRNA_30cpc_Th_RatNr7 | 2867972 | 5735944 reads (99.65%) | 1382945 reads (48.22%) | 12 | 30889 | 11632 |
| mRNA_30cpc_Str_RatNr7 | 1596468 | 3192936 reads (99.62%) | 991144 reads (62.08%) | 15 | 60701 | 21186 |
| mRNA_30cpc_SN_RatNr1 | 1611759 | 3223518 reads (99.67%) | 1361550 reads (84.48%) | 47 | 11471 | 6081 |
| mRNA_30cpc_Ctx_RatNr1 | 1541657 | 3083314 reads (99.67%) | 1203038 reads (78.04%) | 9 | 11542 | 7747 |
| mRNA_30cpc_Th_RatNr1 | 2359538 | 4719076 reads (99.67%) | 1026288 reads (43.50%) | 1 | 17976 | 8587 |
| mRNA_30cpc_Str_RatNr1 | 1505088 | 3010176 reads (99.60%) | 619506 reads (41.16%) | 6 | 37192 | 13397 |
| mRNA_30cpc_SN_RatNr8 | 2054931 | 4109862 reads (99.70%) | 1215848 reads (59.17%) | 80 | 22685 | 11414 |
| mRNA_30cpc_Ctx_RatNr8 | 2102816 | 4205632 reads (99.69%) | 1081123 reads (51.41%) | 18 | 13376 | 7115 |
| mRNA_30cpc_Th_RatNr8 | 2105768 | 4211536 reads (99.69%) | 856547 reads (40.68%) | 9 | 29605 | 12444 |
| mRNA_30cpc_Str_RatNr8 | 1623686 | 3247372 reads (99.64%) | 612950 reads (37.75%) | 9 | 45404 | 20175 |
| mRNA_3cpc_SN_RatNr15 | 1436268 | 2872536 reads (99.64%) | 1435201 reads (99.93%) | 17 | 4226 | 1538 |
| mRNA_3cpc_Ctx_RatNr15 | 1260242 | 2520484 reads (99.65%) | 1062278 reads (84.29%) | 12 | 6434 | 3211 |
| mRNA_3cpc_Th_RatNr15 | 1105966 | 2211932 reads (99.65%) | 944716 reads (85.42%) | 3 | 5642 | 2818 |
| mRNA_3cpc_Str_RatNr15 | 948187 | 1896374 reads (99.65%) | 684628 reads (72.20%) | 2 | 22301 | 6617 |
| mRNA_3cpc_SN_RatNr21 | 1115267 | 2230534 reads (99.75%) | 1112638 reads (99.76%) | 3 | 4489 | 2021 |
| mRNA_3cpc_Ctx_RatNr21 | 1201263 | 2402526 reads (99.73%) | 788753 reads (65.66%) | 2 | 3920 | 1926 |
| mRNA_3cpc_Th_RatNr21 | 1234915 | 2469830 reads (99.73%) | 1063068 reads (86.08%) | 0 | 7414 | 3780 |
| mRNA_3cpc_Str_RatNr21 | 1151549 | 2303098 reads (99.71%) | 845867 reads (73.45%) | 3 | 29847 | 8603 |
| mRNA_3cpc_Ctx_RatNr19 | 1743952 | 3487904 reads (99.61%) | 1738647 reads (99.70%) | 15 | 6568 | 2765 |
| mRNA_3cpc_Th_RatNr19 | 1738722 | 3477444 reads (99.68%) | 1174178 reads (67.53%) | 21 | 10753 | 6535 |
| mRNA_3cpc_Str_RatNr19 | 1645506 | 3291012 reads (99.67%) | 1196277 reads (72.70%) | 5 | 34666 | 12990 |
| mRNA_3cpc_Th_RatNr20 | 788088 | 1576176 reads (99.61%) | 682035 reads (86.54%) | 3 | 5098 | 2470 |
| mRNA_3cpc_Str_RatNr20 | 983263 | 1966526 reads (99.59%) | 827836 reads (84.19%) | 3 | 23426 | 8334 |
| mRNA_3cpc_HEK293Nr2 | 1588150 | 3176300 reads (99.63%) | 1016131 reads (63.98%) | 11 | 7051 | 3030 |
| mRNA_30cpc_HEK293Nr3 | 1920390 | 3840780 reads (99.65%) | 864987 reads (45.04%) | 13 | 20920 | 7730 |
| mRNA_3cpc_pNeuronNr6 | 1143395 | 2286790 reads (99.56%) | 552817 reads (48.35%) | 4 | 6313 | 2553 |
| mRNA_30cpc_pNeuronNr7 | 1652464 | 3304928 reads (99.66%) | 942060 reads (57.01%) | 13 | 24858 | 8032 |
| mRNA_30cpc_4wks_Ctx_RatNr2 | 1955085 | 3910170 reads (99.25%) | 591146 reads (30.24%) | 3 | 5607 | 2561 |
| mRNA_30cpc_4wks_SN_RatNr2 | 1921383 | 3842766 reads (99.39%) | 1210727 reads (63.01%) | 10 | 18161 | 8012 |
| mRNA_30cpc_4wks_Str_RatNr2 | 2101122 | 4202244 reads (100.00%) | 1563614 reads (74.42%) | 24 | 84742 | 59375 |
| mRNA_30cpc_4wks_Th_RatNr2 | 2177483 | 4354966 reads (95.47%) | 752026 reads (34.54%) | 0 | 16068 | 5602 |
| mRNA_3cpc_4wks_Ctx_RatNr13 | 2070763 | 4141526 reads (99.61%) | 183573 reads (8.86%) | 1 | 1403 | 674 |
| mRNA_3cpc_4wks_SN_RatNr13 | 1809233 | 3618466 reads (99.47%) | 105370 reads (5.82%) | 2 | 1696 | 1249 |

*(continued)*

| Name | Reads | Purity | BCs | SCdroppedBC | allBCs | scBCs |
|---|---|---|---|---|---|---|
| mRNA_3cpc_4wks_Str_RatNr13 | 1693037 | 3386074 reads (99.32%) | 1096589 reads (64.77%) | 2 | 17281 | 6400 |
| mRNA_3cpc_4wks_Th_RatNr13 | 1954529 | 3909058 reads (99.61%) | 1079929 reads (55.25%) | 27 | 12266 | 5208 |

```
unlink(paste(tempdir(), "/*", sep = ""), recursive = FALSE, force = FALSE)  #Cleanup of temp files

print("Total execution time:")
```

```
[1] "Total execution time:"
```

```
print(Sys.time() - strt)
```

```
Time difference of 1.177768 hours
```

```
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     2017-11-01
```

```
Packages -----------------------------------------------------------------
```

```
 package               * version  date       source
 acepack                 1.4.1    2016-10-29 CRAN (R 3.4.2)
 AnnotationDbi         * 1.38.2   2017-10-13 Bioconductor
 AnnotationFilter        1.0.0    2017-10-13 Bioconductor
 AnnotationHub           2.8.2    2017-10-13 Bioconductor
 backports               1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                  * 3.4.2    2017-10-06 local
 base64enc               0.1-3    2015-07-28 CRAN (R 3.4.2)
 beanplot              * 1.2      2014-09-19 CRAN (R 3.4.2)
 Biobase               * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics          * 0.22.1   2017-10-13 Bioconductor
 BiocInstaller           1.26.1   2017-10-10 Bioconductor
 BiocParallel          * 1.10.1   2017-10-13 Bioconductor
 biomaRt                 2.32.1   2017-10-13 Bioconductor
 Biostrings            * 2.44.2   2017-10-13 Bioconductor
 biovizBase            * 1.24.0   2017-10-13 Bioconductor
 bit                     1.1-12   2014-04-09 CRAN (R 3.4.2)
 bit64                   0.9-7    2017-05-08 CRAN (R 3.4.2)
 bitops                  1.0-6    2013-08-17 CRAN (R 3.4.2)
 blob                    1.1.0    2017-06-17 CRAN (R 3.4.2)
 BSgenome              * 1.44.2   2017-10-13 Bioconductor
 checkmate               1.8.4    2017-09-25 CRAN (R 3.4.2)
 cluster                 2.0.6    2017-03-16 CRAN (R 3.4.2)
 codetools               0.2-15   2016-10-05 CRAN (R 3.4.2)
 colorspace              1.3-2    2016-12-14 CRAN (R 3.4.2)
 compiler                3.4.2    2017-10-06 local
 curl                    2.8.1    2017-07-21 CRAN (R 3.4.2)
 data.table            * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets              * 3.4.2    2017-10-06 local
 DBI                     0.7      2017-06-18 CRAN (R 3.4.2)
 DelayedArray          * 0.2.7    2017-10-13 Bioconductor
 devtools              * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 dichromat               2.0-0    2013-01-24 CRAN (R 3.4.2)
 digest                  0.6.12   2017-01-27 CRAN (R 3.4.2)
 doParallel            * 1.0.11   2017-09-28 CRAN (R 3.4.2)
```

```
ensembldb                2.0.4    2017-10-13 Bioconductor
evaluate                 0.10.1   2017-06-24 CRAN (R 3.4.2)
foreach                * 1.4.3    2015-10-13 CRAN (R 3.4.2)
foreign                  0.8-69   2017-06-21 CRAN (R 3.4.2)
formatR                * 1.5      2017-04-25 CRAN (R 3.4.2)
Formula                  1.2-2    2017-07-10 CRAN (R 3.4.2)
GenomeInfoDb           * 1.12.3   2017-10-13 Bioconductor
GenomeInfoDbData         0.99.0   2017-10-13 Bioconductor
GenomicAlignments      * 1.12.2   2017-10-13 Bioconductor
GenomicFeatures        * 1.28.5   2017-10-13 Bioconductor
GenomicRanges          * 1.28.6   2017-10-13 Bioconductor
GGally                   1.3.2    2017-08-02 CRAN (R 3.4.2)
ggbio                  * 1.24.1   2017-10-13 Bioconductor
ggplot2                * 2.2.1    2016-12-30 CRAN (R 3.4.2)
graph                    1.54.0   2017-10-13 Bioconductor
graphics               * 3.4.2    2017-10-06 local
grDevices              * 3.4.2    2017-10-06 local
grid                   * 3.4.2    2017-10-06 local
gridExtra                2.3      2017-09-09 CRAN (R 3.4.2)
gtable                   0.2.0    2016-02-26 CRAN (R 3.4.2)
Gviz                   * 1.20.0   2017-10-13 Bioconductor
Hmisc                    4.0-3    2017-05-02 CRAN (R 3.4.2)
hms                      0.3      2016-11-22 CRAN (R 3.4.2)
htmlTable                1.9      2017-01-26 CRAN (R 3.4.2)
htmltools                0.3.6    2017-04-28 CRAN (R 3.4.2)
htmlwidgets              0.9      2017-07-10 CRAN (R 3.4.2)
httpuv                   1.3.5    2017-07-04 CRAN (R 3.4.2)
httr                     1.3.1    2017-08-20 CRAN (R 3.4.2)
hwriter                  1.3.2    2014-09-10 CRAN (R 3.4.2)
interactiveDisplayBase   1.14.0   2017-10-13 Bioconductor
IRanges                * 2.10.5   2017-10-13 Bioconductor
iterators              * 1.0.8    2015-10-13 CRAN (R 3.4.2)
kableExtra             * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr                  * 1.17     2017-08-10 CRAN (R 3.4.2)
lattice                  0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra             0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval                 0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr                 1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                   1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats            * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise                  1.1.0    2017-04-21 CRAN (R 3.4.2)
methods                * 3.4.2    2017-10-06 local
mime                     0.5      2016-07-07 CRAN (R 3.4.2)
munsell                  0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                     7.3-12   2016-02-02 CRAN (R 3.4.2)
OrganismDbi              1.18.1   2017-10-13 Bioconductor
parallel               * 3.4.2    2017-10-06 local
plyr                   * 1.8.4    2016-06-08 CRAN (R 3.4.2)
ProtGenerics             1.8.0    2017-10-13 Bioconductor
R6                       2.2.2    2017-06-17 CRAN (R 3.4.2)
RBGL                     1.52.0   2017-10-13 Bioconductor
RColorBrewer             1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                     0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                    1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                    1.1.1    2017-05-16 CRAN (R 3.4.2)
reshape                  0.8.7    2017-08-06 CRAN (R 3.4.2)
reshape2                 1.4.2    2016-10-22 CRAN (R 3.4.2)
```

```
rlang                    0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown                1.6      2017-06-15 CRAN (R 3.4.2)
rpart                    4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot                1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools              * 1.28.0   2017-10-13 Bioconductor
RSQLite                  2.0      2017-06-19 CRAN (R 3.4.2)
rtracklayer            * 1.36.6   2017-10-13 Bioconductor
rvest                    0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors              * 0.14.7   2017-10-13 Bioconductor
scales                 * 0.5.0    2017-08-24 CRAN (R 3.4.2)
shiny                    1.0.5    2017-08-23 CRAN (R 3.4.2)
ShortRead              * 1.34.2   2017-10-13 Bioconductor
splines                  3.4.2    2017-10-06 local
stats                  * 3.4.2    2017-10-06 local
stats4                 * 3.4.2    2017-10-06 local
stringi                  1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                  1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment   * 1.6.5    2017-10-13 Bioconductor
survival                 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                   1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                    3.4.2    2017-10-06 local
utils                  * 3.4.2    2017-10-06 local
VariantAnnotation        1.22.3   2017-10-13 Bioconductor
withr                    2.0.0    2017-07-28 CRAN (R 3.4.2)
XML                      3.98-1.9 2017-06-19 CRAN (R 3.4.2)
xml2                     1.1.1    2017-01-24 CRAN (R 3.4.2)
xtable                   1.8-2    2016-02-05 CRAN (R 3.4.2)
XVector                * 0.16.0   2017-10-13 Bioconductor
yaml                     2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc                 1.22.0   2017-10-13 Bioconductor
```

# Generate a complete library range object

*Tomas Bjorklund*

*Wed Nov 1 16:14:54 2017*

This short script generates a lookup scoring table of the AAV plasmid library so that it follows the same structure as the mRNA samples so that they can be compared for coverages.

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

## Generate library rage object

```
load("data/alignedLibraries.rda")
load("data/LUTdna.rda")
load("data/multipleContfragmentsComplete.rda")
setkey(output.Table, LUTnr)
setkey(LUT.dna, LUTnr)
output.Table <- output.Table[LUT.dna, nomatch = 0]
output.Table[, `:=`(c("Names", "i.Structure"), NULL)]
setnames(output.Table, "Sequence", "fragment")
setkey(output.Table, fragment)

range.idx <- data.table(fragment = mcols(allFragments.ranges)$Sequence, idxFrag = 1:length(allFragments.rang
    key = "fragment")
output.Table <- output.Table[range.idx, nomatch = 0, allow.cartesian = TRUE]

foundFragments.ranges <- allFragments.ranges[output.Table$idxFrag]
output.Table[, `:=`(c("Reads", "fragment", "idxFrag", "Structure", "LUTnr"),
    NULL)]
output.Table[, `:=`(RNAcount, tCount)]

mcols(foundFragments.ranges) <- c(mcols(foundFragments.ranges), output.Table)

saveRDS(foundFragments.ranges, file = "output/completeLibraryRanges.rds")

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     2017-11-01
Packages -----------------------------------------------------------------
 package             * version  date       source
 backports             1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                * 3.4.2    2017-10-06 local
 Biobase             * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics        * 0.22.1   2017-10-13 Bioconductor
```

```
BiocParallel         * 1.10.1   2017-10-13 Bioconductor
Biostrings           * 2.44.2   2017-10-13 Bioconductor
bitops                 1.0-6    2013-08-17 CRAN (R 3.4.2)
codetools              0.2-15   2016-10-05 CRAN (R 3.4.2)
compiler               3.4.2    2017-10-06 local
data.table           * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
datasets             * 3.4.2    2017-10-06 local
DelayedArray         * 0.2.7    2017-10-13 Bioconductor
devtools             * 1.13.3   2017-08-02 CRAN (R 3.4.2)
digest                 0.6.12   2017-01-27 CRAN (R 3.4.2)
doParallel           * 1.0.11   2017-09-28 CRAN (R 3.4.2)
evaluate               0.10.1   2017-06-24 CRAN (R 3.4.2)
foreach              * 1.4.3    2015-10-13 CRAN (R 3.4.2)
formatR                1.5      2017-04-25 CRAN (R 3.4.2)
GenomeInfoDb         * 1.12.3   2017-10-13 Bioconductor
GenomeInfoDbData       0.99.0   2017-10-13 Bioconductor
GenomicAlignments    * 1.12.2   2017-10-13 Bioconductor
GenomicRanges        * 1.28.6   2017-10-13 Bioconductor
graphics             * 3.4.2    2017-10-06 local
grDevices            * 3.4.2    2017-10-06 local
grid                   3.4.2    2017-10-06 local
hms                    0.3      2016-11-22 CRAN (R 3.4.2)
htmltools              0.3.6    2017-04-28 CRAN (R 3.4.2)
httr                   1.3.1    2017-08-20 CRAN (R 3.4.2)
hwriter                1.3.2    2014-09-10 CRAN (R 3.4.2)
IRanges              * 2.10.5   2017-10-13 Bioconductor
iterators            * 1.0.8    2015-10-13 CRAN (R 3.4.2)
kableExtra           * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr                * 1.17     2017-08-10 CRAN (R 3.4.2)
lattice                0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra           0.6-28   2016-02-09 CRAN (R 3.4.2)
magrittr               1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                 1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats          * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise                1.1.0    2017-04-21 CRAN (R 3.4.2)
methods              * 3.4.2    2017-10-06 local
parallel             * 3.4.2    2017-10-06 local
R6                     2.2.2    2017-06-17 CRAN (R 3.4.2)
RColorBrewer           1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                   0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                  1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                  1.1.1    2017-05-16 CRAN (R 3.4.2)
rlang                  0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown              1.6      2017-06-15 CRAN (R 3.4.2)
rprojroot              1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools            * 1.28.0   2017-10-13 Bioconductor
rvest                  0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors            * 0.14.7   2017-10-13 Bioconductor
ShortRead            * 1.34.2   2017-10-13 Bioconductor
stats                * 3.4.2    2017-10-06 local
stats4               * 3.4.2    2017-10-06 local
stringi                1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment * 1.6.5    2017-10-13 Bioconductor
tibble                 1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                  3.4.2    2017-10-06 local
utils                * 3.4.2    2017-10-06 local
```

```
withr                 2.0.0   2017-07-28 CRAN (R 3.4.2)
xml2                  1.1.1   2017-01-24 CRAN (R 3.4.2)
XVector             * 0.16.0  2017-10-13 Bioconductor
yaml                  2.1.14  2016-11-12 CRAN (R 3.4.2)
zlibbioc              1.22.0  2017-10-13 Bioconductor
```

# Normalize Library counts

*Tomas Bjorklund*

*Wed Nov 1 16:16:12 2017*

This workflow normalizes read counts between samples to compensate for variable read depth.

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

## Generate load list and grouping names

```
strt <- Sys.time()
in.names.all <- list.files("output", pattern = "*.rds", full.names = TRUE)
load.list <- read.table("input/loadlist.txt", header = FALSE, skip = 0, sep = "\t",
    stringsAsFactors = FALSE, fill = TRUE)
colnames(load.list) <- c("Name", "BaseName", "GroupName")
load.list <- rbind(load.list, c("completeLibraryRanges", "", "DNA_pscAAVlib"))
load.list <- load.list[!grepl("Untreat", load.list$Name), ]

select.Cases <- c(unlist(sapply(load.list$Name, function(x) grep(x, in.names.all),
    simplify = TRUE)))

(in.names.all <- in.names.all[select.Cases])
```

```
 [1] "output/found.DNA_pscAAVlib_Prep2.rds"
 [2] "output/found.DNA_AAVlib_DNAse_3cpc.rds"
 [3] "output/found.DNA_AAVlib_DNAse_30cpc.rds"
 [4] "output/found.mRNA_30cpc_SN_RatNr7.rds"
 [5] "output/found.mRNA_30cpc_Ctx_RatNr7.rds"
 [6] "output/found.mRNA_30cpc_Th_RatNr7.rds"
 [7] "output/found.mRNA_30cpc_Str_RatNr7.rds"
 [8] "output/found.mRNA_30cpc_SN_RatNr1.rds"
 [9] "output/found.mRNA_30cpc_Ctx_RatNr1.rds"
[10] "output/found.mRNA_30cpc_Th_RatNr1.rds"
[11] "output/found.mRNA_30cpc_Str_RatNr1.rds"
[12] "output/found.mRNA_30cpc_SN_RatNr8.rds"
[13] "output/found.mRNA_30cpc_Ctx_RatNr8.rds"
[14] "output/found.mRNA_30cpc_Th_RatNr8.rds"
[15] "output/found.mRNA_30cpc_Str_RatNr8.rds"
[16] "output/found.mRNA_3cpc_SN_RatNr15.rds"
[17] "output/found.mRNA_3cpc_Ctx_RatNr15.rds"
[18] "output/found.mRNA_3cpc_Th_RatNr15.rds"
[19] "output/found.mRNA_3cpc_Str_RatNr15.rds"
[20] "output/found.mRNA_3cpc_SN_RatNr21.rds"
[21] "output/found.mRNA_3cpc_Ctx_RatNr21.rds"
[22] "output/found.mRNA_3cpc_Th_RatNr21.rds"
[23] "output/found.mRNA_3cpc_Str_RatNr21.rds"
[24] "output/found.mRNA_3cpc_Ctx_RatNr19.rds"
[25] "output/found.mRNA_3cpc_Th_RatNr19.rds"
[26] "output/found.mRNA_3cpc_Str_RatNr19.rds"
[27] "output/found.mRNA_3cpc_Th_RatNr20.rds"
[28] "output/found.mRNA_3cpc_Str_RatNr20.rds"
[29] "output/found.mRNA_3cpc_HEK293Nr2.rds"
[30] "output/found.mRNA_30cpc_HEK293Nr3.rds"
```

```
[31] "output/found.mRNA_3cpc_pNeuronNr6.rds"
[32] "output/found.mRNA_30cpc_pNeuronNr7.rds"
[33] "output/found.mRNA_30cpc_4wks_Ctx_RatNr2.rds"
[34] "output/found.mRNA_30cpc_4wks_SN_RatNr2.rds"
[35] "output/found.mRNA_30cpc_4wks_Str_RatNr2.rds"
[36] "output/found.mRNA_30cpc_4wks_Th_RatNr2.rds"
[37] "output/found.mRNA_3cpc_4wks_Ctx_RatNr13.rds"
[38] "output/found.mRNA_3cpc_4wks_SN_RatNr13.rds"
[39] "output/found.mRNA_3cpc_4wks_Str_RatNr13.rds"
[40] "output/found.mRNA_3cpc_4wks_Th_RatNr13.rds"
[41] "output/completeLibraryRanges.rds"
```

```
grouping <- data.frame(Sample = gsub("-", "_", gsub("found.|(output/)|(.rds)",
    "", in.names.all)), Group = load.list[match(names(select.Cases), load.list$Name),
    "GroupName"], stringsAsFactors = FALSE)
```

## Load the desired alignment files and annotating group

```
loadRDS <- function(in.name) {
    # in.name <- in.names.all[42]
    this.sample <- readRDS(in.name)
    this.name <- gsub("-", "_", gsub("found.|(output/)|(.rds)", "", in.name))
    this.group <- grouping[match(this.name, grouping$Sample), "Group"]
    mcols(this.sample) <- cbind(mcols(this.sample), data.frame(Sample = this.name,
        Group = this.group, stringsAsFactors = FALSE))

    return(this.sample)
}

all.samples <- lapply(in.names.all, loadRDS)

all.samples <- do.call(GAlignmentsList, unlist(all.samples))
all.samples <- cbind(unlist(all.samples))[[1]]

names(all.samples) <- make.names(names(all.samples), unique = TRUE)
length.Table <- data.table(seqnames = names(seqlengths(all.samples)), seqlength = seqlengths(all.samples),
    key = "seqnames")
all.samples <- data.table(as.data.frame(all.samples), key = "seqnames")
all.samples[, `:=`(c("strand", "qwidth", "cigar", "njunc", "end"), NULL)]
all.samples <- all.samples[length.Table]  #A data.table merge to match seqlengths to their respective seqna
all.samples[, `:=`(c("Category", "Protein", "Origin", "Extra", "Number", "GeneName"),
    tstrsplit(seqnames, ",", fixed = TRUE))]
all.samples[, `:=`(c("seqnames", "Protein", "Origin", "Extra", "Number"), NULL)]
all.samples[, `:=`(GeneName, gsub("/|_", "-", GeneName))]
```

## Normalizing read counts to correct for variable read depth

```
setkey(all.samples, Group)
all.samples <- all.samples[RNAcount > 1, ]  #Filters out single count reads
readCounts <- all.samples[, list(GroupCount = sum(RNAcount)), by = "Group"]
readCounts[, `:=`(GroupCount, GroupCount/max(GroupCount))]
setkey(readCounts, Group)
all.samples <- all.samples[readCounts]  #Merge with normalizing factor
```

```
all.samples[, `:=`(RNAcount, RNAcount/GroupCount)]
setkey(all.samples, Mode)
all.samples <- all.samples["Def"]

setkey(all.samples, Group)
total.AAV.samples <- all.samples[Group != "DNA_pscAAVlib" & Group != "DNA_pscAAVlib_Prep2" &
    Group != "DNA_AAVlib_DNAse_3cpc" & Group != "DNA_AAVlib_DNAse_30cpc"]
# total.AAV.samples <-
# total.AAV.samples[!grepl('4wks',total.AAV.samples$Group)]
transported.AAV.samples.30cpc <- total.AAV.samples[grepl("mRNA_30cpc_SN|mRNA_30cpc_Th|mRNA_30cpc_Ctx",
    total.AAV.samples$Group)]
transported.AAV.samples.3cpc <- total.AAV.samples[grepl("mRNA_3cpc_SN|30cpc_Th|mRNA_3cpc_Ctx",
    total.AAV.samples$Group)]
total.AAV.samples[, `:=`(Group, "mRNA_All")]
transported.AAV.samples.30cpc[, `:=`(Group, "mRNA_30cpc_Trsp")]
transported.AAV.samples.3cpc[, `:=`(Group, "mRNA_3cpc_Trsp")]

all.samples <- rbind(all.samples, total.AAV.samples, transported.AAV.samples.30cpc,
    transported.AAV.samples.3cpc)

rm(total.AAV.samples, transported.AAV.samples.30cpc, transported.AAV.samples.3cpc)

setkeyv(all.samples, c("Group", "Category", "GeneName", "structure", "start",
    "width", "Sequence", "seqlength"))

all.samples <- all.samples[, j = list(bitScore = sum(bitScore * tCount)/sum(tCount),
    mismatches = median(mismatches), mCount = sum(mCount), tCount = sum(tCount),
    BC = paste(unique(BC), collapse = ","), Animals = paste(unique(Sample),
        collapse = ","), LUTnrs = paste(unique(LUTnr), collapse = ","), RNAcount = sum(RNAcount),
    NormCount = log2(sum(RNAcount) + 1) * .N), by = c("Group", "Category", "GeneName",
    "structure", "start", "width", "Sequence", "seqlength")]

all.samples[, `:=`(start, floor((start + 2)/3))]
all.samples[, `:=`(width, ceiling((width)/3))]
all.samples[, `:=`(seqlength, ceiling(seqlength/3))]
all.samples[, `:=`(AA, floor(start + (width/2)))]
all.samples[, `:=`(AAproc, AA/seqlength * 100)]
```

## Remove overhangs on the sequence based on the Structure annotation

```
all.samples[structure == "14aa", `:=`("Sequence", substr(Sequence, 3, 44))]
all.samples[structure == "22aa", `:=`("Sequence", substr(Sequence, 3, 68))]
all.samples[structure == "14aaG4S", `:=`("Sequence", substr(Sequence, 15, 56))]
all.samples[structure == "14aaA5", `:=`("Sequence", substr(Sequence, 15, 56))]

# Change the default behavior to induce start codons and Methionine
GENETIC_CODE_ALT <- GENETIC_CODE
attr(GENETIC_CODE_ALT, "alt_init_codons") <- c("TAA", "TAG")

all.samples[, `:=`(Peptide, mclapply(Sequence, function(x) as.character(Biostrings::translate(DNAString(x),
    genetic.code = GENETIC_CODE_ALT, if.fuzzy.codon = "solve")), mc.cores = detectCores()))]
all.samples[, `:=`(Peptide, as.character(Peptide)), ]
saveRDS(all.samples, file = "data/allSamplesDataTable.RDS")

print("Total execution time:")
```

```
[1] "Total execution time:"
```

```
print(Sys.time() - strt)
```

```
Time difference of 1.321298 hours
```

```
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     2017-11-01
```

```
Packages -----------------------------------------------------------------
```

```
 package              * version  date       source
 acepack                1.4.1    2016-10-29 CRAN (R 3.4.2)
 ade4                   1.7-8    2017-08-09 CRAN (R 3.4.2)
 backports              1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                 * 3.4.2    2017-10-06 local
 base64enc              0.1-3    2015-07-28 CRAN (R 3.4.2)
 Biobase              * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics         * 0.22.1   2017-10-13 Bioconductor
 BiocParallel         * 1.10.1   2017-10-13 Bioconductor
 Biostrings           * 2.44.2   2017-10-13 Bioconductor
 bitops                 1.0-6    2013-08-17 CRAN (R 3.4.2)
 checkmate              1.8.4    2017-09-25 CRAN (R 3.4.2)
 cluster                2.0.6    2017-03-16 CRAN (R 3.4.2)
 colorspace             1.3-2    2016-12-14 CRAN (R 3.4.2)
 compiler               3.4.2    2017-10-06 local
 data.table           * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets             * 3.4.2    2017-10-06 local
 DelayedArray         * 0.2.7    2017-10-13 Bioconductor
 devtools             * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 digest                 0.6.12   2017-01-27 CRAN (R 3.4.2)
 evaluate               0.10.1   2017-06-24 CRAN (R 3.4.2)
 foreign                0.8-69   2017-06-21 CRAN (R 3.4.2)
 formatR                1.5      2017-04-25 CRAN (R 3.4.2)
 Formula              * 1.2-2    2017-07-10 CRAN (R 3.4.2)
 GenomeInfoDb         * 1.12.3   2017-10-13 Bioconductor
 GenomeInfoDbData       0.99.0   2017-10-13 Bioconductor
 GenomicAlignments    * 1.12.2   2017-10-13 Bioconductor
 GenomicRanges        * 1.28.6   2017-10-13 Bioconductor
 ggplot2              * 2.2.1    2016-12-30 CRAN (R 3.4.2)
 graphics             * 3.4.2    2017-10-06 local
 grDevices            * 3.4.2    2017-10-06 local
 grid                   3.4.2    2017-10-06 local
 gridExtra              2.3      2017-09-09 CRAN (R 3.4.2)
 gtable                 0.2.0    2016-02-26 CRAN (R 3.4.2)
 Hmisc                * 4.0-3    2017-05-02 CRAN (R 3.4.2)
 hms                    0.3      2016-11-22 CRAN (R 3.4.2)
 htmlTable              1.9      2017-01-26 CRAN (R 3.4.2)
 htmltools              0.3.6    2017-04-28 CRAN (R 3.4.2)
```

```
htmlwidgets            0.9      2017-07-10 CRAN (R 3.4.2)
httr                   1.3.1    2017-08-20 CRAN (R 3.4.2)
hwriter                1.3.2    2014-09-10 CRAN (R 3.4.2)
IRanges              * 2.10.5   2017-10-13 Bioconductor
kableExtra           * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr                * 1.17     2017-08-10 CRAN (R 3.4.2)
lattice              * 0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra           0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval               0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr               1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                 1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats          * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise                1.1.0    2017-04-21 CRAN (R 3.4.2)
methods              * 3.4.2    2017-10-06 local
multicore            * 0.2      2014-05-17 url
munsell                0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                   7.3-12   2016-02-02 CRAN (R 3.4.2)
parallel             * 3.4.2    2017-10-06 local
plyr                 * 1.8.4    2016-06-08 CRAN (R 3.4.2)
R6                     2.2.2    2017-06-17 CRAN (R 3.4.2)
RColorBrewer           1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                   0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                  1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                  1.1.1    2017-05-16 CRAN (R 3.4.2)
rlang                  0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown              1.6      2017-06-15 CRAN (R 3.4.2)
rpart                  4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot              1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools            * 1.28.0   2017-10-13 Bioconductor
rvest                  0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors            * 0.14.7   2017-10-13 Bioconductor
scales               * 0.5.0    2017-08-24 CRAN (R 3.4.2)
seqinr               * 3.4-5    2017-08-01 CRAN (R 3.4.2)
ShortRead            * 1.34.2   2017-10-13 Bioconductor
splines                3.4.2    2017-10-06 local
stats                * 3.4.2    2017-10-06 local
stats4               * 3.4.2    2017-10-06 local
stringi                1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment * 1.6.5    2017-10-13 Bioconductor
survival             * 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                 1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                  3.4.2    2017-10-06 local
utils                * 3.4.2    2017-10-06 local
withr                  2.0.0    2017-07-28 CRAN (R 3.4.2)
xml2                   1.1.1    2017-01-24 CRAN (R 3.4.2)
XVector              * 0.16.0   2017-10-13 Bioconductor
yaml                   2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc               1.22.0   2017-10-13 Bioconductor
```

# Pairwise sample analysis output

*Tomas Bjorklund*

*Tue Nov 7 10:55:17 2017*

This script presents overview ploits and top candidates.

```
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

```
plotPair <- function(topSample, bottomSample, size.bin = 1, winWidth = 1, NormalizePlot = TRUE,
    PlotBC = TRUE) {
    # Select samples ===================

    # topSample <- 'mRNA_3cpc_Ctx' bottomSample <- 'mRNA_30cpc_Ctx' filterBC <-
    # FALSE filterAnimal <- FALSE AnimaladjustPlot <- FALSE NormalizePlot <-
    # TRUE size.bin <- 1 winWidth=1 PlotBC=TRUE

    fill.values <- eval(parse(text = paste("c(", topSample, "= rgb(38,64,135, maxColorValue = 255), ",
        bottomSample, "= rgb(157,190,217, maxColorValue = 255))", sep = "")))
    setkey(all.samples, Group)
    select.samples <- all.samples[J(names(fill.values))]  #Select the two compared groups
    select.samples[, `:=`(RNAcount, log2(RNAcount + 1))]

    # if (PlotBC){
    # select.samples[,c('meanCount','SDcount','minCount'):=list(mean(RNAcount),
    # sd(RNAcount), min(RNAcount)),by='Group'] select.samples <-
    # select.samples[RNAcount>=minCount+(2*SDcount),] #-SDcount
    # select.samples[,c('meanCount','SDcount','minCount'):=NULL] }




    if (winWidth > 0) {
        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(width - winWidth + 1) + start - 1),
            by = c("GeneName", "start", "width")]
        setnames(windowTable, "V1", "winStart")
        windowTable[, `:=`(winEnd, winStart + winWidth - 1)]
        setkeyv(windowTable, c("GeneName", "start", "width"))
        setkeyv(select.samples, c("GeneName", "start", "width"))
```

```
    select.samples.windowBin <- select.samples[windowTable, allow.cartesian = TRUE]
    select.samples.windowBin[, `:=`(AAproc, winStart/seqlength * 100)]

    setkey(select.samples.windowBin, Group)
    select.samples.windowBin <- select.samples.windowBin[J(names(fill.values))]  #Select the two compared
    setkeyv(select.samples.windowBin, c("Group", "GeneName", "winStart",
        "winEnd"))
    select.samples.windowBin <- select.samples.windowBin[, list(Overlaps = .N,
        seqlength = min(seqlength), AAproc = min(AAproc), BC = paste(t(BC),
            collapse = ","), Animals = paste(t(Animals), collapse = ","),
        LUTnrs = paste(t(LUTnrs), collapse = ","), RNAcount = sum(RNAcount)),
        by = c("Group", "GeneName", "winStart", "winEnd")]

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

} else {
    plot.data.dt <- data.table::copy(select.samples)
}

# ================== Binning of data ==================
FullLength <- 100
position <- seq(0, FullLength, size.bin)
plot.data.dt[, `:=`(bin, findInterval(AAproc, position))]

plot.data.bin <- plot.data.dt[, list(.N, seqlength = min(seqlength), AAproc = position[findInterval(mean(
    position)], BCsum = length(table(strsplit(paste(t(BC), collapse = ","),
    ","))), AnimalCount = length(table(strsplit(paste(t(Animals), collapse = ","),
    ","))), LUTnrs = paste(unique(names(table(strsplit(paste(t(LUTnrs),
    collapse = ","), ","))))), collapse = ","), NormCount = sum(RNAcount)/seqlength *
    FullLength), 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) {
    plot.data.bin[plot.data.bin$Group == names(fill.values)[1]]$NormCount <- plot.data.bin[plot.data.bin$
        names(fill.values)[1]]$NormCount/max(plot.data.bin[plot.data.bin$Group ==
        names(fill.values)[1]]$NormCount)
    plot.data.bin[plot.data.bin$Group == names(fill.values)[2]]$NormCount <- plot.data.bin[plot.data.bin$
        names(fill.values)[2]]$NormCount/max(plot.data.bin[plot.data.bin$Group ==
        names(fill.values)[2]]$NormCount)
}


if (PlotBC && NormalizePlot) {
    plot.data.bin[plot.data.bin$Group == names(fill.values)[1]]$BCanim <- plot.data.bin[plot.data.bin$Gro
        names(fill.values)[1]]$BCanim/max(plot.data.bin[plot.data.bin$Group ==
        names(fill.values)[1]]$BCanim)
    plot.data.bin[plot.data.bin$Group == names(fill.values)[2]]$BCanim <- plot.data.bin[plot.data.bin$Gro
        names(fill.values)[2]]$BCanim/max(plot.data.bin[plot.data.bin$Group ==
        names(fill.values)[2]]$BCanim)
}
```

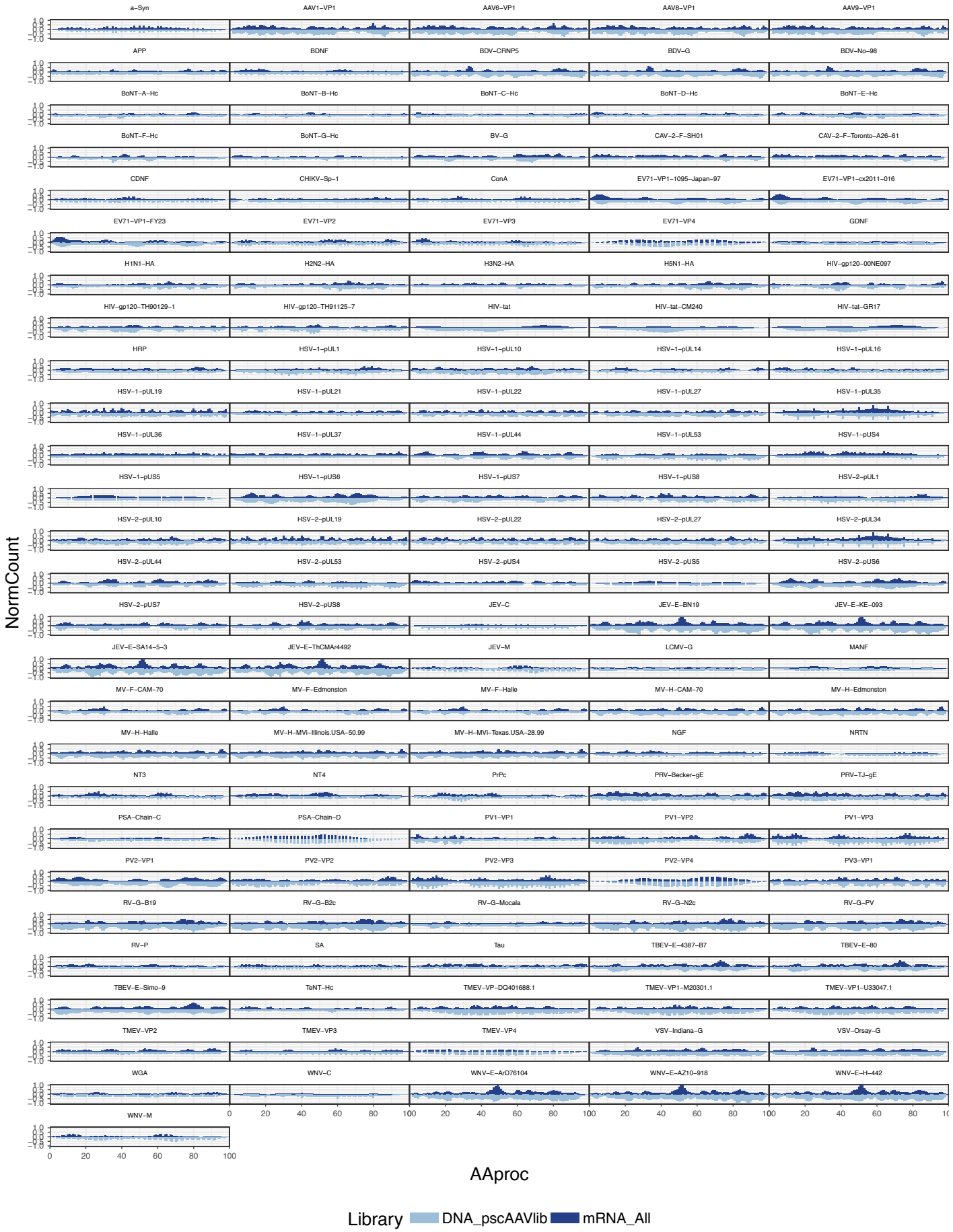
```
plot.data.bin[plot.data.bin$Group == names(fill.values)[2]]$NormCount <- plot.data.bin[plot.data.bin$Grou
    names(fill.values)[2]]$NormCount * -1  #This line flips the values for the second group

plot.data.bin[plot.data.bin$Group == names(fill.values)[2]]$BCanim <- plot.data.bin[plot.data.bin$Group =
    names(fill.values)[2]]$BCanim * -1  #This line flips the values for the second group

# ================== Output plot ==================

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

plot.out <- eval(parse(text = paste("ggplot(plot.data.bin,aes(x=AAproc,y=",
    outVar, ", fill = Group, vjust=-20))", 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(limit = c(0,
    100), breaks = c(seq(0, 100, 20)), expand = c(0, 0)) + facet_wrap(~GeneName,
    ncol = 5) + 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 = 3, vjust = -20), strip.background = element_blank(),
    panel.spacing.y = unit(0, "cm"), panel.spacing.x = unit(0, "cm"))


# ================== Sort and select top samples ==================

select.samples.binPos <<- select.samples
setkeyv(select.samples.binPos, c("Group", "structure", "Sequence"))
setorder(select.samples.binPos, Group, structure, Sequence, GeneName)
select.samples.binPos <- unique(select.samples.binPos, by = c("Group", "structure",
    "Sequence"))
# Due to key, this removes replicates if identical sequence mapped to
# multiple genes

setkeyv(select.samples.binPos, c("Group", "Category", "GeneName", "AA"))
select.samples.binPos[, `:=`(c("BCcount", "NormCount", "AnimalCount", "LUTnrs",
    "mainStruct", "mismatches"), list(length(table(strsplit(paste(t(BC),
    collapse = ","), ","))), sum(NormCount), length(table(strsplit(paste(t(Animals),
    collapse = ","), ","))), paste(unique(names(table(strsplit(paste(t(LUTnrs),
    collapse = ","), ",")))), collapse = ","), paste(unique(structure),
    collapse = ","), median(mismatches))), by = key(select.samples.binPos)]

select.samples.binPos <- unique(select.samples.binPos, by = c("Group", "NormCount",
    "LUTnrs"))
select.samples.binPos <- select.samples.binPos[, c("Group", "GeneName",
    "AA", "NormCount", "BCcount", "AnimalCount", "LUTnrs", "mainStruct",
    "mismatches"), with = FALSE]

if (PlotBC) {
    select.samples.binPos[, `:=`(BCanim, BCcount * AnimalCount)]
    setorder(select.samples.binPos, Group, -BCanim, -BCcount, -AnimalCount,
        -NormCount)
} else {
```

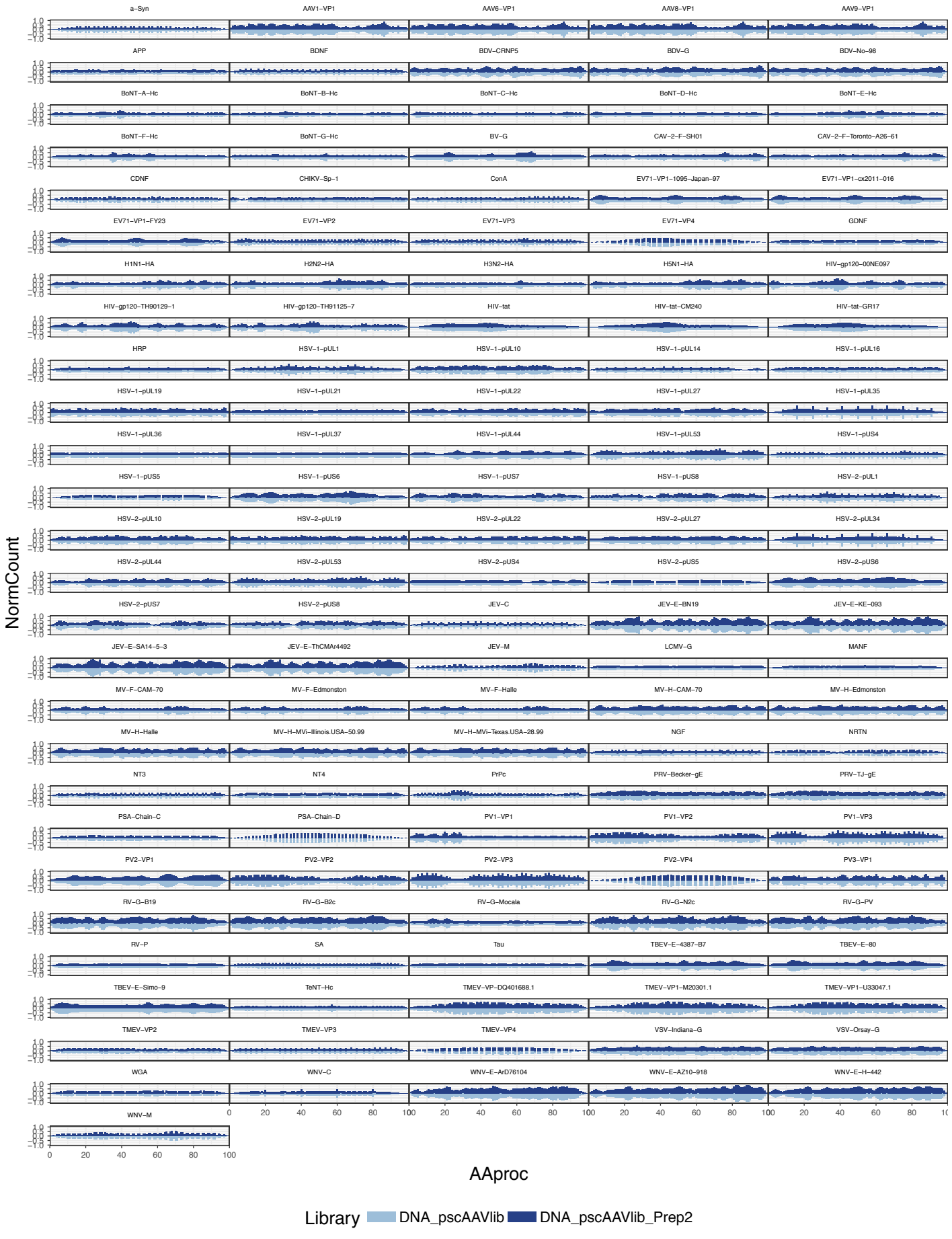
```
        setorder(select.samples.binPos, Group, -NormCount, -BCcount, -AnimalCount)
    }
    setkey(select.samples.binPos, Group)
    select.samples.top <- select.samples.binPos[, head(.SD, 25), by = Group]
    top.sample <- select.samples.top[J(names(fill.values)[1])]
    bottom.sample <- select.samples.top[J(names(fill.values)[2])]
    top.sample[, `:=`(c("Group"), NULL)]
    setnames(top.sample, "GeneName", names(fill.values)[1])
    bottom.sample[, `:=`(c("Group"), NULL)]
    setnames(bottom.sample, "GeneName", names(fill.values)[2])

    out.list <- list(plot = plot.out, plotBin = plot.data.bin, top = top.sample,
        bottom = bottom.sample)

