# VisualMethyl: a shiny web application for interactive DNA methylation analysis and visualization

Irene Fernández Rebollo

Scientific director: Manuel Castro de Moura, mcastro@carrerasresearch.org

Josep Carreras Leukaemia Research Institute, Ctra de Can Ruti, Camí de les Escoles, s/n, 08916 Badalona, Barcelona

## Abstract

The study of DNA methylation as an epigenetic biomarker for detection, diagnosis and prediction of response to treatment is one of the key fields in research. The DNA methylation data analysis process requires the use of tools and pipelines that involve/implicate programming skills and also is highly time-consuming for non-expert bioinformatics scientists; for this reason, it is necessary to create applications that provide results in a simple, interactive way that allows the user to visualize and understand the data clearly. There are already some tools that perform DNA methylation analyses but is needed one that encompasses all of them in a comprehensive workflow, integrating DNA methylation with clinical patient data.

VisualMethyl is a *shiny* web application that provides the analysis and interactive visualizations of DNA methylation arrays (450k and EPIC). This tool performs quality control, exploratory analysis, differentially methylated CpG sites and regions, functional enrichment, and survival analysis; using R packages like *minfi* and *limma* to inspect the data provided and with options in each step to choose the methods and parameters preferred. The results obtained are presented in an interactive way through the use of dynamic graphs and tables in the shiny application which can be downloaded. VisualMethyl is implemented in the R programming language and requires the R package *shiny*, the user interface is customized with Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript. To demonstrate its utility, a dataset of DNA methylation from The Cancer Genome Atlas (TCGA) is analysed.

Supplementary information: https://github.com/ireneefr/VisualMethyl
https://drive.google.com/file/d/1tnlTBFGyLSGycCpHaPo2dOwgRySx9-kk/view?usp=sharing

## Introduction

Epigenetics is the study of heritable changes in gene expression without alterations in the DNA sequence. Epigenetic mechanisms include DNA methylation, histone modifications, nucleosome positioning, and microRNAs (miRNAs). DNA methylation is the best-known epigenetic process involving the addition of a methyl group to the C-5 position of the cytosine pyrimidine ring of DNA. In humans, this modification takes place mainly in cytosines (C) followed by a guanine (G), CpG, and tends to cluster in regions known as CpG islands, usually located at the 5' end of genes (promoter regions). These epigenetic changes regulate gene expression and are one of the key fields in research. Anomalous DNA methylation patterns have been associated with several diseases and carcinogenesis. These modifications can be used as epigenetic biomarkers for detection, diagnosis, and prediction of response to treatment. [1-2]

There exist different technologies to measure DNA methylation but the Illumina Infinium HumanMethylation BeadChips, based on sodium bisulphite conversion of DNA, have been the most popular to analyse genome-scale DNA methylation in human samples (www.illumina.com) since they have been widely used in many large studies, such as The Cancer Genome Atlas (TCGA) and The COSMIC Cell Lines Project. There are two Illumina Infinium HumanMethylation BeadChips: the MethylationEPIC BeadChip (EPIC) and its predecessor, the HumanMethylation450 BeadChip (450k). The Infinium HumanMethylation450 BeadChip (450k) includes 485,764 CpG sites of the human genome, while the Infinium MethylationEPIC BeadChip (EPIC) includes 853,307, incorporating 333,265 CpG sites located in enhancer regions. [4-6]

Some tools like *minfi* [7], *RnBeads* [8], *ChAMP* [9], or *SeSAMe* [10] were developed to analyze DNA methylation data but, although some of them produce plots, these types of software are not oriented to generate simple high-quality visualizations. To use these pipelines is required programming skills for some steps and their interfaces are not very intuitive for non-bioinformatics clinical researchers. Recently, some applications to analyse and visualize DNA methylation have been published like MethSurv [11], SMART [12], and shinyÉPICo [13]. MethSurv provides a survival analysis from DNA methylation data of The Cancer Genome Atlas (TCGA) project, SMART provides several analyses like differential methylation analysis, correlation analysis, and survival analysis also for TCGA DNA methylation data, and shinyÉpico provides analyses for Illumina Methylation arrays (450k and EPIC) like quality control and differential methylation analysis. These web-based tools are user-friendly and produce useful graphs, but, still, they contain a very basic workflow or lack some advanced analyses and graphs more oriented to real clinical applications.

For this reason, we developed VisualMethyl, which performs different analyses, without the need to code, like quality control, exploratory analysis, differentially methylated CpG sites and regions (DMPs and DMRs), functional enrichment, and survival. This application provides an intuitive interface to follow the pipeline easily and generates a lot of interactive plots to represent the results; in this way, the user can perform complex analyses and obtain results without being an expert bioinformatician. In each step the default parameters can be modified to adapt the analyses for different projects, and also is possible to download a customized report with the results obtained in the application.

## Objectives

The aim of this project is to develop a user-friendly application to analyse and visualize DNA methylation arrays (450k and EPIC). This application should perform complete analyses and show the results with high-quality visualizations for non-bioinformatics clinical researchers, in an interactive way, without programming. The sections that has to contain are:

- Data: Upload at least 100 samples in a flexible and fast way.
- Quality Control and Exploratory Analysis: Perform high-quality interactive visualizations.
- DMPs and DMRs: Calculate differently methylated CpGs and regions with several parameters in an user-friendly manner.
- Other sections: Independent analyses to relate DNA methylation data with clinical data and integrate it with an omics database.

