

SOFTWARE

Open Access



# EpiVisR: exploratory data analysis and visualization in epigenome-wide association analyses

Stefan Röder<sup>1\*</sup> , Gunda Herberth<sup>1</sup>, Ana C. Zenclussen<sup>1,2</sup> and Mario Bauer<sup>1</sup>

\*Correspondence:  
stefan.roeder@ufz.de

<sup>1</sup> Department of Environmental Immunology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

<sup>2</sup> Perinatal Research Group, Saxonian Incubator for Clinical Translation (SIKT), Medical Faculty, Leipzig University, Leipzig, Germany

## Abstract

**Background:** With the widespread availability of microarray technology for epigenetic research, methods for calling differentially methylated probes or differentially methylated regions have become effective tools to analyze this type of data. Furthermore, visualization is usually employed for quality check of results and for further insights. Expert knowledge is required to leverage capabilities of these methods. To overcome this limitation and make visualization in epigenetic research available to the public, we designed EpiVisR.

**Results:** The EpiVisR tool allows to select and visualize combinations of traits (i.e., concentrations of chemical compounds) and differentially methylated probes/regions. It supports various modes of enriched presentation to get the most knowledge out of existing data: (1) enriched Manhattan plot and enriched volcano plot for selection of probes, (2) trait-methylation plot for visualization of selected trait values against methylation values, (3) methylation profile plot for visualization of a selected range of probes against selected trait values as well as, (4) correlation profile plot for selection and visualization of further probes that are correlated to the selected probe. EpiVisR additionally allows exporting selected data to external tools for tasks such as network analysis.

**Conclusion:** The key advantage of EpiVisR is the annotation of data in the enriched plots (and tied tables) as well as linking to external data sources for further integrated data analysis. Using the EpiVisR approach will allow users to integrate data from traits with epigenetic analyses that are connected by belonging to the same individuals. Merging data from various data sources among the same cohort and visualizing them will enable users to gain more insights from existing data.

**Keywords:** Visualization, EWAS, DNAm, Profile plot, Shiny application

## Background

DNA methylation has been demonstrated to be highly dynamic and essential for many biological processes. The importance of DNA methylation on biological processes has been shown for gene regulation and cell differentiation amongst others [1, 2]. With the advent of microarray technology in epigenetic research, researchers become supported with a deep and comprehensive overview on changes in DNA methylation applicable for



epigenome-wide association analysis (EWAS). The Illumina Infinium HumanMethylation450 BeadChip (GPL13534 platform) became widely used to measure DNA methylation in roughly 485,000 CpG (cytosine followed by guanine nucleotide) sites [3]. Using the later developed Illumina MethylationEPIC BeadChip (EPIC array, GPL23976 platform) the coverage was expanded to about 850,000 CpG locations [4, 5].

A lot of EWAS have been performed using these techniques, which results in the increased availability of epigenetic datasets [6–12]. Usually, these results become reported and documented as differentially methylated CpG related to a certain trait [12, 13]. In environmental sciences dealing with the exposome [14], a common approach is based on searching for an association between (genome-wide) DNA methylation changes and the internal load on various chemicals (multi-trait analyses). Such an approach evokes a large number of results that have to be screened thoroughly. Common methods applied after calculating regression models are mainly based on filtering the results from corrected p-values. Large scale databases that collect and analyze such data exist [15]; they also allow enrichment as well as network analyses. In addition, approaches based on visualization by heatmap, such as *Complex Heatmap*, are in use [16, 17].

Many R based computational tools for analysis of DNA methylation data have been developed. An overview of existing toolsets can be found in [10]. In recent years online toolsets like EWAS Atlas [12], EWAS Data Hub [18] or EWAS Open Platform [15] emerged. They enable new interaction methods with already existing data, but lack specific interactive annotation. EpiVisR was conceived and developed to fill this gap.

