EBioMedicine 66 (2021) 103339

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Epigenome-wide association study of COVID-19 severity with respiratory failure

Manuel Castro de Moura[a,1], Veronica Davalos[a,1], Laura Planas-Serra[b], Damiana Alvarez-Errico[a], Carles Arribas[a], Montserrat Ruiz[b], Sergio Aguilera-Albesa[e], Jesús Troya[h], Juan Valencia-Ramos[i], Valentina Vélez-Santamaria[b,c], Agustí Rodríguez-Palmero[b,d], Judit Villar-Garcia[g], Juan P. Horcajada[g], Sergiu Albu[f], Carlos Casasnovas[b,c], Anna Rull[j], Laia Reverte[j], Beatriz Dietl[k], David Dalmau[l], Maria J. Arranz[m], Laia Llucià-Carol[n], Anna M. Planas[o], Jordi Pérez-Tur[p], Israel Fernandez-Cadenas[n], Paula Villares[q], Jair Tenorio[r,s], Roger Colobran[t], Andrea Martin-Nalda[u], Pere Soler-Palacin[u], Francesc Vidal[j], Aurora Pujol[b,s,w,*], Manel Esteller[a,v,w,x,*]

[a] *Josep Carreras Leukaemia Research Institute (IJC), 08916 Badalona, Barcelona, Catalonia, Spain*
[b] *Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain*
[c] *Bellvitge University Hospital, L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain*
[d] *University Hospital Germans Trias i Pujol, Badalona, Barcelona, Catalonia, Spain*
[e] *Navarra Health Service Hospital, Pamplona, Spain*
[f] *Institut Guttmann Foundation, Badalona, Barcelona, Catalonia, Spain*
[g] *Hospital del Mar - IMIM Biomedical Research Institute, Barcelona, Catalonia, Spain*
[h] *Infanta Leonor University Hospital, Madrid, Spain*
[i] *University Hospital of Burgos, Burgos, Spain*
[j] *Hospital Universitari de Tarragona Joan XXIII, IISPV, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain*
[k] *Servei de malalties infeccioses Hospital Universitari MutuaTerrassa, Universitat de Barcelona, Barcelona, Catalonia, Spain*
[l] *MutuaTerrassa Research and Innovation Foundation, HIV/AIDS Unit Hospital Universitari MutuaTerrassa, University of Barcelona, Barcelona, Catalonia, Spain*
[m] *Fundaciò Docència i Recerca Mutua Terrassa i Hospital Universitari Mutua Terrassa, Barcelona, Catalonia, Spain*
[n] *Stroke Pharmacogenomics and Genetics Group, Sant Pau Institute of Research, Sant Pau Hospital, Barcelona, Catalonia, Spain*
[o] *Department of Brain Ischemia and Neurodegeneration, Institut d'Investigacions Biomèdiques de Barcelona (IIBB), Consejo Superior de Investigaciones Científicas (CSIC), Area of Neurosciences, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Catalonia, Spain*
[p] *Institut de Biomedicina de València-CSIC, CIBERNED, Unitat Mixta de Neurologia i Genètica, IIS La Fe, Vallencia, Spain*
[q] *Internal Medicine Department, Hospital HM Sanchinarro, HM Hospitales, Madrid, Spain*
[r] *INGEMM-Instituto de Genética Médica y Molecular, Hospital Universitario La Paz, Madrid, Spain*
[s] *Center for Biomedical Research on Rare Diseases (CIBERER), ISCIII, Madrid, Spain*
[t] *Immunology Division, Genetics Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Vall d'Hebron Barcelona Hospital Campus, UAB, Barcelona, Catalonia, Spain*
[u] *Pediatric Infectious Diseases and Immunodeficiencies Unit, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Catalonia, Spain*
[v] *Centro de Investigación Biomédica en Red de Cancer (CIBERONC), Spain*
[w] *Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain*
[x] *Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Catalonia, Spain*

ARTICLE INFO

Article History:
Received 5 January 2021
Revised 16 March 2021
Accepted 26 March 2021
Available online 15 April 2021

ABSTRACT

Background: Patients infected with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for the coronavirus disease 2019 (COVID-19), exhibit a wide spectrum of disease behaviour. Since DNA methylation has been implicated in the regulation of viral infections and the immune system, we performed an epigenome-wide association study (EWAS) to identify candidate loci regulated by this epigenetic mark that could be involved in the onset of COVID-19 in patients without comorbidities.

* Corresponding authors.
E-mail addresses: apujol@idibell.cat (A. Pujol), mesteller@carrerasresearch.org (M. Esteller).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ebiom.2021.103339

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

Methods: Peripheral blood samples were obtained from 407 confirmed COVID-19 patients ≤ 61 years of age and without comorbidities, 194 (47.7%) of whom had mild symptomatology that did not involve hospitalization and 213 (52.3%) had a severe clinical course that required respiratory support. The set of cases was divided into discovery (n = 207) and validation (n = 200) cohorts, balanced for age and sex of individuals. We analysed the DNA methylation status of 850,000 CpG sites in these patients.
Findings: The DNA methylation status of 44 CpG sites was associated with the clinical severity of COVID-19. Of these loci, 23 (52.3%) were located in 20 annotated coding genes. These genes, such as the inflammasome component Absent in Melanoma 2 (AIM2) and the Major Histocompatibility Complex, class I C (HLA-C) candidates, were mainly involved in the response of interferon to viral infection. We used the EWAS-identified sites to establish a DNA methylation signature (EPICOVID) that is associated with the severity of the disease.
Interpretation: We identified DNA methylation sites as epigenetic susceptibility loci for respiratory failure in COVID-19 patients. These candidate biomarkers, combined with other clinical, cellular and genetic factors, could be useful in the clinical stratification and management of patients infected with the SARS-CoV-2.
Funding: The Unstoppable campaign of the Josep Carreras Leukaemia Foundation, the Cellex Foundation and the CERCA Programme/Generalitat de Catalunya.

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus was identified in Wuhan, China, in late 2019 [1], and has provoked an ongoing global pandemic of the resulting illness, COVID-19. At the time of the writing (March 15th, 2021) there have been more than 120 million confirmed cases worldwide and 2.6 million deaths (https://origin-coronavirus.jhu.edu/map.html. COVID-19 has a wide spectrum of clinical manifestations, with the majority of infected subjects showing only mild symptoms or being asymptomatic [2,3]. The principal group of patients with high mortality rates comprises those with severe respiratory failure associated with acute respiratory distress syndrome (ARDS) and interstitial pneumonia. These high-risk patients require early and prolonged support by mechanical ventilation to compensate for their respiratory failure [4,5]. The reasons for the varied clinical repertoire of COVID-19 are largely unknown. Only three risk factors have been consistently related to life-threatening COVID-19-associated respiratory failure: male sex, old age and concomitant medical conditions, such as diabetes, obesity, hypertension and cardiovascular pathology [6,7]. However, even considering all these components, there is great inter-individual variability in each demographic and epidemiological group. Thus, given the immense impact of COVID-19 on morbidity and mortality, there is an unmet medical need to discover endogenous cellular and molecular biomarkers that predict the expected clinical course of the disease. In this regard, striking recent data from genome-wide association studies (GWAS) in large populations of COVID-19 patients and genome sequencing in particular subsets with life-threatening pneumonia indicate that loci at the chromosome 3p21.31 region [8,9], antiviral restriction enzyme activators clustered at the chromosome 12q24.13 locus [9] and the ABO blood-group system are linked to COVID-19 susceptibility [8], and inborn genetic defects of type I interferon (IFN) immunity are associated with extreme severe pneumonia [10,11].

