



# **Analysis of Bisulfite Sequencing PCR**

## **User guide**

Version 1.0.0

Marie Denoulet and Chann Lagadec  
June 3, 2022

# Contents

---

List of Figures	3
List of Code Listings	3
1 General information	4
1.1 What ABSP does do?	4
1.2 What are the advantages of using ABSP?	4
1.3 Bisulfite Sequencing PCR	5
1.4 How does ABSP work?	7
1.5 License	9
2 How to proceed with analysis using ABSP?	12
2.1 Open ABSP for the first time	12
2.1.1 Download and installations	12
2.1.2 Content of the ABSP main directory	12
2.1.3 Launch the app	13
2.2 Open ABSP (not for the first time)	14
2.2.1 Launch the app	14
2.3 Update ABSP	14
2.3.1 Launch the app	14
2.4 Individual analysis	14
2.4.1 Input files requirements	14
2.4.2 Procedure	16
2.4.3 Output report	17
2.4.4 Output files	22
2.5 Grouped analysis	24
2.5.1 Procedure	24
2.5.2 Output report	25
2.5.3 Output files	27
2.6 Multiple analyses	33
2.6.1 Input files requirements	33
2.6.2 Procedure	34
2.6.3 Output files	34
3 Complementary information	35
3.1 Some recommendations for the BSP experiment	35
3.2 Detailed workflow of ABSP individual analysis	35
3.3 Code modifications	37
3.3.1 List of reference genomes	37
3.3.2 Modify the default thresholds	38
3.3.3 Modify the plots colors and point shapes	38

4	Troubleshooting guide	41
4.1	General . . . . .	41
4.2	Individual analysis . . . . .	43
4.3	Grouped analysis . . . . .	46
5	Acknowledgements	47

## List of Figures

---

1	Bisulfite Sequencing PCR experimental principle . . . . .	5
2	Bisulfite Sequencing PCR analysis strategies for both direct-BSP and cloning-BSP . . . . .	6
3	Workflow of the ABSP analytic process . . . . .	8
4	Detailed workflow of the ABSP analytic process . . . . .	10
5	Diagram of the possible ways to launch ABSP analyses . . . . .	11
6	Example of <i>.fasta</i> file for the reference DNA input required for ABSP individual analysis . . . . .	15
7	IGV (Integrative Genomics Viewer) software window . . . . .	16
8	Example of trimming plot from the output report . . . . .	20
9	Diagram of output directories to locate output files from individual analysis . . . . .	23
10	Visualization plots of all replicates (direct-BSP) . . . . .	28
11	Visualization plots of all clones from one sample (cloning-BSP) . .	29
12	Visualization plots of groups (means of replicates/ clones per samples)	30
13	Boxplots and methylation profile plots . . . . .	32
14	Diagram of output directories to locate output files from individual analysis . . . . .	32
15	Detailed workflow of the individual analysis . . . . .	36

## List of Code Listings

---

1	List of reference genomes displayed in the drop-down lists . . . . .	37
2	Genome name selected by default in the individual analysis tab. . .	37
3	Genome name selected by default in the grouped analysis tab. . .	38
4	Default thresholds in the individual analysis script . . . . .	38
5	Default thresholds in the grouped analysis script . . . . .	39
6	Colors and shapes setting for plots in the grouped analysis script .	40

# 1 General information

## 1.1 What ABSP does do?

ABSP, standing for "[Analysis of Bisulfite Sequencing PCR](#)", is an R based tool to analyze CpG methylation profiles using data from Bisulfite Sequencing PCR (BSP) experiment results. It was developed to help researchers to estimate and compare methylation percentages of a DNA region studied using BSP experiments. It provides a complete automated workflow, from trace file sequencing results to data visualization and statistics.

## 1.2 What are the advantages of using ABSP?

- **A complete workflow.** ABSP uses as input the chromatogram trace files as the sequencing results, and through a two-steps analysis, it (1) computes the methylation percentages of individual samples after validating the sequencing quality and (2) gathers the methylation levels from all samples to summarize methylation data, generate publication-ready figures and perform comparative statistics to answer to the experiment hypothesis on the DNA methylation differences between conditions.
- **A fully automated process.** ABSP uses a shiny app on R to provide a user-friendly interface. To launch the analytic process the user is guided to provide the required inputs and can launch the desired analysis with one click. For each analysis, an HTML report file is generated to visualize the results and keep a record of them. Additionally, output files, such as tables and figures, are automatically saved in the corresponding result folders. For an even more automated use, several analyses can be launched with the help of pre-filled input tables (spreadsheet document to fill) in the special tab "Multiple analyses", which is useful for large amounts of samples.
- **Analyses of both direct-BSP and cloning-BSP sequencing data.** ABSP can analyze results from both BSP methods. No existing tool is currently able to analyze both. It allows continuity in the experiment analytic process, as the direct-BSP approach can be performed before cloning the PCR products to have preliminary insights on DNA methylation, and then further confirmed/validated using cloning-BSP.
- **Accessible and flexible.** ABSP is coded using R, a cross-platform tool language increasingly used in biology research, making it very accessible to any researchers. Additionally, for researchers accustomed to R coding, as the entire scripts are provided, ABSP is fully upgradeable. Also, we provide specific guidelines to easily modify some features to adapt ABSP to experiment needs, as adjusting quality thresholds or changing graphical parameters (see section 3.3 Code modifications at page 37).

### 1.3 Bisulfite Sequencing PCR

The Bisulfite Sequencing PCR (BSP) is an experimental technique aiming to estimate methylation levels of CpG sites on a specific DNA region of interest, among a population of DNA molecules. The method was originally developed by Frommer *et al.* in 1992<sup>1</sup> and Clark *et al.* in 1994<sup>2</sup> and was named BSP in opposition to the methylation specific PCR (MSP) method by Li *et al.* in 2002.<sup>3</sup>

This method is composed of three steps, described in figure 1:

1. A DNA bisulfite conversion
2. A PCR amplification and an optional cloning
3. A sequencing of either PCR products or individual subclones

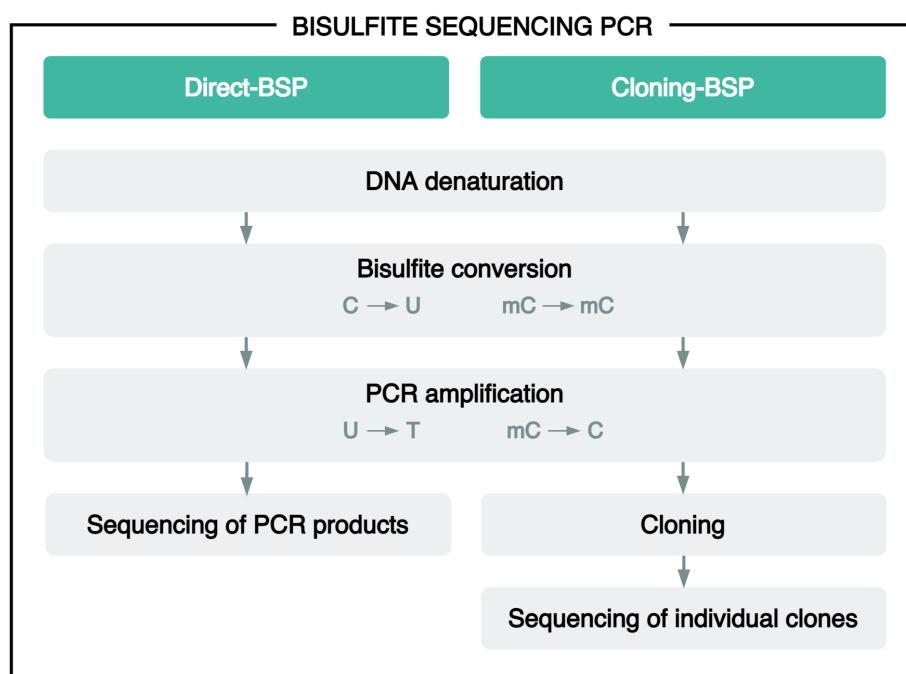


Figure 1. Bisulfite Sequencing PCR experimental principle.

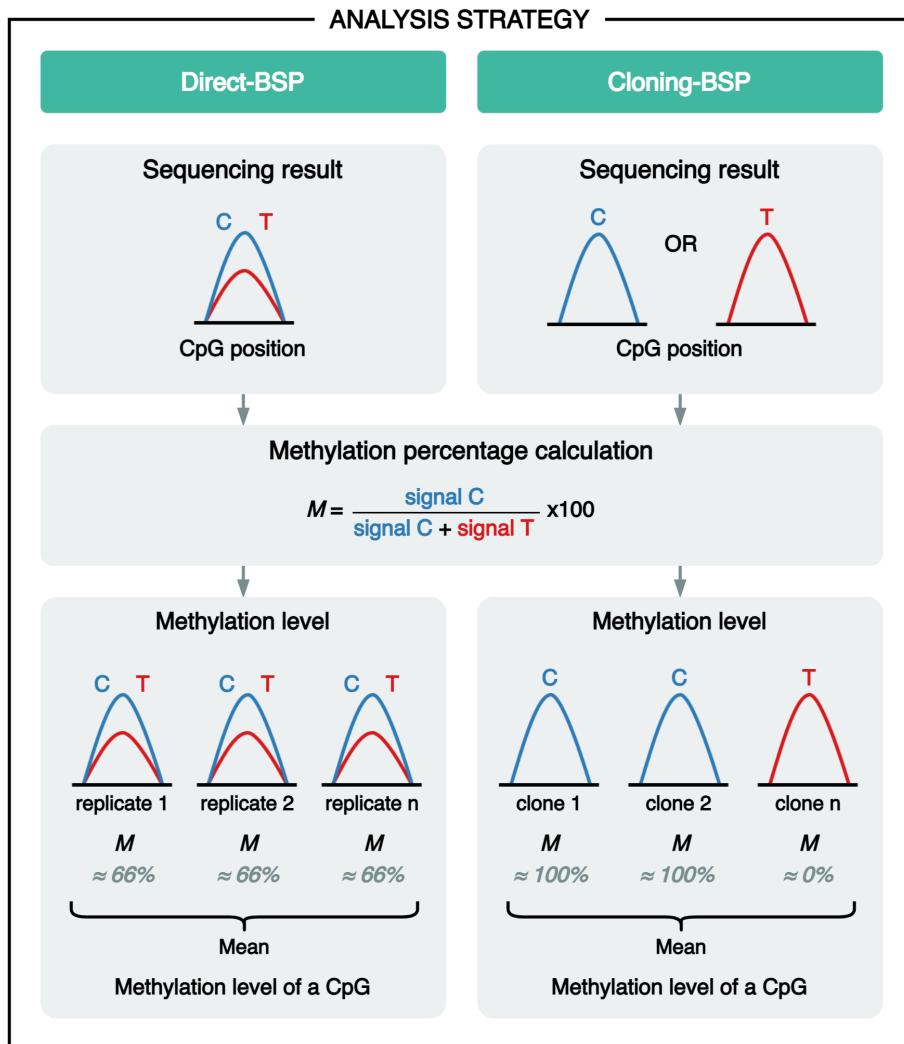
Two approaches of BSP could be used (figure 1). The *direct-BSP* method is characterized by the direct sequencing of PCR products, whereas the *cloning-BSP* consists in cloning PCR products within a specific vector, use it to transform and select bacteria, select several clones, and sequence the individual clones.

<sup>1</sup>M Frommer et al. "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands." In: *Proceedings of the National Academy of Sciences* 89.5 (1992), pp. 1827–1831. ISSN: 0027-8424. doi: [10.1073/pnas.89.5.1827](https://doi.org/10.1073/pnas.89.5.1827).

<sup>2</sup>S J Clark et al. "High sensitivity mapping of methylated cytosines." In: *Nucleic acids research* 22.15 (1994), pp. 2990–7. ISSN: 0305-1048. doi: [10.1093/nar/22.15.2990](https://doi.org/10.1093/nar/22.15.2990).

<sup>3</sup>Long-Cheng Li and Rajvir Dahiya. "MethPrimer: designing primers for methylation PCRs". In: *Bioinformatics* 18.11 (2002), pp. 1427–1431. ISSN: 1367-4803. doi: [10.1093/bioinformatics/18.11.1427](https://doi.org/10.1093/bioinformatics/18.11.1427).

As described in [figure 2](#), the analysis strategies of these two sub-methods are different on some points.



[Figure 2](#). Bisulfite Sequencing PCR analysis strategies for both direct-BSP and cloning-BSP.

First, in **direct-BSP**, mix of DNA molecules with different unknown methylation statuses are sequenced, thereby, at each CpG site, two base signals can co-exist: the methylated signal (C) and the unmethylated signal (T). By calculating the signal ratio, the methylation level of a CpG in the DNA population can directly be estimated, but for reproducibility and statistical significance purposes, it still needs to be repeated in several biological replicates to obtain the final methylation level of a CpG (direct-BSP results are considered less quantitative than cloning-BSP ones).

Secondly, in **cloning-BSP**, as each clone represents only one PCR product, the methylation status of a CpG can be either methylated (C) or unmethylated (T),

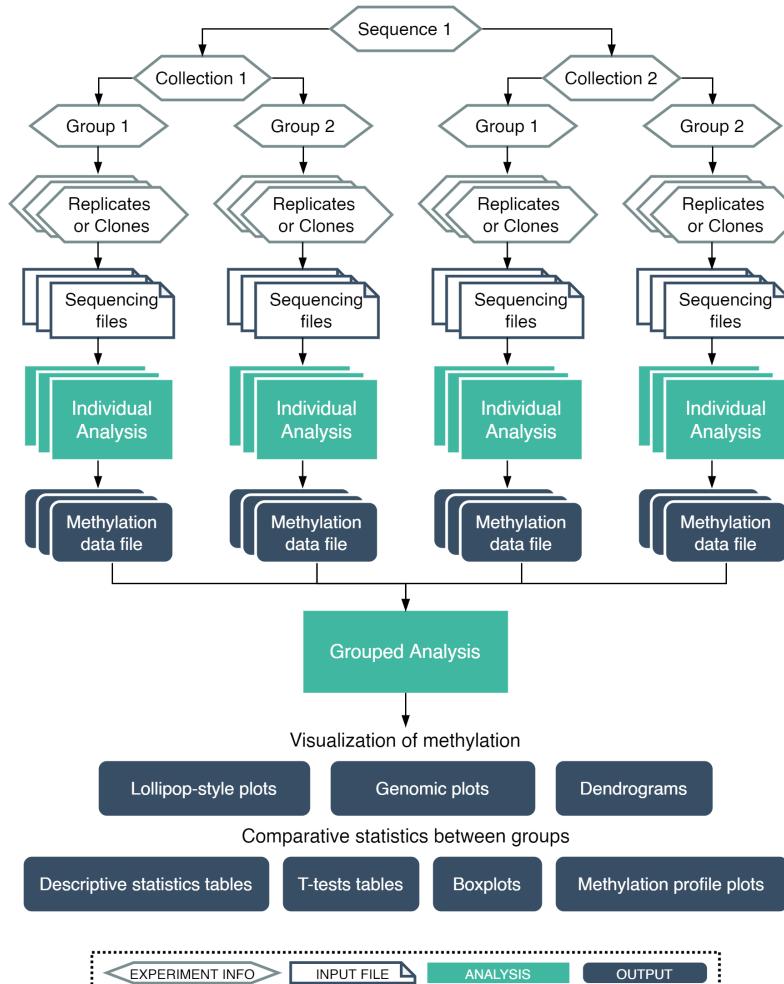
thereby only one of these two signals can exist. Accordingly, the signal ratio can only give either 0% or 100% of methylation (partial methylation is considered biased results), revealing the methylation status. For each CpG, the proportion of clones with a methylated status reveals the methylation level of the CpG in the original DNA population, in general, a minimum of 10 clones is recommended to have a 10% accuracy of the methylation level.

## 1.4 How does ABSP work?

As a first step, each BSP sequencing result is defined by a combination of experiment information ([figure 3](#)):

- **Sequence.** The sequence identifier refers to a unique amplicon sequence produced by the BSP experiment, using a unique set of primer. For example, if several regions of a gene are analyzed by BSP, as each region corresponds to a unique amplicon they must have distinct sequence name (e.g. *CDH1promoter* and *CDH1exon1*; *CDH1-1*, *CDH1-2*, and *CDH1-3*). Make sure the sequence name is strictly identical for all samples of the same sequence.
- **Collection.** The collection corresponds to a separation of samples above groups. Samples from different collections can not be compared, even if they belong to the same group. For example, collections can be different cell lines, organs, or patients, in which the same groups are compared but not between the different collections. To compare these types of samples, consider them as groups. Make sure the collection name is strictly identical for all samples of the same collection.
- **Group.** The group corresponds to the condition that will be compared with other groups/conditions in the grouped analysis. For example, groups can be the "control" and "treated" conditions. Make sure the group name is strictly identical for all samples of the same group.
- **Replicate.** Information has to be provided only when using the direct-BSP approach. The replicate number refers to the repetition identifier number of the sequencing. In order to have robust and reproducible data and to perform comparative statistics, each sample needs to be sequenced at least three times (in both directions).
- **Clones.** Information has to be provided only when using the cloning-BSP approach. In order to estimate the methylation levels among the DNA population, the methylation statuses of several individual clones needs to be sequenced. The ratio of methylated and unmethylated clones for a CpG position will gives the methylation level estimation.
- **Sequencing files.** The sequencing read of each sample is performed in both directions, using a forward primer and a reverse primer.
- **Experiment.** The term "experiment (data)" refers to the unique combination

of collection, group and replicate or clone, for a specific sequence, corresponding to the sample information for the sequencing read.



**Figure 3.** Workflow of the ABSP analytic process.

Then, the ABSP analysis is divided into two steps, corresponding to two scripts (R markdown scripts), as illustrated in the workflow in [figure 3](#).

- **Individual analysis.** For each individual experiment point, two sequencing .ab1 files (one from forward direction, one from reverse direction) are used as input for the individual analysis. First, the sequencing reads are trimmed based on quality to get the correct sequence for alignments with the reference DNA input. Then, results from the alignments go through a quality control step to check for mismatches, gaps, length of aligned sequences, and bisulfite conversion rates (calculated on cytosines outside CpG that should be thymines). If the results are defined as correct, the methylation levels of CpG can be calculated and visualized on a genomic plot. Several output files (e.g. chromatograms, sequences, tables) are saved in folders, especially the methylation data file as a result of the individual sequencing experiment

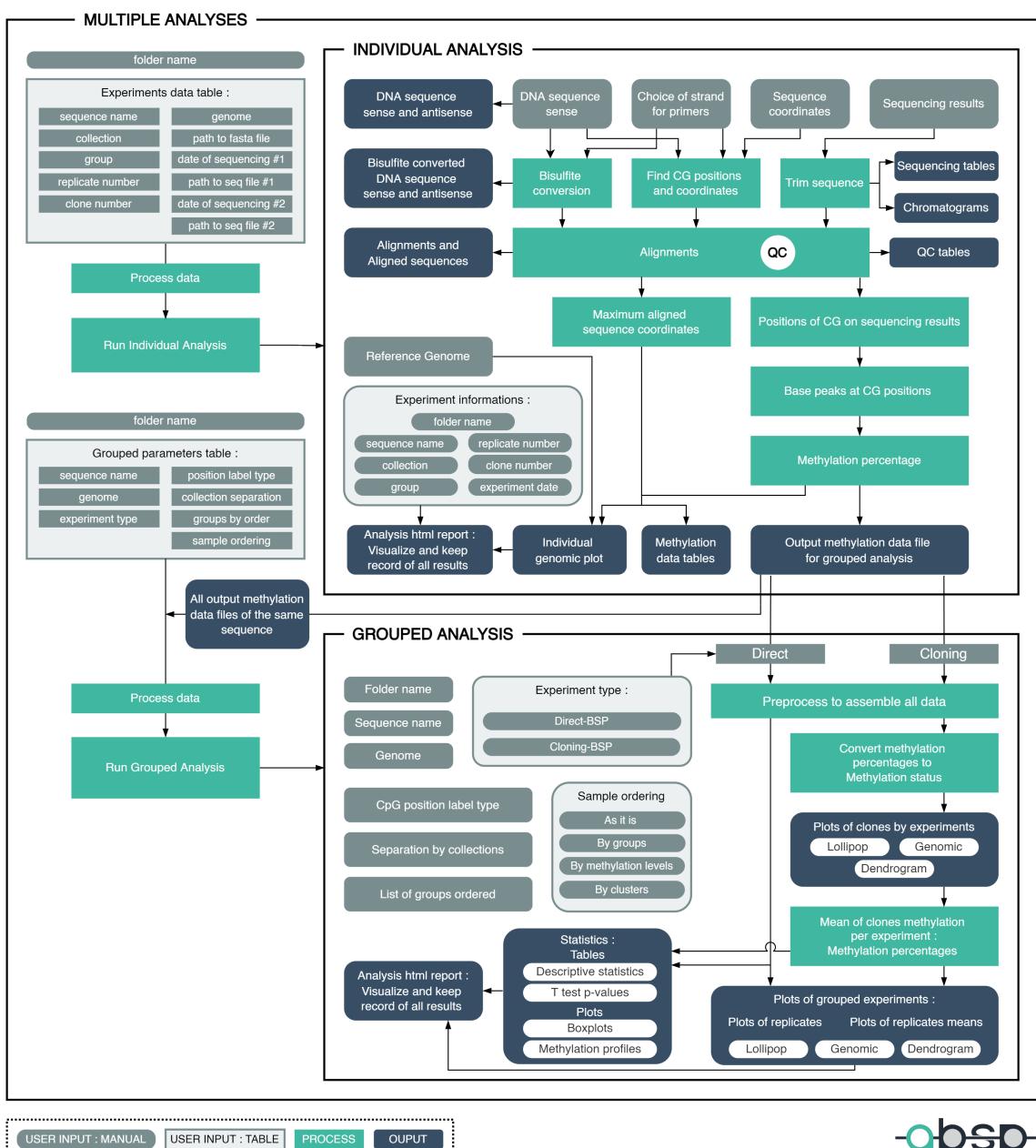
analyzed.

