

genomation - a toolkit for annotation and visualization of genomic data

Altuna Akalin
altuna.akalin@fmi.ch

Vedran Franke
vedran.franke@gmail.com

October 30, 2013

Contents

1	Introduction	1
2	Access the data	1
3	Data input	2
4	Use cases for genomation package	2
4.1	Annotation of HTS data by functional regions	2
4.2	Visualization of CHiP sequencing data	2
4.3	Annotation of bam files	2

1 Introduction

Recent advances in sequencing technologies have enabled a downpour of biological data. The sheer amount of data has impeded the extraction of useful knowledge and novel hypothesis generation. (genomation) is a toolkit for annotation and in bulk visualization of genomic features over predefined regions. The genomic features the package can handle can be anything with a minimal information of chromosome,start and end. The features could have any length and can be associated with a score. Typical examples of such data sets include aligned from reads high-throughput sequencing (HTS) experiments, percent methylation values for CpGs (or other cytosines), locations of transcription factor binding site motifs, and so on.

The pipeline for computational knowledge extraction consists of three steps: data filtering, integration of data from multiple sources or generation of predictive models and biological interpretation of produced models, which leads to novel hypotheses that can be tested in the wetlab. (genomation) aims to facilitate the integration of multiple sources of high throughput data with genomic annotation or already published experimental results.

2 Access the data

High throughput data which will be used to show the functionality of the (genomation) is located in two places. The annotation and cap analysis of gene expression (CAGE) data comes

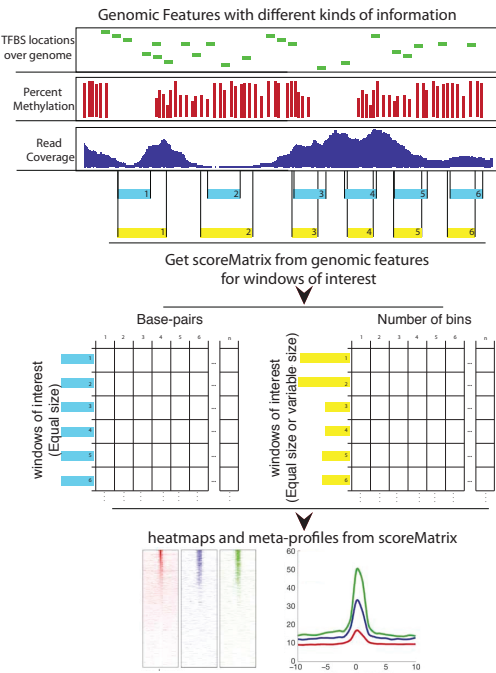


Figure 1: Bulk visualization for different genomic feature datasets flowchart

prepared with the genomation package, while the raw HTS data can be found in the sister package (genomationData).

To install the data package from bioconductor c/p the following lines into your R interpreter:

```
library(devtools)
install_github("genomationData")
```

The (genomationData) vignette contains a verbose description of contained files. To list the available data:

```
list.files(system.file(package = "genomationData"))
```

To see the descriptions of the files:

```
r <- read.table(system.file("SampleInfo.txt", package = "genomationData"),
  header = T, sep = "\t")
head(r)
```

3 Data input

4 Use cases for genomation package

4.1 Annotation of HTS data by functional regions

4.2 Visualization of CHiP sequencing data

4.3 Annotation of bam files