



OPEN ACCESS

EDITED BY

Shuangbo Kong,
Xiamen University, China

REVIEWED BY

Ayse Banu Demir,
İzmir University of Economics, Turkey
Fei Guo,
Chinese Academy of Medical Sciences &
Peking Union Medical College, China

*CORRESPONDENCE

Hong-gang Li,
lhgyx@hotmail.com
Ximiao He,
ximiaohe@hust.edu.cn
Li-quan Zhou,
zhouliquan@hust.edu.cn

[†]These authors have contributed equally
to this work

SPECIALTY SECTION

This article was submitted to Molecular
and Cellular Pathology,
a section of the journal
Frontiers in Cell and Developmental
Biology

RECEIVED 23 July 2022

ACCEPTED 15 September 2022

PUBLISHED 03 October 2022

CITATION

Yin Y, Liu X-z, Tian Q, Fan Y-x, Ye Z,
Meng T-q, Wei G-h, Xiong C-l, Li H-g,
He X and Zhou L-q (2022),
Transcriptome and DNA methylome
analysis of peripheral blood samples
reveals incomplete restoration and
transposable element activation after 3-
months recovery of COVID-19.
Front. Cell Dev. Biol. 10:1001558.
doi: 10.3389/fcell.2022.1001558

COPYRIGHT

© 2022 Yin, Liu, Tian, Fan, Ye, Meng,
Wei, Xiong, Li, He and Zhou. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Transcriptome and DNA methylome analysis of peripheral blood samples reveals incomplete restoration and transposable element activation after 3-months recovery of COVID-19

Ying Yin^{1,2,3,4†}, Xiao-zhao Liu^{1,3,4†}, Qing Tian², Yi-xian Fan^{1,3,4},
Zhen Ye², Tian-qing Meng², Gong-hong Wei⁵,
Cheng-liang Xiong², Hong-gang Li^{2*}, Ximiao He^{1,3,4*} and
Li-quan Zhou^{2*}

¹Department of Physiology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, ²Institute of Reproductive Health, Center for Reproductive Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, ³Center for Genomics and Proteomics Research, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, ⁴Hubei Key Laboratory of Drug Target Research and Pharmacodynamic Evaluation, Huazhong University of Science and Technology, Wuhan, China, ⁵MOE Key Laboratory of Metabolism and Molecular Medicine, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Fudan University Shanghai Cancer Center, Shanghai Medical College of Fudan University, Shanghai, China

Comprehensive analyses showed that SARS-CoV-2 infection caused COVID-19 and induced strong immune responses and sometimes severe illnesses. However, cellular features of recovered patients and long-term health consequences remain largely unexplored. In this study, we collected peripheral blood samples from nine recovered COVID-19 patients (median age of 36 years old) from Hubei province, China, 3 months after discharge as well as 5 age- and gender-matched healthy controls; and carried out RNA-seq and whole-genome bisulfite sequencing to identify hallmarks of recovered COVID-19 patients. Our analyses showed significant changes both in transcript abundance and DNA methylation of genes and transposable elements (TEs) in recovered COVID-19 patients. We identified 425 upregulated genes, 214 downregulated genes, and 18,516 differentially methylated regions (DMRs) in total. Aberrantly expressed genes and DMRs were found to be associated with immune responses and other related biological processes,

Abbreviations: WGBS, whole-genome bisulfite sequencing; DMR, differentially methylated region; TEs, transposable elements; LINES, long interspersed nuclear elements; SINES, short interspersed nuclear elements; LTRs, long terminal repeats; DNAs, DNA transposons; DEGs, differentially expressed genes; DETEs, differentially expressed TEs.

implicating prolonged overreaction of the immune system in response to SARS-CoV-2 infection. Notably, a significant amount of TEs was aberrantly activated and their activation was positively correlated with COVID-19 severity. Moreover, differentially methylated TEs may regulate adjacent gene expression as regulatory elements. Those identified transcriptomic and epigenomic signatures define and drive the features of recovered COVID-19 patients, helping determine the risks of long COVID-19, and guiding clinical intervention.

KEYWORDS

SARS-cov-2, recovery, whole-genome bisulfite sequencing, transposable elements, differentially methylated regions

Introduction

Emerging SARS-CoV-2 coronavirus which causes coronavirus disease 2019 (COVID-19) and results in complicated health issues has expanded rapidly and swept the whole world, threatening global public health (Guan et al., 2020; Huang et al., 2020). Identified key receptors for SARS-CoV-2 infection include ACE2 (Zhou et al., 2020), TMPRSS2 (Hoffmann et al., 2020), and NRP1 (Cantuti-Castelvetri et al., 2020) which are widely expressed in different tissues of the human body. Symptoms of COVID-19 patients include fever, cough, fatigue, headache, diarrhea, and in severe cases even organ failure. Different patients have various symptoms after SARS-CoV-2 infection, and most of them develop mild to moderate illness. Despite intensive multi-omics investigations (Bojkova et al., 2020; Shen et al., 2020; Xiong et al., 2020; Delorey et al., 2021; Wu et al., 2021), the impact of SARS-CoV-2 on the human body and its long-term effect remains largely unexplored. Notably, SARS patients were tracked after the outbreak of SARS at 2003, identifying a significant incidence of sequelae, including pulmonary fibrosis and limited body function (Gomersall et al., 2004). During the current pandemic, COVID-19 patients were reported to suffer from fatigue, sleep difficulties, and anxiety/depression several months after recovery (Chen et al., 2021). Meanwhile, studies on COVID-19 patients showed severely impaired gut microbiota up to 3–6 months after recovery (Chen et al., 2021; Tian et al., 2021). Moreover, pulmonary dysfunction and ‘plasma metabolites’ remain incompletely restored 3 months after recovery (Chen et al., 2021). Understanding the progress of convalescence of COVID-19 patients is therefore valuable for clinical intervention.

Transposable elements (TEs) are mobile DNA elements and comprise about 40% of human genome (Dewannieux et al., 2003). Four major TE classes are long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), long terminal repeats (LTRs) and DNA transposons. They are involved in many cellular processes, such as: transcriptional regulation (Percharde et al., 2018), chromatin structure organization (Fadloun et al., 2013), development and cell differentiation (Bernardes et al., 2020; Padmanabhan Nair et al., 2021). Retrotransposons are active TEs capable of “copy and paste” themselves into the human genome through RNA intermediates. Well-known retrotransposons include LINEs, SINEs,

and LTRs. LINEs are the most common autonomous retrotransposons, and the mobilization activity of SINEs relies on LINEs (Dewannieux et al., 2003; Cordaux and Batzer, 2009). Retrotransposons can cause insertion, deletion, and inversion in the human genome and therefore their increased expression may lead to reduced genome stability (Gilbert et al., 2002; Symer et al., 2002; Newkirk et al., 2017; Malki et al., 2019). We have reported inappropriate upregulation of TEs especially retrotransposons upon SARS-CoV-2 infection in human cell lines and its potential harm including impaired genome stability, enhanced susceptibility of aged people and cancer patients, aberrant expression of retrotransposon-adjacent genes, and induction of inflammation (Yin et al., 2021). Previous in-depth transcriptome analysis revealed aberrant inflammatory responses in COVID-19 patients (Blanco-Melo et al., 2020). Regarding the impact of SARS-CoV-2 on the human body, a recent study using peripheral blood from COVID-19 patients showed impaired transcriptional network and epigenetic profiles which might be useful for targeted treatment and provided promising hallmarks to predict clinical outcomes (Bernardes et al., 2020). However, how SARS-CoV-2 impacts TEs in the human body remains unclear.

Gene expression pattern and epigenetic profile of peripheral blood reflect the whole body’s metabolic status. In the current study, we collected peripheral blood from COVID-19 patients 3 months after recovery from COVID-19 and carried out transcriptome and DNA methylome studies (Figure 1A). Significant amounts of misregulated genes/TEs and differentially methylated regions (DMRs) were identified, indicating incomplete restoration of the human body. Additionally, we identified transcriptome and epigenome signatures which will help identify the long-term impact of COVID-19 on health and provide suggestions for clinical treatment.

Material and methods

Patient recruitment and blood sample collection

14 Participants were recruited from the Center for Reproductive Medicine, Tongji Medical College, Huazhong

University of Science and Technology. Inclusion criteria for COVID-19 patients included age from 25 to 45 years, having detailed medical records of hospitalization and discharge. Diagnosis of COVID-19 was determined by the New Coronavirus Pneumonia Prevention and Control Program (7th edition) published by the National Health Commission of China (<http://www.nhc.gov.cn/xcs/zhengcwj/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>). Exclusion criteria included asymptomatic cases, taking antibiotics within 2 months, gastrointestinal diseases, and severe basic diseases. Age- and gender-matched healthy controls were recruited during regular physical check-ups in the same hospital with none of them receiving antibiotics within 2 months before collection of blood samples. Ethical approval for the study was obtained by the Ethics Committee for Clinical Research of Reproductive Medicine Center, Tongji Medical College, Huazhong University of Science and Technology. All participants included in the study gave informed consent. 9 male patients in the recovery stage (5 Mild, 4 Moderate) and 5 male healthy controls were recruited for this study. All nine patients meet confirmatory laboratory evidence by detection of SARS-CoV-2 RNA in a clinical specimen using a diagnostic molecular amplification test. No pneumonia symptoms were seen in the imaging of 5 patients with mild symptoms. In 4 patients with moderate symptoms, fever and respiratory tract symptoms can be seen, and pneumonia symptoms were seen on imaging. Age, body mass index (BMI) and other demographic information can be identified in [Table 1](#). Blood (5 ml) was withdrawn from each patient into an Ethylenediamine tetraacetic acid (EDTA)-K2 tube to decelerate blood coagulation. Total RNA was isolated from 2.5 ml blood including PBMC using total RNA isolation kit from SIMGEN. Isolated total RNA and remaining blood samples were subjected to mRNA sequencing and whole-genome bisulfite sequencing (WGBS) by Annoroad Gene Technology Co. Ltd. (Beijing).

mRNA isolation for sequencing

Total RNA was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (#E7530L, NEB, United States) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. The clustering of the index-coded samples was performed on a cBot cluster generation system using HiSeq PE Cluster Kit v4-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the libraries were sequenced on an Illumina HiSeq 2500 system. Approximately 25M of paired-end reads (150 bp ×2) for each sample was generated.