    return(out.list)
}
```

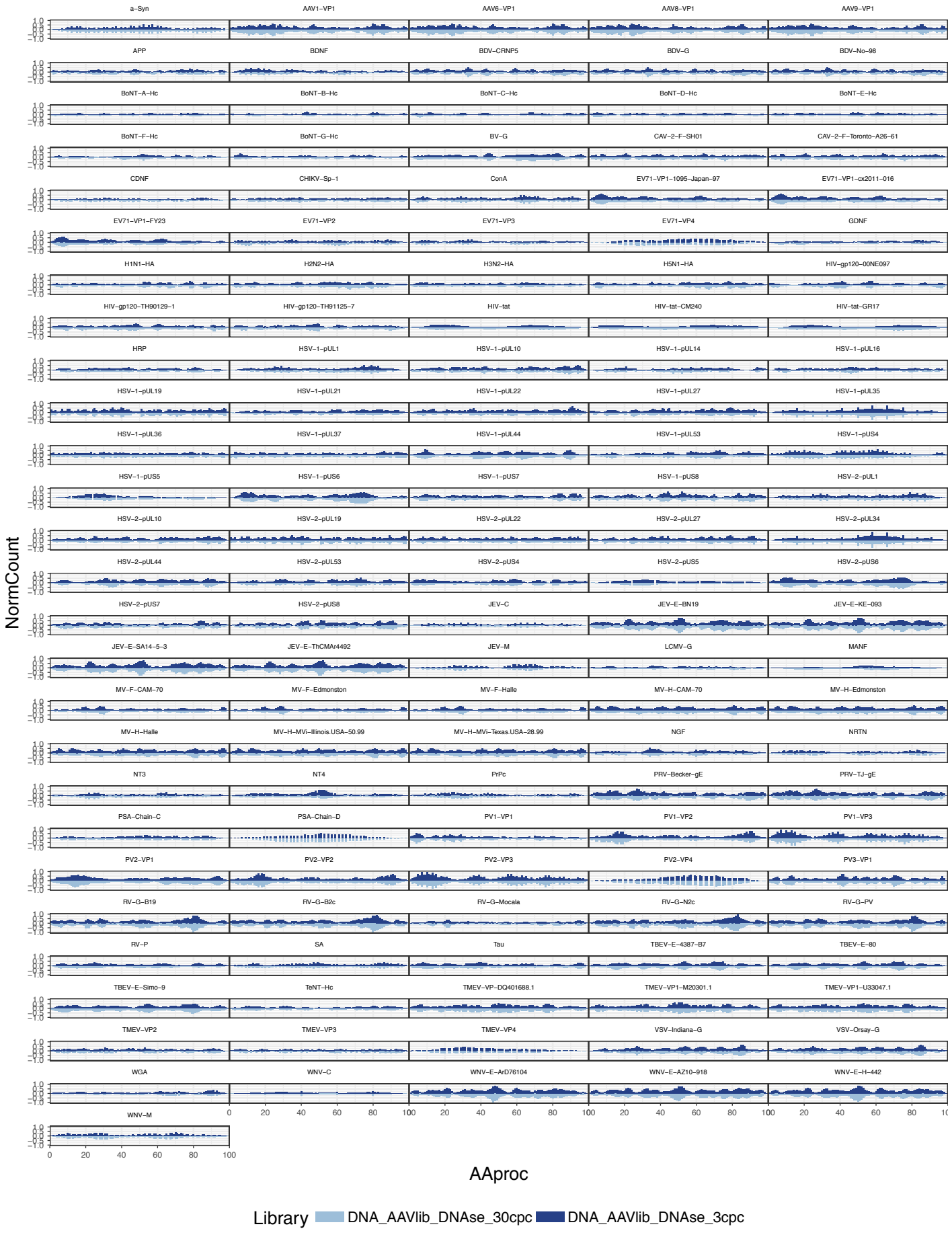
## Analyze samples

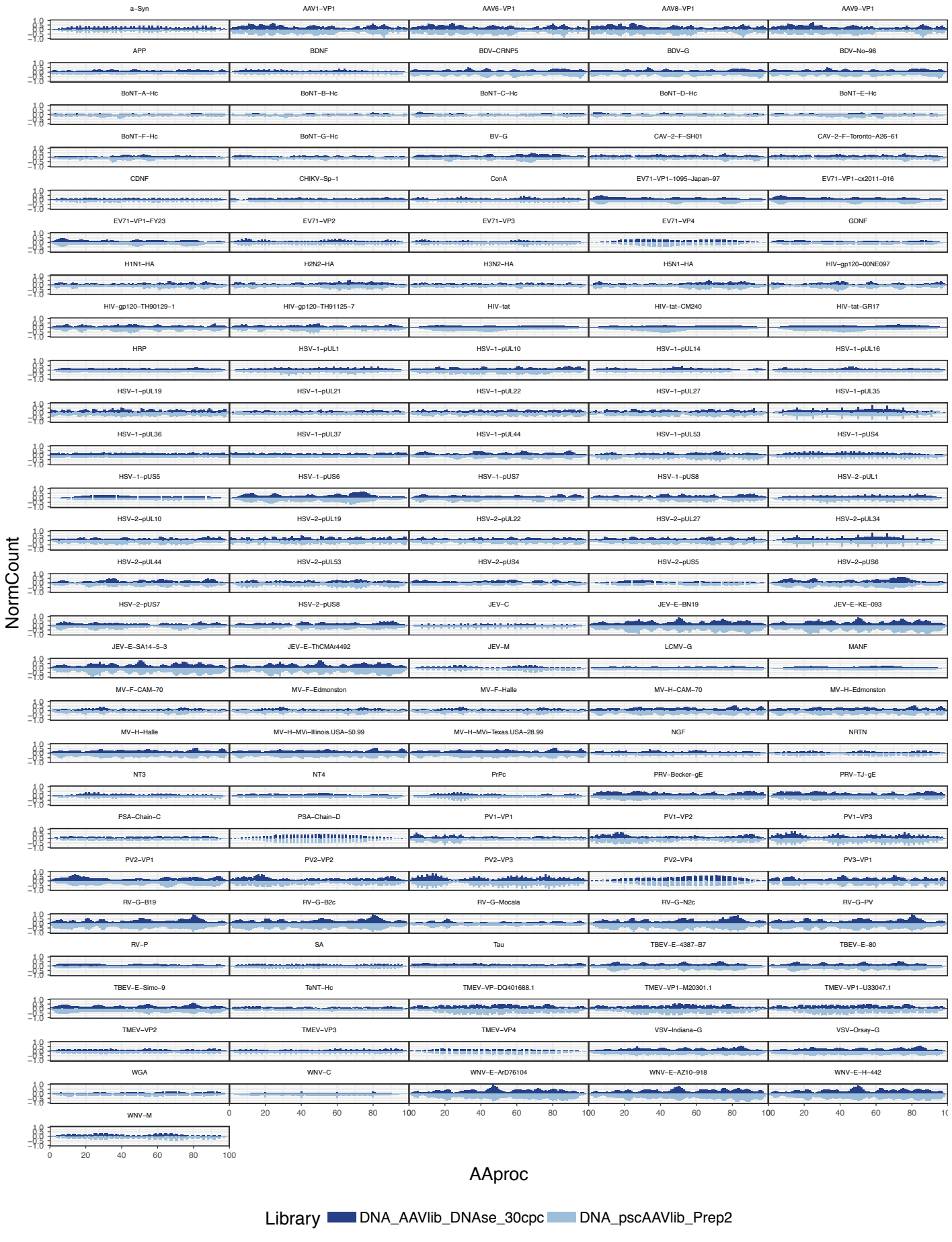
```
# ================== Sample plotting ==================
```

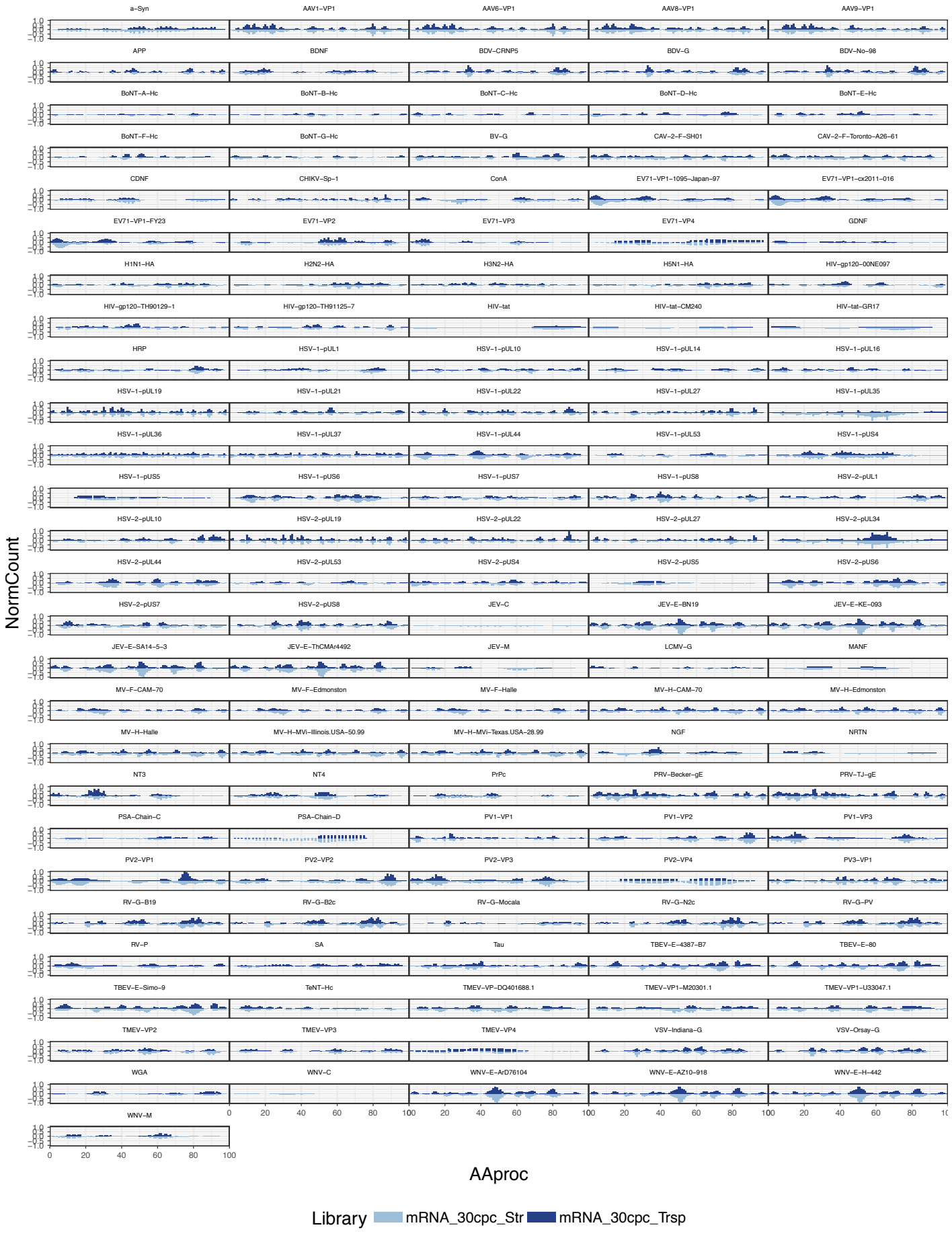
Binning analysis version 1

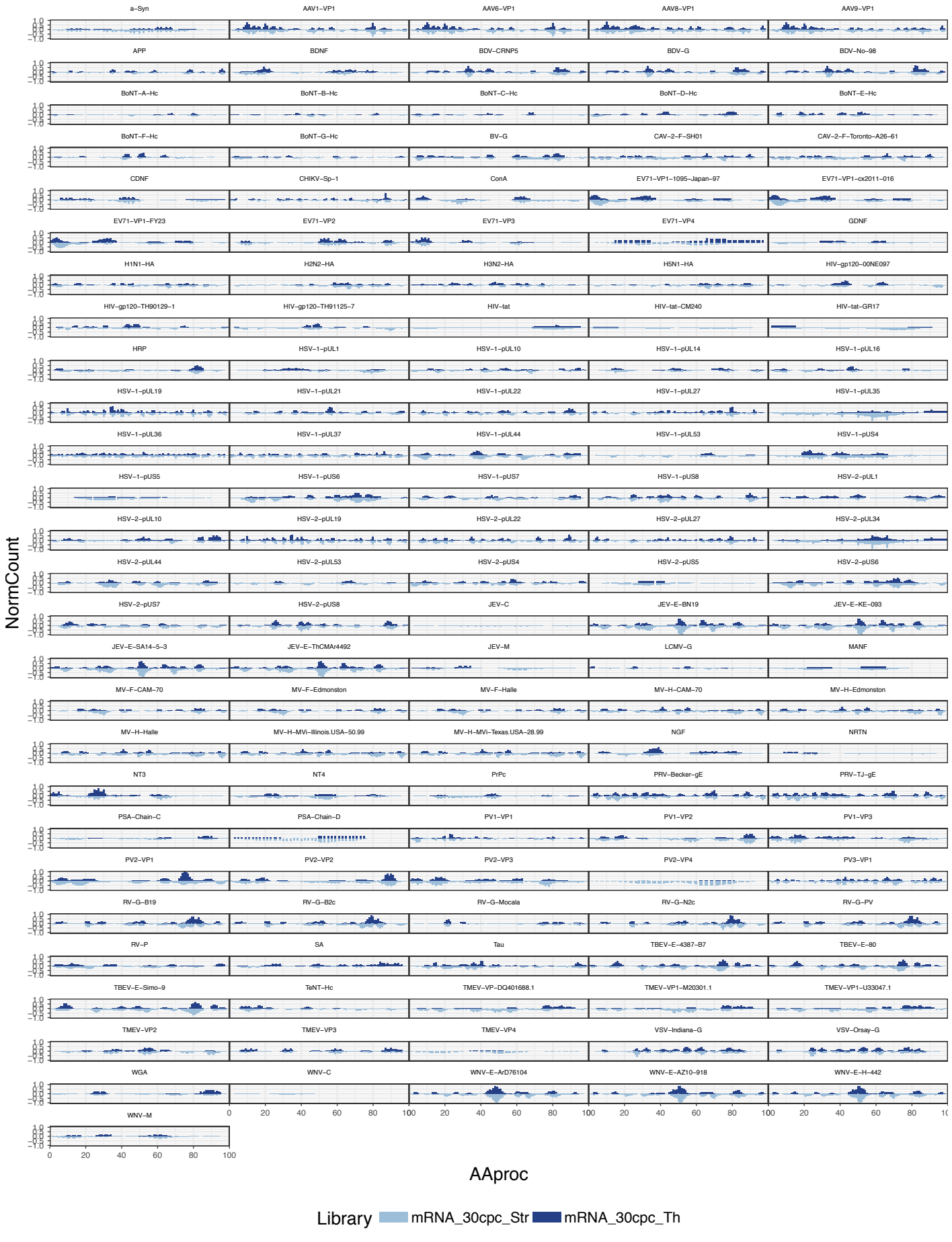

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
DNA_pscAAVlib
mRNA_All

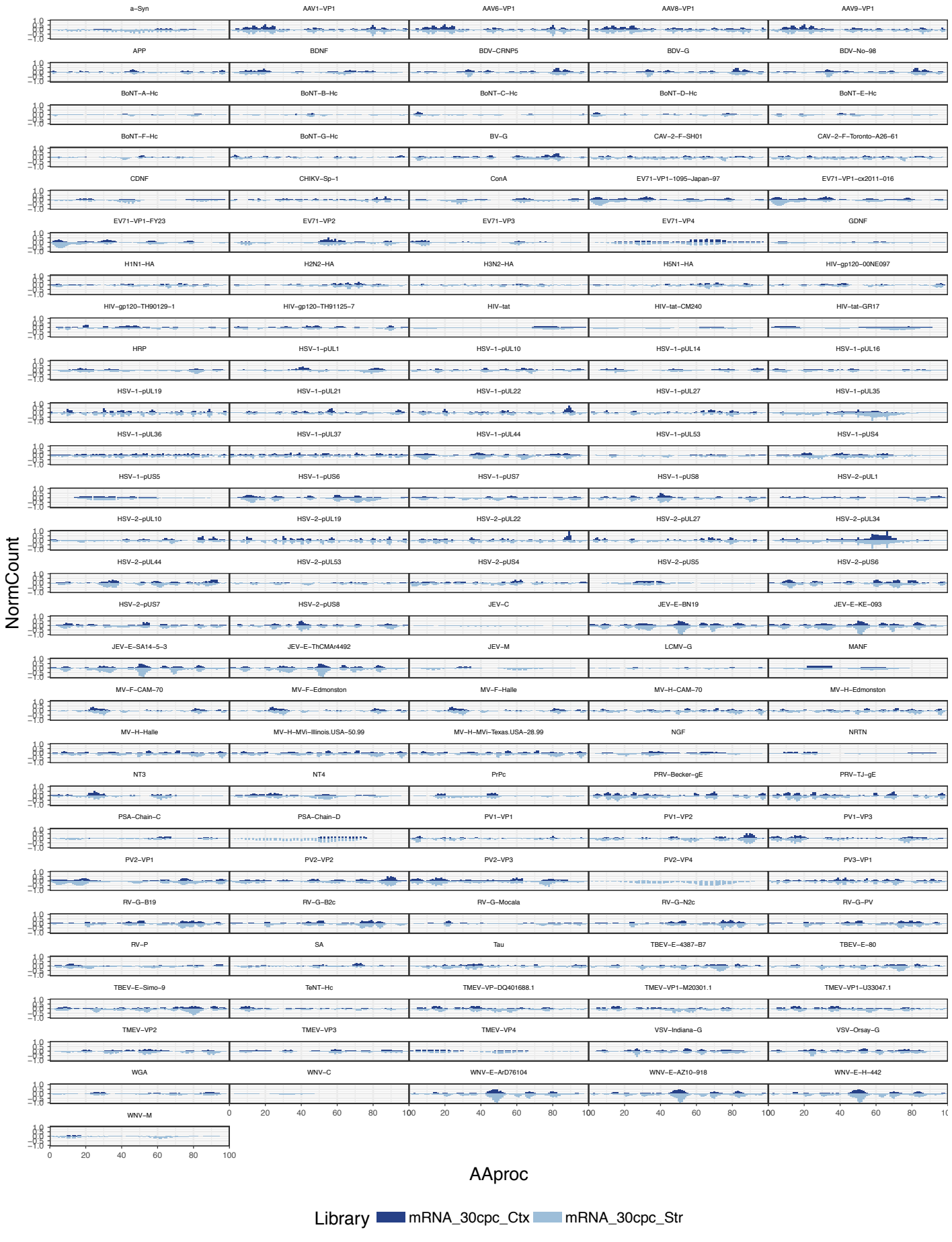

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
DNA_pscAAVlib
DNA_pscAAVlib_Prep2

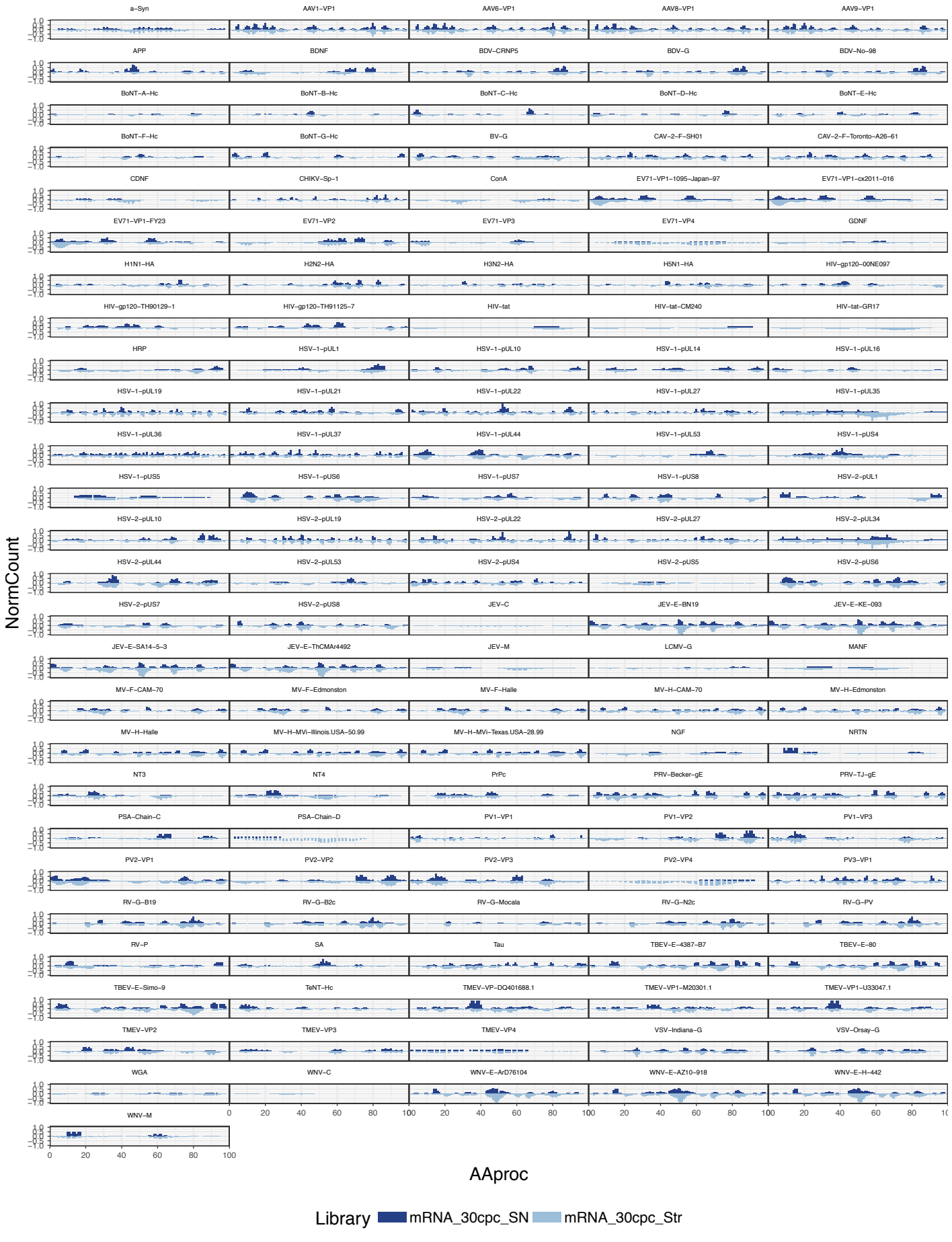

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
DNA_AAVlib_DNAse_30cpc
DNA_AAVlib_DNAse_3cpc

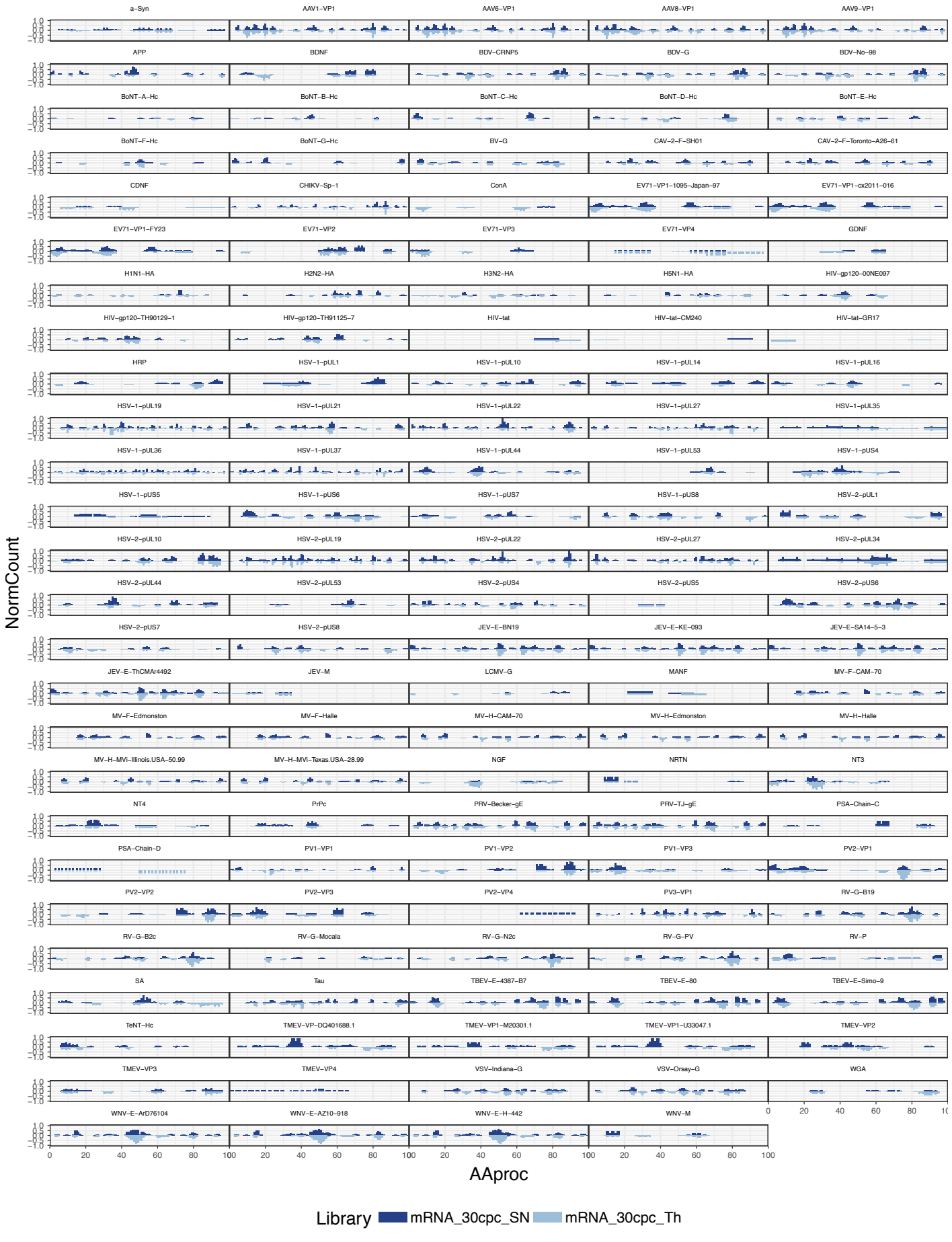

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
DNA_AAVlib_DNAse_30cpc
DNA_pscAAVlib_Prep2

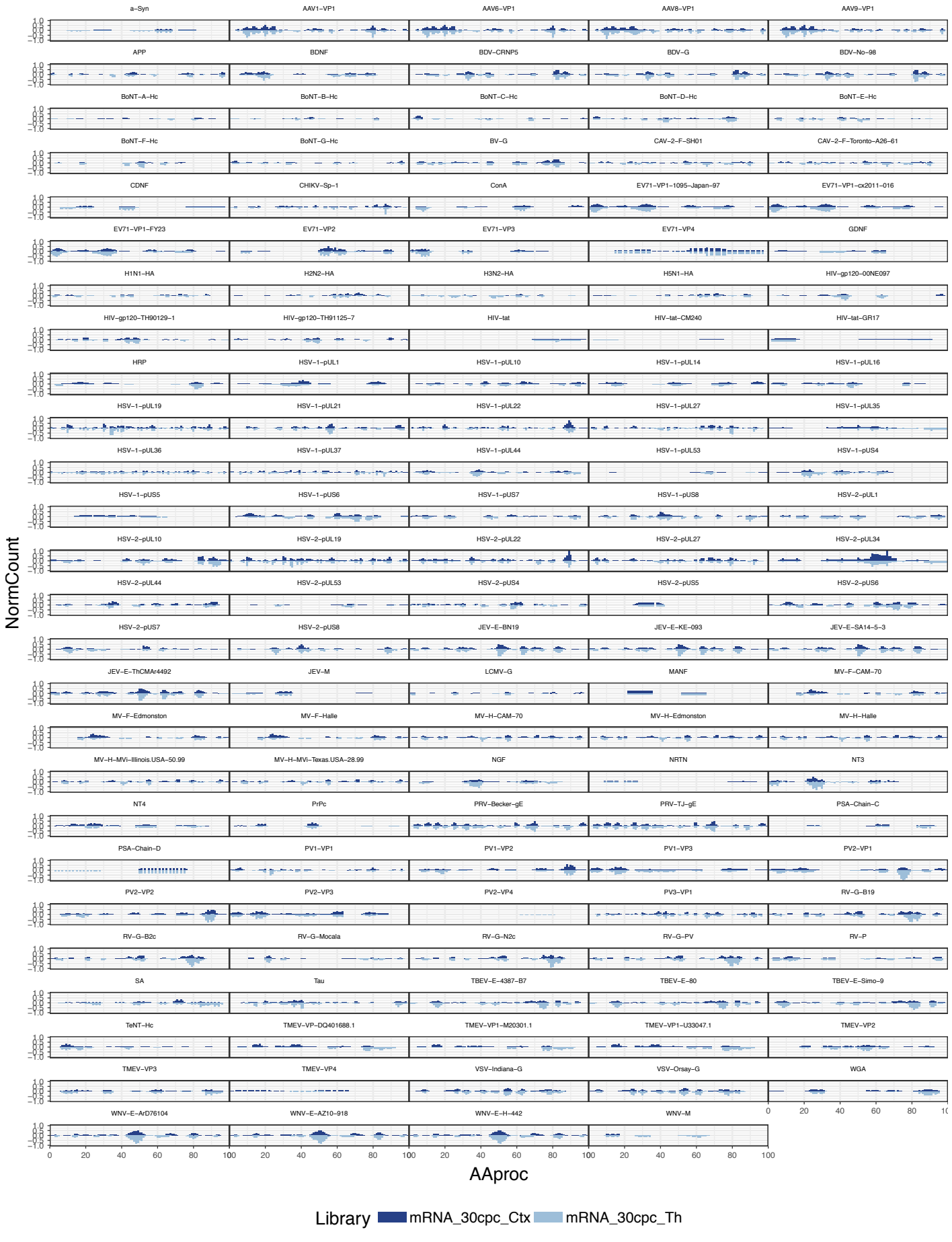

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Trsp

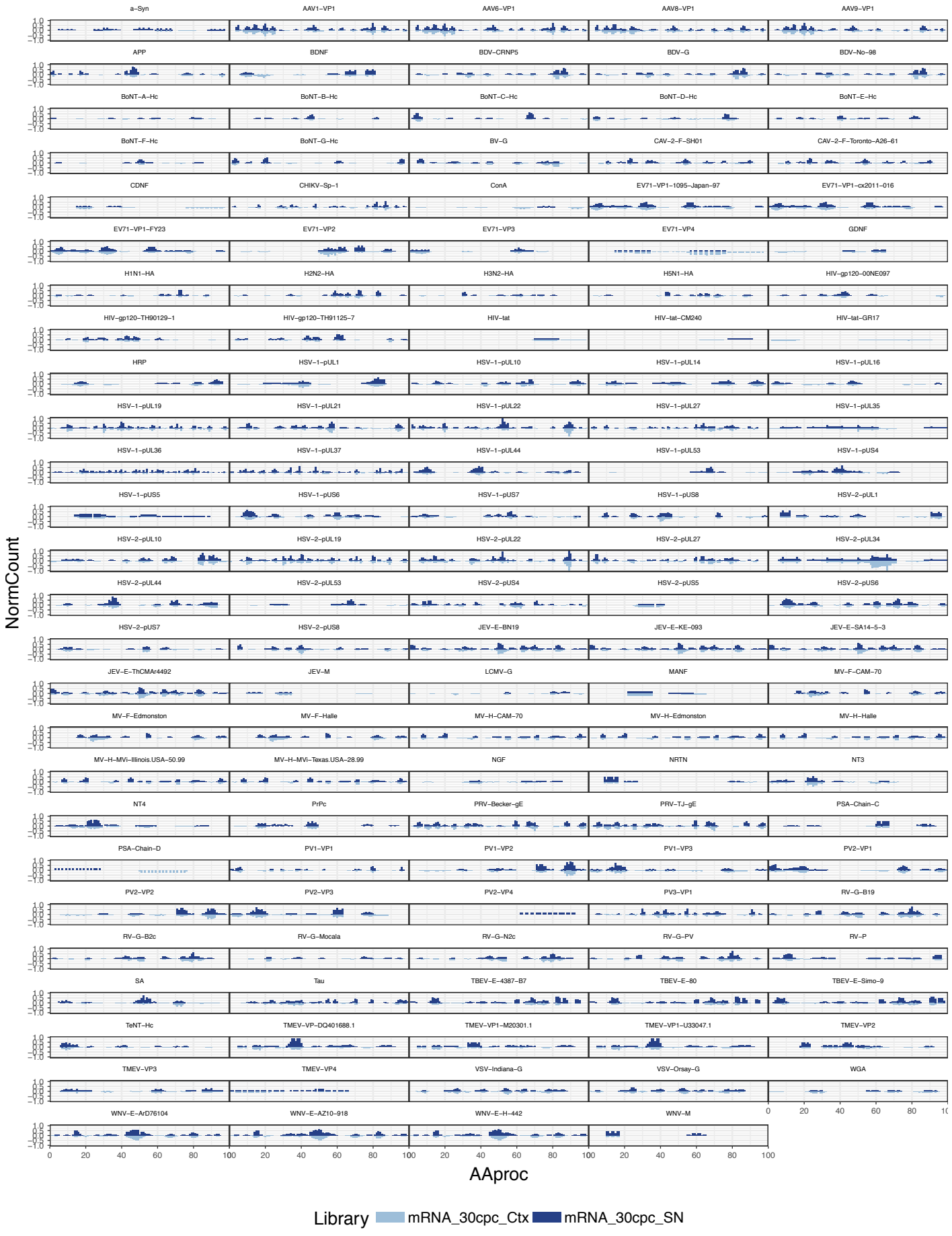

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Th

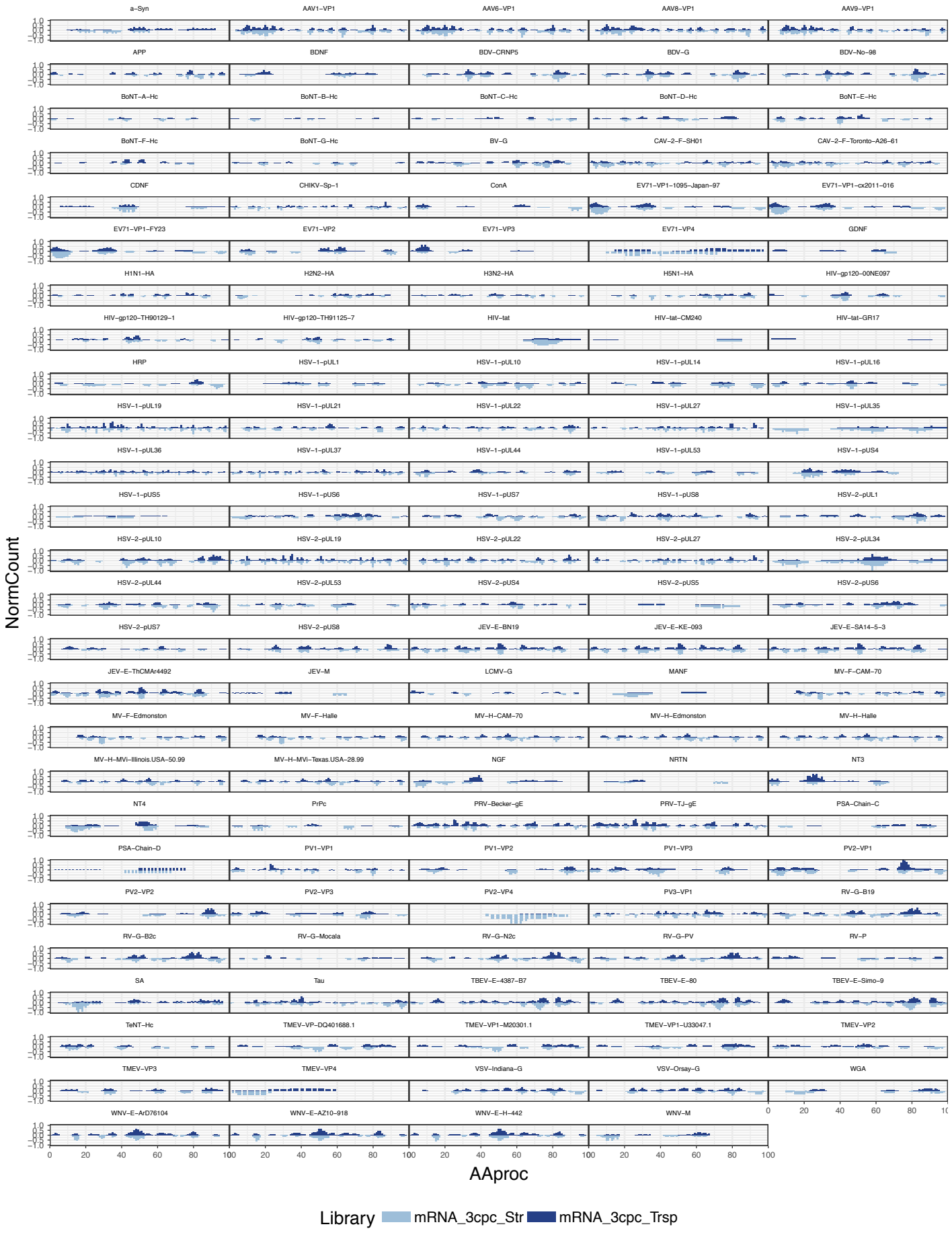

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Str

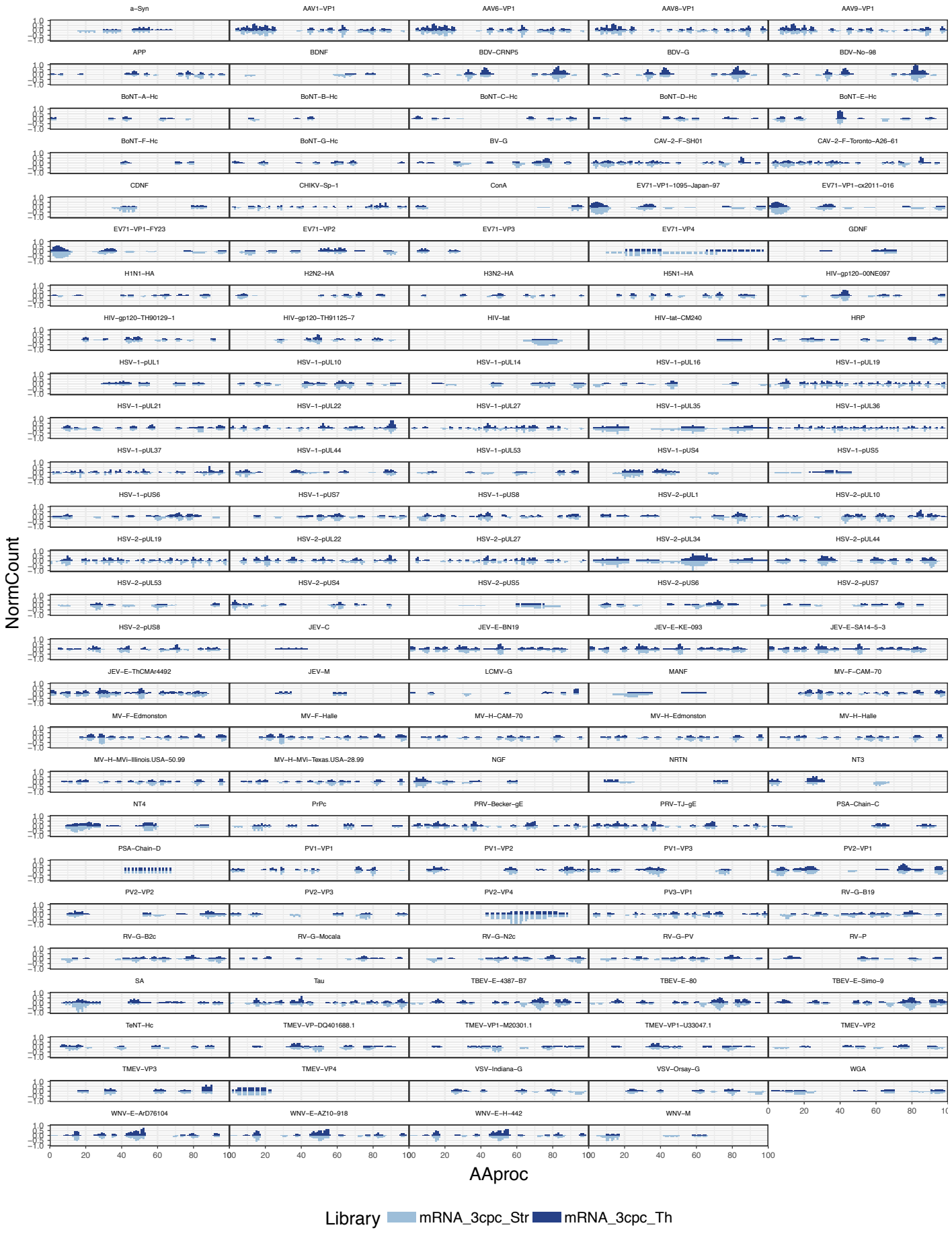

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_SN
mRNA_30cpc_Str

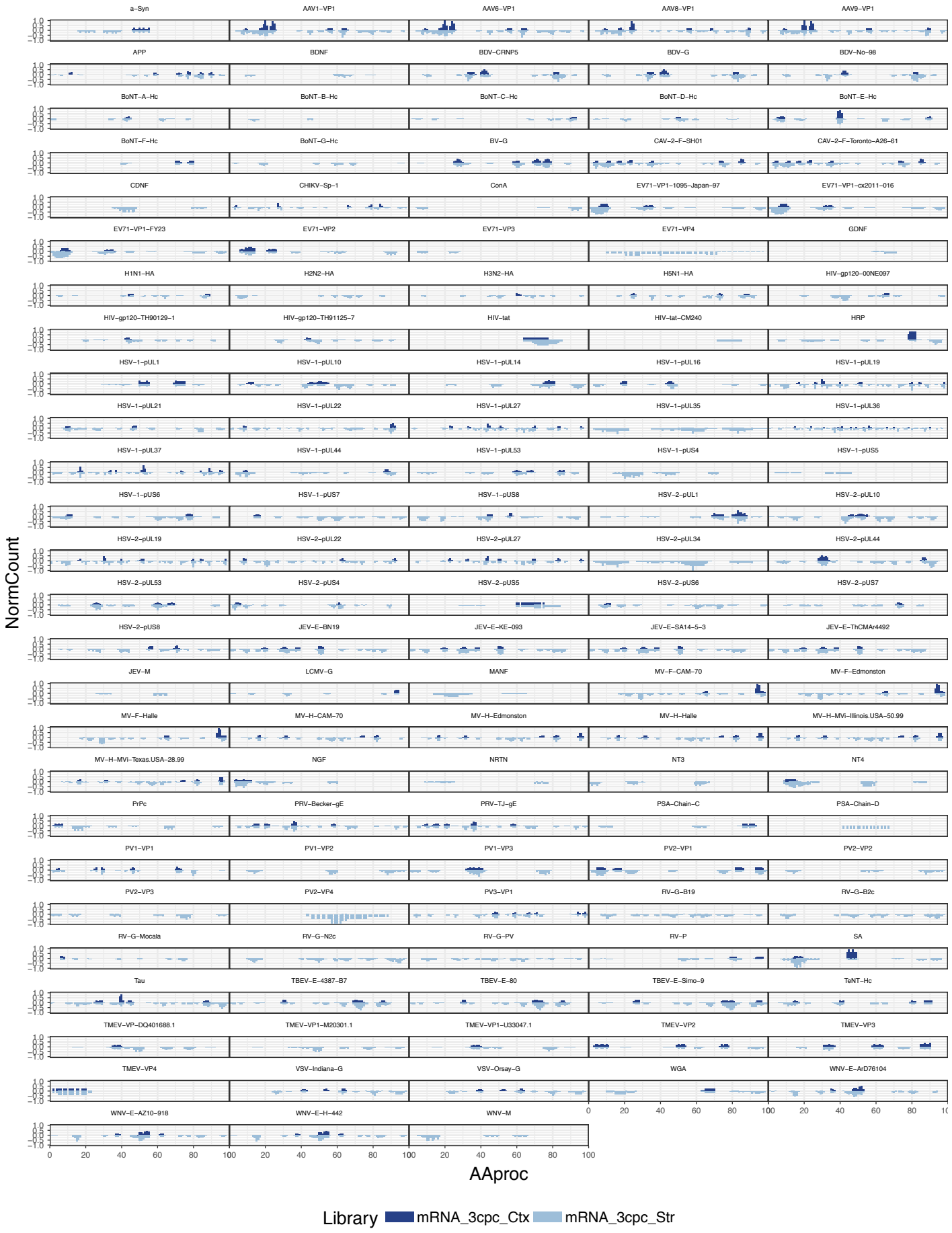

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_SN
mRNA_30cpc_Th

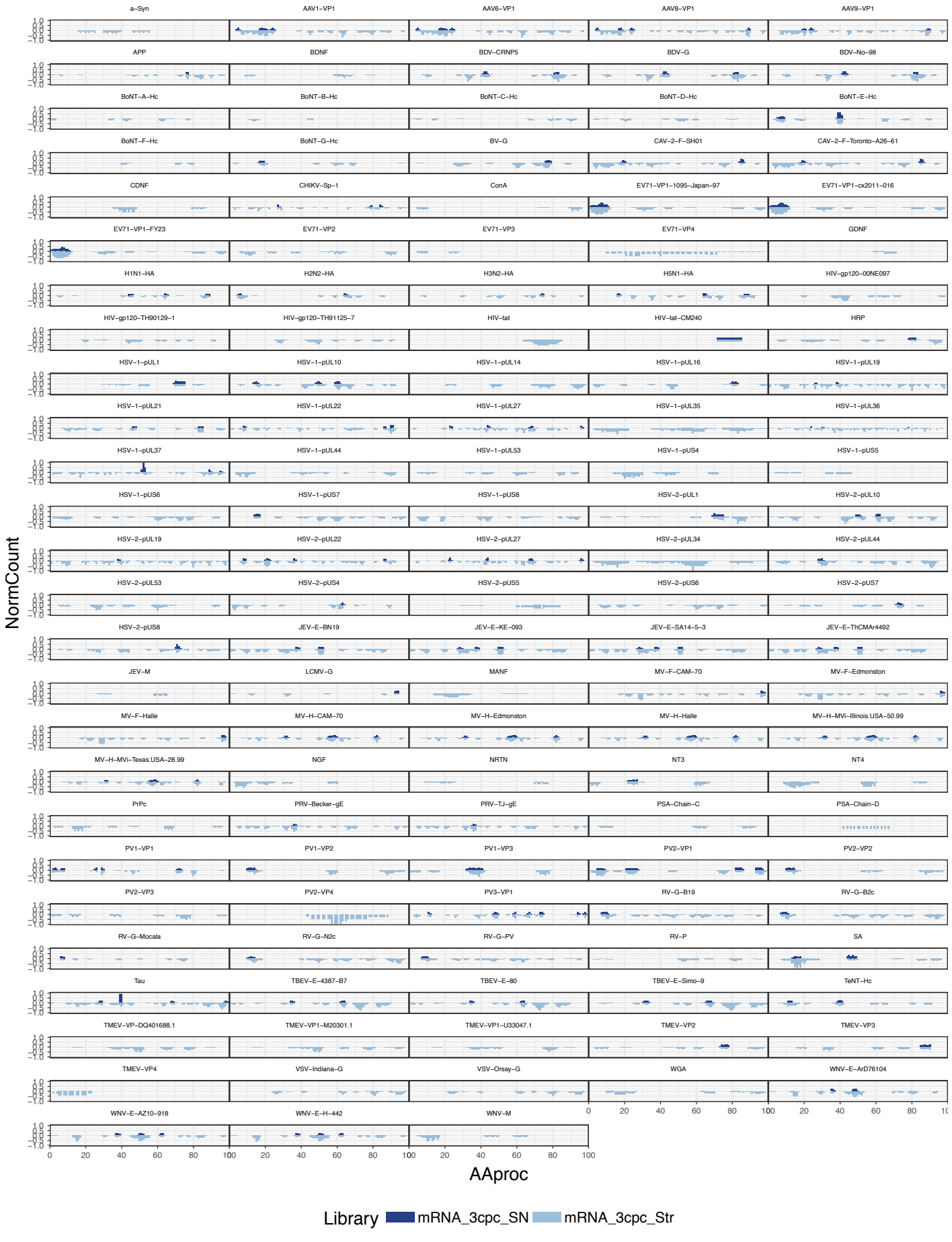

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Th

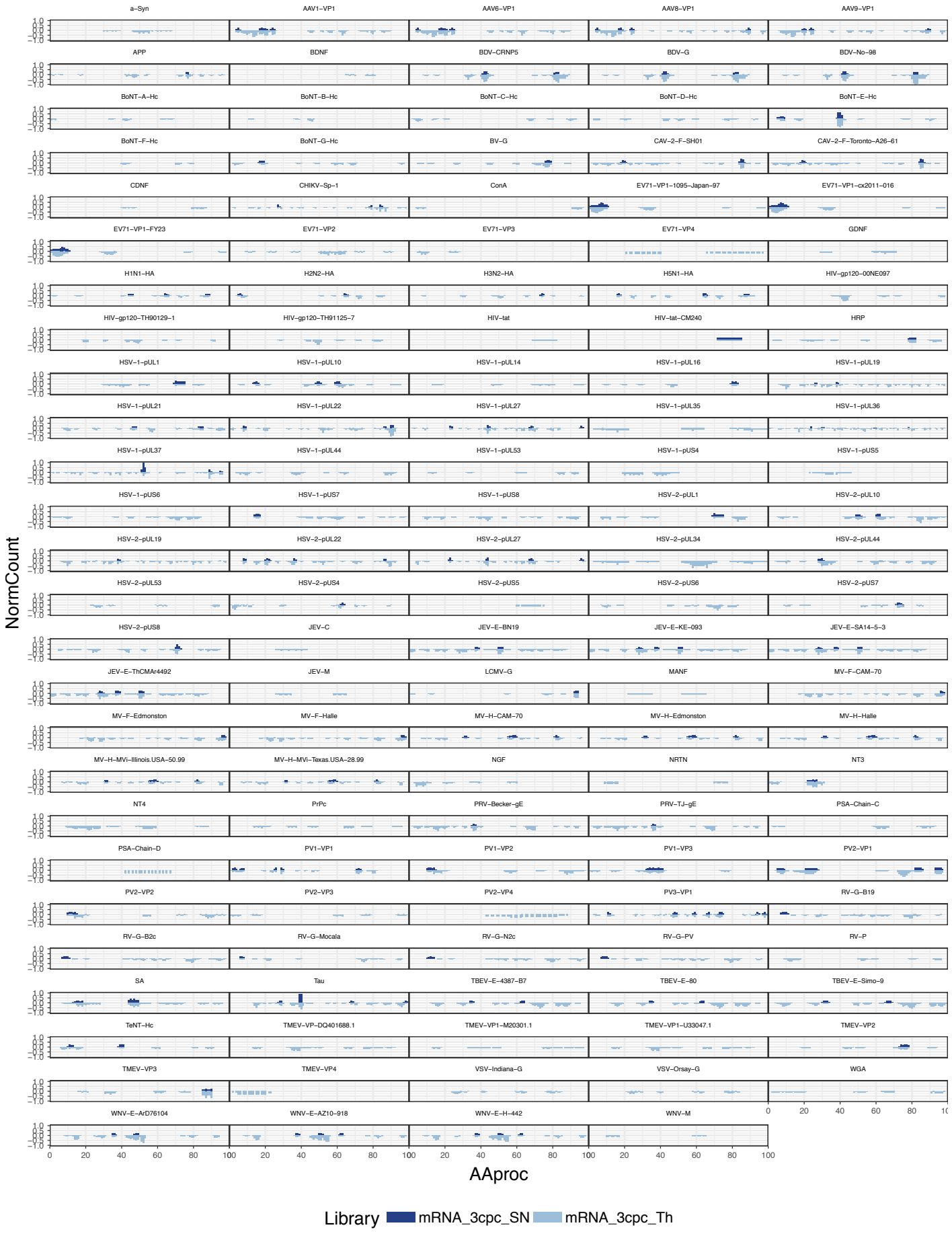

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN

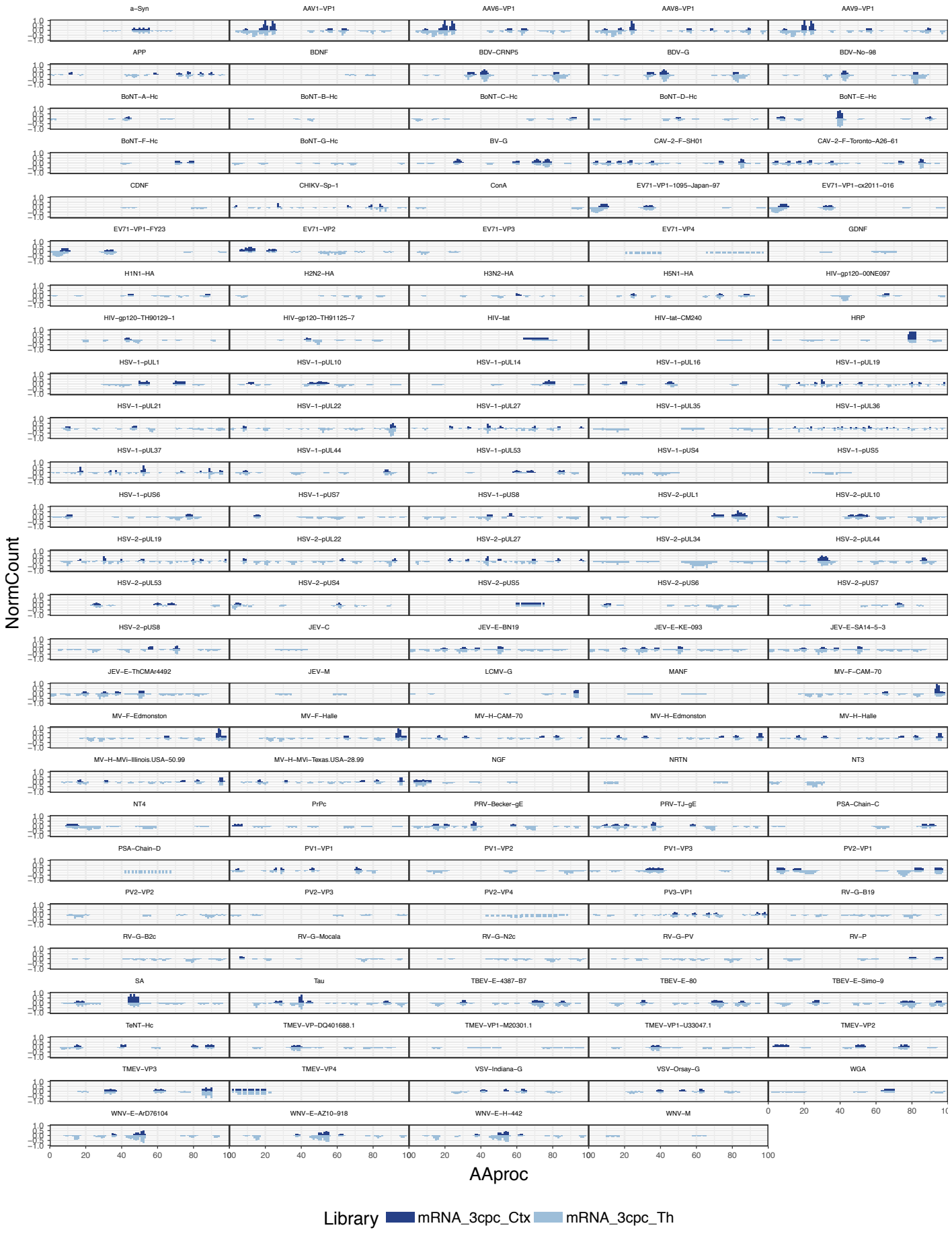

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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–E–BN19
JEV–E–KE–093
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
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
WGA
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Trsp

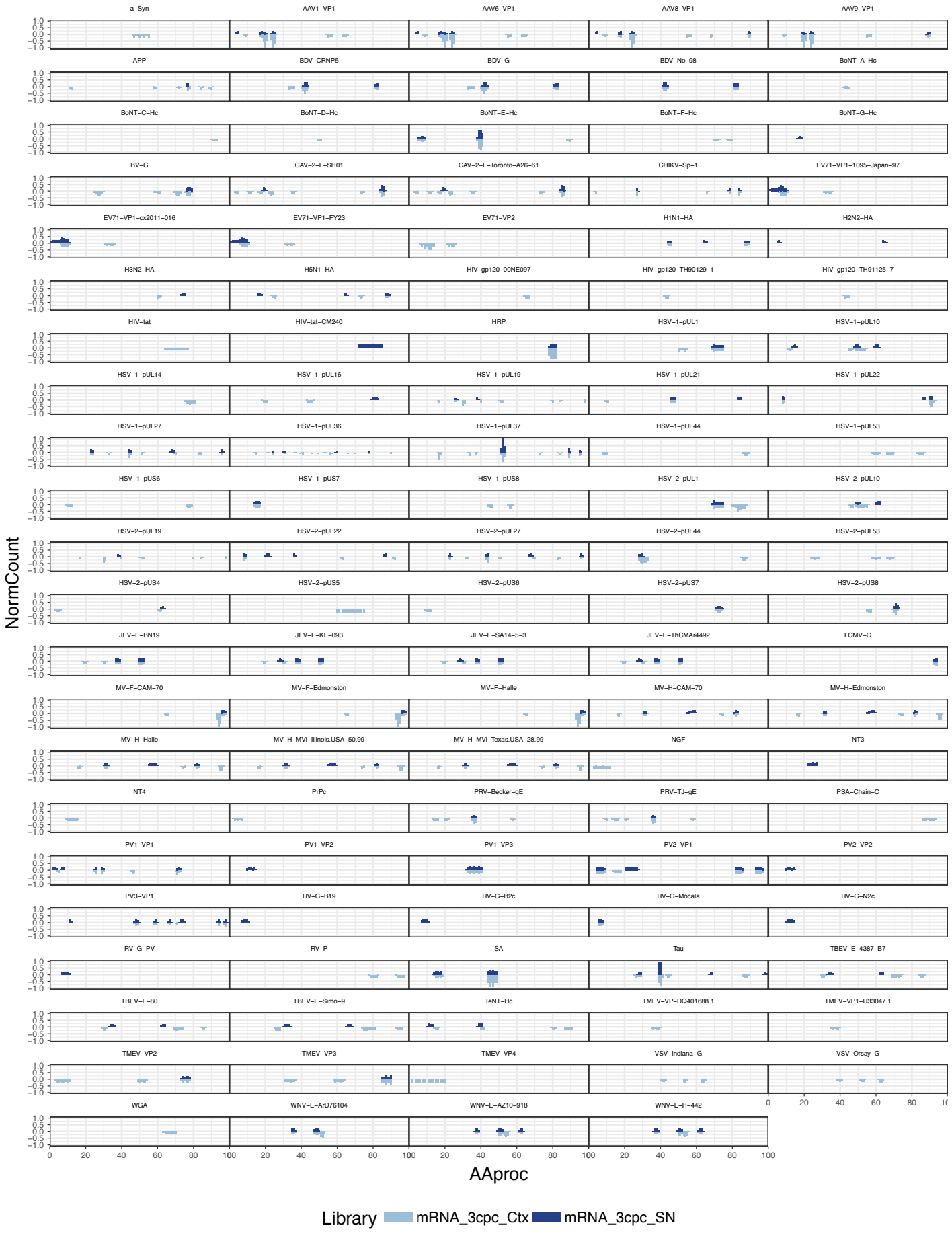

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Th

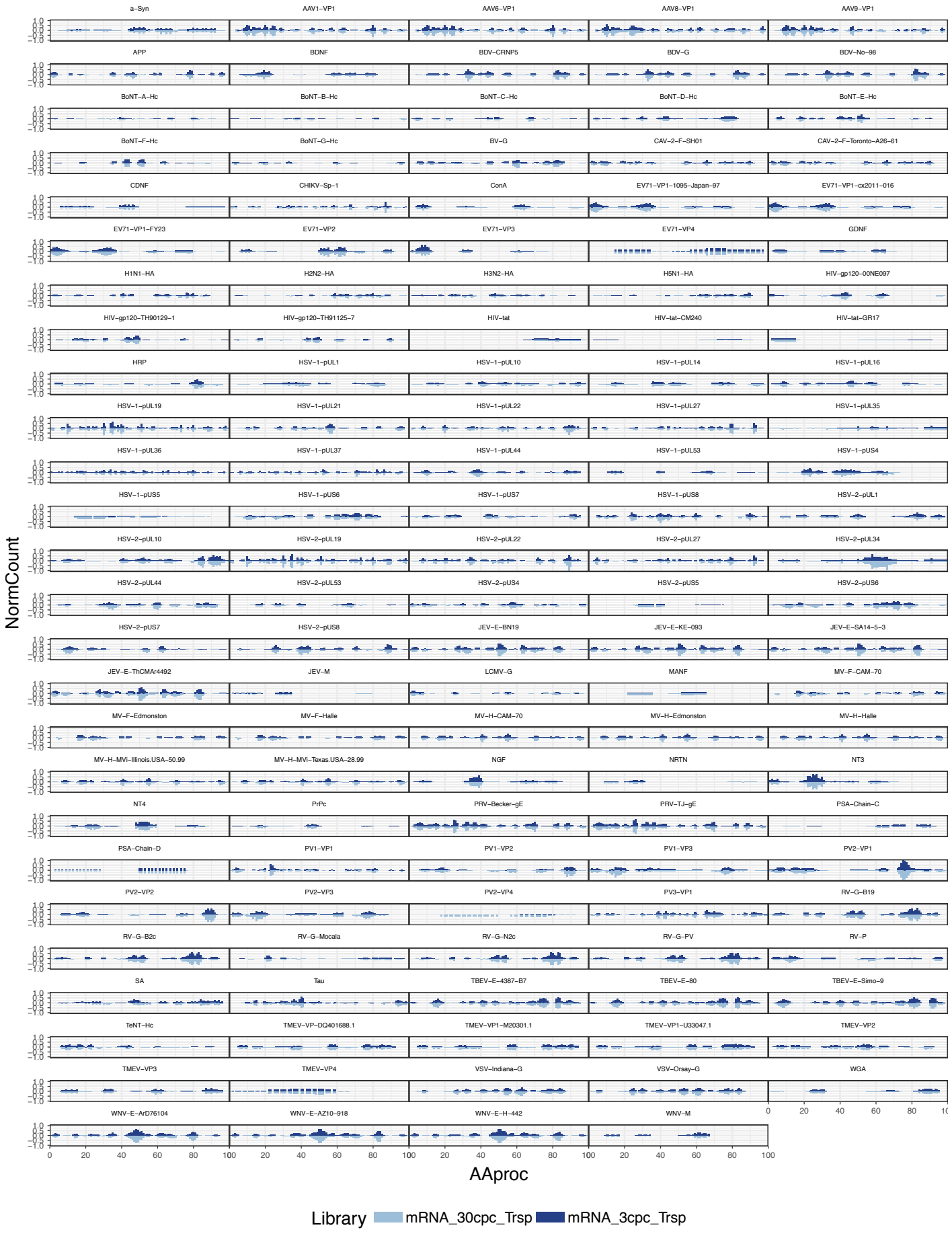

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Str

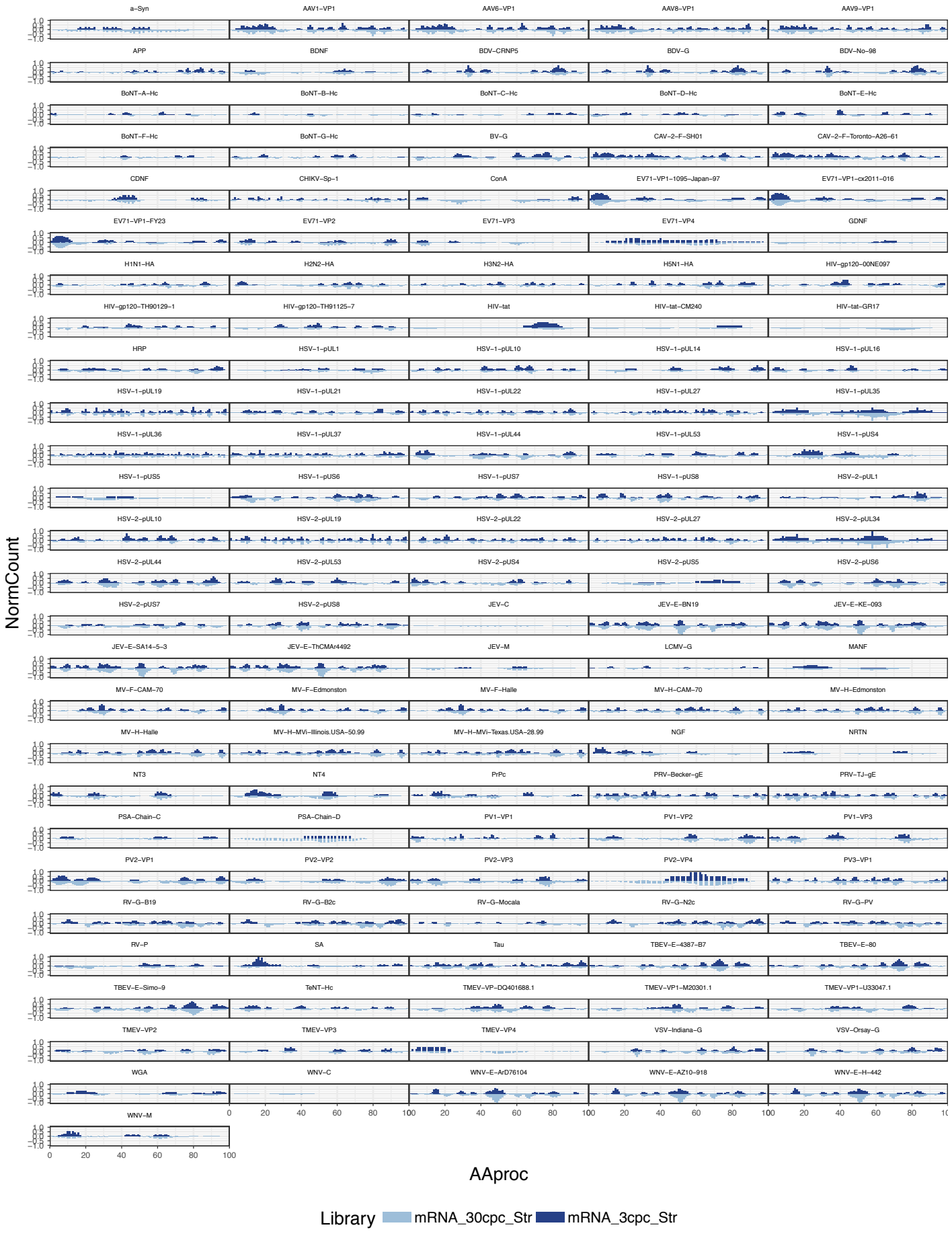

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_SN
mRNA_3cpc_Str

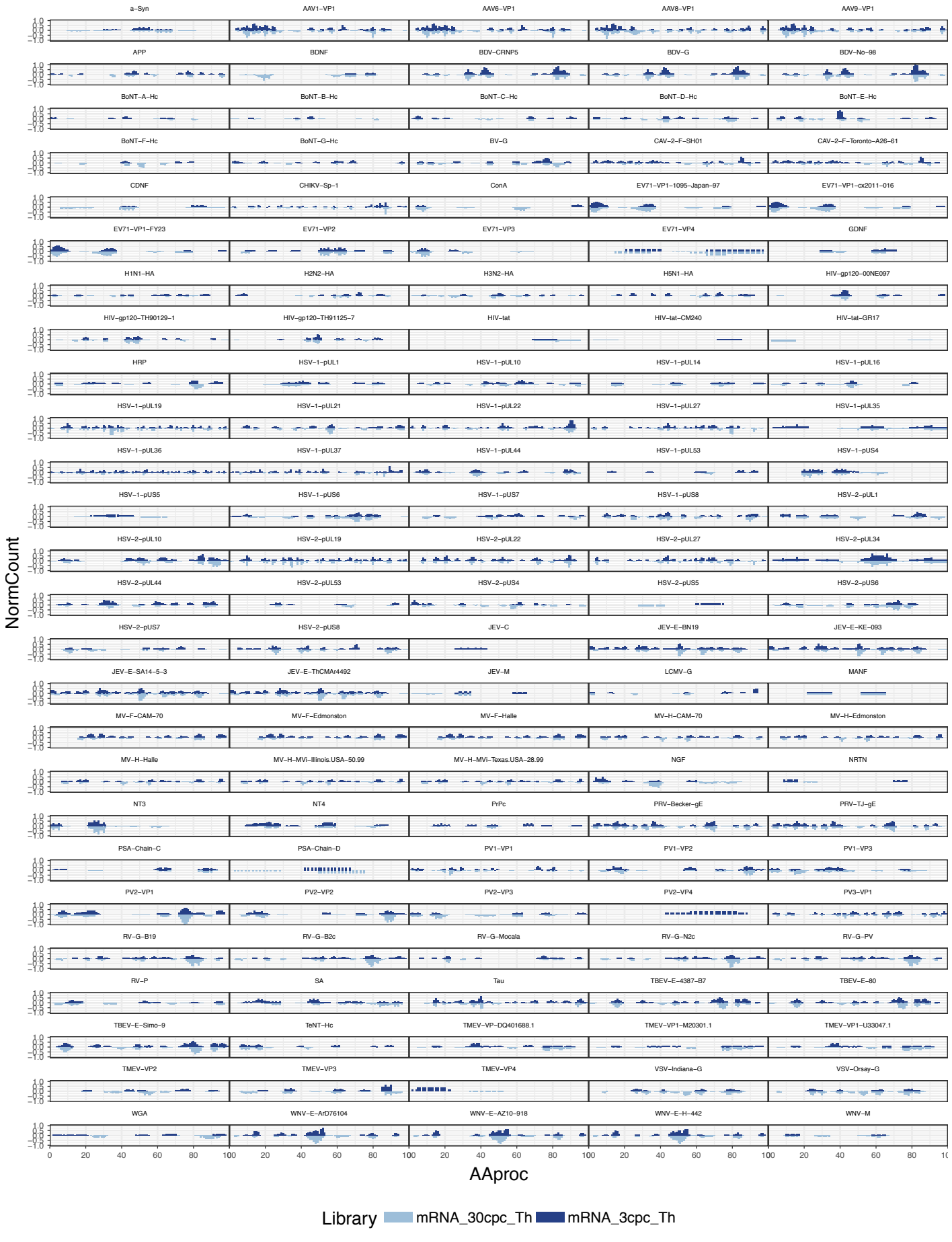

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_SN
mRNA_3cpc_Th

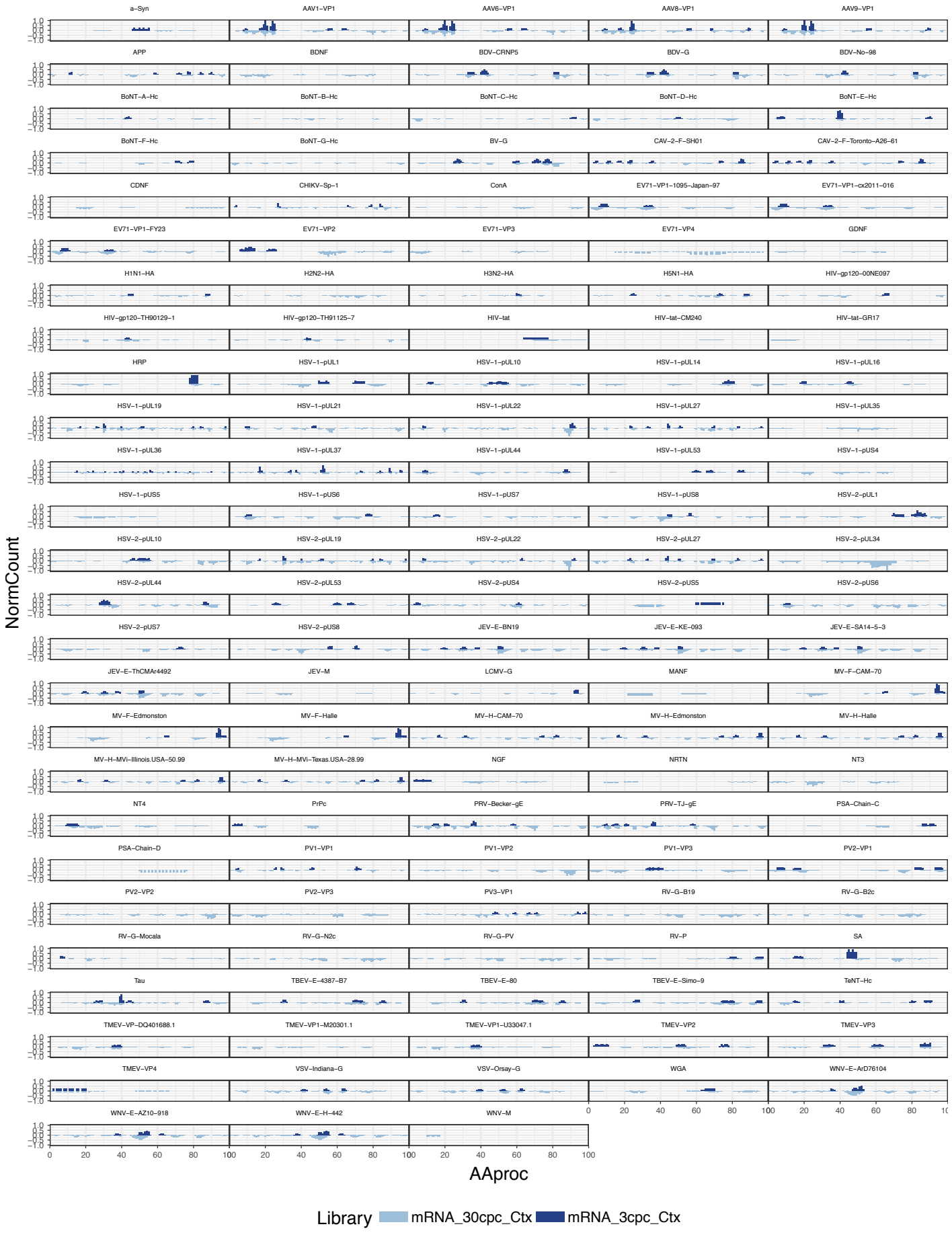

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Th

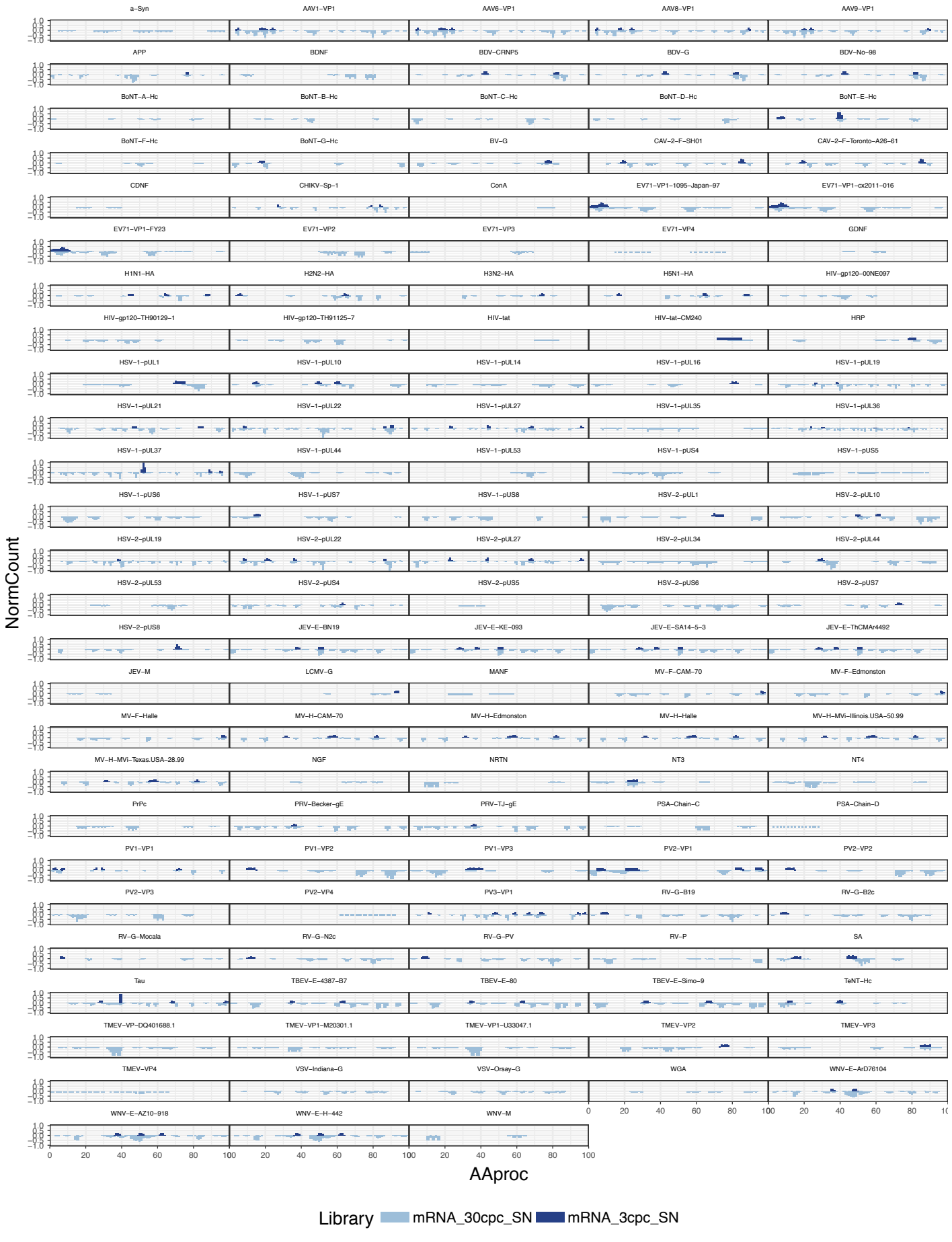

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
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−pUL44
HSV−2−pUL53
HSV−2−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
LCMV−G
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
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PSA−Chain−C
PV1−VP1
PV1−VP2
PV1−VP3
PV2−VP1
PV2−VP2
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−U33047.1
TMEV−VP2
TMEV−VP3
TMEV−VP4
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_SN

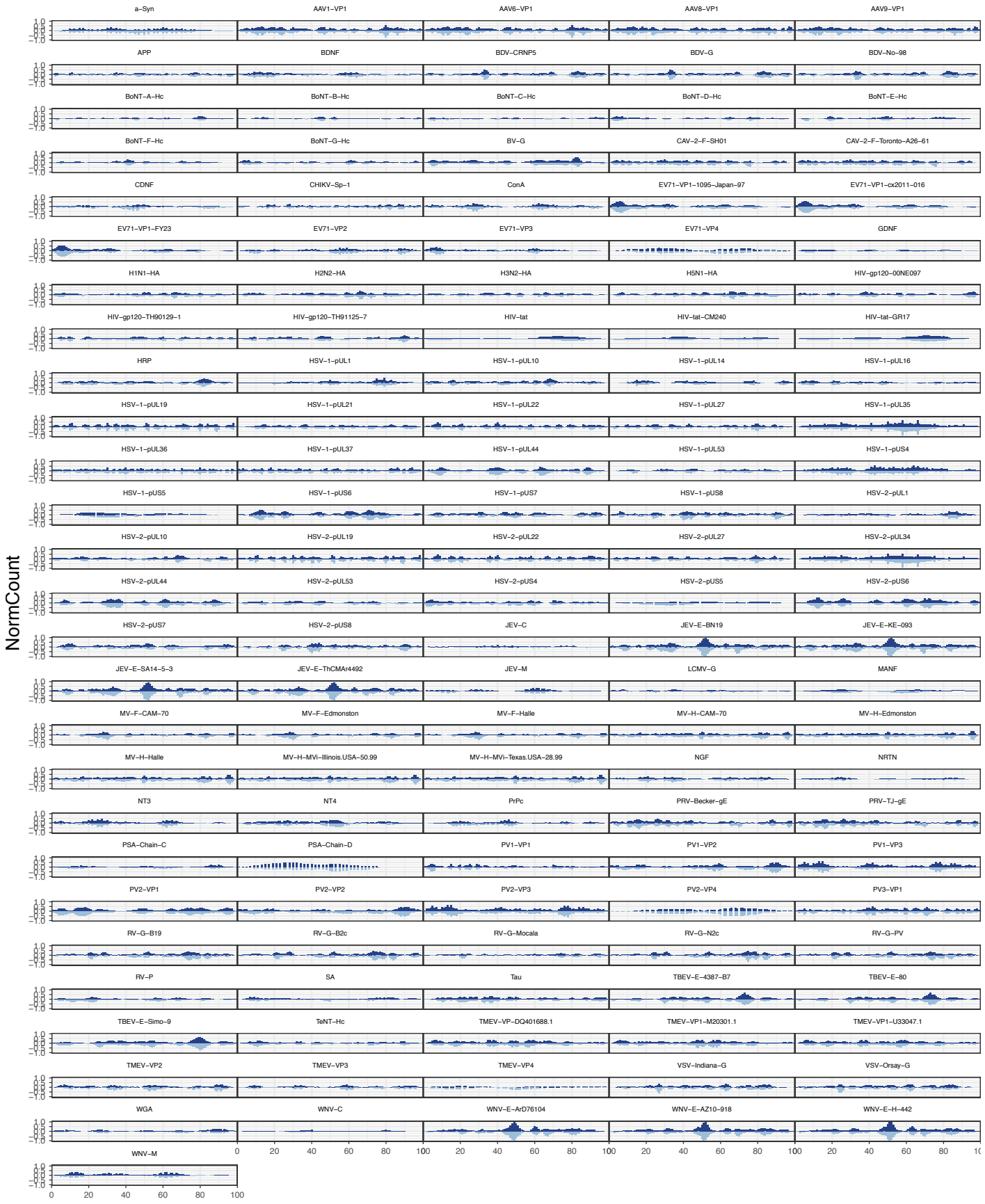

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Trsp
mRNA_3cpc_Trsp

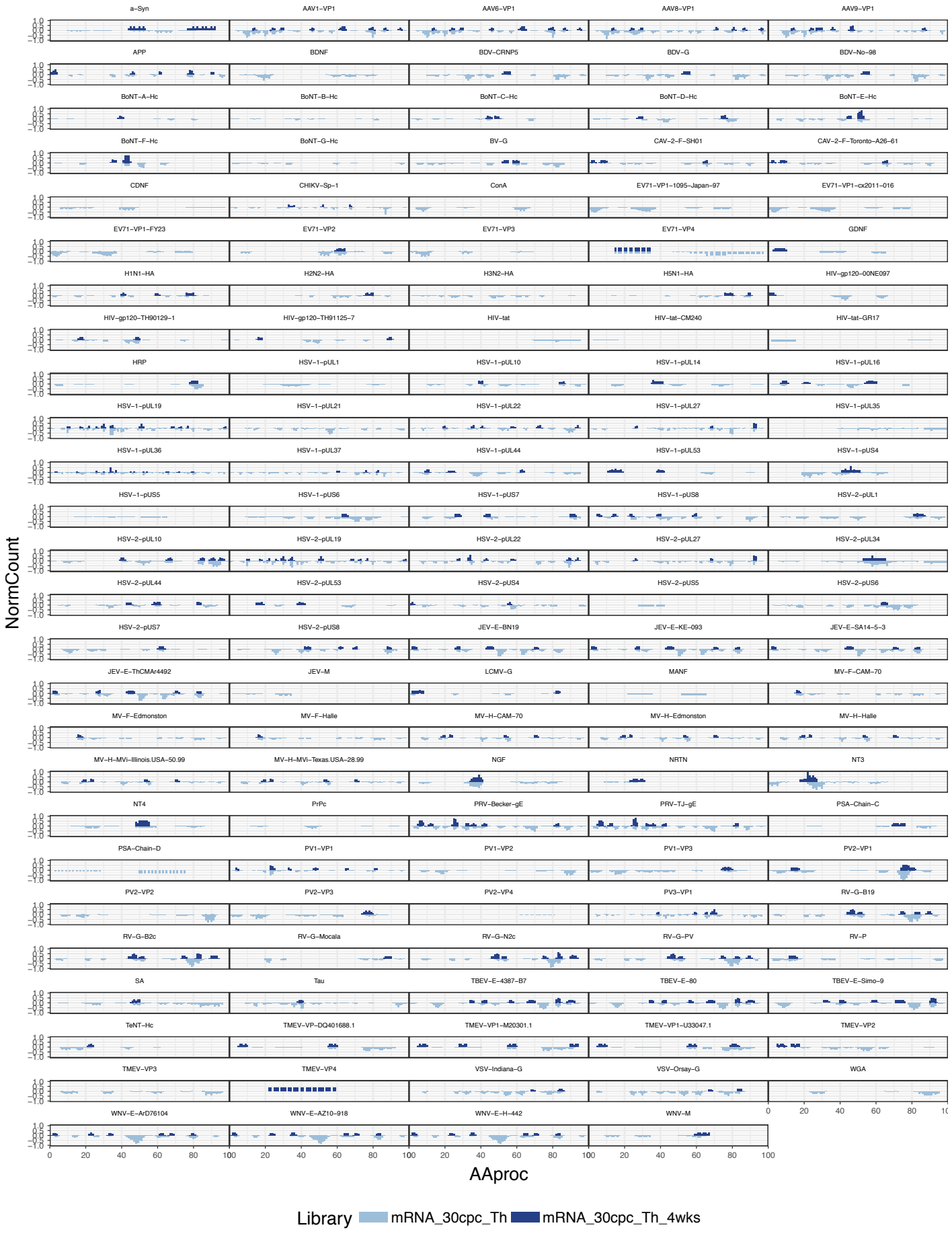

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_3cpc_Str

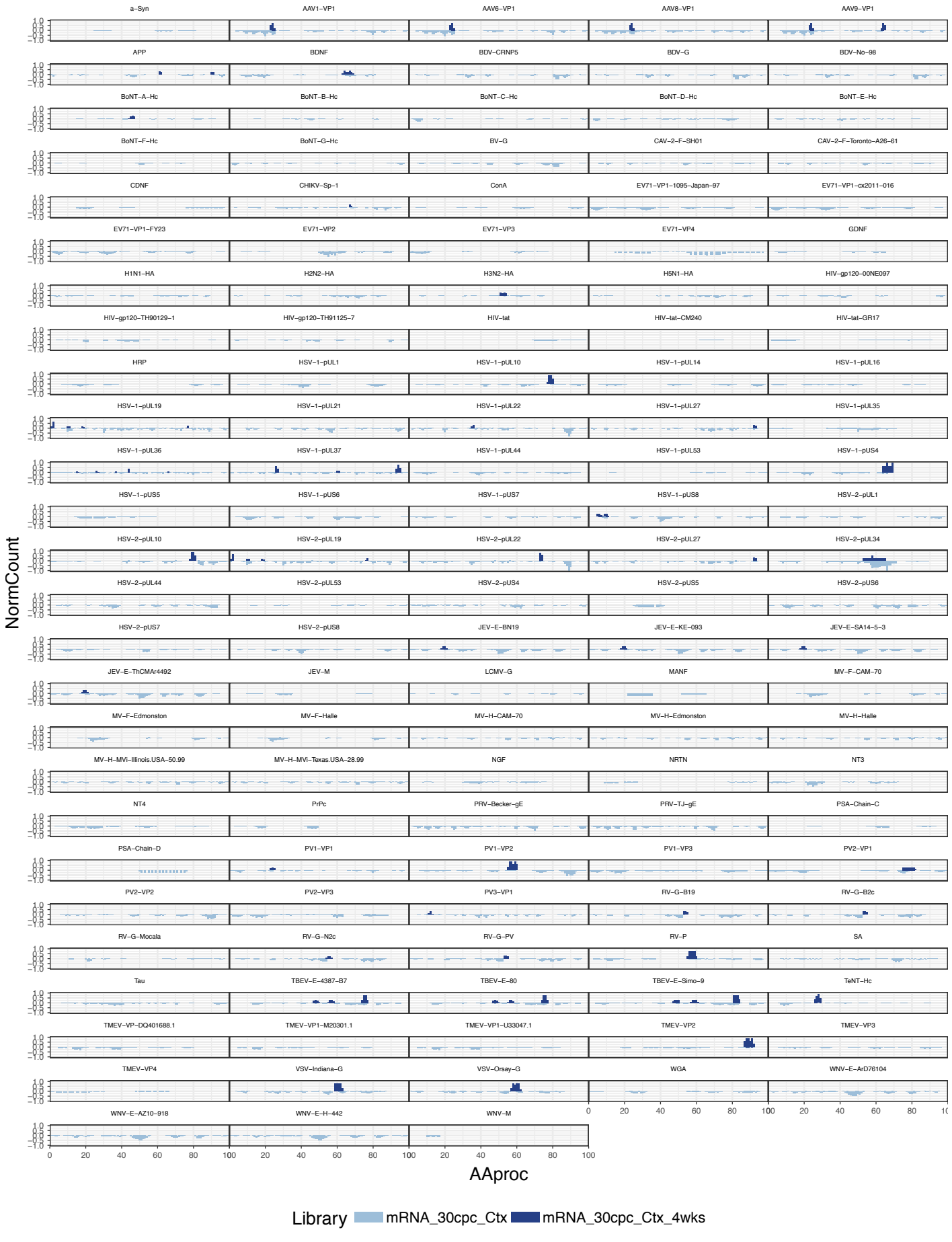

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Th
mRNA_3cpc_Th

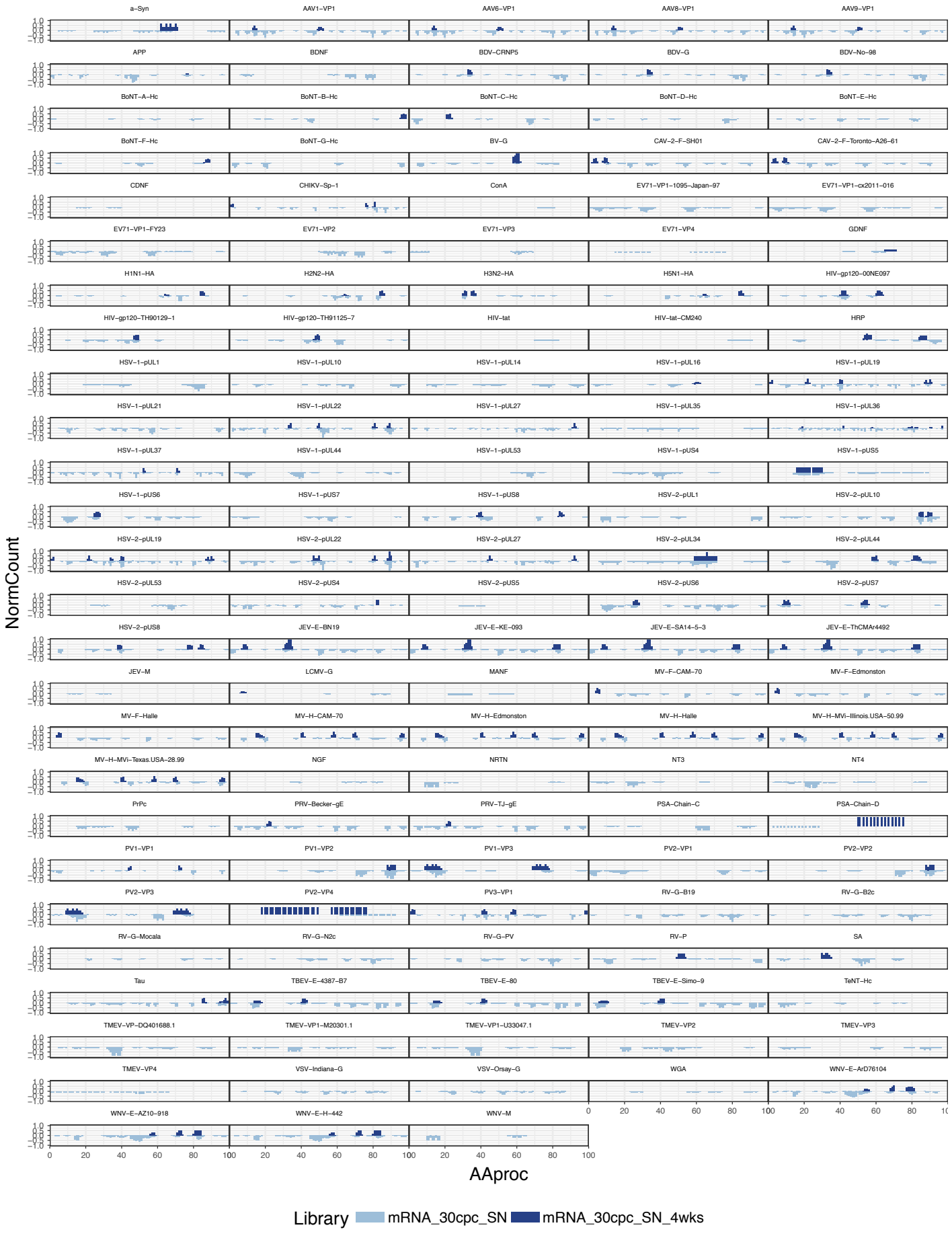

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
PV2−VP1
PV2−VP2
PV2−VP3
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_3cpc_Ctx

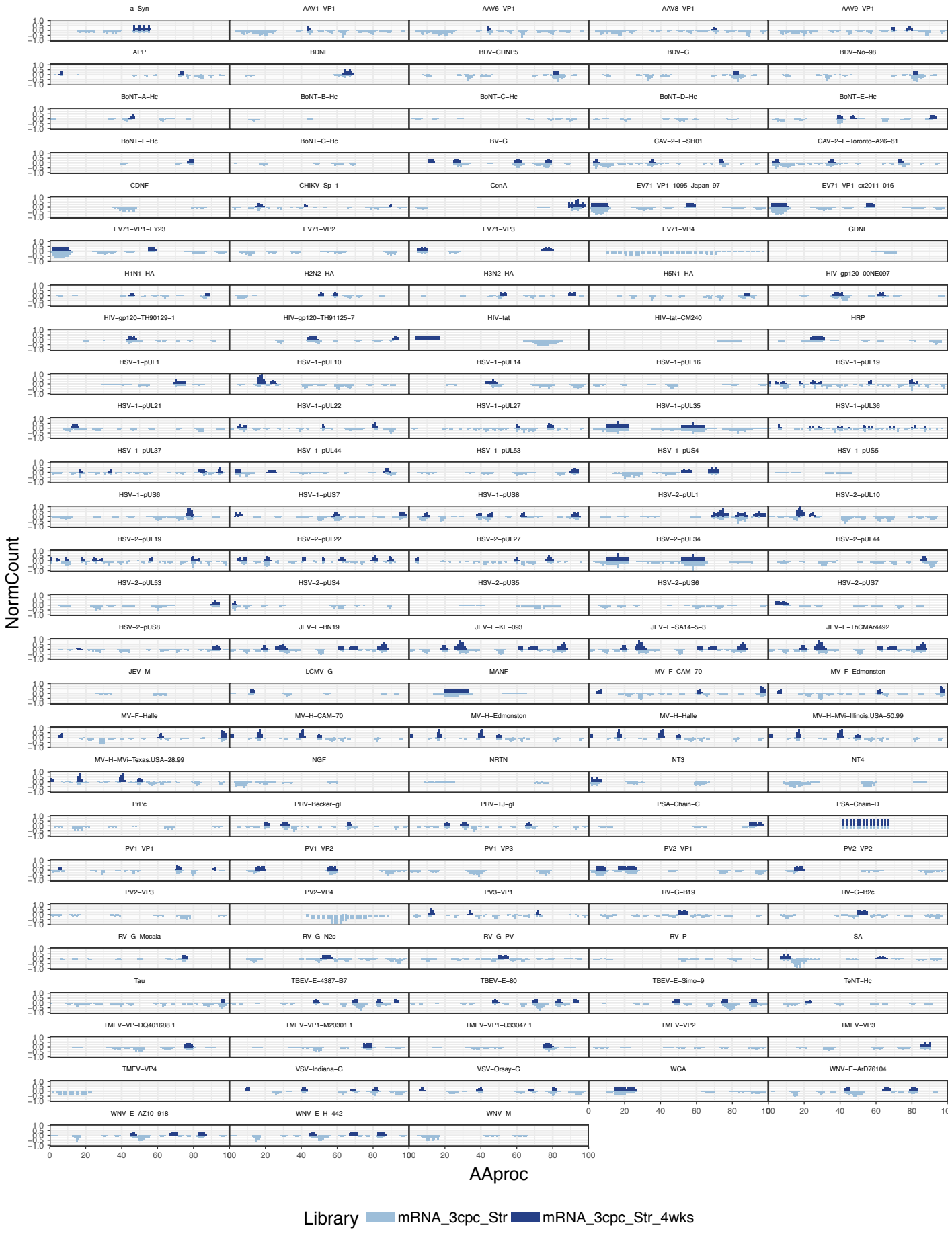

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_SN
mRNA_3cpc_SN

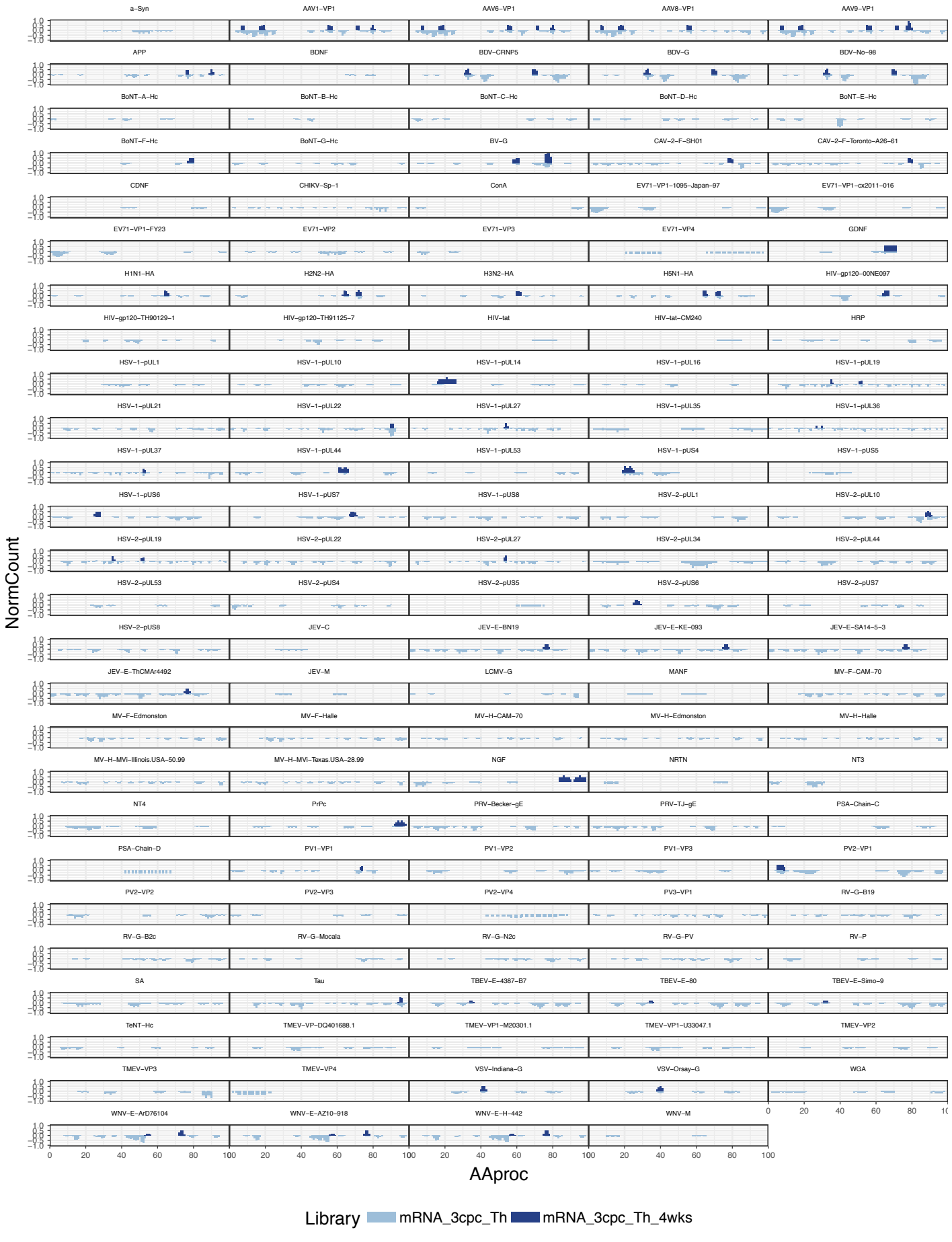

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Str_4wks

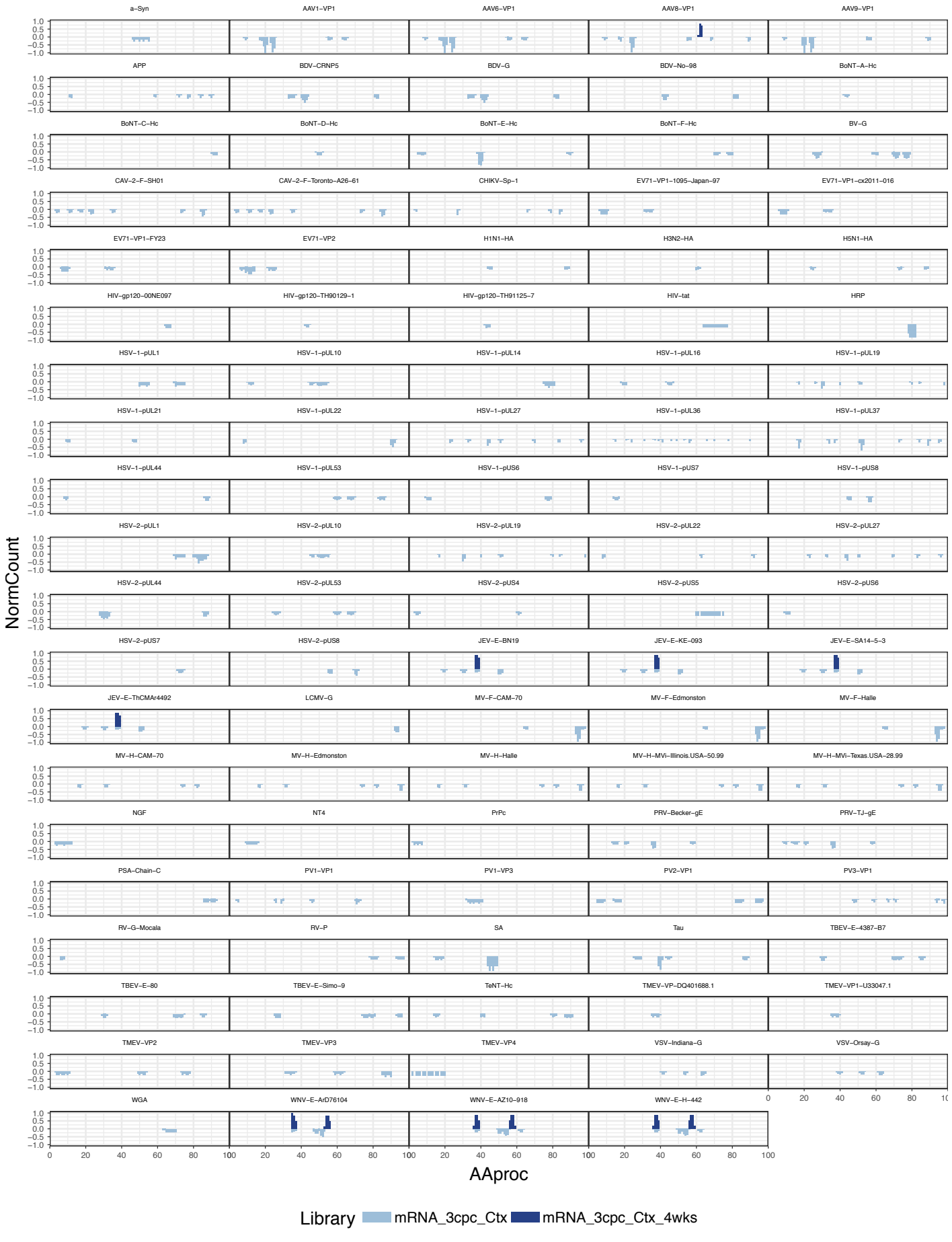

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Th
mRNA_30cpc_Th_4wks

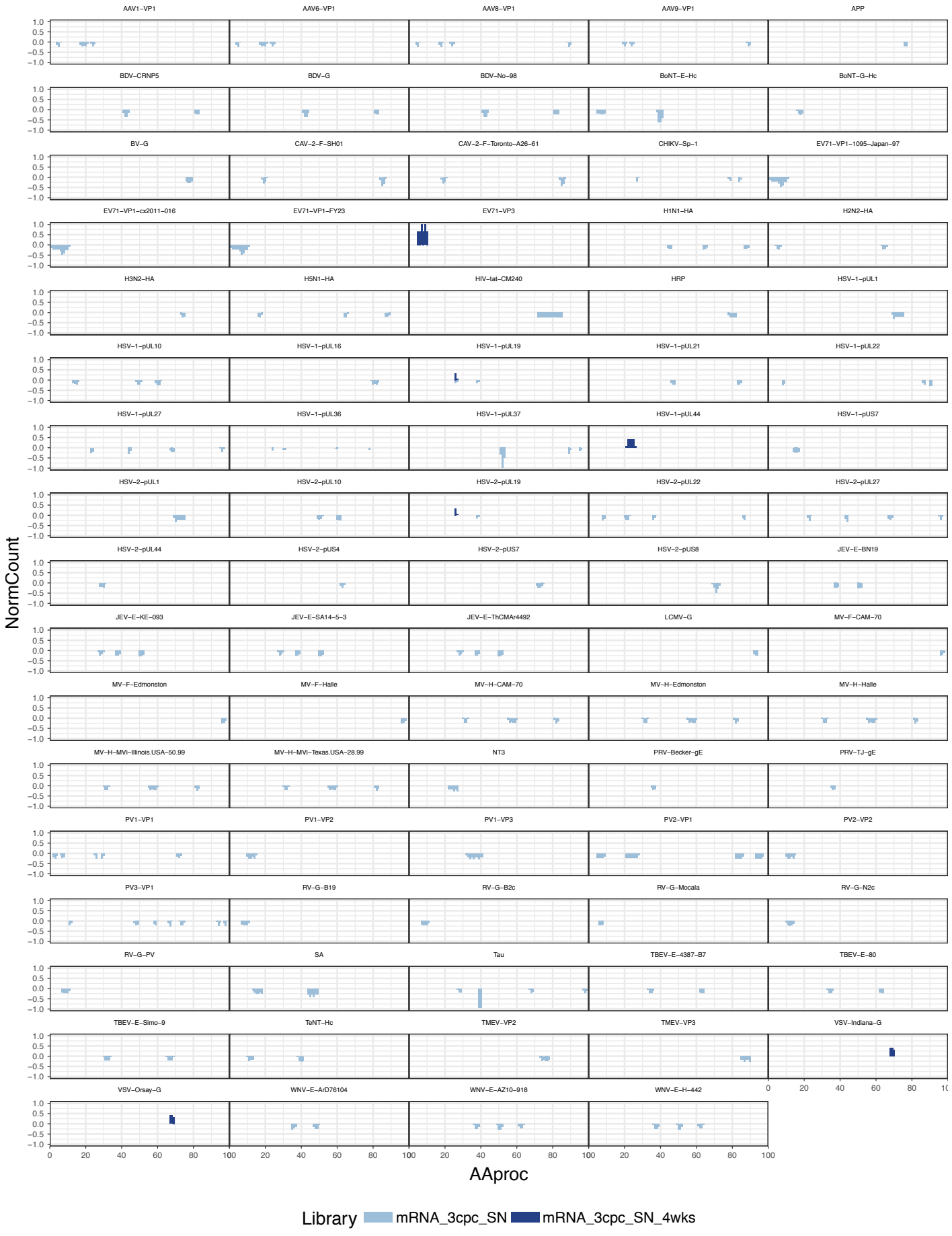

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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–E–BN19
JEV–E–KE–093
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
PV2–VP1
PV2–VP2
PV2–VP3
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
WGA
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Ctx_4wks

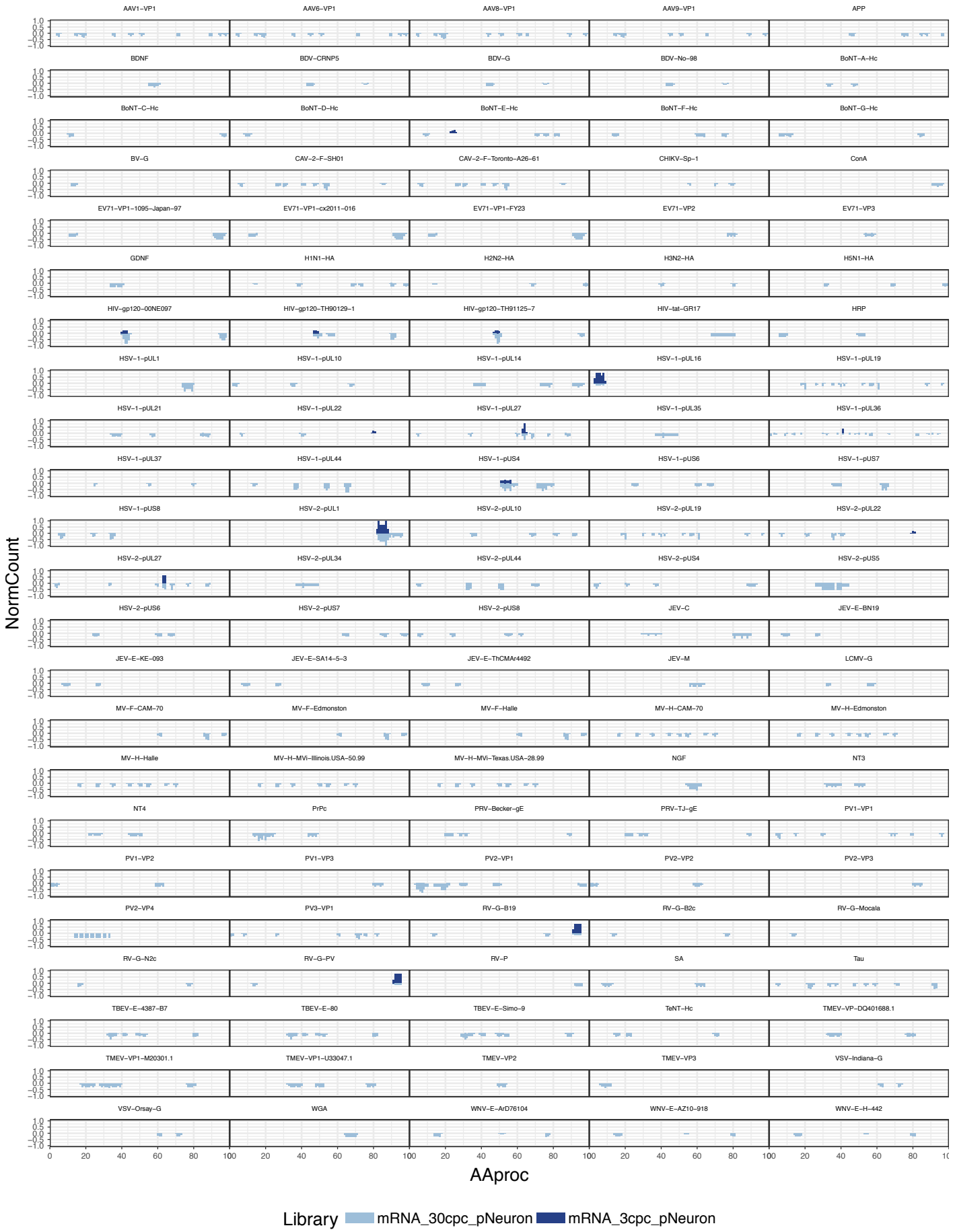

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_SN
mRNA_30cpc_SN_4wks

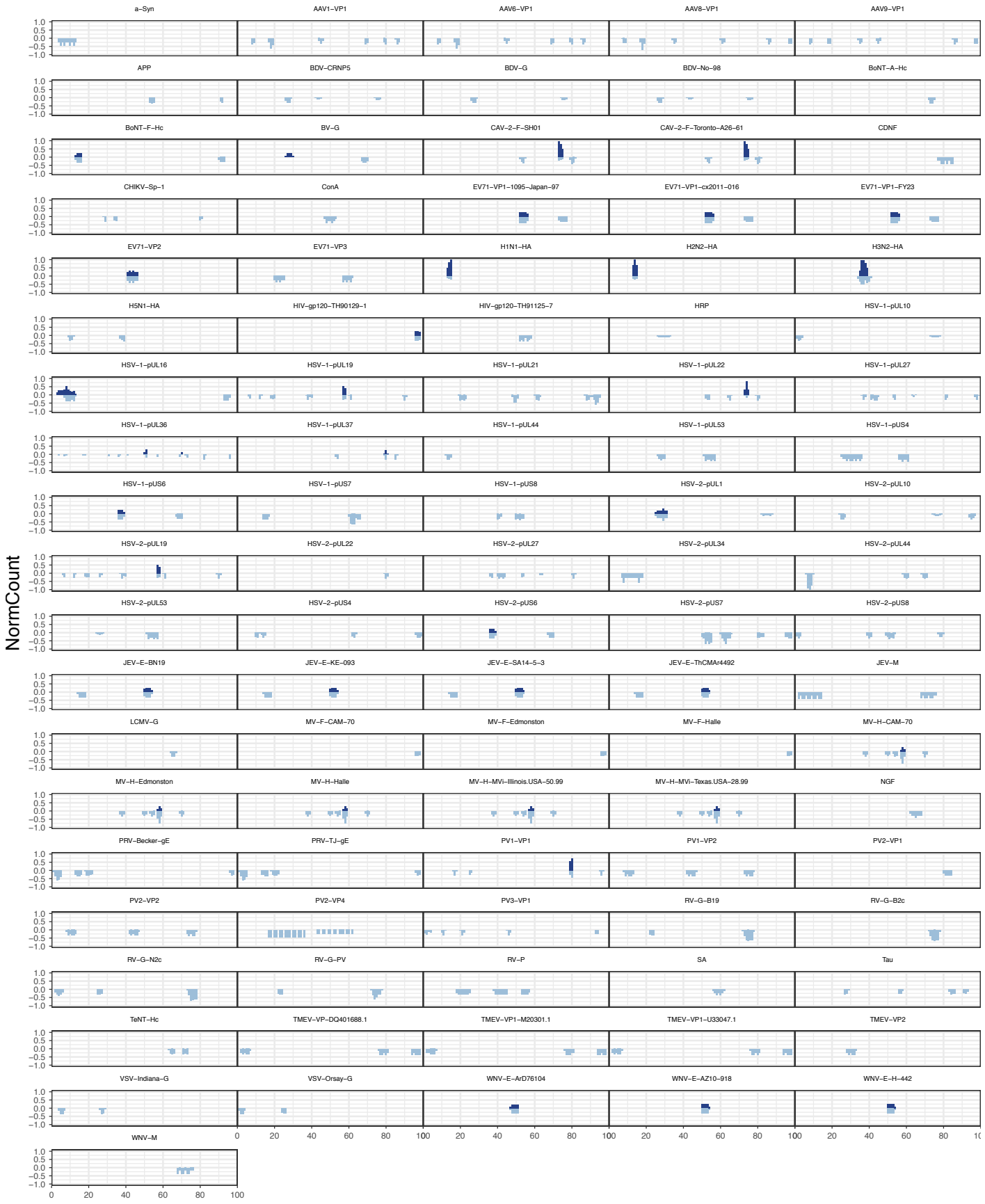

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Str_4wks

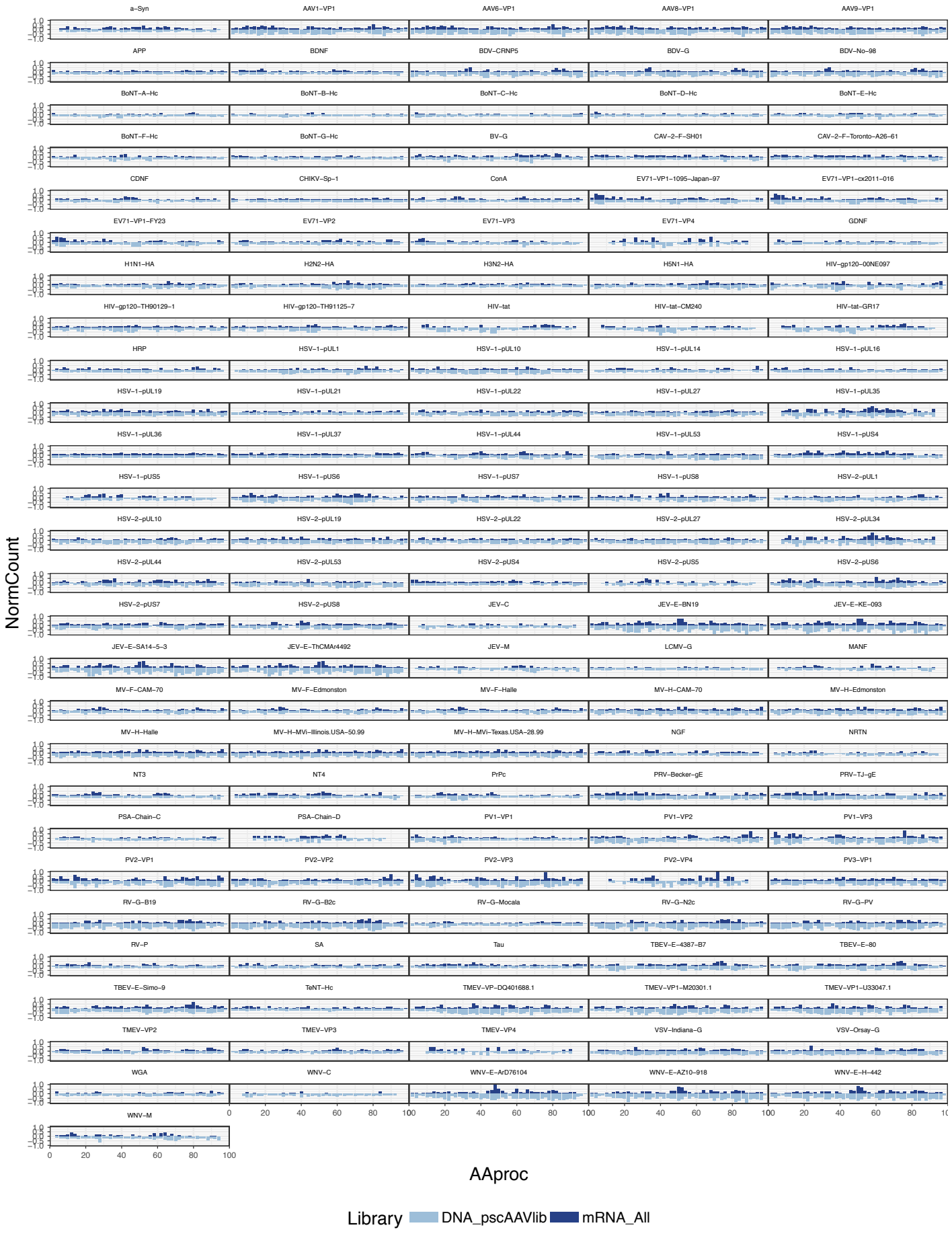

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Th
mRNA_3cpc_Th_4wks

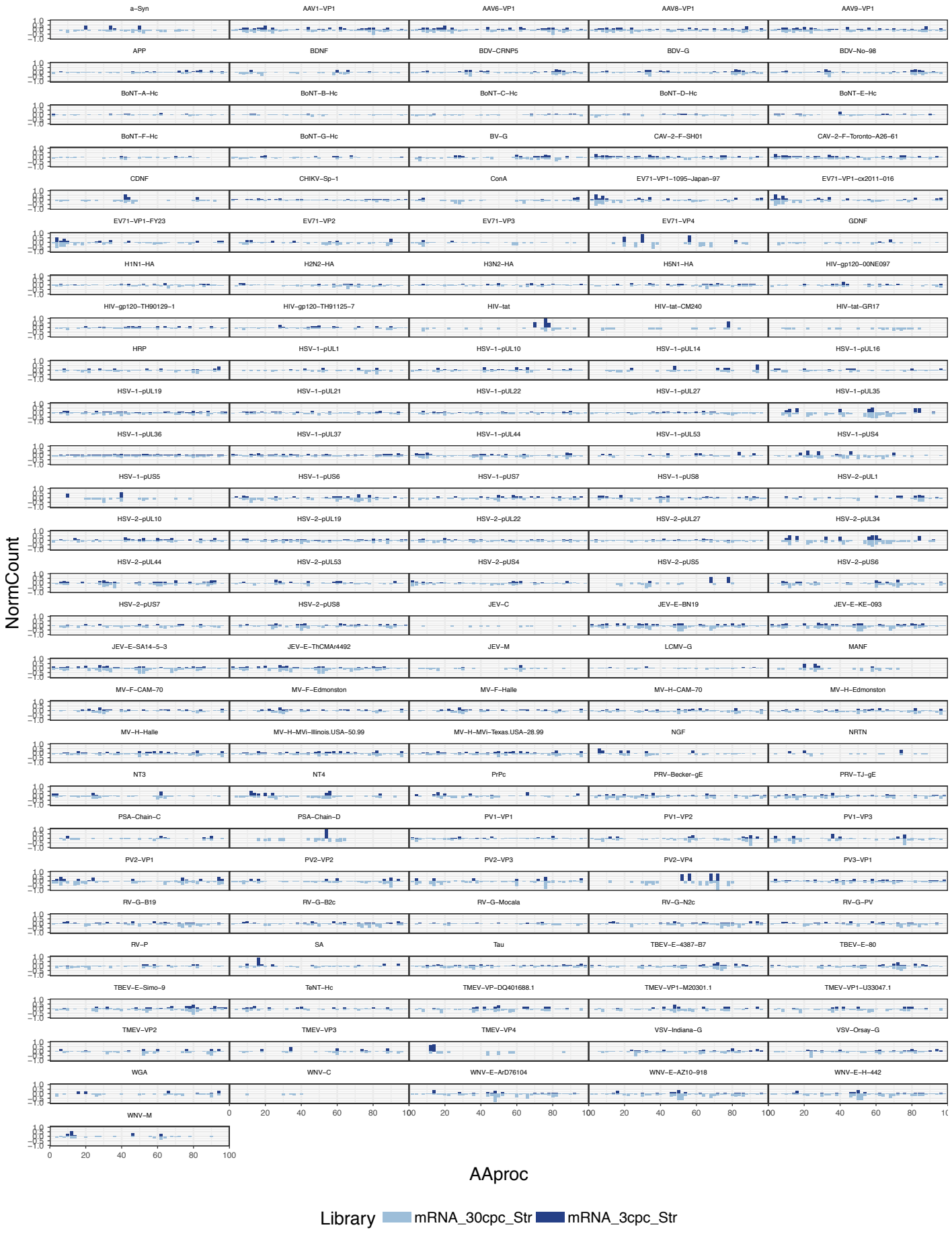

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CHIKV–Sp–1
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
H1N1–HA
H3N2–HA
H5N1–HA
HIV–gp120–00NE097
HIV–gp120–TH90129–1
HIV–gp120–TH91125–7
HIV–tat
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–pUL36
HSV–1–pUL37
HSV–1–pUL44
HSV–1–pUL53
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–pUL44
HSV–2–pUL53
HSV–2–pUS4
HSV–2–pUS5
HSV–2–pUS6
HSV–2–pUS7
HSV–2–pUS8
JEV–E–BN19
JEV–E–KE–093
JEV–E–SA14–5–3
JEV–E–ThCMAr4492
LCMV–G
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
NT4
PrPc
PRV–Becker–gE
PRV–TJ–gE
PSA–Chain–C
PV1–VP1
PV1–VP3
PV2–VP1
PV3–VP1
RV–G–Mocala
RV–P
SA
Tau
TBEV–E–4387–B7
TBEV–E–80
TBEV–E–Simo–9
TeNT–Hc
TMEV–VP–DQ401688.1
TMEV–VP1–U33047.1
TMEV–VP2
TMEV–VP3
TMEV–VP4
VSV–Indiana–G
VSV–Orsay–G
WGA
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Ctx_4wks

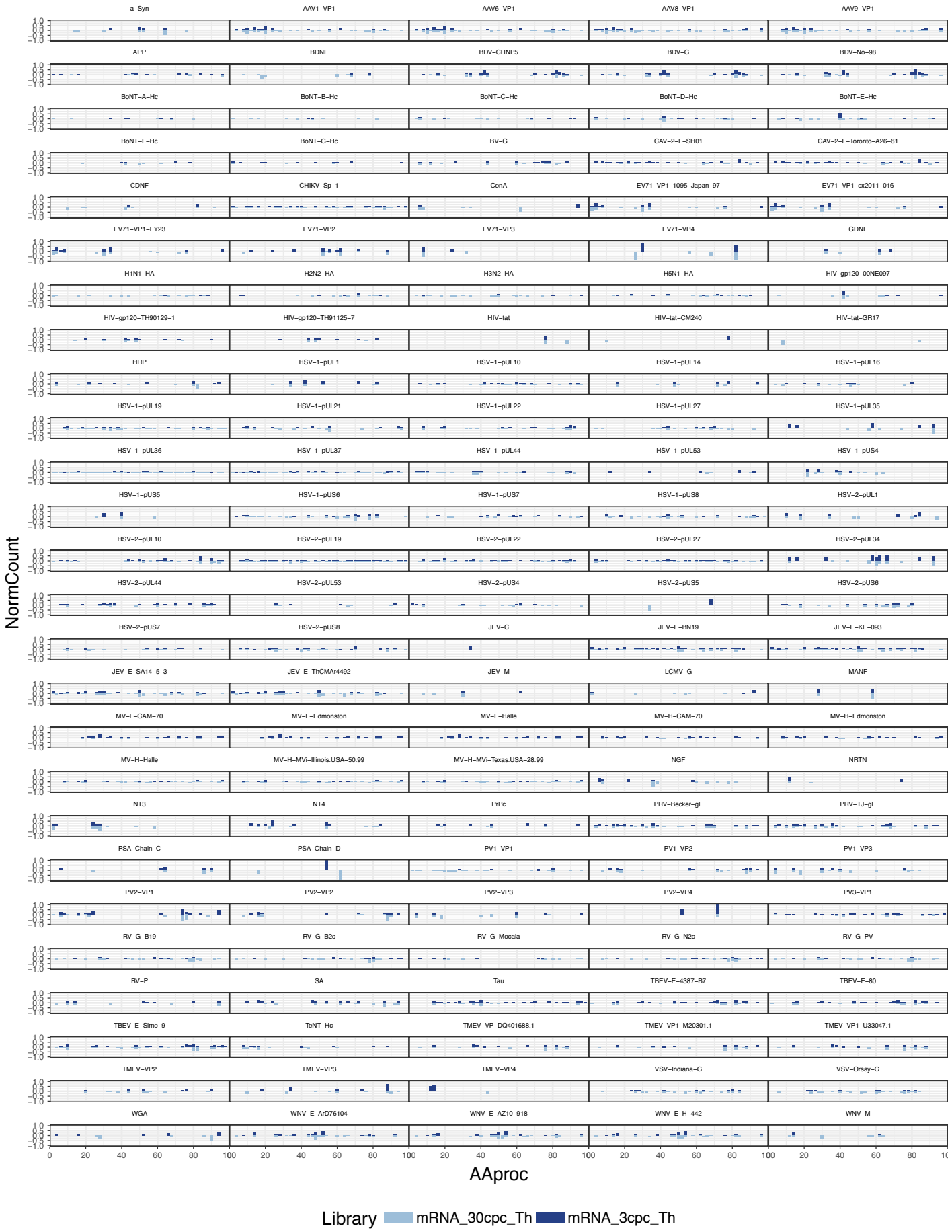

AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−E−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP3
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−tat−CM240
HRP
HSV−1−pUL1
HSV−1−pUL10
HSV−1−pUL16
HSV−1−pUL19
HSV−1−pUL21
HSV−1−pUL22
HSV−1−pUL27
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUS7
HSV−2−pUL1
HSV−2−pUL10
HSV−2−pUL19
HSV−2−pUL22
HSV−2−pUL27
HSV−2−pUL44
HSV−2−pUS4
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
LCMV−G
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
NT3
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV1−VP3
PV2−VP1
PV2−VP2
PV3−VP1
RV−G−B19
RV−G−B2c
RV−G−Mocala
RV−G−N2c
RV−G−PV
SA
Tau
TBEV−E−4387−B7
TBEV−E−80
TBEV−E−Simo−9
TeNT−Hc
TMEV−VP2
TMEV−VP3
VSV−Indiana−G
VSV−Orsay−G
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_SN
mRNA_3cpc_SN_4wks

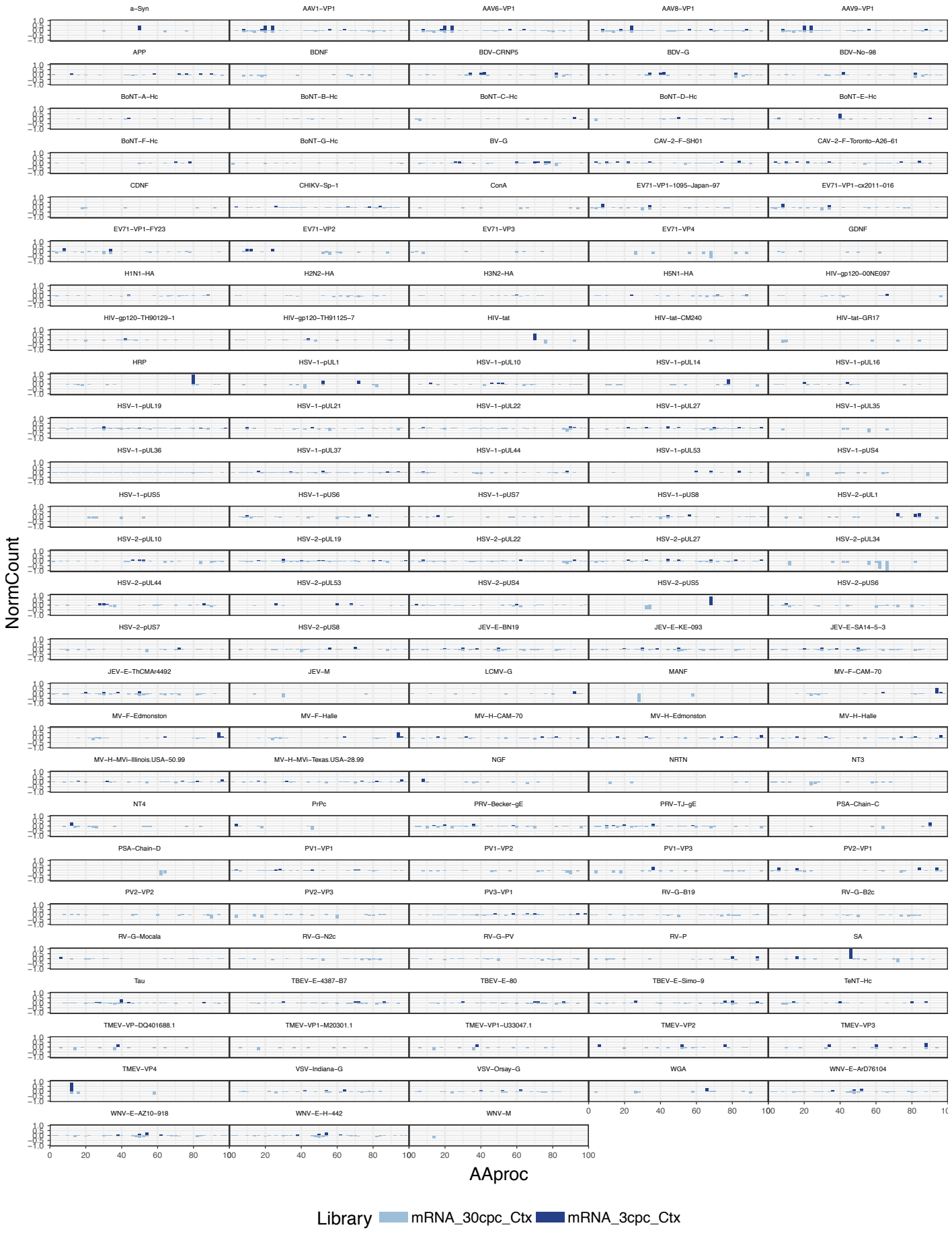

AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
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
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUS4
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−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−C
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
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
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV1−VP3
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
