

> {epistack}

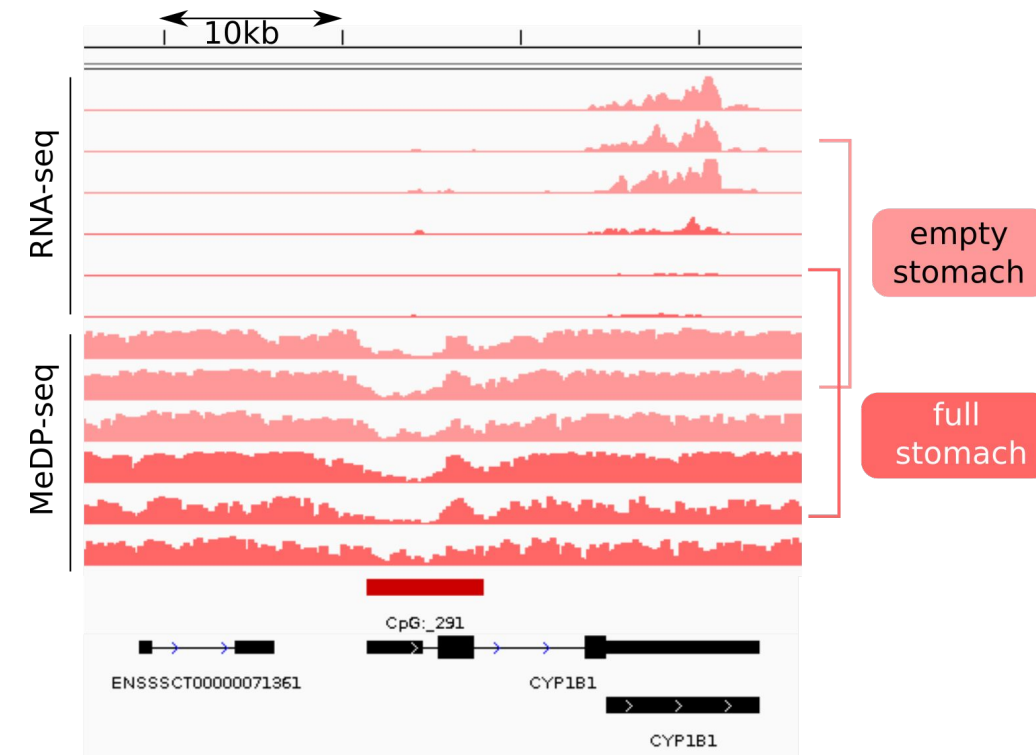
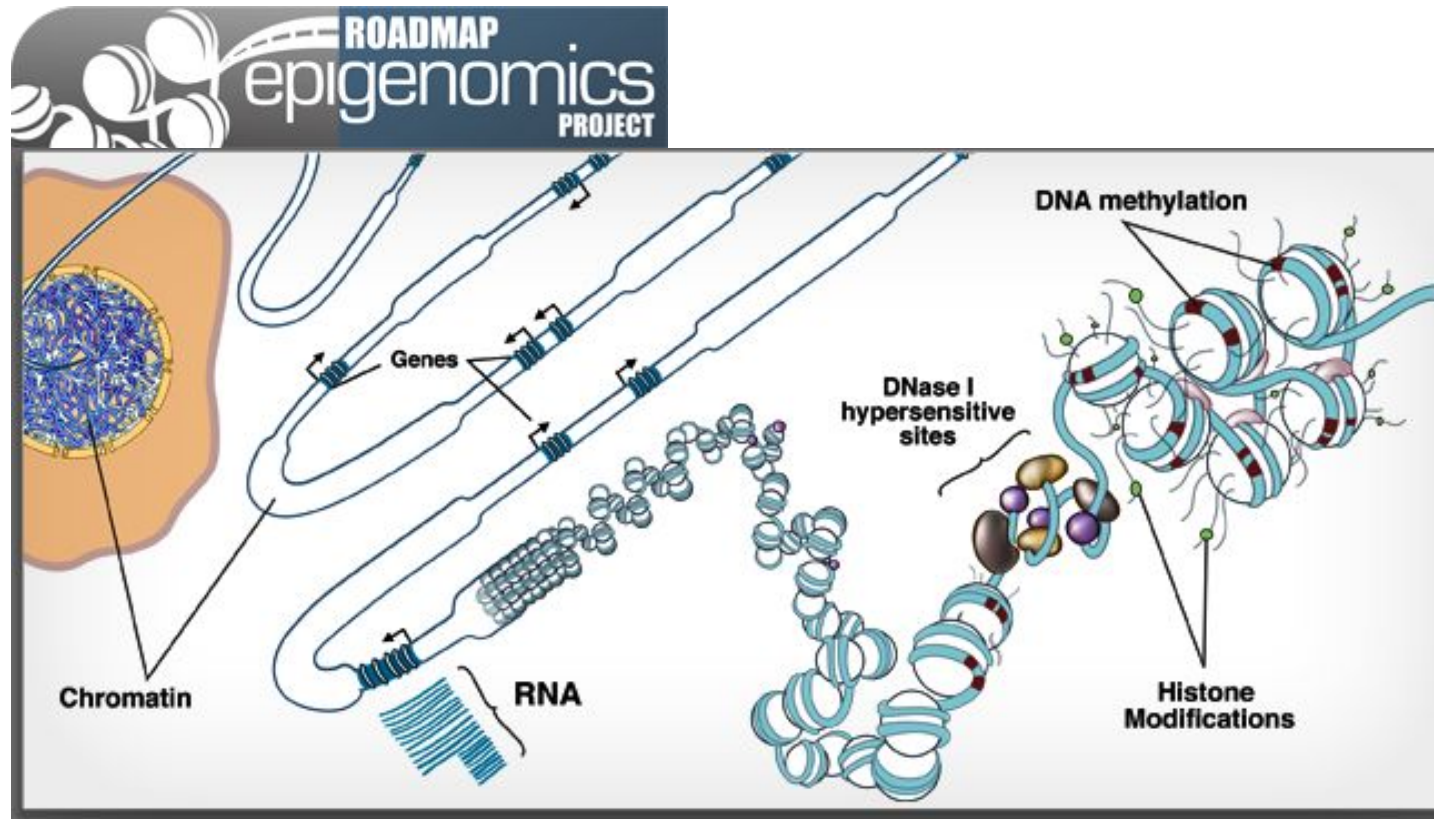
An R package to visualise stack profiles of epigenomic signals

Safia Saci & Guillaume Devailly

Genotoul Biostat Bioinfo day 2021

Epigenomic data

Tracks of genomic scores



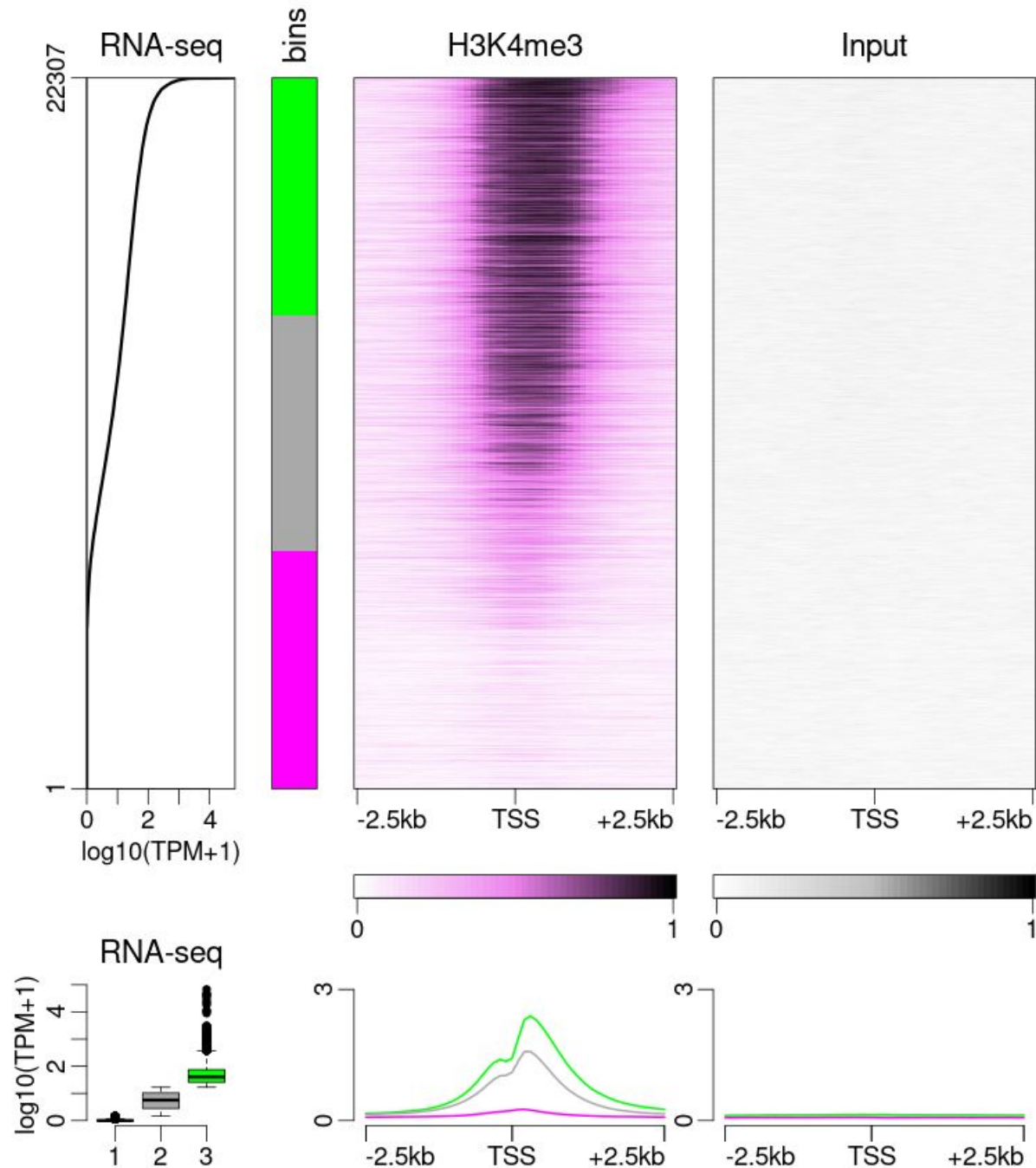
{epistack} epigenetic stacks

Visualise stacks of epigenetic signals on anchor regions:

- ◇ Gene starts
- ◇ Peak center
- ◇ CpG islands
- ◇ Differentially methylated regions
- ◇ ...

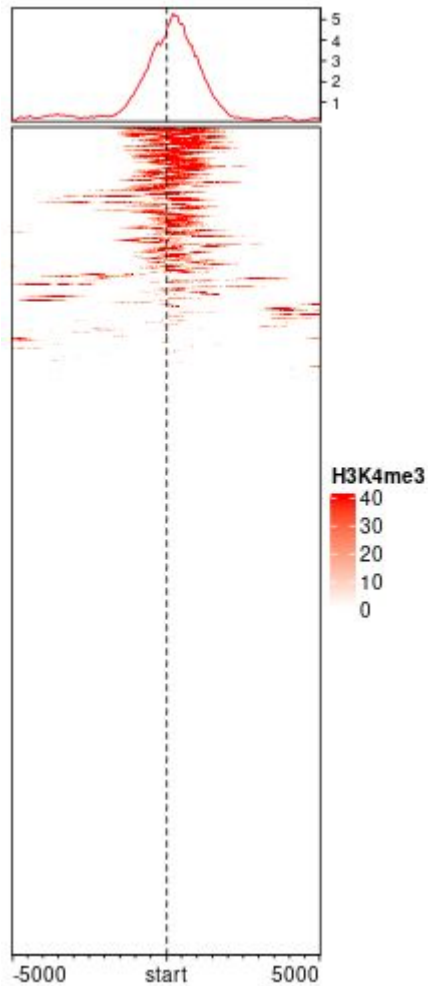
Sort regions:

- ◇ Gene expression levels
- ◇ P-values
- ◇ Fold changes
- ◇ Clustering
- ◇ Region widths
- ◇ Distance to closest TSS
- ◇ Gene types
- ◇ ...

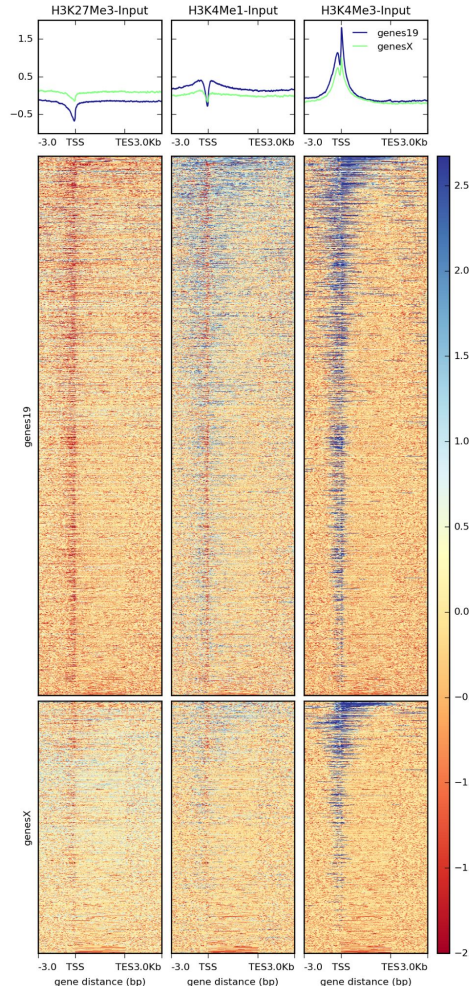


{epistack} alternatives

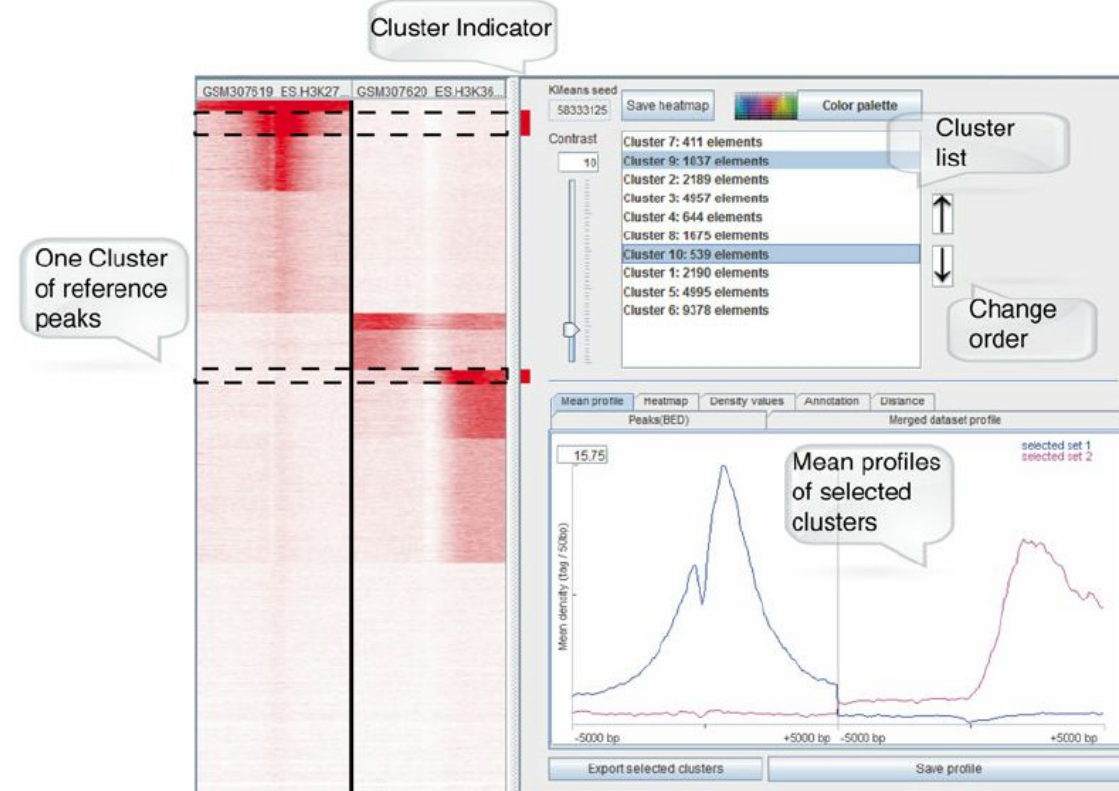
EnrichedHeatmap
(bioconductor)



deepTools plotHeatmap
(CLI - python)



seqMINER
(GUI - java)



And also: seqplots, Repitools, ChIPseeker, ...
@Bioconductor

{epistack} is a visualisation package

Input: A GRanges object, with signal matrices embedded as metadata-columns

```
library(GenomicRanges)
library(epistack)

data("stackepi")
dim(mcols(stackepi))
#> [1] 693 54
stackepi[, 1:6]
#> GRanges object with 693 ranges and 6 metadata columns:
#>
#>      seqnames      ranges strand |      gene_id
#>      <Rle>        <IRanges> <Rle> |      <character>
#>  ENSSSCG00000016737      18 50563134-50566857      - | ENSSSCG00000016737
#>  ENSSSCG00000036350      18 46205448-46215858      + | ENSSSCG00000036350
#>  ENSSSCG00000037869      18 23881967-23924133      + | ENSSSCG00000037869
#>  ENSSSCG00000016444      18  6334193-6348959      + | ENSSSCG00000016444
#>  ENSSSCG00000016714      18 47169867-47173866      + | ENSSSCG00000016714
#>
#>      ...      ...      ...      ...      ...
#>      exp      score window_1 window_2 window_3
#>      <numeric> <numeric> <numeric> <numeric> <numeric>
#>  ENSSSCG00000016737 1213.478      0  0.910984  0.889177  0.879568
#>  ENSSSCG00000036350  771.328      0  0.388921  0.371259  0.425591
#>  ENSSSCG00000037869  270.641      0  0.595544  0.621201  0.616013
#>  ENSSSCG00000016444  261.168      0  0.881246  0.843490  0.706214
#>  ENSSSCG00000016714  193.318      0  0.877790  0.884266  0.899063
#>
#>      ...      ...      ...      ...      ...
```



