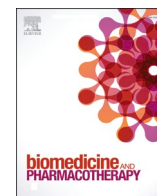




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Epigenetic activation of antiviral sensors and effectors of interferon response pathways during SARS-CoV-2 infection

Jan Bińkowski<sup>a,1</sup>, Olga Taryma-Leśniak<sup>a,1</sup>, Karolina Łuczowska<sup>b</sup>, Anna Niedzwiedz<sup>c</sup>, Kacper Lechowicz<sup>d</sup>, Dominik Strapagiel<sup>e</sup>, Justyna Jarczak<sup>e,f</sup>, Veronica Davalos<sup>g</sup>, Aurora Pujol<sup>h,i,j</sup>, Manel Esteller<sup>g,k,l,m</sup>, Katarzyna Kotfis<sup>d</sup>, Bogusław Machaliński<sup>b</sup>, Miłosz Parczewski<sup>n</sup>, Tomasz K. Wojdacz<sup>a,o,\*</sup>

<sup>a</sup> Independent Clinical Epigenetics Laboratory, Pomeranian Medical University in Szczecin, Szczecin, Poland

<sup>b</sup> Department of General Pathology, Pomeranian Medical University in Szczecin, Szczecin, Poland

<sup>c</sup> Department of Internal Medicine and Diabetology with the Subdepartment of Endocrinology, Independent Public Regional Hospital in Szczecin, Szczecin, Poland

<sup>d</sup> Department of Anaesthesiology, Intensive Therapy and Acute Intoxications, Pomeranian Medical University in Szczecin, Szczecin, Poland

<sup>e</sup> Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland

<sup>f</sup> Laboratory of Molecular Basis of Behavior, Nencki Institute of Experimental Biology, Polish Academy of Science, Warsaw, Poland

<sup>g</sup> Cancer Epigenetics Group, Josep Carreras Leukaemia Research Institute (LJC), Badalona, Barcelona, Catalonia, Spain

<sup>h</sup> Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain

<sup>i</sup> Center for Biomedical Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain

<sup>j</sup> Institutio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

<sup>k</sup> Centro de Investigación Biomédica en Red Cáncer (CIBERONC), Madrid, Spain

<sup>l</sup> Institutio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

<sup>m</sup> Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain

<sup>n</sup> Department of Infectious, Tropical Diseases and Immune Deficiency, Pomeranian Medical University in Szczecin, Szczecin, Poland

<sup>o</sup> Department of Biomedicine, Aarhus University, Aarhus, Denmark

## ARTICLE INFO

**Keywords:**  
COVID-19  
Coronavirus  
SARS-CoV-2  
Epigenetics  
DNA methylation

## ABSTRACT

Recent studies have shown that methylation changes identified in blood cells of COVID-19 patients have a potential to be used as biomarkers of SARS-CoV-2 infection outcomes. However, different studies have reported different subsets of epigenetic lesions that stratify patients according to the severity of infection symptoms, and more importantly, the significance of those epigenetic changes in the pathology of the infection is still not clear. We used methylomics and transcriptomics data from the largest so far cohort of COVID-19 patients from four geographically distant populations, to identify casual interactions of blood cells' methylome in pathology of the COVID-19 disease. We identified a subset of methylation changes that is uniformly present in all COVID-19 patients regardless of symptoms. Those changes are not present in patients suffering from upper respiratory tract infections with symptoms similar to COVID-19. Most importantly, the identified epigenetic changes affect the expression of genes involved in interferon response pathways and the expression of those genes differs between patients admitted to intensive care units and only hospitalized. In conclusion, the DNA methylation changes involved in pathophysiology of SARS-CoV-2 infection, which are specific to COVID-19 patients, can not only be utilized as biomarkers in the disease management but also present a potential treatment target.

## 1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected almost 510 million individuals and claimed over 6.2 million

lives worldwide as of April 25, 2022 (<https://origin-coronavirus.jhu.edu/map.html>). The vast majority of SARS-CoV-2 infections is characterized by a range of mild symptoms including fever, cough, and general malaise [1]. About 14% of patients progress to severe disease (with

\* Corresponding author at: Independent Clinical Epigenetics Laboratory, Pomeranian Medical University in Szczecin, Szczecin, Poland.

E-mail address: [tomasz.wojdacz@pum.edu.pl](mailto:tomasz.wojdacz@pum.edu.pl) (T.K. Wojdacz).

<sup>1</sup> These authors contributed equally to this work

<https://doi.org/10.1016/j.bioph.2022.113396>

Received 16 May 2022; Received in revised form 1 July 2022; Accepted 7 July 2022

Available online 11 July 2022

0753-3322/© 2022 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

dyspnoea, hypoxia, or greater than 50% lung involvement on imaging tests) and further 5% develop critical disease, characterized by acute respiratory distress syndrome (ARDS), and multi-organ failure (MOF) requiring mechanical ventilation in the Intensive Care Units (ICUs) [2]. The severity of COVID-19 disease, caused by SARS-CoV-2 infection, has been generally associated with dysregulation of immune activity [3], including increased levels of cytokines [4,5] produced by a subset of inflammatory monocytes [6,7], lymphopenia [8], and T cell exhaustion [9,10]. The epidemiological studies repeatedly show that sex and older age plus one of the following conditions: diabetes, obesity, hypertension, and cardiovascular pathology are the risk factors for severe outcome of the disease [7,11].

Those risk factors, however, cannot explain the high heterogeneity and unpredictability of the COVID-19 outcomes observed in clinics. Prediction of individual response to the SARS-CoV-2 infection is one of the biggest challenges of the COVID-19 patient clinical management [12–15]. A number of genome-wide association studies (GWAS) have identified genetic loci that appear to modulate the risk of severe outcomes of the infection [16–18]. The prevalence of one of those loci, a haplotype derived from Neandertal DNA, was associated with the COVID-19 mortality in the populations with a high prevalence of that haplotype [19]. However, the populations in which this haplotype is almost absent also display high mortality rates. That suggests a key role of the environmental factors in the disease outcomes.

DNA methylation is one of the major epigenetic mechanisms of gene expression regulation, essential during development and for the maintenance of cell-type identity [20]. In principle, it is a process of covalent addition of methyl groups to the cytosines in DNA strand and cell specific genome-wide methylation pattern is referred to as methylome. DNA methylation is an enzymatic process that can be influenced by environmental factors and several studies have reported significant differences between methylomes of specific populations, but also individuals (as reviewed for example in: [21,22]). Moreover, the acquisition of specific DNA methylation changes in blood has already been shown by us and others to predispose to the disease [23]. Recent studies have confirmed that SARS-CoV-2 infection influences the methylomes of the blood cells [24–26]. However, the results of those studies agree only to some extent, suggesting that repose of the methylome to the infection may be host specific and vary between patients and populations of patients.

We have consolidated methylomics and transcriptomics data from studies that aimed to elaborate SARS-CoV-2 virus influence on blood cells including: two studies from USA [24,26], a study from Spain [25] and a patient cohort from our institution with the aim to investigate the causal effect of the host methylome interaction with the pathogen. Overall, we observed uniform response of methylome of blood cells to SARS-CoV-2 virus in 478 hospitalized COVID-19 patients from geographically distant and ethnically mixed populations. The identified methylation signature was specific to COVID-19 patients and not present in blood of healthy controls and patients with upper respiratory tract infections. We then assessed the expression of genes associated with those methylation changes and shown that expression of at least a subset of those genes is regulated by SARS-CoV-2 induced methylation changes.

## 2. Material and methods

### 2.1. Data collection

All used datasets are deposited at Gene Expression Omnibus (GEO) database. Methylation data of COVID-19 cohorts at: GSE168739, GSE174818, GSE167202, and GSE202296; other infectious respiratory disease cohorts at: GSE174818 and GSE167202; and healthy individuals at: GSE167202, GSE112618, GSE123914, and GSE153211. Expression (RNA-seq) data of COVID-19 and non-COVID-19 patients at: GSE157103.

### 2.2. Ethics

Written consent was obtained from all participants from Polish cohort before enrollment. All procedures were in accordance with the ethical standards approved by the Bioethics Committee of the Pomeranian University, Szczecin, with approval number KB-0012/101/2020 -A.

### 2.3. Methylation profiling data processing

Raw microarray files (.idat) were processed using the ChAMP R package [27]. As the data exceeded available computational abilities, we prepared an R script to process chunks of data in sequential manner. Each chunk ( $n = 50$ ) was drawn randomly without replacement from the sampling frame and then processed into a beta-matrix using standard ChAMP pipeline for EPIC array. Finally, all beta-matrices were combined into one data matrix file for further analysis. Code available from: <https://github.com/ClinicalEpigeneticsLaboratory/distChAMP>. Raw beta values were corrected for cell-fraction proportions using the refBase function of the ChAMP package. However, to predict WBC fractions proportion we used outperforming robust partial correlation method implemented in EpiDISH R package [28–30], instead of the default reference-based algorithm. As a reference methylation profiles, we used “centDHSbloodDMC.m” dataset for seven main WBC types (B-cells, CD4T, CD8T, NK, Monocytes, Neutrophils and Eosinophils) [28]. We defined differentially methylated probe (DMP) as a probe with more than 0.05 absolute average methylation difference between compared groups and we used ANOVA test to calculate p-value for specific CpG if the distribution of beta values in each of the comparison groups was normal (Shapiro-Wilk test) and variance was equal (Bartlett test). If the beta values were not normally distributed or variances between groups were different, we used non-parametric Kruskal-Wallis test. To include additional covariates in analysis we used logit or multinomial logit regression models, then significance of methylation level change was assessed using the Wald test. We controlled for false discovery rate at the level 0.05 using step-up Benjamini-Hochberg procedure.

### 2.4. Gene expression data processing

The RNA-seq data from COVID-19 and other respiratory infections patients (GSE157103) were analysed as TPM (Transcript Per Million). The logit regression model was used to test statistical significance of observed expression differences between studied groups of patients. We considered the expression to be altered if absolute value of  $\log_2$  fold-change of studied groups was  $\geq 1$  and FDR corrected p-value calculated for gene expression variable in logit model was  $\leq 0.05$  (Wald test). The models were adjusted for: sex, age, WBC fractions and steroids intake. To fulfill independence between covariates we removed all covariates with VIF (variance inflation factor)  $> 6$  (in case of continuous variables) or p-value  $\leq 0.05$  (Fisher exact test) in pairwise comparison of covariates (in case of categorical variables).

