# Report on Methylation Data Analysis

Group 4: Daniel Soto and Laura Sudupe

## Introduction

The study aims to understand the epigenetic and transcriptomic remodeling in monocytes of severe COVID-19 patients, comparing the observed changes to healthy donors. The focus is on DNA methylation alterations and their association with cytokine production and immune response.

## Materials and Methods

Public Data Accession: GSE188573
Organism: Homo sapiens
Experiment Type: Methylation profiling by array
Platforms Used: Infinium MethylationEPIC

The study involved monocytes from 48 severe COVID-19 patients and 11 healthy donors. The DNA methylation profiling was carried out using the Infinium MethylationEPIC array.

## Data Processing Pipeline

- Quality Control (QC): The initial step included quality assessment of the methylation data, ensuring reliable measurements.

- Differential Methylation Position (DMP) Analysis: The following step focused on identifying specific CpG sites that exhibited significant methylation differences between the patients and healthy donors.

## Bioinformatics Tools and Resources

Bioconductor packages: Used for the analysis of methylation array data.
Genome databases: Utilized for annotation and interpretation of methylation sites.

## Results

1. Quality Control

This analysis involves a comparative review of DNA methylation signals between two distinct groups: healthy donors (HD) and patients with severe COVID-19. Detection p-values are statistical measures that help distinguish genuine methylation signals from background noise,

and the bar plot visualizes these values across all probes used in the study. Lower p-values are indicative of a more reliable detection of methylation.

The bar plot compares the mean detection p-values for each group. The green bars represent the mean detection p-values for probes from healthy donors, while the orange bars correspond to those from severe COVID-19 patients. Each bar reflects the aggregation of p-values for probes within each sample, providing insight into the overall data quality and signal reliability from the methylation profiling process. The uniformity and distribution of the bars within each group show a summary of the assay performance and suggest potential variations in methylation patterns between healthy individuals and severe COVID-19 patients.

