

MethylAid: Visual and Interactive quality control
of large Illumina 450k data sets

BioC Europe 2015

Maarten van Iterson

Leiden University Medical Center
Department of Molecular Epidemiology

January 9, 2015

Epigenome-wide association studies (EWAS)

- DNA methylation
- cytosine of CpG sites can be converted to 5-methylcytosine
- smoking, bmi and several autoimmune diseases
- sample sizes are hundreds to several thousands
- Illumina 450K HumanMethylation array

Illumina 450K HumanMethylation array

- genotyping of bisulfite-converted genomic DNA
- 480K CpG sites
- 99% of RefSeq genes, CpG island, shores and shelves
- bisulfite conversion, amplification, hybridization, extending, staining and scanning
- several control probes to monitor different aspects of the protocol and quality of the DNA

MethylAid: Visual and Interactive quality control of large Illumina 450k data sets

- *wateRmelon*, *minfi*, *methylumi*, *lumi*, *COHCAP*, *ChAMP*, *shinyMethyl*, ...
- detect bad quality samples/runs using predefined thresholds
- fast and efficient: using *BiocParallel* and an option for reading data in batches
- interactive graphics: using *shiny*

Using *MethylAid*

```
library(minfiData)
baseDir <- system.file("extdata", package = "minfiData")
targets <- read.450k.sheet(baseDir)
library(MethylAid)
sdata <- summarize(targets)
visualize(sdata) ##this will launch the web application
```

summarizing in parallel using *BiocParallel* and the `bathSize`-option

```
library(BiocParallel)
confFile <- system.file("scripts/config.R", package="MethylAid")
BPPARAM <- BatchJobsParam(workers = 10,
  progressBar = FALSE,
  confFile = confFile)
summarize(targets, batchSize = 50, BPPARAM = BPPARAM)
```

demo: <http://shiny.bioexp.nl/MethylAid>

Further information

- vignette shows how to use data from TCGA and from GEO
- and gives more details on the parallel summarization
- application note¹
- a larger demo (approx. 6000 samples) running at:
<http://shiny.bioexp.nl/BIOS>

¹van Iterson, M., Tobi, E., Slieker, R., den Hollander, W., Luijk, R., Slagboom, P., and Heijmans, B. (2014). Methylaid: Visual and interactive quality control of large illumina 450k data sets. *Bioinformatics*