### 2.5. High-dimensional data visualisation

Data with more than three dimensions were visualised in the form of heatmaps clustered using unsupervised Ward's algorithm or alternatively using scatterplots based on data transformed by t-SNE. In t-SNE algorithm, we set perplexity equal to the smallest number of samples in analysed groups and the gradient calculation was performed using “exact” method. The rest of parameters were set as default, implemented in scikit-learn library. In both methods, we assumed Euclidian distance.

### 2.6. The functional enrichment analysis

Gene Set Enrichment Analysis was performed using GENE2FUNC function of FUMA GWAS (Functional Mapping and Annotation of

Genome-Wide Association Studies) [31] with Molecular Signatures Database (MSigDB) as a reference [32,33]. This database contains “hallmark” collection of gene sets [34], which are one of the most reliable databases for GSEA as they were generated by *in silico* identification of the overlaps between gene sets in other MSigDB collections, followed by manual evaluation and annotation of the genes [34]. The evidence channels used in STRING analysis were: (1) textmining, which searches for protein names in PubMed abstracts, in an in-house collection of more than three million full text articles, and in other text collections [35,36], (2) experiments, which includes evidence from actual laboratory experiments, based on interaction databases organized in the IMEx consortium [37] and BioGRID [38], (3) database, which collects evidence imported from pathway databases (asserted by a human expert curator), and (4) co-expression, in which gene expression data originating from a variety of expression experiments are normalized, filtered and then correlated [39] (Table 3).

### 3. Results

#### 3.1. Blood cell composition of clinical sample needs to be considered in analyses of causal methylation changes

In response to SARS-CoV-2 infection the ratios of sub-populations of blood cells may temporally change and the magnitude of those changes differs between infected individuals [5,40]. Increasing number of data suggests that to identify causal methylation pattern alterations in individual cell types, the correction for the proportion of different cells in the individual sample should be performed especially when analysing data from independent studies and populations [28, 41–43]. Thus, we first evaluated to what extent the analysis of SARS-CoV-2 infection-related methylation changes in our study can potentially be affected by sample specific white blood cells (WBC) proportions. Using reference-based Robust Partial Correlations-RPC (EpiDISH) [28] we inferred the proportions of the cells from EPIC microarray data for each of the blood samples studied. The levels of each of the cell types significantly varied between cohorts of COVID-19 patients as well as healthy controls (Supplementary Fig. 1a). Moreover, mean or variance of methylation levels, were significantly different ( $p$ -value  $\leq 0.05$ , Kruskal-Wallis and Levene test, respectively) for all cell fractions except eosinophils, between the groups. We then performed cell-fraction correction (CFC) of individual methylation profiles with refBase function of ChAMP package [27,29,30], modified to use reference-based Robust Partial Correlations-RPC (implemented in EpiDISH package) [28,29]. The proportions of specific blood cell fractions in patients as well as healthy controls showed relatively minor differences after CFC, indicating the need to perform this type of correction before the analysis of the causal methylation changes (Supplementary Fig. 1b).

To confirm positive effect of cell-fraction correction we also regressed beta values for 10,000 randomly selected CpG sites with percentage of blood cell fractions across all methylation profiles in the study to calculate adjusted R-square value of that regression. This value in principle reflects the strength of linear association between beta values (methylation levels) and proportions of the cells in the sample. In our data, up to 94% (median: 0.06; interquartile range (IQR): 0.20) of methylation level at specific CpG sites was potentially attributed to different cell proportions and the influence of cell fractions on beta values is significantly reduced by cell-fraction correction that we performed (Supplementary Fig. 1c). These results again show the need to perform cell fraction correction when studying the methylome blood cells and our subsequent analyses described below are based on corrected data.

#### 3.2. Uniform response of blood methylome to SARS-CoV-2 infection

Majority of the COVID-19 patients in our study, including all patients in Polish ( $n = 32$ ) and USA 1 cohorts ( $n = 102$ ) [24], as well as 213

patients in Spanish [25] and 131 in USA 2 [26] were patients with symptoms that required hospitalization. We began our analysis of the effect of SARS-CoV-2 on methylomes of blood cells with a comparison of methylation profiling data of those patients with the data from 119 healthy blood samples that we considered a healthy control group (the outline of this analysis is shown in Supplementary Fig. 2a and Table 1 includes clinical characteristics of the cohorts). Differentially methylated probe (DMP) in our study was a probe with statistically significant (FDR corrected  $p$ -value  $\leq 0.05$ ; Kruskal-Wallis or ANOVA) average methylation level change (hyper- or hypomethylation) in each of the studied cohorts in reference to controls and the same direction of that change across all cohorts. Additionally, as we used data from different studies and expected a certain level of the batch effect and technological noise, we included in this analysis only probes with methylation level difference of more than 5% points. This analysis identified 1773 DMPs (Supplementary Table 1) of which 1601 (90.3%) were hypo- and 172 (9.7%) hypermethylated. Not surprisingly, the unsupervised clustering based on identified DMPs was able to effectively distinguish two analysed groups (Fig. 1a), but importantly patients did not cluster in experiment dependent fashion indicating lack of batch effect in our data set. Both clustering and t-Distributed Stochastic Neighbour Embedding (t-SNE) based visualisation (Fig. 1b) showed high heterogeneity of methylation levels at identified DMPs across patients with even some of the methylation profiles of healthy individuals mixing with the COVID-19 patients.

The identified DMPs annotated to 1112 genes. To approximate biological function of those genes we used GENE2FUNC function of FUMA GWAS (Functional Mapping and Annotation of Genome-Wide Association Studies) [31] and performed gene set enrichment analysis (GSEA) with Molecular Signatures Database (MSigDB) as a reference [32,33], which contains “hallmark” collection of gene sets [34]. Despite that our analyses identified rather large gene set, those genes were statistically significantly enriched in only two terms: interferon alpha (adjusted  $p$  value =  $2.74E-07$ ) and interferon gamma response pathways (adjusted  $p$  value =  $1.51E-02$ ) (Fig. 1c), with enrichment of 21 of the genes in each of the pathways and 15 genes shared between both pathways. Not surprisingly, this suggests that identified methylation changes take part in interferon response activated upon viral infection (as reviewed e. g. [44]). However, given the accuracy of GSEA analysis normally observed in similar experiments, identification of only two pathways in our analysis indicates a very high specificity of the methylation changes induced by SARS-CoV-2.

#### 3.3. Majority of methylome changes observed in COVID-19 patients is present in patients with other respiratory infections

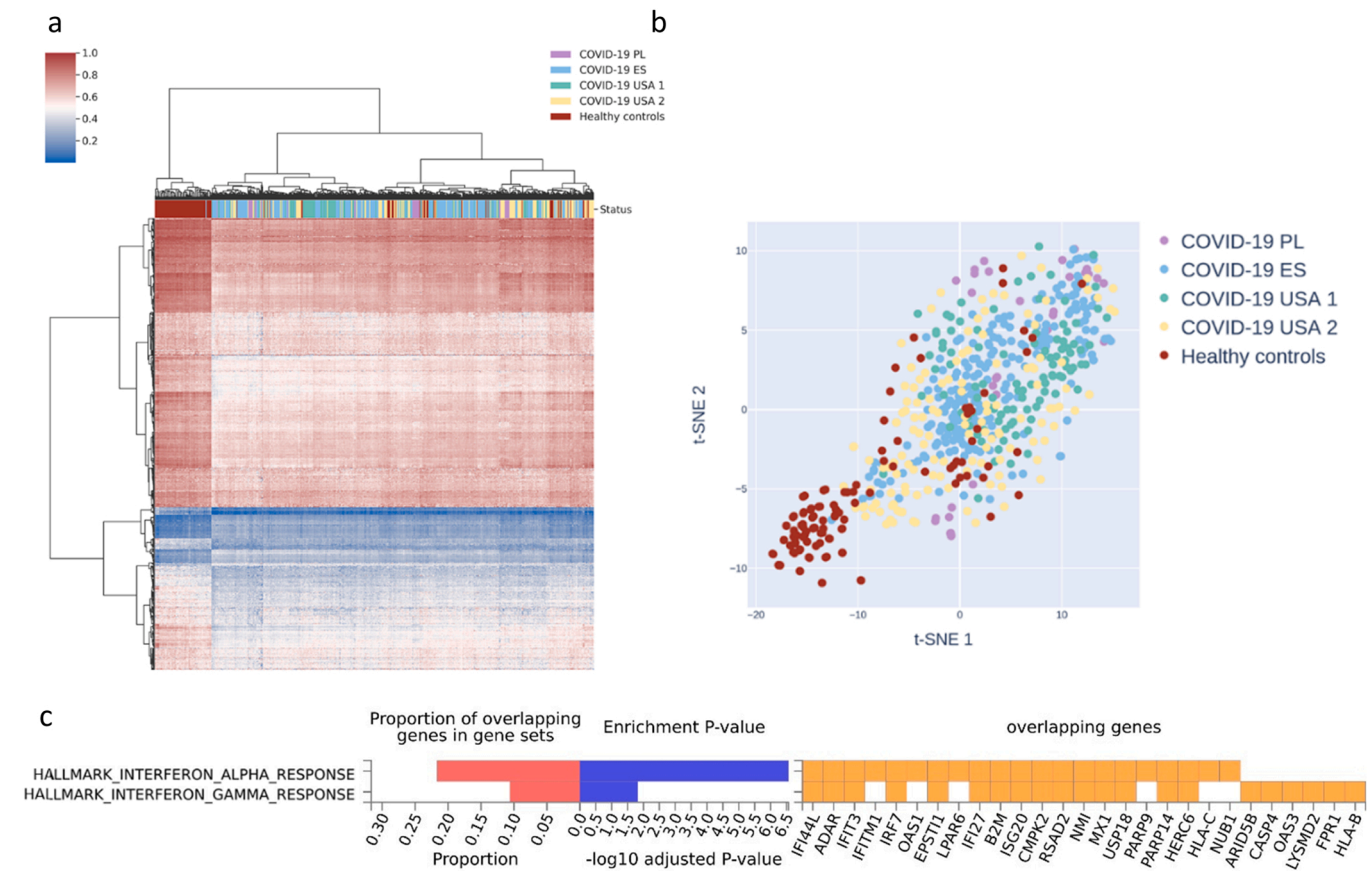
Two studies: USA 1 ( $n = 26$ ) [24] and USA 2 ( $n = 65$ ) [26], reported blood cells methylation profiling data from patients that were tested negative for SARS-CoV-2 but developed similar to COVID-19 disease symptoms. The disease etiology of patients in USA 1 cohort was unknown but most of the cases developed pneumonia and in the USA 2 cohort all patients were positive for other than SARS-CoV-2 acute upper respiratory viral infections (for details see Supplementary Table 2). Considering the differences in diseases' etiology in both cohorts, we analysed methylomes patients from each of those cohorts separately. In general, unsupervised clustering (Fig. 2a-b), as well as t-SNE (Fig. 2c-d) analysis showed that COVID-19 patients and patients with other respiratory infections had a very similar methylation status at the subset of DMPs we identified between hospitalized COVID-19 patients and controls. That suggests that most of the findings in our initial analysis methylation changes are a part of the general response of the immune system to the respiratory tract infections. Nevertheless, especially in case of USA 1 cohort, both unsupervised clustering (Fig. 2a) and t-SNE (Fig. 2c) suggested certain level of clustering of patients with other respiratory disease from COVID-19 patients and hence the presence of the subset of DMPs differentially methylated between these two groups.



