



ARTICLE OPEN

Longitudinal transcriptome analyses show robust T cell immunity during recovery from COVID-19

Hong-Yi Zheng¹, Min Xu¹, Cui-Xian Yang², Ren-Rong Tian¹, Mi Zhang², Jian-Jian Li², Xi-Cheng Wang², Zhao-Li Ding³, Gui-Mei Li³, Xiao-Lu Li³, Yu-Qi He³, Xing-Qi Dong², Yong-Gang Yao^{1,4,5} and Yong-Tang Zheng^{1,4}

Understanding the processes of immune regulation in patients infected with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is crucial for improving treatment. Here, we performed longitudinal whole-transcriptome RNA sequencing on peripheral blood mononuclear cell (PBMC) samples from 18 patients with coronavirus disease 2019 (COVID-19) during their treatment, convalescence, and rehabilitation. After analyzing the regulatory networks of differentially expressed messenger RNAs (mRNAs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) between the different clinical stages, we found that humoral immunity and type I interferon response were significantly downregulated, while robust T-cell activation and differentiation at the whole transcriptome level constituted the main events that occurred during recovery from COVID-19. The formation of this T cell immune response might be driven by the activation of activating protein-1 (AP-1) related signaling pathway and was weakly affected by other clinical features. These findings uncovered the dynamic pattern of immune responses and indicated the key role of T cell immunity in the creation of immune protection against this disease.

Signal Transduction and Targeted Therapy (2020)5:294

; <https://doi.org/10.1038/s41392-020-00457-4>

INTRODUCTION

Since December 2019, a new zoonotic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has swept the world, causing a variety of clinical syndromes collectively termed coronavirus disease 2019 (COVID-19).^{1–3} The World Health Organization declared a pandemic in March 2020. The symptoms of COVID-19 are fever, dry cough, fatigue, diarrhea, conjunctivitis, and pneumonia.¹ Most people do seem to be less affected, either remaining totally asymptomatic or having only mild symptoms. However, some people develop a severe pneumonia, acute respiratory distress syndrome (ARDS) or multiple organ failure.^{2,4} It is currently believed that severe COVID-19 pathogenesis may be mediated by a unique immune response disorder, and the host antiviral immune response affects the severity of the disease and the clinical outcome.^{5,6}

The immune pathology caused by SARS-CoV-2 and the immune protection against COVID-19 had received extensive attention. Recent studies had shown that immune system disorders, such as lymphocytopenia and inflammatory cytokine storm, are associated with the severity of the SARS-CoV-2 infection.^{1,7} Type I interferon (IFN-I) not only has the ability to clear the virus, but also can cooperate with inflammatory factors to promote the severe development of COVID-19.^{8,9} In particular, the inflammatory, exhausted and activated state of T cells affects the severity of COVID-19 symptoms, whereas a robust T cell immune response in patients may affect the predisposition to disease and also prevents re-infection.^{10–12} The humoral immunity against SARS-CoV-2 works

rapidly but may not provide long-lasting immunity, as revealed by recent evaluation of the decay of anti-SARS-CoV-2 antibodies in patients with COVID-19.^{13,14} These studies had greatly expanded our understanding of COVID-19 pathophysiology and immunology. However, we are still unclear about the main events of the immune regulation process of patients during their recovery from illness.

Transcriptome analyses are very suitable for the study of viral infection immunology and allow for an understanding of the immune response dynamics and gene regulatory networks. Recent transcriptome analyses of patients with COVID-19 have shown the dynamics of the immune responses following infection with SARS-CoV-2.^{6,7,9,15,16} Here we characterized the longitudinal transcriptome changes in peripheral blood mononuclear cells (PBMC) of 18 COVID-19 patients with mild, moderate or severe symptoms at three clinical stages (treatment, convalescence, and rehabilitation). Our results showed the immune remodeling processes in patients at the different stages of their illness and revealed the core role of T cell immunity.

RESULTS

Clinical features of SARS-CoV-2 infected patients with mild, moderate or severe symptoms during recovery from COVID-19 From January 17 to February 19, 2020, 18 patients (11 male and 7 female) at the Yunnan Infectious Disease Hospital, Kunming, China, with COVID-19 and SARS-CoV-2 infection confirmed by

¹Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Center for Biosafety Mega-Science, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China; ²Yunnan Infectious Disease Hospital, Kunming 650301, China; ³Kunming Biological Diversity Regional Center of Large Apparatus and Equipments, Public Technical Service Center, Chinese Academy of Sciences, Kunming, Yunnan 650223, China; ⁴Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650204, China and ⁵CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China

Correspondence: Xing-Qi Dong (dongxq801@126.com) or Yong-Gang Yao (yaoyg@mail.kiz.ac.cn) or Yong-Tang Zheng (zhengyt@mail.kiz.ac.cn)

