Altered DNA methylation pattern contributes to differential epigenetic immune signaling in the upper respiratory airway of COVID-19 patients

Melissa Govender,[1] Jyotirmoy Das,[2,3] Francis R. Hopkins,[1] Cecilia Svanberg,[1] Johan Nordgren,[1] Marie Hagbom,[1] Jonas Klingström,[1] Åsa Nilsdotter-Augustinsson,[4,5] Yean K. Yong,[6] Vijayakumar Velu[7], Sivadoss Raju[8], Johanna Sjöwall,[4,5] Esaki M. Shankar,[9#] Sofia Nyström,[1,10#] and Marie Larsson[1*]

[1]Division of Molecular Medicine and Virology, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden, [2]Bioinformatics Unit (Core Facility), Linköping University, Linköping, Sweden, [3]Clinical Genomics Linköping, SciLife Laboratory, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden, [4]Division of Infection and Inflammation, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden, [5]Department of Infectious Diseases, Vrinnevi Hospital, Norrköping, Sweden, [6]Laboratory Centre, Xiamen University Malaysia, 43900 Sepang, Selangor, Malaysia, [7]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Division of Microbiology and Immunology, Emory National Primate Research Center, Emory Vaccine Center, Atlanta, GA 30329, USA, [8]State Public Health Laboratory, Directorate of Public Health and Preventive Medicine, DMS Campus, Teynampet 600 018, Chennai, Tamil Nadu, India, [9]Infection and Inflammation, Department of Biotechnology, Central University of Tamil Nadu, Thiruvarur 610 005, India, [10]Department of Clinical Immunology and Transfusion Medicine, Linköping University, Linköping, Sweden.

[#]Equal contributions

Correspondence: Marie Larsson, Division of Molecular Medicine and Virology, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden. Email: marie.larsson@liu.se; Tel.: 00-46-1328-2987

Abstract

The emergence of SARS-CoV-2 has had a profound adverse impact on global health and continues to remain a threat worldwide. The disease spectrum of COVID-19 ranges from asymptomatic to fatal clinical outcomes especially in the elderly population and in individuals with underlying medical conditions. The impact of COVID-19 on host immune responses and immune cells at the protein and DNA levels remains largely ambiguous. In a case-control study, here we explored the impact of COVID-19 on DNA methylation patterns in the upper respiratory airway to determine how SARS-CoV-2 infection altered the immune status of individuals requiring hospitalization for COVID-19. We performed DNA methylation arrays on nasopharyngeal swabs at inclusion/hospitalization as well as 6 weeks post-inclusion. Our study reveals a distinct DNA methylation pattern in COVID-19 patients compared to healthy controls, characterized by 317 779 differentially methylated CpGs. Notably, within the transcription start sites and gene body, COVID-19 patients exhibited a higher number of genes/CpGs with elevated methylation levels. Enrichment analysis of methylated genes at transcription start sites highlighted the impact on genes associated with inflammatory responses and immune functions. Some SARS-CoV-2 -induced CpG methylations were transient, returning to normal levels by 6 weeks post-inclusion. Enriched genes of interest included IL-17A, a pivotal cytokine implicated with inflammation and healing, and NUP93, associated with antiviral innate immunity. Further, six genes in our data set, OAS1, CXCR5, APP, CCL20, CNR2, and C3AR1, were found in enrichment analysis with previous COVID-19 studies. Additionally, RNAse1 and RNAse2 emerged as key regulators, while IL-18 played a role in various biological processes in COVID-19 patients. Overall, our results demonstrates that COVID-19 has a major impact on the upper airway by modifying the methylation pattern of many genes and this could have implications for the conditioning of the airways and how the individual response to future airway infections.

Introduction

The rapid spread and consistent evolution of SARS-CoV-2 variants continues to widen the magnitude of health threat inflicted by COVID-19 on humankind. Since the onset of the COVID-19 pandemic in 2020, substantial advances have been achieved in the realm of therapeutics, together with the development of a wide array of vaccines to combat the SARS-CoV-2 pandemic [1]. Despite increased control measures, the world has witnessed an high death toll due to SARS-CoV-2 infection largely among non-vaccinated individuals [2]. Furthermore, the spectrum of symptoms and the long lasting negative effects among people with severe COVID-19 [3, 4] together with the viral evolution leading to possible new SARS-CoV-2 variants of concern [5, 6], continue to pose formidable challenges to effective disease prevention. The severity of COVID-19 has been linked to dysregulation of immune responses induced by an exaggerated and prolonged inflammation [7]. Hence, a thorough understanding of the nature of immune responses elicited against SARS-CoV-2 is of paramount importance and relevance.

A large proportion of COVID-19 research has been devoted to phenotype, proteomic [8, 9] and/or transcriptomic studies [10, 11, 12], which has given vital information on the dynamics of the immune cell landscape heretofore. In a previous study, we revealed persistent alterations within the monocytic, dendritic cell, and T cell compartments induced by SARS-CoV-2 infection, which manifested even six to seven months post-hospitalization [13, 14]. This underscores the enduring impact of the virus on immune cell dynamics, highlighting the necessity for continued exploration into the intricacies of the immune responses against SARS-CoV-2.

More recently, there have been multiple epigenetic studies aimed to unravel the complexities of SARS-CoV-2 pathogenesis and its enduring impact on the host [15, 16, 17]. Of the various epigenetic modifications, DNA alterations, particularly methylation, exhibit remarkable stability, and form an integral component of a cell's programming, persisting throughout its life cycle and divisions. The activation or silencing of certain genes by DNA modifications is one of the immune regulatory mechanisms implicated in the control of infection and attrition of inflammatory responses to avoid tissue damage in the host [18]. The regulatory influence of DNA methylation extends to transcriptional accessibility of genes and

the subsequent effects, i.e., activity and expression, depending on the region in the genome/gene affected [19, 20]. Many biological and environmental factors appear to influence the DNA methylation patterns such as age, biological sex, and body mass index (BMI) [21, 22]. Likewise, infections, both viral and bacterial, can elicit rapid epigenetic alterations, such as DNA methylation, and thereby regulate gene expression in cells [5, 6, 23].

Blood DNA methylation patterns in asymptomatic and mild COVID-19 cases differed from patterns found in healthy individuals, and genes such as Wnt, and signaling pathways such as muscarinic acetylcholine receptor signaling, and gonadotropin-releasing hormone receptor pathways were enriched in COVID-19 [24]. Unique SARS-CoV-2-specific methylation patterns were found in COVID-19 patients compared to pre-pandemic healthy controls and to patients with other upper respiratory infections due to rhinovirus and influenza B virus [25]. In addition, the immune functions in SARS-CoV-2-infected individuals are also affected by DNA methylation, and certain methylation profiles have been linked to disease severity, such as respiratory distress and intensive care unit (ICU) admission [25, 26].

COVID-19 patients with acute respiratory distress syndrome (ARDS) had 14% differently methylated genes in the promoter regions compared to healthy controls. These promoter regions regulate genes involved in regulating immune pathways, such as IFN-γ and IFN-α [27]. In addition, higher methylation, i.e., hypermethylation of genes in the 'apoptotic execution pathway' has been linked to higher mortality risk in COVID-19 patients [27]. Furthermore, several DNA methylations occurred in the inflammasome component absent in melanoma (AIM) gene during COVID-19 progression, which were linked to heightened immune responses among patients who recovered from SARS-CoV-2 infection [27].

Several epigenetic studies have investigated the methylation patterns in blood samples of individuals with mild to severe COVID-19 [24] [25], [26] [27]. Given that SARS-CoV-2 infection is initiated in the upper airway, establishing a microenvironment involving the respiratory airway mucosa, and that immune cells of the airway mucosa considerably differ from those distributed in the blood, we aimed to explore the methylation status in the upper respiratory airway using nasopharyngeal specimens obtained from hospitalized patients with moderate to severe COVID-19. We observed a significant difference in the methylation patterns among

COVID-19 patients at hospitalization/inclusion and at 6 weeks post-inclusion relative to healthy controls. We found that several genes were differentially methylated revealing significance at hospitalization as compared to 6 weeks post-inclusion. Unsupervised hierarchical clustering of DNA methylation β-values at both overall and transcription start sites revealed distinct differences between COVID-19 patients and controls. Enrichment analyses of methylated genes at transcription start sites in COVID-19 patients highlighted impacts on inflammatory response and immune processes. Some SARS-CoV-2 -induced CpG methylations were transient, reverting to normal levels within 6 weeks post-infection. At inclusion, certain unique gene ontology biological processes were observed, including response to stimulus and cellular process. Notably, the pro-inflammatory factor IL-17A was enriched at inclusion and the antiviral NUP93 factor was enriched at 6-week post inclusion. RNAse1 and RNAse2 emerged as top regulators as well as IL-18, a vital factor that influences various biological processes in COVID-19. This is, to our knowledge, the first instance of investigation of DNA methylation patterns in the nasal area in COVID-19 patients, providing new insight to immune events in the upper airways during COVID-19.

Materials and methods

Participants and sample collection

The study was approved by the Swedish Ethical Review Authority (Ethics No. 2020-02580) and was carried out in accordance with Good Clinical Laboratory Practices, the International Conference on Harmonization Guidelines, and the Declaration of Helsinki. The patient cohort included hospitalized COVID-19 patients (N=27; age range 26-91 years), and SARS-CoV-2-negative healthcare workers as healthy controls (HC) (N=12; age range 26-62 years), which were selected from recruited individuals previously described [13, 14]. All individuals provided written informed consent prior to enrolment. The COVID-19 patient clinical data and additional information are described in **Table 1**.

