

Host methylation predicts SARS-CoV-2 infection and clinical outcome

Iain R. Konigsberg

University of Colorado

Bret Barnes

Illumina

Monica Campbell

University of Colorado

Elizabeth Davidson

University of Colorado

Yingfei Zhen

University of Colorado

Olivia Pallisard

University of Colorado

Meher Boorgula

University of Colorado

Corey Cox

University of Colorado

Debmalya Nandy

University of Colorado

Souvik Seal

University of Colorado

Kristy Crooks

University of Colorado

Evan Sticca

University of Colorado

Genelle F. Harrison

University of Colorado

Andrew Hopkinson

University of Colorado

Alexis Vest

University of Colorado

Cosby G. Arnold

University of Colorado

Michael G. Kahn

University of Colorado

David P. Kao

University of Colorado

Brett R. Peterson

University of Colorado

Stephen J. Wicks

University of Colorado

Debashis Ghosh

University of Colorado

Steve Horvath

University of California Los Angeles

Wanding Zhou

University of Pennsylvania

Rasika A. Mathias

Johns Hopkins University

Paul J. Norman

University of Colorado

Rishi Porecha

Illumina

Ivana V. Yang

University of Colorado

Christopher R. Gignoux

University of Colorado

Andrew Monte

University of Colorado

Alem Taye

Illumina

Kathleen C. Barnes (✉ KATHLEEN.BARNES@cuanschutz.edu)

University of Colorado

Research Article

Keywords: SARS-CoV-2, COVID-19, epigenetics, disease severity, predictive classifier, precision medicine

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-334297/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Host methylation predicts SARS-CoV-2 infection and clinical outcome

Iain R. Konigsberg^{1*}, Bret Barnes^{2*}, Monica Campbell¹, Elizabeth Davidson¹, Yingfei Zhen¹, Olivia Pallisard¹, Meher Boorgula¹, Corey Cox¹, Debmalya Nandy³, Souvik Seal³, Kristy Crooks¹, Evan Sticca¹, Genelle F. Harrison¹, Andrew Hopkinson¹, Alexis Vest¹, Cosby G. Arnold¹, Michael G. Kahn¹, David P. Kao¹, Brett R. Peterson¹, Stephen J. Wicks¹, Debashis Ghosh³, Steve Horvath⁴, Wanding Zhou⁵, Rasika A. Mathias^{1,6}, Paul J. Norman¹, Rishi Porecha², Ivana V. Yang^{1,**}, Christopher R. Gignoux^{1,**}, Andrew A. Monte^{1,**}, Alem Taye^{2,**}, and Kathleen C. Barnes^{1,**,#}

¹School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO; ²Illumina, Inc., San Diego, CA; ³School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO; ⁴University of California Los Angeles, Los Angeles, CA; ⁵The Children's Hospital of Philadelphia, Philadelphia, PA; ⁶Johns Hopkins University, Baltimore, MD

*Contributed equally to this work

**Equally contributing senior author

#Corresponding author

Abstract

Since the onset of the SARS-CoV-2 pandemic, most clinical testing has focused on RT-PCR¹. Host epigenome manipulation post coronavirus infection²⁻⁴ suggests that DNA methylation signatures may differentiate patients with SARS-CoV-2 infection from uninfected individuals, and help predict COVID-19 disease severity, even at initial presentation. We customized Illumina's Infinium Methylation EPIC array to enhance immune response detection and profiled peripheral blood samples from 164 COVID-19 patients with longitudinal measurements of disease severity and 296 patient controls. Epigenome-wide association analysis revealed 13,033 genome-wide significant methylation sites for case-vs-control status. Genes and pathways involved in interferon signaling and viral response were significantly enriched among differentially methylated sites. We observe highly significant associations at genes previously reported in genetic association studies (*e.g.* *IRF7*, *OAS1*). Using machine learning techniques, models built using sparse regression yielded highly predictive findings: cross-validated best fit AUC was 93.6% for case-vs-control status, and 79.1%, 80.8%, and 84.4% for hospitalization, ICU admission, and progression to death, respectively. In summary, the strong COVID-19-specific epigenetic signature in peripheral blood driven by key immune related pathways related to infection status, disease severity, and clinical deterioration provides insights useful for diagnosis and prognosis of patients with viral infections.

Main

Coronaviruses (CoV) comprise a large group of human and animal pathogens, including the novel enveloped RNA betacoronavirus referred to as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)⁵. This pathogen is associated with coronavirus disease 2019 (COVID-19) first identified in Wuhan, China in 2019⁶ and declared a pandemic on March 11, 2020⁷. Since the onset of the pandemic, multiple tests for diagnosing COVID-19 have been launched, including real-time reverse transcriptase–polymerase chain reaction (RT-PCR), specific antibody detection, and next-generation sequencing assays that query for current or past infections¹. With the exception of next-generation sequencing, which can discern viral subtypes, most diagnostic tests are viral strain dependent, can carry a high false negative rate, do not discern if the virus is viable and replicating, and do not predict clinical outcomes of infection^{1,8,9}. For example, pre-symptomatic patients may test negative^{10,11} while patients who have recovered may continue to test positive though they are no longer infectious¹². Accurate diagnostics are urgently required to control continued communal spread, to better understand host response, and for the development of vaccines and antivirals¹³.

Individuals infected with SARS-CoV-2 have a variable course of infection, ranging from asymptomatic to death. Although the fatality rate varies tremendously according to demographic characteristics and co-morbidities¹⁴, the U.S ranks as one of the countries with the highest COVID-19 mortality rates¹⁵. Identification of which SARS-CoV-2 infected patients are most likely to develop severe disease would enable clinicians to triage patients via augmented clinical decision support. Having more information on disease severity has recently become critical due to widespread lack of hospital and ICU capacity, necessitating difficult decisions about resource triage. To our knowledge, no test can predict COVID-19 clinical course or severity, although work on cytokine abundance ratios after hospitalization has been proposed as a prognostic indicator of severe outcomes¹⁶.

There is considerable evidence that enveloped RNA viruses such as coronaviruses can manipulate the host's epigenome via evolved functions that antagonize and regulate the host innate immune antiviral defense processes^{2,3}, specifically via DNA methylation. Viral-mediated antagonism of antigen-presentation gene expression in the case of Middle East respiratory syndrome coronavirus (MERS-CoV) was shown to occur via DNA methylation⁴. DNA methylation changes at cytosine-phosphate-guanine (CpG) sites have been increasingly leveraged in the emerging field of clinical epigenetics to characterize unique epigenetic signatures that diagnose disease. To date, considerable success has been demonstrated in developing highly accurate and robust machine learning-based disease classifiers using DNA methylation patterns to differentiate Mendelian disorders¹⁷, behavior disorders¹⁸, coronary artery disease¹⁹, and some cancers^{20,21,22}. Consequently integration of a methylation-based disease classification can result in relevant improvement in clinical practice^{23,24}.

With a goal to leverage Illumina's Infinium Methylation EPIC Array to classify differential methylation signatures of SARS-CoV-2 positive (hereafter referred to as SARS-Cov-2+, regardless of additional symptoms) and control peripheral blood DNA samples (either confirmed SARS-CoV-2 negative or samples collected prior to the SARS-CoV-2 pandemic), we conducted this study to determine if DNA methylation patterns could differentiate SARS-CoV-2 infected patients from non-infected patients. Our secondary objective was to determine if DNA methylation patterns could differentiate patients with SARS-CoV-2 infection who go on to develop severe disease.

RESULTS

Study Cohort

We identified 675 patients tested for either SARS-CoV-2 or other acute upper respiratory infections. Of these, 164 were SARS-Cov-2+ by RT-PCR; 58 historical EMSB patients had positive (non-SARS-Cov-2) acute upper respiratory viral RT-PCR tests, 7 had positive (non-SARS-Cov-2) acute upper respiratory viral RT-PCR tests during the pandemic, and 296 were negative for all viral infections and thus served as controls. We excluded 32 samples from the dataset as these were derived from a run with failed hybridization and removed 8 duplicates, resulting in a final cohort of 525 (**Fig. 1**). **Extended Data Table 1** summarizes the demographics and clinical outcomes of patients tested, including proportion of patients with other acute upper respiratory infections.

Disease-Specific DNA Methylation Signature and Differentially Methylated Probes

We first performed an epigenome-wide association study (EWAS) to identify biological signals associated with COVID-19 disease status. After adjustment for age, sex, array position (batch effect), cell proportions via ReFACToR and ancestry via EPISTRUCTURE components, EWAS of COVID-19 disease status in 164 SARS-Cov-2+ compared to 296 controls yielded 13,033 significant CpGs at false discovery (FDR)-adjusted P -value <0.05 (**Fig. 2 and Supplementary Table 1**), with moderate inflation that is typical of EWAS²⁵. In total, we observed 35 probes with a marginal P -value $<10^{-20}$, and 183 with a P -value $<10^{-10}$. We observed that 52.1% of all significant probes are hypermethylated; however, 78% of the top 100 probes sorted by adjusted p -value are hypomethylated. Custom probes on the EPIC+ chip are enriched in significant EWAS results (P -value = 9.94×10^{-7} , Fisher's Exact Test): specifically, 1.72% of EPIC probes are significant as opposed to 2.51% of custom probes. Principal component analysis of top associations reveals clustering by COVID-19 disease status (**Extended Data Fig. 1**). Because of concerns that population admixture may confound results, the COVID-19 disease status EWAS was repeated with EHR-defined race and ethnicity as additional covariates beyond that modeled via EPISTRUCTURE and mixed effects modeling. This had a minimal effect on results (data not shown).

Top hypomethylated CpG sites show strong enrichment for interferon and viral response-related pathways including Type I Interferon Signaling Pathway (KEGG, adjusted P -value = 7.40×10^{-10}) and Negative Regulation of Viral Genome Replication (GO:BP, adjusted P -value = 1.93×10^{-6} ; **Extended Data Fig. 2a**). Hypermethylated CpG sites also show enrichment for relevant biological processes such as Focal Adhesion (GO:CC, adjusted P -value = 0.0187; **Extended Data Fig. 2b**). cg17114584, the third most significant probe with an adjusted P -value of 1.78×10^{-43} , shows 16.9% hypomethylation in cases. This CpG is located in exon 6 of the interferon regulatory factor 7 (*IRF7*). *IRF7* encodes a transcription factor that regulates expression of interferon α and β , as well as interferon-stimulated genes. Other top CpGs are in genes relevant to viral response: *OAS1* (2'-5'-Oligoadenylate Synthetase 1) is interferon-induced and activates RNase L, which degrades viral (and cellular) RNA (adjusted P -value 1.05×10^{-21} , 3.8% methylation change). *MX1* encodes an interferon-induced GTPase that inhibits viral replication. *DTX3L* and *PARP9* form a complex that is involved in interferon-mediated antiviral defenses. This complex has also been shown to promote M1 polarization in macrophages by preventing STAT1 phosphorylation²⁶. *IFIT3* encodes another interferon induced antiviral protein. Overall, we observe strong hypomethylation of interferon and viral response related pathways, which is expected as these pathways are activated transcriptionally in SARS-Cov-2+ individuals²⁷.

Specificity of the COVID-19 Disease Signature from Other Respiratory Infections

We next compared 164 SARS-Cov-2+ samples to 65 samples with other upper respiratory infections to determine the specificity of the methylation signature to SARS-CoV-2. This analysis yielded 1,501 significant CpGs (adjusted P -value <0.05) (**Supplementary Table 2**), of which 780 (52%) were present in the SARS-Cov-2+ compared to controls analysis (**Figure 3**). Comparison of 65 other (non- SARS-Cov-2) upper respiratory infection samples to controls yielded 516 significant CpGs (**Supplementary Table 3**), of which 116 (22%) were present in the SARS-Cov-2+ compared to controls analysis. Furthermore, examination of the strength of the signal demonstrates that the shared probes in the SARS-Cov-2+ versus control and SARS-Cov-2+ versus other upper respiratory infections analysis have low p -values and high effect sizes, whereas this is not the case for probes shared by SARS-Cov-2+ versus control and other upper respiratory infections versus control analyses (**Extended Data Fig. 3a**). These comparisons suggest high specificity of the COVID-19 disease epigenetic signature. To further investigate this, we examined the significant CpGs from our COVID-19 disease signature compared to control EWAS. We observe the same trend of high correlation of effect sizes (methylation change) in SARS-Cov-2+ compared to control and SARS-Cov-2+ compared to other respiratory infections (Pearson $R=0.87$; $p<2.2\times 10^{-16}$) and very low correlations of effect sizes in SARS-Cov-2+ compared to control and other upper respiratory infections compared to control analyses (Pearson $R=-0.027$; $p=0.0022$) (**Extended Data Fig 3b**). While we do not have sufficient power to examine the specific viruses (other coronaviruses, influenza, etc.), these results strongly point to the specificity of our COVID-19 disease epigenetic signature to detect SARS-CoV-2 infection.

Development and Validation of a Classification Model for Prediction of Disease Classes and Disease Severity

To combine methylation data across the genome into a single predictor, we employed machine learning models of sparse regression trained via cross-validated *glmnet*²⁸ as described in Methods. To determine the sensitivity of our model, 460 subjects (SARS-Cov-2+ vs controls) from the testing cohort were supplied to the classification model, with prediction optimized after the approach defined in Methods. All models showed relative stability across iterations (**Extended Data Fig 4**), and yielded sparse results. Details of each top model are available in **Extended Data Table 2**. The best-fitting model has a performance of 93.6% in cross-validation for detecting SARS-CoV-2 infection (**Fig 4a,b**). Model performance was similar in females and males (93.7% and 93.5%, respectively). In addition, model performance on older individuals and younger individuals (median age=56 years) was comparable: 94.4% and 92.8%, respectively. Similarly, race/ethnicity information was not significantly correlated with case-control score (all groups $p>0.05$). When age and race/ethnicity categories were included in a multivariable model along with our prediction score, no additional covariates significantly predicted COVID-19 disease status (all other $p>0.4$). Similarly, BMI was not associated ($p\sim 0.4$).