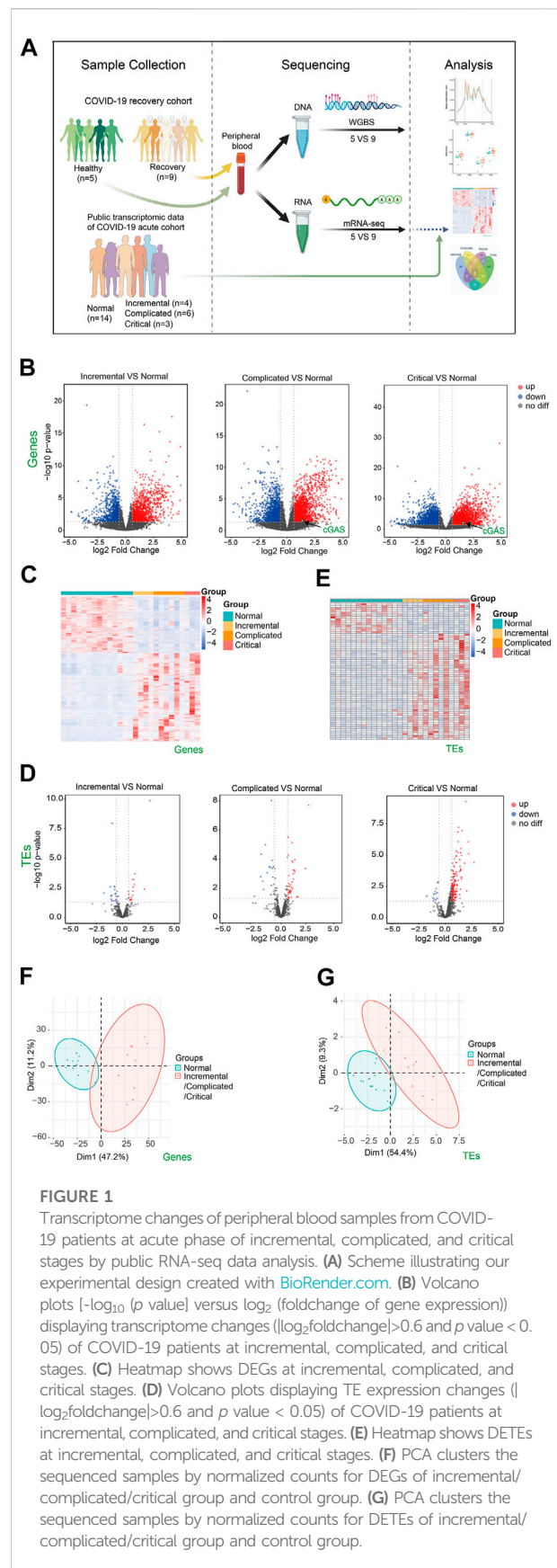


TABLE 1 Clinical Characteristics of COVID-19 Patients at recovery stage (R1-R9) and Healthy controls (H1-H5) in this study.

No	Gender	Ethnicity	Age (yrs)	Disease severity	Disease duration (days)	Time from discharge to sampling (days)	BMI	CoV-IgM	CoV-IgG
R1	Male	Chinese	31	Mild1	26	70	25.95	+	+
R2	Male	Chinese	29	Mild2	25	79	20.05	+	+
R3	Male	Chinese	24	Moderate1	17	76	28.07	-	+
R4	Male	Chinese	36	Moderate2	13	83	26.23	-	+
R5	Male	Chinese	43	Mild3	12	89	23.66	-	+
R6	Male	Chinese	40	Mild4	32	96	35.43	-	+
R7	Male	Chinese	39	Moderate3	9	98	32.1	-	+
R8	Male	Chinese	36	Moderate4	24	78	26.78	-	+
R9	Male	Chinese	43	Mild5	21	107	25.33	-	+
H1	Male	Chinese	28	Healthy	N/A	N/A	26.4	-	-
H2	Male	Chinese	28	Healthy	N/A	N/A	21.3	-	-
H3	Male	Chinese	32	Healthy	N/A	N/A	21.3	-	-
H4	Male	Chinese	39	Healthy	N/A	N/A	23.0	-	-
H5	Male	Chinese	37	Healthy	N/A	N/A	22.9	-	-

Comparison between COVID-19 group and control group showed no significant differences of age ($p > 0.05$) or BMI ($p > 0.05$). Note that average time from discharge to sampling of COVID-19 patients (R1–R9) was 86 days (about 3 months), and median age was 36 years.

DNA isolation for methylation profiling

For constructing WGBS libraries, the genomic DNA was fragmented to a mean size of 350 bp, followed by blunt-ending, dA-Tailing, and adaptor ligation. Insert fragments with different sizes were excised from a 2% agarose gel and purified using the QIAquick Gel Extraction Kit (QIAGEN). Purified DNA was bisulfite converted using the EZ DNA Methylation-Gold Kit (#D5006, ZYMO Research, CA, United States) and PCR amplified. The WGBS libraries were sequenced at 20×depth on an Illumina HiSeq 2500 system as paired-end reads (150 bp ×2). Approximately 15–20 × mean coverage was generated for each sample.

RNA-seq data processing

Raw reads were processed with Trim Galore (v0.6.4) to remove adaptor sequences and poor-quality bases with “--q 20 --phred33 --stringency 5 --length 20 --paired.” To include as many non-uniquely mapped reads as possible, trimmed reads were firstly aligned to the human genome (hg19) by STAR (v2.7.5b) (Dobin et al., 2013) (Dobin, Davis et al., 2013) with default settings including parameters ‘--winAnchorMultimapmax 2000 --outFilterMultimapNmax 1000’. SAMtools (v1.3.1) was used to sort bam files by genomic coordination and make a bam file index. RSEM (v1.2.28) (Li and Dewey, 2011) was used to calculate the FPKM value of

genes. TETranscript (Jin and Hammell, 2018) with default parameters was used to get counts for different transposable elements. UCSC genome browser was used for snapshots of the transcriptome. R package Deseq2 (v1.28.1) (Love et al., 2014) was used to obtain differentially expressed genes (DEGs) and differentially expressed TEs (DETEs). Principal component analysis (PCA) was performed using DESeq2 normalized counts for DEGs/DETEs. Metascape (Zhou et al., 2019) was used to visualize functional profiles of genes and gene clusters. rMATS (v4.1.1) (Shen et al., 2014) was used to identify alternative splicing events with “--readLength 150” and other default parameters. We conducted protein-protein interaction network (PPI network) analysis using STRING database to explore gene interaction at protein level (Szkarczyk et al., 2019). Graphs were created by R. Images were organized by Adobe Illustrator.

ReMap2022 database for transcription factor binding enrichment analysis

DETEs were transformed to genomic range objects (GRanges) using the GRanges R package. The DETE loci were then compared against the ReMap2022 annotated TFBSs (hg19) for enrichment of specific transcription factors and plotted as enrichment dot plots with the ReMapEnrich R package (Hammal et al., 2022).

WGBS data processing and quality control

Raw reads were processed with Trim Galore (v0.6.4) to remove adaptor sequences and poor quality bases with “--q 20 --phred33 --stringency 5 --length 20 --paired.” Trimmed reads were then aligned to the human genome (hg19) by Bismark (v0.22.3) (Krueger and Andrews, 2011) using the parameters “-p 6 --parallel 1 -N 0 -L 20 --quiet --un --ambiguous --bam.” SAMtools (v1.3.1) was used to sort bam files by genomic coordination and make a bam file index. PCR duplicates were removed using Picard (v2.23.3). The methylation ratio at each CpG site was constructed using bismark_methylation_extractor model with the parameters “-p --comprehensive --no_overlap --bedgraph -counts --report --cytosine_report --gzip --buffer --size 30G.” For all samples, the average bisulfite conversion success ratio is >99.2%, the alignment ratio is around 80% of read pairs aligning uniquely, and the duplication rate is <5%. For each CpG site, methylation levels were calculated by (methylation reads/total coverage reads). For a more robust analysis, we applied the minimum threshold 3× coverage and also selected CpGs that all samples had their methylation levels. This screening process gave 18 M of CpGs with confident methylation levels. Methylation profiles were calculated by deeptools (v3.5.1) (Ramirez et al., 2014).