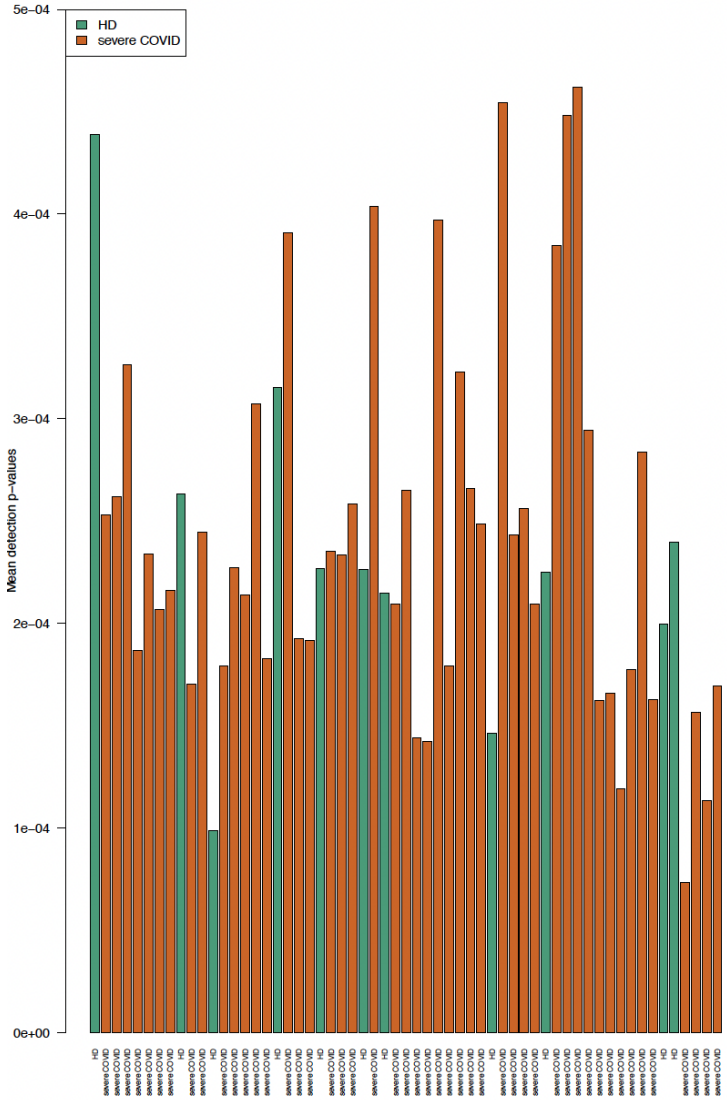


Figure 1: Mean detection p-values summarise the quality of the signal across all the probes in each sample

This analysis illustrates the comparison of the methylation profile between healthy donors and severe COVID-19 patients, both before and after normalization. The density plots depict the distribution of methylation beta values across all CpG sites within the samples.

The pre- and post-normalization density plots validate the normalization process, essential in ensuring that downstream analyses are not confounded by technical variation. The plots confirm that the methylation status of the two groups is similarly distributed after normalization, allowing for a more accurate biological comparison.

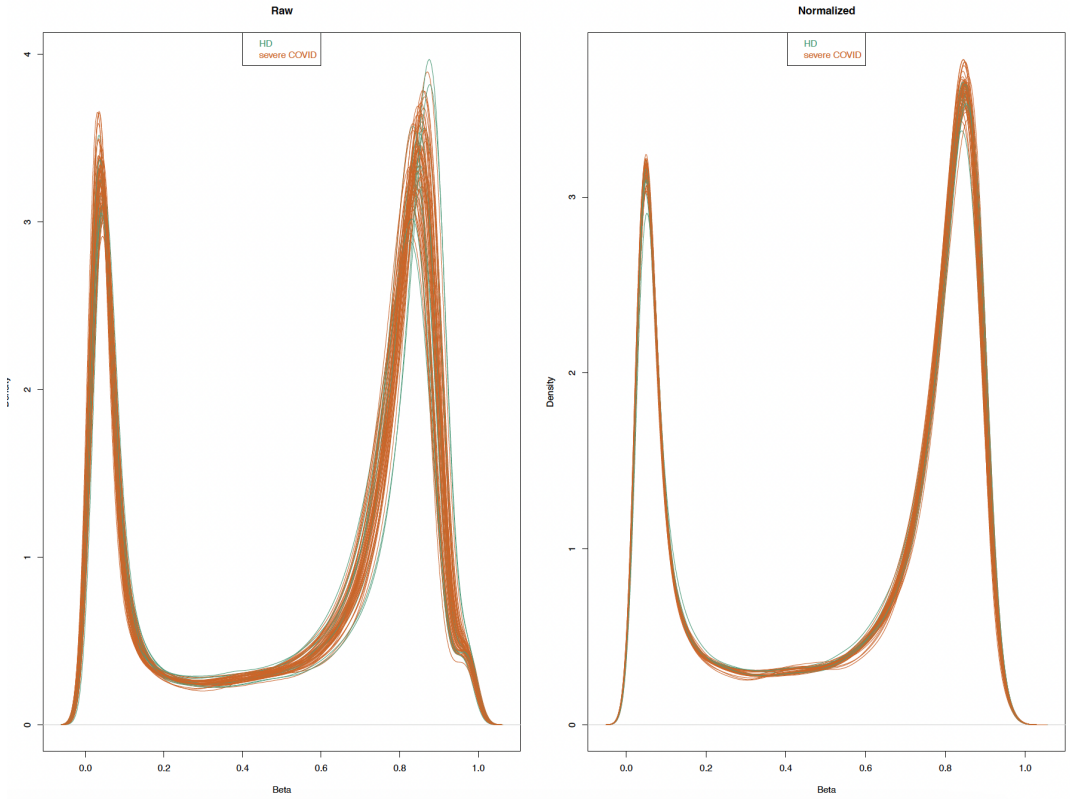


Figure 2: Comparison of methylation beta value distributions in healthy donors (HD) and severe COVID-19 patients before (Raw) and after (Normalized) data normalization, indicating successful adjustment for technical variance while preserving inherent methylation patterns.