To determine the direct association of methylation with clinical outcomes, an additional logistic regression was performed for the subset of individuals with complete blood cell count (CBC) data (341 individuals total). The inclusion of additional blood cell count data did not impact the association between the methylation score and outcome (P -value $<2\times 10^{-16}$ with or without adjustment) and in the larger CBC model (including total hematocrit, white blood cell count, platelets, neutrophils, lymphocytes, monocytes, eosinophils, and basophils), only hematocrit ($p\sim 0.05$) approached nominal significance. The inclusion of hematocrit moderately improved AIC

in logistic regression, but with limited performance increase in multivariable modeling AUC (93.6% vs 94.1%).

Severity analysis focused on hospital length of stay (median duration: 6 days, IQR 3-11, max 53 days), as well as across the spectrum of severity (34 discharged from ER, 84 hospitalized, 35 admitted to ICU, and 11 deaths). The best-fitting model for hospital duration had a cutpoint at 20 days, yielding an AUC of 79.6% with 14 individuals with longer stays versus 135 with shorter stays (or 0 days in hospital) (**Fig 4c**). Dichotomizing the best-fit severity measurements yields AUCs of 79.1%, 80.8%, and 84.4% for hospital admission vs discharge, floor hospital admission vs ICU, and survival vs death, respectively (**Fig 4d**).

DISCUSSION

Here we report DNA methylation profiling in conjunction with analysis using machine learning techniques to identify a SARS-CoV-2 specific epigenetic signature in peripheral blood from a large cohort of individuals tested using conventional RT-PCR technology. We also describe the development of a classification algorithm which has high sensitivity and specificity in predicting infection and in-hospital clinical deterioration, and which confidently rejects the probability of healthy individuals to be affected by SARS-CoV-2 infection. While any predictive signal invites concern of potential confounding, the signal we observe is not driven by confounding either from demographics or typical laboratory measurements (*e.g.* blood cell counts, BMI). Our findings suggest that measurement of methylation signatures that arise during and after SARS-CoV-2 infection may provide clinicians the ability to detect viral infection as well as predict patient clinical course after viral challenge. Unlike sequencing, RT-PCR, and antibody tests, the methylation array is able to predict the severity of SARS-CoV-2 infection and ultimately could provide clinicians with information on how to manage patients infected with SARS-CoV-2.

Our results support the hypothesis that the host epigenome, as measured in peripheral blood, is modified by infection from SARS-CoV-2 and can be used to identify novel biology and it is useful for clinical diagnosis, prognosis, and triage. Despite being a heterogeneous tissue, we relied on peripheral blood as the target tissue because it has proven a reliable source for generating epigenetic signatures and disease classifiers in other settings²⁹⁻³⁵. In this study, we observed many methylation changes that are, on average, more than 10% differentially methylated in the SARS-CoV-2+ group, including *IRF7* and *MX1* interferon-related genes. These are much larger effect sizes than typically observed in EWAS in peripheral blood³⁶ and similar to the clinical utility of epigenetics observed in cancer²². We did not observe confounding by cell proportions, measured by CBC from the electronic health record, providing strong support for the epigenetic signature of SARS-CoV-2. Although cell type heterogeneity can be a strong confounder in epigenetic studies³⁷⁻³⁹, we did not pursue adjustment for cell proportions beyond adjustment for cell type proportions using ReFACTo⁴⁰ because our primary objective is to develop a COVID-19 disease-specific diagnostic methylation platform, rather than interrogate the underlying pathology.

To validate the customized EPIC methylation platform as a reliable tool for the clinical diagnosis of COVID-19 disease, we performed an EWAS with SARS-CoV-2 infection status. We observed that the epigenetic signature of SARS-CoV-2 infection is enriched for pathways related to host viral response, and specifically for Type I Interferon signaling that is a hallmark of host response to this virus⁴⁴. Our findings of altered DNA methylation in interferon response genes are in concordance with published results of changes in expression of interferon response genes by SARS-CoV and MERS-CoV viruses through changes in histone modifications^{2,3}. One of the most significant probes (adjusted $P=1.77 \times 10^{-43}$, 16.9% hypomethylation) is located in the gene encoding *IRF7*; loss-of-function variants in 13 genes including *IRF7* were recently found to be associated

with life-threatening COVID-19 associated pneumonia⁴¹. Another interferon-induced gene, *OAS1*, was similarly significant (adjusted P-value 1.05×10^{-21} , 3.8% methylation change. In a recent GWAS on critical illness due to SARS-CoV-2, significant associations and replication were observed for variants in the oligoadenylate synthetase (OAS) gene cluster, which includes *OAS1*⁴², for which variants had previously been associated in candidate gene studies of SARS-CoV infection^{43,44}. Also, in a Mendelian randomization study it was recently shown that increased circulating OAS1 proteins were associated with reduced SARS-CoV-2 susceptibility and disease severity⁴⁵. Collectively, published genomics studies support several of the strongest associations observed in our study.

Previous work also demonstrated that viruses that cause severe disease (*e.g.*, MERS-CoV, H5N1) alter host response by changing methylation landscape of antigen-presenting genes in the HLA region⁴. While we did not observe genome-wide significant signals at classical HLA alleles, we observed six FDR $q < 0.05$ probes in the region, in *HLA-V*, *DOA*, *DQA1*, *DQA2*, and *DRA*, albeit with attenuated significance compared to top CpGs (minimum $q \sim 0.0109$), suggesting that the mechanism of host manipulation by SARS-CoV-2 may be different. However, these results should be interpreted with caution as interrogation of the HLA region is complex; *HLA-V* for example is a pseudogene⁴⁶.

As the signatures identified in this study appear to be reactive to the disease, aspects of the disease process are expected to impact these results. Namely, we anticipate these changes to be time-sensitive, as the infection will need to have spread enough to induce methylation changes. Similarly, our case-control variables were defined by RT-PCR, which can carry a high false negative rate depending on the stage of infection and timing of sample collection⁴⁷, and may have reduced the classification accuracy. However, we have follow-up EHR information for the patients in this cohort which minimizes the risk of misclassification bias. We do not expect this potential confounder to affect the measures of severity used in this study as these were determined directly from chart review, but we acknowledge that for the initial analysis, the numbers of cases may have limited the statistical power and prognostic ability of machine learning. With additional cases that account for inherent genetic variability within the population, methylation patterns will become more refined and the AUC of these machine learning models to predict disease severity is likely to increase. While ‘duration of hospital stay’ may not be as immediately actionable as predicting ICU admittance or ventilator use, the increased variability in the continuous outcome provides improved signal as observed both in our EWAS and our ML modeling. For this analysis, the 11 individuals who died were removed from downstream analyses, as their duration would be difficult to compare to those who survived. Although the emerging field of epigenetics has demonstrated actionable classification with much smaller sample sizes in contrast to traditional GWAS in other common disease domains⁴⁸, we recognize that additional cases, and in particular understanding the less-severe end of the spectrum (which are likely to be under-reported in data from health systems), will improve our understanding of outcomes across the spectrum of disease severity. We note that, even in our limited sample sizes, the AUCs for ICU and ventilator use still indicate there is signal that can be resolved through future collections. Another limitation of our work is the specificity of the epigenetic signature to SARS-CoV-2 over other respiratory infections. Initial targeted epigenetic analyses demonstrate a trend toward differential methylation, though these findings are limited by low numbers. Currently we are targeting the collection of biospecimens from patients with respiratory infections other than SARS-CoV-2 for these follow-up studies.

Researchers have previously compared the robustness of DNA methylation profiling versus RNA transcriptome profiling in developing classifiers for different disease states^{24,49-51}. One of the advantages of DNA methylation analysis compared to RNA analysis arises from the relative

stability of deoxyribonucleic acid over ribonucleic acid^{9,52}. The inherent instability of RNA, due to its 2'-OH group and the ubiquitous presence of ribonucleases, requires the use of plasticware, buffers, and processing reagents that are devoid of chemical and enzymatic species that stimulate RNA hydrolysis. Contamination even with a small amount of ribonuclease can degrade RNA samples to the degree where they cannot be analyzed.

The strong signature of viral-driven epigenetic changes may have the ability to detect SARS-CoV-2 infection in patients who never develop symptoms (asymptomatic) and in patients who are not yet symptomatic (pre-symptomatic)⁵³. The current testing strategy in the U.S. is primarily targeted to symptomatic patients despite estimates that asymptomatic patients represent 40-45% of infected individuals^{10,53}. Transmission during the incubation period has been reported, and the viral load of symptomatic and asymptomatic patients is similar⁵⁴⁻⁵⁷. The relationship between SARS-CoV-2 viral shedding and risk of transmission is unclear, and the percentage of transmission attributable to asymptomatic or pre-symptomatic infection SARS-CoV-2 infection is unknown⁵⁸. We believe that the epigenetics platform may efficiently identify asymptomatic and pre-symptomatic infections which may, if applied broadly, aid in limiting the spread of SARS-CoV-2

Due to the widespread occurrence of SARS-CoV-2 and progression to COVID-19 disease, there is the need for scalable testing technologies that can be deployed on the national level for surveillance, screening, and prognosis for those infected. The purpose of this study was to identify high-confidence host methylation biomarkers that are able to indicate SARS-CoV-2 infection and predict clinical course of the viral disease in a given patient. This study is a first step towards selecting biomarkers for inclusion on a high-throughput methylation beadchip array specifically for the clinical diagnosis of COVID-19 disease that is also cost-effective given the added value of predicting subsequent clinical outcomes. To that end, we focused on sparse predictive models. Notably, these models are not significantly confounded by demographics or blood cell count information, denoting their specificity to the current infection of the patient, and reducing concern of overfitting to one patient sub-population. These biomarkers can also be used in risk stratification of SARS-CoV-2 infected patients, an unmet need given that none of the existing testing modalities (nucleic acid amplification tests, antigen tests, serology/antibody tests) can achieve this level of specificity. By identifying DNA methylation patterns associated with critical illness, we contend that a methylation test will provide patient-specific treatment targets before critical illness ensues. Pre-emptive dexamethasone^{11,59}, anticoagulation¹², or new pharmacologic targets may prevent mortality, guided by these epigenetics patterns. Although our findings must be complemented with further clinical assessment, our model has shown its capacity to leverage methylation quantification as an innovative strategy to generate epigenetic signatures that assess host response to SARS-CoV-2, which is scalable and may have the ability to confirm positive tests in asymptomatic patients and entire communities, and may ultimately differentially diagnose other viruses causing similar symptoms all within in a comprehensive high-throughput manner.

ACKNOWLEDGMENTS

We thank all study subjects at the University of Colorado Anschutz Medical Campus (CU-AMC) for their participation, along with Drs. Thomas Flaig, Adrie Van Bokhoven, and Alison Lakin for their leadership and technical support through the University of Colorado COVID-19 Biorepository, Dr. Richard Zane for his support on behalf of UCHHealth, and Michelle Edelmann, Andrew Hadd, and Olivia Tintea for technical assistance. From the Illumina, Inc. team, we thank Krishna Bose, Anita Pottekat, Steven Gruber, John Picuri, Jay Kaufman, and Jason Johnson. This work was supported in part by the Health Data Compass Data Warehouse project (healthdatacompass.org), the Biobank at the Colorado Center for Personalized Medicine, and by grant funding from the University of Colorado Anschutz Medical Campus, Chancellor Discovery Innovation Fund.

FIGURES

Figure 1. Flowchart of the study samples.

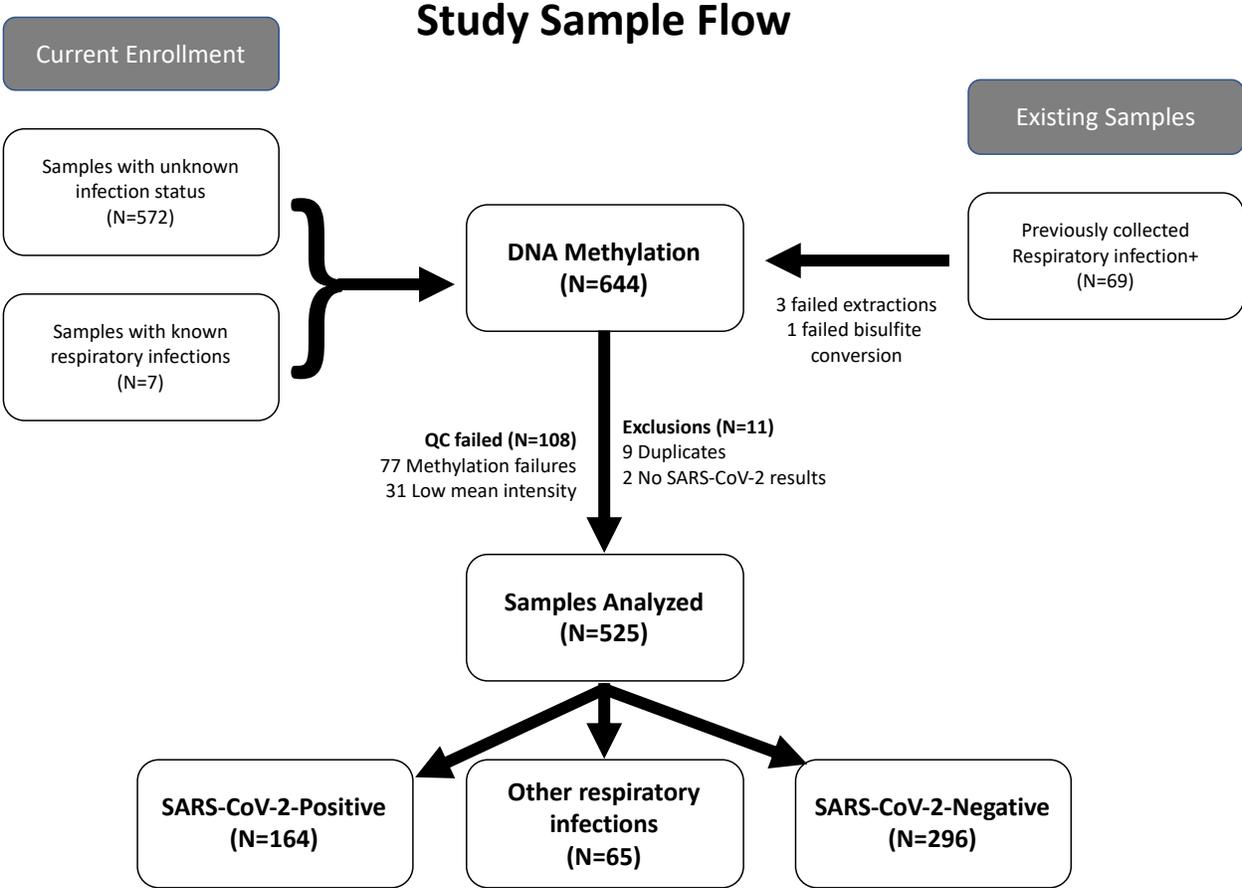


Figure 2. a, Miami plot (top panel) of hypermethylated (top) and hypomethylated (bottom) probes in SARS-Cov-2+ compared to control samples. Significance lines represent FDR-adjusted P -value <0.05 threshold. **b**, Volcano plot of significant (red; FDR-adjusted P -value <0.05) CpG sites (blue CpG sites have FDR-adjusted P -value >0.05). Change in % methylation on the x axis represents the difference in average beta value at a site between cases and controls. Probes for intergenic CpG sites do not have gene annotations.