## Material and methods

VisualMethyl is developed in the R programming language using *shiny*, an R package to build interactive web apps from R. The user interface is customized with Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript.

The application analyzes DNA methylation arrays, Illumina HumanMethylation450 BeadChip (450K) and Illumina HumanMethylationEPIC BeadChip (EPIC). This data has to be uploaded in .zip format which must contain the iDAT (raw Intensity Data) files and the sample sheet (Table 1).

The analyses are computed with *minfi* for reading the data and normalization process, *limma* [14] to generate the linear model and DMPs calculation, *mCSEA* [15] to calculate DMRs, *clusterProfiler* [16] for functional enrichment, and *survival* [17], *survminer* [18] and *compareGroups* [19] for survival analysis. There are different options in each section with default parameters that the user can modify in both calculation steps and some plots.

To generate the graphs *ggplot2* [20] package is used and to make interactive visualizations *plotly* [21] is applied. Also, there are more packages to build specific plots like *pheatmap* [22], *gplots* [23] and *heatmaply* [24] for heatmaps, *qqman* [25] for manhattan plots, *MultiDataSet* [26] for volcano plots, *enrichplot* [27] for dotplots, and *survminer* [18] for Kaplan-Meier plot.

Other packages are used to customize the user interface components of the application such as *shinydashboard* [28], *shinyWidgets* [29] and *shinyjs* [30], this provides a modern interface and makes the application more intuitive and easy to use. And *introjs* [31] to build a help tour that navigates through all the sections.

The code is available at GitHub link: https://github.com/ireneefr/VisualMethyl

| Sample_Name | Sample_Group | Pool_ID | Sentrix_ID | Sentrix_Position | Barcode |
|---|---|---|---|---|---|
| TCGA-E9-A1N5-11A-41D-A14H-05 | Solid Tissue Normal | 450k | 6057833101 | R06C02 | 6057833101_R06C02 |
| TCGA-E2-A1IF-11A-22D-A145-05 | Solid Tissue Normal | 450k | 6026818122 | R01C02 | 6026818122_R01C02 |
| TCGA-E9-A1N6-11A-32D-A145-05 | Solid Tissue Normal | 450k | 6026818111 | R01C01 | 6026818111_R01C01 |
| TCGA-BH-A0DG-11A-43D-A12R-05 | Solid Tissue Normal | 450k | 6042324072 | R03C01 | 6042324072_R03C01 |
| TCGA-BH-A1F6-11B-94D-A13T-05 | Solid Tissue Normal | 450k | 6026818144 | R05C02 | 6026818144_R05C02 |
| TCGA-A7-A0CE-11A-21D-A10Q-05 | Solid Tissue Normal | 450k | 6005486013 | R02C02 | 6005486013_R02C02 |
| TCGA-BH-A0B3-11B-21D-A10Q-05 | Solid Tissue Normal | 450k | 6005486011 | R03C01 | 6005486011_R03C01 |
| TCGA-BH-A0H9-11A-22D-A093-05 | Solid Tissue Normal | 450k | 6005486025 | R04C01 | 6005486025_R04C01 |
| TCGA-A7-A0D9-11A-53D-A10Q-05 | Solid Tissue Normal | 450k | 6005486013 | R01C02 | 6005486013_R01C02 |
| TCGA-BH-A0C3-11A-23D-A12R-05 | Solid Tissue Normal | 450k | 6042324070 | R02C01 | 6042324070_R02C01 |
| TCGA-AC-A6IV-01A-12D-A33F-05 | Primary Tumor | 450k | 9422491072 | R04C02 | 9422491072_R04C02 |
| TCGA-BH-A8FZ-01A-11D-A357-05 | Primary Tumor | 450k | 9426020160 | R04C01 | 9426020160_R04C01 |
| TCGA-S3-AA11-01A-31D-A41Q-05 | Primary Tumor | 450k | 9993943017 | R02C02 | 9993943017_R02C02 |
| TCGA-UU-A93S-01A-21D-A41Q-05 | Primary Tumor | 450k | 9993943017 | R05C01 | 9993943017_R05C01 |
| TCGA-EW-A1J1-01A-11D-A13K-05 | Primary Tumor | 450k | 6026818135 | R01C02 | 6026818135_R01C02 |
| TCGA-B6-A0WX-01A-11D-A10A-05 | Primary Tumor | 450k | 6005486018 | R06C02 | 6005486018_R06C02 |
| TCGA-S3-AA10-01A-21D-A41Q-05 | Primary Tumor | 450k | 9993943013 | R05C02 | 9993943013_R05C02 |
| TCGA-E2-A10C-01A-21D-A10N-05 | Primary Tumor | 450k | 6005486012 | R01C02 | 6005486012_R01C02 |
| TCGA-OL-A5D7-01A-11D-A27Y-05 | Primary Tumor | 450k | 9298768172 | R02C01 | 9298768172_R02C01 |
| TCGA-AR-A1AY-01A-21D-A12R-05 | Primary Tumor | 450k | 6042324037 | R06C01 | 6042324037_R06C01 |