Differentially methylated regions by methylKit

The R package methylKit (v1.14.2) (Akalin et al., 2012) was used to identify DMRs between healthy and recovery groups. The methylation levels at CpG sites were firstly calculated by “methRead” function with mincov = 3. Methylation across the genome was tiled with the ‘tileMethylCounts’ function using the parameters “win.size = 500, step.size = 500, cov.bases = 5”, then ‘unite’ function was used to unite tiled regions with the “destrand = TRUE” parameter. At last, “calculateDiffMeth” function was used to calculate DMRs. DMRs with a minimum of 3 CpG sites and absolute methylation mean difference >10% and q-value < 0.05 were used for further analysis. DMRs were annotated by R package “ChIPseeker” (v1.24.0).

Statistical methods

Plotting and statistical tests were performed using R (v4.0.2). All statistical tests performed in this study were two-sided. Box plots were generated using the R packages “ggplot2” (v3.3.2) and “ggpubr” (v0.4.0) to show median, first and third quartiles and outliers were shown if outside

the 1.5× interquartile range. A two-sided Wilcoxon signed-rank test was used to assess differences between the two groups. Enrichment scores were analyzed using chi-square tests, enrichment score >1 and *p*-value < 0.05 was defined as enrichment.

Results

The SARS-cov-2 infection had a profound impact on transcriptome and TE activation which was positively correlated with COVID-19 severity

To identify how the human body responds to SARS-CoV-2, we downloaded and analyzed public RNA-seq data of peripheral blood from COVID-19 patients at acute phase including incremental, complicated, and critical stages, as well as healthy controls (Bernardes et al., 2020) (Supplementary Table S1). A total of 2392, 4,241 and 3,655 misregulated genes were identified in these three stages respectively (Figures 1B,C). DEGs were mainly enriched in immune response, cell cycle, and DNA repair (Supplementary Figures S1A,B). We previously reported that expression of TEs was upregulated upon SARS-CoV-2 infection in human cell lines. This may result in impaired genome stability, increased susceptibility of aged people and cancer patients, misregulation of retrotransposon-adjacent genes, and probably leading to worse clinical outcomes for patients with underlying diseases (Yin et al., 2021). Therefore, we examined the expression of TEs in the acute phase, and observed a gradual upregulation of TEs from incremental (16 upregulated) to complicated (41 upregulated) to critical (164 upregulated) stage (Figures 1D,E; Supplementary Figure S2), indicating that TE expression levels reflected COVID-19 severity. In agreement with the identification of TE activation, we also noticed cGAS upregulation at complicated and critical stages (Figure 1B), while retrotransposon upregulation is one of the major reasons for stimulating the cGAS-STING pathway (Gorbunova et al., 2021). The cGAS-STING pathway drives immune response to cytosolic DNA, activates IFN responses and induces inflammation. Therefore, retrotransposon is a promising therapeutic target for reducing inflammation in COVID-19 patients. Previous study reported release of mitochondrial DNA into the cytoplasm in SARS-CoV2-infected cells, and this may be another reason of activation of cGAS-STING pathway (Domizio et al., 2022). PCA effectively clustered the samples by either DEGs (Figure 1F) or DETEs (Figure 1G). Collectively, our results indicated that SARS-CoV-2 infection triggered immune responses and activated TEs in human peripheral blood.

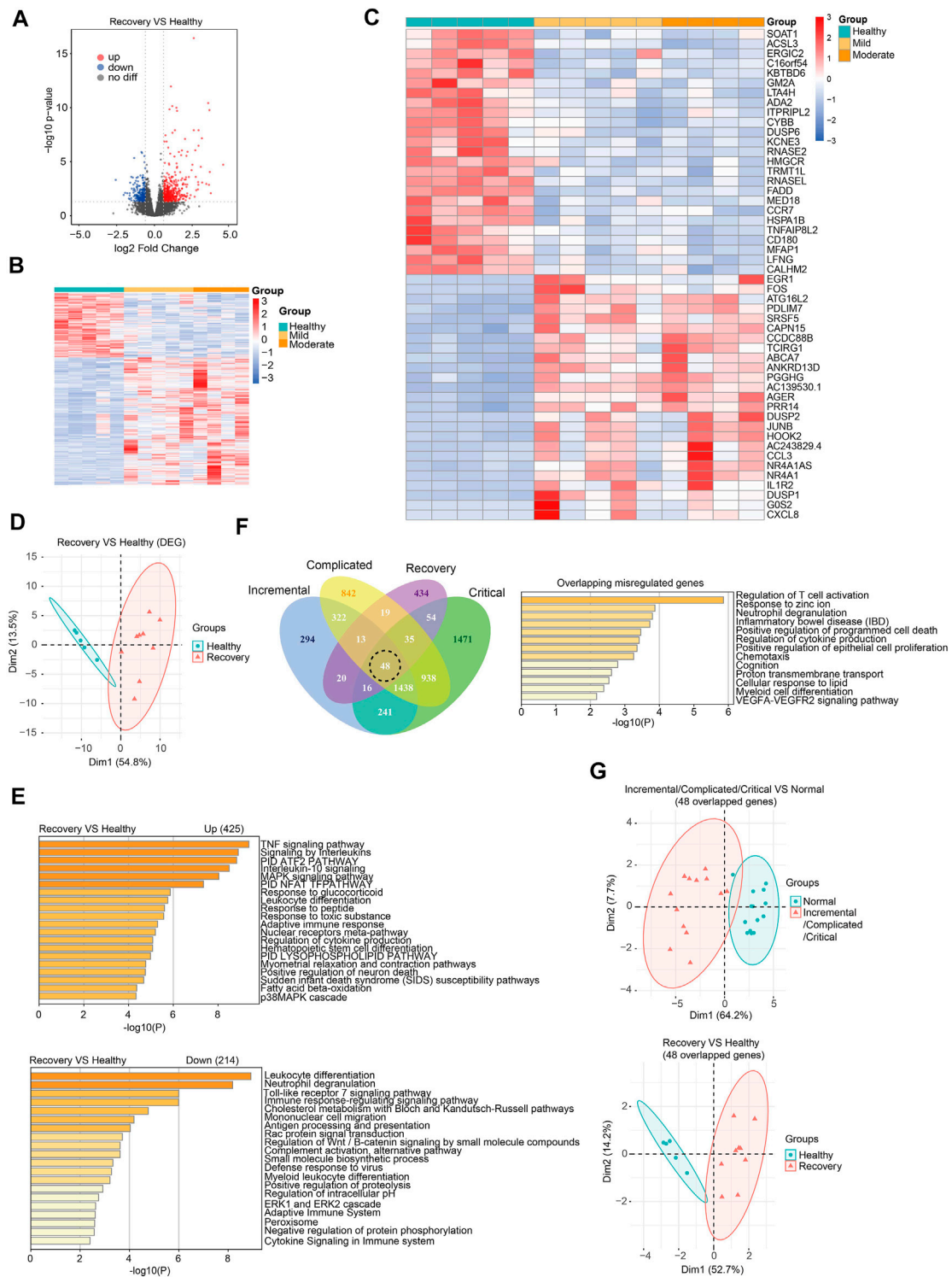


FIGURE 2

Transcriptome analysis identified gene expression changes of peripheral blood samples from COVID-19 patients after 3-months recovery. **(A)** Volcano plots [$-\log_{10}(p \text{ value})$ versus $\log_2(\text{foldchange of gene expression})$] displaying transcriptome changes ($|\log_2(\text{foldchange})| > 0.6$ and $p \text{ value} < 0.05$) at recovery stage. **(B)** Heatmap shows DEGs in COVID-19 patients at recovery stage. **(C)** Heatmap shows top 25 downregulated and top 25 upregulated genes at recovery stage. **(D)** PCA clusters the sequenced samples by normalized counts for DEGs of recovery group and control group. **(E)** GO analysis of upregulated/downregulated genes for functional enrichment at recovery stage by Metascape. **(F)** Venn diagram identifies 48 overlapping misregulated genes among incremental, complicated, critical stages and recovery stage. GO analysis was further performed to identify functional enrichment of the 48 genes by Metascape. **(G)** PCA of incremental/complicated/critical group and control group (left) and PCA of recovery group and control group (right), using 48 overlapped genes. PCA clusters the sequenced samples by normalized counts for 48 overlapping misregulated genes among incremental, complicated, critical stages and recovery stage.

