Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Transcriptome Profiling of Peripheral Blood Mononuclear Cells Reveals COVID-19 Patients Are Not Recovered to Normal After Discharge for 5 Months

Misbah Abbas
University of Science and Technology of China
Deng Shasha
University of Science and Technology of China
Dan Zhao
USTC: University of Science and Technology of China
Kexing Han
USTC: University of Science and Technology of China
Zunera Khalid
USTC: University of Science and Technology of China
Hongliang He
USTC: University of Science and Technology of China
Huan Ma
USTC: University of Science and Technology of China
tengchuan jin (JINT@ustc.edu.cn)
University of Science and Technology of China https://orcid.org/0000-0002-1395-188X

Research

Keywords: Cytokine storm, Transcriptome, PBMCs, COVID-19, SARS-CoV-2, DEG

Posted Date: October 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-966635/v1

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

Abstract

Background

Highly pathogenic coronavirus disease-2019 (COVID-19) initiated by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection has swiftly expanded throughout the world, and the fatality rate is still expanding due to the second wave in 2020 winters. This ongoing epidemic threatens public health with its new strain that emerged in some countries and might cause devastating deaths. Therefore, the host transcriptomic profile from patients during recovery is important for understanding this disease.

Methods

We performed transcriptome profiling of the RNAs isolated from the peripheral blood mononuclear cells (PBMCs) of recovered COVID-19 patients at hospital discharge of three months and five months respectively.

Results

Our results exposed diverse inflammatory genes and cytokine profiles to infection in recovered patients, and emphasize the highly expressed genes in COVID-19 patients like CCL4, CCL3, CXCL9, CXCL16, IL10, CSF2, VEGFA showed a decreasing trend in recovered patients. Furthermore, the integrated analysis predicted that JUN, CTSL, DDIT4, RRAS, BIRC5, CTSZ, CCNB2, CDK1, OAS1/2, IFIT3, RSAD2, and TP53I3 genes may be valuable for the recovery of COVID-19 patients.

Conclusions

Our analysis confirms the presence of some inflammatory genes in recovered patients, suggesting COVID-19 patients did not return to their normal expression even after 5-months of discharge. Identification of transcriptome profiling of recovered patients provides useful information regarding its pathogenesis and might help for the development of better treatment for COVID-19.

Background

SARS-CoV-2 is an enveloped, easily transmittable fatal single-stranded RNA virus consisted of 29,903 nucleotides, which caused an outbreak along worldwide since 2019. It belongs to the order Nidovirales, the family Coronaviridae, and subfamily Coronavirinae of viruses that become a major reason for minor to destructive respiratory tract disorders in humans [1-3]. The clinical manifestations of the COVID-19 caused by this virus depend on the host immune response and viral pathogenesis lead from asymptomatic to minor infections (fever, sneezing, anosmia, etc.), to severe disorders (ARDS, organ failure, etc.), and leading to death [4-8]. Patients with older age and other comorbidities (diabetes, COPD, hypertension, etc.) and patients in immuno-compromised states were faced complications and at utmost risk with severe COVID-19 [9-11]. While worldwide scientists and clinicians have made countless efforts to fight against the disease, several vaccines, anti-viral and immunomodulatory drugs are in different

phases of clinical trials [12-14]. In November, the pharma giant Pfizer and BioNTech (got temporary authorization), Moderna, and a Russian institute have all suggested initial evidence that spike-based vaccines can attain >90% efficacy [15, 16].

Recently, several new variants of SARS-CoV-2 have been reported by the UK, South Africa, and Nigeria, including a new variant with N501Y mutation on S protein, VUI-202012/01 from the UK, which has about 71% higher transmissibility than the existing strains [17]. The development of the COVID-19 is quite similar to the 1918 “Spain Flu”, especially for the second wave with new variants and increasing spread rate globally. In the second wave of Spain flu, the mutated variants become more deadly than the original strain and mortality were very high [18, 19]. Predictably more and more new variants will emerge with the rapidly expanding epidemic. Thus, a deep understanding of anti-viral innate and adaptive immune responses is a prerequisite for accurate diagnosis and adopting a therapeutic strategy. Dysregulation of the immune system is thought to be correlated with the acuteness of COVID-19 infections, and enhanced expression of proinflammatory chemokines and cytokines is exhibited in mild and severe patients [4, 20, 21]. Different studies have revealed that PBMCs are indulgent to SARS-CoV disease and help in the replication of the virus, and are used for transcriptomics evaluations, including changes in the expression of genes in response to infection or treatment [22, 23].