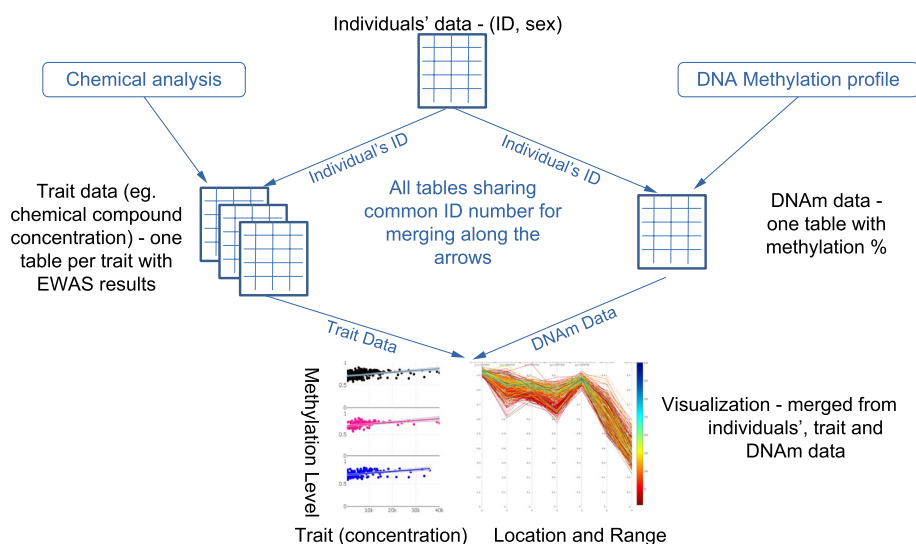
Deeper analysis of results found in EWAS can be tedious and time consuming. To get a faster overview about the biological importance of significant findings in EWAS. We propose EpiVisR, a toolset that integrates visualization as well as annotation and linkage to external data sources for further analyses. It enables the users to inspect results found in an a priori regression process (i.e. from *meffil* package [19]) visually. The visualization described here relies on individuals' data from three domains (individuals' data, trait data, DNAm data) as shown in Fig. 1.

The individuals table of the dataset in use (shown in Fig. 1) provides key information for joining data as well as information for stratification by sex. Trait data becomes merged to DNAm data to show up in various types of visualization. DNAm data itself is shown in the annotated Manhattan and volcano plots of EpiVisR.

## Implementation

EpiVisR provides a visualization module that shows data calculated by a prior calculation step. Due to heterogeneity of the data to be analyzed, the calculation step is implemented outside the EpiVisR package. Therefore, it can be easily adopted to the variable structure needed. Most notable here is the adjustment for cell type proportions as described in [5] and [20]. Also, adjustment for DNA methylation affecting the subject's constitutions such as sex, gestational age, ethnicity, and obesity [21] should be taken into account.

The EWAS model calculation does not necessarily need a visual interface. We recommend using a software package such as *meffil* for model calculation [19]. The calculation module should provide results files with <P\_VAL>, <Beta>, and <DeltaMeth> columns for each CpG. The  $\beta$ -value here is defined as



**Fig. 1** Relationship of trait, DNAm and EWAS analyses; each DNAm probe is available for many individuals as well as each trait; this allows merging of DNAm data with trait data by individuals Id as key attribute; visualization described here is based on this merge step

$$\beta = \frac{\textit{intensity methylated}}{\textit{intensity methylated} + \textit{intensity unmethylated}}$$

for each CpG site [22]. Delta methylation is defined as the difference of methylation between highest and lowest found  $\beta$ -value at each CpG site:

$$\Delta_{\textit{methylation}} = \max(\beta) - \min(\beta)$$

Other definitions for delta methylation (e.g. IQR) might be feasible and depend on the research question as well as on the used toolset for EWAS model calculation. Results should be stored as one file per trait. These files are then used in the visualization module: In a first step, all results from a certain folder are read into a data frame. Minimum as well as maximum values for all descriptive metrics are calculated. These aggregated values can be used for a first selection of trait under investigation out of a possible long list of traits.

Because of the static nature of the results, we implemented a caching strategy. Therefore it is not necessary to load .csv files over and over again. Instead, the data frame with all result summaries is stored in a separate .rds file for later and faster reuse.

EpiVisR was developed as web app on top of the shiny web application framework version 1.7.1 [23]. We used a modularized approach implementing shiny modules for easier maintaining, reuse, and future improvement. Even though the shiny framework offers the concept of reactive programming with seamless updates of related objects, we implemented EpiVisR in an event-driven approach. Otherwise, the users would have to wait at certain processing steps due to large amounts of data that have to be loaded between those processing steps. In consequence, every loading process must be started explicitly by the users.

Data tables in the user interface (UI) were built using the shiny DT module [24]. All plots are built on top of the widely used plotly framework [25]. Therefore, it is possible to drag, to move, and to zoom within the plots as well as to export these plots.

EpiVisR is controlled by a configuration file `<config.yml>`, which is stored in the same folder as the R scripts for visualization. The description of configuration parameters is included in the configuration file itself. Details on how to store data and how to set config file are given in the package vignette on GitHub (see the section “[Availability and requirements](#)”).