Here we have explored another layer of biological information that has not been comprehensively addressed so far in the COVID-19 field: the impact of epigenetic variation on disease severity. Epigenetics, defined as the study of changes in gene function that are heritable and that do not entail a change in DNA sequence, plays a major role in tissue homoeostasis. In particular, DNA methylation, the most widely studied epigenetic mark, is altered in many human diseases [12,13]. This knowledge has been applied clinically in the field of oncology, which uses DNA methylation biomarkers for the clinical management of gliomas [14,15] and cancers of unknown primary origin [16], and for the pathological classification of brain tumors [17]. In the viral arena, it is known that the activity of DNA and RNA viruses, such as HPV, HBV, EBV, KS and HIV, is regulated by DNA methylation changes [18-21]. Most importantly, in the case of COVID-19, the recognition that the activity of the adaptive immune system, including that of B- and T-cells, is significant for the provision of pre-existing immunity to SARS-CoV-2 [22-24]. This means that epigenetic modification is a potentially powerful mechanism that determines the development of severe symptoms. In this regard, the DNA methylation landscape is central to the homoeostasis of the immune system and, in the clinical setting, DNA methylation changes in cancer affect the success of immunotherapy [25,26]. Examples for the latter scenario are the observation that DNA methylation signatures are associated with the clinical response to PD-1 checkpoint blockade in human tumors [27,28]. Similarly to the way GWAS studies were developed for COVID-19 [8,9], and our previous analyses of DNA methylation risk variants for other common human diseases [29,30], we have performed a comprehensive epigenome-wide association study (EWAS) in COVID-19 patients to identify candidate DNA methylation loci linked to the severity of the disease, particularly with respect to respiratory failure.

2. Methods

2.1. Study design and participants

Whole blood samples and clinical data from 407 patients with confirmed COVID-19 were retrospectively collected between March 7th 2020 and September 14th 2020 from fourteen Hospitals in Spain. Patients were eligible if they did not present the risk factors of comorbidities (obesity with a BMI ≥ 30, diabetes, hypertension, autoimmune disorders, and chronic cardiovascular or lung diseases), smoking habit or advanced age (> 61 years). Clinicopathological characteristics of the COVID-19 patients studied are summarized in Table 1. Our study was designed to find epigenetic biomarkers associated with COVID-19 severity in patients ≤ 61 years of age and without comorbidities, thus, we did not recruit all consecutive cases and it was not intended to be representative of the spectrum of the disease as a whole. Following the general guidelines of the National Institutes of Health for COVID-19 clinical staging (https://www.covid19treatmentguidelines.nih.gov/overview/clinical-presentation/), the patients were categorized in two groups: those asymptomatic or with a paucisymptomatic clinical status who were non-hospitalized (Group G3) compared with those who required hospitalized oxygen therapy, such as patients with nasal mask or nasal prongs non-invasive ventilation or high-flow oxygen ventimask (Group G2) and those requiring mechanical ventilation; additional organ support such as vasopressors; or extra corporeal membrane oxygenation (Group G1).

Research in context

Evidence before this study

The clinical spectrum of COVID-19 ranges from asymptomatic individuals and patients with mild symptomatology not requiring hospitalization to those admitted to the ward for respiratory failure due to pneumonia that may eventually led to acute respiratory distress syndrome (ARDS). Risk factors for worse clinical course include older age, male sex, and comorbidities such as diabetes, obesity and hypertension. Genetic susceptibility loci, such as blood group A, and inborn errors of type I IFN immunity have more recently been recognized as risk factors. To our knowledge, this is the first epigenome-wide association study (EWAS) of COVID-19 considering the severity of respiratory failure. Our search of PubMed on March 15th 2021, limited to articles in English, but not by date, using the terms "COVID-19″, "severity", "respiratory failure", "EWAS", "epigenomics", "DNA methylation, "biomarker" and "prediction", yielded no studies addressing this exact topic.

Added value of this study

Our findings demonstrate the existence of differential DNA methylation sites that distinguish COVID-19 patients with paucisymptomatic clinical status from those who will require hospitalized oxygen therapy, including mechanical ventilation and additional organ support measures. The epigenetic loci identified were mostly located within genes associated with the interferon response pathway. Using these DNA methylation biomarkers, we obtained an epigenomic signature that we have called "EPICOVID" that showed great accuracy in predicting COVID-19 severity.

Implications of all the available evidence

This study has identified new biomarkers associated with COVID-19 clinical severity that can be assessed for particular DNA methylation sites or included within a designed overall epigenomic signature. The determination of the DNA methylation status of these sites can be easily added to other molecular, cellular and clinical parameters to enable the more accurate prediction of the COVID-19 patients who might suffer the worst clinical outcomes, so that early intervention strategies may be devised to improve the cure rate and relieve the burden on the healthcare system.

The study protocol and the statistical analysis plan is described in Supplementary Methods.

2.2. Ethics

The protocol of this retrospective study was approved by the institutional ethics review boards of the participating institutions. Written informed consent was obtained from all participants.

2.3. DNA methylation data

The DNA methylation status of the COVID-19 samples was established using the Infinium MethylationEPIC Array (~850,000 CpG sites) following the manufacturer's instructions for the automated processing of arrays with a liquid handler (Illumina Infinium HD Methylation Assay Experienced User Card, Automated Protocol 15019521 v01), as previously described [31]. The DNA methylation beta values were obtained from the raw IDAT files by using the minfi package in R. Briefly, the methylation data pre-processing performed with the ChAMP package in R included normalization with the ssNoob procedure, filtering probes with a detection P-value greater than 0.01, NoCG Start, probes with SNPs, multihit start probes and XY chromosome probes. Confounding covariates in the methylation dataset were identified by the single-value decomposition (SVD) method [32]. The genomic analysis presented in the study was performed using the GRCh37 – hg19 human genome reference build, as described in the Illumina manifest file associated with the DNA methylation EPIC microarray. CpG methylation status was further validated by pyrosequencing using PyroMark Q48 system. Primers were designed using Qiagen's PyroMark Assay Design 2.0 software. Primer sequences are shown in Supplementary Methods.

2.4. Computational analyses

The EWAS analysis was performed by identifying the probes differentially methylated between asymptomatic/paucisymptomatic and severe COVID-19 cases. This involved deriving a linear model with the limma package in R, adjusted by the age and gender covariables, using the methylation beta values of the discovery dataset. We checked the data to ensure that the assumptions for the methods used (such as linearity) were met (Supplementary Methods). CpG probes with a Benjamini-Hochberg adjusted-P value less than 0.05 and an absolute difference in methylation beta greater than 0.10 were selected for further analysis. An additional feature selection step involved applying the elastic net method to the differentially methylated markers in order to select the best markers for prediction. A ten-fold cross-validation was used to identify the best alpha tuning parameter and the minimum lambda value. Gene Ontology (GO) terms associated to the candidate genes derived from EWAS, were examined using BioMart data-mining tool (Ensembl) to classify the genes according to biological processes. Significant GO terms were determined according to the hypergeometric test. The significantly differential DNA methylation sites were used to generate a predictive model of the COVID-19 clinical outcome. A supervised classification model based on the combination of six machine-learning algorithms was used to predict asymptomatic/paucisymptomatic and severe COVID-19 progression status. The classification methods used were random forest (rf), adaptive boosting (adaboost), multivariate adaptive regression splines (earth), k-nearest neighbour (knn), radial kernel support vector classifier (svmRadial) and logistic regression model (glmnet). The caretList function from the caretEnsembl package was used to fit the different models with the same resampling parameters (ten-fold cross-validation and five repeats). The final glm-weighted meta-model was generated from the six predictive models using the caretStack function from caretEnsembl. Model performance was assessed in terms of AUC, accuracy and kappa values. Finally, the prediction model was tested in the validation cohort and the corresponding confusion matrix was obtained. The hierarchical clustering analysis was performed using the Ward.D2 clustering method with Euclidean distances in the pheatmap package in R.

2.5. Role of funders

This work was mainly funded by the Unstoppable campaign of the Josep Carreras Leukaemia Foundation, the Cellex Foundation and the CERCA Programme of the Generalitat de Catalunya. Additional support was provided by the grant COVID19-PoC-6_17, Direcció General de Recerca i Innovació en Salut (DGRIS), Departament de Salut, Generalitat de Catalunya; the Spanish AIDS Research Network "RD16/0025/0006″ Co-funded by European Regional Development Fund/European Social Fund; "A way to make Europe"/"Investing in your future"; Programa de Intensificación de Investigadores INT 20/00031, Instituto de Salud Carlos III; PFIS FI18/00141 from Instituto de Salud Carlos III; Rio Hortega grant CM18/00145; the centre for Biomedical

Table 1
Clinical characteristics of the studied discovery, validation and entire cohorts of COVID-19 patients.