These authors contributed equally: Hong-Yi Zheng, Min Xu, Cui-Xian Yang

Received: 26 September 2020 Revised: 17 November 2020 Accepted: 23 November 2020

Published online: 24 December 2020

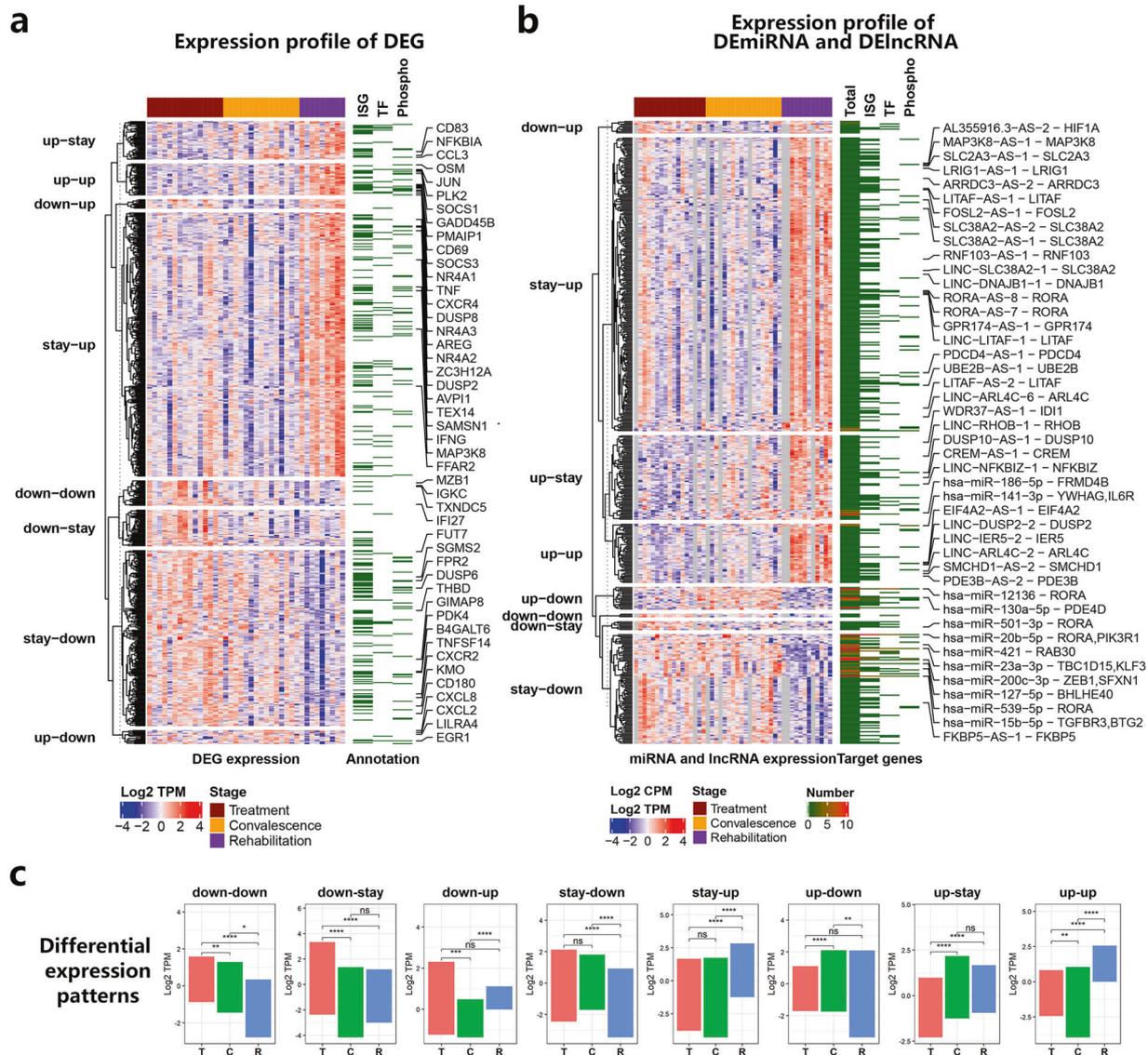


Fig. 2 Differentially expressed genes (DEG), miRNAs (DEmiRNA), and lncRNAs (DElncRNA) in PBMC samples from patients with COVID-19 during the recovery. **a** A heatmap showing the expression profiles of 643 DEGs. The blue-red gradient square maps the scaled \log_2 value of transcripts per kilobase million (\log_2 TPM) for each DEG in a sample. The grouping information on the left indicates genes with different differential expression patterns. The colored bars above the heatmap refer to the clinical stage of each sample. The gene function annotation on the right marks the interferon-stimulated genes (ISG), transcription factor (TF) and phosphorylation regulatory genes (Phospho) in DEGs. The hub genes with a more significant differential expression in the function networks constructed by all DEGs are marked with the respective gene names. **b** A heatmap showing the expression profiles of 67 DEmiRNAs and 405 DElncRNAs. The blue-red gradient square maps the scaled \log_2 TPM of each DElncRNA, or the scaled \log_2 value of counts per million (\log_2 CPM) of each DEmiRNA in a sample. The gene function annotation shows the number of target genes for each non-coding RNA. When the target gene is a DEG, we annotate the regulation module in the format of “non-coding RNA name - target gene name” on the right. **c** By comparing the P value and TPM/CPM at three different clinical stages, DEG, DEmiRNA or DElncRNA can be further clustered into 8 expression patterns. Each pattern reflects the characteristics of its RNAs being upregulated (up), downregulated (down) or stayed for no change (stay) at the convalescence (C) and rehabilitation (R) stages relative to the treatment (T) stage. Ns, *, **, ***, ****: $p > 0.05, \leq 0.05, \leq 0.001, \leq 0.0001$, respectively

IFI27, OAS1, IFIT3, FADD, and RSAD2 (Fig. 3b). These observations suggested that immune regulation plays a key role in the remodeling of innate and adaptive immunity during the recovery from COVID-19.

A large number of gene sets involved in immune cell differentiation were also found to be upregulated during recovery from COVID-19, including autophagy, cell cycle, cytokine production, regulation of T cell differentiation, response to interleukin-1, immune system development, and regulation of NIK/NF- κ B signaling. Furthermore, decrease of the enrichment scores of the

gene sets related to blood coagulation and platelet activation also reflected that these comorbidities of abnormal blood coagulation are being reduced in these patients (Fig. 3a; Supplementary Fig. 2).

Elevated T cell activation and differentiation are core events during recovery from COVID-19
A total of 8 differential expression patterns were clearly discerned in patients with COVID-19 at the treatment, convalescence, and rehabilitation stages, respectively (Fig. 2c). Using enrichment analysis grouped by these expression patterns, we found that

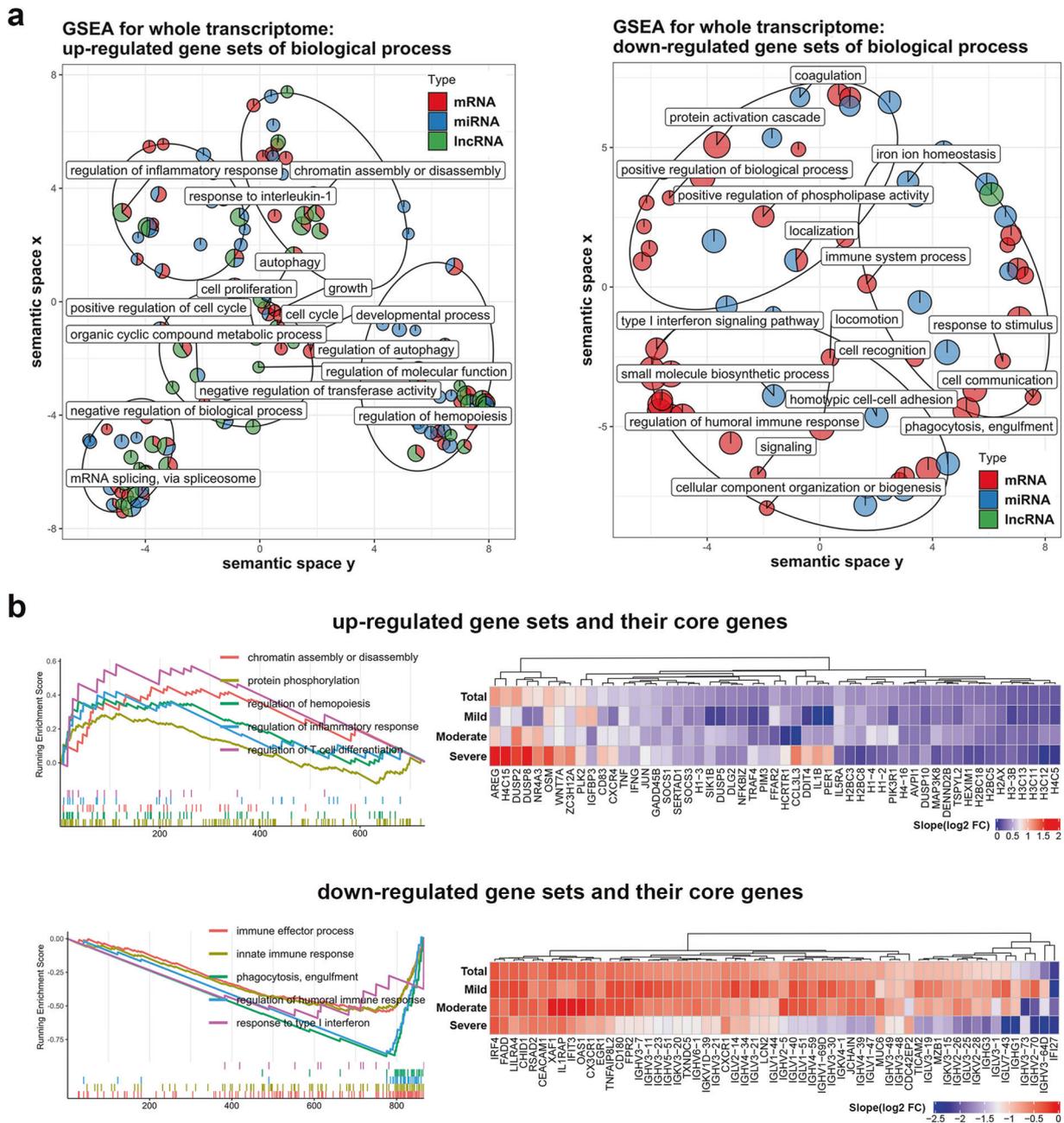


Fig. 3 Gene set enrichment analysis (GSEA) showing an immune system remodeling at the whole transcriptome level. a The bubble-pie graphs displaying upregulated and downregulated gene ontology (GO) gene sets enriched by the differentially expressed mRNAs, miRNAs, and lncRNAs. The color refers to the indicated type of transcripts enriched for a gene set. The bubble size maps the sum enrichment scores of all RNA types in a gene set. The X and Y values of a gene set are calculated by REVIGO's GO semantic space algorithm.⁶⁹ The ellipses refer to clusters formed by gene sets with a high semantic similarity. **b** GSEA plot showing the distribution of the upregulated or downregulated gene sets and the enrichment scores based on DEGs for all samples. The heatmap shows the core genes in these gene sets of different clinical stages for all samples, mild, moderate and severe groups, respectively. The blue-red gradient square maps the slope of the linear regression equation of Log₂ fold change of DEG

DEGs, which are downregulated only at the rehabilitation stage (stay-down), enriched by a large number of antiviral genes and IFN-I signaling genes (e.g., *OAS1*, *IFIT3*, *AIM2*, *IFIT1*, and *FADD*), positive regulatory genes of cytokine signaling pathway (e.g., *TLR2*, *IL6R*, *P2RX7*, *RSAD2* and *AIM2*), as well as some chemokine receptor genes (e.g., *CCR1*, *CCR5*, *CMKLR1*, *CX3CR1*, and *CXCR2*). However, humoral immunity was attenuated earlier than innate immunity. DEGs that downregulated from the convalescence stage to the rehabilitation stage (down-stay or down-down) were enriched in GO terms such as humoral immunity, B cell immunity,

B cell activation and Fc receptor signaling (Fig. 4a, Supplementary Fig. 3).