COVID-19 disease score was defined according to the criteria advocated by the National Institutes of Health [4, 28], and slightly modified based on the requirement for supplementary oxygen and the highest level of care required. Patients in this cohort were classified mostly as moderate to severe based on the above criteria **(Table 1)**. Nasopharyngeal specimens were collected at inclusion (baseline) (timepoint 1) and at 6 weeks post-inclusion (timepoint 2). The samples were transferred into microfuge tubes and centrifuged at 1800 RPM for 6 minutes at 4°C, and the pellets were frozen at -80°C for the extraction of genomic DNA (gDNA) for later use in the experiments.

Genomic DNA isolation

Genomic DNA (gDNA) was isolated using the Qiagen AllPrep DNA/RNA Mini kit (Qiagen) according to the manufacturer's instructions. Briefly, 600 µl of lysis (RLY) buffer was added to thawed pellets and mixed. The lysate was transferred to the AllPrep DNA column and centrifuged at 11000 g for 1 min. The column was washed with 500 µl AW1 buffer (centrifuged at 11000 g for 1 min) and washed with 500 µl AW2 buffer. Finally, the columns were placed in 1.5 ml microfuge tubes, and 30 µl extraction buffer (EB) was added for a 1 min incubation, before eluting the gDNA by centrifuging at 11 000 g for 1 min. The elution step was repeated with the eluant to increase the gDNA concentration.

Quantification of genomic DNA

The quality and concentration of the genomic DNA was measured using the broad range (BR) dsDNA Quantitation Qubit assay (Invitrogen, USA) according to the manufacturer's instructions. Briefly, the samples were prepared with the Qubit working solution (Qubit reagent diluted 1:200 in Qubit buffer). The standards as well as nasopharyngeal samples were prepared in working solution and were vortexed for a few seconds before incubation for 2 min at room temperature. The amount of genomic DNA was then quantified by measuring the fluorescence with the Qubit™ dsDNA BR Assay Kit (Invitrogen, USA) on a Qubit^{TM} 2.0 Fluorometer (Invitrogen^{TM} by Life Technologies^{TM}, Thermofisher, USA).

Sample preparation for DNA methylation microarray

Quantified genomic DNA samples (69-552 ng) were transported to the Bioinformatics and Expression Analysis Core facility (BEA) based at the Karolinska Institute, Stockholm, Sweden. Here, the samples were subjected to bisulfite conversion using the Zymo Research EZ-96 DNA Methylation™ Kit (D5004). The samples were quantified with the Qubit assay, and 70-250ng was used as input for the Illumina Infinium Methylation EPIC 850K BeadChip array (Illumina Inc., San Diego, USA) following the protocol outlined by Illumina. QC, normalized data, and IDAT files were obtained and used for further analysis.

Data processing and statistical analysis: Differential methylation analysis

The IDAT files from Illumina® HumanMethylation EPIC arrays were analyzed using R (v4.2.1)[32] and Bioconductor packages (v3.16), Chip Analysis Methylation Pipeline (ChAMP)[33] analysis package (v2.28.0). The data initially contained 865918 probes. The files were pre-processed to filter out CpGs with detection *p* value >0.01 (removing 32515 probes), with a bead count of <3 in at least 5% of samples (removing 5192 probes), all non CPG probes and only keeping CpGs (removing 5192 probes); those with SNPs [30] (probes that align to multiple locations as identified in Nordlund et al [31]; and finally, filtering probes located on X and Y chromosomes. 715410 probes and all 48 samples were left remaining for analysis.

. A quality assessment on the filtered data was performed and using beta-mixture quantile normalization (BMIQ) function and normalized dataset was calculated. The β- and M-values for each CpG per sample was estimated in a distribution plot. To reduce the batch effect in relation to biological variation on the data matrix, deconvolution (singular value

decomposition, SVD) was performed on the normalized data using *runCombat* function and corrected against the confounding factors lactate dehydrogenase (LDH), and BMI. The differential methylation analysis was done on the corrected data with the linear modeling (*lmFit*) and *eBayes* algorithm between two sample groups. The differentially methylated CpGs were considered significant at the Bonferroni-Hochberg (BH)-corrected p value (p value$_{BH}$) <0.05. The hierarchical cluster analysis was performed using the Euclidean distance calculation within the *ape* package [34](v5.7). The principal component analysis was performed using *FactoMineR* [35](v2.8) and *factoExtra* [36] (v1.0.7) packages with *in house* R script.

Downstream analysis: Feature analysis of differentially methylated CpGs

The resulted differentially methylated CpGs were annotated using *AnnotationDbi* package [37] (v1.60.2) (Human Genome version 38) using the *in house* script to visualize the genomic distribution. The volcano plot displayed the distribution of hypermethylated and hypomethylated (log$_2$FC cut-off <0.3) significant differentially methylated CpGs (p value$_{BH}$ <0.05) using *in house* R script with ggplot2 package [38, 39] (v3.4.2). The cut-off score was calculated using the β-value distribution of all samples with mean±2SD. Heatmaps were calculated using in-house R script with Complex. Heatmap package [40](v2.14.0) from individual β-value. The differentially methylated CPGs result was filtered based on the genomic location, selecting the transcription start site (TSS) regions (TSS200 and TSS1500) and gene body regions using log2FC cut-off score of >0.3 for hypermethylated and <0.3 for hypomethylated and genomic location. Separated heatmaps were created to visualize the β-value distribution. The differentially methylated CpGs were annotated with their respective official gene symbols.

Enrichment analysis was performed using the Metascape database (http://metascape.org){Zhou, 2019 #3} to determine any biological significance of the pooled 1000 differentially methylated genes, i.e. top 500 hypomethylated and 500 hypermethylated genes, in the transcription start site regions TSS1500 and TSS200 in the COVID-19 patients. Further analysis was performed using the CORONASCAPE database (https://metascape.org/COVID/){Zhou, 2019 #3} to identify any similarities between our data, i.e. the pooled 1000 differentially methylated genes, and differentially methylated genes from other COVID-19 datasets. The Ingenuity Pathway Analysis (IPA) software was

used for pathway, top regulator analysis and network summary of the differently methylated genes in the TSS1500 and TSS200 regions. The data set included 5077 mapped ID that gave 2831 analysis-ready molecules with log FC -0.5503 to 0.5475.

Statistical analysis

All differences with a p value$_{BH}$<0.05 were considered significant, if not stated otherwise. We calculated family-wise error rate (FWER) using the Bonferroni-Hochberg (BH)-correction method. All analyses were performed using R (v4.2.1) with the aforementioned packages.

Results

Patients and clinical characteristics

To assess the DNA methylation profiles in the upper respiratory airways of hospitalized COVID-19 patients (N=27), a total of 36 nasopharyngeal swabs were collected, with 21 samples from hospitalization/inclusion (T1), and 15 samples 6 weeks post-inclusion (T2). In addition, nasopharyngeal swabs from healthy subjects (N=12) were used as controls **(Figure 1)**. Participants were recruited between July 2020 to October 2021, and represented hospitalized patients with moderate to severe COVID-19 manifestations, as determined by the Guidelines of the National Institutes of Health, based on the maximum oxygen required and the highest level of care needed [28]. At this time the following SARS-CoV-2 strains circulated in Sweden: alpha, beta, gamma, and delta. Of the total patients, 25.9% had at least two of the following underlying conditions, i.e., cardiovascular disease, pulmonary disease, and diabetes mellitus. The median number of days with symptoms prior to inclusion into the study was nine **(Table 1)**. Nasopharyngeal swabs (N=12) were obtained from healthy controls at enrollment. Standard clinical and blood parameters were assessed at inclusion for COVID-19 patients **(Table 1)** as well as controls.

Raw DNA methylation data revealed high quality and beta distribution

Genomic DNA (gDNA) from nasopharynx samples were prepared for bisulfite conversion, and we performed a DNA methylation array (Illumina Infinium Methylation EPIC 850K). The data were processed and filtered with bioinformatics analysis to identify methylation patterns and determine their biological significance **(Figure 1)**. Initial quality control assessment of the raw methylation data indicated a good continuous probability distribution, i.e., beta distribution, between samples **(Supplementary Figure 1A)**, and separation of healthy control samples and inclusion and 6-week time-points samples from COVID-19 patients in the multidimensional scaling plot **(Supplementary Figure 1B)**. Lactate dehydrogenase and BMI were estimated as significant components of variation/confounding factors in the methylation dataset, and the effect of these factors were corrected using singular value decomposition, i.e., deconvolution. LDH was disregarded due to being a patient-related factor, and the dataset was adjusted for the BMI **(Supplementary Figure 1C)**.