EpiVisR subdivides into several components: each component containing a UI as well as the server logic in one R file:

- `<inputTrait.R>` for selection of the folder containing the result files and for selection of trait under investigation;
- `<plotManhattanVolcano.R>` for visualization of enriched Manhattan and volcano plots as well as for the selection of differentially methylated probes. Gene names as well as further genomic features of probes are annotated to probes in plots and tables. These gene names were annotated from the *meffil* R package [19] (where available) and become merged together with delta methylation values as well as p-values from the scenario of the currently analyzed trait under investigation;
- `<plotTraitDNAm.R>` for showing the relationship between trait under investigation and DNAm, stratified by sex;
- `<plotDNAmProfile.R>` for visualization of the genomic region around the differentially methylated probe;
- `<DTCorrelatingProbes.R>` for identification and plotting of probes somewhere on the epigenome that correspond to the probe currently under investigation;
- `<server.R>` containing and referring to the module’s server components;
- `<ui.R>` containing and referring to the module’s UI components;
- `<util.R>` containing several service functions not relating to a particular module.

EpiVisR works with current versions of Firefox, Edge and RStudio integrated browser. It can be installed from GitHub into an interactive R environment (RStudio) using devtools: `devtools::install_github("steroe/EpiVisR")`.

## Results and discussion

Depending on the original research question, various formats are used in different EWAS databases. To distinguish between important and irrelevant CpG (probe) locations, at least a p-value as well as a methylation difference or methylation range (depending on the scale level of the trait under investigation) should be available in the dataset under investigation.

We propose the use of delta methylation values for differentially methylated probes (DMP) in addition to usually reported  $\beta$  values from regression models [21]. They allow selecting candidate probes with a methylation difference above a certain threshold. This enables further validation of findings by other methods than microarray

technology (e.g., pyro sequencing). EpiVisR recognizes these attributes from their column labels: <P\_VAL>, <Beta>, and <DeltaMeth>.

Example data was drawn from the LiNA cohort study [26, 27] and anonymized afterwards.

**Origin of data, preprocessing, and quality checking**

Results from EWAS are usually generated by adjusted regression models and shared in long tables with probeIDs, p-values and delta methylation  $\beta$  for each trait under investigation that was analyzed. Thus, all common preprocessing steps to generate these  $\beta$  values (using *meffil* [19] or similar packages) were done in prior steps. EpiVisR starts with a table containing probeIDs, delta methylation values, and p-values. Unusual small p-values (near lower bound of the float data type in IEEE 754, the technical standard for floating-point computation [28, 29]) can be excluded [18]. Whether this is necessary depends on which system delivered the p-values from regression models. The same applies for models with small case counts (n). Both might be not reliable and therefore omitted.

EpiVisR can exclude multimodal probes from further processing. This works based on a pre-defined list of those probes, which is stored in a file named <MultiMod-ProbesFileName> and referenced in the config file.

All selected probes are internally annotated with information provided by the *meffil* package [19]. This annotation is used in all visualization steps for annotation with gene symbols as well as other genomic features available from *meffil*.

**Selection of traits**

EpiVisR is able to process folders with results from many traits. They can be ordered by various measures such as the p-value, delta methylation, and other graph-based measures. The scatter plot of trait vs. methylation is available at least in the memory of software during runtime while building the regression models during the screening process. From this scatter image it is easy to calculate graph-based measures by using the scagnostics method [30, 31]. If these measures were calculated during the screening process, they can be used to filter and select traits for visual inspection. In case of a huge number of significant findings, these graph-based measures allow filtering for

trait	MaxN	MinP_VAL	MinFDR	MaxBETA	MaxDeltaMeth	MinDeltaMeth	MaxOutlying	MinOutlying	MaxSkewed	MinSkewed
Creas_PARACETAMOL	21	3.6e-75	8.8e-70	-4.2	0.39	-0.34	1.2	0.0	0.98	0.12
Creas_NICOTINE_M_COTININE	19	5.2e-61	1.3e-56	-0.77	0.28	-0.27	0.83	0.0	0.95	0.086
ENDOGENOUS_BIOMOLECULE_220	3.1e+2	1.3e-60	6.4e-55	-0.000017	0.15	1.0e-8	0.48	0.0	0.84	0.51
XYLOMETAZOLINE	17	7.7e-55	6.2e-50	-2.4e-7	0.35	-0.32	1.2	0.0	0.97	0.090
ENDOGENOUS_BIOMOLECULE_285_2	4.1e+2	2.5e-63	1.2e-47	-9.5e-7	0.13	-0.12	0.67	0.0	0.87	0.55
PARACETAMOL	24	5.7e-49	9.3e-44	-0.000029	0.39	-0.21	1.2	0.0	0.97	0.15
MICONAZOLE	15	6.7e-46	5.4e-41	-0.0000024	0.26	-0.27	1.0	0.0	0.99	0.032
NICOTINE_M_N_OXIDE	14	5.9e-35	3.6e-30	-0.0000048	0.38	-0.28	0.80	0.0	0.98	0.021