Table 1. Example Sample Sheet

# Development and Results

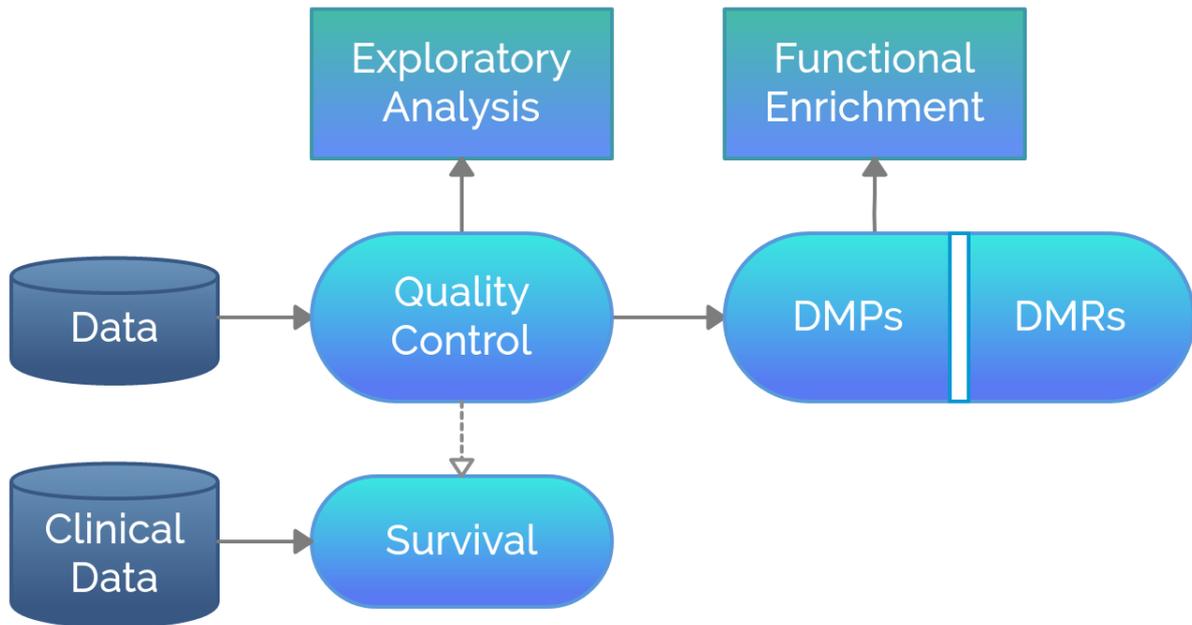The application pipeline (Figure 1) has different sections described below.



Figure 1. VisualMethyl pipeline

## Data

Data is the first section of the application where the user has to upload the data in .zip format; it must contain the iDAT files and the sample sheet in .csv format (Table 1 and Figure S1). When the data is loaded, appear the samples table and some options to select the data correctly:

- Sample Name: Select the sample name column; this column should not contain duplicate values.

- Grouping Variable: Select the column with the variable of interest.

- Donor/Patient: Select donor or patient variable column in case of having it; if there is no donor or patient, the user has to select the sample name column.

- Samples to Process: Select samples (rows) to analyse.

- Sex: Select the sex column in case of having it.

- Age: Select the age column in case of having it.

The columns selected in each option must be different. Then, given the options, the samples are processed to continue with the analysis.

## Quality Control

This step is the most important starting point to identify samples with low quality or different types of bias. This process also includes the normalization step and finally some plots computed with raw and normalized data appear to show the effect of the normalization in the samples (Figure S2). Before running the normalization, the user can select some options:

- Normalization: Select the type of normalization: Raw, Illumina, Funnorm, Noob, SWAN, Quantile, or Noob+Quantile.

- Drop CpHs: Remove non-CpG methylation probes when is selected.

- Drop SNPs: Remove positions with SNPs when is selected.

- Drop X/Y Chr: Remove positions of sex chromosomes when is selected.

When the normalization finishes, some plots are generated:

- Intensities boxplots: Green and red channel intensities by sample. Samples should have similar values between them.

- Failed probes: Probe failure rate (%) by sample with two thresholds (5% and 10%) to observe if there is any bad sample. If there is any sample that exceeds the thresholds, the user should consider removing it from the analysis.
- Control types: Green and red channel log2(intensity) values for the control type selected; the user can select the control type and the slide to plot. There is a threshold to show if the intensity values are correct (above the threshold).

- Predicted sex: The sex of the samples is predicted and can be compared with the sex column of the sample sheet in the case of having it.

- Density plots: Two density plots to compare the samples before and after the normalization process. The user can choose the probe type to plot (I-Green, I-Red, or II). There should be two peaks well represented, and if after the normalization this shape is not obtained, the user has to consider removing the samples that do not follow the correct form.

- SNP analysis: An interactive heatmap computed with the SNPs beta-values of the array. There should be similarities between groups.

- Batch effects: P-values or correlation values can be plotted between principal components and variables. These plots can be useful to find covariables in the model.

## Exploratory Analysis

This section serves to visualize the main characteristics of the data with graphical representations (Figure S3):

- Violin plots: Similar to the density plots, to compare the samples before and after the normalization process.

- Principal Component Analysis: Two Principal Components are plotted and a table with the main statistics of the Principal Components is shown. The user can select the Principal Components to plot and the variable by which the samples are coloured.