**Table 1**  
Clinical characteristics of COVID-19 patients of all studied populations, non-COVID-19 patients, as well as healthy controls.

|                            | COVID-19 PL      | COVID-19 ES [25] | COVID-19 USA-1 [24] | non-COVID-19 USA-1 [24] | COVID-19 USA-2 [26] | non-COVID-19 USA-2 [26] | Healthy controls [26, 56–58] |
|----------------------------|------------------|------------------|---------------------|-------------------------|---------------------|-------------------------|------------------------------|
| Total, n                   | 32               | 407              | 102*                | 26                      | 164                 | 65                      | 119                          |
| Age (years), mean (95% CI) | 48.7 (44.5–52.8) | 42.1 (41.1–43.1) | 61.3 (58.1–64.5)    | 63.8 (57.3–70.4)        | 50.5 (47.9–53.2)    | 54.1 (50.0–58.2)        | 53.2 (50.6–55.7)             |
| Sex, n(%)                  |                  |                  |                     |                         |                     |                         |                              |
| Male                       | 19 (59.4%)       | 185 (45.5%)      | 64 (62.7%)          | 13 (50%)                | 93 (56.7)           | 35 (53.8)               | 18 (15.1%)                   |
| Female                     | 13 (40.6%)       | 222 (54.5%)      | 38 (37.3%)          | 13 (50%)                | 71 (43.3)           | 30 (46.2)               | 101 (84.9%)                  |
| BMI, mean (95% CI)         | 28.6 (26.9–28.7) | < 30             | 30.4 (28.4–32.4)    | 30.4 (26.7–34.0)        | N/A                 | N/A                     | N/A                          |
| Smoking history            | 14 (43.8%)       | N/A              | 18 (17.6%)          | 10 (38.5%)              | N/A                 | N/A                     | N/A                          |
| Diabetes mellitus, n(%)    | 2 (6.3%)         | 0 (0%)           | 36 (35.3%)          | 6 (23.1%)               | N/A                 | N/A                     | N/A                          |
| Hypertension, n(%)         | 9 (28.1%)        | 0 (0%)           | N/A                 | N/A                     | N/A                 | N/A                     | N/A                          |
| Pulmonary disease, n(%)    | 2 (9.4%)         | 0 (0%)           | 21 (20.6%)          | 4 (15.4%)               | 21 (12.8%)          | 8 (12.3%)               | N/A                          |
| ICU, n(%)                  | 0 (0%)           | 99 (24.3%)       | 51 (50%)            | 16 (61.5%)              | N/A                 |                         |                              |
| Severity group, n(%)       |                  |                  |                     |                         |                     |                         |                              |
| Hospitalised               | 32 (100%)        | 213 (52.3%)      | 102 (100%)          | N/A                     | 131 (79.9%)         | N/A                     | N/A                          |
| Asymptomatic/mild symptoms | 0 (0%)           | 194 (47.7%)      | 0 (0%)              | N/A                     | 33 (20.1%)          | N/A                     | N/A                          |
| COVID-19 pneumonia         |                  |                  |                     |                         |                     |                         |                              |
| Yes                        | 25 (78.1%)       | 203 (49.9%)      | N/A                 | N/A                     | N/A                 | N/A                     | N/A                          |
| No                         | 7 (21.9%)        | 184 (45.2%)      | N/A                 | N/A                     | N/A                 | N/A                     | N/A                          |
| Unknown                    | 0 (0%)           | 20 (4.9%)        | N/A                 | N/A                     | N/A                 | N/A                     | N/A                          |

\* Expression data with clinical characteristics used as covariates in logistic regression were available for 99 patients

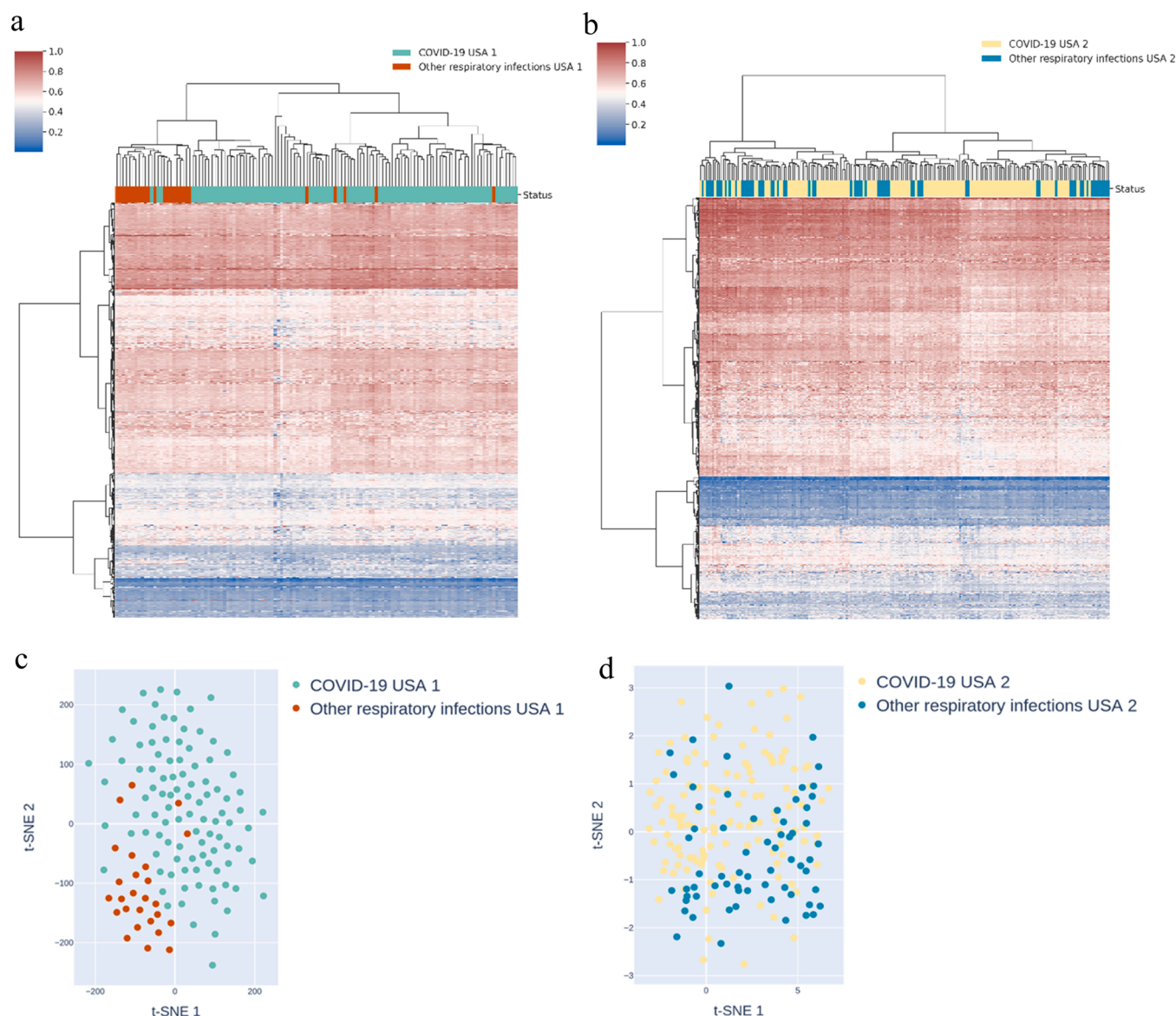


**Fig. 1.** Comparison of methylation levels in 1773 DMPs displaying identical methylation changes in all COVID-19 cohorts and GSEA analysis of the genes annotated to those DMPs.

(a) Heatmap illustrating unsupervised clustering of beta values at identified DMPs in COVID-19 patients and healthy controls.

(b) t-SNE based visualisation of beta values of identified DMPs in COVID-19 patients and healthy controls.

(c) illustration of GSEA results from FUMA GWAS platform with hallmark MSigDB database as a reference.



**Fig. 2.** Comparison of methylation changes at DMPs identified in all COVID-19 cohorts between COVID-19 patients and patients with other respiratory tract infections. a-b Heatmap illustrating unsupervised clustering based on beta values of this subset of DMPs for USA 1 (a) and USA 2 (b) cohort. c-d t-SNE based visualisation of beta values of this subset of DMPs for USA 1 (a) and USA 2 (b) cohort.

### 3.4. A subset of methylation changes is specific to SARS-CoV-2 infection

Given the specificity of the GSEA analysis and as the above results suggested that there might be a subset of methylation changes specific to SARS-CoV-2 infection, we modified our initial analysis and compared methylomes of the COVID-19 patients in each cohort with both healthy blood and two of the other respiratory infections patient cohorts (Supplementary Fig. 2b).

This analysis identified 20 hypo- and 4 hypermethylated DMPs (Fig. 3a and Supplementary Table 3) displaying methylation changes identical in each of the analysed cohorts of COVID-19 patients and different from both healthy controls and patients with other respiratory tract infections. Those DMPs mapped to 17 genes, which are listed in Table 2. Twelve of those genes including: *AIM2*, *CYSTM1*, *DTX3L*, *PARP9*, *CMPK2*, *EPSTI1*, *IFI44L*, *IFIT3*, *IRF7*, *LRBA*, *MX1*, and *TRIM22* have been described to be associated with immune response (source: GeneCards) and all except *LRBA* were annotated as interferon-regulated genes (Interferome database [45]). All the DMPs mapped to those genes were hypomethylated. Interestingly, three of the identified COVID-19

specific DMPs (cg22930808, cg00959259, cg07815522) were located at *PARP9/DTX3L* promoter region (TSS1500, 5'UTR) in N-shore of CpG island present in that genomic region (Fig. 3b and Supplementary Table 3), and four DMPs (cg16785077, cg13155430, cg21549285, cg22862003) mapped to the shores or shelves of two CpG islands presents in promoter of *MX1* (Fig. 3c and Supplementary Table 3). In case of *PARP/DTX3L* observed methylation changes affected adjacent CpG site and formed a differentially methylated region (DMR) (Fig. 3b), while at *MX1* gene DMPs were distributed across the promoter region (Fig. 3c). The four of the remaining genes including: *AQP8*, *CDC42EP3*, *GPR176*, and *PLBD1*, were associated with hypermethylated DMPs and had no described function in immune system. One hypomethylated DMP mapped to *FLJ43663* (*LINC-PINT*) gene which is non-protein coding RNA that has been described to have neuroprotective function in the brains of patients with neurodegenerative disorders [46]. Those five genes also did not show expression changes specific to COVID-19 patients in the gene expression analyses described below.