Characteristics	COVID-19 cohorts		
	Discovery cohort (*N* = 207)	Validation cohort (*N* = 200)	Entire cohort (*N* = 407)
Age (years) - Median [range]	43 [19 - 60]	42 [22 - 61]	42 [19 - 61]
Gender - Frequency (%)			
Female	108 (52.2%)	114 (57.0%)	222 (54.5%)
Male	99 (47.8%)	86 (43.0%)	185 (45.5%)
Oxygen therapy- Frequency (%)			
No oxygen support	105 (50.7%)	90 (45.0%)	195 (47.9%)
Mask or nasal prongs	30 (14.5%)	57 (28.5%)	87 (21.4%)
Non-invasive ventilation or high-flow oxygen	30 (14.5%)	20 (10.0%)	50 (12.3%)
Mechanical ventilation	33 (15.9%)	1 (0.5%)	34 (8.4%)
ECMO, pressors	9 (4.3%)	21 (10.5%)	30 (7.4%)
Any oxygen therapy (not specified)	0 (0%)	11 (5.5%)	11 (2.7%)
Severity group - Frequency (%)			
G1 (ICU)	75 (36.2%)	24 (12.0%)	99 (24.3%)
G2 (Hospitalization with O_2 support)	28 (13.5%)	86 (43.0%)	114 (28.0%)
G3 (Mild symptomatology, home)	104 (50.2%)	90 (45%)	194 (47.7%)
COVID-19 pneumonia - Frequency (%)			
No	93 (44.9%)	91 (45.5%)	184 (45.2%)
Yes	114 (55.1%)	89 (44.5%)	203 (49.9%)
Unknown	0 (0%)	20 (10%)	20 (4.9%)

Abbreviations: ECMO = Extracorporeal membrane oxygenation; ICU = Intensive care unit.

Research on Rare Diseases (CIBERER); “la Caixa” Banking Foundation (LCF/PR/GN18/51140001); and the Programa de Suport als Grups de Recerca AGAUR (2017SGR1080 and 2017SGR948). AP and ME are ICREA Research Professors. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript. The authors collected the data, and had full access to all of the data in the study. They also had the final decision and responsibility to submit the study results for publication.

3. Results

3.1. Patients and epigenotyping

Between March 7th 2020 and September 14th 2020, we collected whole blood samples from 407 adult patients diagnosed with COVID-19 that did not present the risk factors of comorbidities (obesity with a BMI $\geq$ 30, diabetes, hypertension, autoimmune disorders, and chronic cardiovascular and lung diseases), smoking habit or advanced age (>61 years) (Table 1). These patients were divided into a discovery cohort of 207 cases and a validation cohort of 200 cases (Table 1). The two cohorts did not present significant differences related to age (Mann-Whitney U test, P = 0.673), gender (Fisher’s exact test, P = 0.370), requirement of oxygen therapy (Fisher’s exact test, P = 0.275), severity of the disease (G3 vs G2/G1, Fisher’s exact test, P = 0.321) and the occurrence of pneumonia (Fisher’s exact test, P = 0.308). Related to ethnicity, although 72% of our samples were from the West-Eurasia group, Central-South America population was higher in the validation (33%) vs discovery (18%) cohort. DNA from the whole blood samples was extracted in all cases and hybridized to a DNA methylation microarray that interrogates almost 850,000 CpG sites of the human genome [31]. The goal of the study was to identify genomic loci with a differential DNA methylation status in patients with asymptomatic or paucisymptomatic clinical status who were non-hospitalized (Group G3, see Methods) compared with those who required hospitalized oxygen therapy (Groups G2 and G1, see Methods). A graphical schema representing the population of interest and the screening strategy used to identify epigenetic biomarkers associated with the severity of the disease is shown in Figure S1.

3.2. Epigenome-wide association analysis in the discovery cohort

The DNA methylation analysis of the 207 COVID-19 cases of the discovery cohort identified 51 CpG sites with a differential methylation status between the mild (G3, n = 104) and severe (G2 + G1, n = 103) groups (Table S1). Because male patients tend to have a worse clinical disease course [6,7] and the male patients in our entire cohort also exhibited more clinical severity (Pearson’s Chi-squared test with Yate’s continuity correction, $P = 1.35e^{-20}$), the candidate CpG sites were corrected for sex. In a similar manner, although we limited our study to patients $\leq$ 61 years of age to reduce the effect of the established association between older age and disease severity [6,7], our younger patients performed much better than the older patients (one-tailed Wilcoxon rank sum test, $P = 2.47e^{-07}$), so age correction also featured in the analysis. Overall, after correction for sex and age in the analysis of the originally identified 51 CpG sites, we found 44 CpGs whose DNA methylation status differed significantly between asymptomatic/paucisymptomatic and severe COVID-19 patients (Table S2). The Manhattan plot of the fully adjusted values of P from the EWAS for COVID-19 severity in the discovery cohort is shown in Fig. 1.

The genomic context of these distinctly methylated 44 CpG sites is shown in Table S3. Fifteen of these loci were located in genomic human regions with no currently described gene sequence; six were associated with non-coding RNA loci; and the other 23 CpG sites were located within 20 known coding genes (Table 2). RNA and/or protein for the studied COVID-19 patients was not available, but using a collection of 62 blood derived cell lines previously characterized for DNA methylation and expression patterns [33], we observed that for 17 of 20 (85%) coding genes the presence of hypermethylation located in the 5′-end regulatory region was significantly associated with transcript downregulation (Pearson's test, $P < 0.05$). To determine the activities of the identified 20 candidate genes derived from EWAS, we data-mined the Gene Ontology (GO) terms associated to each gene using BioMart tool (Ensembl). Significantly overrepresented genes (hypergeometric test, P value = $2.14e^{-06}$) associated with GO biological processes included those of “immune response”, “type I interferon signalling pathway”, “interferon-gamma-mediated signalling pathway”, “antigen processing and presentation”, “defence response to virus”, “cytokine-mediated signalling pathway” and “inflammatory response”. These processes are highly relevant to the degree of potential response to COVID-19 infection, since they are related to the capacity of the immune system to respond to viral infections. In this regard, the 20 genes containing the CpG methylation variants associated with COVID-19 severity include an overrepresentation of genes that mediate the response to interferon, a key pathway in the physiopathological pathway of the disease [8,9,34]. This reinforces the biological plausibility of the results of the EWAS.

Overall, 35% (7 of 20) of the genes identified were effectors of interferon signalling, implying that they are potential participants in regulating the efficacy of the immune system to deal with viral exposition. The interferon-related candidate genes with DNA methylation status associated with the disease were: AIM2 (absent in melanoma 2), HLA-C (major histocompatibility complex, class I, C), IFI44L (interferon-induced protein 44-like), CXCR2 (C-X-C motif chemokine receptor 2), KIFAP3 (Kinesin Associated Protein 3), SGMS1 (Sphingomyelin Synthase 1) and VIM (vimentin) (Table 2). The cases of AIM2 and HLA-C are particularly interesting because each gene is represented by two differentially methylated CpG sites in our analyses, making it more likely that they contribute to the onset of the disorder. AIM2 is a member of the innate immune system that exerts vital activity during viral infections, and is responsible for the assembly of a macromolecular complex called the inflammasome that unleashes