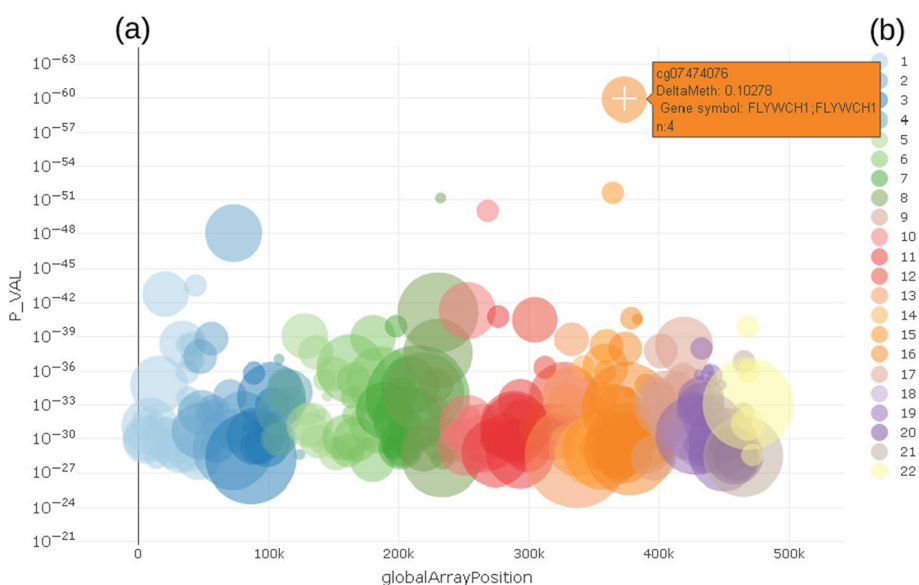
**Fig. 2** Trait selection using a list of filtering criteria; most important are the columns p-value and delta methylation; graph based scagnostics measures are also available

a further criterion before visual inspection of the results. Besides, measures such as the p-value or delta methylation can serve as selection criteria (Fig. 2).

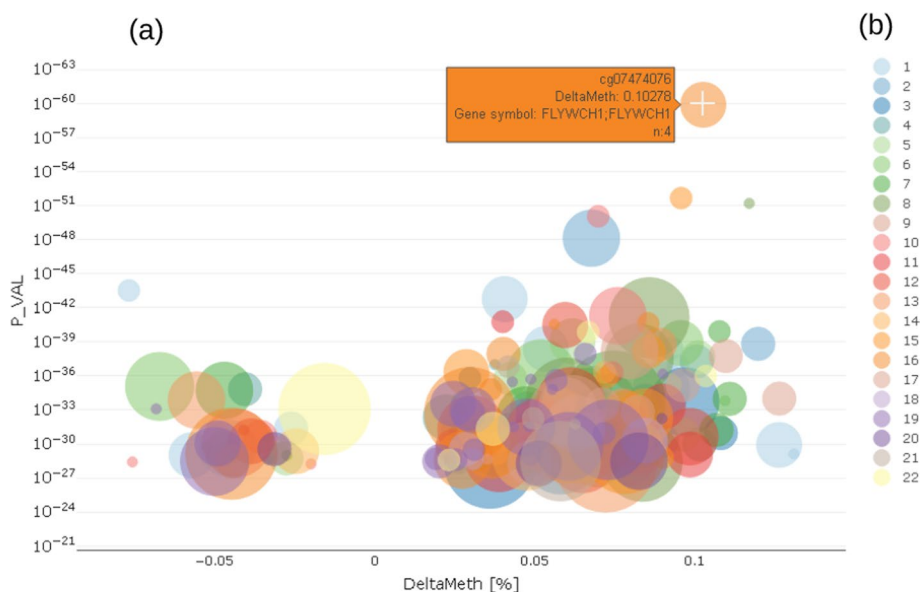
**Visualization**

EpiVisR has a user-friendly design to support the users in finding new insights. It offers the flexibility to use various types of visualization, supporting plots, and showing the relationship of a certain trait under investigation to a single CpG methylation as well as the methylation profile over a certain genomic region. The plots shown here use a data set which was distorted by batch effects for demonstration purposes.