One prominent feature of the enrichment analysis was that many DEGs upregulated in the convalescence, and rehabilitation stages (stay-up or up-stay) are involved in T cell activation and differentiation, such as *PIK3R1*, *IFNG*, *CYLD*, *RHOH*, *BCL3*, *MAP3K8*, and *ICOS*. Moreover, a group of DEGs that upregulated in the convalescence stage and then continued to be upregulated in the rehabilitation stage (up-up) were mainly composed of genes that regulate protein kinase activity,

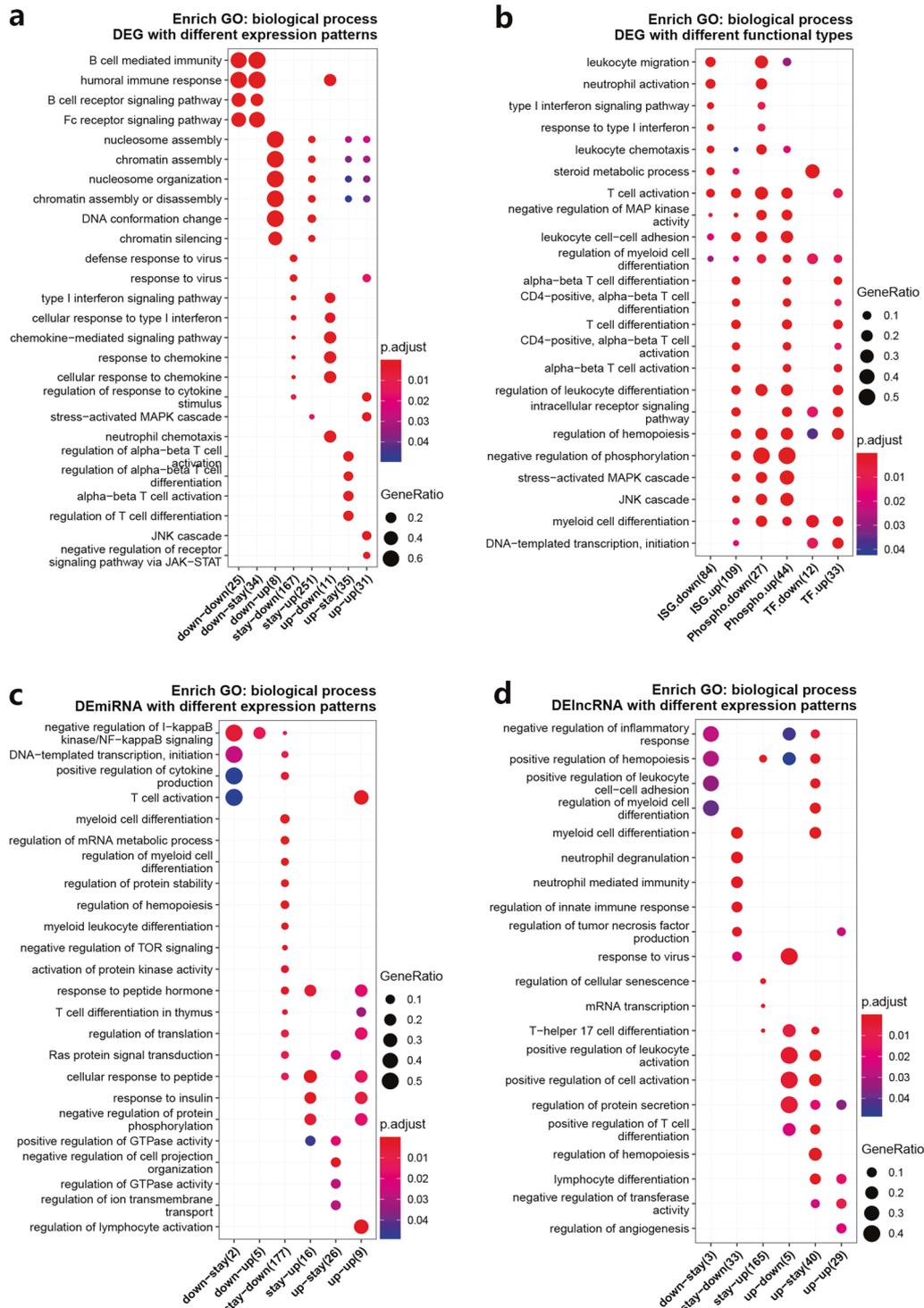


Fig. 4 Gene ontology (GO) enrichment analysis identified T cell immunity as the main regulatory target during the recovery of COVID-19. **a** GO enrichment results of DEGs in 8 differential expression clusters defined in Fig. 2b. The numbers of significantly enriched genes in each cluster were included in parentheses. **b** GO enrichment results of interferon-stimulated genes (ISG), transcription factor genes (TF), and phosphorylation regulatory genes (Phospho) in DEGs. **c, d** GO enrichment results of the genes targeted by DEmiRNAs (**c**) and DElncRNAs (**d**) in 6 differential expression clusters defined in Fig. 2c. The bubble size indicates the gene enrichment ratio (Generatio) of a biological process GO term, with color maps the FDR value (p.adjust) of the enrichment analysis

including *TNF*, *CXCR4*, *SOCS1*, *JUN*, *SOCS3*, *PDCD4*, *DUSP8*, etc. These genes are actively involved in the regulation of MAP kinase activity and JAK-STAT signaling pathway (Fig. 4a, Supplementary Fig. 3), suggesting that the T-cell immunity was elevated during recovery from COVID-19 under a

chronological regulation. Only 69 DEGs showed significant differences between each comparison of mild, moderate and severe groups. These genes were mainly enriched in GO terms of negative regulation of phosphorylation, leukocyte migration and regulation of hematopoiesis (Supplementary Fig. 4).

Multiple levels of regulation participate T cell activation and differentiation during recovery from COVID-19. Further enrichment analysis of DEGs related to interferon, transcription factor and phosphorylation regulation revealed that the activation and differentiation of T cells are the main targets of their regulation. *PRDM1* (up-stay), *RORA* (stay-up), *BCL6* (up-stay), and *ZEB1* (stay-up) acted as important transcription factors for controlling the differentiation of T cells and mediating the differentiation of tissue-resident T cells, Th17 cells, follicular helper T cells, and memory T cells, respectively.^{18–21} For phosphorylation regulatory DEGs, *IFNG* (stay-up), *DUSP10* (up-stay) and *SOCS1* (up-up) were enriched in regulatory T cell differentiation, while stay-down genes *FCGR2B*, *CEACAM1*, and *P2RX7* were enriched in T cell mediated cytotoxicity. Upregulated ISGs, such as *RIPK2* (up-stay), *IFNG* (stay-up), *CD83* (up-stay) and *SOCS1* (up-up), were mostly enriched in T cell activation. Among them, *RORA* (stay-up), *NFKBIZ* (up-stay) and *ZC3H12A* (stay-up) were enriched in Th17 differentiation, while *BCL3* (stay-up) and *BCL6* (up-stay) were enriched in Th2 differentiation (Fig. 4b).