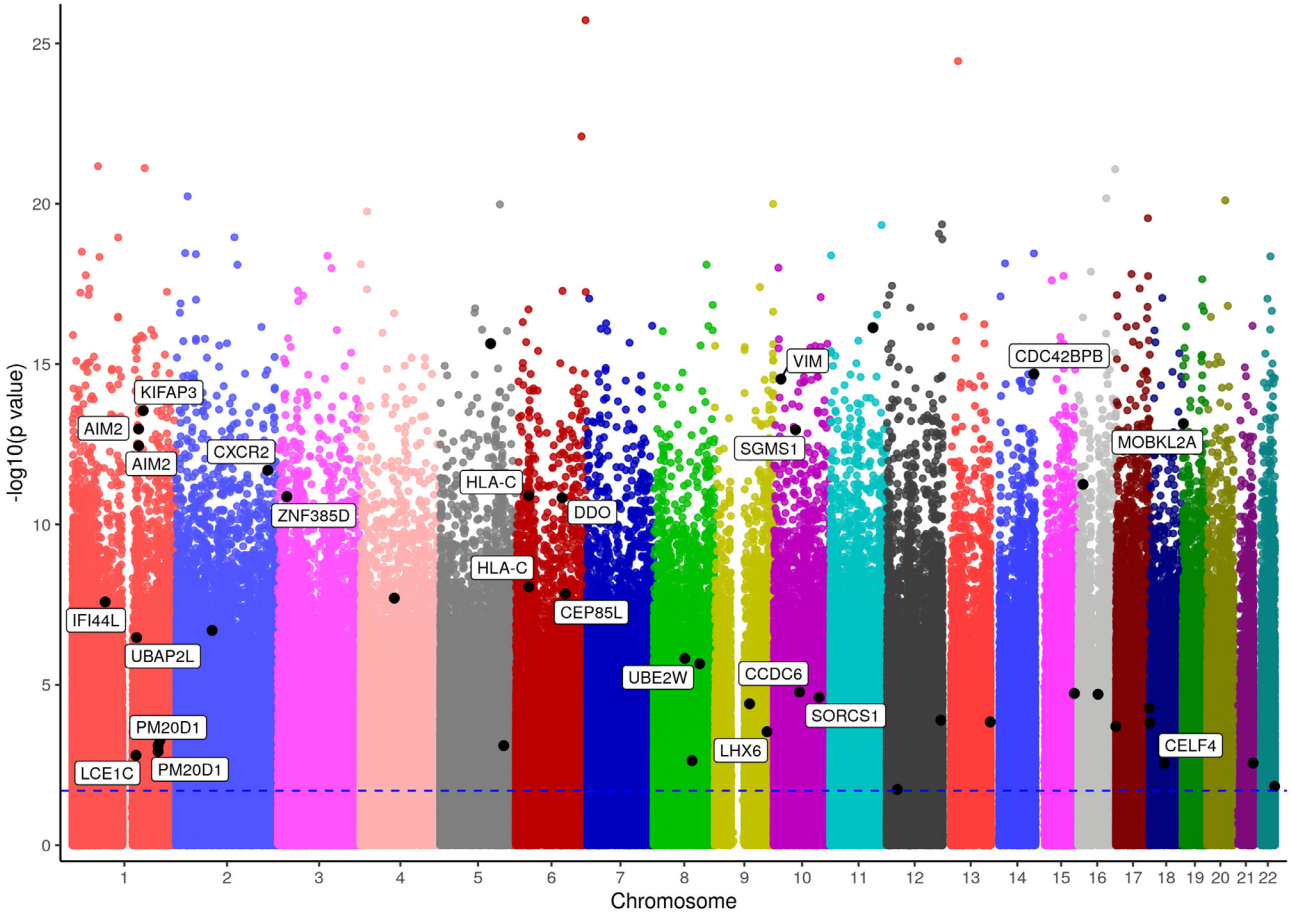

Fig. 1. Manhattan plot from the EWAS performed in the discovery cohort. The plot shows the result of the COVID-19 severity association test over 735.349 CpG positions. The blue dashed line indicates the genome-wide significance threshold of a Benjamini-Hochberg corrected P value lower than 0.05. Black dots represent the 44 CpGs signature labelled with their corresponding genes.

caspase-1 and triggers pro-inflammatory cytokines such as IL-1β and IL-18 [35]. HLA-C is another example of a candidate for mediating COVID-19 severity since specific HLA subtypes have been associated with susceptibility for many viral infections, including HIV, HBV, H1N1, HCV and HPV [36]. From a functional standpoint, HLA-C is involved in both innate and adaptive immunity, but most importantly, HLA-C haplotypes have also been linked to the clinical course of COVID-19 [37-40].

Only one gene, in addition to AIM2 and HLA-C, was represented by two CpG sites in our COVID-19 EWAS: PM20D1 (peptidase M20 domain-containing 1). PM20D1 has most often been categorized as a metabolic disease-associated gene, particularly in relation to obesity and diabetes [41], but genetic and epigenetic variants have also emerged as being a risk factor for neurodegenerative disorders such as Parkinson's [42] and Alzheimer's disease [43,44]. Thus, its additional link as a potential modulator of the severity of SARS-CoV-2 infection, described here, suggests that this peptidase lies at the crossroads of many cellular and pathological pathways.

We also examined in detail whether the DNA methylation status of asymptomatic/paucisymptomatic patients and those that eventually developed the severe symptomatology differed with respect to the genes that previous studies have identified as being candidates for COVID-19 infection susceptibility and/or prognosis. We did not observe any significant DNA methylation content in the 40 genes studied (Table S4), including the well-characterized ACE2 receptor and TMPRSS2 protease [45]; genes derived from GWAS [8,9]; genes obtained from the screening of inborn errors of type I IFN immunity in patients with life-threatening COVID-19 [10,11,34]; and other genes involved in host-cell recognition and binding (Table S4). Thus, there is no evident association between epigenetic events at these genes and the clinical course of COVID-19, reinforcing the specificity of our approach for identifying susceptibility to DNA methylation loci for COVID-19 severity that can easily be complemented by the additional genetic biomarkers.

3.3. Testing EWAS markers in the validation cohort and development of the EPICOVID signature

The predictive value of the single CpG sites whose DNA methylation status was associated with COVID-19 severity in the EWAS of our discovery cohort was confirmed in the validation cohort. Overall, of the 44 CpGs whose DNA methylation status differed significantly between asymptomatic/paucisymptomatic and severe COVID-19 patients after correction for sex and age, 37 (84.1%) were also significantly associated with worse clinical outcome in the validation cohort (Table S2). Of these 37 CpG sites, 19 loci were located in the aforementioned 20 gene-containing regions (95%) (Table 2). The Manhattan plot of the fully adjusted values of P from the EWAS for COVID-19 severity in the validation cohort is shown in Figure S2. Most importantly, when we analysed all the cases as a whole cohort, comprising the discovery and validation sets (n = 407), 95.5% of the CpG sites (42 of 44) still appeared to be associated with COVID-19 severity when controlling for sex and age (Table S2). This was the case even when we included all the 20 candidate genes identified (Table 2). The Manhattan plot of the fully adjusted values of P from the EWAS for COVID-19 severity in the entire cohort is shown in

Table 2
Epigenetic susceptibility loci in genes associated with severe COVID-19 with respiratory failure.