The PCA presents a multi-dimensional scaling (MDS) plot, comparing the methylation data of healthy donors (HD) and severe COVID-19 patients. MDS is a means to visualize the level of similarity of individual cases of a dataset, with the plot demonstrating the relative clustering of the two groups along the principal components, which are the axes accounting for the most variation in the data.

The MDS plot reveals the intrinsic methylation variability between the two groups, underlining the distinct epigenetic landscape characteristic of severe COVID-19 cases compared to healthy individuals. This clustering could reflect differences in immune response, treatment effects, or disease progression.

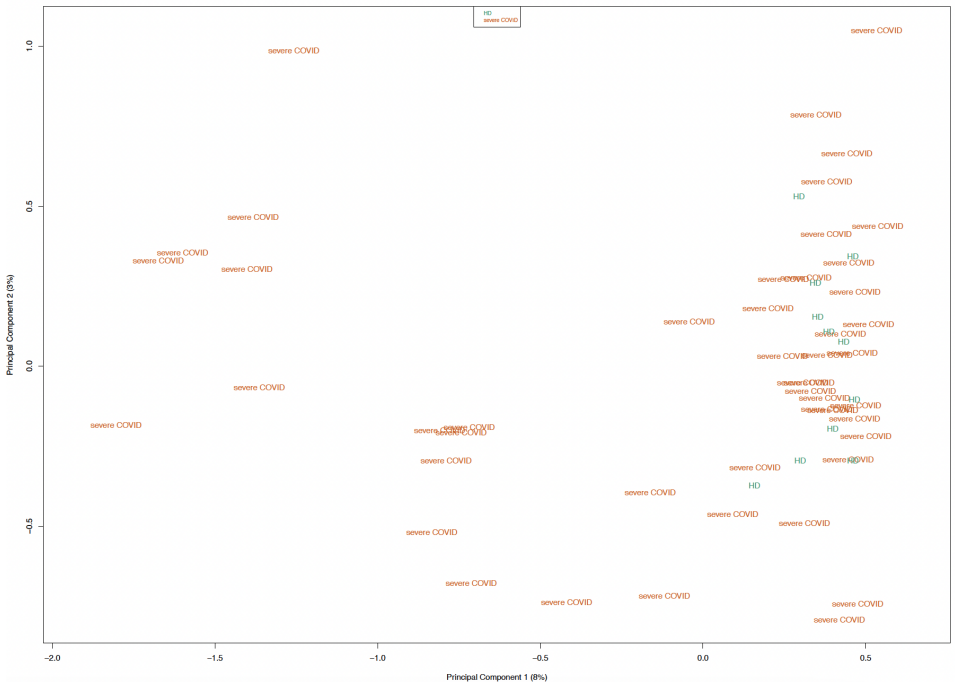


Figure 3: MDS plot illustrating the epigenetic variance between healthy donors and severe COVID-19 patients, emphasizing distinct methylation profiles which may correlate with immune response disparities and potential disease impact on the epigenome.

In the process of methylation data analysis, a crucial step is to filter out low-quality probes to ensure the integrity of downstream results. This is achieved by setting a threshold for detection p-values, which reflect the confidence that a given probe's signal is distinguishable from background noise. In this dataset, probes with a detection p-value less than 0.01 are considered to have passed the quality threshold (TRUE) and are suitable for inclusion in further analysis. Probes failing to meet this criterion (FALSE) are deemed unreliable, likely due to poor hybridization or signal interference, and are thus excluded. After applying this stringent quality control measure, we retain 859,010 high-confidence probes, ensuring that subsequent analytical steps are based on robust data. This filtration step is a standard practice to enhance the validity of the biological interpretations drawn from the study.