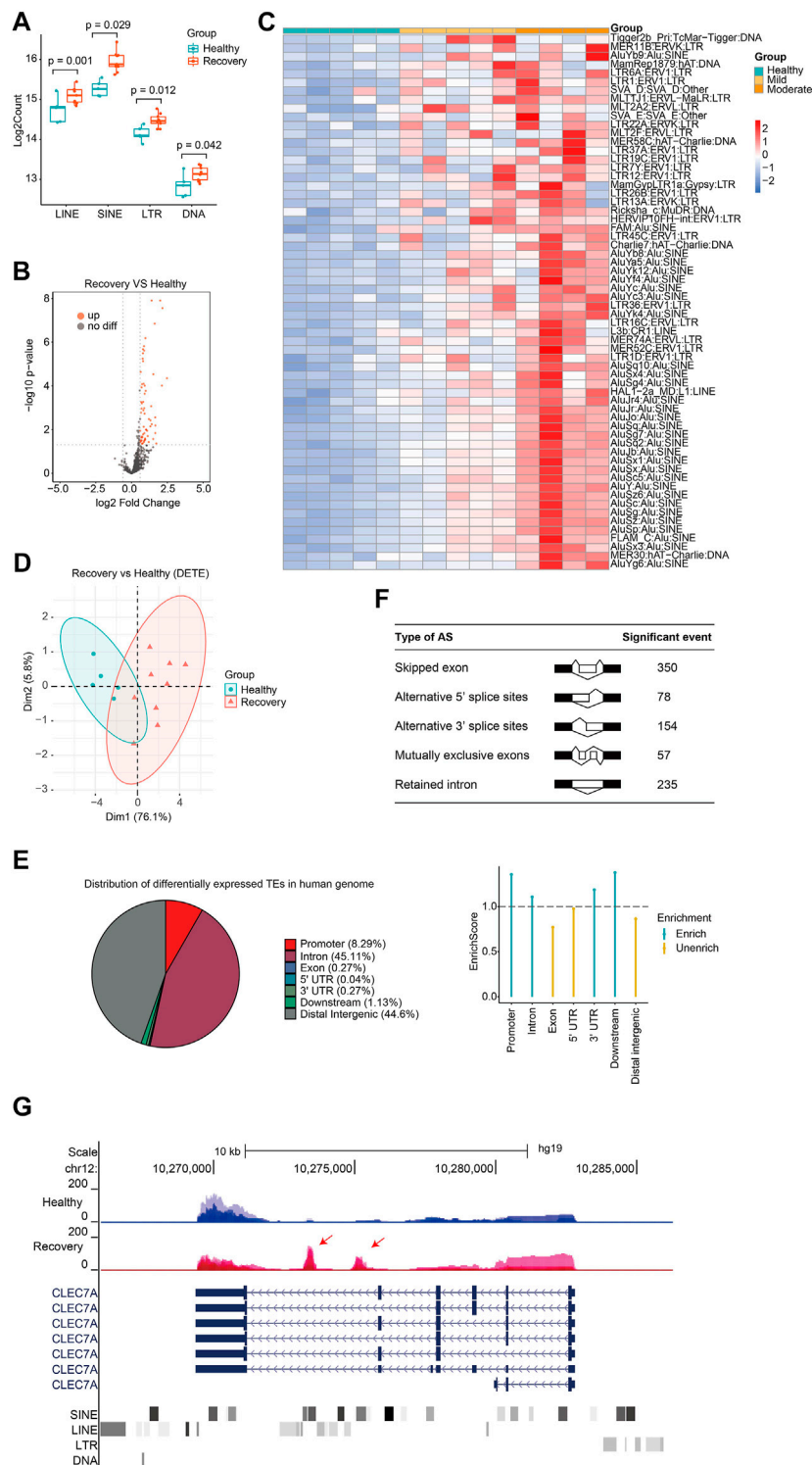


FIGURE 3

Discovery of aberrantly overexpressed TEs in peripheral blood samples from COVID-19 patients after 3-months recovery. **(A)** Boxplot displaying significantly upregulation of four major TE classes. LINE, SINE, LTR, and DNA transposons are shown, and each dot represents one sample. The median, first, and third quartiles are shown. Two-sided Wilcoxon signed-rank test were used for the comparisons. **(B)** Volcano plots [$-\log_{10}$ (p value) versus \log_2 (foldchange of gene expression)] displaying TE expression changes ($|\log_2 \text{foldchange}| > 0.6$ and p value < 0.05) at recovery stage. **(C)** Heatmap demonstrates all 62 upregulated TEs in COVID-19 patients at recovery stage. **(D)** PCA clusters the sequenced samples by normalized counts for DETEs of recovery group and control group. **(E)** Pie chart shows the distribution of DETE in different gene features of human genome.

(Continued)

FIGURE 3 (Continued)

Enrich score was calculated by (Feature ratio of DETE)/(Feature ratio of TE), chi-square tests was used for statistics, enrichment score >1 and p -value < 0.05 was defined as enrichment. (F) Change of alternative splicing (AS) events (FDR<0.05) was observed at recovery stage. (G) UCSC genome browser view of RNA-seq data demonstrates aberrantly increased TE expression from an intron of CLEC7A gene initiates a novel transcript within CLEC7A gene locus.

Whole-blood transcriptome analysis revealed that SARS-cov-2 impacted the expression of multiple genes and alternative splicing events even after a 3-months recovery

To understand the recovery progress of the human body after 3-months convalescence from SARS-CoV-2 infection, we recruited COVID-19 patients with mild/moderate symptoms and controls at similar ages in Hubei province, China (Table 1). Generally, patients had a median age of 36 years and the average follow-up time after hospital discharge was 86 days (approximately 3 months). Peripheral blood was collected for subsequent RNA-seq analysis (see Supplementary Table S2 for quality control information). Generally, 425 genes were upregulated and 214 genes were downregulated (Figures 2A–C). PCA effectively clustered the samples by DEGs (Figure 2D). DEGs were enriched in immune response-related biological processes (Figure 2E). We then ask whether the recovery group and acute phase groups share commonly misregulated genes. We found that among 639 misregulated genes in the recovery group, 205 genes were also misregulated in acute phase groups (Figure 2F). There were 48 genes misregulated in all four disease stages and involved in T cell activation and other immune response-related processes (Figures 2F; Supplementary Figure S3). Notably, we found that these genes play various roles in immune responses, and include transmembrane receptors, voltage-gated channel proteins, and transcription factors. As expected, these 48 overlapping misregulated genes can be used to discriminate COVID-19 patients from controls, no matter whether the patients were at acute or recovery stages (Figure 2G), and are therefore useful for clinical diagnosis and treatment of COVID-19.

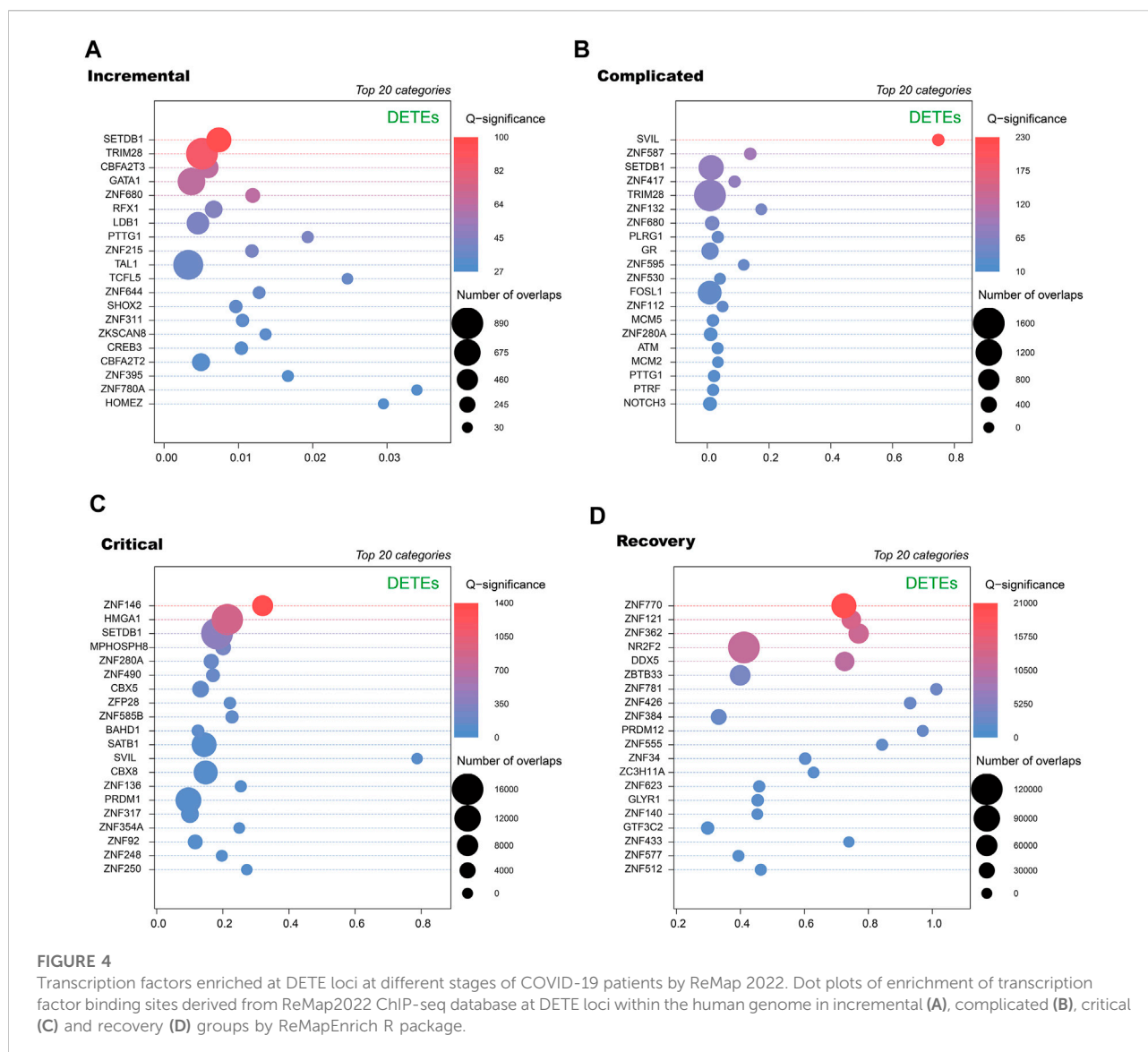
Next, we analyzed expression changes of TEs in the recovery group. When we added up all reads for each class of transposon, we observed that global expression of LINE, SINE, LTR, and DNA transposon were all significantly increased (Figure 3A). There were 62 upregulated TEs with most of them belonging to SINE and LTR, while no downregulated TE subfamilies were observed (Figures 3B,C; Supplementary Table S3). Notably, those with moderate illness seemed to have higher upregulation of DETEs than those with mild illness, indicating that TE levels reflected the severity of COVID-19 at the recovery stage. TEs were not randomly upregulated because differentially expressed TEs can well cluster the samples (Figure 3D). Surprisingly, 56 of the 62 upregulated TEs in the recovery group were absent in patients from the acute phase group, indicating that distinct TE subfamilies were activated during the recovery progress.

