nature portfolio

Peer Review File

Identification of reactive CpGs and RNA expression in early COVID-19 through cis-eQTM analysis reflecting disease severity and recovery

Corresponding Author: Professor Jong Bhak

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this manuscript, the authors focused on the link between early gene regulation and COVID-19 severity in terms of epigenomics and transcriptomics. By analyzing 51 cases of COVID-19 through cis-eQTM analysis, the authors found and selected six key genes and their nine regulatory CpGs associated with COVID-19 infection and severity, as shown in Table 1. Then they simply explained their distinct change patterns in different phases of infection with COVID-19 and took a discussion about limitations. However, this research seems insufficient in combination with clinical factors and still remain several concerns that need to be addressed:
1.The baseline characteristics and outcomes of COVID-19 patients are not clear enough in Table 2. There is a high possibility that infected patients have comorbidities. For COVID-19 patients, usually accompanied by a higher mortality rate than healthy people, are the 51 cases outcomes are Complete recovery? It is insufficient to support the claim without dead samples for critical COVID-19 patients, which may bring selection bias. If possible, more samples are needed for further analysis.
2.COVID-19 disease is highly variable in symptoms and commonly accompanied by complications. It seems too narrow for the research to focus only on the comparison within severe, mild, and convalescent groups. It would be better if the results of the stratified analysis in different clinical factors and further analysis of patients with common complications such as pneumonia and acute respiratory distress syndrome were added to the analysis.
3.Please pay more attention to check and add the information for samples in “an independent convalescent group (N=90) and healthy group (N=310)” for Figure S4 and S5. Please pay more attention to check the reference list. The 32nd reference seems not a perspective that related to the patient classification of COVID-19 severity.
4.Please clearly explain the abbreviation used in figures and tables, especially the y-axis names for supplementary figure 1. Please explain the definitions of the “box” and the “bar” of the Box plots.

Reviewer #2

(Remarks to the Author)
Review of COMMSBIO-24-7574
Ryu and colleagues presents a multi-omics study utilizing cis-eQTM analysis to investigate the relationship between DNA methylation and gene expression in COVID-19 patients. In a clinical cohort including mild, moderate and severe acute COVID-19 cases, PAXgene samples were collected and DNA methylation assessed by whole genome bi-sulfide sequencing. In addition, RNA-sequencing was performed in a matched-pair design to study the regulatory effects of gene methylations. Six key genes were identified which fulfilled the criteria of differential methylation between severe and mild and linked gene expression. Follow-up of patients provided additional insights. Convalescence was linked to changes of CpG sites to levels found in mild or healthy controls.

While this approach is valuable and relevant, there are several critical issues that require addressing. The manuscript requires substantial revision before publication.

Results

- The characterization of patient groups (mild, severe, convalescent) lacks clarity. The convalescent group requires further elaboration regarding its classification, treatment status, and potential confounders. Additionally, demographic details (e.g., smoking status, medication use) should be included in Table 2 to enhance the cohort's description.
- In Table 2, which presents baseline characteristics, the convalescent group appears to be missing. This makes it challenging to understand the distinction between the convalescent group, mild (recovery), and severe (recovery) groups. Providing clarity here would greatly enhance the manuscript.

- The study does not provide sufficient information on how potential confounding factors, such as medication use in acute COVID-19 cases, were accounted for. Some medications may influence DNA methylation patterns, which should be discussed.

- The inclusion of cytokine measurements would enhance the study, given the focus on immune response and cytokine storm mechanisms. If cytokine data were not collected, this should be stated explicitly as a limitation.

- The manuscript provides limited insight into clinical outcomes. Although it stratifies patients based on severity, it does not evaluate associations with long-term recovery markers such as residual symptoms, lung function, or chronic fatigue. These outcomes would strengthen the study's clinical relevance.

page 6, line 124, the phrasing of the sentence "Out of the twelve DEGs regulated their expression by the DMCs" could be improved for clarity. Consider rephrasing this for better readability.

Methods
- The study does not mention whether DNA methylation data were adjusted for cellular composition. Given the known impact of acute infection on immune cell populations (e.g., neutrophilia, T-cell depletion), the results may reflect changes in cell proportions rather than true epigenetic modifications. The use of methylation deconvolution algorithms should be considered (e.g., Accomando et al., 2014; Kaushal et al., 2017). During acute phases of infection neutrophilia or neutropenia are reported. Likewise T-cell and monocyte cell counts vary during acute infection and recovery / convalence. These changes affect large proportions of cells and easily may be read as expression or methylation changes. For example IL18R is expressed by monocytes, B and T cells and NK cells but not by neutrophils. High neutrophil counts are confirmed in your supplemental table Fig 1.

- The manuscript acknowledges the impact of SARS-CoV-2 variants (Delta vs. Omicron) but does not attempt to stratify patients accordingly. Given that the mild-moderate group includes both Delta- and Omicron-infected patients, the authors should explore stratification or discuss the feasibility of such an analysis.

- The classification of patient groups could be improved. The term "mild" currently subsumes mild and moderate cases. To avoid confusion, the manuscript should consistently refer to this group as "mild-moderate" (MM) in all sections.

- The statistical approach requires further justification. The manuscript states that a two-sample t-test was used despite a small and unequal sample size (n=10 severe vs. n=41 mild). The authors should clarify whether assumptions of normality were tested and, if necessary, replace t-tests with non-parametric alternatives.

- Technical details require improvement. The description of the sequencing process includes unclear phrasing (e.g., "Sequenced DNA reads were removed their adapters"). Additionally, product codes for kits and reagents should be provided for reproducibility.

- Normalized gene expression was estimated using RSEM, while DESeq2 was used in the figure. Could you please explain the rationale behind this apparent discrepancy?

On page 18, lines 318–319, kindly mention the specific parameters assessed. This additional detail would make this section more informative.

On page 19, line 340, the sentence starting with "We utilized the LOO..." is somewhat unclear. The phrasing "iteratively collecting DEGs without a sample" is particularly difficult to interpret. Do you mean that one sample is excluded per round? If so, does this mean a single sample is removed per group in each iteration? Providing a clearer explanation of this methodology would significantly improve the reader's understanding.

On page 6, line 132, it is mentioned that an independent convalescent and healthy group was used. Providing more details about the demographics of this group would be valuable. Additionally, this sentence could be rephrased to enhance its clarity.

Include the product codes for all kits and materials used in the study. This will improve reproducibility and clarity for readers.

Have the authors considered conducting a pathway enrichment analysis for the identified DEGs and DMGs? Including the results of such an analysis would provide valuable insights into the biological processes and pathways implicated in the study.

Figures and Supplementary Information
- Figure 1B lacks a clear explanation of how patients were classified into mild, severe, and convalescent groups. The figure legend should explicitly define each group.

- Table 1 should include citations to support the reported associations of key genes with COVID-19 severity.

- Figure S1 requires improvements: the y-axis should include units, and significant indicators (color coding) may not be necessary.

- Figure 2B is missing the healthy control group for RNA expression analysis. This omission limits comparative insights into baseline gene expression.

- Figure 3 should better clarify the relationships between methylation patterns, gene expression, and recovery trajectories.

- Some supplementary figures (e.g., Supplementary Figure 6) contain ambiguous labels that could be misinterpreted as the number of patients rather than gene counts. Clarification is needed.

The inclusion of a graphical abstract illustrating the study design is greatly appreciated. However, some modifications could enhance its clarity and effectiveness. For instance, ensure all groups are clearly labeled and consistent with the main text. Furthermore, in figure 1B, a sample (C19-014) stated to have been removed due to missing data or QC failure is still shown. Could you clarify this? Additionally, it would be helpful to include the number of patients per group in the figure legend to provide a clearer overview. Please also include a brief explanation about the convalescent group in the legend.

Discussion
- The discussion should begin with a summary of key findings rather than limitations.

- The section "Key Genes and Molecular Mechanisms Influencing COVID-19 Severity" should be integrated into the discussion, rather than presented as a standalone section.

- The manuscript acknowledges the impact of SARS-CoV-2 variants (Delta vs. Omicron) but does not attempt to stratify patients accordingly. Given that the mild-moderate group includes both Delta- and Omicron-infected patients, the authors should explore stratification or discuss the feasibility of such an analysis.

- The potential role of identified genes in long-term COVID-19 outcomes is not explored. Since the study mentions rapid epigenetic recovery, it would be valuable to discuss whether residual methylation changes could contribute to post-acute sequelae.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
None

Reviewer #2

(Remarks to the Author)
Dear Hyojung Ryu et al.,

Thank you very much for addressing all my concerns with such accuracy and completeness. I have no further request with one minor aspect. Suppl. Fig5 has p-value indicators that are not present in the graphic. Please remove

Supplementary Fig 5. Baseline leukocyte proportion and its changes one week later in Mild-Moderate and Severe-Critical groups. Boxplots show the proportions (%) of neutrophils, lymphocytes, monocytes, eosinophils, and basophils measured in whole blood at baseline (left panel) and after a week later (right panel) for patients with Mild–Moderate (green) and Severe–Critical (red) disease. Boxplots display the median (centre line), IQR (box limits), and 1.5×IQR (whiskers), and outliers (open circles). Statistical comparisons were conducted using the Wilcoxon rank-sum test. Statistical significance is denoted as follows: ns ($P \geq 0.05$), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), **** ($P < 0.0001$).

Again, thank you very much for your comprehensive revision!

Title : Identification of reactive CpGs and RNA expression in early COVID-19 through *cis*-eQTM analysis reflecting disease severity and recovery

Reviewer #1

Comment 1 :

The **baseline characteristics** and outcomes of COVID-19 patients are not clear enough in Table 2. There is a high possibility that infected patients have **comorbidities**. For COVID-19 patients, usually accompanied by a **higher mortality rate** than healthy people, are the 51 cases outcomes are **Complete recovery**? It is insufficient to support the claim without dead samples for critical COVID-19 patients, which may bring selection bias. If possible, more samples are needed for further analysis.

Response : Thank you for your valuable comment regarding the baseline characteristics and potential selection bias. As per your suggestion, we have updated the baseline characteristics in Table 2 to include additional clinical features. The revised table now includes medication history, presence of pneumonia and ARDS. Additionally, to reflect comorbidities as you mentioned, we have added the Charlson Comorbidity Index (CCI). In the process of revising the manuscript, we identified an error in the originally reported sample size. While 51 patients were initially described, the correct number of patients analyzed in this study is 46, and this has been corrected throughout the revised manuscript. All 46 patients included were fully recovered cases. In this study, we did not intentionally exclude fatal cases; however, none of the severe COVID-19 patients enrolled experienced death during the study period. As a result, all patients achieved complete recovery.

We believe that the absence of fatal cases in our severe group can be attributed to several clinical and public health factors. First, vaccination—which is a key factor in reducing mortality—was relatively common in the severe group (66.7%, vaccination rate). Additionally, in South Korea, antiviral drugs such as remdesivir were rapidly available through government support, and 66.7% of

patients in the severe group received such medications in a timely manner. Second, following the publication of the RECOVERY trial in February 2021(N Engl J Med 2021;384:693–704), dexamethasone became the first treatment shown to improve survival in hospitalized patients requiring oxygen or mechanical ventilation. In accordance with updated global treatment guidelines, 77.8% of patients in the severe-critical group in our study received early steroid therapy, which likely contributed to the absence of fatal cases. Lastly, since the emergence of the Omicron variant in late 2021, the overall mortality rate has decreased markedly. Although a few extremely severe cases did exist during this period, it was difficult to obtain consent from such patients or their caregivers for study participation. We fully agree that incorporating fatal cases could offer more robust insights into the molecular features associated with critical COVID-19 outcomes. In future studies, we will strive to include fatal cases and expand the cohort size to minimize potential selection bias and strengthen the generalizability of our findings.

	COVID-19 Hospitalized patients		**Convalescent (n=90)**	**Healthy controls (n=310)**
	Mild-Moderate group (n=37)	**Severe-Critical group (n=9)**		
Age, mean(min, max)	44(19,71)	50(24,66)	42(19,66)	42(19, 66)
Sex , n(%)				
Male	24(64.9)	4(44.4)	45(50)	188(60)
Female	13(35.1)	5(55.6)	45(50)	122(40)
BMI	24.8	25.5	24.4	23.45
Smoking status, n(%)				
Current	3(8.1)	3(33.3)		46(14.8)
Former	3(8.1)	1(11.1)		79(25.5)
Never	31(83.8)	5(55.6)		173(55.8)
NA	0	0	90	12(3.9)

Oxygen therapy, n(%)				
Yes				
Nasal prong only	8(21.6)	5(55.6)		
HFNC* only	0	2(22.2)		
Both (Nasal prong and HFNC*)	0	2(22.2)		
No	26(70.3)	0		
Intensive care unit, n(%)				
Yes	0 (0)	3(33.3)		
No	37(100)	6(66.7)		
Use of antiviral drugs, n(%)				
Yes				
Regdanvimab	3(8.1)	2(22.2)		
Remdesivir	10(27)	6(66.7)		
No	24(64.9)	1(11.1)		
Use of steroid, n(%)				
Yes	11(29.7)	7(77.8)		
No	26(70.3)	2(22.2)		
Pneumonia, n(%)				
Yes	10 (27)	7(77.8)		
No	27 (73.0)	2(22.2)		
ARDS, n(%)				
Yes	0 (0)	2(22.2)		
No	37(100)	7(77.8)		
Vaccination status, n(%)				
Yes	13(35.1)	6(66.7)		

No	24(64.9)	3(33.3)		
CCI, n(%)				
0-1	24(64.9)	1(11.1)		
2	4(10.8)	4(44.4)		
≥ 3	8(21.6)	3(33.3)		

Table 2. Baseline Characteristics of COVID-19 Hospitalized Patients, Convalescent Patients, and Non-infected Healthy Individuals. Demographic and clinical data are summarized for the clinical cohorts. Continuous variables (e.g., age and BMI) are presented as mean values with minimum and maximum ranges in parentheses. Categorical variables (e.g., sex, smoking status, and CCI) are expressed as counts with corresponding percentages. Smoking status was self-reported and categorized as current, former, or never smoker. Abbreviations: HFNC, high-flow nasal cannula; ARDS, acute respiratory distress syndrome; CCI, Charlson Comorbidity Index.