COVID-19 patient DNA methylation pattern differs from that of healthy controls

For the identification of differentially methylated CpG sites, an assessment of the samples from COVID-19 patients, both at inclusion and 6-week timepoints, and healthy controls samples was performed. A total of 317779 statistically significant differentially methylated CpGs were identified in the pooled COVID-19 patient inclusion (T1) and 6-week timepoint (T2) samples as compared to healthy controls. Inclusion versus control samples showed 321428 significant differentially methylated CpGs, and 6-week timepoint samples versus control samples gave 235708 significant differentially methylated CpGs. Hierarchical clustering was performed to visualize global correlation among T1, T2, and healthy control samples. Two major clusters obtained from the analysis showed a cluster with healthy controls and a cluster with patients, which indicated a general separation between COVID-19 patients and healthy controls **(Figure 2A)**. Principal component analysis, an unsupervised learning method, to show variation in the data sets revealed a clear separation of hospitalized COVID-19 patients from healthy controls, revealing a highest variation in data set (PC1) i.e., 32.2% and second most variation (PC2) of 8.8% **(Figure 2B)**. There were exceptions with two patient samples at inclusion (P02, and P03) and three samples at the 6-week time point (P07, P09, and P24), which clustered with the healthy controls, as well as one healthy control sample (HC01) that clustered with COVID-19 patients. The reason for patients clustering with healthy controls or overlap of some patient samples with healthy controls could be the underlying conditions that had led to moderate/severe COVID-19 that was not linked to altered airway/lung pathology or medications that protected the airway compared to the other COVID-19 patients. The latter appears to be the case for all patient outliers besides one since they had been on anti-inflammatory medication for chronic pulmonary conditions. Regarding the healthy control individuals that clustered with the patient samples, we could not find an obvious explanation from the clinical data/history.

Majority of the COVID-19 altered methylated CpGs remained stable over 6 weeks, but a substantial fraction was found only at inclusion

Next, we compared the difference and overlap in CpGs between inclusion versus healthy controls (T1-HC), and between 6-week timepoint versus healthy controls (T2-HC). A unique set of 94382 differentially methylated CpGs were found in the T1-HC that did not exist in the T2-HC, and a unique set of 8609 differentially methylated CpGs in the T2-HC that was not found in the T1-HC **(Figure 3A)**, indicating that most of the differentially methylated CpGs

still persisted more than 6-week post infection. This suggests a major difference in DNA methylation patterns between COVID-19 patients and healthy individuals, and that most of the alterations in CpG methylation remained for at least 6-week post infection **(Figure 2 and 3A)**. Based on these findings, we used the pooled inclusion and 6-week timepoint samples versus control samples for additional analysis.

Next, we identified the amount of significant differentially methylated CpGs in the various genomic regions, i.e., gene body (body), intergenic region (IGR), exon boundaries (ExonBnd), 1st Exon, 3'UTR, 5'UTR, and the transcription start sites (TSS) 1500 and 200, in COVID-19 patients versus HCs **(Figure 3B)**. The greatest number of differentially methylated CpGs in the different genomic regions appeared in the body region **(Figure 3B)**. There were no major differences in the levels of the significant differentially methylated CpGs in the different genomic regions between groups, although in general, the 6-week post samples had less significantly differently methylated CpGs in the different regions compared to the inclusion timepoint **(Figure 3B)**.

Differently hypomethylated and hypermethylated sites/genes with highest significance diverge in the COVID-19 patients between the inclusion and 6-week time points

Methylation can affect gene transcription either by enhancing, decreasing, or silencing transcription [19, 20]. Hypomethylated and hypermethylated CpG sites specific for COVID-19 were identified based on log_2FC values with a set cut-off ±0.3 and p value$_{BH}$ <0.05 and the top 20 differentially hypomethylated- and hypermethylated CpG sites/genes were emphasized **(Figure 3C, Table 2)**. The hypermethylated genes with the lowest -log_{10}adjusted p value included, SGIP1, SH3GL3, RAD54L2, TMEM63C, and PACRG, associated with the differentially methylated CpGs for hospitalized COVID-19 patients at inclusion and at 6-weeks post-inclusion versus HCs, while the top five hypomethylated genes were, TMEM2, MECOM, XBP1, DLX5, and GRP77 **(Figure 3C, T1 & T2 vs HC)**. The hypermethylated genes with the lowest -log_{10} adjusted p value at inclusion (T1) encompassed SPOCK3, SMC6, SHANK2, RAD54L2, and RUNX1T1, and top hypomethylated genes, DLX5, TMEM2, ZNF664-FAM101A, PLEKHM3, MECOM **(Figure 3C)**. By 6-weeks post inclusion (T2), COVID-19 hypermethylation of genes with lowest -log_{10} adjusted p value included FBN1, TSHZ2, GNA13, CRYL1, and EEPD1, and hypomethylated genes included TNFAIP8, PARL, NISCH, INPP5D, and NOD2 **(Figure 3C)**. There were no overlapping methylated genes when

12

comparing genes with the highest p value$_{BH}$ between inclusion and 6-weeks post-inclusion **(Figure 3C, T1 vs T2)** in COVID-19 patients versus healthy controls. Of note, there was a remarkable decline in the hypomethylated and hypermethylated genes' -log$_{10}$ adjusted p values between the inclusion and 6-weeks post-inclusion, with a decline from ~5 to 3, indicating that the altered CpG methylation induced by COVID-19 was getting restored to the levels found in healthy, i.e., as before the infection. When exploring the top hypermethylated genes according to the highest fold change in the patients at inclusion compared to healthy controls we found BAG3, EHF, FAM178B, TMX2-CTNND1, and MCC, and for top hypomethylated genes we found NISCH, PMP22, SFMBT2, FAM124A, and FCER1G **(Supplementary Table 1)**. By 6 weeks post-inclusion, COVID-19 hypermethylation of genes with highest fold-change compared to healthy controls included RALGAPA2, PDGFD, C1orf228, MCC and FAM178B, while the top five hypomethylated genes were LOC400867, TNFAIP8, GIT2, CANX and PARL **(Supplementary Table 1)**.

Different methylated CpGs patterns in COVID-19 patients versus healthy controls

For a snapshot of the overall pattern of differently methylated CpGs covering all the different genomic regions in hospitalized COVID-19 patient at T1 and T2 versus HCs, a heatmap with hierarchical clustering of the DNA methylation beta-values was constructed with top 20000 methylated CpGs. This showed a clear difference in methylation between the COVID-19 patients and controls **(Figure 4)**. Following this, we focused on the specific methylation patterns for the transcription start sites TSS1500 and TSS200, as methylations in these areas are good predictor of activation versus silencing of genes [41]. The DNA methylation beta-value was analyzed using hierarchical clustering with the top 1000 differentially methylated CpGs and the distribution of other associated parameters, timepoint, age and biological sex; and clinical parameters associated with disease, i.e., LDH and C-reactive protein (CRP) **(Figure 5)**. There was a distinct pattern of differently methylated CpGs in the TSS1500 and TSS200, but we found no specific relation to age or biological sex, CRP, or LDH, with the pattern of differently methylated CpGs **(Figure 5)**. The gene body methylation had similar patterns as seen for transcription start sites **(Supplementary Figure 2)**. Of note, the COVID-19 patients had, within the selected transcription start sites TSS1500, TSS200, and the gene body, more genes/CpGs with high beta values than the healthy controls, i.e., higher level of methylated CpG sites.

13

Enrichment analysis of genes methylated in the transcription start sites in COVID-19 patients indicated an effect on genes involved in inflammatory and immune responses

To investigate the biological significance of methylated genes in COVID-19 patients, we pooled the top 500 hypomethylated and 500 hypermethylated transcription start site regions TSS1500 and TSS200 encoding genes and performed an enrichment analysis using Metascape database (http://metascape.org) [42]. The top enriched ontology terms included regulation of cell activation, leukocyte activation, and inflammatory responses **(Figure 6A),** and top biological processes that included multicellular organismal processes, immune system process, and response to stimulus **(Figure 6B)**. The large number of unique differentially methylated genes in the TSSs between the inclusion and 6-week post timepoint (T1 & T2) **(Figure 3A, Supplementary Table 2)**, showed that the initial effects of the viral infection and immune responses occurring early during COVID-19 (T1) did not persist at the later stage (T2). This finding indicates that some of the CpG methylations in the TSSs induced by COVID-19 were short-term. Further enrichment analysis of the unique genes affected at T1 was performed, using the top 50 differently methylated CpGs at the TSSs, according to adjusted *p* value, and fold-change. For T1 versus HC the top-level gene ontology biological processes were Response to stimulus and Cellular process (adjusted *p* value), whereas the top-level gene ontology biological processes according to fold-change differences were responsible for stimulus and growth **(Figure 8D)**.

Enrichment analysis in DisGeNET also highlighted delayed myelination, which is commonly caused by viral infections [43]. Whilst looking at the enrichment according to fold-changes, we found Immune system process and Localization to be the top-level gene ontology biological processes, and an enrichment for infective cystitis, which has been reported to occur in some COVID-19 patients[46]. Further enrichment analysis in TRRUST[44] indicated the top predicted biological processes for T2 versus HCs, as Multicellular organismal process and Negative regulation of biological process (adjusted *p* value), and an enrichment of the regulation by the tumor suppressor gene TP53. Of note, there was an enrichment for the regulation by IRF-1, highlighting the key role of interferons in COVID-19 [47]. The two genes of interest that were enriched included at T1 IL-17A, a key cytokine involved in inflammation and healing [48], and NUP93 at T2, which is associated with antiviral innate immunity [49].