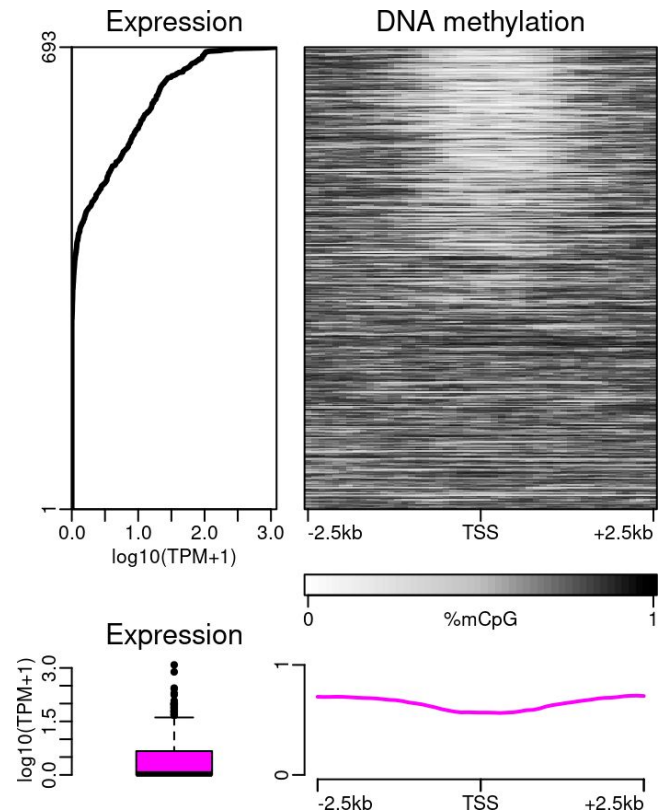
{epistack} is a visualisation package

Input: A GRanges object, with signal matrices embedded as metadata-columns

```
library(GenomicRanges)
library(epistack)

data("stackepi")
dim(mcols(stackepi))
#> [1] 693 54
stackepi[, 1:6]
#> GRanges object with 693 ranges and 6 metadata columns:
#>
#>           seqnames           ranges strand |           gene_id
#>           <Rle>             <IRanges> <Rle> |           <character>
#> ENSSSCG00000016737      18 50563134-50566857   - | ENSSSCG00000016737
#> ENSSSCG00000036350      18 46205448-46215858   + | ENSSSCG00000036350
#> ENSSSCG00000037869      18 23881967-23924133   + | ENSSSCG00000037869
#> ENSSSCG00000016444      18  6334193-6348959    + | ENSSSCG00000016444
#> ENSSSCG00000016714      18 47169867-47173866   + | ENSSSCG00000016714
#>
#>           ...           ...           ...           ...           ...
#>           exp          score window_1 window_2 window_3
#>           <numeric> <numeric> <numeric> <numeric> <numeric>
#> ENSSSCG00000016737 1213.478         0 0.910984 0.889177 0.879568
#> ENSSSCG00000036350  771.328         0 0.388921 0.371259 0.425591
#> ENSSSCG00000037869  270.641         0 0.595544 0.621201 0.616013
#> ENSSSCG00000016444  261.168         0 0.881246 0.843490 0.706214
#> ENSSSCG00000016714  193.318         0 0.877790 0.884266 0.899063
#>
#>           ...           ...           ...           ...           ...
```

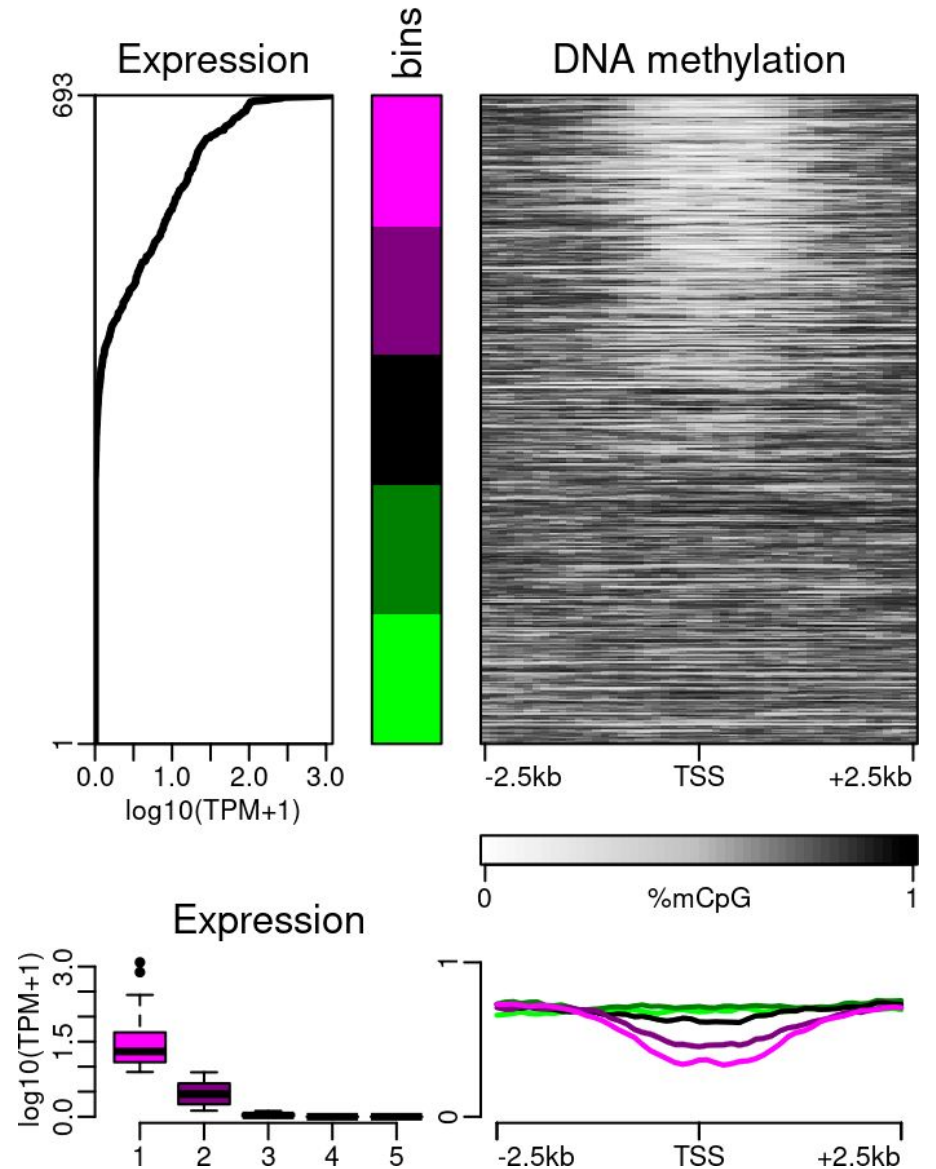
```
plotEpistack(
  stackepi,
  pattern = "^window_", metric_col = "exp",
  ylim = c(0, 1), zlim = c(0, 1),
  x_labels = c("-2.5kb", "TSS", "+2.5kb"),
  titles = "DNA methylation", legends = "%mCpG",
  metric_title = "Expression", metric_label = "log10(TPM+1)",
  metric_transfunc = function(x) log10(x+1)
)
```