Though, how significant immune cells population change and basic molecular procedures of the inflammatory responses and pathological processes of the COVID-19 have remained unclear. However, to investigate the biological process and host transcriptional fluctuations in response to disease and after the infection is necessary, especially with the emergence of the novel variants from different regions and countries.

Here, we performed transcriptome profiling of PMBCs from recovered COVID-19 individuals, to map the major expression profile of pro-inflammatory genes and cytokines induced by infection. We observed lower expressions of CCL4, CCL3, CXCL9, CXCL16, IL10, CSF2, and VEGFA. CXCL3 was down-regulated in all recovered patients except X13-X14. Our study represented a high-resolution transcriptome landscape after the COVID-19 infection in different periods. It helps the researchers to get a better understanding of the host immune response, abnormal expression levels, and cytokine storm fluctuations throughout the recovery stage, which might be helpful for the production of vaccines and antibodies for therapeutics.

Methodology

Sample Isolation, Sequencing, and Filtering

The purple EDTA anticoagulant tubes were used to collect the whole blood from patients recovering from SARS-COV-2 in the First Affiliated Hospital of USTC, Anhui, China during their checkups after discharge. Blood samples were collected for each patient three months and five months after recovery. Ficoll-Hypaque density gradient centrifugation was applied to separate lymphocytes in peripheral blood. In brief, under the action of centrifugation, Ficoll-Hypaque solution with a density of 1.077 can separate PBMCs from Red blood cells, granulocytes, plasma, and other components. The collected PBMCs were

washed twice with DMEM medium and counted under a microscope. One million cells from each sample were sent to BGI (Shenzhen, China) for transcriptome sequencing. A large number of sequencing data was generated in a single run and quality control was then performed to determine whether the sequencing data is suitable for subsequent analysis. The original data were purified by SOAPnuke-v1.5.2 (*https://github.com/BGI-flexlab/SOAPnuke*) with parameters -q 0.2 -l 15 -n 0.05. It was performed to remove adaptor pollution, low-quality reads and unknown "N" bases greater than 5%. In the end, clean reads were saved in FASTQ format [24].

Sequencing Data Analysis

The filtered clean-reads were mapped to the reference sequence *Homo sapiens* (GRCh38.p12) from NCBI, by using HISTAT2-v2.0.4 (*http://www.ccb.jhu.edu/software/hisat*) software with parameter: –sensitive –phred64 –no-mixed –no-discordant -I 1 -X 1000 [25]. The aligned reads were submitted to Bowtie2-v2.2.5 (*http://bowtie-bio.sourceforge.net/bowtie2/index.shtml*) to map the clean reads to the reference gene sequence with parameters: -q –phred64 –sensitive –dpad 0 –mp 1, –gbar 99999999,1 –np 1 –score-min L,0, -0.1 -k 200 -p 16. Meanwhile, the overall analysis of the transcriptome sequencing was performed, including the randomness, coverage, and distribution of the transcripts. Moreover, Perl script was used for sequencing saturation analysis to evaluate whether the amount of sequencing data meets the requirements to a certain extent.

Expression Quantification

Then output was fed to RSEM-v1.2.8 (*http://deweylab.biostat.wisc.edu/rsem/rsem-calculate-expression.html*) to calculate the gene expression level of each sample with default parameters. After that, it is converted by FPKM or TPM to obtain the standardized gene or transcript expression level. After obtaining the Read Counts, gene or transcript expression differential analysis was performed. Here, gene quantification analysis and other analyses based on gene expression like principal component analysis (PCA), sample correlation heat map, expression quantification distribution (boxplot, density map, stacked histogram), and differential gene screening were conducted.

Differential Gene Detection

PossionDis method was used for the detection of differential genes with parameters: Fold Change >= 2 and FDR <= 0.001**.** The sequencing-based differential gene detection method used a rigorous process to screen DEGs between samples [26]. R-software pheat-map (*https://cran.r-project.org/web/packages/pheatmap/*) was utilized to perform hierarchical-clustering analysis on the union set differential genes with default parameters. Additionally, we showed the differential expression of genes between different comparison groups and the intersection/union via VENN diagrams.