**Fig. 3.** Comparison of methylation levels at COVID-19 specific DMPs and results of GSEA analysis based on genes annotated to those DMPs. (a) Comparison of average methylation levels at COVID-19 specific DMPs between all studied cohorts. (b-c) Median methylation levels at all CpG sites targeted by EPIC array in the promoters of PARP9/DTX3L (b) and MX1 (c) genes. The arrows indicate identified DMPs.



**Table 2**

Biological function of proteins encoded by genes associated with COVID-19 specific DMPs (source: GeneCards).

| Gene_Name                   | Full gene name   | Encoded protein function   |
|-----------------------------|--|--|
| <i>IRF7</i>                 | Interferon regulatory factor 7                                 | Key transcriptional regulator of type I interferon (IFN)-dependent immune responses and plays a critical role in the innate immune response against DNA and RNA viruses  |
| <i>MX1</i>                  | MX Dynamin Like GTPase 1                                       | Cellular antiviral response: induced by type I and type II interferons and antagonizes the replication process of several different RNA and DNA viruses  |
| <i>AIM2</i>                 | Absent In Melanoma 2   | Involved in innate immune response by recognizing cytosolic double-stranded DNA and inducing caspase-1-activating inflammasome formation in macrophages  |
| <i>IFI44L</i>               | Interferon Induced Protein 44 Like                             | Exhibits a low antiviral activity against hepatitis C virus  |
| <i>PARP9</i>                | Poly(ADP-Ribose) Polymerase Family Member 9                    | Plays a role in DNA damage repair and in immune responses including interferon-mediated antiviral defences; in macrophages, positively regulates pro-inflammatory cytokines production in response to IFNG stimulation |
| <i>IFIT3</i>                | Interferon-induced protein with tetratricopeptide repeats 3    | Acts as an inhibitor of cellular as well as viral processes, cell migration, proliferation, signalling, and viral replication  |
| <i>TRIM22</i>               | Tripartite Motif Containing 22                                 | Participates in antiviral cell innate immunity; it is interferon-induced   |
| <i>DTX3L</i>                | Deltex E3 Ubiquitin Ligase 3L                                  | Plays a role in DNA damage repair and in interferon-mediated antiviral responses; in association with PARP9, plays a role in antiviral responses   |
| <i>AQP8</i>                 | Aquaporin 8  | Facilitates the transport of water across biological membranes along an osmotic gradient   |
| <i>LRBA</i>                 | Lipopolysaccharide-responsive and beige-like anchor protein    | May be involved in coupling signal transduction and vesicle trafficking to enable polarized secretion and/or membrane deposition of immune effector molecules  |
| <i>FLJ43663 (LINC-PINT)</i> | Long Intergenic Non-Protein Coding RNA, P53 Induced Transcript | .  |
| <i>PLBD1</i>                | Phospholipase B Domain Containing 1                            | May act as an amidase or a peptidase (By similarity)   |
| <i>EPSTI1</i>               | Epithelial-stromal interaction protein 1                       | Plays a role in M1 macrophage polarization and is required for the proper regulation of gene expression during M1 versus M2 macrophage differentiation (By similarity)   |
| <i>CMPK2</i>                | Cytidine/Uridine Monophosphate Kinase 2                        | May participate in dUTP and dCTP synthesis in mitochondria   |
| <i>CDC42EP3</i>             | Cdc42 effector protein 3                                       | Probably involved in the organization of the actin cytoskeleton  |
| <i>CYSTM1</i>               | Cysteine Rich Transmembrane Module Containing 1                | Among its related pathways are Innate Immune System  |
| <i>GPR176</i>               | G protein-coupled receptor 176                                 | Orphan receptor involved in normal circadian rhythm behaviour  |

### 3.5. Most of the genes associated with COVID-19 specific DMPs interact in interferon response pathways

The GSEA analysis with MSigDB hallmark gene sets as a reference again linked those genes to interferon alpha (adjusted p value = 2.73e-11) and interferon gamma response pathways (adjusted p value = 1.57e-7) (Fig. 4a), with six of the genes including *CMPK2*, *EPSTI1*, *IFI44L*, *IFIT3*, *IRF7*, and *MX1* to take part in both processes.

Next, we used STRING platform [47] to assess potential protein-protein interaction (PPI) between proteins encoded by the identified genes. This analysis found evidence for the interactions between ten proteins from our gene set (Fig. 4b) at statistically significant level (Benjamini-Hochberg-adjusted p-value < 1.0e-16), and enrichment of those genes in six local network clusters (ontology terms used in STRING) associated with interferon alpha/beta signalling pathways (Supplementary Table 4). Moreover, four “evidence channels”, which are the source databases utilized in STRING to find the evidence for the interaction (for detailed description of those data bases see Material and methods (Section 2.6.)) also indicated interactions between those genes (Supplementary Table 5).

### 3.6. Identified COVID-19 specific methylation changes are likely a proxy of broader methylation changes

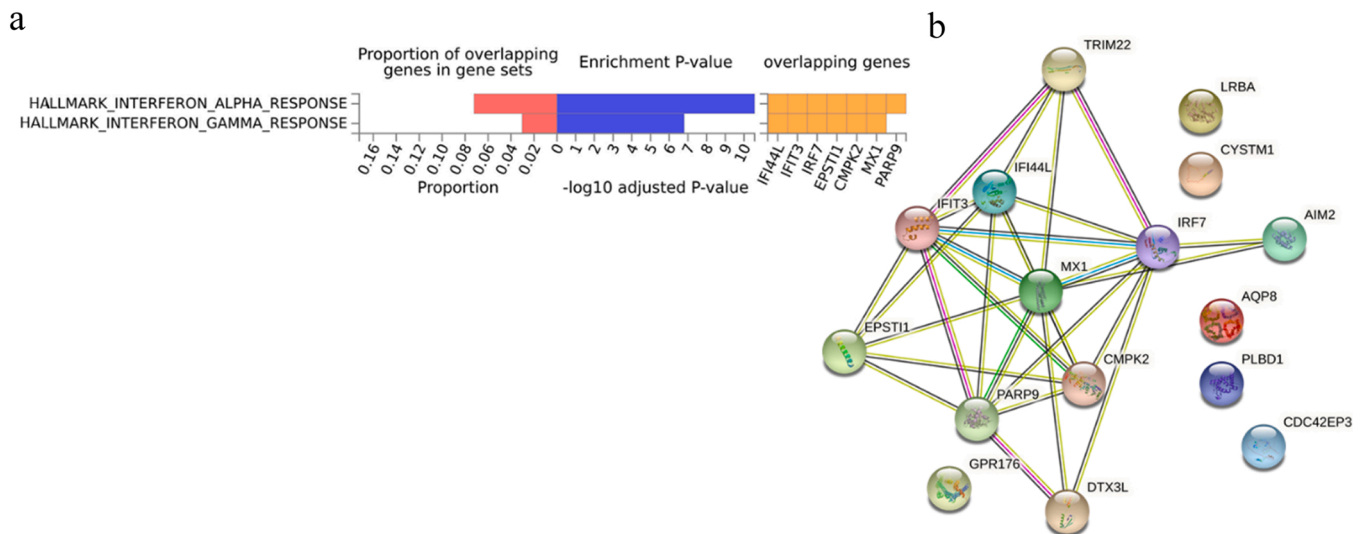
Considering high heterogeneity of studied cohorts, technical limitations of methylation screening technology and bioinformatics data analysis of this type of data, it is plausible that a number of the infection specific methylation changes did not reach minimum methylation difference threshold (five percent points) in our analysis. We attempted to overcome those limitations and assessed whether methylation levels of COVID-19 specific DMPs are correlated with methylation levels of any other CpGs screened by microarray we used. We performed this analysis for each of the patients' cohorts separately and found that in all analysed cohorts methylation changes at 30 additional probes were strongly correlated ( $|r| \geq 0.7$ ; Supplementary Table 6) with at least one of the COVID-19 specific DMPs. Interestingly, nine of the probes mapped to already identified COVID-19 specific genes, including: *MX1*, *IFI44L*, *FLJ43663*, *PARP9/DTX3L*, *AQP8*, *IRF7*, and *AIM2*. Both unsupervised

clustering (Supplementary Fig. 3a-b), as well as t-SNE-based visualisation (Supplementary Fig. 3c-d) based on this subset of 30 DMPs clearly separated COVID-19 patients from patients with other respiratory diseases in both USA 1 and for USA 2 cohorts, indicating that methylation changes within identified probes are also specific to SARS-CoV-2 infection.

Moreover, FUMA based GSEA showed again, significant enrichment of those genes in interferon alpha (adjusted p value = 1.56e-25) and interferon gamma response pathways (adjusted p value = 5.23e-19) with additional seven genes *OAS1*, *IRF9*, *BST2*, *MVB12A*, *IFIH1*, *DDX60*, and *PSMB8* identified in this analysis also associated with interferon pathways (Supplementary Fig. 3e). These results suggest that methylation changes may not reflect a full picture of interaction between SARS-CoV-2 and methylome of blood cells and studies based on more accurate clinical designs are necessary to elaborate the full impact of the virus on blood cells methylome.

### 3.7. SARS-CoV-2 infection specific methylation changes are present in blood cells of asymptomatic patients and patients with mild symptoms

Two of the cohorts, in our study included patients tested positive for SARS-CoV-2 but displaying no or mild infection symptoms and not requiring hospitalization (Spanish cohort: n = 194 and USA 2: n = 33). We compared methylation levels at the 24 COVID-19 specific DMPs between those patient groups and healthy controls using logit regression model (adjusted for age and sex, for the details of this analysis see Material and methods (Section 2.3.) and Supplementary Table 7). The methylation levels at majority of those DMPs except for four (cg10728454, cg07992500, cg18519762, cg22488164) were statistically significantly different (FDR corrected p-value  $\leq 0.05$ ) and close to or higher than 5% point in asymptomatic SARS-CoV-2 positive patients than in healthy controls. Most importantly the methylation changes at almost all of those DMPs displayed a general trend of increasing methylation difference from healthy controls to hospitalized patients with not hospitalized patients showing intermediate methylation levels between those two groups (Fig. 5).