plotEpistack() parameters highlights: bins

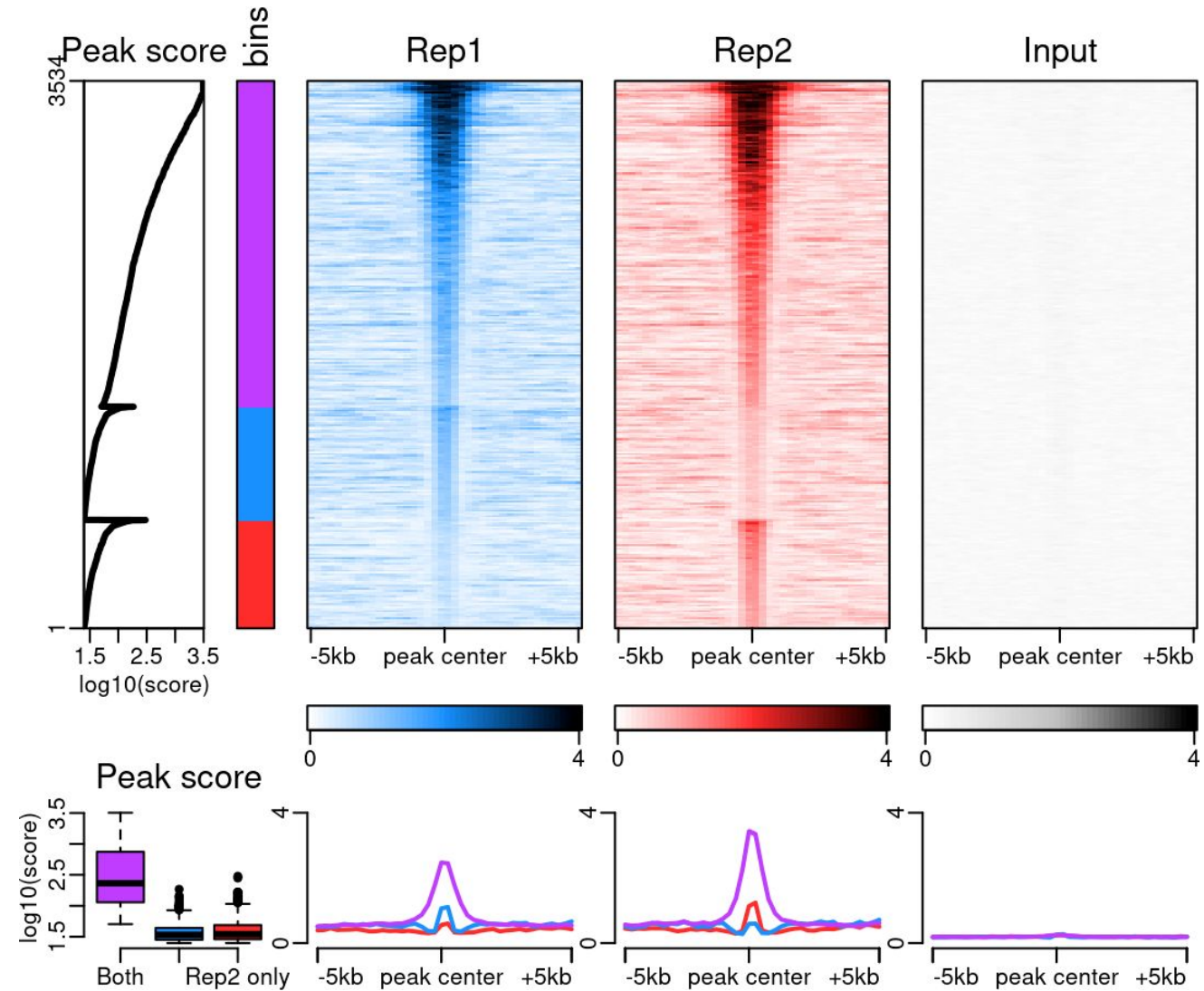
```
stackepi <- addBins(stackepi, nbins = 5)

plotEpistack(
  stackepi,
  pattern = "^window_", metric_col = "exp",
  ylim = c(0, 1), zlim = c(0, 1),
  x_labels = c("-2.5kb", "TSS", "+2.5kb"),
  titles = "DNA methylation", legends = "%mCpG",
  metric_title = "Expression", metric_label = "log10(TPM+1)",
  metric_transfunc = function(x) log10(x+1)
)
```



plotEpistack() parameters highlights: several tracks

```
plotEpistack(  
  meDP,  
  pattern = c("Rep1", "Rep2", "input"),  
  ...  
)
```



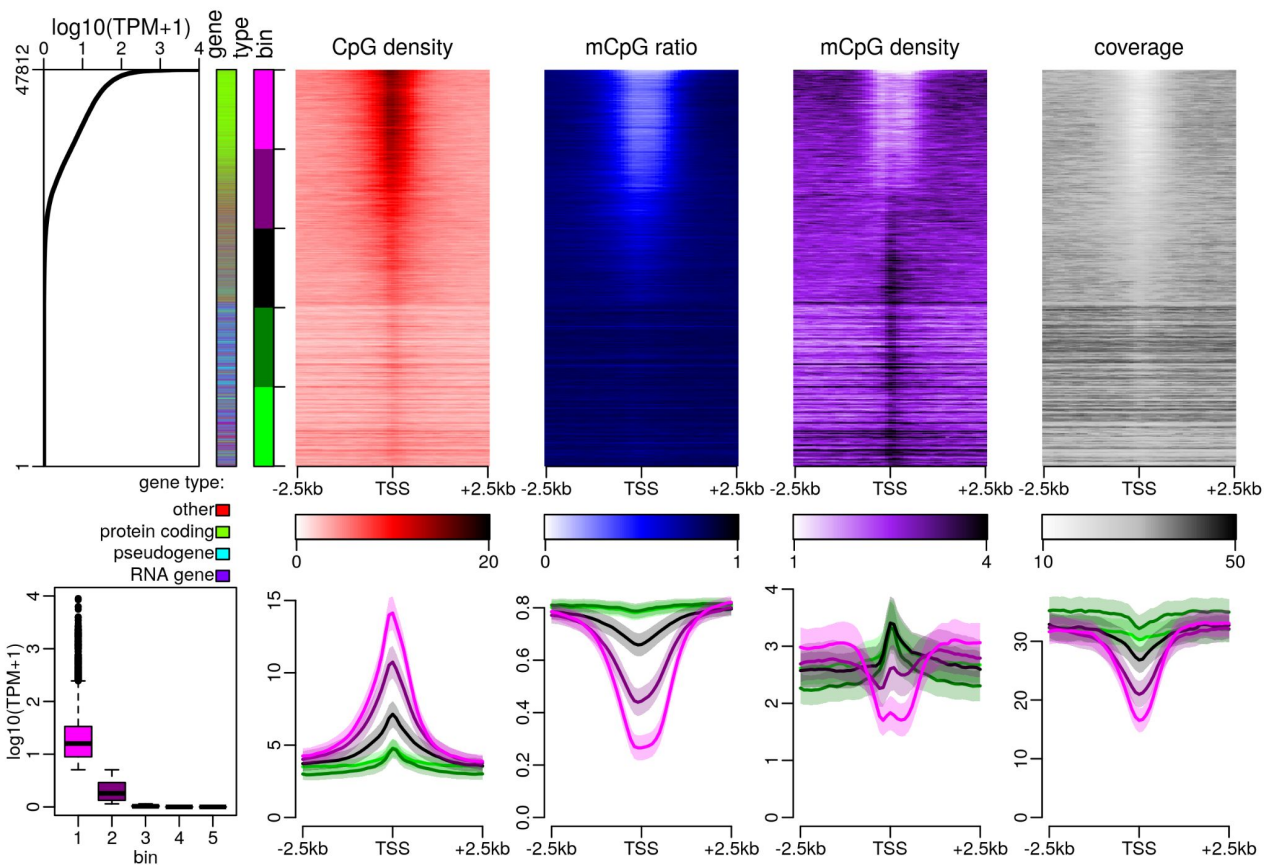
Building epistack's input GRanges

- epigenetic tracks:
 - .bam, .bigwig
 - load into R with `GenomicAlignment::readGAlignment()` or `rtracklayer::import()`
- anchors:
 - .bed, .gtf/.gff
 - load into R with `rtracklayer::import()`
- get epigenetic stacks matrices
 - `EnrichedHeatmap::normalizeToMatrix()`
 - `Repitools::annotationCounts()`
 - `ChIPseeker::getTagMatrix()`
- an additional ordering vector
 - gene expression, p-value, fold change, etc.
 - `epistack::addMetricAndArrangeGRanges()`