- (i) The enriched Manhattan plot (Fig. 3) shows the  $-\log_{10}$  p-value (P\_VAL) of the regression model for a trait against the location of CpG on the microarray (globalArrayPosition) (Fig. 3). Each dot represents a CpG. The color coding visualizes the chromosome a certain CpG belongs to. To show and select only most important CpG locations, it is possible to limit the number CpG by % of top ranked p-values. Compared to a classical Manhattan plot [32, 33] the enrichment allows to identify important CpG in terms of prior knowledge.
- (ii) In the enriched volcano plot [34] (Fig. 4) the  $-\log_{10}$  p-value (P\_VAL) of the regression model for a trait is shown against the delta methylation measure (DeltaMeth), provided by prior screening algorithm. Exemplarily, Fig. 4 shows a greater variance in hyper-methylated probes (DeltaMeth > 0) than hypo-methylated ones (DeltaMeth < 0), which is represented by a larger group of CpG on the right-hand side of the enriched volcano plot.
- (iii) Both the Manhattan and the volcano plot can be enriched with further information, shown here by the size of the dots. This can be used to show the number of already known effects on a certain probe from an annotation table in order to decide more easily which effects are worth deeper inspection. In the examples



**Fig. 3** CpG selection using enriched Manhattan plot **a** showing p-values of differentially methylated CpG in relation to location on genomic region (chromosome). Color coding **b** shows chromosome location of CpG



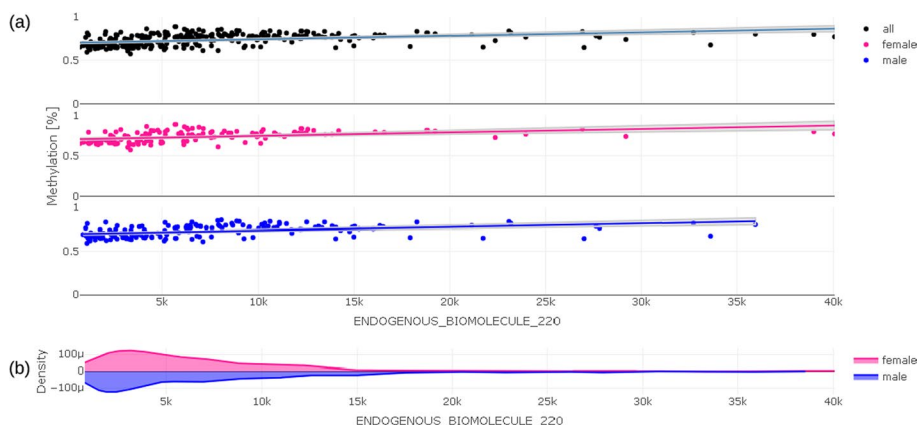
**Fig. 4** CpG selection using enriched volcano plot **a** showing relationship between p-values of differentially methylated CpG and delta methylation of those CpG; left side negative deltas, right side positive deltas; color coding **b** shows chromosome location of CpG

shown here, we used data from the MRC-IEU EWAS catalog [35], showing the number of prior findings concerning a certain probe visualized by the diameter of each dot. From either the Manhattan or volcano plot, the users can select a single probe for deeper inspection of the relationship between the probe and trait under investigation by clicking on a certain dot (probe).

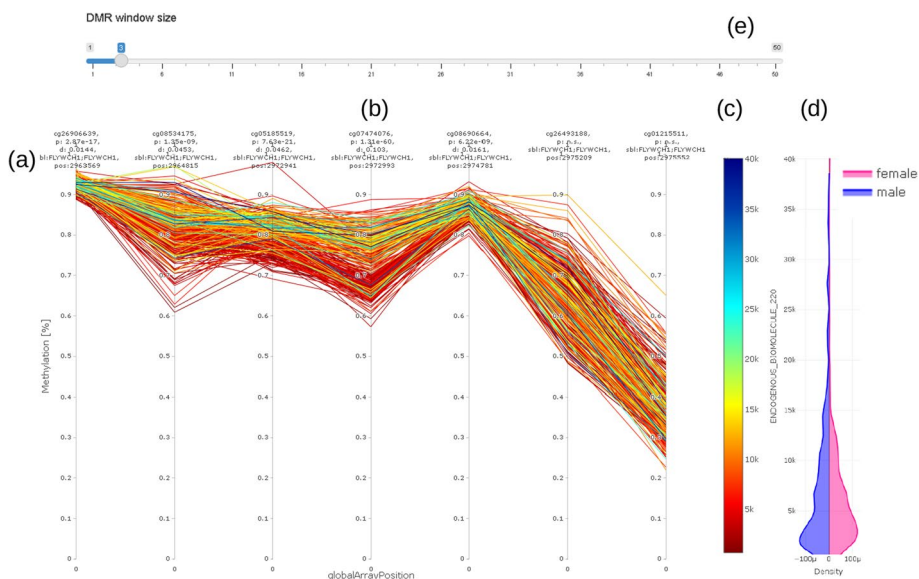
The example shown in Fig. 3 highlights cg07474076, which is located on chr16 in gene FLYWCH1 and has n=4 prior reports in the MRC-IEU catalog. Chromosome coordinates as well as further biological relevant information can be obtained from a table tied with the plots, which holds annotated information as text.