Functional Enrichment and Variation Analysis

Functional enrichment analysis was executed on all the transcripts and DEGs found from this RNA-Seq, which were compared with KEGG

(*http://www.genome.jp/kegg/)* and GO (*http://www.geneontology.org/*) databases. According to the annotated classification of KEGG Pathway and GO analysis, the phyper-function in R was utilized to execute the enrichment study, compute the P-value, and then carry-out FDR-correction. Generally, the function with Qvalue <= 0.05 was considered as significant enrichment. The rMATS-v3.2.5 (*http://rnaseq-mats.sourceforge.net*) statistical model was used to quantify the expression of alternative splicing events with parameters: -analysis U -t paired -a 8. Additionally, gene fusion events were detected by aligning the sequences at the paired-end relationship in the genome and transcript via Ericscript-v0.5.5 (*http://ericscript.sourceforge.net/*) with default parameters.

Cytokine Storm Detection

Cytokines were screened using Perl and Python scripts and performed GO enrichment, KEGG pathway, and disease enrichment analysis, in which Qvalue <= 0.05 was viewed as a significant enrichment.

Results

To explore the influence and mechanism of SARS-CoV-2 disease on the host immune system after the recovery, we utilized RNA-sequence to find transcriptome modifications in PMBC samples from recovered individuals in two different periods. A total of 10 blood samples {5-Control {3-MONTHS: X11, X13, X5, X7, X9} and 5-Treated {5-MONTHS: X12, X14, X6, X8, X10}} and 5 groups (2-females, 3-males) of 3- and 5-months patients' data {group-1 (X11-vs-X12) group-2 (X13-vs-X14), group-3 (X5-vs-X6), group-4 (X7-vs-X8), group-5 (X9-vs-X10)} were verified using the DNBSEQ platform, with an average output of 6.52G data per sample. The clinical information of recovered individuals was showing in Table 1.

The sequencing dataset passed stringent high-quality filtering and plotted the statistics of reads mapped to the genome and genes. The mapping rate for the mapped reads of each sample with the selected reference genome and genes total ranges from 87.8 % to 88.62 % and 52.05 % to 60.79 %, respectively. The average alignment ratio of the observed sample comparison genome was 88.13%, and the gene set was 57.03%. Whereas the proportion of reads that map to the unique position of the reference genome and genes ranges from 78.2% to 83.11 % and 46.86% to 56.1%, respectively (Table. S1). We conducted a Randomness analysis, most of the transcripts were entirely covered, and reads were evenly scattered in various regions of the transcript. Additionally, the coverage of the transcripts of each sample suggested that most of the transcripts were fully covered, and reads were regularly distributed in each area of the transcript. Finally, the sequencing saturation analysis for each sample suggested that the amount of sequencing data meets the desired requirements (Fig. S1A/B/C/D/E/F).

Transcriptome Profiling of PBMCs

To obtain the gene signature, we found a total of 17686 expressed genes and 87324 expressed transcripts in the PBMCs samples of all recovered individuals. Gene expression overview in the PBMCs comparisons group was examined in the PCA (Fig. 1A). To signify the correlation of gene expression between samples, the Pearson correlation coefficients between every two samples were calculated and

reflected in the form of a heatmap (Fig. 1B). To get an overview of gene expression quantification in each sample, we formed a boxplot from which the dispersed degree of the data distribution can be observed (Fig. 1C). The density map was constructed to show the trend of gene abundance as the expression level changes (Fig. 1D). Also, the statistical analysis was used to visualize the number of genes in the diverse FPKM range of each sample (Fig. 1E).

Gene and Transcripts Analysis in Comparison Groups

To get a comparison overview of all genes and transcripts in control and treated groups, we performed the intersection/union analysis of each sample and group. In the first group of a total of 16309 genes, in group-2, 16387, group-3, 16399, group-4, 16320, and group-5, consisted of 16272 total expressed genes. Overall, a little more gene was present in control individuals (17,186) compared to treated groups (17,134), and in case of transcripts 77,980 were present in control and 78,434 in other group (Fig. [2A (control), B (treated)-genes], [C (control), D (treated)-transcripts]). While in each sample, total intersects and union genes (Fig. S2A/B/C/D/E) and transcripts (Fig. S3/A/B/C/D/E) were shown in supplementary figures.