- **Grouped analysis.** All methylation data files from the same sequence are gathered by the grouped analysis. First, a preprocessing step is performed to organize data. Then, visualization plots, lollipop-style plots, and genomic plots (with associated clustering dendograms) are generated to view methylation data differences. Finally, a statistical analysis is performed, descriptive statistics tables and Student's t-test p-values tables are generated, as well as boxplots with t-test p-values and methylation profile plots with Kruskal–Wallis test p-values, to display the significant methylation differences.

For more details, the inputs, processes, and outputs of these two steps, individual analysis, and grouped analysis, are displayed in the [figure 4](#), where the 3 tabs of ABSP are represented: "**Individual analysis**", "**Grouped analysis**", and "**Multiple analyses**".

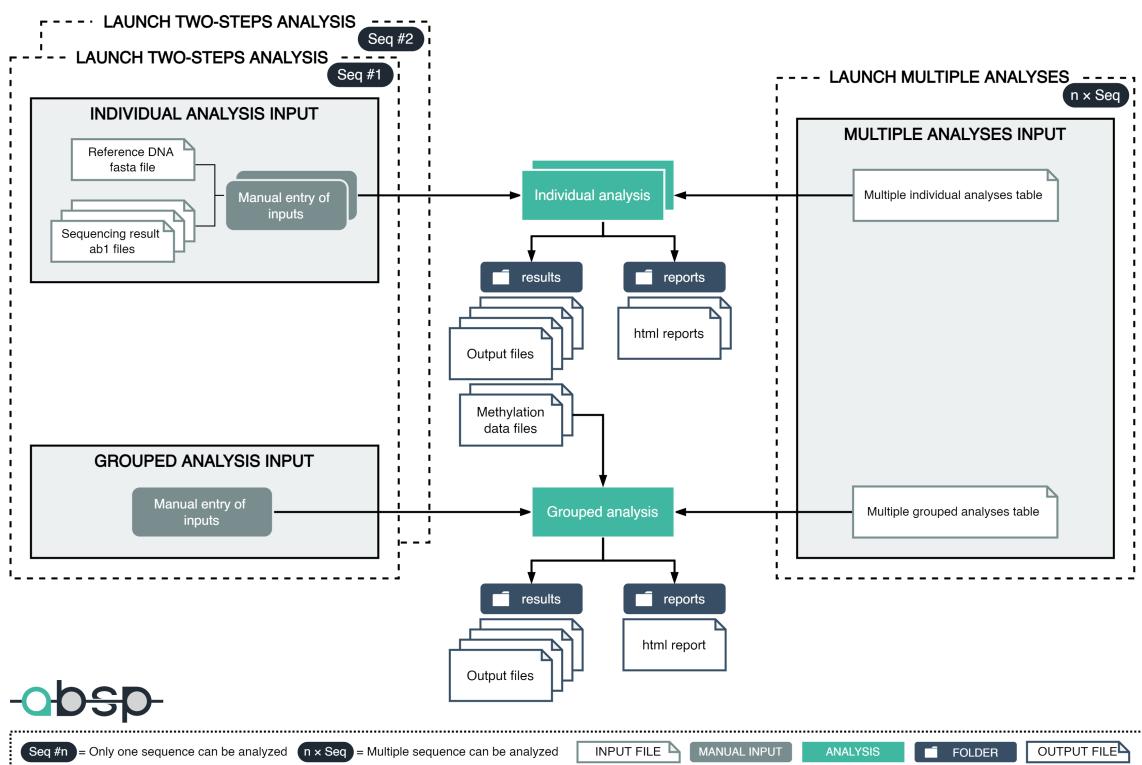
The two steps, individual analysis and grouped analysis, require the manual entry of input data in the corresponding tab, and therefore only one analysis can be launched at the same time ([figure 5](#)). An additional tab has been implemented to launch multiple analyses all at once. This multiple analyses tab can be used to launch either several individual analyses and/or several grouped analyses at the same time, by using tables (.x/sx files or .csv files) as input instead of the manual entry of input for unique analysis.

## 1.5 License



**Figure 4.** Detailed workflow of the ABSP analytic process.

-absp-



**Figure 5.** Diagram of the possible ways to launch ABSP analyses.

## 2 How to proceed with analysis using ABSP?

### 2.1 Open ABSP for the first time

#### 2.1.1 Download and installations

- Install R: <https://www.r-project.org/>
- Install RStudio: <https://www.rstudio.com/>
- Download the ABSP files on github: <https://github.com/ABSP-methylation-tool/ABSP>

#### 2.1.2 Content of the ABSP main directory

The main ABSP folder is organized as follows:

- **documents** *folder for documents available to the user*
  - List of BSgenomes.xlsx *file listing the available genomes*
  - multiple\_grouped\_parameters\_table.xlsx *table of inputs to launch multiple grouped analyses*
  - multiple\_grouped\_parameters\_table.ods *table of inputs to launch multiple grouped analyses*
  - multiple\_individual\_analyses\_table.xlsx *table of inputs to launch multiple individual analyses*
  - multiple\_individual\_analyses\_table.ods *table of inputs to launch multiple individual analyses*
- **renv** *folder for R environment and packages*
  - **library** *folder for the private project library*
  - activate.R *activation script run by the project .Rprofile*
- **reports** *folder for analysis reports*
- **results** *folder for analysis results (data, tables, graphics...)*
- **scripts** *folder for scripts and associated files required to run analysis*
  - ABSP\_functions.R *R script providing the necessary functions for ABSP*
  - ABSP\_grouped\_analysis.RMD *R markdown script of grouped analysis*
  - ABSP\_individual\_analysis.RMD *R markdown script of individual analysis*
  - custom.css *CSS script for custom theme settings of the .html report files*
  - logo.svg *ABSP logo vector image*

- **WWW** *folder for files necessary for the shiny app*
  - ABSP - Analysis.svg *diagram of ABSP analysis strategy*
  - ABSP - BSP.svg *diagram of BSP experiment principle*
  - ABSP - fasta file.png *image of a reference sequence .fasta file example*
  - ABSP - Launch analysis.svg *diagram of the different ways to launch ABSP analyses*
  - ABSP - Workflow simple.svg *diagram of the BSP workflow*
  - custom\_app.css *CSS script for custom theme settings of the app interface*
  - logo.svg *ABSP logo vector image*
- .Rprofile *file used to activate renv for new R sessions launched in the project*
- ABSP RProject.Rproj *R project file*
- ABSP User Guide.pdf *reference manual to use ABSP*
- app.R *shiny app file*
- renv.lock *lockfile describing the state of the project's library at some point in time*

For ABSP to function properly, all the aforementioned files must be downloaded and present in the ABSP main folder with the same structure.

Make sure not to rename, move or delete the provided folders and files. If you want to reorganize files or folders it is better to copy to other directories than to modify the files. However, new folders can be added to the ABSP main folder without causing issues.

### 2.1.3 Launch the app

- Open the ABSP Rproject.Rproj file with RStudio (It might take a few minutes to open).
- Open the app.R file with RStudio.
- Find the "Run App" button in the upper right corner, click on the arrow right next to it and select "Run external" (for a better display).
- Click on the "Run App" button to launch the app.
- A pop-up window should appear if the Shiny package was not already installed on your device, click on "Yes" to accept the Shiny installation.

Once this procedure is done, the package installation can start and it might take a few minutes until the app can be opened in the default web browser.

## 2.2 Open ABSP (not for the first time)

### 2.2.1 Launch the app

- Open the ABSP Rproject.Rproj file with RStudio.
- Open the app.R file with RStudio.
- Click on the "Run App" button to launch the app.

To launch the different analyses refer to the following sections below which describes the individual analysis, grouped analysis and multiples analyses ([figure 5](#)).

## 2.3 Update ABSP

Download the latest version of ABSP files on github: <https://github.com/ABSP-methylation-tool/ABSP>.

### 2.3.1 Launch the app

Use the new ABSP main directory (latest version) to open ABSP.

- Open the ABSP Rproject.Rproj file with RStudio (It might take a few minutes to open).
- Open the app.R file with RStudio.
- Click on the "Run App" button to launch the app.

You can use your previous input files, but make sure they are compatible with the new version (e.g. the *multiple\_individual\_analyses\_table* and *multiple\_grouped\_analyses\_table* files templates for multiple launch of analyses might have been changed).

As the R project and app used are the one from the new ABSP main directory (latest version), the results and reports will be saved in this directory.

## 2.4 Individual analysis

### 2.4.1 Input files requirements

**Sequencing result .ab1 files** As input, ABSP requires the chromatogram trace file (.ab1) from the sequencing run (Sanger) using the bisulfite converted DNA PCR products as templates. It is highly recommended to have both directions sequenced, by a primer on each side of the PCR product: a forward primer and a reverse primer. Although, the analysis can be run using only one trace file instead of both. The directions do not need to be specified as the analysis will determine it automatically.

**Reference DNA sequence and information .fasta file** ABSP also requires a *.fasta* file containing information about the reference DNA. A *.fasta* file is composed of a **header** and a **body**. The header must contain both the **genomic coordinates** of this sequence and the **choice of strand**, the one used for primer design, the one that will be amplified by PCR. Indeed, as both strands are no longer complementary after bisulfite conversion, the primers have to be designed on only one bisulfite converted strand as template DNA. The body must contain the **nucleotide sequence** of the reference DNA from the **plus strand** (upper/sense strand) of the genome.

Formats for the *.fasta* file header:

- The **genomic coordinates** must be written in the format *chr#:# #####-#####* (e.g. *chr16:68771087-68771462*).
- The **choice of strand amplified** must be either "*primers=plus*" or "*primers=minus*". If none of these characters strings are present in the *.fasta* file, by default the plus strand is chosen.

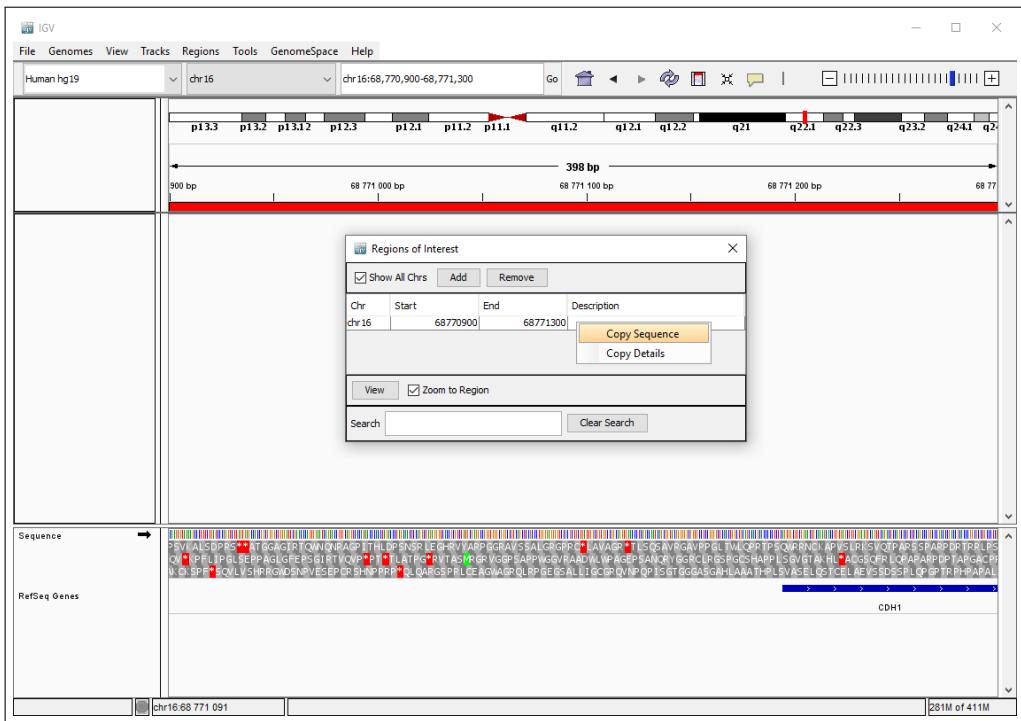
An example of a *.fasta* file content is depicted in [figure 6](#). Note that any other information in the header, such as the sequence name, for example, can be added without consequences if they do not interfere with the previously described formats.

```
> CDH1 chr16:68770900-68771299 primers=plus
CCAAGTGTAAAAGCCCTTTCTGATCCCAGGTCTTAGTGAGCCACCGCGGGGCTGGGATTGAAACCCAGTGGAAATCAGAAC
CGTGCAGGTCCCATAACCCACCTAGACCTAGCAACTCCAGGCTAGAGGGTCACCGCGTCTATGCGAGGCCGGTGGCGG
GCCGTCAGCTCCGCCCTGGGGAGGGGTCGGCGCTGCTGATTGGCTGTGGCCGGCAGGTGAACCCCTAGCCAATCAGCGTA
CGGGGGGGCGGTGCCCTCCGGGGCTCACCTGGCTGCAGCCACGCACCCCCCTCTCAGTGGCGTCGGAACTGCAAAGCACCTGTG
AGCTTGCAGAAGTCAGTTCAGACTCCAGCCCGCTCCAGCCCCGGCCGACCCGACCGCACCCGGCGCCTGCCCTCGC
```

**Figure 6.** Example of *.fasta* file for the reference DNA input required for ABSP individual analysis.

As a recommendation, to help with the creation of the *.fasta* file, the [IGV \(Integrative Genomics Viewer\)](#)<sup>4</sup> software can be easily used to navigate the genome and to get the nucleotide sequence of a specific region, alongside with its genomic coordinates. After navigating on the genome, the current viewed region can be added as a region of interest by selecting the "*Regions*" > "*Region Navigator*" > "*Add*" ([figure 7](#)). In this "*Regions of Interest*" panel, the added region will appear in the list of regions, its coordinates can be adjusted and it can be annotated with a description. By right-clicking on a region from this list, a context menu appears and two options can be selected, "*Copy Sequence*" or "*Copy Details*". The first one copies to the clipboard the nucleotide sequence of the region, and the second one copies the genomic coordinates (in the correct format for the *.fasta* file) as well as the description associated with the region. Make sure to properly verify the first and last nucleotides, as there can be a one nucleotide difference between the coordinates and the actual sequence.

<sup>4</sup><https://software.broadinstitute.org/software/igv/>



**Figure 7.** IGV (Integrative Genomics Viewer) software window.

#### 2.4.2 Procedure

In the **individual analysis tab**, the left panel is to launch analysis, and the right panel provides entries information.

##### Experiment information

- **Select an existing folder or enter a new folder name.** Located in the *AB-SP/results* folder, all of the analysis results will be saved in this folder. Having different folders of results can be used to separate the different analyses by projects, experiments, or users. Note that the six first letters of the folder name will appear in the report file name.
- **Select an existing sequence folder or enter a new sequence name.** The sequence folder is located in the previously selected folder. All of the analysis results will be saved in this folder corresponding to the same sequence. Note that this sequence name will be used in output files (tables, plots...) to refer to the sequence.
- **Enter the collection name.** The collection corresponds to a separation of samples above groups (*details in section 1.4 How does ABSP work? at page 7*). Make sure the collection name is strictly identical for all samples of the same collection.
- **Enter the group name.** The group corresponds to the condition to compare (*details in section 1.4 How does ABSP work? at page 7*). Make sure the group

name is strictly identical for all samples of the same group.

- **For direct-BSP only: Enter replicate number.** In the case of direct sequencing of PCR products only, the replicate number corresponds to the repetition identifier of the sample (*details in section 1.4 How does ABSP work? at page 7*).
- **For cloning-BSP only: Enter clone number.** In the case of clone sequencing only, the clone number corresponds to the identifier number of each clone from the same condition (*details in section 1.4 How does ABSP work? at page 7*).

#### Reference DNA sequence

- **Select the reference genome.** It will only be used to display the genomic sequence in the genomic plot. Make sure to click on the "Pre-install genome" button if the selected genome is used for the first time. Only a short list of genomes is displayed in the drop-down list but more genome assemblies are available. Go to section 3.3 *Code modifications at page 37* to get information on how to add another genome in the drop-down list. The complete list of available genomes can be found in the "List of BSgenomes.xlsx" file in the "ABSP/documents" folder.
- **Select .fasta file of reference DNA sequence.** As described above in the Input files requirements section, the .fasta file of the reference DNA sequence needs to be selected from your folders.

#### 2.4.3 Output report

The HTML report file of the analysis is automatically saved in the *reports* folder in the ABSP directory.

**Header** First, in the top panel, the information about the sample experimental conditions is displayed in a table.

- Folder name
- Sequence name
- Collection
- Group
- Replicate number (for direct-BSP only)
- Cloning number (for cloning-BSP only)
- Date of sequencing #1
- Date of sequencing #2
- Date of analysis
- Prefix of output files

**Reference DNA** This tab summarizes all the data computed from the reference DNA sequence .fasta file.

- **Reference DNA sequence** General information on the sequence (name, strand used for primer design, length, and genomic coordinates) and genomic sequence from plus strand (given by the reference DNA .fasta file) and minus strand (reverse complement).
- **Localization of CG dinucleotides** Detection of CpG sites on the reference DNA (on plus and minus strand) with attribution of the CG number, from 1 to n on the plus strand.
- **Bisulfite converted sequences** Sequences of reference DNA after theoretical bisulfite conversion (CpG sites considered as methylated). The bisulfite conversion is performed on the strand used for PCR primer design, as only this strand is amplified during PCR. The PCR regenerates the opposite strand, corresponding to the reverse complement of the bisulfite converted DNA template.

**Sequencing trimming** This tab summarizes the trimming of sequencing reads based on quality. Two parameters are used: the Phred quality score of each base retrieved from the sequencing file, and the mixed base peak ratio.

- **Summary** In the first tab, the default thresholds used to trim the sequencing results are displayed.
  - Minimum length of the trimmed sequence (default is 30 bp)
  - Minimum Phred quality score (default is 30, corresponding to a base-calling error probability of 0.001%)
  - Minimum ratio of primary peak (default is 0.75)
  - Minimum percentage of non-mixed positions (default is 75%)
- Below the threshold table, the trimming summary for both sequencing reads is displayed and indicates whether or not the trimming was successful (correct trimmed sequence quality) or failed (incorrect trimmed sequence quality).
- **Details per sequencing**
  - **Raw sequence** The sequence, chromatogram, and data table of the sequencing results are displayed.
  - **Quality report** The first trimming is based on the base-calling quality as it uses the Phred quality scores of each base to find the best sequence to trim. This step is provided by the [SangeranalyseR package<sup>5,6</sup>](#). The thresholds and the results of this quality trimming are displayed.
  - **Mixed base peak report** The second trimming is based on the primary peak ratio over the other peaks for each position. At each position, the signal ratio of the primary peak is computed using the peak height

---

<sup>5</sup><https://sangeranalyser.readthedocs.io/en/latest/index.html>

<sup>6</sup>Kuan-Hao Chao et al. "sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R". in: *bioRxiv* (2020), p. 2020.05.18.102459. doi: [10.1101/2020.05.18.102459](https://doi.org/10.1101/2020.05.18.102459).

values for each base, with formula: if  $peak_C > \{peak_A, peak_T, peak_G\}$

$$\text{Primary peak ratio} = \frac{peak_C}{peak_A + peak_T + peak_G + peak_C}$$

If the ratio is above the threshold (default is 0.75), the position is considered non-mixed; if the ratio is below the threshold, the position is considered mixed. All the possible trimmed sequences are obtained by selecting the sequence between n (from 3 to 15) consecutive non-mixed positions. For each one of the possible trimmed sequences, the percentage of non-mixed positions is calculated. Among those, the trimmed sequence which is selected corresponds to the one with a percentage of non-mixed positions above the threshold (default is 75%) with the minimum of consecutive non-mixed positions at extremities (this number is displayed).