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_pNeuron
mRNA_3cpc_pNeuron

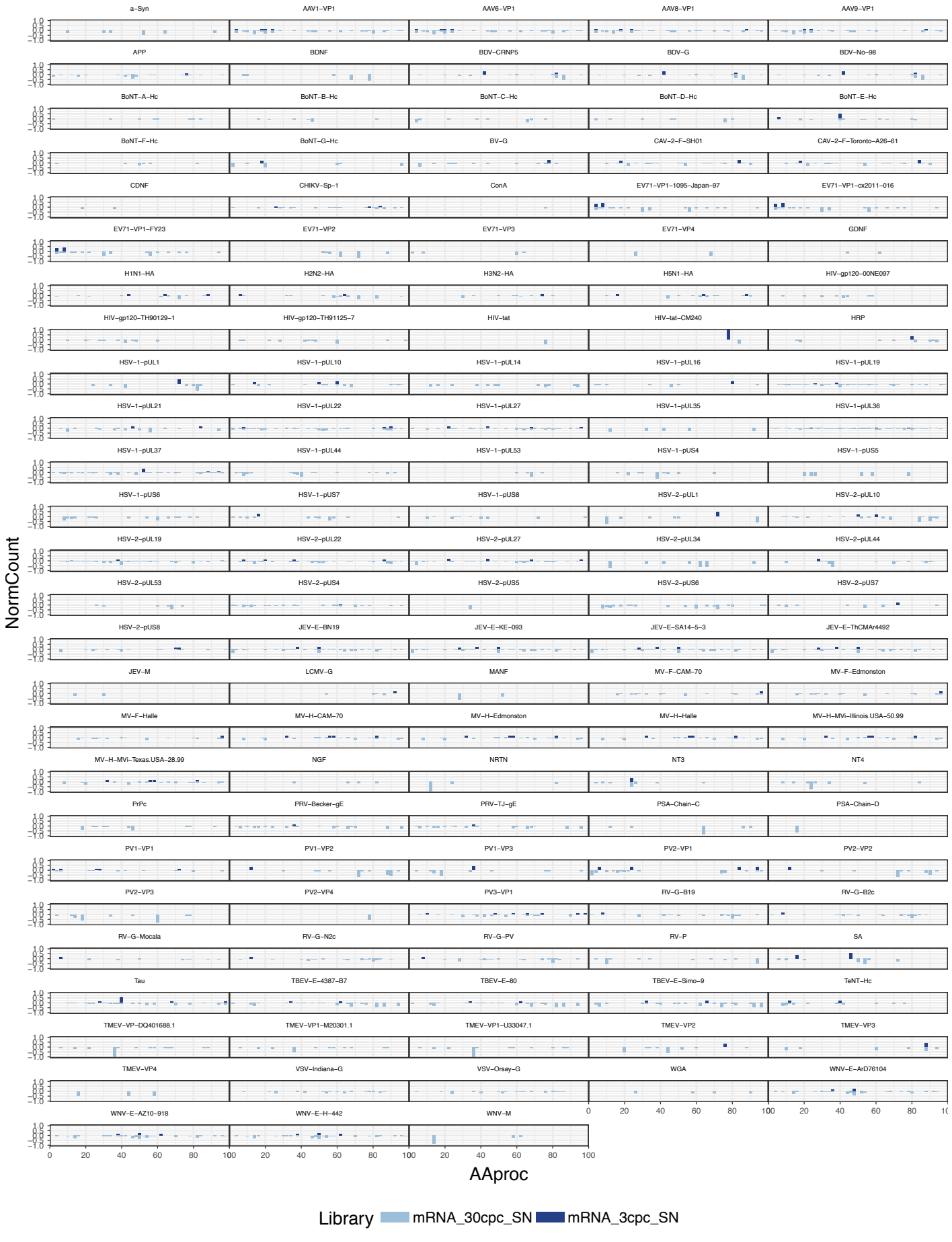

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−F−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HRP
HSV−1−pUL10
HSV−1−pUL16
HSV−1−pUL19
HSV−1−pUL21
HSV−1−pUL22
HSV−1−pUL27
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
HSV−1−pUS4
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−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
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
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV2−VP1
PV2−VP2
PV2−VP4
PV3−VP1
RV−G−B19
RV−G−B2c
RV−G−N2c
RV−G−PV
RV−P
SA
Tau
TeNT−Hc
TMEV−VP−DQ401688.1
TMEV−VP1−M20301.1
TMEV−VP1−U33047.1
TMEV−VP2
VSV−Indiana−G
VSV−Orsay−G
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_HEK293T
mRNA_3cpc_HEK293T

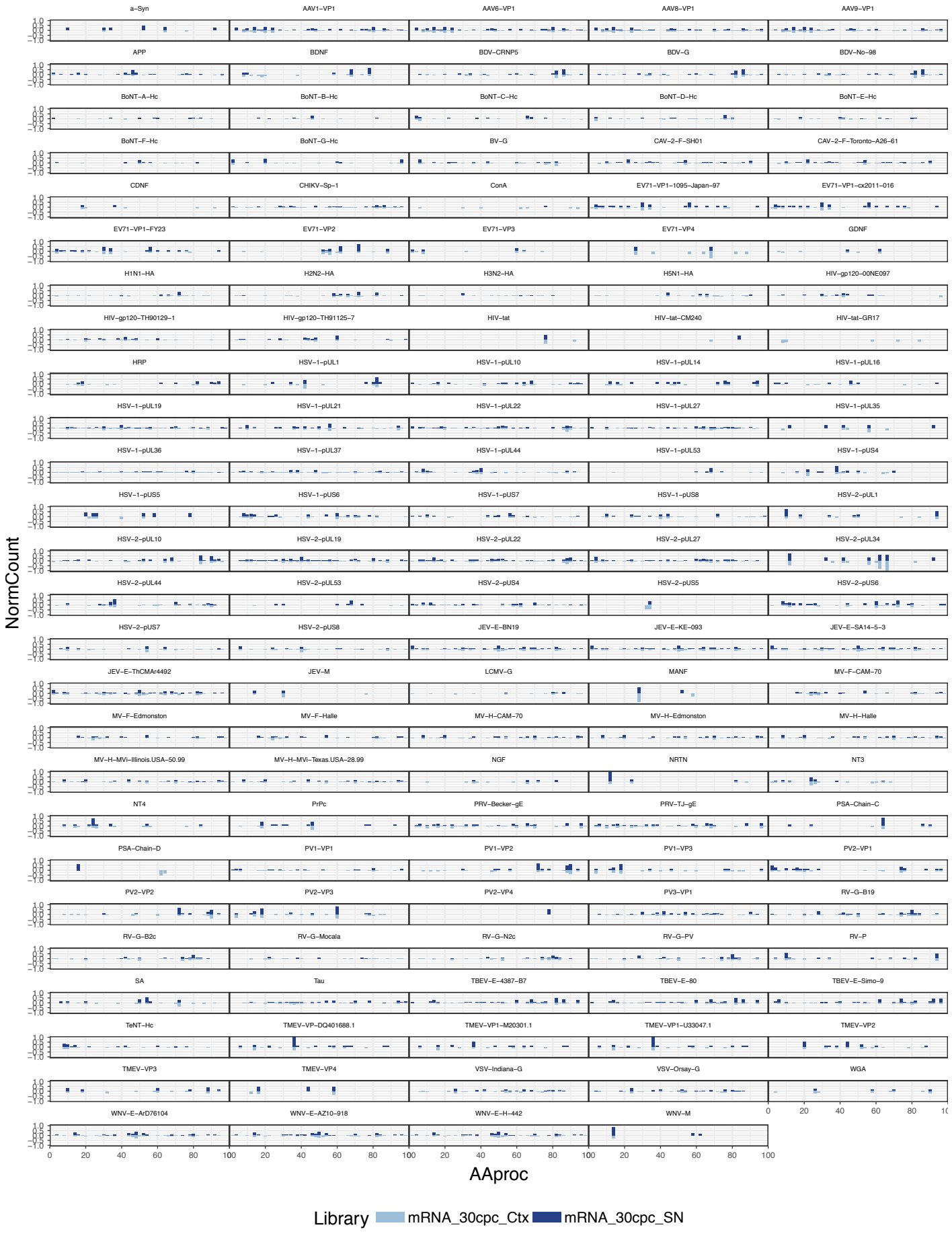

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
DNA_pscAAVlib
mRNA_All

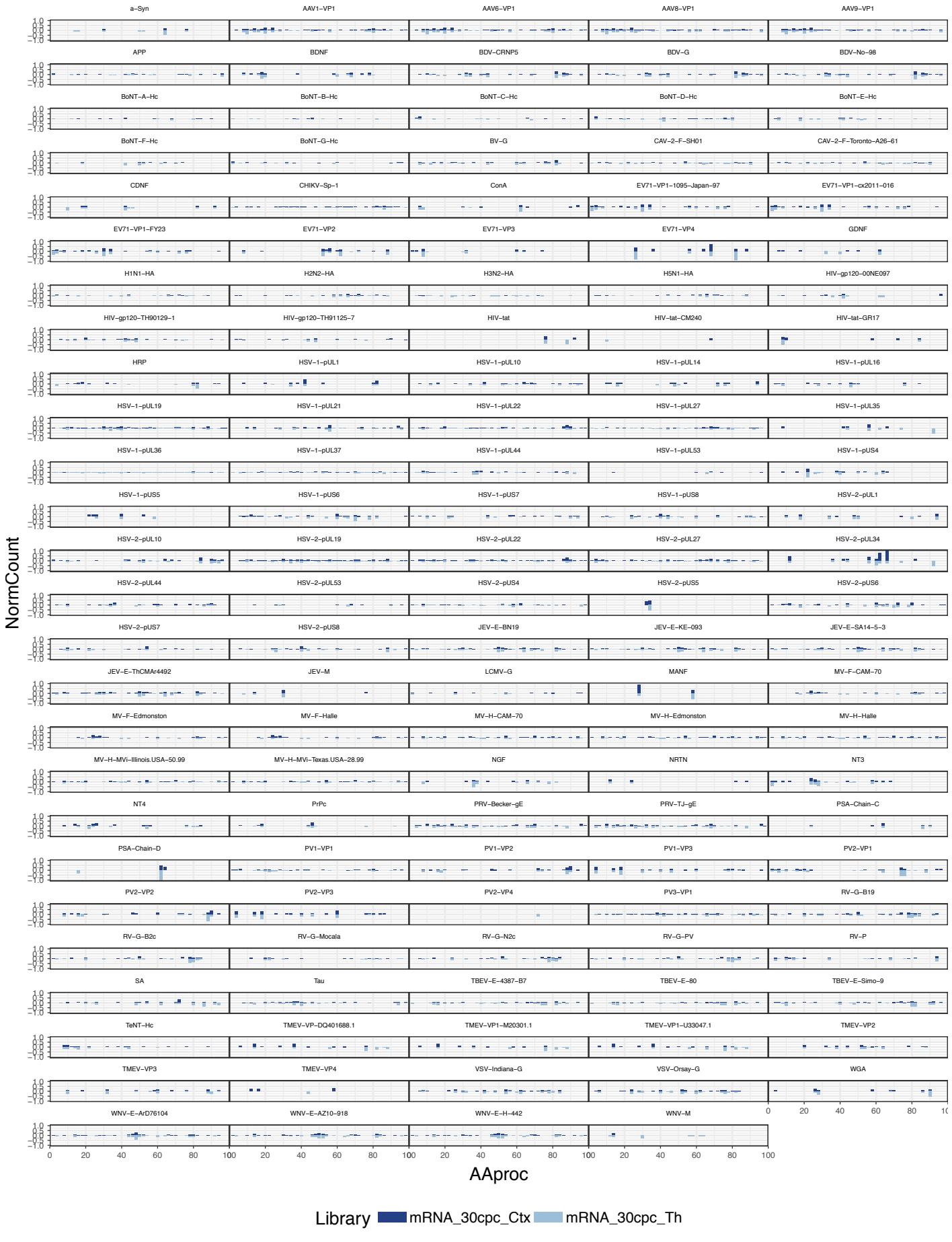

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_3cpc_Str

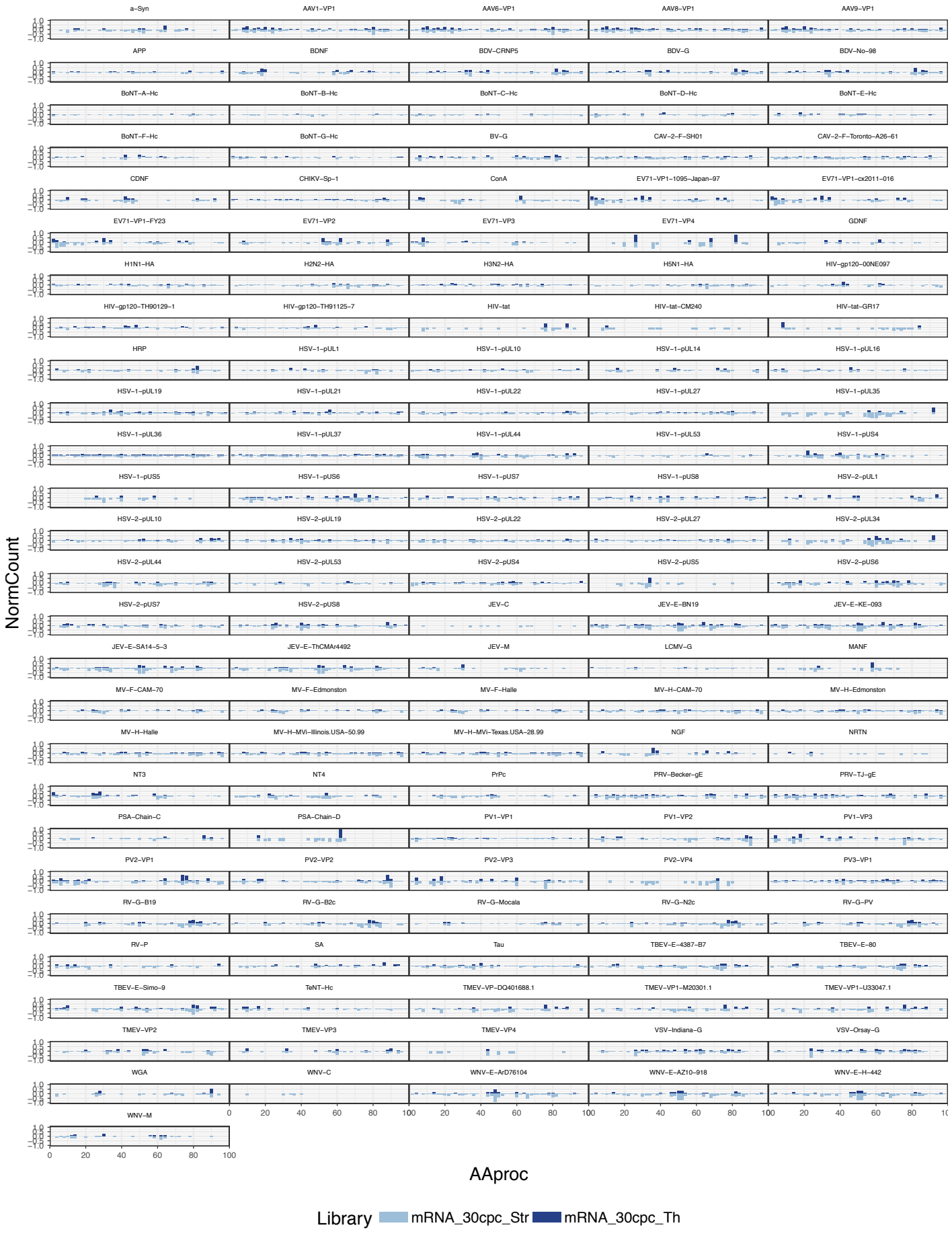

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Th
mRNA_3cpc_Th

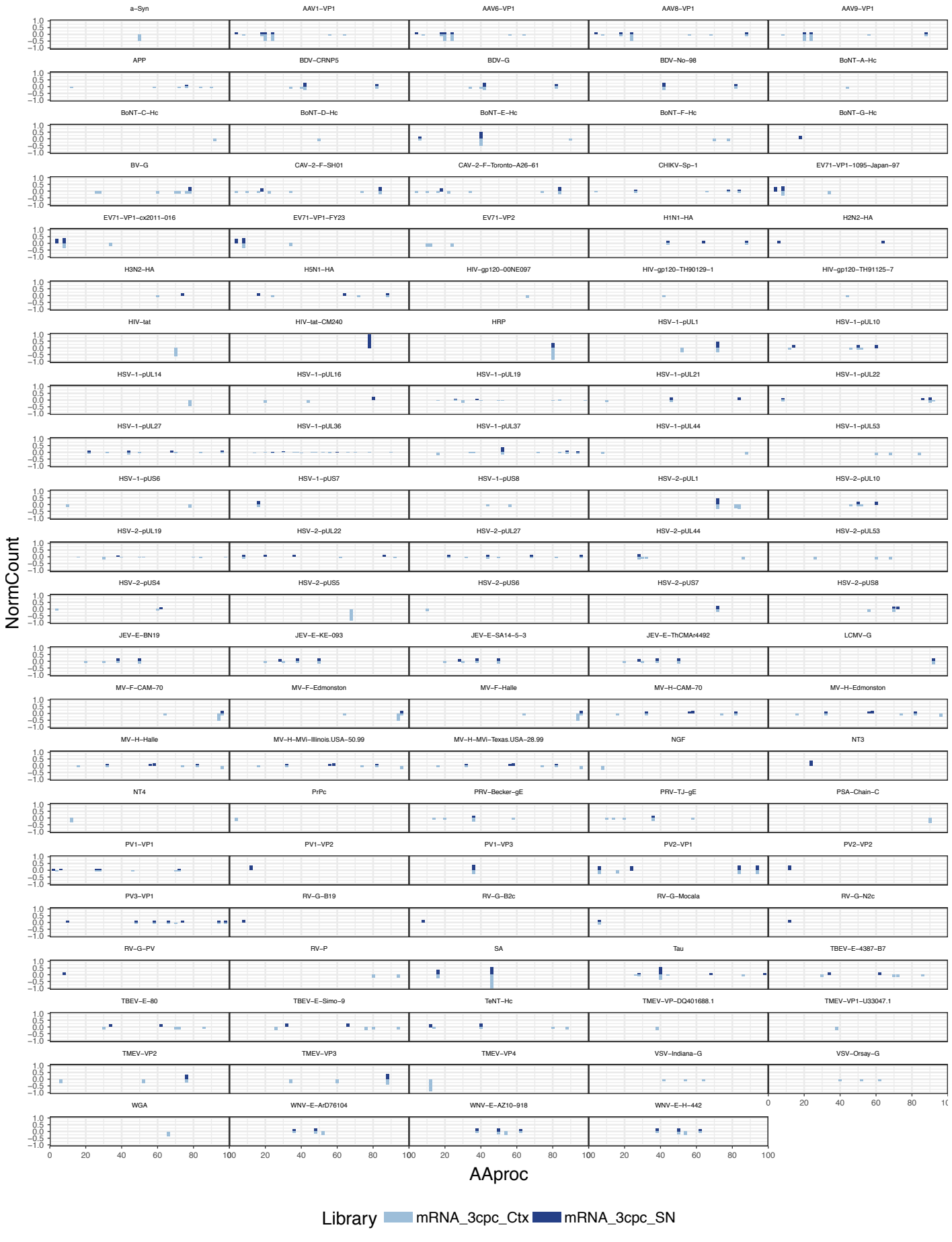

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
PV2−VP1
PV2−VP2
PV2−VP3
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_3cpc_Ctx

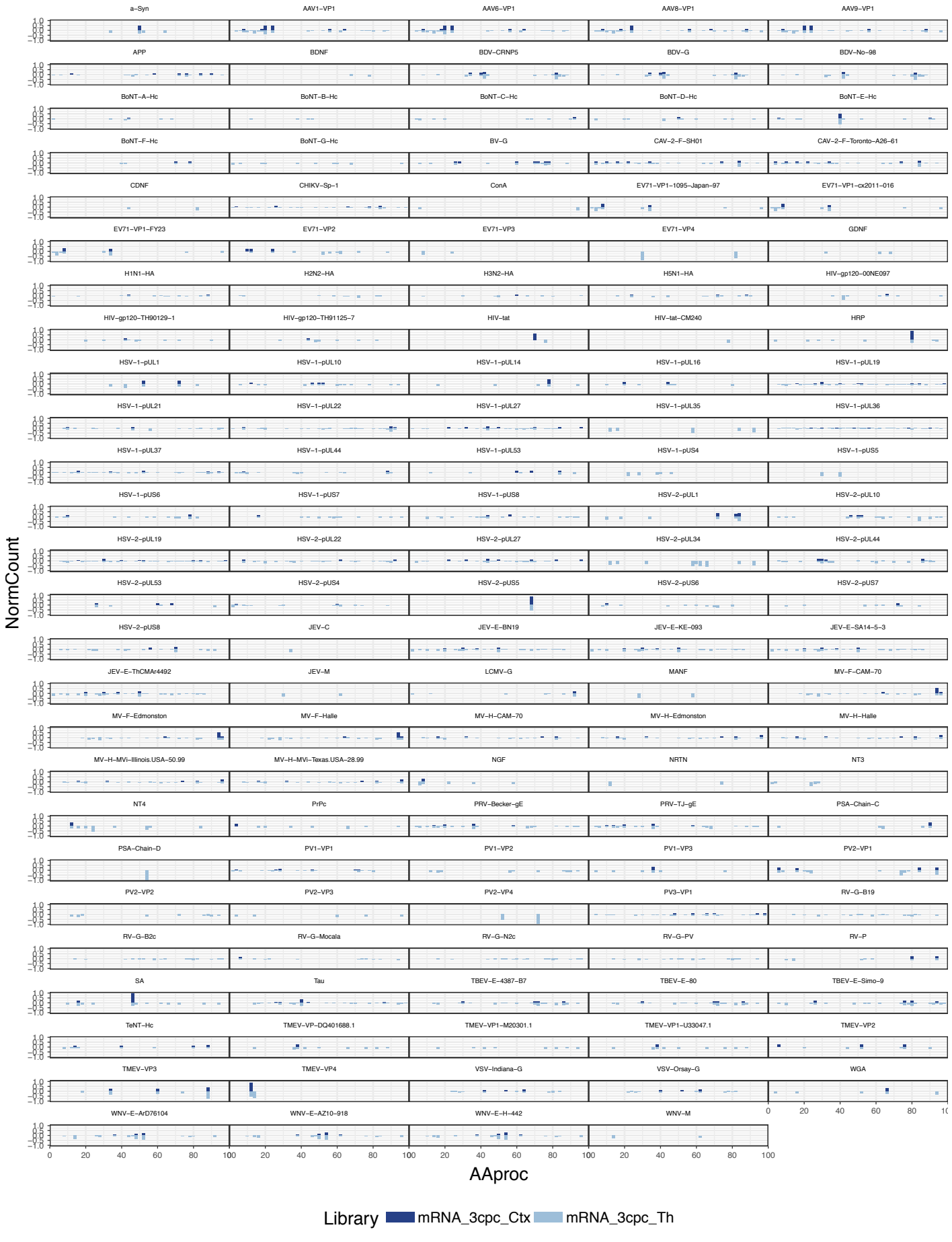

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
NormCount
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
AAproc
Library
mRNA_30cpc_SN
mRNA_3cpc_SN

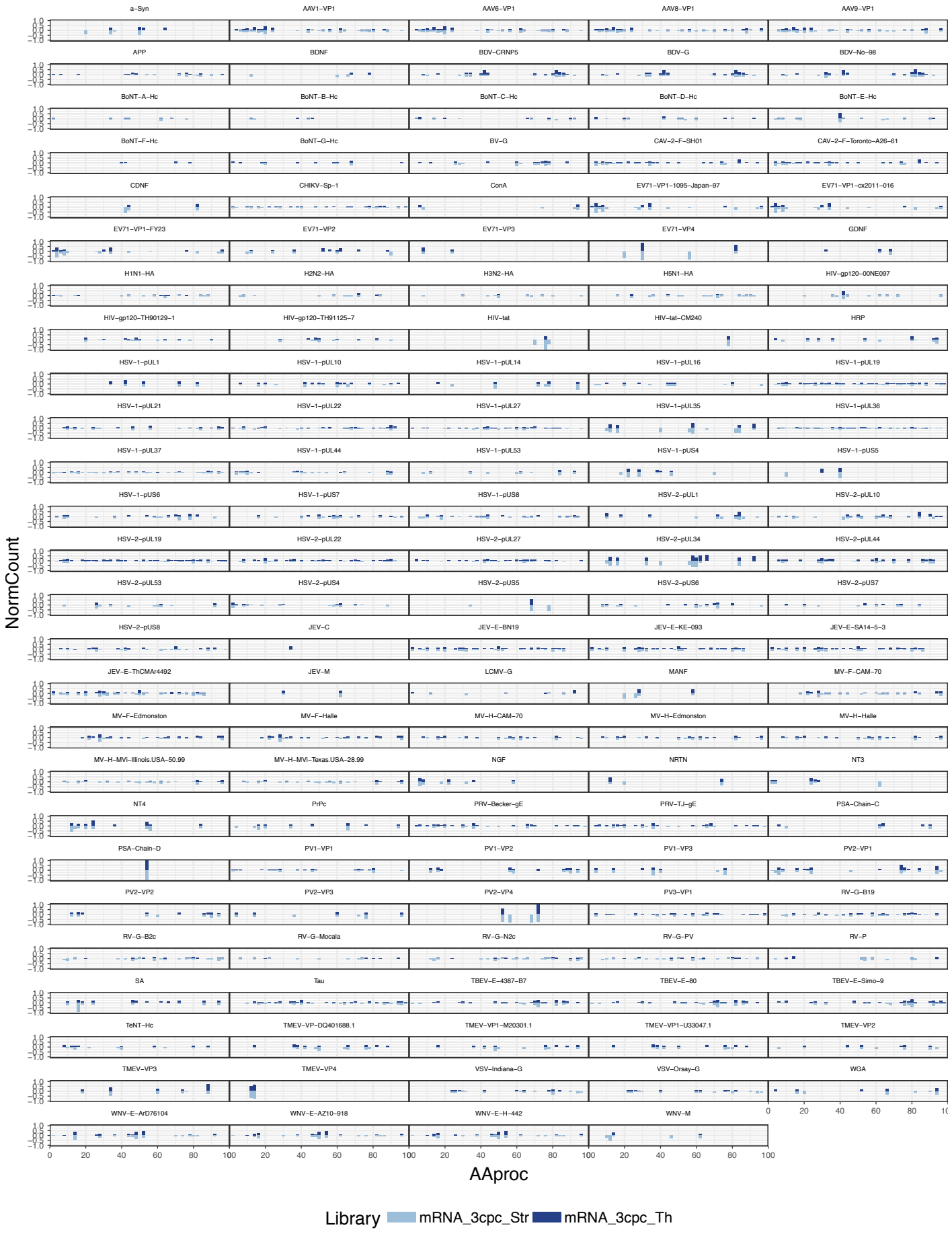

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN

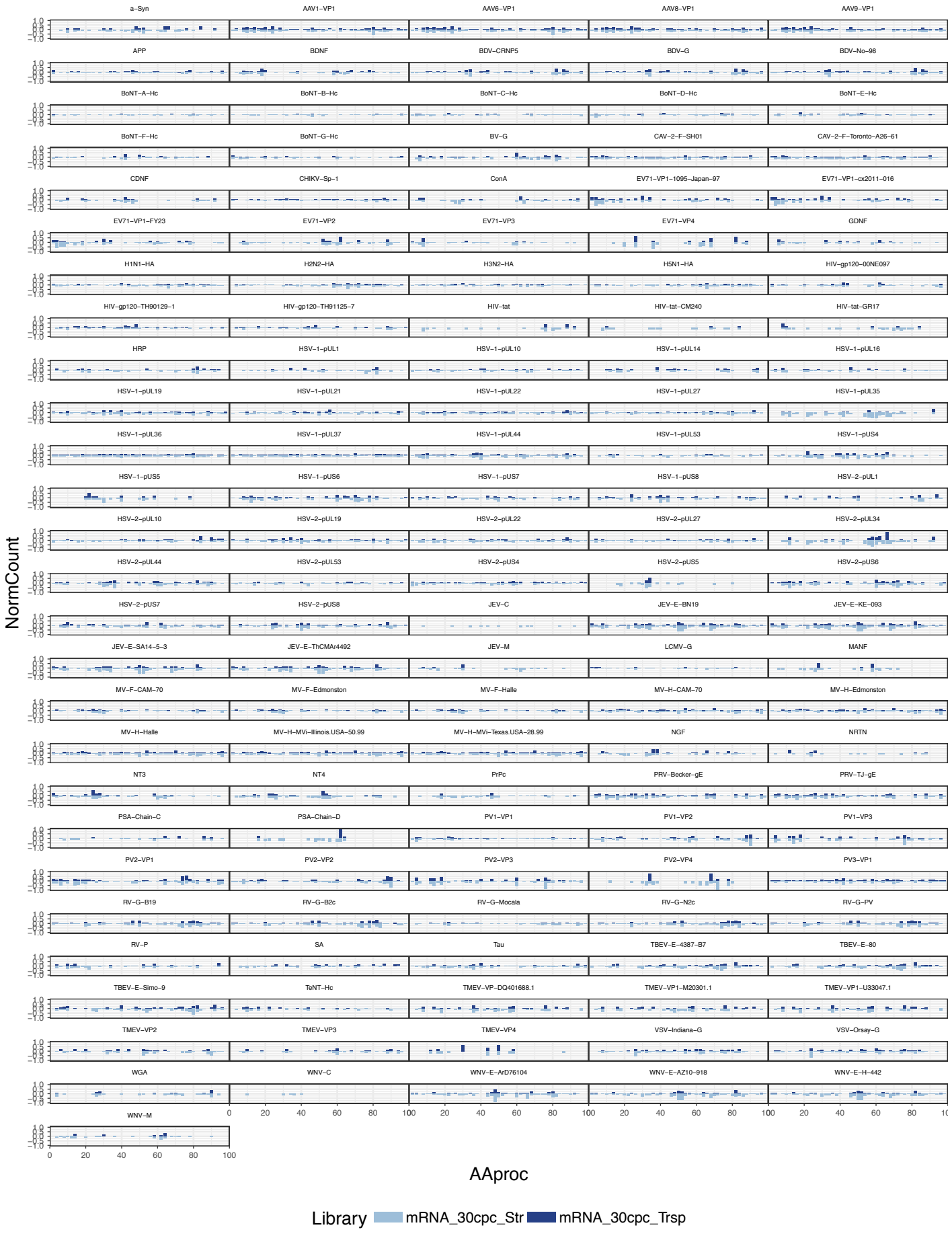

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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–E–BN19
JEV–E–KE–093
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
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
WGA
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Th

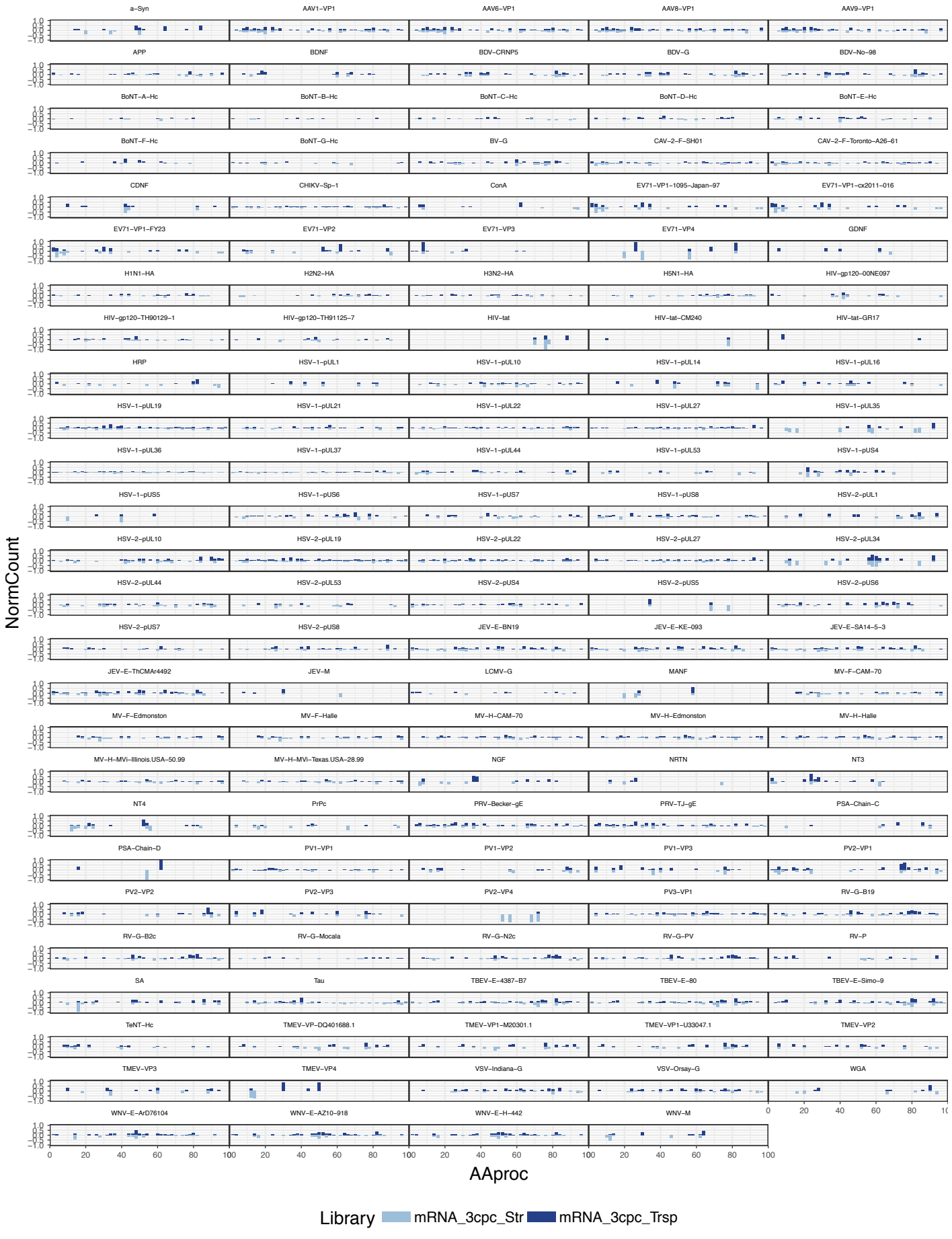

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Th

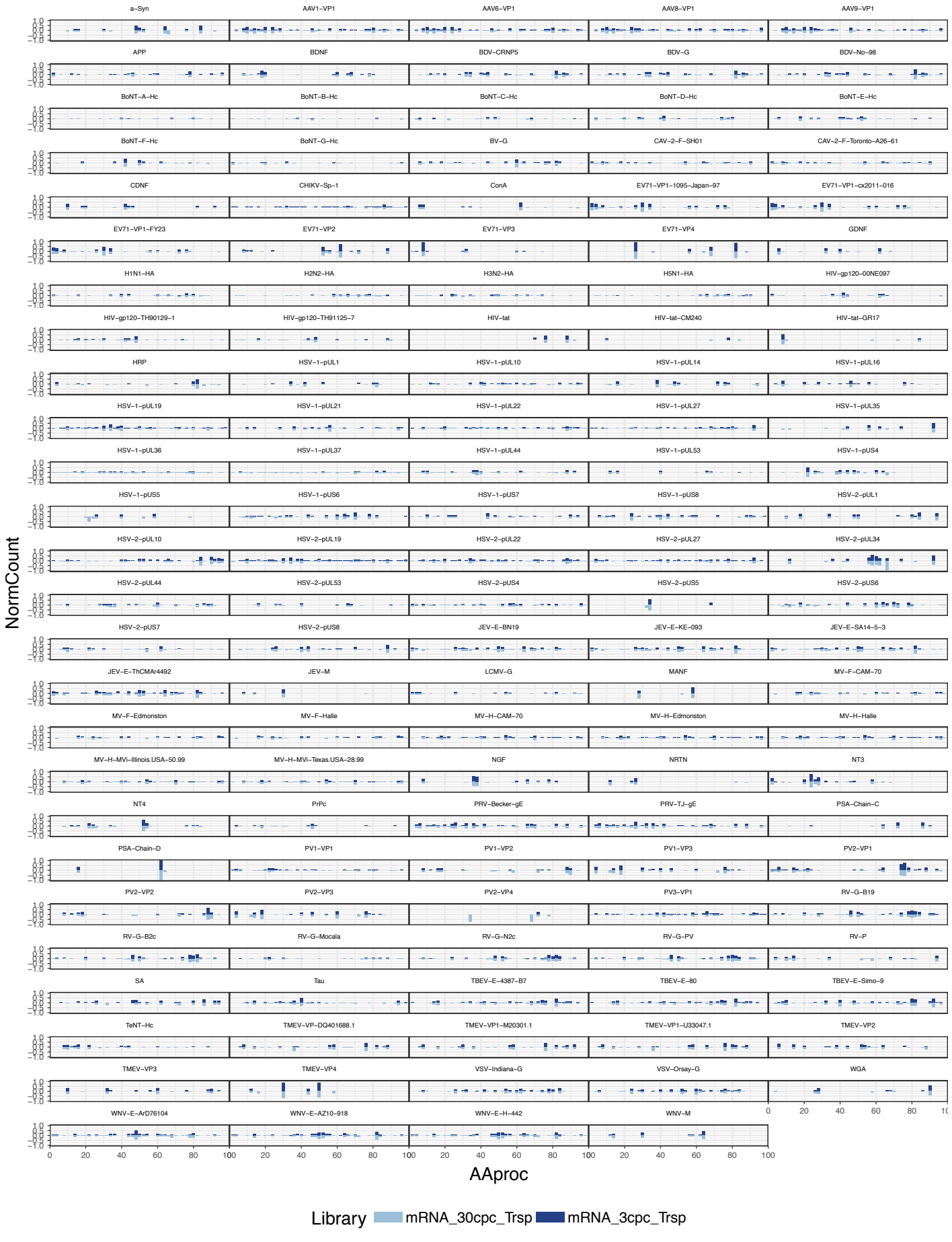

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
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−pUL44
HSV−2−pUL53
HSV−2−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
LCMV−G
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
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PSA−Chain−C
PV1−VP1
PV1−VP2
PV1−VP3
PV2−VP1
PV2−VP2
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−U33047.1
TMEV−VP2
TMEV−VP3
TMEV−VP4
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_SN

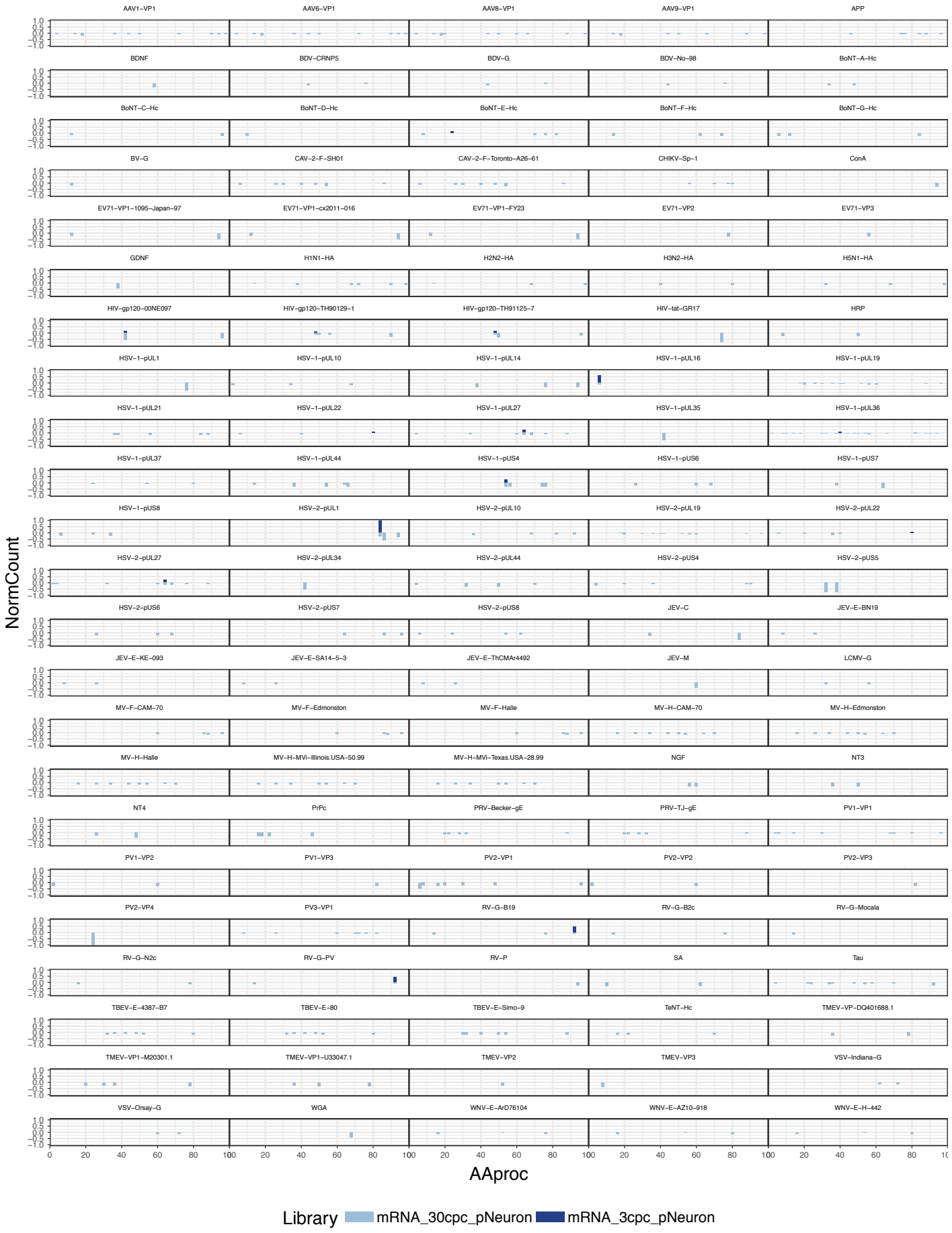

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Th

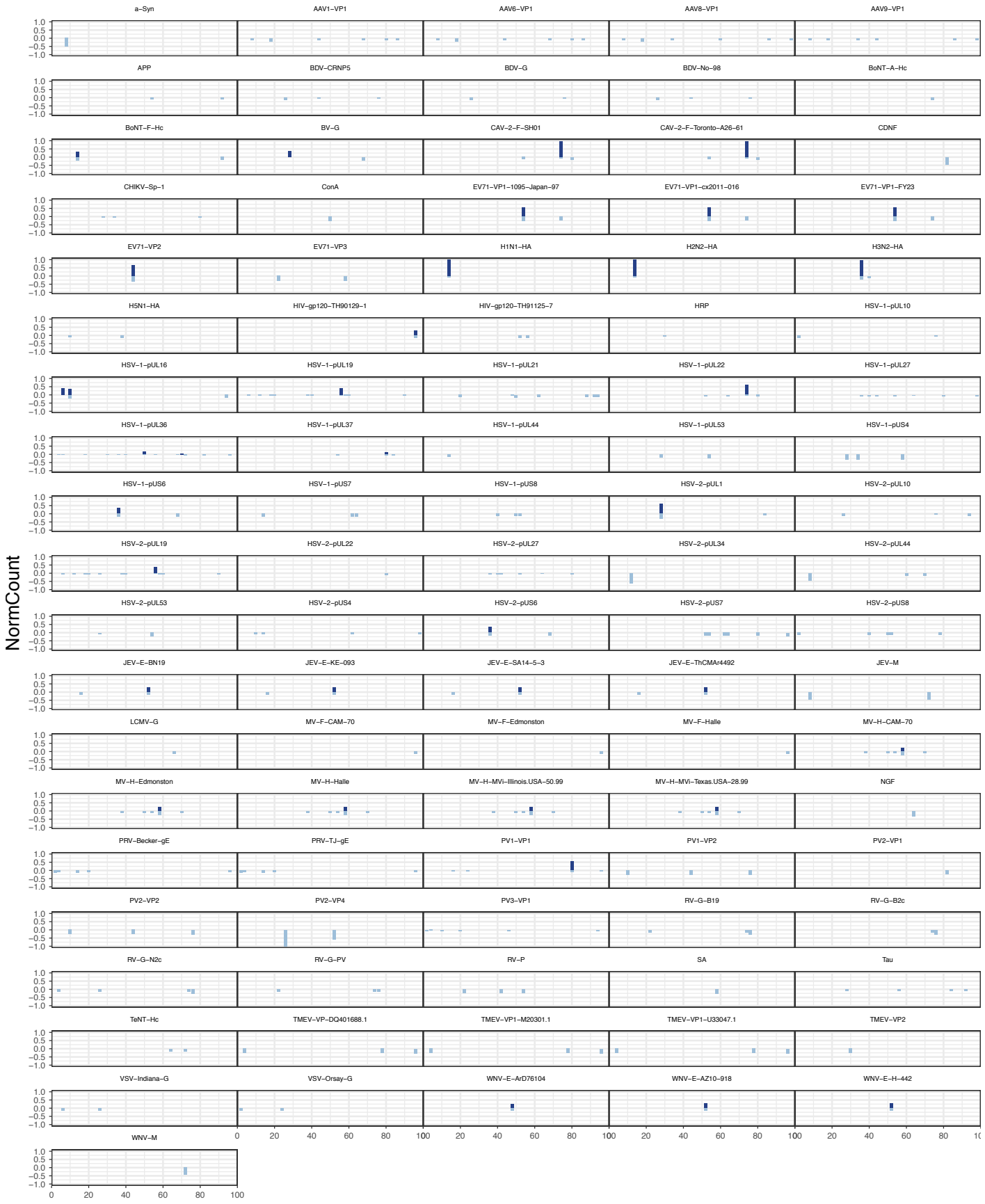

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Th

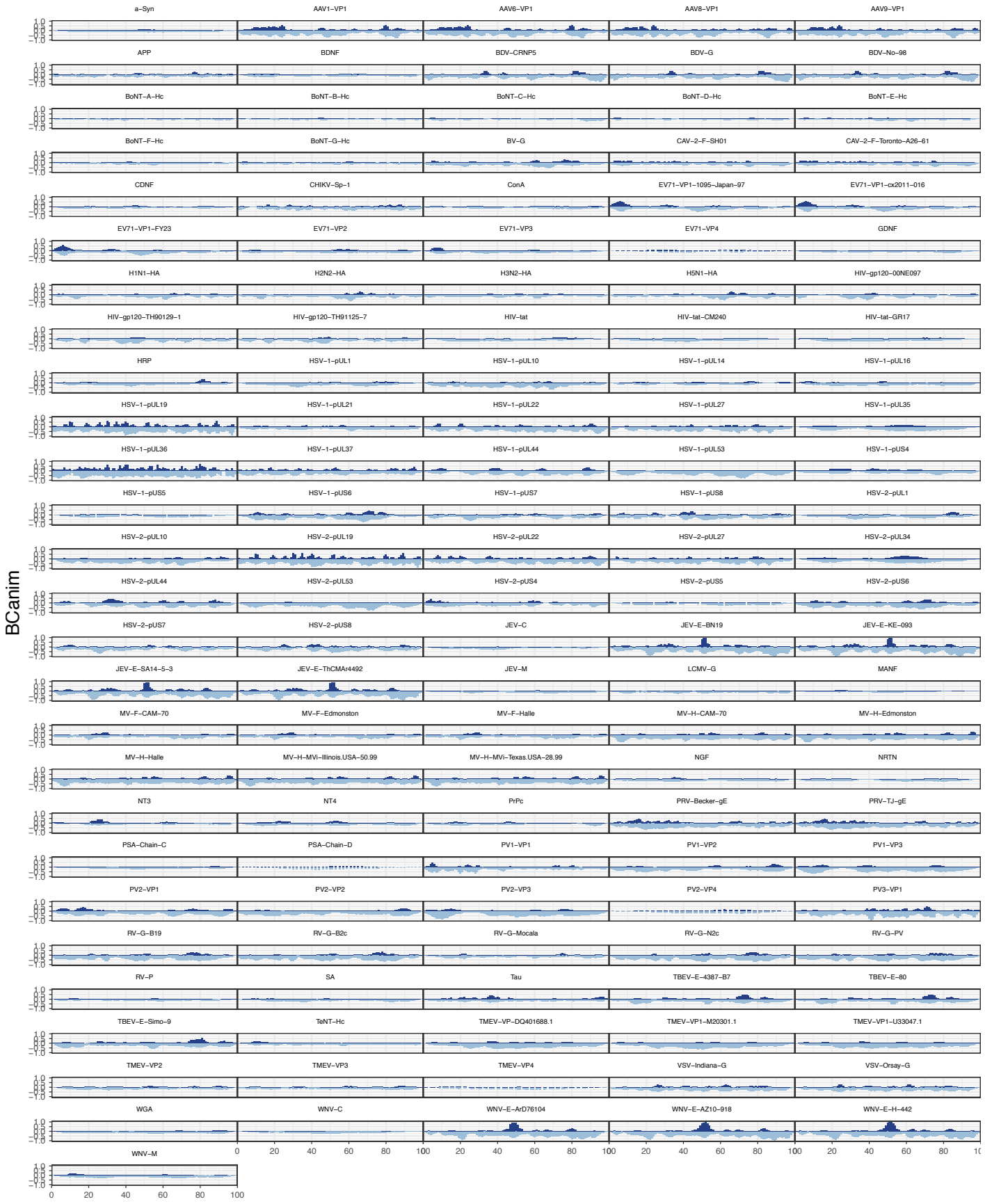

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Trsp

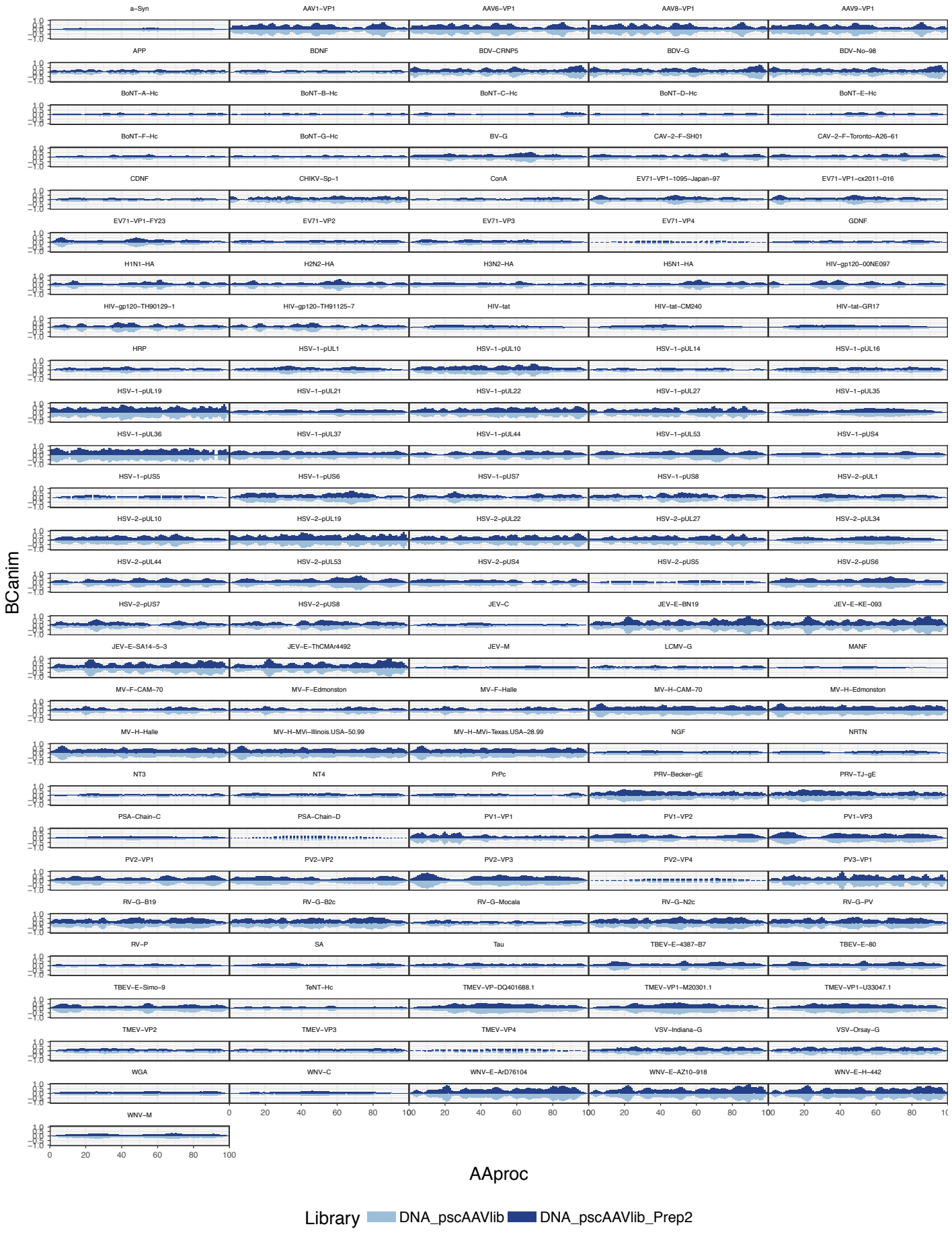

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
NormCount
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Trsp

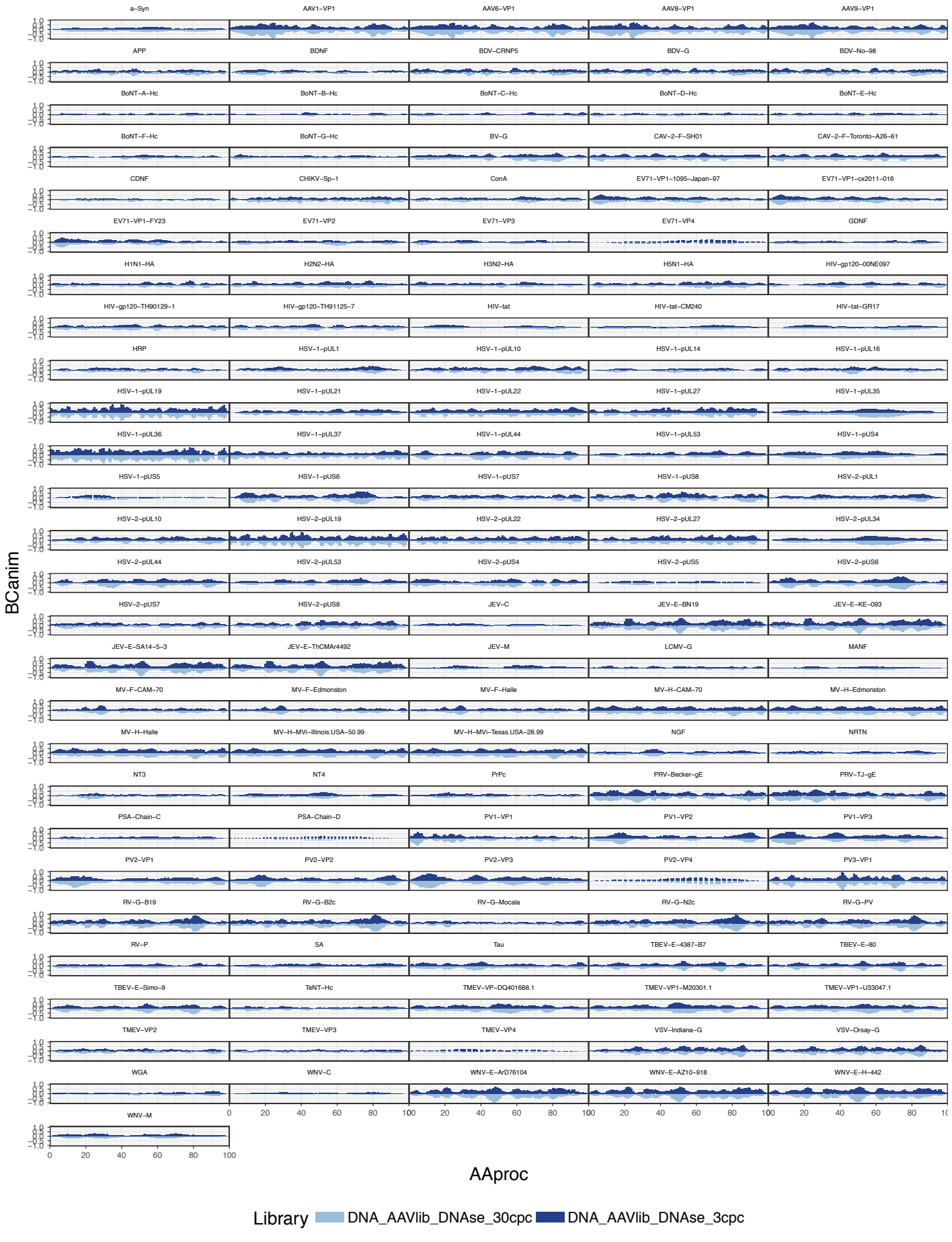

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_Trsp
mRNA_3cpc_Trsp

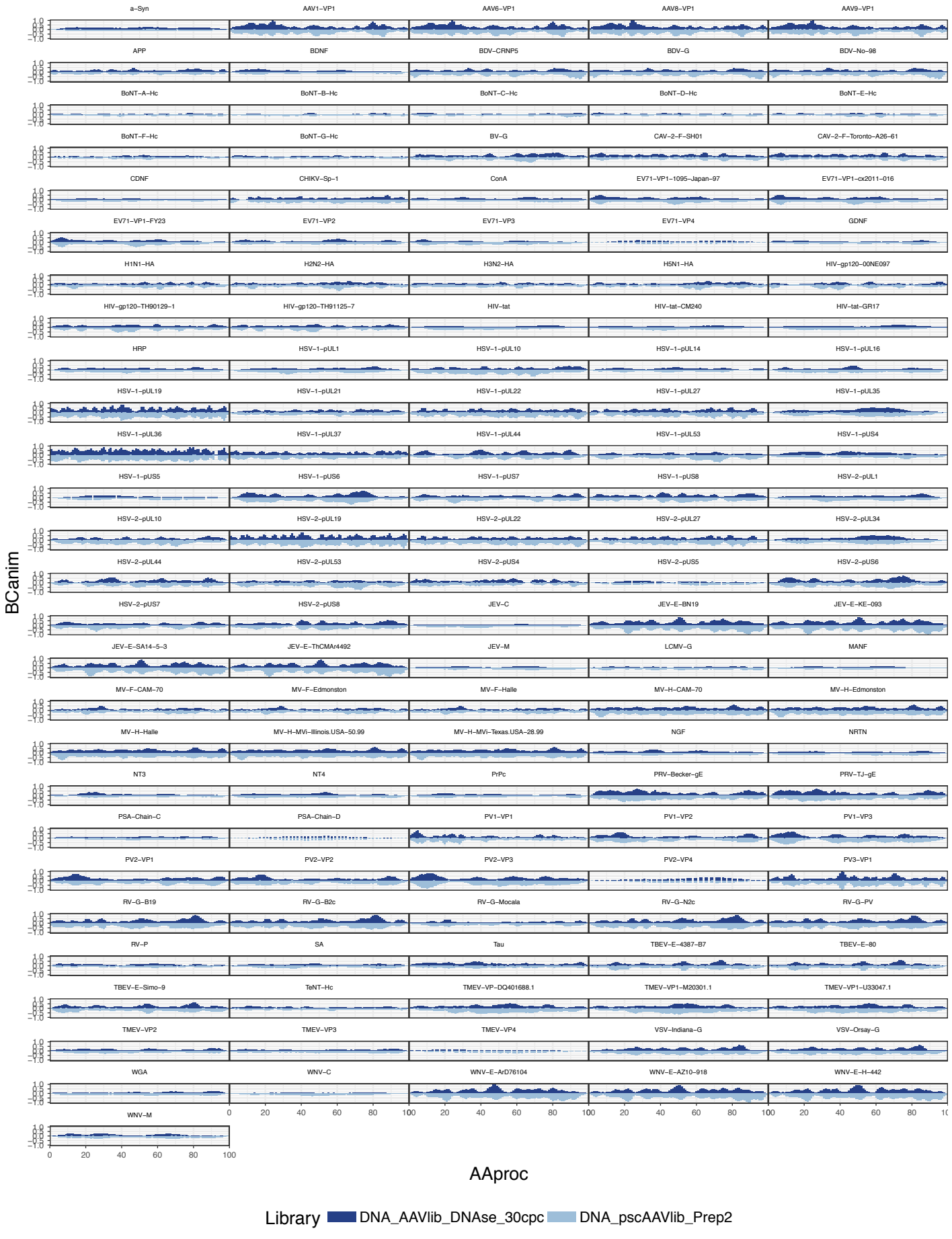

AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
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
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUS4
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−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−C
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
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
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV1−VP3
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
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_pNeuron
mRNA_3cpc_pNeuron

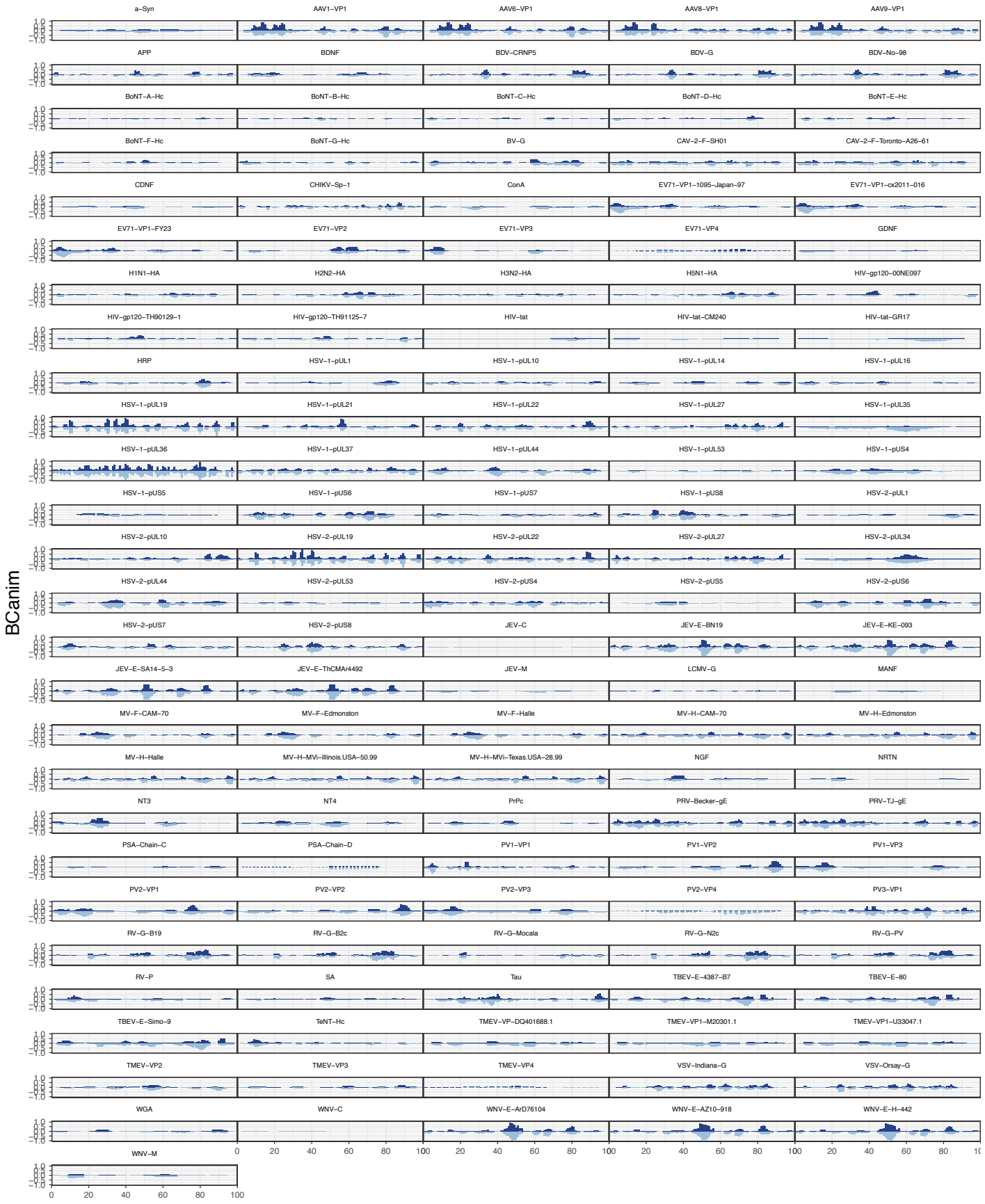

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−F−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HRP
HSV−1−pUL10
HSV−1−pUL16
HSV−1−pUL19
HSV−1−pUL21
HSV−1−pUL22
HSV−1−pUL27
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
HSV−1−pUS4
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−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
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
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV2−VP1
PV2−VP2
PV2−VP4
PV3−VP1
RV−G−B19
RV−G−B2c
RV−G−N2c
RV−G−PV
RV−P
SA
Tau
TeNT−Hc
TMEV−VP−DQ401688.1
TMEV−VP1−M20301.1
TMEV−VP1−U33047.1
TMEV−VP2
VSV−Indiana−G
VSV−Orsay−G
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
NormCount
AAproc
Library
mRNA_30cpc_HEK293T
mRNA_3cpc_HEK293T

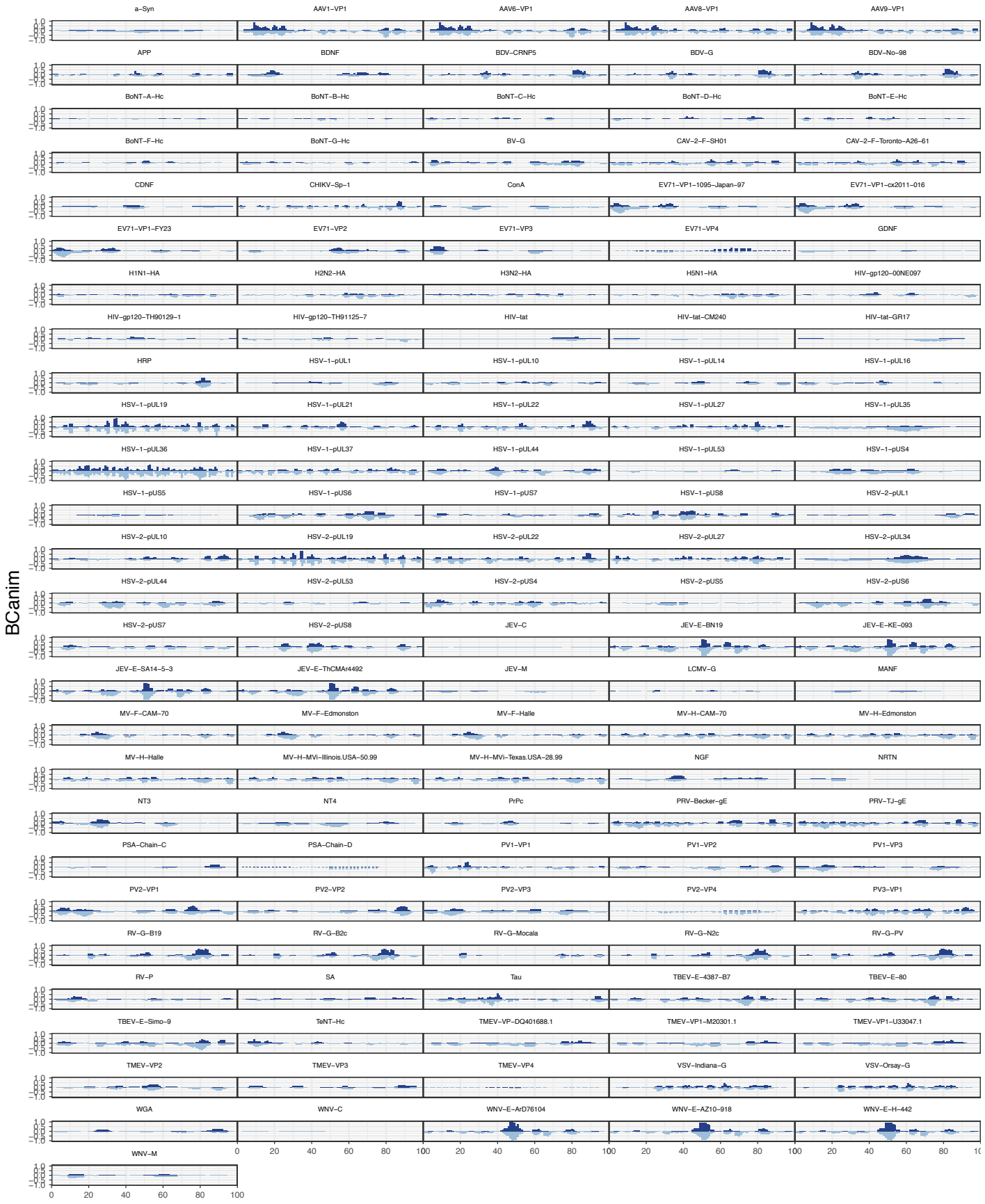

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
DNA_pscAAVlib
mRNA_All

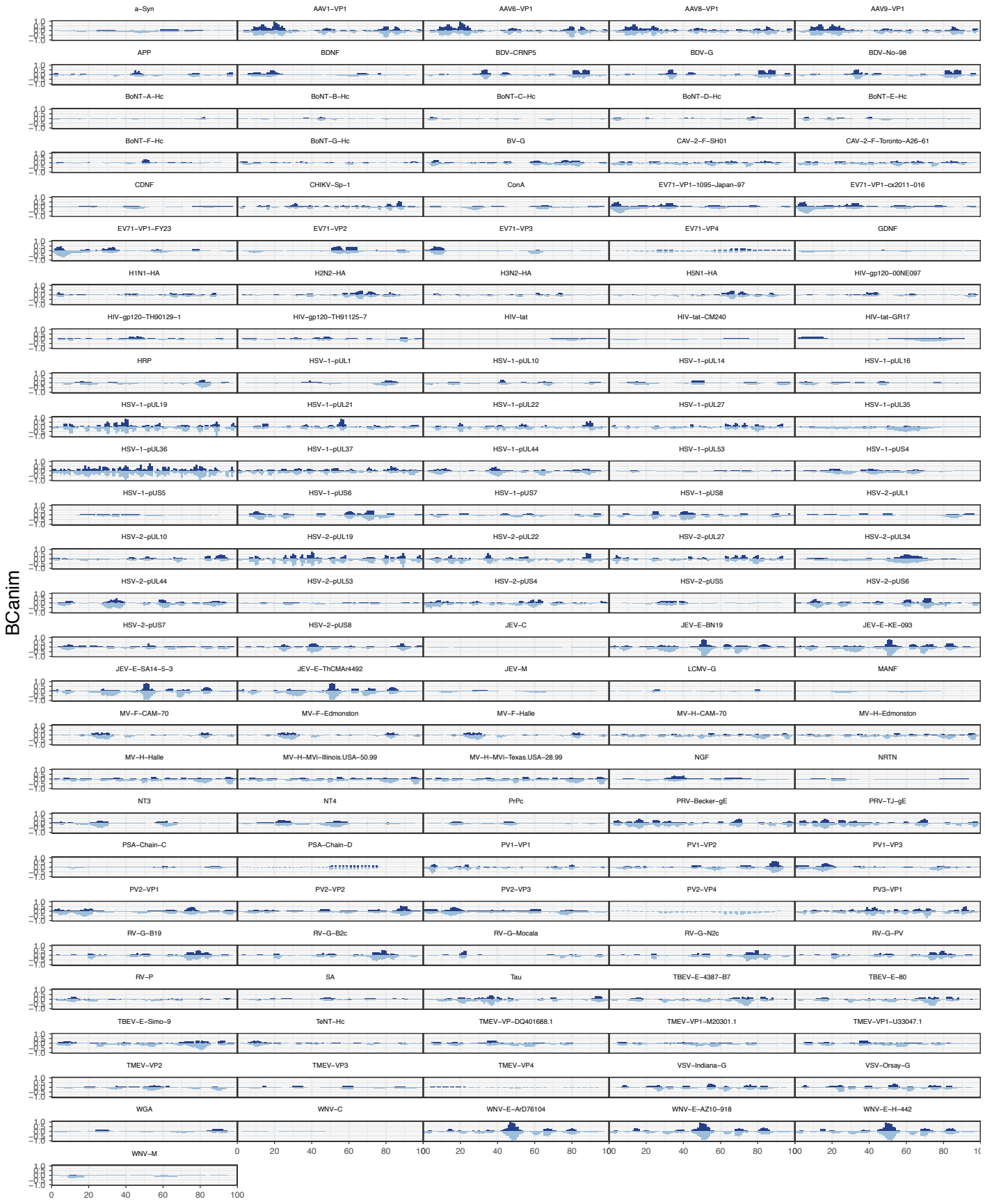

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
DNA_pscAAVlib
DNA_pscAAVlib_Prep2

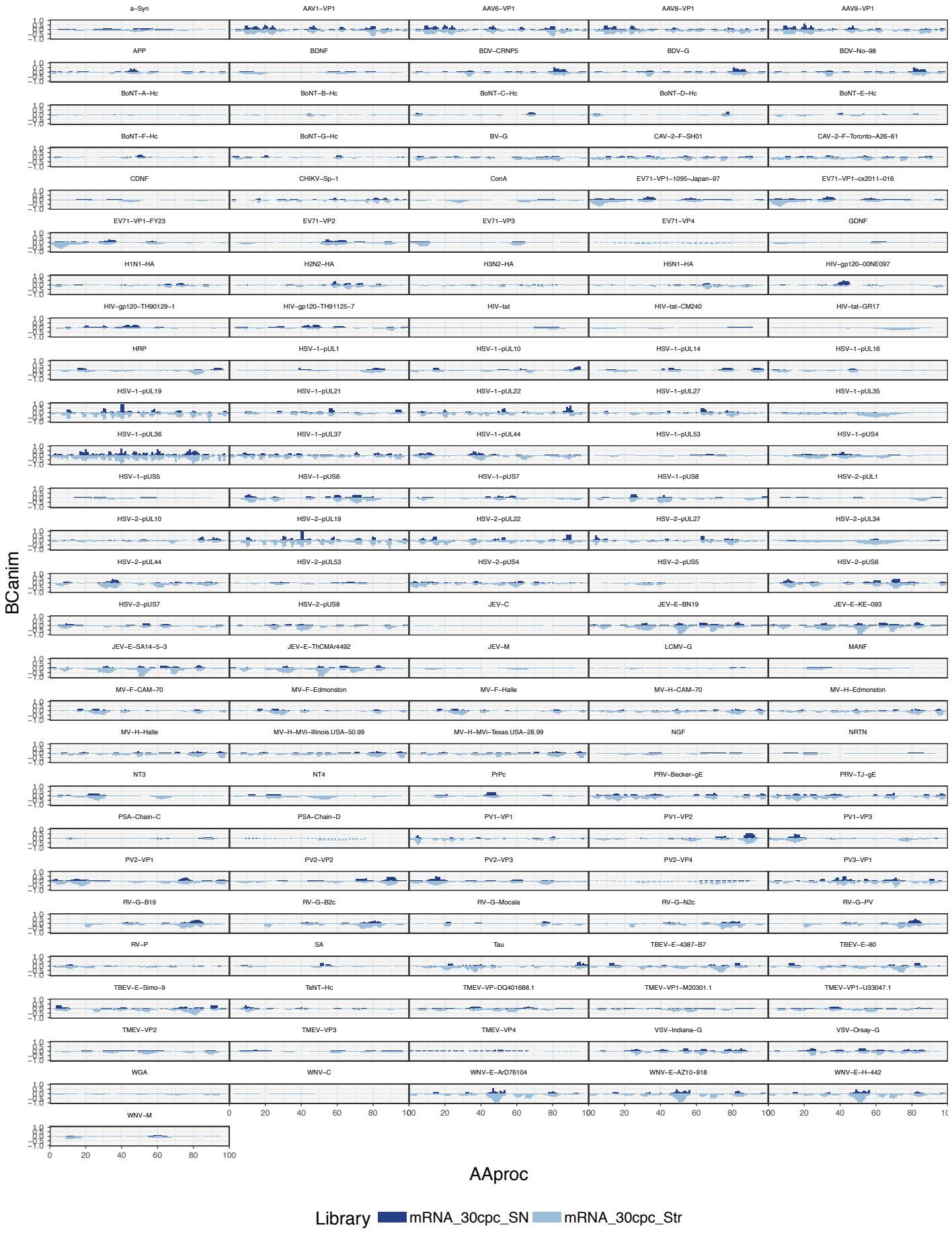

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
DNA_AAVlib_DNAse_30cpc
DNA_AAVlib_DNAse_3cpc

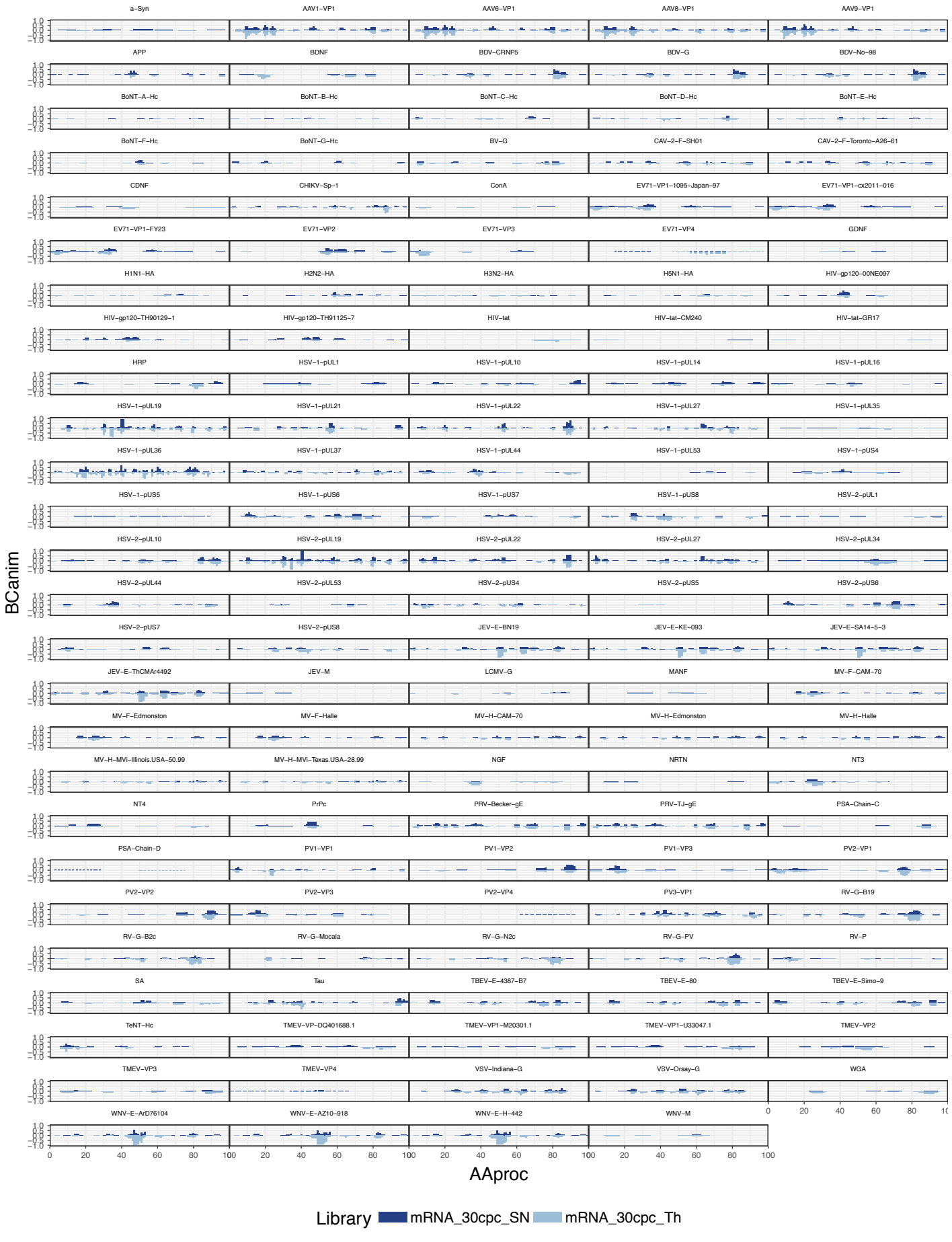

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
DNA_AAVlib_DNAse_30cpc
DNA_pscAAVlib_Prep2

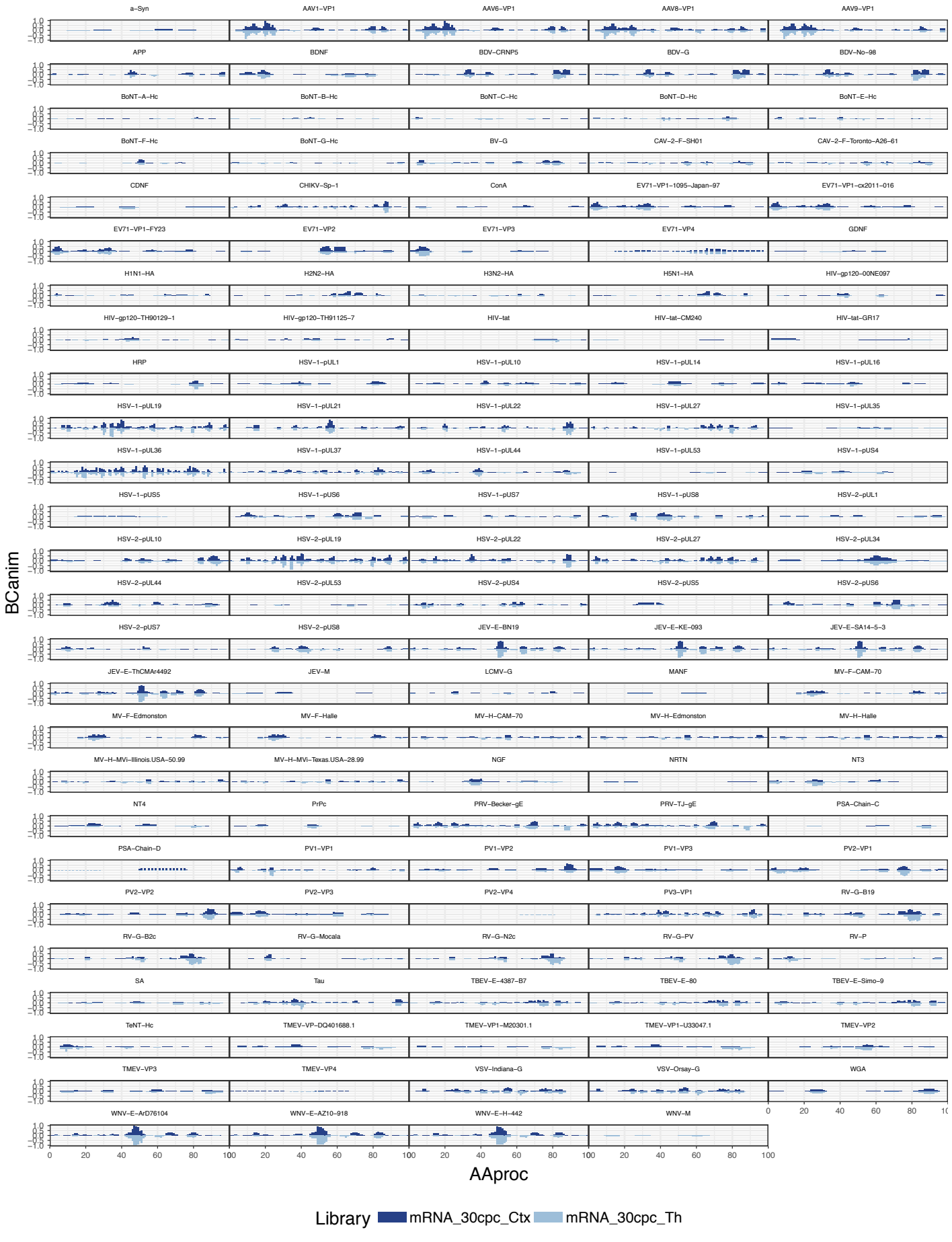

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Trsp

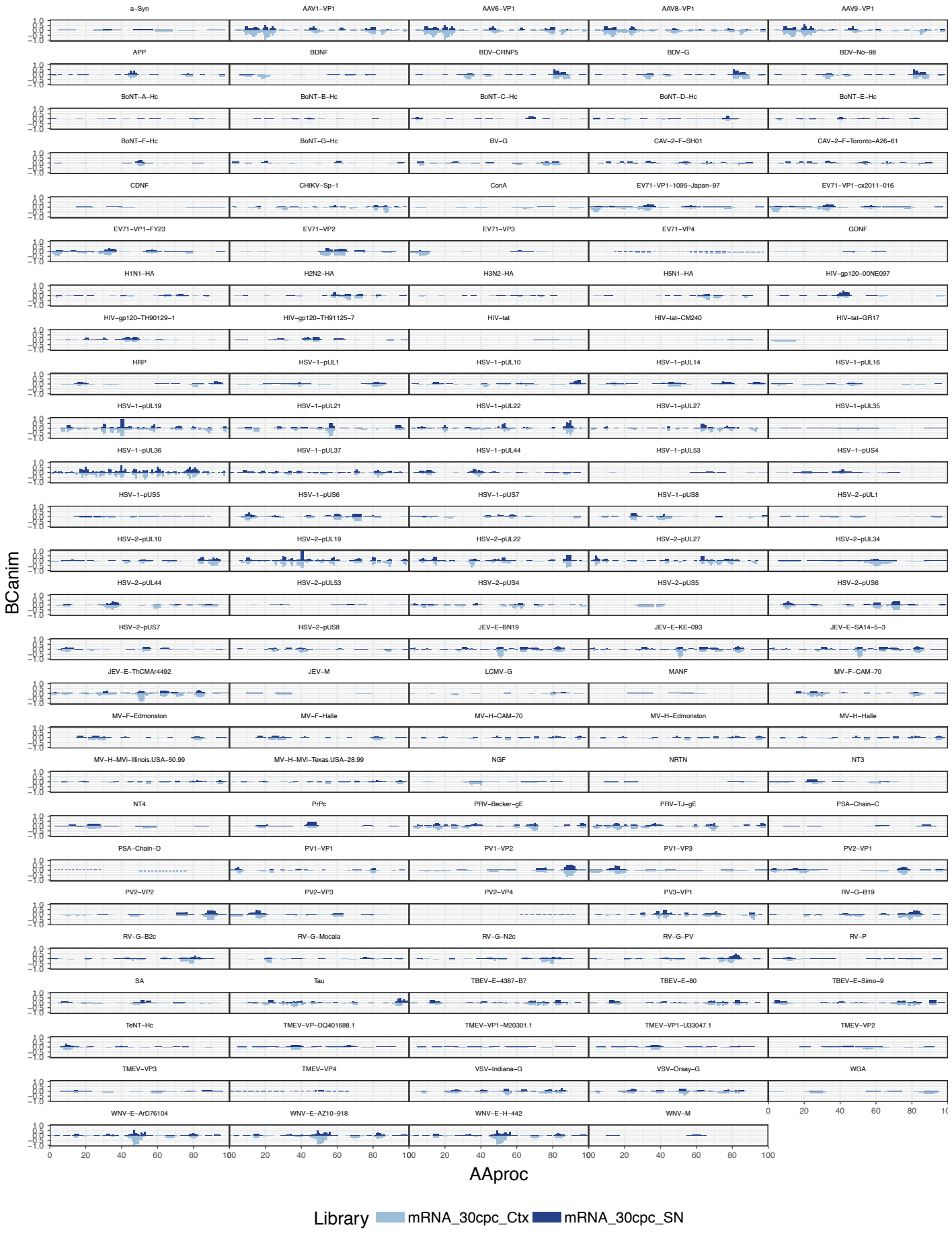

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Th

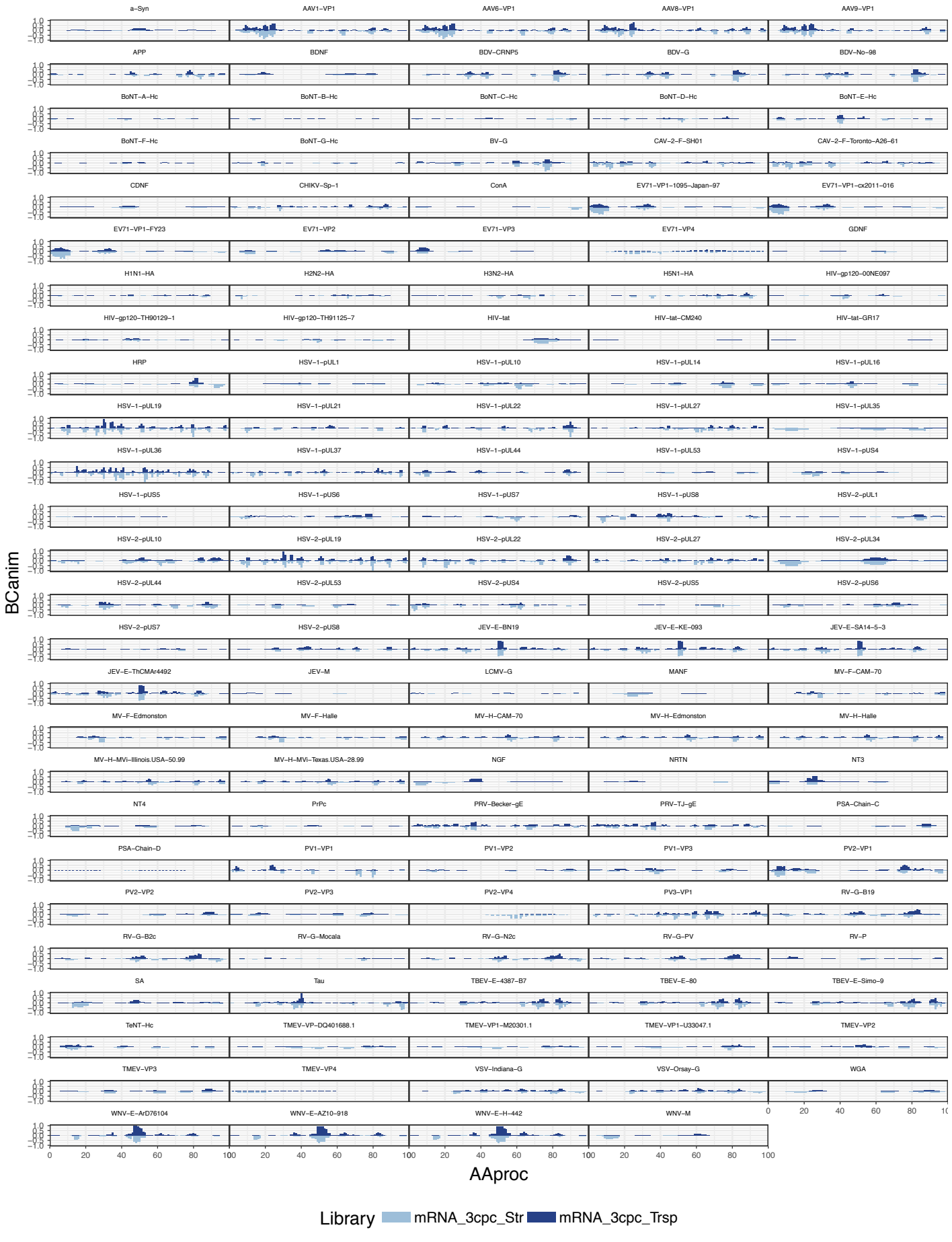

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
EV71–VP1–FY23
EV71–VP2
EV71–VP3
EV71–VP4
GDNF
H1N1–HA
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
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
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
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
WGA
WNV–C
WNV–E–ArD76104
WNV–E–AZ10–918
WNV–E–H–442
WNV–M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Str

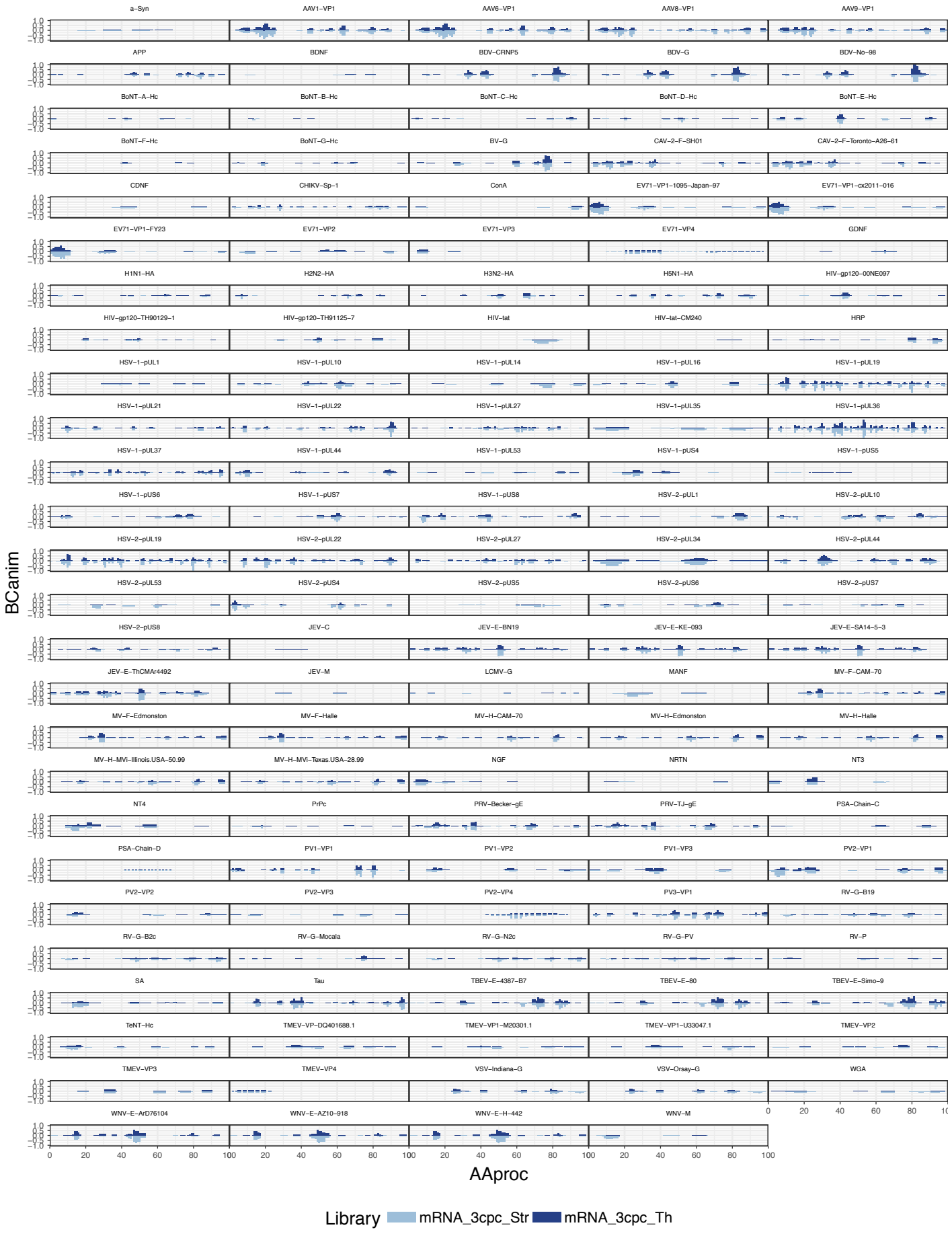

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_SN
mRNA_30cpc_Str

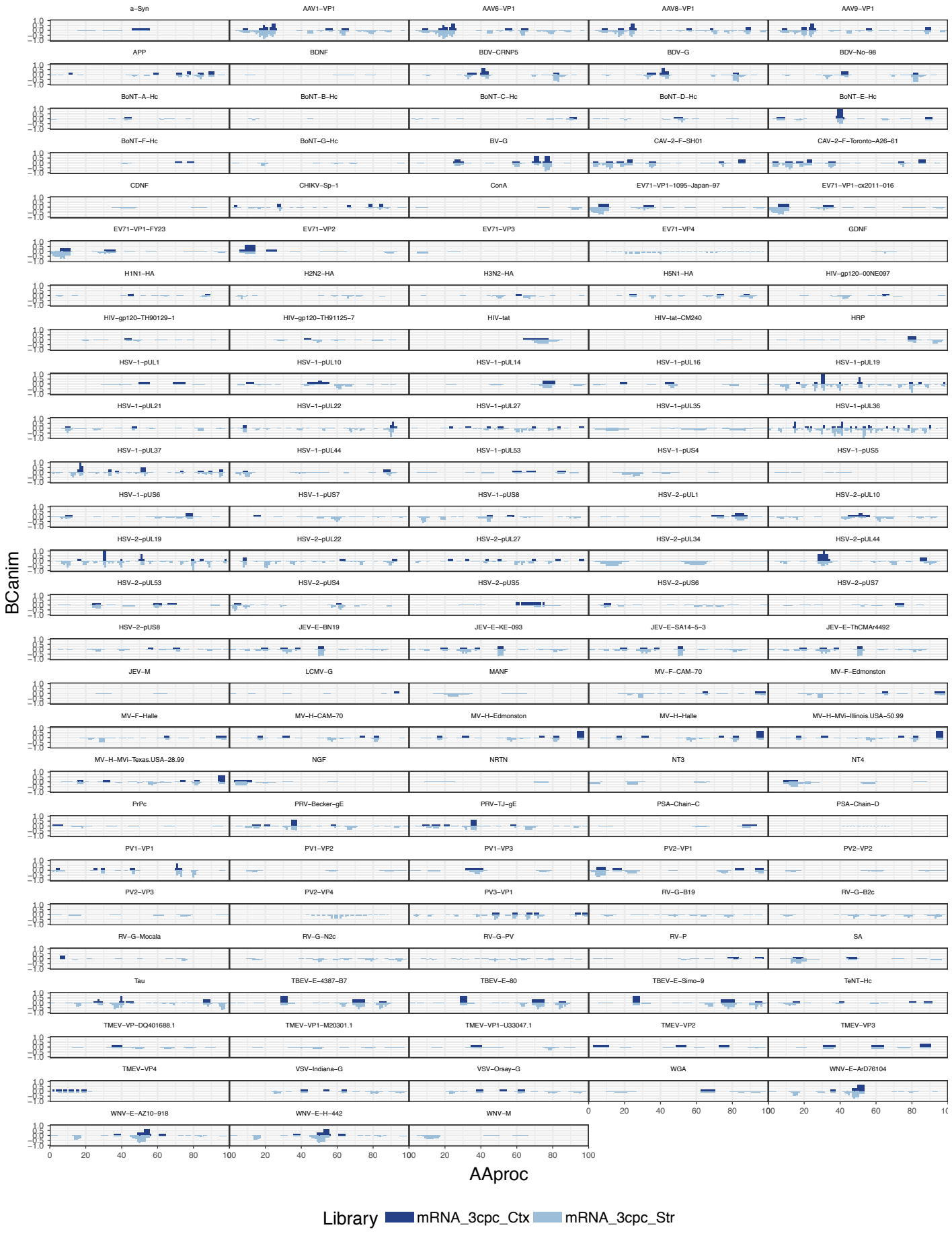

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_SN
mRNA_30cpc_Th

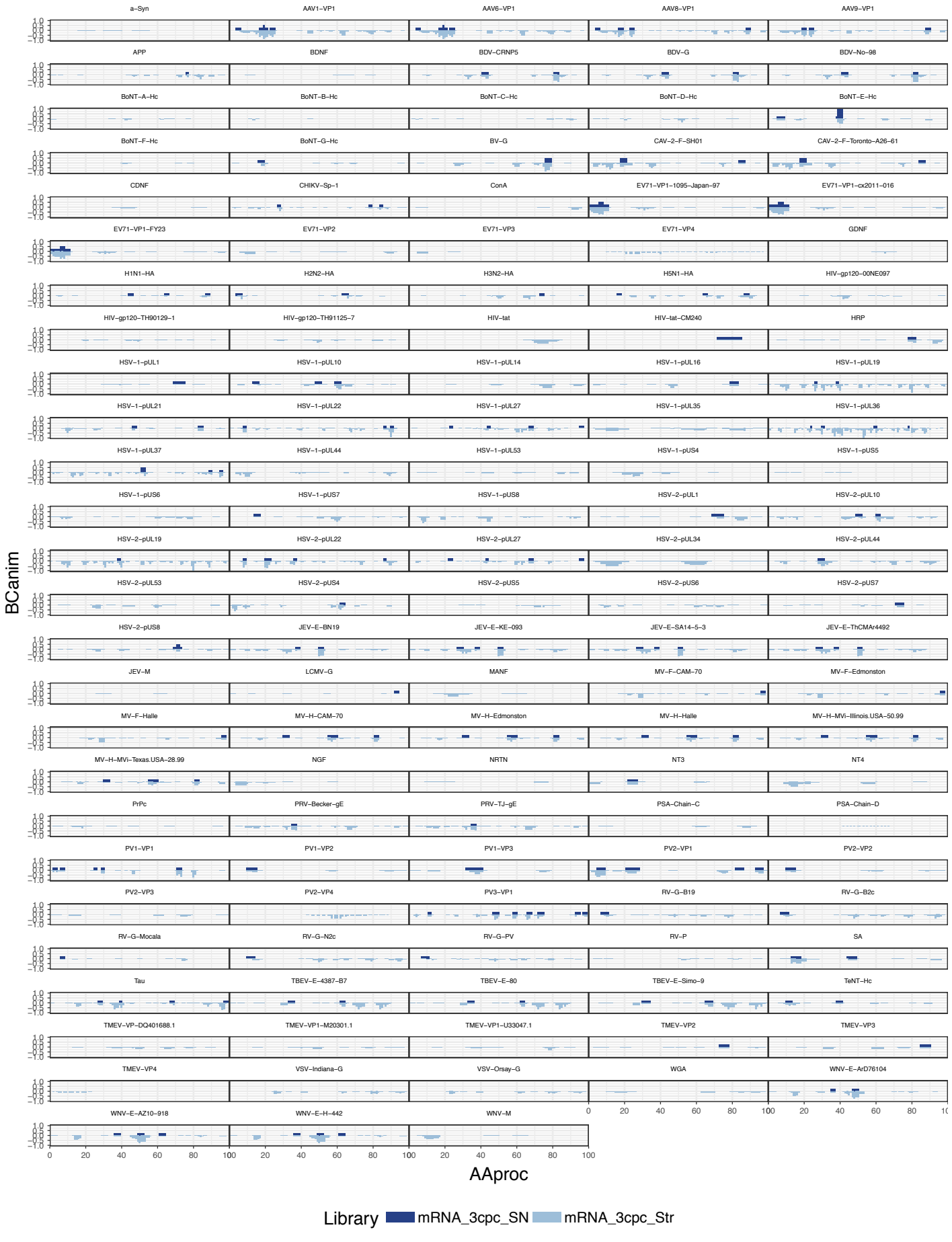

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Th

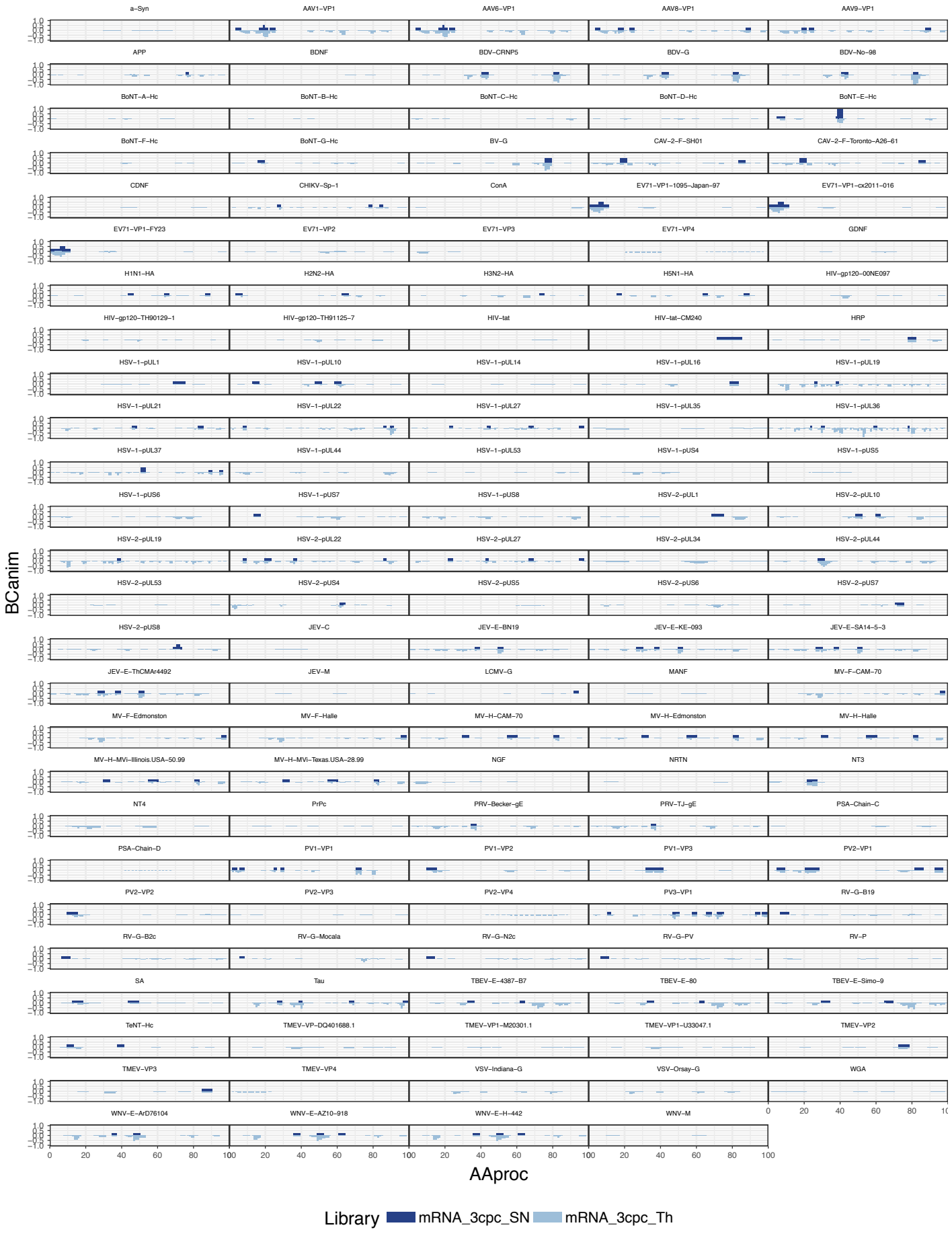

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN

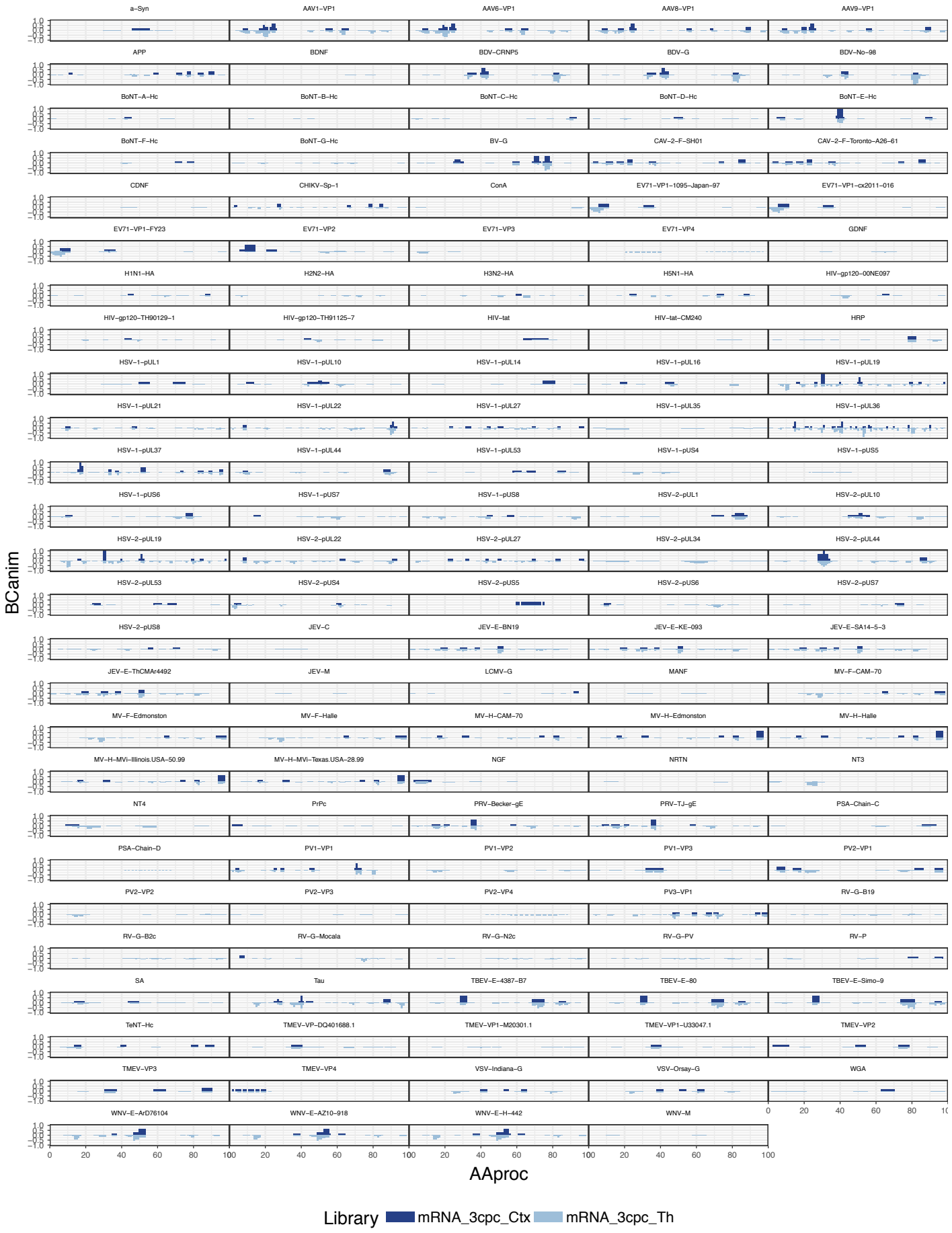

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Trsp

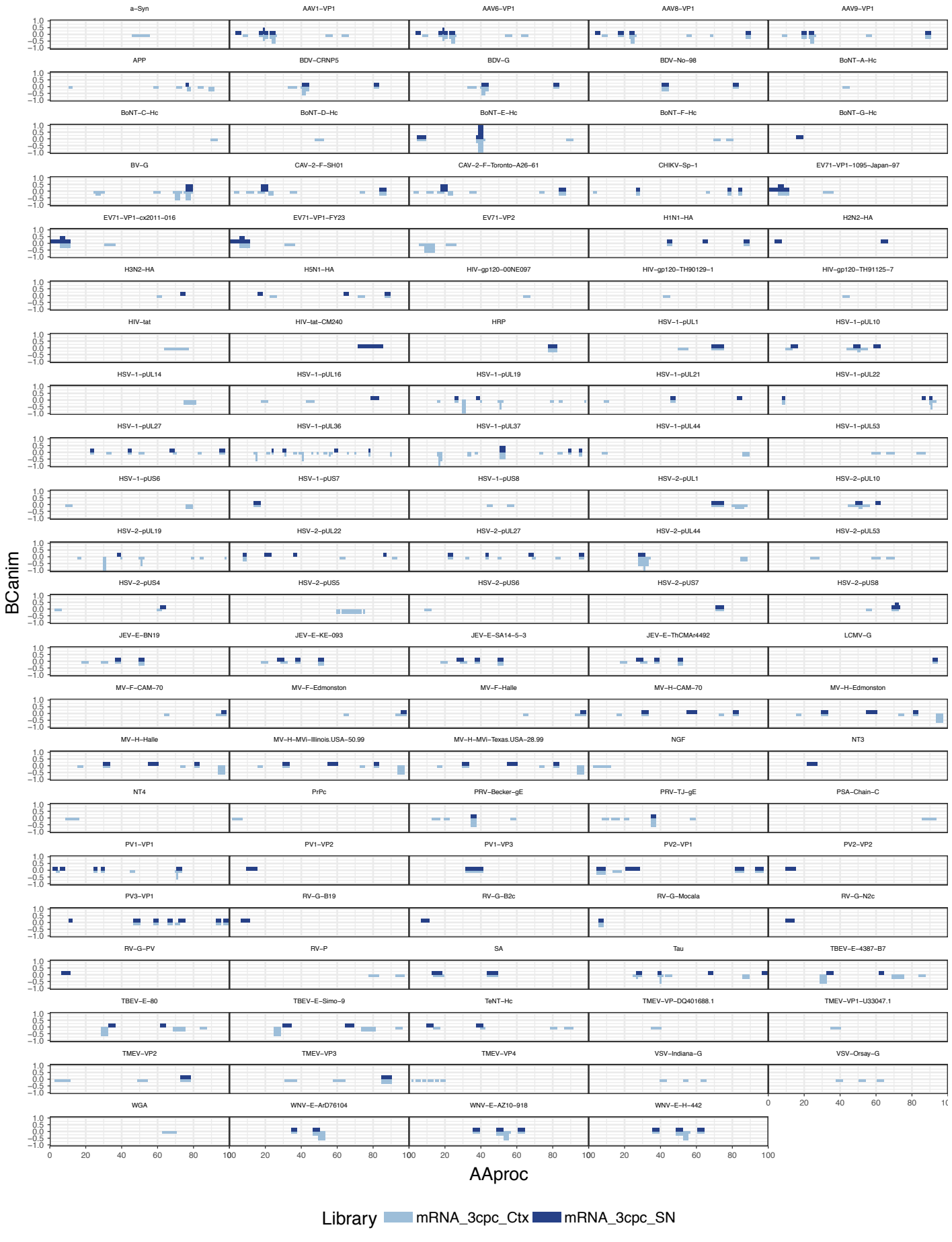

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Th

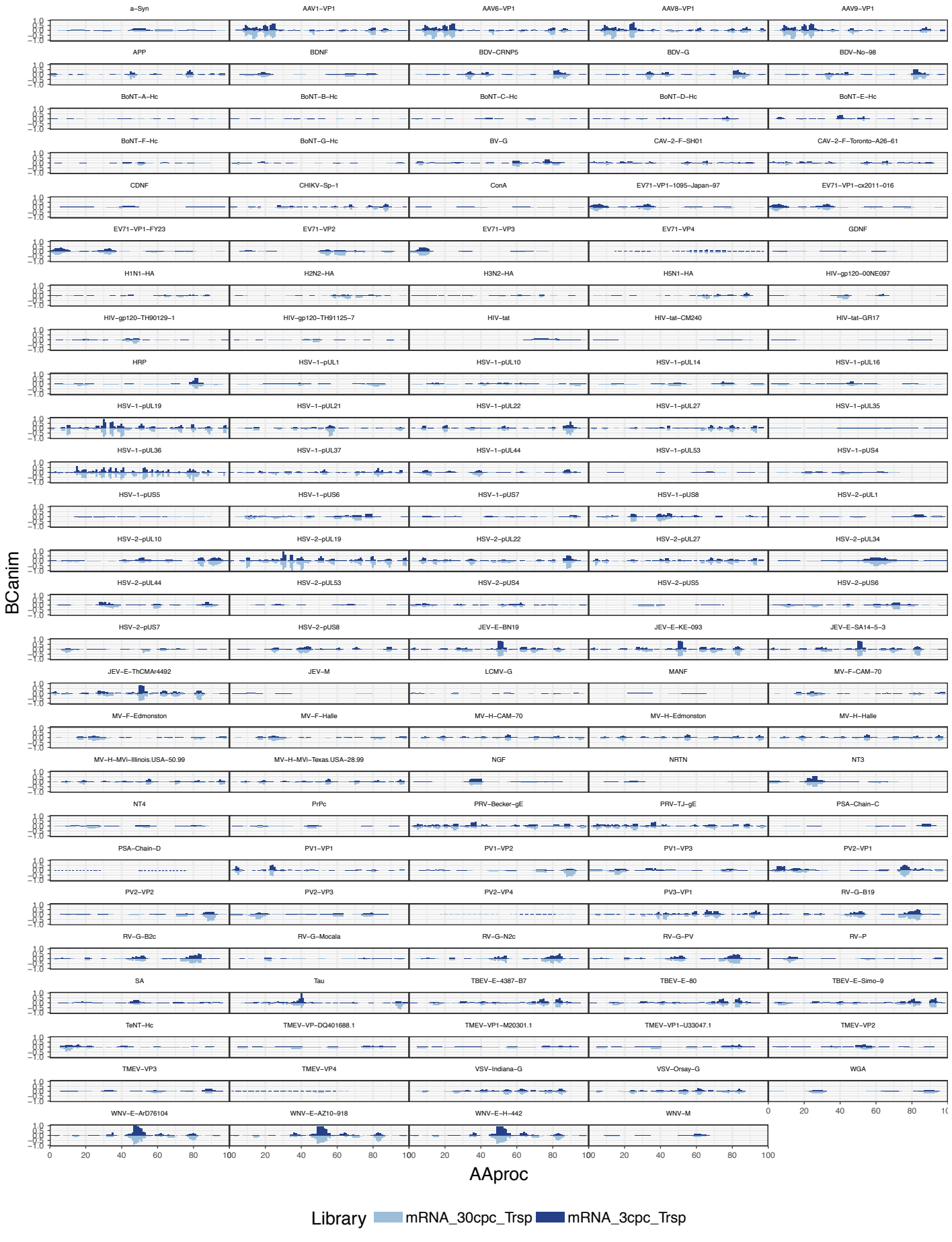

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Str

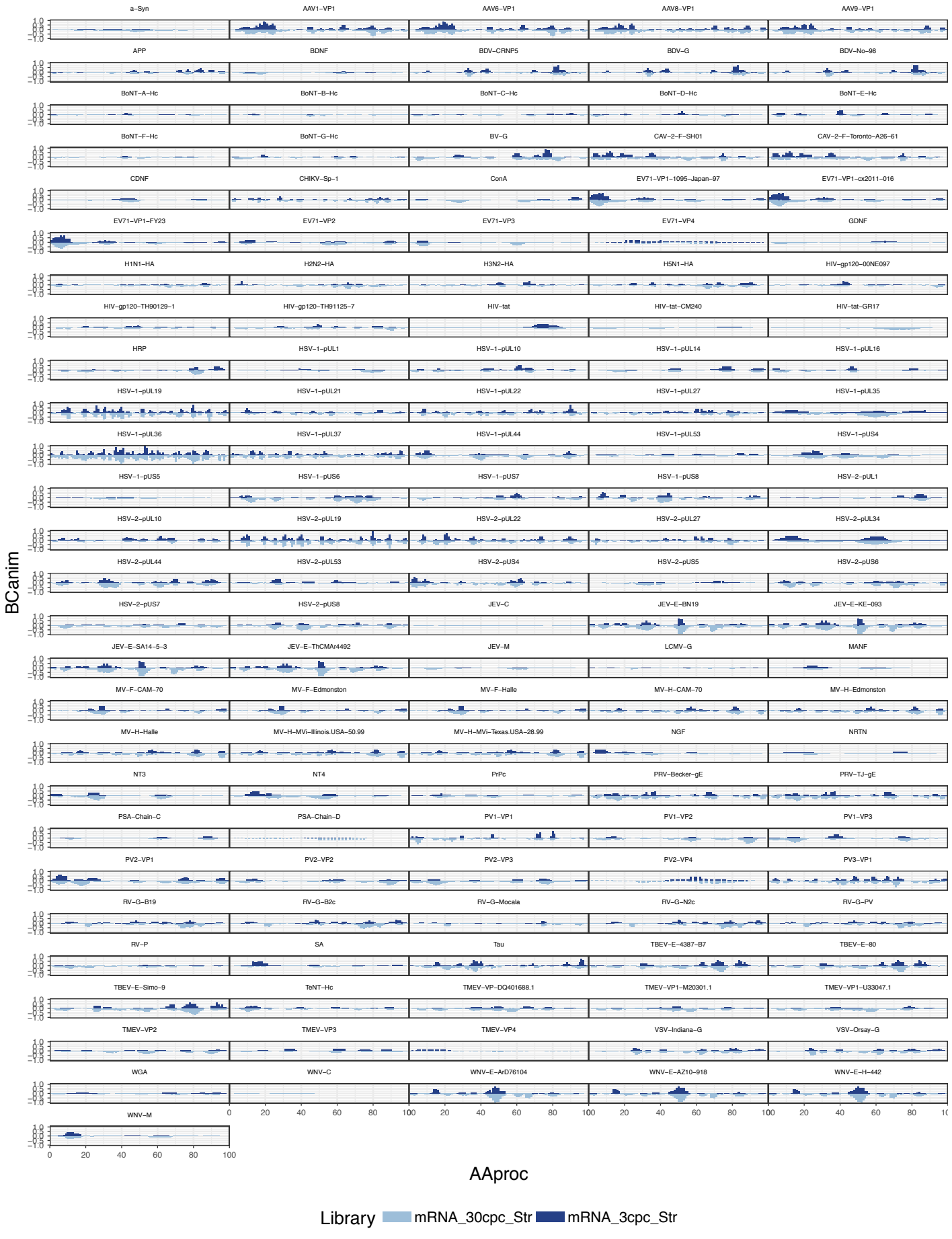

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_SN
mRNA_3cpc_Str

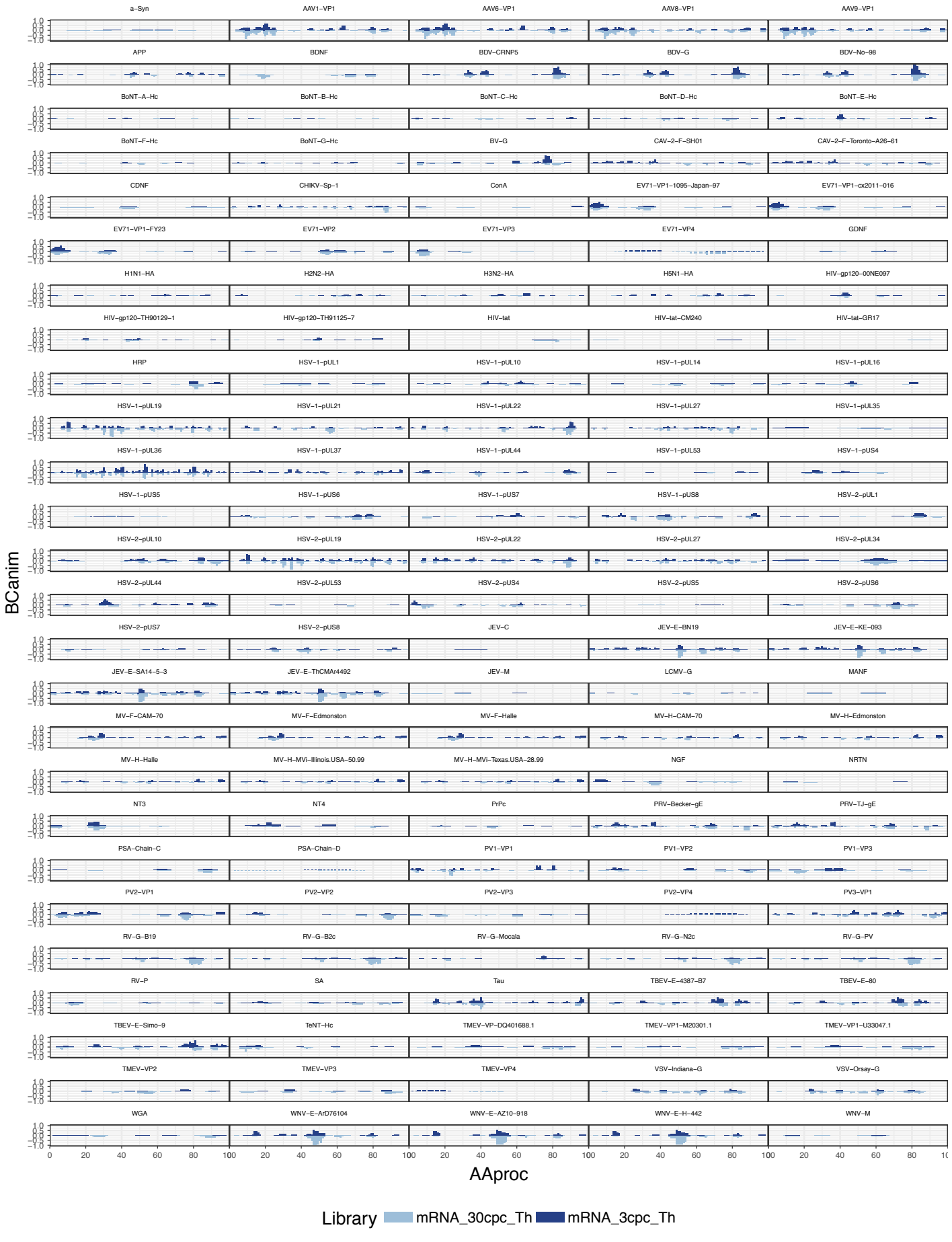

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_SN
mRNA_3cpc_Th

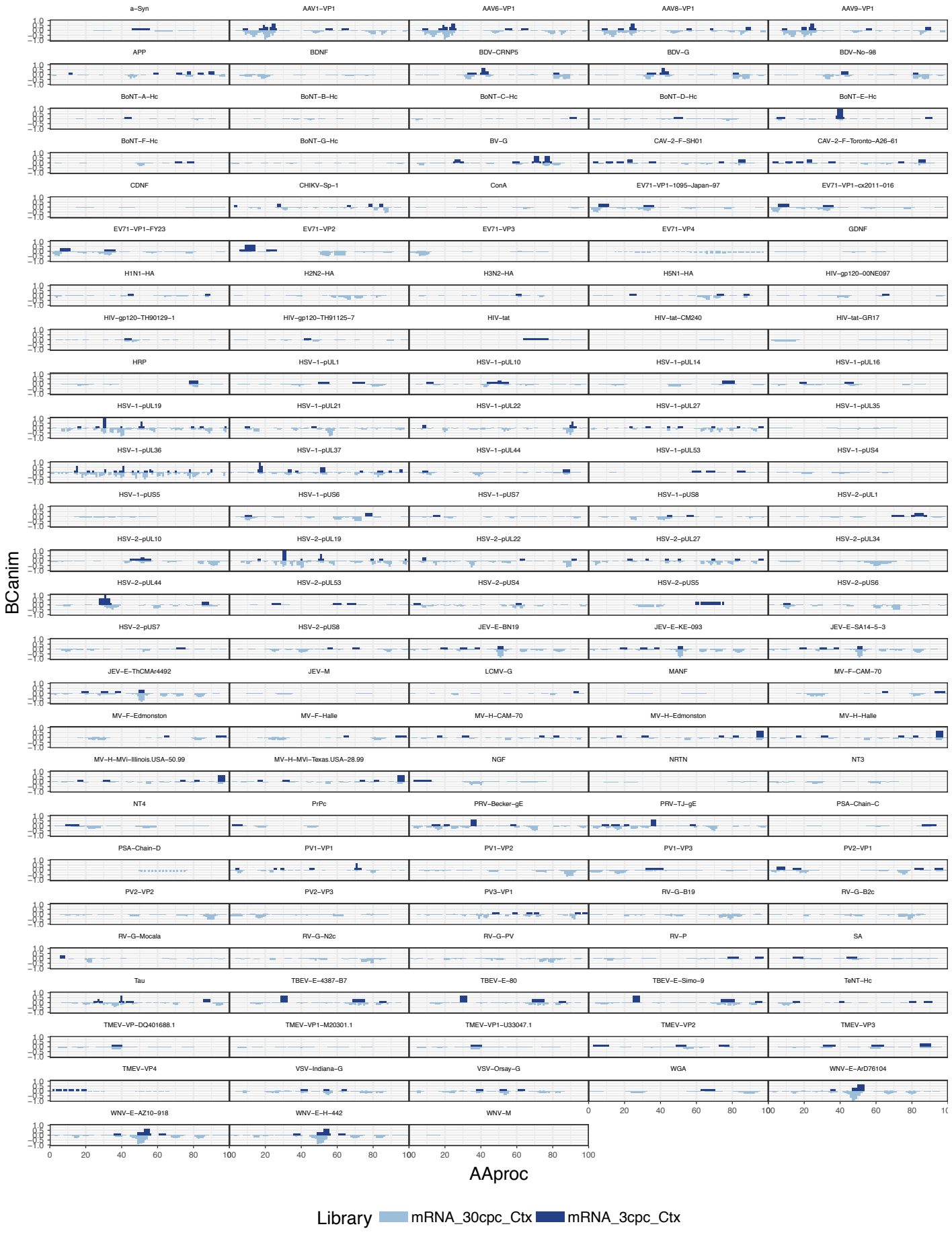