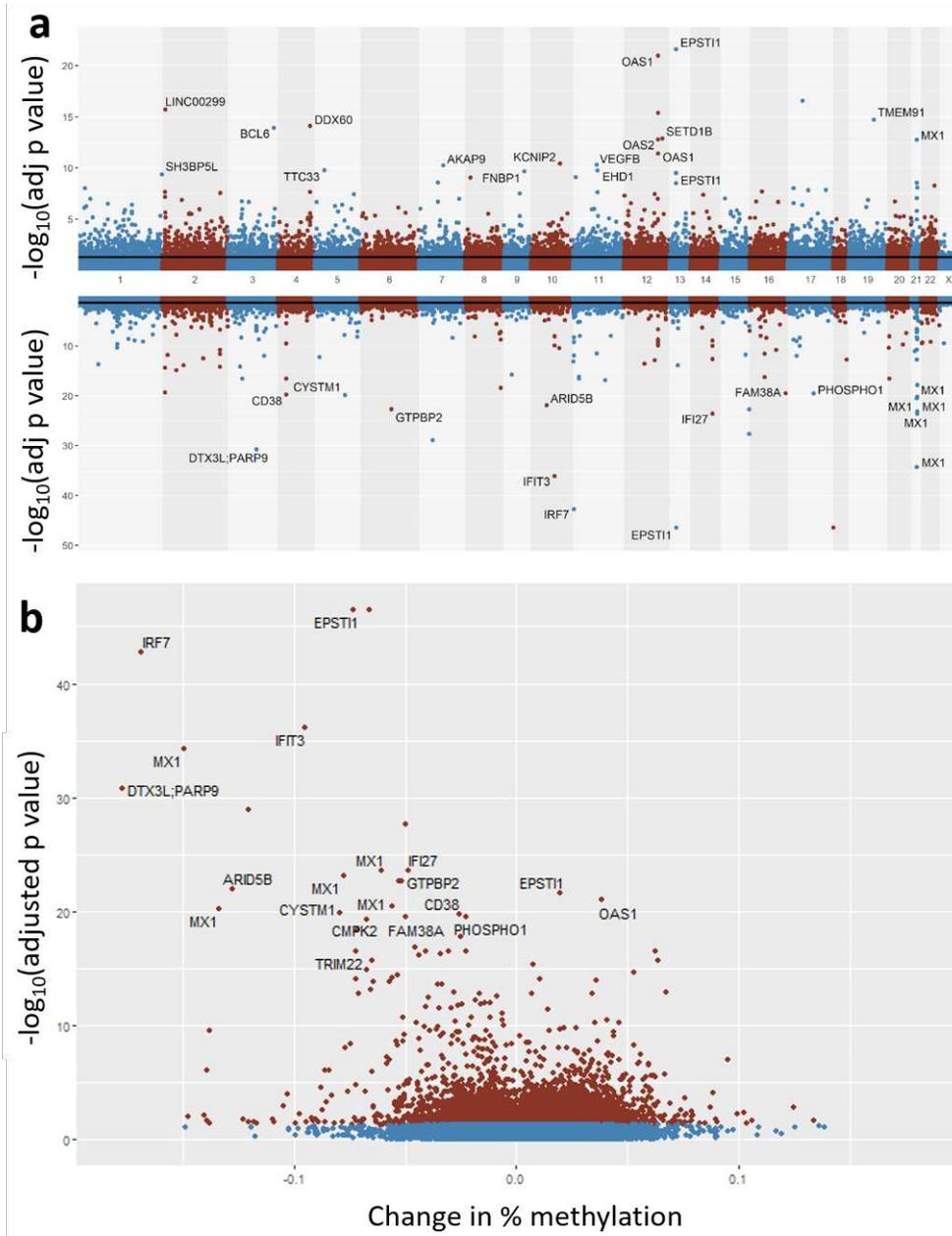


Figure 3. Venn diagram of overlaps between COVID-Control EWAS (13,033 significant probes), COVID-viral infection EWAS (516 significant probes), and respiratory infection-Control EWAS (1,501 significant probes).

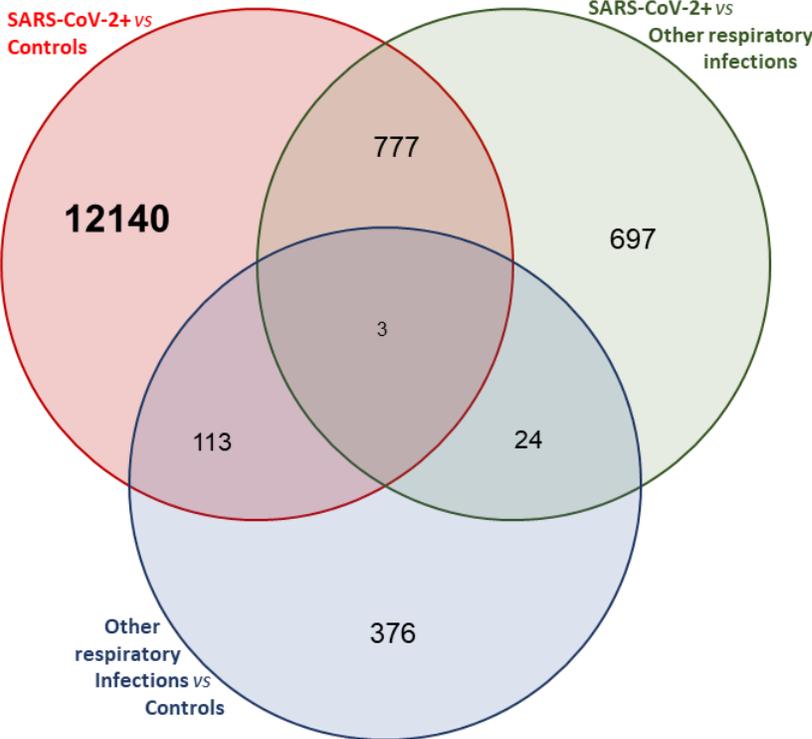
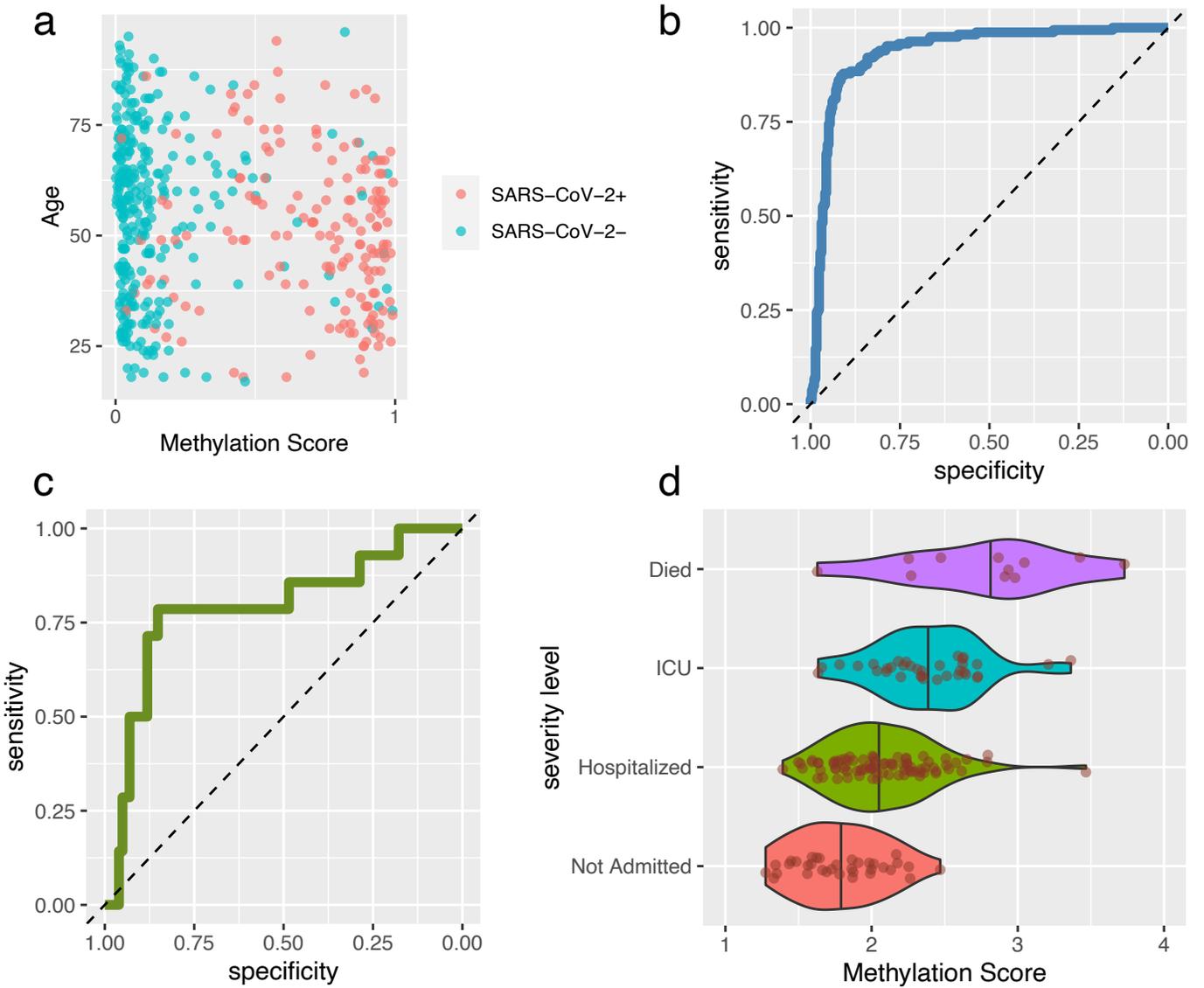


Figure 4. **a**, Out-of-sample case-control methylation score for all 460 individuals compared to case-control status, plotted by biological age. **b**, Receiver-Operating Characteristic (ROC) Curve for data in **a**. **c**, ROC curve of cross-validated prediction of long hospital duration **d**, Violin and jittered scatter plots of severity methylation scores for each outcome in cases.



MATERIAL AND METHODS

Source of Data

This protocol was reviewed and approved by the Colorado Multiple Institutional Review Board (COMIRB) and the research adheres to the ethical principles of research outlined in the US Federal Policy for the Protection of Human Subjects. SARS-CoV-2+ were defined as those patients who tested positive for SARS-CoV-2 infection via a RT-PCR assay in the Biobank at the Colorado Center for Personalized Medicine (Thermo Fisher Scientific, Waltham, MA) or in the UHealth University of Colorado Hospital Clinical Laboratory (Roche Diagnostics, Indianapolis, IN) of a nasopharyngeal swab collected in viral transport media; controls were defined as those who tested negative. Peripheral blood DNA samples were collected in EDTA tubes from patients seen at the UHealth University of Colorado Hospital and tested for SARS-CoV-2 epigenetic signatures starting on March 1, 2020. Blood specimens were collected from patients consented to the University of Colorado COVID-19 Biorepository (<https://research.cuanschutz.edu/university-research/covid-19-clinical-research/covid-19-biobank-specimen-repository>) or the University of Colorado Emergency Medicine Specimen Bank (EMSB)⁶⁰. Control subjects included patients from each study who tested negative for SARS-CoV-2 infection during the index visit. Through the University of Colorado COVID-19 Biorepository and the EMSB, patients tested were consented for blood collection, and data abstraction from their electronic health record (EHR). Data obtained from EHR abstraction included demographics, past medical history, laboratory testing, treatments, vital signs, hospital disposition, and clinical outcomes. In addition, previously-collected samples from patients with acute upper respiratory viral infections (SARS-CoV-2 negative / pan-negative for upper respiratory viral infections / positive for non- SARS-CoV-2 upper respiratory viral infections) between February 5, 2018 and January 1, 2020 were obtained through the EMSB as SARS-CoV-2 negative controls. Additional biospecimens included discarded clinical samples from patients not approached for biorepository enrollment through the UHealth University of Colorado Hospital Clinical Laboratory. Discarded samples were linked to a limited EHR dataset through the Colorado Center for Personalized Medicine's health data warehouse, Health Data Compass, and then deidentified. The limited dataset included age, gender, race, ethnicity, viral test status (SARS-CoV-2 and other upper respiratory viruses), and clinical outcomes. All samples were frozen at -20 °C after collection prior to processing for methylation analyses.