Use case: WGBS coverage at TSS

ROADMAP, H1 cell line



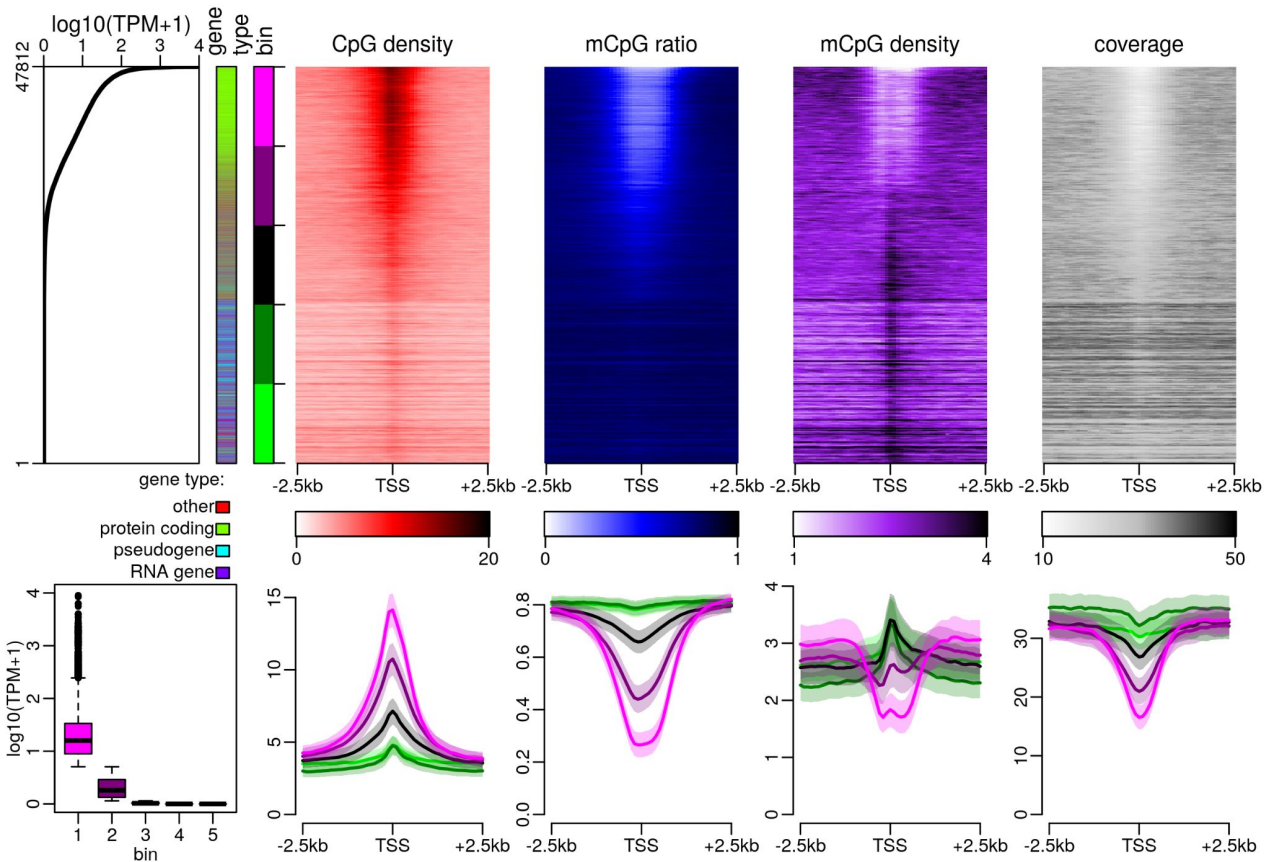
INRAE

{epistack}

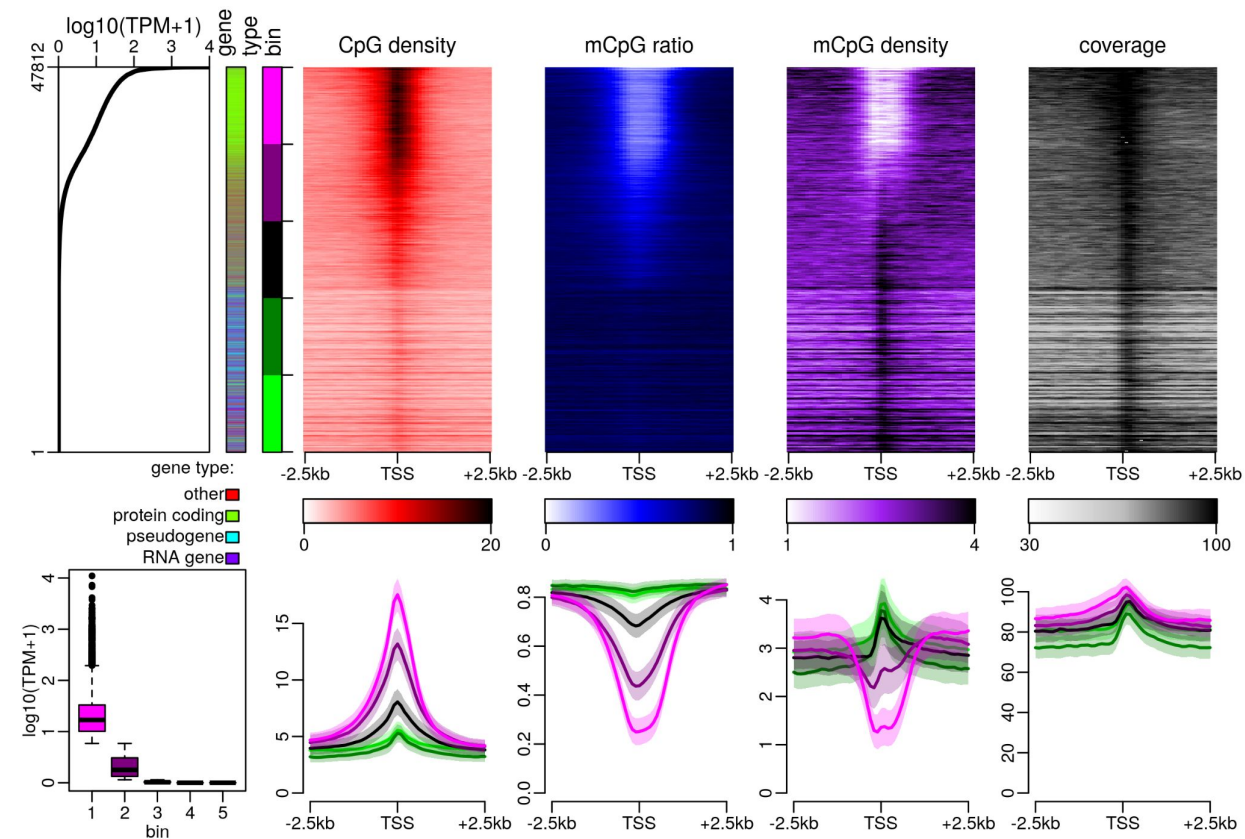
2021-10-12 / Genotoul biostat bioinfo dat / Guillaume Devailly