Comment 2 :

COVID-19 disease is highly variable in symptoms and commonly accompanied by complications. It seems too narrow for the research to focus only on the comparison within severe, mild, and convalescent groups. It would be better if the results of the **stratified analysis** in different clinical factors and further analysis of patients with common complications such as **pneumonia** and **acute respiratory distress syndrome** were added to the analysis.

Response : Thank you for your insightful comments. As recommended, we have considered performing additional stratified analyses based on common clinical factors such as pneumonia and acute respiratory distress syndrome (ARDS). In this study, we divided COVID-19 patients into severe and mild groups according to the FDA guidelines and confirmed pneumonia/ARDS status using the patients' clinical information. However, we found that the number of patients with these specific complications (e.g., pneumonia, ARDS) in each subgroup was too small to yield statistically meaningful stratified analyses. For instance, only two patients in the severe group were diagnosed

with ARDS, making further subgroup analysis difficult to interpret. We acknowledge this as a limitation of our current study, and we plan to address it in future investigations by recruiting a larger cohort or collaborating with multiple centers to increase the sample size. We appreciate your suggestion, as it will guide our efforts to refine and extend our analysis in subsequent research.

Comment 3:

Please pay more attention to check and add the information for samples in "an independent **convalescent group (N=90)** and **healthy group (N=310)**" for Figure S4 and S5.

Response : Thank you for your comment. As suggested, we have added cohort descriptions for the 'Convalescent group (N=90)' and 'Healthy controls (N=310, now updated as N=344)'—both of which were used for comparison with the Severe and Mild groups in our marker analyses—into the Methods sections. In addition, we revised Figure 1A to include a dedicated 'Comparison Cohort' part, clearly outlining the composition and role of these two groups.Furthermore, we updated the figure legends of all supplementary figures that include these cohorts to provide clear descriptions of the 'Convalescent group' and' Healthy controls', thereby ensuring that the characteristics of each cohort are accurately and explicitly presented.

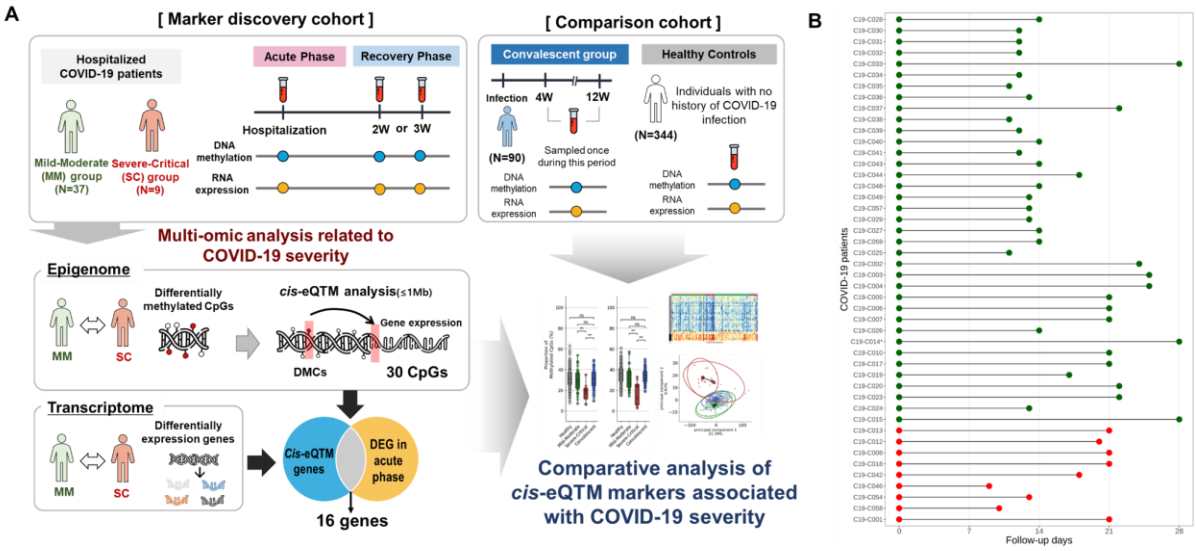
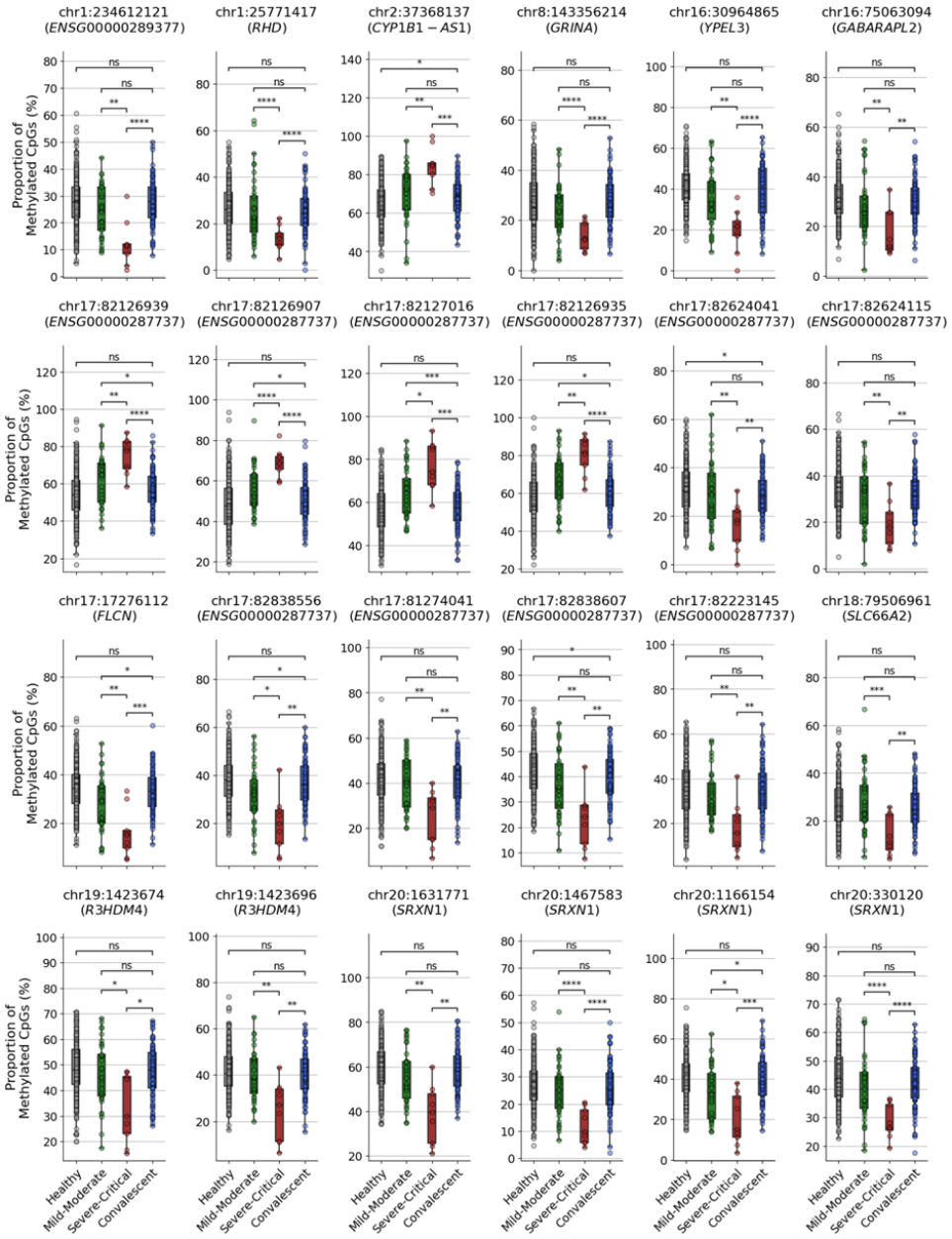


Figure 1. Overall study design and follow-up blood sample collection for *cis*-eQTM analysis in COVID-19 patients. A) Overview of blood sampling and analysis design; The marker discovery cohort consisted of hospitalized COVID-19 patients, stratified into Mild-Moderate (MM) and Severe-Critical (SC) groups, and was used to identify severity-associated *cis*-eQTM markers. The comparison cohort was composed of two independent reference groups : the 'Convalescent group' comprising individuals who recovered from COVID-19 and provided a single blood sample 4 to 12 weeks post-infection, and the 'Healthy Controls' group, consisting of pre-pandemic individuals with no history of SARS-CoV-2 infection. **B)** Clinical follow-up duration of 46 Hospitalized COVID-19 patients. The green and red points indicate MM and SC groups, respectively. Sample C19-C014(*), excluded DEG analysis due to QC failure, was retained in methylation analysis.

[Page 16, lines 324-335 (Revised manuscript)]: For the hospitalized patient group, whole blood samples were collected twice : once during the acute phase (at the time of hospital admission) and once during the recovery phase (2 to 3 weeks after hospitalization). For the convalescent patient group, whole blood samples were obtained between 4 to 12 weeks after COVID-19 diagnosis. These individuals had been previously hospitalized due to COVID-19 but had fully recovered at the time of sample collection. Our sampling period is divided into two halves. The first was during 2021 which was followed by the second in 2022. The acute phase of infection was defined as the period from hospitalization due to COVID-19 to one week thereafter, while the subsequent period of up to three weeks was defined as the recovery phase. A healthy control group consisted of 344 whole blood samples collected (309 samples for methylation, 34 for gene expression studies, and one sample for both omics) from the Korean Genome Project (KGP), approved by the IRB at UNIST in Ulsan, South Korea (IRB No.: UNISTIRB-21-66-A).



Supplementary Fig 3. DNA Methylation Levels of 24 cis-eQTM CpGs Associated with COVID-19 Severity During the Acute Phase. DNA methylation levels (% methylated CpGs) are shown for 24 cis-eQTM CpGs across four clinical groups: Healthy controls (n = 310), patients with Mild-Moderate (n = 37) or Severe-Critical (n = 9) COVID-19 during the acute phase, and Convalescent individuals (n = 90). The x-axis denotes severity groups. The y-axis represents the proportion of methylation at each CpG site. Boxplots display the median (centre line), IQR (box limits), 1.5×IQR (whiskers), and individual data points (closed circles). Statistical comparisons were conducted using Welch's t-test. Statistical significance is indicated as follows: ns ($P \geq 0.05$), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), **** ($P < 0.0001$).

Comment 4:

Please pay more attention to check the reference list. The **32nd reference** seems not a perspective that related to the patient classification of COVID-19 severity.

Response : We thank the reviewer's comment regarding the appropriateness of Reference 32. In our study, we classified COVID-19 patients into four severity levels—mild, moderate, severe, and critical—based on the U.S. FDA's Guidance for Industry: COVID-19: Developing Drugs and Biological Products for Treatment or Prevention (Feb 2021). We have added a supplementary table (Supplementary Table 5) that includes the severity definitions from the FDA guidance.

Severity Level	Criteria
Mild	• Positive testing by virologic test (i.e., a nucleic acid amplification test of an antigen test) • Symptoms of mild illness with COVID-19 that could include fever, cough, sore throat, malaise, headache, muscle pain, nausea, vomiting, diarrhea, and loss of taste or smell, without shortness of breath or dyspnea • No clinical signs indicative of Moderate, Severe, or Critical Severity
Moderate	• Positive testing by virologic test (i.e., a nucleic acid amplification test of an antigen test) • Symptoms of moderate illness with COVID-19, which could include any symptom of mild illness or shortness of breath with exertion • Clinical signs suggestive of moderate illness with COVID-19, such as respiratory rate ≥ 20 breaths per minute, heart rate ≥ 90 beats per minute; with saturation of oxygen (SpO2) > 93% on room air at sea level • No clinical signs indicative of Severe or Critical Illness Severity
Severe	• Positive testing by standard RT-PCR assay or an equivalent test • Symptoms suggestive of severe systemic illness with COVID-19, which could include any symptom of moderate illness or shortness of breath at rest, or respiratory distress • Clinical signs indicative of severe systemic illness with COVID-19, such as respiratory rate ≥ 30 per minute, heart rate ≥ 125 per minute, SpO2 ≤ 93% on room air at sea level or PaO2/FiO2 < 300 • No criteria for Critical Severity
Critical	• Positive testing by virologic test (i.e., a nucleic acid amplification test of an antigen test)

	• Evidence of critical illness, defined by at least one of the following: – Respiratory failure defined based on resource utilization requiring at least one of the following: ▪ Endotracheal intubation and mechanical ventilation, oxygen delivered by high flow nasal cannula (heated, humidified, oxygen delivered via reinforced nasal cannula at flow rates > 20 L/min with fraction of delivered oxygen ≥ 0.5), noninvasive positive pressure ventilation, ECMO, or clinical diagnosis of respiratory failure (i.e., clinical need for one of the preceding therapies, but preceding therapies not able to be administered in setting of resource limitation) – Shock (defined by systolic blood pressure < 90 mm Hg, or diastolic blood pressure < 60 mm Hg or requiring vasopressors) – Multi-organ dysfunction/failure

Supplementary Table 5. COVID-19 Severity Classification According to U.S. FDA Guidance (Feb 2021). The classification of COVID-19 severity used in this study is based on the U.S. FDA's Guidance for Industry: COVID-19: Developing Drugs and Biological Products for Treatment or Prevention.

To further support the use of this classification system, we cited Reference 32 (Sima S. Toussi *et al.*, Nat Microbiol, 2023), which applies the similar severity definitions that incorporate key elements of the FDA's guidance in its Main section. While Sima S. Toussi et al. is not a formal regulatory document, it includes major criteria—such as respiratory rate thresholds, oxygen saturation cutoffs, and definitions of respiratory failure and systemic complications—that are consistent with those outlined by the FDA. Therefore, we believe it remains relevant to our study context and complements the clinical validity of the severity framework we applied. Table R1 provides a side-by-side summary of shared components between the two sources across all four severity levels.