- Heatmaps: Two heatmaps are generated, one with 1000 random CpGs and the other with the 1000 top variable CpGs where should be appreciated differences between groups of samples.

- Deconvolution: A simple deconvolution analysis is done and is represented with a heatmap. This type of plot is interesting if the user is analysing blood samples.

- Age methylation: The methylation age of the samples is predicted with the Horvath method [32]. This can be compared with the age column of the sample sheet in the case of having it.

- Hypo/Hyper: Three bar plots are computed to represent the number of CpGs hypermethylated and hypomethylated by chromosome or their genomic location. The user can select the samples to plot and the beta-value threshold.

## DMPs/DMRs

In this step there are two tabs:

- ➢ Differentially Methylated Positions (DMPs): (Figure S4) To perform this analysis there are some options to generate a linear model:

    - Grouping Variable: Variable of interest to perform the contrasts, by default is the grouping variable selected in the data section.

    - Linear model covariables: Select some variable that can affect the differences in methylation, by default the donor/patient variable is selected or nothing in case of not having donor/patient.

    - Linear model interactions: Select some variable that can influence depending on another variable, by default nothing is selected.

    - Array weights: To estimate the relative quality weights of each array.

    - Contrast options: The user can select eBayes Trend and/or eBayes Robust.

When the calculation is done, appear some results:

- ◆ <u>DMP table and options</u>: Appears a table with the number of hypermethylated and hypomethylated positions and some default parameters that can be modified like the min. ΔBeta, the max. FDR and the max. p-value.

- ◆ <u>DMP heatmap</u>: A heatmap is calculated with the DMPs and can be customized with options like the clustering algorithm (average, single, complete, mcquitty, median, or centroid), the distance function (pearson, spearman, kendall, or euclidean), the scale (row or none), and adding column and/or row colors.

- ◆ <u>DMPs annotation</u>: There is a table with the DMPs and the user can select one of them to generate a boxplot that shows the different beta values of the DMP selected between groups.

- ◆ <u>DMP manhattan</u>: A manhattan plot is generated with the DMPs data and the user can select the contrast to plot.

- ◆ <u>DMP volcano</u>: A volcano plot is generated with the DMPs data and the user can select the contrast to plot.

➢ <u>Differentially Methylated Regions (DMRs)</u>: (Figure S5) To perform this analysis there are also some options:

- - <u>Contrasts to calculate</u>: Select the contrasts to calculate the DMRs.

- - <u>Type of DMRs</u>: The user can choose different types of DMRs: promoters, genes and CGI.

- - <u>Array platform</u>: The array platform can be EPIC or 450k.

- - <u>Min. CpGs in DMR</u>: This is the minimum number of CpGs to define a DMR, by default are 5.

After the calculation, some results are generated:

- ◆ <u>DMR table and options</u>: Appears a table with the number of hypermethylated and hypomethylated regions and some default parameters that can be modified like the min. ΔBeta, the max. FDR and the max. p-value.

- ◆ <u>DMR heatmap</u>: A heatmap is calculated with the DMRs and can be customized with options like the clustering algorithm (average, single, complete, mcquitty, median, or centroid), the distance function (pearson, spearman, kendall, or euclidean), the scale (row or none), and adding column and/or row colors.

◆ <u>DMR annotation</u>: There is a table with the DMRs that can be changed by the different DMRs types and the user can select one of the items to generate a genomic graph of the DMR selected.

## Functional Enrichment

The functional enrichment analysis associates the DMPs obtained in the DMPs/DMRs section with a biological theme (Figure S6). To represent this, dotplots are generated for different databases:

◆ <u>KEGG</u>: A database for the gene functions of the biological system. (https://www.kegg.jp/)

◆ <u>Gene Ontology (GO)</u>: A knowledge base of terms corresponding to gene product characteristics for three domains: molecular function, biological process and cellular component. (http://geneontology.org/)

◆ <u>Reactome</u>: A database of biological pathways. (https://reactome.org/)

## Survival

The survival analysis studies the duration of time to achieve an outcome, for example the death of an organism or the response to a treatment. The user has to upload clinical data with at least a sample name column, a time column and a status column; this data can contain other optional columns that can be used for the analysis.

To run this analysis the methylation data (prior normalization) or at least one clinical information is needed, with a maximum of two. The results obtained with the Cox Proportional Hazards Model are represented in a Kaplan-Meier plot with a risk table; also is obtained some statistical information for each variable and, if methylation data is used, frequencies of the variable values for each grouping variable of interest in the methylation analysis (Figure S7).

## Export

The user can download a report in HTML format with the results obtained in the application. This report can be customized by the selection of the sections and plots to be included in it; it is possible to download the report if not all the sections are done, but those that require some computational step will be blocked until this step is done (Figure S8).

## Help

This section explains the content of the application and helps the user to understand and navigate through all the pipeline. There is also a help tour to give an example of use, where the user can do, step by step, all the analyses with an example data (Figure S9).

## Use Case

In order to test VisualMethyl, publicly available data from The Cancer Genome Atlas (TCGA) database were downloaded. We use 20 samples from the TCGA-BRCA project, 10 that are Solid Tissue Normal and 10 Primary Tumor. This DNA methylation data was measured with the Illumina HumanMethylation450 Beadchip (450k). The aim of analysing this data is to find differences between Solid Tissue Normal and Primary Tumor, identifying which positions and regions occur these changes, and provide biological sense with a functional enrichment.

