



max planck institut
informatik

Analyzing Methylome Data Using

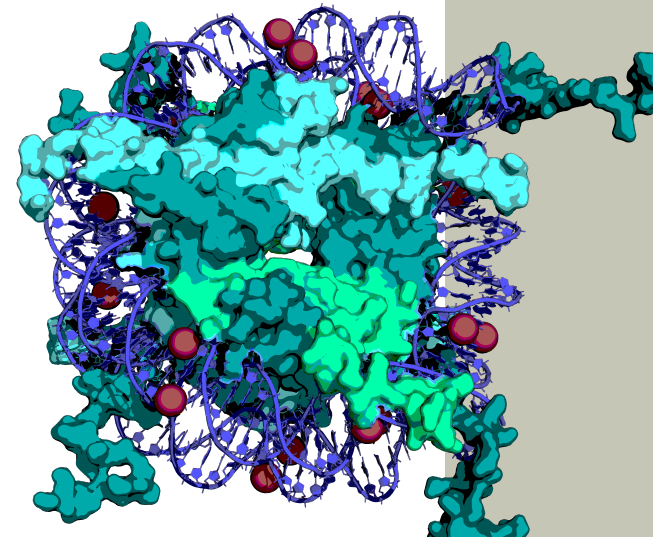
RnBeads

Fabian Müller

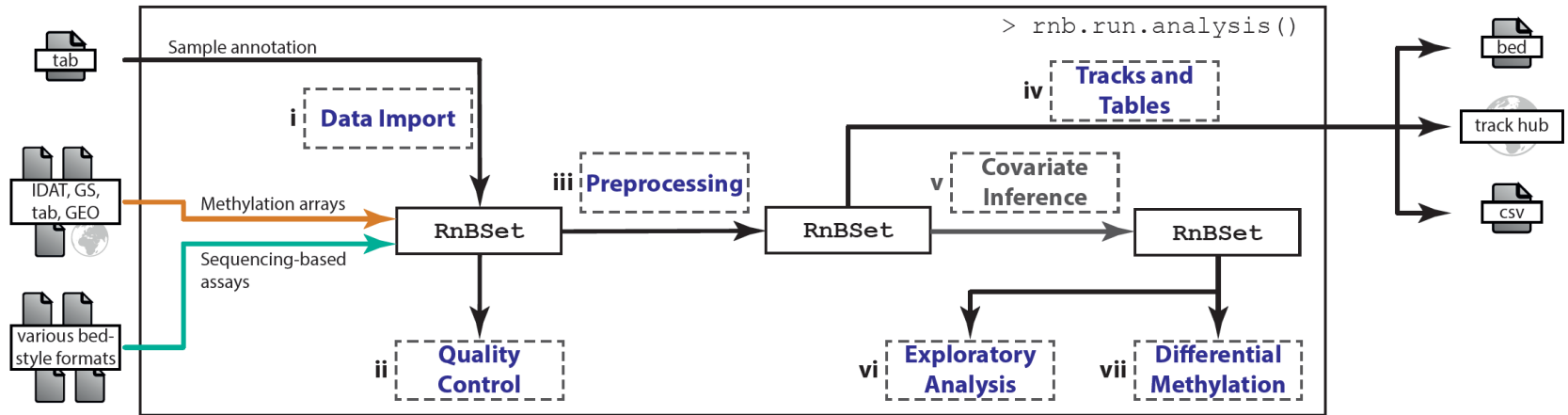
Epigenomics 2016

Puerto Rico

February 2, 2016