Severity Level	FDA Criteria	Reference 32 Criteria
Mild	- Mild symptoms such as fever, cough, sore throat, malaise, headache, muscle pain, nausea, vomiting, diarrhea, or loss of taste or smell - **No shortness of breath or signs of more severe illness**	- Fever, cough, sore throat, loss of taste and/or smell, headache, muscle aches, gastrointestinal symptoms (nausea, vomiting, and/or diarrhea) - **No shortness of breath or clinical signs of progression**

Moderate	- Moderate illness symptoms including **shortness of breath with exertion** - Clinical signs include **respiratory rate ≥ 20 breaths/minute**, **heart rate ≥ 90 beats/minute**, and **oxygen saturation > 93% on room air at sea level**	- Lower respiratory symptoms including **shortness of breath with exertion**; **- Respiratory rate between 20–29 breaths/minute**, able to maintain **oxygen saturation ≥ 94% on room air at sea level**
Severe	- Symptoms include **shortness of breath at rest**, respiratory distress - **Respiratory rate ≥ 30 breaths/minute**, **heart rate ≥ 125 beats/minute**, **oxygen saturation ≤ 93%**, or **arterial partial pressure of oxygen to fraction of inspired oxygen ratio < 300 mmHg**	**- Shortness of breath at rest**, respiratory distress **- Respiratory rate ≥ 30**, **oxygen saturation ≤ 93%**, or **arterial partial pressure of oxygen to fraction of inspired oxygen ratio < 300 mmHg**
Critical	- At least one of the following: (1) **Respiratory failure** (requiring mechanical ventilation, high-flow nasal cannula ≥ 20 L/min with oxygen ≥ 50%, noninvasive ventilation, ECMO), (2) **Shock** (systolic blood pressure < 90 mmHg or diastolic < 60 mmHg, or requiring vasopressors), or (3) **Multi-organ dysfunction/failure**	- Defined by **respiratory failure**, **shock**, or **multi-organ failure** - Patients at this stage are at increased risk of death and complications (e.g., arrhythmias, kidney injury, thromboembolism, septic shock)

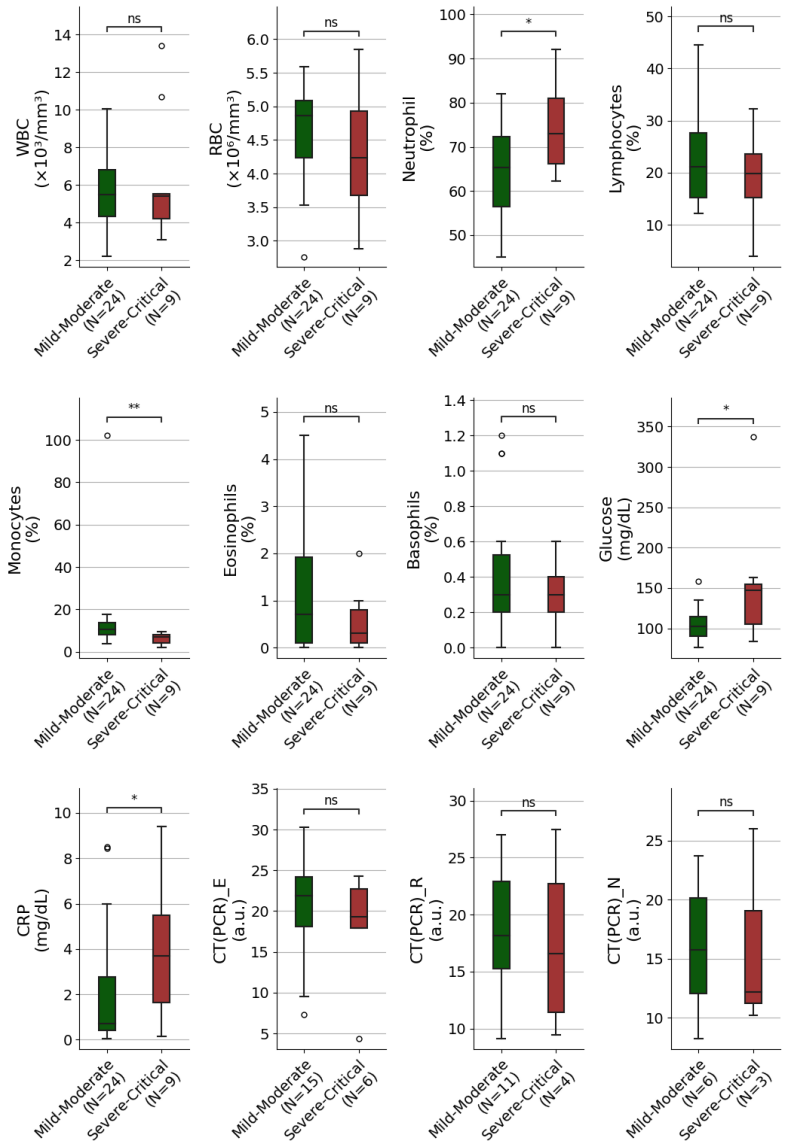
Table R1. Comparison of COVID-19 Severity Classification Criteria between FDA Guidance and Toussi *et al.* (Nat Microbiol, 2023)

Comment 5:

Please clearly explain the abbreviation used in figures and tables, especially the y-axis names for supplementary figure 1. Please explain the definitions of the **"box" and the "bar"** of the Box plots.

Response : Thanks for the suggestion.We have carefully reviewed all figures and tables to ensure that all abbreviations are clearly explained, either within the figures or in the figure legends. We have modified Supplementary Figure 1 and elaborates its figure legend appropriately, including the respective names of the y-axis and abbreviations. In addition, we have explicitly defined the elements of the box plots in the legend: the box represents the interquartile range (IQR), the line within the box

indicates the median, and the whiskers extend to the minimum and maximum values within 1.5 × IQR.



Supplementary Fig 1. Clinical Characteristics of Mild–Moderate and Severe–Critical COVID-19 Patients During the Acute Phase. The x-axis denotes clinical severity groups, with sample sizes indicated in parentheses. The y-axis indicates the clinical parameter measured, with units denoted in parentheses. Box plots display the median (centre line), interquartile range (box limits), and 1.5× interquartile range (whiskers). Individual data points beyond this range are shown as outliers (open circles). Significance thresholds are indicated as follows: ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). Abbreviations: WBC=White Blood Cell count; RBC=Red Blood Cell count; CRP=C-reactive protein concentration; CT(PCR)=Cycle threshold (Ct) values from PCR assays for SARS-CoV-2 targets (E, R, and N genes); a.u.=arbitrary units.

Reviewer #2

Results

Comment 1 :

The characterization of patient groups (**mild, severe, convalescent**) lacks clarity. The convalescent group requires further elaboration regarding its **classification**, **treatment status**, and potential **confounders**. Additionally, **demographic details (e.g., smoking status, medication use)** should be included in Table 2 to enhance the cohort's description.

Response : Thank you for raising this important point. The convalescent group refers to individuals who had fully recovered from COVID-19 and provided a blood sample between 4 and 12 weeks post-infection. Unlike the hospitalized group, blood collection for this group did not occur during hospitalization, which limited our ability to obtain detailed clinical and demographic information such as medication use at the time of infection. However, we have included all available information for this group in the updated Table 2 including smoking status(current, former, never). In addition, we have expanded Table 2 to include demographic variables for the hospitalized COVID-19 group(mild-moderate and severe-critical) as well, to provide a more comprehensive overview of the cohort characteristics.

	COVID-19 Hospitalized patients		**Convalescent (n=90)**	**Healthy controls (n=310)**
	Mild-Moderate group (n=37)	**Severe-Critical group (n=9)**		
Age, mean(min, max)	44(19,71)	50(24,66)	42(19,66)	42(19, 66)
Sex , n(%)				
Male	24(64.9)	4(44.4)	45(50)	188(60)
Female	13(35.1)	5(55.6)	45(50)	122(40)

BMI	24.8	25.5	24.4	23.45
Smoking status, n(%)				
Current	3(8.1)	3(33.3)		46(14.8)
Former	3(8.1)	1(11.1)		79(25.5)
Never	31(83.8)	5(55.6)		173(55.8)
NA	0	0	90	12(3.9)
Oxygen therapy, n(%)				
Yes				
Nasal prong only	8(21.6)	5(55.6)		
HFNC* only	0	2(22.2)		
Both (Nasal prong and HFNC*)	0	2(22.2)		
No	26(70.3)	0		
Intensive care unit, n(%)				
Yes	0 (0)	3(33.3)		
No	37(100)	6(66.7)		
Use of antiviral drugs, n(%)				
Yes				
Regdanvimab	3(8.1)	2(22.2)		
Remdesivir	10(27)	6(66.7)		
No	24(64.9)	1(11.1)		
Use of steroid, n(%)				
Yes	11(29.7)	7(77.8)		
No	26(70.3)	2(22.2)		
Pneumonia, n(%)				
Yes	10 (27)	7(77.8)		

No	27 (73.0)	2(22.2)		
ARDS, n(%)				
Yes	0 (0)	2(22.2)		
No	37(100)	7(77.8)		
Vaccination status, n(%)				
Yes	13(35.1)	6(66.7)		
No	24(64.9)	3(33.3)		
CCI, n(%)				
0-1	24(64.9)	1(11.1)		
2	4(10.8)	4(44.4)		
≥ 3	8(21.6)	3(33.3)		

Table 2. Baseline Characteristics of COVID-19 Hospitalized Patients, Convalescent Patients, and Non-infected Healthy Individuals. Demographic and clinical data are summarized for the clinical cohorts. Continuous variables (e.g., age and BMI) are presented as mean values with minimum and maximum ranges in parentheses. Categorical variables (e.g., sex, smoking status, and CCI) are expressed as counts with corresponding percentages. Smoking status was self-reported and categorized as current, former, or never smoker. Abbreviations: HFNC, high-flow nasal cannula; ARDS, acute respiratory distress syndrome; CCI, Charlson Comorbidity Index.

Comment 2 :

In Table 2, which presents baseline characteristics, the convalescent group appears to be missing. This makes it challenging to understand the distinction between the convalescent group, mild (recovery), and severe (recovery) groups. Providing clarity here would greatly enhance the manuscript.

Response : Thank you for highlighting this important point. In our study, we utilized two independent COVID-19 cohorts with distinct roles in the analysis. The first cohort comprised hospitalized COVID-19 patients, from which the mild and severe groups were derived. Patient classification in this cohort followed the U.S. FDA's *Guidance for Industry: COVID-19: Developing Drugs and*

Biological Products for Treatment or Prevention (May 2023), which defines four clinical categories : mild, moderate, severe, critical. Based on these definitions, we grouped patients into mild-moderate and severe-critical categories and used these groups for the *cis*-eQTM analysis. The second cohort consisted of COVID-19 convalescent individuals, defined as patients who had recovered from COVID-19 infection 4 to 12 weeks post symptom onset. These individuals had also been hospitalized during their acute phase but had fully recovered at the time of sample collection. This convalescent cohort was not used for biomarker discovery, but rather to assess the profiles of severity-associated markers identified from the hospitalized cohort in a recovery phase population.

To address the reviewer's concern, we have updated the Method section to provide clearer explanation of the convalescent cohort. In addition, Table 2 has been updated to include additional demographic and clinical information such as current smoking status, medication use, comorbidity score, to better characterize the patient cohorts and account for potential confounding factors.

[Page 16, lines 324-329 (Revised manuscript)]: For the hospitalized patient group, whole blood samples were collected twice : once during the acute phase(at the time of hospital admission) and once during the recovery phase(2 to 3 weeks after hospitalization). For the convalescent patient group, whole blood samples were obtained between 4 to 12 weeks after COVID-19 diagnosis. These individuals had been previously hospitalized due to COVID-19 but had fully recovered at the time of sample collection.

Comment 3 :

The study does not provide sufficient information on how potential confounding factors, such as medication use in acute COVID-19 cases, were accounted for. Some **medications may influence DNA methylation patterns**, which should be discussed.

Response : We fully agree with the reviewer that certain medications may influence DNA methylation patterns. Based on the clinical records, we confirmed that patients received antiviral and steroid medications during hospitalization, and we have included this information in the updated baseline characteristics table (Table 2). However, we would like to clarify that all blood samples used for DNA methylation and gene expression profiling were collected prior to the administration of any medication. Therefore, the omics data analyzed in this study are unlikely to reflect drug-induced epigenetic changes. We appreciate the reviewer's insightful suggestion, and we agree that analyzing post-treatment samples would be valuable for investigating the potential impact of medications on DNA methylation. If such samples become available, this will be an important direction for future research.

Comment 4 :

The inclusion of **cytokine measurements** would enhance the study, given the focus on immune response and cytokine storm mechanisms. If cytokine data were not collected, this should be stated explicitly as a **limitation**.

Response : We fully agree that such data would strengthen the interpretation of immune responses and cytokine-related mechanisms in COVID-19. However, cytokine measurements were not collected as part of our study design. We have now clearly stated this as a limitation in the Discussion section of the revised manuscript.

[Page 15, lines 311-316 (Revised manuscript)]: Lastly, while cytokine measurements were not performed in this study, the absence of these data limits our ability to directly evaluate the contribution of cytokine-mediated inflammatory responses to disease severity. Given the established role of cytokine storms in severe COVID-19 cases, future studies integrating cytokine profiling with

epigenetic and transcriptomic data would offer a more comprehensive understanding of immune dysregulation in acute infection.

Comment 5 :

The manuscript provides limited insight into **clinical outcomes**. Although it stratifies patients based on severity, it does not evaluate associations with **long-term recovery** markers such as **residual symptoms, lung function, or chronic fatigue.** These outcomes would strengthen the study's clinical relevance.