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Th

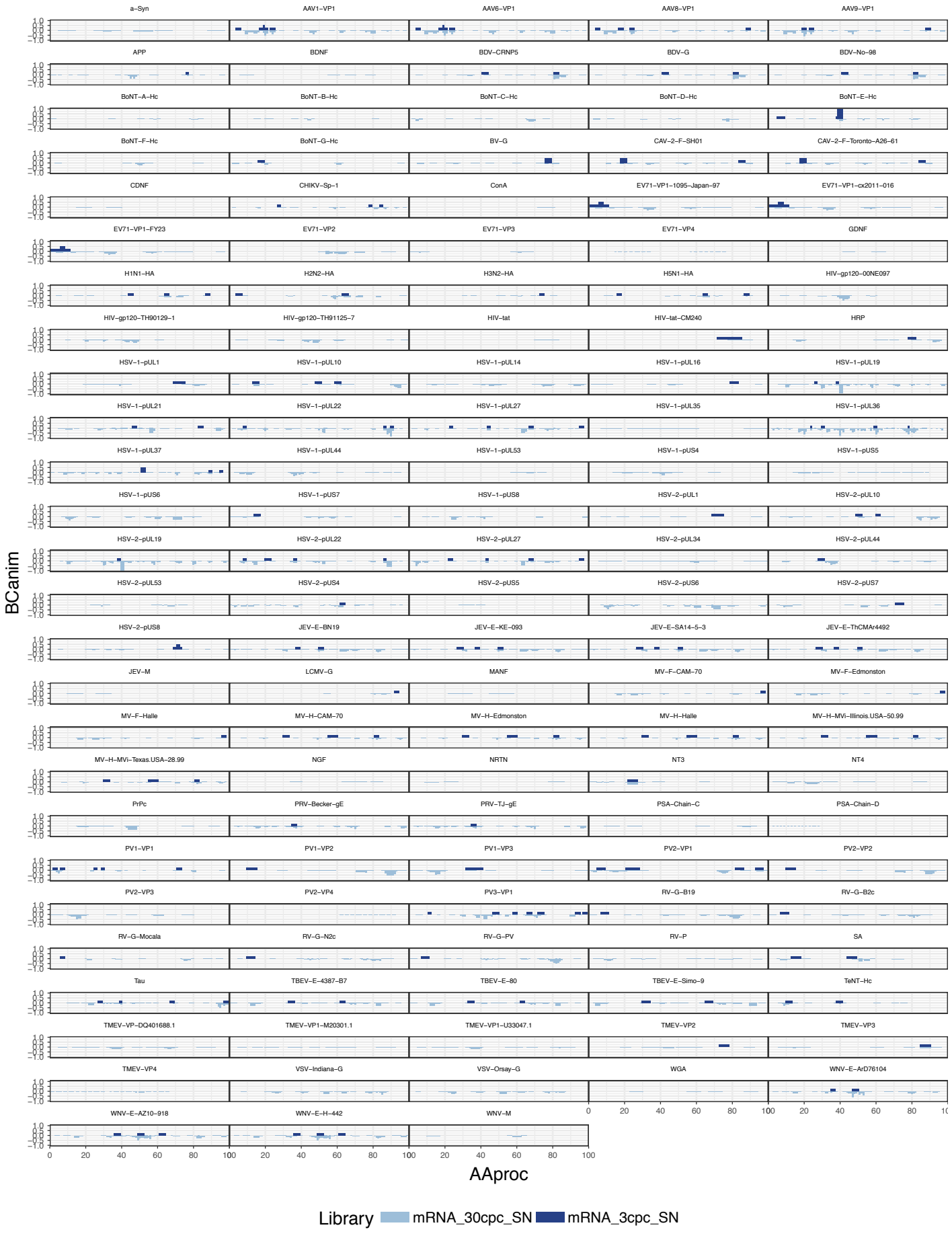

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
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−pUL44
HSV−2−pUL53
HSV−2−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
LCMV−G
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
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PSA−Chain−C
PV1−VP1
PV1−VP2
PV1−VP3
PV2−VP1
PV2−VP2
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−U33047.1
TMEV−VP2
TMEV−VP3
TMEV−VP4
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_SN

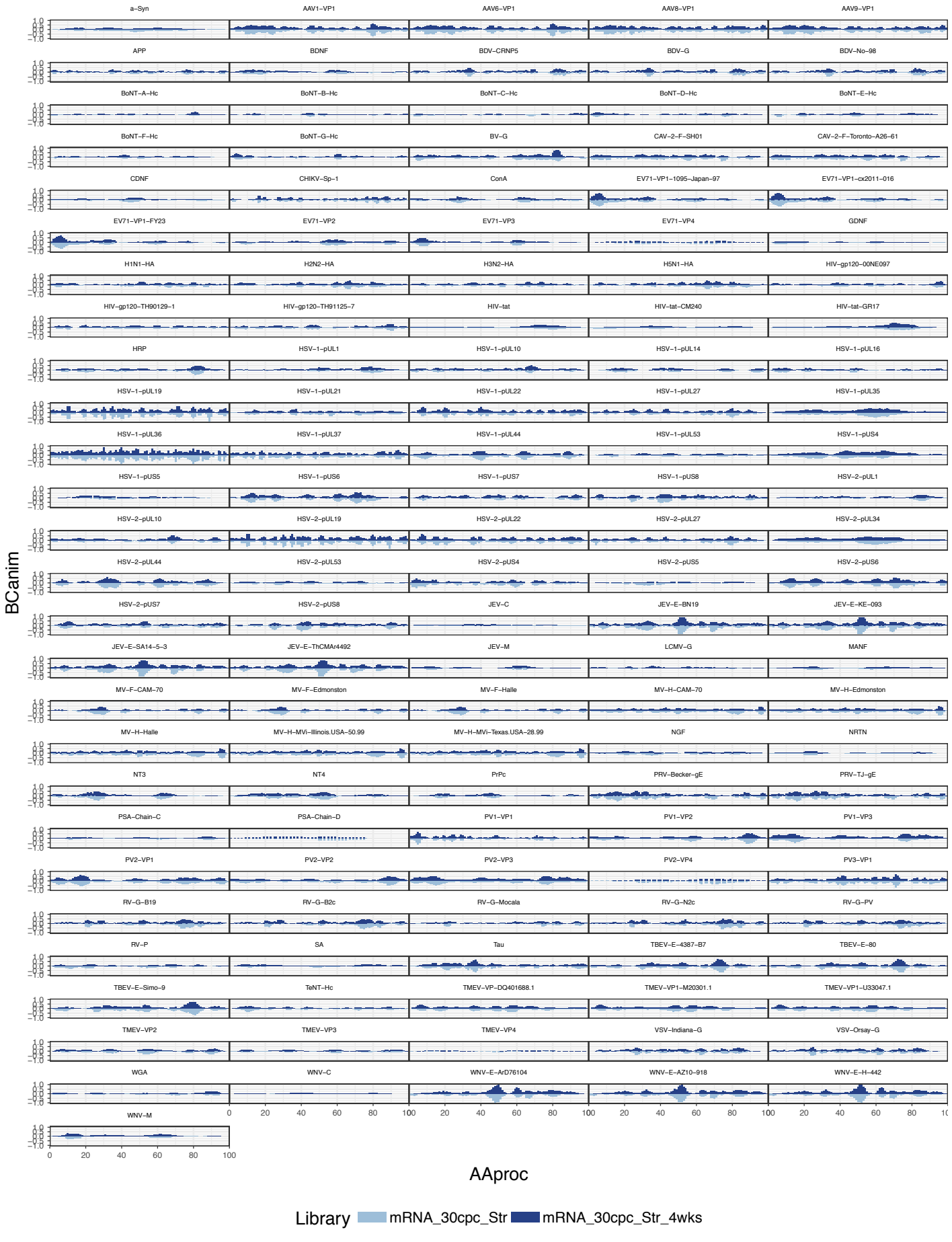

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Trsp
mRNA_3cpc_Trsp

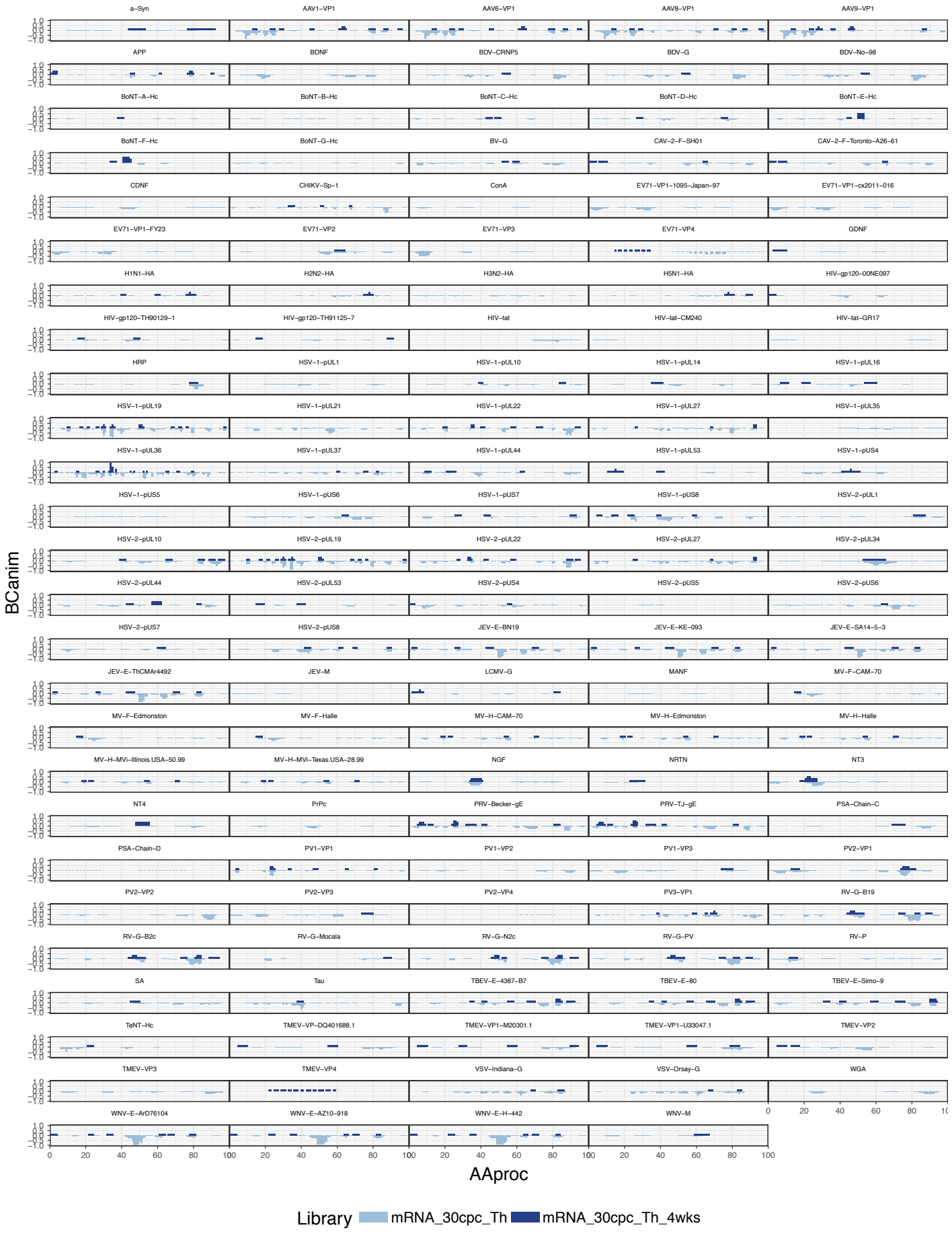

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Str
mRNA_3cpc_Str

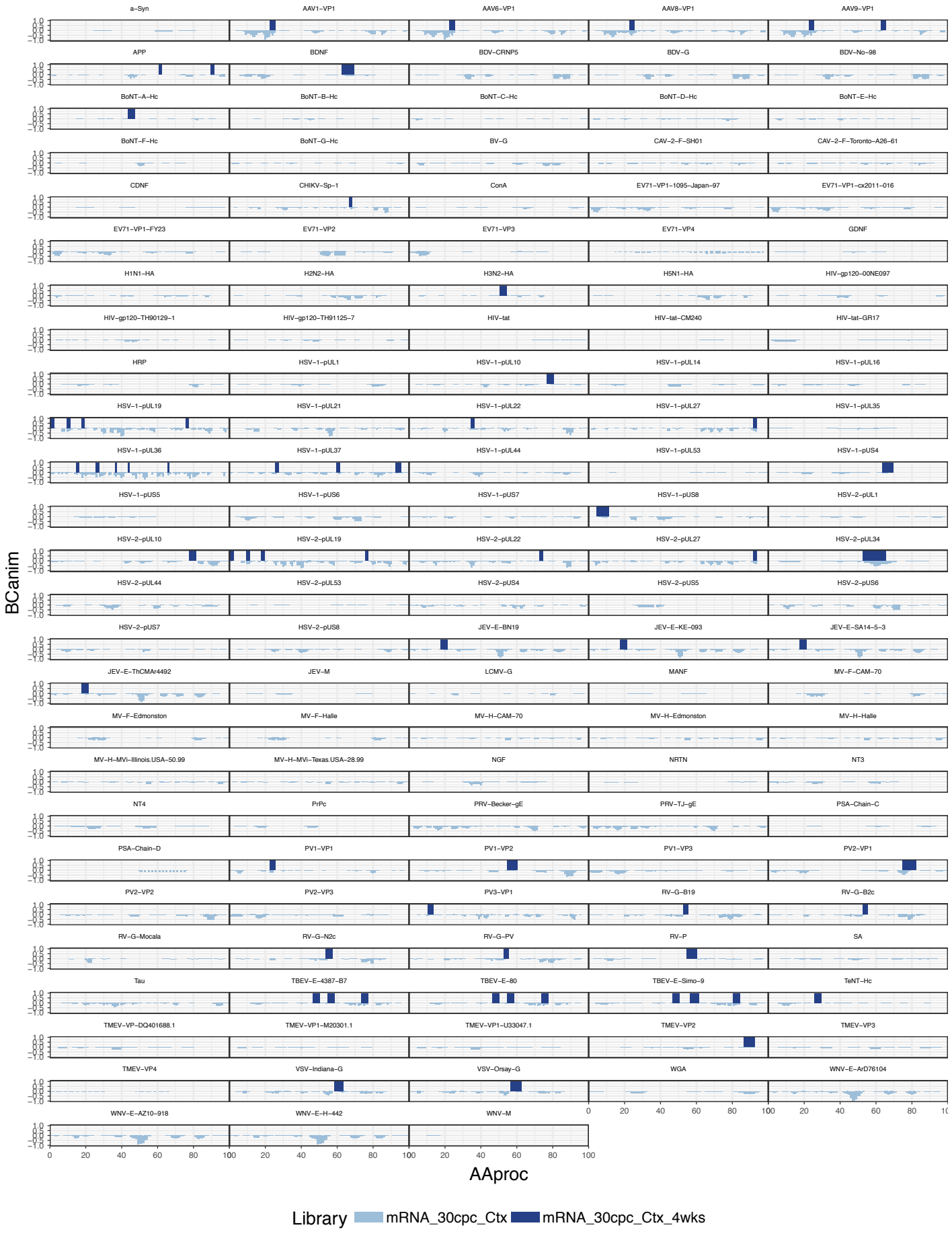

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Th
mRNA_3cpc_Th

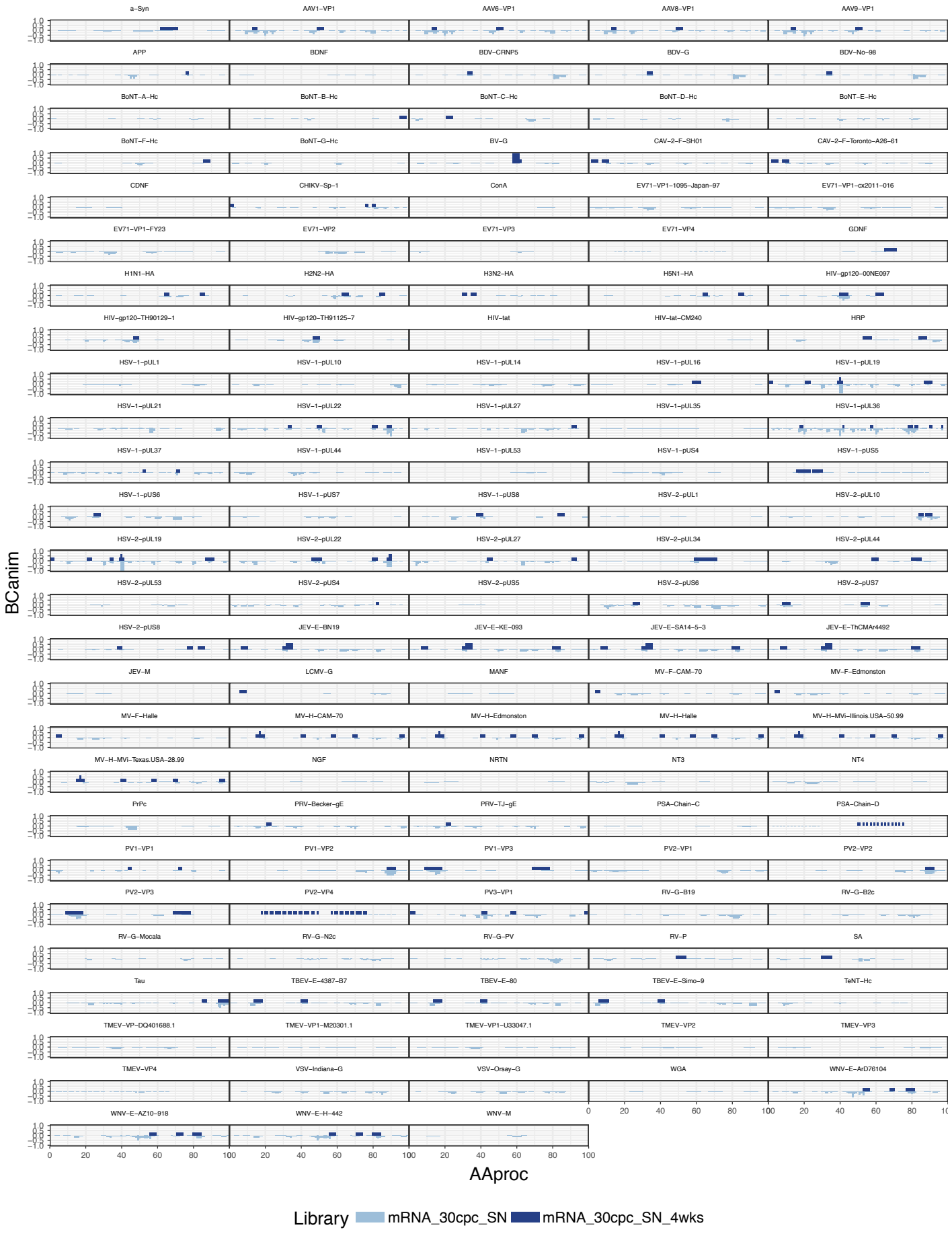

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
PV2−VP1
PV2−VP2
PV2−VP3
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Ctx
mRNA_3cpc_Ctx

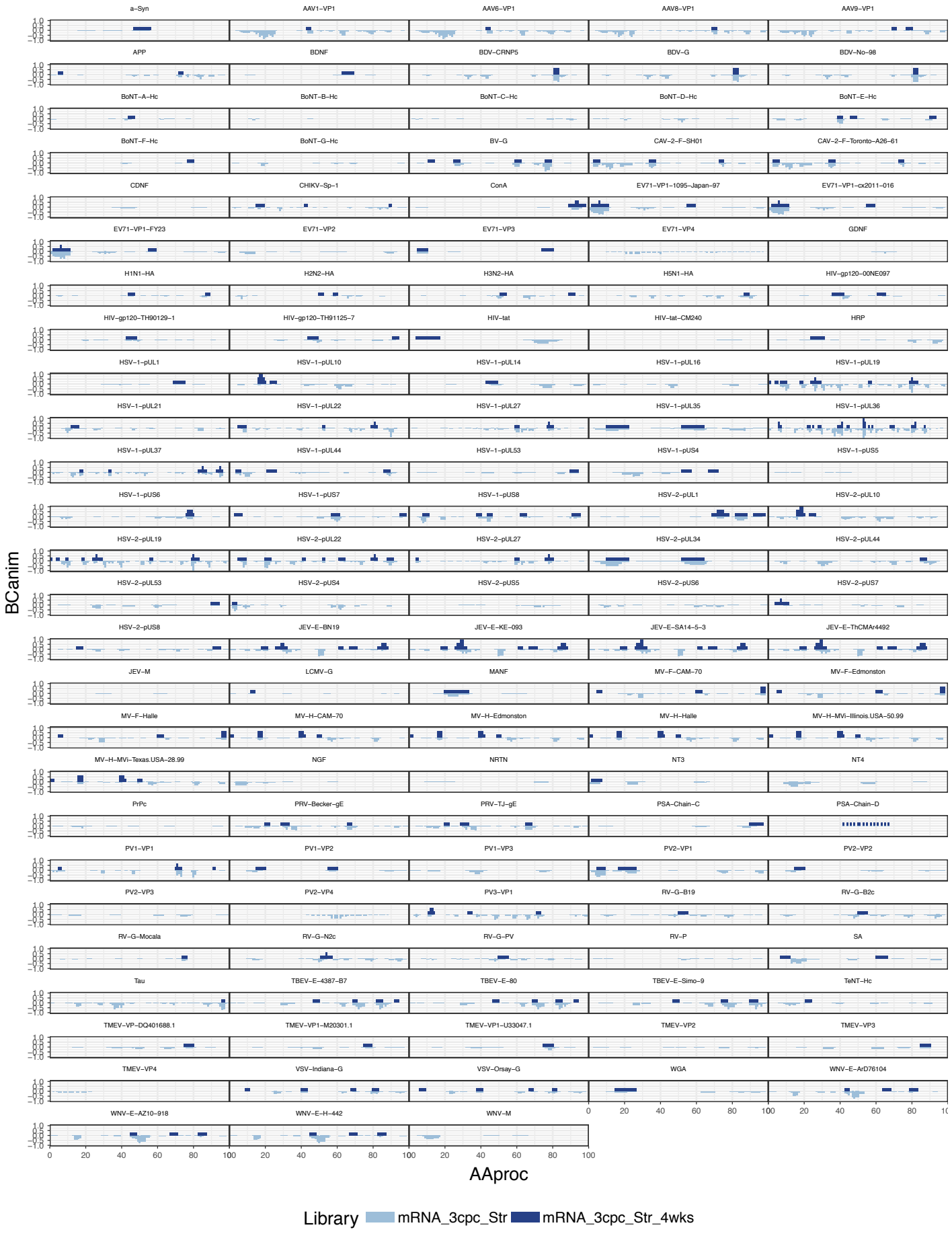

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_SN
mRNA_3cpc_SN

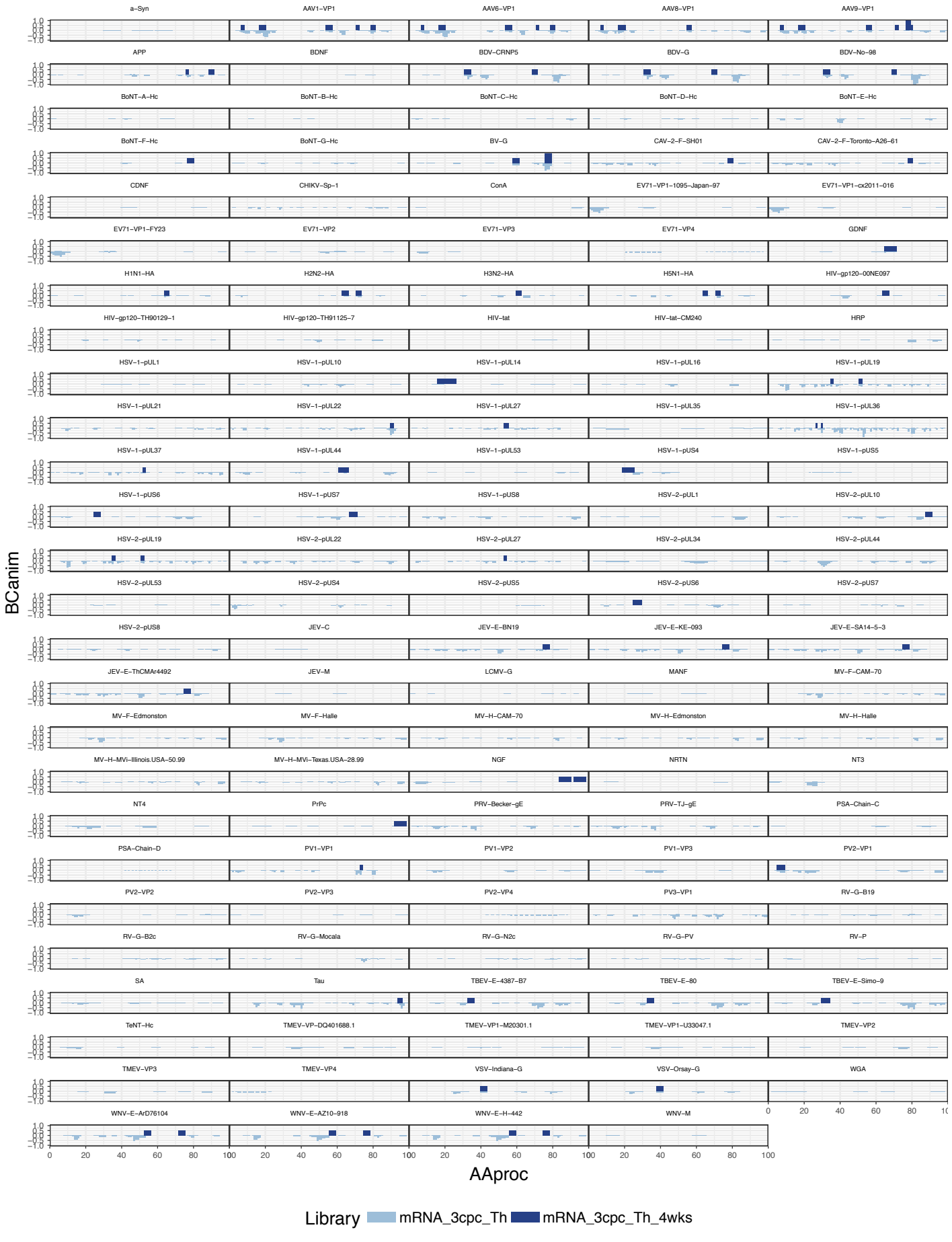

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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
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
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
WGA
WNV−C
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Str
mRNA_30cpc_Str_4wks

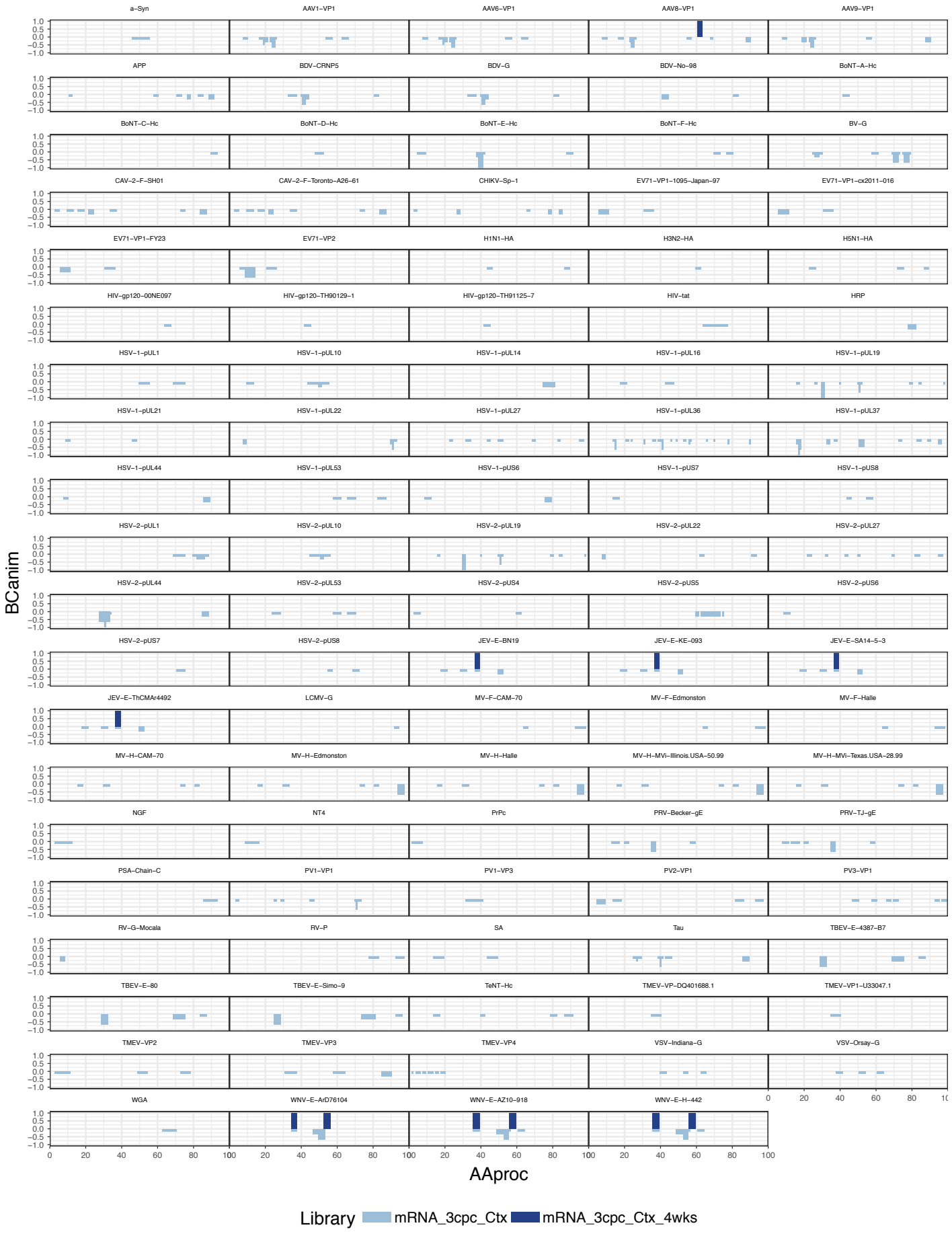

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Th
mRNA_30cpc_Th_4wks

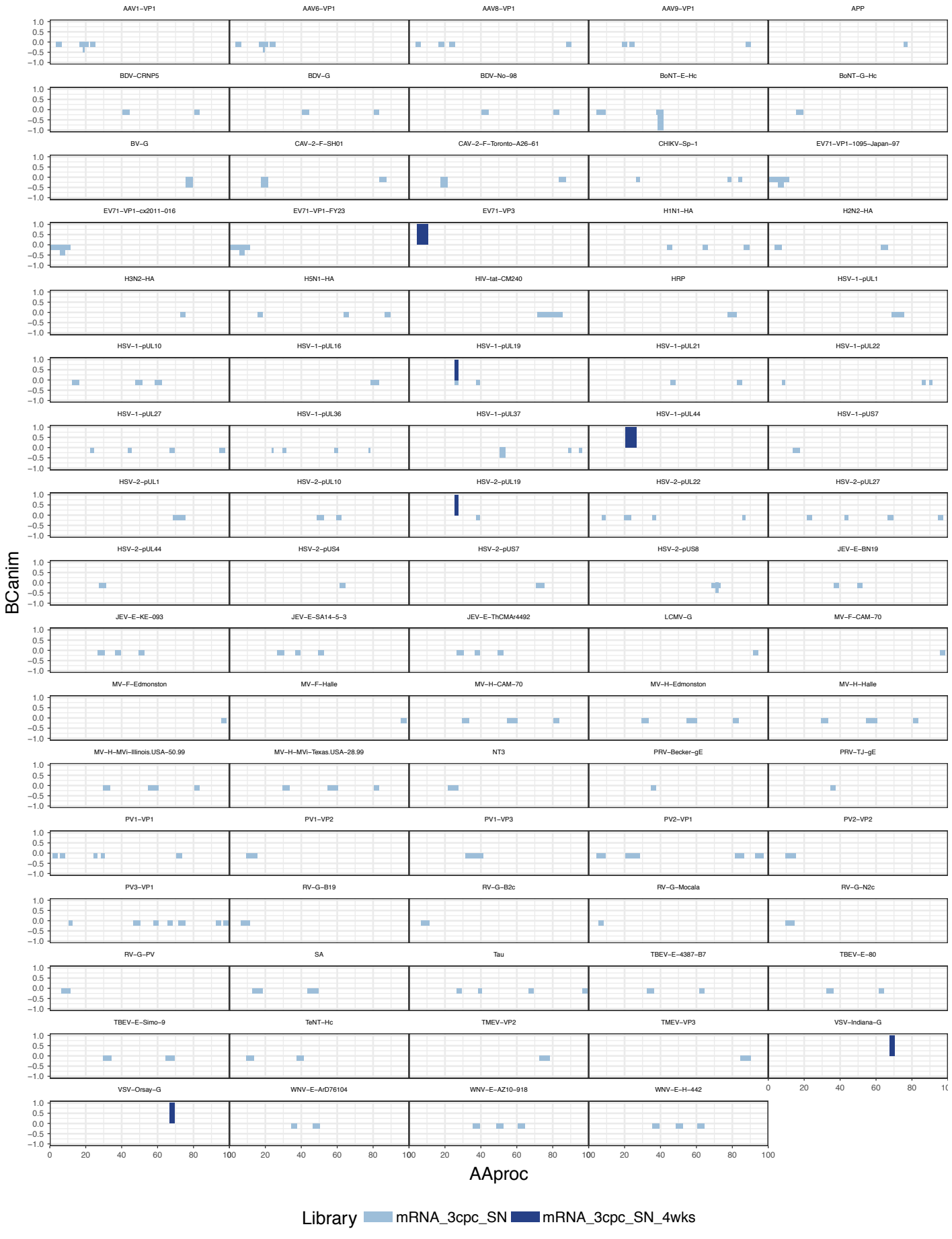

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
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
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−E−BN19
JEV−E−KE−093
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
PV2−VP1
PV2−VP2
PV2−VP3
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_Ctx
mRNA_30cpc_Ctx_4wks

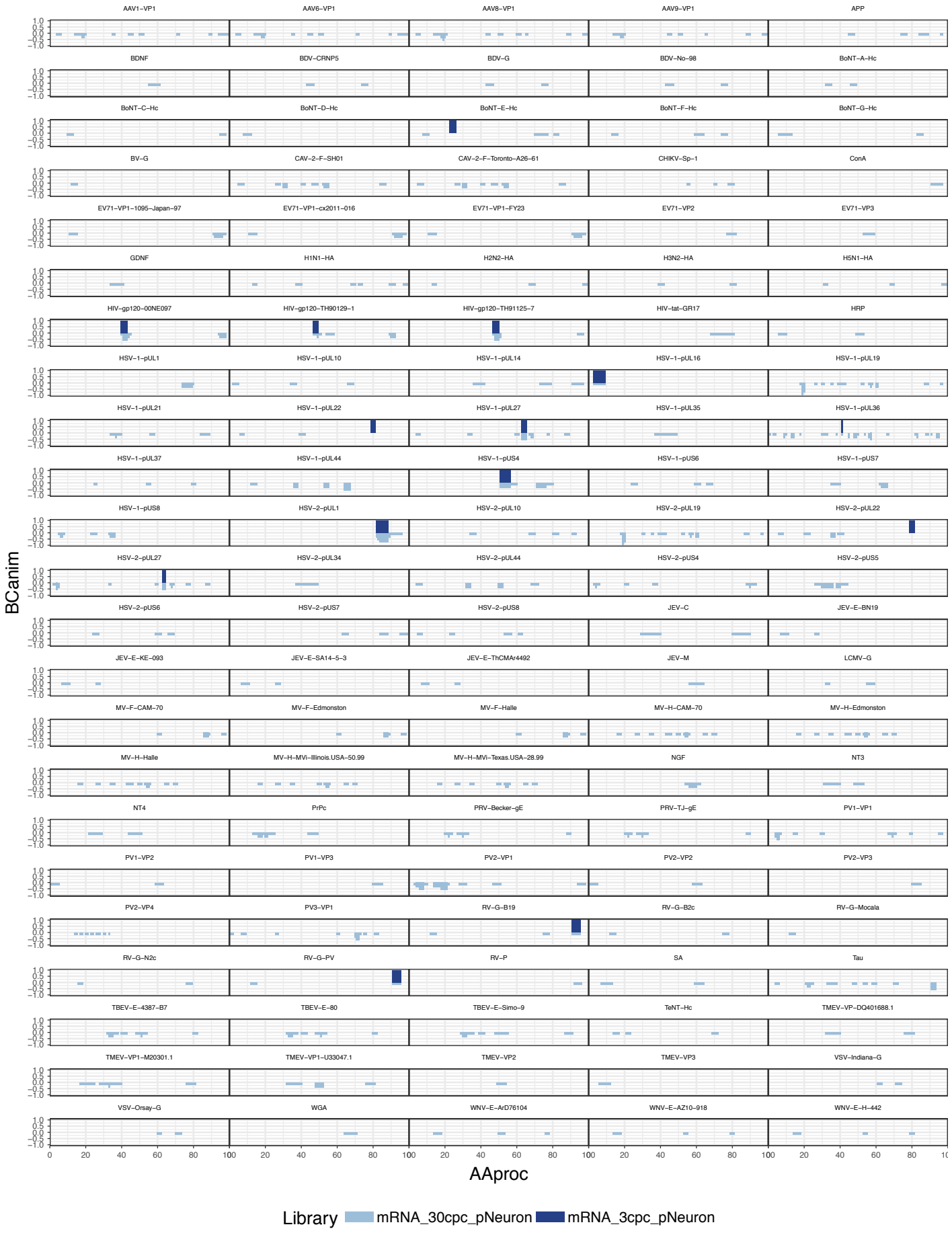

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_SN
mRNA_30cpc_SN_4wks

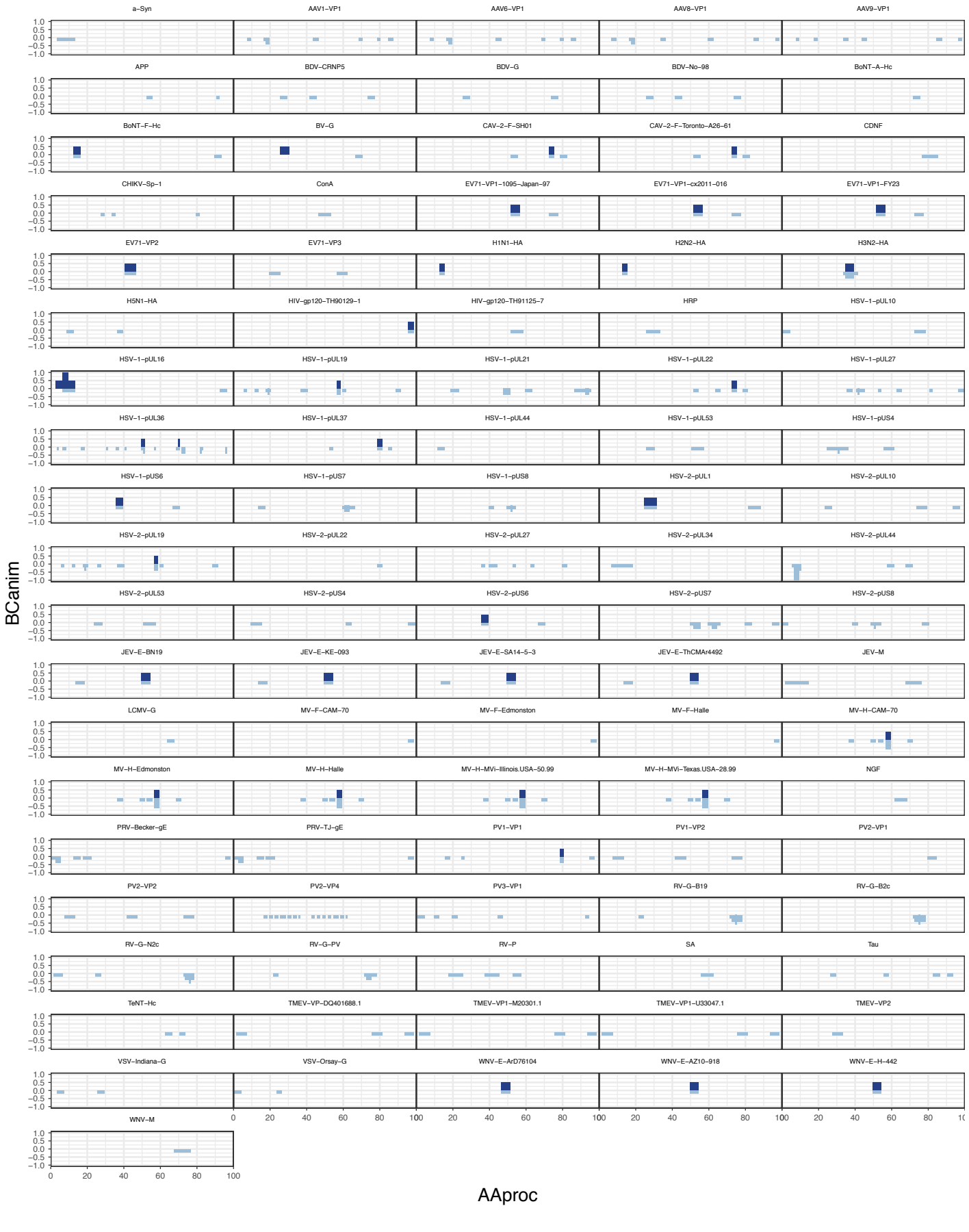

a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−B−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CDNF
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
EV71−VP4
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
HIV−tat−CM240
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
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−E−BN19
JEV−E−KE−093
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Str
mRNA_3cpc_Str_4wks


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
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
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
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
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
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
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Th
mRNA_3cpc_Th_4wks


a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−C−Hc
BoNT−D−Hc
BoNT−E−Hc
BoNT−F−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
H1N1−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HIV−tat
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−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
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−pUL44
HSV−2−pUL53
HSV−2−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
LCMV−G
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
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PSA−Chain−C
PV1−VP1
PV1−VP3
PV2−VP1
PV3−VP1
RV−G−Mocala
RV−P
SA
Tau
TBEV−E−4387−B7
TBEV−E−80
TBEV−E−Simo−9
TeNT−Hc
TMEV−VP−DQ401688.1
TMEV−VP1−U33047.1
TMEV−VP2
TMEV−VP3
TMEV−VP4
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_Ctx
mRNA_3cpc_Ctx_4wks


AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−E−Hc
BoNT−G−Hc
BV−G
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
CHIKV−Sp−1
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP3
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−tat−CM240
HRP
HSV−1−pUL1
HSV−1−pUL10
HSV−1−pUL16
HSV−1−pUL19
HSV−1−pUL21
HSV−1−pUL22
HSV−1−pUL27
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUS7
HSV−2−pUL1
HSV−2−pUL10
HSV−2−pUL19
HSV−2−pUL22
HSV−2−pUL27
HSV−2−pUL44
HSV−2−pUS4
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
LCMV−G
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
NT3
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV1−VP3
PV2−VP1
PV2−VP2
PV3−VP1
RV−G−B19
RV−G−B2c
RV−G−Mocala
RV−G−N2c
RV−G−PV
SA
Tau
TBEV−E−4387−B7
TBEV−E−80
TBEV−E−Simo−9
TeNT−Hc
TMEV−VP2
TMEV−VP3
VSV−Indiana−G
VSV−Orsay−G
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_3cpc_SN
mRNA_3cpc_SN_4wks


AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDNF
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−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
CHIKV−Sp−1
ConA
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
EV71−VP1−FY23
EV71−VP2
EV71−VP3
GDNF
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−00NE097
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
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
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUS4
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−pUS4
HSV−2−pUS5
HSV−2−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−C
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
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
NT3
NT4
PrPc
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV1−VP3
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
VSV−Indiana−G
VSV−Orsay−G
WGA
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_pNeuron
mRNA_3cpc_pNeuron


a−Syn
AAV1−VP1
AAV6−VP1
AAV8−VP1
AAV9−VP1
APP
BDV−CRNP5
BDV−G
BDV−No−98
BoNT−A−Hc
BoNT−F−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
EV71−VP1−FY23
EV71−VP2
EV71−VP3
H1N1−HA
H2N2−HA
H3N2−HA
H5N1−HA
HIV−gp120−TH90129−1
HIV−gp120−TH91125−7
HRP
HSV−1−pUL10
HSV−1−pUL16
HSV−1−pUL19
HSV−1−pUL21
HSV−1−pUL22
HSV−1−pUL27
HSV−1−pUL36
HSV−1−pUL37
HSV−1−pUL44
HSV−1−pUL53
HSV−1−pUS4
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−pUS6
HSV−2−pUS7
HSV−2−pUS8
JEV−E−BN19
JEV−E−KE−093
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
JEV−M
LCMV−G
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
PRV−Becker−gE
PRV−TJ−gE
PV1−VP1
PV1−VP2
PV2−VP1
PV2−VP2
PV2−VP4
PV3−VP1
RV−G−B19
RV−G−B2c
RV−G−N2c
RV−G−PV
RV−P
SA
Tau
TeNT−Hc
TMEV−VP−DQ401688.1
TMEV−VP1−M20301.1
TMEV−VP1−U33047.1
TMEV−VP2
VSV−Indiana−G
VSV−Orsay−G
WNV−E−ArD76104
WNV−E−AZ10−918
WNV−E−H−442
WNV−M
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
100
BCanim
AAproc
Library
mRNA_30cpc_HEK293T
mRNA_3cpc_HEK293T

