Detecting and simulating inheritance of differential methylation

A Lab Rotation report

In partial fullfillment of the requirements for the degree
Master of Science in Computational Biology and Bioinformatics

Hana Pařízková

Supervisors: Deepak K. Tanwar and Prof. Dr. Isabelle Mansuy

October 2019

Contents

Acknowledgment

First of all, I would like to thank Deepak Tanwar for his helpful supervision of this project, and Prof. Isabelle Mansuy for making the project possible and welcoming me into her lab even for such a short time.

Thank you, Mark Robinson and Pierre-Luc Germain, for discussing the ideas on the `metylInheritance` improvements, and thank you, Irina and Martin, for the fruitful discussion about epigenetic inheritance.

Last but not least, I would like to thank all the other members of the Mansuy lab for welcoming me so nicely, as well as to my office colleagues for all their support and encouragement.

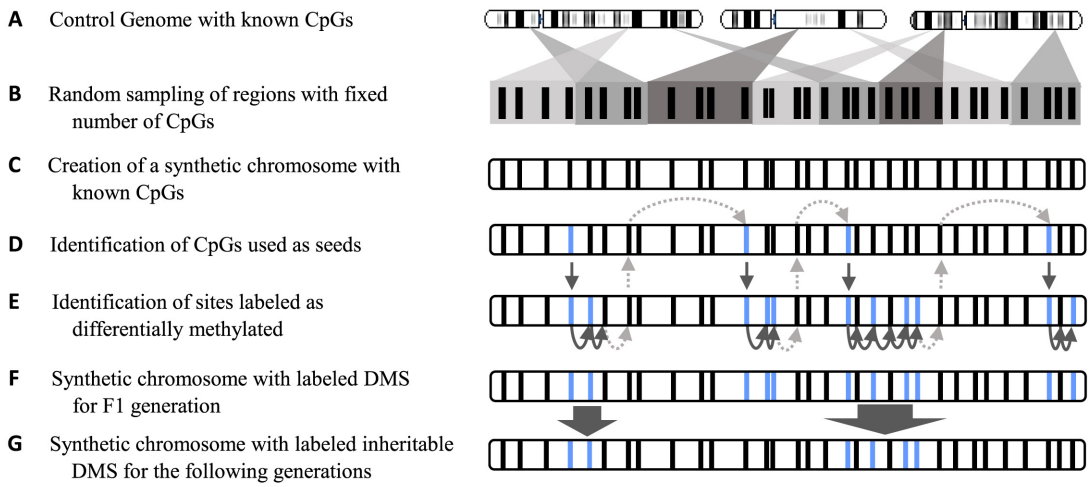


Figure 1: Schematic representation of the methInheritSim method for generating methylated data over several generations. Figure taken from [5].

1 Introduction

DNA methylation is one of the most important epigenetic modifications. In vertebrates, DNA is usually methylated on cytosines in a CpG context, most often in regions rich in CpG dinucleotides, so called CpG islands.

Studies have shown that environmental factors, such as air pollution[1] or cigarette smoking[2], are associated with changes in DNA methylation and subsequent changes in gene expression. Moreover, there is evidence that these changes in methylation level can be transmitted from the parents to their offspring[3][4].

1.1 `methylInheritance` and `methInheritSim` packages

The R package `methylInheritance`[5] provides a statistical framework for testing if a statistically significant trans-generational methylation modification is present in a data set. The input is DNA methylation data from several generations, where only individuals in the first generation were treated. First, differentially methylated regions (DMRs) or differentially methylated sites (DMSs) are identified using `methylKit`[6] in each generation separately. To assess the significance of the findings, a permutation analysis is performed as follows: labels of all samples, i.e. both cases and controls and across all generations, are randomly permuted, DMRs/DMSs are identified and the number of DMRs/DMSs conserved across generations is computed. This is repeated many times to obtain a null distribution of the number of conserved DMRs/DMSs. The observed number of conserved DMRs/DMSs is then compared to this null distribution to obtain a p-value.

In the same paper[5], `methInheritSim`, a package for simulation of methylated data over several generations, was also provided. The pipeline is summarized in figure 1 and explained below:

1. Synthetic chromosome creation: Regions with a fixed number of CpGs in them are randomly selected from the input data provided by the user and combined to create a synthetic chromosome (steps A, B and C in figure 1).

2. Selection of the differentially methylated sites:

   (a) Define seeds: From the last DMS, do a jump of size s, where s is exponentially distributed (step D in figure 1).

   (b) Create the DMR from the seed: Each next CpG within 1000 bp from a previous DMS is called DMS with probability p, p is proportional to $e^{b\log(d)}$, where b is a suitable constant and d distance from the previous DMS (step E in figure 1).

3. Select the inheritable regions: Each DMR is inheritable with a fixed probability (value 0.3 was used in the paper) (step G in figure 1).

4. Assign methylation levels in following generations: For controls and not inheritable regions, the methylation level is beta distributed, $Beta(\alpha, \beta)$, for inheritable DMRs it is beta distributed with a mean shifted by $Diff$,

the effect size: $Beta(\alpha + Diff, \beta)$. The methylation is supposed to be inherited in a Mendelian fashion, i.e. crossing of a treated individual with a control individual leads to offspring being heterozygous for the methylation pattern, and the effect size in heterozygots is half the effect in homozygots.

The `methylInheritance` package was tested on simulated data created by the above described pipeline, and also on real data of male rats exposed in early-life to persistent organic pollutants[5]. For the simulated data, three different effect sizes (0.5, 0.7 and 0.8) and three different sample sizes (6, 12 or 18 individuals per generation, split equally among cases and controls) were tested. As expected, the algorithm works better for data with larger effect size. The sample size seems to have only moderate effect. Unfortunately, the results are not very encouraging: For the smallest (and the most realistic) effect size of 0.5, the authors showed that the power of DMSs detection in the third generation of the simulated data is almost 0, and the false discovery rate (FDR) in the third generation is almost 1. Also, for effect size 0.5, significant epigenetic inheritance was not detected most of the time (see figures 4 and 5 in the paper[5]).

1.2 Summary of the project

In this short project, we have designed a different way of simulating artificial data with inheritance of differential methylation. We have compared our simulation results with the `methInheritSim` package and with one real data set to see whether the simulated data are an accurate representation of the real data. In the end, we suggest how a good simulation could be used to improve the `methylInheritance` package.

2 Methods

2.1 Data simulation – version 1

We decided to design our own simulation procedure based on the `simDMRs` function from `dmrseq` package[7] and the `methInheritSim` procedure described above[5]. The procedure includes the following steps:

1. In the first generation, which is assumed to be the generation directly exposed to the treatment, a given number of DMRs are simulated using functions `simDMRs.RRBS` (for RRBS data) or `simDMRs.WGBS` (for WGBS data). These functions were adopted from the `simDMRs` function from `dmrseq` package.

   For `simDMRs.WGBS`, we start with a set of WGBS data coming only from control samples. These samples are divided into two artificial groups of (if possible) equal size. CpG islands are detected using the `clusterMaker` function from the `bumphunter` package as stretches of genome with high concentration of CpGs. Only CpG islands with at least 5 and at most 500 CpGs are considered further. Given number of CpG islands are then chosen randomly to become differentially methylated regions. For each selected CpG island, the size of the effect is sampled from a beta distribution centered at a specified value (by default 0.3). The effect is then added to the samples in the first group, randomly in up (i.e. more methylation than in control) or down (i.e. less methylation than control) direction, by sampling number of methylated reads from binomial distribution with n equal to the observed coverage and p equal to the mean proportion of methylated reads at a given CpG plus the effect size.

   The only difference in the `simDMRs.RRBS` function is that since not all CpGs are present in the RRBS data, a list of all CpG positions in the given genome must also be provided. The CpG islands are then detected based on this list as described above, positions of DMRs are sampled and the effect is added only to those CpGs that are present in the RRBS data set.

2. The *penetrance* for each differentially methylated site (i.e. each site within a DMR) is sampled from a normal distribution with given mean and standard deviation (by default with mean 0.9 and s.d. 0.1). The penetrance is the probability that a given effect will appear in an individual. In our case, penetrance is the probability that a given DMS will be exhibited in a given individual.

3. When simulating generation i, $i \geq 2$, each differentially methylated site in generation $(i-1)$ (for $i = 2$ every site within a DMR simulated in the first generation) is marked as inheritable with a given probability (0.3 by default). For inheritable sites, each individual is chosen to exhibit differential methylation at site j with probability $\frac{p_j}{2^{i-2}}$ with p_j the penetrance of site j. The denominator represents the expected proportion of individuals heterozygous for site j in generation i, accounting for the 'washing out' of the effect due to crossing with controls: In F2, all individuals have one differentially methylated copy and one normal copy of

site j, in F3 half of the individuals will be heterozygous and half of them will have two normal copies, etc. Here we assume that the methylation is inherited in a Mendelian fashion.

Coverage for each sample at each site is sampled from Poisson distribution with mean equal to the mean coverage at this site in the provided data. The use of Poisson distribution for read coverage was adapted from `WGBSSuite`[8].

Since all individuals in F2 and following generations are heterozygous for the DMSs (assuming Mendelian inheritance), the observed effect size is only half of the effect size observed in the first generation. The proportion by which the effect size changes in heterozygous individuals is controlled by `propHetero` parameter (by default its value is 0.5).