Response : We appreciate your valuable insight regarding the importance of evaluating long-term recovery markers such as residual symptoms, lung function, and chronic fatigue. In the current study, our primary focus was on the acute phase of COVID-19 and the immediate differences in clinical severity. Consequently, we did not systematically collect data on long-term outcomes like residual symptoms, detailed pulmonary function tests, or chronic fatigue indices. We acknowledge that incorporating such long-term recovery outcomes would significantly enhance the clinical relevance of our findings. We plan to address this limitation in future investigations by incorporating longer follow-up periods and more detailed clinical assessments, which will allow us to explore these critical aspects of post-COVID-19 recovery in greater depth.

Comment 6 :

page 6, line 124, the phrasing of the sentence "**Out of the twelve DEGs regulated their expression by the DMCs**" could be improved for clarity. Consider rephrasing this for better readability.

Response : We agree that the sentence needs better readability and have rephrased the sentence. In addition, since our revised DEGs now include 16 genes (from 18 genes), we now have ten genes of no known association to COVID-19. Only one of them had no significant difference between SC and MM in expression.

[Page 7, lines 146-148 (Revised manuscript)]: Among ten DEGs with no past association with COVID-19 severity, nine genes displayed significant expression changes between MM and SC (Figure S4).

Methods

Comment 7 :

The study does not mention whether DNA methylation data were adjusted for **cellular composition**. Given the known impact of acute infection on immune cell populations (e.g., neutrophilia, T-cell depletion), the results may reflect changes in cell proportions rather than true epigenetic modifications. The use of methylation deconvolution algorithms should be considered (e.g., Accomando et al., 2014; Kaushal et al., 2017). During acute phases of **infection neutrophilia or neutropenia** are reported. Likewise T-cell and monocyte cell counts vary during acute infection and recovery / convalence.

These changes affect large proportions of cells and easily may be read as expression or methylation changes. For example **IL18R** is expressed by monocytes, B and T cells and NK cells but **not by neutrophils**. **High neutrophil counts** are confirmed in your supplemental table Fig 1.

Response :

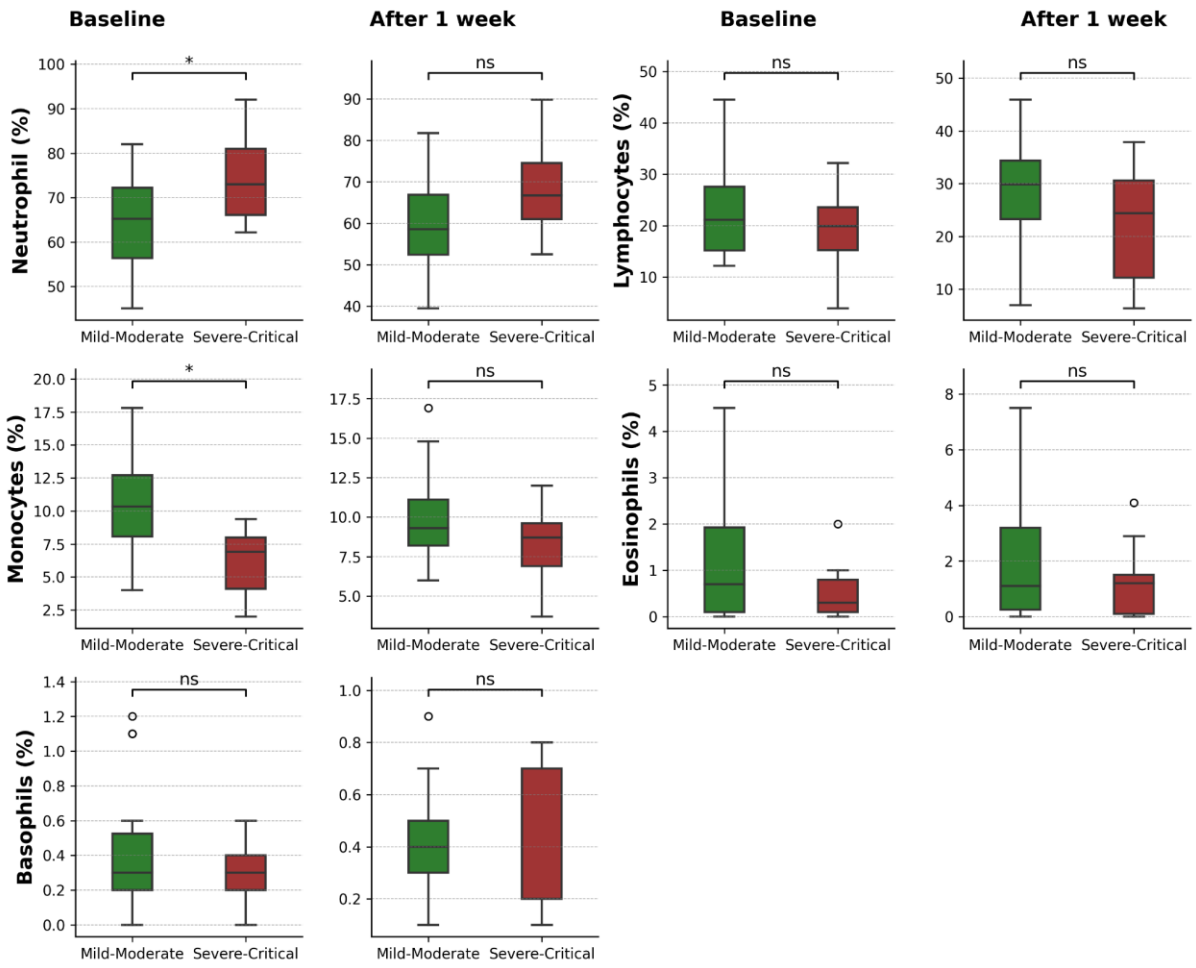
We appreciate the reviewer's insightful comment. As suggested, we examined whether immune cell proportions during the acute phase of infection influenced DNA methylation signals by referring to clinical white blood cell differential counts. At baseline, among the five measured leukocyte types (neutrophils, lymphocytes, monocytes, eosinophils, and basophils), we observed significant differences in neutrophil and monocyte levels between the severe and mild groups. However, these differences were no longer significant one week later, suggesting their transient nature. Based on these findings, we reanalyzed the methylation data by including baseline clinical leukocyte counts as covariates in the model. This adjustment resulted in the loss of previously identified CpG markers, with only a single site—chr6:36697843, associated with *FKBP5*—remaining significant (Table S4).

These results have been incorporated into the revised manuscript (Results section, page 7, *Leukocyte-adjusted cis-eQTM analysis reveals FKBP5 methylation as a cell-type proportion-independent marker of severe COVID-19*) and are presented in Figure 2C,D, Figure S5, and Table S4. Additionally, we expanded the 'Discussion' to reflect the potential role of *FKBP5* methylation at the intersection of stress and inflammation in the context of severe COVID-19.

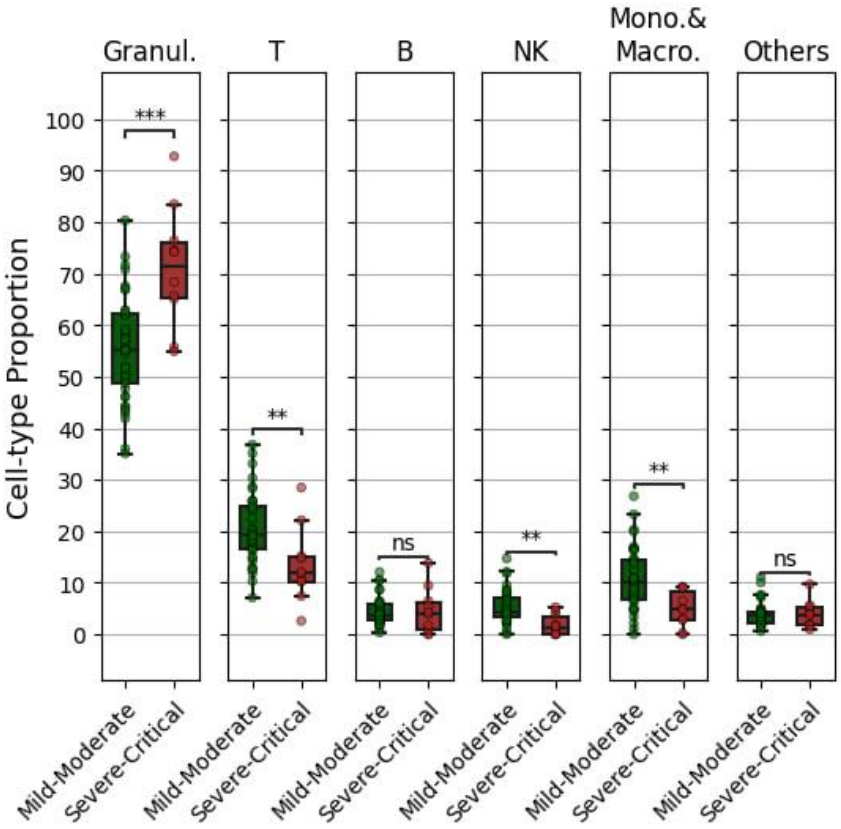
Furthermore, to validate our findings using an independent approach, we performed cell-type deconvolution using the UXM algorithm(Loyfer, Netanel, *et al*. "A DNA methylation atlas of normal human cell types." *Nature* 613.7943 (2023): 355-364). This analysis focused specifically on blood-derived immune cells. Among the five deconvoluted cell types, we found that granulocytes, T cells, NK cells, and monocytes/macrophages differed significantly in proportion between the severe and mild groups, while the 'Blood-B' cell population did not (Revision Fig 1). All remaining minor cell types were grouped as "others" and showed no statistically significant differences. These findings further support that the immune landscape in severe COVID-19 is markedly altered and may confound epigenetic signals if not properly accounted for.

Direction of methylation	Position	Gene symbol	Direction of mRNA expression	Correlation methylation with mRNA expression	
				Spearman's ρ	P-value
Hypomethylation	chr6:36697843	*FKBP5*	Up regulation	-0.55129	0.00879

Supplementary Table 4. Key genes associated with the severity of COVID-19 in the acute phase after adjustment for leukocyte composition



Supplementary Fig 5. Baseline leukocyte proportion and its changes one week later in Mild-Moderate and Severe-Critical groups. Boxplots show the proportions (%) of neutrophils, lymphocytes, monocytes, eosinophils, and basophils measured in whole blood at baseline (left panel) and after a week later (right panel) for patients with Mild–Moderate (green) and Severe–Critical (red) disease. Boxplots display the median (centre line), IQR (box limits), and 1.5×IQR (whiskers), and outliers (open circles). Statistical comparisons were conducted using the Wilcoxon rank-sum test. Statistical significance is denoted as follows: ns ($P \geq 0.05$), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), **** ($P < 0.0001$).

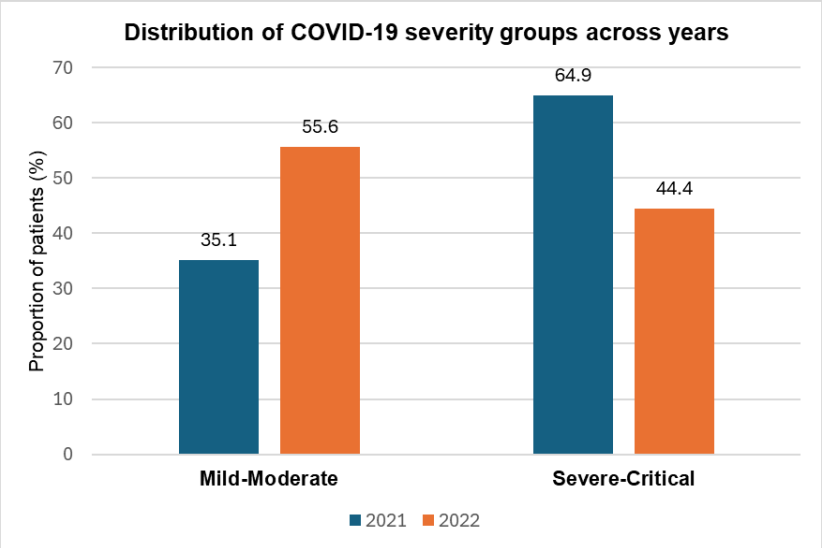


Revision Fig 1. Immune cell-type proportion shifts in COVID-19 patients stratified by clinical severity. Estimated immune cell-type proportions derived from whole blood DNA methylation profiles of COVID-19 patients with mild-to-moderate (green) or severe-to-critical (red) disease. Cell-type deconvolution was performed using a fragment-level reference-based method (UXM algorithm), categorizing cells into following classes: granulocytes (Granul.), T cells (T), B cells (B), natural killer cells (NK), monocytes and macrophages (Mono. & Macro.), and other minor cell populations (Others). Each dot represents an individual sample. Box plots display the median and interquartile range (IQR), with whiskers extending to 1.5× IQR. Statistical comparisons between severity groups were performed using two-sided unpaired t-tests. Significance is denoted as follows: n.s. ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$).

Comment 8 :

The manuscript acknowledges the impact of SARS-CoV-2 variants (Delta vs. Omicron) but does not attempt to **stratify** patients accordingly. Given that the mild-moderate group includes both **Delta- and Omicron-**infected patients, the authors should explore stratification or discuss the feasibility of such an analysis.

Response : Thank you for pointing this out. Our patient samples were collected between 2021 and 2022, during which distinct SARS-CoV-2 variants were predominant in South Korea. Specifically, the Delta variant was the major circulating strain in 2021, while the Omicron variant became dominant in 2022[1,2]. However, in Korea, the viral variant type is typically not determined for each patient sample as part of routing practice, so we do not have direct information on each sample's variant. Although we intended to perform stratified analysis based on variant type, the number of available samples was insufficient to allow for reliable classification. Instead, we examined the collection years of the mild-moderate and severe-critical groups to assess whether the distribution of samples was biased toward a particular year, which could reflect the influence of specific variants. As shown in Revision Fig 2, the samples are evenly distributed across both years without skew toward a particular year. This suggests that the variant effect is unlikely to have systematically biased the grouping. Nevertheless, we acknowledge that biological differences between SARS-CoV-2 variants could still potentially influence immune responses and epigenetic patterns, and future studies with precise variant identification would help address this limitation more directly.



