Methylation profile study of CD14+ monocytes of multiple sclerosis-affected individuals.

Martina Gallinaro [1], Moron Dalla Tor. L[2], Ronchi. J[3], Foti. M[3], Patuzzo. C[1], Malerba.G[1]

[1] Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona
[2] Department of Medicine and Surgery, University of Parma
[3] Department of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza (MB)

Introduction

Epigenetic mechanisms, which include DNA methylation, histone modification, and microRNA (miRNA), can produce heritable phenotypic changes with no changes in DNA sequence. Disruption of gene expression patterns regulated by epigenetics can result in autoimmune diseases, cancers, and various other diseases. Methylation is one of the most studied epigenetic mechanisms known to affect gene expression. It refers to the covalent binding of a methyl group to the fifth position of cytosine residues in the CpG dinucleotide context in mammals. Genome-wide analyses such as microarrays and next-generation sequencing technologies have been used to assess large fractions of the methylome. Several different quantitative approaches have also been established to map the DNA epigenomes with single-base resolution, as represented by the bisulfite-based methods, such as classical bisulfite sequencing. Although the investigation of the role of epigenetic mechanisms in the development of common human diseases first focused for the most part on oncological diseases, more and more of researchers' attention is currently focused on different pathologies, in particular autoimmune and neurodegenerative ones. Identification of the features of the epigenetic regulation characteristic of these pathologies can help in our understanding of the mechanisms of their development and contribute to the creation of new effective therapeutic drugs.

Aim

In this study we analysed 26 CD14+ monocyte samples coming from relapsing-remitting multiple sclerosis (MS) patients and controls. The aim was to estimate the epigenetic profile and investigate differentially methylated regions between cases and controls. Since we noticed that enzymatic fragmentation protocol for sample preparation led to the disruption of the epigenetic profile with almost a total loss of the methylation signals; we hence used physical fragmentation protocol as a second and independent method to evaluate whether the method of DNA fragmentation had an impact on the observed results. After we evaluated differentially methylated loci (DML) among controls and cases MS patients.

Materials and Methods

Sample preparation and sequencing:

In our analysis we worked with 26 samples coming from CD14+ relapsing-remitting multiple sclerosis patients (pool 1). DNA libraries of pool 1 samples were prepared with enzymatic fragmentation using endonucleases cocktail, while in a second round of experiment 8 samples, coming from pool 1, named as pool 2, were subjected to physical fragmentation by using acoustic shearing. After the fragmentation both the pools were subjected to bisulfite conversion, target capturing (CpGiant probe) and then sequenced by Illumina Next-Generation Sequencing platform. 400M total reads output were produced with sequence length of 150bp.

Bioinformatics analysis:

After the production of the raw fastq files the data were pre-processed with fastp software and then the reads were aligned against hg38 human reference genome with bsbolt software. Once the aligned reads are obtained, the methylation extraction step is performed and the methylation values (β values) of each CpG with a single base resolution (CpG report file) are produced.

Methylation profile study:

In order to infer any possible correlation between the methylation signals of the two pools, we searched for a proper cutoff for the β value of pool 1 (β_1) to make pool 1 and pool2 methylation profile as much concordant as possible. We set as hypermethylated (β=1) all the loci in pool 2 having a β value greater or equal to 0.8, and as hypomethylated all the loci in pool 2 having a β value lower or equal to 0.2 (β=0). We then searched for a cutoff for the β_1 value to define a locus as methylated or non-methylated, according to pool 2 profiles (β_2). After having choosen the proper threshold X, we applyed at pool 1 loci the following rule:

if β≥X → β=1 (hypermethylation)

if β<X → β=0 (hypomethylation)

Formula 1

Once the rule is applied, all loci can be labelled as either methylated (β=1) or non-methylated (β=0) and association test can be performed to look for differentially methylated regions between cases and controls (exact fisher test).

