tcgaR:

Jean-Philippe Fortin, Kasper Daniel Hansen

April 6, 2015

1 Introduction

Up to now, more than 10,000 methylation samples from the state-of-the-art 450K microarray have been made available through The Cancer Genome Atlas portal [1] and the Gene Expression Omnibus (GEO) [2]. Large-scale comparison studies, for instance between cancers or tissues, become possible epigenome-widely. These large studies often require a substantial amount of time spent on preprocessing the data and performing quality control. For such studies, it is not rare to encounter significant batch effects, and those can have a dramatic impact on the validity of the biological results [3, 4]. With that in mind, we developed *shinyMethyl* to make the preprocessing of large 450K datasets intuitive, enjoyable and reproducible. *shinyMethyl* is an interactive visualization tool for Illumina 450K methylation array data based on the packages *minfi* and *shiny* [5, 6].

comment: Some comment

2 Data

Data: April 6 2015

DNA Methylation 27k data

Get methylation mappings

The function `getMethMappings` allows to get all the mappings.

Clinical data

The function `getClinicalData` allows to get all the available data for each cancer type.

Abb.	Full name	27k	450k	Reference
ACC	Adrenocortical carcinoma		✓	
BLCA	Bladder Urothelial Carcinoma		✓	
BRCA	Breast invasive carcinoma	✓	✓	
CESC	Cervical squamous cell carcinoma		✓	
CHOL	Cholangiocarcinoma		✓	
COAD	Colon adenocarcinoma	✓	✓	
DLBC	Large B-cell Lymphoma		✓	
ESCA	Esophageal carcinoma		✓	
GBM	Glioblastoma multiforme	✓	✓	
HNSC	Head and Neck squamous cell carcinoma		✓	
KICH	Kidney Chromophobe		✓	
KIRP	Kidney renal papillary cell carcinoma	✓	✓	
KIRC	Kidney renal clear cell carcinoma	✓	✓	
LAML	Acute Myeloid Leukemia	✓	✓	
LCML	Chronic Myelogenous Leukemia			
LGG	Lower grade glioma		✓	
LIHC	Liver hepatocellular carcinoma		✓	
LUAD	Lung adenocarcinoma	✓	✓	
LUSC	Lung squamous cell carcinoma	✓	✓	
MESO	Mesothelioma		✓	
OV	Ovarian serous cystadenocarcinoma	✓	✓	
PAAD	Pancreatic adenocarcinoma		✓	
PCPG	Pheochromocytoma and Paraganglioma		✓	
PRAD	Prostate adenocarcinoma		✓	
READ	Rectum adenocarcinoma	✓	✓	
SARC	Sarcoma		✓	
SKCM	Skin Cutaneous Melanoma		✓	
STAD	Stomach adenocarcinoma	✓	✓	
TGCT	Testicular Germ Cell Tumors		✓	
THCA	Thyroid carcinoma		✓	
THYM	Thymoma		✓	
UCEC	Uterine Corpus Endometrial Carcinoma	✓	✓	
UCS	Uterine Carcinosarcoma		✓	
UVM	Uveal Melanoma		✓	

Table 1: Hi-C data sources.

comment: some comment

Session info

Here is the output of `sessionInfo` on the system on which this document was compiled:

Abb.	Full name	N Normal	N Tumor	Reference
BRCA	Breast invasive carcinoma	27	318	
COAD	Colon adenocarcinoma	37	166	
GBM	Glioblastoma multiforme	0	296	
KIRC	Kidney renal clear cell carcinoma	199	219	
KIRP	Kidney renal papillary cell carcinoma	5	16	
LAML	Acute Myeloid Leukemia	0	194	
LUAD	Lung adenocarcinoma	24	127	
LUSC	Lung squamous cell carcinoma	27	134	
OV	Ovarian serous cystadenocarcinoma	14	603	
READ	Rectum adenocarcinoma	5	68	
STAD	Stomach adenocarcinoma	59	82	
UCEC	Uterine Corpus Endometrial Carcinoma	1	117	

Table 2: Hi-C data sources.

References

[1] The Cancer Genome Atlas. *Data portal*, 2014. Online. URL: https://tcga-data.nci.nih.gov/tcga/.

[2] R. Edgar, M. Domrachev, and A. E. Lash. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, 30(1):207–210, Jan 2002.

[3] Jeffrey T Leek, Robert B Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739, 2010. doi:10.1038/nrg2825.

[4] Kristin N Harper, Brandilyn A Peters, and Mary V Gamble. Batch effects and pathway analysis: two potential perils in cancer studies involving dna methylation array analysis. *Cancer Epidemiol Biomarkers Prev*, 22(6):1052–60, 2013. doi:10.1158/1055-9965.EPI-13-0114.

[5] Martin J Aryee, Andrew E Jaffe, Hector Corrada Bravo, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and Rafael A Irizarry. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30(10):1363–1369, 2014. doi:10.1093/bioinformatics/btu049, PMID:24478339.

[6] RStudio and Inc. *shiny: Web Application Framework for R*, 2014. R package version 0.9.1. URL: http://CRAN.R-project.org/package=shiny.