- **Trimming plot** The two previous report steps give two different trimmed sequences that can be viewed on the trimming plot ([figure 8](#)). The top panel represents a dot plot of the Phred quality score per position, values in green are above the threshold and values in red below, in which the start and end positions of the trimmed sequence are represented by orange vertical lines. The second panel is also a dot plot but it represents the primary peak ratio per position, values in green are above the threshold (considered as non-mixed) and values in red below (considered as mixed), in which the start and end positions of the trimmed sequence are represented by cyan vertical lines. In the last panel, the two trimmed sequences are represented in the same color, orange, and cyan. The raw sequence is displayed in red. The overlapping of the two previous trimmed sequences gives the final trimmed sequence, in green, which is the one kept for the rest of the analysis and corresponds to the information given in the summary tab and the final trimmed sequence tab.
- **Final trimmed sequence** The sequence, chromatogram, and data table of the final trimmed sequencing results are displayed.

Quality score and primary base peak ratio plots

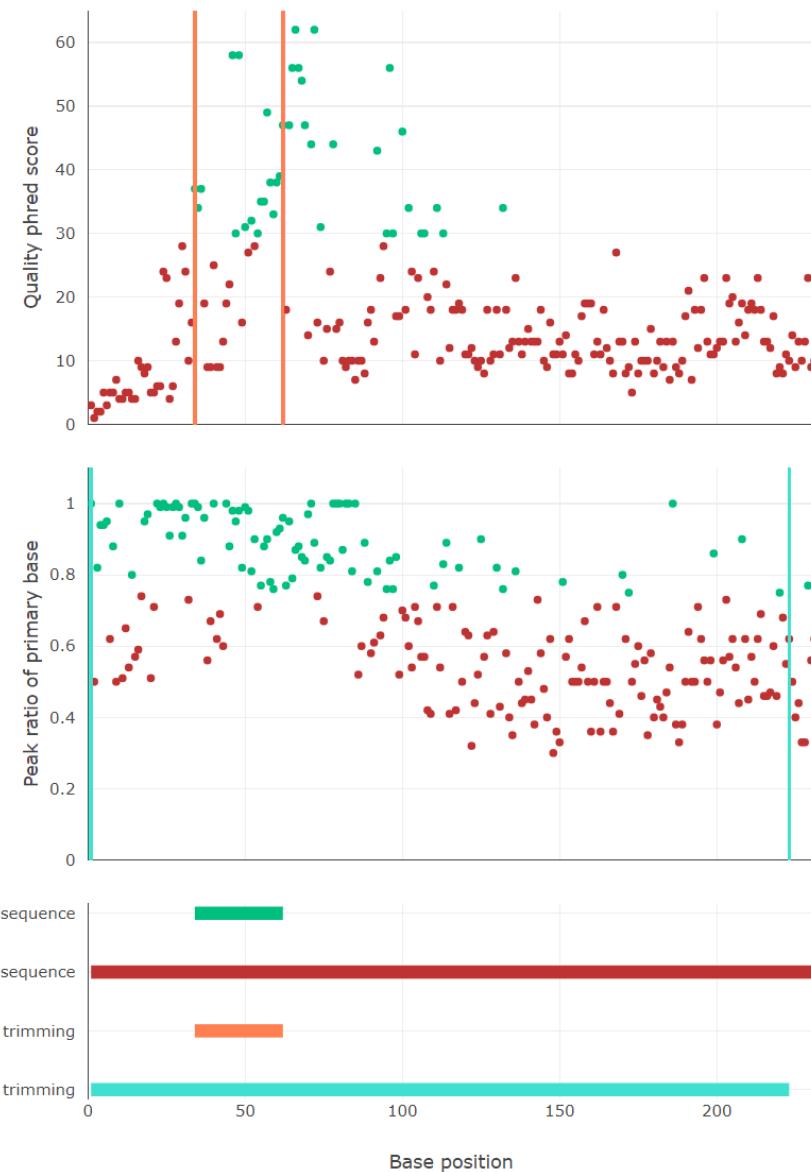


Figure 8. Example of trimming plot from the output report.

**Alignments** In the first two tabs, each sequencing read is aligned with either the sense sequence (bisulfite converted sequence from the template strand) as if it is a forward sequencing, or the antisense sequence (reverse complement of bisulfite converted sequence from the template strand) as if it is a reverse sequencing. The direction of each sequencing result is determined based on the aligned sequence length: the alignment which gives the longest aligned sequence is considered the correct one. If for one sequencing read the aligned sequences are equal between alignment as forward and as reverse, the direction determination depends on the other one. The two last tabs display the correct alignments and aligned sequences.

**Quality control** This tab summarizes the quality of the trimmed sequencing result aligned with the reference DNA.

- **Summary** In the first tab, the default thresholds are used to control the quality of the aligned sequencing result.

- Minimum length of the aligned sequence (default is 30 bp)
- Minimum identity percentage of alignment (default is 75%)
- Minimum of bisulfite conversion rate mean (default is 0.90, corresponding to 90% conversion efficacy)

Below the threshold table, the quality summaries for both sequencing results are displayed and indicate whether or not the aligned sequencing results have a sufficient quality (correct) or insufficient (incorrect) relative to thresholds.

- **Mismatch positions** For both sequencing reads, a table indicates the mismatched positions and nucleotides, on both the sequencing read and reference DNA sequence.
- **Insertions/deletions** For both sequencing reads, a table indicates the insertions/deletions (gaps) found in either the sequencing read or the reference DNA sequence.
- **Conversion rates** For both sequencing reads, a table indicates the bisulfite conversion rate for each cytosine outside a CpG in the aligned sequence. The first column corresponds to the identifier number of the cytosine on the reference DNA sequence, in the second one is displayed its position on the reference DNA sequence, in the third one its position on the trimmed sequencing result, and in the fourth its position on the raw sequencing result. The position matching is obtained thanks to the alignment of sequences. For each position on the raw sequencing result, the peak height values of signals are extracted and used to compute the conversion rate, with the following formulas (for forward and reverse sequencing respectively):

$$\text{Bisulfite conversion rate} = \frac{\text{peak}_T}{\text{peak}_C + \text{peak}_T}$$

$$\text{Bisulfite conversion rate} = \frac{\text{peak}_A}{\text{peak}_G + \text{peak}_A}$$

For each position a bisulfite conversion is obtained. The mean of rates from

all positions is indicated in the summary table in the first tab, as well as standard deviation.

- **Maximum aligned sequence** The maximum aligned sequence corresponds to the sequence covered by at least one of the two sequencing results. Its information such as its length, its coordinates, and its nucleotide sequence are displayed.

### Methylation

- For both sequencing reads, the tables of computed **methylation percentages** are displayed. The first column corresponds to the CpG site identifier number on the reference DNA sequence (list of all CpG sites in the *Reference DNA* tab, *Localization of CG dinucleotides* tab). In the next three columns, its coordinates are specified. Then, the position of the methylated cytosine (the C in forward, the G in reverse) is displayed, as well as its position in the raw and trimmed sequencing results. The position matching is obtained thanks to the alignment of sequences. For each position on the raw sequencing result, the peak height values of signals are extracted and used to compute the methylation percentage, with the following formulas, for forward and reverse sequencing respectively:

$$\text{Methylation percentage} = \frac{\text{peak}_C}{\text{peak}_C + \text{peak}_T} \times 100$$

$$\text{Methylation percentage} = \frac{\text{peak}_G}{\text{peak}_G + \text{peak}_A} \times 100$$

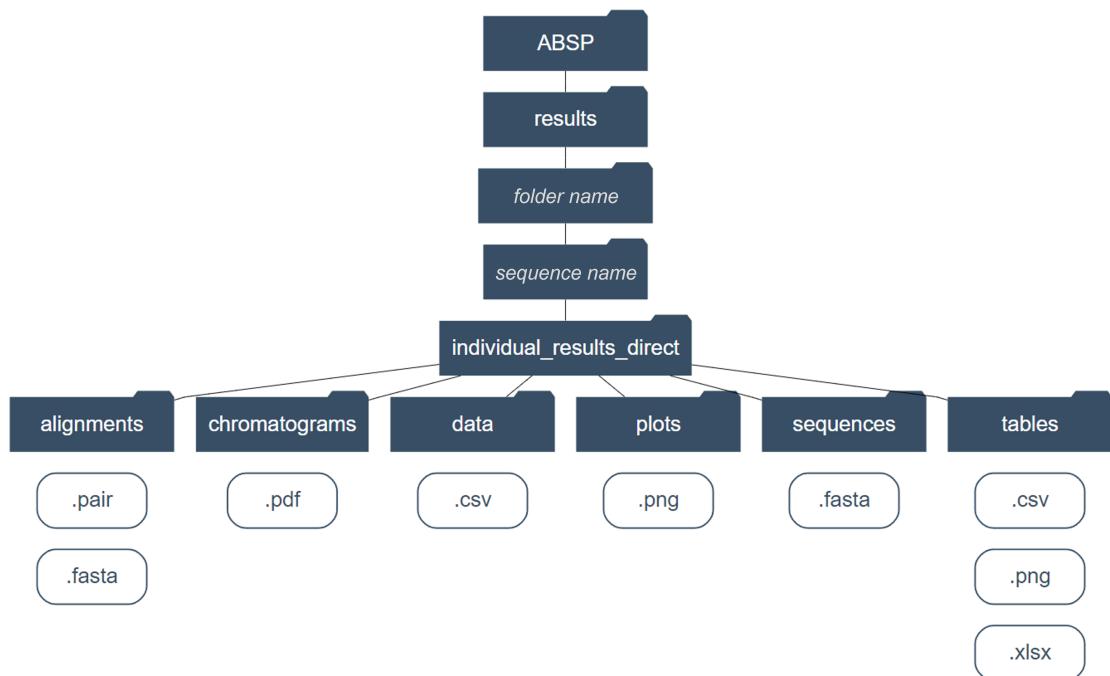
- **Combined** Methylation results from both sequencing results are then combined in a unique table with the calculation of the average methylation and standard deviation per position.
- **Individual methylation plot** Finally, a plot is generated to visualize results relative to the genomic sequence.

### Output data

- **Directories** A diagram of output files directories is displayed ([figure 9](#)).
- **Files** A list of all the output files with links to local folders is displayed.
- **Methylation data file preview** The output methylation data file which will be used as input for the grouped analysis is displayed as a table (The "alg\_coord\_start" and "alg\_coord\_end" columns contain a unique value corresponding to the start and end coordinates of the maximum aligned sequence for the individual analysis).

#### 2.4.4 Output files

All the output files are located in the *results* directory, as depicted by the diagram in [figure 9](#).



**Figure 9.** Diagram of output directories to locate output files from the individual analysis. The “*folder name*” and “*sequence name*” depend on the input entries when launching the analysis.

As for the same sequence (same primer set) results from both direct-BSP and cloning-BSP can be generated, the two types of outputs are separated into two subfolders: “*individual\_results\_direct*” and “*individual\_results\_cloning*”.

- **alignments** In subfolders specific to each individual analysis, the .pair files of the alignments and the .fasta files of the aligned sequence are saved.
- **chromatograms** In subfolders specific to each individual analysis, the chromatograms of raw sequencing reads and the chromatograms of the trimmed sequencing are saved as .pdf files.
- **data** It contains all the methylation data .csv files for the grouped analysis.
- **plots** It contains all the individual methylation plots as well as legends for plots as .png image files.
- **sequences** The reference DNA sequences from plus and minus strands and the bisulfite converted sequence of the template strand are saved as .fasta files.
- **tables** In subfolders specific to each individual analysis, the data tables of raw sequencing results, the data tables of trimmed sequencing results, the summary table of sequencing trimming, the summary table of quality control, the bisulfite conversion rates table and the methylation percentages tables

are saved as .csv .png and .xlsx files.

## 2.5 Grouped analysis

### 2.5.1 Procedure

In the **grouped analysis tab**, the left panel is to launch analysis, and the right panel provides entries information.

#### Experiment information

- **Select an existing folder.** Located in the "ABSP/results" folder, all of the analysis results will be saved in this folder. Having different folders of results can be used to separate the different analyses by projects, experiments, or users. Note that the six first letters of the folder name will appear in the report file name.
- **Select an existing sequence folder.** Located in the "ABSP/results/previous folder" folder, all of the analysis results will be saved in this folder for this sequence. This sequence name will be used in output files (tables, plots...) to refer to the sequence.
- **Select the reference genome.** It will only be used to display the genomic sequence in the genomic plot. See section 3.3 Code modifications at page 37 to add another genome in the drop-down list. The complete list of available genomes can be found in the "List of BSgenomes.xlsx" file in the "ABSP/documents" folder.
- **Select the experiment type.** The choice of the experiment can be either Direct-BSP or Cloning-BSP. The correct experiment type entry is essential to retrieve the methylation data files either in the "individual\_results\_direct" or "individual\_results\_cloning" folders.

#### Plot parameters

- **Select position labels for plots.** The CpG positions on plots can be referred to by different label types:
  - The **CpG coordinates** label type displays the genomic coordinates of the CpG site in the format *chr#:#####-#####* (e.g. "chr16:68771230-68771231").
  - The **CpG numbers** label type displays the CpG site identifier number on the represented sequence, from 1 to n.
  - The **None** label type displays blank labels, which can be a suitable alternative in case of extremely close CpG positions as labels may overlap.
- **Choose to separate plots by collections.** For lollipop plots, genomic plots, and boxplots and only for display purposes (it does not affect data). If this parameter is not ticked (default), all samples from different collections will be displayed on the same plot. If this parameter is ticked, a plot is generated

by collection, displaying samples from one collection only.

- **Indicate the order of groups for display.** For this input to work, the folder, sequence, and experiment type have to be selected and correct. In this case, the group names are extracted from methylation data files corresponding to the previously selected entries. The groups must all be selected, in the desired order for display.
- **Select the types of sample ordering for plots.** Four different sample ordering (ordinate axis ordering) are available for visualization plots, each provides a specific way of ordering samples on the ordinate axis based on different parameters. At least one must be selected, up to the four of them.
  - **As it is** arranges samples by alphabetic order of collections. If none or one collection is present, this order is equivalent to the *By groups* one.
  - **By groups** arranges samples by the provided group order above.
  - **By methylation levels** arranges samples depending on their methylation mean.
  - **By clusters** arranges samples depending on the hierarchical clustering calculated and represented by an associated dendrogram.

### 2.5.2 Output report

The HTML report *.html* file of the analysis is automatically saved in the "reports" folder in the ABSP directory.

**Header** First, in the top panel, the information about the sample experimental conditions are displayed in a table.

- Folder name
- Sequence name
- Reference genome
- Type of experiment
- Group order
- Date of analysis

**Files content** This tab summarizes all the data that have been used for this analysis.

- **Data files content** General information about the data is listed: sequence name, collections, groups and replicates or clones.
- **Data files paths** Paths of the methylation files that were found and used for the analysis are listed.

**Methylation data** This tab regroups the methylation data in tables from the retrieved files.

- **Methylation data of replicates/clones** For each replicate or clone, depending on the experiment type, the methylation percentages of each CpG sites

are displayed in a table, with the mean of all positions and the associated standard deviation by replicate/clone.

For clones: Methylation percentages calculated from sequencing results are converted to 0% or 100% methylation status.

- A CpG site is considered unmethylated (0%) when methylation percentage is between 0% and the defined threshold (default is 20%).
- A CpG site is considered methylated (100%) when methylation percentage is between the defined threshold (default is 80%) and 100%.
- A CpG site partially methylated, with methylation percentage between 20% and 80%, is removed and annotated as NA (Not Available).
- For one clone, if more than a threshold percentage (default is 20%) of CpG sites are partially methylated, the clone is considered as defective and all of its CpG sites are annotated as NA (Not Available).

The thresholds can be modified in the script, please refer to section 3.3 Code modifications at page 37.

- **Methylation data of groups** For each group, the mean of methylation percentages of each CpG sites are displayed in a table, with the mean of all positions and the associated standard deviation by groups.

**Plots of replicates** Only for direct-BSP. This tab provides plots to visualize the methylation data of replicates.

Lollipop plots (condensed and proportional), genomic plot and cluster dendrogram plot are as illustrated in [figure 10](#).

**Plots of clones** Only for cloning-BSP. This tab provides plots to visualize the methylation data of clones from each sample in separated plots.

Lollipop plots (condensed and proportional), genomic plot and cluster dendrogram plot are illustrated in [figure 11](#).

**Plots of groups** This tab provides plots to visualize the methylation data of groups per collection. For each sample, the mean of replicates or clones is calculated per CpG and results are displayed in this tab.

Lollipop plots (condensed and proportional), genomic plot and cluster dendrogram plot are illustrated in [figure 12](#).

**Statistics** This tab aims to compare methylation between groups, to find statistical significant differences, either by comparing CpG site per CpG site, or by comparing the all region.

- **Descriptive statistics of groups** Two tables of methylation data descriptive statistics are displayed: one for methylation data by CpG positions, and one for methylation data of all CpG positions for each sample.
- **Student's T test** Two tables display T tests p-values between groups 2 by 2, one table for methylation data by CpG positions, and one for methylation data of all CpG positions for each sample.

- **Boxplots** Boxplots of methylation data with T tests p-values as numbers or symbols are generated, one plot details data of each CpG position ([figure 13 A](#) for direct-BSP data and [figure 13 D](#) for cloning-BSP), and another one data from methylation mean of the region ([figure 13 B](#) for direct-BSP data and [figure 13 E](#) for cloning-BSP).
- **Methylation profile plots** Methylation profile plots with Kruskal-Wallis tests p-values are generated for each collection ([figure 13 C and F](#)).

The [Methylation plotter](#)<sup>7,8</sup> tool served as a base to develop the different types of plots with different sample ordering.

#### Output data

- **Directories** A diagram of output files directories is displayed ([figure 14](#)).
- **Files** A list of all the output files with links to local folders is displayed.

#### 2.5.3 Output files

All the output files are located in the *results* directory, as depicted by the diagram in [figure 14](#).

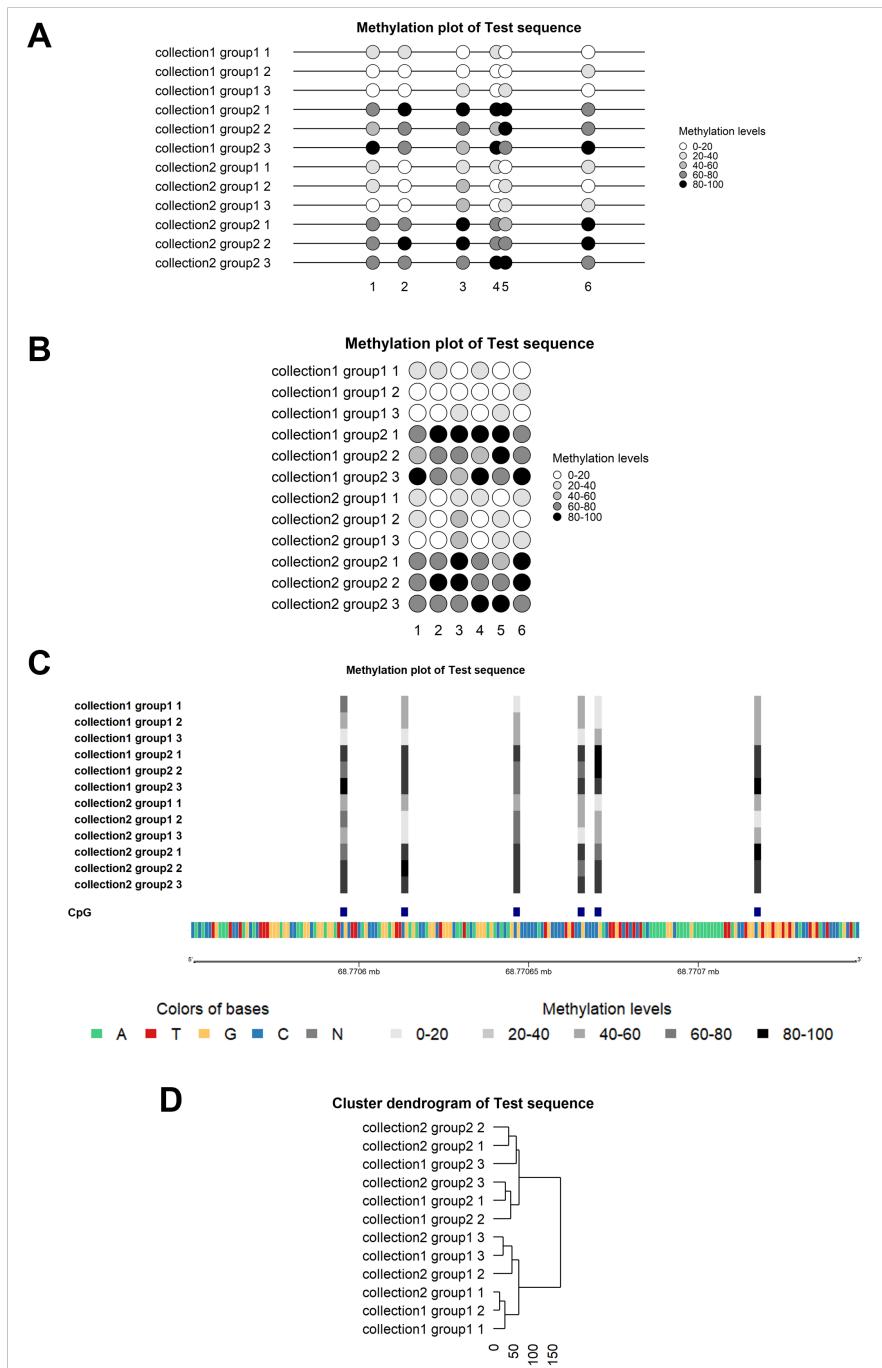
As for the same sequence results from both direct-BSP and cloning-BSP can be generated, the two types of outputs are separated into two subfolders: "*grouped\_results\_direct*" and "*grouped\_results\_cloning*".