```
print("Total analysis time:")
```

```
[1] "Total analysis time:"
```

```
print(Sys.time() - strt1)
```

```
Time difference of 2.132019 hours
```

```
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     2017-11-07
```

```
Packages -----------------------------------------------------------------
```

```
 package                * version  date       source
 acepack                  1.4.1    2016-10-29 CRAN (R 3.4.2)
 AnnotationDbi            1.38.2   2017-10-13 Bioconductor
 AnnotationFilter         1.0.0    2017-10-13 Bioconductor
 AnnotationHub            2.8.2    2017-10-13 Bioconductor
 backports                1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                   * 3.4.2    2017-10-06 local
 base64enc                0.1-3    2015-07-28 CRAN (R 3.4.2)
 Biobase                * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics           * 0.22.1   2017-10-13 Bioconductor
 BiocInstaller            1.26.1   2017-10-10 Bioconductor
 BiocParallel             1.10.1   2017-10-13 Bioconductor
 biomaRt                  2.32.1   2017-10-13 Bioconductor
 Biostrings             * 2.44.2   2017-10-13 Bioconductor
 biovizBase               1.24.0   2017-10-13 Bioconductor
 bit                      1.1-12   2014-04-09 CRAN (R 3.4.2)
 bit64                    0.9-7    2017-05-08 CRAN (R 3.4.2)
 bitops                   1.0-6    2013-08-17 CRAN (R 3.4.2)
 blob                     1.1.0    2017-06-17 CRAN (R 3.4.2)
 BSgenome                 1.44.2   2017-10-13 Bioconductor
 checkmate                1.8.4    2017-09-25 CRAN (R 3.4.2)
 cluster                  2.0.6    2017-03-16 CRAN (R 3.4.2)
 codetools                0.2-15   2016-10-05 CRAN (R 3.4.2)
 colorspace               1.3-2    2016-12-14 CRAN (R 3.4.2)
 compiler                 3.4.2    2017-10-06 local
 curl                     2.8.1    2017-07-21 CRAN (R 3.4.2)
 data.table             * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets               * 3.4.2    2017-10-06 local
 DBI                      0.7      2017-06-18 CRAN (R 3.4.2)
 DelayedArray           * 0.2.7    2017-10-13 Bioconductor
 devtools               * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 dichromat                2.0-0    2013-01-24 CRAN (R 3.4.2)
 digest                   0.6.12   2017-01-27 CRAN (R 3.4.2)
 doParallel             * 1.0.11   2017-09-28 CRAN (R 3.4.2)
 ensembldb                2.0.4    2017-10-13 Bioconductor
 evaluate                 0.10.1   2017-06-24 CRAN (R 3.4.2)
 foreach                * 1.4.3    2015-10-13 CRAN (R 3.4.2)
 foreign                  0.8-69   2017-06-21 CRAN (R 3.4.2)
```

```
formatR                   1.5      2017-04-25 CRAN (R 3.4.2)
Formula                   1.2-2    2017-07-10 CRAN (R 3.4.2)
GenomeInfoDb            * 1.12.3   2017-10-13 Bioconductor
GenomeInfoDbData          0.99.0   2017-10-13 Bioconductor
GenomicAlignments       * 1.12.2   2017-10-13 Bioconductor
GenomicFeatures           1.28.5   2017-10-13 Bioconductor
GenomicRanges           * 1.28.6   2017-10-13 Bioconductor
GGally                    1.3.2    2017-08-02 CRAN (R 3.4.2)
ggbio                   * 1.24.1   2017-10-13 Bioconductor
ggplot2                 * 2.2.1    2016-12-30 CRAN (R 3.4.2)
graph                     1.54.0   2017-10-13 Bioconductor
graphics                * 3.4.2    2017-10-06 local
grDevices               * 3.4.2    2017-10-06 local
grid                    * 3.4.2    2017-10-06 local
gridExtra                 2.3      2017-09-09 CRAN (R 3.4.2)
gtable                    0.2.0    2016-02-26 CRAN (R 3.4.2)
Hmisc                     4.0-3    2017-05-02 CRAN (R 3.4.2)
hms                       0.3      2016-11-22 CRAN (R 3.4.2)
htmlTable                 1.9      2017-01-26 CRAN (R 3.4.2)
htmltools                 0.3.6    2017-04-28 CRAN (R 3.4.2)
htmlwidgets               0.9      2017-07-10 CRAN (R 3.4.2)
httpuv                    1.3.5    2017-07-04 CRAN (R 3.4.2)
httr                      1.3.1    2017-08-20 CRAN (R 3.4.2)
interactiveDisplayBase    1.14.0   2017-10-13 Bioconductor
IRanges                 * 2.10.5   2017-10-13 Bioconductor
iterators               * 1.0.8    2015-10-13 CRAN (R 3.4.2)
kableExtra              * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr                   * 1.17     2017-08-10 CRAN (R 3.4.2)
labeling                  0.3      2014-08-23 CRAN (R 3.4.2)
lattice                   0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra              0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval                  0.2.0    2016-06-12 CRAN (R 3.4.2)
magrittr                  1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                    1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats             * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise                   1.1.0    2017-04-21 CRAN (R 3.4.2)
methods                 * 3.4.2    2017-10-06 local
mime                      0.5      2016-07-07 CRAN (R 3.4.2)
munsell                   0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                      7.3-12   2016-02-02 CRAN (R 3.4.2)
OrganismDbi               1.18.1   2017-10-13 Bioconductor
parallel                * 3.4.2    2017-10-06 local
plyr                    * 1.8.4    2016-06-08 CRAN (R 3.4.2)
ProtGenerics              1.8.0    2017-10-13 Bioconductor
R6                        2.2.2    2017-06-17 CRAN (R 3.4.2)
RBGL                      1.52.0   2017-10-13 Bioconductor
RColorBrewer              1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                      0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                     1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                     1.1.1    2017-05-16 CRAN (R 3.4.2)
reshape                   0.8.7    2017-08-06 CRAN (R 3.4.2)
reshape2                  1.4.2    2016-10-22 CRAN (R 3.4.2)
rlang                     0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown                 1.6      2017-06-15 CRAN (R 3.4.2)
rpart                     4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot                 1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools               * 1.28.0   2017-10-13 Bioconductor
RSQLite                   2.0      2017-06-19 CRAN (R 3.4.2)
rtracklayer               1.36.6   2017-10-13 Bioconductor
```

```
rvest                    0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors              * 0.14.7   2017-10-13 Bioconductor
scales                   0.5.0    2017-08-24 CRAN (R 3.4.2)
shiny                    1.0.5    2017-08-23 CRAN (R 3.4.2)
splines                  3.4.2    2017-10-06 local
stats                  * 3.4.2    2017-10-06 local
stats4                 * 3.4.2    2017-10-06 local
stringi                  1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                  1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment   * 1.6.5    2017-10-13 Bioconductor
survival                 2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                   1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                    3.4.2    2017-10-06 local
utils                  * 3.4.2    2017-10-06 local
VariantAnnotation        1.22.3   2017-10-13 Bioconductor
withr                    2.0.0    2017-07-28 CRAN (R 3.4.2)
XML                      3.98-1.9 2017-06-19 CRAN (R 3.4.2)
xml2                     1.1.1    2017-01-24 CRAN (R 3.4.2)
xtable                   1.8-2    2016-02-05 CRAN (R 3.4.2)
XVector                * 0.16.0   2017-10-13 Bioconductor
yaml                     2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc                 1.22.0   2017-10-13 Bioconductor
```

# Heatmap analysis output

*Tomas Bjorklund*

*Wed Sep 27 16:23:32 2017*

MNM021

A secondary approach to visualizing the top candidates and their relative abundance.

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

```
opts_chunk$set(fig.width = 8, fig.height = 10.2)
```

## Selection of relevant samples

```
select.samples <- readRDS("data/allSamplesDataTable.RDS")
select.samples$Group[select.samples$Group== "mRNA_3cpc_HEK293T"] <- "mRNA_3cpc_HEK293T"
select.samples$Group[select.samples$Group== "mRNA_30cpc_HEK293T"] <- "mRNA_30cpc_HEK293T"
select.samples <- select.samples[-grep("4wks|mRNA_3cpc_pNeuron_RNA",select.samples$Group),]

select.samples.binCat <- data.table::copy(select.samples)
setkeyv(select.samples.binCat,c("Group","Category"))
```

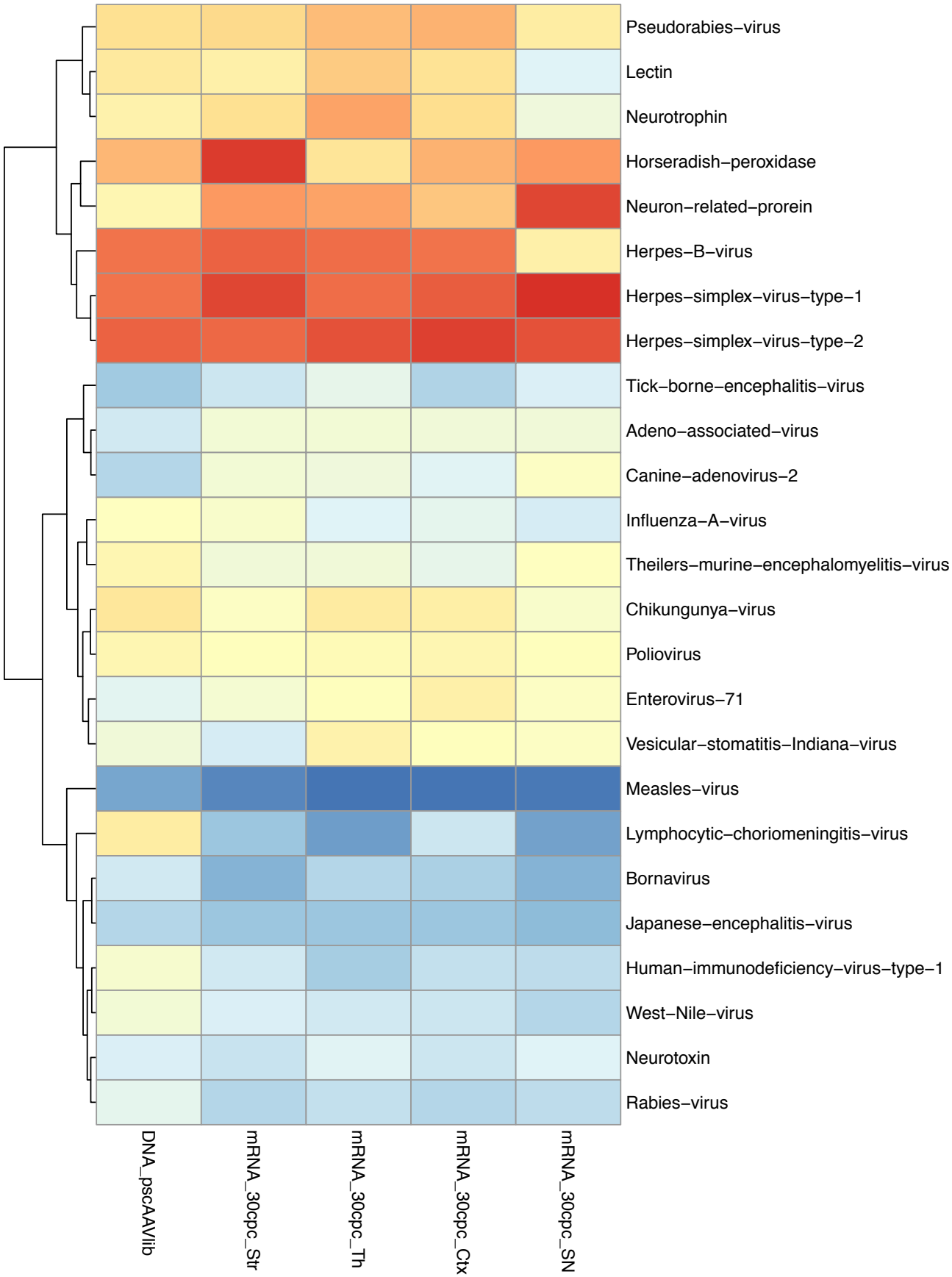
*MNM004