CpG ID	Chromosome Location	Gene name	Discovery Cohort				Validation Cohort				Entire Cohort			
			P value	Odds Ratio	P value Analysis corrected for age and sex	Odds Ratio Analysis corrected for age and sex	P value	Odds Ratio	P value Analysis corrected for age and sex	Odds Ratio Analysis corrected for age and sex	P value	Odds Ratio	P value Analysis corrected for age and sex	Odds Ratio Analysis corrected for age and sex
cg24145401	1:159,047,177	AIM2	2.70×10^{-10}	6.17×10^{7}	1.31×10^{-5}	3.34×10^{2}	3.57×10^{-21}	7.87×10^{18}	8.59×10^{-20}	3.49×10^{16}	1.65×10^{-30}	4.65×10^{28}	6.33×10^{-27}	3.25×10^{23}
cg17515347	1:159,047,163	AIM2	1.11×10^{-10}	2.03×10^{8}	7.51×10^{-6}	1.10×10^{3}	8.25×10^{-20}	9.06×10^{16}	1.01×10^{-17}	1.57×10^{14}	8.19×10^{-30}	6.39×10^{27}	1.40×10^{-25}	7.59×10^{21}
cg04736673	10:61,647,141	CCDC6	1.99×10^{-4}	1.90	1.45×10^{-3}	4.52×10^{-1}	7.20×10^{-8}	4.97×10^{3}	3.29×10^{-6}	1.59×10^{2}	7.20×10^{-12}	4.71×10^{7}	1.86×10^{-10}	1.69×10^{6}
cg02003183	14:103,415,882	CDC42BPB	6.08×10^{-12}	1.03×10^{10}	2.29×10^{-4}	7.79	2.24×10^{-9}	1.84×10^{5}	5.97×10^{-6}	7.48×10^{1}	8.41×10^{-21}	2.27×10^{17}	3.59×10^{-11}	1.03×10^{7}
cg15355235	18:35,001,518	CELF4	9.54×10^{-3}	1.53×10^{-2}	2.28×10^{-2}	2.10×10^{-2}	1.47×10^{-3}	1.38×10^{-1}	1.11×10^{-2}	2.30×10^{-2}	2.52×10^{-5}	6.81	5.12×10^{-4}	2.79×10^{-1}
cg10947500	6:118,985,774	CEP85L	9.99×10^{-7}	1.77×10^{3}	3.20×10^{-5}	9.67×10^{1}	9.32×10^{-8}	3.80×10^{3}	2.37×10^{-8}	3.15×10^{4}	1.30×10^{-14}	3.63×10^{10}	1.52×10^{-14}	3.17×10^{10}
cg19225688	2:218,990,043	CXCR2	1.04×10^{-9}	1.09×10^{7}	3.93×10^{-4}	2.70	1.37×10^{-15}	6.56×10^{11}	2.53×10^{-12}	2.96×10^{8}	1.01×10^{-24}	5.62×10^{21}	1.77×10^{-17}	2.97×10^{13}
cg02872426	6:110,736,772	DDO	4.73×10^{-9}	1.55×10^{6}	1.29×10^{-2}	3.80×10^{-2}	2.14×10^{-5}	1.27×10^{1}	3.29×10^{-6}	1.67×10^{2}	1.41×10^{-14}	3.33×10^{10}	1.68×10^{-8}	1.54×10^{4}
cg08309069	6:31,240,651	HLA-C	4.28×10^{-9}	1.77×10^{6}	1.11×10^{-3}	6.07×10^{-1}	2.25×10^{-2}	6.96×10^{-3}	5.22×10^{-2}	5.00×10^{-3}	1.28×10^{-9}	2.14×10^{5}	1.71×10^{-5}	8.44
cg05030953	6:31,241,000	HLA-C	6.79×10^{-7}	2.90×10^{3}	2.09×10^{-3}	2.83×10^{-1}	2.13×10^{-1}	6.13×10^{-4}	2.86×10^{-1}	1.00×10^{-3}	6.73×10^{-6}	2.76×10^{1}	5.12×10^{-4}	2.84×10^{-1}
cg13452062	1:79,088,559	IFI44L	1.55×10^{-6}	1.01×10^{3}	8.40×10^{-4}	9.77×10^{-1}	5.31×10^{-11}	9.10×10^{6}	1.01×10^{-7}	6.60×10^{3}	1.75×10^{-16}	3.73×10^{12}	1.11×10^{-11}	3.64×10^{7}
cg26931608	1:170,036,455	KIFAP3	4.10×10^{-11}	7.52×10^{8}	8.97×10^{-4}	8.18×10^{-1}	6.54×10^{-8}	5.50×10^{3}	4.93×10^{-5}	9.11	1.06×10^{-19}	1.37×10^{16}	5.17×10^{-11}	6.48×10^{6}
cg07796016	1:152,779,584	LCE1C	6.14×10^{-3}	2.58×10^{-2}	2.62×10^{-3}	1.93×10^{-1}	4.46×10^{-4}	5.01×10^{-1}	1.21×10^{-3}	2.36×10^{-1}	4.13×10^{-6}	4.62×10^{1}	3.21×10^{-6}	5.38×10^{1}
cg13571460	9:124,989,337	LHX6	1.69×10^{-3}	1.26×10^{-1}	3.40×10^{-4}	4.64	5.91×10^{-2}	2.41×10^{-3}	4.53×10^{-2}	5.00×10^{-3}	1.53×10^{-4}	9.81×10^{-1}	2.52×10^{-5}	5.53
cg07381806	19:2094,327	MOBKL2A	8.19×10^{-11}	3.00×10^{8}	3.93×10^{-4}	3.28	4.78×10^{-8}	7.62×10^{3}	4.91×10^{-6}	9.82×10^{1}	1.31×10^{-18}	8.42×10^{14}	2.11×10^{-10}	1.37×10^{6}
cg17178900	1:205,818,956	PM20D1	3.60×10^{-3}	4.94×10^{-2}	1.13×10^{-2}	4.40×10^{-2}	3.76×10^{-3}	4.95×10^{-2}	1.84×10^{-3}	1.44×10^{-1}	1.41×10^{-5}	1.26×10^{1}	7.50×10^{-6}	2.09×10^{1}
cg14893161	1:205,819,251	PM20D1	4.97×10^{-3}	3.34×10^{-2}	2.99×10^{-2}	1.60×10^{-2}	1.43×10^{-3}	1.42×10^{-1}	4.66×10^{-4}	6.54×10^{-1}	7.95×10^{-6}	2.32×10^{1}	5.46×10^{-6}	3.02×10^{1}
cg10188795	10:52,158,244	SGMS1	1.15×10^{-10}	1.92×10^{8}	1.30×10^{-4}	1.57×10^{1}	4.31×10^{-20}	2.23×10^{17}	1.45×10^{-16}	7.55×10^{12}	7.58×10^{-30}	7.09×10^{27}	1.32×10^{-21}	5.88×10^{17}
cg24795173	10:108,751,940	SORCS1	2.64×10^{-4}	1.32	2.10×10^{-3}	2.72×10^{-1}	1.23×10^{-2}	1.34×10^{-2}	1.17×10^{-2}	2.10×10^{-2}	5.33×10^{-6}	3.53×10^{1}	8.79×10^{-6}	1.74×10^{1}
cg14859874	1:154,238,265	UBAP2L	1.07×10^{-5}	8.28×10^{1}	4.90×10^{-4}	1.99	1.72×10^{-2}	9.34×10^{-3}	3.80×10^{-2}	6.00×10^{-3}	3.78×10^{-7}	5.68×10^{2}	3.30×10^{-6}	5.06×10^{1}
cg12682382	8:74,787,918	UBE2W	3.27×10^{-5}	1.95×10^{1}	2.37×10^{-3}	2.26×10^{-1}	1.42×10^{-8}	2.70×10^{4}	1.52×10^{-6}	4.15×10^{2}	1.82×10^{-13}	2.23×10^{9}	7.83×10^{-12}	5.86×10^{7}
cg26063719	10:17,273,187	VIM	7.68×10^{-12}	7.09×10^{9}	3.93×10^{-4}	2.77	6.58×10^{-18}	2.68×10^{14}	6.57×10^{-14}	1.35×10^{10}	2.66×10^{-29}	1.40×10^{27}	3.69×10^{-19}	1.69×10^{15}
cg06760111	3:21,551,522	ZNF385D	4.45×10^{-9}	1.68×10^{6}	1.40×10^{-2}	3.50×10^{-2}	9.63×10^{-7}	3.32×10^{2}	3.90×10^{-4}	8.87×10^{-1}	4.95×10^{-16}	1.21×10^{12}	1.96×10^{-7}	1.06×10^{3}

Notes: CpG ID corresponds to the unique CpG site identifier in the HumanMethylationEPIC array (Illumina). Chromosomal location denoted according human reference assembly GRCh37/hg19.
All depicted P values are Benjamini-Hochberg adjusted p values.