RNAse1 and RNAse2 were identified as top regulators, and IL-18 was involved in several biological processes in COVID-19 patients

To identify activated and inhibited canonical pathways, and to identify top regulator networks based on all significant hypomethylated and hypermethylated transcription start site regions encoding genes, we used Ingenuity Pathway Analysis (IPA) **(Figure 7)**. The canonical pathways predicted to be activated included PI3K/AKT Signaling, Antioxidant Action of Vitamin C, Th2 Pathway, Natural Killer Cell Signaling, and Neuroinflammation Signaling Pathway. The pathways predicted to be inhibited included S100 Family Signaling Pathway, Leukocyte Extravasation Signaling, Phospholipases, and ERK/MAPK Signaling **(Figure 7A)**. RNAse1 and RNAse2, hydrolytic enzymes catalyzing degradation of RNA, were identified as top regulators in the regulator network with the highest consistency score **(Figure 7B)**. RNAse1 and RNAse2 regulate factors such as pro-inflammatory and anti-inflammatory cytokines leading to the homing and chemotaxis of several immune cells and cell death. Next, we explored another top regulator network, with four regulators, i.e., DNA methyltransferase 3 alpha (DNMT3A), an enzyme important for methylation; Ubiquitin specific peptidase (USP22), a ubiquitin-specific processing protease; the enzyme Caspase 4 (CASP4), autophagy-pyroptosis-related function; and Colony Stimulating Factor 3 (CFS3), cytokine controls the production, differentiation, and function of granulocytes. These regulators altogether contribute to increased interactions with multiple enzymes, and transmembrane proteins, which have an effect on functions relating to degranulation of phagocytes and myeloid cells **(Figure 7C)**. A summary of the major biological pathways and factors demonstrating the role of IL-18, an important pro-inflammatory Th1 polarizing cytokine linked to various processes including chemotaxis and migration of leukocytes, including neutrophils, and degranulation **(Figure 7D)**. IL-18 plays a major role in various infectious, metabolic, and inflammatory diseases by inducing IFN-γ thereby promoting Th1 cell activation and enhancing the cytotoxic activity of CD8+ T cells and natural killer (NK) cells [50].

Enrichment analysis match of genes methylated in the transcription start sites in COVID-19 patients versus healthy controls compared to other COVID-19 datasets

To examine the similarities with other COVID-19 datasets, we performed an enrichment analysis match of the top 1000 affected genes, consisting of pooled top 500 hypomethylated and 500 hypermethylated transcription start sites regions encoding genes, found in our dataset **(Figure 8)**. Several COVID-19 datasets in the CORONASCAPE database [42] were predicted to match, and five datasets were selected based on gene overlap and adjusted logP values. All datasets were from transcriptomic studies on blood and airway samples from SARS-CoV-2 infected patients and in vitro SARS-CoV-2 infected cell line such asprimary human airway epithelial and hypotriploid alveolar basal epithelial cells. There was an overlap in genes between all datasets and many genes were found in several of the datasets **(Figure 8A)**. Of note, there were many genes unique to our data set **(Figure 8A)**. Our DNA methylation data shared enrichment of six genes, i.e., OAS1, CXCR5, APP, CCL20, CNR2, and C3AR1 with the other datasets **(Figure 8B),** of which several have been defined in previous COVID-19 studies [51, 52, 53, 54]. The enriched top terms defined by the analysis match included regulation of cell activation, leukocyte activation, inflammatory response, regulation of immune effector processes, and neutrophil degranulation **(Figure 8C)**. Furthermore, the top-level biological processes included multicellular organismal processes, immune system process, response to stimuli, and regulation of biological processes **(Figure 8D)**. Taken together, our findings clearly indicate a multilayered activation of responses by SARS-CoV-2 infection.

Discussion

Severe infections and conditions will alter the DNA methylation pattern in cells and tissues. Here were nasopharyngeal samples from patients hospitalized with COVID-19 used to identify SARS-CoV-2-induced epigenetic signatures by profiling the DNA methylation patterns at inclusion/hospitalization as well as 6-weeks post-inclusion. We utilized upper respiratory airway samples to explore the effect on the airway as there is evidence that epigenetic alterations that occur in one location or cell type cannot directly be extrapolated for another cell/tissue [55]. There was a clear separation in the DNA methylation pattern in the airway samples between COVID-19 patients and healthy controls. The high impact exerted by COVID-19 on the DNA methylation pattern in the airway, is in line with other studies that explored this in blood samples from COVID-19 patients [15, 24, 25, 26, 27, 56]. Interestingly, patients who had undergone anti-inflammatory treatment such as oral/nasal inhalation glucocorticoids for previous/chronic airway manifestations had a different DNA methylation pattern and clustered with healthy controls in the PCA. This suggests that the initial inflammatory response played a major part of the imprinted DNA methylation pattern in the airways and that this did not occur to the same degree among individuals on anti-inflammatory treatment due to less inflammation in the airway. Another explanation could be that the anti-inflammatory drugs alter the DNA methylating effects exerted by SARS-CoV-2 on the airway and the local infiltration of immune cells.

We did not see a clear distinction in DNA methylation between COVID-19 patients with moderate and severe disease when exploring hierarchical clustering. Separate DNA methylation patterns have been demonstrated between mild and severe COVID-19 [57, 58].We found some evidence reflecting a restoration of the epigenetic profile in the patients, but even if certain of the COVID-19-induced hypomethylated and hypermethylated CpGs/genes returned to levels found in healthy, there was still major lasting effects on the DNA methylation pattern in the COVID-19 patients. One explanation for the changes in the airway methylation patterns could be altered cell compositions, such as infiltration of immune cells, proliferation of tissue resident cells and damaged epithelia at the initial phases or during active infection versus a more healing phase a few weeks after COVID-19 [59, 60].

It's evident that some of the differently methylated genes that were present only at inclusion in the airway samples, but not at the 6-week post inclusion time point, such as OAS1 and

OAS3 were part of the innate type 1 interferon responses to viruses. These genes have been confirmed to be affected in COVID-19 patients and are linked to clinical outcome [10, 25]. In blood samples, others have shown that some DNA methylation alterations are present one year after SARS-CoV-2 infection in hospitalized patients [61]. Genome areas with genes and CpGs that returned to normal levels were connected to viral responses such as type I IFN signaling, whereas areas associated with cell activation, leukocyte activation, lymphocyte activation, immune system remained altered over time [61]. Our findings are in line with these previous findings in blood[61] with regards to DNA methylation returning to normal levels in genes involved in viral responses such as the OAS1, whereas genes involved in e.g. leukocyte activation, and immune system remained altered.

We saw alterations in DNA methylation in all genomic regions in the COVID-19 patients, but focused on the deeper analysis of the DNA hypermethylation and hypomethylation patterns in the transcription promotor regions as this more easily can be translated to silenced and activated genes, respectively [62]. Further, a previous study found that differential patterns of COVID-19 DNA methylation in blood occurred primarily in the promoter regions of immune-related genes [27].

The top enriched ontology and biological processes predicted by the pathway analysis in the COVID-19 data set included regulation of cell activation, leukocyte activation, inflammatory responses, leukocyte migration, immune system process, and response to stimuli, which is in accordance with previous COVID-19 studies [25, 63]. Signaling pathways predicted to be affected included the S100 signaling pathway and PI3K/AKT Signaling. S100 family genes have been identified as prognostic markers of severe COVID-19 [64] and the PI3K/Akt intracellular pathway might be an important signaling pathway in the cytokine storm induced by SARS-CoV-2 and also in COVID-19 coagulopathies [65].

The endogenous antimicrobial polypeptides RNAse1 and RNAse2 were defined as top regulators in the COVID-19 DNA methylation data set. Zechendorf et al reported that elevated RNASE1 levels were linked to renal dysfunction among ICU-admitted COVID-19 cases [66]. Others found increased RNAse2 levels in COVID-19 patients with critical disease [67]. The role(s) RNAses may play in SARS-CoV-2 infection is largely unknown. RNASEs might play a

role regulating the innate immune response activated by SARS-CoV2 by recognizing and degrading single- and double-stranded RNA molecules and thereby limiting the TLR3 and TLR7/8 ligands, i.e., extracellular viral RNA, released by Infected dying cells [68]

An analysis of the major biological pathways and factors demonstrated a key role for IL-18, an important proinflammatory cytokine induced by IFN-γ and linked to various processes including chemotaxis and migration of leukocytes, inflammasome activation, pyroptosis and cytotoxic activity of CD8+ T cells and NK cells [69]. IL-18 has been shown to play a major role in various infectious, metabolic, and inflammatory disorders [50]. COVID-19 with elevated levels of IL-18 has been associated with increased severity and mortality [70, 71]. Furthermore, IL-18 SNPs/mutations, such as IL18-105G>A, have been shown to be protective against severe COVID-19 [72].

Whilst exploring some of the other genes affected by COVID-19, we found that the IL-10 gene was hypomethylated in COVID-19 patients at inclusion/hospitalization, indicating an active IL-10 response early on during infection and it has been shown to play a part in COVID-19 pathogenesis, both during acute infection and post/long COVID-19 [73, 74]. Genes involved in innate immune responses such as NOD2, KLK7 and genes involved immune regulation, such as SEMA4D, INPP5D, TOX, IKZF3, were identified among the differently methylated genes with the highest fold change in COVID-19 patients. The epigenetic imprinting by the SARS-CoV-2 infection on these genes involved in innate responses and immune regulation is important determinates for the disease development [75].