Different TE subfamilies were misregulated in patients from various stages, and this may be explained by crosstalk among different TE subfamilies and complicated response inside human body. Therefore, we analyzed previously reported transcriptome data of LINE1 knockdown in mouse embryonic stem cells (mESCs) (Percharde et al., 2018) and indeed found abnormally expressed subfamilies like SINE and LTR (Supplementary Figure S4). Another possible reason for different misregulated TE subfamilies between the acute group and recovery group is the diversity of human race and genetic background. Next, we analyzed the distribution of upregulated TEs in the human genome and found that they were mainly enriched in promoter, intron, and downstream regions (Figure 3E). Differentially expressed TEs were overrepresented in the human genome, indicating that upregulation of TEs plays important role in SARS-CoV-2-induced transcriptome changes and serves as a transcriptome signature in COVID-19 patients at the recovery stage. Meanwhile, our transcriptome analysis identified significant changes in alternative splicing events in the recovery group (Figure 3F). Although it is unclear how TEs were activated in the recovery group, increased expression of some of these may induce the generation of novel transcripts inside gene loci and impact alternative splicing patterns (Figure 3G). Beyond transcriptional disturbance, enhanced TE expression may also reduce genome stability, induce inflammation, and cause age-associated disorders.

To further annotate DETEs, we used ReMap2022 to obtain transcription factor binding sites (TFBSs) from publicly available ChIP-seq dataset (Hammal et al., 2022). By comparing loci of the DETEs with annotated TFBSs, we observed distinct categories of transcription factors enriched at TE loci in different stages (Figure 4). Notably, zinc finger proteins (ZNFs) are key regulators of TE expression, and are frequently identified in enriched transcription factors but diverse ZNFs are seen in these stages. Therefore, alteration of DETEs at acute and recovery stages may be caused by dynamic expression of ZNFs at these stages.

Whole-blood DNA methylome analysis of the recovery group identified genome-wide DMRs which mainly localized at TEs' loci between healthy and recovery groups

Gene expression can be regulated by DNA methylation which can be long-term memorized. Thus, we next examined



the expression of DNA Methyltransferase (DNMT) and Ten-eleven Translocation (TET) family genes which play key roles in controlling DNA methylation status. Almost all of these genes showed altered expression levels in patients at the recovery stage (Supplementary Figure S5). This prompted us to further investigate changes of in genome-wide DNA methylation in COVID-19 patients after a 3-months recovery.

To study how DNA methylome was changed between healthy and recovery groups and how it may correlate with transcriptome alteration, we examined whole-blood DNA methylome by WGBS (see Supplementary Table S4 for quality control information). Generally, whole-genome CG methylation levels showed no significant alterations (Figure 5A). Analysis of genomic regions from 2 kb upstream of transcriptional start sites (TSSs) to 2 kb

downstream of transcriptional end sites (TESs) indicated no significant CG methylation changes (Figure 5B). Next, we examined loci of TEs and identified minimum alteration of global CG methylation level (Figures 5C,D).

To explore changes in CG methylation patterns between healthy and recovery groups, we analyzed DMRs and identified 18,516 DMRs in total (absolute methylation mean difference >10% and q-value < 0.05). 8,724 DMRs had hypo-methylation (hypo-DMRs) and 9,792 DMRs had hyper-methylation (hyper-DMRs) in the recovery group. Identified DMRs were mainly enriched at gene promoter, intron, and certain TE regions (Figure 5E). Although we didn't detect significant effects of COVID-19 on DNA methylome in genome-wide analysis, GO analysis showed that hyper-DMRs and hypo-DMRs in gene promoter/body regions were involved



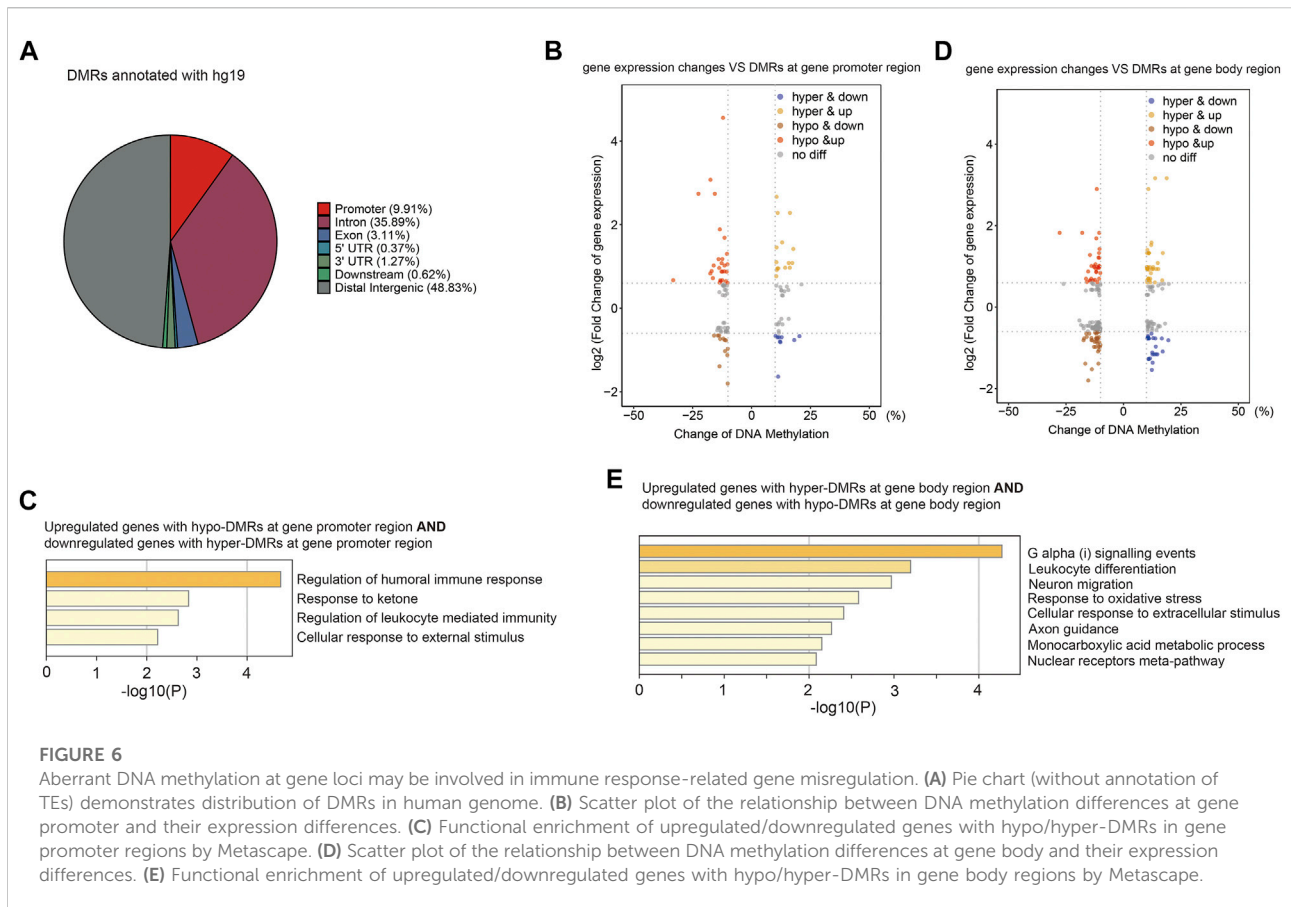
FIGURE 5

WGBS analysis of recovery and control group revealed comparable DNA methylation profiles and genome-wide DMRs which mainly localized at TE loci. (A) The whole mean methylation level per sample of 18M CpG sites for Healthy and Recovery group, and each dot represents one sample. Boxplot shows the mean methylation status per samples of Healthy and Recovery group, each dot represents mean methylation status of one sample. (B) Methylation profile of gene features per sample. Methylation levels were measured in each 200 bp interval of a 2 kb region upstream and downstream of all annotated genes. Methylation levels were measured in 10 equally sized bins for CDSs and introns, and 5 equally sized bins for UTRs. (C) Methylation profile of TE subfamilies per sample. Methylation levels were measured in each 200 bp interval of a 2 kb region upstream and downstream of all annotated TEs, then for each TEs methylation levels were measured in 20 equally sized bins. (D) Boxplot of mean methylation status of four major TE classes. LINE, SINE, LTR, and DNA transposons are shown, and each dot represents one sample. (E) Pie chart demonstrates distribution of DMRs in different gene features of human genome and TE subfamilies. Enrich score was calculated by (Feature ratio of DMR)/(Feature ratio of reference genome), sliding windows was made across reference genome with 500 bps consistent with our call DMR strategy. Chi-square tests was used for statistics, enrichment score >1 and p -value < 0.05 was defined as enrichment. (F) Functional enrichment of genes with hyper-DMRs identified at promoter/gene body by Metascape. (G) Functional enrichment of genes with hypo-DMRs identified at promoter/gene body by Metascape. For above boxplots, the median, first, and third quartiles are shown. Two-sided Wilcoxon signed-rank test were used for all the comparisons.

in various signal pathways, immune response, and metabolism (Figures 5F,G).