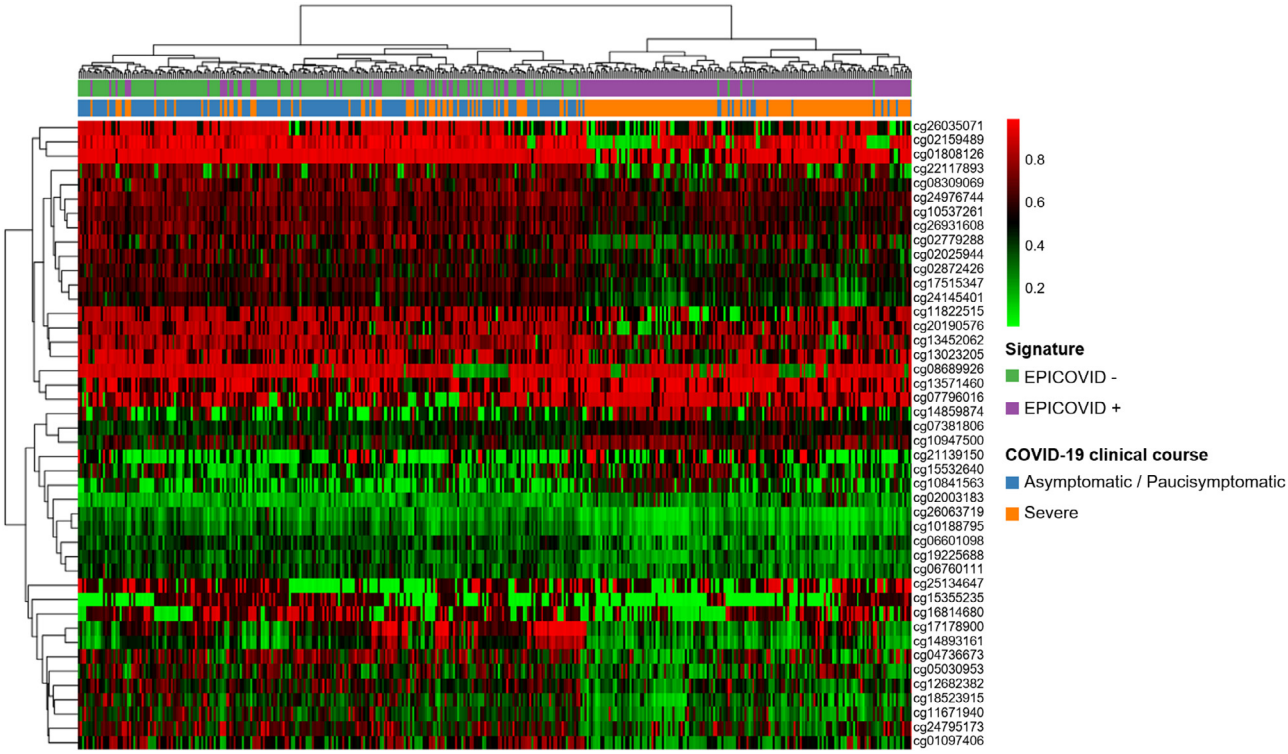

Fig. 2. Heatmap representing the entire cohort clustered by methylation beta values of the 44 CpGs defining the EPICOVID signature. Cluster analysis was performed using the Ward.D2 clustering method and assuming Euclidean distances.

Figure S3. Finally, although our study was specifically designed to identify epigenetic biomarkers associated with patients that required hospitalized oxygen therapy vs those asymptomatic or paucisymptomatic who were non-hospitalized (Table S2), we also show the differentially methylated CpG sites in all possible comparisons between G1/G2/G3 for the discovery, validation and entire cohort (Table S5) to foster future research efforts in the field.

Identifying single DNA methylation loci associated with the onset of COVID-19 disease could be very useful, but the development an overall epigenomic signature based on our EWAS data would also be extremely helpful for early stratification of COVID-19 patients according to their clinical risk. To achieve this, we selected the 44 methylation sites that, adjusted by gender and age, were correlated with COVID-19 severity (Table S2) to train in our discovery cohort (n = 207), using a metamodel generated by combinating six different machine learning algorithms (see Methods). Using this approach, we obtained an epigenomic signature, hereafter referred to as the EPICOVID signature, that predicted COVID-19 severity with 90.18% accuracy (mean Kappa = 0.804). Supervised hierarchical clustering using the COVID-19 signature also clearly distinguished two branches that were significantly enriched with respect to each condition: those who were asymptomatic/paucisymptomatic and those with respiratory failure (Fisher's exact test, $P = 7.29e^{-18}$) (Figure S4), providing additional evidence of the relevance of this epigenomic signature to disease severity.

Once we had found that EPICOVID signature was a predictor of COVID-19 disease severity in the discovery cohort, we examined whether the signature retained its predictive value in our validation cohort of COVID-19 patients (n = 200). We observed that the EPICOVID-positive signature was also associated with worse COVID-19 clinical course with a specificity of 88.18%, a sensitivity of 77.78%, and positive and negative predictive values (PPV and NPV) of 84.34% and 82.91%, respectively. The accuracy was 83.5% and the mean Kappa 0.6643. We also plotted the Receiver Operating Characteristic (ROC) curve and calculated the Area Under the Curve (AUC=92.1%) to further assess and visualize the model performance (Figure S5). As in the discovery cohort, supervised hierarchical clustering of the COVID-19 signature in the validation cohort also distinguished two branches that were significantly enriched with respect to paucisymptomatology compared with respiratory failure onset (Fisher's exact test, $P = 8.8e^{-16}$) (Figure S6). The use of the EPICOVID signature in the supervised hierarchical clustering for the entire cohort of COVID-19 cases (n = 407) confirmed the existence of two branches that classified patients as those exhibiting mild symptoms or a worse clinical course (Fisher's exact test, $P = 2.9e^{-50}$) (Fig. 2), further strengthening the value of the discovered DNA methylation signature to indicate the risk of disease severity. Importantly, the EPICOVID signature was not only associated with clinical parameters of COVID-19 severity, but also with its corresponding laboratory findings [46,47] such as the levels of the proinflammatory cytokine interleukin-6, c-reactive protein, ferritin, fibrinogen and D-dimer, and the total lymphocyte count (for all cases Fisher's exact test, $P < 0.01$) (Table S6).

To further demonstrate the specificity of the EPICOVID signature for COVID-19, we determined if the identified epigenomic profile was overrepresented in DNA methylation datasets available in the public functional genomics data repository GEO for other different respiratory inflammatory pathologies and diseases that involve hyperactivation of the inflammatory cascade. We observed that the EPICOVID signature was not enriched in other respiratory inflammatory diseases such as tuberculosis [48], chronic obstructive pulmonary disease (COPD) [49], asthma [50] and other respiratory allergies [51] (For all cases, hypergeometric test $P > 0.05$). In addition, the EPICOVID signature was also not enriched in other inflammatory or viral infectious diseases such as rheumatoid arthritis [52], multiple sclerosis [53], systemic lupus erythematosus [54], Sjögren's syndrome [55], inflammatory bowel disease [56], hepatitis C infection [57] and amongst HIV-infected individuals [58] (For all cases, hypergeometric

test $P > 0.05$). Thus, these results further support the specificity of the EPICOVID signature for COVID-19 cases.

We validated the DNA methylation data derived from the microarray platform with a different method, pyrosequencing (Methods). The pyrosequencing results of the methylation levels of 21 CpG sites (from the 44 CpG sites that define the EPICOVID signature) comparing asymptomatic/paucisymptomatic and severe COVID-19 cases (a total of 39 patients) matched the DNA methylation microarray data. Illustrative examples are shown in Figure S7. The use of more user-friendly PCR approaches, such as the described pyrosequencing technology, could facilitate the analyses at the common hospital laboratory level. In this regard, it is worth noting that, if instead of the comprehensive EPICOVID signature we selected the top five CpG sites associated with severity according to P-value (Table S2), its single differential methylation status was still associated with COVID-19 severity (69%−76% accuracy range). Thus, a more restricted signature derived from the heatmap and clustering analyses might be useful in future prospective multicentre studies.

Finally, having demonstrated the EPICOVID signature to be associated with patients who developed severe COVID-19 symptomatology, we set out to measure the frequency of the epigenomic profile in the general population. To achieve this goal, we determined the DNA methylation status of whole blood samples of 338 healthy individuals interrogated with the same epigenomic microarray platform [59-62]. All of them were $\leq$ 61-year-old adults, as in our discovery and validation cohorts. Most importantly, all these samples were collected before the emergence of COVID-19, so these donors were never exposed to the SARS-CoV-2 virus. Using the described cohort, we observed that only 13.3% (45 of 338) of individuals exhibited the EPICOVID+ signature, whereas 86.7% (293 of 338) did not carry the DNA methylation profile derived from our study that was associated with a severe COVID-19 clinical course. Without any definitive epidemiological data indicating which percentage of the described worldwide adult population infected by the SARS-CoV-2 virus will eventually require hospitalization with oxygen therapy, the epigenomic signature could be helpful to provide such estimation. These findings reinforce the aforementioned association between the EPICOVID+ signature and the severity of the disease, and its potential role as a useful biomarker in the design of early prevention and therapeutic strategies to reduce the morbidity and mortality of the disorder.

4. Discussion

To our knowledge, this is the first study to report an epigenome-wide association with COVID-19 severity. Using a large collection of individuals who were categorized as positive for the SARS-CoV-2 virus and achieved very different outcomes, from asymptomatic or mild symptomatology to ward hospitalization with respiratory support that could even require invasive mechanical ventilation, we interrogated close to 850,000 DNA methylation sites of the human genome and identified epigenetic loci that were associated with the worse clinical course. The epigenomic landscape of the severe COVID-19 patients obtained reflects an enrichment of genes involved in interferon-related pathways, involving innate and adaptive immunity, from interleukin and chemokine activity and viral response networks to the major histocompatibility and inflammasome complexes.