Use case: WGBS coverage at TSS

ROADMAP, H1 cell line



ROADMAP, ESC derived CD56+ Ectoderm



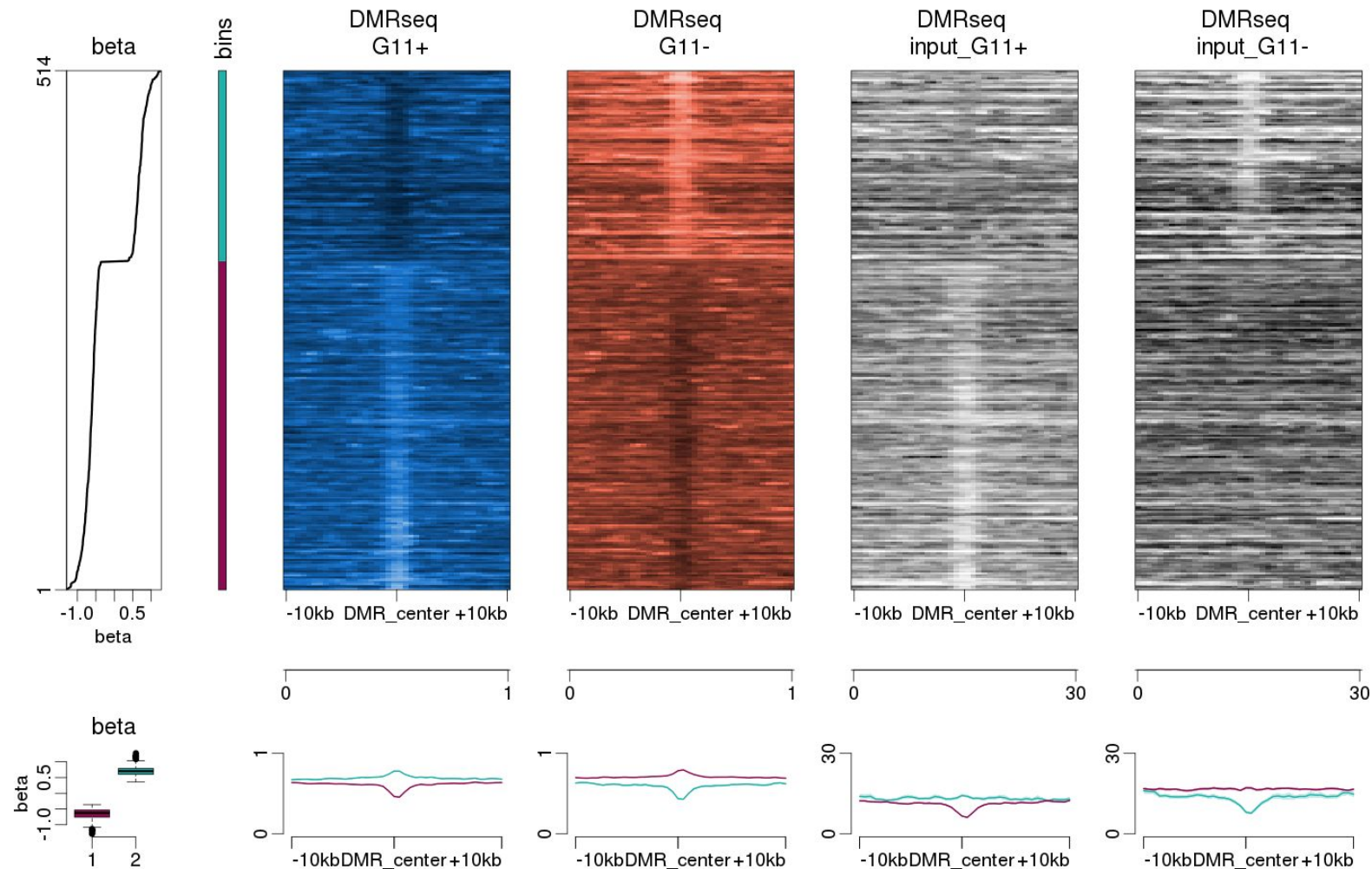
INRAE

{epistack}

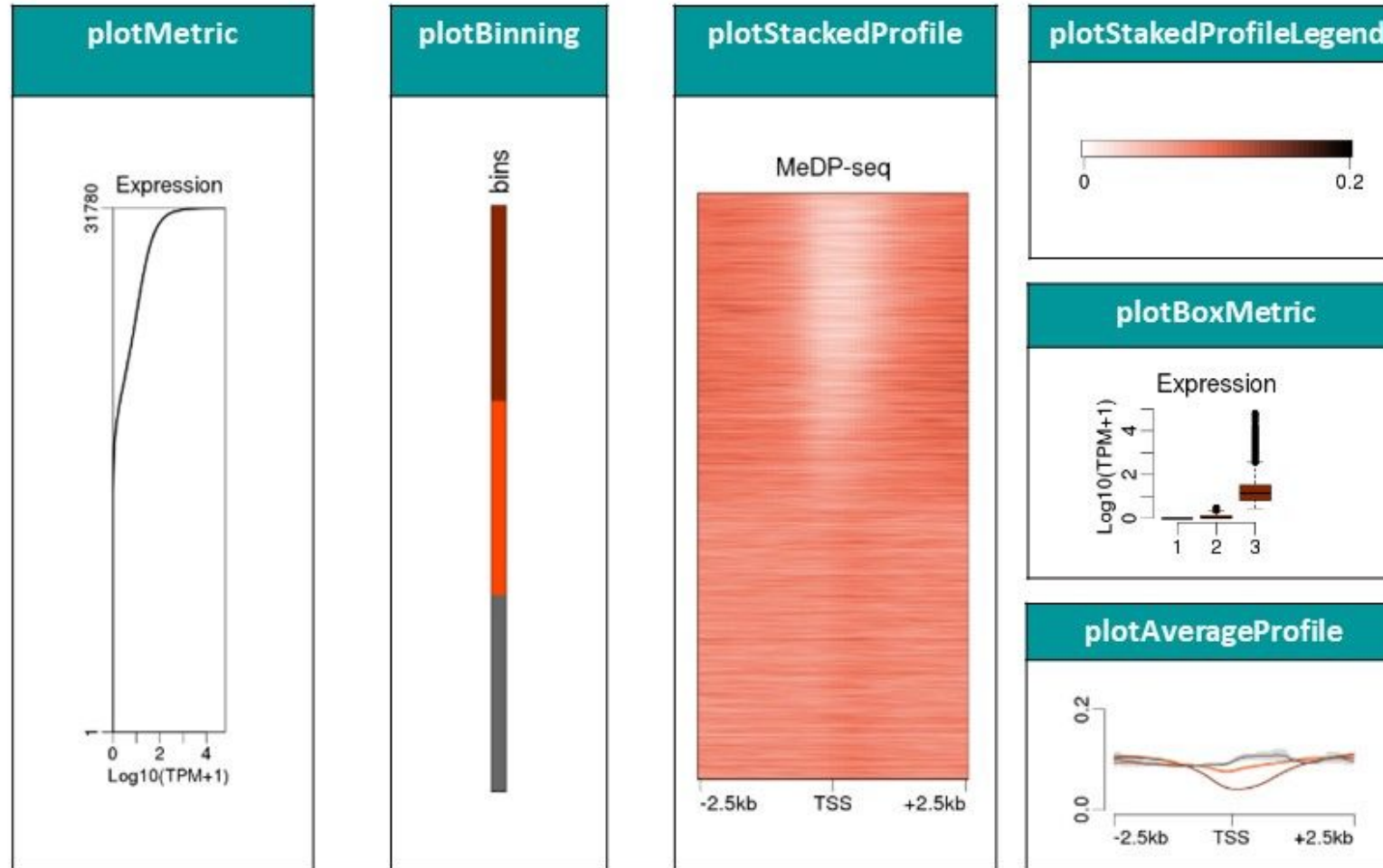
2021-10-12 / Genotoul biostat bioinfo dat / Guillaume Devailly

Use case: visualisations of DMR

Differentially Methylated Regions



Individual plotting functions

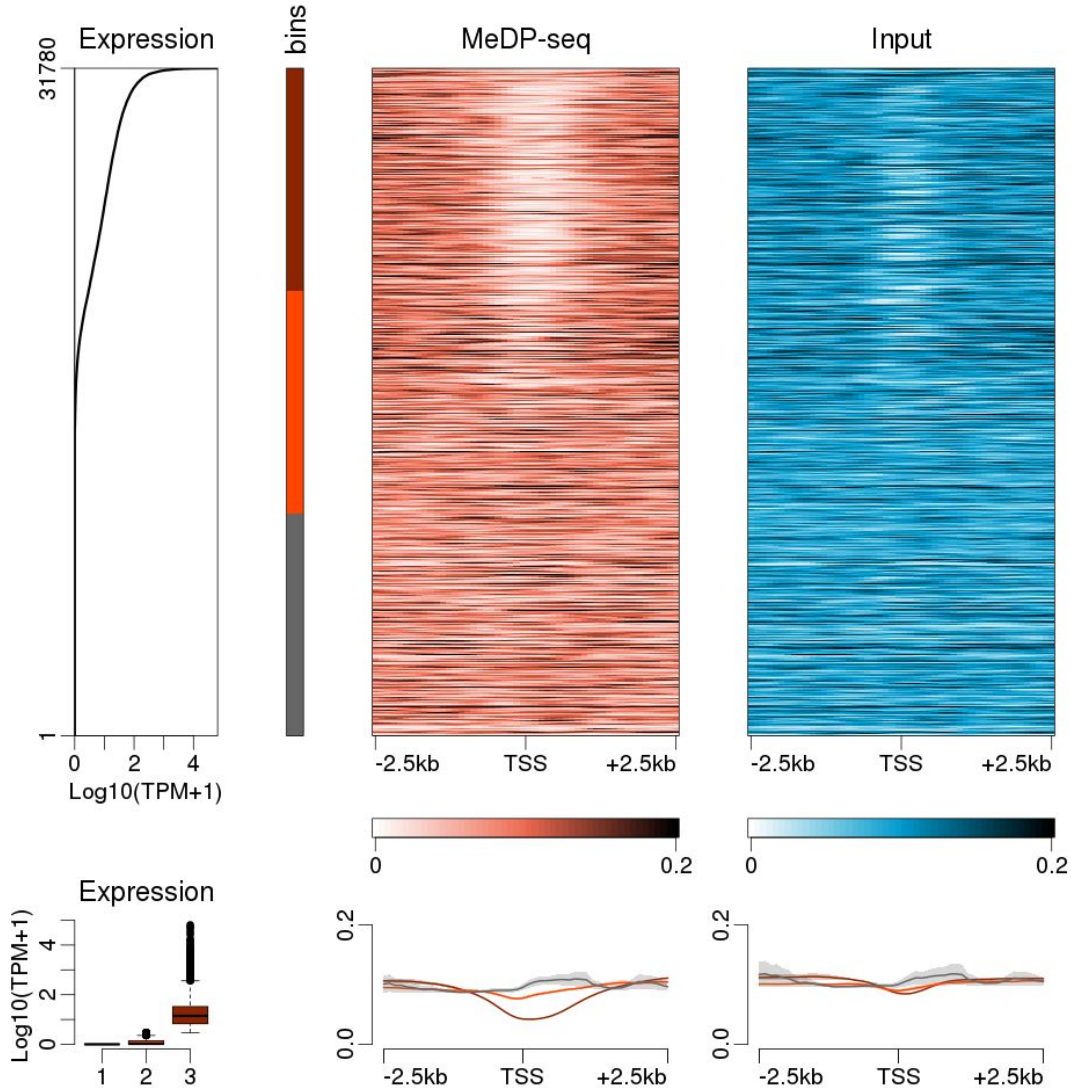


Assemble panels in the order
you wish

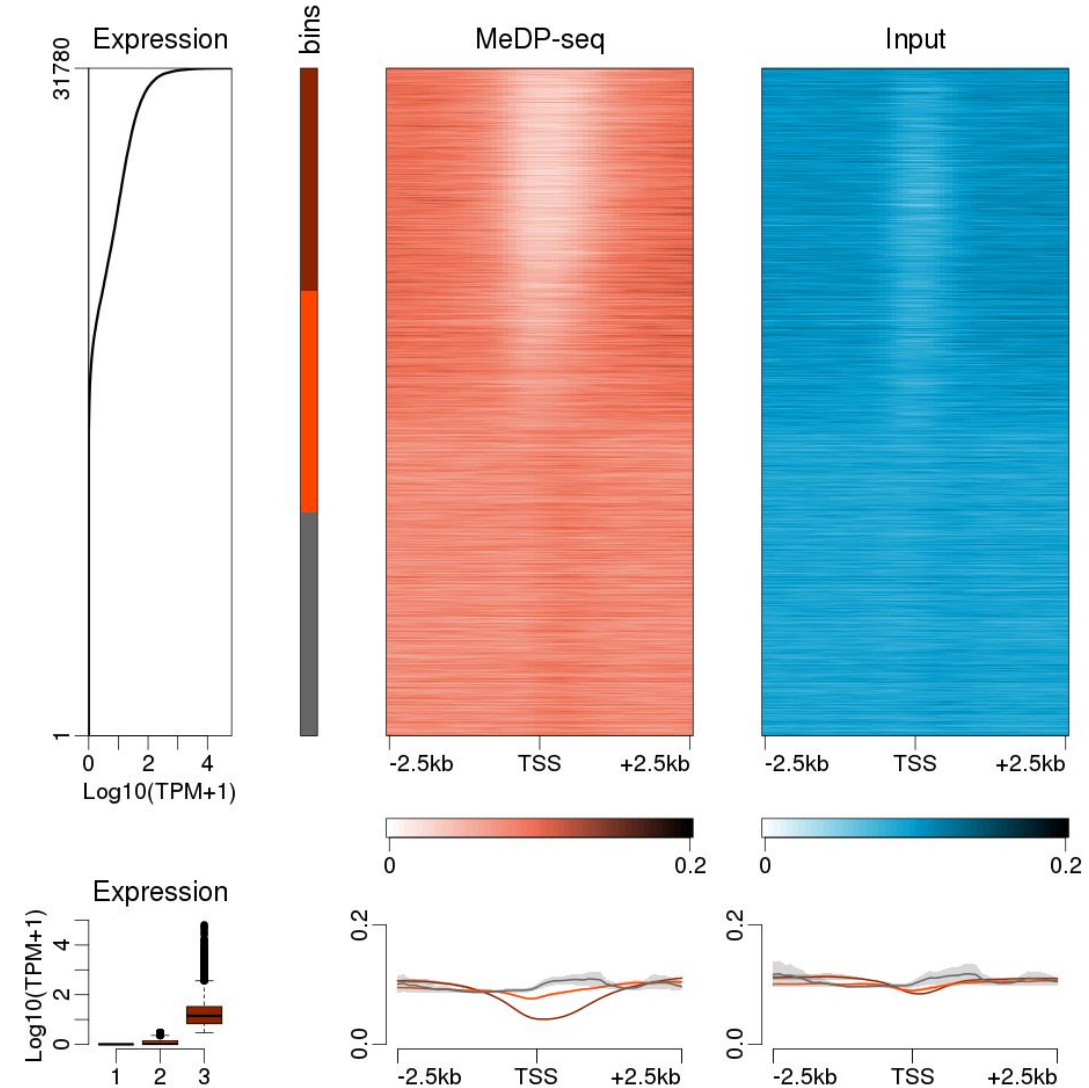
```
layout()  
{gridExtra}  
{patchwork}
```