Customization of the Infinium MethylationEPIC Array

Following a literature review of known epigenetic associations with respiratory viral infections from recent coronavirus outbreaks, we selected additional content to enrich Illumina's Infinium MethylationEPIC Array⁶¹. We specifically enriched for known HLA alleles accounting for known genomic variation⁶² as well as multiple alternative haplotypes and unpublished reference sequences spanning the MHC genomic region, the Natural Killer Cell Immunoreceptor (KIR) and other immunogenetic loci (*e.g.* cytokines, interferon response genes), to enhance the sensitivity of immune response detection. The custom panel targeted 262 genes with 7,831 additional probes. While the majority of the additional probes targeted unique sequences within the genome, a number of probes were intentionally designed to target genomic sequences with a limited degree of repetitiveness. The list of genes and the Illumina IDs for the probes that target these genes are given in **Supplementary Table 4**.

Methylation Array and Quality Assessment

DNA Extraction: Biospecimens were accessioned and tracked via the Colorado Anschutz Research Genetics Organization (CARGO) laboratory information management system (LIMS). Genomic DNA was extracted from SARS-Cov-2+ peripheral blood on the bead-based, automated extraction Maxwell(R) RSC System (Promega) in a biological safety cabinet in compliance with CDC safety guidelines and procedures for handling SARS-CoV-2 biospecimens (biospecimens from SARS-Cov-2+ cases), and from controls on the Autogen FlexSTAR+ using Autogen's FlexiGene Blood extraction kit (Holliston, MA). All DNA samples were quantified using both absorbance (NanoDrop 2000; Thermo Fisher Scientific, Waltham, MA) and fluorescence-based methods (Qubit; Thermo Fisher Scientific, Waltham, MA) using standard dyes selective for dsDNA, minimizing the effects of contaminants that affect the quantitation. DNA quality was assessed using an Agilent TapeStation (Agilent, Santa Clara, CA.). Samples were then uploaded to CARGO's LIMS, barcoded and labeled.

Bisulfite Conversion and Amplification: Purified DNA samples were processed using Zymo EZ-96 DNA Methylation bisulfite conversion kits (Zymo, Irvine, CA) as described previously⁶³. The product of this process contains cytosine converted to uracil if it was previously unmethylated. The bisulfite treated DNA was subjected to whole genome amplification (WGA) via random hexamer priming and Phi29 DNA polymerase, and the amplification products were then enzymatically fragmented, purified from dNTPs, primers, and enzymes, and applied to the Illumina chip as described elsewhere⁶⁴.

Hybridization and Single-base Extension: The bisulfite converted amplified DNA products were denatured into single strands and hybridized to the customized Infinium 850K BeadChip (EPIC+; Illumina Inc., San Diego, CA) via allele-specific annealing to either the methylation specific probe or the non-methylation probe. Hybridization to the chip was followed by single base extension with labeled di-deoxynucleotides according to Illumina's Infinium protocol at the CARGO laboratory⁶³.

Fluorescence Staining and Scanning of Chip: The hybridized BeadChips were stained, washed, and scanned to show the intensities of the un-methylated and methylated bead types using Illumina's iScan System.

Data Processing and Quality Control

IDAT files were processed, filtered, and normalized using the SeSAME R package⁶⁵. Type I probe channel was empirically determined from signal intensities. Probe detection p-values (representing the ability to differentiate true signal from background fluorescence) were calculated for each color channel using pOOBAH, which leverages the fluorescence of out-of-band (OOB) probes. Normalization was performed using noob, which uses OOB probes to perform a normal-exponential deconvolution of fluorescent intensities⁶⁶. Finally, a common dye bias that results in greater intensities in the Red color channel was corrected to ensure that the distribution of intensities in the two color channels were equal. Probes with detection p-values > 0.05 were removed, as well as probes overlapping SNPs with global MAF >1% in dbSNP, probes with poor mapping, and probes containing non-unique sequence according to Zhou *et al*⁶⁷. Beta values were logit-transformed into M values for modeling. Probes with >25% missingness were removed. Remaining missing values were then imputed with mean probe M-value.

Selection of Discovery/Training and Testing Cohorts and Controls

Case-control analyses were performed using the entire genotyped dataset passing epigenetics QC, with SARS-CoV-2 infection status determined as described above (see **Fig. 1** for a summary of the workflow). Analyses were repeated including and excluding controls with other upper respiratory infections validated by clinical respiratory panels. Measurements of disease severity and progression (*e.g.* hospitalization, ICU admittance, ventilator use) were extracted from chart review within the UCHealth EHR.

Control for batch effect and robustness of the identified epigenetic signatures

To minimize possible batch effects and other sources of variability, samples were split into SARS-CoV-2+ and SARS-CoV-2 negative control sets, randomized within sets to account for unavailable phenotypes, and then distributed across chips. To reduce batch and plating effects a minimum of two SARS-CoV-2+ and two SARS-CoV-2 negative control samples were run on each chip (12 chips per plate, 8 samples each) and positive/negative status was randomized across the chip.

Epigenome-wide association study (EWAS) with COVID-19 disease status

Preprocessing was performed using the GLINT⁶⁸ package for association testing and estimating components to adjust for population structure (EPISTRUCTURE⁶⁹) and we used ReFACTOR⁴⁰ to account for cell type proportions. We chose ReFACTOR to account for cell proportion information in a data-driven fashion. The linear mixed-effects model (LMM) in GLINT was fit to each probe, testing for differences based on COVID-19 disease status while correcting for age, sex, chip position, 6 ReFACTOR components, 1 EPISTRUCTURE component, and a variance component representing individual covariance⁷⁰. Enrichment of top hits in common databases was performed using enrichR⁷¹. Probes were sorted by adjusted p-value and the top 800 genes to which differentially methylated probes map were used as input to perform overrepresentation enrichment analysis within Gene Ontology (GO) categories, Kyoto Encyclopedia of Genes and Genomes pathways (KEGG), BioPlanet, and WikiPathways⁷²⁻⁷⁵.

Clinical Outcome Stratification

Clinical data were abstracted via detailed chart review for all EMSB patients. COVID-19 disease severity was determined by an ordered severity score of 1) discharged from Emergency Department; 2) admitted to inpatient care; 3) progressed to ICU; and 4) death. We also determined a hospital duration variable, where individuals without a measured hospital stay (*i.e.* discharged from the Emergency Department) were assigned 0 and individuals who died were removed from the cohort for length of stay analysis to minimize bias associated with timing of decisions to withdraw care.

Construction and Validation of a Prediction Model

Predictive modeling was performed using the Lasso⁷⁶ and Elastic Net⁷⁷ algorithms for sparse penalized regression modeling available in the *glmnet* software package²⁸. For each prediction model, all probes passing QC were included, allowing the algorithm to pick sites appropriate to the level of penalization of the hyperparameters across the entire least angle regression path. For each trait of interest, a separate model was created and best-fitting parameters were chosen after 10-fold cross-validation either by maximizing area under the receiver-operator characteristic curve (AUC for dichotomous traits) or minimizing mean-squared error (MSE for quantitative traits). Each was fit across a grid of parameters representing various strengths of penalization and combination of L1 and L2 penalties under the weighted elastic net model. Both the days of

hospitalization and case severity were modeled as continuous outcomes. To assess performance for quantitative traits in a manner comparable to dichotomous traits, we swept across potential cutpoints to estimate AUCs for this newly-derived dichotomous variable. While case-control status was the primary phenotype of interest, measures of severity were assessed in SARS-Cov-2+ cases only.

To estimate stability of estimation in parameters, we performed 100 iterations of model training and testing. Within each iteration for case-control and severity outcomes we employed 10-fold cross validation to derive the model and a held-out set of 30% removed from train/test to gauge out-of-sample performance of the best-fitting model. Our train/test and validation splits were created within each stratum to preserve representation across all outcomes and reflect the distribution across the total dataset. For hospitalization duration, the train/test/validation models had instability in convergence and so we reverted to a train/test model using the 10-fold cross-validation within the default `cv.glmnet()` function. We assessed overall performance for the dichotomous COVID+/COVID- case-control status using out-of-sample AUC, the F1 score (a measure of the relationship between precision and recall), the distribution of best-fit λ penalty via cross-validation, and the number of probes chosen in the final model. For the quantitative outcomes we assessed overall performance using out-of-sample R^2 , the slope of the model, and λ number of probes. Finally, these were stratified each across the elastic net weights (α) from 0.01-1, representing the proportion of ridge (L2) vs Lasso (L1) penalty to choose a final model. All models included nonzero λ to encourage sparsity (a L2-only model would include prediction from the entire array). Final models described in results were chosen based on best-performing (maximum R^2 or AUC) versus median values for each chosen set of hyperparameters. The final, out-of-sample best-fit prediction for each outcome was considered the “methylation score” used in downstream modeling, characterization of association, and determination of potential confounding with demographic and blood cell proportion characteristics.

Supplementary Information

Supplementary Table 1. Significant probes in COVID-control EWAS (N=13,033).

Supplementary Table 2. Significant probes in COVID-upper respiratory infection EWAS (N=516).

Supplementary Table 3. Significant probes in upper respiratory infection-control EWAS (N=1,501).

Supplementary Table 4. Added probes, targeted genes, and Illumina IDs.

Extended Data

Extended Data Table 1: Demographic summary of all samples analyzed.

Trait	Total (N=525)	COVID -ve (N=296)	COVID +ve (N=164)	Other Viral Infections (N=65)
Age (years)				
Mean (SD)	54.1 (18.2)	56.1 (18.8)	50.5 (17.0)	54.1 (17.0)
Gender				
Female	251 (47.8%)	145 (49.0%)	71 (43.3%)	35 (53.8%)
Male	274 (52.2%)	151 (51.0%)	93 (56.7%)	30 (46.2%)
Ancestry				
American Indian or Alaskan Native				
Not Hispanic or Latino	2 (0.4%)	1 (0.3%)	1 (0.6%)	0 (0%)
Asian				
Not Hispanic or Latino	18 (3.4%)	6 (2.0%)	12 (7.3%)	0 (0%)
Black or AA				
Not Hispanic or Latino	92 (17.5%)	50 (16.9%)	25 (15.2%)	17 (26.2%)
Native Hawaiian or Pacific Islander				
Hispanic or Latino	1 (0.2%)	0 (0%)	1 (0.6%)	0 (0%)
Not Hispanic or Latino	2 (0.4%)	1 (0.3%)	1 (0.6%)	0 (0%)
White				
Hispanic or Latino	28 (5.3%)	13 (4.4%)	13 (7.9%)	2 (3.1%)
Not Hispanic or Latino	239 (45.5%)	171 (57.8%)	32 (19.5%)	36 (55.4%)
Missing	1 (0.2%)	0 (0%)	1 (0.6%)	0 (0%)
Other				
Hispanic or Latino	116 (22.1%)	39 (13.2%)	68 (41.5%)	9 (13.8%)
Not Hispanic or Latino	9 (1.7%)	7 (2.4%)	2 (1.2%)	0 (0%)
Missing				
Hispanic or Latino	12 (2.3%)	5 (1.7%)	6 (3.7%)	1 (1.5%)
Not Hispanic or Latino	2 (0.4%)	1 (0.3%)	1 (0.6%)	0 (0%)
Missing	3 (0.6%)	2 (0.7%)	1 (0.6%)	0 (0%)

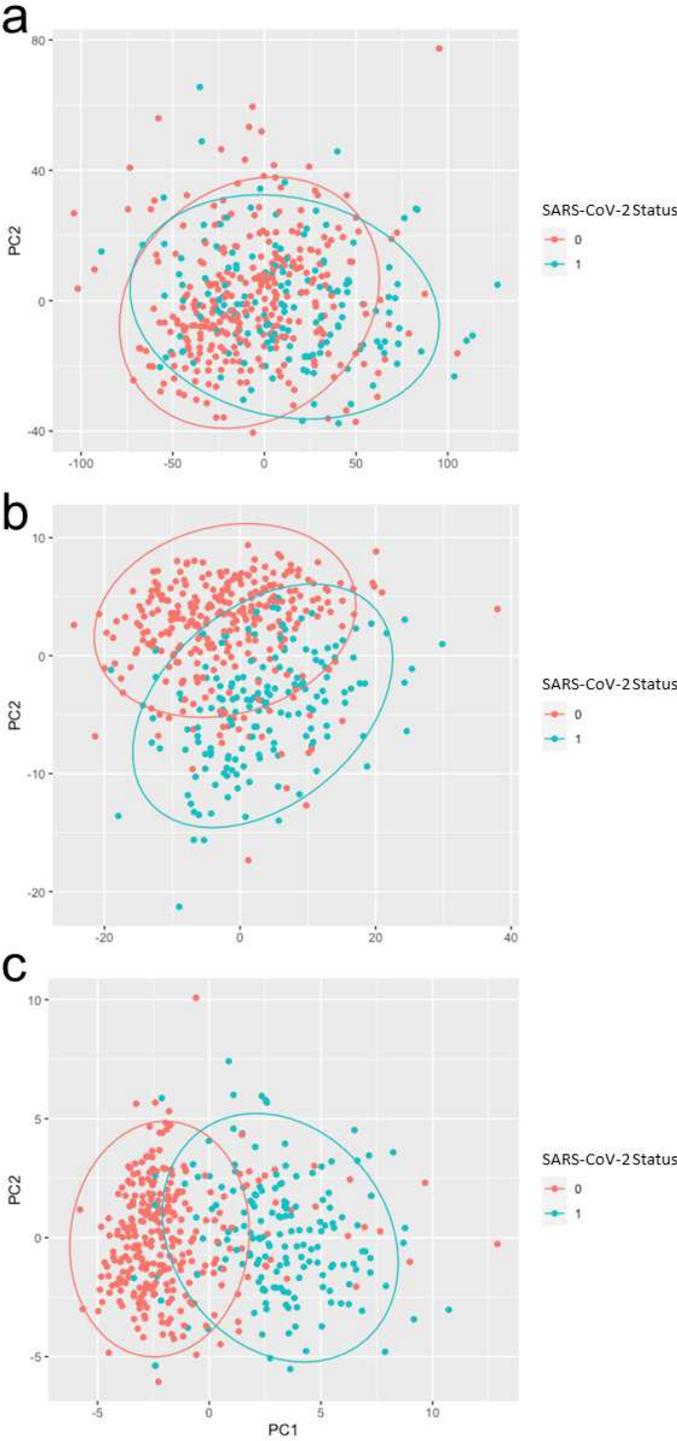
Extended Data Table 2: Top classification parameters for prediction of disease classes and disease severity for elastic net models, based on train/test/validation (SARS-CoV-2 infection and COVID severity), and train/test (hospitalization duration). Parameters chosen for best-performing median prediction accuracy (AUC for case-control, R2 for duration and severity) across the 100 iterations of training and 10-fold cross-validation. Within best α , λ and number of nonzero probes chosen across the range of best-fitting parameters.

