RnBeadsDJ – A Quickstart Guide to the RnBeads Data Juggler

Fabian Müller, Yassen Assenov, Pavlo Lutsik

Contact: rnbeads@mpi-inf.mpg.de

Package version: 1.12.2

September 25, 2018

RnBeads is an R package for the comprehensive analysis of genome-wide DNA methylation data with single basepair resolution. RnBeadsDJ is a graphical user interface (GUI) guiding through RnBeads analyses.

Contents

1 Getting Started **1**

1.1 Installation . . . 1
1.2 Launching the Data Juggler . . . 1
1.3 Example Datasets . . . 2

2 Tutorial: Running RnBeads Analysis **2**

2.1 Setting Up the Analysis . . . 2
2.2 Selecting the Input Data . . . 3
2.3 Specifying Analysis Options . . . 4
2.4 Running the Analysis . . . 5
2.5 Inspecting the Analysis Results . . . 5

1 Getting Started

1.1 Installation

A detailed protocol on how to install RnBeads and its dependencies can be found on here: http://rnbeads.mpi-inf.mpg.de/publication/data/tutorial/00_prelim.html.

1.2 Launching the Data Juggler

To launch the Data Juggler (DJ), you just have to load the RnBeads package in an R session and then run the launch command:

```
suppressPackageStartupMessages(library(RnBeads))
rnb.run.dj()
```

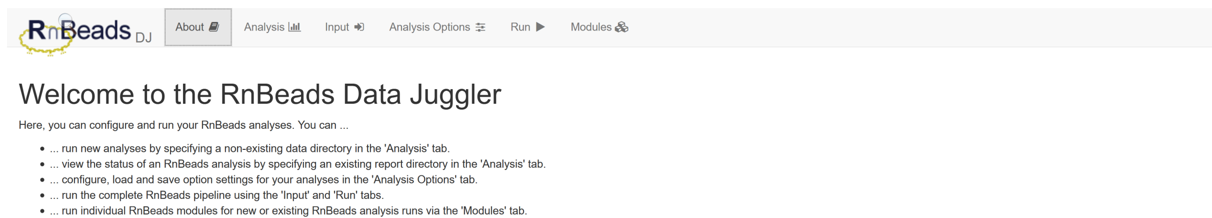

Figure 1: Starting page of RnBeadsDJ.

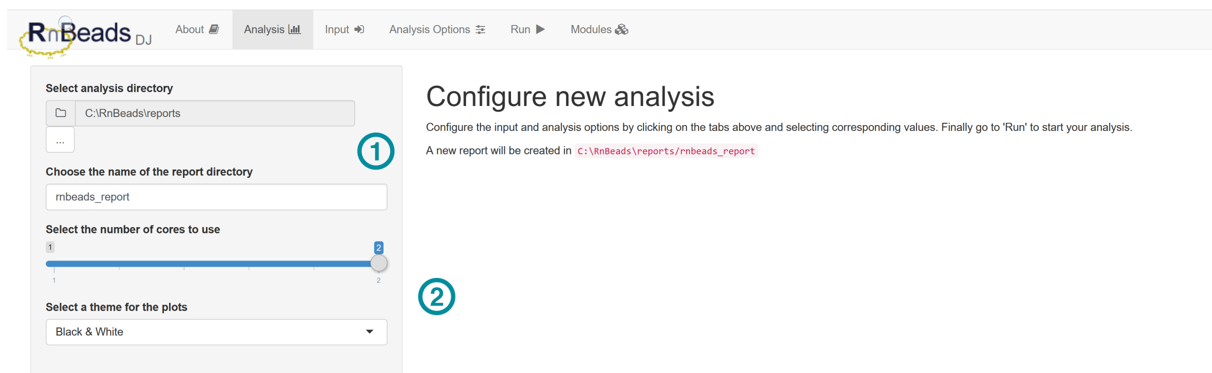

Figure 2: Analysis setup.

This will open your web browser and present you with the starting page of the DJ.