Consistent with the changes of DEGs, there were many upregulated DElncRNAs that positively regulate the differentiation of lymphocytes and T cells at the convalescence and rehabilitation stages. The regulation of T cell differentiation by DEMiRNAs mainly occurred at the rehabilitation stage. The stay-down miRNAs let-7b-5p, miR-103a-2-5p, miR-200c-3p and miR-2115-3p were significantly downregulated, while their target genes (*RASGRP1*, *CDK6*, *ZEB1*, and *ATG5*, respectively) were significantly upregulated. The stay-up DElncRNAs *RORA-AS-7*, *RORA-AS-8*, *ITPKB-AS-1*, *RUNX1-AS-8*, *STAT3-AS-1*, and *MALT1-AS-2* were enriched in differentiation of T-helper cells. In addition, the stay-up DElncRNAs *MAP3K8-AS-1*, *RASGRP1-AS-1*, *RASGRP1-AS-2*, *LINC-CD47-1*, *LINC-CD44-1*, *PREX1-AS-2*, *TNFRSF1B-AS-1*, and *TNFRSF1B-AS-2* were enriched in T cell activation (Fig. 4c, d).

The AP-1 linked signaling pathway plays a key role in T cell activation and differentiation during recovery from COVID-19. Using the functional network analyses of DEGs, DEMiRNAs and DElncRNAs, we found that the TNF, MAPK, and NF- κ B signaling pathways are the most significantly changed pathways during recovery from COVID-19 (Supplementary Fig. 3–6). These networks are closely linked with inflammatory factor genes *TNF* and *IL1B*, and transcription factor AP-1 subunit gene *JUN*. The AP-1 connected nodes were at the core of the enriched KEGG maps, including Toll-like receptor signaling pathway, IL-17 signaling pathway and PD-L1 expression and PD-1 checkpoint pathway (Fig. 5a). The activation of AP-1 and NF- κ B signals caused an increased expression of inflammatory factors such as TNF- α , IL-1 β , IL-8, MIP-1 α and MIP-1 β according to the KEGG enrichment analysis (Supplementary Fig. 7), which might be detrimental to COVID-19 rehabilitation. Concordantly, for downstream genes of AP-1, the expression levels of some chemokine genes (e.g., *CCL2*, *CXCL1*, *CXCL2*, and *CXCL10*) were downregulated, while the expression levels of immune negative regulation genes (e.g. *BCL3*, *NFKBIA*, *SOCS3*, and *TNFAIP3*) were upregulated during the COVID-19 recovery (Fig. 5b). These results indicated that the activation of the AP-1 linked pathway does not cause excessive inflammation and was more likely to play a role in immune regulation.

The transcription factor AP-1 subunit genes *FOSL2* and *JUNB* were also upregulated at the rehabilitation stage, and the non-coding RNAs (e.g., *FOSL2-AS-1*, *LINC-JUND-1*, and *miR-494-3p*) that regulate these genes were also significantly differentially expressed. Some genes that interact with *JUN*, such as *HIF1A*, can enhance the effector function of T cells and promote virus clearance,²² and were regulated by DElncRNAs *HIF1A-AS-1* and *LINC-HSP90AA1-1*. Other genes need to work with AP-1 for controlling T cell differentiation, such as *RORA*,²³ were regulated by a variety of DElncRNAs, including *RORA-AS-7*, *RORA-AS-8*, *miR-20b-5p*, *miR-12136*, *miR-539-5p*, and *miR-501-3p* (Fig. 5c). In

addition, we found that *JUN*, *PIK3R1*, *IFNG*, *CXCR4*, *SOCS3*, *HIF1A*, and other significantly regulated genes had the highest connectivity in the GO, KEGG, and protein–protein interaction (PPI) networks (Fig. 5d). Therefore, the AP-1 linked signaling pathway acted as a core to regulate T cell activation and differentiation at whole transcriptome level during recovery from COVID-19.

The activation and differentiation of T cells are associated with various clinical features in patients with COVID-19. To explore the impact of other clinical features on recovery-related whole-transcriptome, we made an analysis of the whole transcriptome using a weighted correlation network analysis (WGCNA). A total of 18 expression modules were discerned, in which the black and tan modules were significantly positively correlated with the clinical stage, but had no significant correlations with other clinical features. Oppositely, significant negative correlations with age, clinical type, and comorbidities were found in the light cyan module, but it did not show a significant correlation with the clinical stage (Fig. 6a, b). It seemed that the transcriptome changes during the recovery process are weakly affected by other clinical features. Specifically, the black module contained many genes and non-coding RNAs related to the regulation of phosphorylation, hematopoiesis and transferase activity, such as *ZC3H12A*, *DUSP8*, *CD83*, *JUNB*, *NR4A3*, *OSM*, *SOCS3*, *JUN*, *TNF*, *MAP3K8-AS-1*, *RIPK1-AS-1*, and *FOSL2-AS-1*. The transcripts grouped into the tan module were mostly involved in nucleosome assembly and cell cycle, such as *PLK2*, *ID2*, *H4C15*, *H3C1*, *H1-2*, *H1-3*, *LINC-H2BC12-1*, *RCC1-AS-1*, and *ID2-AS-1*. Many genes in these two modules were differentially expressed genes as described above. Although the GO terms enriched by RNAs in the light cyan module overlapped with the black module, it was more concentrated on GO terms related to T cell activation, T cell differentiation, lymphocyte proliferation and myeloid leukocyte differentiation (Fig. 6c). We also discerned the co-expression pattern of genes belonging to the light cyan module. The hub genes of the light cyan module were enriched in T cell differentiation, such as *SATB1*, *LEF1*, *CCR7*, *CAMK4*, *IL6ST*, *TCF7*, *IL7R*, *SIRPG*, and *TMIGD2*, and lncRNAs of *CCR7-AS-1*, *LEF1-AS-1*, *LINC-CCR7-2*, *LINC-TCF7-1* and *TCF7-AS-1*, and most of them were mainly expressed in low-level differentiated T cells (Fig. 6d).^{24–27} The results showed that various clinical features affect the patient's T cell immunity, but cannot change its tendency to become robust during recovery from COVID-19.

DISCUSSION

In this study, we systematically analyzed the whole transcriptome characteristics of PBMC samples from COVID-19 patients with mild, moderate, and severe symptoms at three different time points during their treatment, convalescence, and rehabilitation. The results showed a robust T-cell immune response, but a weakening innate and humoral immunity during recovery from illness, regardless of the clinical types or disease severity, as demonstrated by the altered levels of mRNAs, miRNAs and lncRNAs involved in T cell activation and differentiation.

Many viruses, including SARS-CoV-2, have evolved escape mechanisms against the antiviral activity of IFN-I.²⁸ Recent studies had shown that although the production of TNF and IL-6 was increased, and the NF- κ B-driving inflammatory response was highly activated, the IFN-I response in seriously ill COVID-19 patients and rhesus monkeys infected with SARS-CoV-2 were still impaired.^{9,29} These findings indicated that IFN-I should be treated as a target for the treatment of COVID-19. In fact, the patients in our study received IFN- α treatment immediately on their admission to hospital and achieved good results. The IFN-I treatment could enhance the cytotoxic function of the patient's NK and CD8+ T cells,^{30,31} inhibited hematopoietic function and caused the alteration of peripheral blood neutrophils and