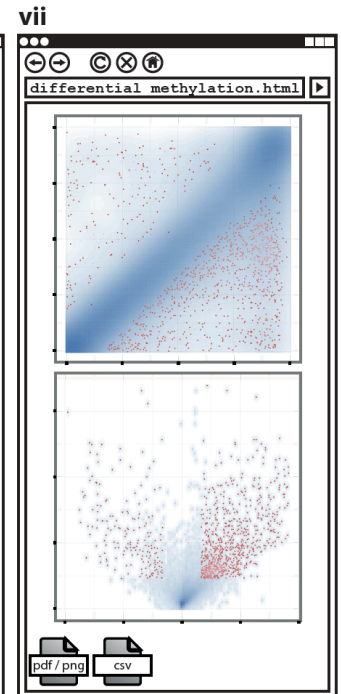
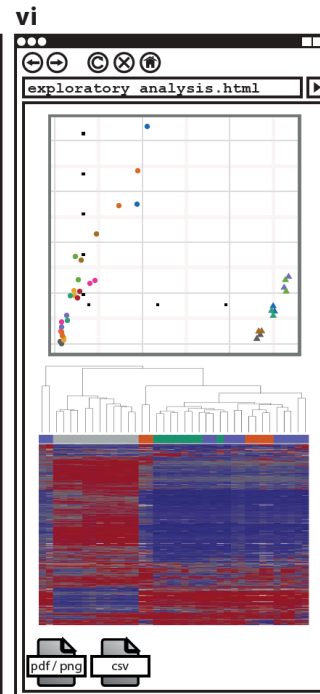
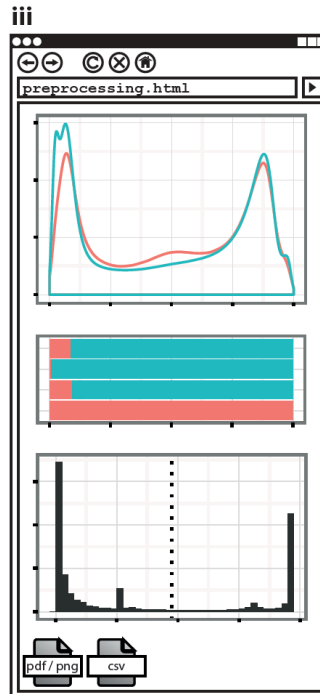
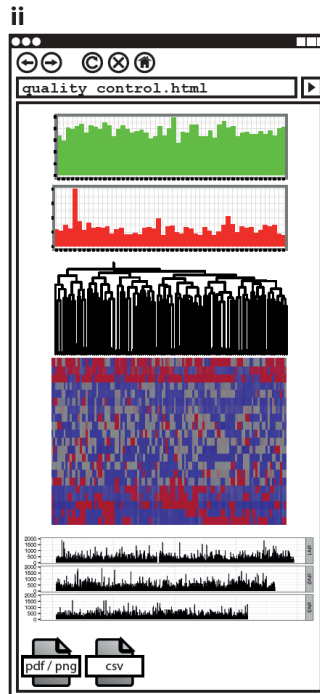
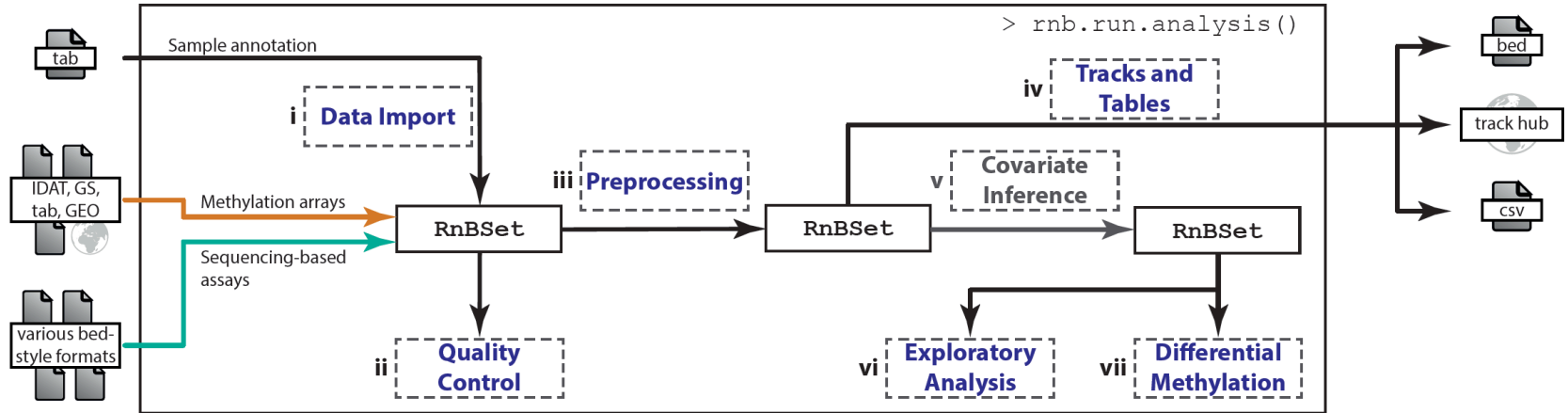