- (iv) In a trait-methylation plot, the range of values in the trait under investigation vs. methylation at a certain CpG location can be shown. Lines represent the trend of trait – CpG relationship together with confidence interval as shadows. The trait-methylation plot can additionally be stratified to other attributes such as <gender>, which are stored in the traits table (Fig. 5). An additional horizontal violin plot shown below the trait-methylation plot visualizes the distribution of sex-stratified trait under investigation.
- (v) The methylation profile shows the methylation of all samples (connected by straight lines) covered by the EWAS on a single selected CpG in the center of the plot and its neighboring CpG (Fig. 6). The value (concentration) of the selected trait under investigation (e.g., chemical compound) is shown on a color scale to the right (c). The length of the visualization window can be adjusted according to the selected probe by using the input bar (a) above the plot (b). Distribution of values is visualized using the density plot (d). Most notably in this example is the clearly visible separation of values with high concentration values in trait under investigation (blueish/greenish lines) from the lower concentrations (reddish/ yellowish lines) over a range of three CpGs from the 3rd to the 5th probe in Fig. 6.





**Fig. 5** Trait-methylation plot (a) showing relationship between delta methylation and value of trait under investigation for all individuals (black), for females (pink), for males (blue); Distribution of values of trait under investigation is given in violin plot (b) below



**Fig. 6** Methylation (DNAm) profile plot (a) showing methylation level (y axis) for each individual over a certain genomic range (globalArrayPosition on x axis) with selected CpG in the center (b); value (concentration) of trait under investigation is color coded (c) (red color for low concentration, blue for high concentration); violin plot (d) on the right hand side shows distribution of trait under investigation stratified by sex of individuals. DMR window size can be adjusted using the slider bar showing selected number of CpG up- and downstream of CpG under consideration (e). For better readability, we do not show real distances between CpG. Distances Information on location, annotation, and effect size is given above the methylation profile plot (a)

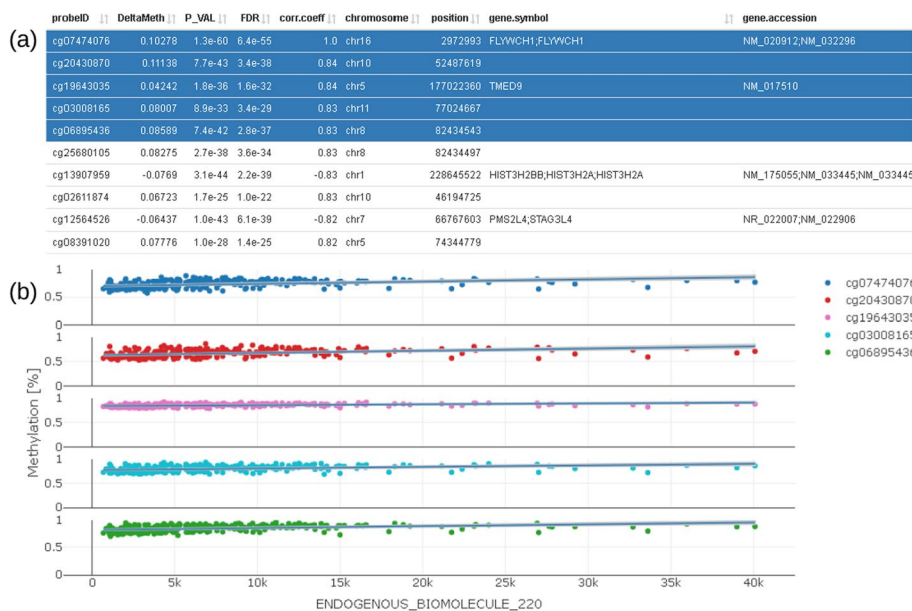
The methylation profile plot does not show real distances between probes. It shows only the CpGs available on the microarray in use (in this example, the GPL23976 platform).

(vi) An additional table with all probes shown in the methylation profile plot can be used for exporting methylation data of selected probes together with trait data into external tools (Fig. 7).