Finally, the number of methylated reads is sampled. For sites that are not marked as differentially methylated in this generation and for individuals which do not exhibit a given DMS, the number of methylated reads is sampled from a binomial distribution with n being the coverage and p being the average proportion of methylation at a given site observed in the data before DMRs were added. For differentially methylated sites, the number of methylated reads is sampled from a binomial distribution with n the coverage and p the average proportion of methylation plus the effect size (adjusted as described in the previous paragraph).

2.2 Data simulation – version 2

After discussion with Martin Roszkowski and Irina Lazar-Contes, PhD students in the Mansuy lab, about what is known and assumed about inheritance of differential methylation, we designed a second version of the simulation.

Again, DMRs are simulated in F1 using `simDMRs.WGBS` or `simDMRs.RRBS` (see section 2.1 for details). A small percentage of the simulated DMRs, specified by parameter `percImprinted` (default 0.02), is denoted as "imprinted". These regions are inherited as a whole. They are inherited in a Mendelian fashion, i.e. the effect observed in heterozygotes is half the original effect size, and the expected proportion of heterozygotes in generation i is $1/2^{i-2}$.

From the rest of the DMSs, each one can be inherited to the following generation with probability `propInherite`. By default, the probability of inheritance from F1 to F2 is 0.3, and from F2 onwards it is 1, i.e. the inheritable DMSs are inheritable forever. For the methylation pattern, we assume that the treated individuals are crossed with controls and that the methylation state at a given CpG is copied randomly either from the mother or from the father. I.e., if an individual has one parent with the site differentially methylated, with probability 0.5 we keep the same effect size (if the methylation pattern is copied from the treated parent) and with probability 0.5 there is no more an effect (if the methylation pattern is copied from the untreated parent). The probability that an individual in generation $i, i \geq 2$ will have a parent with a given site differentially methylated is $1/2^{i-2}$. In total, the probability that a given individual in generation $i, i \geq 2$ will have a given site differentially methylated is $1/2 \cdot 1/2^{i-2} = 1/2^{i-1}$.

All penetrances are supposed to be 1. The coverage, number of methylated and unmethylated reads are sampled the same way as described in section 2.1.

2.3 Comparison with a real data set

To see if our simulation procedures give reasonable results, the simulation results were compared with a real RRBS data set of rats exposed in early life to organic pollutants[5] (GEO accession number GSE109056). This dataset includes four generations – the directly exposed generation and three future generations of offspring.

Makefiles, log files, and results of all following steps are in the `/mnt/IM/RRBS/SRP128759` folder. The pipeline is summarized in figure 2.

The real data were downloaded from GEO (folder `data/raw`) and the quality was checked using FastQC (`20190926-01_FastQC_raw_data`). Then the data were trimmed using TrimGalore with parameters `-q 30 --length 30 --rrbs --paired` (`20190926-02_trimming_raw_data`). The quality was checked again (`20190926-03_FastQC_trimmed_data`) and the data were aligned using Bismark in a paired-end mode (`20190926-04_alignment_Bismark`). Finally, the methylation levels were extracted using Bismark methylation extractor (`20190926-05_methylation_extraction`).

Simulated data were created using the control samples from generation 1 by adding 2000 simulated DMRs, using either the simulation scheme described in section 2.1 (`20191003-08_simulation`), or the one described in section 2.2 (`20191017-16_simulation2`).

Further, simulation function from package `methInheritSim`[5] was used to create another artificial data set. We created one artificial chromosome of 150,000 CpGs (`nbBlock=3000, nbCpG=50`). We simulated four generations, each with three control and three treated individuals. The mean effect size was fixed to 0.3 (`vDiff=0.3`). Other

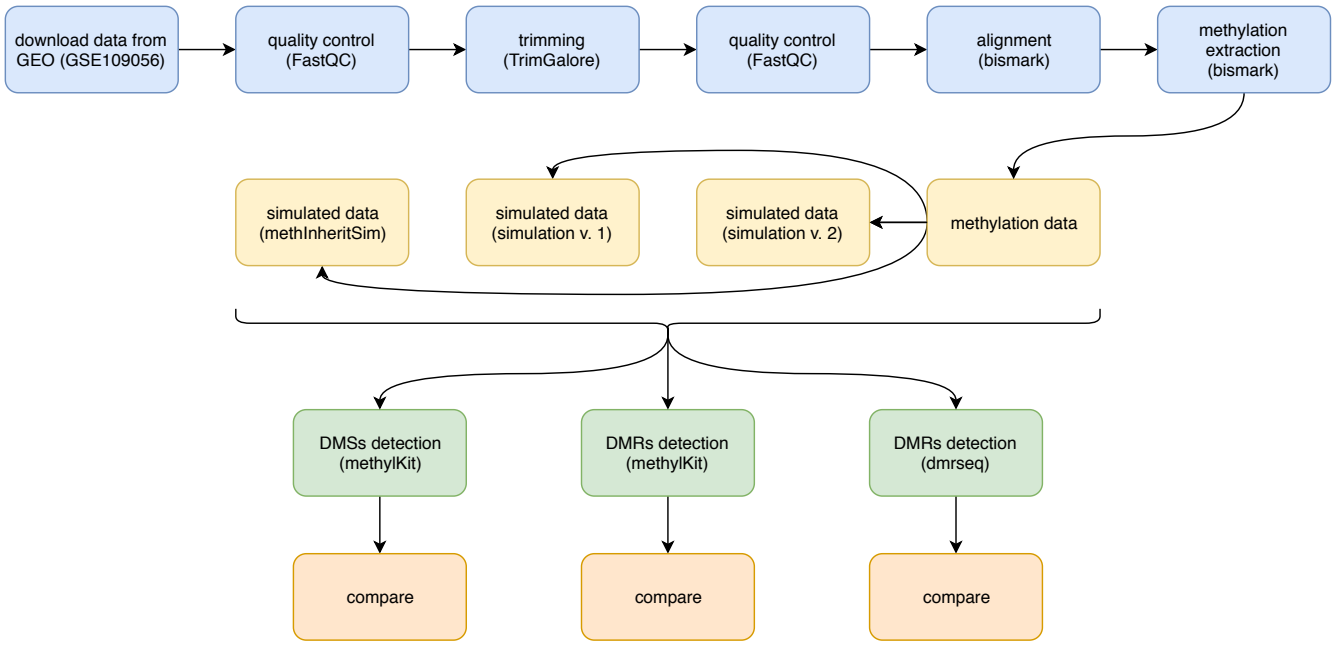


Figure 2: Summary of the data processing and analysis pipeline

parameters were kept to their default values (the simulation is part of the analysis done in folder `20191011-14_methylKit_methInheritSimData`).

DMSs were detected in both real and simulated data in each generation separately using `methylKit` as described in [5], i.e. with `difference=20` and otherwise default values (folders `20191004-10_methylKit_realData`, `20191007-12_methylKit_simulatedData`, `20191011-14_methylKit_methInheritSimData`, and `20191017-17_methylKit_simulatedData_v2`).

DMRs detection was done using `dmrseq` (folders `20191004-09_dmrseq_realData`, `20191007-11_dmrseq_simulatedData`, and `20191011-13_dmrseq_methInheritSimData`) and using `methylKit` (folders `20191017-19_methylKit_regions_realData`, `20191017-20_methylKit_regions_simulatedData`, and `20191017-21_methylKit_regions_methInheritSimData`).

For `dmrseq`, the analysis was performed with three different sets of parameters (see `dmrseq` reference manual[9] for the meaning and default values of the parameters):

- default parameters
- `minNumRegion = 3, minInSpan = 10, cutoff = 0.05, bpSpan = 5000, maxGapSmooth = 10000, maxGap = 5000`
- `minNumRegion = 3, minInSpan = 10, cutoff = 0.05, bpSpan = 10000, maxGapSmooth = 100000, maxGap = 10000`

The non-default parameters try to account for the fact that `dmrseq` was designed for WGBS data, while our data is RRBS. The values of the parameters were suggested by the authors of the `dmrseq` package in issue nr. 14 of the GitHub repository of the package[10], and are actually not crucial for our purpose, since our goal was only to show that the `dmrseq` results on the simulated and real dataset are comparable.

DMRs were also detected using methylKit tiling analysis as described in [5], i.e. with `difference=20` and otherwise default values.

In order to have equal number of samples in all data sets, the detection was done only on three control and three treated samples in each generation in the real data set, although more samples would be theoretically available.

Data set	F1	F2	F3	F4
Real data	2,053	364	1,044	1,014
`methInheritSim`	85,592	1,605	1,050	1,025
our simulation version 1	8,613	799	32	1
our simulation version 2	8,613	1,848	497	158

Table 1: Number of `methylKit` detected DMSs in individual data sets.

Data set	F1	F2	F3	F4
Real data	327	147	421	353
methInheritSim	6804	145	125	122
our simulation version 2	1058	82	49	36

Table 2: Number of `methylKit` detected DMRs (1000 bp windows) in individual data sets.