Once the iDAT files and the sample sheet is obtained (Table 1), this data is uploaded to the application in .zip format. For this example, the grouping variable is Sample_Group column and there is no donor/patient, sex, or age column (Figure S1).

In the quality control section, the Noob normalization is performed and we can observe that all the samples are correct (Figure S2), for example, the Density Plots show the corresponding peaks and how the normalization has worked (Figure 2).
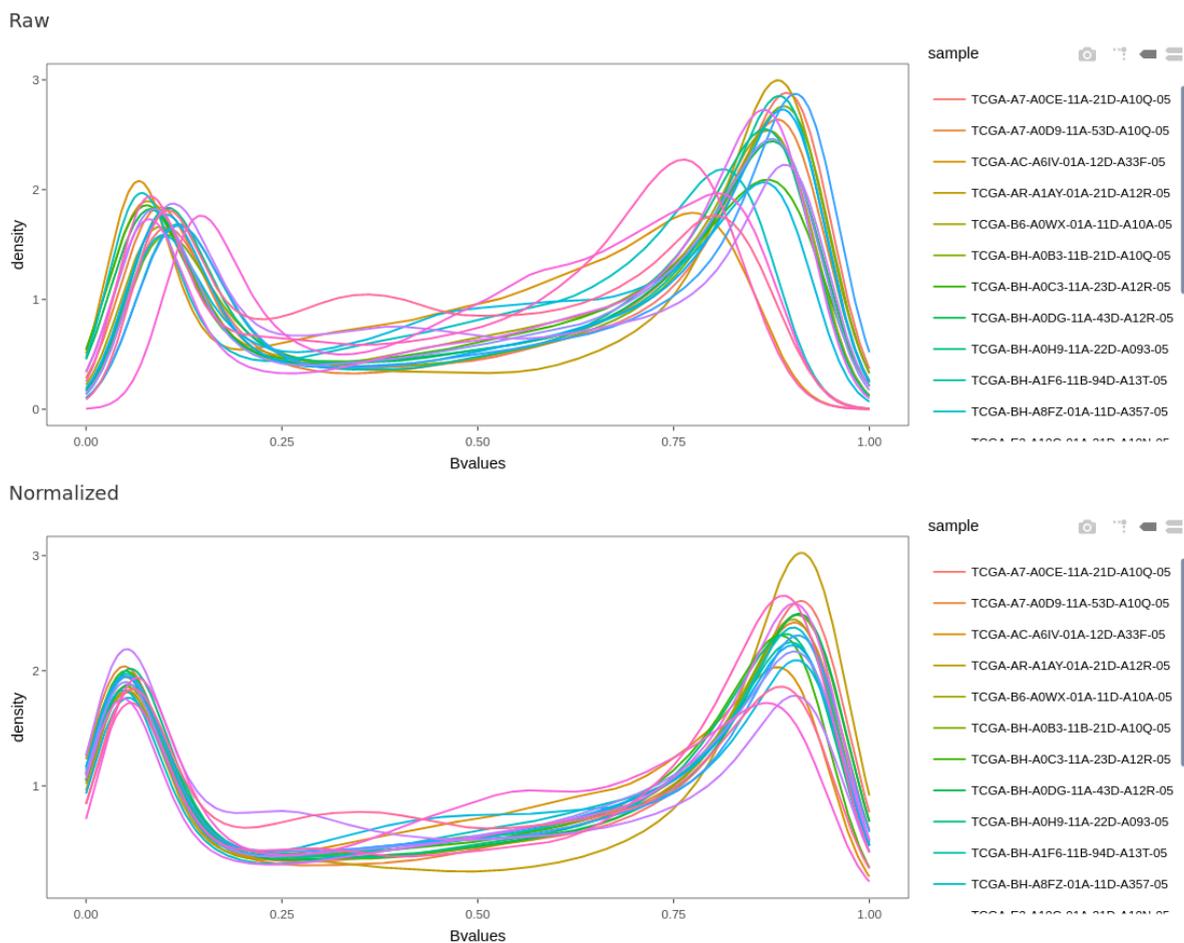


Figure 2. Density Plots with the example data

The exploratory analysis gives us more information about the data (Figure S3); for example, the Principal Component Analysis plot shows how the samples are separated by the principal components (Figure 3).
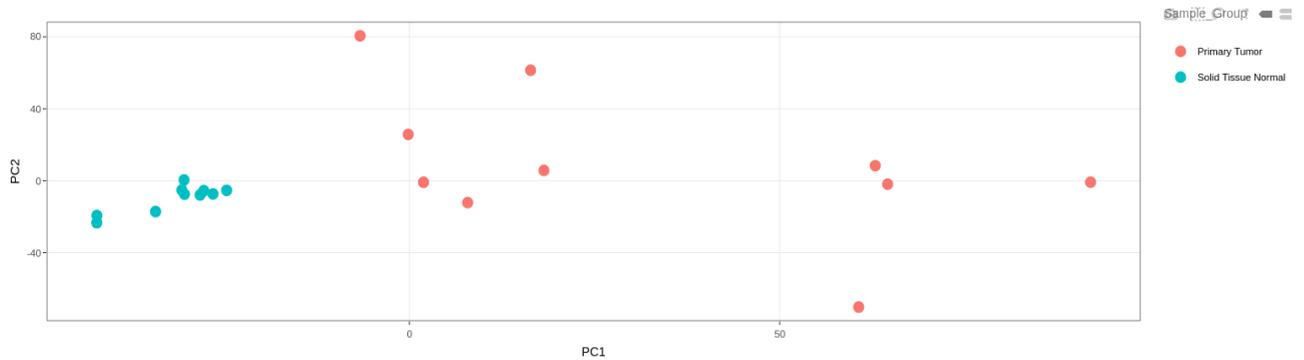