CG methylation at different genomic loci may play different roles. Without annotation of TE loci, the distribution of identified DMRs in the human genome was as follows: 9.91% in promoters, 40.64% in gene bodies, 49.45% in intergenic and other regions (Figure 6A). We then examined expression levels of identified

genes with DMRs at promoters or gene bodies. We found that nine hyper-DMRs at promoters were associated with downregulated genes, while 28 hypo-DMRs at promoters were associated with upregulated genes (Figure 6B; Supplementary Table S5), and these genes were mainly involved in immune responses (Figure 6C). Meanwhile, 27 hypo-DMRs at gene bodies were associated with gene downregulation, and 13 hyper-DMRs at gene bodies were



associated with gene upregulation (Figure 6D), mainly involved in stress response and related signaling pathways (Figure 6E).

Next, we focused on DMRs between healthy and recovery groups within TE loci, and identified 13,233 DMRs. The percentages of these DMRs were as follows: 36.48% in LINE, 38.23% in SINE, 15.71% in LTR, and 9.58% in DNA transposon (Figure 7A). Subfamilies of SINE with altered DNA methylation were mainly Alu, subfamilies of LINE were mainly LINE1, subfamilies of LTR were mainly ERVL-MaLR, and subfamilies of DNA transposon were mainly hAT-Charlie (Figure 7B). We then analyzed expression changes of different TEs with altered DNA methylation and identified upregulated expression of TEs (which are mainly localized at introns) associated with both hyper- and hypo-DMRs in TEs (Figure 7C). Upregulation of certain TEs without hypo-DMRs could be explained by aberrant expression of specific transcriptional regulators. DMRs at TE loci are mainly distributed at intergenic regions, intron, and promoter regions (Figure 7D). To explore the potential impact of DNA methylation changes of intergenic TEs on gene expression, we identified adjacent genes of intergenic TEs with altered DNA methylation and calculated total normalized gene counts for each sample. Interestingly, we observed significant upregulation of genes adjacent to TEs with increased DNA methylation, suggesting

that these intergenic TEs act as distal gene silencers (Figure 7E). Further analysis revealed that 19 hypo-DMRs annotated at TE loci were located at the promoter of upregulated genes, and nine hyper-DMRs annotated at TE loci were located at the promoter of downregulated genes (Figure 7F). GO analysis showed their function in ERK signaling regulation and T cell activation (Figure 7G). Furthermore, we identified DMRs at promoter regions that were annotated as various TE subfamilies, although most of those TE subfamilies were not more enriched in DMRs at the gene promoter region relative to other regions in the human genome (Figure 7H). One example is the SIK1 gene which was upregulated in the recovery group and its promoter contains hypo-DMR overlapped with SINE (Figure 7I). These results indicated that TEs with altered DNA methylation may function as regulatory elements for adjacent gene expression.

Discussion

Our previous study on SARS-CoV-2-infected human cell lines showed viral infection-induced gene misregulation and upregulation of TEs (Yin et al., 2021). However, how TEs behave in the human body remain elusive. To identify how the transcriptional program responds

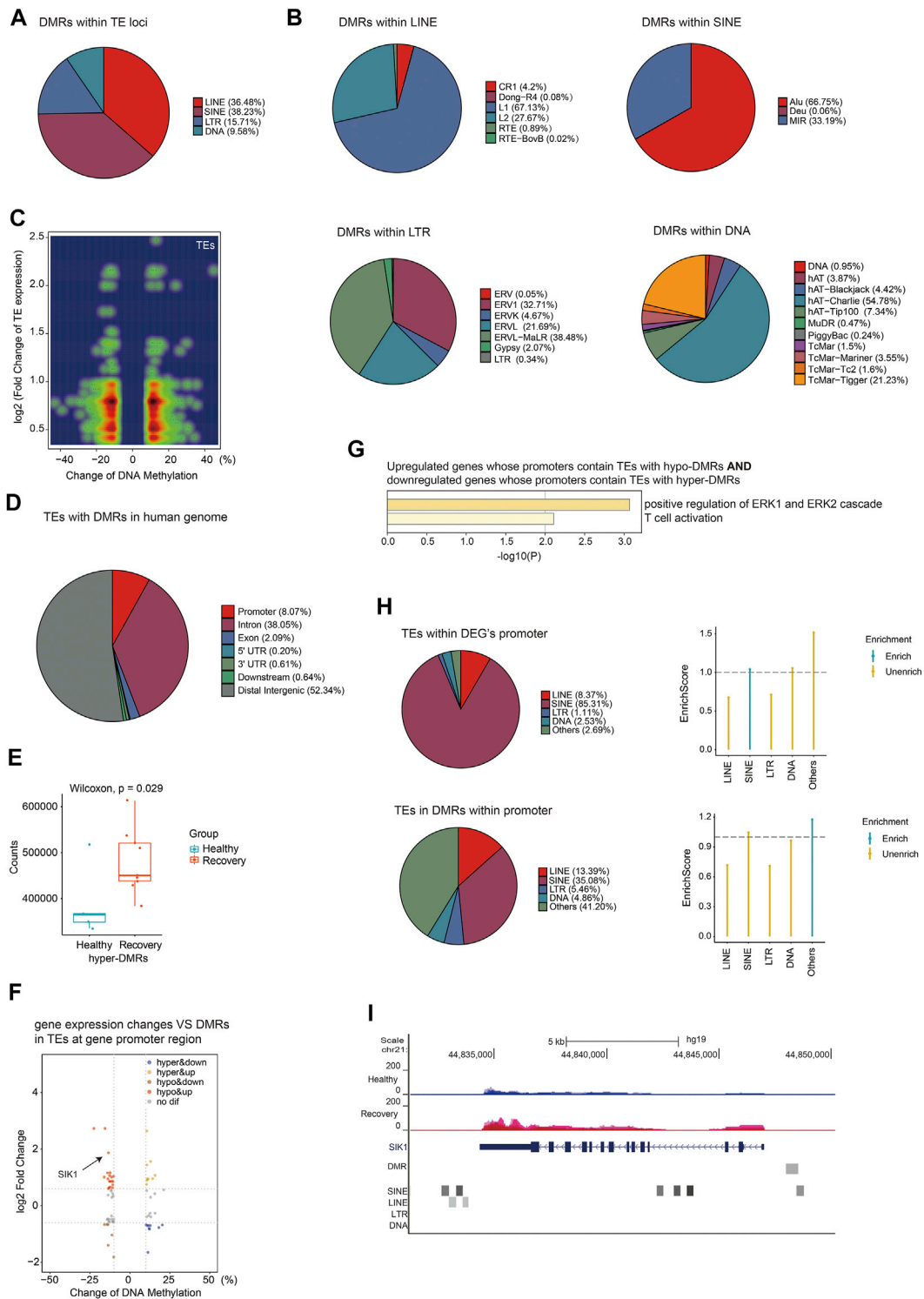


FIGURE 7
 Aberrant DNA methylation at TE loci may be involved in gene regulation. **(A)** Pie chart demonstrates distribution of DMRs in TE subfamilies. **(B)** Pie chart demonstrates distribution of DMRs in LINE/SINE/LTR/DNA transposons. **(C)** Density scatter plot demonstrates relationship between DNA methylation differences of TEs and their expression differences. **(D)** Pie chart (without annotation of TEs) demonstrates distribution of TEs with DMRs in human genome. **(E)** Calculation of total gene counts adjacent to TEs with hyper-DMRs in healthy and recovery group. The median, first, and third quartiles are shown in boxplots. Two-sided Wilcoxon signed-rank test were used for all the comparisons. **(F)** Scatter plot of the relationship between DNA methylation differences at TEs and expression differences of genes with promoters containing TEs with DMRs. **(G)** Functional
 (Continued)

FIGURE 7 (Continued)

enrichment of upregulated/downregulated genes whose promoters contain TEs with hypo/hyper-DMRs by Metascape. **(H)** The upper pie chart demonstrates the distribution of TEs overlapped with DEGs' promoters in TE subfamilies. Enrichment score was calculated by (TE subfamily ratio of TEs overlapped with DEGs' promoters)/(TE subfamily ratio of TEs overlapped with all genes' promoters). The bottom pie chart shows the distribution of DMRs overlapped with genes' promoters in TE subfamilies. Enrichment score was calculated by (TE subfamily ratio of DMRs overlapped with genes' promoters)/(TE subfamily ratio of gene promoters). Sliding windows was made across gene promoters with 500 bps consistent with our call DMR strategy. Chi-square tests was used for statistics, enrichment score >1 and p -value < 0.05 was defined as enrichment. **(I)** UCSC genome browser view of RNA-seq data demonstrated increased SIK1 expression in recovery group and position of hypo-DMR in SIK1 promoter.

to SARS-CoV-2 in the human body, we downloaded and analyzed a public RNA-seq dataset of peripheral blood samples from COVID-19 patients in the acute phase (Bernardes et al., 2020). As anticipated, we observed both misregulation of genes and aberrantly activation of TEs.