3 Results

3.1 Comparison of the simulations

3.1.1 Differentially methylated sites

The numbers of reported differentially methylated sites with q-value after multiple testing correction below 0.01 and methylation difference at least 20 % are summarized in table 1. Notice the high number of DMSs in the first generation of `methInheritSim` data, although this data set was actually much smaller than the others – it contained only 150,000 CpGs, while the real and our simulated data contained more than 1,700,000 CpGs. In particular, more than 50 % of all CpGs in methInheritSim data were marked as DMSs in the first generation, while it was only about 0.1 % for the real data set and 0.5 % in our simulated data.

The numbers of DMSs common to multiple generations are depicted using Venn diagrams in figure 3 and as UpSet plots in figure 4. The proportion of common DMSs inherited to the following generation is depicted in figure 5.

In order to see which of the three simulation methods (`methInheritSim`, our simulation version 1 and version 2) produces results with parameters most similar to the real data set, we also plotted the distribution of q-values (figure 6), distribution of effect sizes (figure 7) and distribution of distance to the closest neighbouring DMS (figure 8) for each generation. Since simulation version 2 showed to produce better results than simulation version 1, only results from version 2 are plotted. Similar plots for simulation version 1 can be found in folder `20191014-15_comparison_DMSs_methylKit`.

3.1.2 Differentially methylated regions

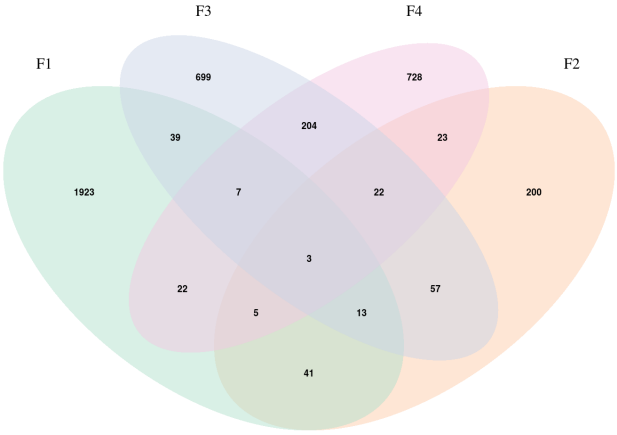
`dmrseq` Unfortunately, no statistically significant differentially methylated regions were identified in the real data using `dmrseq`. Hence, this comparison could not be done.

`methylKit` The regions analysis was from time reasons done only for our simulation version 2.

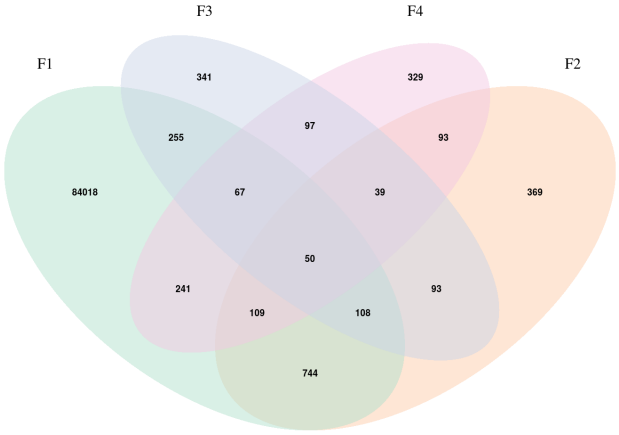
The numbers of reported differentially methylated regions (1000 bp windows) with q-value below 0.01 and methylation difference at least 20 % are summarized in table 2. Notice again the high number of DMRs in the `methInheritSim` data set.

The numbers of DMRs common to multiple generations are depicted using Venn diagrams in figure 9 and using UpSet plots in figure 10. The proportion of F1 DMRs which were also detected in F2 is 2.14 % for the real data, 1.06 % for the `methInheritSim` data and 1.42 % for our simulation version 2. Numbers of DMRs common for more than two generations are however so low, that looking e.g. at proportion of DMRs shared between F1 and F2 that are also present in F3 does not make any sense.

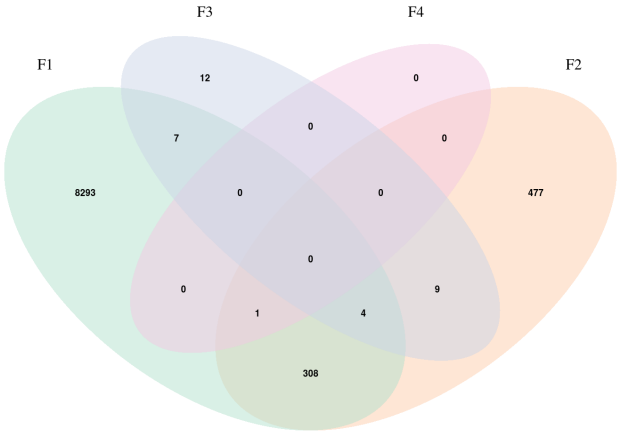
As for DMSs, we again plotted the distribution of q-values (figure 11), effect sizes (figure 12) and distance to the closest neighbouring DMS (figure 13) for each generation.



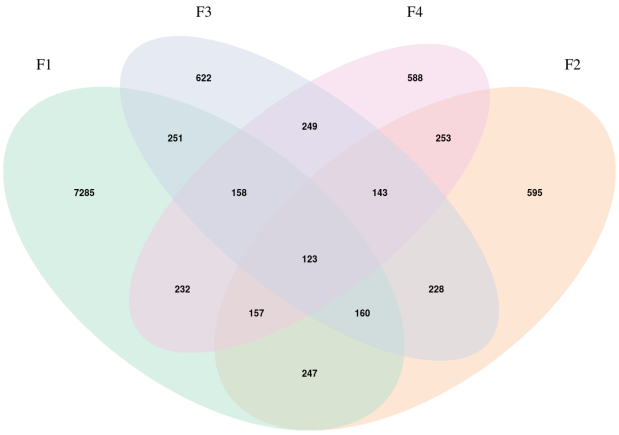
(a) Real data set



(b) methInheritSim simulation

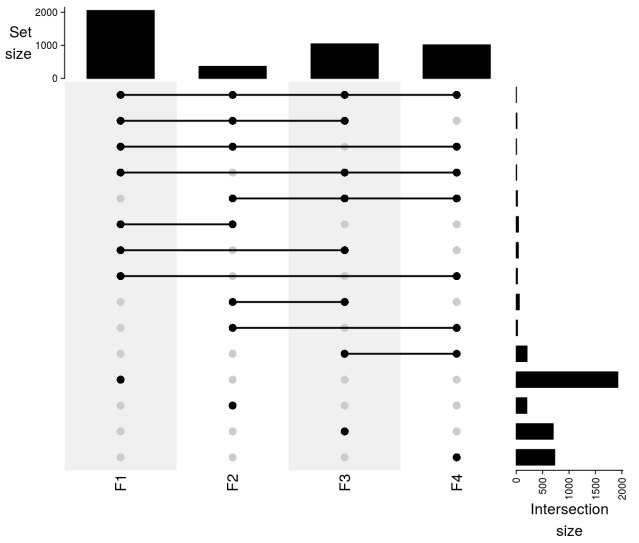


(c) Our simulation, version 1

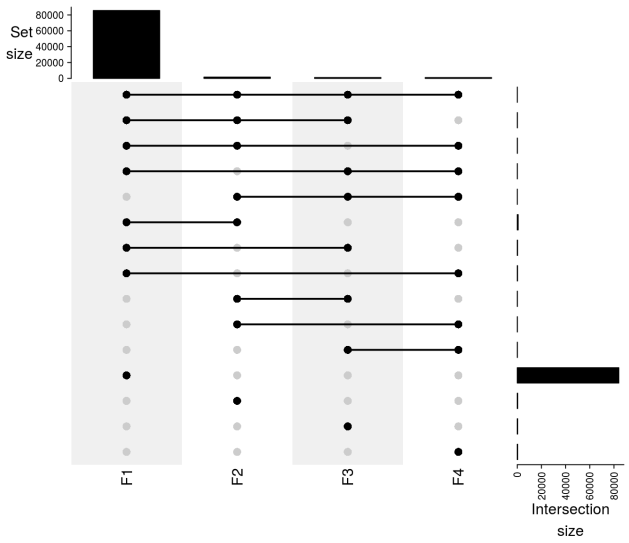


(d) Our simulation, version 2

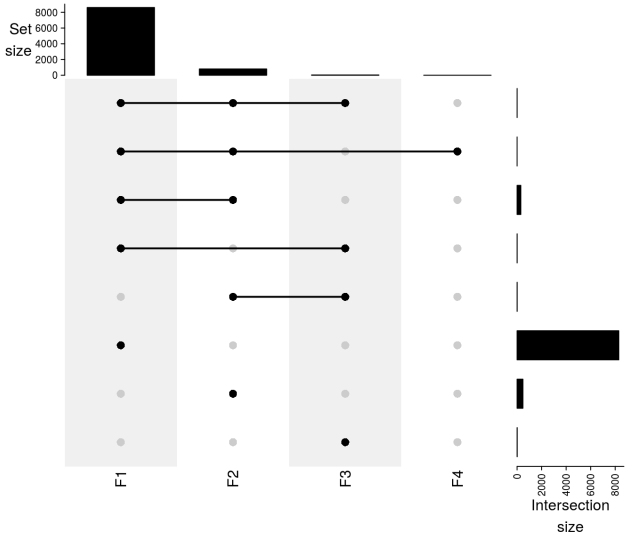
Figure 3: Venn plots of DMSs present in individual generations for all data sets