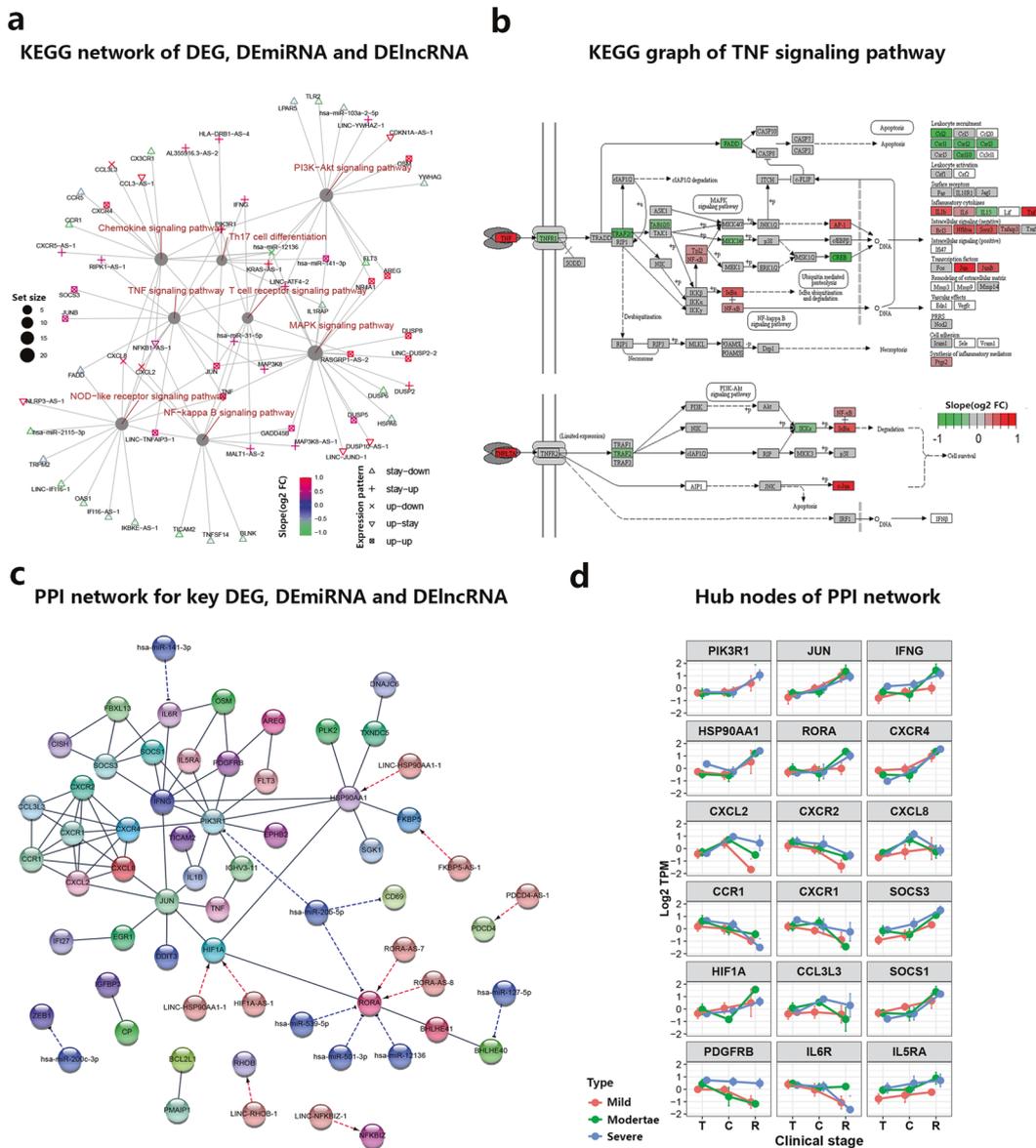


Fig. 5 Gene function network revealed a core role of the AP-1 linked signaling pathways during the recovery of COVID-19. **a** A network showing the relationship between genes and enriched KEGG pathway. The genes displayed are selected from the DEGs and the genes regulated by DE miRNAs or DE lncRNAs with a more significant differential expression and a higher connectivity. The color maps the linear regression slope of Log_2 fold change of a gene. The size of the circle is proportional to the number of genes enriched in a KEGG pathway. **b** A KEGG graph showing the interactions of genes involved in the TNF signaling pathway. The color indicates linear regression slope of Log_2 fold change of a gene. **c** A protein–protein interaction network (PPI) showing the interactions between genes and non-coding RNA. The genes displayed are selected from the top hub genes with the most significant differential expression. **d** A line graph showing the expression profiles of the main hub genes during COVID-19 recovery. The color indicates samples with different clinical types. T, treatment stage; C, convalescence stage; R, rehabilitation stage. The data are shown as mean \pm SD

thrombocytopenia.³² In addition, IFN-I treatment directly or indirectly triggered the signal cascade of JAK-STAT, MAPK, PI3K, and NF- κ B pathways to participate in the regulation of interferon response.^{33–35} Therefore, the changes at the transcriptome levels of our patients might partially come from interferon therapy. However, despite the fact that the patient’s MAPK and NF- κ B signaling pathways were still active, a lot of ISG and the gene sets of positive regulation of T cell cytotoxicity and JAK-STAT signaling pathway were downregulated at the rehabilitation stage, suggesting that the robust regulation of immune response may not directly come from interferon therapy.

Humoral immunity plays an important role in the prevention of infectious diseases, but its role in COVID-19 is still unclear.

A detailed clinical report showed that the circulating antibody-secreting B cells appeared in the blood on the 7th day, and IgM and IgG antibodies gradually increased from the 7th day after infection and maintained high levels until the 20th day.³⁶ However, multiple studies found that in the first 3 months after infection, the level of antibodies against SARS-CoV-2 in patients who have recovered from mild symptoms reduced sharply, with a predicted exhaustion within a year.^{13,37} These findings mean that although antibodies against SARS-CoV-2 appear quickly, most people may not have lasting immunity. The most recent report on re-infection of SARS-CoV-2 in a human seemed to support this speculation.³⁸ In our study, we found that the humoral immune response was attenuated at the convalescence stage