Furthermore, we observed alterations of TSSs for genes involved in protein transcription, translation, folding, quality control, post transcriptional modification affected e.g. CANX, GPN3, RBM47, METTL21A, and BAG3 and in neurological development/disorders e.g. PARL, ADK, TSHZ2, ASAM10, and AATK. Infections and inflammation impact all steps in the production, control, and modification of new proteins and SARS-CoV2 has a direct impact on protein translation, and protein trafficking [76], and need post translational modifications [77]. Regarding neurological effects of COVID-19 there are reports that have demonstrated direct effects on the nervous system and connection between long COVID-19 [78, 79] and exploring our data set's matches with other data sets revealed mutual enrichment of six genes, i.e.,

OAS1, CXCR5, APP, CCL20, CNR2, and C3AR1 [10, 51, 52, 53, 54, 80, 81, 82, 83, 84]. The enriched top terms defined by the analysis match included regulation of cell activation, leukocyte activation, inflammatory response, regulation of immune effector processes, and neutrophil degranulation. The genes and the biological processes/responses defined in the enrichment analysis are commonly found in numerous COVID-19 studies {Blanco-Melo, 2020 #113{Wilk, 2020 #117}{Vanderheiden, 2020 #119} and reflect the effects the viral infection and subsequent immune response exert on the host.

Severe conditions such as sepsis are connected to changes in DNA methylation patterns in white blood cells in areas with genes involved in, e.g., inflammatory pathways, innate and adaptive immune response, which are reflected in altered gene expression [85]. Subsequently, a relevant comparison of our data would be with other severe infectious diseases to determine the DNA methylation alteration that is specific for COVID-19. In addition, a larger cohort could also increase the ability to determine with higher precision unique DNA methylations and patterns for COVID-19 and it would be of interest to verify some findings with functional tests.

In conclusion, we found a clear distinct DNA methylation pattern in the COVID-19 patients compared to healthy controls and that most but not all altered methylations lasted for more than 6 weeks post inclusion. in the COVID-19 patients compared to controls. There was an enrichment of genes with altered methylation at transcription start sites associated with Inflammation and immune system in the COVID-19 data set, which might reflect an attempt to dampen the inflammation and immune reaction.

Author contributions

MG, FH, CS, handling of study samples. ML and MG designed the experiments. MG conducted experiments. MG, ML, JD, SN, YKY, JK, EMS and ML analyzed the data. AN-A, AJH, MH, JS, JN, SN, and ML were involved in establishing the cohort study. MG, JD, EMS, SN, VV, RS, and ML were involved in the writing of the initial manuscript draft. MG, EMS, JD, JK, SN, SR, VV, FH, CS, JN, YKY, AN-A, AH, MH, JS, and ML were involved finalizing the manuscript for submission. All authors read and approved the final manuscript.

Ethical statement

The studies involving human participants were reviewed and approved by Swedish Ethical Review Authority. The patients/participants provided their written informed consent to participate in this study.

References

1. CDC.. COVID-19 Treatments and Medications. *https://www.cdc.gov/coronavirus/2019-ncov/your-health/treatments-for-severe-illness.html* (2023).

2. WHO. WHO Coronavirus (COVID-19) Dashboard. . **https://covid19.who.int/** (2020 7 February 2023).

3. McIntosh, K. COVID-19: Clinical features UpToDate 2023 *https://www.uptodate.com/contents/covid-19-clinical-features*.

4. Phua, J. *et al.* Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *Lancet Respir Med* **8**, 506-517 (2020).

5. Carabelli, A.M. *et al.* SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol* **21**, 162-177 (2023).

6. Kudriavtsev, A.V. *et al.* Immune Escape Associated with RBD Omicron Mutations and SARS-CoV-2 Evolution Dynamics. *Viruses* **14** (2022).

7. Montazersaheb, S. *et al.* COVID-19 infection: an overview on cytokine storm and related interventions. *Virol J* **19**, 92 (2022).

8. Zhong, W. *et al.* Next generation plasma proteome profiling of COVID-19 patients with mild to moderate symptoms. *EBioMedicine* **74**, 103723 (2021).

9. Ciccosanti, F. *et al.* Proteomic analysis identifies a signature of disease severity in the plasma of COVID-19 pneumonia patients associated to neutrophil, platelet and complement activation. *Clin Proteomics* **19**, 38 (2022).

10. Binkowski, J. *et al.* Epigenetic activation of antiviral sensors and effectors of interferon response pathways during SARS-CoV-2 infection. *Biomed Pharmacother* **153**, 113396 (2022).

11. Sarker, B. *et al.* Identification of host genomic biomarkers from multiple transcriptomics datasets for diagnosis and therapies of SARS-CoV-2 infections. *PLoS One* **18**, e0281981 (2023).

12. Bibert, S. *et al.* Transcriptomic Signature Differences Between SARS-CoV-2 and Influenza Virus Infected Patients. *Front Immunol* **12**, 666163 (2021).

13. Govender, M. *et al.* T cell perturbations persist for at least 6 months following hospitalization for COVID-19. *Front Immunol* **13**, 931039 (2022).

14. Hopkins, F.R. *et al.* Major alterations to monocyte and dendritic cell subsets lasting more than 6 months after hospitalization for COVID-19. *Front Immunol* **13**, 1082912 (2022).

15.   Bradic, M. *et al.* DNA methylation predicts the outcome of COVID-19 patients with acute respiratory distress syndrome. *J Transl Med* **20**, 526 (2022).

16.   Dey, A. *et al.* Epigenetic perspectives associated with COVID-19 infection and related cytokine storm: an updated review. *Infection*, 1-16 (2023).

17.   Nikesjo, F. *et al.* Defining post-acute COVID-19 syndrome (PACS) by an epigenetic biosignature in peripheral blood mononuclear cells. *Clin Epigenetics* **14**, 172 (2022).

18.   Qin, W., Scicluna, B.P. & van der Poll, T. The Role of Host Cell DNA Methylation in the Immune Response to Bacterial Infection. *Front Immunol* **12**, 696280 (2021).

19.   Moore, L.D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23-38 (2013).

20.   Dhar, G.A., Saha, S., Mitra, P. & Nag Chaudhuri, R. DNA methylation and regulation of gene expression: Guardian of our health. *Nucleus (Calcutta)* **64**, 259-270 (2021).

21.   Dhingra, R., Nwanaji-Enwerem, J.C., Samet, M. & Ward-Caviness, C.K. DNA Methylation Age-Environmental Influences, Health Impacts, and Its Role in Environmental Epidemiology. *Curr Environ Health Rep* **5**, 317-327 (2018).

22.   Samblas, M., Milagro, F.I. & Martinez, A. DNA methylation markers in obesity, metabolic syndrome, and weight loss. *Epigenetics* **14**, 421-444 (2019).

23.   Khalil, C.A. *Medical Epigenetics* vol. 29, 2021.

24.   Huoman, J. *et al.* Epigenetic rewiring of pathways related to odour perception in immune cells exposed to SARS-CoV-2 in vivo and in vitro. *Epigenetics* **17**, 1875-1891 (2022).

25.   Konigsberg, I.R. *et al.* Host methylation predicts SARS-CoV-2 infection and clinical outcome. *Commun Med (Lond)* **1**, 42 (2021).

26.   Castro de Moura, M. *et al.* Epigenome-wide association study of COVID-19 severity with respiratory failure. *EBioMedicine* **66**, 103339 (2021).

27.   Balnis, J. *et al.* Blood DNA methylation and COVID-19 outcomes. *Clin Epigenetics* **13**, 118 (2021).

28.   NIH. COVID-19 treatment guidelines panel. coronavirus disease 2019. (COVID-19) treatment guidelines. . *https://www.covid19treatmentguidelines.nih.gov/*. (2020).

29.   Morris, T.J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428-430 (2014).

30.   Zhou, W., Laird, P.W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22 (2017).

31.   Nordlund, J. *et al.* Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol* **14**, r105 (2013).

32.   Team., R.C. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria, (2022)*.

33.   Tian, Y. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **33**, 3982-3984 (2017).

34.   Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290 (2004).

35.   Kassambara, A. & Mundt, F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. *https://CRAN.R-project.org/package=factoextra* (2020).

36.   Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25**, 1-18 (2008).

37.   Pagès, H., Carlson, M., Falcon, S. & Li, N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. . *https://bioconductor.org/packages/AnnotationDbi*. (2023).

38.   Volpe, M. & Das, J. methylR: a graphical interface for comprehensive DNA methylation array data analysis. *Bioinformatics* (2023).

39.   Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

40.   Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).

41.   Tirado-Magallanes, R., Rebbani, K., Lim, R., Pradhan, S. & Benoukraf, T. Whole genome DNA methylation: beyond genes silencing. *Oncotarget* **8**, 5629-5637 (2017).

42.   Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

43.   Wouk, J., Rechenchoski, D.Z., Rodrigues, B.C.D., Ribelato, E. & Faccin-Galhardi, L.C. Viral infections and their relationship to neurological disorders. *Archives Virology* **166**, 733–753 (2021).

44. Han, H. *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* **46**, D380-D386 (2018).

45. Pinero, J., Sauch, J., Sanz, F. & Furlong, L.I. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput Struct Biotechnol J* **19**, 2960-2967 (2021).

46. Jain, S. *et al.* De Novo Lower Urinary Tract Symptoms in COVID-19 Patients. *Cureus* **15**, e33947 (2023).

47. Lee, J.S. & Shin, E.C. The type I interferon response in COVID-19: implications for treatment. *Nat Rev Immunol* **20**, 585-586 (2020).

48. Zenobia, C. & Hajishengallis, G. Basic biology and role of interleukin-17 in immunity and inflammation. *Periodontol 2000* **69**, 142-159 (2015).

49. Monwan, W., Kawasaki, T., Hasan, M.Z., Ori, D. & Kawai, T. Identification of nucleoporin 93 (Nup93) that mediates antiviral innate immune responses. *Biochem Biophys Res Commun* **521**, 1077-1082 (2020).