Despite extensive investigations, the outcome of SARS-CoV-2 infection, long-term health consequences (Nabavi, 2020), and long-term recovery progress of COVID-19 patients remain elusive. The majority of the COVID-19-related DMRs are near the gene promoter regions and were hypo-methylated even though the global methylation level remains similar between healthy control and COVID-19 patients (Balmis et al., 2021). However, currently reported results on DNA methylome of COVID-19 patients (Bernardes et al., 2020; Balmis et al., 2021) all depended on 450K or 850K methylation array which covered only a small percent of CpG sites (1%–3%) and could not be used to obtain DNA methylation information of transposable elements. Our WGBS data produced nearly whole-genome CpG sites coverage (coverages of all samples are around 80%) which is valuable for genome-wide identification of differentially methylated regions, especially for transposable elements. Here, we asked whether 3 months was enough to restore the transcriptome and DNA methylation of the patient's peripheral blood cells to normal. Based on our results, even though the global methylation level remains the same between healthy control and 3-months recovered patients, the overall transcriptome or epigenome profile of peripheral blood samples from the recovery group is still sufficiently different from the control group.

Our result shows that genes involved in leukocyte differentiation are enriched in both upregulated and downregulated genes, and this means that leukocyte differentiation was disturbed (Supplementary Table S6). In upregulated genes, we observed the RELB gene which is the key component of the NF-kappa-B complex, and the NF-kappa-B pathway has been reported to play an important role in leukocyte differentiation (Maslova et al., 2020). We also found upregulation of BCL3 and NFKBIZ. BCL3 is a transcriptional activator that promotes the transcription of NF-kappa-B target genes (Bours et al., 1993), and NFKBIZ is involved in the regulation of NF-kappa-B transcription factor complexes (Cowland et al., 2006; Totzke et al., 2006). Notably, we noticed the downregulation of LILRB4 which was reported to interfere with NF-kappa-B upregulation (Chang et al., 2002). Moreover, downregulated genes also include CEBPA and HDAC9 which are coactivators/corepressors and participate in transcriptional regulation to control cell proliferation and differentiation, and they may orchestrate the expression of NF-kappa-B targets in a context-dependent manner. Therefore, some

immune response-related biological processes like leukocyte differentiation were severely disturbed instead of being simply activated or inhibited in recovery COVID-19 patients, through pathways such as the NF-kappa-B pathway. For patients in the acute phase and recovery stage, 48 genes mainly involved in immune responses were consistently misregulated, implicating their long-term involvement in cellular responses to SARS-CoV-2. Interestingly, those genes include transmembrane receptors like PILRB, voltage-gated channel protein HVCN1, and transcription factors like BCL6 and NR4A1, etc. PILRB is a cellular signaling activating receptor and is involved in the regulation of the immune system. Upregulation of PILRB in recovery patients indicates its involvement in orchestrating immune cell types. Voltage-gated channel protein HVCN1 is expressed in immune cells and its mRNA level was downregulated in the recovery group. HVCN1 facilitates reactive oxygen species generation in phagocytosis during immune responses and its downregulation might reflect inefficient immune responses upon SARS-CoV-2 infection, leading to an extended time before full recovery. We also detected upregulated transcription factors BCL6 and NR4A1. BCL6 is a zinc finger transcription factor and regulates the transcriptional activity of STAT-dependent IL-4 responses in B cells. NR4A1 is expressed in human lymphocytes and regulates transcription in response to stress stimuli. These two transcription factors might play key roles in controlling the cell fate of immune cells in COVID-19 patients and contribute to long-term COVID-19 symptoms. Notably, sub-families of TEs remained upregulated 3 months after recovery, suggesting that a longer timeframe may be needed to return to normal levels. Retrotransposons can encode proteins and form retrovirus-like particles (Grow et al., 2015), so it is possible that some virus-like particles visualized by electron microscopy (EM) in COVID-19 patients (Yao et al., 2021) may derive from TEs like LTRs due to their enhanced expression rather than SARS-CoV-2.

While some aberrant gene expression can be interpreted by DNA methylation changes, other mechanisms undermining the transcriptional network in the human body need further exploration. Our WGBS analysis showed no significant changes in global DNA methylation and no significant changes in DNA methylation at TE subtypes, and this may be caused by bulk and heterogeneous levels of cells. Interestingly, we still found 18,516 DMRs between the healthy and recovery groups, and 13,233 DMRs were within TE loci. This supported altered TE expression and changes in DNA methylation at TE loci. Next, we conducted PPI network

analysis using STRING database to explore gene interaction at protein level. We used significantly downregulated/upregulated genes in RNA-seq result which also overlapped with DNA methylation changes. As shown in [Supplementary Figure S6](#), several genes did have strong interactions. For example, we found that the chemokine receptor CCR7 was downregulated with hyper-methylation at the promoter, and transcription factor RUNX3 was upregulated with hyper-methylation in the gene body in recovery COVID-19 patients. This finding supports that RUNX3 protein was increased to repress mRNA level of CCR7, because it was previously reported that RUNX3 protein inhibits *Ccr7* expression immune cells in mice ([Fainaru et al., 2005](#)). It should be noted that there was a contribution of cell population differences between healthy and recovery groups in identified genes/TEs which showed differential expression patterns or DNA methylation patterns. For example, healthy and recovery individuals may have different circulating memory T cells/exhausted T cell features. However, at least part of those changes at TE loci should be due to TE activation in acute and recovered patients. One supporting evidence of TE activation is that our previous study on human cell lines showed upregulation of TE upon SARS-CoV-2 infection ([Yin et al., 2021](#)). Another important supporting evidence is that our analysis of RNA-seq data of acute patients showed activation of the cGAS-STING pathway (see [Figure 1B](#) for cGAS upregulation in acute patients) which can be triggered by upregulation of retrotransposon-derived cytoplasmic DNA ([Decout et al., 2021](#)). Interestingly, a recent preprint manuscript showed that viral load of SARS-CoV-2 was positively correlated with TE upregulation ([Sorek et al., 2021](#)), and this is in agreement with our result that TE activation was positively correlated with COVID-19 severity at the recovery stage. However, the study also showed that TE activation was relatively mild during SARS-CoV-2 infection, and this may be partially explained by the cGAS-STING pathway which can be activated upon SARS-CoV-2 infection and may be responsible for the degradation of cytoplasmic retrotransposon-derived DNA to inactivate retrotransposons. We did not identify significant activation of cGAS gene in recovered patients, probably because TE activation is not severe enough to activate the cGAS-STING pathway. Besides transcriptomes and DNA methylomes, whether COVID-19 leaves other irreversible sequelae requires further investigation, such as telomere length (TL) and mitochondrial DNA (mtDNA) copy number, which is associated with many diseases including cardiovascular diseases, psychiatric disorders, cancers, and inflammatory diseases ([Shay, 2016](#); [Sun and St John, 2016](#)).

Our study reveals genes with aberrant expression and genomic regions with altered epigenetic modification in COVID-19 patients 3 months after recovery. Our results support the long-term disease stage marked by an overactivated immune response. Besides, this report provides potential genomic targets to facilitate the convalescence of

COVID-19 patients. Moreover, we provide potential transcriptional and epigenetic signatures to track SARS-CoV-2 infection history, identify the profound viral impact on human cells and reveal long COVID-19 risks. However, due to genetic mutations of SARS-CoV-2 variants, gender, health, age of patients, and other factors, transcriptional and epigenetic changes may vary and need further validation and investigations.

In summary, we examined the transcriptomes and DNA methylomes of COVID-19 patients 3 months after recovery, and noticed that both genes and TEs were impacted at transcription and DNA methylation levels. Misregulated genes were involved in immune response and other biological functions, such as stress response, and metabolic processes, while TEs in intron and other regions were specifically activated which may disrupt the transcription process and genome integrity and induce inflammation. Furthermore, DNA methylome analysis showed that genes with DMRs were also involved in immune response-related processes; and differentially methylated promoter and distal intergenic region may play important roles in gene regulation. Finally, altered CG methylation can indirectly impact gene expression and may play regulatory roles in stress, illness, and aging. Further studies are needed to track changes in transcriptome and DNA methylome of COVID-19 patients for a longer time to identify how long is required for a full recovery.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Bioproject PRJCA006301, <https://ngdc.cnbc.ac.cn/bioproject/browse/PRJCA006301>.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee for Clinical Research of Reproductive Medicine Center, Tongji Medical College, Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

Author contributions

L-QZ, XH, H-GL, and C-LX conceived and designed the project. YY, QT, ZY, and T-QM collected the blood samples. YY and X-ZL analyzed the data and wrote the manuscript. Y-XF and G-HW helped with data analysis. L-QZ, XH and H-GL revised

the manuscript. All authors have read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [NSFC 32000488, to YY, NSFC 32170820, to L-QZ], Fundamental Research Funds for the Central Universities, HUST [2021GCRC073, to XH, 2019kfyXJJS083, to YY], HUST COVID-19 Rapid Response Call [2020kfyXGYJ057, to H-GL], the National Key R&D Program of China [2018YFC1004500 and 2018YFC1004000, to L-QZ], and program for HUST Academic Frontier Youth Team [to L-QZ].

Acknowledgments

The authors would like to express their thanks to Drs. Hui Zhou, Xing-jie Zhang, and Zhen-Yu Zhong for their assistance on this study.

