metR
Cheat Sheet

Aleksandra Brodecka [aut, cre]
Przemysław Biecek [aut, ths]
Warsaw University of Technology

Introduction

metR package allows comprehensive data analysis from methylation studies. By metR we can create regions, get basic statistics about them and plot methylation rate.
In metR have been implemented 6 different methods to check methylation difference, which based on methylation rate and also on number of methylated and unmethylated citosines.
So you only need methylation data to obtain interesting results!
Example of methylation data - schizophrenia are included in metR.

Preprocess data into one table

```
data <- preprocess(sample.1, sample.2)
```

The **sample.1** and **sample.2** are two data.frames from methylation studies.

Create regions

```
data.tiles <- create_tiles_max_gap(data, gaps.length = 100)
data.tiles <- create_tiles_fixed_length(data, tiles.length = 1000, common = T)
```

The **data** is result of preprocess function.
The **gaps.length** indicates maximum difference position in region.
The **tiles.length** creates regions where *k-region* in chromosome are observations for which position is between *[k * tiles.length;(k + 1)* tiles.length -1]*
The **common** - if TRUE creates second group regions are created where k-region in chromosome are observations for which position is between *[k * tiles.length + tiles.length/2 ; (k + 1)* tiles.length -1 + tiles.length/2]*

Get basic statistics about created regions

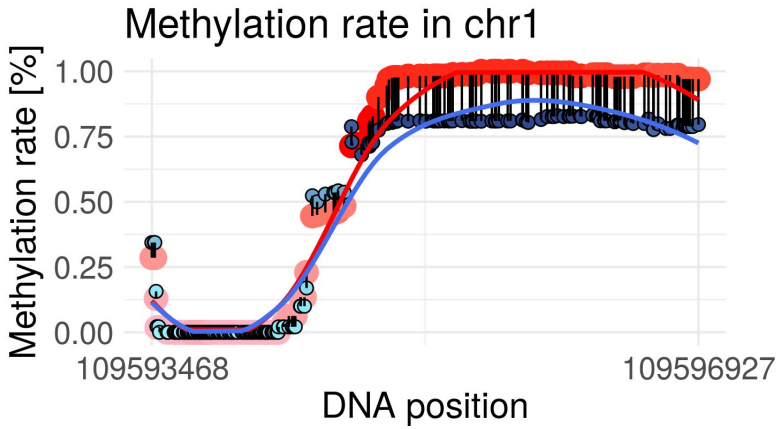
```
stats <- get_stats(data.tiles)
```

The **data.tiles** is result of function: create_tiles_max_gap or create_tiles_fixed_length.

Visualization

```
draw_methylation(data, chr = 'chr1', start = 109593468, end = 109596927, bind.probes = T, smooth.methylation = T)
```

The **data** is result of function preprocessing or create_tiles_max_gap or create_tiles_fixed_length.
The **chr** is chromosome name of region which will be plotted and **start** and **end** are the minimum and maximum position of region.
The **bind.probes** is logi value if observation will be bounded by vertical lines and **smooth.methylation** is logi value if methylation rate by samples will be smoothed.



Find differentially methylated regions

```
result <- find_DMR(data.tiles, methods = c('Wilcoxon', 'Reg.Log', 'Reg.Mixed', 'Reg.Corr.Mixed', p.value.log.reg = 0.001, p.value.reg.mixed = 0.001, p.value.reg.corr.mixed = 0.001)
```

The **data.tiles** is result of function: create_tiles_max_gap or create_tiles_fixed_length.
The **methods** is vector of methods that are used to sorting regions. Possible values: **'Wilcoxon'** - Wilcoxon signed test; **'Ttest'** - t-Student test with unequal variance; **'KS'** - Kolmogorov-Smirnov test; **'Reg.Log'** - Wald test of grouping variable from logistic regression; **'Reg.Mixed'** - Wald test of grouping variable from logistic regression with mixed effects; **'Reg.Corr.Mixed'** - Wald test of grouping variable from logistic regression with mixed effect and estimated previous correlation matrix.
The **p.value.log.reg, p.value.reg.mixed, p.value.reg.corr.mixed** if not NULL regions with p.value of prob variable smaller than p.value.log.reg (p.value.reg.mixed, p.value.reg.corr.mixed) will be returned and decreasingly ordered by absolute value of beta coefficient of prob variable otherwise regions will be increasingly ordered by p.value. This values are used only for regression methods.