Revision Fig2. Proportion of samples by year in the Mild-Moderate(MM) and Severe-Critical(SC) groups.

References

1. Lee, Dong-Wook, *et al.* "Genomic epidemiology of SARS-CoV-2 Omicron variants in the Republic of Korea." *Scientific Reports* 12.1 (2022): 22414.
2. Lee, Kyung-Shin, *et al*. "Risk factors for critical COVID-19 illness during Delta-and Omicron-predominant period in Korea; using K-COV-N cohort in the National health insurance service." *Plos one* 19.3 (2024): e0300306.

Comment 9 :

The classification of patient groups could be improved. The term "mild" currently subsumes mild and moderate cases. To avoid confusion, the manuscript should consistently refer to this group as **"mild-moderate" (MM)** in all sections.

Response : Thank you for your suggestion. As per your recommendation, we have revised the terminology throughout the manuscript to ensure clearer classification. Specifically, we have changed 'Mild' to 'Mild-Moderate(MM)' and 'Severe' to 'Severe-Critical(SC)'.

Comment 10 :

The statistical approach requires further justification. The manuscript states that a **two-sample t-test** was used despite a small and unequal sample size (n=10 severe vs. n=41 mild). The authors should clarify whether **assumptions of normality were tested** and, if necessary, replace **t-tests with non-parametric alternatives**.

Response : Thanks for the question. Indeed, checking appropriate statistical tests is important. We have considered both parametric and non-parametric measures to test significance on both Methylation and

RNA data according to their normality to maximize the power of the statistical experiments. We have thoroughly checked for the normality assumption with the Shaprio-Wilk test as discussed in the Methods (Methods: **Statistical analysis and Visualization**). For the Methylation data, >90% of the distributions in the test had no evidence of significant deviation from the normal distribution; therefore, we proceeded with Welch's *t*-test. In contrast, RNA data did not exhibit normal distributions; hence, we conducted Wilcoxon rank-sum tests. We now provide the full statistics of normality tests in **Table S6, 7**.

CpGs	Healthy	Mild-Moderate	Severe-Critical	Convalescent
chr1:25771417	0.281	0.006	0.965	0.468
chr1:234612121	0.382	0.382	0.382	0.726
chr15:90100633	0.303	0.334	0.303	0.303
chr16:30964865	0.877	0.877	0.877	0.281
chr16:75063094	**0.01**	**0.038**	**0.038**	0.772
chr17:17276112	0.606	0.665	0.606	0.872
chr17:81274041	0.977	0.37	0.511	0.511
chr17:82126907	0.085	0.085	0.653	0.653
chr17:82126935	0.678	0.818	0.678	0.678
chr17:82126939	0.703	0.703	0.703	0.703
chr17:82127016	0.493	0.633	0.633	0.633
chr17:82223145	0.061	0.061	0.572	0.643

chr17:82624041	0.023	0.724	0.724	0.672
chr17:82624115	0.024	0.636	0.636	0.612
chr17:82838556	0.876	0.933	0.876	0.876
chr17:82838607	0.762	0.762	0.762	0.762
chr18:79506961	**0.048**	0.204	0.204	0.204
chr19:1423674	0.386	0.838	0.357	0.386
chr19:1423696	0.262	0.791	0.791	0.908
chr2:37368137	0.647	0.324	0.647	0.647
chr2:102143316	0.048	0.975	0.846	0.846
chr20:330120	0.056	0.536	0.536	0.776
chr20:654151	**0.046**	**0.009**	0.077	0.359
chr20:1166154	0.809	0.646	0.646	0.646
chr20:1467583	0.345	0.712	0.408	0.712
chr20:1631771	0.746	0.746	0.746	0.746
chr5:75320838	0.397	0.397	0.397	0.397
chr6:36697843	0.377	0.961	0.332	0.332
chr6:108562564	0.678	0.678	0.678	0.678
chr8:143356214	0.571	0.849	0.571	0.854

Supplementary Table 6. Summary table displaying significance of Shapiro-Wilk tests for 30 DMC-LOOs. Each value represents an adjusted *P*-value of the normality test after Benjamini-Hochberg correction. Those values that have no evidence of normal distribution have been marked as bold font.

Genes	Healthy	Mild-Moderate	Severe-Critical	Convalescent

ENSG00000287737.3_ENSG00000287737	0.997	0.949	0.949	0.802
ENSG00000232973.14_CYP1B1-AS1	0.556	0.556	0.945	**0.002**
ENSG00000118689.15_FOXO3	0.559	0.559	0.06	**0.019**
ENSG00000154803.13_FLCN	0.79	0.309	**0.022**	0.79
ENSG00000115607.10_IL18RAP	0.853	0.385	0.437	0.385
ENSG00000187010.21_RHD	0.002	0.347	**0.003**	0.132
ENSG00000090238.12_YPEL3	0.663	0.663	0.138	**0.034**
ENSG00000140564.13_FURIN	0.608	0.608	0.055	0.364
ENSG00000096060.15_FKBP5	0.634	0.318	0.634	0.318
ENSG00000271303.2_SRXN1	0.059	0.254	0.114	0.059
ENSG00000176928.7_GCNT4	**0.028**	**0.02**	**0.002**	**0.02**
ENSG00000198858.10_R3HDM4	0.921	0.921	0.136	**0.05**
ENSG00000122490.19_SLC66A2	0.864	0.864	0.063	**0.009**
ENSG00000178719.17_GRINA	0.741	0.793	0.524	0.741
ENSG00000034713.8_GABARAPL2	0.539	0.539	**0.011**	**0.032**
ENSG00000289377.2_ENSG00000289377	**0.004**	**0.025**	**0.002**	**0.037**

Supplementary Table 7. Summary table displaying significance of Shapiro-Wilk tests for 16 DEG-LOOs. Each value represents an adjusted *P*-value of the normality test after Benjamini-Hochberg correction. Those values that have no evidence of normal distribution have been marked as bold font.

Comment 11 :

Technical details require improvement. The description of the sequencing process includes unclear phrasing (e.g., "Sequenced DNA reads were removed their adapters"). Additionally, **product codes** for kits and reagents should be provided for reproducibility.

Response : We agree the writing for the sequencing process needs improvement. It is not explicit in the initial manuscript but "*Sequenced DNA reads were removed their adapters*" should have described a bioinformatic process to trim adapters. We have rephrased and re-structured the relevant sections from beginning for better readability (Methods: **Target Bisulfite Sequencing and Methylation Quantification** & **mRNA Sequencing and Expression Quantification**). We also provide the **product codes** for kits and reagents for reproducibility.

[Page 19, lines 365-380 (Revised manuscript)]:

Target Bisulfite Sequencing and Methylation Quantification

Frozen whole blood samples collected in EDTA tubes (Becton Dickinson, #367856) were used for genomic DNA (gDNA) extraction, performed using the DNeasy Blood & Tissue Kit (Qiagen, #69506) according to the manufacturer's instructions.

Bisulfite-converted DNA libraries were prepared using the SureSelectXT Methyl-Seq Target Enrichment System Kit (Agilent, #G9651), following the manufacturer's protocol. Sequencing was performed on the NovaSeq 6000 platform (Illumina), generating 150 bp paired-end reads.

Raw sequencing reads were processed using Fastp (v0.23.1) to remove adapters and low-quality bases, with the following parameters: --cut_front 1 --cut_right 0 --cut_tail 1 --detect_adapter_for_pe --trim_front1 15 --trim_front2 15 --trim_tail1 0 --trim_tail2 0 --cut_mean_quality 20 --n_base_limit 1 --average_qual 20 34. Read quality was assessed before and after filtering reads using FastQC (v0.11.9) 35.

Filtered reads were aligned to the human reference genome (GRCh38.p13) using Bismark (v0.23.1), a bisulfite-aware aligner 36. Methylation calling and quantification were conducted using the methylKit

R package (v1.20.0) 37. To annotate CpG sites and associate them with proximal genes or regulatory features, we used the annotatr R package (v1.20.0) 38.

[Page 20, lines 382-403 (Revised manuscript)]:

mRNA Sequencing and Expression Quantification

Frozen whole blood samples collected in PAXgene® Blood RNA Tubes (PreAnalytiX, #762174) were used for total RNA extraction. RNA integrity and concentration were assessed using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, #Q32866) and the Qubit RNA HS Assay Kit (Thermo Fisher Scientific, #Q32854). For mRNA enrichment, 200 ng of total RNA per sample was processed using the Dynabeads mRNA Purification Kit (Thermo Fisher Scientific, #01152851), which depletes rRNA and isolates polyadenylated transcripts using oligo(dT) beads.

Library preparation was performed using the MGIEasy RNA Directional Library Prep Set (MGI, #MG1000006386), following the manufacturer's protocol. Fragment size distribution was checked using 4150 TapeStation (Agilent, #G2992A) with the cDNA D1000 ScreenTape (Agilent, #5067-5582). Final libraries were quantitated using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Q32866). RNA sequencing was run on DNBSEQ-T7RS (MGI) platform, generating 150 bp paired-end reads.

RNA-seq reads were trimmed and quality-filtered using Fastp (v0.23.1) with the same parameters as for DNA methylation processing. Read quality was verified using FastQC (v0.11.9).

Filtered reads were aligned to the human reference genome (GRCh38.p13) using STAR (v2.7.10b), a splice-aware aligner, with default settings 39. Gene-level quantification was performed using RSEM (v1.3.3) using default options 40, with gene annotations from GENCODE v42 (GFF3 format). Raw expression values were derived from the expected counts generated by RSEM.

Comment 12 :

Normalized gene expression was estimated using **RSEM**, while **DESeq2** was used in the figure. Could you please explain the rationale behind this apparent discrepancy?

Response : Thanks for the question. We have perhaps failed to clearly show how the normalization was performed on raw RNA counts. Indeed, we used the "expected count" provided by the *RSEM* program; however, we did not use *RSEM* for normalization as written in the initial manuscript. Instead, we harnessed the RLE method provided by *DESeq2* to adjust by sequencing depth and library composition. Following up from your point, we decided to apply VST (variance-stabilizing transformation) on the normalized count, which is effectively applying a log, giving us the fold change, instead of comparing the counts themselves. To avoid any further confusion, we emphasize that the normalized value was used exclusively in the pairwise comparison, as shown in Figure 2C/D. Raw counts from RSEM were used as the input for the DEG discovery, NOT the normalized count. (Materials and methods section : **mRNA Sequencing and Expression Quantification** & **Normalization of Gene Expression Data**)

[Page 20, lines 402 (Revised manuscript)]: Raw expression values were derived from the expected counts generated by RSEM.

[Page 21, lines 415-423 (Revised manuscript)]: To account for differences in sequencing depth and library composition, raw expected gene counts were normalized using the DESeq2 R package (v1.38.3)47. A variance stabilizing transformation (VST) was subsequently applied to the normalized counts. VST approximates a log2 transformation while decoupling mean-variance relationships of RNA-seq data, thereby improving the accuracy and interpretability of downstream statistical analyses.

Comment 13 :

On page 18, lines 318–319, kindly mention the **specific parameters** assessed. This additional detail would make this section more informative.

Response : Thanks for the question. We agree that it is a good idea to show specific parameters used for running *fastp* to remove adaptors and trim and discard low-quality reads. We have used the following parameters uniformly across all of the sequencing data we have used: *--cut_front 1 --cut_right 0 --cut_tail 1 --detect_adapter_for_pe --trim_front1 15 --trim_front2 15 --trim_tail1 0 --trim_tail2 0 --cut_mean_quality 20 --n_base_limit 1 --average_qual 20.*

[Page 19, lines 372-376 (Revised manuscript)]: Raw sequencing reads were processed using Fastp (v0.23.1) to remove adapters and low-quality bases, with the following parameters: --cut_front 1 --cut_right 0 --cut_tail 1 --detect_adapter_for_pe --trim_front1 15 --trim_front2 15 --trim_tail1 0 --trim_tail2 0 --cut_mean_quality 20 --n_base_limit 1 --average_qual 20. Read quality was assessed before and after filtering reads using FastQC (v0.11.9).

Comment 14 :

On page 19, line 340, the sentence starting with "We utilized the LOO..." is somewhat unclear. The phrasing "**iteratively collecting DEGs without a sample**" is particularly difficult to interpret. Do you mean that **one sample is excluded per round**? If so, does this mean a single sample is removed per group in each iteration? Providing a clearer explanation of this methodology would significantly improve the reader's understanding.

Response : Firstly, we want to notify the Reviewers that there have been a few adjustments in the revised DEG analysis. The number of iterations should have been **9 times** instead of 8 times, since there are a total of 9 Severe-Critical (SC) patients with RNA-seq data available. The last iteration was inadvertently missed due to a mistake. Therefore, we have rerun the analysis including all of nine samples, and the threshold from six to **seven overlaps** across the nine iterations. After the adjustments, the number of DEG-LOOs declined from 18 to **16 genes**, removing *IL18R1* and *LINC02217* from our shortlist of biomarkers. With the refined details elaborated above, we have modified the relevant Methods section as follows:

[Page 22, lines 445-455 (Revised manuscript)]: Differentially Expressed Genes (DEGs) associated with COVID-19 severity were identified using the DESeq2 R package (v1.38.3). Given the limited number of Severe-Critical (SC) cases (N=9), we employed a leave-one-out (LOO) cross-validation strategy to enhance the robustness of DEG discovery. In each iteration, one SC sample was excluded, and differential expression analysis was performed between the remaining SC and mild cases. This procedure was repeated nine times, each time omitting a different SC sample. Genes were considered significantly differentially expressed if they satisfied the following thresholds: absolute $\log_2$ fold change ($|\log_2 FoldChange|$) > 1.3 and false discovery rate (FDR) < 0.05. To define robust DEGs (DEG-LOOs), we retained only those genes that exhibited consistent directionality—either upregulation or downregulation in SC patients—in at least seven out of the nine independent LOO iterations.