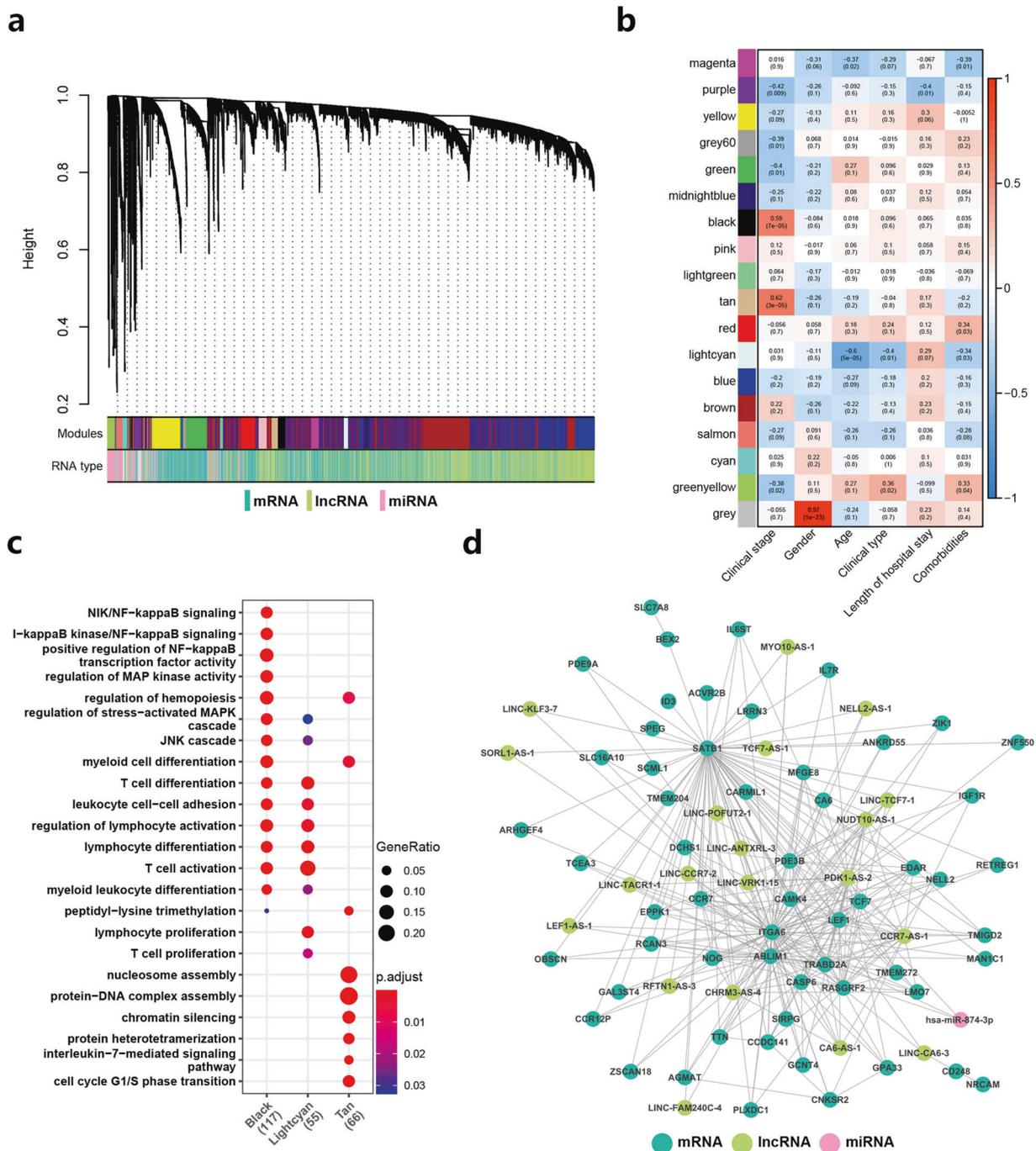


Fig. 6 Weighted correlation network analysis (WGCNA) showed the impact of multiple clinical features on the transcriptomic profiling of peripheral blood cells from patients with COVID-19 during the recovery. **a** A hierarchical clustering tree showing all modules based on top 8000 most variably expressed mRNAs/lncRNAs and top 600 most variably expressed miRNAs in all samples. **b** A matrix plot showing the correlations between modules (Y axis) and clinical factors (X axis). The color and the number outside the brackets of a square indicate the Pearson's correlation r value between a module and the corresponding clinical factor, and the number inside the brackets indicates the Pearson's correlation p value. **c** A bubble plot showing the GO enrichment analysis results of the black, light cyan and tan modules. The numbers in brackets on the X axis indicate the total number of significantly enriched genes in each module. The bubble size indicates the gene enrichment ratio (GeneRatio) of a GO term, with color referring to the FDR value (p.adjust) of enrichment analysis. **d** A correlation network showing the co-expression pattern between genes and non-coding RNAs in the light cyan module

(~20–40 days), and there are no significant differences between patients with varying clinical features. The fast decline of humoral immunity against COVID-19 will pose challenges to vaccination, and the detailed mechanism underlying this awaits further study.

A large number of studies have shown that SARS-CoV-2 infection can cause a more pronounced lymphocytopenia in

patients with moderate and severe COVID-19, as reflected by the decreased level of T cells, including Th1, Treg, and CD8⁺ T cells.^{1,39,40} In addition, neutrophils increased and monocytes decreased in COVID-19 patients.^{39,41} These abnormalities may be reversed in some immune cells after recovery. For example, Sekine, et al.⁴² found that the SARS-CoV-2 specific T cells in acute

infection showed a highly activated cytotoxic phenotype, while these T cells in the recovery stage had multiple functions and exhibited a stem cell-like memory phenotype; Wen, et al.⁴³ showed that the T and B cell clones were highly expanded in COVID-19 recovery. In our study, we also found an enhanced transcription of genes involved in the activation and differentiation of T cells during the recovery from COVID-19, indicating that T cell immunity was strengthened.

Further analysis of the functional network of differentially expressed whole transcriptomes showed that AP-1 signal connects MAPK signal and T cell function and plays a key regulatory role. The MAPK signal cascade is essential for regulating AP-1 transcriptional activation and DNA binding activity, while AP-1 has a pleiotropic effect on the process of T cell activation, differentiation, and exhaustion.⁴⁴ The synergy of NFAT and AP-1 would stimulate gene expression after immune response, including IL-2, IFN- γ , TNF- α , GM-CSF, IL-4, FasL, CD25.⁴⁵ The lack of AP-1 led to the repression of downstream genes and blocked the activation and proliferation of T cells, eventually leading to T cell anergy.⁴⁶ Note that the gene set of chromatin assembly was found to be significantly upregulated during the recovery process. Since AP-1 is present in most activated specific open chromatin regions, the binding of AP-1 and the opening of chromatin during T cell activation are conducive to the formation of super enhancers, and the activity of AP-1 helps the activated T cells to form a specific epigenome.⁴⁷ In our study, we found that active T cell immunity is the main immune feature during recovery from COVID-19, which is compatible with a strong AP-1 signaling.

The impact of the clinical features of COVID-19 patients on their immune response has always received much attention. Several recent studies found that the transcriptome of lung tissue and peripheral blood from severe COVID-19 patients was significantly enriched in a large number of ISGs for IFN-I response and a high level of IFN- α , TNF- α and IL-6 compared to healthy people and mild patients.^{8,48} Regarding immune cell function, the loss of the polyfunctionality and the increased level of exhaustion of T cells were important characteristics that distinguish severe from mild patients.⁴⁹ The expression of genes induced by SARS-CoV-2 infection increased with aging.⁵⁰ At the same time, the activation of the innate immune system, the disorder of the adaptive immune system and the inflammatory signal of the elderly were also increasing.⁵¹ These reasons might account for a higher susceptibility to SARS-CoV-2 in the elderly. There was evidence that during SARS-CoV-2 infection, female patients have stronger levels of T cell activation than male patients, while male patients have higher plasma levels of natural immune cytokines (such as IL-8 and IL-18).⁵² The mechanism by which chronic metabolic comorbidities (such as obesity, type 2 diabetes, and metabolic syndrome) affect COVID-19 has not been fully elucidated, but it was generally believed that it interacts with age and gender to cause immune metabolic disorders and chronic systemic inflammation, which aggravate the excessive inflammation induced by SARS-CoV-2 infection.⁵³ In our study, there were very few DEGs between groups of different clinical types during recovery from COVID-19. The results of WGCNA showed that other clinical features have limited influence on the trend of clinical stage-related transcriptome changes. The pattern that patients would have the same strong T cell immune response after recovery from COVID-19 is very promising for the prevention and treatment of COVID-19 and is conducive to the maintenance of herd immunity.

The current study has several limitations. First, we did not perform a fine-grained immune infiltration analysis of immune cell subpopulations using flow cytometry or single-cell analysis in this study, simply because we did not have sufficient samples for these assays. In addition, the altered expression patterns of the hub genes were not validated by quantitative real-time PCR. Second, as the AP1 signaling is at the core and regulates T cell phenotype, the comparison of T cell phenotypes at different clinical stages,

and correlation analysis between AP1 signal and infiltration abundance or markers of different T cell phenotypes would further enhance the pattern in this study. Third, we had no complete data of the immunological analyses, such as cytokines, antibodies, and biochemical indices, for these patients and could not perform an association analysis of these parameters with the transcriptomic analysis, which would offer some more insights into the dynamic pattern of immune responses during recovery of COVID-19. A focused and well-designed experimental study with rhesus monkey model of SARS-CoV-2 infection would help to clarify these key issues in the future.²⁹

In conclusion, we obtained longitudinal whole transcriptome data from patients with different levels of disease severity and allowed us to present a comprehensive view of the dynamics and transcriptional regulation of peripheral blood immune cells during recovery from COVID-19. The rapid attenuation of innate immunity and humoral immunity at the transcription level were important features, and the regulation of poly-functional memory T cell responses was critical for the treatment. The inability to carry out experimental verification is the main limitation, but the mutual verification at the levels of mRNA, miRNA, and lncRNA improved the credibility of the results. These results have direct implications for modulating T cell immunity in the successful treatment of COVID-19 and the development of more effective ways to prevent SARS-CoV-2 infection.