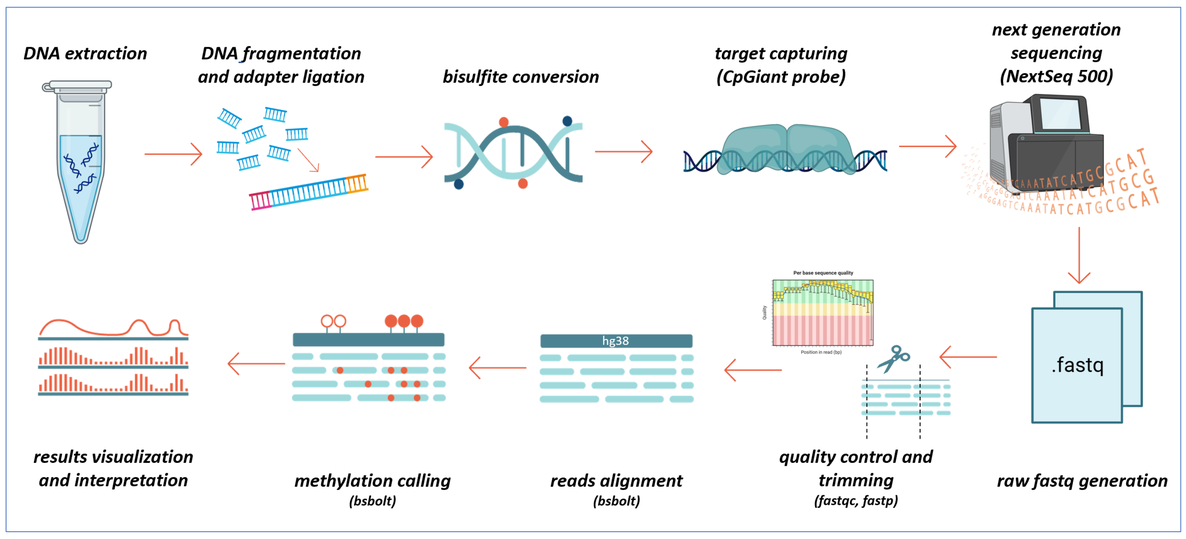

***Fig.1 Schematic representation of the workflow**: DNA libraries were prepared by SeqCap Epi Enrichment System (Roche), then the samples were fragmented with enzymatic (pool1) or mechanical (pool2) fragmentation. Samples were bisulfite converted and target captured, after they were sequenced on Illumina NGS platform (NextSeq 500). After the sequencing, raw fastq files were obtained and the quality was checked with fastQC software. Then the file were trimmed with fastp and the reads were aligned against human reference genome hg38 with bsbolt aligner. Once the reads are aligned, the methylation calling is performed and the β values for each CpG are obtained . In the end β values for each pool are compared for data visualization and interpretation. [Fig. 1 done with Biorender]*

Results

The epigenetic profiles resulting from pool 1 samples show a distribution lacking of many methylation signals(pink curve in **Fig. 2**), with a great abundance of hypomethylated CpG. The epigenetic profiles resulting from pool 2 samples show a bimodal distribution (light blue curve in **Fig. 2**). The measured CpG methylation levels distribution for pool 2 showed the expected results with many loci either fully methylated or non-methylated, so we decided to take the results from pool 2 as our reference for analyses.