```
select.samples.binCat <- select.samples.binCat[,list(BCcount=length(table(strsplit(paste(t(BC),
                                                    collapse=","), ","))), NormCount=mean(log2(RNAcount+1))
                                                        ), by=key(select.samples.binCat)]

setkey(select.samples.binCat,Group)
ref.table <- select.samples.binCat["DNA_pscAAVlib"]
ref.table[,c("Group","NormCount"):=NULL]
setnames(ref.table,"BCcount","libBC")
setkey(ref.table,"Category")
select.samples.binCat[,totBC:=sum(BCcount), by="Group"]
max.count <- max(select.samples.binCat$totBC)
select.samples.binCat[,BCcountN:=BCcount/totBC*max.count]
length.Table <- unique(select.samples, by=c("Category","GeneName"))
length.Table <- length.Table[,list(seqlength=sum(seqlength)), by="Category"]
setkey(length.Table,"Category")
setkey(select.samples.binCat,"Category")
select.samples.binCat <- select.samples.binCat[length.Table,nomatch=0]
select.samples.binCat <- select.samples.binCat[ref.table,nomatch=0]

select.samples.binCat[,Category:=gsub("/|_|'","-",Category)]
select.samples.binCat[,BCcountNseq:=BCcountN/seqlength]
select.samples.binCat[,refNormBC:=BCcountN/libBC]



select.samples.binGene <- data.table::copy(select.samples)
setkeyv(select.samples.binGene,c("Group","Category","GeneName","seqlength"))
select.samples.binGene <- select.samples.binGene[,list(BCcount=length(table(strsplit(paste(t(BC), collapse="
                                                    NormCount=mean(log2(RNAcount+1))), by=key(select.samp
select.samples.binGene[,GeneName:=gsub("/|_|'","-",GeneName)]
setkey(select.samples.binGene,Group)
ref.table <- select.samples.binGene["DNA_pscAAVlib"]
ref.table[,c("Group","Category","NormCount","seqlength"):=NULL]
setnames(ref.table,"BCcount","libBC")
```

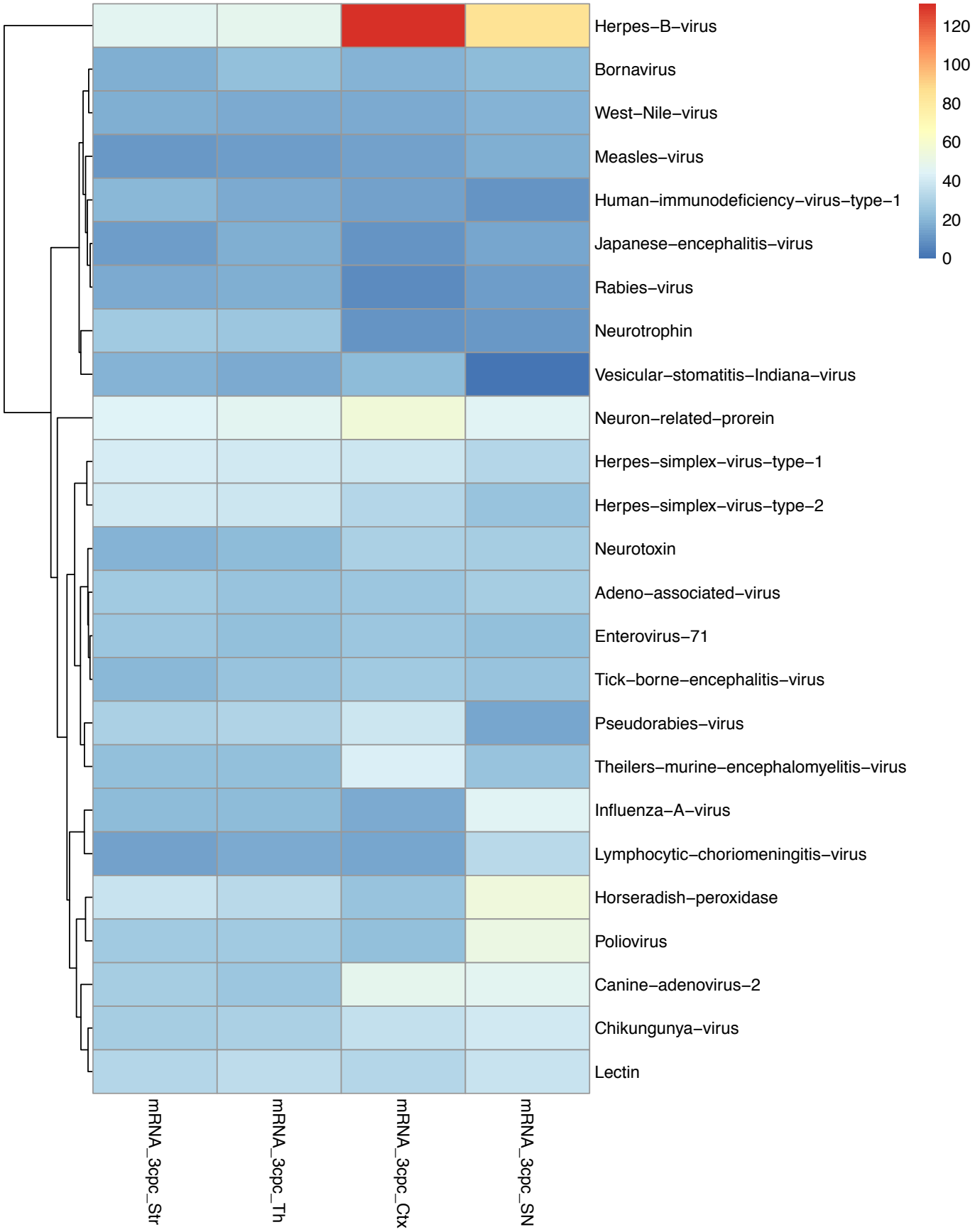
```
setkey(ref.table,"GeneName")
setkey(select.samples.binGene,"GeneName")
select.samples.binGene[,totBC:=sum(BCcount), by="Group"]
select.samples.binGene <- select.samples.binGene[ref.table,nomatch=0]
max.count <- max(select.samples.binGene$totBC)
select.samples.binGene[,BCcountN:=BCcount/totBC*max.count]
select.samples.binGene[,BCcountNseq:=BCcountN/seqlength]
select.samples.binGene[,refNormBC:=BCcountN/libBC]

select.samples.binPos <- data.table::copy(select.samples)
setkeyv(select.samples.binPos,c("Group","structure","Sequence"))
select.samples.binPos <- unique(select.samples.binPos, by=c("Group","structure","Sequence"))
#Due to key, this removes replicates if identical sequence mapped to multiple genes

setkeyv(select.samples.binPos,c("Group","GeneName","AA","seqlength"))
select.samples.binPos <- select.samples.binPos[,list(BCcount=length(table(strsplit(paste(t(BC), collapse=","
                                                    NormCount=mean(log2(RNAcount+1)),
                                                    AnimalCount=length(table(strsplit(paste(t(Animals), col
                                                    LUTnrs=paste(unique(names(table(strsplit(paste(t(LUTnrs
                                                    mainStruct=paste(unique(structure), collapse=","),
                                                    mismatches=median(mismatches)), by=key(select.samples.b
setkey(select.samples.binPos,Group)
ref.table <- select.samples.binPos["DNA_pscAAVlib"]
ref.table[,c("Group","NormCount","AnimalCount","LUTnrs","mainStruct","mismatches","seqlength"):=NULL]
setnames(ref.table,"BCcount","libBC")
setkeyv(ref.table,c("GeneName","AA"))
setkeyv(select.samples.binPos,c("GeneName","AA"))
select.samples.binPos <- select.samples.binPos[ref.table,nomatch=0]
select.samples.binPos[,totBC:=sum(BCcount), by="Group"]
max.count <- max(select.samples.binPos$totBC)
select.samples.binPos[,BCcountN:=BCcount/totBC*max.count]
select.samples.binPos[,libNormBC:=BCcountN/libBC]
select.samples.binPos[,BCcountNanim:=BCcountN+AnimalCount]
select.samples.binPos[,BCcountanim:=BCcount+AnimalCount]
select.samples.binPos[,BCcountNseq:=BCcountN/seqlength]
select.samples.binPos[,NormCountBC:=BCcountNseq*NormCount]


select.samples.binPos[,GeneAA:=paste(GeneName," [",AA,"] - ", mainStruct, sep="")]
```

## Plot Heatmaps split by Category

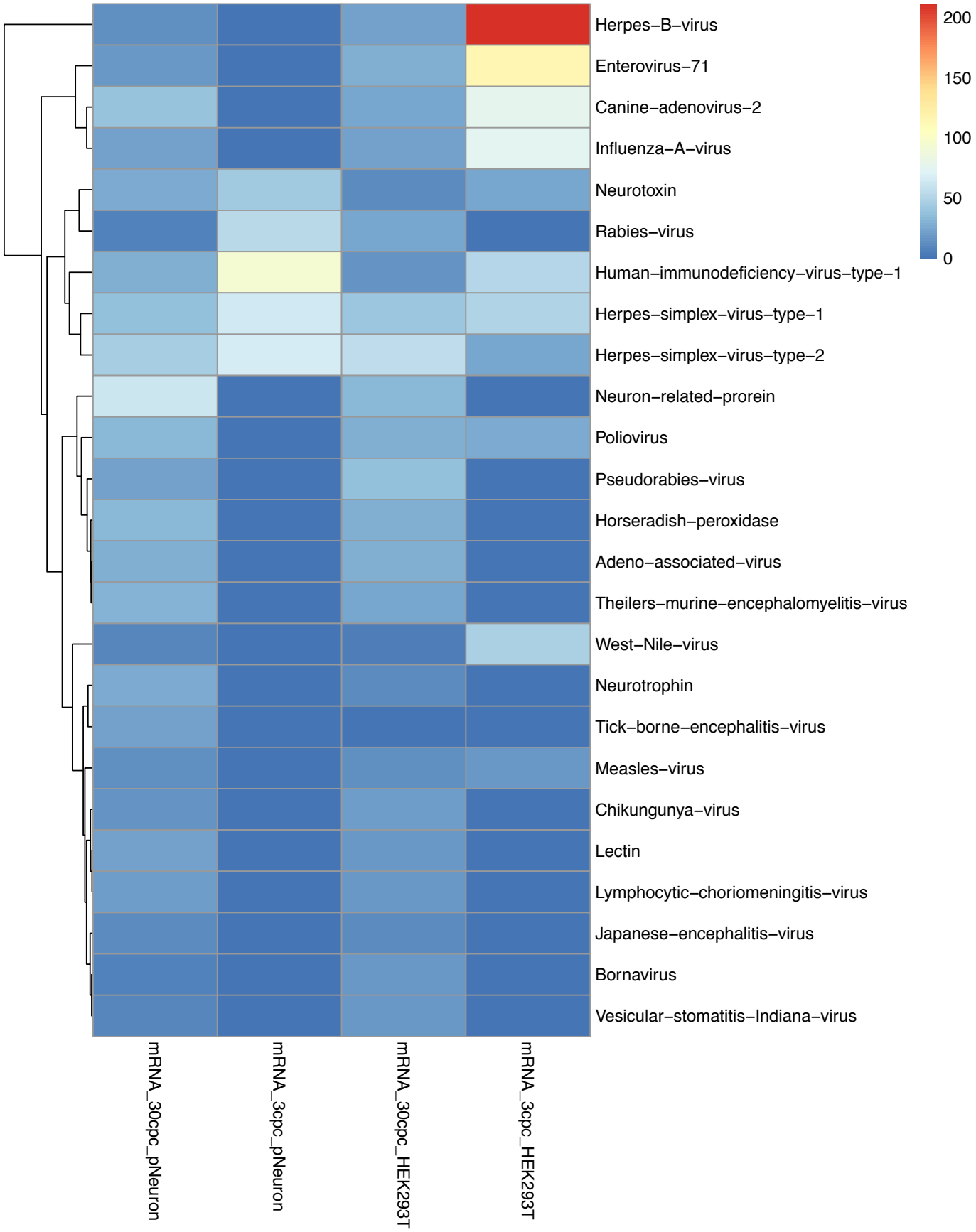
```
plotCategory <- function(select.samples.table,plot.col,sample.select){
  setkey(select.samples.table,Group)
  select.samples.select <- select.samples.table[sample.select]
  eval(parse(text=paste("setorder(select.samples.select,Group, -", plot.col,")", sep="")))
  select.samples.matrix <- acast(select.samples.select, Category~Group, value.var=plot.col)
  #Utilizes reshape 2 to make matrix for heatmap

  select.samples.matrix[is.na(select.samples.matrix)] <- 0
  select.samples.matrix <- select.samples.matrix[,sample.select]
  return(pheatmap(select.samples.matrix, cluster_rows=TRUE, show_rownames=TRUE, cluster_cols=FALSE))
}

plotCategory(select.samples.binCat,"BCcountNseq",c("DNA_pscAAVlib","mRNA_30cpc_Str","mRNA_30cpc_Th","mRNA_30
```


Pseudorabies−virus
Lectin
Neurotrophin
Horseradish−peroxidase
Neuron−related−prorein
Herpes−B−virus
Herpes−simplex−virus−type−1
Herpes−simplex−virus−type−2
Tick−borne−encephalitis−virus
Adeno−associated−virus
Canine−adenovirus−2
Influenza−A−virus
Theilers−murine−encephalomyelitis−virus
Chikungunya−virus
Poliovirus
Enterovirus−71
Vesicular−stomatitis−Indiana−virus
Measles−virus
Lymphocytic−choriomeningitis−virus
Bornavirus
Japanese−encephalitis−virus
Human−immunodeficiency−virus−type−1
West−Nile−virus
Neurotoxin
Rabies−virus
40
35
30
25
20
15
10
DNA_pscAAVlib
mRNA_30cpc_Str
mRNA_30cpc_Th
mRNA_30cpc_Ctx
mRNA_30cpc_SN

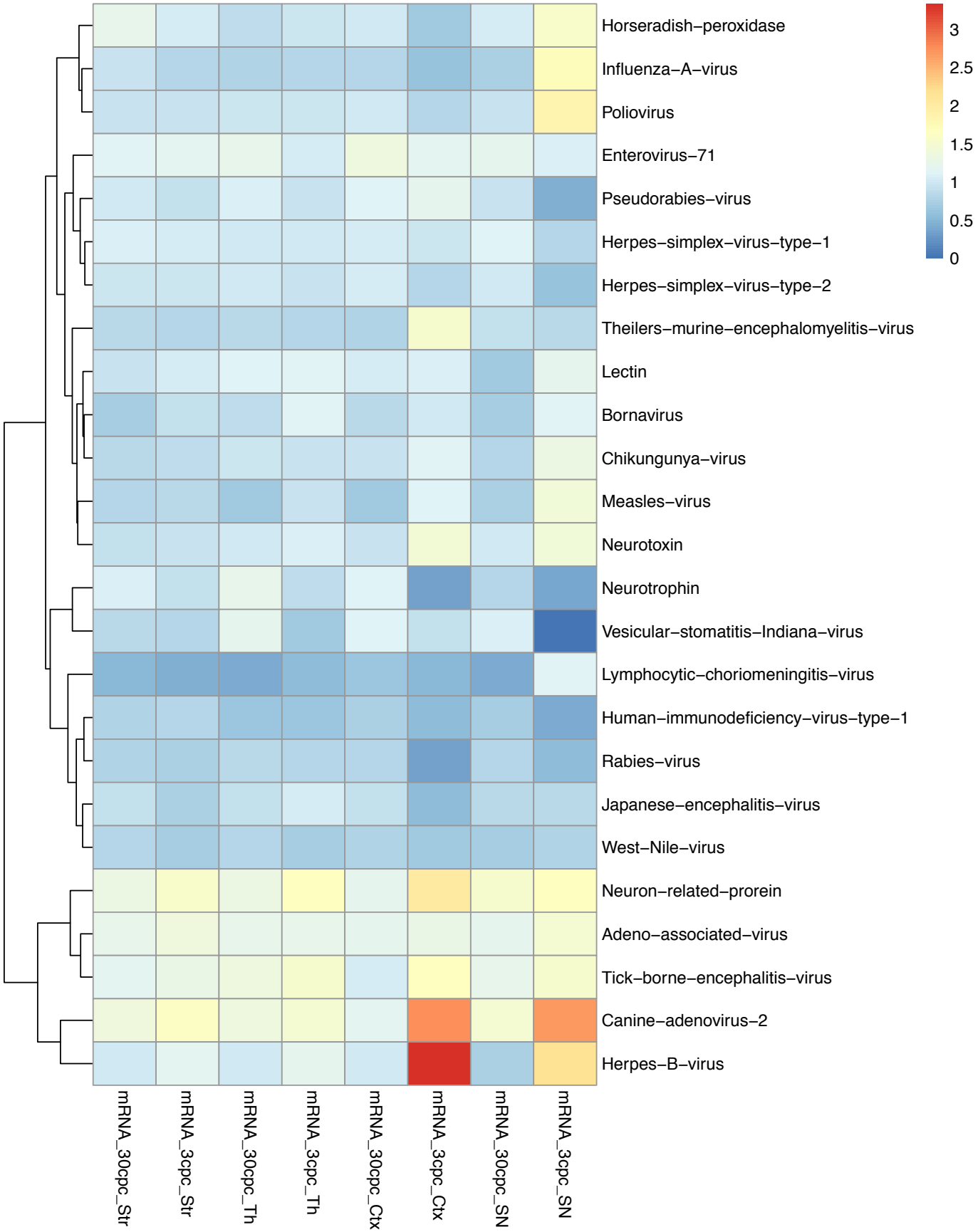
```
plotCategory(select.samples.binCat,"BCcountNseq",c("mRNA_3cpc_Str","mRNA_3cpc_Th","mRNA_3cpc_Ctx","mRNA_3cpc
```



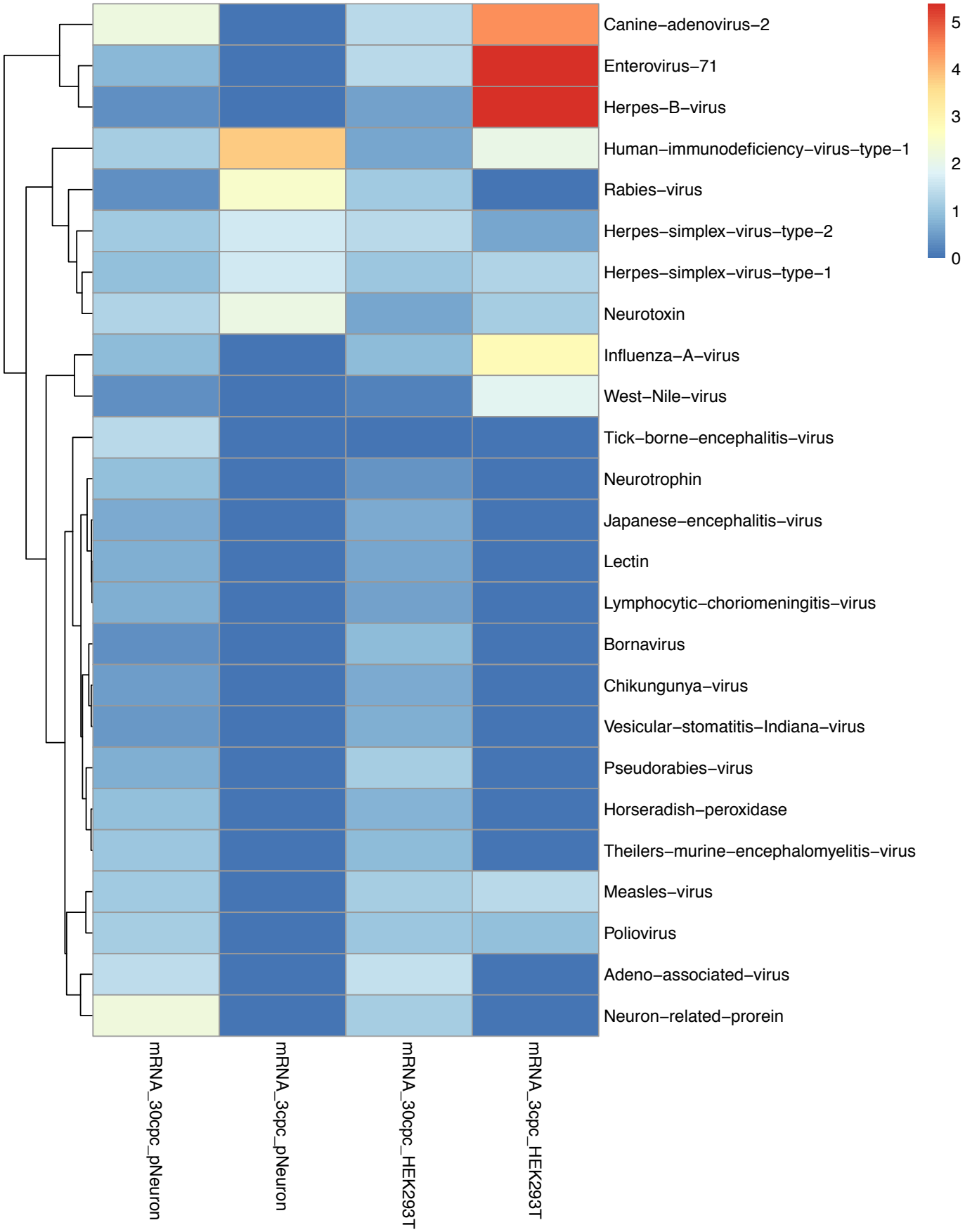
```
plotCategory(select.samples.binCat,"BCcountNseq",c("mRNA_30cpc_pNeuron","mRNA_3cpc_pNeuron","mRNA_30cpc_HEK
```



```
plotCategory(select.samples.binCat,"refNormBC",c("mRNA_30cpc_Str","mRNA_3cpc_Str","mRNA_30cpc_Th","mRNA_3cpc
```

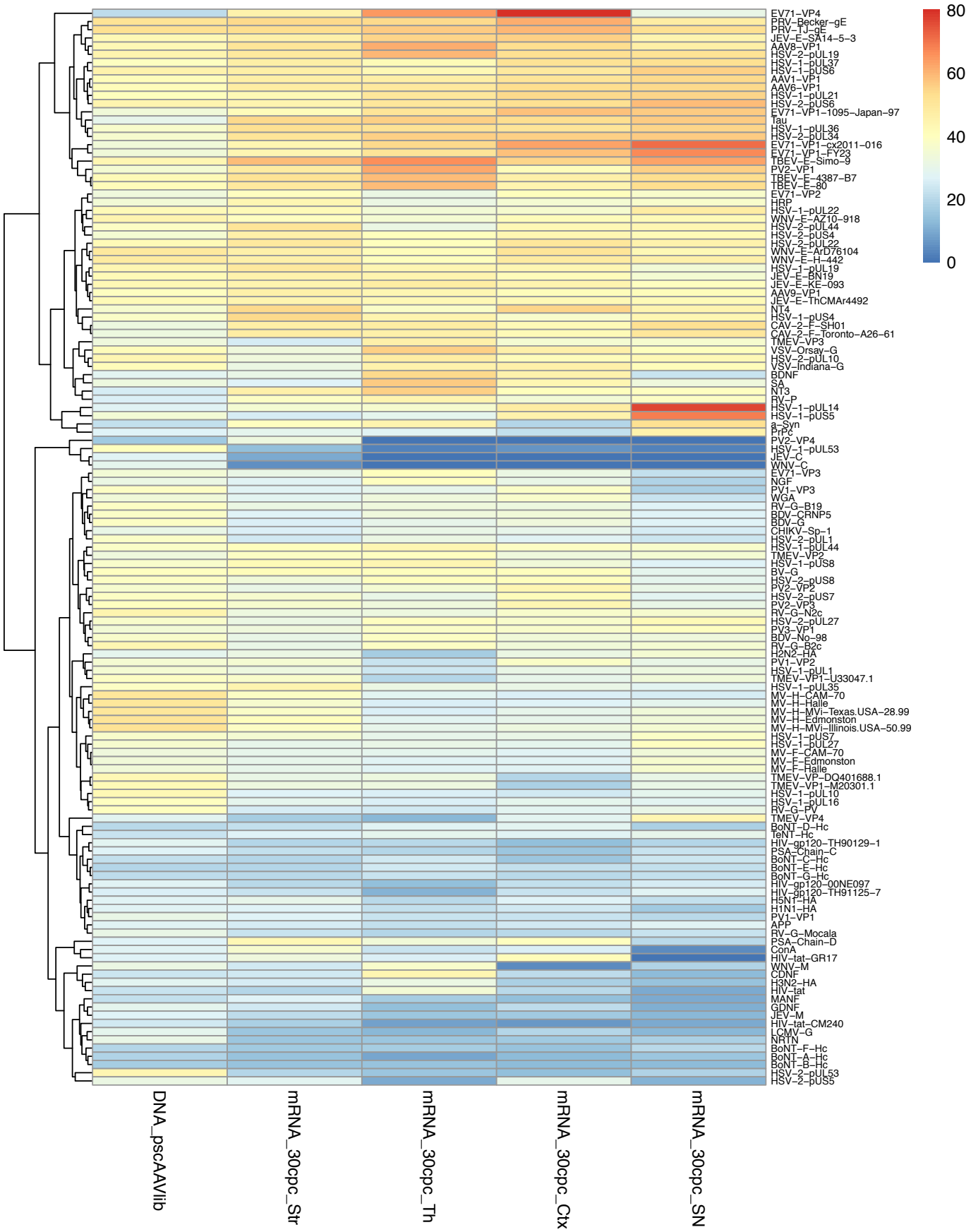


```
plotCategory(select.samples.binCat,"refNormBC",c("mRNA_30cpc_pNeuron","mRNA_3cpc_pNeuron","mRNA_30cpc_HEK293
```

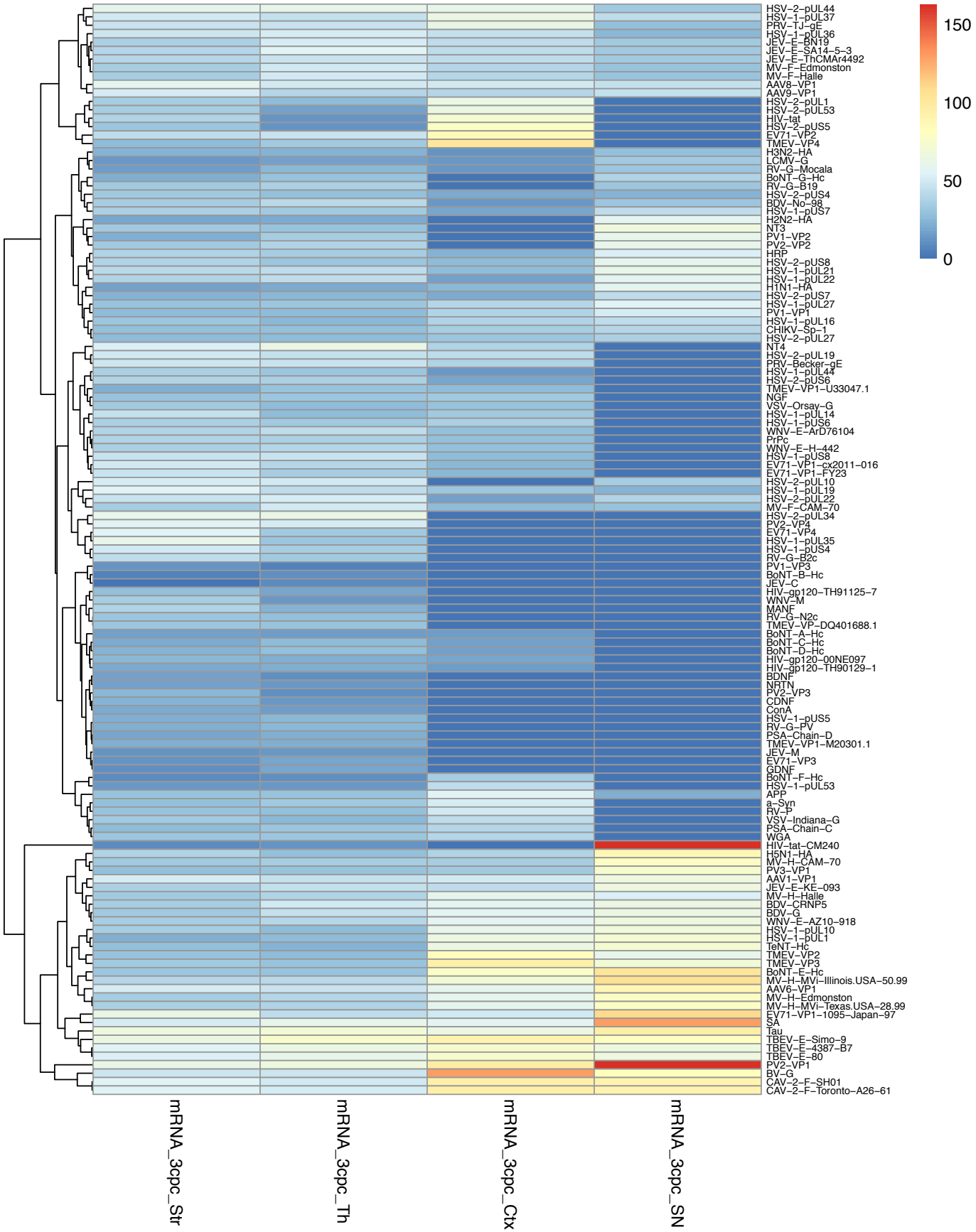


## Plot Heatmaps split by GeneName

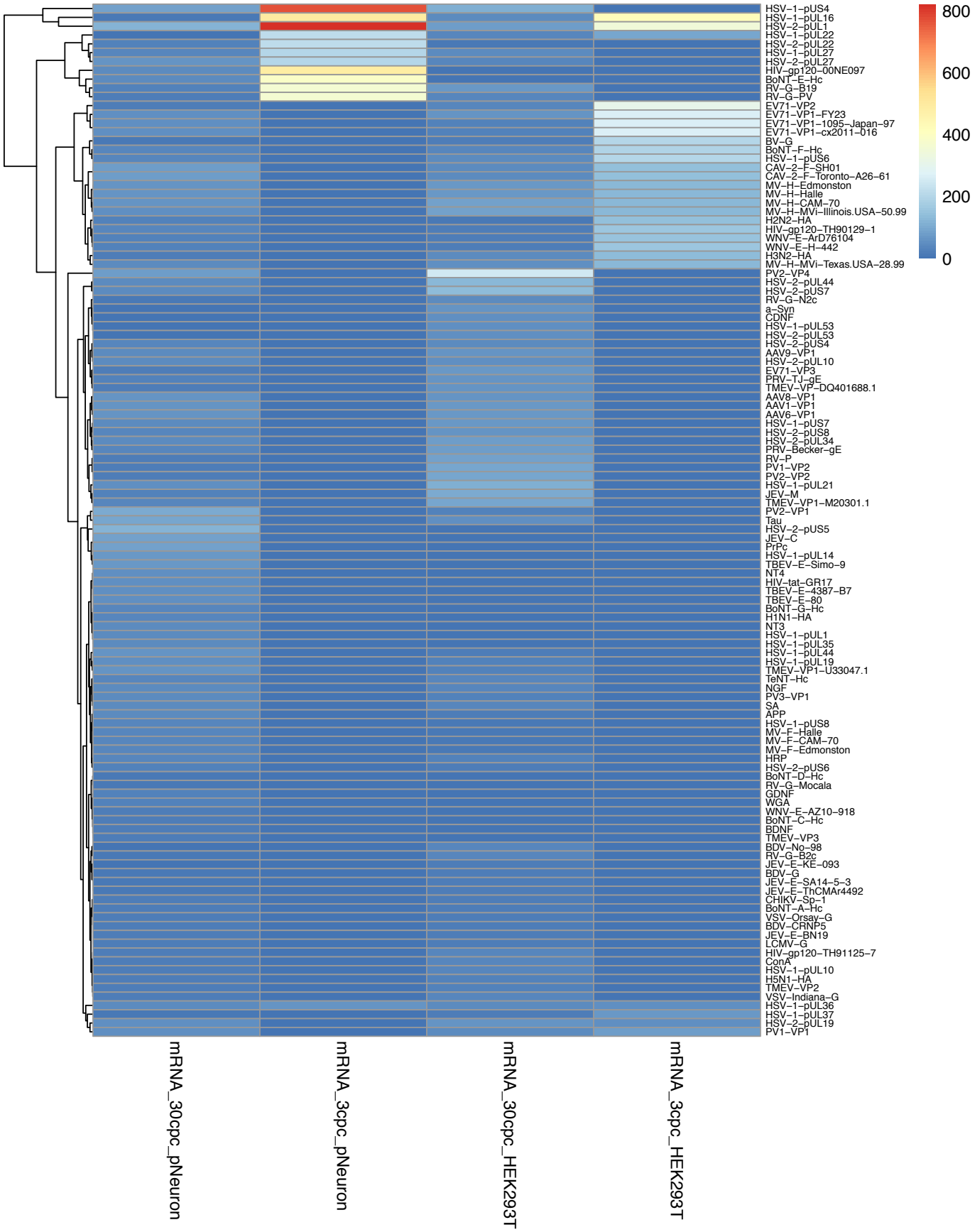
```
plotGene <- function(select.samples.table,plot.col,sample.select){
  setkey(select.samples.table,Group)
  select.samples.select <- select.samples.table[sample.select]
  eval(parse(text=paste("setorder(select.samples.select,Group, -", plot.col,")", sep="")))
  select.samples.matrix <- acast(select.samples.select, GeneName~Group, value.var=plot.col)
  #Utilizes reshape 2 to make matrix for heatmap

  select.samples.matrix[is.na(select.samples.matrix)] <- 0
  select.samples.matrix <- select.samples.matrix[,sample.select]
  return(pheatmap(select.samples.matrix, fontsize_row=5.8, cluster_rows=TRUE, show_rownames=TRUE, cluster_co
}

plotGene(select.samples.binGene,"BCcountNseq",c("DNA_pscAAVlib","mRNA_30cpc_Str","mRNA_30cpc_Th","mRNA_30cpc
```



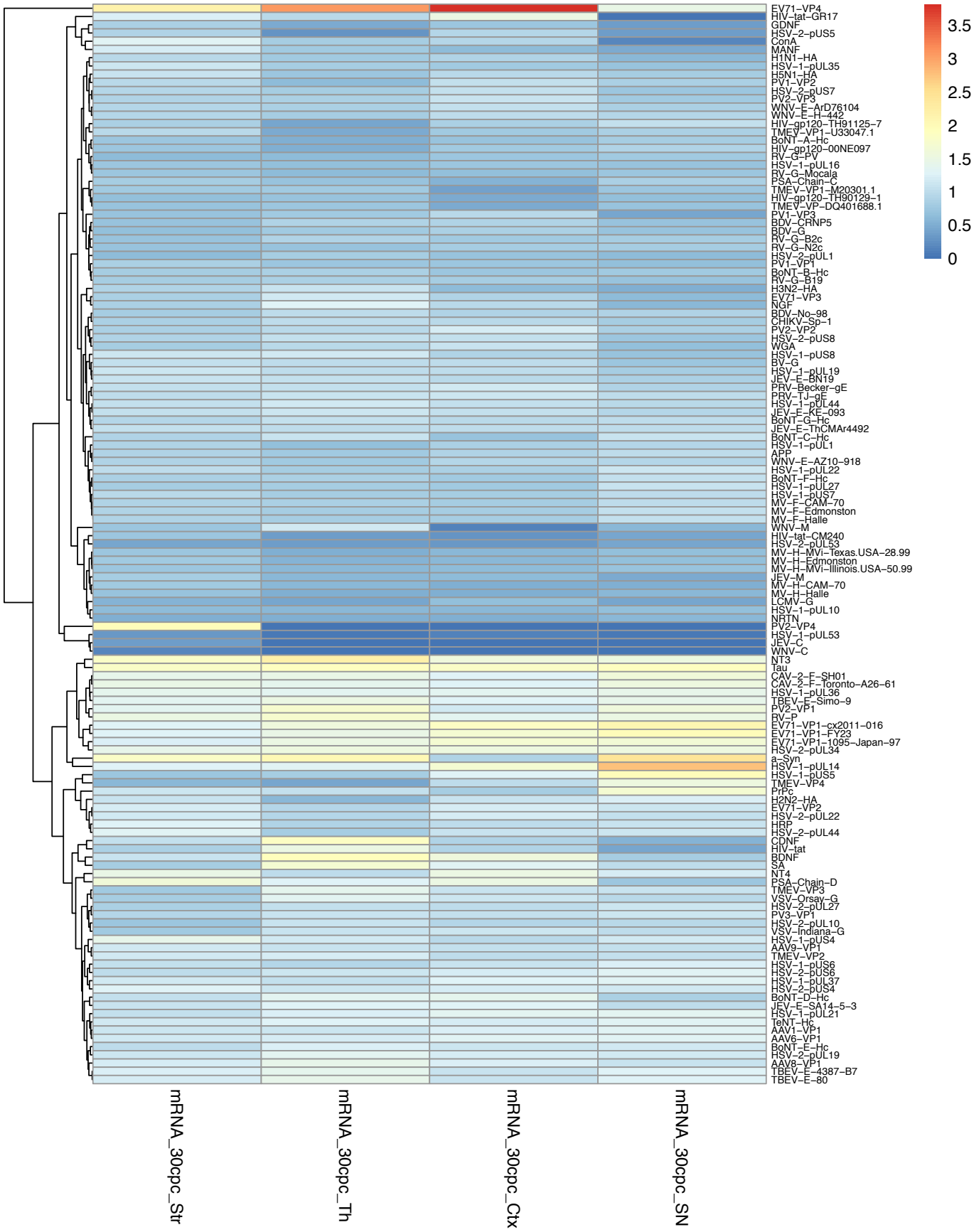

```
plotGene(select.samples.binGene,"BCcountNseq",c("mRNA_3cpc_Str","mRNA_3cpc_Th","mRNA_3cpc_Ctx","mRNA_3cpc_SN
```



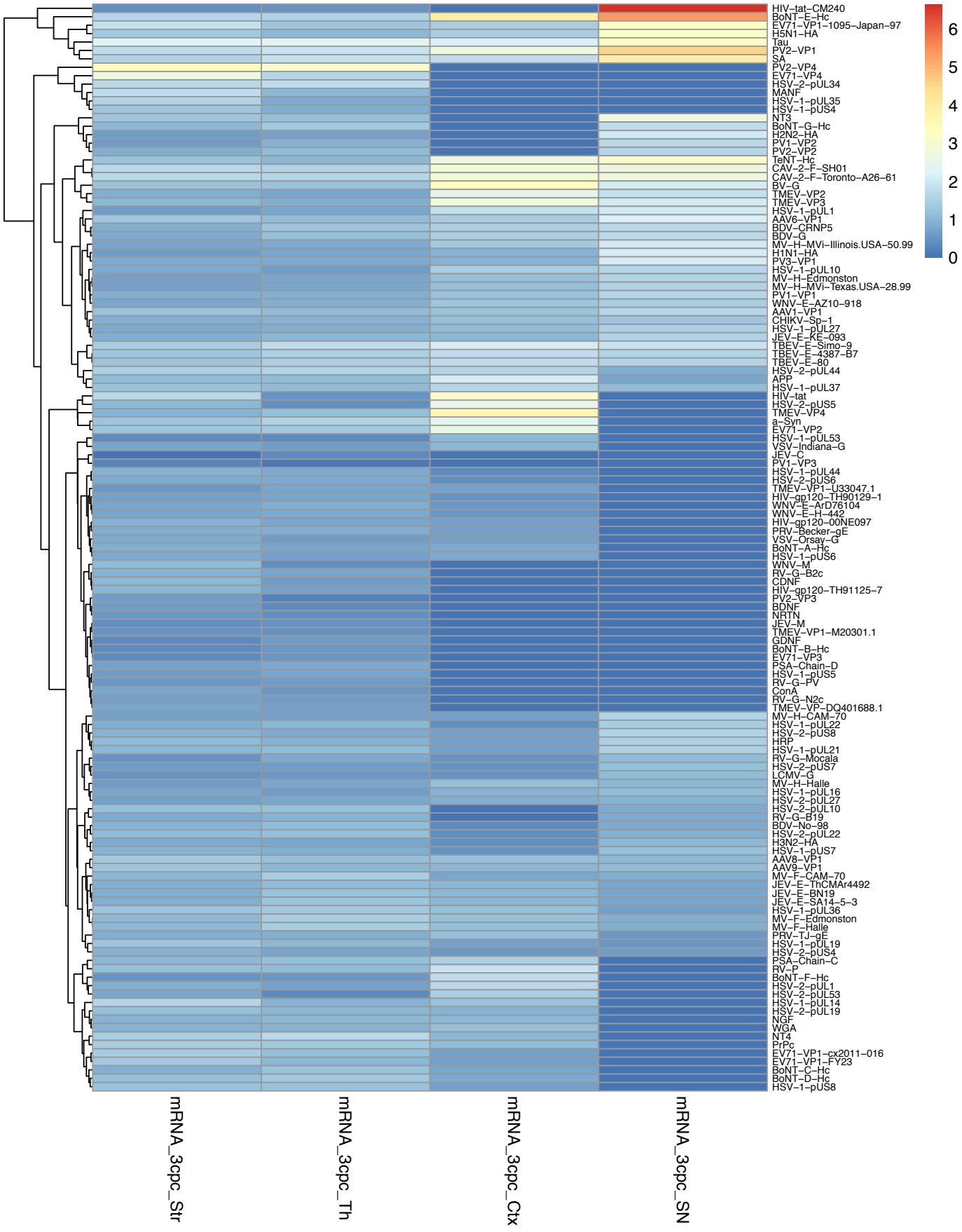

```
plotGene(select.samples.binGene,"BCcountNseq",c("mRNA_30cpc_pNeuron","mRNA_3cpc_pNeuron","mRNA_30cpc_HEK293
```



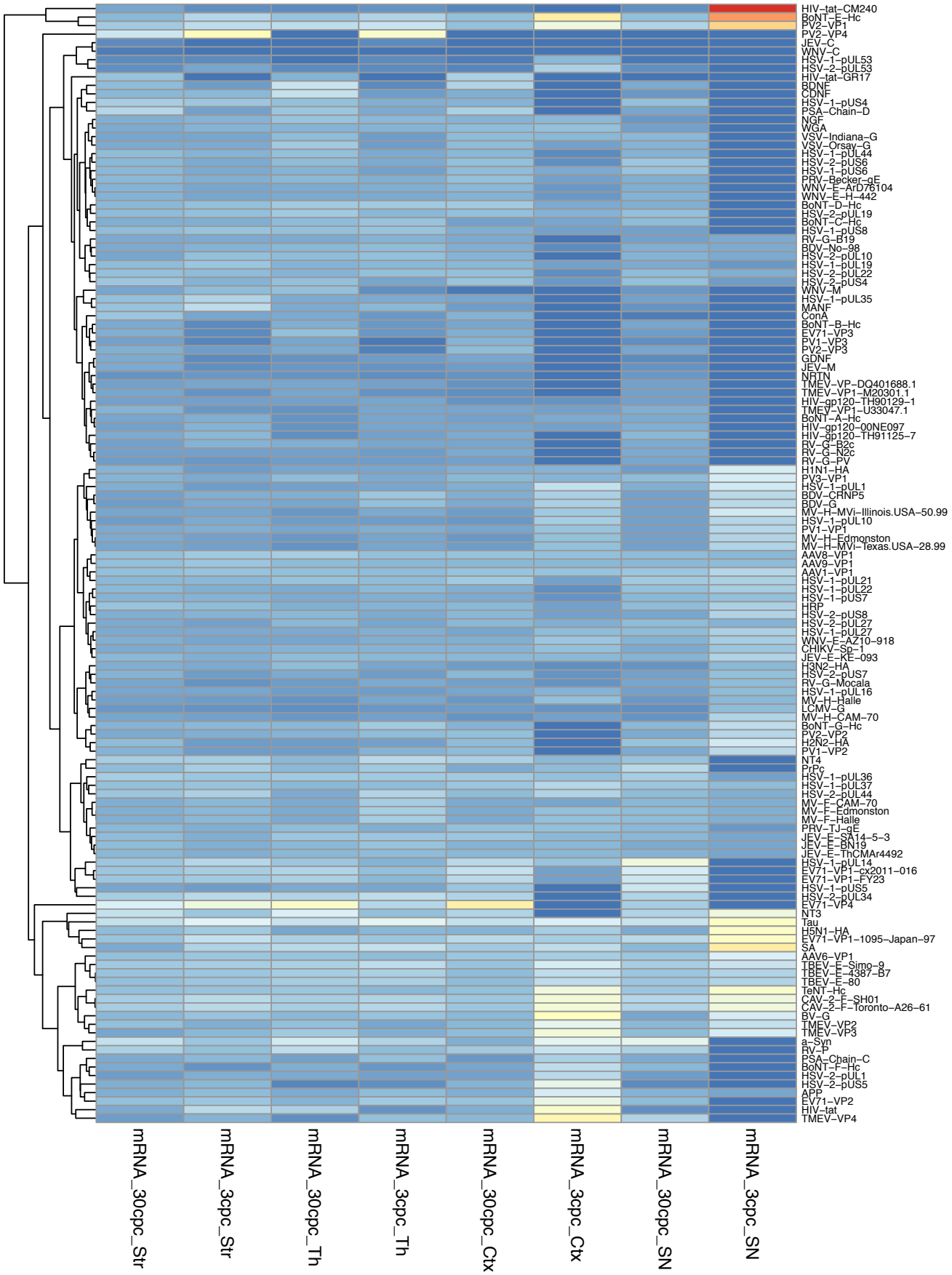

```
plotGene(select.samples.binGene,"refNormBC",c("mRNA_30cpc_Str","mRNA_30cpc_Th","mRNA_30cpc_Ctx","mRNA_30cpc_
```




```
plotGene(select.samples.binGene,"refNormBC",c("mRNA_3cpc_Str","mRNA_3cpc_Th","mRNA_3cpc_Ctx","mRNA_3cpc_SN")
```




```
plotGene(select.samples.binGene,"refNormBC",c("mRNA_30cpc_Str","mRNA_3cpc_Str","mRNA_30cpc_Th","mRNA_3cpc_Th
```




```
plotGene(select.samples.binGene,"refNormBC",c("mRNA_30cpc_pNeuron","mRNA_3cpc_pNeuron","mRNA_30cpc_HEK293T"
```


HSV−1−pUS4
BoNT−E−Hc
HIV−gp120−00NE097
HSV−1−pUL16
HSV−2−pUL1
PV2−VP4
RV−G−B19
RV−G−PV
HSV−1−pUL22
HSV−2−pUL22
HSV−1−pUL27
HSV−2−pUL27
EV71−VP1−FY23
EV71−VP1−1095−Japan−97
EV71−VP1−cx2011−016
BoNT−F−Hc
EV71−VP2
WNV−E−ArD76104
WNV−E−H−442
MV−H−MVi−Texas.USA−28.99
PV1−VP1
MV−H−Edmonston
MV−H−Halle
MV−H−CAM−70
MV−H−MVi−Illinois.USA−50.99
CAV−2−F−SH01
CAV−2−F−Toronto−A26−61
HIV−gp120−TH90129−1
H3N2−HA
BV−G
H2N2−HA
HSV−1−pUS6
CDNF
a−Syn
PV1−VP2
PV2−VP2
HSV−1−pUL21
HSV−2−pUL44
RV−P
HSV−2−pUS7
JEV−M
Tau
HSV−2−pUS5
PV2−VP1
HSV−1−pUL14
JEV−C
PrPc
HSV−1−pUL36
HSV−1−pUL37
HSV−2−pUL19
HSV−2−pUL34
EV71−VP3
TMEV−VP1−M20301.1
AAV6−VP1
AAV1−VP1
TeNT−Hc
AAV8−VP1
NGF
HSV−1−pUS7
AAV9−VP1
HSV−2−pUS8
HSV−2−pUL10
HSV−2−pUS4
PRV−Becker−gE
ConA
HIV−gp120−TH91125−7
APP
TMEV−VP1−U33047.1
BoNT−A−Hc
SA
HSV−1−pUS8
MV−F−CAM−70
MV−F−Edmonston
MV−F−Halle
HRP
HSV−1−pUL19
PV3−VP1
BoNT−C−Hc
GDNF
H1N1−HA
HSV−1−pUL1
HSV−1−pUL35
TBEV−E−4387−B7
TBEV−E−80
HSV−1−pUL44
HIV−tat−GR17
NT4
TBEV−E−Simo−9
BoNT−G−Hc
NT3
RV−G−N2c
HSV−1−pUL53
HSV−2−pUL53
BDV−Nq−98
RV−G−B2c
TMEV−VP−DQ401688.1
H5N1−HA
PRV−TJ−gE
TMEV−VP2
HSV−1−pUL10
VSV−Indiana−G
JEV−E−KE−093
BDV−G
JEV−E−SA14−5−3
JEV−E−ThCMAr4492
HSV−2−pUS6
BDV−CRNP5
JEV−E−BN19
VSV−Orsay−G
CHIKV−Sp−1
LCMV−G
RV−G−Mocala
WNV−E−AZ10−918
BDNF
WGA
BoNT−D−Hc
TMEV−VP3
mRNA_30cpc_pNeuron
mRNA_3cpc_pNeuron
mRNA_30cpc_HEK293T
mRNA_3cpc_HEK293T
20
15
10
5
0

## Selection of top ten fragments per sample

```
setkeyv(select.samples.binPos,c("Group","BCcountNanim","AnimalCount","NormCount"))
setorder(select.samples.binPos,Group,-BCcountanim,-AnimalCount,-BCcount,-NormCount)
setkey(select.samples.binPos,Group)

select.samples.topTwenty <- select.samples.binPos[, head(.SD, 20), by=Group]
select.samples.topTwenty[,c("totBC","seqlength","BCcountN","GeneAA","BCcountNseq"):=NULL]

for (thisGroup in unique(select.samples.topTwenty$Group)){
  out <- select.samples.topTwenty[J(thisGroup)]
  out[,Group:=NULL]
  out[,NormCount:=round(NormCount,digits = 0)]
  setnames(out,c("GeneName","AnimalCount","mismatches","BCcount","NormCount","BCcountNanim"),
           c(thisGroup,"Animal","missM","BCs","nCount","BCsNan"))

  print(knitr::kable(out, format = "latex", booktabs = T) %>%   kable_styling(latex_options = c("striped", "
}
```

```
plotPos <- function(select.samples.table,plot.col,sample.select){
  setkeyv(select.samples.table,"Group")
  select.samples.select <- select.samples.table[J(sample.select)]
  eval(parse(text=paste("setorder(select.samples.select,Group, -", plot.col,",-AnimalCount,-NormCount)", sep
  select.samples.topTen <- select.samples.select[, head(.SD, 10), by=Group]
  select.samples.out <- select.samples.select[select.samples.select$GeneAA %in% select.samples.topTen$GeneAA
  select.samples.out <- acast(select.samples.out, GeneAA~Group, value.var=plot.col) #Utilizes reshape 2 to r
  select.samples.out[is.na(select.samples.out)] <- 0
  select.samples.out <- select.samples.out[,sample.select]
  return(pheatmap(select.samples.out, cluster_rows=TRUE, show_rownames=TRUE, cluster_cols=FALSE))
}

plotPos(select.samples.binPos,"NormCount",c("mRNA_30cpc_pNeuron","mRNA_3cpc_pNeuron","mRNA_30cpc_HEK293T","m
```

```
##  setting  value
##  version  R version 3.3.2 (2016-10-31)
##  system   x86_64, linux-gnu
##  ui       X11
##  language (EN)
##  collate  en_US.UTF-8
##  tz       <NA>
##  date     2017-09-27
```

```
## Packages -------------------------------------------------------------------
```

```
##  package               * version  date       source
##  acepack                 1.4.1    2016-10-29 CRAN (R 3.3.2)
##  annotate                1.52.1   2017-09-13 Bioconductor
##  AnnotationDbi           1.36.2   2017-09-13 Bioconductor
##  AnnotationHub           2.6.5    2017-09-13 Bioconductor
##  backports               1.0.5    2017-01-18 url
##  base64enc               0.1-3    2015-07-28 cran (@0.1-3)
##  Biobase               * 2.34.0   2017-09-13 Bioconductor
##  BiocGenerics          * 0.20.0   2017-09-13 Bioconductor
##  BiocInstaller           1.24.0   2017-09-13 Bioconductor
##  BiocParallel            1.8.2    2017-09-13 Bioconductor
##  biomaRt                 2.30.0   2017-09-13 Bioconductor
##  Biostrings            * 2.42.1   2017-09-13 Bioconductor
##  biovizBase              1.22.0   2017-09-13 Bioconductor
##  bitops                  1.0-6    2013-08-17 CRAN (R 3.3.2)
##  BSgenome                1.42.0   2017-09-13 Bioconductor
##  checkmate               1.8.2    2016-11-02 cran (@1.8.2)
##  cluster                 2.0.5    2016-10-08 url
##  codetools               0.2-15   2016-10-05 CRAN (R 3.3.2)
##  colorspace              1.3-2    2016-12-14 CRAN (R 3.3.2)
##  data.table            * 1.10.4   2017-02-01 CRAN (R 3.3.2)
##  DBI                     0.5-1    2016-09-10 CRAN (R 3.3.2)
##  DESeq2                * 1.14.1   2017-09-13 Bioconductor
##  devtools              * 1.12.0   2016-12-05 CRAN (R 3.3.2)
##  dichromat               2.0-0    2013-01-24 CRAN (R 3.3.2)
##  digest                  0.6.12   2017-01-27 CRAN (R 3.3.2)
##  doParallel            * 1.0.10   2015-10-14 CRAN (R 3.3.2)
##  ensembldb               1.6.2    2017-09-13 Bioconductor
##  evaluate                0.10     2016-10-11 url
##  foreach               * 1.4.3    2015-10-13 CRAN (R 3.3.2)
##  foreign                 0.8-67   2016-09-13 url
##  formatR               * 1.4      2016-05-09 url
##  Formula                 1.2-1    2015-04-07 url
##  genefilter              1.56.0   2017-09-13 Bioconductor
##  geneplotter             1.52.0   2017-09-13 Bioconductor
##  GenomeInfoDb          * 1.10.3   2017-09-13 Bioconductor
##  GenomicAlignments     * 1.10.1   2017-09-13 Bioconductor
##  GenomicFeatures         1.26.4   2017-09-13 Bioconductor
##  GenomicRanges         * 1.26.4   2017-09-13 Bioconductor
##  GGally                  1.3.0    2016-11-13 CRAN (R 3.3.2)
##  ggbio                 * 1.22.4   2017-09-13 Bioconductor
##  ggplot2               * 2.2.1    2016-12-30 CRAN (R 3.3.2)
##  graph                   1.52.0   2017-09-13 Bioconductor
##  gridExtra               2.2.1    2016-02-29 url
##  gtable                  0.2.0    2016-02-26 CRAN (R 3.3.2)
##  Hmisc                   4.0-2    2016-12-31 url
##  hms                     0.3      2016-11-22 CRAN (R 3.3.2)
```

```
##  htmlTable               1.9      2017-01-26 cran (@1.9)
##  htmltools               0.3.5    2016-03-21 url
##  htmlwidgets             0.8      2016-11-09 cran (@0.8)
##  httpuv                  1.3.3    2015-08-04 CRAN (R 3.3.2)
##  httr                    1.2.1    2016-07-03 CRAN (R 3.3.2)
##  interactiveDisplayBase  1.12.0   2017-09-13 Bioconductor
##  IRanges               * 2.8.2    2017-09-13 Bioconductor
##  iterators             * 1.0.8    2015-10-13 CRAN (R 3.3.2)
##  kableExtra            * 0.5.2    2017-09-15 CRAN (R 3.3.2)
##  knitr                 * 1.15.1   2016-11-22 url
##  lattice                 0.20-34  2016-09-06 url
##  latticeExtra            0.6-28   2016-02-09 CRAN (R 3.3.2)
##  lazyeval                0.2.0    2016-06-12 CRAN (R 3.3.2)
##  locfit                  1.5-9.1  2013-04-20 CRAN (R 3.3.2)
##  magrittr                1.5      2014-11-22 CRAN (R 3.3.2)
##  Matrix                  1.2-8    2017-01-20 url
##  memoise                 1.0.0    2016-01-29 url
##  mime                    0.5      2016-07-07 CRAN (R 3.3.2)
##  munsell                 0.4.3    2016-02-13 CRAN (R 3.3.2)
##  nnet                    7.3-12   2016-02-02 CRAN (R 3.3.2)
##  OrganismDbi             1.16.0   2017-09-13 Bioconductor
##  pheatmap              * 1.0.8    2015-12-11 CRAN (R 3.3.2)
##  plyr                  * 1.8.4    2016-06-08 CRAN (R 3.3.2)
##  R6                      2.2.0    2016-10-05 CRAN (R 3.3.2)
##  RBGL                    1.50.0   2017-09-13 Bioconductor
##  RColorBrewer            1.1-2    2014-12-07 CRAN (R 3.3.2)
##  Rcpp                    0.12.9   2017-01-14 url
##  RCurl                   1.95-4.8 2016-03-01 CRAN (R 3.3.2)
##  readr                   1.1.1    2017-05-16 CRAN (R 3.3.2)
##  reshape                 0.8.6    2016-10-21 CRAN (R 3.3.2)
##  reshape2              * 1.4.2    2016-10-22 CRAN (R 3.3.2)
##  rlang                   0.1.2    2017-08-09 CRAN (R 3.3.2)
##  rmarkdown               1.6      2017-06-15 CRAN (R 3.3.2)
##  rpart                   4.1-10   2015-06-29 url
##  rprojroot               1.2      2017-01-16 cran (@1.2)
##  Rsamtools             * 1.26.2   2017-09-13 Bioconductor
##  RSQLite                 1.1-2    2017-01-08 CRAN (R 3.3.2)
##  rtracklayer             1.34.2   2017-09-13 Bioconductor
##  rvest                   0.3.2    2016-06-17 CRAN (R 3.3.2)
##  S4Vectors             * 0.12.2   2017-09-13 Bioconductor
##  scales                  0.4.1    2016-11-09 url
##  shiny                   1.0.0    2017-01-12 CRAN (R 3.3.2)
##  stringi                 1.1.2    2016-10-01 url
##  stringr                 1.2.0    2017-02-18 CRAN (R 3.3.2)
##  SummarizedExperiment  * 1.4.0    2017-09-13 Bioconductor
##  survival                2.40-1   2016-10-30 url
##  tibble                  1.3.4    2017-08-22 CRAN (R 3.3.2)
##  VariantAnnotation       1.20.3   2017-09-13 Bioconductor
##  withr                   1.0.2    2016-06-20 url
##  XML                     3.98-1.5 2016-11-10 CRAN (R 3.3.2)
##  xml2                    1.1.1    2017-01-24 CRAN (R 3.3.2)
##  xtable                  1.8-2    2016-02-05 CRAN (R 3.3.2)
##  XVector               * 0.14.1   2017-09-13 Bioconductor
##  yaml                    2.1.14   2016-11-12 CRAN (R 3.3.2)
##  zlibbioc                1.20.0   2017-09-13 Bioconductor
```

# Analysis plots for AAV plasmid library and coverage

*Tomas Bjorklund*

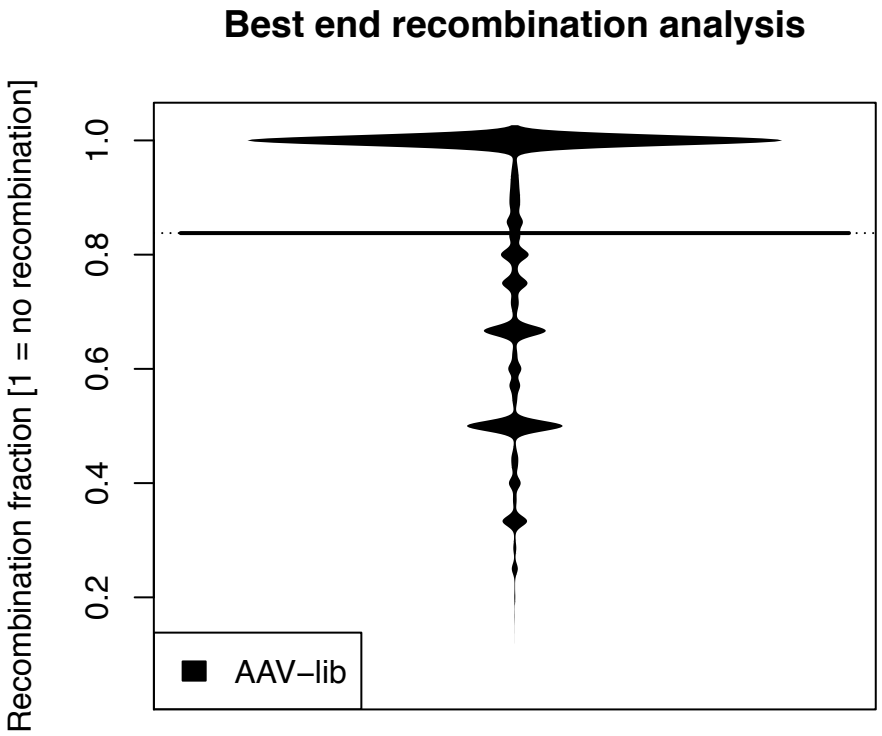
*Thu Nov 2 14:44:12 2017*

This script provides a number of measurements on the AAV plasmid library and the readouts in vitro & in vivo.

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

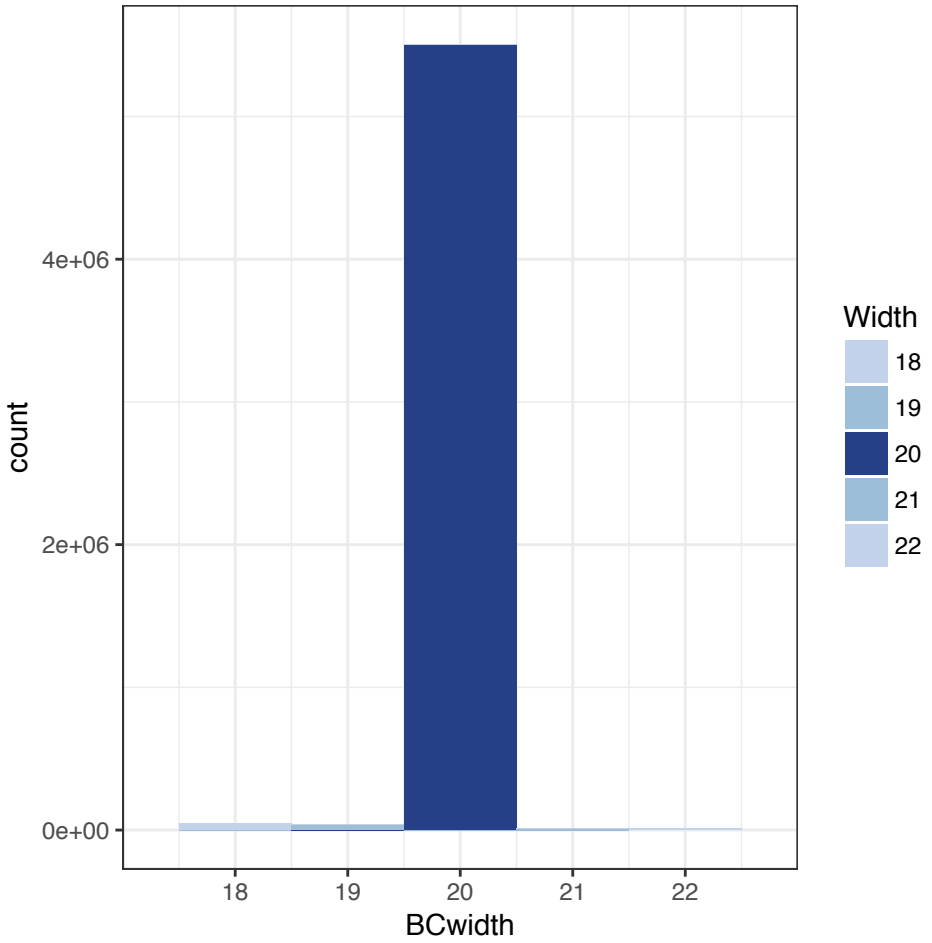
## Generation plots for library purity

```
complete.ranges <- readRDS("output/completeLibraryRanges.rds")
purity.table <- data.table(mcols(complete.ranges)$mCount/mcols(complete.ranges)$tCount)
purity.table$BCwidth <- width(mcols(complete.ranges)$BC)

beanplot(data = purity.table$V1, ll = 0.04, what = c(1, 1, 1, 0), bw = "nrd0",
    log = "", main = "Best end recombination analysis", ylab = "Recombination fraction [1 = no recombinatio
    border = NA, col = list("black", c("grey", "white")))
legend("bottomleft", fill = c("black"), legend = c("AAV-lib"))
```




```
fill.values <- c(A = rgb(193, 210, 234, maxColorValue = 255), B = rgb(157, 190,
    217, maxColorValue = 255), c = rgb(38, 64, 135, maxColorValue = 255), D = rgb(157,
    190, 217, maxColorValue = 255), E = rgb(193, 210, 234, maxColorValue = 255))
```

```
names(fill.values) <- c("18", "19", "20", "21", "22")

ggplot(purity.table, aes(x = BCwidth, fill = as.character(BCwidth))) + geom_histogram(binwidth = 1,
    boundary = 0.5) + theme_bw() + scale_fill_manual(name = "Width", values = fill.values) +
    scale_x_continuous(limit = c(17, 23), breaks = c(seq(18, 22, 1)), expand = c(0,
        0))
```



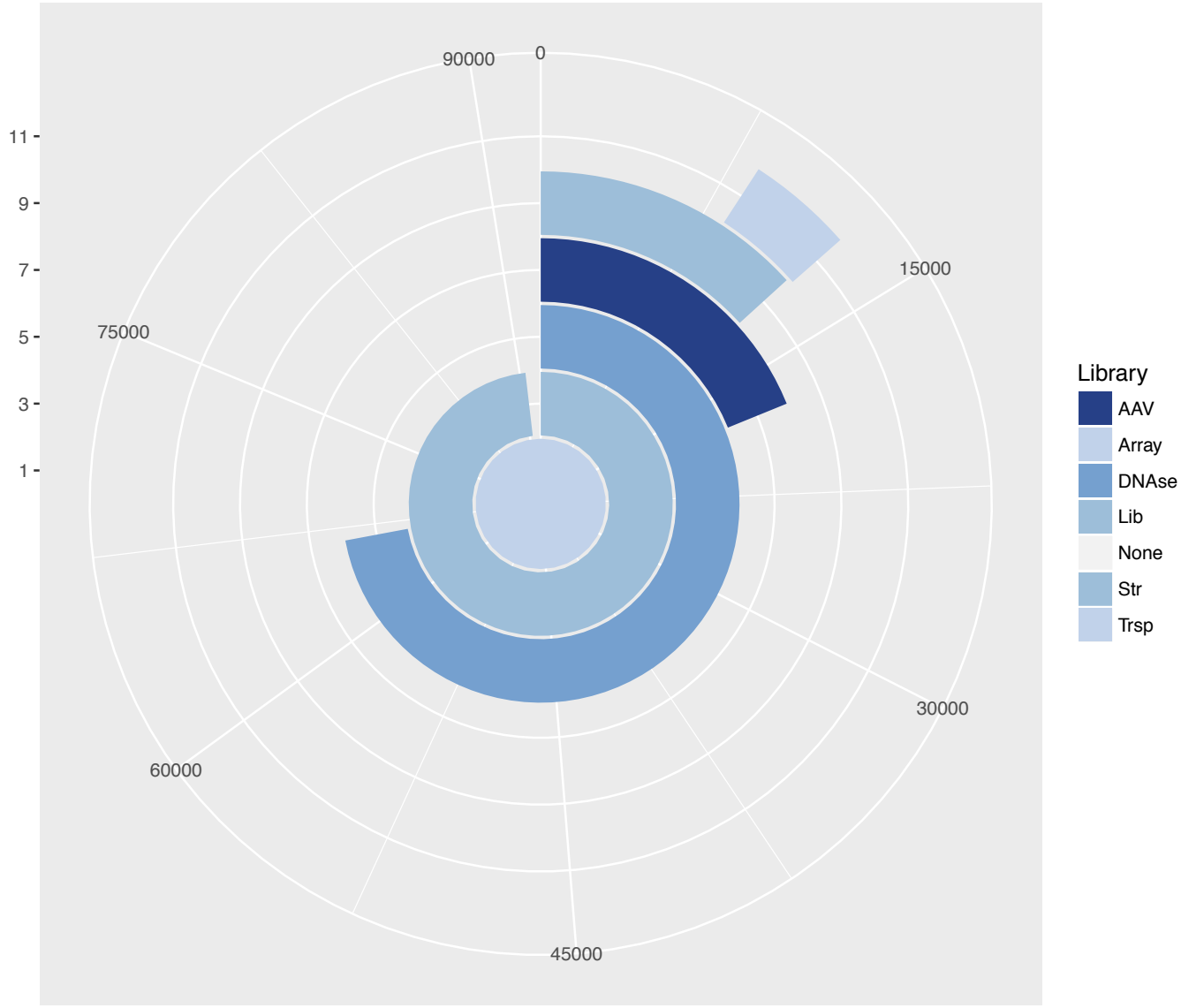

```
opts_chunk$set(fig.width = 8, fig.height = 8)
```

## Plot Venn diagrams of fragments

```
load("data/LUTdna.rda")
complete.library <- readRDS("data/allSamplesDataTable.RDS")
setkey(complete.library, Group)
complete.library <- complete.library[-grep("4wks", Group)]
seq.arry <- LUT.dna$LUTnr
seq.lib <- unique(complete.library[J("DNA_pscAAVlib")]$LUTnr)
seq.AAV <- unique(complete.library[J("mRNA_All")]$LUTnr)
seq.DNAse <- unique(complete.library[grep("DNA_AAVlib_DNAse", Group)]$LUTnr)
seq.str <- unique(complete.library[grep("Str", Group)]$LUTnr)
seq.Trsp <- unique(complete.library[grep("SN|Ctx|Th", Group)]$LUTnr)

venn.area1 <- length(seq.arry)
venn.area2 <- length(seq.lib)
```

```
venn.area3 <- length(seq.DNAse)
venn.area4 <- length(seq.AAV)
venn.area5 <- length(seq.str)
venn.area6 <- length(seq.Trsp)


isect.Str_Trsp <- length(intersect(seq.str, seq.Trsp))

venn.n12 <- length(intersect(seq.arry, seq.lib))
venn.n23 <- length(intersect(seq.lib, seq.DNAse))

venn.n13 <- length(intersect(seq.arry, seq.DNAse))
venn.n123 <- length(intersect(intersect(seq.arry, seq.lib), seq.DNAse))


output.table <- data.frame(NameArray = character(), NameLib = character(), NameDNAse = character(),
    NameAAV = character(), NameStr = character(), NameTrsp = character(), ArrayStart = numeric(),
    ArrayEnd = numeric(), LibStart = numeric(), LibEnd = numeric(), DNAseStart = numeric(),
    DNAseEnd = numeric(), AAVStart = numeric(), AAVend = numeric(), StrStart = numeric(),
    StrEnd = numeric(), TrspStart = numeric(), TrspEnd = numeric(), stringsAsFactors = FALSE)
output.table[1:3, 1] <- c("Array", "None", "None")
output.table[1:3, 2] <- c("Lib", "None", "None")
output.table[1:3, 3] <- c("DNAse", "None", "None")
output.table[1:3, 4] <- c("AAV", "None", "None")
output.table[1:3, 5] <- c("Str", "None", "None")
output.table[1:3, 6] <- c("None", "Trsp", "None")
output.table[1:3, 7:18] <- 0
output.table$ArrayEnd[1] <- output.table$LibEnd[2] <- output.table$DNAseEnd[2] <- output.table$AAVend[2] <-
output.table$LibStart[2] <- output.table$LibEnd[1] <- length(intersect(seq.arry,
    seq.lib))
output.table$DNAseStart[2] <- output.table$DNAseEnd[1] <- length(intersect(seq.lib,
    seq.DNAse))
output.table$AAVStart[2] <- output.table$AAVend[1] <- length(intersect(seq.lib,
    seq.AAV))
output.table$StrStart[2] <- output.table$StrEnd[1] <- length(intersect(seq.AAV,
    seq.str))
output.table$TrspStart[2] <- output.table$TrspEnd[1] <- length(seq.str) - length(intersect(seq.str,
    seq.Trsp))
output.table$TrspStart[3] <- output.table$TrspEnd[2] <- output.table$TrspStart[2] +
    length(seq.Trsp)


fill.values <- c(Array = rgb(193, 210, 234, maxColorValue = 255), Lib = rgb(157,
    190, 217, maxColorValue = 255), DNAse = rgb(117, 160, 207, maxColorValue = 255),
    AAV = rgb(38, 64, 135, maxColorValue = 255), Str = rgb(157, 190, 217, maxColorValue = 255),
    Trsp = rgb(193, 210, 234, maxColorValue = 255), None = rgb(255, 255, 255,
        maxColorValue = 255, alpha = 0))


ggplot(output.table) + scale_x_continuous(limit = c(0, 12), breaks = c(seq(1,
    11, 2)), expand = c(0, 0)) + scale_y_continuous(breaks = c(seq(0, 90000,
    15000))) + scale_fill_manual(name = "Library", values = fill.values) + theme(aspect.ratio = 1) +
    geom_rect(data = output.table, aes(fill = NameTrsp, ymax = TrspEnd, ymin = TrspStart,
        xmax = 11.95, xmin = 10.05)) + geom_rect(data = output.table, aes(fill = NameStr,
    ymax = StrEnd, ymin = StrStart, xmax = 9.95, xmin = 8.05)) + geom_rect(data = output.table,
    aes(fill = NameAAV, ymax = AAVend, ymin = AAVStart, xmax = 7.95, xmin = 6.05)) +
    geom_rect(data = output.table, aes(fill = NameDNAse, ymax = DNAseEnd, ymin = DNAseStart,
```

```
    xmax = 5.95, xmin = 4.05)) + geom_rect(data = output.table, aes(fill = NameLib,
ymax = LibEnd, ymin = LibStart, xmax = 3.95, xmin = 2.05)) + geom_rect(data = output.table,
aes(fill = NameArray, ymax = ArrayEnd, ymin = ArrayStart, xmax = 1.95, xmin = 0)) +
coord_polar(theta = "y")
```



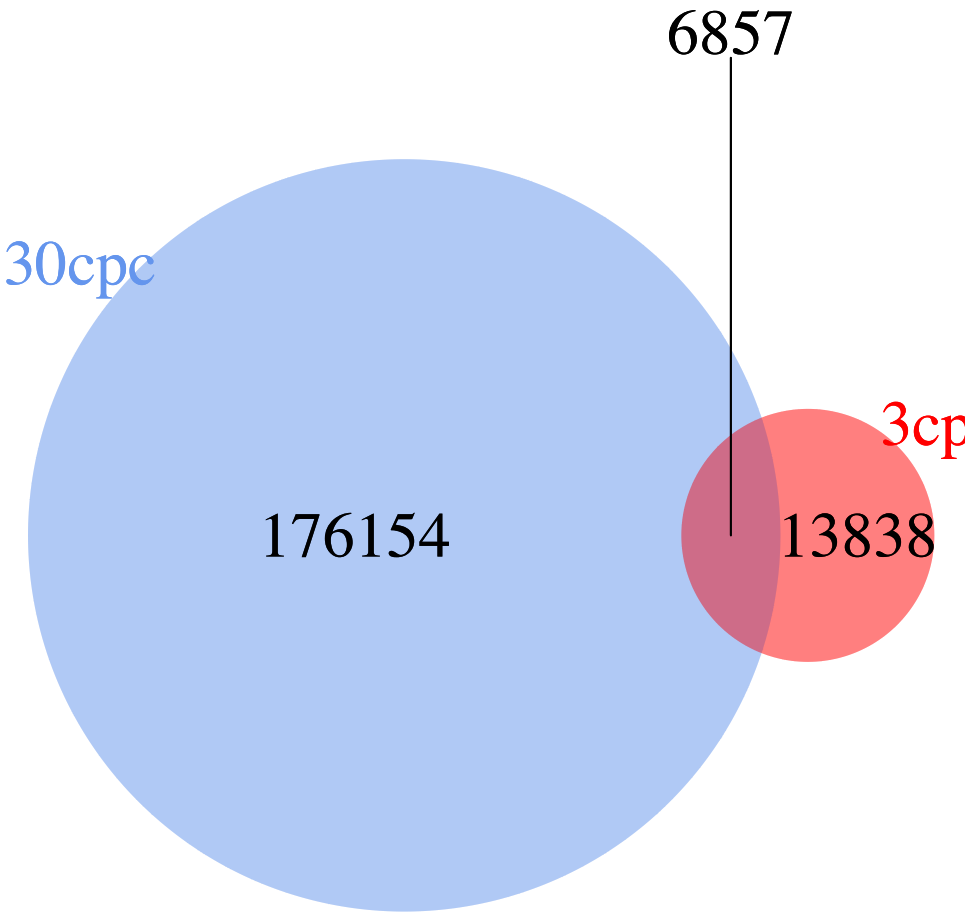