Figure 3. Principal Component Analysis plot with the example data

Then, the DMPs/DMRs analysis can be done. In this case we use the default parameters in both sections. The results show the differences between Solid Tissue Normal and Primary Tumor (Figure 4, Figure S4 and Figure S5).
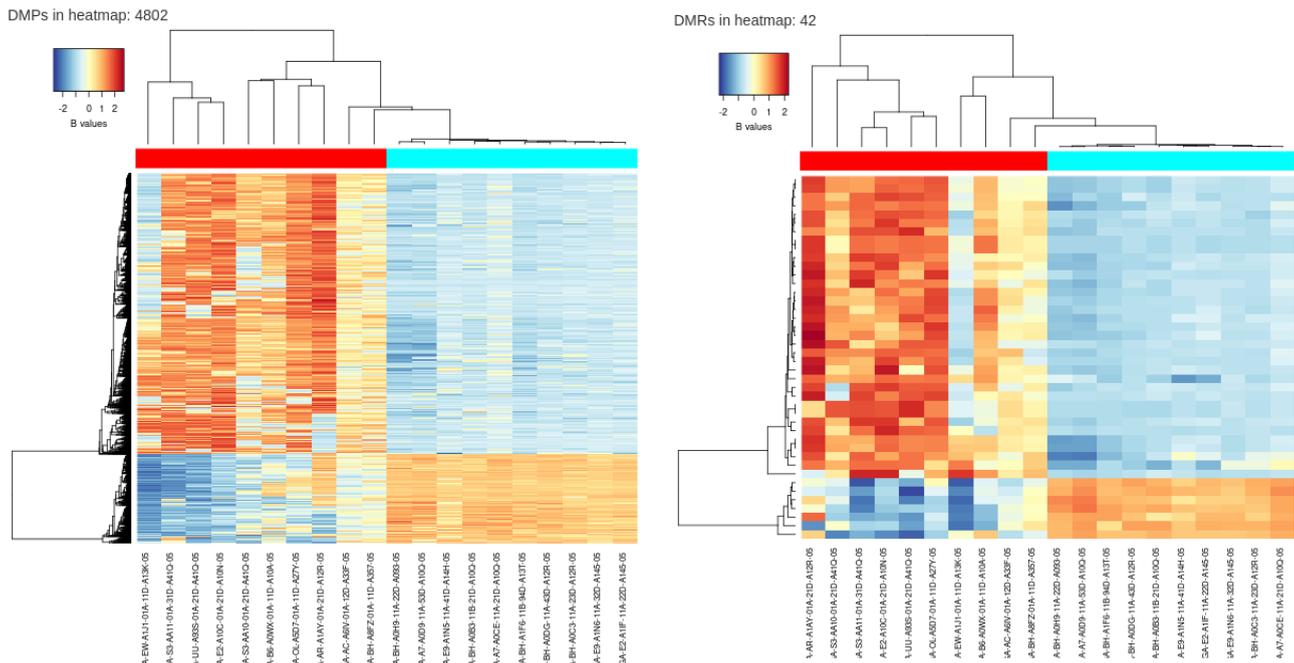