1.3 Example Datasets

In this tutorial, we will use a dataset of Illumina Infinium 450k beadarray methylation data from multiple human embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs) lines. You can download the dataset from the tutorial website[1].

2 Tutorial: Running RnBeads Analysis

If you have not done so already, please download and extract the example dataset from the RnBeads tutorial website. Then launch the DJ as described above:

```
suppressPackageStartupMessages(library(RnBeads))
rnb.run.dj()
```

You will be presented with the starting page of the DJ (Figure 1).

2.1 Setting Up the Analysis

Next, go to the Analysis tab to set up your analysis directory and general options (Figure 2).

[1] https://rnbeads.org/tutorial.html

Sample_ID	Sentrix_ID	Sentrix_Position	Sample_Plate	Sample_Well	Sample_Group	Cell_Line	Passage_No	Tr
hES_HUES13_p47	5815381013.00	R03C01	WG0001341-MSA4	C01	hESC	hES_HUES13	47	N
hiPS_20b_p43	5815381013.00	R05C02	WG0001341-MSA4	C02	hiPSC	hiPS_20b	43	N
hES_HUES1_p29	5815381013.00	R02C01	WG0001341-MSA4	B01	hESC	hES_HUES1	29	N
hiPS_11c_p23	5815381013.00	R04C02	WG0001341-MSA4	B02	hiPSC	hiPS_11c	23	N
hES_HUES1_p28	5815381013.00	R01C01	WG0001341-MSA4	A01	hESC	hES_HUES1	28	N
hiPS_20b_p49_TeSR	5815381013.00	R03C02	WG0001341-MSA4	A02	hiPSC	hiPS_20b	49	Te
hiPS_20b_p49_KOSR	5815381013.00	R02C02	WG0001341-MSA4	H01	hiPSC	hiPS_20b	49	K
hiPS_17b_p35_TeSR	5815381013.00	R01C02	WG0001341-MSA4	G01	hiPSC	hiPS_17b	35	Te
hiPS_27b_p31	5815381013.00	R06C02	WG0001341-MSA4	D02	hiPSC	hiPS_27b	31	N
hES_HUES64_p19	5815381013.00	R04C01	WG0001341-MSA4	D01	hESC	hES_HUES64	19	N
hiPS_17b_p35_KOSR	5815381013.00	R06C01	WG0001341-MSA4	F01	hiPSC	hiPS_17b	35	K
hES_H9_p58	5815381013.00	R05C01	WG0001341-MSA4	E01	hESC	hES_H9	58	N

Figure 3: Input data.

1. Click on the button labeled "..." to select an existing directory as the base directory for the analysis (Note that on some Windows computers the selection dialog box will be opened in the background and that you might need to minimize your browser to find it). Then specify an analysis name. RnBeads will create a new folder named according to this in the base directory, if it does not already exist.

2. Specify general analysis options such as the number of compute cores to be used for the analysis and the style preset for report plots.

2.2 Selecting the Input Data

In the `Input` tab you can configure your input data (Figure 3).

1. Select a sample annotation file. This should be a tabular file (comma or tab separated) that contains information on the samples in your dataset. You can select which separator is used in this file using the dropdown box below. For our example dataset, the sample annotation is contained in `sample_annotation.csv`, which is a comma-separated file. If the annotation can be loaded correctly, a preview of the sample information is shown on the right.

2. Select the appropriate platform that was used to quantify DNA methylation. In our example this is the Illumina Infinium 450k beadarray.

3. Select the directory where the methylation readout can be found (again, on some Windows machines, the corresponding selection box might appear in the background). For Illumina beadarrays this should be a directory containing `idat` intensity files for each sample. The example dataset you downloaded contains such a directory.