RnBeads – Workflow



Assenov, Müller, Lutsik, Walter, Lengauer, Bock (2014), *Nature Methods*, 11(11), 1138–1140

RnBeads – Workflow



Analyzing Methylome Data Using RnBeads

Hands-On

Tutorial website:
<http://rnbeads.mpi-inf.mpg.de/tutorial.php>



System Requirements

- Windows, Linux, MacOS, ...
- Working installation of R (version 3.0 or higher)
 - Or admin rights in order to install R and ghostscript
- ≥ 4 GB of main memory
- ≥ 4 GB of free disk space
- Internet connection



Getting Started

- Visit the tutorial website:
 - <http://rnbeads.mpi-inf.mpg.de>
 - → Tutorial
- Download the tutorial scripts and data
 - Download and unzip
- Install R
- Install Ghostscript and RnBeads
- Run your methylation analyses



The screenshot shows the RnBeads website interface. At the top, there is a navigation bar with the MPI logo and the text 'max planck institut informatik'. Below this, the 'RnBeads' logo is prominently displayed. A sidebar on the left contains a menu with categories such as 'Algorithms & Complexity', 'Computational Biology & Applied Algorithms', 'Software', and 'Offers'. The main content area is titled 'Tutorial' and contains the following text:

This site contains instructions and material for the workshop on "Analyzing Methylation Data using RnBeads" at the Epigenomics 2016 meeting. In order to not overstress the internet bandwidth at the meeting venue, we recommend that you download and install `R` and `RnBeads` as well as the example datasets beforehand if you would like to follow along with the hands-on exercises of the workshop. To do so, please follow steps 1., 2. and 3. described below. We are looking forward to meeting you at the workshop!

1. Download the Tutorial Scripts and Datasets

Download the following files for the tutorial:

- Download the tutorial source files
- Download the Ziller 450K dataset
- Download the Bock RRBS dataset

Extract the `epigenomics2016.zip` archive to a destination of your choice. Extract the downloaded dataset archives to the `data` subdirectory of your `epigenomics2016` directory.

2. Install `R`

Visit the [CRAN website](#) and download the version of `R` corresponding to your platform. Install `R` using the downloaded executable.

3. Installing `RnBeads`

For a basic installation of `RnBeads` simply open an `R` session and run the following command:

```
source("http://rnbeads.mpi-inf.mpg.de/install.R")
```

4. Testing Your `RnBeads` Installation and Further Preparations for the Tutorial

Follow the instructions provided in `src/00_prepare.R` from the tutorial script folder.

5. Analyze Your Methylation Data Using `RnBeads`

Follow the instructions provided in the presentation.

Example Reports

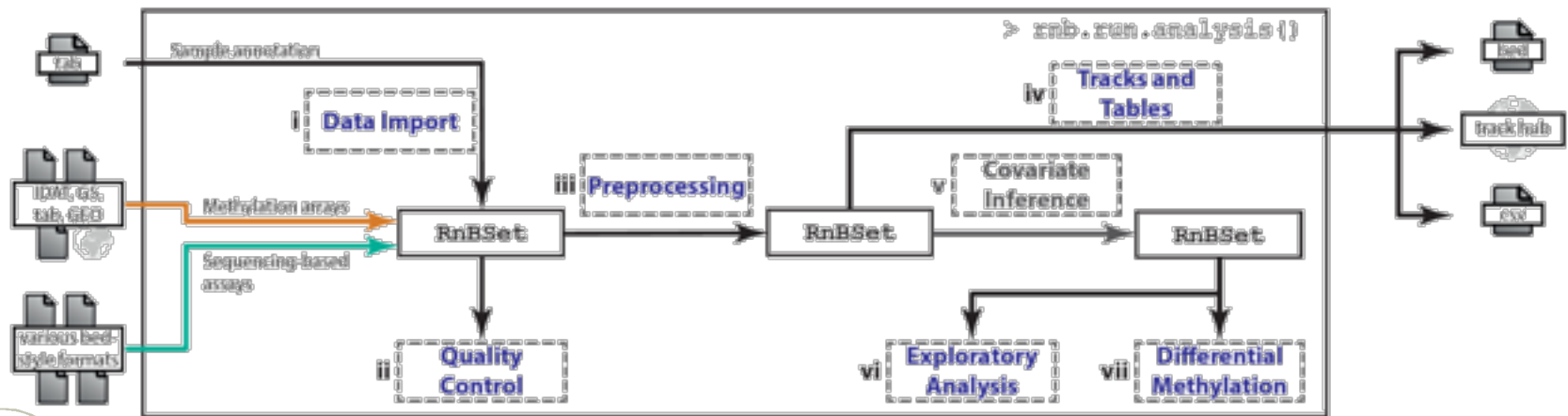
Reports for the examples discussed in the tutorial are available here:

- Ziller2011_450K_data_analysis
- Bock2012_RRBS_data_analysis



Vanilla Analysis

- Runs through the entire RnBeads workflow with specified analysis options
- Get started with `01_vanilla_Ziller450k.R`
 1. Start R
 2. Load RnBeads
 3. Specify analysis options
 4. Run analysis



Ziller, et al. (2011), *PLoS Genetics*, 7(12), e1002389



- Explore RnBeads modules, functions and classes
- See examples in `02_step_by_step_Ziller450k.Rmd`
 - With a few exceptions, examples for Illumina 450K array analysis generalize to bisulfite sequencing data

DNA methylation analysis with RnBeads

Epigenomics 2016 Workshop, February 1, 2016

Step-by-step analysis of the Ziller2011 450K dataset

This document contains examples and recipes for analyzing Illumina HumanMethylation 450 array data. The example data used here are taken from Ziller, et al. (2011), PLoS Genetics, 7(12), e1002389.

Preliminaries

Loading the RnBeads package

```
library(RnBeads)
```

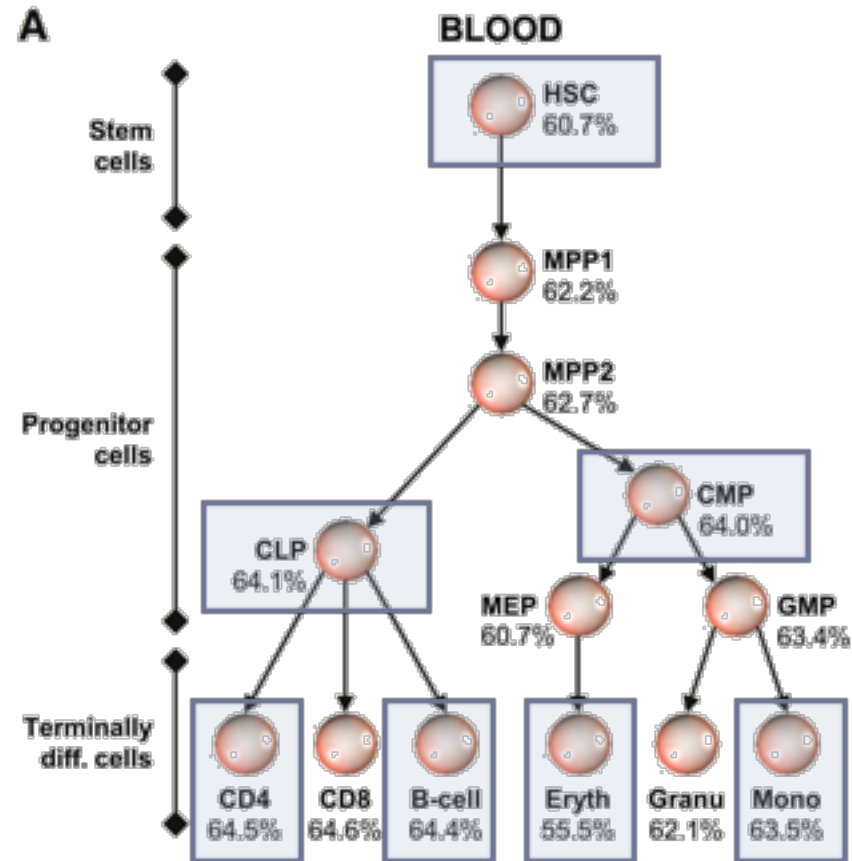
Make sure you are in the `epigenomics2016` directory and set up variables that define where the data is located and where results should be written to.

```
# setwd(".") # set the working directory to 'epigenomics2016'  
dataDir <- file.path(getwd(), "data")  
resultDir <- file.path(getwd(), "results")  
  
datasetDir <- file.path(dataDir, "Ziller2011_PLoSGen_450K")  
idatDir <- file.path(datasetDir, "dataset", "idat")  
sampleSheet <- file.path(datasetDir, "dataset", "sample_annotation.csv")  
reportDir <- file.path(resultDir, "report_Ziller2011_stepByStep")
```