Outcome	Best α	Median λ	Median # Probes
SARS-CoV2 status	0.4	0.095	142
Hospital duration	0.1	20.8	158
Severity	0.01	2.03	2803

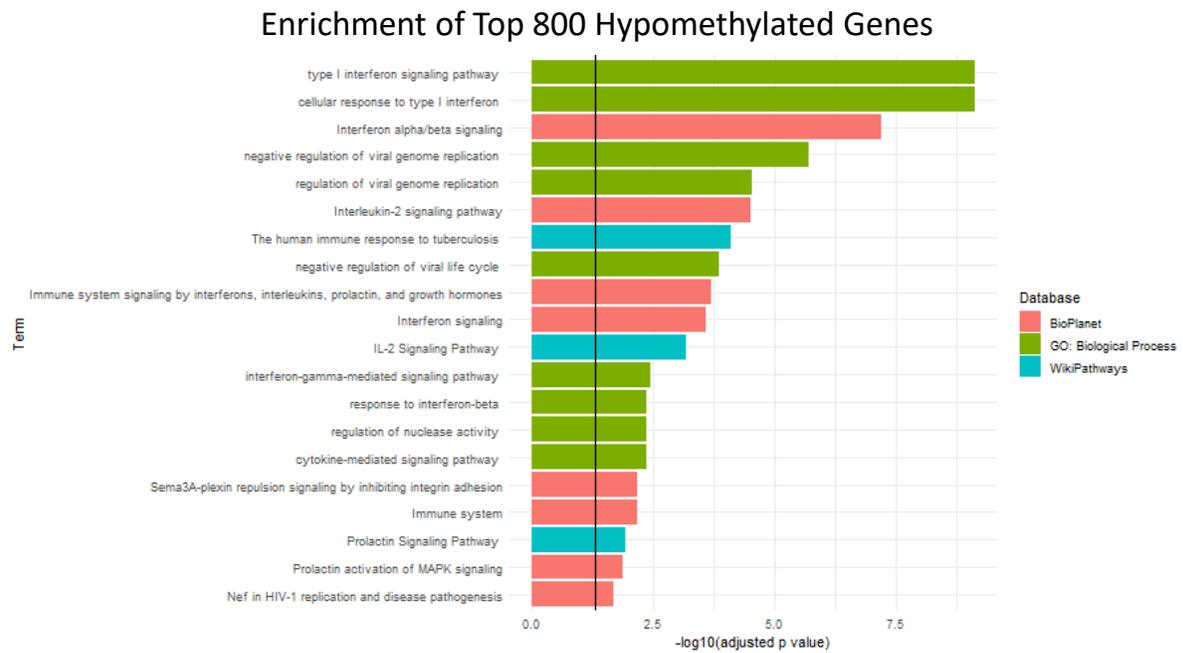
Extended Data Table 3: Frequency of upper respiratory infections in 65 SARS-CoV-2 negative controls with upper respiratory infections

Upper Respiratory Infection	Overall (N=65)
Adenovirus	1 (1.5%)
B Pertussis	1 (1.5%)
Coronavirus HKU1	3 (4.6%)
Coronavirus NL63	2 (3.1%)
Coronavirus OC43	1 (1.5%)
Coronavirus OC43 & Influenza A H1	1 (1.5%)
Human Metapneumovirus	8 (12.3%)
Influenza A	6 (9.2%)
Influenza A H1 2009	2 (3.1%)
Influenza A H3	3 (4.6%)
Influenza B	7 (10.8%)
Parainfluenza 1	2 (3.1%)
Parainfluenza 4	1 (1.5%)
Respiratory Syncytial Virus	6 (9.2%)
Rhinovirus/Enterovirus	21 (32.3%)

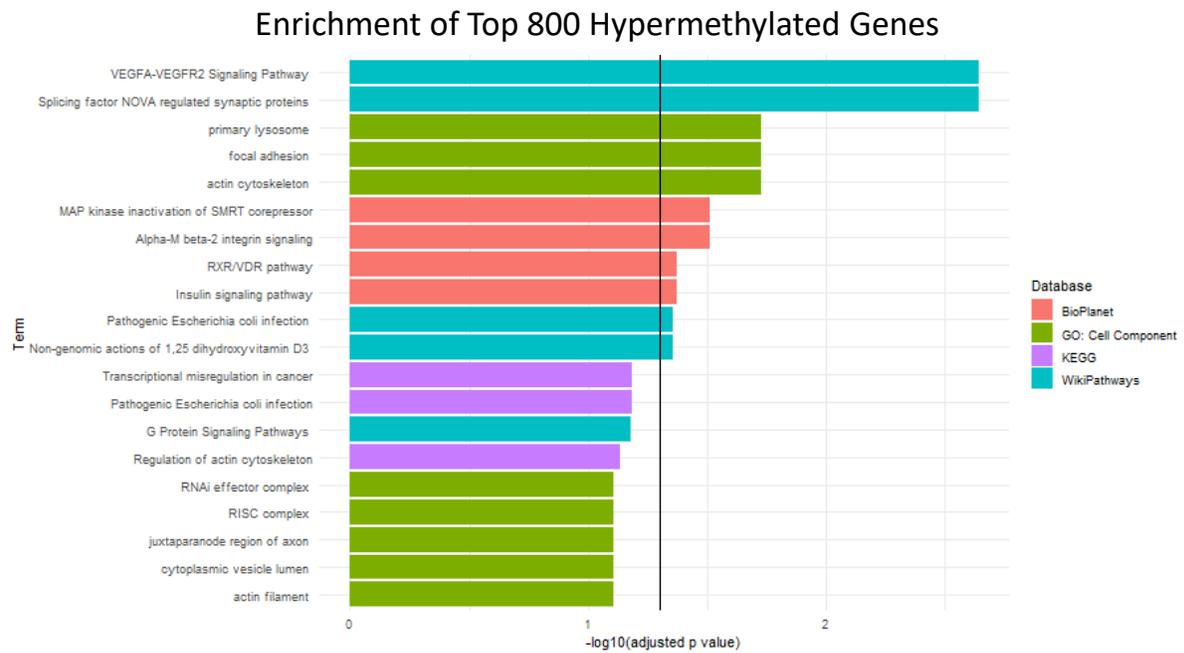
Extended Data Figure 1. Principal component analysis of **a.** 13,033 significant CpGs in COVID-19 disease status EWAS **b.** 1,000 CpGs ranked by p value **c.** 100 CpGs ranked by p value. Ellipses represent 95% confidence intervals assuming a t distribution.



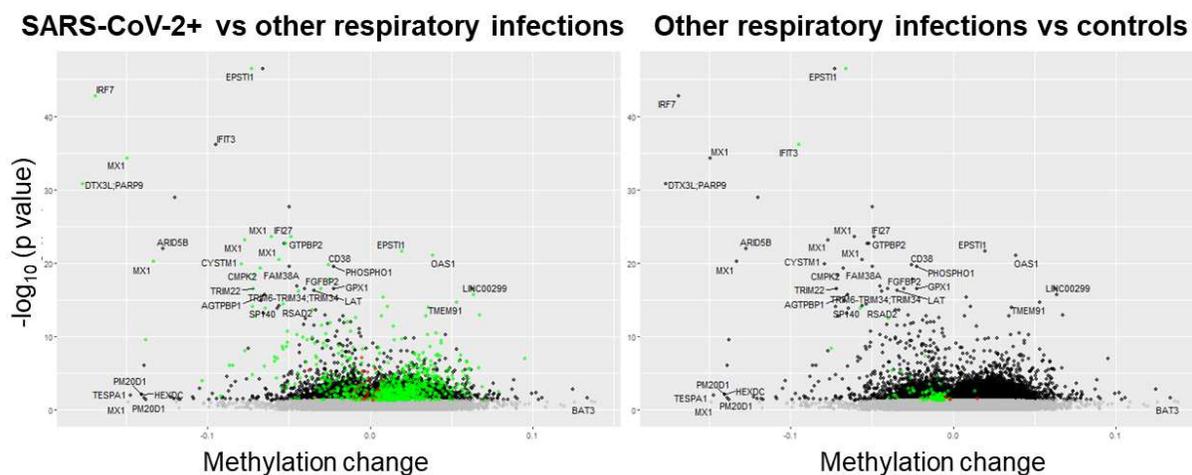
Extended Data Figure 2a. Enriched pathways in top hypomethylated hits from SARS-CoV-2 infection status EWAS.



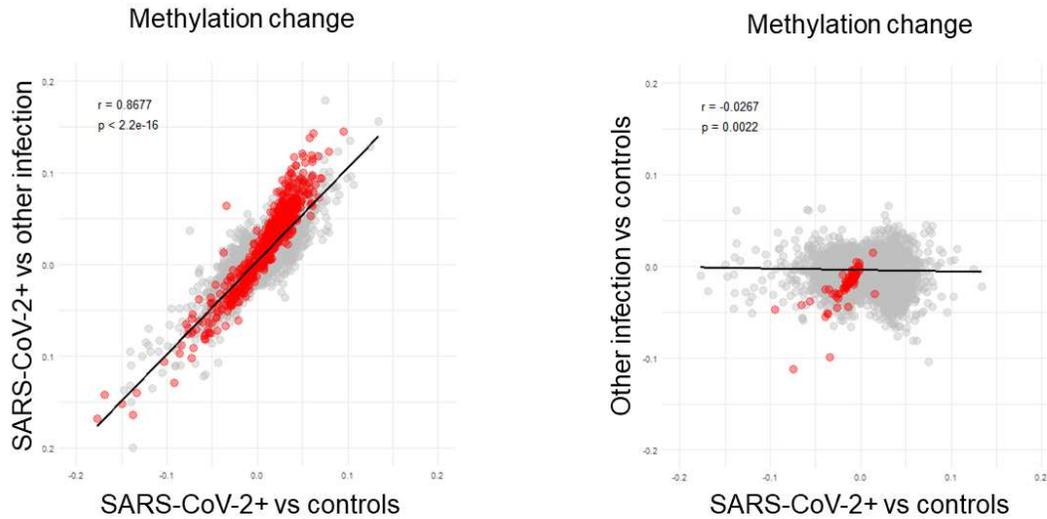
Extended Data Figure 2b. Enriched pathways in top hypermethylated hits from SARS-CoV-2 infection status EWAS.



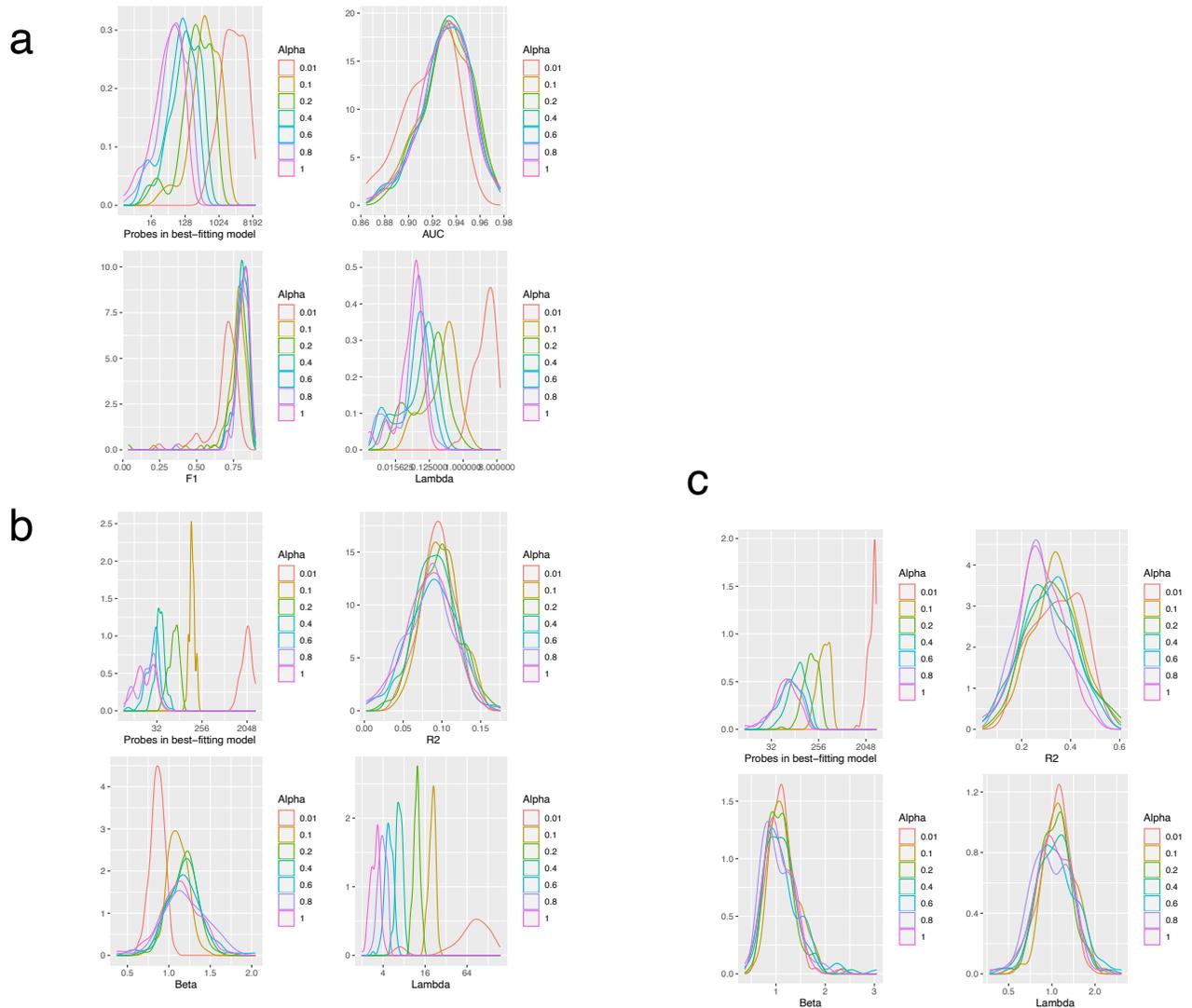
Extended Data Figure 3a. Specificity of the COVID-19 disease EWAS signature. Both volcano plots depict the 13,033 CpGs significant in the EWAS for SARS-Cov-2+ cases compared to controls; the colors are used to represent overlap in those CpGs with EWAS for SARS-Cov-2+ cases compared to other respiratory infections (left panel) and other respiratory viruses compared to controls (right panel). Green dots represent CpGs concordant in directionality between the two EWAS and red dots represent CpGs discordant in directionality between the two EWAS. The black dots are CpGs significant in the SARS-Cov-2+ vs control EWAS but not in the other EWAS.