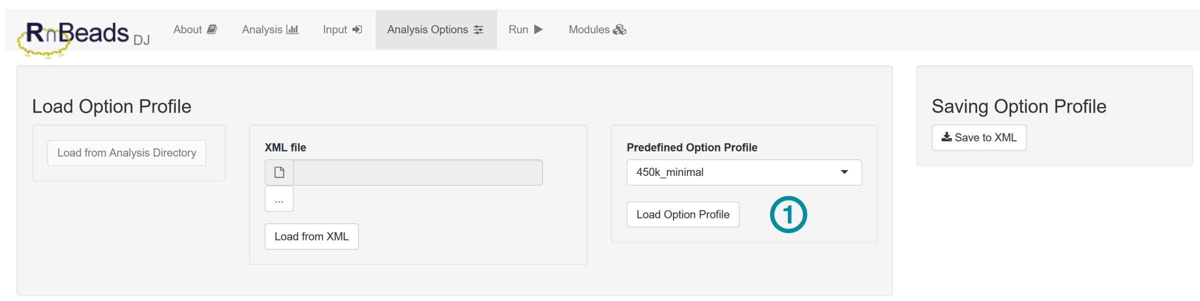

Figure 4: Analysis options – loading option profiles

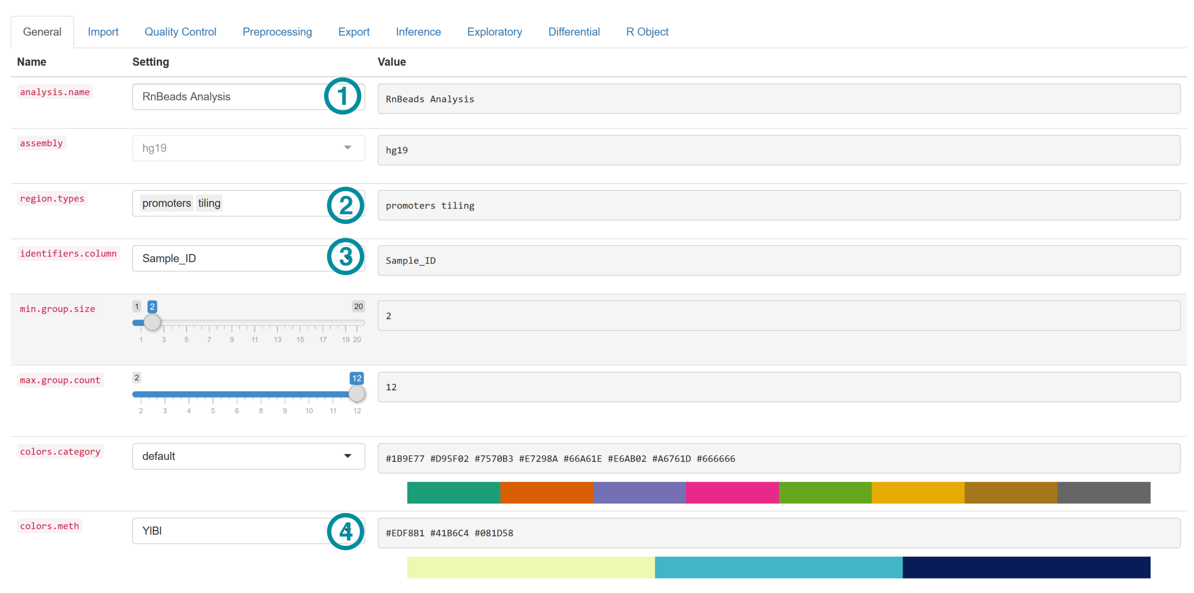

Figure 5: Analysis options – general settings

2.3 Specifying Analysis Options

The `Options` tab allows you to configure your analysis parameters. You can load preset option profiles and save them (Figure 4).

1. For our example analysis, select the "450k_minimal" profile, which will set some options to preset values such that the analysis will run in a reasonable amount of time.

The panel below contains several tabs which let you specify options pertaining to various steps in the `RnBeads` pipeline. Each tab lists the option names, option settings and current values in a tabular format. Note that you can hover over option names to see tooltips containing a more detailed description of individual options. General analysis options can be specified in the `General` tab (Figure 5). Here, let us ...

1. ... assign a reasonable name for the analysis.

2. ... select the region types for which methylation levels will be summarized in the analysis.

3. ... select a column of the sample annotation table that will be used as sample identifiers.

Figure 6: Analysis options – preprocessing

4. ... select one of the predefined color schemes for categorical and methylation values.