The overplotting issue: more regions than pixels

Default R behaviour

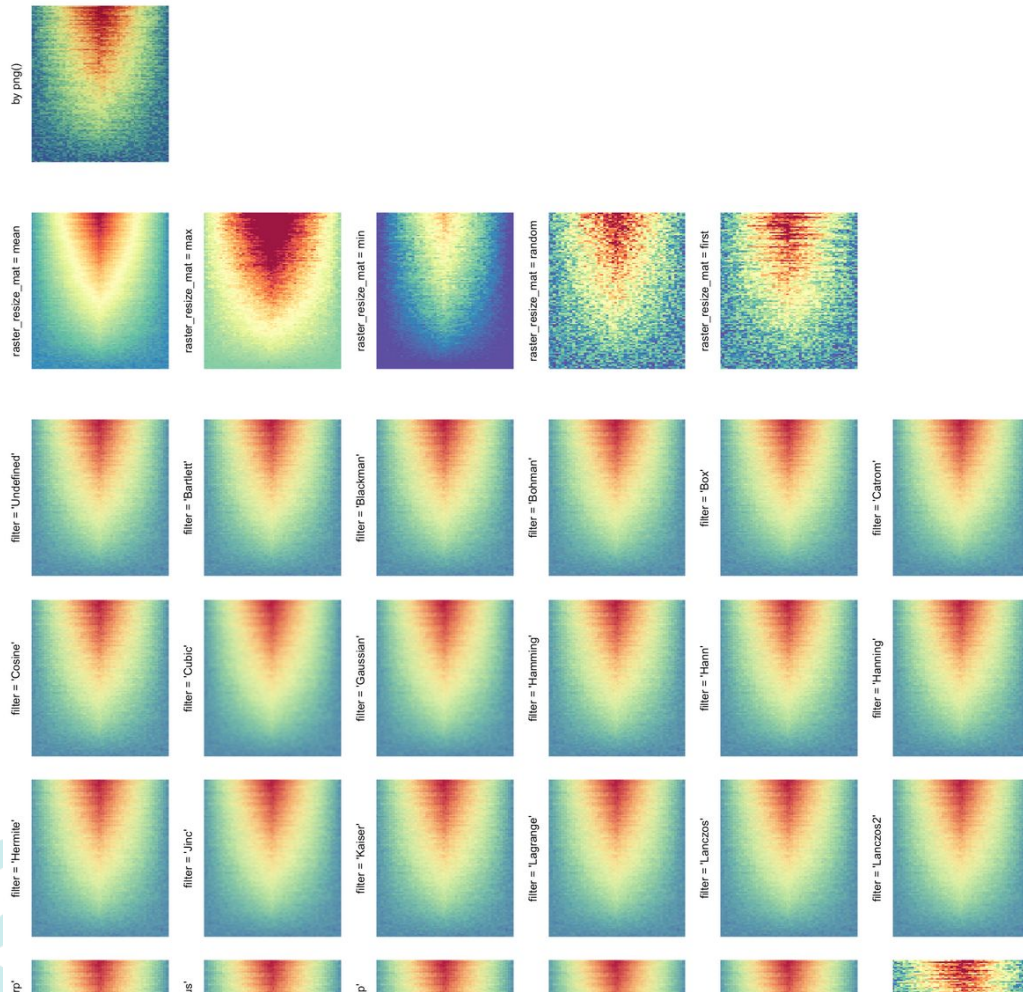


Default {epistack} behaviour

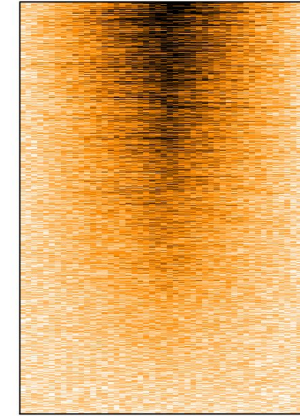


The overplotting issue: more regions than pixels

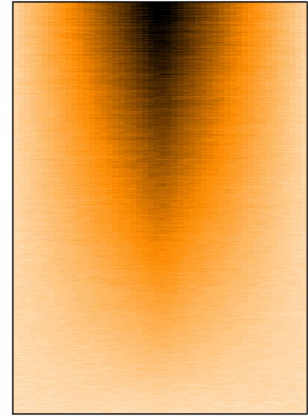
- `epistack::redimMatrix(mat, target_height, target_width, summary_func)`
- bioinfo-fr.net/creer-des-heatmaps-a-partir-de-grosses-matrices-en-r
- jokergoo.github.io/2020/06/30/rasterization-in-complexheatmap/