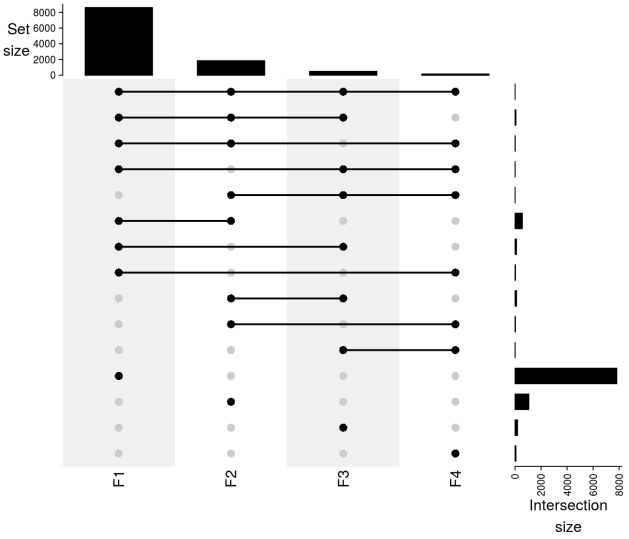
(a) Real data set



(b) methInheritSim simulation



(c) Our simulation, version 1



(d) Our simulation, version 2

Figure 4: UpSet plots of DMSs present in individual generations for all data sets

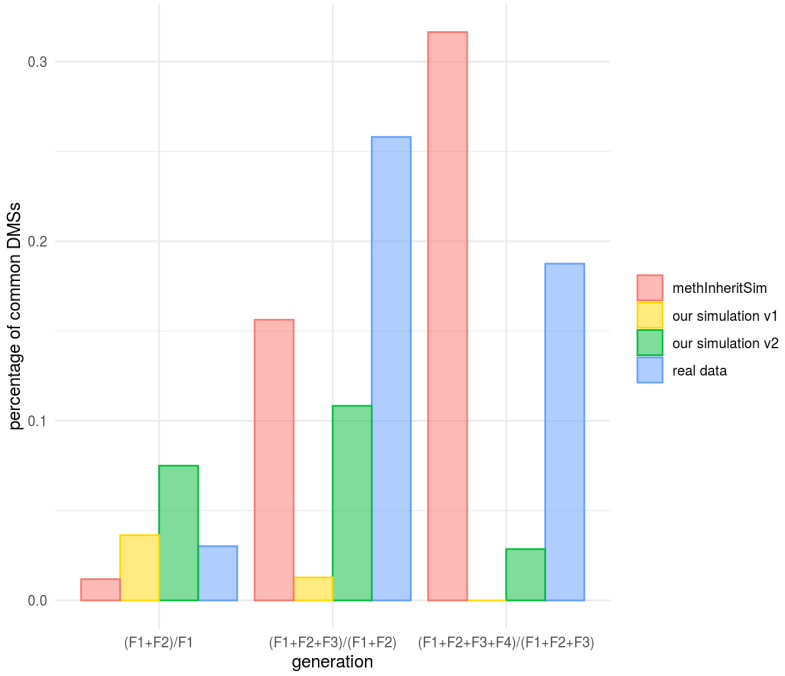


Figure 5: Proportion of DMSs inherited to the following generation in individual data sets. Left – proportion of DMSs present in both F1 and F2 out of all DMSs in F1; middle – proportion of DMSs present in all F1, F2 and F3 out of those present in both F1 and F2; right – proportion of DMSs present in all four generations out of those that are present in all F1, F2 and F3.

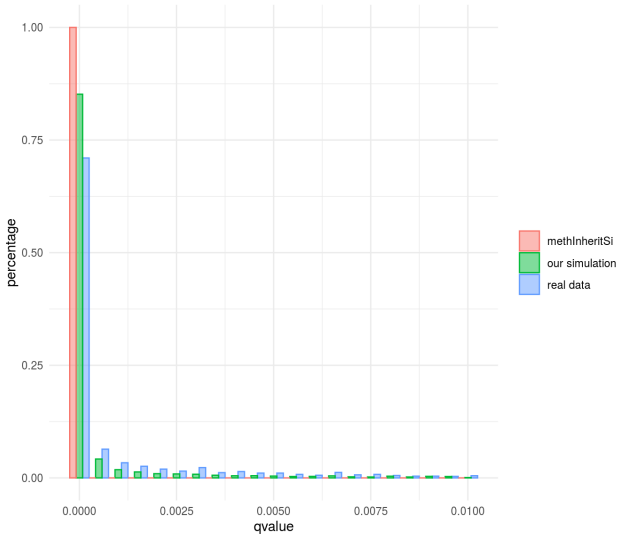
4 Discussion

4.1 Performance of the simulations

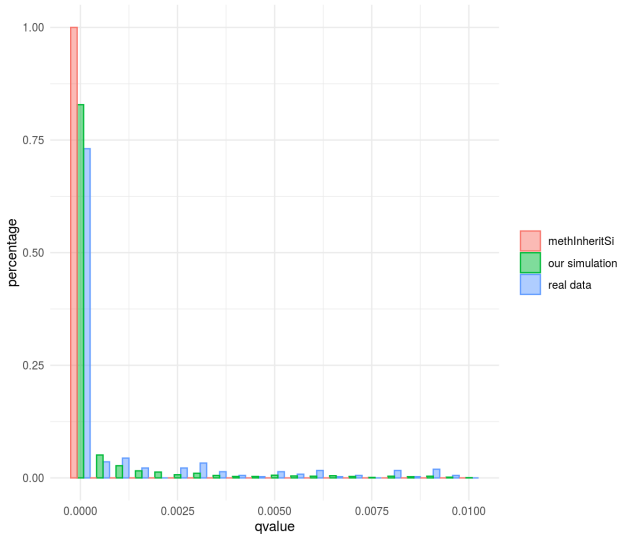
Drawing conclusions from comparison with one real data set is rather dangerous – much more real data sets would be needed to gain confidence that the simulation results correspond well to reality. However, we were not able to find more suitable RRBS or WGBS data sets. (There are some MeDIP-Seq data available (e.g. GEO accession GSE118557), but our simulation, as well as the `methInheritSim` and `methylInheritance` packages are not compatible with MeDIP-Seq data.)

From the comparison with the one available data set it seems that both our simulation and the `methInheritSim` method have certain limitations: Our simulation version 2 produces results for which it holds that for both DMSs and DMRs, the distributions of q-values, effect sizes and distances to the closest neighbouring DMS/DMR in individual generations correspond rather well to the observed distributions. On the other hand, the percentage of DMSs conserved between several generations is quite different. Interestingly, we observe significantly less DMSs conserved for three or four generations than what is observed in the real data (see figure 5), although the default value of the `propInherite` parameter was chosen so that all DMSs that were transmitted from F1 to F2 continue being transmitted to F3, F4, etc. Since all the other parameters (effect sizes, q-values, etc.) seem to be consistent with the real data, it is possible that the solution could be to say, that the methylation pattern is not inherited from both parents with the same probability, but maybe the differentially methylated sites are more likely to be inherited to the offspring. However, it is important to realize that there are a lot of parameters that influence the final result, and also that the phenomenon of epigenetic inheritance is not very well understood, so it is not easy (if not even impossible) to say what causes this discrepancy between our simulation and the real data.

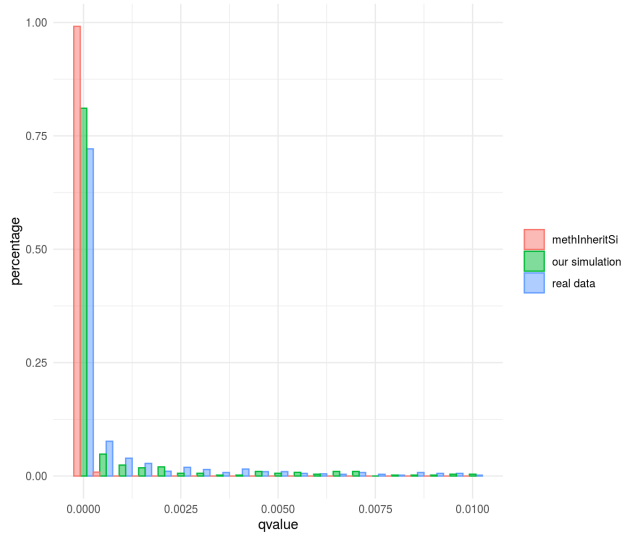
For `methInheritSim` function, some characteristics of the data seem to be rather inconsistent with the real data. The q-values of the simulated DMSs and DMRs are very low (see figures 6 and 11), the shape of the effect size distribution in F1 does not correspond to the real one (see figure 7a) and the DMSs and DMRs seem to be generally in closer proximity than in reality (figures 8 and 13). Also, an obvious weakness of the function is the fact that it creates an "artificial chromosome", which limits the usage of the method. On the other hand, the proportion of the conserved DMSs is more consistent with the real data (figure 5). This is a bit puzzling, as there should not be any differences in the inheritance step between `methInheritSim` and our simulation version 1. A thorough look at the