The `Preprocessing` tab contains option settings for filtering and normalizing the dataset (Figure 6).

1. For our example, enable the removal of methylation probes located on sex chromosomes.
2. Set the normalization method to "bmiq" and disable background normalization by setting the corresponding option value to "none".

For adjusting parameters for differential methylation analysis use the corresponding tab (Figure 7).

1. For our example, select "Sample_Group" and "Treatment" as the columns that define sample groups that will be compared in the analysis.

2.4 Running the Analysis

Now that the analysis directory, all the input data and options have been specified, click on the `Run` tab (Figure 8).

1. If everything was specified correctly, the run button will be enabled. Click it to start the analysis. Note: on our hardware with all the settings specified above, the analysis took about 1 1/2 hours on two compute cores.

2.5 Inspecting the Analysis Results

Upon successful completion an analysis report will be generated (Figure 9). It allows you to browse the generated analysis results interactively in your web browser. The tutorial website[2]

[2] https://rnbeads.org/tutorial.html

General | Import | Quality Control | Preprocessing | Export | Inference | Exploratory | Differential | R Object

Name	Setting	Value
differential	☑ Enable	TRUE
differential.comparison.columns	Sample_Group Treatment ①	Sample_Group Treatment
covariate.adjustment.columns		
differential.site.test.method	limma	limma
differential.report.sites	☑ Enable	TRUE
differential.variability	☐ Enable	FALSE
differential.variability.method	diffVar	diffVar
differential.enrichment.go	☐ Enable	FALSE
differential.enrichment.lola	☐ Enable	FALSE

Figure 7: Analysis options – differential methylation

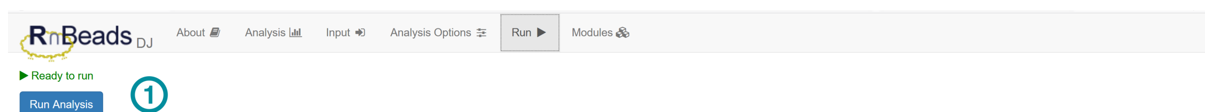

Figure 8: Running the Analysis

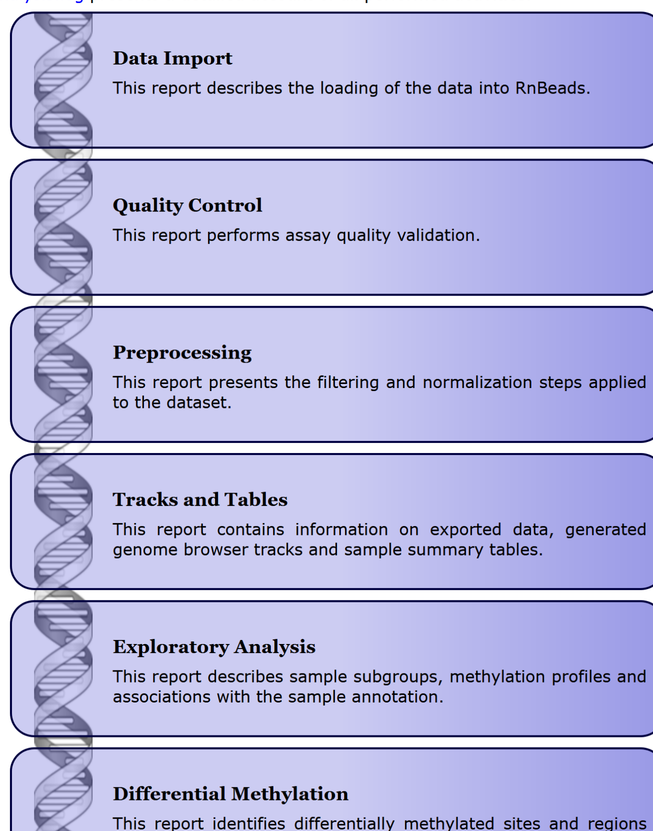

Figure 9: RnBeads analysis report

contains a link to the report that was generated during the writing of this tutorial.