References

- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., et al. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13 (10), R87. doi:10.1186/gb-2012-13-10-r87
- Balnis, J., Madrid, A., Hogan, K. J., Drake, L. A., Chieng, H. C., Tiwari, A., et al. (2021). Blood DNA methylation and COVID-19 outcomes. *Clin. Epigenetics* 13 (1), 118. doi:10.1186/s13148-021-01102-9
- Bernardes, J. P., Mishra, N., Tran, F., Bahmer, T., Best, L., Blase, J. I., et al. (2020). Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* 53 (6), 1296–1314.e9. doi:10.1016/j.immuni.2020.11.017
- Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Moller, R., et al. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* 181 (5), 1036–1045. doi:10.1016/j.cell.2020.04.026
- Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., et al. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 583 (7816), 469–472. doi:10.1038/s41586-020-2332-7
- Bours, V., Franzoso, G., Azarenko, V., Park, S., Kanno, T., Brown, K., et al. (1993). The oncoprotein Bcl-3 directly transactivates through kappa B motifs via association with DNA-binding p50B homodimers. *Cell* 72 (5), 729–739. doi:10.1016/0092-8674(93)90401-b
- Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., et al. (2020). Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 370 (6518), 856–860. doi:10.1126/science.abd2985
- Chang, C. C., Ciubotariu, R., Manavalan, J. S., Yuan, J., Colovai, A. I., Piazza, F., et al. (2002). Tolerization of dendritic cells by T(S) cells: The crucial role of inhibitory receptors ILT3 and ILT4. *Nat. Immunol.* 3 (3), 237–243. doi:10.1038/ni760
- Chen, Y., Gu, S., Chen, Y., Lu, H., Shi, D., Guo, J., et al. (2021). Six-month follow-up of gut microbiota richness in patients with COVID-19. *Gut* 71, 222–225. doi:10.1136/gutjnl-2021-324090
- Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* 10 (10), 691–703. doi:10.1038/nrg2640
- Cowland, J. B., Muta, T., and Borregaard, N. (2006). IL-1beta-specific up-regulation of neutrophil gelatinase-associated lipocalin is controlled by I kappa B-zeta. *J. Immunol.* 176 (9), 5559–5566. doi:10.4049/jimmunol.176.9.5559
- Decout, A., Katz, J. D., Venkatraman, S., and Ablasser, A. (2021). The cGAS-STING pathway as a therapeutic target in inflammatory diseases. *Nat. Rev. Immunol.* 21 (9), 548–569. doi:10.1038/s41577-021-00524-z
- Delorey, T. M., Ziegler, C. G. K., Heimberg, G., Normand, R., Yang, Y., Segerstolpe, A., et al. (2021). COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* 595 (7865), 107–113. doi:10.1038/s41586-021-03570-8
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35 (1), 41–48. doi:10.1038/ng1223
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635
- Domizio, J. D., Gulen, M. F., Saidoune, F., Thacker, V. V., Yatim, A., Sharma, K., et al. (2022). The cGAS-STING pathway drives type I IFN immunopathology in COVID-19. *Nature* 603 (7899), 145–151. doi:10.1038/s41586-022-04421-w
- Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., et al. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* 20 (3), 332–338. doi:10.1038/nsmb.2495
- Fainaru, O., Shseyov, D., Hantisteanu, S., and Groner, Y. (2005). Accelerated chemokine receptor 7-mediated dendritic cell migration in Runx3 knockout mice and the spontaneous development of asthma-like disease. *Proc. Natl. Acad. Sci. U. S. A.* 102 (30), 10598–10603. doi:10.1073/pnas.0504787102
- Gilbert, N., Lutz-Prigge, S., and Moran, J. V. (2002). Genomic deletions created upon LINE-1 retrotransposition. *Cell* 110 (3), 315–325. doi:10.1016/s0092-8674(02)00828-0
- Gomersall, C. D., Joynt, G. M., Lam, P., Li, T., Yap, F., Lam, D., et al. (2004). Short-term outcome of critically ill patients with severe acute respiratory syndrome. *Intensive Care Med.* 30 (3), 381–387. doi:10.1007/s00134-003-2143-y
- Gorbanova, V., Seluanov, A., Mita, P., McKerrow, W., Fenyo, D., Boeke, J. D., et al. (2021). The role of retrotransposable elements in ageing and age-associated diseases. *Nature* 596 (7870), 43–53. doi:10.1038/s41586-021-03542-y
- Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522 (7555), 221–225. doi:10.1038/nature14308

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.1001558/full#supplementary-material>

- Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382 (18), 1708–1720. doi:10.1056/NEJMoa2002032
- Hammal, F., de Langen, P., Bergon, A., Lopez, F., and Ballester, B. (2022). ReMap 2022: A database of human, mouse, Drosophila and arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res.* 50 (1), D316–D325. doi:10.1093/nar/gkab996
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181 (2), 271–280. doi:10.1016/j.cell.2020.02.052
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395 (10223), 497–506. doi:10.1016/S0140-6736(20)30183-5
- Jin, Y., and Hammell, M. (2018). Analysis of RNA-seq data using TETranscripts. *Methods Mol. Biol.* 1751, 153–167. doi:10.1007/978-1-4939-7710-9_11
- Krueger, F., and Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27 (11), 1571–1572. doi:10.1093/bioinformatics/btr167
- Li, B., and Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinforma.* 12, 323. doi:10.1186/1471-2105-12-323
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Malki, S., van der Heijden, G. W., O'Donnell, K. A., Martin, S. L., and Bortvin, A. (2019). A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. *Dev. Cell* 51 (5), 658. doi:10.1016/j.devcel.2019.11.011
- Maslova, A., Ramirez, R. N., Ma, K., Schmutz, H., Wang, C., Fox, C., et al. (2020). Deep learning of immune cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 117 (41), 25655–25666. doi:10.1073/pnas.2011795117
- Nabavi, N. (2020). Long Covid: How to define it and how to manage it. *BMJ* 370, m3489. doi:10.1136/bmj.m3489
- Newkirk, S. J., Lee, S., Grandi, F. C., Gaysinskaya, V., Rosser, J. M., Vanden Berg, N., et al. (2017). Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 114 (28), E5635–E5644. doi:10.1073/pnas.1701069114
- Padmanabhan Nair, V., Liu, H., Ciceri, G., Jungverdorben, J., Frishman, G., Tchieu, J., et al. (2021). Activation of HERV-K(HML-2) disrupts cortical patterning and neuronal differentiation by increasing NTRK3. *Cell Stem Cell* 28, 1566–1581.e8. doi:10.1016/j.stem.2021.04.009
- Percharde, M., Lin, C. J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., et al. (2018). A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* 174 (2), 391–405. e319. doi:10.1016/j.cell.2018.05.043
- Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. doi:10.1093/nar/gku365
- Shay, J. W. (2016). Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* 6 (6), 584–593. doi:10.1158/2159-8290.CD-16-0062
- Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182 (1), 59–72. e15. doi:10.1016/j.cell.2020.05.032
- Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., et al. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* 111 (51), E5593–E5601. doi:10.1073/pnas.1419161111
- Sorek, M., Meshorer, E., and Schlesinger, S. (2021). Impaired activation of transposable elements in SARS-CoV-2 infection. doi:10.1101/2021.02.25.432821v3
- Sun, X., and St John, J. C. (2016). The role of the mtDNA set point in differentiation, development and tumorigenesis. *Biochem. J.* 473 (19), 2955–2971. doi:10.1042/BCJ20160008
- Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., et al. (2002). Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* 110 (3), 327–338. doi:10.1016/s0092-8674(02)00839-5
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47 (1), D607–D613. doi:10.1093/nar/gky1131
- Tian, Y., Sun, K. Y., Meng, T. Q., Ye, Z., Guo, S. M., Li, Z. M., et al. (2021). Gut microbiota may not be fully restored in recovered COVID-19 patients after 3-month recovery. *Front. Nutr.* 8, 638825. doi:10.3389/fnut.2021.638825
- Totzke, G., Essmann, F., Pohlmann, S., Lindenblatt, C., Janicke, R. U., and Schulze-Osthoff, K. (2006). A novel member of the IkappaB family, human IkappaB-zeta, inhibits transactivation of p65 and its DNA binding. *J. Biol. Chem.* 281 (18), 12645–12654. doi:10.1074/jbc.M511956200
- Wu, P., Chen, D., Ding, W., Wu, P., Hou, H., Bai, Y., et al. (2021). The trans-omics landscape of COVID-19. *Nat. Commun.* 12 (1), 4543. doi:10.1038/s41467-021-24482-1
- Xiong, Y., Liu, Y., Cao, L., Wang, D., Guo, M., Jiang, A., et al. (2020). Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* 9 (1), 761–770. doi:10.1080/22221751.2020.1747363
- Yao, X. H., Luo, T., Shi, Y., He, Z. C., Tang, R., Zhang, P. P., et al. (2021). A cohort autopsy study defines COVID-19 systemic pathogenesis. *Cell Res.* 31 (8), 836–846. doi:10.1038/s41422-021-00523-8
- Yin, Y., Liu, X. Z., He, X., and Zhou, L. Q. (2021). Exogenous coronavirus interacts with endogenous retrotransposon in human cells. *Front. Cell. Infect. Microbiol.* 11, 609160. doi:10.3389/fcimb.2021.609160
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798), 270–273. doi:10.1038/s41586-020-2012-7
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6