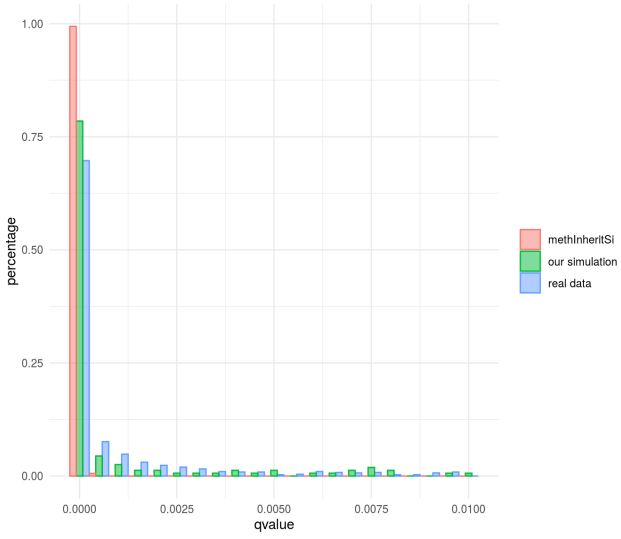
(a) F1



(b) F2

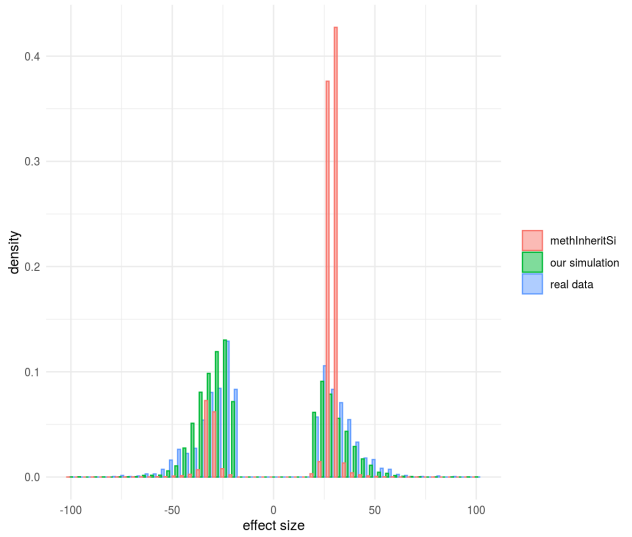


(c) F3

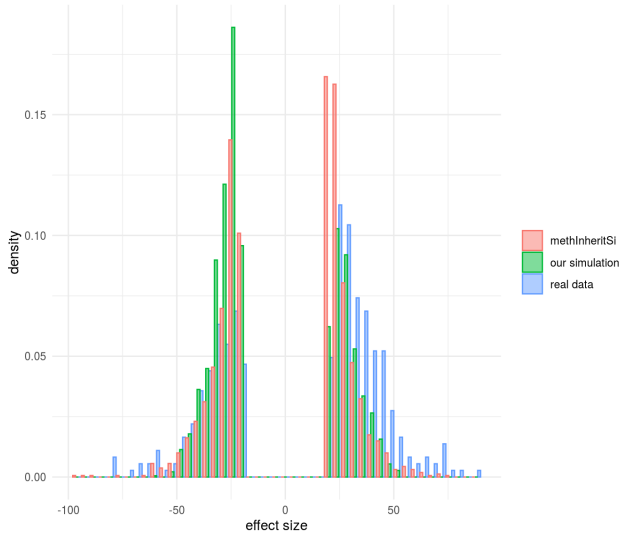


(d) F4

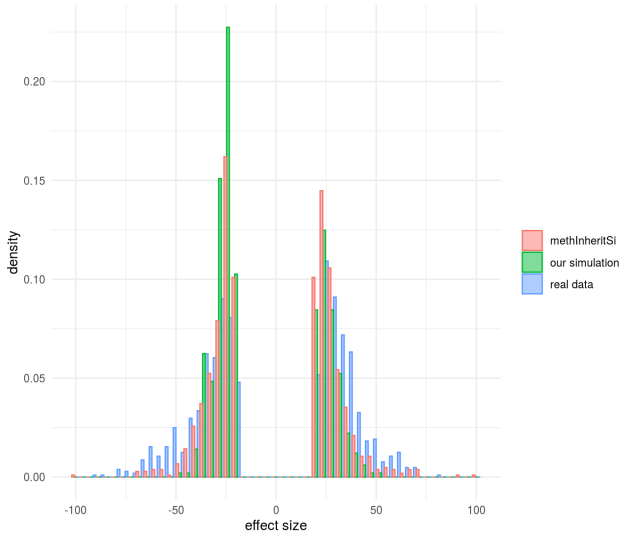
Figure 6: Distribution of q-values of the DMSs in individual generations.



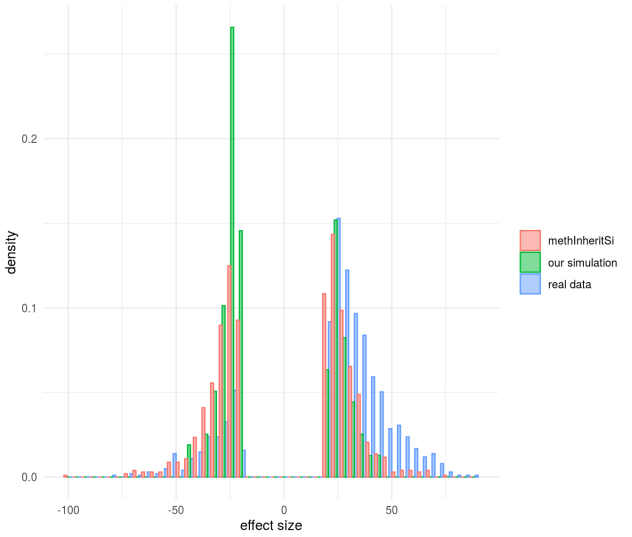
(a) F1



(b) F2

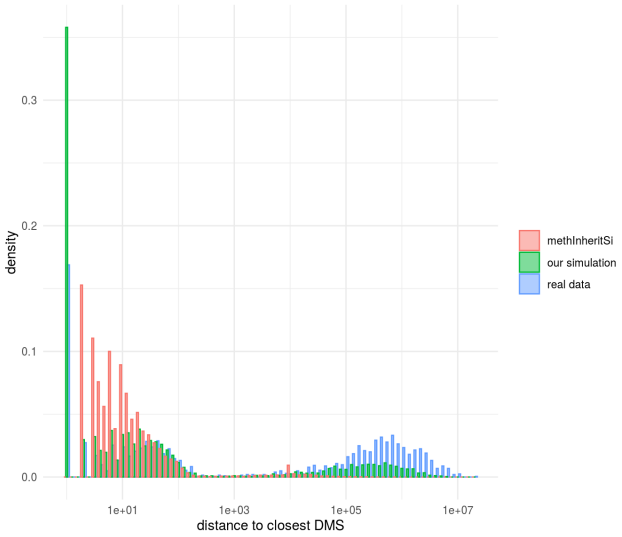


(c) F3

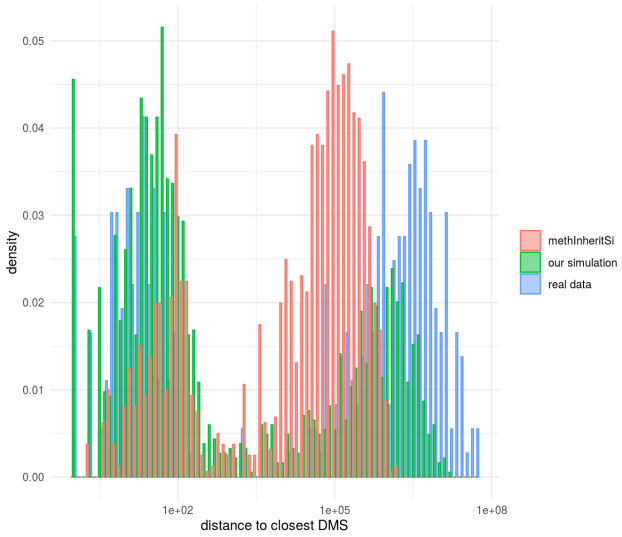


(d) F4

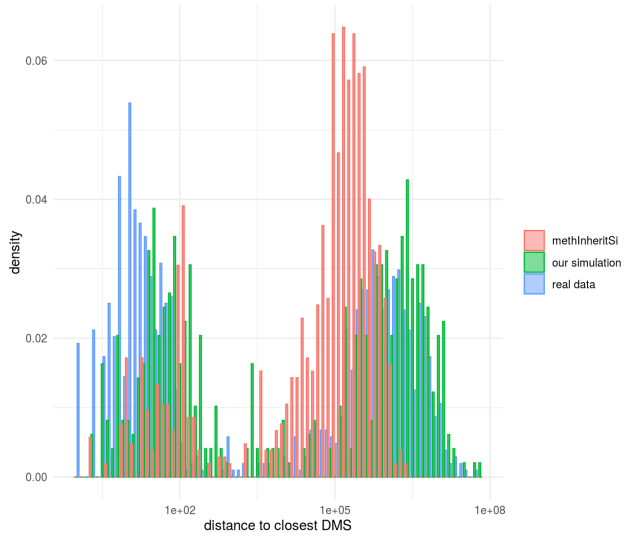
Figure 7: Distribution of effect sizes of the DMSs in individual generations. Notice the empty region between ca -20 and 20 corresponding to non-significant CpGs.