- **boxplots** Boxplots are saved as *.png* files.
- **dendro\_plots** Cluster dendograms are saved as *.png* files.
- **genomic\_plots** Genomic plots are saved as *.png* files, in subfolders for plots of replicates/clones and plots of groups.
- **lollipop\_plots** Lollipop plots are saved as *.png* files, in subfolders for plots of replicates/clones and plots of groups.
- **meth\_profile\_plots** Methylation profile plots are saved as *.png* files
- **tables** The methylation data tables, the descriptive statistics tables of positions or means (means of all positions), and the Student's T test tables are saved as *.csv* and *.xlsx* files.

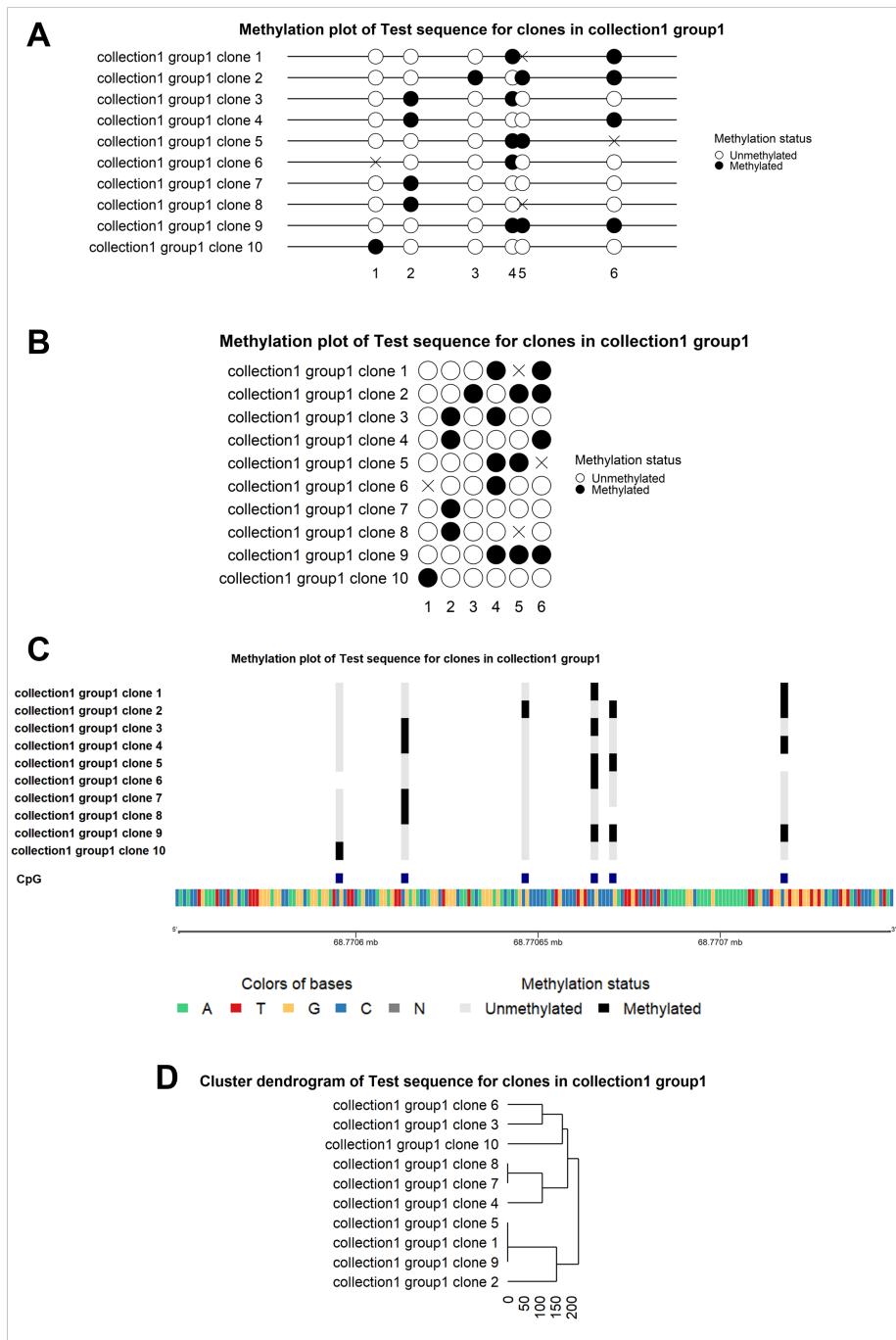
---

<sup>7</sup>[http://maplab.imppc.org/methylation\\_plotter/index.html](http://maplab.imppc.org/methylation_plotter/index.html)

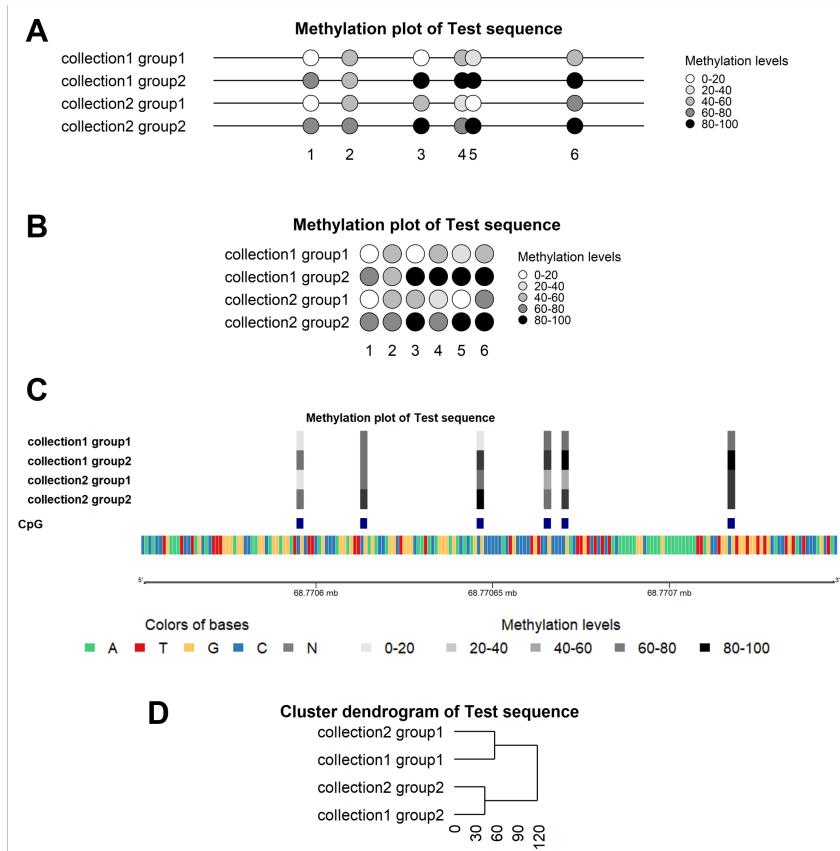
<sup>8</sup>Izaskun Mallona, Anna Díez-Villanueva, and Miguel A Peinado. "Methylation plotter: a web tool for dynamic visualization of DNA methylation data". In: *Source Code for Biology and Medicine* 9.1 (2014). Methylation plotter, p. 11. doi: [10.1186/1751-0473-9-11](https://doi.org/10.1186/1751-0473-9-11).



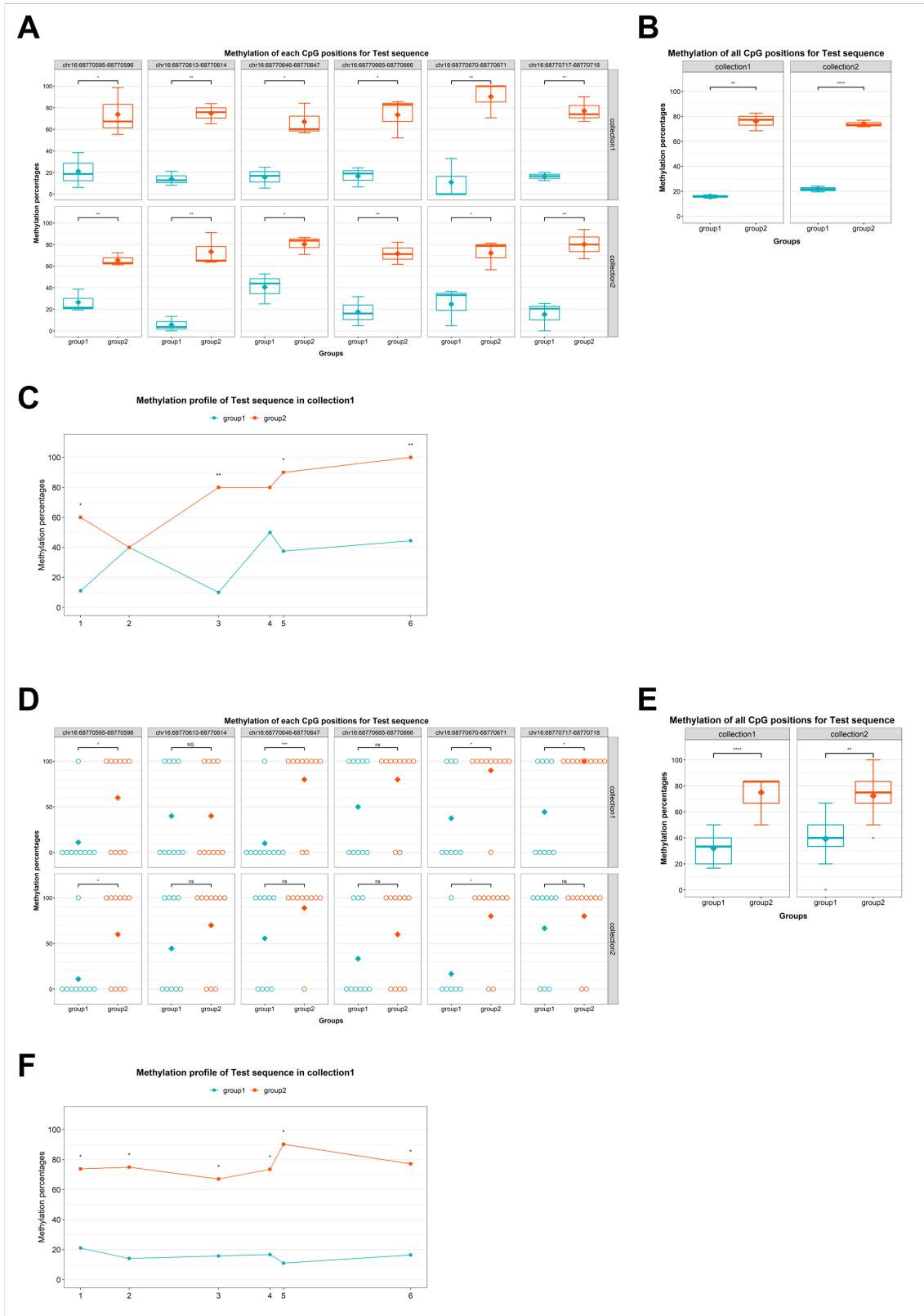
**Figure 10.** Visualization plots of all replicates (direct-BSP). The plots were generated based on mock methylation data for a test sequence. Three replicates (1, 2, and 3) per sample are represented on the plots, from two groups (group1 and group2) and two collections (collection1 and collection2) (sample ordering *as it is*). The methylation levels are given as percentages and corresponds directly to the methylation output data from the individual analyses. **A.** Proportional lollipop plot. **B.** Condensed lollipop plot. **C.** Genomic plot. **D.** Cluster dendrogram (for sample ordering *by clusters*).



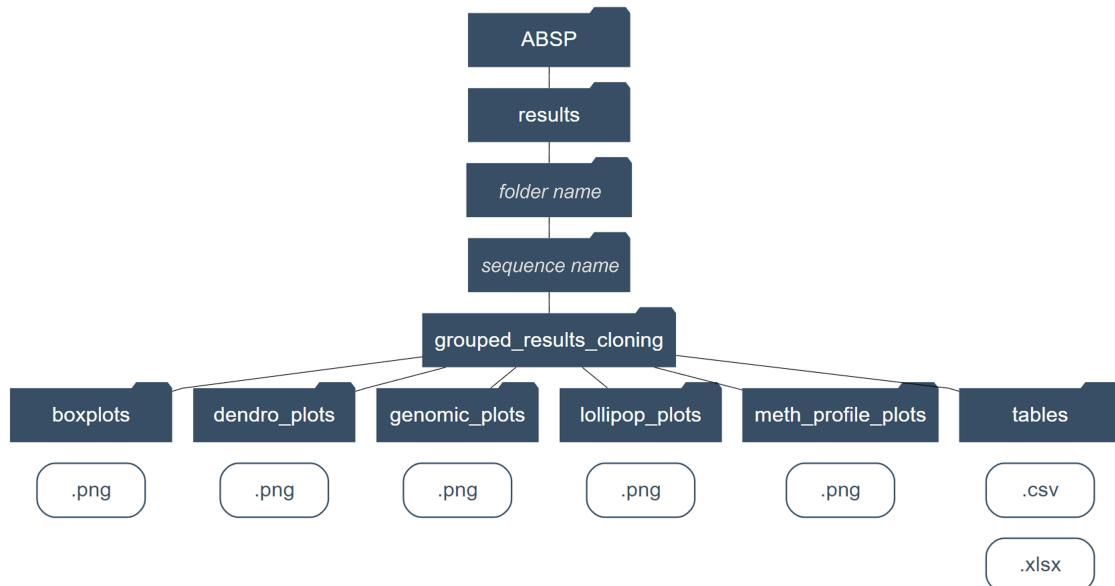
**Figure 11.** Visualization plots of all clones from one sample (cloning-BSP). The plots were generated based on mock methylation data for a test sequence. Ten clones (from 1 to 10) for the sample "*collection1 group1*" are represented on the plots (sample ordering *as it is*). The methylation status corresponds to the conversion of methylation percentages from the individual analyses into unmethylated (0%) or methylated (100%) (or not available) methylation statuses. **A.** Proportional lollipop plot. **B.** Condensed lollipop plot. **C.** Genomic plot. **D.** Cluster dendrogram (for sample ordering *by clusters*).



**Figure 12.** Visualization plots of groups (means of replicates/ clones per samples). The plots were generated based on mock methylation data for a test sequence. Plots represents the means of methylation percentages of ten clones per sample and per CpG position (sample ordering as *it is*). **A.** Proportional lollipop plot. **B.** Condensed lollipop plot. **C.** Genomic plot. **D.** Cluster dendrogram (for sample ordering by clusters).



**Figure 13.** Boxplots and methylation profile plots. The plots were generated based on mock methylation data for a test sequence. **A.** Boxplots of direct-BSP methylation results for each CpG site. **B.** Boxplots of direct-BSP methylation results for means of all CpG sites. **C.** Methylation profile plot of direct-BSP methylation results for one of the two collections. **D.** Boxplots of cloning-BSP methylation results for each CpG site. As methylation levels of CpG from clones can only be either 0% or 100% thereby boxes can't be drawn, instead, each clone is represented in the plot by a circle. **E.** Boxplots of cloning-BSP methylation results for means of all CpG sites. **F.** Methylation profile plot of cloning-BSP methylation results for one of the two collections. In boxplots, symbols represents significance levels of Student's T test p-values (\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ , \*\*\*\*  $p \leq 0.0001$ ). In methylation profile plots, symbols represents significance levels of Kruskal-Wallis test p-values (\*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ , \*\*\*\*  $p \leq 0.0001$ ).



**Figure 14.** Diagram of output directories to locate output files from the individual analysis. The "folder name" and "sequence name" depend on the input entries when launching the analysis.

## 2.6 Multiple analyses

As described in the [figure 5](#), multiple analyses can be launched at the same time using data tables as input with all the required information. Both individual and grouped analyses can be launched, at the same time or separately.

### 2.6.1 Input files requirements

Two files are provided in the "documents" folder. They must be filled with the desired input entries, all information about how to fill the documents are also indicated within the documents as notes.

The documents are provided in the .xlsx (Microsoft Excel Open XML Format Spreadsheet) and .ods (OpenDocument Spreadsheet) formats, but the input file format must be either .xlsx or .csv (Comma-Separated Values) (.ods files must be converted to one of those formats).

- **Experiments data table for individual analyses:**

[\*"multiple\\_individual\\_analyses\\_table.xlsx"\*](#)

- SEQUENCE NAME. The sequence name should be unique and consistent for each amplicon. It must not contain any special character.
- COLLECTION. The collection refers to a separation of samples above the groups/conditions. Leave empty if you do not want to indicate a collection.
- GROUP. The group refers to the condition that is to be compared. It must not contain any special character.
- REPLICATE NUMBER. The replicate number refers to the experiment repetition identifier. It must be an integer  $\geq 1$ . Only for direct-BSP experiments, leave empty otherwise.
- CLONE NUMBER. The clone number refers to the individual clone identifier. It must be an integer  $\geq 1$ . Only for cloning-BSP experiments, leave empty otherwise.
- GENOME. The genome refers to the reference genome assembly used for coordinates and for plots displaying the genomic sequence. Only a short list of genomes are displayed in the cells, but you can use another available genome. Please refer to the provided document: "*List of BSgenomes*" to get the list of available genomes assemblies and the correct spelling.
- PATH TO FASTA FILE OF REFERENCE DNA. Path to the .fasta file of the reference DNA sequence on your computer. On Windows, you can copy the file path by holding down shift then right-clicking on the file and selecting "Copy as path" in the menu. On MacOS, you can copy the file path by right-clicking on the file to display the menu then holding down the option key and selecting "Copy ... as Pathname".
- DATE SEQUENCING #1. Date of the sequencing result #1 for traceability. The date format must be YYYY-MM-DD and the cell format must be set to "date".
- PATH TO SEQUENCING FILE #1. Path to sequencing result #1 .ab1 file.
- DATE OF SEQUENCING #2. Date of the sequencing result #2 for traceability.

The date format must be YYYY-MM-DD and the cell format must be set to "date".

- PATH TO SEQUENCING FILE #2. Path to sequencing result #2 .ab1 file.

- **Parameters table for grouped analyses:**

*"multiple\_grouped\_analyses\_table.xlsx"*.

- SEQUENCE NAME. The sequence name should be unique and consistent for each amplicon. It must not contain any special character.
- GENOME. The genome refers to the reference genome assembly used for coordinates and for plots displaying the genomic sequence. Only a short list of genomes are displayed in the cells, but you can use another available genome. Please refer to the provided document: "List of BSgenomes" to get the list of available genomes assemblies and the correct spelling.
- EXPERIMENT TYPE. The experiment type can be either Direct-BSP or Cloning-BSP.
- CpG POSITION LABEL TYPE. The CpG position label type refers to the displayed element corresponding to the CpG position on plots.
- SEPARATION OF COLLECTIONS. This parameter refers to the generation of plots: if FALSE, all samples from all collections will be displayed on the same plot and if TRUE, samples will be split in different plots, one plot per collection.
- LIST OF GROUPS ORDERED. The cell must contain all of the groups within the experiment in the order you want them to be displayed. Type your groups separated by commas. It must not contain any special character.
- TYPE OF SAMPLE ORDERING. The type of sample ordering refers to the parameter used for ordering samples on plots. Four types of ordering are available: As it is, By groups, By methylation levels and By clusters. Multiple ordering can be chosen by typing the ordering names separated by commas.