Extended Data Figure 3b. Specificity of the COVID-19 disease EWAS signature. Correlation plots for 13,033 CpGs significant in SARS-Cov-2+ vs control EWAS. Correlations with SARS-Cov-2+ compared to other upper respiratory infections (left panel) and with other upper respiratory infections compared to controls (right panel). Red dots represent CpGs significant in both SARS-Cov-2+ vs control and the other EWAS while gray dots represent CpGs only significant in the SARS-Cov-2+ vs control EWAS.



Extended Data Fig 4. Density plots across 100 iterations for each hyperparameter set in glmnet for case-control, hospital duration, and severity. All y-axes reflect density values. For **Ex. Data 4a** as a dichotomous outcome: 1) Number of probes selected in best-fitting model, 2) AUC from out-of-sample model fits across each iteration via Receiver-Operating Characteristic, 3) F1: a statistical relationship between precision and recall, and 4) λ , the best-fitting L1 penalty determined across the least-angle path. For **Ex. Data 4b** and **Ex. Data 4c** as continuous outcomes, we include distributions for number of probes, out-of-sample R^2 predictions, the β value between the predicted score and the true outcome, and the best-fit λ . All plots are stratified by Alpha, the weighted combination of L1 and L2 penalization, where L1 is Lasso regression and L2 is ridge regression. All models smoothed using the default density parameters ggplot2.



References

- 1 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565-574, doi:10.1016/S0140-6736(20)30251-8 (2020).
- 2 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497-506, doi:10.1016/S0140-6736(20)30183-5 (2020).
- 3 Roser, M., Ritchie, H. & Ortiz-Ospina, E. Coronavirus Disease (COVID-19) – Statistics and Research. Our World in Data. (2020). < <https://ourworldindata.org/coronavirus>>.
- 4 Administration, U. F. a. D. *Coronavirus Disease 2019 (COVID-19) Emergency Use Authorizations for Medical Devices*, <<https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations>> (2020).
- 5 Woloshin, S., Patel, N. & Kesselheim, A. S. False Negative Tests for SARS-CoV-2 Infection - Challenges and Implications. *N Engl J Med* **383**, e38, doi:10.1056/NEJMp2015897 (2020).
- 6 Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med* **173**, 262-267, doi:10.7326/M20-1495 (2020).
- 7 Fox-Lewis, S., Muttaiyah, S., Rahnama, F., McAuliffe, G. & Roberts, S. An understanding of discordant SARS-CoV-2 test results: an examination of the data from a central Auckland laboratory. *N Z Med J* **133**, 81-88 (2020).
- 8 Aslam, A. *et al.* SARS CoV-2 Surveillance and Exposure in the Perioperative Setting with Universal testing and Personal Protective Equipment (PPE) Policies. *Clin Infect Dis*, doi:10.1093/cid/ciaa1607 (2020).
- 9 Lan, L. *et al.* Positive RT-PCR Test Results in Patients Recovered From COVID-19. *JAMA : the journal of the American Medical Association* **323**, 1502-1503, doi:10.1001/jama.2020.2783 (2020).
- 10 Vandenberg, O., Martiny, D., Rochas, O., van Belkum, A. & Kozlakidis, Z. Considerations for diagnostic COVID-19 tests. *Nat Rev Microbiol*, doi:10.1038/s41579-020-00461-z (2020).
- 11 Garibaldi, B. T. *et al.* Patient Trajectories Among Persons Hospitalized for COVID-19 : A Cohort Study. *Ann Intern Med*, doi:10.7326/M20-3905 (2020).
- 12 Bilinski, A. & Emanuel, E. J. COVID-19 and Excess All-Cause Mortality in the US and 18 Comparison Countries. *JAMA*, doi:10.1001/jama.2020.20717 (2020).
- 13 Menachery, V. D. *et al.* Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *mBio* **5**, e01174-01114, doi:10.1128/mBio.01174-14 (2014).
- 14 Marazzi, I. *et al.* Suppression of the antiviral response by an influenza histone mimic. *Nature* **483**, 428-433, doi:10.1038/nature10892 (2012).
- 15 Menachery, V. D. *et al.* MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc Natl Acad Sci U S A* **115**, E1012-E1021, doi:10.1073/pnas.1706928115 (2018).
- 16 Aref-Eshghi, E. *et al.* Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet* **106**, 356-370, doi:10.1016/j.ajhg.2020.01.019 (2020).

- 17 Bollepalli, S., Korhonen, T., Kaprio, J., Anders, S. & Ollikainen, M. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. *Epigenomics* **11**, 1469-1486, doi:10.2217/epi-2019-0206 (2019).
- 18 Agha, G. *et al.* Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. *Circulation* **140**, 645-657, doi:10.1161/CIRCULATIONAHA.118.039357 (2019).
- 19 Jurmeister, P. *et al.* Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aaw8513 (2019).
- 20 Benhamida, J. K. *et al.* Reliable Clinical MLH1 Promoter Hypermethylation Assessment Using a High-Throughput Genome-Wide Methylation Array Platform. *J Mol Diagn* **22**, 368-375, doi:10.1016/j.jmoldx.2019.11.005 (2020).
- 21 Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474, doi:10.1038/nature26000 (2018).
- 22 Karimi, S. *et al.* The central nervous system tumor methylation classifier changes neuro-oncology practice for challenging brain tumor diagnoses and directly impacts patient care. *Clin Epigenetics* **11**, 185, doi:10.1186/s13148-019-0766-2 (2019).
- 23 Korshunov, A. *et al.* DNA-methylation profiling discloses significant advantages over NanoString method for molecular classification of medulloblastoma. *Acta Neuropathol* **134**, 965-967, doi:10.1007/s00401-017-1776-9 (2017).
- 24 Saben, J. L. *et al.* The Emergency Medicine Specimen Bank: An Innovative Approach To Biobanking In Acute Care. *Acad Emerg Med* **26**, 639-647, doi:10.1111/acem.13620 (2019).
- 25 Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 208, doi:10.1186/s13059-016-1066-1 (2016).
- 26 Robinson, J. *et al.* The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* **43**, D423-431, doi:10.1093/nar/gku1161 (2015).
- 27 Boorgula, M. P. *et al.* Replicated methylation changes associated with eczema herpeticum and allergic response. *Clin Epigenetics* **11**, 122, doi:10.1186/s13148-019-0714-1 (2019).
- 28 Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics* **37**, 549-554, doi:10.1038/ng1547 (2005).
- 29 Zhou, W., Triche, T. J., Jr., Laird, P. W. & Shen, H. SeSAmE: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic acids research* **46**, e123, doi:10.1093/nar/gky691 (2018).
- 30 Triche, T. J., Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90, doi:10.1093/nar/gkt090 (2013).
- 31 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22, doi:10.1093/nar/gkw967 (2017).
- 32 Rahmani, E. *et al.* GLINT: a user-friendly toolset for the analysis of high-throughput DNA-methylation array data. *Bioinformatics* **33**, 1870-1872, doi:10.1093/bioinformatics/btx059 (2017).
- 33 Rahmani, E. *et al.* Genome-wide methylation data mirror ancestry information. *Epigenetics & chromatin* **10**, 1, doi:10.1186/s13072-016-0108-y (2017).

- 34 Rahmani, E. *et al.* Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nature methods* **13**, 443-445, doi:10.1038/nmeth.3809 (2016).
- 35 Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**, 833-835, doi:10.1038/nmeth.1681 (2011).
- 36 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
- 37 Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-261, doi:10.1093/nar/gkh036 (2004).
- 38 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092 (2017).
- 39 Huang, R. *et al.* The NCATS BioPlanet - An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics. *Front Pharmacol* **10**, 445, doi:10.3389/fphar.2019.00445 (2019).
- 40 Slenter, D. N. *et al.* WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* **46**, D661-D667, doi:10.1093/nar/gkx1064 (2018).
- 41 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
- 42 van Iterson, M., van Zwet, E. W., Consortium, B. & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome biology* **18**, 19, doi:10.1186/s13059-016-1131-9 (2017).
- 43 Iwata, H. *et al.* PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. *Nat Commun* **7**, 12849, doi:10.1038/ncomms12849 (2016).
- 44 Lee, J. S. & Shin, E. C. The type I interferon response in COVID-19: implications for treatment. *Nature reviews. Immunology* **20**, 585-586, doi:10.1038/s41577-020-00429-3 (2020).
- 45 Jiang, H. *et al.* DNA methylation markers in the diagnosis and prognosis of common leukemias. *Signal Transduct Target Ther* **5**, 3, doi:10.1038/s41392-019-0090-5 (2020).
- 46 Ciolfi, A. *et al.* Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. *Clin Epigenetics* **12**, 7, doi:10.1186/s13148-019-0804-0 (2020).
- 47 Wang, L., Ni, S., Du, Z. & Li, X. A six-CpG-based methylation markers for the diagnosis of ovarian cancer in blood. *J Cell Biochem* **121**, 1409-1419, doi:10.1002/jcb.29376 (2020).
- 48 Imgenberg-Kreuz, J. *et al.* Shared and Unique Patterns of DNA Methylation in Systemic Lupus Erythematosus and Primary Sjogren's Syndrome. *Front Immunol* **10**, 1686, doi:10.3389/fimmu.2019.01686 (2019).
- 49 Bend, E. G. *et al.* Gene domain-specific DNA methylation epesignatures highlight distinct molecular entities of ADNP syndrome. *Clin Epigenetics* **11**, 64, doi:10.1186/s13148-019-0658-5 (2019).
- 50 Wang, C., Chen, L., Yang, Y., Zhang, M. & Wong, G. Identification of potential blood biomarkers for Parkinson's disease by gene expression and DNA methylation data integration analysis. *Clin Epigenetics* **11**, 24, doi:10.1186/s13148-019-0621-5 (2019).
- 51 Panagopoulou, M. *et al.* Circulating cell-free DNA in breast cancer: size profiling, levels, and methylation patterns lead to prognostic and predictive classifiers. *Oncogene* **38**, 3387-3401, doi:10.1038/s41388-018-0660-y (2019).

- 52 Breton, C. V. *et al.* Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children's Environmental Health Studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. *Environmental health perspectives* **125**, 511-526, doi:10.1289/EHP595 (2017).
- 53 Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nature reviews. Genetics* **12**, 529-541, doi:10.1038/nrg3000 (2011).
- 54 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).
- 55 Michels, K. B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nature methods* **10**, 949-955, doi:10.1038/nmeth.2632 (2013).
- 56 Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, doi:10.1126/science.abd4570 (2020).
- 57 van der Made, C. I. *et al.* Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA*, doi:10.1001/jama.2020.13719 (2020).
- 58 Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in Covid-19 *medRxiv* **2020.09.24.20200048** (2020).
- 59 Hamano, E. *et al.* Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem Biophys Res Commun* **329**, 1234-1239, doi:10.1016/j.bbrc.2005.02.101 (2005).
- 60 He, J. *et al.* Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: a case-control study. *BMC Infect Dis* **6**, 106, doi:10.1186/1471-2334-6-106 (2006).
- 61 Zhou, S. *et al.* Circulating Proteins Influencing COVID-19 Susceptibility and Severity: a Mendelian Randomization Study. *medRxiv*, 2020.2010.2013.20212092, doi:10.1101/2020.10.13.20212092 (2020).
- 62 Kucirka, L., Lauer, S., Laeyendecker, O., Boon, D. & Lessler, J. Variation in False Negative Rate of RT-PCR Based SARS-CoV-2 Tests by Time Since Exposure. *medRxiv*, 2020.2004.2007.20051474, doi:10.1101/2020.04.07.20051474 (2020).
- 63 Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med*, doi:10.7326/M20-1495 (2020).
- 64 Tsai, P. C. & Bell, J. T. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* **44**, 1429-1441, doi:10.1093/ije/dyv041 (2015).
- 65 Oran, D. P. & Topol, E. J. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med* **173**, 362-367, doi:10.7326/M20-3012 (2020).
- 66 Tan, J. *et al.* Transmission and clinical characteristics of asymptomatic patients with SARS-CoV-2 infection. *Future Virology* **15**, 373-380, doi:10.2217/fvl-2020-0087 (2020).
- 67 Buitrago-Garcia, D. *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Medicine* **17**, e1003346, doi:10.1371/journal.pmed.1003346 (2020).
- 68 Zou, L. *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* **382**, 1177-1179, doi:10.1056/NEJMc2001737 (2020).
- 69 Hu, Z. *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* **63**, 706-711, doi:10.1007/s11427-020-1661-4 (2020).

70 Liu, Y. *et al.* Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect Dis* **20**, 656-657, doi:10.1016/S1473-3099(20)30232-2 (2020).

Additional Information:

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to Kathleen C. Barnes at Kathleen.barnes@cuanscutz.edu.