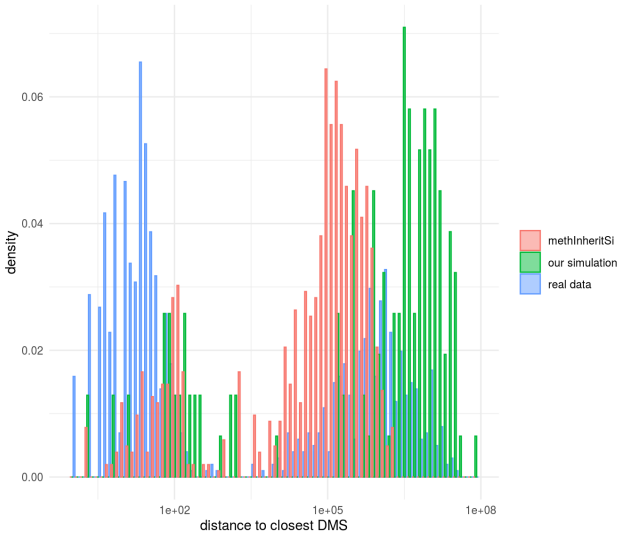
(a) F1



(b) F2

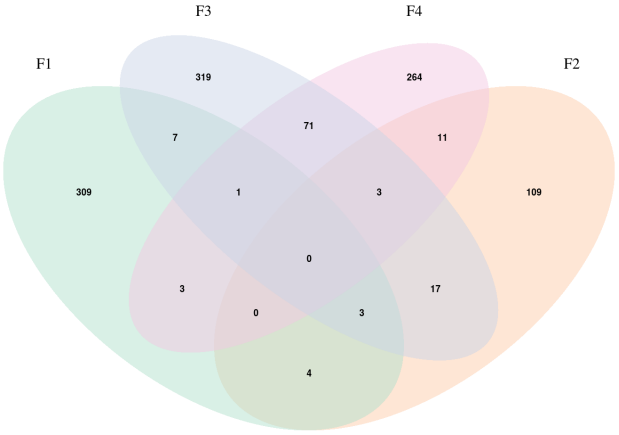


(c) F3

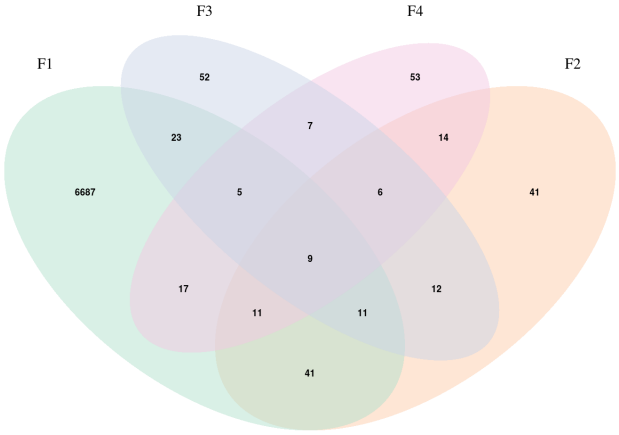


(d) F4

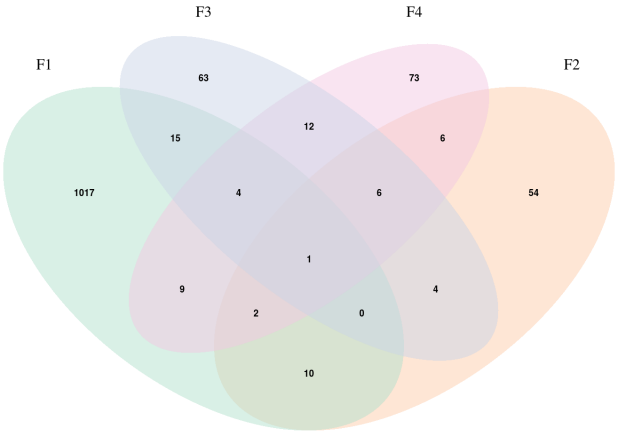
Figure 8: Distribution of distances of the DMSs to the closest neighbouring DMS in individual generations. Notice the logaritmic scale of the x axis.



(a) Real data set

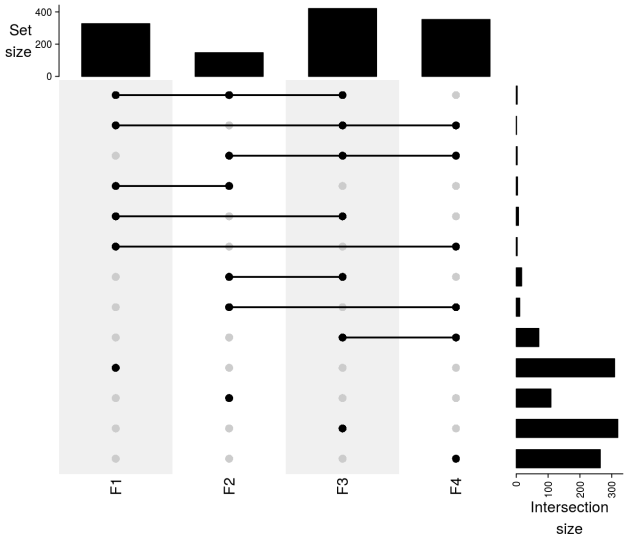


(b) methInheritSim simulation

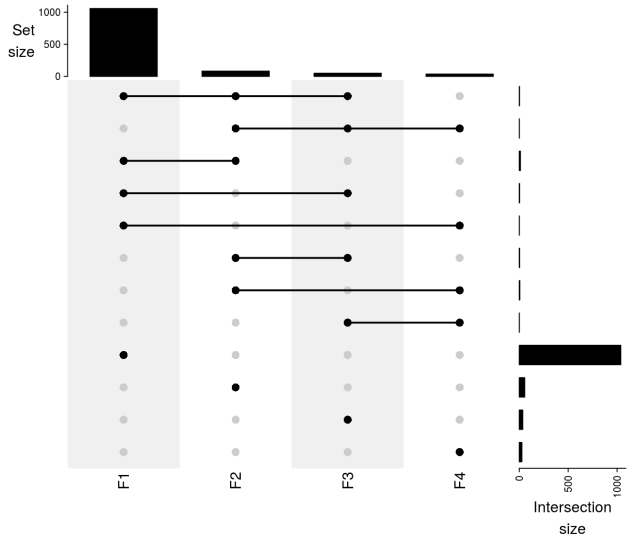


(c) Our simulation, version 2

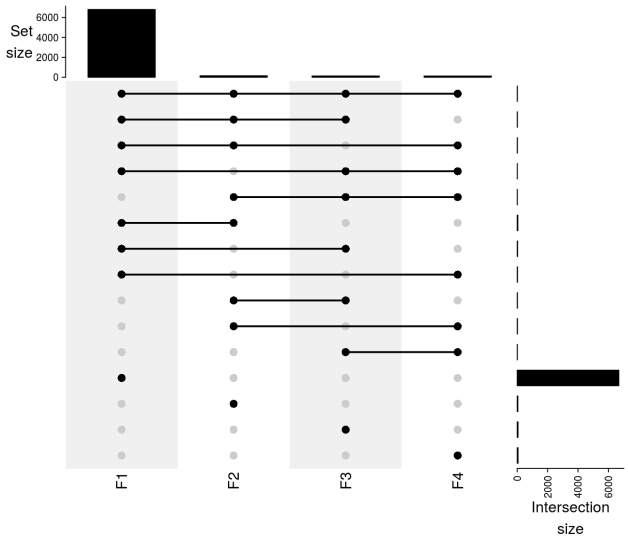
Figure 9: Venn plots of DMRs present in individual generations for all data sets