Comment 15 :

On page 6, line 132, it is mentioned that an independent convalescent and **healthy group** was used. Providing more details about the demographics of this group would be valuable. Additionally, this sentence could be rephrased to enhance its clarity.

Response : Thanks for the suggestion. In short, we have extended Table 2 to include the demographic details of subjects in both convalescent and healthy groups. Additionally, we revised the original sentence(On page 6, line 132) to improve readability and incorporated it into a newly reorganized section of the Results, due to major structural updates made during the revision process.

[Page 10, lines 180-183 (Revised manuscript)]: Methylation and expression profiles were examined in the hospitalized MM and SC groups, an independent convalescent cohort (4–12 weeks post-infection, n = 90), and healthy controls (no viral infection at the time of collection, n = 344).

	COVID-19 Hospitalized patients		**Convalescent (n=90)**	**Healthy controls (n=310)**
	Mild-Moderate group (n=37)	**Severe-Critical group (n=9)**		
Age, mean(min, max)	44(19,71)	50(24,66)	42(19,66)	42(19, 66)
Sex , n(%)				
Male	24(64.9)	4(44.4)	45(50)	188(60)
Female	13(35.1)	5(55.6)	45(50)	122(40)
BMI	24.8	25.5	24.4	23.45
Smoking status, n(%)				
Current	3(8.1)	3(33.3)		46(14.8)
Former	3(8.1)	1(11.1)		79(25.5)
Never	31(83.8)	5(55.6)		173(55.8)
NA	0	0	90	12(3.9)
Oxygen therapy, n(%)				
Yes				

Nasal prong only	8(21.6)	5(55.6)		
HFNC* only	0	2(22.2)		
Both (Nasal prong and HFNC*)	0	2(22.2)		
No	26(70.3)	0		
Intensive care unit, n(%)				
Yes	0 (0)	3(33.3)		
No	37(100)	6(66.7)		
Use of antiviral drugs, n(%)				
Yes				
Regdanvimab	3(8.1)	2(22.2)		
Remdesivir	10(27)	6(66.7)		
No	24(64.9)	1(11.1)		
Use of steroid, n(%)				
Yes	11(29.7)	7(77.8)		
No	26(70.3)	2(22.2)		
Pneumonia, n(%)				
Yes	10 (27)	7(77.8)		
No	27 (73.0)	2(22.2)		
ARDS, n(%)				
Yes	0 (0)	2(22.2)		
No	37(100)	7(77.8)		
Vaccination status, n(%)				
Yes	13(35.1)	6(66.7)		
No	24(64.9)	3(33.3)		
CCI, n(%)				

0-1	24(64.9)	1(11.1)		
2	4(10.8)	4(44.4)		
≥ 3	8(21.6)	3(33.3)		

Table 2. Baseline Characteristics of COVID-19 Hospitalized Patients, Convalescent Patients, and Non-infected Healthy Individuals. Demographic and clinical data are summarized for the clinical cohorts. Continuous variables (e.g., age and BMI) are presented as mean values with minimum and maximum ranges in parentheses. Categorical variables (e.g., sex, smoking status, and CCI) are expressed as counts with corresponding percentages. Smoking status was self-reported and categorized as current, former, or never smoker. Abbreviations: HFNC, high-flow nasal cannula; ARDS, acute respiratory distress syndrome; CCI, Charlson Comorbidity Index.

Comment 16 :

Include the product codes for all kits and materials used in the study. This will improve reproducibility and clarity for readers.

Response : Thanks for pointing this out. We now provide the product codes for kits and reagents for reproducibility, which are written in detail in relevant sections (Materials and methods section: **Target Bisulfite Sequencing and Methylation Quantification** & **mRNA Sequencing and Expression Quantification**).

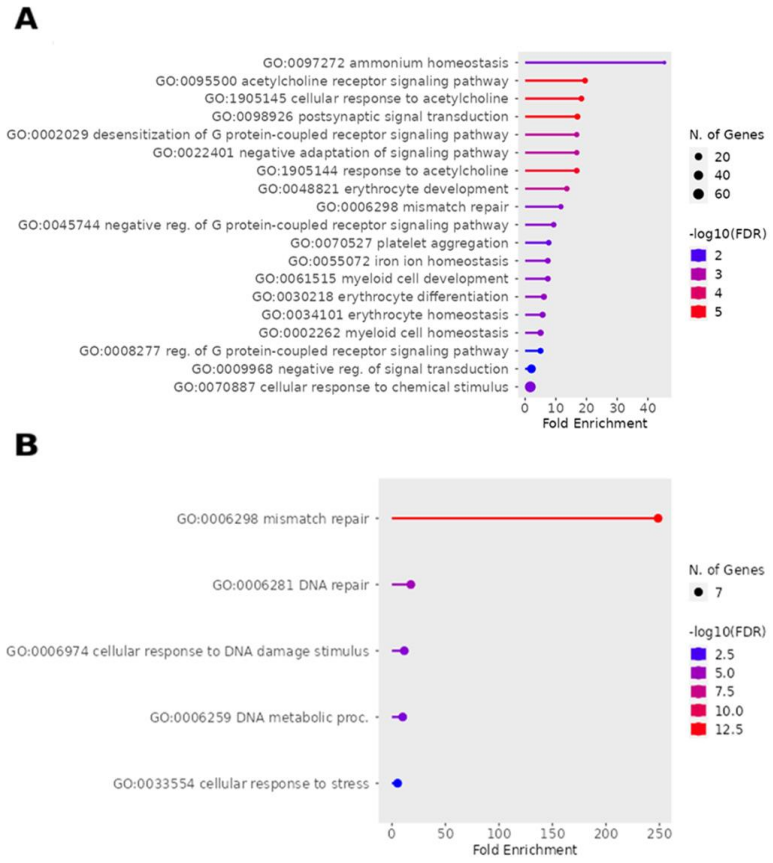
Comment 17 :

Have the authors considered conducting a **pathway enrichment analysis** for the identified DEGs and DMGs? Including the results of such an analysis would provide valuable insights into the **biological processes and pathways** implicated in the study.

Response : Thanks for the suggestion. We now provide the results from the pathway enrichment analysis for both the identified DEGs in our revised manuscript as Supplementary Data. The enrichment results reveal the overall, global molecular changes in the Severe-to-Critical (SC) patients as compared

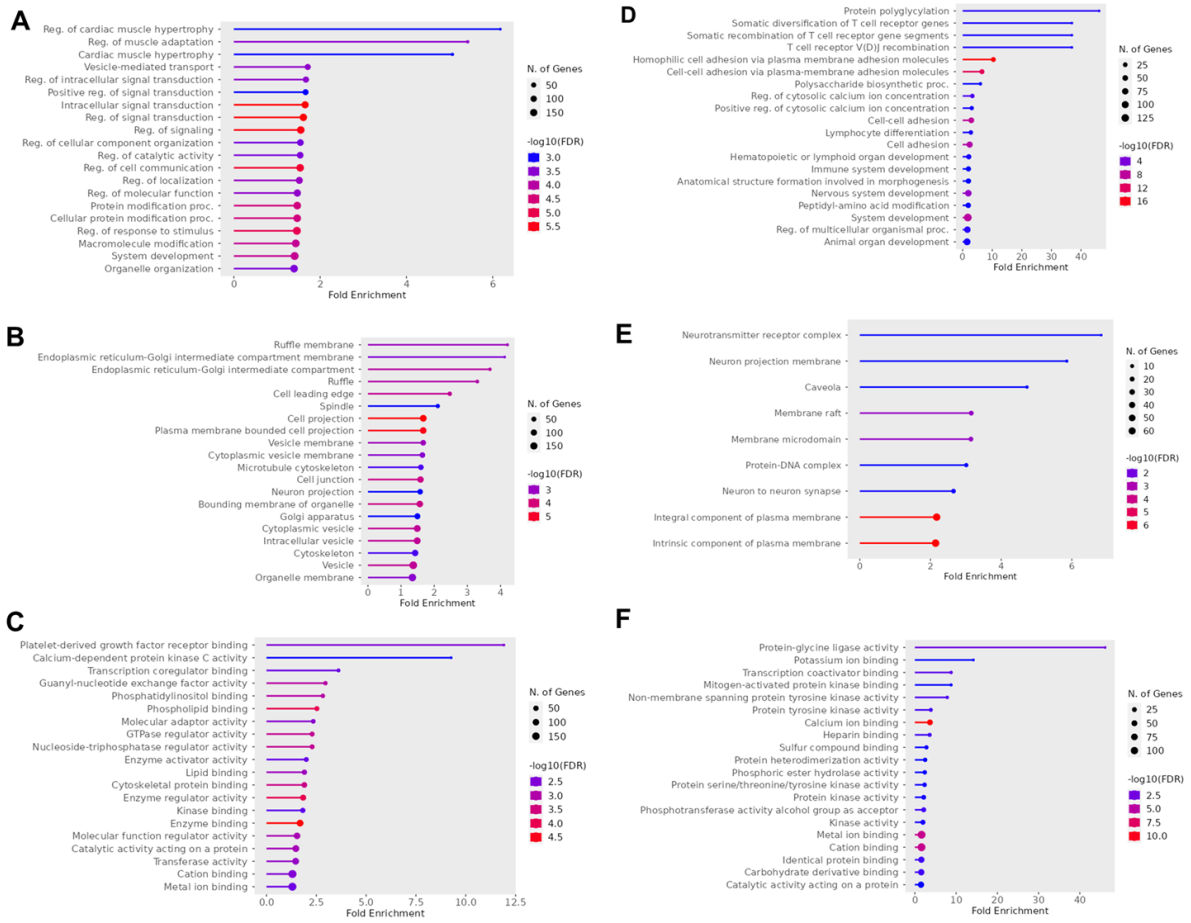
to Mild-to-Moderate (MM) cases. We have included the corresponding methods and results in relevant sections, respectively (Methods: **Gene Ontology Enrichment of DEGs**).

In addition to the DEGs, we performed pathway enrichment analysis for the 1944 initially identified DMCs (hypomethylated CpG : 1296; hypermethylation CpG : 648). As shown in Revision Fig 3, the enrichment results did not reveal any pathways directly associated with COVID-19 severity in either the hypo- or hypermethylated markers. This lack of clear functional association motivated us to pursue a more targeted approach by conducting eQTM analysis. This integrative analysis enabled us to prioritize functionally relevant markers and gain better insight into severity-specific molecular alterations.



Supplementary Fig 2. Gene Ontology Enrichment of 297 Severity-Associated DEG-LOOs During Acute COVID-19. Lollipop plot illustrating Gene Ontology-Biological Process (GO-BP) enrichment for **A)** 283 up-regulated and **B)** 14 down-regulated genes linked to COVID-19 severity during the acute phase. The x-axis denotes fold enrichment, while the y-axis lists enriched GO terms. Dot size corresponds to the number of genes in each GO term, and color indicates statistical significance as –log10(FDR). Abbreviations: GO, Gene Ontology; BP, Biological Process; FDR, False Discovery Rate.

[Page 5, lines 111-115 (Revised manuscript)]: The up-regulated genes were enriched in pathways related to erythrocyte dynamics, innate immune activation, platelet aggregation, and acetylcholine receptor signaling (Figure S2A), whereas down-regulated genes were predominantly involved in DNA mismatch repair (MMR) pathway (Figure S2B).



Revision Fig 3. Gene ontology enrichment of COVID-19 severity-associated DMC-LOOs during acute COVID-19. Lollipop plot illustrating Gene Ontology(GO) enrichment analysis. The analysis was performed separately for genes annotated from hypomethylated (N=908) and hypermethylated DMCs (N=524), based on gene mapping using annotatr. A) GO-BP for hypomethylated genes. B) GO-CC for hypomethylated genes. C) GO-MF for hypomethylated genes. D) GO-BP for hypermethylated genes. E) GO-CC for hypermethylated genes. F) GO-MF for hypermethylated genes. Dot size corresponds to the number of genes in each GO term, and color indicates statistical significance as –log10(FDR). Abbreviations: GO, Gene Ontology; BP, Biological Process; CC, Cellular Component; MF, Molecular Function; FDR, False Discovery Rate.

Discussion

Comment 18 :

The discussion should begin with a **summary** of key findings rather than limitations.

Response : Thank you for the suggestion. We have restructured our discussion to begin with the summary of key findings and limitations at the end in our revised manuscript.

[Page 11, lines 208-221 (Revised manuscript)]: Using early-phase whole-blood multi-omics from clinically well-characterized COVID-19 patients, we mapped CpG-to-transcript links with cis-eQTM analysis and uncovered a focused set of "reactive" CpGs whose methylation shifts reflected both disease severity and subsequent recovery. During the acute phase, coordinated hypo- or hyper-methylation at these sites corresponded to sharp transcriptional changes that distinguished severe-critical from mild-moderate cases. After correcting for transient leukocyte imbalances, only one CpG–RNA pair, FKBP5, remained independently associated with severity, highlighting it as a core epigenetic signal rather than a by-product of immune-cell redistribution. Longitudinal follow-up demonstrated that both methylation and gene expression at the reactive loci returned to baseline within 4–12 weeks, underscoring the dynamic and largely reversible nature of COVID-19-induced epigenetic programming. These findings establish a compact, cell-type composition–independent CpG/RNA signature that captures the trajectory from acute pathology to convalescence, providing possible prognostic biomarkers with mechanistic insight pertaining to host response and recovery during COVID-19.

Comment 19 :

The section "Key Genes and Molecular Mechanisms Influencing COVID-19 Severity" should be integrated into the discussion, rather than presented as a **standalone** section.

Response : Thank you for the suggestion. We have integrated the section "Key Genes and Molecular Mechanisms Influencing COVID-19 Severity" into the Discussion to provide a more cohesive interpretation of the findings within the context of the overall study.