On the other hand, after the data cleaning process, which included the removal of single nucleotide polymorphisms (SNPs) within CpG islands, the dimensions of the dataset now stand at 830,573 by 59. This reduction in dimensionality signifies that from the original dataset, 830,573 probes across 59 samples remain post-filtration. The exclusion of SNPs within CpG islands is a vital step, as SNPs can affect the binding of DNA methylation probes, potentially

confounding the methylation signal. By eliminating these variables, the analysis focuses on true methylation patterns, unobscured by genetic variation within the CpG sites. This careful curation of the dataset is crucial for maintaining the accuracy of methylation status assessments across the samples under study.

The next analysis contrasts DNA methylation patterns between healthy individuals and severe COVID-19 patients using Beta and M-values. Beta values indicate the proportion of methylation, while M-values provide a logit-transformed measure for statistical analysis. The plots reveal the variability within each group and suggest distinct methylation profiles associated with severe COVID-19, useful for identifying biomarkers or understanding disease mechanisms.

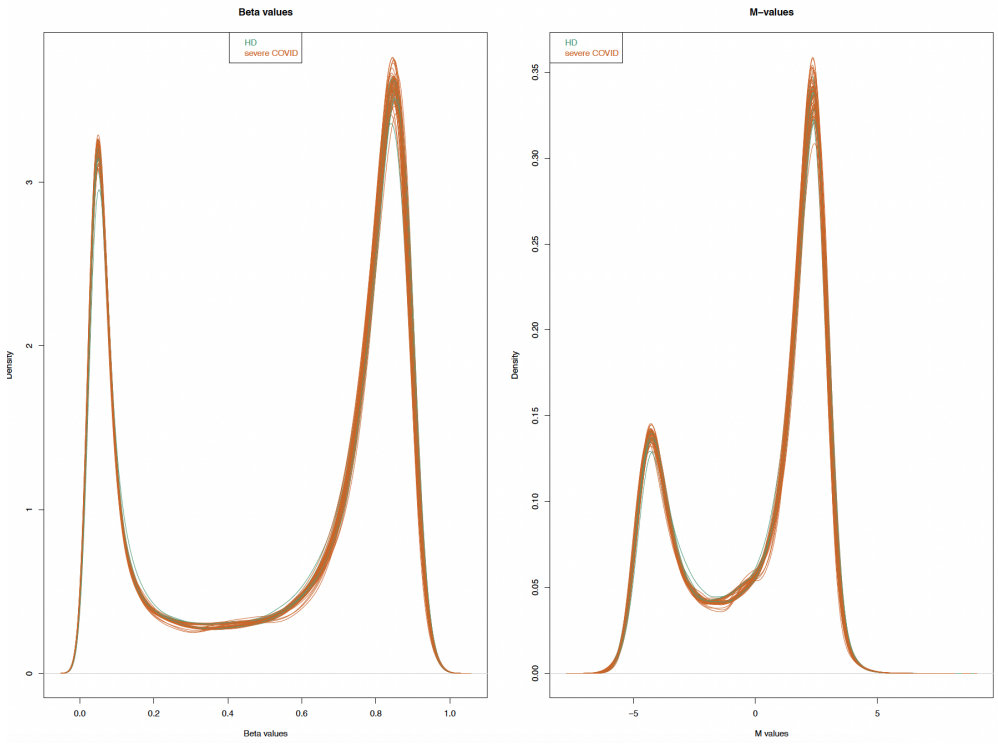


Figure 4: Distribution of Beta and M-values for methylation data in healthy donors (HD) and severe COVID-19 patients, illustrating individual variability and group differences in DNA methylation profiles.

2. DMP Analysis

The DMP analysis revealed a number of CpG sites with significant methylation differences. Discussion of key findings, such as the identification of methylation changes associated with immune response genes.