50. Kaplanski, G. Interleukin-18: Biological properties and role in disease pathogenesis. *Immunol Rev* **281**, 138-153 (2018).

51. Wickenhagen, A. *et al.* A prenylated dsRNA sensor protects against severe COVID-19. *Science* **374**, eabj3624 (2021).

52. Yu, M. *et al.* Delayed generation of functional virus-specific circulating T follicular helper cells correlates with severe COVID-19. *Nat Commun* **14**, 2164 (2023).

53. Chen, J. *et al.* Amyloid precursor protein facilitates SARS-CoV-2 virus entry into cells and enhances amyloid-beta-associated pathology in APP/PS1 mouse model of Alzheimer's disease. *Transl Psychiatry* **13**, 396 (2023).

54. Coperchini, F. *et al.* The cytokine storm in COVID-19: Further advances in our understanding the role of specific chemokines involved. *Cytokine Growth Factor Rev* **58**, 82-91 (2021).

55. Gonzalez-Jaramillo, V. *et al.* Epigenetics and Inflammatory Markers: A Systematic Review of the Current Evidence. *Int J Inflam* **2019**, 6273680 (2019).

56. Konwar, C. *et al.* Risk-focused differences in molecular processes implicated in SARS-CoV-2 infection: corollaries in DNA methylation and gene expression. *Epigenetics Chromatin* **14**, 54 (2021).

57.   Peterson, D.R. *et al.* Gene Expression Risk Scores for COVID-19 Illness Severity. *J Infect Dis* **227**, 322-331 (2023).

58.   Barturen, G. *et al.* Whole blood DNA methylation analysis reveals respiratory environmental traits involved in COVID-19 severity following SARS-CoV-2 infection. *Nat Commun* **13**, 4597 (2022).

59.   Winkley, K. *et al.* Immune cell residency in the nasal mucosa may partially explain respiratory disease severity across the age range. *Sci Rep* **11**, 15927 (2021).

60.   Gallo, O., Locatello, L.G., Mazzoni, A., Novelli, L. & Annunziato, F. The central role of the nasal microenvironment in the transmission, modulation, and clinical progression of SARS-CoV-2 infection. *Mucosal Immunol* **14**, 305-316 (2021).

61.   Balnis, J. *et al.* Whole-Genome Methylation Sequencing Reveals that COVID-19-induced Epigenetic Dysregulation Remains 1 Year after Hospital Discharge. *Am J Respir Cell Mol Biol* **68**, 594-597 (2023).

62.   Ehrlich, M. DNA hypermethylation in disease: mechanisms and clinical relevance. *Epigenetics* **14**, 1141-1163 (2019).

63.   Corley, M.J. *et al.* Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19. *J Leukoc Biol* **110**, 21-26 (2021).

64.   Biji, A. *et al.* Identification of COVID-19 prognostic markers and therapeutic targets through meta-analysis and validation of Omics data from nasopharyngeal samples. *EBioMedicine* **70**, 103525 (2021).

65.   Khezri, M.R., Varzandeh, R. & Ghasemnejad-Berenji, M. The probable role and therapeutic potential of the PI3K/AKT signaling pathway in SARS-CoV-2 induced coagulopathy. *Cell Mol Biol Lett* **27**, 6 (2022).

66.   Zechendorf, E. *et al.* A Potential Association between Ribonuclease 1 Dynamics in the Blood and the Outcome in COVID-19 Patients. *Int J Mol Sci* **24** (2023).

67.   Unterman, A. *et al.* Single-cell multi-omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19. *Nat Commun* **13**, 440 (2022).

68.   Sievers, B.L., Cheng, M.T.K., Csiba, K., Meng, B. & Gupta, R.K. SARS-CoV-2 and innate immunity: the good, the bad, and the "goldilocks". *Cell Mol Immunol* (2023).

69.   Yasuda, K., Nakanishi, K. & Tsutsui, H. Interleukin-18 in Health and Disease. *Int J Mol Sci* **20** (2019).

70.   Nasser, S.M.T. *et al.* Elevated free interleukin-18 associated with severity and mortality in prospective cohort study of 206 hospitalised COVID-19 patients. *Intensive Care Med Exp* **11**, 9 (2023).

71.   Coutinho, L.L. *et al.* Elevated IL-18 predicts poor prognosis in critically ill COVID-19 patients at a Brazilian hospital in 2020-21. *Future Microbiol* **17**, 1287-1294 (2022).

72.   Tiemi Enokida Mori, M. *et al.* Protective effects of IL18-105G > A and IL18-137C > Ggenetic variants on severity of COVID-19. *Cytokine* **174**, 156476 (2024).

73.   Lu, L., Zhang, H., Dauphars, D.J. & He, Y.W. A Potential Role of Interleukin 10 in COVID-19 Pathogenesis. *Trends Immunol* **42**, 3-5 (2021).

74.   Espin, E. *et al.* Cellular and molecular biomarkers of long COVID: a scoping review. *EBioMedicine* **91**, 104552 (2023).

75.   Cheong, J.G. *et al.* Epigenetic memory of coronavirus infection in innate immune cells and their progenitors. *Cell* **186**, 3882-3902 e3824 (2023).

76.   Banerjee, A.K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* **183**, 1325-1339 e1321 (2020).

77.   Fung, T.S. & Liu, D.X. Post-translational modifications of coronavirus proteins: roles and function. *Future Virol* **13**, 405-430 (2018).

78.   Mina, Y. *et al.* Deep Phenotyping of Neurologic Postacute Sequelae of SARS-CoV-2 Infection. *Neurol Neuroimmunol Neuroinflamm* **10** (2023).

79.   Iadecola, C., Anrather, J. & Kamel, H. Effects of COVID-19 on the Nervous System. *Cell* **183**, 16-27 e11 (2020).

80.   Wilk, A.J. *et al.* A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat Med* **26**, 1070-1076 (2020).

81.   Vanderheiden, A. *et al.* Type I and Type III Interferons Restrict SARS-CoV-2 Infection of Human Airway Epithelial Cultures. *J Virol* **94** (2020).

82.   Lieberman, N.A.P. *et al.* In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. *PLoS Biol* **18**, e3000849 (2020).

83.   Sun, J. *et al.* Comparative Transcriptome Analysis Reveals the Intensive Early Stage Responses of Host Cells to SARS-CoV-2 Infection. *Front Microbiol* **11**, 593857 (2020).

84. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045 e1039 (2020).

85. Falcao-Holanda, R.B., Brunialti, M.K.C., Jasiulionis, M.G. & Salomao, R. Epigenetic Regulation in Sepsis, Role in Pathophysiology and Therapeutic Perspective. *Front Med (Lausanne)* **8**, 685333 (2021).

Figure legends

Figure 1: Graphical summary of the study and methods. Nasopharyngeal samples were collected from 27 COVID-19 patients at the beginning of the study, i.e., inclusion, time point 1(T1); and followed up 6 weeks later (T2, 6W post inclusion), (N=36). Healthy controls (HC) (N=12) were included at inclusion. Genomic DNA (gDNA) was isolated and sent for bisulfite treatment and methylation analysis with the Illumina infinium Methylation EPICS 850K. Bioinformatics analysis was performed to identify unique CPGs and differentially methylation genes, followed by further analysis to assess the biological relevance.

Figure 2: Distinct separation of COVID-19 patients compared to healthy controls. Nasopharyngeal samples (N=36) were collected from 27 COVID-19 patients (P) at inclusion (T1) N=21 and 6-weeks post-inclusion (T2) N=15, and from healthy controls (N=12) (HC) at inclusion of the study. **(A)** Hierarchical clustering using the Euclidean distance performed on the β-value data matrix to visualize global correlation was performed and **(B)** Principal component analysis of normalized β-values of COVID-19 patients and healthy controls.

Figure 3: Distribution of differentially methylated CpGs and genes in COVID-19 patients compared to controls. Nasopharyngeal samples (N=36) collected from 27 COVID-19 patients at inclusion (T1) N=21 and 6-weeks post-inclusion (T2) N=15, were assessed for differential methylation within the different genomic regions: Body, Intergenic regions (IGR), 3' UTR and 5' UTR, 1st Exon, Exon boundaries (ExonBnd), and the transcriptions start sites TSS200 and TSS1500, and compared to healthy controls (N=12) (HC). **(A)** Venn diagram of unique and common differently methylated CpGs. **(B)** Genomic region distribution for Patients at T1 & T2 vs HC, Patients at T1 vs HC, and Patients at T2 vs HC. Significant set using the Bonferroni-Hochberg (BH)-corrected p-value (p-value BH) < 0.05 and log2FC cut-off (log2FC) set at ±0.3. **(C)** Volcano plots analyzing the $\log_2$ fold change ($\log_2 FC$) with a cut-off > 0.3 for hypermethylated CpGs and cut-off < 0.3 for hypomethylated CpGs and the use of BH corrected p values < 0.05 cut-off for hypermethylated and hypomethylated CpGs. Top 20 DMGs were annotated using the highest BH corrected p-values highest for hypermethylated and hypomethylated CpGs.

Figure 4: Clear distinctions in the methylation patterns of COVID-19 patients compared to healthy controls. Nasopharyngeal samples (N=36) collected from 27 COVID-19 patients (P) at inclusion (1) N=21 and 6-weeks post-inclusion (2) N=15, and healthy controls (N=12) (HC) were assessed for differential methylation pattern showing the top 20 000 data points hypo and hypermethylated CpGs.