Figure 4. DMPs and DMRs heatmaps with the example data

Finally, a Functional Analysis is performed obtaining some biological information that can be useful to explain the differences between groups (Figure 5 and Figure S6).
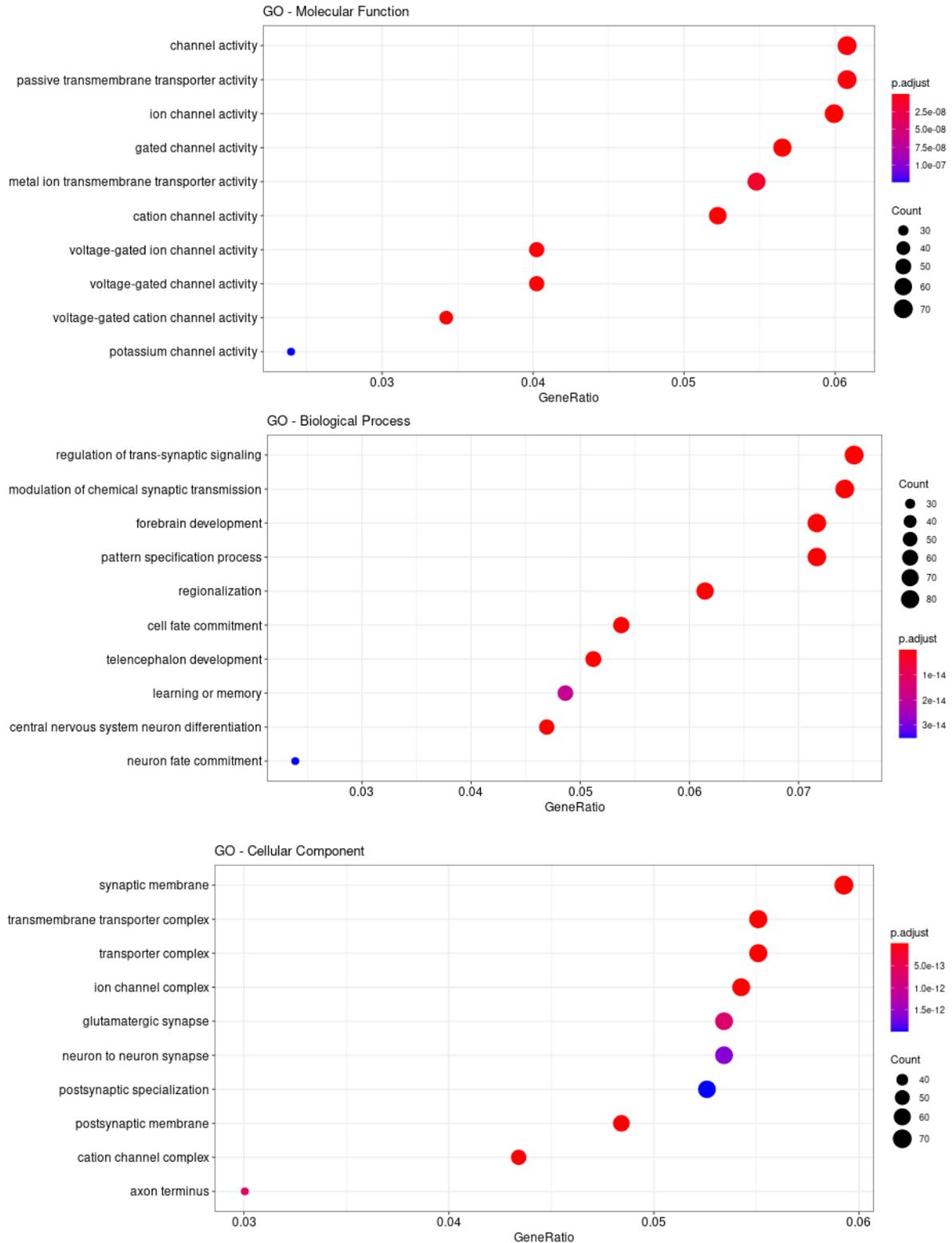


Figure 5. Gene Ontology Enrichment Analysis with the example data

## Conclusion

VisualMethyl is a Shiny web application for interactive DNA methylation analysis and visualization. The application is easy to use because the user does not need programming skills and the results are represented in interactive plots.

This application presents a complete pipeline with analyses that can be very useful for researchers and clinicians without specific bioinformatics skills: quality control, exploratory analysis, differentially methylated CpG sites and regions, functional enrichment, and survival analysis. It also avoids time consuming since all the analyses are incorporated into a single platform where they are computed in a user-friendly interface without programming.

For the moment, only the analyses described above are completely available, but in the new version of the application under development, other modules will be incorporated to create a complete platform like predictive models to find potential biomarkers applying machine learning approaches; these types of analyses can be very useful for future clinical studies involving the research of new DNA methylation prognosis biomarkers. The application is going to be integrated with external databases like The Cancer Genome Atlas (TCGA) to integrate the DNA methylation with other types of omics data as well as other relevant clinical parameters.

## Bibliography

1. Esteller, Manel. 2008. "Epigenetics in Cancer." *The New England Journal of Medicine* 358 (11): 1148–59. https://doi.org/10.1056/NEJMra072067.

2. Moore, Lisa D., Thuc Le, and Guoping Fan. 2013. "DNA Methylation and Its Basic Function." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 38 (1): 23–38. https://doi.org/10.1038/npp.2012.112.

3. Sharma, Shikhar, Theresa K. Kelly, and Peter A. Jones. 2010. "Epigenetics in Cancer." *Carcinogenesis* 31 (1): 27–36. https://doi.org/10.1093/carcin/bgp220.

4. Moran, Sebastian, Carles Arribas, and Manel Esteller. 2016. "Validation of a DNA Methylation Microarray for 850,000 CpG Sites of the Human Genome Enriched in Enhancer Sequences." *Epigenomics* 8 (3): 389–99. https://doi.org/10.2217/epi.15.114.

5. Bibikova, Marina, Jennie Le, Bret Barnes, Shadi Saedinia-Melnyk, Lixin Zhou, Richard Shen, and Kevin L. Gunderson. 2009. "Genome-Wide DNA Methylation Profiling Using Infinium® Assay." *Epigenomics* 1 (1): 177–200. https://doi.org/10.2217/epi.09.14.

6. Li, Yuanyuan, and Trygve O. Tollefsbol. 2011. "DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis." *Methods in Molecular Biology* 791: 11–21. https://doi.org/10.1007/978-1-61779-316-5_2.

7. 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): 1363–69. https://doi.org/10.1093/bioinformatics/btu049.

8. Müller, Fabian, Michael Scherer, Yassen Assenov, Pavlo Lutsik, Jörn Walter, Thomas Lengauer, and Christoph Bock. 2019. "RnBeads 2.0: Comprehensive Analysis of DNA Methylation Data." *Genome Biology* 20 (1): 55. https://doi.org/10.1186/s13059-019-1664-9.