The dot plots display the methylation beta values at various CpG sites, comparing healthy individuals (HD) with those experiencing severe COVID-19. Each plot corresponds to a unique CpG site, labeled at the top. Beta values range from 0 to 1, where values near 0 suggest low methylation and values near 1 suggest high methylation. Each dot is a unique measurement for subjects within each group.

These plots help us to identify specific methylation patterns associated with COVID-19 severity, potentially contributing to the understanding of its pathogenesis or identifying possible biomarkers.

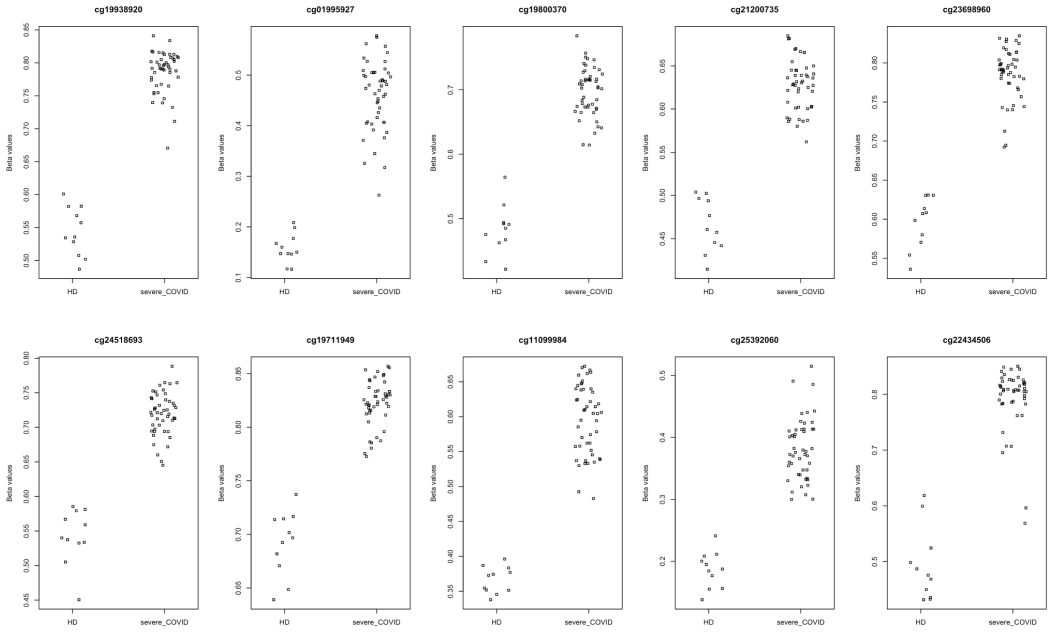


Figure 5: Comparison of beta methylation values between healthy individuals (HD) and patients with severe COVID-19 across various CpG sites. Each point represents an individual's methylation at the specified CpG site..

## **Discussion**

The observed differential methylation patterns across multiple CpG sites between healthy individuals and patients with severe COVID-19 provide a compelling picture of the potential epigenetic regulation in response to the virus. These specific changes in methylation might be influencing the expression of genes related to the immune system, which can affect how the body responds to COVID-19.

Some of the identified CpG sites could be situated in or near genes that are crucial for the activation or suppression of immune pathways. Understanding these changes offers a deeper insight into the pathophysiological mechanisms driven by epigenetic alterations in severe COVID-19 cases.

The comparison with healthy donors highlights the unique epigenetic landscape the infection shapes. This knowledge opens avenues for considering epigenetic therapy as a complementary strategy to modulate the immune response in critically ill patients. Targeted demethylation or the use of epigenetic drugs could reset the dysregulated immune response seen in severe COVID-19.

## **Conclusion**

In conclusion, the differential methylation analysis underscores the significance of epigenetic changes in severe COVID-19 patients, underscoring their potential impact on the immune response. This could pave the way for novel therapeutic strategies that modify these epigenetic alterations to improve patient outcomes in severe cases of COVID-19.