

Practical: basic stats for peak calling

Jacques van Helden

2015-10-01

Peak-calling: Carl's question

What we want to do



Defining the data directory

Open a connection to the lab web browser.

We will first define the URL from which the data can be downloaded, by concatenating the URL fo the course with the path to our dataset.

To concatenate paths, it is *recommended* to use the **R** command `file.path()`.

```
url.course <- "http://jvanheld.github.io/EBA15_stats_RStudio"
url.data <- file.path(url.course, "practicals", "02_peak-ca
```

Loading a data table

R enables to download data directly from the Web.

Load counts per window in chip sample.

```
## Define URL of the ChIP file
chip.bedg.file <- file.path(url.data, "FNR_200bp.bedg")

## Load the file content in an R data.frame
chip.bedg <- read.table(chip.bedg.file)

## Set column names
names(chip.bedg) <- c("chrom", "start", "end", "counts")
```

Exploring a data frame: dim()

Before anything else, let us inspect the size of the data frame, in order to check that it was properly loaded.

```
dim(chip.bedg)
```

```
## [1] 23199      4
```

Exploring a data frame: head()

the function head() displays the first rows of a table.

```
head(chip.bedg)
```

##		chrom	start	end	counts
## 1	gi 49175990 ref NC_000913.2		0	200	1594
## 2	gi 49175990 ref NC_000913.2		200	400	834
## 3	gi 49175990 ref NC_000913.2		400	600	222
## 4	gi 49175990 ref NC_000913.2		600	800	172
## 5	gi 49175990 ref NC_000913.2		800	1000	123
## 6	gi 49175990 ref NC_000913.2		1000	1200	116

Exercise: loading the input counts per window

We already loaded the count table for the FNR ChIP counts per window.

The background level will be estimated by loading counts per window in a genomic input sample. These counts are available in the same directory a file named `input_200bp.bedg`

Load the counts per window in the input sample (genome sequencing).

Exercise: loading the input counts per window

```
## Define URL of the input file
input.bedg.file <- file.path(url.data, "input_200bp.bedg")

## Load the file content in an R data.frame
input.bedg <- read.table(input.bedg.file)

## Set column names
names(chip.bedg) <- c("chrom", "start", "end", "counts")
```