### 2.6.2 Procedure

- **Select an existing folder** within the ABSP results folder to locate all of the analyses results. To create a new folder, select the "*Create new folder*" entry and enter the name of the new folder in the text input. Note that the six first letters will appear in the report file name.
- **Select the filled table as input.** Both the experiments data table and the grouped parameters table can be provided at the same time to launch individual analyses followed by grouped analyses, or only one of the two tables can be provided and will launch the corresponding analyses, either individual analyses or grouped analyses.
- **Launch the analyses** by clicking on the bottom button "*Run analyses*".

### 2.6.3 Output files

The reports and output files of analyses are saved in the *reports* and *results* folders, in the same way as for manual launch of individual and grouped analyses, see

sections 2.4 Individual analysis at page 14 and 2.5 Grouped analysis at page 24.

## 3 Complementary information

### 3.1 Some recommendations for the BSP experiment

The length of the PCR products should be exceed 350-400 bp, as the bisulfite treatment causes DNA strand breakages long amplicon can not be properly amplified, and ABSP plot display for sequences above 400 bp with numerous CpG is not optimal.

Several tools can be used to design BSP primers, several are listed in the [Methtools](#)<sup>9</sup> list of tools, such as [Methprimer](#)<sup>10</sup> to design primers and [BiSearch](#)<sup>11</sup> to check for unintended PCR products, on bisulfite treated DNA.

### 3.2 Detailed workflow of ABSP individual analysis

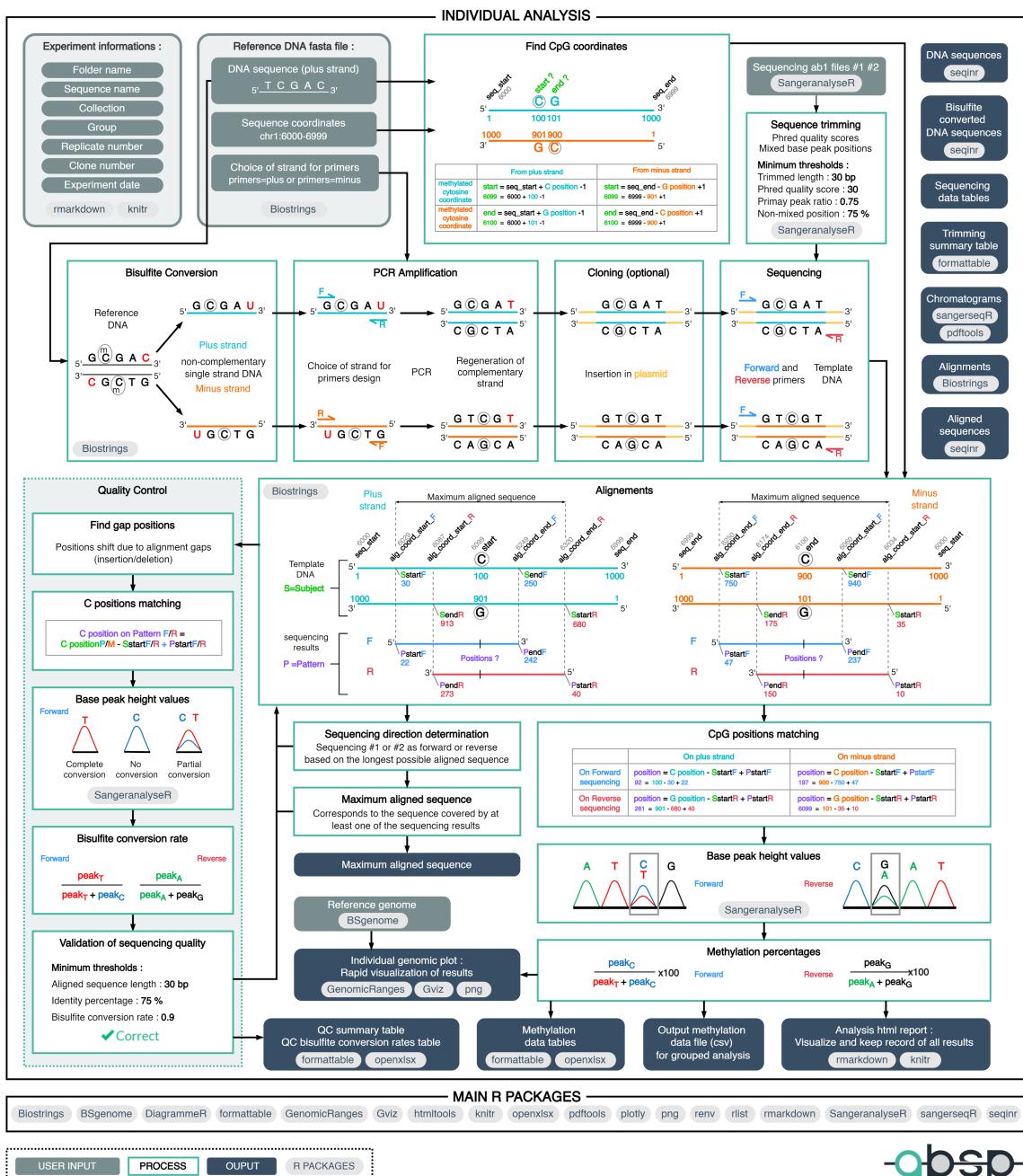
In [figure 15](#), the individual analysis input, steps and outputs are described with more details.

---

<sup>9</sup><http://bigd.big.ac.cn/methbank/methTool/list/>

<sup>10</sup><https://www.urogene.org/methprimer/>

<sup>11</sup><http://bisearch.enzim.hu/>



**Figure 15.** Detailed workflow of the individual analysis.

### 3.3 Code modifications

#### 3.3.1 List of reference genomes

In the individual analysis and grouped analysis tabs of the app, the drop-down list to select the reference genome is limited. If your reference genome does not appear it doesn't mean that it is not available, and can manually added. The complete list of [genomes assemblies](#)<sup>12</sup> with the correct spelling can be found in the "List of BSgenomes.xlsx" file in the "ABSP/documents" folder.

To modify the displayed drop-down list items, the *app.R* script have to be modify. In the [listing 1](#), the "list\_genomes" object corresponds to the vector of character strings listing the displayed reference genome. Any of the other reference genome can be added to the list.

[Listing 1](#). List of reference genomes displayed in the drop-down lists as selectable inputs for analyses. From the "*app.R*" script.

```
12 # ****
13
14 # Here, to simplify, a short list of genomes is displayed but all BSgenome
15 # can be used
16 # To get the list of all genomes of BSgenome package run : 'BSgenome::'
17 # available.genomes()', more information on genomes at https://genome.ucsc.edu/cgi-bin/hgGateway
18 # A new genome can be added to the list displayed just below :
19 list_genomes <- c(
20   "BSgenome.Hsapiens.UCSC.hg19", "BSgenome.Hsapiens.UCSC.hg38", "
21   BSgenome.Mmusculus.UCSC.mm10", "BSgenome.Mmusculus.UCSC.mm39", "
22   "BSgenome.Rnorvegicus.UCSC.rn6", "BSgenome.Rnorvegicus.UCSC.rn7", "
23   BSgenome.Cfamiliaris.UCSC.canFam3", "BSgenome.Mmulatta.UCSC.rheMac8",
24   "BSgenome.Ggallus.UCSC.galGal6", "BSgenome.Drerio.UCSC.danRer11", "
25   BSgenome.Celegans.UCSC.ce11", "BSgenome.Dmelanogaster.UCSC.dm6")
```

To modify the genome name selected by default in the drop-down "Select genome" list of either the individual or grouped analysis tabs, change the name of the genome in the following lines ([listing 2](#) and [listing 3](#)):

[Listing 2](#). Genome name selected by default in the drop-down-list of the individual analysis tab. From the "*app.R*" script.

```
225 selectInput("genomeI", label = "Select genome", choices=list_genomes, selected
25   ="BSgenome.Hsapiens.UCSC.hg19")
```

<sup>12</sup><https://genome.ucsc.edu/cgi-bin/hgGateway>

**Listing 3.** Genome name selected by default in the drop-down-list of the grouped analysis tab. From the "app.R" script.

```
497 selectInput("genomeG", label = "Select genome", choices=list_genomes, selected  
= "BSgenome.Hsapiens.UCSC.hg19"),
```

### 3.3.2 Modify the default thresholds

#### Individual analysis

The default thresholds used for the individual analysis can be modified ([listing 4](#)).

**Listing 4.** Default thresholds in the individual analysis script. From the "ABSP\_individual\_analysis.Rmd" script.

```
607 '``'{r Thresholds, include=F}  
608  
609 # Thresholds  
610  
611 # Maximum base-calling error probability (value per position):  
612 th_quality_error <- 0.001  
613 # Minimum phred quality score, logarithmically linked to error probability  
# (value per position):  
614 th_quality_phred <- (-10*log(th_quality_error,10))  
615  
616 # Minimum ratio of primary peak, corresponding to the primary peak value  
# over the total of peak value, to consider a position as non-mixed (value per position):  
617 th_mixed_position <- 0.75  
618 # Minimum percentage of non-mixed positions in the trimmed sequence to be  
# considered as non-mixed (value for the total trimmed sequence):  
619 th_mixed_perc <- 75 # %  
620  
621 # Minimum length of trimmed sequences  
622 th_min_trim <- 30 # bp  
623  
624 # Minimum length of aligned sequences  
625 th_min_alg <- 30 # bp  
626 '``'
```

#### Grouped analysis

The default thresholds used for the grouped analysis, for clones methylation percentage conversion into methylation status (0% or 100%), can be modified ([listing 5](#)).

### 3.3.3 Modify the plots colors and point shapes

For the grouped analysis, the plot colors and point shapes parameters can be modified. As depicted in [listing 6](#), the color of bases can be changed in the *bases\_colors*

**Listing 5.** Default thresholds in the grouped analysis script. From the "ABSP\_grouped\_analysis.Rmd" script.

```
564 ' '{r Thresholds, include=F}
565 
566 # Thresholds (used for cloning only) :
567 
568 # unmethylated clones : methylation between 0% and 20%
569 th_unmethylated_max <- 20
570 
571 # methylated clones : methylation between 80% and 100%
572 th_methylated_min <- 80
573 
574 # maximum proportion of partial positions allowed : 20% of CpG positions
575 th_partialpos_ratio <- 0.2
576 
577 clone_thresholds <- c(th_unmethylated_max, th_methylated_min, th_
578     partialpos_ratio)
      '''
```

object, the plot colors for each group in the *plot\_colors* object and the point shapes for each group in the *plot\_shapes* object.

In the *ABSP\_grouped\_analysis.Rmd* script, the code chunk below the one depicted in the [listing 6](#) gives examples of different palette colors and points shapes, with functions to visualize them.

**Listing 6.** Colors and shapes setting for plots in the grouped analysis script. From the *ABSP\_grouped\_analysis.Rmd* script

```

364 '``{r colors and shapes settings, include=F}
365 # Colors of sequence track for genomic plots
366 bases_colors <- c(A="#43CD80", T="#D7191C", G="#FDC661", C="#2C7BB6", N="#
367   7F7F7F")
368
369 # Colors of groups for plots
370 plot_colors <- c("aqua"="#00AFBB",
371   "tangerine"="#FC4E07",
372   "sun"="#E7B800",
373   "berry"="#d30446",
374   "lime"="#90c613",
375   "grape"="#7839de",
376   "flamingo"="#d12a97",
377   "jade"="#00b673",
378   "ink"="#1221ed",
379   "terracotta"="#a93b2c")
380
381 # Shapes of groups for methylation profile plots
382 plot_shapes <- c("round"=19,
383   "square"=15,
384   "triangle"=17,
385   "round_border"=21,
386   "square_border"=22,
387   "triangle_border"=24,
388   "diamond"=18,
389   "small_round"=20,
390   "reverse_triangle_border"=25,
391   "diamond_border"=23)
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
878
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
978
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1087
1088
1089
1090
1091
1092
1093
1094
1095
1095
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1157
1158
1159
1160
1161
1162
1163
1164
1165
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1185
1186
1187
1188
1189
1189
1190
1191
1192
1193
1193
1194
1195
1196
1197
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1213
1214
1215
1216
1216
1217
1218
1219
1219
1220
1221
1222
1223
1223
1224
1225
1226
1226
1227
1228
1229
1229
1230
1231
1232
1233
1233
1234
1235
1236
1236
1237
1238
1239
1239
1240
1241
1242
1243
1243
1244
1245
1246
1246
1247
1248
1249
1249
1250
1251
1252
1253
1253
1254
1255
1256
1256
1257
1258
1259
1259
1260
1261
1262
1263
1263
1264
1265
1266
1266
1267
1268
1269
1269
1270
1271
1272
1273
1273
1274
1275
1276
1276
1277
1278
1279
1279
1280
1281
1282
1283
1283
1284
1285
1286
1286
1287
1288
1289
1289
1290
1291
1292
1293
1293
1294
1295
1296
1296
1297
1298
1299
1299
1300
1301
1302
1303
1303
1304
1305
1306
1306
1307
1308
1309
1309
1310
1311
1312
1312
1313
1314
1315
1315
1316
1317
1318
1318
1319
1320
1321
1321
1322
1323
1324
1324
1325
1326
1327
1327
1328
1329
1329
1330
1331
1332
1332
1333
1334
1335
1335
1336
1337
1338
1338
1339
1340
1341
1341
1342
1343
1344
1344
1345
1346
1347
1347
1348
1349
1349
1350
1351
1352
1352
1353
1354
1355
1355
1356
1357
1358
1358
1359
1360
1361
1361
1362
1363
1364
1364
1365
1366
1367
1367
1368
1369
1369
1370
1371
1372
1372
1373
1374
1375
1375
1376
1377
1377
1378
1379
1379
1380
1381
1382
1382
1383
1384
1385
1385
1386
1387
1388
1388
1389
1390
1391
1391
1392
1393
1394
1394
1395
1396
1397
1397
1398
1399
1399
1400
1401
1402
1402
1403
1404
1405
1405
1406
1407
1408
1408
1409
1410
1411
1411
1412
1413
1414
1414
1415
1416
1417
1417
1418
1419
1419
1420
1421
1422
1422
1423
1424
1425
1425
1426
1427
1427
1428
1429
1429
1430
1431
1432
1432
1433
1434
1435
1435
1436
1437
1438
1438
1439
1440
1441
1441
1442
1443
1444
1444
1445
1446
1447
1447
1448
1449
1449
1450
1451
1452
1452
1453
1454
1455
1455
1456
1457
1458
1458
1459
1460
1461
1461
1462
1463
1464
1464
1465
1466
1467
1467
1468
1469
1469
1470
1471
1472
1472
1473
1474
1475
1475
1476
1477
1477
1478
1479
1479
1480
1481
1482
1482
1483
1484
1485
1485
1486
1487
1488
1488
1489
1490
1491
1491
1492
1493
1494
1494
1495
1496
1497
1497
1498
1499
1499
1500
1501
1502
1502
1503
1504
1505
1505
1506
1507
1508
1508
1509
1510
1511
1511
1512
1513
1514
1514
1515
1516
1517
1517
1518
1519
1519
1520
1521
1522
1522
1523
1524
1525
1525
1526
1527
1527
1528
1529
1529
1530
1531
1532
1532
1533
1534
1535
1535
1536
1537
1538
1538
1539
1540
1541
1541
1542
1543
1544
1544
1545
1546
1547
1547
1548
1549
1549
1550
1551
1552
1552
1553
1554
1555
1555
1556
1557
1558
1558
1559
1560
1561
1561
1562
1563
1564
1564
1565
1566
1567
1567
1568
1569
1569
1570
1571
1572
1572
1573
1574
1575
1575
1576
1577
1577
1578
1579
1579
1580
1581
1582
1582
1583
1584
1585
1585
1586
1587
1588
1588
1589
1590
1591
1591
1592
1593
1594
1594
1595
1596
1597
1597
1598
1599
1599
1600
1601
1602
1602
1603
1604
1605
1605
1606
1607
1608
1608
1609
1610
1611
1611
1612
1613
1614
1614
1615
1616
1617
1617
1618
1619
1619
1620
1621
1622
1622
1623
1624
1625
1625
1626
1627
1627
1628
1629
1629
1630
1631
1632
1632
1633
1634
1635
1635
1636
1637
1638
1638
1639
1640
1641
1641
1642
1643
1644
1644
1645
1646
1647
1647
1648
1649
1649
1650
1651
1652
1652
1653
1654
1655
1655
1656
1657
1658
1658
1659
1660
1661
1661
1662
1663
1664
1664
1665
1666
1667
1667
1668
1669
1669
1670
1671
1672
1672
1673
1674
1675
1675
1676
1677
1677
1678
1679
1679
1680
1681
1682
1682
1683
1684
1685
1685
1686
1687
1688
1688
1689
1690
1691
1691
1692
1693
1694
1694
1695
1696
1697
1697
1698
1699
1699
1700
1701
1702
1702
1703
1704
1705
1705
1706
1707
1708
1708
1709
1710
1711
1711
1712
1713
1714
1714
1715
1716
1717
1717
1718
1719
1719
1720
1721
1722
1722
1723
1724
1725
1725
1726
1727
1727
1728
1729
1729
1730
1731
1732
1732
1733
1734
1735
1735
1736
1737
1738
1738
1739
1740
1741
1741
1742
1743
1744
1744
1745
1746
1747
1747
1748
1749
1749
1750
1751
1752
1752
1753
1754
1755
1755
1756
1757
1758
1758
1759
1760
1761
1761
1762
1763
1764
1764
1765
1766
1767
1767
1768
1769
1769
1770
1771
1772
1772
1773
1774
1775
1775
1776
1777
1777
1778
1779
1779
1780
1781
1782
1782
1783
1784
1785
1785
1786
1787
1788
1788
1789
1790
1791
1791
1792
1793
1794
1794
1795
1796
1797
1797
1798
1799
1799
1800
1801
1802
1802
1803
1804
1805
1805
1806
1807
1808
1808
1809
1810
1811
1811
1812
1813
1814
1814
1815
1816
1817
1817
1818
1819
1819
1820
1821
1822
1822
1823
1824
1825
1825
1826
1827
1827
1828
1829
1829
1830
1831
1832
1832
1833
1834
1835
1835
1836
1837
1838
1838
1839
1840
1841
1841
1842
1843
1844
1844
1845
1846
1847
1847
1848
1849
1849
1850
1851
1852
1852
1853
1854
1855
1855
1856
1857
1858
1858
1859
1860
1861
1861
1862
1863
1864
1864
1865
1866
1867
1867
1868
1869
1869
1870
1871
1872
1872
1873
1874
1875
1875
1876
1877
1877
1878
1879
1879
1880
1881
1882
1882
1883
1884
1885
1885
1886
1887
1888
1888
1889
1890
1891
1891
1892
1893
1894
1894
1895
1896
1897
1897
1898
1899
1899
1900
1901
1902
1902
1903
1904
1905
1905
1906
1907
1908
1908
1909
1910
1911
1911
1912
1913
1914
1914
1915
1916
1917
1917
1918
1919
1919
1920
1921
1922
1922
1923
1924
1925
1925
1926
1927
1927
1928
1929
1929
1930
1931
1932
1932
1933
1934
1935
1935
1936
1937
1938
1938
1939
1940
1941
1941
1942
1943
1944
1944
1945
1946
1947
1947
1948
1949
1949
1950
1951
1952
1952
1953
1954
1955
1955
1956
1957
1958
1958
1959
1960
1961
1961
1962
1963
1964
1964
1965
1966
1967
1967
1968
1969
1969
1970
1971
1972
1972
1973
1974
1975
1975
1976
1977
1977
1978
1979
1979
1980
1981
1982
1982
1983
1984
1985
1985
1986
1987
1988
1988
1989
1990
1991
1991
1992
1993
1994
1994
1995
1996
1997
1997
1998
1999
1999
2000
2001
2002
2002
2003
2004
2005
2005
2006
2007
2008
2008
2009
2010
2011
2011
2012
2013
2014
2014
2015
2016
2017
2017
2018
2019
2019
2020
2021
2022
2022
2023
2024
2025
2025
2026
2027
2027
2028
2029
2029
2030
2031
2032
2032
2033
2034
2035
2035
2036
2037
2038
2038
2039
2040
2041
2041
2042
2043
2044
2044
2045
2046
2047
2047
2048
2049
2049
2050
2051
2052
2052
2053
2054
2055
2055
2056
2057
2058
2058
2059
2060
2061
2061
2062
2063
2064
2064
2065
2066
2067
2067
2068
2069
2069
2070
2071
2072
2072
2073
2074
2075
2075
2076
2077
2077
2078
2079
2079
2080
2081
2082
2082
2083
2084
2085
2085
2086
2087
2088
2088
2089
2090
2091
2091
2092
2093
2094
2094
2095
2096
2097
2097
2098
2099
2099
2100
2101
2102
2102
2103
2104
2105
2105
2106
2107
2108
2108
2109
2110
2111
2111
2112
2113
2114
2114
2115
2116
2117
2117
2118
2119
2119
2120
2121
2122
2122
2123
2124
2125
2125
2126
2127
2128
2128
2129
2130
2131
2131
2132
2133
2134
2134
2135
2136
2137
2137
2138
2139
2139
2140
2141
2142
2142
2143
2144
2145
2145
2146
2147
2148
2148
2149
2150
2151
2151
2152
2153
2154
2154
2155
2156
2157
2157
2158
2159
2159
2160
2161
2162
2162
2163
2164
2165
2165
2166
2167
2168
2168
2169
2170
2171
2171
2172
2173
2174
2174
2175
2176
2177
2177
2178
2179
2179
2180
2181
2182
2182
2183
2184
2185
2185
2186
2187
2188
2188
2189
2190
2191
2191
2192
2193
2194
2194
2195
2196
2197
2197
2198
2199
2199
2200
2201
2202
2202
2203
2204
2205
2205
2206
2207
2208
2208
2209
2210
2211
2211
2212
2213
2214
2214
2215
2216
2217
2217
2218
2219
2219
2220
2221
2222
2222
2223
2224
2225
2225
2226
2227
2228
2
```

## 4 Troubleshooting guide

### 4.1 General

Error	Cause	Solution
<b>After following the opening procedure (for the first time) and clicking on the "Run App" button, the app interface is not opening.</b>	Package installation failed or some packages are not fully installed. Plus, to finalize some package installation, R need to be restarted.	In the RStudio console, type "renv::restore(prompt=F)" and press enter to run it. The download and installation of packages should start. Then, in the top bar, click on "Session" and "Restart R". Then try to open the app by clicking on the "Run app" button. If the problem is not solved, please contact us with a description of the issue.
<b>Warning message about RStudio version in the RStudio console.</b> <i>"R graphics engine version X is not supported by this version of RStudio. The Plots tab will be disabled until a newer version of RStudio is installed."</i>	The RStudio installed on device is not up-to-date.	Download and install the newest version of RStudio at <a href="https://www.rstudio.com/products/rstudio/download/">https://www.rstudio.com/products/rstudio/download/</a> .
<b>Prompt message displayed in the RStudio console during the selected BSgenome installation.</b> <i>"Do you want to install from sources the package which needs compilation? (Yes/no/cancel)"</i>	The installation of the BSgenome requires other packages to be installed.	Type "Yes" in the console panel and press enter to run it.