Figure 5: Differentially methylated CpGs between COVID-19 patients and controls in transcription start sites TSS200 and TSS1500. Nasopharyngeal samples (N=36) collected from 27 COVID-19 patients (P) at inclusion (T1) N=21 and 6-weeks post-inclusion (T2) N=15 and healthy controls (N=12) (HC) were assessed in the TSS200 and TSS1500 regions for top 1000 hypo and hypermethylated CpG and 40 differently methylated CPG sites annotated randomly. Hierarchical clustering against sample groups was performed. Top annotation showed the distribution of Sample Groups, Gender, BMI, Age, CRP, LDH.

Figure 6: Enrichment analysis for COVID-19-specific differentially methylated genes in transcription start sites TSS200 and TSS1500. Nasopharyngeal samples (N=36) collected from 27 COVID-19 patients (P) at inclusion (1) N=21 and 6-weeks post-inclusion (2) N=15, and from healthy controls (N=12) (HC), were processed and evaluated for DNA methylation. Analysis was performed with Metascape [42] for **(A)** enriched terms and **(B)** biological processes for the top 1000 hypo- and hypermethylated genes.

Figure 7: Top canonical pathway and top regulatory networks in hospitalized COVID-19 patients. DNA methylation analysis was performed on nasopharyngeal samples from 27 COVID-19 patients (P) at inclusion (1) N=21 and 6-weeks post-inclusion (2) N=15, and from healthy controls (N=12) (HC). **(A)** Canonical pathway analysis was performed with Ingenuity Pathway Analysis to identify top activated and inhibited pathways linked to the methylated genes located in the transcription start sites 1500 and 200, with a z-score cut-off set to 1. Positive z-score indicated in orange, while negative z-scored indicated in blue. Significant TSS differentially methylated CpGs with a cut-off of +/- 0.3-fold-change was assessed in Ingenuity Pathway Analysis for **(B-C)** top regulator effect networks, and for **(D)** graphical summary.

Figure 8: Enrichment match analysis for top 1000 differently methylated genes across COVID-19-specific studies. DNA methylation analysis was performed on nasopharyngeal samples from 27 COVID-19 patients (P) at inclusion (1) N=21 and 6-weeks post-inclusion (2) N=15, and from healthy controls (N=12) (HC). Analysis match was performed against the top 5 selected datasets based on logP values and overlapping genes from COVID-19 studies to assess similarities between enrichment of genes. **(A)** Circos plot for overlapping gene lists and shared term level, where blue curves link genes belonging to the same enriched ontology term. The inner circle represents gene lists, where hits are arranged along the arc. Genes that hit multiple lists are colored in dark orange, and genes unique to a list are shown in light orange. **(B)** Network nodes identifying neighborhoods where proteins are densely connected. Heatmap of **(C)** enriched terms and **(D)** top-level gene ontology biological processes across input gene lists. Each data set has been numbered and color-coded: 1, RNA_Blanco-Melo_Lung_Down; 2, RNA_Sun_Calu-3_0h_Up; 3, RNA Vanderheiden-pHAE_48h_Up; 4, RNA_Lieberman_Nasopharynx_High_vs_Low_Down; 5, gene list from the present study; and 6, RNA_Wilk_B-cells_patient-C6_Up. Enrichment analysis was performed with Metascape and Coronascape [42].

Table 1. Clinical and-demographical characteristics of hospitalized COVID-19 patients.

Variable	Clinical data	Reference range
Number of COVID-19 patients	27	
Age, median (range)	56 (26–91)	
Body mass index, median (range)	28.7 (19.4–41.5)	
Biological sex, % (N)	44.4 F/55.6 M (12F/15M)	
Days in hospital, median (range)	6 (2-22)	
ICU/pandemic Ward %, (N)	7.4/92.6 (2/25)	
Days with symptoms before inclusion, median (range)	9 (2-20)	
Spike IgG antibody positive at inclusion, % (N)	63 (17)	
Nucleocapsid IgG antibody positive at inclusion, % (N)	77.8 (21)	
Viral load at inclusion (copies/ml), median (range)	8721.43 (1071.43-3.4x10^7)	
Antiviral treatment, % (N)	22.2 (6)	
Corticosteroid treatment, % (N)	59.3 (16)	
No/oxygen/HFNOT:CPAP[1]/mechanical ventilation, % (N)	7.4/44.4/48.1/3.7 (2/12/13/1)	
Cardiovascular disease, % (N)	63 (17)	
Pulmonary disease, % (N)	33.3 (9)	
Diabetes mellitus, % (N)	25.9 (7)	
Two of the underlying conditions, % (N)	25.9 (7)	
Disease score: moderate/severe, % (N)	92.6 (25)/7.4 (2)	
Ongoing smoking/snus, % (N)	0	
Previous history of smoking/snus, % (N)	51.9 (14)	
Leukocytes (x 10^9/L), median (range)	6.7 (2.8-21.2)	3.5-8.8
Thrombocytes (x 10^9/L), median (range)	239 (134–458)	150–400
Lymphocytes (x 10^9/L), median (range)	1.1 (0.4–2.1)	1.1–4.8
Monocytes (x 10^9/L), median (range)	0.4 (0.1-2.24)	0.1–1
Lactate dehydrogenase (µKat/L), median (range)	5.3 (2.5–16)	>70 years <3.5, <70 years < 4.3
C-reactive protein (mg/L), median (range)	40 (0–318)	0–10

Table 2: Top hyper- and hypomethylated genes with unique differentially methylated CpGs in COVID-19 patients compared to controls.

Comparison	Significantly* Hypermethylated genes	Significantly* Hypomethylated genes
COVID-19 patients at inclusion (T1) and 6-week post inclusion (T2) vs HC	SGIP1, **RAD54L2**, SH3GL3, **TMEM63C**, PACRG, **RUNX1T1**, **CTNNA2**, **PRLR**, LOC101929153, **SPOCK3**, CORIN, **HPS5**, MAP7, LINC0147, **TMEM212**, **SHANK2**, MOBP, PADI4, WT1, **LOC285780**	TMEM2, **MECOM**, XBP1, **DLX5**, GPR77, **DLX5, DLX5**, SLC39A10, PDS5A, XBP1, **DLX5**, VIM, **DLX5**, COL4A1, INTS9, **PLEKHM3,** PPP1CA, FILIP1
Patients T1 vs HC	**RAD54L2**, **SPOCK3**, SMC6, **SHANK2**, **RUNX1T1**, MALSU1, SAR1B, GMDS, **HPS5**, SPATA6L, **CTNNA2**, **TMEM212**, **LOC285780,** TIPARP, LOC339862, **PRLR**, **TMEM63C**, RASGRP4, NEK7, TSPAN5	**DLX5,** ZNF664-FAM101A, LPIN1, **TMEM2**, **PLEKHM3,** STAT5B, YY1, WHSC1L1, EXT1, **DLX5**, LOC101928650, SGK1, **COL4A1,** LIMS1, **MECOM**, DENND5A, ZNF622, C8orf75, ZFAND3, PINX1
Patients T2 vs HC	FBN1, TSHZ2, GNA13, CRYL1, EEPD1, CRYL1, SPIDR, MCL1, RPS6KA2, NSUN6, BAG3, TK2, LOC101927630, TC39C, FAM3B, IKZF3, MAP3K13, RHOU, FYN, PPM1H	TNFAIP8, PARL, NISCH, INPP5D, NOD2, PMP22, EML4, FILIP1L, LOC100506844, GIT2, SIK3, VPS33A, AATK, SORL1, CANX, DGKG, TNIK, SNTN1, LOC400867