9. Tian, Yuan, Tiffany J. Morris, Amy P. Webster, Zhen Yang, Stephan Beck, Andrew Feber, and Andrew E. Teschendorff. 2017. "ChAMP: Updated Methylation Analysis Pipeline for Illumina BeadChips." *Bioinformatics* 33 (24): 3982–84. https://doi.org/10.1093/bioinformatics/btx513.

10. Zhou, Wanding, Timothy J. Triche Jr, Peter W. Laird, and Hui Shen. 2018. "SeSAMe: Reducing Artifactual Detection of DNA Methylation by Infinium BeadChips in Genomic Deletions." *Nucleic Acids Research* 46 (20): e123. https://doi.org/10.1093/nar/gky691.

11. Modhukur, Vijayachitra, Tatjana Iljasenko, Tauno Metsalu, Kaie Lokk, Triin Laisk-Podar, and Jaak Vilo. 2018. "MethSurv: A Web Tool to Perform Multivariable Survival Analysis Using DNA Methylation Data." *Epigenomics* 10 (3): 277–88. https://doi.org/10.2217/epi-2017-0118.

12. Li, Yin, Di Ge, and Chunlai Lu. 2019. "The SMART App: An Interactive Web Application for Comprehensive DNA Methylation Analysis and Visualization." Epigenetics & Chromatin 12 (1): 71. https://doi.org/10.1186/s13072-019-0316-3.

13. Morante-Palacios, Octavio, and Esteban Ballestar. 2021. "shinyÉPICo: A Graphical Pipeline to Analyze Illumina DNA Methylation Arrays." *Bioinformatics* , January. https://doi.org/10.1093/bioinformatics/btaa1095.

14. Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47. https://doi.org/10.1093/nar/gkv007.

15. Martorell-Marugán, Jordi, Víctor González-Rumayor, and Pedro Carmona-Sáez. 2019. "mCSEA: Detecting Subtle Differentially Methylated Regions." *Bioinformatics* 35 (18): 3257–62. https://doi.org/10.1093/bioinformatics/btz096.

16. Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "clusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters." *Omics: A Journal of Integrative Biology* 16 (5): 284–87. https://doi.org/10.1089/omi.2011.0118.

17. Therneau, Terry, and T. Lumley. 2013. "R Survival Package." https://rweb.webapps.cla.umn.edu/R/library/survival/doc/survival.pdf.

18. Kassambara, Alboukadel, Marcin Kosinski, Przemyslaw Biecek, and S. Fabian. 2017. "Survminer: Drawing Survival Curves using'ggplot2'." *R Package Version 0. 3* 1.

19. Subirana, Isaac, Héctor Sanz, Joan Vila, and Others. 2014. "Building Bivariate Tables: The compareGroups Package for R." *Journal of Statistical Software* 57 (12): 1–16. https://www.jstatsoft.org/htaccess.php?volume=57&type=i&issue=12&paper=true.

20. Wickham, Hadley. 2011. "ggplot2." *Wiley Interdisciplinary Reviews. Computational Statistics* 3 (2): 180–85. https://doi.org/10.1002/wics.147.

21. Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. CRC Press. https://play.google.com/store/books/details?id=7zPNDwAAQBAJ.

22. Kolde, Raivo. 2012. "Pheatmap: Pretty Heatmaps." *R Package Version* 1 (2).

23. Warnes, Gregory R., Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, et al. 2009. "Gplots: Various R Programming Tools for Plotting Data." *R Package Version* 2 (4): 1.

24. Galili, Tal, Alan O'Callaghan, Jonathan Sidi, and Carson Sievert. 2018. "Heatmaply: An R Package for Creating Interactive Cluster Heatmaps for Online Publishing." Bioinformatics  34 (9): 1600–1602. https://doi.org/10.1093/bioinformatics/btx657.

25. Turner, S. 2017. "Qqman: QQ and Manhattan Plots for GWAS Data. R Package Version 0.1. 4."

26. Hernandez-Ferrer, Carles, Carlos Ruiz-Arenas, Alba Beltran-Gomila, and Juan R. González. 2017. "MultiDataSet: An R Package for Encapsulating Multiple Data Sets with Application to Omic Data Integration." *BMC Bioinformatics* 18 (1): 36. https://doi.org/10.1186/s12859-016-1455-1.

27. Yu, G. 2018. "Enrichplot: Visualization of Functional Enrichment Result." *R Package Version* 1 (2).

28. Chang, Winston, and Barbara Borges Ribeiro. 2018. "Shinydashboard: Create Dashboards with 'Shiny.'" *R Package Version 0. 7* 1.

29. Perrier, V., F. Meyer, and D. Granjon. 2019. "shinyWidgets: Custom Inputs Widgets for Shiny." *R Package Version*.

30. Perrier, V., F. Meyer, and D. Granjon. 2019. "shinyWidgets: Custom Inputs Widgets for Shiny." *R Package Version*.
31. Arasteh, Ehsan, and Afshin Mehrabani. 2013. *Instant IntroJs*. Packt Publishing Ltd.
32. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* 14, 3156 (2013). https://doi.org/10.1186/gb-2013-14-10-r115

Supplementary data:

https://drive.google.com/file/d/1tnITBFGyLSGycCpHaPo2dOwgRySx9-kk/view?usp=sharing