**Fig. 4.** GSEA analysis of the genes associated with COVID-19 specific DMPs. (a) GSEA analysis based on “hallmark” collection in Molecular Signatures Database (MSigDB). (b) PPI networks analysis of genes associated with COVID-19 specific DMPs. Lines show interaction between proteins according to: textmining (green), experiments (pink), databases (blue), and co-expression (black) “evidence channels” (see Material and methods (Section 2.6.) for the definitions of the evidence channels).

### 3.8. Hypomethylation of genes associated with COVID-19 specific DMPs activates those genes expression

The STRING analysis indicated that proteins encoded by ten of the genes with COVID-19 specific DMPs interact (Fig. 4b). To assess the association between methylation and expression of the genes in this gene set, we first compared their expression levels between COVID-19 patients ( $n = 62$ ) and healthy blood samples ( $n = 24$ ) [48]. This analysis showed that nine of the genes including *MX1*, *CMPK2*, *PARP9*, *TRIM22*, *AIM2*, *IFI44L*, *IFIT3*, *IRF7*, and *EPSTI1* were significantly upregulated in blood of COVID-19 patients ( $p \leq 0.05$ , ANOVA;  $|\log_2(\text{FC})| \geq 1$ ) (Fig. 6a). The expression of *DTX3L* was also statistically significantly upregulated ( $p \leq 0.05$ , ANOVA) but the effect size was slightly lower than in case of other genes (Fig. 6a).

Then, we compared the expression of those genes between COVID-19 patients ( $n = 102$ ) and patients with other respiratory diseases ( $n = 26$ ) using RNA-seq data from: [49]. The logistic regression analysis adjusted for WBC fraction, age, sex, and steroids intake (for the details of this analysis see Material and methods (Section 2.4.)) showed that the expression was statistically significantly upregulated (FDR corrected  $p$ -value  $\leq 0.05$ ) specifically in COVID-19 patients, with  $|\log_2(\text{FC})| \geq 1$  observed for all the genes except *AIM2* for which  $|\log_2(\text{FC})|$  was 0.92 (Supplementary Table 8). Only one gene *IRF7* did not show statistically significant expression change in this analysis.

Interestingly, despite the strong association of methylation and expression, the expression of all genes in this gene set was markedly differed between COVID-19 patients (Fig. 6b), suggesting potential association of the expression levels with the disease outcome. We, therefore, compared the expression levels between 50 patients described as only hospitalized and 46 as admitted to intensive care units (ICU). The expression of all of those genes significantly differed between non-ICU and ICU patients (FDR corrected  $p$ -value  $\leq 0.05$ ), however the effect size ( $|\log_2(\text{FC})| \geq 1$ ) was significant only for four of the genes, including *IFI44L*, *MX1*, *CMPK2*, and *IFIT3* (Fig. 6c and Supplementary Table 9). Together these results indicate that there are other factors influencing the expression of the genes epigenetically activated during SARS-CoV-2 infection, what also potentially explains remarkable unpredictability and heterogeneity of COVID-19 clinical outcomes in the presence of uniform methylation changes.

## 4. Discussion

We combined methylomics and transcriptomics data from four geographically distant and ethnically heterogeneous cohorts of COVID-19 patients, with the aim to assess causal interaction of SARS-CoV-2 virus with host methylome. The study involved the largest so far cohort of COVID-19 patients from geographically distant and to some extent ethnically heterogeneous populations. Our analysis identified 1773 CpG sites that displayed remarkably stable methylation changes in blood of hospitalized 478 COVID-19 patients and those changes were independent of individual WBC proportions. The pathway enrichment analysis based on this methylation signature showed that identified methylation changes mapped to genes significantly enriched in terms related to only: interferon alpha and interferon gamma response. Nevertheless, the majority of those changes were also present in patients displaying symptoms of upper tract respiratory infection similar to SARS-CoV-2 infection, indicating that this methylation signature is a part of general response of methylome of blood cells to the infection. However, the methylation status of 24 of CpG sites from this signature, which annotated to 17 genes, was different in blood of COVID-19 patients than in healthy blood and blood of patients with other upper tract respiratory infections but displaying symptoms similar to COVID-19. This suggests that methylome of blood cells is involved in pathogenesis of COVID-19 in virus specific manner and epigenetic changes regulate a specific subset of the genes during SARS-CoV-2 infection.

Twelve out of seventeen of the genes associated with COVID-19 specific methylation changes including: *AIM2*, *CYSTM1*, *DTX3L*, *PARP9*, *CMPK2*, *EPSTI1*, *IFI44L*, *IFIT3*, *IRF7*, *LRBA*, *MX1*, and *TRIM22* have been associated with immune response and eleven (except for *LRBA*) were shown to be regulated by interferon [45]. Moreover, proteins encoded by ten of those genes, including *AIM2*, *DTX3L*, *PARP9*, *CMPK2*, *EPSTI1*, *IFI44L*, *IFIT3*, *IRF7*, *MX1*, and *TRIM22* have been shown to interact. Detailed analysis of function of those genes showed that all of them have been previously shown to be effectors in the interferon response pathways [45] and *PARP9* gene has additionally been reported as a non-canonical sensor for RNA virus to initiate and amplify type I interferon production [50]. Interestingly, *MX1* and *PARP9/DTX3L* promoter regions respectively harboured four and three CpG sites hypomethylated in COVID-19 patients. The epigenetic regulation of *MX1* has been shown in viral infections but to our best knowledge involvement the DNA methylation-mediated epigenetic regulation of *PARP9/DTX3L*



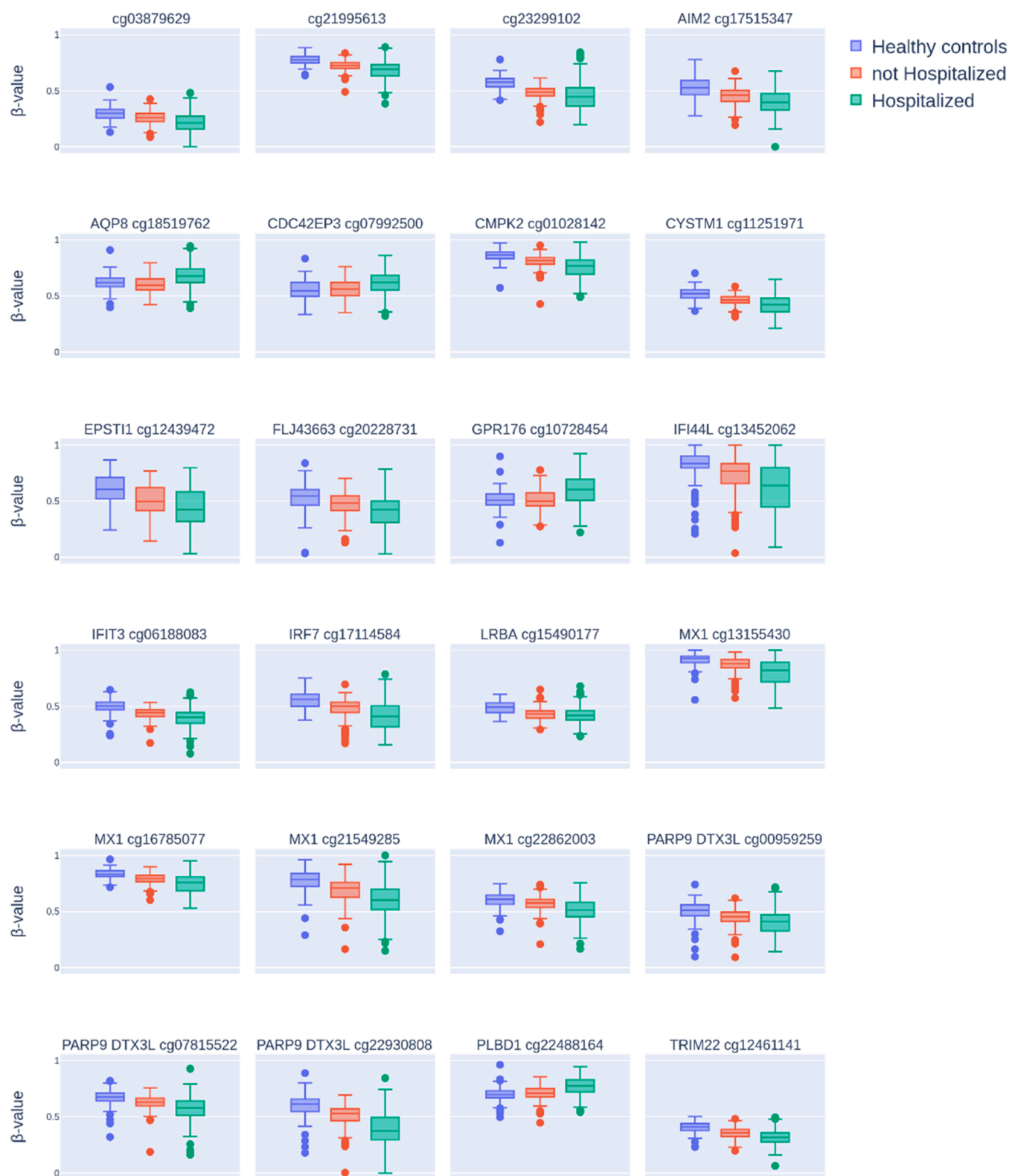
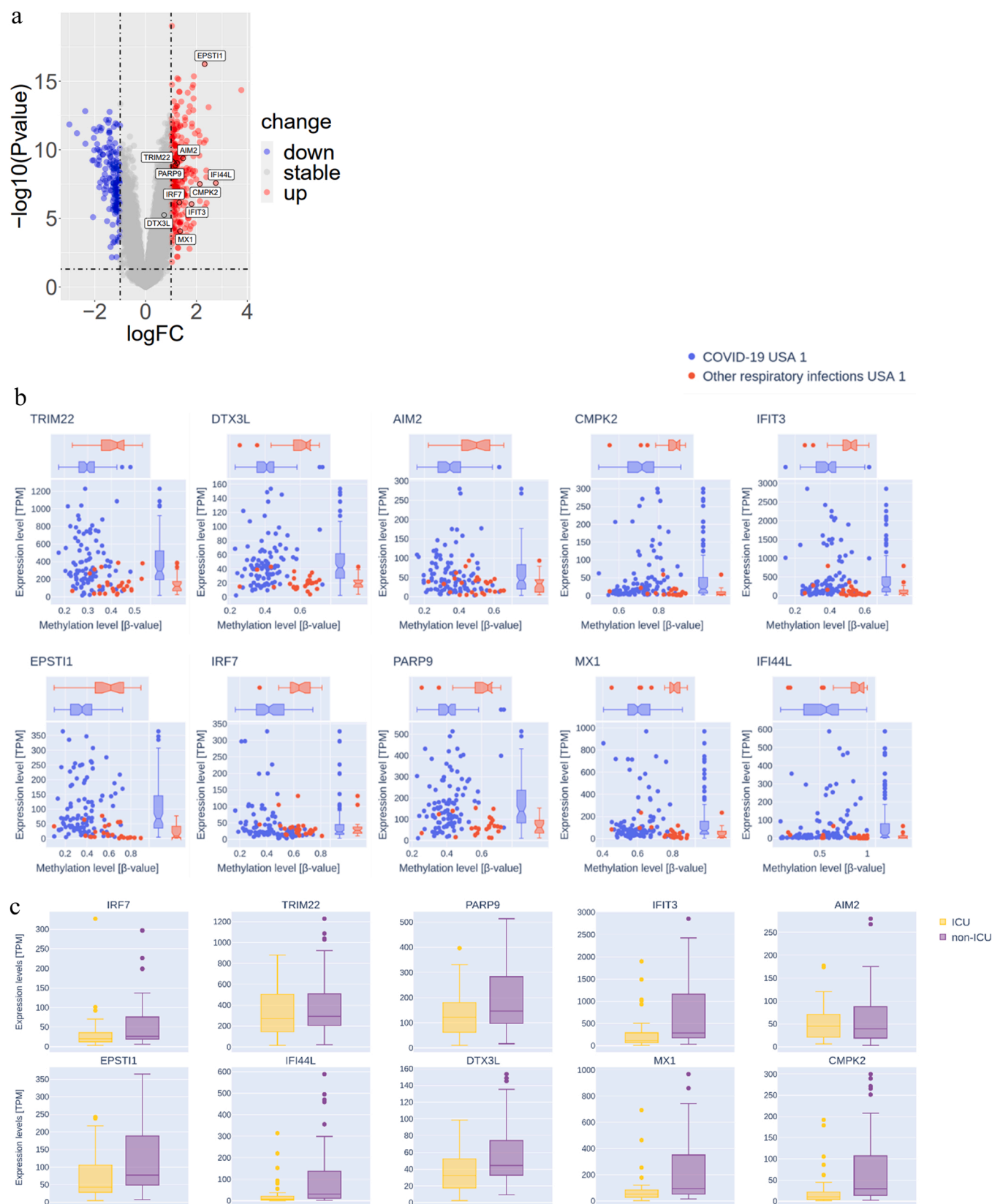