The most prominent cases are those revealed by the DNA methylation susceptibility sites discovered in the AIM2 and HLA-C genes. The AIM2 gene is expressed in the cellular cytosol and, upon detection of host-DNA released in the context of viral infections, initiates the assembly of the inflammasome complex, subsequent activation of the caspase-1 and a marked innate immune response involving release of cytokines IL-1β and IL-18. AIM2 cannot distinguish between viral DNA and self-DNA, so, whereas this lack of specificity provides a good defence against a wide spectrum of pathogens, the activation of AIM2 also affects host cells that produce exacerbated immune responses [35]. In this regard, viral infections of the lung associated with abnormal pulmonary inflammation, pneumonitis, fibrosis and respiratory tract sequelae have related AIM2 hyperactivity [63,64]. The EWAS-derived candidate HLA-C also probably contributes to the severity of COVID-19. HLA-C is found on the cell surface and its main function is to display intracellular peptide fragments to CD8+ cytotoxic T cells, triggering an immune response [36], a mechanism used for virus-infected host cells to induce immune surveillance and eliminate infected cells. However, as also occurred with AIM2, the overactivation of HLA-C can cause autoimmune reactions and the excessive release of cytokines leading to an exacerbated immune response. In this regard, HLA-C genetic variants are known to be associated with COVID-19 clinical outcome [37-40]. Importantly for the natural history of COVID-19 infection, GWAS have also shown that HLA-C constitutes a susceptibility locus for human pneumonia [65,66]. Similar links between the other candidate epigenetic targets identified here and immune imbalance associated with lung inflammation can be found, such as the case of IFI44L, which is overexpressed in the bronchoalveolar lavage of severe COVID-19 patients [67] and SARS-CoV-2-infected bronchial epithelial cells [68]. Thus, the genes identified by EWAS are *bona fide* players in COVID-19 pathophysiology that merit further functional and translational attention.

The development of the EPICOVID signature to predict the risk of COVID-19 severity upon SARS-CoV-2 infection could also be of value in the clinical setting. Given the pandemic extension of the disease, all the actions that might improve patient care by optimizing the efficiency of the medical resources and costs are welcome additions to our prognostic portfolio. In this regard, early categorization and stratification of COVID-19 patients according to their potential clinical severity could also help reduce associated morbidity and mortality. The epigenetic biomarker-based COVID-19 risk assessment signature developed in our study, combined with other genetic, cellular, serological and clinical parameters, could identify patients who require close monitoring and early active treatments to prevent the progression of the disease as far as possible. Our findings might also have implications for COVID-19 vaccination. In these pressing times, where the newly developed vaccines against COVID-19 are not yet universally available, the identification of particular sets of individuals who become critically ill when infected, compared with those who experience no or mild symptoms could be of great benefit to public health triage. This is an exciting area in which comprehensive and rigorous studies, dealing with more biological questions, are warranted in the immediate future. For example, vaccine adjuvants activate inflammasomes, amongst them the one assembled by AIM2 [69], and it is possible that epigenetic variants of this gene and others influence the efficacy of COVID-19 vaccination.

In conclusion, our results indicate that the epigenomic background of the individuals infected by SARS-CoV-2 affects the degree of COVID-19 severity. Together with other laboratory and clinical characteristics, it would be helpful to be able to predict who will experience a pure asymptomatic onset or a severe form that compromises respiratory function in patients with no associated comorbidities. For other COVID-19 populations, such as older patients and those with previous pathologies, the extrapolation of these data would require further research. The DNA methylation landscape discovered relates to the immune function of the host, suggesting that when the infection has not been cleared by the immune system, it elicits an inflammatory response in the lung and other organs mediated by inflammasome, major histocompatibility complex, cytokine, interleukin and interferon-induced genes that are associated with the clinical severity of COVID-19.

Contributors

MCdM, VD and ME designed the study, contributed to the analysis, and wrote the first draft of the manuscript. LPS and AP did further

data analysis. DAE and CA performed the DNA methylation microarray assay. In-depth clinical and pathological characterization and recruitment of patients were carried out by MR, SAA, JT, JVR, VVS, ARP, JVG, JPH, SA, CC, AR, LR, BD, DD, MJA, LLC, AMP, JPT, IFC, PV, JT, RC, AMN, PSP, FV and AP. All authors helped draft the manuscript or revise it critically for significant intellectual content, and made substantial contributions to the concept and design of the study and acquisition, analysis and interpretation of data.

Data sharing

The complete DNA methylation raw data of the 407 COVID-19 cases is available on the GEO repository under accession number GSE168739 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168739.

Declaration of Competing Interest

Dr. Esteller reports grants from Ferrer International, personal fees from Quimatryx, outside the submitted work. The other authors declare no conflicts of interest.

Acknowledgements

We thank the Health Department and Centres de Recerca de Catalunya (CERCA) Programme of the Generalitat de Catalunya, the Josep Carreras Leukaemia Foundation and the Cellex Foundation for institutional support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2021.103339.