[Page 12, lines 230-255 (Revised manuscript)]: This study is among the first to apply cis-eQTM analysis for dissecting gene regulatory programs linked to COVID-19 severity. We identified a distinct pattern of hypomethylation and concurrent upregulation of six genes—*SRXN1, FURIN, IL18RAP, FOXO3, GCNT4*, and *FKBP5*—involved in key processes, including host cell invasion, oxidative stress response, interleukin-18-mediated signaling, and lung fibrosis. These five genes are not only epigenetically regulated but also biologically implicated in known pathogenic mechanisms which are recognized as drivers of COVID-19 severity. Notably, SRXN1, which was hypomethylated and up-regulated in severe cases across five distinct CpG loci, encodes Sulfiredoxin 1, a critical antioxidant mitigating reactive oxygen species (ROS) damage. Its dysregulation amplifies oxidative stress and has been repeatedly reported to be highly expressed in both blood and lung tissues of severe COVID-19 patients [16-20]. FURIN, another gene up-regulated in severe cases, facilitates viral entry by cleaving the S1/S2 junction of the SARS-CoV-2 spike protein21. Its expression correlates with disease progression and has been identified as a potential target for antiviral therapy[22,23]. Our results also revealed significant upregulation of *IL18RAP* and *FOXO3* in severe COVID-19 cases—genes that are mechanistically linked to hyper-inflammatory responses and respiratory dysfunction. *IL18RAP*, a key mediator of IL-18 signaling, likely contributes to excessive immune activation and cardiopulmonary complications through inflammasome-driven IL-18 release, a pathway implicated in macrophage activation syndrome and multi-organ failure[24-26]. In parallel, *FOXO3*, which regulates immune balance and oxidative stress, was associated with increased oxygen demand and inflammatory lung injury, suggesting that its overexpression may exacerbate pulmonary damage in severe disease[27-29]. Interestingly, *GCNT4* was the only gene found to be epigenetically downregulated in severe patients. A prior proteomic study reported its elevation

in respiratory failure among severe COVID-19 cases[30]. This discrepancy may reflect differences in regulatory control between the transcriptome and proteome or patient cohort characteristics.

Comment 20 :

The manuscript acknowledges the impact of SARS-CoV-2 variants (Delta vs. Omicron) but does not attempt to stratify patients accordingly. Given that the mild-moderate group includes both Delta- and Omicron-infected patients, the authors should explore stratification or discuss the feasibility of such an analysis.

Response : We fully agree with your comment. As noted in our response to Comment 8, both mild-moderate and severe-critical patients were enrolled across 2021 and 2022, which correspond to the periods when the Delta and Omicron variants, respectively, were predominant in South Korea. However, since we did not directly determine the variant type for each patient, we cannot definitively classify individual by variant. Furthermore, the number of samples in each group was insufficient to perform meaningful stratified analysis (2021 samples: MM group, n = 13; SC group, n = 5; 2022 samples: MM group, n = 24; SC group, n = 4). We have acknowledged this limitation in the Discussion section.

[Page 15, lines 311-316 (Revised manuscript)]: Lastly, while cytokine measurements were not performed in this study, the absence of these data limits our ability to directly evaluate the contribution of cytokine-mediated inflammatory responses to disease severity. Given the established role of cytokine storms in severe COVID-19 cases, future studies integrating cytokine profiling with epigenetic and transcriptomic data would offer a more comprehensive understanding of immune dysregulation in acute infection.

Comment 21 :

The potential role of identified genes in **long-term COVID-19** outcomes is not explored. Since the study mentions rapid epigenetic recovery, it would be valuable to discuss whether **residual methylation changes** could contribute to **post-acute sequelae**.

Response : We appreciate this important point. We agree that investigating the potential contribution of residual methylation changes to long-term COVID-19 outcomes is a valuable direction. We observed that eQTM genes and CpGs of severity generally returned to healthy-like patterns during convalescence, regardless of initial disease severity. However, a subset of molecular features showed delayed normalization. While these may hold potential relevance for post-acute sequelae, our current cohort does not include individuals with clinically confirmed long-term symptoms. Therefore, we are unable to directly assess whether these residual molecular changes are linked to long COVID or other persistent outcomes. Future studies involving patients with documented post-acute sequelae will be essential to address this important question.

Figures and Supplementary Information

Comment 22 :

Figure 1B lacks a clear explanation of how patients were classified into mild, severe, and convalescent groups. The figure legend should explicitly define each group.

Response : Thank you for your helpful comment. In response, we have revised Figure 1 to clarify the classification of patient groups. Specifically, the 'mild' and 'severe' groups correspond to individuals in the 'Hospitalized COVID-19 patient' group, whereas the 'Convalescent group' comprises a separate set of recovered COVID-19 individuals who were not part of the hospitalized group. We have revised a visual representation of these groups along with the 'Healthy controls' group in Figure 1A. Additionally, we have updated figure legend to include explicit definitions of each group.

Furthermore, we would like to clarify that Figure 1B depicts the follow-up duration specifically for patients within the Hospitalized COVID-19 patient group, separated into Mild-Moderate(MM) and Severe-Critical(SC) subgroups. This explanation has also been added to the revised figure legend.

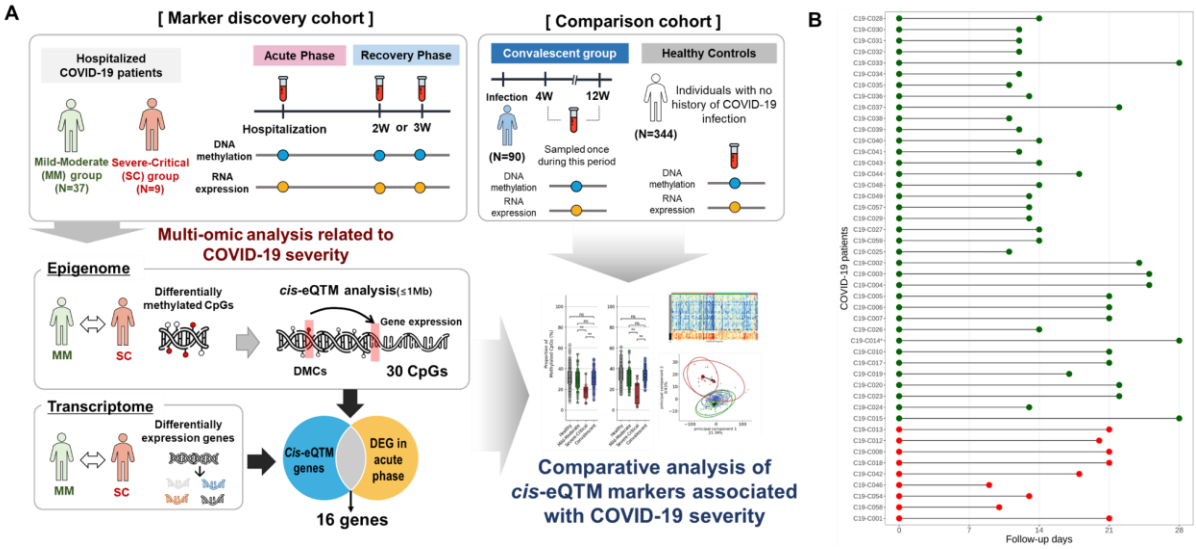


Figure 1. Overall study design and follow-up blood sample collection for *cis*-eQTM analysis in COVID-19 patients. A) Overview of blood sampling and analysis design; The marker discovery cohort consisted of hospitalized COVID-19 patients, stratified into Mild-Moderate (MM) and Severe-Critical (SC) groups, and was used to identify severity-associated *cis*-eQTM markers. The comparison cohort was composed of two independent reference groups : the 'Convalescent group' comprising individuals who recovered from COVID-19 and provided a single blood sample 4 to 12 weeks post-infection, and the 'Healthy Controls' group, consisting of pre-pandemic individuals with no history of SARS-CoV-2 infection. **B)** Clinical follow-up duration of 46 Hospitalized COVID-19 patients. The green and red points indicate MM and SC groups, respectively. Sample C19-C014(*), excluded DEG analysis due to QC failure, was retained in methylation analysis.

Comment 23 :

Table 1 should include citations to support the reported associations of key genes with COVID-19 severity.

Response : In response to your comment, we have added a new column to Table 1 containing references that support the reported associations of each gene with COVID-19 severity. Specifically, we have cited the relevant studies by indicating their reference numbers as listed in the Reference section of the manuscript.

Direction of methylation	**Position**	**Gene symbol**	**Direction of mRNA expression**	**Correlation methylation with mRNA expression**		**Association with COVID-19**	**References**
				Spearman's *ρ*	***P*-value**		
Hypo-methylation	chr2:102143316	*IL18RAP*	Up regulation	-0.503	0.024	Immune response by mediating the activity of interleukin-18	[24-26]
	chr5:75320838	*GCNT4*	Down regulation	0.528	0.015	Respiratory failure	[30]
	chr6:36697843	*FKBP5*	Up regulation	-0.551	0.009	Stress-mediated immune response	[31-35]
	chr6:108562564	*FOXO3*	Up regulation	-0.501	0.025	Lung abnormalities, oxidative stress, development and maturation of T and B lymphocytes, secretion of inflammatory cytokines	[27-29]

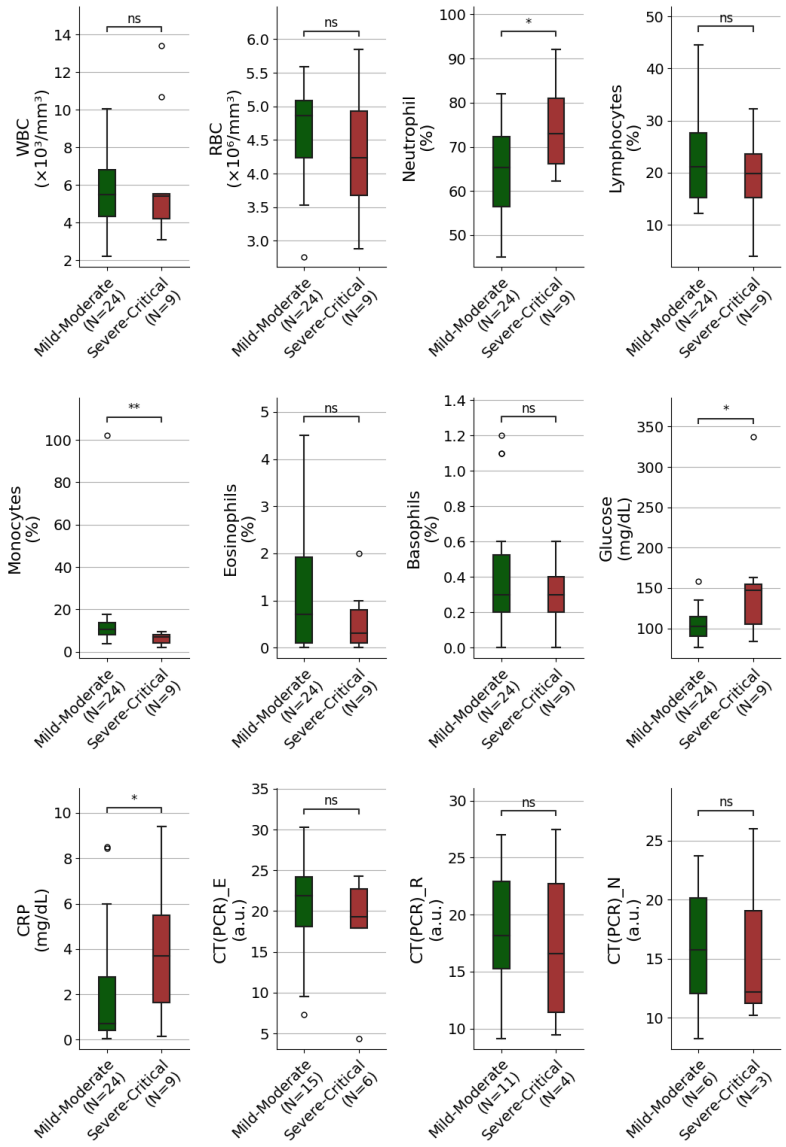
chr15:90100633	*FURIN*	Up regulation	-0.602	0.003	Viral infection	[21-23]
chr20:330120	*SRXN1*	Up regulation	-0.582	0.004	Oxidative stress related to multi-organ damage	[16-20]
chr20:654151	*SRXN1*	Up regulation	-0.586	0.004		
chr20:1166154	*SRXN1*	Up regulation	-0.547	0.010		
chr20:1467583	*SRXN1*	Up regulation	-0.538	0.012		
chr20:1631771	*SRXN1*	Up regulation	-0.510	0.021		

Table 1. **Six key genes associated with the severity of COVID-19 in the acute phase.** Columns indicate the direction of DNA methylation change (hypo- or hypermethylation); CpG site genomic position (hg38); associated gene symbol; direction of mRNA expression change (up- or downregulation); direction of correlation between DNA methylation and gene expression, Spearman's correlation coefficient (ρ), and associated *P*-value from *cis*-eQTM analysis; previously reported associations of each gene with COVID-19 and relevant references.

Comment 24 :

Figure S1 requires improvements: the y-axis should include units, and significant indicators (color coding) may not be necessary.

Response : Thanks for the suggestion. In the revised Figure S1, we have provided the units of each clinical value at the respective y-axis of the boxplots, and removed the color coding scheme for significance.



Supplementary Fig 1. Clinical Characteristics of Mild–Moderate and Severe–Critical COVID-19 Patients During the Acute Phase. The x-axis denotes clinical severity groups, with sample sizes indicated in parentheses. The y-axis indicates the clinical parameter measured, with units denoted in parentheses. Box plots display the median (centre line), interquartile range (box limits), and 1.5× interquartile range (whiskers). Individual data points beyond this range are shown as outliers (open circles). Significance thresholds are indicated as follows: ns ($P > 0.05$), * ($P \leq 0.05$), ** ($P \leq 0.01$), *** ($P \leq 0.001$), **** ($P \leq 0.0001$). Abbreviations: WBC=White Blood Cell count; RBC=Red Blood Cell count; CRP=C-reactive protein concentration; CT(PCR)=Cycle threshold (Ct) values from PCR assays for SARS-CoV-2 targets (E, R, and N genes); a.u.=arbitrary units.