Differential Expression Pattern in Comparison Groups

To characterize the genomic changes in SARS-CoV-2 recovered individuals, we evaluated transcriptomic data from PBMCs and found a total of 5317 DEGs, of which 3170 genes were up-regulated and 2147 genes were down-regulated (Fig. 3A/B). We also identified 85,207 differentially expressed transcripts (41,865 down-regulated and 43,342 up-regulated) (Fig. S3). The up-regulated and down-regulated genes of each comparison group were listed up in Table S2. An overview of statistical significance among the gene expression in each sample group was observed, which assisted rapid visual recognition of genes with large fold changes (Fig. 4A). Cluster the expression of the DEGs gave a general overview of gene expression variation among each group (Fig. 4B). The number of DEGs was illustrated graphically which exhibited overlapped genes/transcripts in the comparison group (Fig. 3C/D).

Characterization of Longitudinal Analysis

We did a longitudinal analysis (3rd to 5th months) of individuals one by one, which clarified the number of expressed genes in each sample and those genes that were specific to each patient. Overall, in control-group (3rd-month) 5254 genes (3107-up, 2147-down), and treated-group (5th-month) 5273 (3170-up, 2103-down) genes were differentially expressed. While in the comparison of X11 to X12, we identified a total of 885 and 889 DEGs, respectively, from which 9 genes were specifically down-regulated in X11, while 13 genes were specifically up-regulated in X12. In X13, DEGs were 15626 from which 10 genes were particularly down-regulated, whereas X14 contained 15855 genes and only 12 were uniquely showed higher expression compared to X13. From the third group, X5 had 794 and X6 had 755 DEGs, where 5 down-regulated genes were specific to X5 and 11 up-regulated genes, were specific to X6. The 3rd month's sample of X7 differentially expressed 1081 and during 5th months (X8) 1085 different expressions of genes, from which X7 specific down-regulated genes were 10 uniquely up-regulated genes. DEGs in X9

were 1408 and X10 was 1411, where 10 genes showed down and 13 genes showed a high expression that specific to X9 and X10, respectively (Table. S3).

Functional Enrichment and Gene Annotation

We then performed the functional enrichment and classification analysis to identify the changes in five different groups' DEGs. GO annotation of DEGs was classified and mapped. Among the up-regulated genes of group-1, 536 genes were involved in biological process, 569 in a cellular component, and 537 in molecular function. While from the up-regulated genes of the other groups (2,3,4,5), 569, 525, 624, 681 (biological process), 583,548, 656, 699 (cellular component) and 563,530,618, 677 (molecular function) genes were involved respectively. Based on GO enrichment analysis, up-regulated genes were enriched in a cellular component like "mitochondrial inner membrane, main axon, phagocytic vesicle, plasma membrane, lysosome, cytoskeleton, etc.". While in-case of molecular function genes were significantly enriched in "ATPase activity, peroxisome proliferator-activated receptor binding, cytokine/chemokine activity, heme and IgG and nucleotide-binding, etc." According to the Go_P annotation classification, they were significantly enriched in the biological process including "blood coagulation, oxygen transport, neutrophil chemotaxis, cellular response to DNA damage stimulus, positive regulation of interferon, immune system process, neutrophil degranulation and so on". In group-2, GO annotation analysis revealed that up-regulated genes were annotated and involved in biological processes, cellular components, and molecular functions, but not significantly enriched (Fig. S5/A/B/C/D).

Whereas, down-regulated genes in each group were significantly annotated in GO terms such as biological processes (ranges from 170-667 genes), cellular component (187-682 genes), and molecular functions (173-664 genes). Additionally, down-regulated genes were also significantly enriched in cellular components (membrane, cell, transcription factor complex, cytoplasmic vesicle lumen, etc.), biological processes (inflammatory response, neutrophil chemotaxis, immune response, apoptotic process, lymphocyte/monocyte, so on), and in molecular functions (cytokine and chemokine activities, tumor necrosis factor receptor binding, etc.) (Fig. S6/A/B/C/D).