(a) Real data set

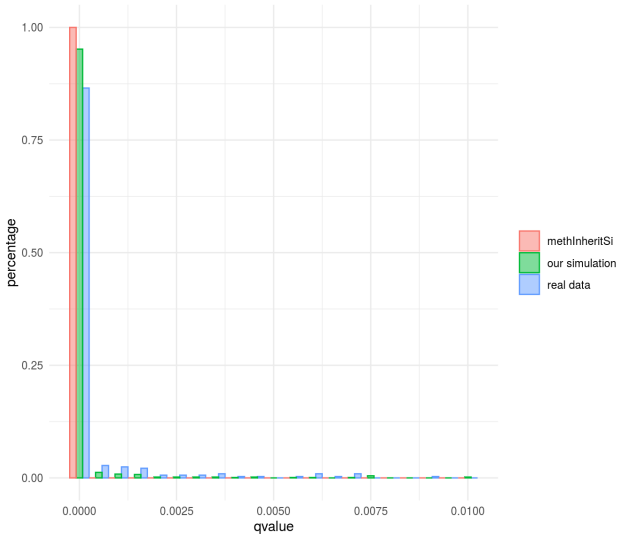


(b) methInheritSim simulation

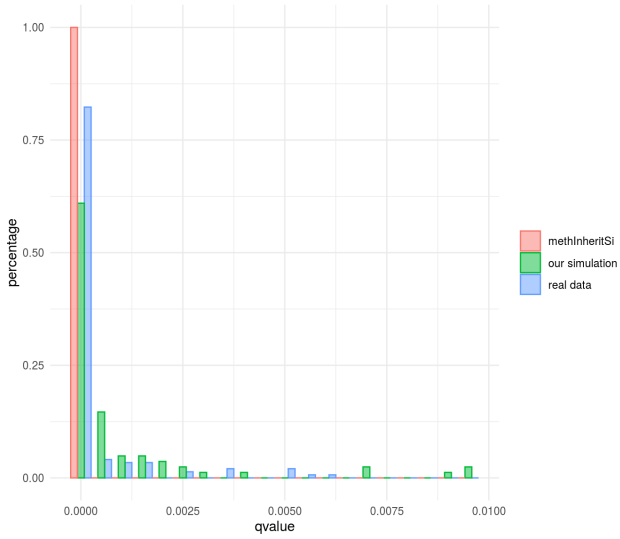


(c) Our simulation, version 2

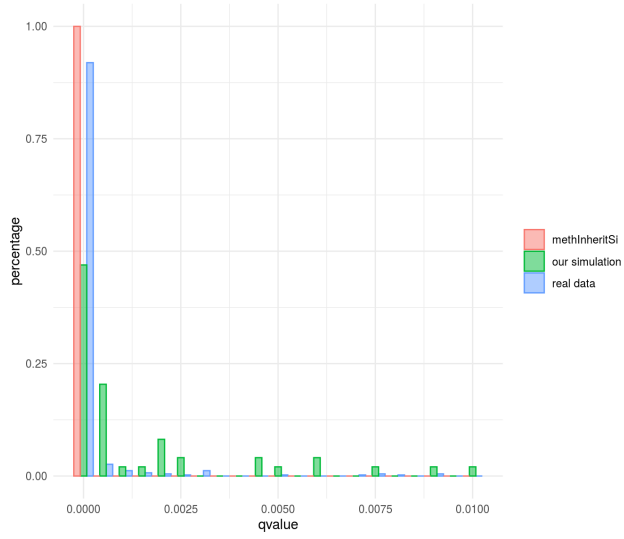
Figure 10: UpSet plots of DMRs present in individual generations for all data sets



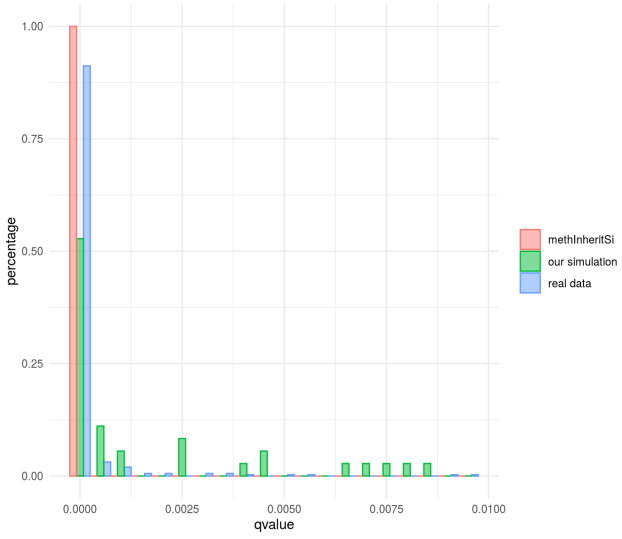
(a) F1



(b) F2

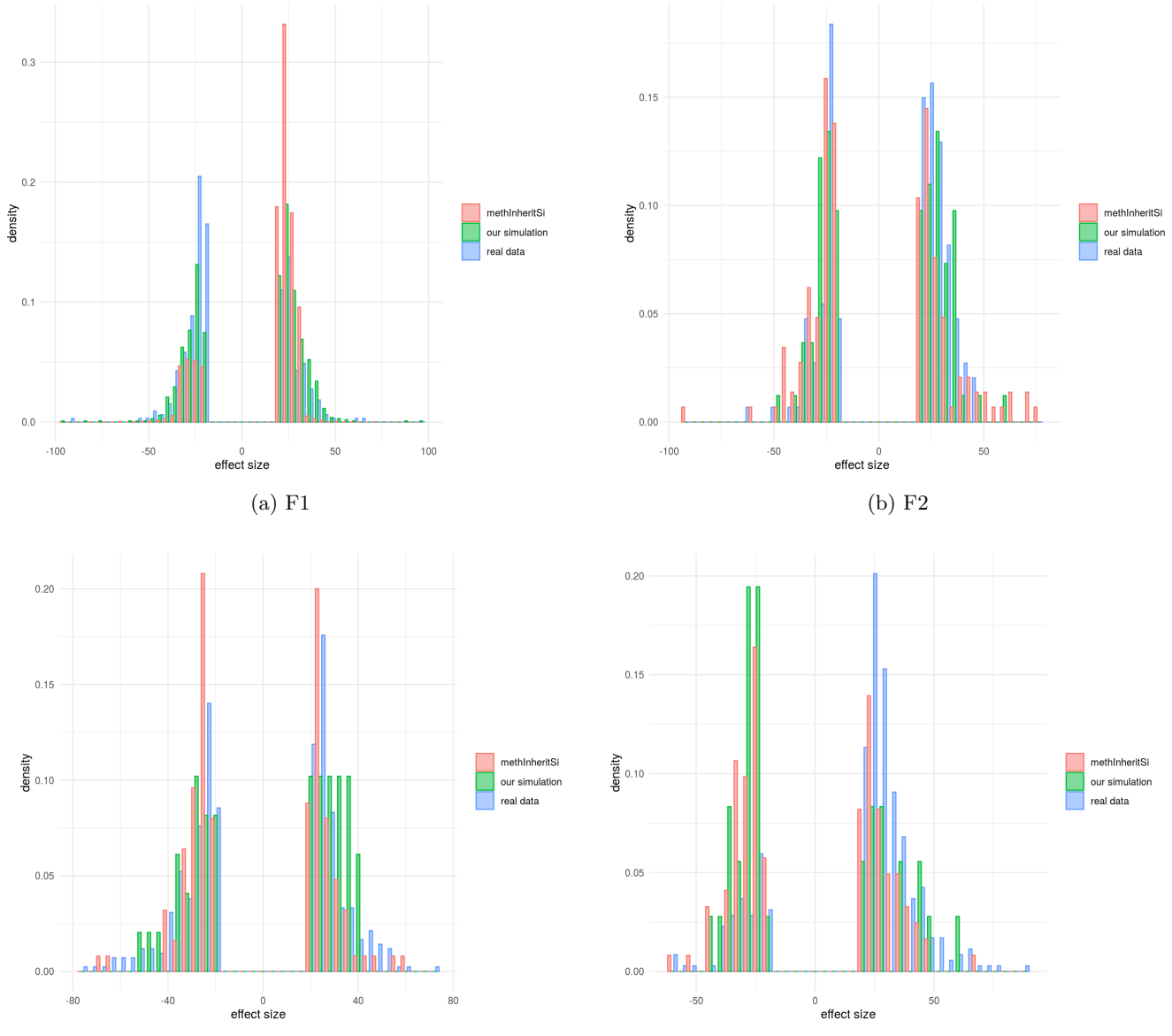


(c) F3



(d) F4

Figure 11: Distribution of q-values of the DMRs in individual generations.