<p><b>Warning message when opening the ABSP about project configuration and R version (and the app does not function).</b></p> <p><i>"This project is configured to use R version 'X.X.X', but 'X.X.X' is currently being used."</i></p>	<p>The R version does not match with the R project configuration, the R version is either older or newer than the one used to generate the ABSP project.</p>	<p>Either download the latest version of R to match the one of the project configuration (<a href="https://www.r-project.org/">https://www.r-project.org/</a>), or download the latest version of ABSP to match the R version (<a href="https://github.com/ABSP-methylation-tool/ABSP">https://github.com/ABSP-methylation-tool/ABSP</a>). If the currently available ABSP version does not match the currently available R version, download the corresponding version of R and contact us, in order for us to update ABSP as soon as possible to match the current R version.</p>
--	--	---

## 4.2 Individual analysis

Error	Cause	Solution
<p><b>Analysis report aborted after the reference DNA step.</b>  <i>Error: Length of reference DNA sequence does not match with the provided genomic coordinates. Please verify concordance between the reference DNA sequence and genomic coordinates.</i></p>	The provided reference sequence in the fasta file has not the same length as the one calculated based on genomic coordinates provided in the fasta file header. So the sequence and coordinates might not match.	Check the reference and its coordinates. If the IGV (Integrative Genomics Viewer) software is used to get them, make sure to properly verify the first and last nucleotides, as there can be a one nucleotide difference between the coordinates and the actual sequence.
<p><b>Analysis report aborted after the sequencing trimming step.</b>  <i>Error: Analysis has been stopped as none of the sequencing results are of sufficient quality to be used.</i></p>	The provided sequencing results are not of good quality, the trimming step was not able to find a trimmed sequence long enough (below length threshold) and/or with poor quality (below quality score and/or non-mixed positions thresholds).	Lowering a bit the trimming thresholds might solve the issue, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.
<p><b>Analysis report aborted after the alignment step.</b>  <i>Error: Analysis has been stopped as none of the possible alignments are of sufficient length (&lt; N bp) to be used.</i></p>	Even if the sequencing results were of good quality and successfully trimmed, the alignments of them with the reference DNA sequence gives too short aligned sequences to pursue the analysis.	Check if the reference DNA sequence is the correct one. Check if the sequencing results passed the trimming step with values just above thresholds that might explain the alignment results. In this case, lowering some thresholds may solve the issue. If not, the best option is to provide new results from a new sequencing run.

<p><b>Analysis report aborted after the alignment step.</b></p> <p><i>Error: Analysis has been stopped as sequencing result direction (forward or reverse) could not be found.</i></p>	<p>This error occurs when the determined direction for both sequencing results happened to be the same, or when only one sequencing has been successfully trimmed and aligned, but alignments as forward or reverse have the same length.</p>	<p>For the first case, the results might have passed the alignment step with values just below thresholds, explaining the identical length of aligned sequence in both direction, still the best option should be to provide new sequencing results to have better trimmed and aligned results. For the second case, lowering a bit some thresholds might help for the other sequencing to pass the trimming and alignment steps, still the best option should be to provide new sequencing results to have better trimmed and aligned results.</p>
<p><b>Analysis report aborted after the alignment step.</b></p> <p><i>Error: Analysis has been stopped as none of the provided sequencing are of sufficient quality to be used.</i></p>	<p>The provided sequencing results are not of good quality, they passed the trimming step but not the alignment step. The aligned sequences were not long enough (below length threshold) or direction determination failed.</p>	<p>Lowering a bit the trimming and alignment thresholds might solve the issue, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>
<p><b>Analysis report aborted after the Quality Control (QC) step.</b></p> <p><i>Error: Analysis has been stopped as sequencing results are defined as incorrect by Quality Control. The analysis can not be performed.</i></p>	<p>The provided sequencing results are not of good quality when compared to the reference DNA; they passed the trimming and alignment steps but the quality control found the sequencing results incorrect: either the identity percentage or the mean bisulfite conversion rates is are below the threshold.</p>	<p>Lowering a bit the thresholds might solve the issue, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>

<p><b>Analysis report aborted after the Quality Control (QC) step.</b></p> <p><i>Error: Analysis has been stopped as no CpG sites were found covered by sequencing results.</i></p>	<p>Even if the sequencing results passed the trimming, alignment and quality control steps, the aligned results is not long enough and does not cover any CpG sites on the sequence, thereby methylation levels can not be computed.</p>	<p>Lowering a bit the thresholds might solve the issue to get longer aligned sequences, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>
<p><b>Analysis failed and no report has been generated</b></p>	<p>Unexpected error.</p>	<p>Please contact us and send us the error message appearing in the RStudio console.</p>

### 4.3 Grouped analysis

Error	Cause	Solution
<b>Analysis report aborted after the header.</b> <i>Error: No file methylation data files found. Check inputs: folder name and sequence name, and check the 'data' directory in the individual results folder.</i>	No methylation data files from the individual analysis, corresponding to the input information, were able to be retrieved.	Check if the individual analysis have already been run, if the files were not moved in another folder, they should be in the 'data' directory in the individual results folder, or if the input information as folder name, sequence name, and experiment type are correct. If the issue is not solved by these recommendations, please contact us.
<b>Analysis failed and no report has been generated.</b>	Unexpected error.	Please contact us and send us the error message appearing in the RStudio console.

## 5 Acknowledgements

An important part of the plot generation work for the grouped analysis was based on the R code provided by the [Methylation plotter](#)<sup>13,14</sup> tool.

Main R packages used:

- arrangements
- BiocManager
- Biostrings
- BSgenome
- compareGroups
- DiagrammeR
- dplyr
- formattable
- GenomeInfoDb
- GenomicRanges
- ggdendro
- ggplot2
- ggpubr
- Gviz
- htmltools
- htmlwidgets
- knitr
- openxlsx
- pdfTools
- plotly
- png
- purrr
- RColorBrewer
- readr
- renv
- rlist
- rmarkdown
- Rmisc
- rstatix
- sangeranalyseR
- sangerseqR
- seqinr
- shiny
- shinythemes
- webshot

---

<sup>13</sup>[http://maplab.imppc.org/methylation\\_plotter/index.html](http://maplab.imppc.org/methylation_plotter/index.html)

<sup>14</sup>Mallona, Díez-Villanueva, and Peinado, “Methylation plotter: a web tool for dynamic visualization of DNA methylation data”.

List of all R packages used:

- **abind 1.4.5** - Plate T, Heiberger R (2016). *\_abind: Combine Multidimensional Arrays\_*. R package version 1.4-5, <https://CRAN.R-project.org/package=abind>.
- **ade4 1.7.19** - Dray S, Dufour A (2007). "The ade4 Package: Implementing the Duality Diagram for Ecologists." *\_Journal of Statistical Software\_*, \*22\*(4), 1-20. doi:10.18637/jss.v022.i04 <https://doi.org/10.18637/jss.v022.i04>.
- **AnnotationDbi 1.58.0** - Pagès H, Carlson M, Falcon S, Li N (2022). *\_AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor\_*. R package version 1.58.0, <https://bioconductor.org/packages/AnnotationDbi>.
- **AnnotationFilter 1.20.0** - Morgan M, Rainer J (2022). *\_AnnotationFilter: Facilities for Filtering Bioconductor Annotation Resources\_*. R package version 1.20.0, <https://github.com/Bioconductor/AnnotationFilter>.
- **arrangements 1.1.9** - Lai R (2020). *\_arrangements: Fast Generators and Iterators for Permutations, Combinations, Integer Partitions and Compositions\_*. R package version 1.1.9, <https://CRAN.R-project.org/package=arrangements>.
- **askpass 1.1** - Ooms J (2019). *\_askpass: Safe Password Entry for R, Git, and SSH\_*. R package version 1.1, <https://CRAN.R-project.org/package=askpass>.
- **assertthat 0.2.1** - Wickham H (2019). *\_assertthat: Easy Pre and Post Assertions\_*. R package version 0.2.1, <https://CRAN.R-project.org/package=assertthat>.
- **backports 1.4.1** - Lang M, R Core Team (2021). *\_backports: Reimplementations of Functions Introduced Since R-3.0.0\_*. R package version 1.4.1, <https://CRAN.R-project.org/package=backports>.
- **base64enc 0.1.3** - Urbanek S (2015). *\_base64enc: Tools for base64 encoding\_*. R package version 0.1-3, <https://CRAN.R-project.org/package=base64enc>.
- **BH 1.78.0.0** - Eddelbuettel D, Emerson JW, Kane MJ (2021). *\_BH: Boost C++ Header Files\_*. R package version 1.78.0-0, <https://CRAN.R-project.org/package=BH>.
- **Biobase 2.56.0** - Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Ole's AK, Pag'es H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M (2015). "Orchestrating high-throughput genomic analysis with Bioconductor." *\_Nature Methods\_*, \*12\*(2), 115-121. <http://www.nature.com/nmeth/journal/v12/n2/full/nmeth.3252.html>.
- **BiocGenerics 0.42.0** - Huber, W., Carey, J. V., Gentleman, R., Anders, S., Carlson, M., Carvalho, S. B, Bravo, C. H, Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, D. K, Irizarry, A. R, Lawrence, M., Love, I. M, MacDonald, J., Obenchain, V., Ole's, K. A, Pag'es, H., Reyes, A., Shannon, P., Smyth, K. G, Tenenbaum, D., Waldron, L., Morgan, M. (2015). "Orchestrating high-throughput genomic analysis with Bioconductor." *\_Nature Methods\_*, \*12\*(2), 115-121. <http://www.nature.com/nmeth/journal/v12/n2/full/nmeth.3252.html>.
- **BiocManager 1.30.17** - Morgan M (2022). *\_BiocManager: Access the Bioconductor Project Package Repository\_*. R package version 1.30.17, <https://CRAN.R-project.org/package=BiocManager>.
- **BiocParallel 1.30.0** - Morgan M, Wang J, Obenchain V, Lang M, Thompson R, Turaga N (2022). *\_BiocParallel: Bioconductor facilities for parallel evaluation\_*. R package version 1.30.0, <https://github.com/Bioconductor/BiocParallel>.
- **BiocStyle 2.24.0** - Oleś A (2022). *\_BiocStyle: Standard styles for vignettes and other Bioconductor documents\_*. R package version 2.24.0, <https://github.com/Bioconductor/BiocStyle>

## [BiocStyle](#).

- **Biostrings 2.64.0** - Pagès H, Aboyoun P, Gentleman R, DebRoy S (2022). *\_Biostrings: Efficient manipulation of biological strings\_*. R package version 2.64.0, <https://bioconductor.org/packages/Biostrings>.
- **bit 4.0.4** - Oehlschlägel J, Ripley B (2020). *\_bit: Classes and Methods for Fast Memory-Efficient Boolean Selections\_*. R package version 4.0.4, <https://CRAN.R-project.org/package=bit>.
- **bit64 4.0.5** - Oehlschlägel J, Silvestri L (2020). *\_bit64: A S3 Class for Vectors of 64bit Integers\_*. R package version 4.0.5, <https://CRAN.R-project.org/package=bit64>.
- **bitops 1.0.7** - port SobSDiR, revised ebMM, Dutky mbS (2021). *\_bitops: Bitwise Operations\_*. R package version 1.0-7, <https://CRAN.R-project.org/package=bitops>.
- **blob 1.2.3** - Wickham H (2022). *\_blob: A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')\_*. R package version 1.2.3, <https://CRAN.R-project.org/package=blob>.
- **bookdown 0.26** - Xie Y (2022). *\_bookdown: Authoring Books and Technical Documents with R Markdown\_*. R package version 0.26, <https://github.com/rstudio/bookdown>.
- **brio 1.1.3** - Hester J, Csárdi G (2021). *\_brio: Basic R Input Output\_*. R package version 1.1.3, <https://CRAN.R-project.org/package=brio>.
- **broom 0.8.0** - Robinson D, Hayes A, Couch S (2022). *\_broom: Convert Statistical Objects into Tidy Tibbles\_*. R package version 0.8.0, <https://CRAN.R-project.org/package=broom>.
- **BSgenome 1.64.0** - Pagès H (2022). *\_BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs\_*. R package version 1.64.0, <https://bioconductor.org/packages/BSgenome>.
- **bslib 0.3.1** - Sievert C, Cheng J (2021). *\_bslib: Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'\_*. R package version 0.3.1, <https://CRAN.R-project.org/package=bslib>.
- **cachem 1.0.6** - Chang W (2021). *\_cachem: Cache R Objects with Automatic Pruning\_*. R package version 1.0.6, <https://CRAN.R-project.org/package=cachem>.
- **callr 3.7.0** - Csárdi G, Chang W (2021). *\_callr: Call R from R\_*. R package version 3.7.0, <https://CRAN.R-project.org/package=callr>.
- **car 3.0.13** - Fox J, Weisberg S (2019). *\_An R Companion to Applied Regression\_*, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- **carData 3.0.5** - Fox J, Weisberg S, Price B (2022). *\_carData: Companion to Applied Regression Data Sets\_*. R package version 3.0-5, <https://CRAN.R-project.org/package=carData>.
- **checkmate 2.1.0** - Lang M (2017). "checkmate: Fast Argument Checks for Defensive R Programming." *\_The R Journal\_*, \*9\*(1), 437-445. doi:10.32614/RJ-2017-028 <https://doi.org/10.32614/RJ-2017-028>.
- **chron 2.3.56** - James D, Hornik K (2020). *\_chron: Chronological Objects which Can Handle Dates and Times\_*. R package version 2.3-56. S original by David James, R port by Kurt Hornik., <https://CRAN.R-project.org/package=chron>.
- **cli 3.3.0** - Csárdi G (2022). *\_cli: Helpers for Developing Command Line Interfaces\_*. R package version 3.3.0, <https://CRAN.R-project.org/package=cli>.

- **clipr 0.8.0** - Lincoln M (2022). *\_clipr: Read and Write from the System Clipboard\_*. R package version 0.8.0, <https://CRAN.R-project.org/package=clipr>.
- **colorspace 2.0.3** - Zeileis A, Fisher JC, Hornik K, Ihaka R, McWhite CD, Murrell P, Stauffer R, Wilke CO (2020). "colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes." *\_Journal of Statistical Software\_*, \*96\*(1), 1-49. doi:10.18637/jss.v096.i01 <https://doi.org/10.18637/jss.v096.i01>.
- **commonmark 1.8.0** - Ooms J (2022). *\_commonmark: High Performance CommonMark and Github Markdown Rendering in R\_*. R package version 1.8.0, <https://CRAN.R-project.org/package=commonmark>.
- **compareGroups 4.5.1** - Subirana I, Sanz H, Vila J (2014). "Building Bivariate Tables: The compareGroups Package for R." *\_Journal of Statistical Software\_*, \*57\*(12), 1-16. <https://www.jstatsoft.org/v57/i12/>.
- **corrplot 0.92** - Wei T, Simko V (2021). *\_R package 'corrplot': Visualization of a Correlation Matrix\_*. (Version 0.92), <https://github.com/taiyun/corrplot>.
- **cowplot 1.1.1** - Wilke C (2020). *\_cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'\_*. R package version 1.1.1, <https://CRAN.R-project.org/package=cowplot>.
- **cpp11 0.4.2** - Hester J, François R (2021). *\_cpp11: A C++11 Interface for R's C Interface\_*. R package version 0.4.2, <https://CRAN.R-project.org/package=cpp11>.
- **crayon 1.5.1** - Csárdi G (2022). *\_crayon: Colored Terminal Output\_*. R package version 1.5.1, <https://CRAN.R-project.org/package=crayon>.
- **crosstalk 1.2.0** - Cheng J, Sievert C (2021). *\_crosstalk: Inter-Widget Interactivity for HTML Widgets\_*. R package version 1.2.0, <https://CRAN.R-project.org/package=crosstalk>.
- **curl 4.3.2** - Ooms J (2021). *\_curl: A Modern and Flexible Web Client for R\_*. R package version 4.3.2, <https://CRAN.R-project.org/package=curl>.
- **data.table 1.14.2** - Dowle M, Srinivasan A (2021). *\_data.table: Extension of 'data.frame'\_*. R package version 1.14.2, <https://CRAN.R-project.org/package=data.table>.
- **DBI 1.1.2** - R Special Interest Group on Databases (R-SIG-DB), Wickham H, Müller K (2021). *\_DBI: R Database Interface\_*. R package version 1.1.2, <https://CRAN.R-project.org/package=DBI>.
- **dbplyr 2.1.1** - Wickham H, Girlich M, Ruiz E (2021). *\_dbplyr: A 'dplyr' Back End for Databases\_*. R package version 2.1.1, <https://CRAN.R-project.org/package=dbplyr>.
- **DelayedArray 0.22.0** - Pagès H, Hickey wcfP, Lun A (2022). *\_DelayedArray: A unified framework for working transparently with on-disk and in-memory array-like datasets\_*. R package version 0.22.0, <https://bioconductor.org/packages/DelayedArray>.
- **desc 1.4.1** - Csárdi G, Müller K, Hester J (2022). *\_desc: Manipulate DESCRIPTION Files\_*. R package version 1.4.1, <https://CRAN.R-project.org/package=desc>.
- **DiagrammeR 1.0.9** - Iannone R (2022). *\_DiagrammeR: Graph/Network Visualization\_*. R package version 1.0.9, <https://CRAN.R-project.org/package=DiagrammeR>.
- **dichromat 2.0.0.1** - Lumley T (2022). *\_dichromat: Color Schemes for Dichromats\_*. R package version 2.0-0.1, <https://CRAN.R-project.org/package=dichromat>.
- **diffobj 0.3.5** - Gaslam B (2021). *\_diffobj: Diffs for R Objects\_*. R package version 0.3.5, <https://CRAN.R-project.org/package=diffobj>.
- **digest 0.6.29** - Lucas DEwcbA, Tuszynski J, Bengtsson H, Urbanek S, Frasca M, Lewis B, Stokely M, Muehlenleisen H, Murdoch D, Hester J, Wu W, Kou Q, Onkelinx T, Lang M, Simko V, Hornik K, Neal R, Bell K, de Queljoe M, Suruceanu I, Denney B, Schumacher D, Chang. aW

(2021). *\_digest: Create Compact Hash Digests of R Objects\_*. R package version 0.6.29, <https://CRAN.R-project.org/package=digest>.