Fig. 5. Comparison of the methylation levels at COVID-19 specific DMPs between non-hospitalized COVID-19 patients, healthy controls and hospitalised COVID-19 patients.

bidirectional promoter has not been previously described, and importantly transcripts of *PARP9* (alias *BAL1*) and *DTX3L* (alias *BBAP*), as well as *MX1*, have well described function in interferon signaling pathways and pathology of SARS-CoV-2 virus infection [50–55].

Most importantly, we showed that all of the genes in this gene set

were differentially expressed between blood cells of COVID-19 patients and healthy controls, and nine of them (except for *IRF7*) were also up regulated in blood of COVID-19 patients but not in patients with other respiratory infections. This indicates that methylation is essential for the activation of those genes and it is likely that they undergo activation



**Fig. 6.** Analysis of association of methylation changes with the expression of genes mapped to COVID-19 specific DMPs. (a) Volcano plot describing the expression of the analysed genes between COVID-19 patients and healthy controls in COVID19db dataset. The vertical lines correspond to two-fold expression change and the horizontal line represents a p-value of 0.05. (b) Scatter plots describing association of methylation levels with expression of genes up regulated in COVID-19 patients. Each dot in the scatterplot represents expression and methylation level of the gene for COVID-19 (green) or other respiratory infections (red) patients. (c) Comparison of expression levels of genes upregulated in COVID-19 patients between non-ICU and ICU COVID-19 patients.

specifically upon SARS-CoV-2 infection. However, despite of the uniform in all studied patient cohort methylation of genes associated with SARS-CoV-2 infection, the expression of the genes mapped to those epigenetic changes differed significantly between patients and interestingly the patients with high expression were significantly less likely to be admitted to ICU. This suggests that the response of methylome of blood cells to the infection appears to be essential for the regulation of the expression of the genes during infection but there are other factors, which influence the response of blood transcriptome of COVID-19 patients to SARS-CoV-2 virus. The heterogeneity of the gene expression could potentially explain the unpredictability of COVID-19 clinical outcomes but also suggests that other mechanisms are needed for the transcriptome response to the infection. Nevertheless, further studies focused on the design driven by the clinical outcomes are needed to elaborate on those interactions.

## 5. Conclusions

In conclusion, we identified methylation changes in blood cells from COVID-19 patients that are specific to SARS-CoV-2 infection and are present in patients from four geographically distant regions. Furthermore, those changes appear to regulate the expression of associated genes, but our results also indicate that there are other, likely host specific, factors modulating the expression of those genes during the infection. One more important conclusion from our study is that the power of the study, as well as consistent clinical description of the patients, is critical to elaborate molecular effects of the SARS-CoV-2 infection. Only by basing our analyses on four cohorts of COVID-19 patients, cohort of healthy subjects, and data from non-COVID-19 respiratory disease, we were able to not only identify disease related methylation changes, but most importantly cross-validate findings between the populations.

## Funding

This study was funded by Polish National Agency for Academic Exchange grant number: PPN/PPO/2018/1/00088/U and European Regional Development Fund of the West Pomeranian Province 2014–2020, PROTO\_LAB/K1/2020/U/18. The funding agencies did not influence on study design, execution and publication of results.

## Data sharing statement

Complete DNA methylation raw data of the 32 COVID-19 patients from Polish cohort are deposited at GEO repository under accession number GSE202296.

## CRedit author statement

**Jan Bińkowski:** Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Olga Taryma-Leśniak:** Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Karolina Łuczowska:** Methodology, Resources, Data curation. **Anna Niedzwiedź:** Methodology, Resources, Data curation. **Kacper Lechowicz:** Methodology, Resources, Data curation. **Dominik Strapagiel:** Resources, Data curation, Writing - review & editing. **Justyna Jarczak:** Methodology, Resources, Data curation. **Veronica Davalos:** Conceptualization, Writing - review & editing. **Aurora Pujol:** Conceptualization, Writing - review & editing. **Manel Esteller:** Conceptualization, Writing - review & editing. **Katarzyna Kotfis:** Methodology, Resources, Data curation, Writing - review & editing. **Bogusław Machaliński:** Methodology, Resources, Data Curation. **Miłosz Parczewski:** Conceptualization, Methodology, Resources, Data curation, Writing - review & editing. **Tomasz K. Wojdacz:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing,

Supervision, Funding acquisition. All authors contributed to interpretation of findings and approved the final version of the manuscript.

## Conflict of interest statement

Manel Esteller is an advisor of Quimatrix and Ferrer International. The remaining authors declare no conflict of interest.

## Acknowledgments

We thank the team members who contributed to these studies.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopha.2022.113396.

## References

- [1] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet* 395 (10223) (2020) 507–513.
- [2] C. Wu, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, S. Wang, Y. Chao, Z. Yang, J. Xu, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, Y. Song, Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 Pneumonia in Wuhan, China, *JAMA Intern Med* 180 (7) (2020) 934–943.
- [3] P. Mehta, D.F. McAuley, M. Brown, E. Sanchez, R.S. Tattersall, J.J. Manson, U. K. Hlh, Across llaboration, COVID-19: consider cytokine storm syndromes and immunosuppression, *Lancet* 395 (10229) (2020) 1033–1034.
- [4] G. Chen, D. Wu, W. Guo, Y. Cao, D. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, X. Zhang, M. Zhang, S. Wu, J. Song, T. Chen, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, Clinical and immunological features of severe and moderate coronavirus disease 2019, *J. Clin. Invest* 130 (5) (2020) 2620–2629.
- [5] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (10223) (2020) 497–506.
- [6] C. Guo, B. Li, H. Ma, X. Wang, P. Cai, Q. Yu, L. Zhu, L. Jin, C. Jiang, J. Fang, Q. Liu, D. Zong, W. Zhang, Y. Lu, K. Li, X. Gao, B. Fu, L. Liu, X. Ma, J. Weng, H. Wei, T. Jin, J. Lin, K. Qu, Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm, *Nat. Commun.* 11 (1) (2020) 3924.
- [7] Y. Zhou, B. Fu, X. Zheng, D. Wang, C. Zhao, Y. Qi, R. Sun, Z. Tian, X. Xu, H. Wei, Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients, *Natl. Sci. Rev.* 7 (6) (2020) 998–1002.
- [8] C. Qin, L. Zhou, Z. Hu, S. Zhang, S. Yang, Y. Tao, C. Xie, K. Ma, K. Shang, W. Wang, D.S. Tian, Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China, *Clin. Infect. Dis.* 71 (15) (2020) 762–768.
- [9] B. Diao, C. Wang, Y. Tan, X. Chen, Y. Liu, L. Ning, L. Chen, M. Li, Y. Liu, G. Wang, Z. Yuan, Z. Feng, Y. Zhang, Y. Wu, Y. Chen, Reduction and functional exhaustion of T cells in patients with corONAVIRUS DISEASE 2019 (COVID-19), *Front Immunol.* 11 (2020) 827.
- [10] H.Y. Zheng, M. Zhang, C.X. Yang, N. Zhang, X.C. Wang, X.P. Yang, X.Q. Dong, Y. T. Zheng, Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients, *Cell Mol. Immunol.* 17 (5) (2020) 541–543.
- [11] X. Li, S. Xu, M. Yu, K. Wang, Y. Tao, Y. Zhou, J. Shi, M. Zhou, B. Wu, Z. Yang, C. Zhang, J. Yue, Z. Zhang, H. Ren, X. Liu, J. Xie, M. Xie, J. Zhao, Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan, *J. Allergy Clin. Immunol.* 146 (1) (2020) 110–118.
- [12] J.K. Baillie, Translational genomics. Targeting the host immune response to fight infection, *Science* 344 (6186) (2014) 807–808.
- [13] J.P. Bernardes, N. Mishra, F. Tran, T. Bahmer, L. Best, J.I. Blase, D. Bordoni, J. Franzenburg, U. Geisen, J. Josephs-Spaulding, P. Kohler, A. Kunstner, E. Rosati, A.C. Aschenbrenner, P. Bacher, N. Baran, T. Boysen, B. Brandt, N. Bruse, J. Dorr, A. Dräger, G. Elke, D. Ellinghaus, J. Fischer, M. Forster, A. Franke, S. Franzenburg, N. Frey, A. Friedrichs, J. Fuss, A. Gluck, J. Hamm, F. Hinrichsen, M.P. Hoepfner, S. Imm, R. Junker, S. Kaiser, Y.H. Kan, R. Knoll, C. Lange, G. Laue, C. Lier, M. Lindner, G. Marinos, R. Markewitz, J. Nattermann, R. Noth, P. Pickers, K. F. Rabe, A. Renz, C. Rocken, J. Rupp, A. Schaffarczyk, A. Scheffold, J. Schulte-Schrepping, D. Schunk, D. Skowasch, T. Ulas, K.P. Wandinger, M. Wittig, J. Zimmermann, H. Busch, B.F. Hoyer, C. Kaleta, J. Heyckendorf, M. Kox, J. Rybníček, S. Schreiber, J.L. Schultze, P. Rosenstiel, H.C.A.L.B. Network, C.-O.I. Deutsche, Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19, *Immunity* 53 (6) (2020), 1296–1314 e9.