References

[1] Zhu N, Zhang D, Wang W, et al. China novel coronavirus investigating and research team. A novel coronavirus from patients with pneumonia in China. *N Engl J Med* 2020 2019;382:727–33.
[2] Coronaviridae study group of the international committee on taxonomy of viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020;5:536–44.
[3] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese center for disease control and prevention. JAMA 2020;323:1239–42.
[4] Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. N Engl J Med 2020 published online May 15.. doi: 10.1056/NEJMcp2009575.
[5] Marini JJ, Gattinoni L. Management of COVID-19 respiratory distress. JAMA 2020;323:2329–30.
[6] Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020;395:1054–62.
[7] Li X, Xu S, Yu M, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol 2020;146:110–8.
[8] Covid-19 GWAS Group Severe, Ellinghaus D, Degenhardt F, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med 2020;383:1522–34.
[9] Pairo-Castineira E, Clohisey S, Klaric L, et al. Genetic mechanisms of critical illness in Covid-19. Nature 2020 published online Dec 11. doi: 10.1038/s41586-020-03065-y.
[10] Zhang Q, Bastard P, Liu Z, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science 2020;370:eabd4570.
[11] Bastard P, Rosen LB, Zhang Q, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science 2020;370:eabd4585.
[12] Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. N Engl J Med 2018;378:1323–34.
[13] Berdasco M, Esteller M. Clinical epigenetics: seizing opportunities for translation. Nat Rev Genet 2019;20:109–27.
[14] Esteller M, Garcia-Foncillas J, Andion E, et al. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. N Engl J Med 2000;343:1350–4.
[15] Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005;352:997–1003.
[16] Moran S, Martínez-Cardús A, Sayols S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol 2016;17:1386–95.
[17] Capper D, Stichel D, Sahm F, et al. Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. Acta Neuropathol 2018;136:181–210.
[18] Fernandez AF, Rosales C, Lopez-Nieva P, et al. The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. Genome Res 2009;19:438–51.
[19] Pei Y, Wong JH, Robertson ES. Herpesvirus Epigenetic Reprogramming and Oncogenesis. Annu Rev Virol 2020;7:309–31.
[20] Moron-Lopez S, Urrea V, Dalmau J, et al. The genome-wide methylation profile of CD4+ T cells from HIV-infected individuals identifies distinct patterns associated with disease progression. Clin Infect Dis 2020 published online Jul 26. doi: 10.1093/cid/ciaa1047.
[21] Oriol-Tordera B, Berdasco M, Llano A, et al. Methylation regulation of antiviral host factors, interferon stimulated genes (ISGs) and T-cell responses associated with natural HIV control. PLoS Pathog 2020;16:e1008678.
[22] Grifoni A., Weiskopf D., Ramirez S.I., et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell*2020; 181: 1489−501.e15.
[23] Kreer C, Zehner M, Weber T, et al. Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients. Cell 2020;182:843–54 e12.
[24] Meckiff BJ, Ramírez-Suástegui C, Fajardo V, et al. Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive CD4+ T cells in COVID-19. Cell 2020;183:1340–53 e16.
[25] Loo Yau H, Ettayebi I, De Carvalho DD. The cancer epigenome: exploiting its vulnerabilities for immunotherapy. Trends Cell Biol 2019;29:31–43.
[26] Villanueva L, Álvarez-Errico D, Esteller M. The contribution of epigenetics to cancer immunotherapy. Trends Immunol 2020;41:676–91.
[27] Duruisseaux M, Martínez-Cardús A, Calleja-Cervantes ME, et al. Epigenetic prediction of response to anti-PD-1 treatment in non-small-cell lung cancer: a multicentre, retrospective analysis. Lancet Respir Med 2018;6:771–81.
[28] Jung H, Kim HS, Kim JY, et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nat Commun 2019;10:4278.
[29] Heyn H, Sayols S, Moutinho C, et al. Linkage of DNA methylation quantitative trait loci to human cancer risk. Cell Rep 2014;7:331–8.
[30] Ongen H, Andersen CL, Bramsen JB, et al. Putative cis-regulatory drivers in colorectal cancer. Nature 2014;512:87–90.
[31] Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics 2016;8:389–99.
[32] Teschendorff AE, Menon U, Gentry-Maharaj A, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS ONE 2009;4:e8274.
[33] Iorio F, Knijnenburg TA, Vis DJ, et al. A landscape of pharmacogenomic interactions in cancer. Cell 2016;166:740–54.
[34] van der Made CI, Simons A, Schuurs-Hoeijmakers J, et al. Presence of genetic variants among young men with severe COVID-19. JAMA 2020;324:1–11.
[35] Kumari P, Russo AJ, Shivcharan S, Rathinam VA. AIM2 in health and disease: inflammasome and beyond. Immunol Rev 2020;297:83–95.
[36] Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nat Rev Immunol 2018;18:325–39.
[37] Sakuraba A, Haider H, Sato T. Population difference in allele frequency of HLA-C*05 and its correlation with COVID-19 mortality. Viruses 2020;12:1333.
[38] Lorente L, Martín MM, Franco A, et al. HLA genetic polymorphisms and prognosis of patients with COVID-19. Med Intensiva 2020 published online Sep 6.. doi: 10.1016/j.medin.2020.08.004.
[39] Pisanti S, Deelen J, Gallina AM, et al. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. J Transl Med 2020;18:352.
[40] Moura RR, Agrelli A, Santos-Silva CA, et al. Immunoinformatic approach to assess SARS-CoV-2 protein S epitopes recognised by the most frequent MHC-I alleles in the Brazilian population. J Clin Pathol 2020 published online Aug 5. doi: 10.1136/jclinpath-2020-206946.
[41] Benson KK, Hu W, Weller AH, et al. Natural human genetic variation determines basal and inducible expression of PM20D1, an obesity-associated gene. Proc Natl Acad Sci U S A 2019;116:23232–42.
[42] Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet 2009;41:1303–7.
[43] Sanchez-Mut JV, Heyn H, Silva BA, et al. PM20D1 is a quantitative trait locus associated with Alzheimer's disease. Nat Med 2018;24:598–603.
[44] Sanchez-Mut JV, Glauser L, Monk D, Gräff J. Comprehensive analysis of PM20D1 QTL in Alzheimer's disease. Clin Epigenetics 2020;12:20.
[45] Singh H, Choudhari R, Nema V, Khan AA. ACE2 and TMPRSS2 polymorphisms in various diseases with special reference to its impact on COVID-19 disease. Microb Pathog 2020;150:104621.
[46] Ji P, Zhu J, Zhong Z, et al. Association of elevated inflammatory markers and severe COVID-19: a meta-analysis. Medicine (Baltimore) 2020;99:e23315.
[47] Chaudhary R, Garg J, Houghton DE, et al. Thrombo-inflammatory biomarkers in COVID-19: systematic review and meta-analysis of 17,052 patients. Mayo Clin Proc Innov Qual Outcomes 2021 Feb 8. doi: 10.1016/j.mayocpiqo.2021.01.009.
[48] Chen YC, Hsiao CC, Chen TW, et al. Whole genome DNA methylation analysis of active pulmonary tuberculosis disease identifies novel epigenotypes: pARP9/miR-505/RASGRP4/GNG12 gene methylation and clinical phenotypes. Int J Mol Sci 2020;21:3180.

[49] Chen YC, Tsai YH, Wang CC, et al. Epigenome-wide association study on asthma and chronic obstructive pulmonary disease overlap reveals aberrant DNA methylations related to clinical phenotypes. Sci Rep 2021;11:5022.
[50] Yang IV, Richards A, Davidson EJ, et al. The nasal methylome: a key to understanding allergic asthma. Am J Respir Crit Care Med 2017;195:829–31.
[51] Langie SA, Szarc Vel Szic K, Declerck K, et al. Whole-genome saliva and blood DNA methylation profiling in individuals with a respiratory allergy. PLoS ONE 2016;11 e0151109.
[52] Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013;31:142–7.
[53] Kular L, Liu Y, Ruhrmann S, et al. DNA methylation as a mediator of HLA-DRB1*15:01 and a protective variant in multiple sclerosis. Nat Commun 2018;9:2397.
[54] Joseph S, George NI, Green-Knox B, et al. Epigenome-wide association study of peripheral blood mononuclear cells in systemic lupus erythematosus: identifying DNA methylation signatures associated with interferon-related genes based on ethnicity and SLEDAI. J Autoimmun 2019;96:147–57.
[55] Brække Norheim K, Imgenberg-Kreuz J, Jonsdottir K, et al. Epigenome-wide DNA methylation patterns associated with fatigue in primary Sjögren's syndrome. Rheumatology (Oxford) 2016;55:1074–82.
[56] Ventham NT, Kennedy NA, Adams AT, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat Commun 2016;7:13507.
[57] Zhang X, Hu Y, Justice AC, Li B, et al. DNA methylation signatures of illicit drug injection and hepatitis C are associated with HIV frailty. Nat Commun 2017;8:2243.
[58] Oriol-Tordera B, Berdasco M, Llano A, et al. Methylation regulation of antiviral host factors, interferon stimulated genes (ISGs) and T-cell responses associated with natural HIV control. PLoS Pathog 2020;16:e1008678.
[59] Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol 2018;19:64.
[60] Novakovic B, Lewis S, Halliday J, et al. Assisted reproductive technologies are associated with limited epigenetic variation at birth that largely resolves by adulthood. Nat Commun 2019;10:3922.
[61] Soriano-Tárraga C, Lazcano U, Giralt-Steinhauer E, et al. Identification of 20 novel loci associated with ischaemic stroke. Epigenome-wide association study. Epigenetics 2020;15:988–97.
[62] Gallego-Fabrega C, Cullell N, Soriano-Tárraga C, et al. DNA methylation of MMPs and TIMPs in atherothrombosis process in carotid plaques and blood tissues. Oncotarget 2020;11:905–12.
[63] Yoo JK, Kim TS, Hufford MM, Braciale TJ. Viral infection of the lung: host response and sequelae. J Allergy Clin Immunol 2013;132:1263–76.
[64] Ma R, Ortiz Serrano TP, Davis J, Prigge AD, Ridge KM. The cGAS-STING pathway: the role of self-DNA sensing in inflammatory lung disease. FASEB J 2020;34:13156–70.
[65] Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun 2017;8:599.
[66] Khadzhieva MB, Kuzovlev AN, Salnikova LE. Pneumonia: host susceptibility and shared genetics with pulmonary function and other traits. Clin Exp Immunol 2019;198:367–80.
[67] Shaath H, Vishnubalaji R, Elkord E, Alajez NM. Single-cell transcriptome analysis highlights a role for neutrophils and inflammatory macrophages in the pathogenesis of severe COVID-19. Cells 2020;9:2374.
[68] Vishnubalaji R, Shaath H, Alajez NM. Protein coding and long noncoding RNA (lncRNA) transcriptional landscape in SARS-CoV-2 infected bronchial epithelial cells highlight a role for interferon and inflammatory response. Genes (Basel) 2020;11:760.
[69] Reinke S, Thakur A, Gartlan C, Bezbradica JS, Milicic A. Inflammasome-mediated immunogenicity of clinical and experimental vaccine adjuvants. Vaccines (Basel) 2020;8:554.