- **downloader 0.4** - Chang W (2015). *\_downloader: Download Files over HTTP and HTTPS\_*. R package version 0.4, <https://CRAN.R-project.org/package=downloader>.
- **dplyr 1.0.9** - Wickham H, François R, Henry L, Müller K (2022). *\_dplyr: A Grammar of Data Manipulation\_*. R package version 1.0.9, <https://CRAN.R-project.org/package=dplyr>.
- **DT 0.23** - Xie Y, Cheng J, Tan X (2022). *\_DT: A Wrapper of the JavaScript Library 'DataTables'\_*. R package version 0.23, <https://CRAN.R-project.org/package=DT>.
- **ellipsis 0.3.2** - Wickham H (2021). *\_ellipsis: Tools for Working with ....\_*. R package version 0.3.2, <https://CRAN.R-project.org/package=ellipsis>.
- **ensemblDb 2.20.1** - Rainer J, Gatto L, Weichenberger CX (2019). "ensemblDb: an R package to create and use Ensembl-based annotation resources." *\_Bioinformatics\_*. doi:10.1093/bioinformatics/btz031 <https://doi.org/10.1093/bioinformatics/btz031>, <https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz031/5301311>.
- **evaluate 0.15** - Wickham H, Xie Y (2022). *\_evaluate: Parsing and Evaluation Tools that Provide More Details than the Default\_*. R package version 0.15, <https://CRAN.R-project.org/package=evaluate>.
- **excelR 0.4.0** - Bista S (2020). *\_excelR: A Wrapper of the 'JavaScript' Library 'jExcel'\_*. R package version 0.4.0, <https://CRAN.R-project.org/package=excelR>.
- **fansi 1.0.3** - Gaslam B (2022). *\_fansi: ANSI Control Sequence Aware String Functions\_*. R package version 1.0.3, <https://CRAN.R-project.org/package=fansi>.
- **farver 2.1.0** - Pedersen T, Nicolae B, François R (2021). *\_farver: High Performance Colour Space Manipulation\_*. R package version 2.1.0, <https://CRAN.R-project.org/package=farver>.
- **fastmap 1.1.0** - Chang W (2021). *\_fastmap: Fast Data Structures\_*. R package version 1.1.0, <https://CRAN.R-project.org/package=fastmap>.
- **fastmatch 1.1.3** - Urbanek S (2021). *\_fastmatch: Fast 'match()' Function\_*. R package version 1.1-3, <https://CRAN.R-project.org/package=fastmatch>.
- **filelock 1.0.2** - Csárdi G (2018). *\_filelock: Portable File Locking\_*. R package version 1.0.2, <https://CRAN.R-project.org/package=filelock>.
- **flextable 0.7.0** - Gohel D (2022). *\_flextable: Functions for Tabular Reporting\_*. R package version 0.7.0, <https://CRAN.R-project.org/package=flextable>.
- **fontawesome 0.2.2** - Iannone R (2021). *\_fontawesome: Easily Work with 'Font Awesome' Icons\_*. R package version 0.2.2, <https://CRAN.R-project.org/package=fontawesome>.
- **formatR 1.12** - Xie Y (2022). *\_formatR: Format R Code Automatically\_*. R package version 1.12, <https://CRAN.R-project.org/package=formatR>.
- **formattable 0.2.1** - Ren K, Russell K (2021). *\_formattable: Create 'Formattable' Data Structures\_*. R package version 0.2.1, <https://CRAN.R-project.org/package=formattable>.
- **Formula 1.2.4** - Zeileis A, Croissant Y (2010). "Extended Model Formulas in R: Multiple Parts and Multiple Responses." *\_Journal of Statistical Software\_*, \*34\*(1), 1-13. doi:10.18637/jss.v034.i01 <https://doi.org/10.18637/jss.v034.i01>.
- **fs 1.5.2** - Hester J, Wickham H, Csárdi G (2021). *\_fs: Cross-Platform File System Operations Based on 'libuv'\_*. R package version 1.5.2, <https://CRAN.R-project.org/package=fs>.
- **futile.logger 1.4.3** - Rowe BLY (2016). *\_futile.logger: A Logging Utility for R\_*. R package version 1.4.3, <https://CRAN.R-project.org/package=futile.logger>.

- **futile.options 1.0.1** - Rowe BLY (2018). *\_futile.options: Futile Options Management\_*. R package version 1.0.1, <https://CRAN.R-project.org/package=futile.options>.
- **gdtools 0.2.4** - Gohel D, Wickham H, Henry L, Ooms J (2022). *\_gdtools: Utilities for Graphical Rendering\_*. R package version 0.2.4, <https://CRAN.R-project.org/package=gdtools>.
- **generics 0.1.2** - Wickham H, Kuhn M, Vaughan D (2022). *\_generics: Common S3 Generics not Provided by Base R Methods Related to Model Fitting\_*. R package version 0.1.2, <https://CRAN.R-project.org/package=generics>.
- **GenomeInfoDb 1.32.1** - Arora S, Morgan M, Carlson M, Pagès H (2022). *\_GenomeInfoDb: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style\_*. R package version 1.32.1, <https://bioconductor.org/packages/GenomeInfoDb>.
- **GenomicAlignments 1.32.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *\_PLoS Computational Biology\_*, \*9\*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **GenomicFeatures 1.48.1** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *\_PLoS Computational Biology\_*, \*9\*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **GenomicRanges 1.48.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *\_PLoS Computational Biology\_*, \*9\*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **ggdendro 0.1.23** - de Vries A, Ripley BD (2022). *\_ggdendro: Create Dendograms and Tree Diagrams Using 'ggplot2'\_*. R package version 0.1.23, <https://CRAN.R-project.org/package=ggdendro>.
- **ggplot2 3.3.6** - Wickham H (2016). *\_ggplot2: Elegant Graphics for Data Analysis\_*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- **ggpubr 0.4.0** - Kassambara A (2020). *\_ggpubr: 'ggplot2' Based Publication Ready Plots\_*. R package version 0.4.0, <https://CRAN.R-project.org/package=ggpubr>.
- **ggrepel 0.9.1** - Slowikowski K (2021). *\_ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'\_*. R package version 0.9.1, <https://CRAN.R-project.org/package=ggrepel>.
- **ggsci 2.9** - Xiao N (2018). *\_ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'\_*. R package version 2.9, <https://CRAN.R-project.org/package=ggsci>.
- **ggsignif 0.6.3** - Constantin A, Patil I (2021). "ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'." *\_PsyArxiv\_*. doi:10.31234/osf.io/7awm6 <https://doi.org/10.31234/osf.io/7awm6>, <https://psyarxiv.com/7awm6>.
- **glue 1.6.2** - Hester J, Bryan J (2022). *\_glue: Interpreted String Literals\_*. R package version 1.6.2, <https://CRAN.R-project.org/package=glue>.
- **gmp 0.6.5** - Lucas A, Scholz I, Boehme R, Jasson S, Maechler M (2022). *\_gmp: Multiple Precision Arithmetic\_*. R package version 0.6-5, <https://CRAN.R-project.org/package=gmp>.

- **gridExtra 2.3** - Auguie B (2017). *\_gridExtra: Miscellaneous Functions for "Grid" Graphics\_*. R package version 2.3, <https://CRAN.R-project.org/package=gridExtra>.
- **gttable 0.3.0** - Wickham H, Pedersen T (2019). *\_gttable: Arrange 'Grobs' in Tables\_*. R package version 0.3.0, <https://CRAN.R-project.org/package=gttable>.
- **Gviz 1.40.1** - Hahne F, Ivanek R (2016). "Statistical Genomics: Methods and Protocols." In Mathé E, Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor, 335-351. Springer New York, New York, NY. ISBN 978-1-4939-3578-9, doi:10.1007/978-1-4939-3578-9\_16 [https://doi.org/10.1007/978-1-4939-3578-9\\_16](https://doi.org/10.1007/978-1-4939-3578-9_16), [http://dx.doi.org/10.1007/978-1-4939-3578-9\\_16](http://dx.doi.org/10.1007/978-1-4939-3578-9_16).
- **HardyWeinberg 1.7.5** - Graffelman J (2015). "Exploring Diallelic Genetic Markers: The HardyWeinberg Package." *\_Journal of Statistical Software\_*, \*64\*(3), 1-23. <https://www.jstatsoft.org/v64/i03/>.
- **highr 0.9** - Xie Y, Qiu Y (2021). *\_highr: Syntax Highlighting for R Source Code\_*. R package version 0.9, <https://CRAN.R-project.org/package=highr>.
- **Hmisc 4.7.0** - Harrell Jr F (2022). *\_Hmisc: Harrell Miscellaneous\_*. R package version 4.7-0, <https://CRAN.R-project.org/package=Hmisc>.
- **hms 1.1.1** - Müller K (2021). *\_hms: Pretty Time of Day\_*. R package version 1.1.1, <https://CRAN.R-project.org/package=hms>.
- **htmlTable 2.4.0** - Gordon M, Gragg S, Konings P (2022). *\_htmlTable: Advanced Tables for Markdown/HTML\_*. R package version 2.4.0, <https://CRAN.R-project.org/package=htmlTable>.
- **htmltools 0.5.2** - Cheng J, Sievert C, Schloerke B, Chang W, Xie Y, Allen J (2021). *\_htmltools: Tools for HTML\_*. R package version 0.5.2, <https://CRAN.R-project.org/package=htmltools>.
- **htmlwidgets 1.5.4** - Vaidyanathan R, Xie Y, Allaire J, Cheng J, Sievert C, Russell K (2021). *\_htmlwidgets: HTML Widgets for R\_*. R package version 1.5.4, <https://CRAN.R-project.org/package=htmlwidgets>.
- **httpuv 1.6.5** - Cheng J, Chang W (2022). *\_httpuv: HTTP and WebSocket Server Library\_*. R package version 1.6.5, <https://CRAN.R-project.org/package=httpuv>.
- **httr 1.4.3** - Wickham H (2022). *\_httr: Tools for Working with URLs and HTTP\_*. R package version 1.4.3, <https://CRAN.R-project.org/package=httr>.
- **igraph 1.3.1** - Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *\_InterJournal\_*, \*Complex Systems\*, 1695. <https://igraph.org>.
- **influenceR 0.1.0.1** - Simon J, Aditya K (2021). *\_influenceR: Software Tools to Quantify Structural Importance of Nodes in a Network\_*. R package version 0.1.0.1, <https://CRAN.R-project.org/package=influenceR>.
- **IRanges 2.30.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *\_PLoS Computational Biology\_*, \*9\*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **isoband 0.2.5** - Wilke C, Pedersen T (2021). *\_isoband: Generate Isolines and Isobands from Regularly Spaced Elevation Grids\_*. R package version 0.2.5, <https://CRAN.R-project.org/package=isoband>.
- **jpeg 0.1.9** - Urbanek S (2021). *\_jpeg: Read and write JPEG images\_*. R package version 0.1-9, <https://CRAN.R-project.org/package=jpeg>.

- **jquerylib 0.1.4** - Sievert C, Cheng J (2021). `_jquerylib`: Obtain 'jQuery' as an HTML Dependency Object\_. R package version 0.1.4, <https://CRAN.R-project.org/package=jquerylib>.
- **jsonlite 1.8.0** - Ooms J (2014). "The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects." `_arXiv:1403.2805 [stat.CO]`\_. <https://arxiv.org/abs/1403.2805>.
- **kableExtra 1.3.4** - Zhu H (2021). `_kableExtra`: Construct Complex Table with 'kable' and Pipe Syntax\_. R package version 1.3.4, <https://CRAN.R-project.org/package=kableExtra>.
- **knitr 1.39** - Xie Y (2022). `_knitr`: A General-Purpose Package for Dynamic Report Generation in R\_. R package version 1.39, <https://yihui.org/knitr/>.
- **labeling 0.4.2** - Talbot, J (2020). `_labeling`: Axis Labeling\_. R package version 0.4.2, <https://CRAN.R-project.org/package=labeling>.
- **lambda.r 1.2.4** - Rowe BLY (2019). `_lambda.r`: Modeling Data with Functional Programming\_. R package version 1.2.4, <https://CRAN.R-project.org/package=lambda.r>.
- **later 1.3.0** - Chang W, Cheng J (2021). `_later`: Utilities for Scheduling Functions to Execute Later with Event Loops\_. R package version 1.3.0, <https://CRAN.R-project.org/package=later>.
- **latticeExtra 0.6.29** - Sarkar D, Andrews F (2019). `_latticeExtra`: Extra Graphical Utilities Based on Lattice\_. R package version 0.6-29, <https://CRAN.R-project.org/package=latticeExtra>.
- **lazyeval 0.2.2** - Wickham H (2019). `_lazyeval`: Lazy (Non-Standard) Evaluation\_. R package version 0.2.2, <https://CRAN.R-project.org/package=lazyeval>.
- **lifecycle 1.0.1** - Henry L, Wickham H (2021). `_lifecycle`: Manage the Life Cycle of your Package Functions\_. R package version 1.0.1, <https://CRAN.R-project.org/package=lifecycle>.
- **lme4 1.1.29** - Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." `_Journal of Statistical Software`, \*67\*(1), 1-48. doi:10.18637/jss.v067.i01 <https://doi.org/10.18637/jss.v067.i01>.
- **logger 0.2.2** - Daróczi G (2021). `_logger`: A Lightweight, Modern and Flexible Logging Utility\_. R package version 0.2.2, <https://CRAN.R-project.org/package=logger>.
- **magrittr 2.0.3** - Bache S, Wickham H (2022). `_magrittr`: A Forward-Pipe Operator for R\_. R package version 2.0.3, <https://CRAN.R-project.org/package=magrittr>.
- **maptools 1.1.4** - Bivand R, Lewin-Koh N (2022). `_maptools`: Tools for Handling Spatial Objects\_. R package version 1.1-4, <https://CRAN.R-project.org/package=maptools>.
- **MatrixGenerics 1.8.0** - Ahlmann-Eltze C, Hickey P, Pagès H (2022). `_MatrixGenerics`: S4 Generic Summary Statistic Functions that Operate on Matrix-Like Objects\_. R package version 1.8.0, <https://bioconductor.org/packages/MatrixGenerics>.
- **MatrixModels 0.5.0** - Bates D, Maechler M (2021). `_MatrixModels`: Modelling with Sparse and Dense Matrices\_. R package version 0.5-0, <https://CRAN.R-project.org/package=MatrixModels>.
- **matrixStats 0.62.0** - Bengtsson H (2022). `_matrixStats`: Functions that Apply to Rows and Columns of Matrices (and to Vectors)\_. R package version 0.62.0, <https://CRAN.R-project.org/package=matrixStats>.
- **memoise 2.0.1** - Wickham H, Hester J, Chang W, Müller K, Cook D (2021). `_memoise`: 'Memoisation' of Functions\_. R package version 2.0.1, <https://CRAN.R-project.org/package=memoise>.

- **mice 3.14.0** - van Buuren S, Groothuis-Oudshoorn K (2011). "mice: Multivariate Imputation by Chained Equations in R." *\_Journal of Statistical Software\_*, \*45\*(3), 1-67. doi:10.18637/jss.v045.i03 <https://doi.org/10.18637/jss.v045.i03>.
- **mime 0.12** - Xie Y (2021). *\_mime: Map Filenames to MIME Types\_*. R package version 0.12, <https://CRAN.R-project.org/package=mime>.
- **minqa 1.2.4** - Bates D, Mullen KM, Nash JC, Varadhan R (2014). *\_minqa: Derivative-free optimization algorithms by quadratic approximation\_*. R package version 1.2.4, <https://CRAN.R-project.org/package=minqa>.
- **munsell 0.5.0** - Wickham C (2018). *\_munsell: Utilities for Using Munsell Colours\_*. R package version 0.5.0, <https://CRAN.R-project.org/package=munsell>.
- **numDeriv 2016.8.1.1** - Gilbert P, Varadhan R (2019). *\_numDeriv: Accurate Numerical Derivatives\_*. R package version 2016.8-1.1, <https://CRAN.R-project.org/package=numDeriv>.
- **officer 0.4.2** - Gohel D (2022). *\_officer: Manipulation of Microsoft Word and PowerPoint Documents\_*. R package version 0.4.2, <https://CRAN.R-project.org/package=officer>.
- **openssl 2.0.1** - Ooms J (2022). *\_openssl: Toolkit for Encryption, Signatures and Certificates Based on OpenSSL\_*. R package version 2.0.1, <https://CRAN.R-project.org/package=openssl>.
- **openxlsx 4.2.5** - Schäuberger P, Walker A (2021). *\_openxlsx: Read, Write and Edit xlsx Files\_*. R package version 4.2.5, <https://CRAN.R-project.org/package=openxlsx>.
- **pbkrtest 0.5.1** - Halekoh U, Højsgaard S (2014). "A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest." *\_Journal of Statistical Software\_*, \*59\*(9), 1-30. <https://www.jstatsoft.org/v59/i09/>.
- **pdftools 3.2.0** - Ooms J (2022). *\_pdftools: Text Extraction, Rendering and Converting of PDF Documents\_*. R package version 3.2.0, <https://CRAN.R-project.org/package=pdftools>.
- **phangorn 2.8.1** - Schliep K (2011). "phangorn: phylogenetic analysis in R." *\_Bioinformatics\_*, \*27\*(4), 592-593. doi:10.1093/bioinformatics/btq706 <https://doi.org/10.1093/bioinformatics/btq706>.
- **pillar 1.7.0** - Müller K, Wickham H (2022). *\_pillar: Coloured Formatting for Columns\_*. R package version 1.7.0, <https://CRAN.R-project.org/package=pillar>.
- **pixmap 0.4.12** - Bivand R, Leisch F, Maechler M (2021). *\_pixmap: Bitmap Images / Pixel Maps\_*. R package version 0.4-12, <https://CRAN.R-project.org/package=pixmap>.
- **pkgconfig 2.0.3** - Csárdi G (2019). *\_pkgconfig: Private Configuration for 'R' Packages\_*. R package version 2.0.3, <https://CRAN.R-project.org/package=pkgconfig>.
- **pkgload 1.2.4** - Wickham H, Chang W, Hester J, Henry L (2021). *\_pkgload: Simulate Package Installation and Attach\_*. R package version 1.2.4, <https://CRAN.R-project.org/package=pkgload>.
- **plogr 0.2.0** - Müller K (2018). *\_plogr: The 'plog' C++ Logging Library\_*. R package version 0.2.0, <https://CRAN.R-project.org/package=plogr>.
- **plotly 4.10.0** - Sievert C (2020). *\_Interactive Web-Based Data Visualization with R, plotly, and shiny\_*. Chapman and Hall/CRC. ISBN 9781138331457, <https://plotly-r.com>.
- **plyr 1.8.7** - Wickham H (2011). "The Split-Apply-Combine Strategy for Data Analysis." *\_Journal of Statistical Software\_*, \*40\*(1), 1-29. <https://www.jstatsoft.org/v40/i01/>.
- **png 0.1.7** - Urbanek S (2013). *\_png: Read and write PNG images\_*. R package version 0.1-7, <https://CRAN.R-project.org/package=png>.