ID_Kind	gender	ENDOGENOUS_BIOMOLECULE_220	cg26906639	cg08534175	cg05185519	cg07474076	cg08080664	cg26493188	cg01215511
w		2.4e+4	0.94	0.88	0.83	0.77	0.91	0.76	0.43
w		5.7e+3	0.95	0.77	0.84	0.89	0.89	0.71	0.36
w		9.7e+3	0.96	0.85	0.79	0.78	0.90	0.62	0.30
w		6.3e+3	0.93	0.90	0.81	0.82	0.88	0.62	0.28
m		1.5e+4	0.94	0.84	0.86	0.76	0.86	0.54	0.26
m		8.3e+3	0.93	0.87	0.80	0.80	0.89	0.66	0.40
m		1.7e+4	0.93	0.89	0.82	0.78	0.85	0.68	0.38
w		6.9e+3	0.96	0.71	0.86	0.75	0.92	0.69	0.26
w		8.4e+3	0.91	0.87	0.84	0.78	0.87	0.63	0.36
m		1.4e+4	0.93	0.84	0.85	0.74	0.90	0.73	0.43
m		2.2e+3	0.94	0.80	0.83	0.74	0.89	0.72	0.43

**Fig. 7** Traits data merged with methylation data; this represents the base data for plotting the methylation profile plot in Fig. 6 together with the selected DMR window and can be used for exporting data to external tools



**Fig. 8** Correlation profile plot (b) showing CpG which are correlated to the selected CpG; correlated CpG are selected in the table on the top (a) and plotted below using different colors

The above example (Fig. 3 to Fig. 6) clearly shows a methylation difference of delta methylation = 0.102 (~ 10%) for CpG cg07474076 between the lowest and highest values of the endogenous biomolecule under investigation.

(vii) We furthermore provide a correlation profile plot for selected CpG and its most correlated CpGs of interest (Fig. 8). The table shows correlation coefficients (corr. coeff) together with annotated CpG information (gene.symbol, genomic feature, etc.). Selected CpGs can be visualized in the plot below the table.

All these visualization tools should help to obtain a fast impression from prior calculated regression models to make trustworthy conclusions. A detailed description of the workflow and interaction with the software is given in the packages vignette on GitHub.

### Interfacing to external tools

All probeIDs are enriched with additional hyperlinks that link to the EWAS data hub [18] and allow further inspection against prior findings related to this particular probe.

Comparison of standard deviation (SD) at a certain CpG location with findings from other studies can also be used to determine whether a DMP is related with a certain trait under investigation or not. Given a DMP was found in a single cohort and no other cohort with the same trait is available, the SD at this particular probe can be calculated and visualized also without information on the trait under investigation. If the DMP does persist in the compared cohort, then there is a high probability that the DMP was not found by chance.

A further important source for identifying useful associations of DNAm to trait under investigation are annotations leading to pathways, linking together functionally associated and differentially methylated probes at various locations besides correlation. EpiVisR supports the export of probe lists into external tools for pathway analysis by compiling together all necessary data (gene.symbol, p-value, delta methylation) from significant probes. In contrast to gene analysis, where the change in gene expression is measured as an 'n-fold change', we use the measure 'delta methylation' for results of EWAS data.

### Strengths and limitations

EpiVisR is solely based on R programming language and is based on a local file-system for storage of results, with no local database required. Both were designed this way to ensure high performance. Simultaneously, this is a limitation since interdependencies between different traits under investigation cannot be systematically recognized.

### Conclusions

EpiVisR enables the users to visualize results from regression models, describing methylation profiles for certain trait under investigation. It furthermore allows inspecting the relationship between trait under investigation and methylation in a visual way in order to produce high-quality plots by using the plotly framework and to generate lists of differentially methylated probes for subsequent external analysis (e.g., pathway analysis).

To the best of our knowledge, EpiVisR is the first implementation that covers the aspects of result visualization, enrichment, and linking to external databases for EWAS with multiple trait under investigation. This in turn will allow easier selection of relevant trait under investigation for further inspection.

### Availability and requirements

Project name: EpiVisR. Download location: EpiVisR can be downloaded from <https://github.com/SteRoe/EpiVisR>. Operating systems: various, platform independent. Programming language: R (version 4.1.2). Additional libraries: *shiny* library (version 1.7.1), *plotly* library (version 4.10.0), *QCEWAS* library (version 1.2–2), *data.table* library (version 1.14.2) [24], *tidyverse* meta package (1.3.1) [36], *RcolorBrewer* library (1.1–2) [37] License: MIT. Restrictions for non-academics use: none.

### Abbreviations

DMP	Differentially methylated CpG probe
DMR	Differentially methylated CpG region
DNAm	CpG methylated DNA
EWAS	Epigenome-Wide Association Studies

SD Standard deviation  
 UI User interface

#### Acknowledgements

Not applicable.

#### Author contributions

MB and SR developed the original idea. SR wrote the software and the manuscript. GH and AZ served as major contributors in writing the manuscript. All authors read and approved the final manuscript.

#### Funding

Open Access funding enabled and organized by Projekt DEAL. Open access funding provided by Helmholtz Centre for Environmental Research – UFZ. This research was financed by intramural funding from the Department of Environmental Immunology of the Helmholtz Centre for Environmental Research – UFZ in Leipzig, Germany and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The funder had no role in the design of the study and interpretation of results.