Reprints and permissions information is available at www.nature.com/reprints.

- 1 *Coronavirus Disease 2019 (COVID-19) Emergency Use Authorizations for Medical Devices*, <<https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations>> (2020).
- 2 Menachery, V. D. *et al.* Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses. *mBio* **5**, e01174-01114, doi:10.1128/mBio.01174-14 (2014).
- 3 Marazzi, I. *et al.* Suppression of the antiviral response by an influenza histone mimic. *Nature* **483**, 428-433, doi:10.1038/nature10892 (2012).
- 4 Menachery, V. D. *et al.* MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc Natl Acad Sci U S A* **115**, E1012-E1021, doi:10.1073/pnas.1706928115 (2018).
- 5 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565-574, doi:10.1016/S0140-6736(20)30251-8 (2020).
- 6 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497-506, doi:10.1016/S0140-6736(20)30183-5 (2020).
- 7 Roser, M., Ritchie, H., Ortiz-Ospina, E. & Hasell, J. *Coronavirus Disease (COVID-19) – Statistics and Research. Our World in Data.*, <<https://ourworldindata.org/coronavirus>> (2020).
- 8 Woloshin, S., Patel, N. & Kesselheim, A. S. False Negative Tests for SARS-CoV-2 Infection - Challenges and Implications. *N Engl J Med* **383**, e38, doi:10.1056/NEJMp2015897 (2020).
- 9 Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med* **173**, 262-267, doi:10.7326/M20-1495 (2020).
- 10 Fox-Lewis, S., Muttaiyah, S., Rahnema, F., McAuliffe, G. & Roberts, S. An understanding of discordant SARS-CoV-2 test results: an examination of the data from a central Auckland laboratory. *N Z Med J* **133**, 81-88 (2020).
- 11 Aslam, A. *et al.* SARS CoV-2 Surveillance and Exposure in the Perioperative Setting with Universal testing and Personal Protective Equipment (PPE) Policies. *Clin Infect Dis*, doi:10.1093/cid/ciaa1607 (2020).

- 12 Lan, L. *et al.* Positive RT-PCR Test Results in Patients Recovered From COVID-19. *Jama* **323**, 1502-1503, doi:10.1001/jama.2020.2783 (2020).
- 13 Vandenberg, O., Martiny, D., Rochas, O., van Belkum, A. & Kozlakidis, Z. Considerations for diagnostic COVID-19 tests. *Nat Rev Microbiol*, doi:10.1038/s41579-020-00461-z (2020).
- 14 Garibaldi, B. T. *et al.* Patient Trajectories Among Persons Hospitalized for COVID-19 : A Cohort Study. *Ann Intern Med*, doi:10.7326/M20-3905 (2020).
- 15 Bilinski, A. & Emanuel, E. J. COVID-19 and Excess All-Cause Mortality in the US and 18 Comparison Countries. *Jama* **324**, 2100-2102, doi:10.1001/jama.2020.20717 (2020).
- 16 McElvaney, O. J. *et al.* A linear prognostic score based on the ratio of interleukin-6 to interleukin-10 predicts outcomes in COVID-19. *EBioMedicine* **61**, 103026, doi:10.1016/j.ebiom.2020.103026 (2020).
- 17 Aref-Eshghi, E. *et al.* Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet* **106**, 356-370, doi:10.1016/j.ajhg.2020.01.019 (2020).
- 18 Bollepalli, S., Korhonen, T., Kaprio, J., Anders, S. & Ollikainen, M. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. *Epigenomics* **11**, 1469-1486, doi:10.2217/epi-2019-0206 (2019).
- 19 Agha, G. *et al.* Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. *Circulation* **140**, 645-657, doi:10.1161/CIRCULATIONAHA.118.039357 (2019).
- 20 Jurmeister, P. *et al.* Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aaw8513 (2019).
- 21 Benhamida, J. K. *et al.* Reliable Clinical MLH1 Promoter Hypermethylation Assessment Using a High-Throughput Genome-Wide Methylation Array Platform. *J Mol Diagn* **22**, 368-375, doi:10.1016/j.jmoldx.2019.11.005 (2020).
- 22 Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474, doi:10.1038/nature26000 (2018).
- 23 Karimi, S. *et al.* The central nervous system tumor methylation classifier changes neuro-oncology practice for challenging brain tumor diagnoses and directly impacts patient care. *Clin Epigenetics* **11**, 185, doi:10.1186/s13148-019-0766-2 (2019).
- 24 Korshunov, A. *et al.* DNA-methylation profiling discloses significant advantages over NanoString method for molecular classification of medulloblastoma. *Acta Neuropathol* **134**, 965-967, doi:10.1007/s00401-017-1776-9 (2017).
- 25 van Iterson, M., van Zwet, E. W., Consortium, B. & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18**, 19, doi:10.1186/s13059-016-1131-9 (2017).
- 26 Iwata, H. *et al.* PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. *Nature communications* **7**, 12849, doi:10.1038/ncomms12849 (2016).
- 27 Lee, J. S. & Shin, E. C. The type I interferon response in COVID-19: implications for treatment. *Nat Rev Immunol* **20**, 585-586, doi:10.1038/s41577-020-00429-3 (2020).
- 28 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
- 29 Jiang, H. *et al.* DNA methylation markers in the diagnosis and prognosis of common leukemias. *Signal Transduct Target Ther* **5**, 3, doi:10.1038/s41392-019-0090-5 (2020).

- 30 Ciolfi, A. *et al.* Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. *Clin Epigenetics* **12**, 7, doi:10.1186/s13148-019-0804-0 (2020).
- 31 Wang, L., Ni, S., Du, Z. & Li, X. A six-CpG-based methylation markers for the diagnosis of ovarian cancer in blood. *Journal of cellular biochemistry* **121**, 1409-1419, doi:10.1002/jcb.29376 (2020).
- 32 Imgenberg-Kreuz, J. *et al.* Shared and Unique Patterns of DNA Methylation in Systemic Lupus Erythematosus and Primary Sjogren's Syndrome. *Front Immunol* **10**, 1686, doi:10.3389/fimmu.2019.01686 (2019).
- 33 Bend, E. G. *et al.* Gene domain-specific DNA methylation epigenatures highlight distinct molecular entities of ADNP syndrome. *Clin Epigenetics* **11**, 64, doi:10.1186/s13148-019-0658-5 (2019).
- 34 Wang, C., Chen, L., Yang, Y., Zhang, M. & Wong, G. Identification of potential blood biomarkers for Parkinson's disease by gene expression and DNA methylation data integration analysis. *Clin Epigenetics* **11**, 24, doi:10.1186/s13148-019-0621-5 (2019).
- 35 Panagopoulou, M. *et al.* Circulating cell-free DNA in breast cancer: size profiling, levels, and methylation patterns lead to prognostic and predictive classifiers. *Oncogene* **38**, 3387-3401, doi:10.1038/s41388-018-0660-y (2019).
- 36 Breton, C. V. *et al.* Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children's Environmental Health Studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. *Environ Health Perspect* **125**, 511-526, doi:10.1289/EHP595 (2017).
- 37 Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nature reviews. Genetics* **12**, 529-541, doi:10.1038/nrg3000 (2011).
- 38 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).
- 39 Michels, K. B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* **10**, 949-955, doi:10.1038/nmeth.2632 (2013).
- 40 Rahmani, E. *et al.* Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat Methods* **13**, 443-445, doi:10.1038/nmeth.3809 (2016).
- 41 Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, doi:10.1126/science.abd4570 (2020).
- 42 Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in Covid-19. *medRxiv* (2020).
- 43 Hamano, E. *et al.* Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem Biophys Res Commun* **329**, 1234-1239, doi:10.1016/j.bbrc.2005.02.101 (2005).
- 44 He, J. *et al.* Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: a case-control study. *BMC Infect Dis* **6**, 106, doi:10.1186/1471-2334-6-106 (2006).
- 45 Zhou, Y., Liu, Z., Yang, H., Wang, J. & Li, M. J. Coagulation factors and COVID-19 severity: Mendelian randomization analyses and supporting evidence. *medRxiv* (2020).
- 46 Horton, R. *et al.* Gene map of the extended human MHC. *Nature reviews. Genetics* **5**, 889-899, doi:10.1038/nrg1489 (2004).

- 47 Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in False Negative Rate of RT-PCR Based SARS-CoV-2 Tests by Time Since Exposure. *medRxiv* (2020).
- 48 Tsai, P. C. & Bell, J. T. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* **44**, 1429-1441, doi:10.1093/ije/dyv041 (2015).
- 49 Sadikovic, B., Aref-Eshghi, E., Levy, M. A. & Rodenhiser, D. DNA methylation signatures in mendelian developmental disorders as a diagnostic bridge between genotype and phenotype. *Epigenomics* **11**, 563-575, doi:10.2217/epi-2018-0192 (2019).
- 50 Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol* **125**, 913-916, doi:10.1007/s00401-013-1126-5 (2013).
- 51 Vasudevan, H. N. *et al.* DNA methylation profiling demonstrates superior diagnostic classification to RNA-sequencing in a case of metastatic meningioma. *Acta Neuropathol Commun* **8**, 82, doi:10.1186/s40478-020-00952-3 (2020).
- 52 Voet, D. & Voet, J. *Biochemistry*. 4th edn, 85 (Wiley & Sons, 2011).
- 53 Oran, D. P. & Topol, E. J. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med* **173**, 362-367, doi:10.7326/M20-3012 (2020).
- 54 Tan, F. *et al.* Viral Transmission and Clinical Features in Asymptomatic Carriers of SARS-CoV-2 in Wuhan, China. *Front Med (Lausanne)* **7**, 547, doi:10.3389/fmed.2020.00547 (2020).
- 55 Buitrago-Garcia, D. *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med* **17**, e1003346, doi:10.1371/journal.pmed.1003346 (2020).
- 56 Zou, L. *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N Engl J Med* **382**, 1177-1179, doi:10.1056/NEJMc2001737 (2020).
- 57 Hu, Z. *et al.* Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* **63**, 706-711, doi:10.1007/s11427-020-1661-4 (2020).
- 58 Liu, Y. *et al.* Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect Dis* **20**, 656-657, doi:10.1016/S1473-3099(20)30232-2 (2020).
- 59 Group, R. C. *et al.* Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*, doi:10.1056/NEJMoa2021436 (2020).
- 60 Saben, J. L. *et al.* The Emergency Medicine Specimen Bank: An Innovative Approach To Biobanking In Acute Care. *Acad Emerg Med* **26**, 639-647, doi:10.1111/acem.13620 (2019).
- 61 Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 208, doi:10.1186/s13059-016-1066-1 (2016).
- 62 Robinson, J. *et al.* The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* **43**, D423-431, doi:10.1093/nar/gku1161 (2015).
- 63 Boorgula, M. P. *et al.* Replicated methylation changes associated with eczema herpeticum and allergic response. *Clin Epigenetics* **11**, 122, doi:10.1186/s13148-019-0714-1 (2019).

- 64 Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**, 549-554, doi:10.1038/ng1547 (2005).
- 65 Zhou, W., Triche, T. J., Jr., Laird, P. W. & Shen, H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res* **46**, e123, doi:10.1093/nar/gky691 (2018).
- 66 Triche, T. J., Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90, doi:10.1093/nar/gkt090 (2013).
- 67 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22, doi:10.1093/nar/gkw967 (2017).
- 68 Rahmani, E. *et al.* GLINT: a user-friendly toolset for the analysis of high-throughput DNA-methylation array data. *Bioinformatics* **33**, 1870-1872, doi:10.1093/bioinformatics/btx059 (2017).
- 69 Rahmani, E. *et al.* Genome-wide methylation data mirror ancestry information. *Epigenetics Chromatin* **10**, 1, doi:10.1186/s13072-016-0108-y (2017).
- 70 Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**, 833-835, doi:10.1038/nmeth.1681 (2011).
- 71 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
- 72 Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-261, doi:10.1093/nar/gkh036 (2004).
- 73 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092 (2017).
- 74 Huang, R. *et al.* The NCATS BioPlanet - An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics. *Front Pharmacol* **10**, 445, doi:10.3389/fphar.2019.00445 (2019).
- 75 Slenter, D. N. *et al.* WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* **46**, D661-D667, doi:10.1093/nar/gkx1064 (2018).
- 76 Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288 (1996).
- 77 Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320, doi:<https://doi.org/10.1111/j.1467-9868.2005.00503.x> (2005).