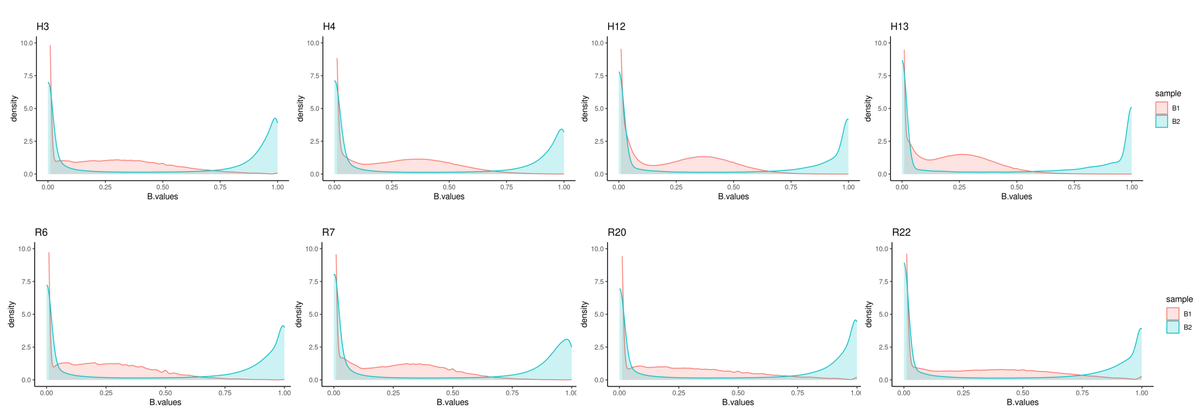

***Fig. 2**: Epigenetic profiles of the eight samples are shown: the β values for pool 1 (β_1) in pink, β values for pool 2 (β_2) in light blue. x-axis: methylation rate (β-values) from 0 (hypomethylation) to 1 (hypermthylation). y-axis: density*

In **Fig. 3** β_2 values for a subset of 500 hypermethylated loci (β≥80%) per chromosome for sample H4 are shown. The β_1 values of the same set of loci are depicted in **Fig. 4**. **Fig. 3** and **Fig. 4** illustrate that loci that should be hypermethylated according to pool 2, instead show a wide range of β values in pool 1, with only few loci in a hypermethylated state. They range from 0 to near 80%. This can suggest us that the treatment with enzymatic protocol led to a loss of the methylation signal in hypermethylated loci in pool 1.

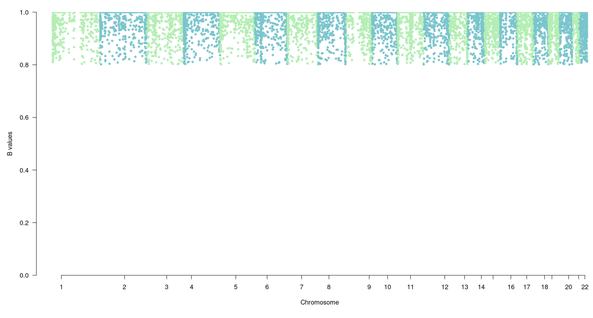

***Figure 3**: β_2 values of CpG sites hypermethylated in sample H4 ($\beta_2 \geq 80\%$)*

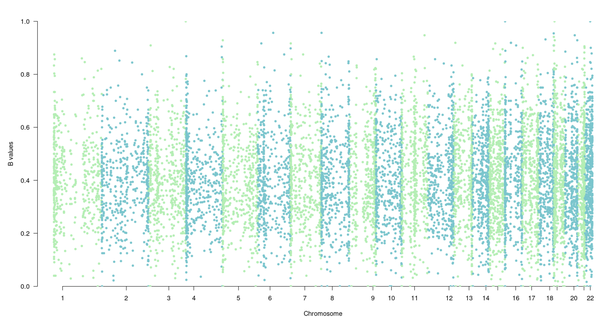

***Figure 4**: β_1 values of CpG sites hypermethylated in sample H4 from pool 2 ($\beta_2 \geq 80\%$)*

According to our results (**Fig.2**), we decided to take pool 2 samples as reference for assign the methylation status to the loci of pool 1 samples, where enzymatic fragmentation affected the methylation signal. To evaluate which cutoff would be more accurate to define a locus as hypermethylated we tested different threshold from 0.01 to 0.22 of β (**Fig. 7** and **Fig. 8**).

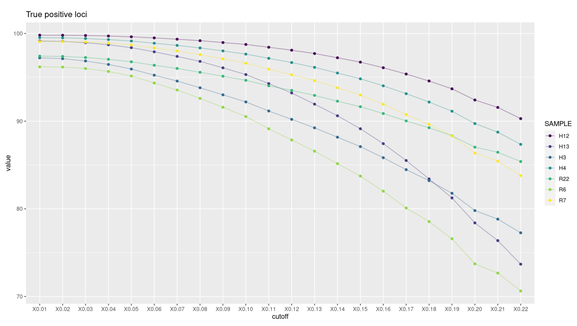

***Fig. 7:** Percentage of true positive methylated loci from pool 1 taking pool 2 hypermethylated loci as reference. x-axis: cutoff. y-axis: percentage of true positive loci*