MATERIALS AND METHODS

Patient cohort and sample collection

A total of 18 patients with COVID-19 were admitted at the Yunnan Infectious Disease Hospital, Kunming, China, and were enrolled in the study from January 17 to April 9, 2020. Peripheral venous blood samples from patients were obtained at the three clinical stages (treatment, convalescence and rehabilitation). PBMC samples were isolated from the fresh peripheral blood by Ficoll-Paque (GE Healthcare) density gradient centrifugation, and were stored at -80°C until the use for total RNA extraction. The detailed clinical features of all patients and the detailed sampling time are shown in Fig. 1 and Supplementary Table 1. All samples are processed in a BSL-2 laboratory qualified for SARS-CoV-2 testing, and in accordance with the laboratory biosafety guide for the novel coronavirus (2nd edition) issued by the National Health Commission of China.

RNA extraction and library preparation

Total RNA was extracted by using the RNAeasy kit (TianGen, Beijing, China) according to the manufacturer's instructions. The purity, concentration, and integrity of total RNA were checked using the NanoPhotometer spectrophotometer (IMPLEN, CA, USA), the Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, USA), and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 System (Agilent Technologies, CA, USA), respectively. Besides, RNA degradation and contamination were monitored on 1% agarose gels. A total amount of 1 μg total RNA per sample was used to prepare for the rRNA-depleted cDNA library. Ribosomal RNA was removed by Epicentre Ribo-zeroTM rRNA Removal Kit (Epicentre, USA), and rRNA free residue was cleaned up by ethanol precipitation. Subsequently, sequencing libraries were generated using the rRNA-depleted RNA by NEBNext UltraTM Directional RNA Library Prep Kit for Illumina (NEB, USA) and sequenced on an Illumina HiSeq 4000 platform to generate 150 bp paired-end reads. For a small RNA library, a total amount of 2 μg total RNA per sample was used as the input material. Sequencing libraries were generated using NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB, USA) following the manufacturer's recommendations. Small RNA libraries were sequenced on an Illumina HiSeq 2500 platform and 50 bp single-end reads were generated. All the data have

been deposited in the Gene Expression Omnibus database under the accession number GSE157859.

RNA sequencing (RNA-seq) data processing

For mRNA and lncRNA, raw RNA-seq reads generated from rRNA-depleted libraries were trimmed to remove sequencing adapters and low-quality reads. Quality of data were checked by FastQC v0.11.9. Adapters were removed by Trimmomatic v0.39.⁵⁴ The clean reads were then aligned to the primary assembly of the human reference genome, GRCh38, using STAR v2.7.3a.⁵⁵ StringTie v2.1.2 was used to assemble transcripts in a reference-guided manner for each sample.⁵⁶ The reference (GENCODE v33) and assembled transcripts for all samples were merged by StringTie.⁵⁶ Novel transcripts were obtained by comparing the merged transcripts with the reference transcripts by cuffcompare (code=='x' | Code=='u' | code=='i').⁵⁷ Novel lncRNAs were identified through the following filters: (1) transcript length >200 bp and ≥2 exons; (2) lack of coding potential predicted by both CPC2⁵⁸ and CPAT;⁵⁹ (3) no overlap with known lncRNAs from the RefLnc⁶⁰ or LNCipedia databases.⁶¹ Novel lncRNAs, known lncRNAs from RefLnc and LNCipedia,⁶¹ and reference transcripts (GENCODE v33) were merged and used for transcript quantification by kallisto v0.46.1.⁶² For small RNA-seq analysis, known and novel miRNAs were obtained and quantified by miRdeep2⁶³ using human and chimpanzee miRNA from miRBase v22.1⁶⁴ as the references.

Differentially express RNA and target genes for non-coding RNA DESeq2⁶⁵ was used to identify differentially expressed mRNAs (DEGs), lncRNAs (DElncRNAs) or miRNAs (DEmiRNAs) between patients at different clinical stages or with different clinical types (Supplementary Data 1, 2). *P*-value of differential expression was adjusted by the Benjamini–Hochberg's (BH) method (FDR). Because some samples with poor library quality were excluded, DEGs and DElncRNAs were calculated based on 39 samples, while DEmiRNAs was calculated based on all samples. The target genes of lncRNAs was predicted by their nearest and co-expressed (Pearson's correlation $P < 0.05$) protein-coding genes. Target genes of miRNA were obtained by merging the results from two publicly available databases miRDB⁶⁶ and miRWalk 2.0.⁶⁷ Genes predicted by both of the two databases and negatively co-expressed (Pearson's correlation $P < 0.05$ and $r < 0$) with the miRNA were defined as target genes of the miRNA. The functional annotation and analysis of DEmiRNA and DElncRNA in this study are for their target genes, therefore non-coding RNAs without target genes were excluded from functional analysis.

Gene set enrichment analysis (GSEA)

The values of log₂ fold change relative to the treatment stage of DEG and target genes for DEmiRNA and DElncRNA at the convalescence stage and the rehabilitation stage are subjected to linear regression analysis. Then through the GSEA algorithm of DOSE,⁶⁸ the slope values of log₂ fold change of these genes are used to calculate the significantly enriched biological process GO gene sets (FDR < 0.05) and the corresponding enrichment scores. These GSEA results were further classified and summarized by REVIGO,⁶⁹ and converted into two-dimensional lattice data based on GO semantic space. Finally, clustering algorithms are used to analyze the similarities and differences of the signal pathways affected by different types of transcriptomes (mRNA, miRNA, and lncRNA) or different clinical types of samples (Mild, moderate, and severe).

Functional analyses

Since there are multiple comparisons of samples at three clinical stages in our study, there can be as many as 8 differential expression patterns of the differentially expressed transcriptome during COVID-19 recovery. DEGs and target genes for DEmiRNAs or DElncRNAs can be further divided into interferon-stimulated

genes (ISG), transcription factor genes and phosphorylation regulatory genes according to their functions. These differentially expressed RNAs with different expression patterns and functions are respectively analyzed and visualized for GO enrichment by clusterProfiler.⁷⁰ In addition, clusterProfiler and pathview⁷¹ can further construct KEGG enrichment network and functional network of these genes, and visualize the relationship between genes and between genes and signal pathways. All differentially expressed RNAs are constructed to a protein–protein interaction network (PPI) by STRING V11.⁷² According to the network analysis, we select 65 highly differentially expressed (Log₂ fold change >1) genes with top connectivity and 19 related non-coding RNAs to reconstruct a core PPI network.

Weighted correlation network analysis (WGCNA) network construction

The mRNA, lncRNA, and miRNA signed WGCNA co-expression network was constructed on the basis of the top 8000 most variably expressed mRNAs/lncRNAs and top 600 most variably expressed miRNAs in all patients with COVID-19. The correlation matrix was obtained by calculating the Pearson's correlations between all gene pairs across all subjects in the dataset, and then were converted into an adjacency matrix using a power function (power $\beta = 14$). For module detection, the adjacency matrix was further transformed into a topological overlap matrix (TOM), and hierarchical clustering was used to group genes based on the dissimilarity matrix (1-TOM), followed by a dynamic cut-tree algorithm to dynamically cut clustering dendrogram branches into gene modules (Supplementary Fig. 8). A height cutoff of 0.1 was used to merge modules whose expression profiles are highly similar (Supplementary Data 3). Correlations between module eigengenes and traits of patients with COVID-19 were computed by Pearson's correlation.^{73,74} Network was visualized by using the Cytoscape 3.8 software.⁷⁵

DATA AVAILABILITY

All the RNA sequencing data have been deposited in the Gene Expression Omnibus database under the accession number GSE157859.

ACKNOWLEDGEMENTS

We express our gratitude to all donors, health care personnel, medical lab staff, laboratory managers, and lab technicians involved in this work. We are very grateful to Dr. Ian Logan for his help with language editing and helpful comments, and the four anonymous reviewers for their critical comments. This work was supported by the National Key Research and Development Program of China (2020YFC0842000 to Y.T.Z.), and the Yunnan Provincial major science and technology special project (202003AC100007 to X.Q.D.).