Comment 25 :

Figure 2B is missing the healthy control group for RNA expression analysis. This omission limits comparative insights into baseline gene expression.

Response : We thank the Reviewer for improving our paper by pointing this out. Now, we have added the healthy control group in both Figure 2C (previously 2B) and 2D, as well as 3C, consistent with the Methylation data. We have included our healthy control RNA-seq data after batch removal using the ComBat-seq method. The data was processed using the same bioinformatic pipeline and identical parameters with COVID-19 data to prevent any *in silico* batch effects. We have extended the relevant section to provide the full methodological details (Materials and methods section: **Batch Effect Correction of Gene Expression Data**).

[Page 21, lines 406-412 (Revised manuscript)]: To integrate gene expression data from an independent cohort of 35 healthy individuals, we implemented rigorous batch effect correction to ensure comparability. All samples were processed using an identical bioinformatic pipeline as RNA-seq data for COVID-19 patients to avoid potential *in silico* batch effects. We applied the ComBat-seq method implemented in the sva R package (v3.54.0) to remove batch effect by differences in sequencing platforms (i.e., NovaSeq 6000 vs DNBSEQ-T7RS) while preserving the biological condition (i.e., Healthy vs. COVID-19 status).

Comment 26 :

Figure 3 should better clarify the relationships between methylation patterns, gene expression, and recovery trajectories.

Response : We agree that Figure 3 needs better clarification on the relationships between methylation, gene expression, and recovery trajectories. To better explain the Figure:

- **Panel A** shows that severity-dependent methylation differences—particularly in severe-critical patients—are prominent during the acute phase but diminish during recovery.

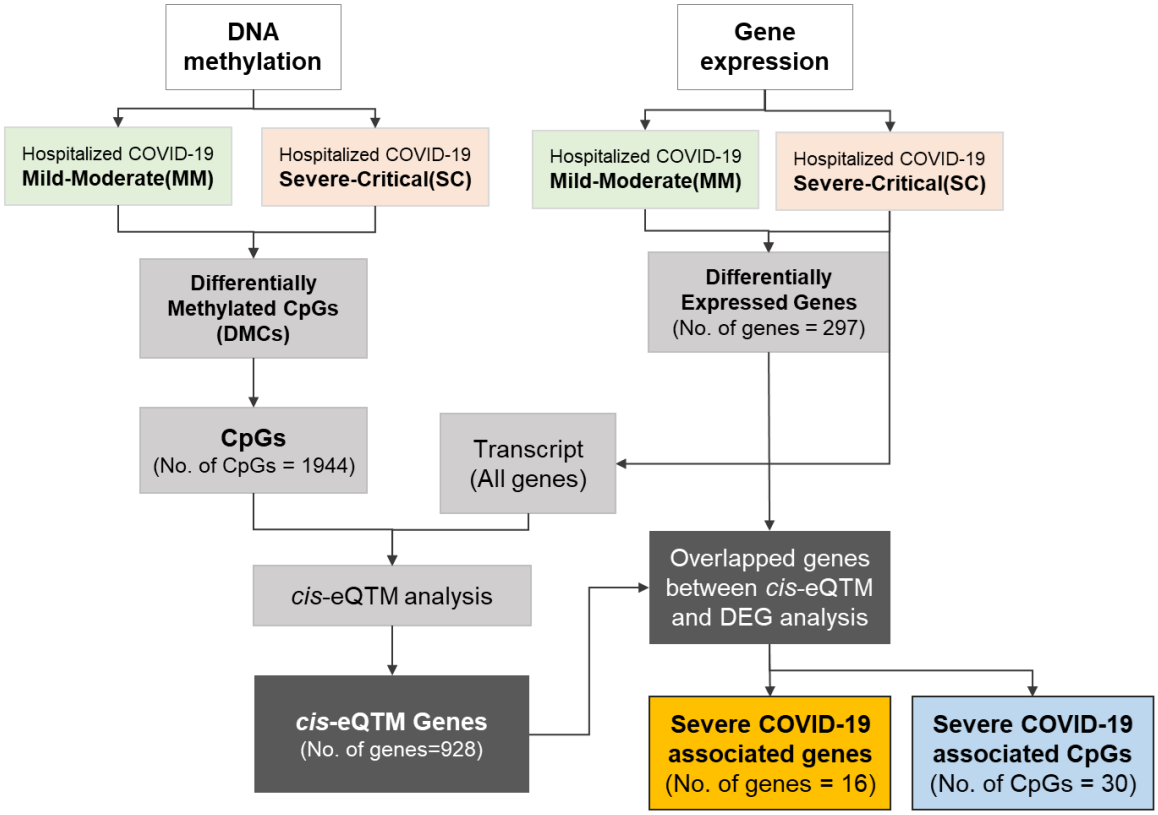
- **Panel B** shows that the PCA reveals partial methylation recovery in severe cases, contrasted with complete recovery in mild cases. We have enlarged the directional arrows and added scatter points of individual patients to highlight the progression and clarify residual differences from healthy controls.
- **Panel C** demonstrates that gene expression normalizes rapidly in both severity groups, aligning with healthy profiles by the recovery phase. Again, we added the individual scatter points to better depict the distances from healthy controls.

We have elaborated on the relationships by adding a separate Results section (Results: **Resolution of acute-phase epigenetic alterations during COVID-19 recovery**) and revised the figure legend of Figure 3 for clarification.

Comment 27 :

Some supplementary figures (e.g., Supplementary Figure 6) contain ambiguous labels that could be misinterpreted as the number of patients rather than gene counts. Clarification is needed.

Response : In accordance with your comment, we have revised the ambiguous labels to avoid potential misinterpretation. Specifically, the label "N" in Supplementary Figure 6 (now updated as Supplementary Figure 9) has been corrected to clearly indicate that it refers to gene counts rather than the number of patients.



Supplementary Fig 9. Overall process of COVID-19 *cis*-eQTM analysis. The figure illustrates the workflow, including the discovery of differentially methylated CpGs and differentially expressed genes. Integration of methylation and gene expression data was achieved through the identification of cis-eQTM associations.

Comment 28 :

The inclusion of a **graphical abstract** illustrating the study design is greatly appreciated. However, some modifications could enhance its clarity and effectiveness.

For instance, ensure all groups are clearly labeled and consistent with the main text. Furthermore, in figure 1B, a sample **(C19-014) stated to have been removed due to missing data or QC failure** is still shown.

Could you clarify this? Additionally, it would be helpful to include the **number of patients per group in the figure legend to provide a clearer overview.** Please also include a brief explanation about the convalescent group in the legend.

Response : Revisiting our initial manuscript, we see that we were not sufficient in details as to which samples were used in the analyses for different omics and clinical data. In the revised manuscript and renewed Graphical Abstract, we have tried our best to be as explicit as possible in which samples are excluded from the analysis due to technical problems. We provide the short summary of the detail below for the ease of understanding.

- **Methylation Data**: **46** samples (**37** MM and **9** SC)
    - **5** subjects (**four** MM and **one** SC) have missing data due to their follow-up loss
- **RNA Data**: **45** samples (**36** MM group and **9** SC group)
    - **5** subjects with no Methylation data
    - **an additional** subject (**one** MM) has missing data due to QC failure during library preparation; C19-014

Because C19-C014 passed methylation QC, it was included in the DMC analysis and therefore appears in Figure 1B, which depicts methylation results. To prevent any confusion, we have added an asterisk next to C19-C014 in Figure 1B with the note "Sample C19-C014(*), excluded DEG analysis due to QC failure, was retained in methylation analysis*"*

Additionally, we have updated Figure 1A to reflect the revised number of samples in the MM and SC groups, and this information is now clearly indicated in the figure legend as well. We also added the convalescent group and healthy controls to Figure 1A to provide a more comprehensive overview of the study design and improve the clarity of the overall analytical workflow.

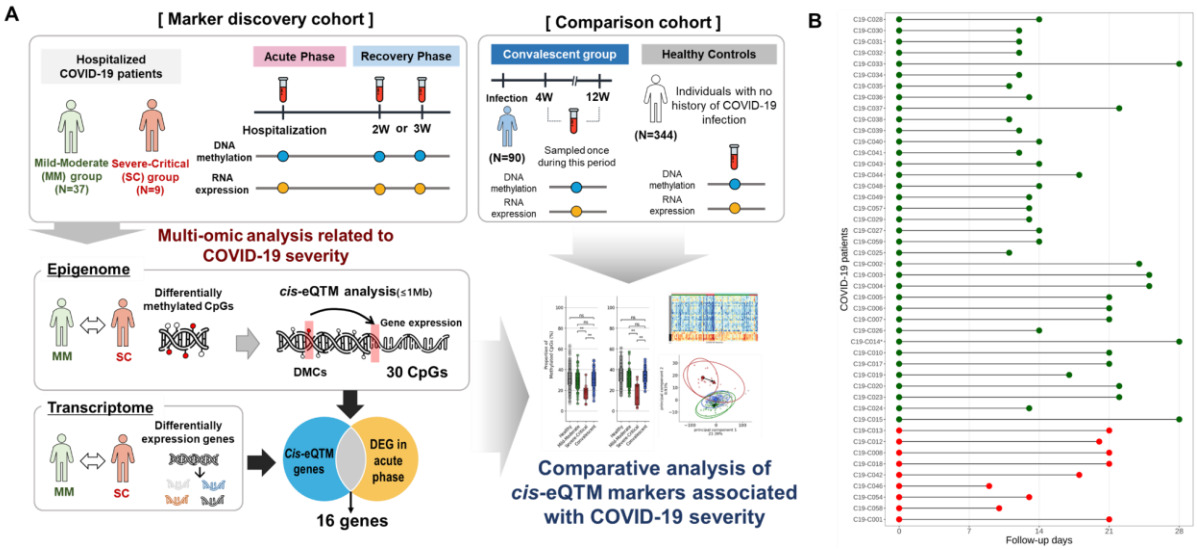


Figure 1. Overall study design and follow-up blood sample collection for *cis*-eQTM analysis in COVID-19 patients. A) Overview of blood sampling and analysis design; The marker discovery cohort consisted of hospitalized COVID-19 patients, stratified into Mild-Moderate (MM) (N=37) and Severe-Critical (SC)(N=9) groups, and was used to identify severity-associated *cis*-eQTM markers. The comparison cohort was composed of two independent reference groups : the 'Convalescent group' (N=90) comprising individuals who recovered from COVID-19 and provided a single blood sample 4 to 12 weeks post-infection, and the 'Healthy Controls' group (N=344), consisting of pre-pandemic individuals with no history of SARS-CoV-2 infection. **B)** Clinical follow-up duration of 46 Hospitalized COVID-19 patients. The green and red points indicate MM and SC groups, respectively. Sample C19-C014(*), excluded DEG analysis due to QC failure, was retained in methylation analysis.

Additional Corrections Initiated by the Authors

In addition to the changes made in response to the reviewers' comments, we made the following corrections and clarifications during our own re-review of the manuscript:

1. Correction of sample size

In the original manuscript, the number of hospitalized COVID-19 patients was reported as N=51 based on the availability of clinical information. Upon careful re-evaluation, we re-checked the samples actually used for molecular analyses and clinical data availability. We confirmed that the final dataset included 37 mild-moderate cases and 9 severe-critical cases, totaling N=46. Accordingly, we have corrected the reported sample numbers to N=46 throughout the manuscript, including in Figure 1, the Abstract, Results, and Methods sections.

2. Revision of DEG identification based on leave-one-out (LOO) analysis

In the original manuscript, we carried out eight iterations of LOO discovery to find the confident set of DEGs. However, the iterations should have been 9 times since there are a total of 9 Severe-Critical (SC) patients with RNA-seq data available. The last iteration was inadvertently missed due to a mistake. Therefore, we have rerun the analysis including all of nine samples, and the threshold from six to seven overlaps across the nine iterations. After the adjustments, the number of DEG-LOOs declined from 18 to 16 genes, removing *IL18R1* and LINC02217 from our shortlist of biomarkers.

Title : Identification of reactive CpGs and RNA expression in early COVID-19 through *cis*-eQTM analysis reflecting disease severity and recovery

Reviewer #2

Comment 1 :

Thank you very much for addressing all my concerns with such accuracy and completeness. I have no further request with one minor aspect. Suppl. Fig5 has p-value indicators that are not present in the graphic. Please remove

Supplementary Fig 5. Baseline leukocyte proportion and its changes one week later in Mild-Moderate and Severe-Critical groups. Boxplots show the proportions (%) of neutrophils, lymphocytes, monocytes, eosinophils, and basophils measured in whole blood at baseline (left panel) and after a week later (right panel) for patients with Mild–Moderate (green) and Severe–Critical (red) disease. Boxplots display the median (centre line), IQR (box limits), and 1.5×IQR (whiskers), and outliers (open circles). Statistical comparisons were conducted using the Wilcoxon rank-sum test. Statistical significance is denoted as follows: ns ($P \geq 0.05$), * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$), **** ($P < 0.0001$).

Response : We thank the reviewer for the detailed comment. As suggested, we have revised the caption of Supplementary Figure 5 to remove the p-value indicators that were not shown in the figure. The caption has been updated accordingly.

[Revised supplementary materials] : Supplementary Fig 5. Baseline leukocyte proportion and its changes one week later in Mild-Moderate and Severe-Critical groups. Boxplots show the proportions (%) of neutrophils, lymphocytes, monocytes, eosinophils, and basophils measured in whole blood at baseline (left panel) and after a week later (right panel) for

patients with Mild–Moderate (green) and Severe–Critical (red) disease. Boxplots display the median (centre line), IQR (box limits), and 1.5×IQR (whiskers), and outliers (open circles). Statistical comparisons were conducted using the Wilcoxon rank-sum test. Statistical significance is denoted as follows: ns ($P \geq 0.05$), * ($P < 0.05$).