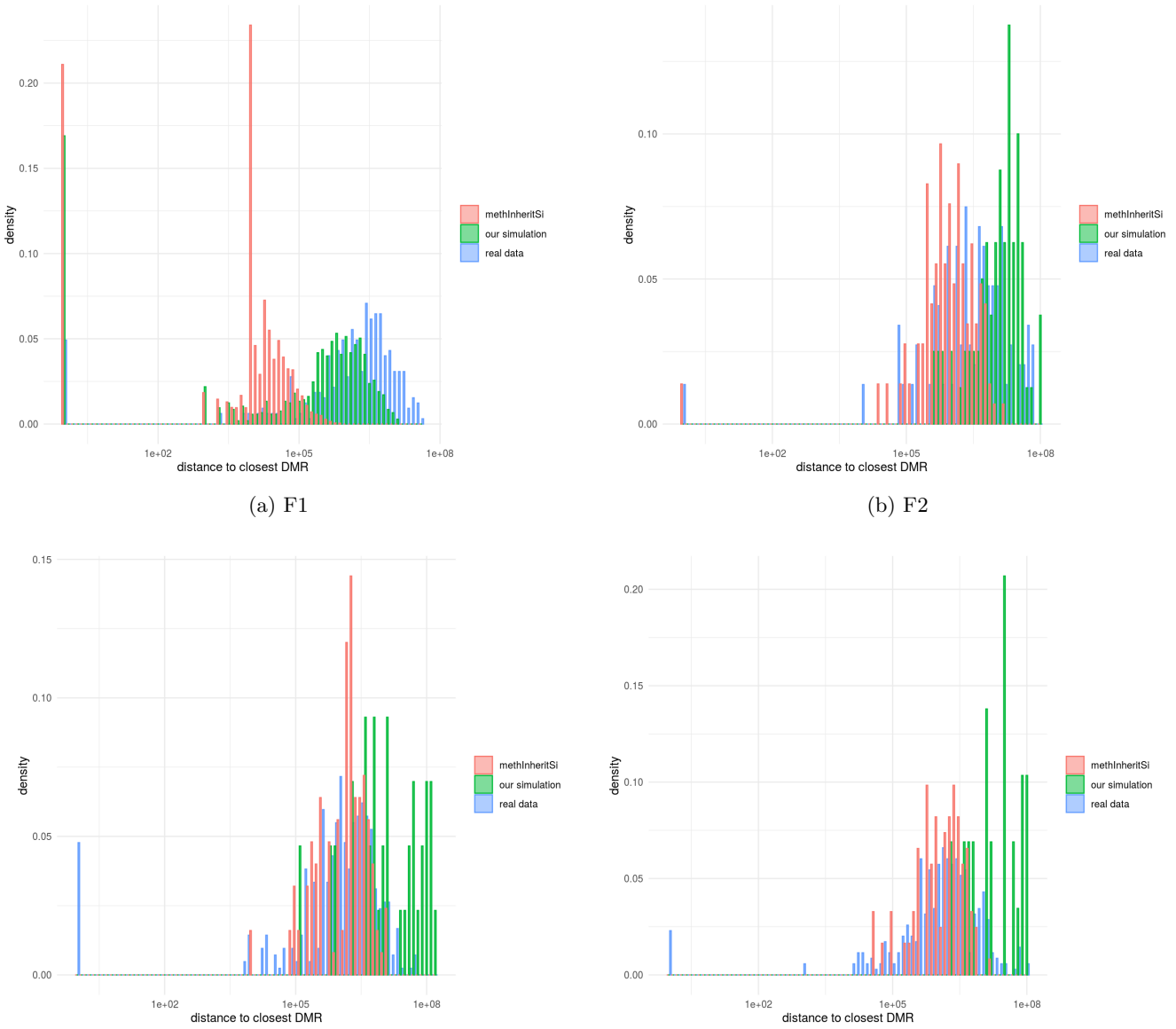
(a) F1

(b) F2

(c) F3

(d) F4

Figure 12: Distribution of effect sizes of the DMRs in individual generations. Notice the empty region between ca -20 and 20 corresponding to non-significant effect sizes.



(a) F1

(b) F2

(c) F3

(d) F4

Figure 13: Distribution of distances of the DMRs to the closest neighbouring DMR in individual generations. Distance 1 corresponds to windows that are right next to each other (`methylKit` does not connect the 1000 bp windows to form larger regions). Notice the logaritmic scale of the x axis.

`methInheritSim` code would be helpful in deciphering where the discrepancy stems from.

4.2 Limitations of the simulations

As all simulations, also our simulation of epigenetic inheritance grossly simplifies the real phenomenon. For example, no functional annotation is taken into account, the sites are considered to be independent of each other, and the population size in each generation is considered to be very large. Also, no new DMSs/DMRs are created after F1, only the current ones are or are not inherited to the offspring, although in reality the DMSs/DMRs in parents could initiate formation of another DMSs/DMRs in the following generations. It will, for sure, be possible to model the process of epigenetic inheritance better in the future, once more details about the process are known. Despite all the limitations, we, however, believe that the presented simulation scheme might be useful.

4.3 Potential `methylInheritance` improvements

The original purpose of creating and testing the simulations was to create a reliable simulated data set which could be used to improve the performance of the `methylInheritance` package. No changes to the `methylInheritance` package were implemented in the end due to lack of time, however, we present here several ideas on improvements of the package:

- **Permutation analysis:** Currently, labels of all the samples, across generations and treatment status, are permuted. However, we believe it might be better if we kept the assignment to generation fixed, and permuted only the treatment labels within each generation. It is also worth considering if the labels of the F1 samples should not be fixed completely, as we might argue that the effect in F1 is fixed and we only want to test if it is transmitted to the following generations.
- **Differential methylation analysis:** Currently, `methylKit` is used for both DMSs and DMRs detection. However, our benchmark from last year (unpublished) showed there are better DMRs detection methods, e.g. `dmrseq`[7]. It is highly probable that including more reliable detection methods would improve the `methylInheritance`'s FDR and power.
- **Analysis on the regions level:** It is known that differential methylation often appears on the level of whole regions. However, if the length of the regions is not fixed (as it is in the case of `methylKit` tiling analysis), it is not straightforward how to implement the `methylInheritance` pipeline on regions level. In particular, number of overlapping DMRs is not a good indicator of the level of inheritance between two consecutive generations, since two short DMRs should not be more important than one long one. Probably, the total length of overlapping regions of DMRs could be used instead of the number of conserved DMSs used on the CpG level.
- **Inheritance of individual CpGs:** Currently, `methylInheritance` only decides if significant level of epigenetic inheritance can be detected in the data. However, for biological purposes, it would be extremely useful if we also knew which particular DMSs/DMRs are significantly transmitted to the following generation. Possibly, a permutation test could be used also for this purpose.

5 Conclusion

This project focused on modelling the phenomenon of epigenetic inheritance. We have implemented our own simulation of the process which should reflect the current knowledge well. We compared the simulation results of our method with results from a simulation method from `methInheritSim` R package, as well as with a real data set. The results of our simulation seem to be consistent with the reality in many parameters, however, improvements are still much needed.

We believe that a good artificial model of epigenetic inheritance could be used not only for benchmarking algorithms and pipelines such as `methylInheritance`, but possibly also for a better understanding of the mechanisms underlying the process.

References

[1] C. V. Breton and A. N. Marutani, "Air pollution and epigenetics: Recent findings," *Current Environmental Health Reports*, vol. 1, no. 1, pp. 35–45, Mar. 2014, ISSN: 2196-5412. DOI: 10.1007/s40572-013-0001-9. [Online]. Available: https://doi.org/10.1007/s40572-013-0001-9.

[2] E. S. Wan, W. Qiu, A. Baccarelli, V. J. Carey, H. Bacherman, S. I. Rennard, A. Agusti, W. Anderson, D. A. Lomas, and D. L. DeMeo, "Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome," *Human Molecular Genetics*, vol. 21, no. 13, pp. 3073–3082, Apr. 2012, ISSN: 0964-6906. DOI: 10.1093/hmg/dds135. eprint: http://oup.prod.sis.lan/hmg/article-pdf/21/13/3073/17255233/dds135.pdf. [Online]. Available: https://doi.org/10.1093/hmg/dds135.

[3] T. B. Franklin, H. Russig, I. C. Weiss, J. Gräff, N. Linder, A. Michalon, S. Vizi, and I. M. Mansuy, "Epigenetic transmission of the impact of early stress across generations," *Biological Psychiatry*, vol. 68, no. 5, pp. 408–415, 2010, Stress, Neuroplasticity, and Posttraumatic Stress Disorder, ISSN: 0006-3223. DOI: https://doi.org/10.1016/j.biopsych.2010.05.036. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0006322310005767.

[4] G. E. Blake and E. D. Watson, "Unravelling the complex mechanisms of transgenerational epigenetic inheritance," *Current Opinion in Chemical Biology*, vol. 33, pp. 101–107, 2016, Chemical genetics and epigenetics * Molecular imaging, ISSN: 1367-5931. DOI: https://doi.org/10.1016/j.cbpa.2016.06.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1367593116300849.

[5] P. Belleau, A. Deschênes, M.-P. Scott-Boyer, R. Lambrot, M. Dalvai, S. Kimmins, J. Bailey, and A. Droit, "Inferring and modeling inheritance of differentially methylated changes across multiple generations," *Nucleic Acids Research*, vol. 46, no. 14, e85–e85, May 2018, ISSN: 0305-1048. DOI: 10.1093/nar/gky362. eprint: http://oup.prod.sis.lan/nar/article-pdf/46/14/e85/25509654/gky362.pdf. [Online]. Available: https://doi.org/10.1093/nar/gky362.

[6] A. Akalin, M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick, and C. E. Mason, "Methylkit: A comprehensive r package for the analysis of genome-wide dna methylation profiles," *Genome Biology*, vol. 13, no. 10, 2012. DOI: 10.1186/gb-2012-13-10-r87.

[7] K. Korthauer, S. Chakraborty, Y. Benjamini, and R. A. Irizarry, "Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing," *Biostatistics*, vol. 20, no. 3, pp. 367–383, Feb. 2018, ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxy007. eprint: http://oup.prod.sis.lan/biostatistics/article-pdf/20/3/367/28852569/kxy007.pdf. [Online]. Available: https://doi.org/10.1093/biostatistics/kxy007.

[8] O. J. L. Rackham, P. Dellaportas, E. Petretto, and L. Bottolo, "WGBSSuite: simulating whole-genome bisulphite sequencing data and benchmarking differential DNA methylation analysis tools," *Bioinformatics*, vol. 31, no. 14, pp. 2371–2373, Mar. 2015, ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv114. eprint: http://oup.prod.sis.lan/bioinformatics/article-pdf/31/14/2371/17086629/btv114.pdf. [Online]. Available: https://doi.org/10.1093/bioinformatics/btv114.

[9] *Reference manual of dmrseq package*, https://bioconductor.org/packages/release/bioc/manuals/dmrseq/man/dmrseq.pdf, Accessed: 2019-10-07.

[10] *Issue nr. 14, dmrseq repository on github*, https://github.com/kdkorthauer/dmrseq/issues/14, Accessed: 2019-10-21.