#### Availability of data and materials

A dataset demonstrating the capabilities of the described software is provided among the software itself on <https://github.com/SteRoe/EpiVisR>.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 8 February 2022 Accepted: 13 July 2022

Published online: 23 July 2022

#### References

- Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science*. 1975;187(4173):226–32.
- Compere SJ, Palmiter RD. DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells. *Cell*. 1981;25(1):233–40.
- Dedeurwaerder S, DeFrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium methylation 450K technology. *Epigenomics*. 2011;3(6):771–84.
- Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the illumina methylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 2016;17(1):208.
- Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol*. 2018;19(1):64.
- Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207–10.
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12(8):529–41.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(D1):D991–5.
- Chadwick LH, Sawa A, Yang IV, Baccarelli A, Breakefield XO, Deng HW, et al. New insights and updated guidelines for epigenome-wide association studies. *Neuroepigenetics*. 2015;1:14–9.
- Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods San Diego Calif*. 2015;15(72):3–8.
- BIOS consortium, Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin Epigenetics*. 2017; 9(1):15.
- Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*. 2019;47(D1):D983–8.
- Liu D, Zhao L, Wang Z, Zhou X, Fan X, Li Y, et al. EWASdb: epigenome-wide association study database. *Nucleic Acids Res*. 2019;47(D1):D989–93.
- Rappaport SM. Implications of the exposome for exposure science. *J Expo Sci Environ Epidemiol*. 2011;21(1):5–9.
- Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res*. 2022;50(D1):D1004–9.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci*. 1998;95(25):14863–8.

17. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847–9.
18. Xiong Z, Li M, Yang F, Ma Y, Sang J, Li R, et al. EWAS Data Hub: a resource of DNA methylation array data and meta-data. *Nucleic Acids Res*. 2020;48(D1):D890–5.
19. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinform Oxf Engl*. 2018;34(23):3983–9.
20. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform*. 2012;13(1):86.
21. Bauer M. Cell-type-specific disturbance of DNA methylation pattern: a chance to get more benefit from and to minimize cohorts for epigenome-wide association studies. *Int J Epidemiol*. 2018;47(3):917–27.
22. Weinhold L, Wahl S, Pechlivanis S, Hoffmann P, Schmid M. A statistical model for the analysis of beta values in DNA methylation studies. *BMC Bioinform*. 2016;17(1):480.
23. Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, et al. shiny: Web Application Framework for R [Internet]. 2021. Available from: <https://CRAN.R-project.org/package=shiny>
24. Dowlle M, Srinivasan A. data.table: Extension of 'data.frame' [Internet]. 2021. Available from: <https://CRAN.R-project.org/package=data.table>
25. Plotly Technologies Inc. Collaborative data science [Internet]. Montreal, QC: Plotly Technologies Inc.; 2015. Available from: <https://plot.ly>
26. Hinz D, Simon JC, Maier-Simon C, Milkova L, Röder S, Sack U, et al. Reduced maternal regulatory T cell numbers and increased T helper type 2 cytokine production are associated with elevated levels of immunoglobulin E in cord blood. *Clin Exp Allergy*. 2010;40(3):419–26.
27. Herberth G, Hinz D, Röder S, Schlink U, Sack U, Diez U, et al. Maternal immune status in pregnancy is related to offspring's immune responses and atopy risk. *Allergy*. 2011;66(8):1065–74.
28. Goldberg D. What every computer scientist should know about floating-point arithmetic. *ACM Comput Surv*. 1991;23(1):5–48.
29. Hough DG. The IEEE Standard 754: One for the History Books. *Computer*. 2019;52(12):109–12.
30. Tukey JW, Tukey PA. Computer graphics and exploratory data analysis: An introduction. *Proc Sixth Annu Conf Expo Comput Graph*. 1985;III(1985).
31. Wang Y, Wang Z, Liu T, Correll M, Cheng Z, Deussen O, et al. Improving the robustness of scagnostics. *IEEE Trans Vis Comput Graph*. 2020;26(1):759–69.
32. Gibson G. Hints of hidden heritability in GWAS. *Nat Genet*. 2010;42(7):558–60.
33. Turner D. S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw*. 2018;3(25):731.
34. Cui X, Churchill GA. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol*. 2003;4(4):210.
35. MRC-IEU EWAS Catalog [Internet]. 2021 [cited 2021 Apr 30]. Available from: <http://www.ewascatalog.org/documentation/>
36. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw*. 2019;4(43):1686.
37. Neuwirth E. RColorBrewer: ColorBrewer Palettes [Internet]. 2014. Available from: <https://CRAN.R-project.org/package=RColorBrewer>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