```
opts_chunk$set(fig.width = 5, fig.height = 5)
```

```
knitr::kable(output.table, format = "latex", booktabs = T) %>% kable_styling(latex_options = c("striped",
    "scale_down", "repeat_header")) %>% landscape()
```

| NameArray | NameLib | NameDNAse | NameAAV | NameStr | NameTrsp | ArrayStart | ArrayEnd | LibStart | LibEnd | DNAseStart | DNAseEnd | AAVStart | AAVend | StrStart | StrEnd | TrspStart | TrspEnd |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Array | Lib | DNAse | AAV | Str | None | 0 | 92343 | 0 | 90635 | 0 | 66531 | 0 | 17401 | 0 | 12240 | 0 | 8483 |
| None | None | None | None | None | Trsp | 0 | 0 | 90635 | 92343 | 66531 | 92343 | 17401 | 92343 | 12240 | 92343 | 8483 | 12474 |
| None | None | None | None | None | None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12474 | 92343 |

## Barcode Venn diagrams for 30cpc and 3cpc DNAse resistant libraries

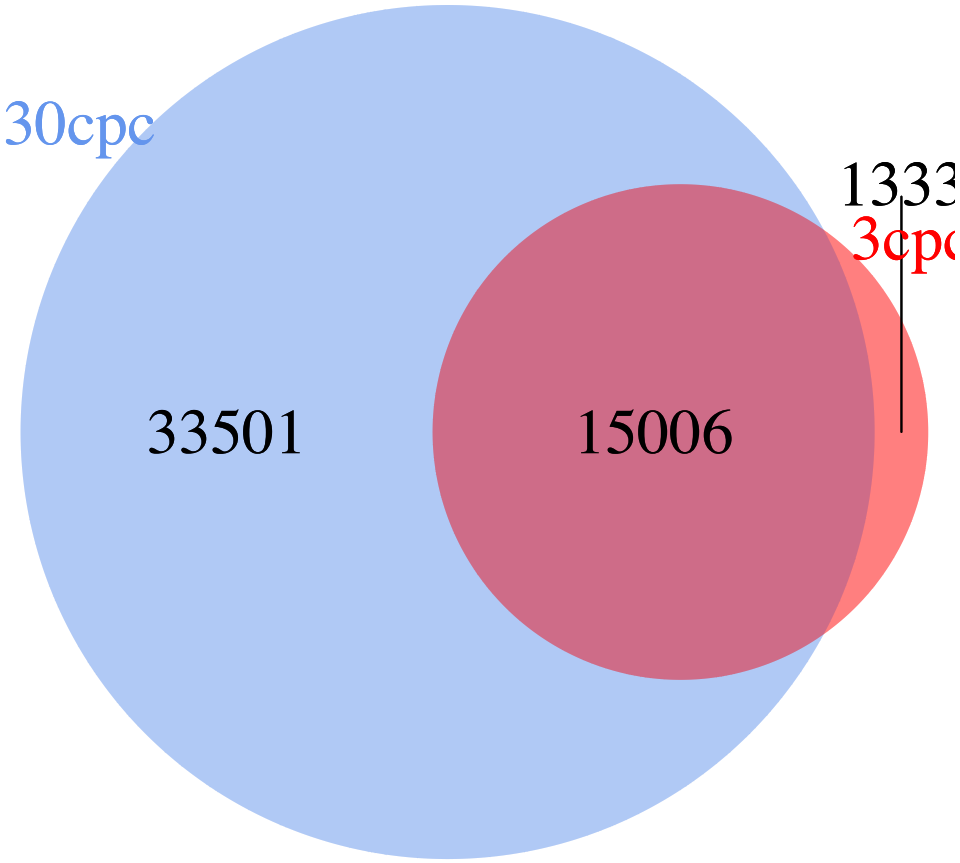
```
total.30cpc <- unique(complete.library[grep("DNA_AAVlib_DNAse_30cpc", Group)]$BC)
total.3cpc <- unique(complete.library[grep("DNA_AAVlib_DNAse_3cpc", Group)]$BC)

total.30cpc <- names(table(strsplit(paste(total.30cpc, collapse = ","), ",")))
total.3cpc <- names(table(strsplit(paste(total.3cpc, collapse = ","), ",")))

venn.area1 <- length(total.30cpc)
venn.area2 <- length(total.3cpc)

venn.n12 <- length(intersect(total.30cpc, total.3cpc))

venn.colors <- c("cornflower blue", "red")
grid.newpage()
venn.plot <- draw.pairwise.venn(area1 = venn.area1, area2 = venn.area2, cross.area = venn.n12,
    scaled = TRUE, fill = venn.colors, alpha = 0.3, lty = "blank", cex = 2,
    cat.cex = 2, cat.col = venn.colors, category = c("30cpc", "3cpc"))
grid.draw(venn.plot)
```




## Fragment Venn diagrams for 30cpc and 3cpc DNAse resistant libraries

```
total.30cpc <- unique(complete.library[grep("DNA_AAVlib_DNAse_30cpc", Group)]$Sequence)
total.3cpc <- unique(complete.library[grep("DNA_AAVlib_DNAse_3cpc", Group)]$Sequence)
```

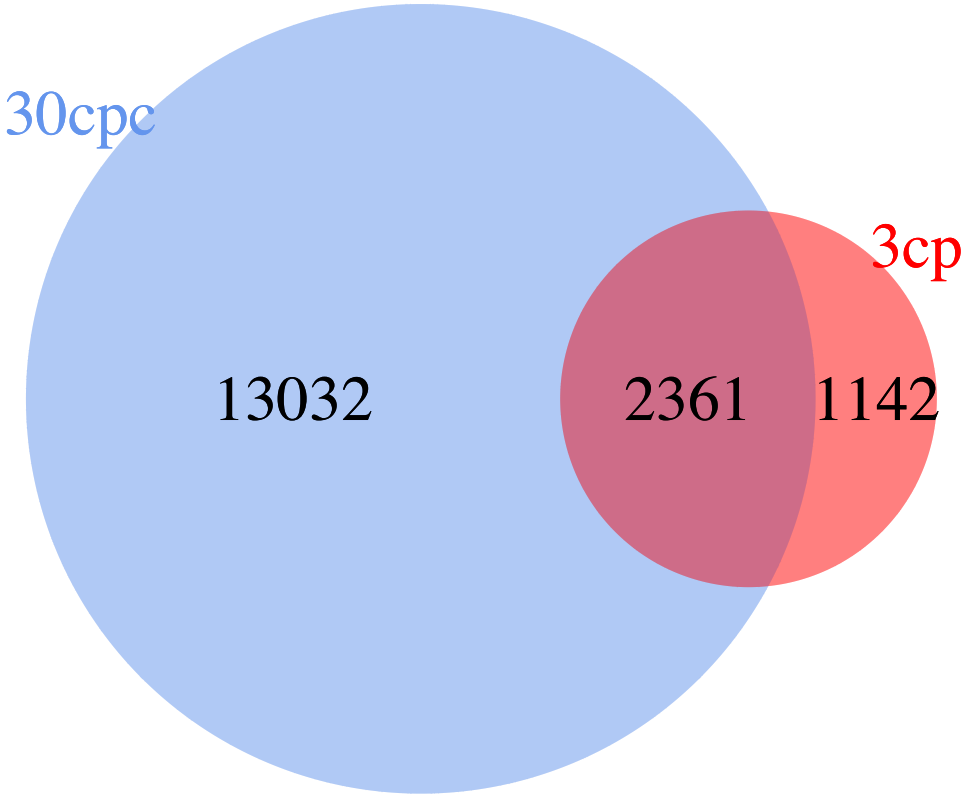
```
venn.area1 <- length(total.30cpc)
venn.area2 <- length(total.3cpc)

venn.n12 <- length(intersect(total.30cpc, total.3cpc))


grid.newpage()
venn.plot <- draw.pairwise.venn(area1 = venn.area1, area2 = venn.area2, cross.area = venn.n12,
    scaled = TRUE, fill = venn.colors, alpha = 0.3, lty = "blank", cex = 2,
    cat.cex = 2, cat.col = venn.colors, category = c("30cpc", "3cpc"))
grid.draw(venn.plot)
```




## Barcode Venn diagrams for 30cpc and 3cpc infective libraries

```
RNA.library <- complete.library[-grep("DNAse", Group)]
total.30cpc <- unique(RNA.library[grep("30cpc", Group)]$BC)
total.3cpc <- unique(RNA.library[grep("3cpc", Group)]$BC)

total.30cpc <- names(table(strsplit(paste(total.30cpc, collapse = ","), ",")))
total.3cpc <- names(table(strsplit(paste(total.3cpc, collapse = ","), ",")))

venn.area1 <- length(total.30cpc)
```

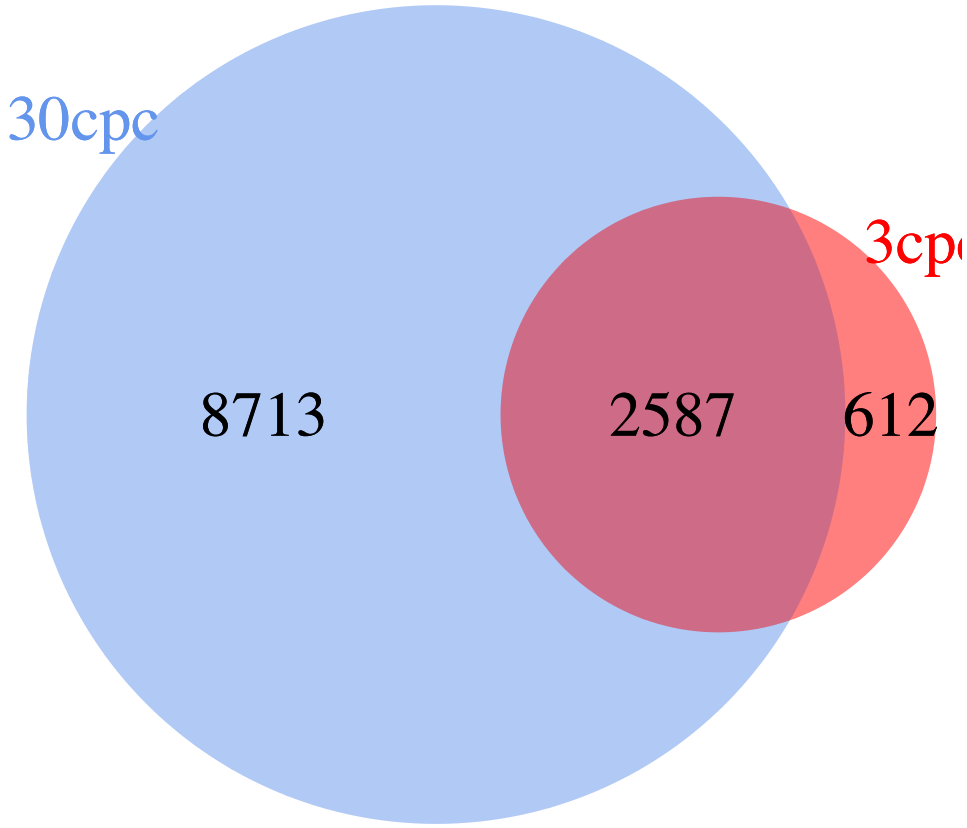
```
venn.area2 <- length(total.3cpc)

venn.n12 <- length(intersect(total.30cpc, total.3cpc))

venn.colors <- c("cornflower blue", "red")
grid.newpage()
venn.plot <- draw.pairwise.venn(area1 = venn.area1, area2 = venn.area2, cross.area = venn.n12,
    scaled = TRUE, fill = venn.colors, alpha = 0.3, lty = "blank", cex = 2,
    cat.cex = 2, cat.col = venn.colors, category = c("30cpc", "3cpc"))
grid.draw(venn.plot)
```




## Fragment Venn diagrams for 30cpc and 3cpc infective libraries

```
total.30cpc <- unique(RNA.library[grep("30cpc", Group)]$Sequence)
total.3cpc <- unique(RNA.library[grep("3cpc", Group)]$Sequence)


venn.area1 <- length(total.30cpc)
venn.area2 <- length(total.3cpc)

venn.n12 <- length(intersect(total.30cpc, total.3cpc))


grid.newpage()
```

```
venn.plot <- draw.pairwise.venn(area1 = venn.area1, area2 = venn.area2, cross.area = venn.n12,
    scaled = TRUE, fill = venn.colors, alpha = 0.3, lty = "blank", cex = 2,
    cat.cex = 2, cat.col = venn.colors, category = c("30cpc", "3cpc"))
grid.draw(venn.plot)
```



# Top 23 connectivity analysis output

*Tomas Bjorklund*

*Wed Nov 1 19:50:32 2017*

This script presents top 23 candidates as connectivity graph.

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

## Selection of relevant samples

```
select.samples <- readRDS("data/allSamplesDataTable.RDS")
select.samples <- select.samples[-grep("4wks|mRNA_3cpc_pNeuron_RNA", select.samples$Group),
    ]

select.samples.merge <- data.table::copy(select.samples)
select.samples.merge$Lib <- "30cpc"
select.samples.merge$Lib[grep("3cpc", select.samples.merge$Group)] <- "3cpc"
select.samples.merge[, `:=`(Group, gsub("_30cpc|_3cpc", "", Group))]

setkey(select.samples.merge, Group)
select.samples.merge.binPos <- select.samples.merge[c("mRNA_Str", "mRNA_Th",
    "mRNA_Ctx", "mRNA_SN")]
```

## Summarize data over each aa

```
setorder(select.samples.merge.binPos, Group, GeneName, start, width)
winWidth = 1
windowTable <- select.samples.merge.binPos[, 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)]
setkeyv(windowTable, c("GeneName", "start", "width"))
setkeyv(select.samples.merge.binPos, c("GeneName", "start", "width"))
select.samples.windowBin <- select.samples.merge.binPos[windowTable, nomatch = 0,
    allow.cartesian = TRUE]

setkeyv(select.samples.windowBin, c("Group", "GeneName", "winStart", "winEnd"))
select.samples.windowBin.out <- select.samples.windowBin[, list(Overlaps = .N,
    BCcount = length(table(strsplit(paste(t(BC), collapse = ","), ","))), NormCount = mean(log2(RNAcount +
        1)), AnimalCount = length(table(strsplit(paste(t(Animals), collapse = ","),
        ","))), LUTnrs = paste(unique(names(table(strsplit(paste(t(LUTnrs),
        collapse = ","), ",")))), collapse = ","), mainStruct = paste(unique(structure),
        collapse = ","), mainLibs = paste(unique(Lib), collapse = ","), libCount = length(unique(Lib)),
    mismatches = median(mismatches)), by = c("Group", "GeneName", "winStart",
    "winEnd", "seqlength")]
```

## Identify local maxima for each gene and sample

```
select.samples.windowBin <- data.table::copy(select.samples.windowBin.out)
select.samples.windowBin[, `:=`(Score, BCcount + AnimalCount + libCount)]
setorder(select.samples.windowBin, Group, -Score, -AnimalCount, -BCcount, -libCount,
    -NormCount, GeneName, winStart)
select.samples.windowBin[, `:=`(Rank, seq(.N)), by = c("Group")]
setorder(select.samples.windowBin, Group, GeneName, winStart, -Score)


windowTable <- select.samples.windowBin[, c("Group", "GeneName", "seqlength"),
    with = FALSE]
windowTable <- unique(windowTable, by = c("Group", "GeneName", "seqlength"))
windowTable <- windowTable[, seq(seqlength), by = c("Group", "GeneName")]
setnames(windowTable, "V1", "winStart")
setkeyv(windowTable, c("Group", "GeneName", "winStart"))
setkeyv(select.samples.windowBin, c("Group", "GeneName", "winStart"))
select.samples.windowBin <- select.samples.windowBin[windowTable]
select.samples.windowBin$Score[is.na(select.samples.windowBin$Score)] <- 0

select.samples.windowBin$LUTnrs[is.na(select.samples.windowBin$LUTnrs)] <- seq(length(which(is.na(select.sar

setorder(select.samples.windowBin, Group, GeneName, -winStart)
select.samples.windowBin.locMax <- data.table::copy(select.samples.windowBin)
select.samples.windowBin.locMax[, `:=`(Peak, as.logical(extract(turnpoints(Score),
    peak = 1, pit = 0))), by = c("Group", "GeneName")]  #,proba=0.0001

setorder(select.samples.windowBin.locMax, Group, GeneName, winStart)
select.samples.windowBin.locMax <- select.samples.windowBin.locMax[select.samples.windowBin.locMax$Peak,
    ]
select.samples.windowBin.locMax[, `:=`(Peak, NULL)]

setorder(select.samples.windowBin.locMax, Group, Rank, -Score, GeneName, winStart)
select.samples.windowBin.locMax <- unique(select.samples.windowBin.locMax, by = c("Group",
    "LUTnrs"))
```

## Make binning for 22aa fragment length

```
# Connect highest scoring fragments

setorder(select.samples.windowBin.locMax, Group, -Score, -AnimalCount, -BCcount,
    -libCount, -NormCount, GeneName, winStart)
setkey(select.samples.windowBin.locMax, Group)
select.samples.windowBin.locMax[, `:=`(Rank2, seq(.N)), by = c("Group")]


select.samples.topSelect <- select.samples.windowBin.locMax[, head(.SD, 35),
    by = Group]


setorder(select.samples.topSelect, Group, GeneName, winStart)
winWidth = 21
windowTable <- select.samples.topSelect[, c("GeneName", "winStart"), with = FALSE]
windowTable <- unique(windowTable, by = c("GeneName", "winStart"))
```

```
windowTable <- windowTable[, seq((winStart - winWidth), (winStart)), by = c("GeneName",
    "winStart")]
setnames(windowTable, "V1", "binBaseStart")
windowTable <- windowTable[binBaseStart > 0, ]
windowTable[, `:=`(binBaseEnd, binBaseStart + winWidth)]
scoreSelect <- select.samples.topSelect[, c("Group", "GeneName", "winStart",
    "Score"), with = FALSE]
setkeyv(windowTable, c("GeneName", "winStart"))
setkeyv(scoreSelect, c("GeneName", "winStart"))
scoreSelect <- scoreSelect[windowTable, allow.cartesian = TRUE]

scoreSelect[, `:=`(mCount, length(unique(Group))), by = c("GeneName", "binBaseStart")]
scoreSelect[, `:=`(oCount, .N), by = c("GeneName", "binBaseStart")]
scoreSelect[, `:=`(tCount, mCount + oCount)]
scoreSelect[, `:=`(offset, abs(binBaseStart + 6 - mean(winStart))), c("GeneName",
    "binBaseStart")]
# scoreSelect <- scoreSelect[mCount>oCount,]
setorder(scoreSelect, GeneName, winStart, -tCount, binBaseStart, -offset)

scoreSelect <- scoreSelect[, head(.SD, 1), by = c("GeneName", "winStart")]

scoreSelect <- scoreSelect[, c("GeneName", "winStart", "binBaseStart", "binBaseEnd"),
    with = FALSE]
setkeyv(scoreSelect, c("GeneName", "winStart"))
setkeyv(select.samples.windowBin.locMax, c("GeneName", "winStart"))


select.samples.windowBin.locMax.bin <- scoreSelect[select.samples.windowBin.locMax,
    nomatch = 0]
setorder(select.samples.windowBin.locMax.bin, GeneName, -binBaseStart, -Score,
    Group)

select.samples.windowBin.locMerge <- select.samples.windowBin.locMax.bin[, head(.SD,
    1), by = c("GeneName", "binBaseStart", "binBaseEnd", "Group")]
select.samples.windowBin.locMerge <- unique(select.samples.windowBin.locMerge,
    by = c("Group", "GeneName", "LUTnrs"))
```

## Selection of top twenty fragments per sample

```
setorder(select.samples.windowBin.locMerge, Group, -Score, -AnimalCount, -BCcount,
    -libCount, -NormCount, GeneName, binBaseStart)
setkey(select.samples.windowBin.locMerge, Group)

select.samples.topTwenty <- select.samples.windowBin.locMerge[, head(.SD, 25),
    by = Group]
select.samples.topTwenty <- select.samples.topTwenty[, c("GeneName", "binBaseStart",
    "binBaseEnd"), with = FALSE]
select.samples.topTwenty <- unique(select.samples.topTwenty, by = c("GeneName",
    "binBaseStart"))

setorder(select.samples.windowBin, Group, -Score, -AnimalCount, -BCcount, -libCount,
    -NormCount, GeneName, winStart)
select.samples.windowBin[, `:=`(Rank, seq(.N)), by = c("Group")]
setkeyv(scoreSelect, c("GeneName", "winStart"))
setkeyv(select.samples.windowBin, c("GeneName", "winStart"))
```

```
select.samples.windowBin <- scoreSelect[select.samples.windowBin, nomatch = 0]
setorder(select.samples.windowBin, Group, GeneName, binBaseStart, -Score)
select.samples.windowBin.Bin <- unique(select.samples.windowBin, by = c("Group",
    "GeneName", "binBaseStart"))

setkeyv(select.samples.topTwenty, c("GeneName", "binBaseStart"))
setkeyv(select.samples.windowBin, c("GeneName", "binBaseStart"))
select.samples.windowBin.allTop <- select.samples.windowBin[select.samples.topTwenty,
    nomatch = 0]
select.samples.windowBin.allTop <- unique(select.samples.windowBin.allTop, by = c("Group",
    "LUTnrs"))
setorder(select.samples.windowBin.allTop, Group, GeneName, binBaseStart, -Score)
select.samples.windowBin.allTop <- unique(select.samples.windowBin.allTop, by = c("Group",
    "GeneName", "binBaseStart"))

select.samples.windowBin.allTop[, `:=`(GeneAA, paste(GeneName, " [", binBaseStart +
    floor(winWidth/2), "]", sep = "")), by = Group]
setorder(select.samples.windowBin.allTop, Group, Rank)
```

## Plotting ranked order

```
setkey(select.samples.windowBin.allTop, Group)

v1 <- select.samples.windowBin.allTop["mRNA_Str"]$GeneAA
v2 <- select.samples.windowBin.allTop["mRNA_Th"]$GeneAA
v3 <- select.samples.windowBin.allTop["mRNA_Ctx"]$GeneAA
v4 <- select.samples.windowBin.allTop["mRNA_SN"]$GeneAA


o <- 0.05
DF <- data.table(x = c(rep(1, length(v1)), rep(2, length(v2)), rep(3, length(v3)),
    rep(4, length(v4))), x1 = c(rep(1, length(v1)), rep(2, length(v2)), rep(3,
    length(v3)), rep(4, length(v4))), y = c(rev(seq_along(v1)), rev(seq_along(v2)),
    rev(seq_along(v3)), rev(seq_along(v4))), g = c(v1, v2, v3, v4))
DF[, `:=`(groupMax, max(y)), by = x]
allMax <- max(DF$y)
DF[, `:=`(y, y - groupMax + allMax)]


ggplot(DF, aes(x = x, y = y, group = g, label = g)) + geom_path(aes(x = x1),
    size = 0.3, color = "blue") + geom_text(size = 1.8) + theme_minimal() +
    theme(axis.title = element_blank(), axis.text = element_blank(), axis.ticks = element_blank(),
        panel.grid = element_blank())
```

# Three sample pairwise sample analysis output

*Tomas Bjorklund*

*Wed Nov 1 19:56:43 2017*

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

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

## Generation of infective library

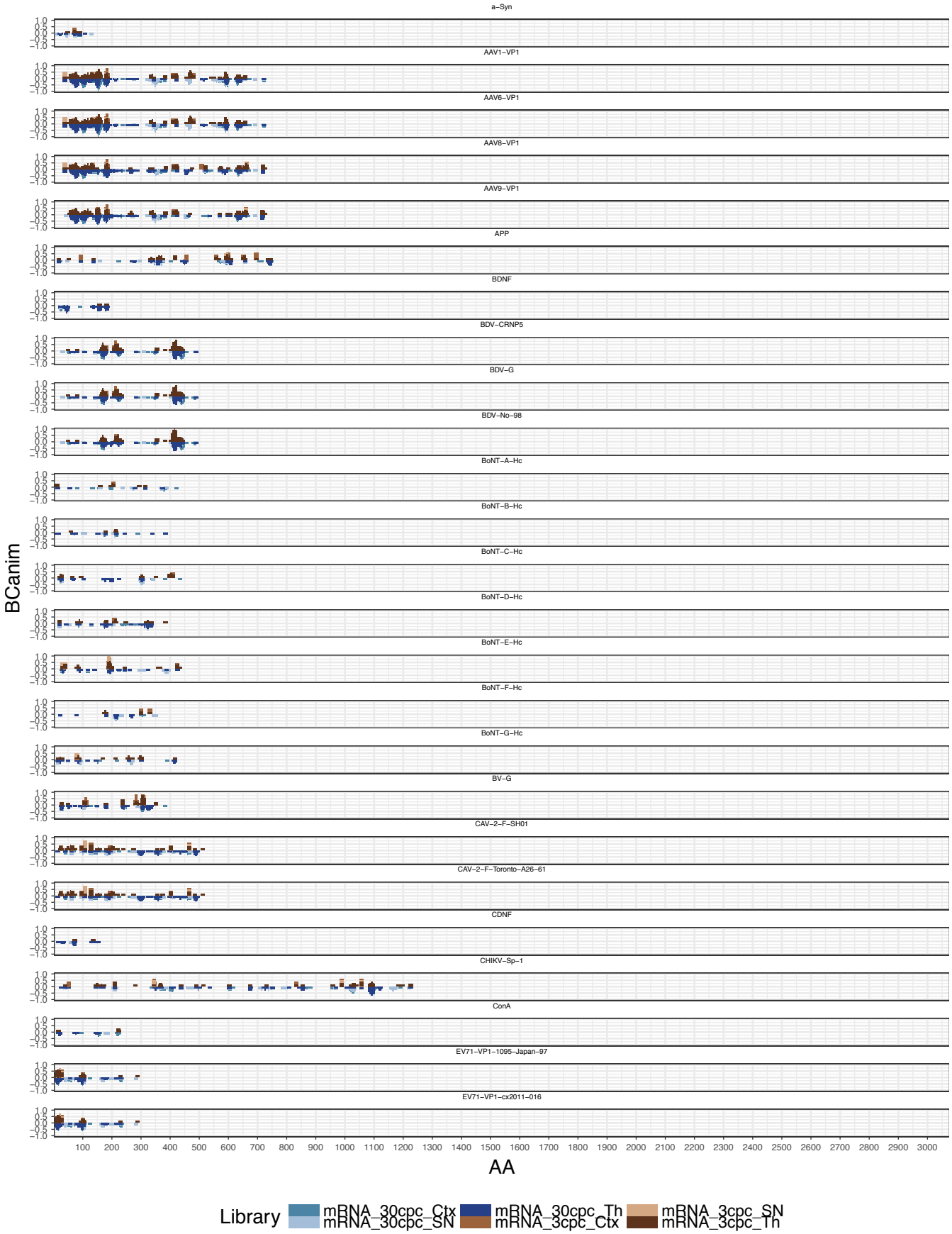
```
all.samples <- readRDS("data/allSamplesDataTable.RDS")

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)]
```

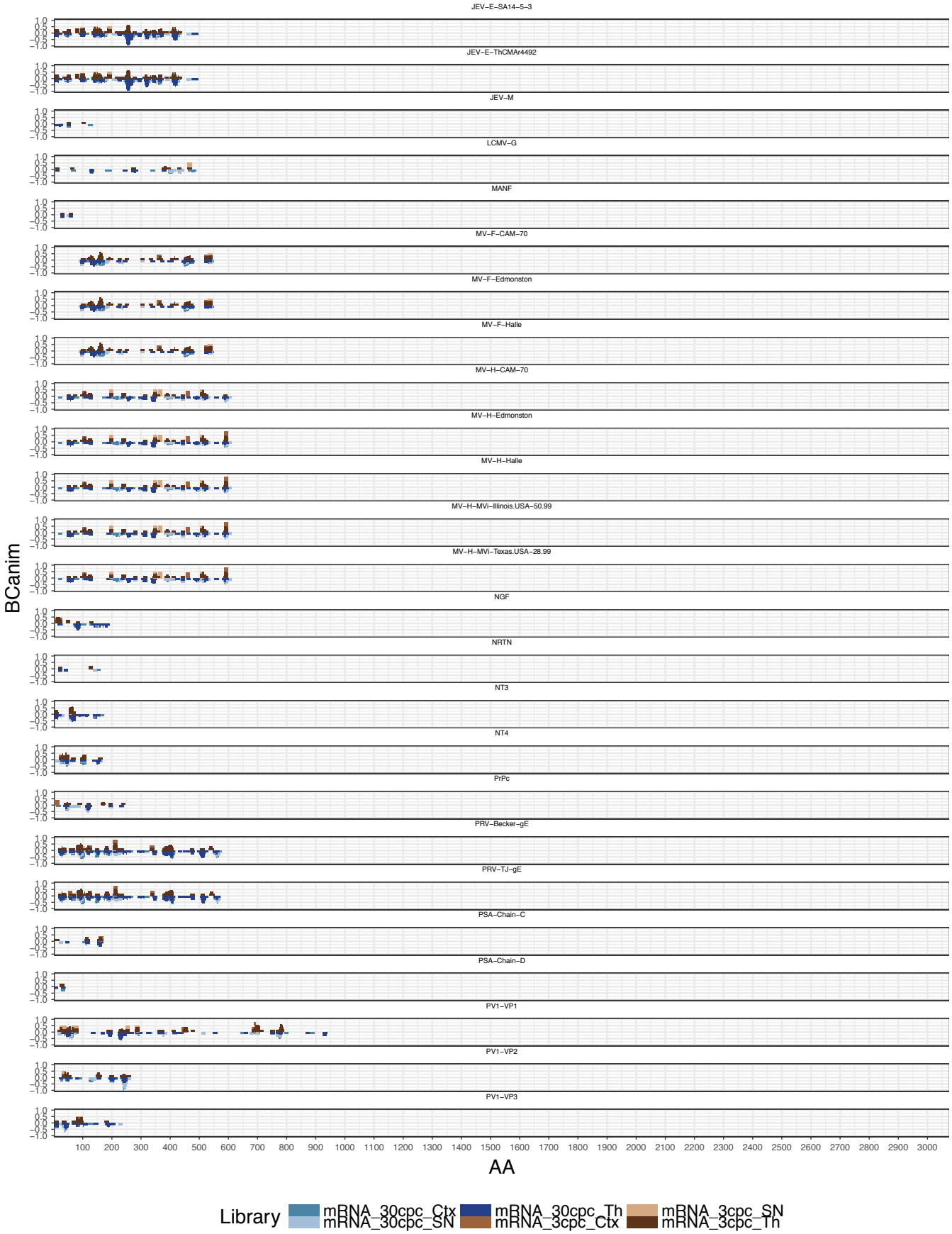
```
setkeyv(windowTable, c("GeneName", "start", "width"))
setkeyv(select.samples, c("GeneName", "start", "width"))
select.samples.windowBin <- select.samples[windowTable, allow.cartesian = TRUE]

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(B
    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)
```

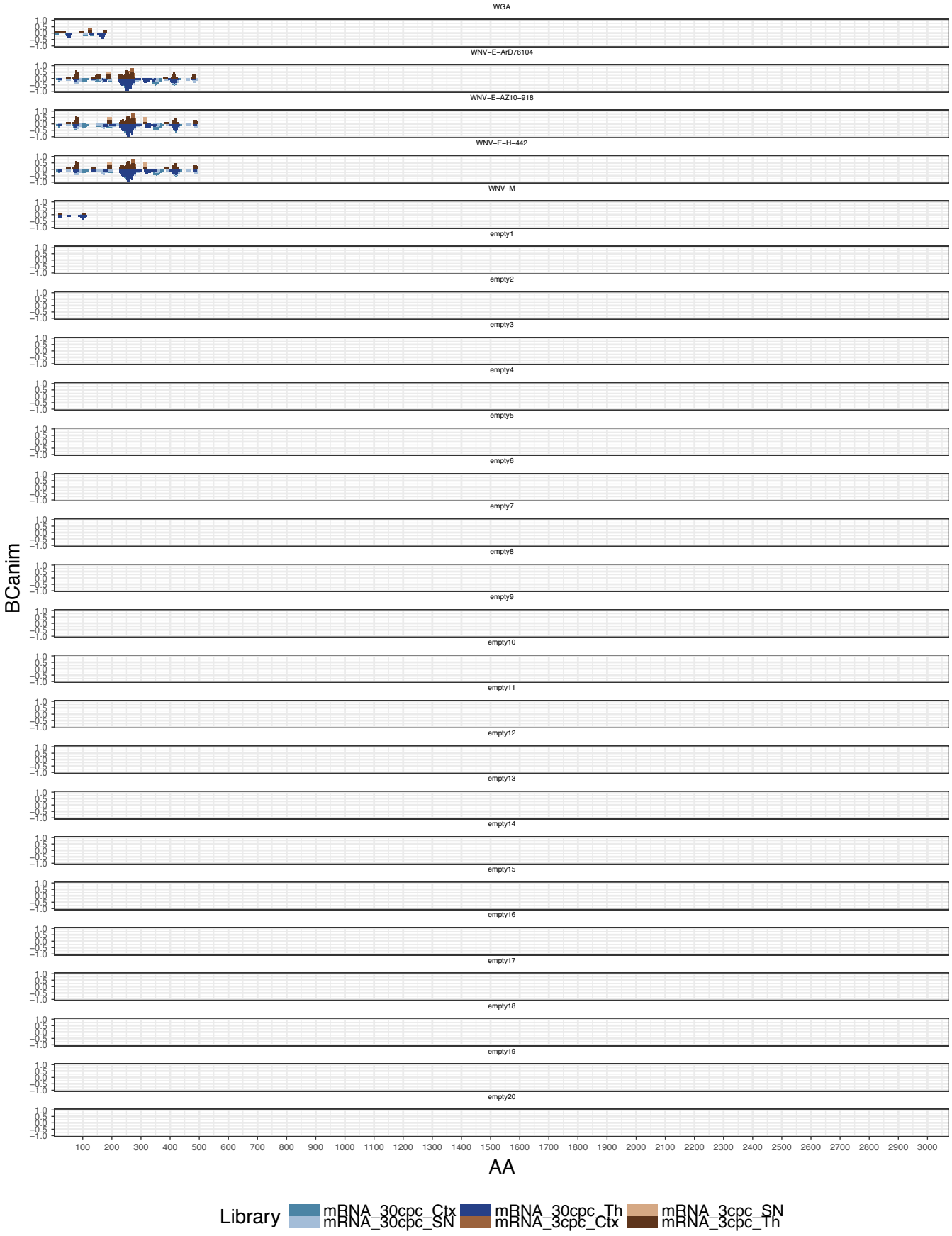

a–Syn
AAV1–VP1
AAV6–VP1
AAV8–VP1
AAV9–VP1
APP
BDNF
BDV–CRNP5
BDV–G
BDV–No–98
BoNT–A–Hc
BoNT–B–Hc
BoNT–C–Hc
BoNT–D–Hc
BoNT–E–Hc
BoNT–F–Hc
BoNT–G–Hc
BV–G
CAV–2–F–SH01
CAV–2–F–Toronto–A26–61
CDNF
CHIKV–Sp–1
ConA
EV71–VP1–1095–Japan–97
EV71–VP1–cx2011–016
BCanim
1.0
0.5
0.0
−0.5
−1.0
100 200 300 400 500 600 700 800 900 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


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
BCanim
1.0
0.5
0.0
−0.5
−1.0
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
BCanim
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th


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
BCanim
1.0
0.5
0.0
−0.5
−1.0
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


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
BCanim
1.0
0.5
0.0
−0.5
−1.0
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


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
BCanim
AA
Library
mRNA_30cpc_Ctx
mRNA_30cpc_SN
mRNA_30cpc_Th
mRNA_3cpc_Ctx
mRNA_3cpc_SN
mRNA_3cpc_Th

```
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     2017-11-01
```

```
Packages -----------------------------------------------------------------
```

```
 package                * version  date
 acepack                  1.4.1    2016-10-29
 AnnotationDbi            1.38.2   2017-10-13
 AnnotationFilter         1.0.0    2017-10-13
 AnnotationHub            2.8.2    2017-10-13
 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   2017-10-13
 BiocGenerics           * 0.22.1   2017-10-13
 BiocInstaller            1.26.1   2017-10-10
 BiocParallel             1.10.1   2017-10-13
 biomaRt                  2.32.1   2017-10-13
 Biostrings             * 2.44.2   2017-10-13
 biovizBase               1.24.0   2017-10-13
 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   2017-10-13
 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    2017-10-13
 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    2017-10-13
 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   2017-10-13
 GenomeInfoDbData         0.99.0   2017-10-13
 GenomicAlignments      * 1.12.2   2017-10-13
 GenomicFeatures          1.28.5   2017-10-13
```

```
GenomicRanges          * 1.28.6   2017-10-13
GGally                   1.3.2    2017-08-02
ggbio                  * 1.24.1   2017-10-13
ggplot2                * 2.2.1    2016-12-30
ggplus                 * 0.1      2017-10-13
graph                    1.54.0   2017-10-13
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   2017-10-13
IRanges                * 2.10.5   2017-10-13
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   2017-08-16
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   2017-10-13
parallel               * 3.4.2    2017-10-06
plyr                   * 1.8.4    2016-06-08
ProtGenerics             1.8.0    2017-10-13
R6                       2.2.2    2017-06-17
RBGL                     1.52.0   2017-10-13
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   2017-10-13
RSQLite                  2.0      2017-06-19
rtracklayer              1.36.6   2017-10-13
rvest                    0.3.2    2016-06-17
S4Vectors              * 0.14.7   2017-10-13
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    2017-10-13
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   2017-10-13
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   2017-10-13
yaml                     2.1.14   2016-11-12
zlibbioc                 1.22.0   2017-10-13
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)
CRAN (R 3.4.2)
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)
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)
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)
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
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
CRAN (R 3.4.2)
CRAN (R 3.4.2)
Bioconductor
CRAN (R 3.4.2)
CRAN (R 3.4.2)
local
local
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)
Bioconductor
```

# Clustering Polypeptide motifs using the Hammock hidden Markov model peptide clustering

*Tomas Bjorklund*

*Wed Nov 8 10:50:14 2017*

This script clusters Polypeptide motifs using the Hammock hidden Markov model peptide clustering.

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

## Loading samples

```
all.samples <- readRDS("data/allSamplesDataTable.RDS")
all.samples[, `:=`(Peptide, as.character(Peptide)), ]

setkey(all.samples, Group)

select.samples <- all.samples[J(c("mRNA_30cpc_SN", "mRNA_30cpc_Th", "mRNA_30cpc_Ctx",
    "mRNA_3cpc_SN", "mRNA_3cpc_Th", "mRNA_3cpc_Ctx", "mRNA_30cpc_SN_4wks", "mRNA_30cpc_Th_4wks",
    "mRNA_30cpc_Ctx_4wks", "mRNA_3cpc_SN_4wks", "mRNA_3cpc_Th_4wks", "mRNA_3cpc_Ctx_4wks"))]

select.samples[, `:=`(BCcount, as.integer(mclapply(BC, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Animalcount, as.integer(mclapply(Animals, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Score, BCcount + Animalcount - 1), ]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))

fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/trspSamplesPeptides.fasta",
    open = "w", nbchar = 60, as.string = TRUE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/Hammock", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar full -i /home/rstudio/data/
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23 --max_shift
# 13 --max_aln_length 37 --count_threshold 50 --max_inner_gaps 0
# --assign_thresholds 14.1,10.5,7.0
hammock.log <- data.table(readLines("data/Hammock/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

Hammock log file

2017-11-08 10:50:21.165:
Hammock version 1.1.1 Run with –help for a brief description of command line parameters.

2017-11-08 10:50:21.264: Program started in mode “full”.
Command-line arguments:
full -i /home/rstudio/data/trspSamplesPeptides.fasta -d /home/rstudio/data/Hammock –max_shift 7 -c 65 -t 32

Complete list of input/output parameters:
-i, –input /home/rstudio/data/trspSamplesPeptides.fasta
-d, –output_directory /home/rstudio/data/Hammock
-t, –thread 32
-l, –labels null

Complete list of clinkage clustering parameters:
-f, –file_format fasta
-m, –matrix /home/rstudio/Hammock_v_1.1.1/matrices/blosum62.txt
-g, –alignment_threshold (–greedy_threshold)null
-x, –max_shift 7
-p, –gap_penalty 0
-C, –cache_size_limit 1

2017-11-08 10:50:21.265: Loading input sequences...
2017-11-08 10:50:21.304: 4143 unique sequences loaded.
2017-11-08 10:50:21.308: 9028 total sequences loaded.
2017-11-08 10:50:21.308: 4143 unique sequences after non-specified labels filtered out
2017-11-08 10:50:21.313: 9028 total sequences after non-specified labels fileterd out
2017-11-08 10:50:21.314: Shortest sequence: 14 AA. Longest sequence: 22 AA.
2017-11-08 10:50:21.314: Up to 10 000 unique sequences. Using clinkage clustering. Use –use_greedy to force greedy clustering
2017-11-08 10:50:21.321: Generating input statistics...
2017-11-08 10:50:21.331: Clinkage clustering threshold not set. Setting automatically to: 26
2017-11-08 10:50:21.333: Clinkage clustering...
2017-11-08 10:50:40.209: Ready. Clustering time: 18876
2017-11-08 10:50:40.209: Resulting clusers: 1668
2017-11-08 10:50:40.210: Building MSAs...
2017-11-08 10:50:40.688: Ready. Total time: 19355
2017-11-08 10:50:40.688: Saving results to output files...
2017-11-08 10:50:40.907: Clinkage clustering results in: /home/rstudio/data/Hammock/initial_clusters.tsv
2017-11-08 10:50:40.907: and: /home/rstudio/data/Hammock/initial_clusters_sequences.tsv
2017-11-08 10:50:40.908: and: /home/rstudio/data/Hammock/initial_clusters_sequences_original_order.tsv
2017-11-08 10:50:40.908:
Loading clusters...
2017-11-08 10:50:40.931: Maximal alignment length not set. Setting automatically to: 30
2017-11-08 10:50:40.931: Minimal number of match states not set. Setting automatically to: 5
2017-11-08 10:50:40.964: Assign threshold sequence not set. Setting automatically to:
2017-11-08 10:50:40.966: 14.43,11.39,8.36,
2017-11-08 10:50:40.966: Overlap threshold not set. Setting automatically to:
2017-11-08 10:50:40.967: 10.63,6.08,0.0,
2017-11-08 10:50:40.967: Merge threshold not set. Setting automatically based on average sequence length to:
2017-11-08 10:50:40.968: 15.19,13.67,12.15,

Complete list of HMM-based clustering parameters:
-a, –part_threshold null

-s, –size_threshold null
-c, –count_threshold 65
-n, –assign_thresholds 14.43,11.39,8.36,
-v, –overlap_thresholds 10.63,6.08,0.0,
-r, –merge_thresholds 15.19,13.67,12.15,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 30
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-08 10:50:40.991:
Clustering in 3 rounds...
2017-11-08 10:50:40.992:
2017-11-08 10:50:40.992: Round 1:

2017-11-08 10:50:40.993: 65 clusters remaining
2017-11-08 10:50:40.993: Building hmms and searching database...
2017-11-08 10:50:41.427: Extending clusters...
2017-11-08 10:50:41.431: 52 sequences to be inserted into clusters
2017-11-08 10:50:41.431: 27 clusters to be extended
2017-11-08 10:50:41.452: 35 sequences rejected
2017-11-08 10:50:41.453: 1 cluster pairs to check and merge.
2017-11-08 10:50:41.453: Merging clusters from 1 groups...
2017-11-08 10:50:41.460: Buiding hhs...
2017-11-08 10:50:41.469: HH clustering...
2017-11-08 10:50:41.658:
2017-11-08 10:50:41.658: Round 2:

2017-11-08 10:50:41.658: 65 clusters remaining
2017-11-08 10:50:41.658: Building hmms and searching database...
2017-11-08 10:50:42.040: Extending clusters...
2017-11-08 10:50:42.044: 73 sequences to be inserted into clusters
2017-11-08 10:50:42.044: 34 clusters to be extended
2017-11-08 10:50:42.074: 58 sequences rejected
2017-11-08 10:50:42.075: 90 cluster pairs to check and merge.
2017-11-08 10:50:42.075: Merging clusters from 4 groups...
2017-11-08 10:50:42.082: Buiding hhs...
2017-11-08 10:50:42.119: HH clustering...
2017-11-08 10:50:43.158:
2017-11-08 10:50:43.158: Round 3:

2017-11-08 10:50:43.159: 64 clusters remaining
2017-11-08 10:50:43.159: Building hmms and searching database...
2017-11-08 10:50:43.520: Extending clusters...
2017-11-08 10:50:43.523: 208 sequences to be inserted into clusters
2017-11-08 10:50:43.523: 52 clusters to be extended
2017-11-08 10:50:43.631: 132 sequences rejected
2017-11-08 10:50:43.632: Overlap threshold is 0. Running full cluster merging.
2017-11-08 10:50:43.638: Buiding hhs...
2017-11-08 10:50:43.664: HH clustering...
2017-11-08 10:50:45.287:
Ready. Clustering time : 4296

Hammock log file

```
2017-11-08 10:50:45.287: Resulting clusers: 61
2017-11-08 10:50:45.287: Containing 471 unique sequences and 1535 total sequences.
2017-11-08 10:50:45.290: Unique sequences not assigned: 3672, total sequences not assigned: 7493
2017-11-08 10:50:45.291: Saving results to outupt files...
2017-11-08 10:50:45.340: Results in: /home/rstudio/data/Hammock/final_clusters_sequences.tsv
2017-11-08 10:50:45.341: and: /home/rstudio/data/Hammock/final_clusters.tsv
2017-11-08 10:50:45.341: and: /home/rstudio/data/Hammock/final_clusters_sequences_original_order.tsv
2017-11-08 10:50:45.341:
Calculating KLD...
2017-11-08 10:50:45.426: Final system KLD over match state MSA positions: 13.360165786988897
2017-11-08 10:50:45.426: Final system KLD over all MSA positions: 19.27055070815488
2017-11-08 10:50:45.426: Program successfully ended.
```

## Generation of Weblogo visualization

```
ham.clusters <- data.table(read.table("/home/rstudio/data/Hammock/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
id.order <- as.list(ham.clusters$cluster_id)
ham.clusters.all <- data.table(read.table("/home/rstudio/data/Hammock/final_clusters_sequences.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
ham.clusters.all[, `:=`(alignment, gsub("\\-", "\\_", alignment))]
setkey(select.samples, Peptide)
setkey(select.samples.trsp, Peptide)

unlink("/home/rstudio/data/WEBlogos", recursive = TRUE, force = FALSE)
dir.create(file.path("/home/rstudio/data/", "WEBlogos"), showWarnings = FALSE)
dir.create(file.path("/home/rstudio/data/Hammock/", "alignments_final_Scored"),
    showWarnings = FALSE)

setkey(ham.clusters.all, cluster_id)
setkey(ham.clusters, cluster_id)
setkey(select.samples.pepMerge, Peptide)


opts_chunk$set(out.width = "100%", fig.align = "center")
generateWeblogo <- function(in.name) {
    # in.name <- ham.clusters$cluster_id[2] in.name <- 6777
    this.fa <- read.fasta(file = paste("/home/rstudio/data/Hammock/alignments_final/",
        in.name, ".aln", sep = ""))
    allSeqs <- unlist(getSequence(this.fa, as.string = TRUE))
    allSeqs <- data.table(unlist(lapply(allSeqs, function(x) gsub("([-])", "",
        toupper(x)))))
    allSeqs.out <- select.samples.pepMerge[J(allSeqs)]
    allSeqs.out$Annot <- data.table(getName(this.fa))
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Score, sep = ""))]
    allSeqs.out$Alignment <- data.table(toupper(unlist(getSequence(this.fa,
        as.string = TRUE))))

    allSeqs.out <- allSeqs.out[rep(1:.N, Score)][, `:=`(Indx, 1:.N), by = Peptide]
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Indx, sep = ""))]

    write.fasta(as.list(allSeqs.out$Alignment), allSeqs.out$Annot, nbchar = 60,
        paste("/home/rstudio/data/Hammock/alignments_final_Scored/", in.name,
            ".aln", sep = ""), open = "w")
```

```
this.main <- ham.clusters[J(in.name)]
main.gene <- select.samples.trsp[J(this.main$main_sequence)]$GeneName[1]
this.title <- paste("## Peptide", this.main$main_sequence, "from", main.gene,
    "with cluster number", in.name, sep = " ")

tmp <- system(paste("weblogo --format PDF --sequence-type protein --size large --errorbars NO --resolutio
    this.title, "' < /home/rstudio/data/Hammock/alignments_final_Scored/",
    in.name, ".aln > /home/rstudio/data/WEBlogos/", in.name, ".pdf", sep = ""),
    intern = TRUE, ignore.stdout = FALSE)

cat("\n")
cat(this.title, "\n")
cat("\n")
cat("\n")
cat(paste0("![Peptide: ", this.main$main_sequence, " from ", main.gene,
    " with cluster number ", in.name, "](/home/rstudio/data/WEBlogos/",
    in.name, ".pdf)"))
cat("\n")
out.table <- knitr::kable(this.main[, c(1:8, 10)], format = "latex")
print(column_spec(out.table, 1:9, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
    "scale_down", "repeat_header")))
cat("\n")
this.cluster <- ham.clusters.all[J(in.name)]
out.table <- knitr::kable(this.cluster[, c(1:7)], format = "latex")
print(column_spec(out.table, 1:7, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
    "scale_down", "repeat_header")))

this.found <- select.samples[J(this.cluster$sequence)]
setnames(this.cluster, "sequence", "Peptide")
this.found <- merge(this.found, this.cluster[, 2:3], by = "Peptide", all = FALSE)
cat("\n")
cat("\n")
output.order <- c("alignment", "LUTnrs", "GeneName", "start", "structure",
    "Group", "Score")
if (nrow(this.found) >= 48) {
    this.found.p1 <- this.found[1:47, ]
    out.table <- knitr::kable(this.found.p1[, ..output.order], format = "latex")
    print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
        "repeat_header")))
    cat("\n")
    cat("\n\n\\pagebreak\n")
    cat("\n\n\\clearpage\n")
    this.found <- this.found[48:nrow(this.found), ]
    if (nrow(this.found) >= 48) {
        this.found.p2 <- this.found[1:47, ]
        out.table <- knitr::kable(this.found.p2[, ..output.order], format = "latex")
        print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
            "repeat_header")))
        cat("\n")
        cat("\n\n\\pagebreak\n")
        cat("\n\n\\clearpage\n")
        this.found <- this.found[48:nrow(this.found), ]
    }
    if (nrow(this.found) >= 48) {
        this.found.p2 <- this.found[1:47, ]
        out.table <- knitr::kable(this.found.p2[, ..output.order], format = "latex")
        print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
```

```
# setkey(select.samples.trsp,Peptide) select.samples.trsp.select <-
# select.samples.trsp[J(c('PPDELNLTTASLPL'))]



print("Total analysis time:")
```

```
[1] "Total analysis time:"
```

```
print(Sys.time() - strt1)
```

```
Time difference of 1.221091 mins
```

```
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     2017-11-08
```

```
Packages -----------------------------------------------------------------
```

```
 package    * version  date       source
 ade4         1.7-8    2017-08-09 CRAN (R 3.4.2)
 backports    1.1.1    2017-09-25 CRAN (R 3.4.2)
 base       * 3.4.2    2017-10-06 local
 compiler     3.4.2    2017-10-06 local
 data.table * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets   * 3.4.2    2017-10-06 local
 devtools   * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 digest       0.6.12   2017-01-27 CRAN (R 3.4.2)
 evaluate     0.10.1   2017-06-24 CRAN (R 3.4.2)
 formatR      1.5      2017-04-25 CRAN (R 3.4.2)
 graphics   * 3.4.2    2017-10-06 local
 grDevices  * 3.4.2    2017-10-06 local
 highr        0.6      2016-05-09 CRAN (R 3.4.2)
 hms          0.3      2016-11-22 CRAN (R 3.4.2)
 htmltools    0.3.6    2017-04-28 CRAN (R 3.4.2)
 httr         1.3.1    2017-08-20 CRAN (R 3.4.2)
 kableExtra * 0.5.2    2017-09-15 CRAN (R 3.4.2)
 knitr      * 1.17     2017-08-10 CRAN (R 3.4.2)
 magrittr     1.5      2014-11-22 CRAN (R 3.4.2)
 memoise      1.1.0    2017-04-21 CRAN (R 3.4.2)
 methods      3.4.2    2017-10-06 local
 parallel   * 3.4.2    2017-10-06 local
 plyr         1.8.4    2016-06-08 CRAN (R 3.4.2)
 R6           2.2.2    2017-06-17 CRAN (R 3.4.2)
 Rcpp         0.12.13  2017-09-28 CRAN (R 3.4.2)
 readr        1.1.1    2017-05-16 CRAN (R 3.4.2)
 reshape2   * 1.4.2    2016-10-22 CRAN (R 3.4.2)
 rlang        0.1.2    2017-08-09 CRAN (R 3.4.2)
 rmarkdown    1.6      2017-06-15 CRAN (R 3.4.2)
 rprojroot    1.2      2017-01-16 CRAN (R 3.4.2)
 rvest        0.3.2    2016-06-17 CRAN (R 3.4.2)
 seqinr     * 3.4-5    2017-08-01 CRAN (R 3.4.2)
```

```
stats     * 3.4.2   2017-10-06 local
stringi     1.1.5   2017-04-07 CRAN (R 3.4.2)
stringr     1.2.0   2017-02-18 CRAN (R 3.4.2)
tibble      1.3.4   2017-08-22 CRAN (R 3.4.2)
tools       3.4.2   2017-10-06 local
utils     * 3.4.2   2017-10-06 local
withr       2.0.0   2017-07-28 CRAN (R 3.4.2)
xml2        1.1.1   2017-01-24 CRAN (R 3.4.2)
yaml        2.1.14  2016-11-12 CRAN (R 3.4.2)
```

# Clustering Polypeptide motifs using the Hammock hidden Markov model peptide clustering

*Tomas Bjorklund*

*Wed Nov 8 14:46:45 2017*

This script clusters Polypeptide motifs using the Hammock hidden Markov model peptide clustering.

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

## Loading samples

```
all.samples <- readRDS("data/allSamplesDataTable.RDS")
all.samples[, `:=`(Peptide, as.character(Peptide)), ]

setkey(all.samples, Group)

select.samples <- all.samples[J(c("mRNA_30cpc_SN", "mRNA_30cpc_Th", "mRNA_30cpc_Ctx",
    "mRNA_3cpc_SN", "mRNA_3cpc_Th", "mRNA_3cpc_Ctx", "mRNA_30cpc_SN_4wks", "mRNA_30cpc_Th_4wks",
    "mRNA_30cpc_Ctx_4wks", "mRNA_3cpc_SN_4wks", "mRNA_3cpc_Th_4wks", "mRNA_3cpc_Ctx_4wks"))]

select.samples[, `:=`(BCcount, as.integer(mclapply(BC, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Animalcount, as.integer(mclapply(Animals, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Score, BCcount + Animalcount - 1), ]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))
setkey(select.samples.trsp, Sequence)

selected.peptides <- data.table(read.table("input/selectedPeptides.txt", header = TRUE,
    skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
# Removed the first 2 serotypes selected for in vitro properties
selected.peptides <- selected.peptides[3:25, ]
select.samples.main <- select.samples.trsp[J(selected.peptides$Sequence)]

select.samples.main[, `:=`(Group, as.character(lapply(Group, function(x) gsub("(_4wks)",
    "", x))))]
select.samples.main <- select.samples.main[, sum(Score), by = c("Sequence",
    "Group", "Peptide")]
setnames(select.samples.main, "V1", "Score")
select.samples.main <- merge(selected.peptides, select.samples.main, by = "Sequence",
    all = FALSE)
select.samples.main[, `:=`(Peptide, as.character(Peptide))]

out.table <- data.table(dcast(select.samples.main, Name + Peptide ~ Group, fun = sum,
    value.var = "Score"))

out.table[, `:=`(alignment, paste("----------------", Peptide, "----------------",
    sep = ""))]
out.table <- transform(out.table, sum = rowSums(out.table[, 3:8]))
# reorder for output
out.table <- out.table[, c(1, 2, 9, 10, 3:8)]
setnames(out.table, c("Name", "Peptide"), c("cluster_id", "sequence"))
write.table(out.table, file = "data/selectedPeptides.tsv", quote = FALSE, sep = "\t",
```

```
    row.names = FALSE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")

# removing the selected peptides from all sequences to allow for clustering
select.samples.trsp <- select.samples.trsp[!(select.samples.trsp$Peptide %in%
    out.table$sequence), ]


fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/trspSamplesPeptidesNonSelect.fasta",
    open = "w", nbchar = 60, as.string = TRUE)
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/HammockSelected", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar cluster -i /home/rstudio/da
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23 --max_shift
# 13 --max_aln_length 42 --count_threshold 50 --max_inner_gaps 0
# --assign_thresholds 14.1,10.5,7.0
hammock.log <- data.table(readLines("data/HammockSelected/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

| Hammock log file |
|---|
| 2017-11-08 14:46:53.943: |
| Hammock version 1.1.1 Run with –help for a brief description of command line parameters. |
| 2017-11-08 14:46:54.082: Program started in mode “cluster”. |
| Command-line arguments: |
| cluster -i /home/rstudio/data/selectedPeptides.tsv -as /home/rstudio/data/trspSamplesPeptidesNonSelect.fasta -d |
| /home/rstudio/data/HammockSelected –max_shift 9 –assign_thresholds 15,12,10 –max_aln_length 42 -c 23 -t 32 |
| Complete list of input/output parameters: |
| -i, –input /home/rstudio/data/selectedPeptides.tsv |
| -d, –output_directory /home/rstudio/data/HammockSelected |
| -t, –thread 32 |
| -l, –labels null |
| 2017-11-08 14:46:54.082: |
| Loading clusters... |
| 2017-11-08 14:46:54.088: Loading additional sequences... |
| 2017-11-08 14:46:54.128: Generating input statistics... |
| 2017-11-08 14:46:54.131: Minimal number of match states not set. Setting automatically to: 5 |
| 2017-11-08 14:46:54.143: Overlap threshold not set. Setting automatically to: |
| 2017-11-08 14:46:54.144: 10.04,5.74,0.0, |

2017-11-08 14:46:54.144: Merge threshold not set. Setting automatically based on average sequence length to:
2017-11-08 14:46:54.145: 14.35,12.91,11.48,

Complete list of HMM-based clustering parameters:
-a, –part_threshold null
-s, –size_threshold null
-c, –count_threshold 23
-n, –assign_thresholds 15.0,12.0,10.0,
-v, –overlap_thresholds 10.04,5.74,0.0,
-r, –merge_thresholds 14.35,12.91,11.48,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 42
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-08 14:46:54.157:
Clustering in 3 rounds...
2017-11-08 14:46:54.159:
2017-11-08 14:46:54.159: Round 1:

2017-11-08 14:46:54.160: 23 clusters remaining
2017-11-08 14:46:54.160: Building hmms and searching database...
2017-11-08 14:46:54.461: Extending clusters...
2017-11-08 14:46:54.464: 90 sequences to be inserted into clusters
2017-11-08 14:46:54.465: 23 clusters to be extended
2017-11-08 14:46:54.506: 3 sequences rejected
2017-11-08 14:46:54.508: 1 cluster pairs to check and merge.
2017-11-08 14:46:54.508: Merging clusters from 1 groups...
2017-11-08 14:46:54.511: Buiding hhs...
2017-11-08 14:46:54.522: HH clustering...
2017-11-08 14:46:54.743:
2017-11-08 14:46:54.743: Round 2:

2017-11-08 14:46:54.743: 23 clusters remaining
2017-11-08 14:46:54.743: Building hmms and searching database...
2017-11-08 14:46:54.917: Extending clusters...
2017-11-08 14:46:54.919: 17 sequences to be inserted into clusters
2017-11-08 14:46:54.919: 9 clusters to be extended
2017-11-08 14:46:54.929: 2 sequences rejected
2017-11-08 14:46:54.930: 17 cluster pairs to check and merge.
2017-11-08 14:46:54.930: Merging clusters from 2 groups...
2017-11-08 14:46:54.934: Buiding hhs...
2017-11-08 14:46:54.952: HH clustering...
2017-11-08 14:46:55.205:
2017-11-08 14:46:55.206: Round 3:

2017-11-08 14:46:55.206: 23 clusters remaining
2017-11-08 14:46:55.206: Building hmms and searching database...
2017-11-08 14:46:55.348: Extending clusters...
2017-11-08 14:46:55.350: 20 sequences to be inserted into clusters
2017-11-08 14:46:55.350: 10 clusters to be extended
2017-11-08 14:46:55.359: 6 sequences rejected

Hammock log file

2017-11-08 14:46:55.360: Overlap threshold is 0. Running full cluster merging.
2017-11-08 14:46:55.362: Buiding hhs...
2017-11-08 14:46:55.375: HH clustering...
2017-11-08 14:46:55.654:
Ready. Clustering time : 1496
2017-11-08 14:46:55.654: Resulting clusers: 23
2017-11-08 14:46:55.654: Containing 139 unique sequences and 537 total sequences.
2017-11-08 14:46:55.657: Unique sequences not assigned: 4004, total sequences not assigned: 8491
2017-11-08 14:46:55.657: Saving results to outupt files...
2017-11-08 14:46:55.682: Results in: /home/rstudio/data/HammockSelected/final_clusters_sequences.tsv
2017-11-08 14:46:55.682: and: /home/rstudio/data/HammockSelected/final_clusters.tsv
2017-11-08 14:46:55.682:
Calculating KLD...
2017-11-08 14:46:55.716: Final system KLD over match state MSA positions: 11.744046931714234
2017-11-08 14:46:55.716: Final system KLD over all MSA positions: 19.319796797087644
2017-11-08 14:46:55.716: Program successfully ended.

## Generation of Weblogo visualization

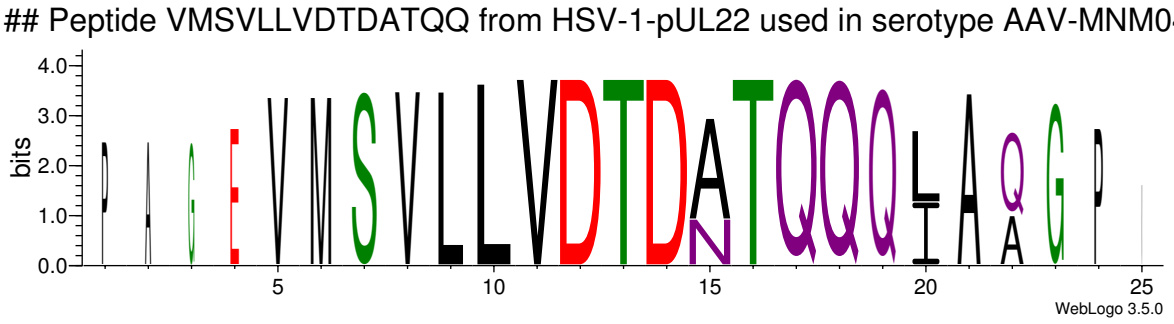
```
ham.clusters <- data.table(read.table("/home/rstudio/data/HammockSelected/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
id.order <- as.list(ham.clusters$cluster_id)
ham.clusters.all <- data.table(read.table("/home/rstudio/data/HammockSelected/final_clusters_sequences.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
ham.clusters.all[, `:=`(alignment, gsub("\\-", "\\_", alignment))]
setkey(select.samples, Peptide)
setkey(select.samples.trsp, Peptide)
setkey(select.samples.pepMerge, Peptide)

unlink("/home/rstudio/data/WEBlogosSelected", recursive = TRUE, force = FALSE)
dir.create(file.path("/home/rstudio/data/", "WEBlogosSelected"), showWarnings = FALSE)
dir.create(file.path("/home/rstudio/data/HammockSelected/", "alignments_final_Scored"),
    showWarnings = FALSE)

setkey(ham.clusters.all, cluster_id)
setkey(ham.clusters, cluster_id)


opts_chunk$set(out.width = "100%", fig.align = "center")
generateWeblogo <- function(in.name) {
    # in.name <- ham.clusters$cluster_id[2] in.name <- 67777

    this.fa <- read.fasta(file = paste("/home/rstudio/data/HammockSelected/alignments_final/",
        in.name, ".aln", sep = ""))
    allSeqs <- unlist(getSequence(this.fa, as.string = TRUE))
    allSeqs <- data.table(unlist(lapply(allSeqs, function(x) gsub("([-])", "",
        toupper(x)))))
    allSeqs.out <- select.samples.pepMerge[J(allSeqs)]
    allSeqs.out$Annot <- data.table(getName(this.fa))
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Score, sep = ""))]
    allSeqs.out$Alignment <- data.table(toupper(unlist(getSequence(this.fa,
        as.string = TRUE))))
    align.all <- strsplit(allSeqs.out$Alignment, "[A-Z]")
    tail.length <- min(unlist(lapply(lapply(align.all, tail, n = 1L), nchar)))
```

```
    head.length <- min(unlist(lapply(lapply(align.all, head, n = 1L), nchar)))
    total.length <- max(nchar(allSeqs.out$Alignment))
    allSeqs.out[, `:=`(Alignment, lapply(Alignment, function(x) substr(x, head.length +
        1, total.length - tail.length)))]
    allSeqs.out <- allSeqs.out[rep(1:.N, Score)][, `:=`(Indx, 1:.N), by = Peptide]
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Indx, sep = ""))]

    write.fasta(as.list(allSeqs.out$Alignment), allSeqs.out$Annot, nbchar = 60,
        paste("/home/rstudio/data/HammockSelected/alignments_final_Scored/",
            in.name, ".aln", sep = ""), open = "w")

    this.main <- ham.clusters[J(in.name)]
    main.gene <- select.samples[J(this.main$main_sequence)]$GeneName[1]
    this.title <- paste("## Peptide ", this.main$main_sequence, " from ", main.gene,
        " used in serotype AAV-MNM0", in.name, sep = "")
    tmp <- system(paste("weblogo --format PDF --sequence-type protein --size large --stacks-per-line 48 --err
        this.title, "' < /home/rstudio/data/HammockSelected/alignments_final_Scored/",
        in.name, ".aln > /home/rstudio/data/WEBlogosSelected/", in.name, ".pdf",
        sep = ""), intern = TRUE, ignore.stdout = FALSE)

    cat("\n")
    cat(this.title, "\n")
    cat("\n")
    cat("\n")
    cat(paste0("![Peptide: ", this.main$main_sequence, " from ", main.gene,
        " with cluster number ", in.name, "](/home/rstudio/data/WEBlogosSelected/",
        in.name, ".pdf)"))
    cat("\n")
    out.table <- knitr::kable(this.main[, c(1:9)], format = "latex")
    print(column_spec(out.table, 1:9, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
        "scale_down", "repeat_header")))
    cat("\n")
    this.cluster <- ham.clusters.all[J(in.name)]
    align.all <- strsplit(this.cluster$alignment, "[A-Z]")
    tail.length <- min(unlist(lapply(lapply(align.all, tail, n = 1L), nchar)))
    head.length <- min(unlist(lapply(lapply(align.all, head, n = 1L), nchar)))
    total.length <- max(nchar(this.cluster$alignment))
    this.cluster[, `:=`(alignment, lapply(alignment, function(x) substr(x, head.length +
        1, total.length - tail.length)))]

    out.table <- knitr::kable(this.cluster[, c(1:7)], format = "latex")
    print(column_spec(out.table, 1:7, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
        "scale_down", "repeat_header")))

    this.found <- select.samples[J(this.cluster$sequence)]
    setnames(this.cluster, "sequence", "Peptide")
    this.found <- merge(this.found, this.cluster[, 2:3], by = "Peptide", all = FALSE)
    cat("\n")
    cat("\n")
    output.order <- c("alignment", "LUTnrs", "GeneName", "start", "structure",
        "Group", "Score")
    if (nrow(this.found) >= 48) {
        this.found.p1 <- this.found[1:47, ]
        out.table <- knitr::kable(this.found.p1[, ..output.order], format = "latex")
        print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
            "repeat_header")))
        cat("\n")
        cat("\n\n\\pagebreak\n")
```

```
        cat("\n\n\\clearpage\n")
        this.found <- this.found[48:nrow(this.found), ]
        if (nrow(this.found) >= 48) {
            this.found.p2 <- this.found[1:47, ]
            out.table <- knitr::kable(this.found.p2[, ..output.order], format = "latex")
            print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
                "repeat_header")))
            cat("\n")
            cat("\n\n\\pagebreak\n")
            cat("\n\n\\clearpage\n")
            this.found <- this.found[48:nrow(this.found), ]
        }
        if (nrow(this.found) >= 48) {
            this.found.p2 <- this.found[1:47, ]
            out.table <- knitr::kable(this.found.p2[, ..output.order], format = "latex")
            print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
                "repeat_header")))
            cat("\n")
            cat("\n\n\\pagebreak\n")
            cat("\n\n\\clearpage\n")
            this.found <- this.found[48:nrow(this.found), ]
        }
        out.table <- knitr::kable(this.found[, ..output.order], format = "latex")
        print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
            "repeat_header")))
        cat("\n")
        cat("\n\n\\clearpage\n")
        cat("\n")
    } else {
        out.table <- knitr::kable(this.found[, ..output.order], format = "latex")
        print(column_spec(out.table, 1, monospace = TRUE) %>% kable_styling(latex_options = c("striped",
            "scale_down", "repeat_header")))
        cat("\n")
        cat("\n\n\\pagebreak\n")
        cat("\n\n\\clearpage\n")
    }
}
```

```
invisible(lapply(id.order, generateWeblogo))
```

## Peptide VMSVLLVDTDATQQ from HSV-1-pUL22 used in serotype AAV-MNM04




| cluster_id | main_sequence | sum | mRNA_3cpc_Th | mRNA_30cpc_Ctx | mRNA_30cpc_Th | mRNA_30cpc_SN | mRNA_3cpc_Ctx | mRNA_3cpc_SN |
|---|---|---|---|---|---|---|---|---|
| 4 | VMSVLLVDTDATQQ | 48 | 6 | 14 | 14 | 11 | 0 | 0 |

```
# setkey(select.samples.trsp,Peptide) select.samples.trsp.select <-
# select.samples.trsp[J(c('PPDELNLTTASLPL'))]



print("Total analysis time:")
```

```
[1] "Total analysis time:"
```

```
print(Sys.time() - strt1)
```

```
Time difference of 25.50499 secs
```

```
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     2017-11-08
```

```
Packages -----------------------------------------------------------------
```

```
 package    * version  date       source
 ade4         1.7-8    2017-08-09 CRAN (R 3.4.2)
 backports    1.1.1    2017-09-25 CRAN (R 3.4.2)
 base       * 3.4.2    2017-10-06 local
 compiler     3.4.2    2017-10-06 local
 data.table * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets   * 3.4.2    2017-10-06 local
 devtools   * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 digest       0.6.12   2017-01-27 CRAN (R 3.4.2)
 evaluate     0.10.1   2017-06-24 CRAN (R 3.4.2)
 formatR      1.5      2017-04-25 CRAN (R 3.4.2)
 graphics   * 3.4.2    2017-10-06 local
 grDevices  * 3.4.2    2017-10-06 local
 highr        0.6      2016-05-09 CRAN (R 3.4.2)
 hms          0.3      2016-11-22 CRAN (R 3.4.2)
 htmltools    0.3.6    2017-04-28 CRAN (R 3.4.2)
 httr         1.3.1    2017-08-20 CRAN (R 3.4.2)
 kableExtra * 0.5.2    2017-09-15 CRAN (R 3.4.2)
 knitr      * 1.17     2017-08-10 CRAN (R 3.4.2)
 magrittr     1.5      2014-11-22 CRAN (R 3.4.2)
 memoise      1.1.0    2017-04-21 CRAN (R 3.4.2)
 methods      3.4.2    2017-10-06 local
 parallel   * 3.4.2    2017-10-06 local
 plyr         1.8.4    2016-06-08 CRAN (R 3.4.2)
 R6           2.2.2    2017-06-17 CRAN (R 3.4.2)
 Rcpp         0.12.13  2017-09-28 CRAN (R 3.4.2)
 readr        1.1.1    2017-05-16 CRAN (R 3.4.2)
 reshape2   * 1.4.2    2016-10-22 CRAN (R 3.4.2)
 rlang        0.1.2    2017-08-09 CRAN (R 3.4.2)
 rmarkdown    1.6      2017-06-15 CRAN (R 3.4.2)
 rprojroot    1.2      2017-01-16 CRAN (R 3.4.2)
 rvest        0.3.2    2016-06-17 CRAN (R 3.4.2)
 seqinr     * 3.4-5    2017-08-01 CRAN (R 3.4.2)
```

```
stats      * 3.4.2    2017-10-06 local
stringi      1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr      1.2.0    2017-02-18 CRAN (R 3.4.2)
tibble       1.3.4    2017-08-22 CRAN (R 3.4.2)
tools        3.4.2    2017-10-06 local
utils      * 3.4.2    2017-10-06 local
withr        2.0.0    2017-07-28 CRAN (R 3.4.2)
xml2         1.1.1    2017-01-24 CRAN (R 3.4.2)
yaml         2.1.14   2016-11-12 CRAN (R 3.4.2)
```

# Heatmaps generated from HMM peptide clustering

*Tomas Bjorklund*

*Wed Nov 1 20:03:56 2017*

This script clusters Polypeptide motifs using the Hammock hidden Markov model peptide clustering and generates Heatmaps for most functional motifs.

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

## Loading samples

```
all.samples <- readRDS("data/allSamplesDataTable.RDS")
all.samples[, `:=`(Peptide, as.character(Peptide)), ]

setkey(all.samples, Group)
```

## Generation of heatmaps for in vivo transported samples

```
select.samples <- all.samples[J(c("mRNA_30cpc_Str", "mRNA_30cpc_SN", "mRNA_30cpc_Th",
    "mRNA_30cpc_Ctx", "mRNA_3cpc_Str", "mRNA_3cpc_SN", "mRNA_3cpc_Th", "mRNA_3cpc_Ctx",
    "mRNA_30cpc_Str_4wks", "mRNA_30cpc_SN_4wks", "mRNA_30cpc_Th_4wks", "mRNA_30cpc_Ctx_4wks",
    "mRNA_3cpc_Str_4wks", "mRNA_3cpc_SN_4wks", "mRNA_3cpc_Th_4wks", "mRNA_3cpc_Ctx_4wks"))]

select.samples[, `:=`(BCcount, as.integer(mclapply(BC, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Animalcount, as.integer(mclapply(Animals, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Score, BCcount + Animalcount - 1), ]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))

fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/invivoSamplesPeptides.fasta",
    open = "w", nbchar = 60, as.string = TRUE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/HammockInVivo", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar full -i /home/rstudio/data/
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23 --max_shift
# 13 --max_aln_length 37 --count_threshold 50 --max_inner_gaps 0
# --assign_thresholds 14.1,10.5,7.0
```

```
hammock.log <- data.table(readLines("data/HammockInVivo/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

| Hammock log file |
|---|
| 2017-11-01 20:04:22.029: |
| Hammock version 1.1.1 Run with –help for a brief description of command line parameters. |
| |
| 2017-11-01 20:04:22.145: Program started in mode “full”. |
| Command-line arguments: |
| full -i /home/rstudio/data/invivoSamplesPeptides.fasta -d /home/rstudio/data/HammockInVivo –max_shift 7 -c 250 –alignment_threshold 26 –assign_thresholds 50,40,30 -t 32 |
| |
| Complete list of input/output parameters: |
| -i, –input /home/rstudio/data/invivoSamplesPeptides.fasta |
| -d, –output_directory /home/rstudio/data/HammockInVivo |
| -t, –thread 32 |
| -l, –labels null |
| |
| Complete list of clinkage clustering parameters: |
| -f, –file_format fasta |
| -m, –matrix /home/rstudio/Hammock_v_1.1.1/matrices/blosum62.txt |
| -g, –alignment_threshold (–greedy_threshold)26 |
| -x, –max_shift 7 |
| -p, –gap_penalty 0 |
| -C, –cache_size_limit 1 |
| |
| 2017-11-01 20:04:22.146: Loading input sequences... |
| 2017-11-01 20:04:22.250: 15768 unique sequences loaded. |
| 2017-11-01 20:04:22.263: 39001 total sequences loaded. |
| 2017-11-01 20:04:22.263: 15768 unique sequences after non-specified labels filtered out |
| 2017-11-01 20:04:22.278: 39001 total sequences after non-specified labels fileterd out |
| 2017-11-01 20:04:22.282: Shortest sequence: 14 AA. Longest sequence: 22 AA. |
| 2017-11-01 20:04:22.282: More than 10 000 unique sequences. Using greedy clustering. Use –use_clinkage to force clinkage clustering |
| 2017-11-01 20:04:22.309: Generating input statistics... |
| 2017-11-01 20:04:22.357: Initial greedy clusters limit not set. Setting automatically to: 394 |
| 2017-11-01 20:04:22.358: Greedy clustering... |
| 2017-11-01 20:04:27.362: Ready. Clustering time: 5003 |
| 2017-11-01 20:04:27.362: Resulting clusers: 12939 |
| 2017-11-01 20:04:27.362: Building MSAs... |
| 2017-11-01 20:04:27.657: Ready. Total time: 5299 |
| 2017-11-01 20:04:27.657: Saving results to output files... |
| 2017-11-01 20:04:27.993: Greedy clustering results in: /home/rstudio/data/HammockInVivo/initial_clusters.tsv |
| 2017-11-01 20:04:27.993: and: /home/rstudio/data/HammockInVivo/initial_clusters_sequences.tsv |
| 2017-11-01 20:04:27.993: and: /home/rstudio/data/HammockInVivo/initial_clusters_sequences_original_order.tsv |
| 2017-11-01 20:04:27.993: |
| Loading clusters... |
| 2017-11-01 20:04:28.059: Maximal alignment length not set. Setting automatically to: 31 |
| 2017-11-01 20:04:28.063: Minimal number of match states not set. Setting automatically to: 5 |
| 2017-11-01 20:04:28.170: Overlap threshold not set. Setting automatically to: |
| 2017-11-01 20:04:28.177: 10.83,6.19,0.0, |
| 2017-11-01 20:04:28.177: Merge threshold not set. Setting automatically based on average sequence length to: |

2017-11-01 20:04:28.182: 15.47,13.92,12.38,

Complete list of HMM-based clustering parameters:
-a, –part_threshold null
-s, –size_threshold null
-c, –count_threshold 250
-n, –assign_thresholds 50.0,40.0,30.0,
-v, –overlap_thresholds 10.83,6.19,0.0,
-r, –merge_thresholds 15.47,13.92,12.38,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 31
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-01 20:04:28.244:
Clustering in 3 rounds...
2017-11-01 20:04:28.246:
2017-11-01 20:04:28.246: Round 1:

2017-11-01 20:04:28.246: 250 clusters remaining
2017-11-01 20:04:28.246: Building hmms and searching database...
2017-11-01 20:04:30.188: Extending clusters...
2017-11-01 20:04:30.213: 0 sequences to be inserted into clusters
2017-11-01 20:04:30.213: 0 clusters to be extended
2017-11-01 20:04:30.214: 0 sequences rejected
2017-11-01 20:04:30.216: 102 cluster pairs to check and merge.
2017-11-01 20:04:30.216: Merging clusters from 33 groups...
2017-11-01 20:04:30.239: Buiding hhs...
2017-11-01 20:04:30.330: HH clustering...
2017-11-01 20:04:32.646:
2017-11-01 20:04:32.647: Round 2:

2017-11-01 20:04:32.647: 242 clusters remaining
2017-11-01 20:04:32.647: Building hmms and searching database...
2017-11-01 20:04:34.190: Extending clusters...
2017-11-01 20:04:34.199: 0 sequences to be inserted into clusters
2017-11-01 20:04:34.200: 0 clusters to be extended
2017-11-01 20:04:34.200: 0 sequences rejected
2017-11-01 20:04:34.209: 2152 cluster pairs to check and merge.
2017-11-01 20:04:34.209: Merging clusters from 1 groups...
2017-11-01 20:04:34.233: Buiding hhs...
2017-11-01 20:04:34.311: HH clustering...
2017-11-01 20:04:41.225:
2017-11-01 20:04:41.225: Round 3:

2017-11-01 20:04:41.225: 229 clusters remaining
2017-11-01 20:04:41.226: Building hmms and searching database...
2017-11-01 20:04:42.594: Extending clusters...
2017-11-01 20:04:42.612: 7 sequences to be inserted into clusters
2017-11-01 20:04:42.612: 5 clusters to be extended
2017-11-01 20:04:42.619: 0 sequences rejected
2017-11-01 20:04:42.620: Overlap threshold is 0. Running full cluster merging.