- [14] D.A. Dorward, C.D. Russell, I.H. Um, M. Elshani, S.D. Armstrong, R. Penrice-Randal, T. Millar, C.E.B. Lerpiniere, G. Tagliavini, C.S. Hartley, N.P. Randle, N. N. Gachanja, P.M.D. Potey, X. Dong, A.M. Anderson, V.L. Campbell, A.J. Duguid, W. Al Qsous, R. BouHaidar, J.K. Baillie, K. Dhaliwal, W.A. Wallace, C.O.C. Bellamy, S. Prost, C. Smith, J.A. Hiscox, D.J. Harrison, C.D. Lucas, Tissue-specific immunopathology in fatal COVID-19, *Am. J. Respir. Crit. Care Med* 203 (2) (2021) 192–201.
- [15] R.C. Group, P. Horby, W.S. Lim, J.R. Emberson, M. Mafham, J.L. Bell, L. Linsell, N. Staplin, C. Brightling, A. Ustianowski, E. Elmahi, B. Prudon, C. Green, T. Felton, D. Chadwick, K. Rege, C. Fegan, L.C. Chappell, S.N. Faust, T. Jaki, K. Jeffery, A. Montgomery, K. Rowan, E. Juszczak, J.K. Baillie, R. Haynes, M.J. Landray, Dexamethasone in hospitalized patients with Covid-19, *N. Engl. J. Med* 384 (8) (2021) 693–704.
- [16] S. Kasela, Z. Daniloski, S. Bollepalli, T.X. Jordan, B.R. tenOever, N.E. Sanjana, T. Lappalainen, Integrative approach identifies SLC6A20 and CXCR6 as putative causal genes for the COVID-19 GWAS signal in the 3p21.31 locus, *Genome Biol.* 22 (1) (2021) 242.
- [17] E. Pairo-Castineira, S. Cline, L. Klaric, A.D. Bretherick, K. Rawlik, D. Pasko, S. Walker, N. Parkinson, M.H. Fourman, C.D. Russell, J. Furniss, A. Richmond, E. Gountouna, N. Wrobel, D. Harrison, B. Wang, Y. Wu, A. Meynert, F. Griffiths, W. Oosthuizen, A. Kousathanas, L. Moutsianas, Z. Yang, R. Zhai, C. Zheng, G. Grimes, R. Beale, J. Millar, B. Shih, S. Keating, M. Zechner, C. Haley, D. J. Porteous, C. Hayward, J. Yang, J. Knight, C. Summers, M. Shankar-Hari, P. Klennerman, L. Turtle, A. Ho, S.C. Moore, C. Hinds, P. Horby, A. Nichol, D. Maslove, L. Ling, D. McAuley, H. Montgomery, T. Walsh, A.C. Pereira, A. Renieri, O.I. Gen, I.C. Investigators, C.-H.G. Initiative, I. and Me, B. Investigators, C.I. Gen, X. Shen, C.P. Ponting, A. Fawkes, A. Tenesa, M. Caulfield, R. Scott, K. Rowan, L. Murphy, P.J.M. Openshaw, M.G. Semple, A. Law, V. Vitart, J. F. Wilson, J.K. Baillie, Genetic mechanisms of critical illness in COVID-19, *Nature* 591 (7848) (2021) 92–98.
- [18] G.G. Severe Covid, D. Ellinghaus, F. Degenhardt, L. Bujanda, M. Buti, A. Alballos, P. Invernizzi, J. Fernandez, D. Prati, G. Baselli, R. Asselta, M.M. Grimsrud, C. Milani, F. Aziz, J. Kassens, S. May, M. Wendorff, L. Wienbrandt, F. Uellendahl-Werth, T. Zheng, X. Yi, R. de Pablo, A.G. Chercoles, A. Palom, A.E. Garcia-Fernandez, F. Rodriguez-Frias, A. Zanella, A. Bandera, A. Protti, A. Aghemo, A. Lleo, A. Biondi, A. Caballero-Garralda, A. Gori, A. Tanck, A. Carreras Nolla, A. Latiano, A.L. Fracanzani, A. Peschuck, A. Julia, A. Pesenti, A. Voza, D. Jimenez, B. Mateos, B. Nafria Jimenez, C. Quereda, C. Paccapelo, C. Gassner, C. Angelini, C. Cea, A. Solier, D. Pestana, E. Muniz-Diaz, E. Sandoval, E.M. Paraboschi, E. Navas, F. Garcia Sanchez, F. Ceriotti, F. Martinelli-Boneschi, F. Peyvand, F. Blasi, L. Tellez, A. Blanco-Grau, G. Hemmrich-Stanisak, G. Grasselli, G. Costantino, G. Cardamone, G. Foti, S. Aneli, H. Kurihara, H. Elabd, I. My, I. Galvan-Femenia, J. Martin, J. Erdmann, J. Ferrusquia-Acosta, G. Garcia-Etxebarria, L. Izquierdo-Sanchez, L.R. Bettini, L. Sumoy, L. Terranova, L. Moreira, L. Santoro, L. Scudeller, F. Mesonero, L. Roade, M.C. Ruhlemann, M. Schaefer, M. Carrabba, M. Riveiro-Barciela, M.E. Figuera Basso, M.G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angio, M. Baldini, M. Cazzaniga, M. Schulzky, M. Cecconi, M. Wittig, M. Ciccarelli, M. Rodriguez-Gandia, M. Boccione, M. Miozzo, N. Montano, N. Braun, N. Sacchi, N. Martinez, O. Ozer, O. Palmieri, P. Faverio, P. Preatoni, P. Bonfanti, P. Omodei, P. Tentorio, P. Castro, P.M. Rodrigues, A. Blandino Ortiz, R. de Cid, R. Ferrer, R. Gualtierotti, R. Nieto, S. Goerg, S. Badalamenti, S. Marsal, G. Matullo, S. Pelusi, S. Juzenas, S. Aliberti, V. Monzani, V. Moreno, T. Wesse, T.L. Lenz, T. Pumarola, V. Rimoldi, S. Bosari, W. Albrecht, W. Peter, M. Romero-Gomez, M. D'Amato, S. Duga, J.M. Banales, J. R. Hov, T. Folseraas, L. Valenti, A. Franke, T.H. Karlsen, Genomewide association study of severe covid-19 with respiratory failure, *N. Engl. J. Med.* 383 (16) (2020) 1522–1534.
- [19] H. Zeberg, S. Paabo, The major genetic risk factor for severe COVID-19 is inherited from Neanderthals, *Nature* 587 (7835) (2020) 610–612.
- [20] S.T. Lee, Y. Xiao, M.O. Muench, J. Xiao, M.E. Fomin, J.K. Wiencke, S. Zheng, X. Dou, A. de Smith, A. Chokkalingam, P. Buffler, X. Ma, J.L. Wiemels, A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network, *Nucleic Acids Res.* 40 (22) (2012), 11339–51.
- [21] H. Heyn, S. Moran, I. Hernandez-Herrera, S. Sayols, A. Gomez, J. Sandoval, D. Monk, K. Hata, T. Marques-Bonet, L. Wang, M. Esteller, DNA methylation contributes to natural human variation, *Genome Res* 23 (9) (2013) 1363–1372.
- [22] E.M. Martin, R.C. Fry, Environmental influences on the epigenome: exposure-associated DNA methylation in human populations, *Annu. Rev. Public Health* 39 (2018) 309–333.
- [23] K. Prajzdzanc, P. Domagala, J. Hybiak, J. Rys, T. Huzarski, M. Zwicz, J. Tomczek-Zwicz, W. Redelbach, A. Sejda, J. Gronwald, T. Kluz, R. Wisniewski, C. Cybulski, A. Lukomska, K. Bialkowska, G. Sukiennicki, K. Kulczycka, S.A. Narod, T.K. Wojdacz, J. Lubinski, A. Jakubowska, BRCA1 promoter methylation in peripheral blood is associated with the risk of triple-negative breast cancer, *Int J. Cancer* 146 (5) (2020) 1293–1298.
- [24] J. Balnis, A. Madrid, K.J. Hogan, L.A. Drake, H.C. Chieng, A. Tiwari, C.E. Vincent, A. Chopra, P.A. Vincent, M.D. Robek, H.A. Singer, R.S. Alisch, A. Jaitovich, Blood DNA methylation and COVID-19 outcomes, *Clin. Epigenetics* 13 (1) (2021) 118.
- [25] M. Castro de Moura, V. Davalos, L. Planas-Serra, D. Alvarez-Erillo, C. Arribas, M. Ruiz, S. Aguilera-Albesa, J. Troya, J. Valencia-Ramos, V. Velaz-Santamaria, A. Rodriguez-Palmero, J. Villar-Garcia, J.P. Horcajada, S. Albu, C. Casasnovas, A. Rull, L. Reverte, B. Dietl, D. Dalmau, M.J. Arranz, L. Lucía-Carol, A.M. Planas, J. Perez-Tur, I. Fernandez-Cadenas, P. Villares, J. Tenorio, R. Colobran, A. Martin-Nalda, P. Soler-Palacin, F. Vidal, A. Pujol, M. Esteller, Epigenome-wide association study of COVID-19 severity with respiratory failure, *EBioMedicine* 66 (2021), 103339.
- [26] I.R. Konigsberg, B. Barnes, M. Campbell, E. Davidson, Y. Zhen, O. Pallisard, M. P. Boorgula, C. Cox, D. Nandy, S. Seal, K. Crooks, E. Sticca, G.F. Harrison, A. Hopkinson, A. Vest, C.G. Arnold, M.G. Kahn, D.P. Kao, B.R. Peterson, S.J. Wicks, D. Ghosh, S. Horvath, W. Zhou, R.A. Mathias, P.J. Norman, R. Porecha, I.V. Yang, C.R. Gignoux, A.A. Monte, A. Taye, K.C. Barnes, Host methylation predicts SARS-CoV-2 infection and clinical outcome, *Commun. Med (Lond.)* 1 (1) (2021) 42.
- [27] T.J. Morris, L.M. Butcher, A. Feber, A.E. Teschendorff, A.R. Chakravarty, T. K. Wojdacz, S. Beck, ChAMP: 450k chip analysis methylation pipeline, *Bioinformatics* 30 (3) (2014) 428–430.
- [28] A.E. Teschendorff, C.E. Breeze, S.C. Zheng, S. Beck, A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies, *BMC Bioinforma.* 18 (1) (2017) 105.
- [29] E.A. Houseman, W.P. Accomando, D.C. Koestler, B.C. Christensen, C.J. Marsit, H. Nelson, J.K. Wiencke, K.T. Kelsey, DNA methylation arrays as surrogate measures of cell mixture distribution, *BMC Bioinforma.* 13 (2012) 86.
- [30] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, A.A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (5) (2015) 453–457.
- [31] K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA, *Nat. Commun.* 8 (1) (2017) 1826.
- [32] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. USA* 102 (43) (2005), 15545–50.
- [33] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdottir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0, *Bioinformatics* 27 (12) (2011) 1739–1740.
- [34] A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J.P. Mesirov, P. Tamayo, The molecular signatures database (MSigDB) hallmark gene set collection, *Cell Syst.* 1 (6) (2015) 417–425.
- [35] J.S. Amberger, C.A. Bocchini, F. Schiettecatte, A.F. Scott, A. Hamosh, OMIM.org: online mendelian inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders, *Nucleic Acids Res* 43(Database Issue) (2015) D789–D798.
- [36] J.M. Cherry, E.L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E.T. Chan, K. R. Christie, M.C. Costanzo, S.S. Dwight, S.R. Engel, D.G. Fisk, J.E. Hirschman, B. C. Hitz, K. Karra, C.J. Krieger, S.R. Miyasato, R.S. Nash, J. Park, M.S. Skrzypek, M. Simson, S. Weng, E.D. Wong, Saccharomyces Genome Database: the genomics resource of budding yeast, *Nucleic Acids Res* 40(Database Issue) (2012) D700–D705.
- [37] S. Orchard, S. Kerrien, S. Abbani, B. Aranda, J. Bhate, S. Bidwell, A. Bridge, L. Briganti, F.S. Brinkman, G. Cesareni, A. Chatr-aryamontri, E. Chautard, C. Chen, M. Dumousseau, J. Goll, R.E. Hancock, L.I. Hannick, I. Jurisica, J. Khadake, D. J. Lynn, U. Mahadevan, L. Perfetto, A. Raghunath, S. Ricard-Blum, B. Roehert, L. Salwinski, V. Stumpflen, M. Tyers, P. Uetz, I. Xenarios, H. Hermjakob, Protein interaction data curation: the International Molecular Exchange (IMEx) consortium, *Nat. Methods* 9 (4) (2012), 345–50.
- [38] C. Stark, B.J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, *Nucleic Acids Res* 34(Database Issue) (2006) D535–D539.
- [39] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids Res* 43(Database Issue) (2015) D447–D452.
- [40] A. Anurag, P.K. Jha, A. Kumar, Differential white blood cell count in the COVID-19: A cross-sectional study of 148 patients, *Diabetes Metab. Syndr.* 14 (6) (2020) 2099–2102.
- [41] A.E. Jaffe, R.A. Irizarry, Accounting for cellular heterogeneity is critical in epigenome-wide association studies, *Genome Biol.* 15 (2) (2014) R31.
- [42] Y. Liu, M.J. Aryee, L. Padyukov, M.D. Fallin, E. Hesselberg, A. Runarsson, L. Reinius, N. Acevedo, M. Taub, M. Ronninger, K. Shchetynsky, A. Scheynius, J. Kere, L. Alfredsson, L. Klareskog, T.J. Ekstrom, A.P. Feinberg, Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis, *Nat. Biotechnol.* 31 (2) (2013) 142–147.
- [43] L. Qi, A.E. Teschendorff, Cell-type heterogeneity: Why we should adjust for it in epigenome and biomarker studies, *Clin. Epigenetics* 14 (1) (2022) 31.
- [44] J. Parkin, B. Cohen, An overview of the immune system, *Lancet* 357 (9270) (2001), 1777–89.
- [45] I. Rusinova, S. Forster, S. Yu, A. Kannan, M. Masse, H. Cumming, R. Chapman, P. J. Hertzog, Interferome v2.0: an updated database of annotated interferon-regulated genes, *Nucleic Acids Res* 41(Database Issue) D1040–6 (2013).
- [46] A. Simchovitz, M. Hanan, N. Yayon, S. Lee, E.R. Bennett, D.S. Greenberg, S. Kadener, H. Soreg, A lncRNA survey finds increases in neuroprotective LINC-PINT in Parkinson's disease substantia nigra, *Aging Cell* 19 (3) (2020), e13115.
- [47] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic Acids Res* 45 (D1) (2017) D362–D368.
- [48] W. Zhang, Y. Zhang, Z. Min, J. Mo, Z. Ju, W. Guan, B. Zeng, Y. Liu, J. Chen, Q. Zhang, H. Li, C. Zeng, Y. Wei, G.C. Chan, COVID19db: a comprehensive database platform to discover potential drugs and targets of COVID-19 at whole transcriptomic scale, *Nucleic Acids Res* 50 (D1) (2022) D747–D757.
- [49] K.A. Overmyer, E. Shishkova, L.J. Miller, J. Balnis, M.N. Bernstein, T.M. Peters-Clarke, J.G. Meyer, Q. Quan, L.K. Muehlbauer, E.A. Trujillo, Y. He, A. Chopra, H.

