The minfi User's Guide

Kasper D. Hansen, Jean-Phillipe Fortin

27 July 2016

Abstract

A comprehensive guide to using the minfi package for analyzing DNA methylation microarrays from Illumina.vignettes.

Package version: minfi 1.18.6

Contents

Introduction

The *minfi* package provides tools for analyzing Illumina's Methylation arrays, specifically the 450k and EPIC (also known as the 850k) arrays. We have partial support for the older 27k array.

The tasks addressed in this package include preprocessing, QC assessments, identification of interesting methylation loci and plotting functionality. Analyzing these types of arrays is ongoing research in ours and other groups.

The input data to this package are IDAT files, representing two different color channels prior to normalization. This is the most complete data type, because IDAT files includes measurements on control probes. It is possible to use Genome Studio files together with the data structures contained in this package, but only some functionality is available because Genome Studio output does not contain control probe information. In addition, usually Genome Studio output is normalized using the methods implemented in Genome Studio and these are generally considered inferior.

Citing the minfi package

The MINFI package contains methods which are described across multiple manuscripts, by different non-overlapping authors. This makes citing the package a bit difficult, so here is a guide.

- If you are using MINFI in a publication, please cite (Aryee et al. 2014). This publication includes details on sex estimation using `getSex()`, quality control using `getQC()`, quantile normalization using `preprocessQuantile()`, bump hunting using `bumphunter()` and block finding using `blockFinder()`.
- If you are using `preprocessQuantile()`, it would be considerate to also cite (Touleimat and Tost 2012), since this publication describes a method essentially identical to `preprocessQuantile()`.
- If you are using `bumphunter()`, you should also cite the original bump hunter publication (Jaffe et al. 2012).
- If you are using SWAN normalization as implemented in `preprocessSWAN()` please cite (**???**).
- If you are using noob background correction as implemented in `preprocessNoob()`, please cite (Triche et al. 2013).
- If you are using functional normalization as implemented in `preprocessFunnorm()`, please cite (Jean-Philippe Fortin et al. 2014). The default in `preprocessFunnorm()` is to do noob background correction. If this is used, please also cite (Triche et al. 2013).
- If you are estimating A/B compartments as implemented in `compartments()` and `extractAB()`, please cite (Jean-Philippe Fortin and Kasper D Hansen 2015).

minfi classes

Reading data

Manifest / annotation

Quality control

- use shinyMethyl
- minfiQC
- getSex checks
- mds plots
- plot conversion probes

Preprocessing

- preprocessRaw
- preprocessIllumina
- preprocessSWAN
- preprocessQuantile
- preprocessNoob
- preprocessFunnorm

SNPs and other issues

- SNPs
- cross reactive probes

- remove bad probes
- detection P
- Gap Hunting

Identifying differentially methylated regions

- DMP finding
- Bump hunting
- Block finding
- plot DMRs with annotation.

Correcting for cell type heterogenity

- estimateCellTypes
- reference packages

Other stuff

- Horvath age estimation
- other packages

Sessioninfo

- R version 3.3.1 (2016-06-21), `x86_64-apple-darwin13.4.0`
- Locale: `C/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8`
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: BiocStyle 2.0.2
- Loaded via a namespace (and not attached): Rcpp 0.12.6, digest 0.6.9, evaluate 0.9, formatR 1.4, htmltools 0.3.5, knitr 1.13, magrittr 1.5, rmarkdown 1.0, stringi 1.1.1, stringr 1.0.0, tools 3.3.1, yaml 2.1.13

References

Aryee, Martin J, Andrew E Jaffe, Hector Corrada Bravo, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and Rafael A Irizarry. 2014. "Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays." *Bioinformatics* 30 (10): 1363–9. doi:10.1093/bioinformatics/btu049.

Jaffe, Andrew E, Peter Murakami, Hwajin Lee, Jeffrey T Leek, M Daniele Fallin, Andrew P Feinberg, and Rafael A Irizarry. 2012. "Bump Hunting to Identify Differentially Methylated Regions in Epigenetic Epidemiology Studies." *International Journal of Epidemiology* 41 (1): 200–209. doi:10.1093/ije/dyr238.

Jean-Philippe Fortin, and Kasper D Hansen. 2015. "Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data." *Genome Biology* 16: 180. doi:10.1186/s13059-015-0741-y.

Jean-Philippe Fortin, Aurélie Labbe, Mathieu Lemire, Brent W Zanke, Thomas J Hudson, Elana J Fertig, Celia MT Greenwood, and Kasper D Hansen. 2014. "Functional normalization of 450k methylation array data improves replication

in large cancer studies." *Genome Biology* 15 (11): 503. doi:10.1186/s13059-014-0503-2.

Touleimat, Nizar, and Jörg Tost. 2012. "Complete Pipeline for Infinium() Human Methylation 450K BeadChip Data Processing Using Subset Quantile Normalization for Accurate DNA Methylation Estimation." *Epigenomics* 4 (3): 325–41. doi:10.2217/epi.12.21.

Triche, Timothy J, Daniel J Weisenberger, David Van Den Berg, Peter W Laird, and Kimberly D Siegmund. 2013. "Low-level processing of Illumina Infinium DNA Methylation BeadArrays." *Nucleic Acids Research* 41 (7): e90. doi:10.1093/nar/gkt090.