Hammock log file

```
2017-11-01 20:04:42.639: Buiding hhs...
2017-11-01 20:04:42.649: HH clustering...
2017-11-01 20:04:53.917:
Ready. Clustering time : 25673
2017-11-01 20:04:53.917: Resulting clusers: 203
2017-11-01 20:04:53.917: Containing 2428 unique sequences and 10180 total sequences.
2017-11-01 20:04:53.926: Unique sequences not assigned: 13340, total sequences not assigned: 28821
2017-11-01 20:04:53.926: Saving results to outupt files...
2017-11-01 20:04:54.050: Results in: /home/rstudio/data/HammockInVivo/final_clusters_sequences.tsv
2017-11-01 20:04:54.050: and: /home/rstudio/data/HammockInVivo/final_clusters.tsv
2017-11-01 20:04:54.051: and: /home/rstudio/data/HammockInVivo/final_clusters_sequences_original_order.tsv
2017-11-01 20:04:54.051:
Calculating KLD...
2017-11-01 20:04:54.242: Final system KLD over match state MSA positions: 18.828779438507368
2017-11-01 20:04:54.243: Final system KLD over all MSA positions: 29.927410491876895
2017-11-01 20:04:54.243: Program successfully ended.
```

## Generation of Weblogo visualization

```
ham.clusters <- data.table(read.table("/home/rstudio/data/HammockInVivo/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
id.order <- as.list(ham.clusters$cluster_id)
ham.clusters.all <- data.table(read.table("/home/rstudio/data/HammockInVivo/final_clusters_sequences.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
ham.clusters.all[, `:=`(alignment, gsub("\\-", "\\_", alignment))]
setkey(select.samples, Peptide)
setkey(select.samples.trsp, Peptide)

unlink("/home/rstudio/data/WEBlogosInVivo", recursive = TRUE, force = FALSE)
dir.create(file.path("/home/rstudio/data/", "WEBlogosInVivo"), showWarnings = FALSE)
dir.create(file.path("/home/rstudio/data/HammockInVivo/", "alignments_final_Scored"),
    showWarnings = FALSE)

setkey(ham.clusters.all, cluster_id)
setkey(ham.clusters, cluster_id)
setkey(select.samples.pepMerge, Peptide)


opts_chunk$set(out.width = "100%", fig.align = "center")
generateWeblogo <- function(in.name) {
    # in.name <- ham.clusters$cluster_id[12] in.name <- 67777
    this.fa <- read.fasta(file = paste("/home/rstudio/data/HammockInVivo/alignments_final/",
        in.name, ".aln", sep = ""))
    allSeqs <- unlist(getSequence(this.fa, as.string = TRUE))
    allSeqs <- data.table(unlist(lapply(allSeqs, function(x) gsub("([-])", "",
        toupper(x)))))
    allSeqs.out <- select.samples.pepMerge[J(allSeqs)]
    allSeqs.out$Annot <- data.table(getName(this.fa))
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Score, sep = ""))]
    allSeqs.out$Alignment <- data.table(toupper(unlist(getSequence(this.fa,
        as.string = TRUE))))

    allSeqs.out <- allSeqs.out[rep(1:.N, Score)][, `:=`(Indx, 1:.N), by = Peptide]
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Indx, sep = ""))]
```

```
    write.fasta(as.list(allSeqs.out$Alignment), allSeqs.out$Annot, nbchar = 60,
        paste("/home/rstudio/data/HammockInVivo/alignments_final_Scored/", in.name,
            ".aln", sep = ""), open = "w")


    this.main <- ham.clusters[J(in.name)]
    main.gene <- select.samples.trsp[J(this.main$main_sequence)]$GeneName[1]
    this.title <- paste("## Peptide", this.main$main_sequence, "from", main.gene,
        "with cluster number", in.name, sep = " ")

    tmp <- system(paste("weblogo --format PDF --sequence-type protein --size large --errorbars NO --resolutio
        this.title, "' < /home/rstudio/data/HammockInVivo/alignments_final_Scored/",
        in.name, ".aln > /home/rstudio/data/WEBlogosInVivo/", in.name, ".pdf",
        sep = ""), intern = TRUE, ignore.stdout = FALSE)
}

invisible(mclapply(id.order, generateWeblogo, mc.cores = detectCores()))


ham.clusters.merged <- ham.clusters

ham.clusters.merged[, `:=`(mRNA_Str, mRNA_30cpc_Str + mRNA_3cpc_Str + mRNA_30cpc_Str_4wks +
    mRNA_3cpc_Str_4wks)]
ham.clusters.merged[, `:=`(mRNA_SN, mRNA_30cpc_SN + mRNA_3cpc_SN + mRNA_30cpc_SN_4wks +
    mRNA_3cpc_SN_4wks)]
ham.clusters.merged[, `:=`(mRNA_Th, mRNA_30cpc_Th + mRNA_3cpc_Th + mRNA_30cpc_Th_4wks +
    mRNA_3cpc_Th_4wks)]
ham.clusters.merged[, `:=`(mRNA_Ctx, mRNA_30cpc_Ctx + mRNA_3cpc_Ctx + mRNA_30cpc_Ctx_4wks +
    mRNA_3cpc_Ctx_4wks)]
ham.clusters.merged[, `:=`(c("mRNA_30cpc_Str", "mRNA_30cpc_SN", "mRNA_30cpc_Th",
    "mRNA_30cpc_Ctx", "mRNA_3cpc_Str", "mRNA_3cpc_SN", "mRNA_3cpc_Th", "mRNA_3cpc_Ctx",
    "mRNA_30cpc_Str_4wks", "mRNA_30cpc_SN_4wks", "mRNA_30cpc_Th_4wks", "mRNA_30cpc_Ctx_4wks",
    "mRNA_3cpc_Str_4wks", "mRNA_3cpc_SN_4wks", "mRNA_3cpc_Th_4wks", "mRNA_3cpc_Ctx_4wks"),
    NULL)]


ham.clusters.merged.melt <- melt(ham.clusters.merged, id = c("cluster_id", "main_sequence",
    "sum"))
setkeyv(ham.clusters.merged.melt, "variable")
ham.clusters.topTen <- setorder(setDT(ham.clusters.merged.melt), -value)[, head(.SD,
    14), keyby = variable]
# ham.clusters.topTen <- ham.clusters.merged.melt[, head(.SD, 15),
# by=variable]
ham.clusters.select <- ham.clusters.merged.melt[ham.clusters.merged.melt$cluster_id %in%
    unique(ham.clusters.topTen$cluster_id)]

ham.clusters.select[, `:=`(geneName, lapply(main_sequence, function(x) select.samples.trsp[J(x)]$GeneName[1]))
ham.clusters.select[, `:=`(listName, paste("Pep:", main_sequence, "from", geneName,
    "cl:", cluster_id, sep = " "))]

select.samples.out <- acast(ham.clusters.select, listName ~ variable, value.var = "value")  #Utilizes reshape
select.samples.out[is.na(select.samples.out)] <- 0
select.samples.out <- select.samples.out[, c(3, 4, 2, 1)]
select.samples.out.scaled <- scale(select.samples.out, center = FALSE, scale = colSums(select.samples.out))
# select.samples.out.scaled <-
# select.samples.out.scaled[order(round(select.samples.out.scaled[,1],digits
# = 2),round(select.samples.out.scaled[,2],digits =
# 2),round(select.samples.out.scaled[,3],digits =
```

```
    collapse = ","), ","))), mc.cores = detectCores()))]
select.samples[, `:=`(Score, BCcount + Animalcount - 1), ]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))

fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/invitroSamplesPeptides.fasta",
    open = "w", nbchar = 60, as.string = TRUE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/HammockInVitro", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar full -i /home/rstudio/data/
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23 --max_shift
# 13 --max_aln_length 37 --count_threshold 50 --max_inner_gaps 0
# --assign_thresholds 14.1,10.5,7.0
hammock.log <- data.table(readLines("data/HammockInVitro/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

| Hammock log file |
|---|
| 2017-11-01 20:05:17.318: |
| Hammock version 1.1.1 Run with –help for a brief description of command line parameters. |
| 2017-11-01 20:05:17.410: Program started in mode “full”. |
| Command-line arguments: |
| full -i /home/rstudio/data/invitroSamplesPeptides.fasta -d /home/rstudio/data/HammockInVitro –max_shift 7 -c 50 -t 32 |
| Complete list of input/output parameters: |
| -i, –input /home/rstudio/data/invitroSamplesPeptides.fasta |
| -d, –output_directory /home/rstudio/data/HammockInVitro |
| -t, –thread 32 |
| -l, –labels null |
| Complete list of clinkage clustering parameters: |
| -f, –file_format fasta |
| -m, –matrix /home/rstudio/Hammock_v_1.1.1/matrices/blosum62.txt |
| -g, –alignment_threshold (–greedy_threshold)null |
| -x, –max_shift 7 |
| -p, –gap_penalty 0 |
| -C, –cache_size_limit 1 |

Hammock log file

2017-11-01 20:05:17.410: Loading input sequences...
2017-11-01 20:05:17.420: 462 unique sequences loaded.
2017-11-01 20:05:17.421: 518 total sequences loaded.
2017-11-01 20:05:17.421: 462 unique sequences after non-specified labels filtered out
2017-11-01 20:05:17.424: 518 total sequences after non-specified labels fileterd out
2017-11-01 20:05:17.424: Shortest sequence: 14 AA. Longest sequence: 22 AA.
2017-11-01 20:05:17.424: Up to 10 000 unique sequences. Using clinkage clustering. Use –use_greedy to force greedy clustering
2017-11-01 20:05:17.427: Generating input statistics...
2017-11-01 20:05:17.429: Clinkage clustering threshold not set. Setting automatically to: 26
2017-11-01 20:05:17.430: Clinkage clustering...
2017-11-01 20:05:17.887: Ready. Clustering time: 457
2017-11-01 20:05:17.888: Resulting clusers: 348
2017-11-01 20:05:17.888: Building MSAs...
2017-11-01 20:05:17.988: Ready. Total time: 558
2017-11-01 20:05:17.988: Saving results to output files...
2017-11-01 20:05:18.021: Clinkage clustering results in: /home/rstudio/data/HammockInVitro/initial_clusters.tsv
2017-11-01 20:05:18.021: and: /home/rstudio/data/HammockInVitro/initial_clusters_sequences.tsv
2017-11-01 20:05:18.021: and: /home/rstudio/data/HammockInVitro/initial_clusters_sequences_original_order.tsv
2017-11-01 20:05:18.021:
Loading clusters...
2017-11-01 20:05:18.026: Maximal alignment length not set. Setting automatically to: 30
2017-11-01 20:05:18.026: Minimal number of match states not set. Setting automatically to: 5
2017-11-01 20:05:18.042: Assign threshold sequence not set. Setting automatically to:
2017-11-01 20:05:18.045: 14.35,11.33,8.31,
2017-11-01 20:05:18.045: Overlap threshold not set. Setting automatically to:
2017-11-01 20:05:18.046: 10.58,6.04,0.0,
2017-11-01 20:05:18.046: Merge threshold not set. Setting automatically based on average sequence length to:
2017-11-01 20:05:18.046: 15.11,13.6,12.09,

Complete list of HMM-based clustering parameters:
-a, –part_threshold null
-s, –size_threshold null
-c, –count_threshold 50
-n, –assign_thresholds 14.35,11.33,8.31,
-v, –overlap_thresholds 10.58,6.04,0.0,
-r, –merge_thresholds 15.11,13.6,12.09,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 30
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-01 20:05:18.062:
Clustering in 3 rounds...
2017-11-01 20:05:18.063:
2017-11-01 20:05:18.063: Round 1:

2017-11-01 20:05:18.063: 50 clusters remaining
2017-11-01 20:05:18.063: Building hmms and searching database...
2017-11-01 20:05:18.359: Extending clusters...
2017-11-01 20:05:18.359: 0 sequences to be inserted into clusters
2017-11-01 20:05:18.359: 0 clusters to be extended

Hammock log file

2017-11-01 20:05:18.360: 0 sequences rejected
2017-11-01 20:05:18.361: 0 cluster pairs to check and merge.
2017-11-01 20:05:18.361: Merging clusters from 0 groups...
2017-11-01 20:05:18.367: Buiding hhs...
2017-11-01 20:05:18.367: HH clustering...
2017-11-01 20:05:18.373:
2017-11-01 20:05:18.373: Round 2:

2017-11-01 20:05:18.373: 50 clusters remaining
2017-11-01 20:05:18.374: Building hmms and searching database...
2017-11-01 20:05:18.611: Extending clusters...
2017-11-01 20:05:18.612: 3 sequences to be inserted into clusters
2017-11-01 20:05:18.612: 3 clusters to be extended
2017-11-01 20:05:18.617: 3 sequences rejected
2017-11-01 20:05:18.618: 9 cluster pairs to check and merge.
2017-11-01 20:05:18.618: Merging clusters from 5 groups...
2017-11-01 20:05:18.624: Buiding hhs...
2017-11-01 20:05:18.638: HH clustering...
2017-11-01 20:05:18.835:
2017-11-01 20:05:18.835: Round 3:

2017-11-01 20:05:18.836: 50 clusters remaining
2017-11-01 20:05:18.836: Building hmms and searching database...
2017-11-01 20:05:19.082: Extending clusters...
2017-11-01 20:05:19.082: 19 sequences to be inserted into clusters
2017-11-01 20:05:19.082: 15 clusters to be extended
2017-11-01 20:05:19.089: 11 sequences rejected
2017-11-01 20:05:19.090: Overlap threshold is 0. Running full cluster merging.
2017-11-01 20:05:19.094: Buiding hhs...
2017-11-01 20:05:19.125: HH clustering...
2017-11-01 20:05:21.650:
Ready. Clustering time : 3588
2017-11-01 20:05:21.650: Resulting clusers: 44
2017-11-01 20:05:21.650: Containing 108 unique sequences and 161 total sequences.
2017-11-01 20:05:21.650: Unique sequences not assigned: 354, total sequences not assigned: 357
2017-11-01 20:05:21.650: Saving results to outupt files...
2017-11-01 20:05:21.661: Results in: /home/rstudio/data/HammockInVitro/final_clusters_sequences.tsv
2017-11-01 20:05:21.662: and: /home/rstudio/data/HammockInVitro/final_clusters.tsv
2017-11-01 20:05:21.662: and: /home/rstudio/data/HammockInVitro/final_clusters_sequences_original_order.tsv
2017-11-01 20:05:21.662:
Calculating KLD...
2017-11-01 20:05:21.662: 13 clusters omitted from KLD calculation because each of them only contains a single unique sequence.
2017-11-01 20:05:21.685: Final system KLD over match state MSA positions: 7.501060786231158
2017-11-01 20:05:21.685: Final system KLD over all MSA positions: 7.227012879783852
2017-11-01 20:05:21.686: Program successfully ended.

## Generation of Weblogo visualization

```
ham.clusters <- data.table(read.table("/home/rstudio/data/HammockInVitro/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
id.order <- as.list(ham.clusters$cluster_id)
ham.clusters.all <- data.table(read.table("/home/rstudio/data/HammockInVitro/final_clusters_sequences.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))
```

```
ham.clusters.all[, `:=`(alignment, gsub("\\-", "\\_", alignment))]
setkey(select.samples, Peptide)
setkey(select.samples.trsp, Peptide)

unlink("/home/rstudio/data/WEBlogosInVitro", recursive = TRUE, force = FALSE)
dir.create(file.path("/home/rstudio/data/", "WEBlogosInVitro"), showWarnings = FALSE)
dir.create(file.path("/home/rstudio/data/HammockInVitro/", "alignments_final_Scored"),
    showWarnings = FALSE)

setkey(ham.clusters.all, cluster_id)
setkey(ham.clusters, cluster_id)
setkey(select.samples.pepMerge, Peptide)


opts_chunk$set(out.width = "100%", fig.align = "center")
generateWeblogo <- function(in.name) {
    # in.name <- ham.clusters$cluster_id[12] in.name <- 6777
    this.fa <- read.fasta(file = paste("/home/rstudio/data/HammockInVitro/alignments_final/",
        in.name, ".aln", sep = ""))
    allSeqs <- unlist(getSequence(this.fa, as.string = TRUE))
    allSeqs <- data.table(unlist(lapply(allSeqs, function(x) gsub("([-])", "",
        toupper(x)))))
    allSeqs.out <- select.samples.pepMerge[J(allSeqs)]
    allSeqs.out$Annot <- data.table(getName(this.fa))
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Score, sep = ""))]
    allSeqs.out$Alignment <- data.table(toupper(unlist(getSequence(this.fa,
        as.string = TRUE))))

    allSeqs.out <- allSeqs.out[rep(1:.N, Score)][, `:=`(Indx, 1:.N), by = Peptide]
    allSeqs.out[, `:=`(Annot, paste(Annot, "_", Indx, sep = ""))]

    write.fasta(as.list(allSeqs.out$Alignment), allSeqs.out$Annot, nbchar = 60,
        paste("/home/rstudio/data/HammockInVitro/alignments_final_Scored/",
            in.name, ".aln", sep = ""), open = "w")


    this.main <- ham.clusters[J(in.name)]
    main.gene <- select.samples.trsp[J(this.main$main_sequence)]$GeneName[1]
    this.title <- paste("## Peptide", this.main$main_sequence, "from", main.gene,
        "with cluster number", in.name, sep = " ")

    tmp <- system(paste("weblogo --format PDF --sequence-type protein --size large --errorbars NO --resolutio
        this.title, "' < /home/rstudio/data/HammockInVitro/alignments_final_Scored/",
        in.name, ".aln > /home/rstudio/data/WEBlogosInVitro/", in.name, ".pdf",
        sep = ""), intern = TRUE, ignore.stdout = FALSE)
}

invisible(mclapply(id.order, generateWeblogo, mc.cores = detectCores()))


ham.clusters.merged <- ham.clusters

# ham.clusters.merged[,mRNA_Str := mRNA_30cpc_Str + mRNA_3cpc_Str +
# mRNA_30cpc_Str_4wks + mRNA_3cpc_Str_4wks] ham.clusters.merged[,mRNA_SN :=
# mRNA_30cpc_SN + mRNA_3cpc_SN + mRNA_30cpc_SN_4wks + mRNA_3cpc_SN_4wks]
# ham.clusters.merged[,mRNA_Th := mRNA_30cpc_Th + mRNA_3cpc_Th +
# mRNA_30cpc_Th_4wks + mRNA_3cpc_Th_4wks] ham.clusters.merged[,mRNA_Ctx :=
# mRNA_30cpc_Ctx + mRNA_3cpc_Ctx + mRNA_30cpc_Ctx_4wks + mRNA_3cpc_Ctx_4wks]
```

```
# ham.clusters.merged[,c('mRNA_30cpc_Str',
# 'mRNA_30cpc_SN','mRNA_30cpc_Th','mRNA_30cpc_Ctx','mRNA_3cpc_Str',
# 'mRNA_3cpc_SN','mRNA_3cpc_Th','mRNA_3cpc_Ctx','mRNA_30cpc_Str_4wks',
# 'mRNA_30cpc_SN_4wks','mRNA_30cpc_Th_4wks','mRNA_30cpc_Ctx_4wks','mRNA_3cpc_Str_4wks','mRNA_3cpc_SN_4wks','m
# := NULL]

library(reshape)
ham.clusters.merged.melt <- melt(ham.clusters.merged, id = c("cluster_id", "main_sequence",
    "sum"))
setkeyv(ham.clusters.merged.melt, "variable")
ham.clusters.topTen <- setorder(setDT(ham.clusters.merged.melt), -value)[, head(.SD,
    8), keyby = variable]
# ham.clusters.topTen <- ham.clusters.merged.melt[, head(.SD, 15),
# by=variable]
ham.clusters.select <- ham.clusters.merged.melt[ham.clusters.merged.melt$cluster_id %in%
    unique(ham.clusters.topTen$cluster_id)]

ham.clusters.select[, `:=`(geneName, lapply(main_sequence, function(x) select.samples.trsp[J(x)]$GeneName[1]))
ham.clusters.select[, `:=`(listName, paste("Pep:", main_sequence, "from", geneName,
    "cl:", cluster_id, sep = " "))]

select.samples.out <- acast(ham.clusters.select, listName ~ variable, value.var = "value")  #Utilizes reshape
select.samples.out[is.na(select.samples.out)] <- 0
select.samples.out <- select.samples.out[, c(2, 3, 1, 4)]
select.samples.out.scaled <- scale(select.samples.out, center = FALSE, scale = colSums(select.samples.out))
# select.samples.out.scaled <-
# select.samples.out.scaled[order(round(select.samples.out.scaled[,1],digits
# = 2),round(select.samples.out.scaled[,2],digits =
# 2),round(select.samples.out.scaled[,3],digits =
# 2),round(select.samples.out.scaled[,4],digits = 2),decreasing=TRUE),]
pheatmap(select.samples.out.scaled, cluster_rows = TRUE, show_rownames = TRUE,
    cluster_cols = FALSE)
```

```
select.samples[, `:=`(Score, BCcount)]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))

fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/DNAsePeptides.fasta",
    open = "w", nbchar = 60, as.string = TRUE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/HammockDNAse", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar full -i /home/rstudio/data/
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23
# --alignment_threshold 26 --assign_thresholds 50,40,30
hammock.log <- data.table(readLines("data/HammockDNAse/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

| Hammock log file |
|---|
| 2017-11-01 20:06:04.268: |
| Hammock version 1.1.1 Run with –help for a brief description of command line parameters. |
| 2017-11-01 20:06:04.398: Program started in mode “full”. |
| Command-line arguments: |
| full -i /home/rstudio/data/DNAsePeptides.fasta -d /home/rstudio/data/HammockDNAse –max_shift 7 -c 2000 -t 32 |
| Complete list of input/output parameters: |
| -i, –input /home/rstudio/data/DNAsePeptides.fasta |
| -d, –output_directory /home/rstudio/data/HammockDNAse |
| -t, –thread 32 |
| -l, –labels null |
| Complete list of clinkage clustering parameters: |
| -f, –file_format fasta |
| -m, –matrix /home/rstudio/Hammock_v_1.1.1/matrices/blosum62.txt |
| -g, –alignment_threshold (–greedy_threshold)null |
| -x, –max_shift 7 |
| -p, –gap_penalty 0 |
| -C, –cache_size_limit 1 |
| 2017-11-01 20:06:04.399: Loading input sequences... |
| 2017-11-01 20:06:04.567: 49840 unique sequences loaded. |
| 2017-11-01 20:06:04.587: 203706 total sequences loaded. |

Hammock log file

2017-11-01 20:06:04.588: 49840 unique sequences after non-specified labels filtered out
2017-11-01 20:06:04.616: 203706 total sequences after non-specified labels fileterd out
2017-11-01 20:06:04.624: Shortest sequence: 14 AA. Longest sequence: 22 AA.
2017-11-01 20:06:04.625: More than 10 000 unique sequences. Using greedy clustering. Use –use_clinkage to force clinkage clustering
2017-11-01 20:06:04.670: Generating input statistics...
2017-11-01 20:06:04.747: Greedy clustering threshold not set. Setting automatically to: 27
2017-11-01 20:06:04.748: Initial greedy clusters limit not set. Setting automatically to: 1246
2017-11-01 20:06:04.749: Greedy clustering...
2017-11-01 20:06:35.369: Ready. Clustering time: 30620
2017-11-01 20:06:35.370: Resulting clusers: 35413
2017-11-01 20:06:35.370: Building MSAs...
2017-11-01 20:06:36.253: Ready. Total time: 31504
2017-11-01 20:06:36.253: Saving results to output files...
2017-11-01 20:06:37.044: Greedy clustering results in: /home/rstudio/data/HammockDNAse/initial_clusters.tsv
2017-11-01 20:06:37.045: and: /home/rstudio/data/HammockDNAse/initial_clusters_sequences.tsv
2017-11-01 20:06:37.045: and: /home/rstudio/data/HammockDNAse/initial_clusters_sequences_original_order.tsv
2017-11-01 20:06:37.045:
Loading clusters...
2017-11-01 20:06:37.169: Maximal alignment length not set. Setting automatically to: 32
2017-11-01 20:06:37.176: Minimal number of match states not set. Setting automatically to: 5
2017-11-01 20:06:37.428: Assign threshold sequence not set. Setting automatically to:
2017-11-01 20:06:37.430: 15.14,11.95,8.77,
2017-11-01 20:06:37.431: Overlap threshold not set. Setting automatically to:
2017-11-01 20:06:37.435: 11.16,6.38,0.0,
2017-11-01 20:06:37.435: Merge threshold not set. Setting automatically based on average sequence length to:
2017-11-01 20:06:37.439: 15.94,14.34,12.75,
2017-11-01 20:06:37.557: 6 clusters rejected because of match states and information content constraints.

Complete list of HMM-based clustering parameters:
-a, –part_threshold null
-s, –size_threshold null
-c, –count_threshold 2000
-n, –assign_thresholds 15.14,11.95,8.77,
-v, –overlap_thresholds 11.16,6.38,0.0,
-r, –merge_thresholds 15.94,14.34,12.75,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 32
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-01 20:06:37.746:
Clustering in 3 rounds...
2017-11-01 20:06:37.748:
2017-11-01 20:06:37.748: Round 1:

2017-11-01 20:06:37.748: 1994 clusters remaining
2017-11-01 20:06:37.748: Building hmms and searching database...
2017-11-01 20:06:59.186: Extending clusters...
2017-11-01 20:06:59.343: 13792 sequences to be inserted into clusters
2017-11-01 20:06:59.350: 1553 clusters to be extended
2017-11-01 20:07:03.859: 10675 sequences rejected

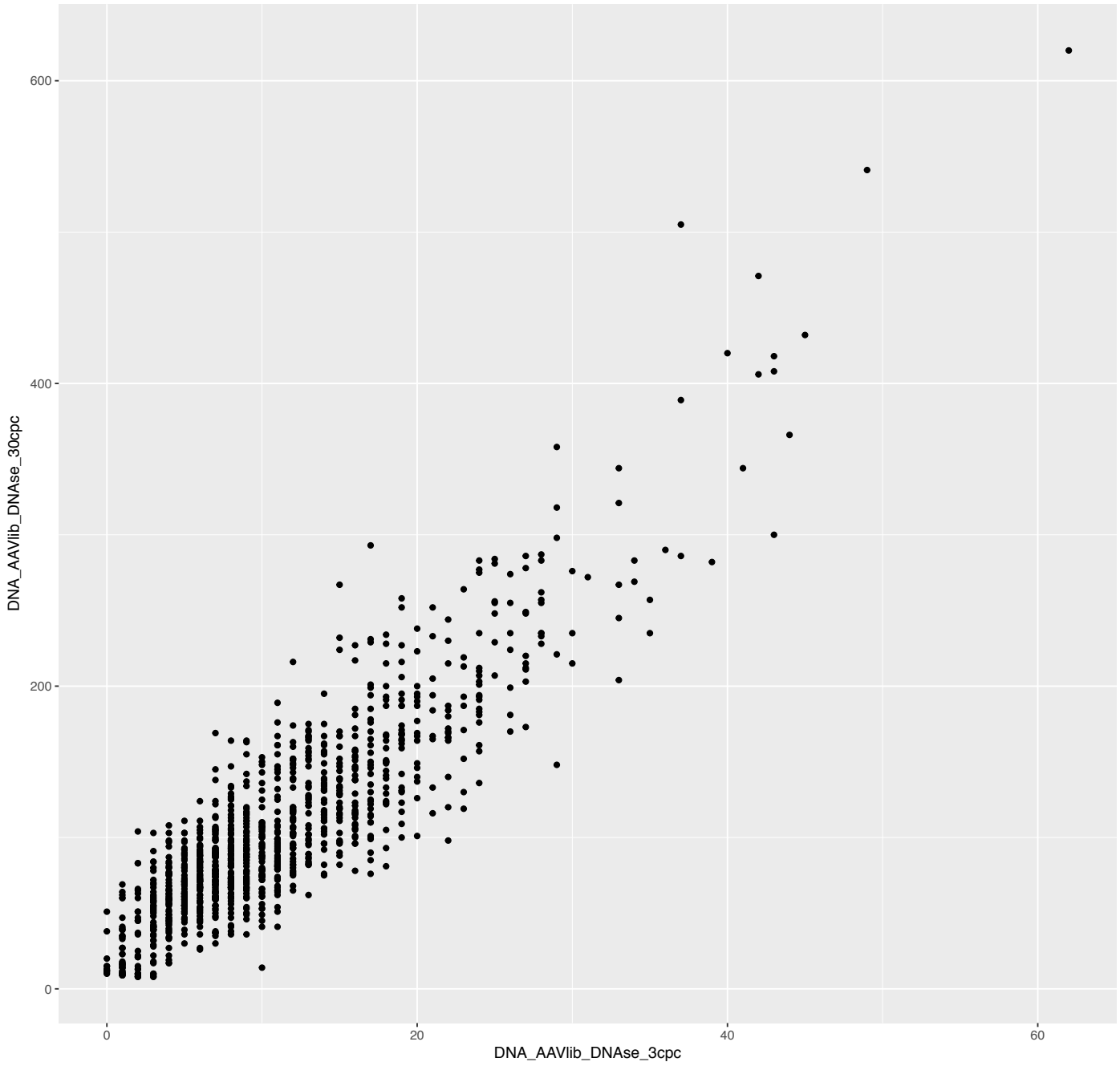
Hammock log file

```
2017-11-01 20:07:03.907: 5089 cluster pairs to check and merge.
2017-11-01 20:07:03.908: Merging clusters from 98 groups...
2017-11-01 20:07:04.070: Buiding hhs...
2017-11-01 20:07:05.077: HH clustering...
2017-11-01 20:09:50.897:
2017-11-01 20:09:50.897: Round 2:

2017-11-01 20:09:50.897: 1530 clusters remaining
2017-11-01 20:09:50.897: Building hmms and searching database...
2017-11-01 20:10:04.082: Extending clusters...
2017-11-01 20:10:04.169: 11386 sequences to be inserted into clusters
2017-11-01 20:10:04.173: 1184 clusters to be extended
2017-11-01 20:10:08.780: 9109 sequences rejected
2017-11-01 20:10:08.963: 51972 cluster pairs to check and merge.
2017-11-01 20:10:08.963: Merging clusters from 1 groups...
2017-11-01 20:10:09.113: Buiding hhs...
2017-11-01 20:10:09.657: HH clustering...
2017-11-01 20:11:29.880:
2017-11-01 20:11:29.881: Round 3:

2017-11-01 20:11:29.881: 1374 clusters remaining
2017-11-01 20:11:29.881: Building hmms and searching database...
2017-11-01 20:11:41.643: Extending clusters...
2017-11-01 20:11:41.693: 14554 sequences to be inserted into clusters
2017-11-01 20:11:41.697: 1204 clusters to be extended
2017-11-01 20:11:48.153: 10703 sequences rejected
2017-11-01 20:11:48.156: Overlap threshold is 0. Running full cluster merging.
2017-11-01 20:11:48.264: Buiding hhs...
2017-11-01 20:11:48.715: HH clustering...
2017-11-01 20:13:36.448:
Ready. Clustering time : 418702
2017-11-01 20:13:36.448: Resulting clusers: 1127
2017-11-01 20:13:36.448: Containing 25589 unique sequences and 134333 total sequences.
2017-11-01 20:13:36.463: Unique sequences not assigned: 24251, total sequences not assigned: 69373
2017-11-01 20:13:36.463: Saving results to outupt files...
2017-11-01 20:13:36.931: Results in: /home/rstudio/data/HammockDNAse/final_clusters_sequences.tsv
2017-11-01 20:13:36.931: and: /home/rstudio/data/HammockDNAse/final_clusters.tsv
2017-11-01 20:13:36.931: and: /home/rstudio/data/HammockDNAse/final_clusters_sequences_original_order.tsv
2017-11-01 20:13:36.931:
Calculating KLD...
2017-11-01 20:13:36.932: 21 clusters omitted from KLD calculation because each of them only contains a single unique
sequence.
2017-11-01 20:13:38.320: Final system KLD over match state MSA positions: 20.23129789753592
2017-11-01 20:13:38.320: Final system KLD over all MSA positions: 36.0292448781723
2017-11-01 20:13:38.320: Program successfully ended.
```

## Generation of Scatter plot generation

```
ham.clusters <- data.table(read.table("/home/rstudio/data/HammockDNAse/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))



pred.points <- ggplot(data = ham.clusters, aes(x = DNA_AAVlib_DNAse_3cpc, y = DNA_AAVlib_DNAse_30cpc)) +
```

```
    labs(x = "DNA_AAVlib_DNAse_3cpc", y = "DNA_AAVlib_DNAse_30cpc") + geom_point()
print(pred.points)
```




## Clustering DNAse resistant virions with library

```
select.samples <- all.samples[J(c("DNA_pscAAVlib", "DNA_pscAAVlib_Prep2"))]

select.samples[, `:=`(BCcount, as.integer(mclapply(BC, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Score, BCcount)]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))
```

```
fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/LibDNAsePeptides.fasta",
    open = "w", nbchar = 60, as.string = TRUE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/HammockLibDNAse", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar full -i /home/rstudio/data/
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23
# --alignment_threshold 26 --assign_thresholds 50,40,30
hammock.log <- data.table(readLines("data/HammockLibDNAse/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

| Hammock log file |
|---|
| 2017-11-01 20:14:05.475: |
| Hammock version 1.1.1 Run with –help for a brief description of command line parameters. |
| 2017-11-01 20:14:05.586: Program started in mode “full”. |
| Command-line arguments: |
| full -i /home/rstudio/data/LibDNAsePeptides.fasta -d /home/rstudio/data/HammockLibDNAse –max_shift 7 -c 2000 -t 32 |
| Complete list of input/output parameters: |
| -i, –input /home/rstudio/data/LibDNAsePeptides.fasta |
| -d, –output_directory /home/rstudio/data/HammockLibDNAse |
| -t, –thread 32 |
| -l, –labels null |
| Complete list of clinkage clustering parameters: |
| -f, –file_format fasta |
| -m, –matrix /home/rstudio/Hammock_v_1.1.1/matrices/blosum62.txt |
| -g, –alignment_threshold (–greedy_threshold)null |
| -x, –max_shift 7 |
| -p, –gap_penalty 0 |
| -C, –cache_size_limit 1 |
| 2017-11-01 20:14:05.587: Loading input sequences... |
| 2017-11-01 20:14:05.825: 60179 unique sequences loaded. |
| 2017-11-01 20:14:05.848: 2906509 total sequences loaded. |
| 2017-11-01 20:14:05.849: 60179 unique sequences after non-specified labels filtered out |
| 2017-11-01 20:14:05.884: 2906509 total sequences after non-specified labels fileterd out |

Hammock log file

2017-11-01 20:14:05.893: Shortest sequence: 14 AA. Longest sequence: 22 AA.
2017-11-01 20:14:05.893: More than 10 000 unique sequences. Using greedy clustering. Use –use_clinkage to force clinkage clustering
2017-11-01 20:14:05.945: Generating input statistics...
2017-11-01 20:14:06.040: Greedy clustering threshold not set. Setting automatically to: 27
2017-11-01 20:14:06.040: Initial greedy clusters limit not set. Setting automatically to: 1504
2017-11-01 20:14:06.042: Greedy clustering...
2017-11-01 20:14:54.443: Ready. Clustering time: 48401
2017-11-01 20:14:54.443: Resulting clusers: 40062
2017-11-01 20:14:54.444: Building MSAs...
2017-11-01 20:14:55.625: Ready. Total time: 49583
2017-11-01 20:14:55.625: Saving results to output files...
2017-11-01 20:14:56.481: Greedy clustering results in: /home/rstudio/data/HammockLibDNAse/initial_clusters.tsv
2017-11-01 20:14:56.482: and: /home/rstudio/data/HammockLibDNAse/initial_clusters_sequences.tsv
2017-11-01 20:14:56.482: and:
/home/rstudio/data/HammockLibDNAse/initial_clusters_sequences_original_order.tsv
2017-11-01 20:14:56.482:
Loading clusters...
2017-11-01 20:14:56.622: Maximal alignment length not set. Setting automatically to: 32
2017-11-01 20:14:56.629: Minimal number of match states not set. Setting automatically to: 5
2017-11-01 20:14:56.895: Assign threshold sequence not set. Setting automatically to:
2017-11-01 20:14:56.897: 15.3,12.08,8.86,
2017-11-01 20:14:56.897: Overlap threshold not set. Setting automatically to:
2017-11-01 20:14:56.901: 11.28,6.44,0.0,
2017-11-01 20:14:56.901: Merge threshold not set. Setting automatically based on average sequence length to:
2017-11-01 20:14:56.905: 16.11,14.5,12.89,
2017-11-01 20:14:57.020: 3 clusters rejected because of match states and information content constraints.

Complete list of HMM-based clustering parameters:
-a, –part_threshold null
-s, –size_threshold null
-c, –count_threshold 2000
-n, –assign_thresholds 15.3,12.08,8.86,
-v, –overlap_thresholds 11.28,6.44,0.0,
-r, –merge_thresholds 16.11,14.5,12.89,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 32
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-01 20:14:57.203:
Clustering in 3 rounds...
2017-11-01 20:14:57.204:
2017-11-01 20:14:57.204: Round 1:

2017-11-01 20:14:57.205: 1997 clusters remaining
2017-11-01 20:14:57.205: Building hmms and searching database...
2017-11-01 20:15:19.930: Extending clusters...
2017-11-01 20:15:20.097: 15752 sequences to be inserted into clusters
2017-11-01 20:15:20.106: 1560 clusters to be extended
2017-11-01 20:15:24.592: 10906 sequences rejected
2017-11-01 20:15:24.643: 4665 cluster pairs to check and merge.

Hammock log file

2017-11-01 20:15:24.644: Merging clusters from 84 groups...
2017-11-01 20:15:24.793: Buiding hhs...
2017-11-01 20:15:25.796: HH clustering...
2017-11-01 20:17:11.700:
2017-11-01 20:17:11.700: Round 2:

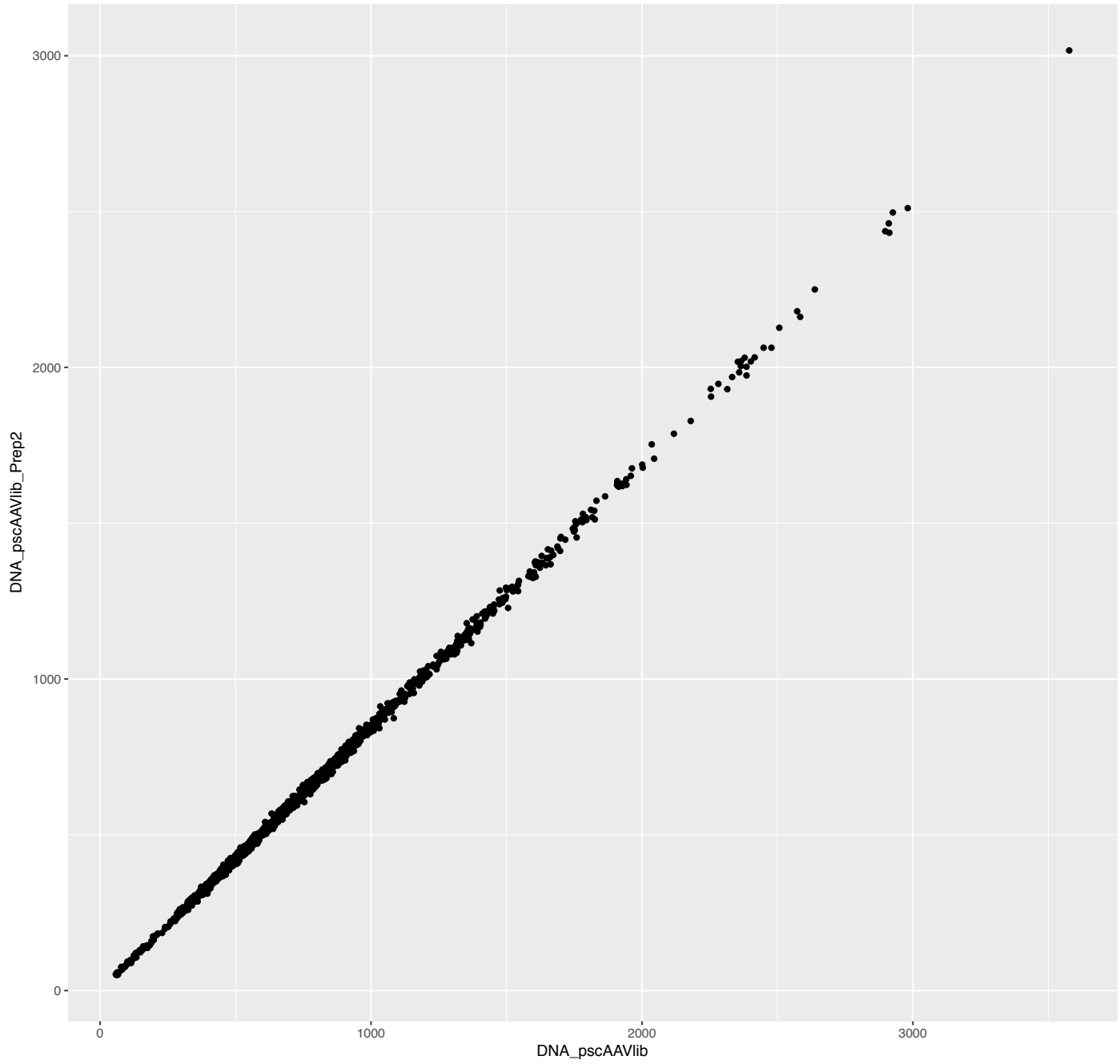
2017-11-01 20:17:11.700: 1739 clusters remaining
2017-11-01 20:17:11.700: Building hmms and searching database...
2017-11-01 20:17:27.477: Extending clusters...
2017-11-01 20:17:27.572: 12758 sequences to be inserted into clusters
2017-11-01 20:17:27.575: 1313 clusters to be extended
2017-11-01 20:17:32.974: 10118 sequences rejected
2017-11-01 20:17:33.166: 67740 cluster pairs to check and merge.
2017-11-01 20:17:33.167: Merging clusters from 1 groups...
2017-11-01 20:17:33.296: Buiding hhs...
2017-11-01 20:17:33.939: HH clustering...
2017-11-01 20:19:09.955:
2017-11-01 20:19:09.956: Round 3:

2017-11-01 20:19:09.956: 1558 clusters remaining
2017-11-01 20:19:09.956: Building hmms and searching database...
2017-11-01 20:19:24.858: Extending clusters...
2017-11-01 20:19:24.921: 16600 sequences to be inserted into clusters
2017-11-01 20:19:24.925: 1331 clusters to be extended
2017-11-01 20:19:32.199: 11821 sequences rejected
2017-11-01 20:19:32.203: Overlap threshold is 0. Running full cluster merging.
2017-11-01 20:19:32.363: Buiding hhs...
2017-11-01 20:19:32.838: HH clustering...
2017-11-01 20:21:16.119:
Ready. Clustering time : 378916
2017-11-01 20:21:16.119: Resulting clusers: 1340
2017-11-01 20:21:16.119: Containing 34315 unique sequences and 1936731 total sequences.
2017-11-01 20:21:16.129: Unique sequences not assigned: 25864, total sequences not assigned: 969778
2017-11-01 20:21:16.130: Saving results to outupt files...
2017-11-01 20:21:16.787: Results in: /home/rstudio/data/HammockLibDNAse/final_clusters_sequences.tsv
2017-11-01 20:21:16.787: and: /home/rstudio/data/HammockLibDNAse/final_clusters.tsv
2017-11-01 20:21:16.787: and: /home/rstudio/data/HammockLibDNAse/final_clusters_sequences_original_order.tsv
2017-11-01 20:21:16.787:
Calculating KLD...
2017-11-01 20:21:16.788: 21 clusters omitted from KLD calculation because each of them only contains a single unique sequence.
2017-11-01 20:21:18.589: Final system KLD over match state MSA positions: 21.284280102889714
2017-11-01 20:21:18.589: Final system KLD over all MSA positions: 39.68035578266766
2017-11-01 20:21:18.589: Program successfully ended.

## Generation of Scatter plot generation

```
ham.clusters <- data.table(read.table("/home/rstudio/data/HammockLibDNAse/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))



pred.points <- ggplot(data = ham.clusters, aes(x = DNA_pscAAVlib, y = DNA_pscAAVlib_Prep2)) +
    labs(x = "DNA_pscAAVlib", y = "DNA_pscAAVlib_Prep2") + geom_point()
```

```
print(pred.points)
```




## Clustering DNAse resistant virions with library

```
select.samples <- all.samples[J(c("DNA_pscAAVlib_Prep2", "DNA_AAVlib_DNAse_3cpc",
    "DNA_AAVlib_DNAse_30cpc"))]

select.samples[, `:=`(BCcount, as.integer(mclapply(BC, function(x) length(table(strsplit(paste(t(x),
    collapse = ","), ","))), mc.cores = detectCores())))]
select.samples[, `:=`(Score, BCcount)]
select.samples.trsp <- unique(select.samples, by = c("Animals", "BC", "LUTnrs"))
```

```
fasta.names <- paste(1:nrow(select.samples.trsp), select.samples.trsp$Score,
    select.samples.trsp$Group, sep = "|")
write.fasta(as.list(select.samples.trsp$Peptide), fasta.names, "data/LibDNAsePeptides.fasta",
    open = "w", nbchar = 60, as.string = TRUE)

# Generate Scoring table for Weblogo Weighting
select.samples.pepMerge <- select.samples.trsp[, sum(Score), by = c("Peptide")]
setnames(select.samples.pepMerge, "V1", "Score")
```

## Executing Hammock Clustering

```
Sys.setenv(PATH = paste("/root/HMMER/binaries", Sys.getenv("PATH"), sep = ":"),
    HHLIB = "/home/rstudio/Hammock_v_1.1.1/hhsuite-2.0.16/lib/hh/")
unlink("/home/rstudio/data/HammockLibDNAse", recursive = TRUE, force = FALSE)
sys.out <- system(paste("java -jar /home/rstudio/Hammock_v_1.1.1/dist/Hammock.jar full -i /home/rstudio/data/
    detectCores(), sep = ""), intern = TRUE, ignore.stdout = TRUE)
# Alternative parameters --use_clinkage --alignment_threshold 23
# --alignment_threshold 26 --assign_thresholds 50,40,30
hammock.log <- data.table(readLines("data/HammockLibDNAse/run.log"))

colnames(hammock.log) <- c("Hammock log file")
knitr::kable(hammock.log, longtable = T)
```

| Hammock log file |
|---|
| 2017-11-01 20:21:44.692: |
| Hammock version 1.1.1 Run with –help for a brief description of command line parameters. |
| 2017-11-01 20:21:44.816: Program started in mode “full”. |
| Command-line arguments: |
| full -i /home/rstudio/data/LibDNAsePeptides.fasta -d /home/rstudio/data/HammockLibDNAse –max_shift 7 -c 2000 -t 32 |
| Complete list of input/output parameters: |
| -i, –input /home/rstudio/data/LibDNAsePeptides.fasta |
| -d, –output_directory /home/rstudio/data/HammockLibDNAse |
| -t, –thread 32 |
| -l, –labels null |
| Complete list of clinkage clustering parameters: |
| -f, –file_format fasta |
| -m, –matrix /home/rstudio/Hammock_v_1.1.1/matrices/blosum62.txt |
| -g, –alignment_threshold (–greedy_threshold)null |
| -x, –max_shift 7 |
| -p, –gap_penalty 0 |
| -C, –cache_size_limit 1 |
| 2017-11-01 20:21:44.817: Loading input sequences... |
| 2017-11-01 20:21:45.109: 60086 unique sequences loaded. |
| 2017-11-01 20:21:45.138: 1535104 total sequences loaded. |
| 2017-11-01 20:21:45.138: 60086 unique sequences after non-specified labels filtered out |
| 2017-11-01 20:21:45.182: 1535104 total sequences after non-specified labels fileterd out |

Hammock log file

2017-11-01 20:21:45.194: Shortest sequence: 14 AA. Longest sequence: 22 AA.
2017-11-01 20:21:45.194: More than 10 000 unique sequences. Using greedy clustering. Use –use_clinkage to force clinkage clustering
2017-11-01 20:21:45.260: Generating input statistics...
2017-11-01 20:21:45.380: Greedy clustering threshold not set. Setting automatically to: 27
2017-11-01 20:21:45.380: Initial greedy clusters limit not set. Setting automatically to: 1502
2017-11-01 20:21:45.381: Greedy clustering...
2017-11-01 20:22:34.960: Ready. Clustering time: 49578
2017-11-01 20:22:34.961: Resulting clusers: 39583
2017-11-01 20:22:34.961: Building MSAs...
2017-11-01 20:22:36.116: Ready. Total time: 50734
2017-11-01 20:22:36.117: Saving results to output files...
2017-11-01 20:22:37.032: Greedy clustering results in: /home/rstudio/data/HammockLibDNAse/initial_clusters.tsv
2017-11-01 20:22:37.033: and: /home/rstudio/data/HammockLibDNAse/initial_clusters_sequences.tsv
2017-11-01 20:22:37.033: and:
/home/rstudio/data/HammockLibDNAse/initial_clusters_sequences_original_order.tsv
2017-11-01 20:22:37.033:
Loading clusters...
2017-11-01 20:22:37.176: Maximal alignment length not set. Setting automatically to: 32
2017-11-01 20:22:37.184: Minimal number of match states not set. Setting automatically to: 5
2017-11-01 20:22:37.458: Assign threshold sequence not set. Setting automatically to:
2017-11-01 20:22:37.461: 15.3,12.08,8.86,
2017-11-01 20:22:37.461: Overlap threshold not set. Setting automatically to:
2017-11-01 20:22:37.467: 11.27,6.44,0.0,
2017-11-01 20:22:37.467: Merge threshold not set. Setting automatically based on average sequence length to:
2017-11-01 20:22:37.472: 16.1,14.49,12.88,
2017-11-01 20:22:37.602: 6 clusters rejected because of match states and information content constraints.

Complete list of HMM-based clustering parameters:
-a, –part_threshold null
-s, –size_threshold null
-c, –count_threshold 2000
-n, –assign_thresholds 15.3,12.08,8.86,
-v, –overlap_thresholds 11.27,6.44,0.0,
-r, –merge_thresholds 16.1,14.49,12.88,
-e, –relative_thresholds false
-b, –absolute_thresholds true
-h, –min_conserved_positions 5
-y, –max_gap_proportion 0.05
-k, –min_ic 1.2
-j, –max_aln_length 32
-u, –max_inner_gaps 0
-q, –extension_increase_length false

2017-11-01 20:22:37.792:
Clustering in 3 rounds...
2017-11-01 20:22:37.794:
2017-11-01 20:22:37.794: Round 1:

2017-11-01 20:22:37.794: 1994 clusters remaining
2017-11-01 20:22:37.794: Building hmms and searching database...
2017-11-01 20:22:59.916: Extending clusters...
2017-11-01 20:23:00.059: 16358 sequences to be inserted into clusters
2017-11-01 20:23:00.069: 1568 clusters to be extended
2017-11-01 20:23:06.086: 11310 sequences rejected
2017-11-01 20:23:06.152: 4703 cluster pairs to check and merge.

Hammock log file

2017-11-01 20:23:06.152: Merging clusters from 84 groups...
2017-11-01 20:23:06.364: Buiding hhs...
2017-11-01 20:23:07.377: HH clustering...
2017-11-01 20:24:53.748:
2017-11-01 20:24:53.748: Round 2:

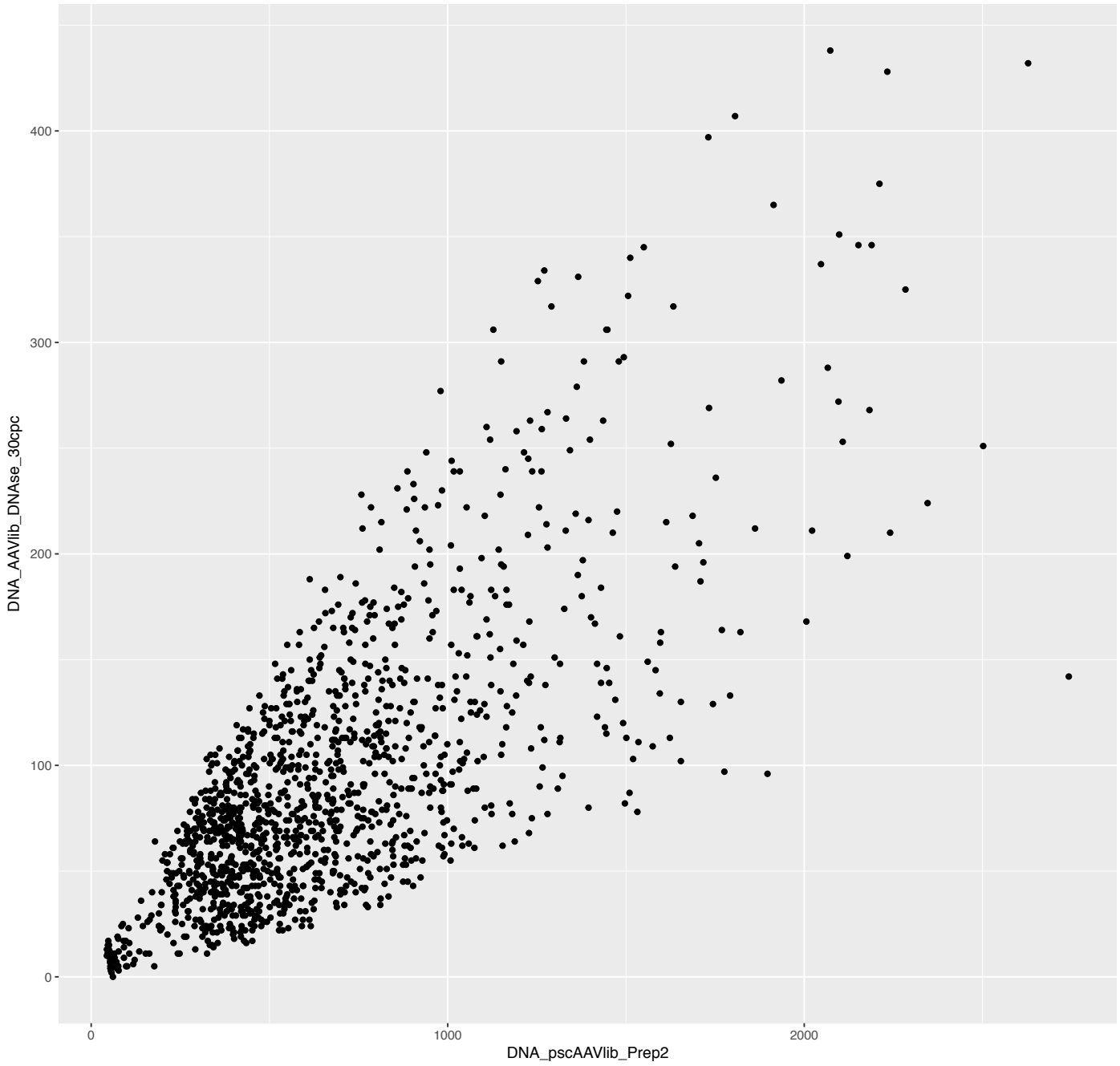
2017-11-01 20:24:53.749: 1735 clusters remaining
2017-11-01 20:24:53.749: Building hmms and searching database...
2017-11-01 20:25:09.238: Extending clusters...
2017-11-01 20:25:09.343: 12559 sequences to be inserted into clusters
2017-11-01 20:25:09.346: 1300 clusters to be extended
2017-11-01 20:25:14.878: 10059 sequences rejected
2017-11-01 20:25:15.073: 72128 cluster pairs to check and merge.
2017-11-01 20:25:15.074: Merging clusters from 1 groups...
2017-11-01 20:25:15.204: Buiding hhs...
2017-11-01 20:25:15.794: HH clustering...
2017-11-01 20:26:42.914:
2017-11-01 20:26:42.914: Round 3:

2017-11-01 20:26:42.914: 1571 clusters remaining
2017-11-01 20:26:42.914: Building hmms and searching database...
2017-11-01 20:26:58.109: Extending clusters...
2017-11-01 20:26:58.175: 16562 sequences to be inserted into clusters
2017-11-01 20:26:58.179: 1360 clusters to be extended
2017-11-01 20:27:05.805: 11873 sequences rejected
2017-11-01 20:27:05.809: Overlap threshold is 0. Running full cluster merging.
2017-11-01 20:27:05.940: Buiding hhs...
2017-11-01 20:27:06.451: HH clustering...
2017-11-01 20:28:52.048:
Ready. Clustering time : 374255
2017-11-01 20:28:52.048: Resulting clusers: 1345
2017-11-01 20:28:52.049: Containing 34643 unique sequences and 1030396 total sequences.
2017-11-01 20:28:52.057: Unique sequences not assigned: 25443, total sequences not assigned: 504708
2017-11-01 20:28:52.057: Saving results to outupt files...
2017-11-01 20:28:52.678: Results in: /home/rstudio/data/HammockLibDNAse/final_clusters_sequences.tsv
2017-11-01 20:28:52.678: and: /home/rstudio/data/HammockLibDNAse/final_clusters.tsv
2017-11-01 20:28:52.678: and: /home/rstudio/data/HammockLibDNAse/final_clusters_sequences_original_order.tsv
2017-11-01 20:28:52.678:
Calculating KLD...
2017-11-01 20:28:52.679: 31 clusters omitted from KLD calculation because each of them only contains a single unique sequence.
2017-11-01 20:28:54.606: Final system KLD over match state MSA positions: 21.314423213245473
2017-11-01 20:28:54.606: Final system KLD over all MSA positions: 39.98132228244945
2017-11-01 20:28:54.606: Program successfully ended.

## Generation of Scatter plot generation

```
ham.clusters <- data.table(read.table("/home/rstudio/data/HammockLibDNAse/final_clusters.tsv",
    header = TRUE, skip = 0, sep = "\t", stringsAsFactors = FALSE, fill = TRUE))



pred.points <- ggplot(data = ham.clusters, aes(x = DNA_pscAAVlib_Prep2, y = DNA_AAVlib_DNAse_30cpc)) +
    labs(x = "DNA_pscAAVlib_Prep2", y = "DNA_AAVlib_DNAse_30cpc") + geom_point()
```

```
print(pred.points)
```




```
print("Total analysis time:")
```

```
[1] "Total analysis time:"
```

```
print(Sys.time() - strt1)
```

```
Time difference of 24.77801 mins
```

```
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     2017-11-01

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

 package              * version  date       source
 acepack                1.4.1    2016-10-29 CRAN (R 3.4.2)
 ade4                   1.7-8    2017-08-09 CRAN (R 3.4.2)
 annotate               1.54.0   2017-10-13 Bioconductor
 AnnotationDbi          1.38.2   2017-10-13 Bioconductor
 AnnotationFilter       1.0.0    2017-10-13 Bioconductor
 AnnotationHub          2.8.2    2017-10-13 Bioconductor
 backports              1.1.1    2017-09-25 CRAN (R 3.4.2)
 base                 * 3.4.2    2017-10-06 local
 base64enc              0.1-3    2015-07-28 CRAN (R 3.4.2)
 Biobase              * 2.36.2   2017-10-13 Bioconductor
 BiocGenerics         * 0.22.1   2017-10-13 Bioconductor
 BiocInstaller          1.26.1   2017-10-10 Bioconductor
 BiocParallel           1.10.1   2017-10-13 Bioconductor
 biomaRt                2.32.1   2017-10-13 Bioconductor
 Biostrings           * 2.44.2   2017-10-13 Bioconductor
 biovizBase             1.24.0   2017-10-13 Bioconductor
 bit                    1.1-12   2014-04-09 CRAN (R 3.4.2)
 bit64                  0.9-7    2017-05-08 CRAN (R 3.4.2)
 bitops                 1.0-6    2013-08-17 CRAN (R 3.4.2)
 blob                   1.1.0    2017-06-17 CRAN (R 3.4.2)
 BSgenome               1.44.2   2017-10-13 Bioconductor
 checkmate              1.8.4    2017-09-25 CRAN (R 3.4.2)
 cluster                2.0.6    2017-03-16 CRAN (R 3.4.2)
 codetools              0.2-15   2016-10-05 CRAN (R 3.4.2)
 colorspace             1.3-2    2016-12-14 CRAN (R 3.4.2)
 compiler               3.4.2    2017-10-06 local
 curl                   2.8.1    2017-07-21 CRAN (R 3.4.2)
 data.table           * 1.10.4-2 2017-10-12 CRAN (R 3.4.2)
 datasets             * 3.4.2    2017-10-06 local
 DBI                    0.7      2017-06-18 CRAN (R 3.4.2)
 DelayedArray         * 0.2.7    2017-10-13 Bioconductor
 DESeq2               * 1.16.1   2017-10-13 Bioconductor
 devtools             * 1.13.3   2017-08-02 CRAN (R 3.4.2)
 dichromat              2.0-0    2013-01-24 CRAN (R 3.4.2)
 digest                 0.6.12   2017-01-27 CRAN (R 3.4.2)
 doParallel           * 1.0.11   2017-09-28 CRAN (R 3.4.2)
 ensembldb              2.0.4    2017-10-13 Bioconductor
 evaluate               0.10.1   2017-06-24 CRAN (R 3.4.2)
 foreach              * 1.4.3    2015-10-13 CRAN (R 3.4.2)
 foreign                0.8-69   2017-06-21 CRAN (R 3.4.2)
 formatR              * 1.5      2017-04-25 CRAN (R 3.4.2)
 Formula                1.2-2    2017-07-10 CRAN (R 3.4.2)
 futile.logger        * 1.4.3    2016-07-10 CRAN (R 3.4.2)
 futile.options         1.0.0    2010-04-06 CRAN (R 3.4.2)
 genefilter             1.58.1   2017-10-13 Bioconductor
 geneplotter            1.54.0   2017-10-13 Bioconductor
 GenomeInfoDb         * 1.12.3   2017-10-13 Bioconductor
 GenomeInfoDbData       0.99.0   2017-10-13 Bioconductor
 GenomicAlignments    * 1.12.2   2017-10-13 Bioconductor
 GenomicFeatures        1.28.5   2017-10-13 Bioconductor
 GenomicRanges        * 1.28.6   2017-10-13 Bioconductor
```

```
GGally                   1.3.2    2017-08-02 CRAN (R 3.4.2)
ggbio                  * 1.24.1   2017-10-13 Bioconductor
ggplot2                * 2.2.1    2016-12-30 CRAN (R 3.4.2)
graph                    1.54.0   2017-10-13 Bioconductor
graphics               * 3.4.2    2017-10-06 local
grDevices              * 3.4.2    2017-10-06 local
grid                   * 3.4.2    2017-10-06 local
gridExtra                2.3      2017-09-09 CRAN (R 3.4.2)
gtable                   0.2.0    2016-02-26 CRAN (R 3.4.2)
highr                    0.6      2016-05-09 CRAN (R 3.4.2)
Hmisc                    4.0-3    2017-05-02 CRAN (R 3.4.2)
hms                      0.3      2016-11-22 CRAN (R 3.4.2)
htmlTable                1.9      2017-01-26 CRAN (R 3.4.2)
htmltools                0.3.6    2017-04-28 CRAN (R 3.4.2)
htmlwidgets              0.9      2017-07-10 CRAN (R 3.4.2)
httpuv                   1.3.5    2017-07-04 CRAN (R 3.4.2)
httr                     1.3.1    2017-08-20 CRAN (R 3.4.2)
interactiveDisplayBase   1.14.0   2017-10-13 Bioconductor
IRanges                * 2.10.5   2017-10-13 Bioconductor
iterators              * 1.0.8    2015-10-13 CRAN (R 3.4.2)
kableExtra             * 0.5.2    2017-09-15 CRAN (R 3.4.2)
knitr                  * 1.17     2017-08-10 CRAN (R 3.4.2)
labeling                 0.3      2014-08-23 CRAN (R 3.4.2)
lambda.r                 1.2      2017-09-16 CRAN (R 3.4.2)
lattice                  0.20-35  2017-03-25 CRAN (R 3.4.2)
latticeExtra             0.6-28   2016-02-09 CRAN (R 3.4.2)
lazyeval                 0.2.0    2016-06-12 CRAN (R 3.4.2)
locfit                   1.5-9.1  2013-04-20 CRAN (R 3.4.2)
magrittr                 1.5      2014-11-22 CRAN (R 3.4.2)
Matrix                   1.2-11   2017-08-16 CRAN (R 3.4.2)
matrixStats            * 0.52.2   2017-04-14 CRAN (R 3.4.2)
memoise                  1.1.0    2017-04-21 CRAN (R 3.4.2)
methods                * 3.4.2    2017-10-06 local
mime                     0.5      2016-07-07 CRAN (R 3.4.2)
munsell                  0.4.3    2016-02-13 CRAN (R 3.4.2)
nnet                     7.3-12   2016-02-02 CRAN (R 3.4.2)
OrganismDbi              1.18.1   2017-10-13 Bioconductor
parallel               * 3.4.2    2017-10-06 local
pheatmap               * 1.0.8    2015-12-11 CRAN (R 3.4.2)
plyr                   * 1.8.4    2016-06-08 CRAN (R 3.4.2)
ProtGenerics             1.8.0    2017-10-13 Bioconductor
R6                       2.2.2    2017-06-17 CRAN (R 3.4.2)
RBGL                     1.52.0   2017-10-13 Bioconductor
RColorBrewer             1.1-2    2014-12-07 CRAN (R 3.4.2)
Rcpp                     0.12.13  2017-09-28 CRAN (R 3.4.2)
RCurl                    1.95-4.8 2016-03-01 CRAN (R 3.4.2)
readr                    1.1.1    2017-05-16 CRAN (R 3.4.2)
reshape                * 0.8.7    2017-08-06 CRAN (R 3.4.2)
reshape2               * 1.4.2    2016-10-22 CRAN (R 3.4.2)
rlang                    0.1.2    2017-08-09 CRAN (R 3.4.2)
rmarkdown                1.6      2017-06-15 CRAN (R 3.4.2)
rpart                    4.1-11   2017-04-21 CRAN (R 3.4.2)
rprojroot                1.2      2017-01-16 CRAN (R 3.4.2)
Rsamtools              * 1.28.0   2017-10-13 Bioconductor
RSQLite                  2.0      2017-06-19 CRAN (R 3.4.2)
rtracklayer              1.36.6   2017-10-13 Bioconductor
rvest                    0.3.2    2016-06-17 CRAN (R 3.4.2)
S4Vectors              * 0.14.7   2017-10-13 Bioconductor
scales                   0.5.0    2017-08-24 CRAN (R 3.4.2)
```

```
seqinr                * 3.4-5    2017-08-01 CRAN (R 3.4.2)
shiny                   1.0.5    2017-08-23 CRAN (R 3.4.2)
splines                 3.4.2    2017-10-06 local
stats                 * 3.4.2    2017-10-06 local
stats4                * 3.4.2    2017-10-06 local
stringi                 1.1.5    2017-04-07 CRAN (R 3.4.2)
stringr                 1.2.0    2017-02-18 CRAN (R 3.4.2)
SummarizedExperiment  * 1.6.5    2017-10-13 Bioconductor
survival                2.41-3   2017-04-04 CRAN (R 3.4.2)
tibble                  1.3.4    2017-08-22 CRAN (R 3.4.2)
tools                   3.4.2    2017-10-06 local
utils                 * 3.4.2    2017-10-06 local
VariantAnnotation       1.22.3   2017-10-13 Bioconductor
VennDiagram           * 1.6.17   2016-04-18 CRAN (R 3.4.2)
withr                   2.0.0    2017-07-28 CRAN (R 3.4.2)
XML                     3.98-1.9 2017-06-19 CRAN (R 3.4.2)
xml2                    1.1.1    2017-01-24 CRAN (R 3.4.2)
xtable                  1.8-2    2016-02-05 CRAN (R 3.4.2)
XVector               * 0.16.0   2017-10-13 Bioconductor
yaml                    2.1.14   2016-11-12 CRAN (R 3.4.2)
zlibbioc                1.22.0   2017-10-13 Bioconductor
```