AUTHOR CONTRIBUTIONS

X.Q.D., Y.G.Y., and Y.T.Z. conceived the project; C.X.Y., R.R.T., Z.L.D., G.M.L., X.L.L., and Y.Q.H. designed and performed experiments; H.Y.Z. and M.X. analyzed data; M.Z., J.J.L., X.C.W., and X.Q.D. provided critical resources; H.Y.Z., M.X., and C.X.Y. drafted the manuscript; Y.G.Y. and Y.T.Z. edited the manuscript. All authors contributed intellectually and approved the manuscript.

ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41392-020-00457-4>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

REFERENCES

1. Wiersinga, W. J. et al. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA* **324**, 782–793 (2020).
2. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).

3. Wong, G. et al. Zoonotic origins of human coronavirus 2019 (HCoV-19 / SARS-CoV-2): why is this work important? *Zool. Res.* **41**, 213–219 (2020).
4. Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514–523 (2020).
5. Zhang, J. Y. et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat. Immunol.* **21**, 1107–1118 (2020).
6. Ong, E. Z. et al. A dynamic immune response shapes COVID-19 progression. *Cell Host Microbe* **27**, 879–882. e872 (2020).
7. Xiong, Y. et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* **9**, 761–770 (2020).
8. Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).
9. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
10. Zhou, R. et al. Acute SARS-CoV-2 infection impairs dendritic cell and T cell responses. *Immunity* **53**, 864–877.e5 (2020).
11. De Biasi, S. et al. Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia. *Nat. Commun.* **11**, 3434 (2020).
12. Le Bert, N. et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).
13. Ibarondo, F. J. et al. Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild covid-19. *N. Engl. J. Med.* **383**, 1085–1087 (2020).
14. Long, Q. X. et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med.* **26**, 845–848 (2020).
15. Zhou, Z. et al. Heightened innate immune responses in the respiratory tract of COVID-19 patients. *Cell Host Microbe* **27**, 883–890. e882 (2020).
16. Lee, J. S. et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci. Immunol.* **5**, eabd1554 (2020).
17. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
18. Liu, Y., Ma, C. & Zhang, N. Tissue-specific control of tissue-resident memory T cells. *Crit. Rev. Immunol.* **38**, 79–103 (2018).
19. Yang, X. O. et al. T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR alpha and ROR gamma. *Immunity* **28**, 29–39 (2008).
20. Yu, D. et al. The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31**, 457–468 (2009).
21. Guan, T. et al. ZEB1, ZEB2, and the miR-200 family form a counterregulatory network to regulate CD8(+) T cell fates. *J. Exp. Med.* **215**, 1153–1168 (2018).
22. Doedens, A. L. et al. Hypoxia-inducible factors enhance the effector responses of CD8(+) T cells to persistent antigen. *Nat. Immunol.* **14**, 1173–1182 (2013).
23. Yamazaki, S. et al. The AP-1 transcription factor JunB is required for Th17 cell differentiation. *Sci. Rep.* **7**, 17402 (2017).
24. Nussing, S. et al. Divergent SATB1 expression across human life span and tissue compartments. *Immunol. Cell Biol.* **97**, 498–511 (2019).
25. Willinger, T. et al. Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo. *J. Immunol.* **176**, 1439–1446 (2006).
26. Campbell, J. J. et al. CCR7 expression and memory T cell diversity in humans. *J. Immunol.* **166**, 877–884 (2001).
27. Helgeland, H. et al. Transcriptome profiling of human thymic CD4+ and CD8+ T cells compared to primary peripheral T cells. *BMC Genomics.* **21**, 350 (2020).
28. Lee, J. S. & Shin, E. C. The type I interferon response in COVID-19: implications for treatment. *Nat. Rev. Immunol.* **20**, 585–586 (2020).
29. Song, T. Z. et al. Delayed severe cytokine storm and immune cell infiltration in SARS-CoV-2-infected aged Chinese rhesus macaques. *Zool. Res.* **41**, 503–516 (2020).
30. Welsh, R. M., Bahl, K., Marshall, H. D. & Urban, S. L. Type 1 interferons and antiviral CD8 T-cell responses. *PLoS Pathog.* **8**, e1002352 (2012).
31. Paoletti, R., Bernardini, G., Molfetta, R. & Santoni, A. NK cells and interferons. *Cytokine Growth Factor Rev.* **26**, 113–120 (2015).
32. Peck-Radosavljevic, M. et al. Rapid suppression of hematopoiesis by standard or pegylated interferon-alpha. *Gastroenterology* **123**, 141–151 (2002).
33. Hervas-Stubbs, S. et al. Direct effects of type I interferons on cells of the immune system. *Clin. Cancer Res.* **17**, 2619–2627 (2011).
34. Platanias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386 (2005).
35. Du, Z. et al. Non-conventional signal transduction by type 1 interferons: the NF-kappaB pathway. *J. Cell Biochem.* **102**, 1087–1094 (2007).
36. Thevarajan, I. et al. Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nat. Med.* **26**, 453–455 (2020).
37. Long, Q. X. et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).
38. To, K. K. et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. *Clin. Infect. Dis. ciae1275* (2020).
39. Zhou, Y., Zhang, Z., Tian, J. & Xiong, S. Risk factors associated with disease progression in a cohort of patients infected with the 2019 novel coronavirus. *Ann. Palliat. Med.* **9**, 428–436 (2020).
40. Zheng, M. et al. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell. Mol. Immunol.* **17**, 533–535 (2020).
41. Guan, W. J. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
42. Sekine, T. et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* **183**, 158–168.e14 (2020).
43. Wen, W. et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.* **6**, 31 (2020).
44. Atsaves, V., Leventaki, V., Rassidakis, G. Z. & Claret, F. X. AP-1 Transcription factors as regulators of immune responses in cancer. *Cancers (Basel)*. **11**, 1037 (2019).
45. Macian, F., Lopez-Rodriguez, C. & Rao, A. Partners in transcription: NFAT and AP-1. *Oncogene* **20**, 2476–2489 (2001).
46. Li, W., Whaley, C. D., Mondino, A. & Mueller, D. L. Blocked signal transduction to the ERK and JNK protein kinases in anergic CD4+ T cells. *Science* **271**, 1272–1276 (1996).
47. Barski, A. et al. AP-1 transcription factor remodels chromatin during T cell activation. *J. Immunol.* **200**, 110.116–110.116 (2018).
48. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e1039 (2020).
49. Zheng, H. Y. et al. Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. *Cell. Mol. Immunol.* **17**, 541–543 (2020).
50. Chow, R. D. & Chen, S. The aging transcriptome and cellular landscape of the human lung in relation to SARS-CoV-2. *bioRxiv*, 2020.2004.2007.030684. Preprint at <https://doi.org/10.1101/2020.04.07.030684> (2020).
51. Nidadavolu, L. & Walston, J. Underlying vulnerabilities to the cytokine storm and adverse COVID-19 outcomes in the aging immune system. *J. Gerontol. A Biol. Sci. Med. Sci.* glaa209 (2020).
52. Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315–320 (2020).
53. Mauvais-Jarvis, F. Aging, male sex, obesity, and metabolic inflammation create the perfect storm for COVID-19. *Diabetes* **69**, 1857–1863 (2020).
54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
55. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
56. Perte, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
57. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
58. Kang, Y. J. et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
59. Wang, L. et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74 (2013).
60. Jiang, S. et al. An expanded landscape of human long noncoding RNA. *Nucleic Acids Res.* **47**, 7842–7856 (2019).
61. Volders, P. J. et al. LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* **47**, D135–D139 (2019).
62. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
63. Friedlander, M. R. et al. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).
64. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).
65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
66. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131 (2020).
67. Dweep, H. & Gretz, N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods* **12**, 697 (2015).
68. Yu, G., Wang, L. G., Yan, G. R. & He, Q. Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).
69. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

70. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
71. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
72. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
73. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17, (2005).
74. Langfelder, P., Horvath, S. & Fast, R. Functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46**, i11 (2012).
75. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021