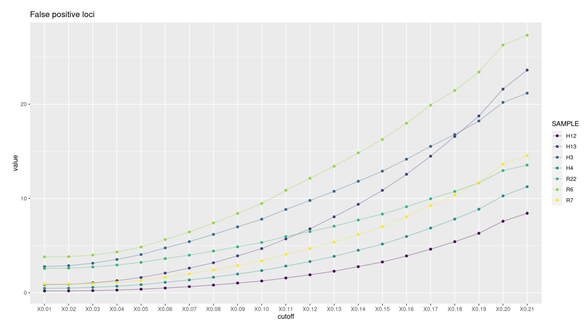

***Fig. 8:** Percentage of false positive methylated loci from pool 1 taking pool 2 hypermethylated loci as reference. x-axis: cutoff. y-axis: percentage of false positive loci*

Table 2 show the 5 genes with the highest number of differentially methylated loci among all the 1.178.239 analysed loci, the p-values are shown in **Fig. 9**, (9.668 with a nominal p-value lower than 0.01) calculated with exact fisher test. These genes mapped all on chromosomes 5 and they are associated with protocadherin proteins and lncRNA.

Gene name	DML	min p-value	max p-value
PCDHA9	29	7,62E-03	9,09E-02
PCDHA5	42	7,62E-03	9,09E-02
PCDHGA3	50	3,37E-03	9,49E-02
PCDHA1	53	2,88E-03	9,09E-02
PCDHGA1	54	3,37E-02	9,49E-02

***Table 1:** Gene name, number of differentially methylated loci, min p-value from exact fisher test, max p-value from exact fisher test*

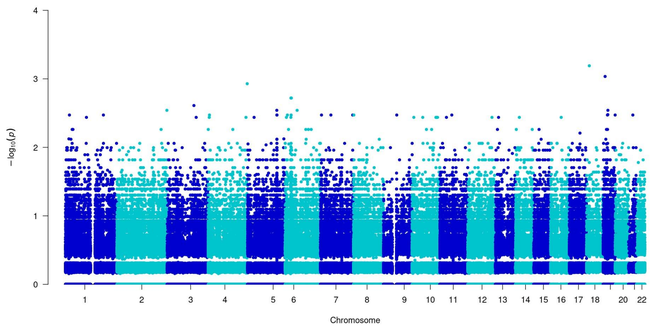

***Figure. 9**: Manhattan plot of p-values for differentially methylated loci*

Discussion and Conclusions

From the obtained results our first conclusion was that the fragmentation methods during the library preparation step affected the methylation signals for what concerned pool 1 samples, as seen from **Fig. 2**. Since we saw that the hypermethylated cytosines (according to pool 2) were mostly detected as hypomethylated in pool 1 sample (**Fig 3** and **Fig. 4**) we decided to infer the correct methylation status of the loci common to pool 1 and pool 2 by testing different cut-off to state which is the threshold that produce less false positive loci. **Fig. 7** shows that a high number of true positive methylated loci (~95%) is reached when a low cut-off is used (β between 0.01 and 0.05), with a low number of false positive loci (less than 5%). Thus, we used a cutoff of 0.04 to applied **Formula 1**. The resulting β values were then tested searched for DML between cases and controls with exact fisher test. Our methods highlighted the presence of DML (**Table 1**) in a set of genes associated with protocadherin proteins. Since these genes are also associated with lncRNA, they could suggest us that gene regulation mechanisms, rather than defects in coding genes and direct epigenetic regulation mechanisms on mRNAs, are important in the aetiopathogenesis of the disease. Further studies need to be done to investigate the role of these DML and the epigenetic mechanism in the multiple sclerosis.

References

- Kiselev IS, Kulakova OG, Boyko AN, Favorova OO. DNA Methylation As an Epigenetic Mechanism in the Development of Multiple Sclerosis. Acta Naturae. 2021 Apr-Jun;13(2):45-57. doi: 10.32607/actanaturae.11043. PMID: 34377555; PMCID: PMC8327151.
- Citation: Zhang L, Gu C, Yang L, Tang F, Gao YQ(2017) The sequence preference of DNAmethylation variation in mammalians. PLoS ONE12(10): e0186559. https://doi.org/10.1371/journal.pone.0186559.
- Feng, L., Lou, J. (2019). DNA Methylation Analysis. In: Zhang, Q. (eds) Nanotoxicity. Methods in Molecular Biology, vol 1894. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-8916-4_12

martina.gallinaro@univr.it