

GenomicTools - Vignette

Daniel Fischer

August 11, 2016

Introduction

The R-package GenomicTools is designed for the analysis of so-called omics data, and here especially on gene expression and SNP data. The focus is on performing an eQTL, a QTL or a Multifactor dimensionality reduction (MDR). Although MDR is not limited to genomic field and all other kinds of categorical data can be used with it, the implementation is here tailored for genomic data and currently the generalization to other variables proves to be difficult. The package comes with a couple of example dataset, further datasets can be downloaded from the project page, links are given for that below. The following chapters explain in detail, how the package can be applied in different scenarios and how the output is to be interpreted.

Installation of GenomicTools

The latest stable version of GenomicTools is located on Cran and can be installed via

```
install.packages("GenomicTools")
```

No special dependices are required and the dependencies should also be automatically installed by R.

The latest developed version, including the latest bugfixes is located in GitHub and can be installed like this:

```
library(devtools)
install_github("fischuu/GenomicTools")
```

The GitHub page is located here

<https://github.com/fischuu/GenomicTools>

and bugfixes and comments can easily be handed in via that platform. The package has also an own webpage where additional information may be posted and is located here:

<http://genomictools.danielfischer.name/>

Once the package is installed, it can be loaded into the workspace by typing

```
library("GenomicTools")
```

Included datasets and import functions

An overview of the example datasets. Currently many of them are simulated, but they will be moved to real datasets gradually.

Annotation files

Simulated Example File

An example annotation track is the `annotTrack` object. It can be loaded to the workspace via

```
data("annotTrack")
```

and the first rows and columns of it look like this

```
annotTrack[1:6,1:9]
```

```
[1] 1 2 3 4 5 6 7 8 9
```

A gtf file

GTF files are provided e.g. from Ensembl and can be downloaded from the corresponding webpage. For example the human annotation for Ensembl build 85 can be found here:

ftp://ftp.ensembl.org/pub/release-85/gtf/homo_sapiens/Homo_sapiens.GRCh38.85.gtf.gz

After downloading this file, it can be imported to R with

```
ensGTF <- importGTF(file="/home/ejo138/ownCloud/R-Packages-Pages/GenomicTools/Datasets/Homo_sapiens.GRC
```

General background for eQTL/QTL analyses

There are two methods implemented to perform an (e)QTL that be be picked with the `method=` option in the `eQTL/QTL` function. The two options are `LM` and `directional`. In case of `LM` a classical linear model is fitted to the data and it is tested if the slope is zero or not. This is the same method that is practically implemented in all (e)QTL software tools. The second option `directional`, however, uses a directional test based on probabilistic indices as it was presented in (Fischer et al. 2014). For the directional test, there are still another tuning option. The p-values can be either determined using a permutation type test, or using asymptotic results. The options to set this are either `testType="permutation"` or `testType="asymptotic"`. Currently the required asymptotic test is not implemented in the used R-package `gMWT`, but this will happen during August 2016 and is then also available in `GenomicTools`.

Perform an eQTL analysis

To run an eQTL first a couple of data objects has to be prepared. In the most simplest case there is only a single gene that should be tested against

References

Fischer, Daniel, Hannu Oja, Johanna Schleutker, Pranab K. Sen, and Tiina Wahlfors. 2014. "Generalized Mann–Whitney Type Tests for Microarray Experiments." *Scandinavian Journal of Statistics* 41 (3): 672–92.