Figures

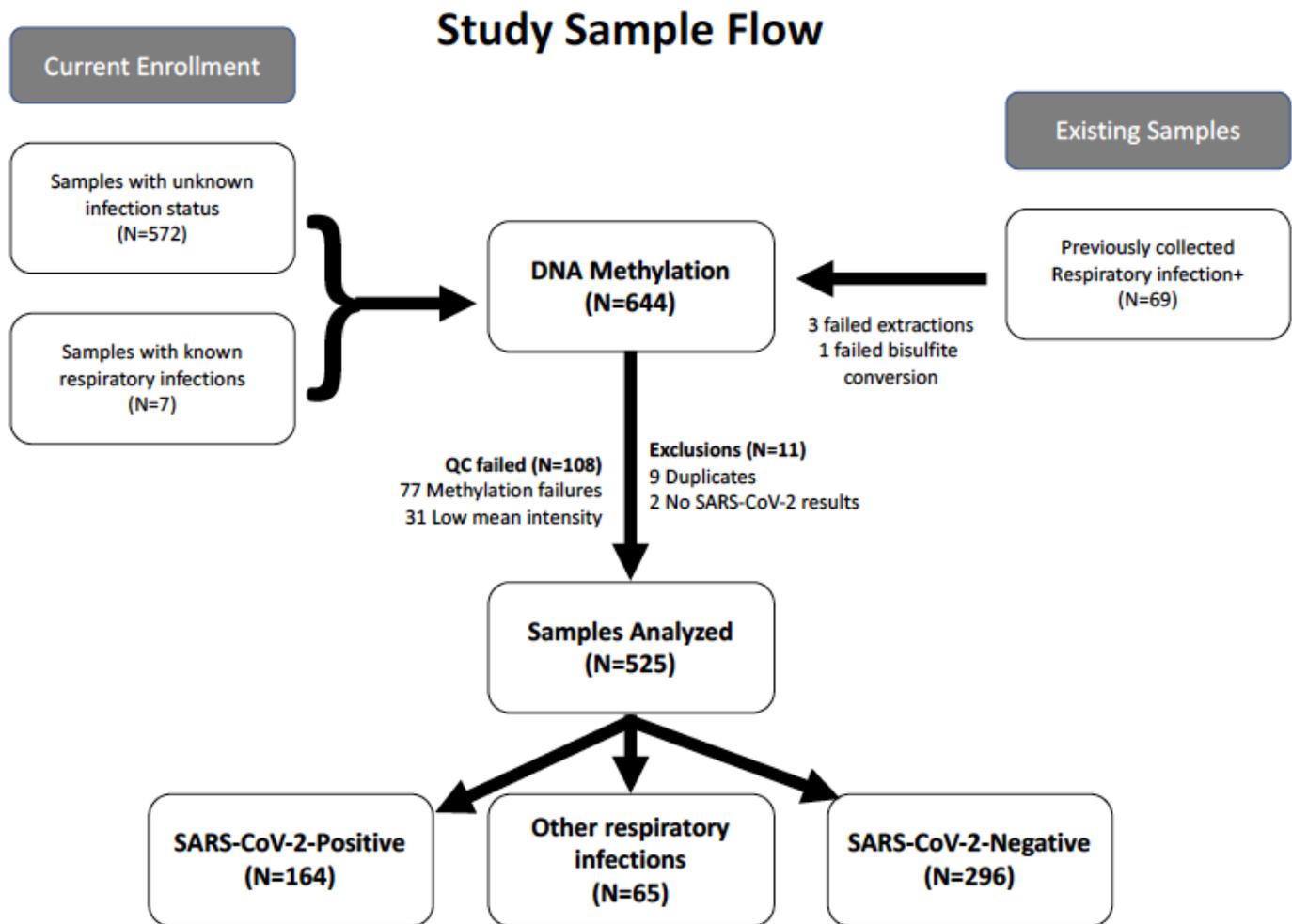


Figure 1

Flowchart of the study samples.

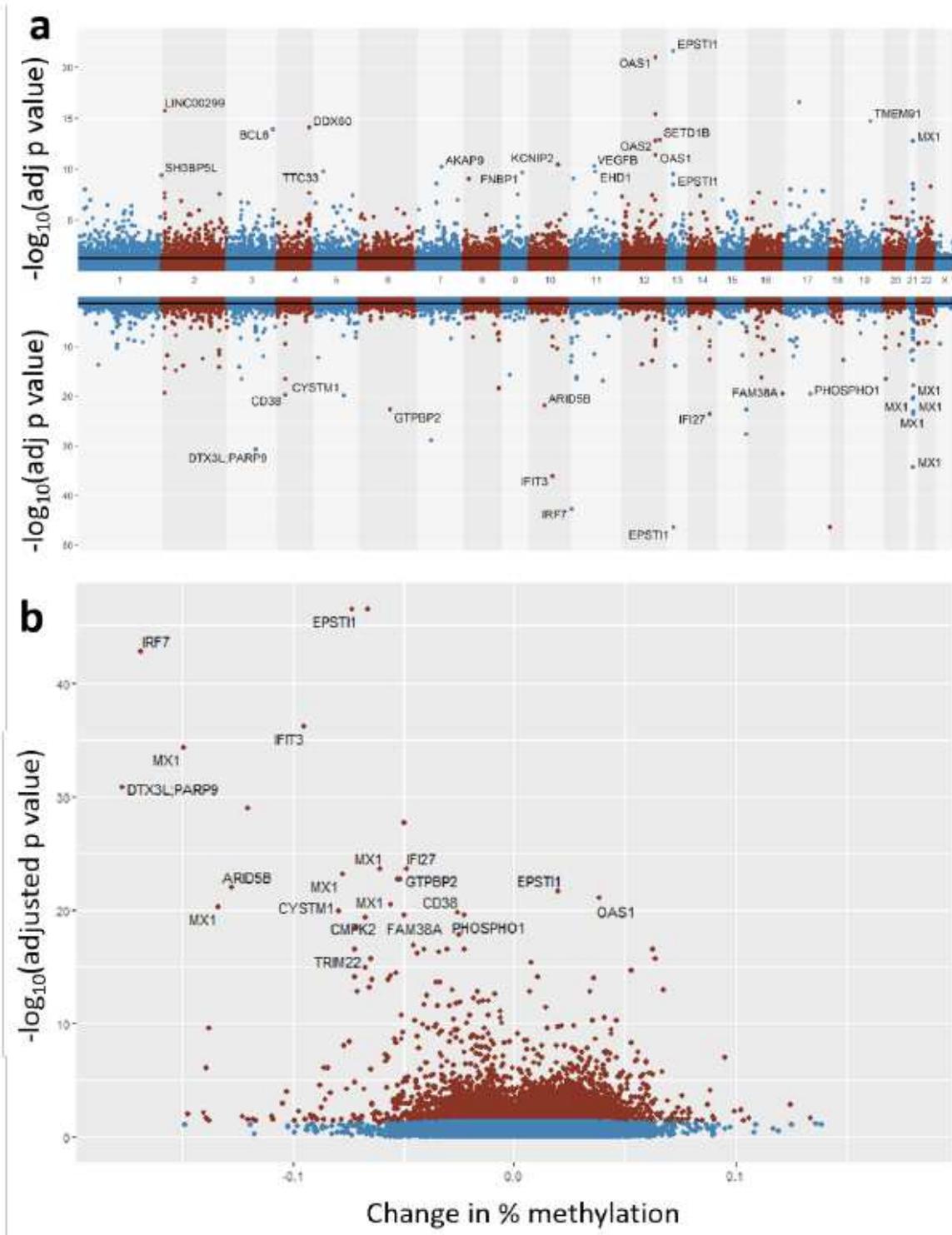


Figure 2

a, Miami plot (top panel) of hypermethylated (top) and hypomethylated (bottom) probes in SARS-Cov-2+ compared to control samples. Significance lines represent FDR-adjusted P-value < 0.05 threshold. b, Volcano plot of significant (red; FDR-adjusted P-value < 0.05) CpG sites (blue CpG sites have FDR-adjusted P-value > 0.05). Change in % methylation on the x axis represents the difference in average beta value at a site between cases and controls. Probes for intergenic CpG sites do not have gene annotations.

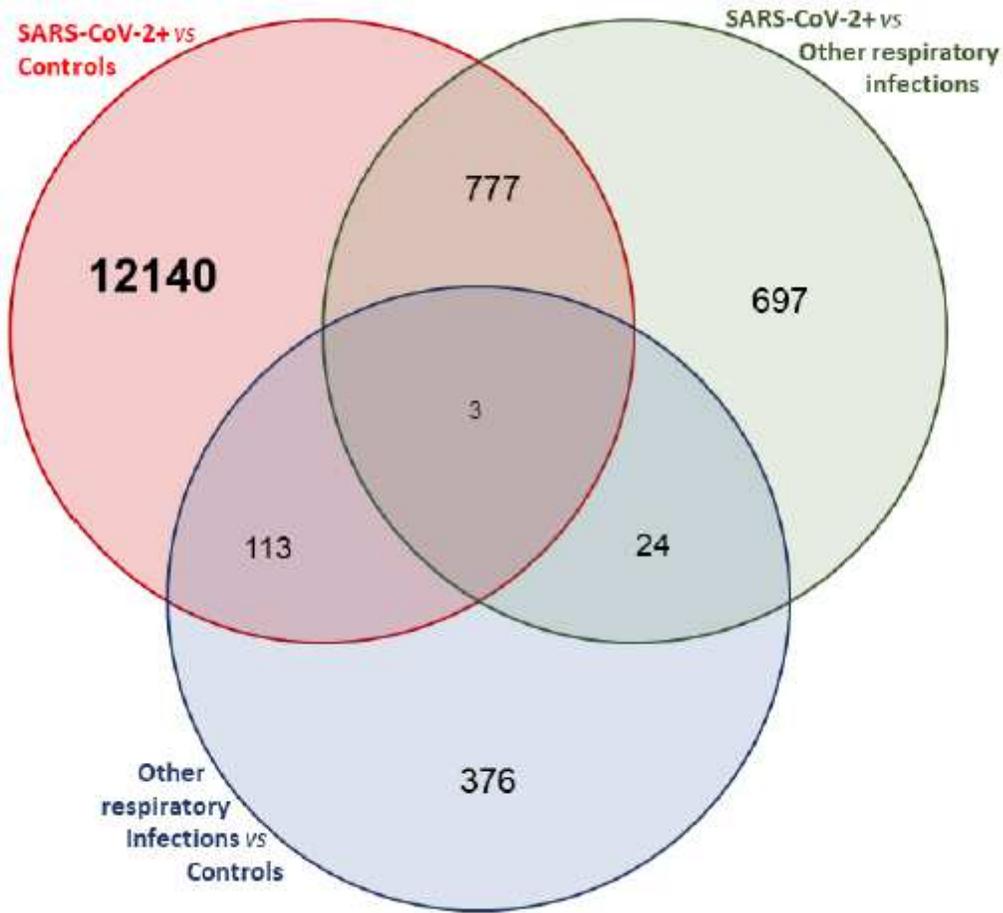


Figure 3

Venn diagram of overlaps between COVID-Control EWAS (13,033 significant probes), COVID-viral infection EWAS (516 significant probes), and respiratory infection- Control EWAS (1,501 significant probes).

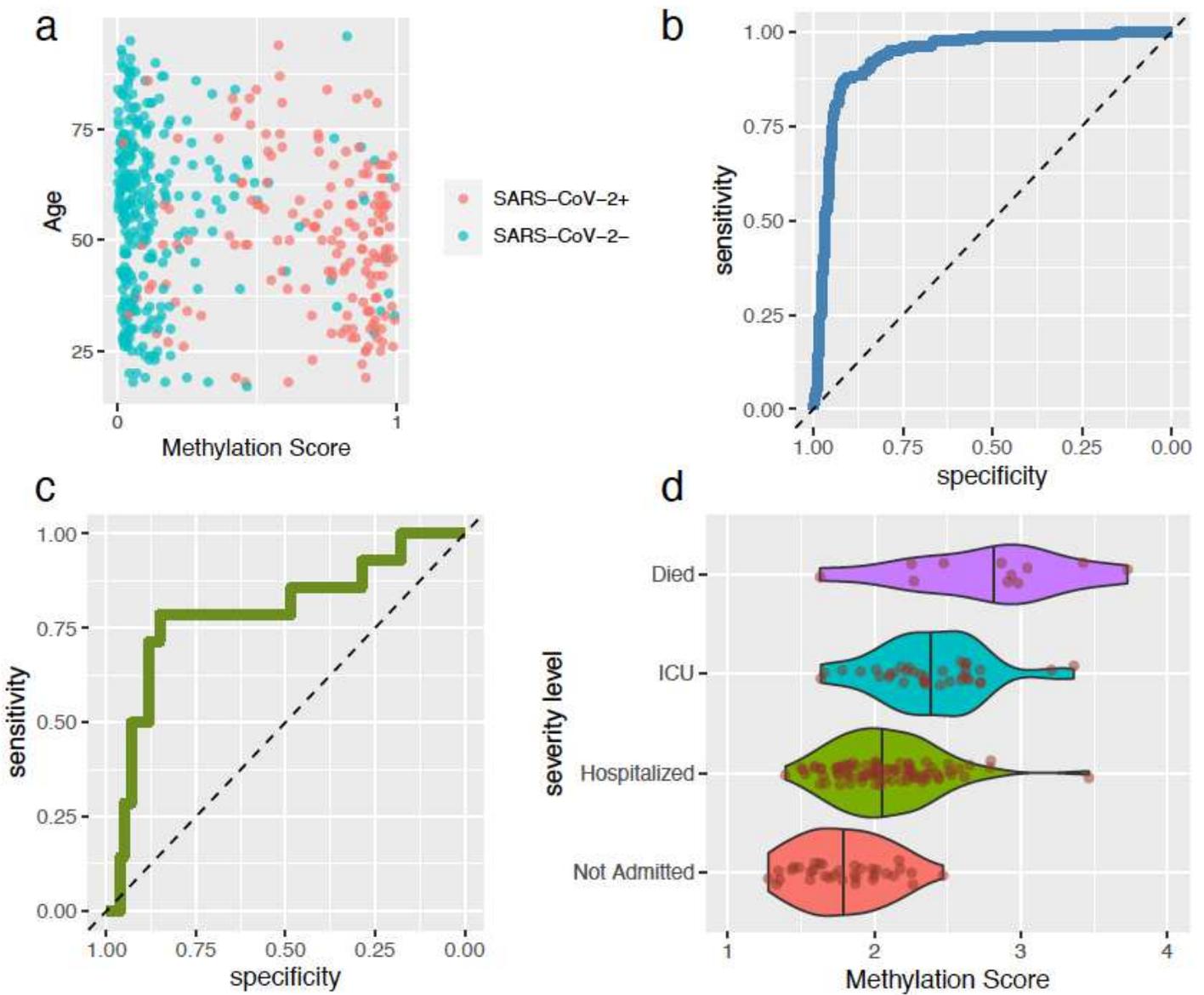


Figure 4

a, Out-of-sample case-control methylation score for all 460 individuals compared to case-control status, plotted by biological age. b. Receiver-Operating Characteristic (ROC) Curve for data in a. c, ROC curve of cross-validated prediction of long hospital duration d. Violin and jittered scatter plots of severity methylation scores for each outcome in cases.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1COVIDvsControlsSignificantAssociations.csv](#)

- [SupplementaryTable2COVIDvsResplnfectionSignificantAssociations.csv](#)
- [SupplementaryTable3ResplnfectionvsControlsSignificantAssociations.csv](#)
- [SupplementaryTable4COVICprobes.csv](#)