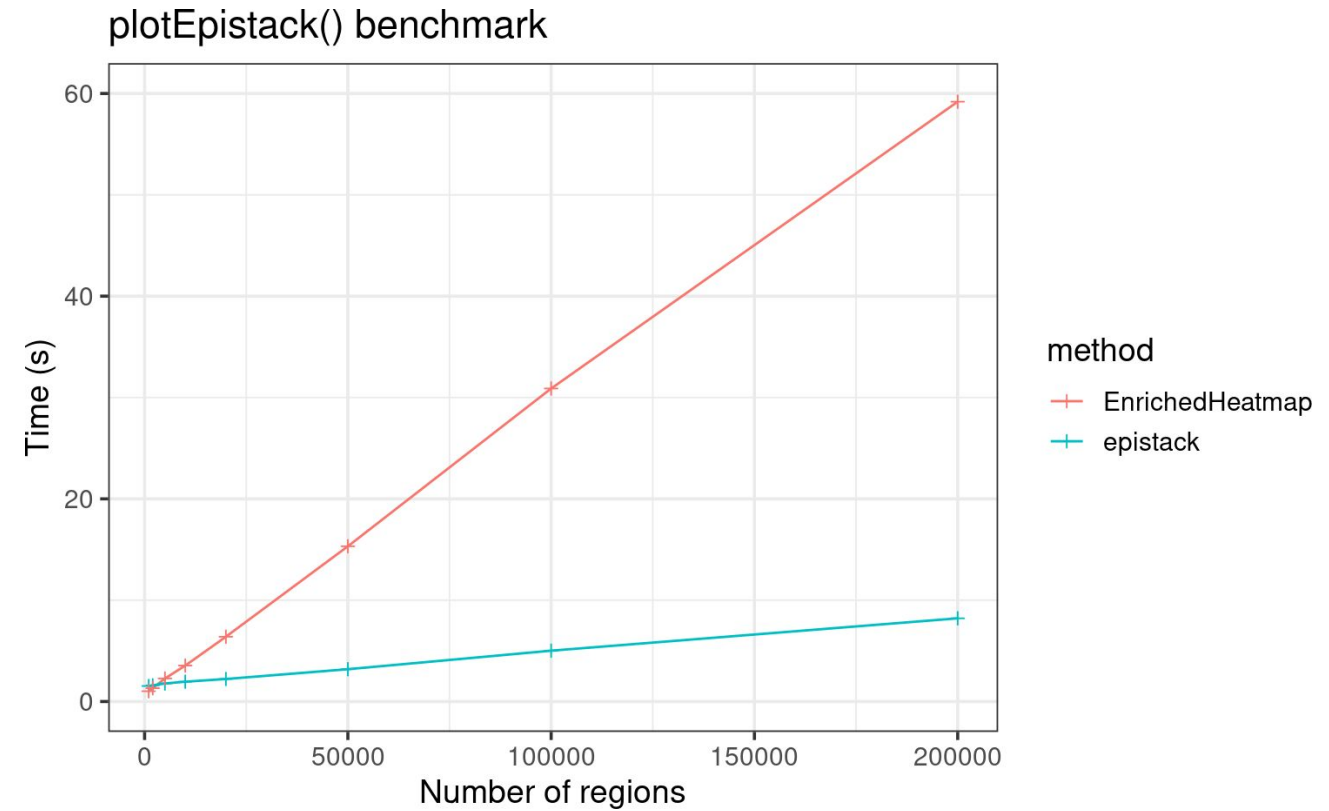
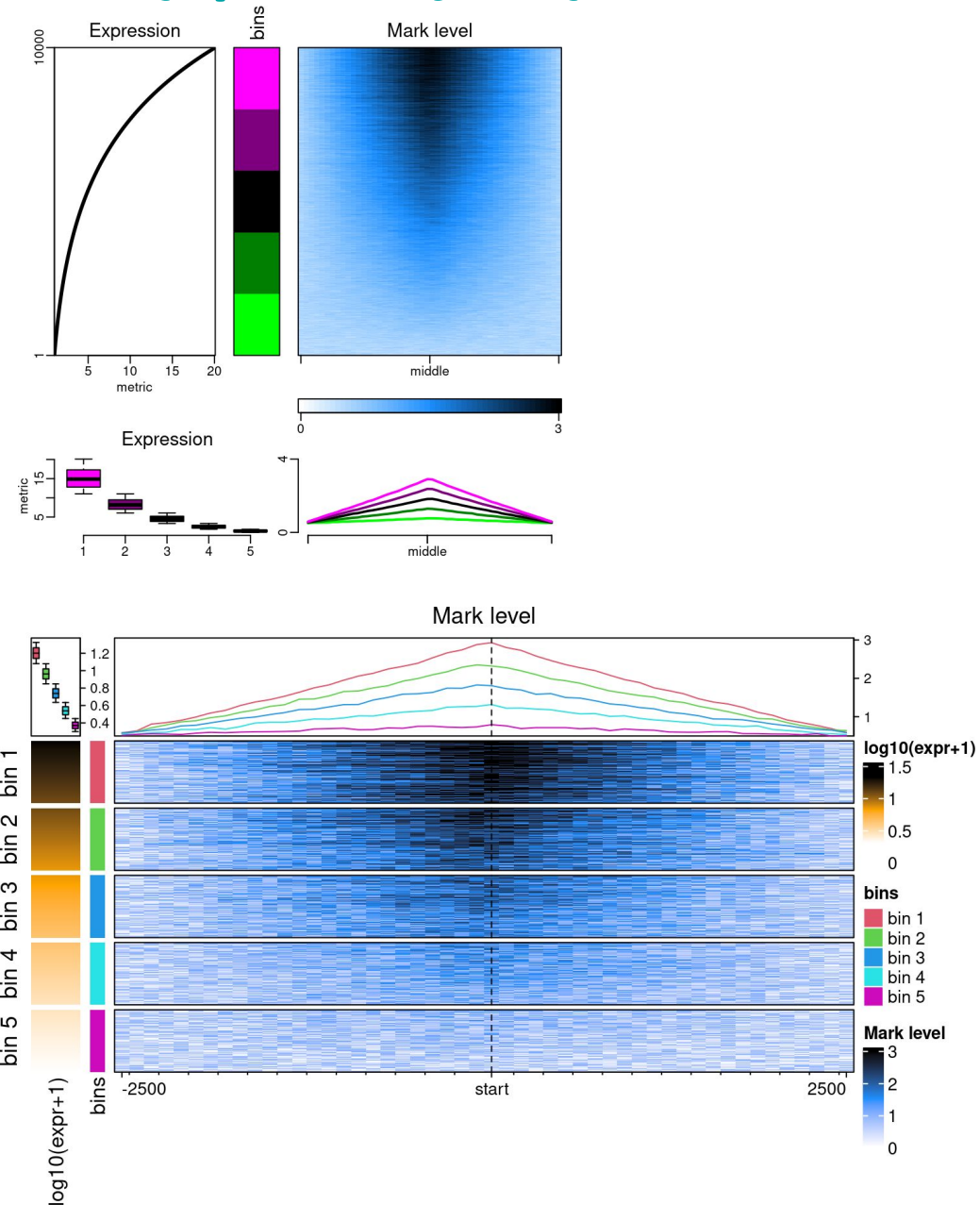
Original matrix



Reduced matrix

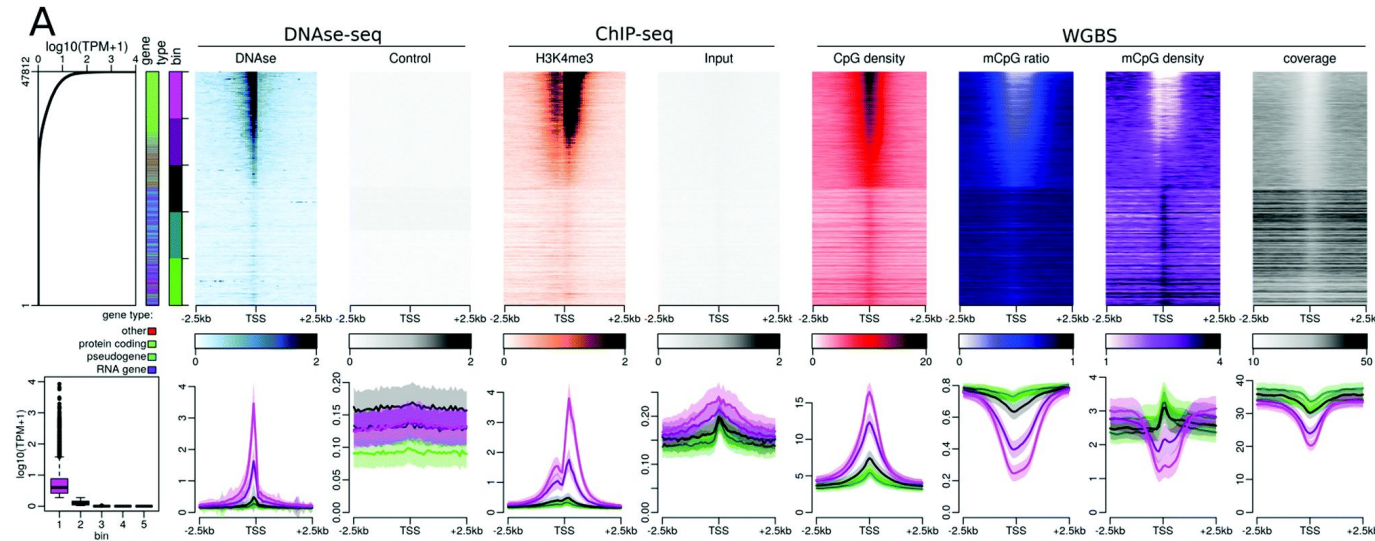


{epistack} vs {EnrichedHeatmap} benchmark



{epistack} in the real world

- inspired by our work on ROADMAP data
doi.org/10.1039/d0mo00130a
ioshiapps.cbu.uib.no/perepigenomics_app/
- Used internally to visualise DMR in various team projects
- To be applied on ALL FAANG ChIP-seq / ATAC-seq / WGBS data in the ANR VizFaDa



INRAE

{epistack}

2021-10-12 / Genotoul biostat bioinfo dat / Guillaume Devailly

{epistack}

Available on github:

github.com/GenEpi-GenPhySE/epistack

```
remotes::install_github( "GenEpi-GenPhySE/epistack" )
```

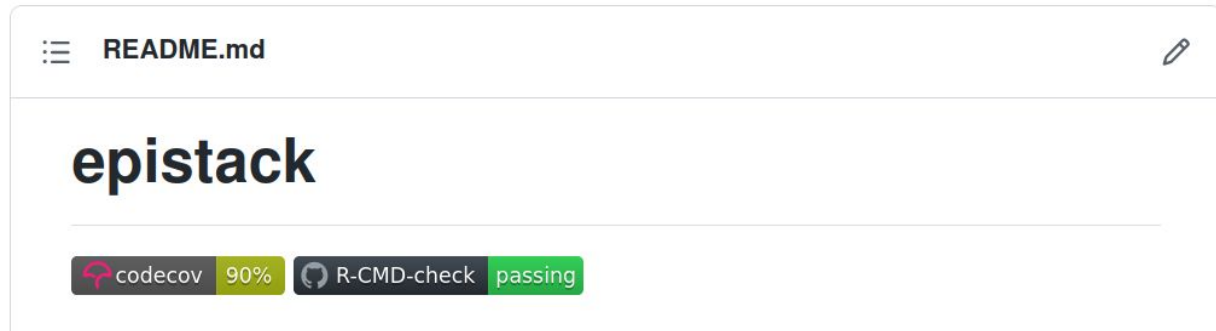
Submitted to Bioconductor:

github.com/Bioconductor/Contributions/issues/2362

Vignette:

gdevailly.github.io/using_epistack.html

Built using {devtools}, {testthat}, {roxygen2}, Github Actions



INRAE

{epistack}

2021-10-12 / Genotoul biostat bioinfo dat / Guillaume Devailly