Analyzing Bisulfite-Sequencing Data

- Explore RnBeads modules, functions and classes
- Vanilla analysis
 - `03_vanilla_BockRRBS.R`
- Custom analysis
 - See examples in `04_step_by_step_BockRRBS.Rmd`



Bock, et al. (2012), *Mol. Cell*, 47(4), 633-647



Performance Tuning

- Parallel computation
 - foreach and doParallel packages
- Disk-based storage of large data matrices
 - ff package
- Deploying RnBeads on a scientific computing cluster

```
#####  
# (1) Parallel computing  
#####  
parallel.isEnabled()  
num.cores <- 2  
# enable parallel computation for critical steps in the analysis  
# in the back, RnBeads uses the "foreach" and "doParallel" packages  
parallel.setup(num.cores)  
  
parallel.isEnabled()  
parallel.getNumWorkers()  
  
# disable parallel computation again  
parallel.teardown()  
parallel.isEnabled()
```



Beyond DNA Methylation

- Logging
- HTML reports
- See examples in
`05_beyond_methylation.R`

RnBeads

An Eye Opener

▼ Adding a text section

Here is some text for our awesome report

▼ A subsection

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas vestibulum placerat lobortis. Ut viverra fringilla urna at rutrum. In hac habitasse platea dictumst. Vestibulum adipiscing rutrum libero et interdum. Etiam sed odio ac nibh ullamcorper congue. Proin ac ipsum elit. Ut porta lorem sed lorem molestie pharetra. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Cras ac augue eu turpis dignissim aliquam. Vivamus at arcu ligula, vel scelerisque nisi. Vivamus ac lorem libero, quis venenatis metus. Fusce et lectus at lectus vestibulum faucibus ac id sem.

Nunc congue molestie fringilla. Aliquam erat volutpat. Integer consequat turpis nec dolor pulvinar et vulputate magna adipiscing. Curabitur purus dolor, porttitor vel dapibus quis, eleifend at lacus. Cras at mauris est, quis aliquam libero. Nulla facilisi. Nam facilisis placerat aliquam. Morbi in odio non ligula mollis rhoncus et et erat. Maecenas ut dui nisl. Mauris consequat cursus augue quis euismod.

TODO: Add content here

To be or not to be, that is the question

▼ Lists and Tables

1. AAA
2. BBB
3. CCC
4. DDD
5. EEE
6. FFF
7. GGG
8. HHH
9. III
10. JJJ

| | Col 1 | Col 2 | Col 3 | Col 4 |
|-------|-------|-------|-------|-------|
| Row 1 | AAA | GGG | MMM | SSS |
| Row 2 | BBB | HHH | NNN | TTT |
| Row 3 | CCC | III | OOO | UUU |
| Row 4 | DDD | JJJ | PPP | VVV |
| Row 5 | EEE | KKK | QQQ | WWW |
| Row 6 | FFF | LLL | RRR | XXX |



Summary

- RnBeads ...
 - ... is an R package for start-to-finish DNA methylation analysis
 - ... enables large-scale analyses for virtually any methylation platform providing basepair resolution (methylation arrays and bisulfite sequencing)
 - ... supports a wide range of input and export formats
 - ... performs analysis on individual CpGs as well as predefined or custom genomic regions
 - ... generates informative and comprehensive HTML reports containing method descriptions, publication grade plots and data tables
 - ... does not require expert knowledge to run
 - ... enables configurable for custom analysis due to its modular design and programming interface



Acknowledgements



<http://rnbeads.mpi-inf.mpg.de>

Yassen Assenov (MPII, DKFZ)
Pavlo Lutsik (Saarland Univ.)
Fabian Müller



Pavlo Lutsik
Karl Nordstrøm
Jörn Walter



Felipe Albrecht
Peter Ebert
Markus List
Michael Scherer
Georg Friedrich
Joachim Büch
Christoph Bock (MPII, CeMM)
Thomas Lengauer