- **polynom 1.4.1** - Venables B, Hornik K, Maechler M (2022). *\_polynom: A Collection of Functions to Implement a Class for Univariate Polynomial Manipulations\_*. R package version 1.4-1. S original by Bill Venables, packages for R by Kurt Hornik and Martin Maechler., <https://CRAN.R-project.org/package=polynom>.
- **praise 1.0.0** - Csardi G, Sorhus S (2015). *\_praise: Praise Users\_*. R package version 1.0.0, <https://CRAN.R-project.org/package=praise>.
- **prettyunits 1.1.1** - Csardi G (2020). *\_prettyunits: Pretty, Human Readable Formatting of Quantities\_*. R package version 1.1.1, <https://CRAN.R-project.org/package=prettyunits>.
- **processx 3.5.3** - Csárdi G, Chang W (2022). *\_processx: Execute and Control System Processes\_*. R package version 3.5.3, <https://CRAN.R-project.org/package=processx>.
- **progress 1.2.2** - Csárdi G, FitzJohn R (2019). *\_progress: Terminal Progress Bars\_*. R package version 1.2.2, <https://CRAN.R-project.org/package=progress>.
- **promises 1.2.0.1** - Cheng J (2021). *\_promises: Abstractions for Promise-Based Asynchronous Programming\_*. R package version 1.2.0.1, <https://CRAN.R-project.org/package=promises>.
- **ProtGenerics 1.28.0** - Gatto L, Rainer J (2022). *\_ProtGenerics: Generic infrastructure for Bioconductor mass spectrometry packages\_*. R package version 1.28.0, <https://github.com/RforMassSpectrometry/ProtGenerics>.
- **ps 1.7.0** - Loden J, Daeschler D, Rodola' G, Csárdi G (2022). *\_ps: List, Query, Manipulate System Processes\_*. R package version 1.7.0, <https://CRAN.R-project.org/package=ps>.
- **purrr 0.3.4** - Henry L, Wickham H (2020). *\_purrr: Functional Programming Tools\_*. R package version 0.3.4, <https://CRAN.R-project.org/package=purrr>.
- **qpdf 1.1** - Ooms J (2019). *\_qpdf: Split, Combine and Compress PDF Files\_*. R package version 1.1, <https://CRAN.R-project.org/package=qpdf>.
- **quadprog 1.5.8** - dpodi/LINPACK) SobBATRpbAW<FcfcCM (2019). *\_quadprog: Functions to Solve Quadratic Programming Problems\_*. R package version 1.5-8, <<https://CRAN.R-project.org/package=quadprog>>
- **quantreg 5.93** - Koenker R (2022). *\_quantreg: Quantile Regression\_*. R package version 5.93, <https://CRAN.R-project.org/package=quantreg>.
- **R6 2.5.1** - Chang W (2021). *\_R6: Encapsulated Classes with Reference Semantics\_*. R package version 2.5.1, <https://CRAN.R-project.org/package=R6>.
- **rappdirs 0.3.3** - Ratnakumar S, Mick T, Davis T (2021). *\_rappdirs: Application Directories: Determine Where to Save Data, Caches, and Logs\_*. R package version 0.3.3, <https://CRAN.R-project.org/package=rappdirs>.
- **RColorBrewer 1.1.3** - Neuwirth E (2022). *\_RColorBrewer: ColorBrewer Palettes\_*. R package version 1.1-3, <https://CRAN.R-project.org/package=RColorBrewer>.
- **Rcpp 1.0.8.3** - Eddelbuettel D, François R (2011). "Rcpp: Seamless R and C++ Integration." *\_Journal of Statistical Software\_*, \*40\*(8), 1-18. doi:10.18637/jss.v040.i08 <https://doi.org/10.18637/jss.v040.i08>.
- **RcppEigen 0.3.3.9.2** - Bates D, Eddelbuettel D (2013). "Fast and Elegant Numerical Linear Algebra Using the RcppEigen Package." *\_Journal of Statistical Software\_*, \*52\*(5), 1-24. doi:10.18637/jss.v052.i05 <https://doi.org/10.18637/jss.v052.i05>.
- **RCurl 1.98.1.6** - Temple Lang D (2022). *\_RCurl: General Network (HTTP/FTP/...) Client Interface for R\_*. R package version 1.98-1.6, <https://CRAN.R-project.org/package=RCurl>.

- **readr 2.1.2** - Wickham H, Hester J, Bryan J (2022). *\_readr: Read Rectangular Text Data\_*. R package version 2.1.2, <https://CRAN.R-project.org/package=readr>.
- **rematch2 2.1.2** - Csárdi G (2020). *\_rematch2: Tidy Output from Regular Expression Matching\_*. R package version 2.1.2, <https://CRAN.R-project.org/package=rematch2>.
- **renv 0.15.4** - Ushey K (2022). *\_renv: Project Environments\_*. R package version 0.15.4, <https://CRAN.R-project.org/package=renv>.
- **reshape2 1.4.4** - Wickham H (2007). "Reshaping Data with the reshape Package." *\_Journal of Statistical Software\_*, \*21\*(12), 1-20. <http://www.jstatsoft.org/v21/i12/>.
- **restfulr 0.0.13** - Lawrence M (2017). *\_restfulr: R Interface to RESTful Web Services\_*. R package version 0.0.13, <https://CRAN.R-project.org/package=restfulr>.
- **rjson 0.2.21** - Couture-Bell A (2022). *\_rjson: JSON for R\_*. R package version 0.2.21, <https://CRAN.R-project.org/package=rjson>.
- **rlang 1.0.2** - Henry L, Wickham H (2022). *\_rlang: Functions for Base Types and Core R and 'Tidyverse' Features\_*. R package version 1.0.2, <https://CRAN.R-project.org/package=rlang>.
- **rlist 0.4.6.2** - Ren K (2021). *\_rlist: A Toolbox for Non-Tabular Data Manipulation\_*. R package version 0.4.6.2, <https://CRAN.R-project.org/package=rlist>.
- **rmarkdown 2.14** - Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2022). *\_rmarkdown: Dynamic Documents for R\_*. R package version 2.14, <https://github.com/rstudio/rmarkdown>.
- **Rmisc 1.5.1** - Hope RM (2022). *\_Rmisc: Ryan Miscellaneous\_*. R package version 1.5.1, <https://CRAN.R-project.org/package=Rmisc>.
- **rprojroot 2.0.3** - Müller K (2022). *\_rprojroot: Finding Files in Project Subdirectories\_*. R package version 2.0.3, <https://CRAN.R-project.org/package=rprojroot>.
- **Rsamtools 2.12.0** - Morgan M, Pagès H, Obenchain V, Hayden N (2022). *\_Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import\_*. R package version 2.12.0, <https://bioconductor.org/packages/Rsamtools>.
- **RSQlite 2.2.14** - Müller K, Wickham H, James DA, Falcon S (2022). *\_RSQlite: SQLite Interface for R\_*. R package version 2.2.14, <https://CRAN.R-project.org/package=RSQlite>.
- **rstatix 0.7.0** - Kassambara A (2021). *\_rstatix: Pipe-Friendly Framework for Basic Statistical Tests\_*. R package version 0.7.0, <https://CRAN.R-project.org/package=rstatix>.
- **rstudioapi 0.13** - Ushey K, Allaire J, Wickham H, Ritchie G (2020). *\_rstudioapi: Safely Access the RStudio API\_*. R package version 0.13, <https://CRAN.R-project.org/package=rstudioapi>.
- **rtracklayer 1.56.0** - Lawrence M, Gentleman R, Carey V (2009). "rtracklayer: an R package for interfacing with genome browsers." *\_Bioinformatics\_*, \*25\*, 1841-1842. doi:10.1093/bioinformatics/btp328 <https://doi.org/10.1093/bioinformatics/btp328>, <http://bioinformatics.oxfordjournals.org/content/25/14/1841.abstract>.
- **rvest 1.0.2** - Wickham H (2021). *\_rvest: Easily Harvest (Scrape) Web Pages\_*. R package version 1.0.2, <https://CRAN.R-project.org/package=rvest>.
- **S4Vectors 0.34.0** - Pagès H, Lawrence M, Aboym P (2022). *\_S4Vectors: Foundation of vector-like and list-like containers in Bioconductor\_*. R package version 0.34.0, <https://bioconductor.org/packages/S4Vectors>.
- **sangeranalyseR 1.6.1** - Chao K, Barton K, Palmer S, Lanfear R (2021). "sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R." *\_Genome Biology and*

Evolution\_. doi:10.1093/gbe/evab028 <https://doi.org/10.1093/gbe/evab028>.

- **sangerseqR 1.32.0** - Hill JT, Demarest BL, Bisgrove BW, Su Y, Smith M, Yost HJ (2014). "Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products." \_Developmental Dynamics\_. doi:10.1002/dvdy.24183. <https://doi.org/10.1002/dvdy.24183..>
- **sass 0.4.1** - Cheng J, Mastny T, Iannone R, Schloerke B, Sievert C (2022). \_sass: Syntactically Awesome Style Sheets ('Sass')\_. R package version 0.4.1, <https://CRAN.R-project.org/package=sass>.
- **scales 1.2.0** - Wickham H, Seidel D (2022). \_scales: Scale Functions for Visualization\_. R package version 1.2.0, <https://CRAN.R-project.org/package=scales>.
- **selectr 0.4.2** - Potter S (2012). "Introducing the selectr Package." The University of Auckland, Auckland, New Zealand. <http://stattech.wordpress.fos.auckland.ac.nz/2012-10-introducing-the-selectr-package/>.
- **sessioninfo 1.2.2** - Wickham H, Chang W, Flight R, Müller K, Hester J (2021). \_sessioninfo: R Session Information\_. R package version 1.2.2, <https://CRAN.R-project.org/package=sessioninfo>.
- **shiny 1.7.1** - Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2021). \_shiny: Web Application Framework for R\_. R package version 1.7.1, <https://CRAN.R-project.org/package=shiny>.
- **shinycssloaders 1.0.0** - Sali A, Attali D (2020). \_shinycssloaders: Add Loading Animations to a 'shiny' Output While It's Recalculating\_. R package version 1.0.0, <https://CRAN.R-project.org/package=shinycssloaders>.
- **shinydashboard 0.7.2** - Chang W, Borges Ribeiro B (2021). \_shinydashboard: Create Dashboards with 'Shiny'\_. R package version 0.7.2, <https://CRAN.R-project.org/package=shinydashboard>.
- **shinyjs 2.1.0** - Attali D (2021). \_shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds\_. R package version 2.1.0, <https://CRAN.R-project.org/package=shinyjs>.
- **shinythemes 1.2.0** - Chang W (2021). \_shinythemes: Themes for Shiny\_. R package version 1.2.0, <https://CRAN.R-project.org/package=shinythemes>.
- **shinyWidgets 0.7.0** - Perrier V, Meyer F, Granjon D (2022). \_shinyWidgets: Custom Inputs Widgets for Shiny\_. R package version 0.7.0, <https://CRAN.R-project.org/package=shinyWidgets>.
- **snow 0.4.4** - Tierney L, Rossini AJ, Li N, Sevcikova H (2021). \_snow: Simple Network of Workstations\_. R package version 0.4-4, <https://CRAN.R-project.org/package=snow>.
- **sourcetools 0.1.7** - Ushey K (2018). \_sourcetools: Tools for Reading, Tokenizing and Parsing R Code\_. R package version 0.1.7, <https://CRAN.R-project.org/package=sourcetools>.
- **sp 1.4.7** - Pebesma EJ, Bivand RS (2005). "Classes and methods for spatial data in R." \_R News\_, \*5\*(2), 9-13. <https://CRAN.R-project.org/doc/Rnews/>.
- **SparseM 1.81** - Koenker R (2021). \_SparseM: Sparse Linear Algebra\_. R package version 1.81, <https://CRAN.R-project.org/package=SparseM>.
- **stringr 1.4.0** - Wickham H (2019). \_stringr: Simple, Consistent Wrappers for Common String Operations\_. R package version 1.4.0, <https://CRAN.R-project.org/package=stringr>.

- **SummarizedExperiment 1.26.1** - Morgan M, Obenchain V, Hester J, Pagès H (2022). *\_SummarizedExperiment: SummarizedExperiment container\_*. R package version 1.26.1, <https://bioconductor.org/packages/SummarizedExperiment>.
- **svglite 2.1.0** - Wickham H, Henry L, Pedersen T, Luciani T, Decorde M, Lise V (2022). *\_svglite: An 'SVG' Graphics Device\_*. R package version 2.1.0, <https://CRAN.R-project.org/package=svglite>.
- **sys 3.4** - Ooms J (2020). *\_sys: Powerful and Reliable Tools for Running System Commands in R\_*. R package version 3.4, <https://CRAN.R-project.org/package=sys>.
- **systemfonts 1.0.4** - Pedersen T, Ooms J, Govett D (2022). *\_systemfonts: System Native Font Finding\_*. R package version 1.0.4, <https://CRAN.R-project.org/package=systemfonts>.
- **testthat 3.1.4** - Wickham H (2011). "testthat: Get Started with Testing." *\_The R Journal\_, \*3\*, 5-10.* [https://journal.r-project.org/archive/2011-1/RJournal\\_2011-1\\_Wickham.pdf](https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf).
- **tibble 3.1.7** - Müller K, Wickham H (2022). *\_tibble: Simple Data Frames\_*. R package version 3.1.7, <https://CRAN.R-project.org/package=tibble>.
- **tidy 1.2.0** - Wickham H, Girlich M (2022). *\_tidy: Tidy Messy Data\_*. R package version 1.2.0, <https://CRAN.R-project.org/package=tidy>.
- **tidyselect 1.1.2** - Henry L, Wickham H (2022). *\_tidyselect: Select from a Set of Strings\_*. R package version 1.1.2, <https://CRAN.R-project.org/package=tidyselect>.
- **tinytex 0.39** - Xie Y (2022). *\_tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents\_*. R package version 0.39, <https://github.com/rstudio/tinytex>.
- **truncnorm 1.0.8** - Mersmann O, Trautmann H, Steuer D, Bornkamp B (2018). *\_truncnorm: Truncated Normal Distribution\_*. R package version 1.0-8, <https://CRAN.R-project.org/package=truncnorm>.
- **tzdb 0.3.0** - Vaughan D (2022). *\_tzdb: Time Zone Database Information\_*. R package version 0.3.0, <https://CRAN.R-project.org/package=tzdb>.
- **utf8 1.2.2** - Perry PO (2021). *\_utf8: Unicode Text Processing\_*. R package version 1.2.2, <https://CRAN.R-project.org/package=utf8>.
- **uuid 1.1.0** - Urbanek S, Ts'o T (2022). *\_uuid: Tools for Generating and Handling of UUIDs\_*. R package version 1.1-0, <https://CRAN.R-project.org/package=uuid>.
- **VariantAnnotation 1.42.0** - Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M (2014). "VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants." *\_Bioinformatics\_, \*30\*(14), 2076-2078. doi:10.1093/bioinformatics/btu168* <https://doi.org/10.1093/bioinformatics/btu168>.
- **vctrs 0.4.1** - Wickham H, Henry L, Vaughan D (2022). *\_vctrs: Vector Helpers\_*. R package version 0.4.1, <https://CRAN.R-project.org/package=vctrs>.
- **viridis 0.6.2** - Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro A, Sciaini, Marco, Scherer, Cédric (2021). *\_viridis - Colorblind-Friendly Color Maps for R\_*. doi:10.5281/zenodo.4679424 <https://doi.org/10.5281/zenodo.4679424>, R package version 0.6.2, <https://sjmgarnier.github.io/viridis/>.
- **viridisLite 0.4.0** - Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro A, Sciaini, Marco, Scherer, Cédric (2021). *\_viridis - Colorblind-Friendly Color Maps for R\_*. doi:10.5281/zenodo.4679424 <https://doi.org/10.5281/zenodo.4679424>, R package version 0.4.0, <https://sjmgarnier.github.io/viridis/>.

- **visNetwork 2.1.0** - Almende B.V. and Contributors, Thieurmel B, Robert T (2021). `_visNetwork: Network Visualization using 'vis.js' Library_`. R package version 2.1.0, <https://CRAN.R-project.org/package=visNetwork>.
- **vroom 1.5.7** - Hester J, Wickham H, Bryan J (2021). `_vroom: Read and Write Rectangular Text Data Quickly_`. R package version 1.5.7, <https://CRAN.R-project.org/package=vroom>.
- **waldo 0.4.0** - Wickham H (2022). `_waldo: Find Differences Between R Objects_`. R package version 0.4.0, <https://CRAN.R-project.org/package=waldo>.
- **webshot 0.5.3** - Chang W (2022). `_webshot: Take Screenshots of Web Pages_`. R package version 0.5.3, <https://CRAN.R-project.org/package=webshot>.
- **withr 2.5.0** - Hester J, Henry L, Müller K, Ushey K, Wickham H, Chang W (2022). `_withr: Run Code 'With' Temporarily Modified Global State_`. R package version 2.5.0, <https://CRAN.R-project.org/package=withr>.
- **writexl 1.4.0** - Ooms J (2021). `_writexl: Export Data Frames to Excel 'xlsx' Format_`. R package version 1.4.0, <https://CRAN.R-project.org/package=writexl>.
- **xfun 0.31** - Xie Y (2022). `_xfun: Supporting Functions for Packages Maintained by 'Yihui Xie'_`. R package version 0.31, <https://CRAN.R-project.org/package=xfun>.
- **XML 3.99.0.9** - Temple Lang D (2022). `_XML: Tools for Parsing and Generating XML Within R and S-Plus_`. R package version 3.99-0.9, <https://CRAN.R-project.org/package=XML>.
- **xml2 1.3.3** - Wickham H, Hester J, Ooms J (2021). `_xml2: Parse XML_`. R package version 1.3.3, <https://CRAN.R-project.org/package=xml2>.
- **xtable 1.8.4** - Dahl D, Scott D, Roosen C, Magnusson A, Swinton J (2019). `_xtable: Export Tables to LaTeX or HTML_`. R package version 1.8-4, <https://CRAN.R-project.org/package=xtable>.
- **XVector 0.36.0** - Pagès H, Aboyoun P (2022). `_XVector: Foundation of external vector representation and manipulation in Bioconductor_`. R package version 0.36.0, <https://bioconductor.org/packages/XVector>.
- **yaml 2.3.5** - Garbett SP, Stephens J, Simonov K, Xie Y, Dong Z, Wickham H, Horner J, reikoch, Beasley W, O'Connor B, Warnes GR, Quinn M, Kamvar ZN (2022). `_yaml: Methods to Convert R Data to YAML and Back_`. R package version 2.3.5, <https://CRAN.R-project.org/package=yaml>.
- **zeallot 0.1.0** - Teetor N (2018). `_zeallot: Multiple, Unpacking, and Destructuring Assignment_`. R package version 0.1.0, <https://CRAN.R-project.org/package=zeallot>.
- **zip 2.2.0** - Csárdi G, Podgórski K, Geldreich R (2021). `_zip: Cross-Platform 'zip' Compression_`. R package version 2.2.0, <https://CRAN.R-project.org/package=zip>.
- **zlibbioc 1.42.0** - Morgan M (2022). `_zlibbioc: An R packaged zlib-1.2.5_`. R package version 1.42.0, <https://bioconductor.org/packages/zlibbioc>.
- **base 4.2.0** - R Core Team (2022). `_R: A Language and Environment for Statistical Computing_`. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **class 7.3.20** - Venables WN, Ripley BD (2002). `_Modern Applied Statistics with S_`, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- **cluster 2.1.3** - Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2022). `_cluster: Cluster Analysis Basics and Extensions_`. R package version 2.1.3 - For new features, see the 'Changelog' file (in the package source), <https://CRAN.R-project.org/package=cluster>.

- **codetools 0.2.18** - Tierney L (2020). *\_codetools: Code Analysis Tools for R\_*. R package version 0.2-18, <https://CRAN.R-project.org/package=codetools>.
- **compiler 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **datasets 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **foreign 0.8.82** - R Core Team (2022). *\_foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...\_*. R package version 0.8-82, <https://CRAN.R-project.org/package=foreign>.
- **graphics 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **grDevices 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **grid 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **KernSmooth 2.23.20** - Wand M (2021). *\_KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)\_*. R package version 2.23-20, <https://CRAN.R-project.org/package=KernSmooth>.
- **lattice 0.20.45** - Sarkar D (2008). *\_Lattice: Multivariate Data Visualization with R\_*. Springer, New York. ISBN 978-0-387-75968-5, <http://lmdvr.r-forge.r-project.org>.
- **MASS 7.3.56** - Venables WN, Ripley BD (2002). *\_Modern Applied Statistics with S\_*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- **Matrix 1.4.1** - Bates D, Maechler M, Jagan M (2022). *\_Matrix: Sparse and Dense Matrix Classes and Methods\_*. R package version 1.4-1, <https://CRAN.R-project.org/package=Matrix>.
- **methods 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **nlme 3.1.157** - Pinheiro J, Bates D, R Core Team (2022). *\_nlme: Linear and Nonlinear Mixed Effects Models\_*. R package version 3.1-157, <https://CRAN.R-project.org/package=nlme>.
- **nnet 7.3.17** - Venables WN, Ripley BD (2002). *\_Modern Applied Statistics with S\_*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- **parallel 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **rpart 4.1.16** - Therneau T, Atkinson B (2022). *\_rpart: Recursive Partitioning and Regression Trees\_*. R package version 4.1.16, <https://CRAN.R-project.org/package=rpart>.
- **spatial 7.3.15** - Venables WN, Ripley BD (2002). *\_Modern Applied Statistics with S\_*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/>

[MASS4/](#).

- **splines 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **stats 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **stats4 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **survival 3.3.1** - Therneau T (2022). *\_A Package for Survival Analysis in R\_*. R package version 3.3-1, <https://CRAN.R-project.org/package=survival>.
- **tcltk 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **tools 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **utils 4.2.0** - R Core Team (2022). *\_R: A Language and Environment for Statistical Computing\_*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.