- C. Chieng, A. Tiwari, M.A. Judson, B. Paulson, D.R. Brademan, Y. Zhu, L. R. Serrano, V. Linke, L.A. Drake, A.P. Adam, B.S. Schwartz, H.A. Singer, S. Swanson, D.F. Mosher, R. Stewart, J.J. Coon, A. Jaitovich, Large-Scale Multi-omic Analysis of COVID-19 Severity, *Cell Syst.* 12 (1) (2021), 23–40 e7.
- [50] J. Xing, A. Zhang, Y. Du, M. Fang, L.J. Minze, Y.J. Liu, X.C. Li, Z. Zhang, Identification of poly(ADP-ribose) polymerase 9 (PARP9) as a noncanonical sensor for RNA virus in dendritic cells, *Nat. Commun.* 12 (1) (2021) 2681.
- [51] P. Juszczynski, J.L. Kutok, C. Li, J. Mitra, R.C. Aguiar, M.A. Shipp, BAL1 and BBAP are regulated by a gamma interferon-responsive bidirectional promoter and are overexpressed in diffuse large B-cell lymphomas with a prominent inflammatory infiltrate, *Mol. Cell Biol.* 26 (14) (2006) 5348–5359.
- [52] Y. Zhang, D. Mao, W.T. Roswit, X. Jin, A.C. Patel, D.A. Patel, E. Agapov, Z. Wang, R.M. Tidwell, J.J. Atkinson, G. Huang, R. McCarthy, J. Yu, N.E. Yun, S. Paessler, T. G. Lawson, N.S. Omattage, T.J. Brett, M.J. Holtzman, PARP9-DTX3L ubiquitin ligase targets host histone H2BJ and viral 3C protease to enhance interferon signaling and control viral infection, *Nat. Immunol.* 16 (12) (2015) 1215–1227.
- [53] L.C. Russo, R. Tomasin, I.A. Matos, A.C. Manucci, S.T. Sowa, K. Dale, K. W. Caldecott, L. Lehtio, D. Schechtman, F.C. Meotti, A. Bruni-Cardoso, N.C. Hoch, The SARS-CoV-2 Nsp3 macrodomain reverses PARP9/DTX3L-dependent ADP-ribosylation induced by interferon signaling, *J. Biol. Chem.* 297 (3) (2021), 101041.
- [54] D. Blanco-Melo, B.E. Nilsson-Payant, W.C. Liu, S. Uhl, D. Hoagland, R. Moller, T. X. Jordan, K. Oishi, M. Panis, D. Sachs, T.T. Wang, R.E. Schwartz, J.K. Lim, R. A. Albrecht, B.R. tenOever, imbalanced host response to SARS-CoV-2 drives development of COVID-19, *Cell* 181 (5) (2020), 1036–1045 e9.
- [55] J. Bizzotto, P. Sanchis, M. Abbate, S. Lage-Vickers, R. Lavignolle, A. Toro, S. Olszevicki, A. Sabater, F. Cascardo, E. Vazquez, J. Cotignola, G. Gueron, SARS-CoV-2 infection boosts MX1 antiviral effector in COVID-19 patients, *iScience* 23 (10) (2020), 101585.
- [56] L.A. Salas, D.C. Koestler, R.A. Butler, H.M. Hansen, J.K. Wiencke, K.T. Kelsey, B. C. Christensen, An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray, *Genome Biol.* 19 (1) (2018) 64.
- [57] I. Zaimi, D. Pei, D.C. Koestler, C.J. Marsit, I. De Vivo, S.S. Tworoger, A.E. Shields, K.T. Kelsey, D.S. Michaud, Variation in DNA methylation of human blood over a 1-year period using the Illumina MethylationEPIC array, *Epigenetics* 13(10–11) (2018) 1056–1071.
- [58] M.V. Cubellis, L. Pignata, A. Verma, A. Sparago, R. Del Prete, M. Monticelli, L. Calzari, V. Antona, D. Melis, R. Tenconi, S. Russo, F. Cerrato, A. Riccio, Loss-of-function maternal-effect mutations of PADI6 are associated with familial and sporadic Beckwith-Wiedemann syndrome with multi-locus imprinting disturbance, *Clin. Epigenetics* 12 (1) (2020) 139.