log$_2$FC cut-off was set to ±0.3. Significant set as Bonferroni-Hochberg corrected p value <0.05.
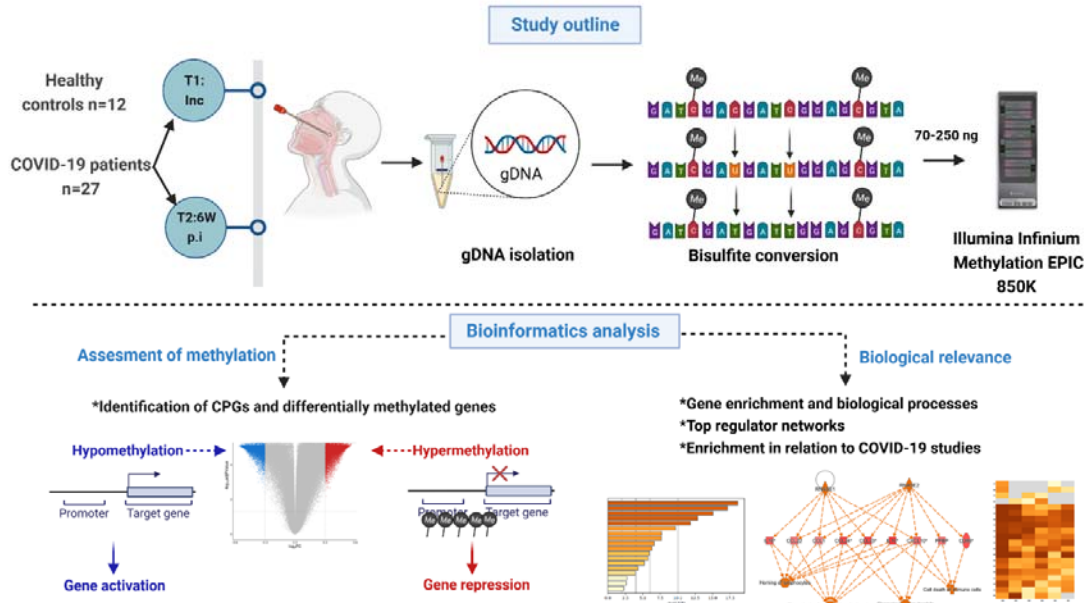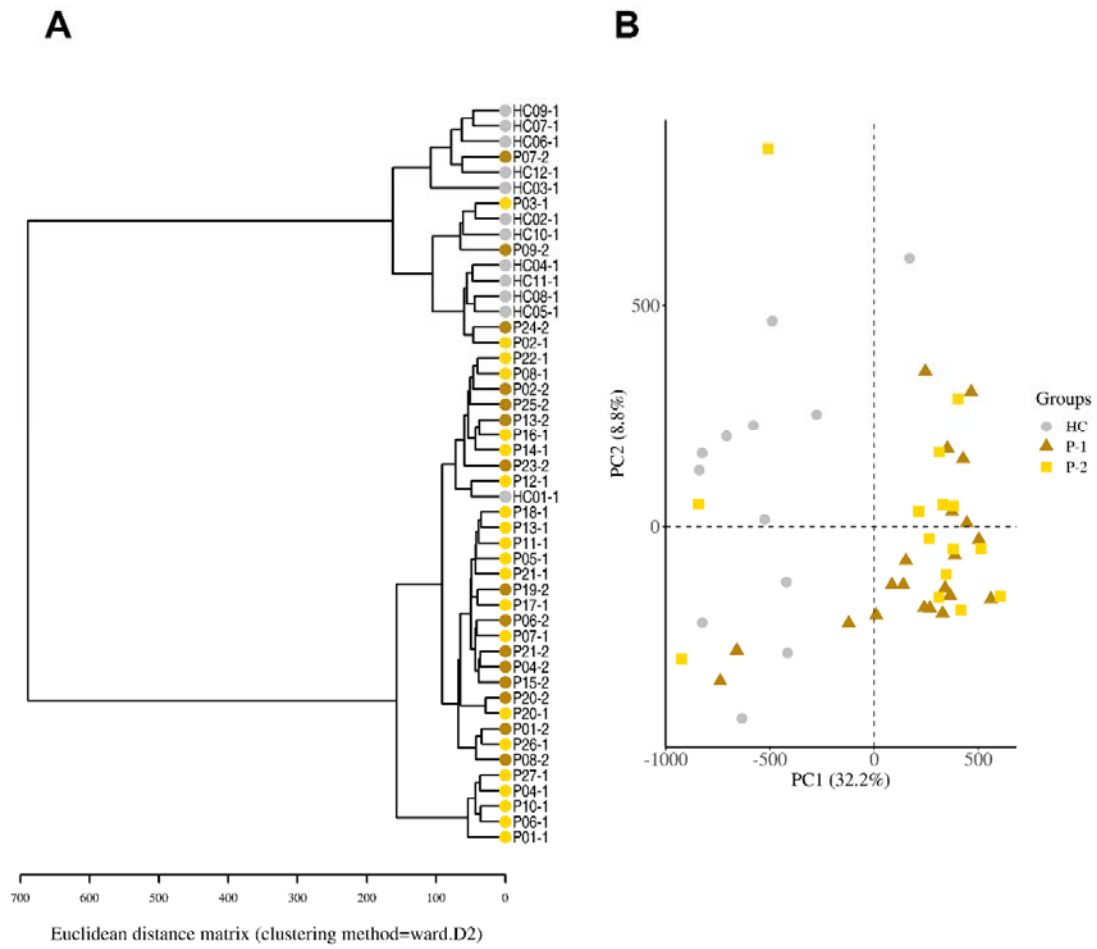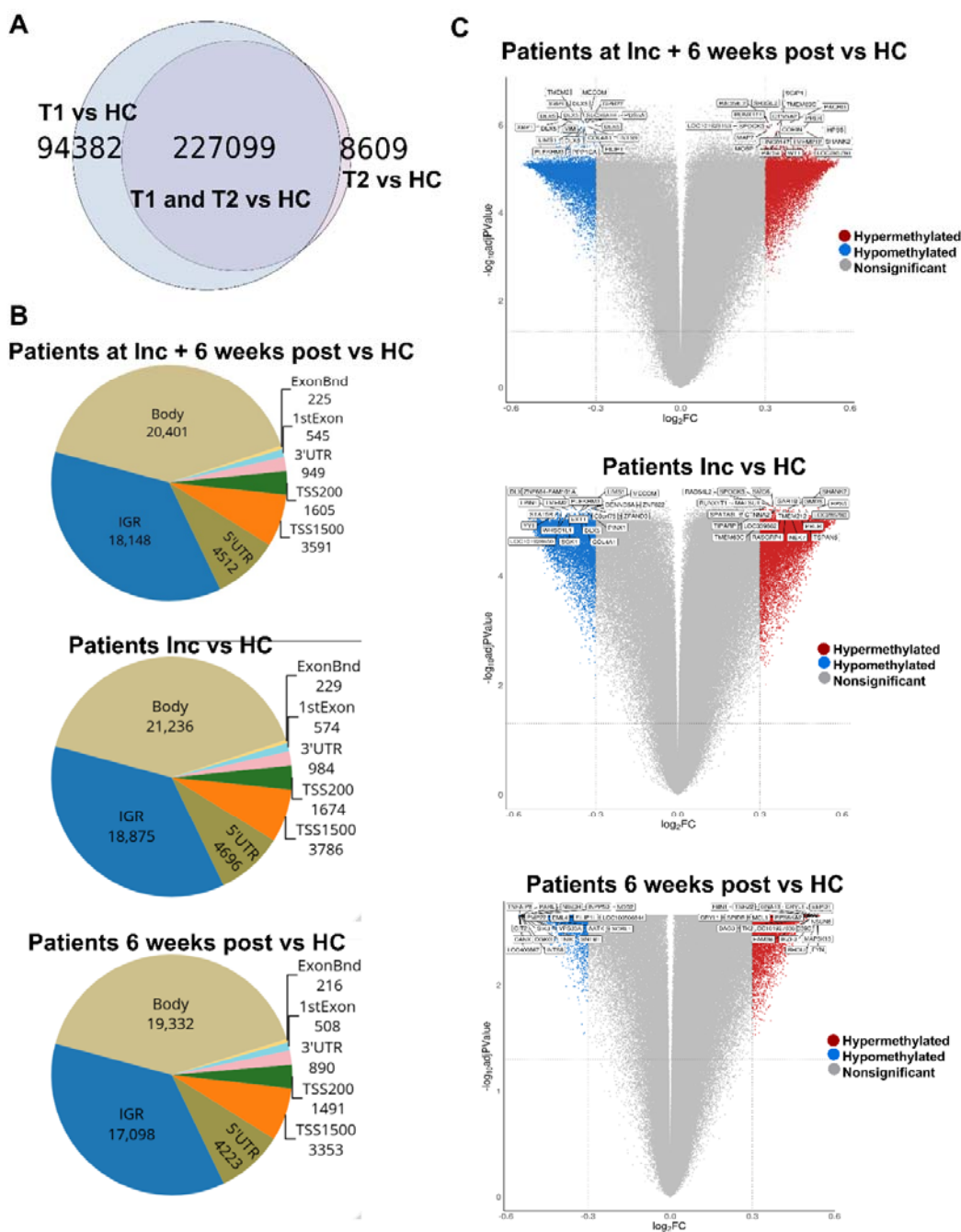
Figure 1

Figure 2

A

B

Figure 3

Figure 4

Figure 5

Figure 6

A   Top enriched Ontology terms

GO:0050865: regulation of cell activation
GO:0045321: leukocyte activation
GO:0006954: inflammatory response
R-HSA-6798695: Neutrophil degranulation
GO:0001819: positive regulation of cytokine production
GO:0050900: leukocyte migration
GO:0002697: regulation of immune effector process
GO:0002274: myeloid leukocyte activation
GO:0050778: positive regulation of immune response
GO:0071345: cellular response to cytokine stimulus
WP3937: Microglia pathogen phagocytosis pathway
GO:0032655: regulation of interleukin-12 production
R-HSA-1280215: Cytokine Signaling in Immune system
GO:1902105: regulation of leukocyte differentiation
GO:0043408: regulation of MAPK cascade
GO:1903555: regulation of tumor necrosis factor superfamily cytokine production
GO:0006909: phagocytosis
GO:0030029: actin filament-based process
GO:0002687: positive regulation of leukocyte migration
GO:0032753: positive regulation of interleukin-4 production

-log10(P)

B   Top enriched Biological processes

GO:0032501: multicellular organismal process
GO:0002376: immune system process
GO:0050896: response to stimulus
GO:0048518: positive regulation of biological process
GO:0050789: regulation of biological process
GO:0032502: developmental process
GO:0048519: negative regulation of biological process
GO:0044419: biological process involved in interspecies interaction between organisms
GO:0023052: signaling
GO:0040011: locomotion
GO:0051179: localization
GO:0009987: cellular process
GO:0065007: biological regulation
GO:0008152: metabolic process
GO:0040007: growth
GO:0016032: viral process
GO:0042592: homeostatic process
GO:0098754: detoxification

-log10(P)

Figure 7

Figure 8