Pathway and Disease Enrichment Analysis

KEGG Pathway-based evaluation was performed to further figure out the biological function of DEGs in each group. KEGG Pathway classification involved in genes was divided into several branches. A classified statistic was further made for the metabolic pathways under each branch. Interestingly, the KEGG enrichment study discovered that up-regulated genes were enriched in many pathways (Leishmaniasis, Williams-Beuren/Stickler syndrome, and PPAR signaling, etc.) and also showed enrichment for several diseases (NK cell defects, Neutropenic disorders, Agammaglobulinemias, TNDM, and polydactyly disorders, etc.) Functional profiling of the PBMCs up-regulated DEGs were recognized important terms like an inflammatory response, cytokine activity, chemokine activity, and type-1-interferon signaling pathway signifying a cellular-mediated response to the immune system (Fig. S5A/B/C/D). While down-regulated genes were also enriched in many pathways (TNF/IL-17 and Chemokine signaling, and involved in malaria and cancer pathways, etc.) and significantly enriched in

many diseases like Hypertrophic, cardiomyopathy, etc. While group-2 down-regulated genes were not significantly enriched in any pathway (Fig. S6A/B/C/D).

Variation Analysis

We also performed statistics of alternative splicing events for each sample and each comparison group (Fig. S7). Gene fusion events were detected by aligning the sequences at the paired-end relationship in the genome and transcript and also displayed their location on chromosomes (Fig. S8).

Regulated Cytokines Analysis

To evaluate the immunological response in patients after recovery, we also fetched genes involved in cytokine storm and identified a total of 108 cytokines (59-up and 49-down). While in the case of comparison groups (1, 2, 3, 4, 5), a total of 14, 11, 38, 7, 7 up-regulated and 17, 12, 2, 29,27 down-regulated cytokines were observed, respectively (Table. S4). Moreover, JUN, FOS, IRF1, DUSP1, MKI67, IFI6, TNFSF13, ISG15, CSF1, and S100A8 genes were up-regulated in some individuals. CCL4, CCL3, CXCL9, CXCL16, IL10, CSF2, VEGFA, IL10, CXCL3, FOSB, FOSL2, IFNG, and many other genes showed less expression. We also observed some unusual behaviors of most cytokines in all groups, because they show different expression levels in different groups such as IL2R, IL1B, and IL6 were up-regulated in one and were down-regulated in another. KEGG pathway and GO enrichment, and disease enrichment analysis for the regulated cytokines exposed certain important pathways that were involved in severe immune degenerative disorders (Fig. 5A/B/C/D/E).

Table 1: Clinical Information of Recovered Individuals

Control group (3-months)	Treat group (5-months)	Comparison group	Gender	Age	First symptoms	Admission time	Discharge time
X11	X12	X11-vs-X12	male	58	1/16/2020	1/26/2020	2/18/2020
X13	X14	X13-vs-X14	male	37	1/30/2020	2/3/2020	2/27/2020
X5	X6	X5-vs-X6	female	38	1/23/2020	1/27/2020	2/15/2020
X7	X8	X7-vs-X8	male	56	1/18/2020	1/26/2020	2/14/2020
X9	X10	X9-vs-X10	female	8	1/31/2020	2/1/2020	2/18/2020

Discussion

This project was proposed to investigate host transcriptional and immune response variations in recovered COVID-19 patients, to participate in the information of the biological processes and functional enrichment triggering the COVID-19 pathogenesis and the immune responses prompted. Severe

respiratory disorders and other difficulties with more significant morbidities and mortalities were triggered by SARS-CoV-2 infection, and there was no most favorable treatment for it because the knowledge about host immune response to this infection was limited. Massive variations in the host transcriptome generally caused by a viral infection, lead to abnormal metabolism and modified host immune response [27, 28]. However, the RNA virus tends to evolve with different variants after generations [29], so, SARS-CoV-2 can generate various mutations, which will increase the transmissibility and might increase mortality rate or affect specific people like youngers. Therefore, host transcriptomic profile from different patients is important for accurate diagnosis and therapeutics. Our results exhibited a diverse immune response of recovered COVID-19 patients, as shown by the differential expression pattern of genes and cytokines. We identified gene signatures from each recovered individual and each comparison group and found 5317 genes were differentially expressed (3170-up and 2147-down). In the case of each comparison group, we found 898, 1143, 3760, 41095, and 1421 DEGs in group-1,2,3,4, and 5 respectively. The overaggressive response of the immune system leading to immunopathology was observed in previous studies [30]. We also did a longitudinal analysis of each 3 to 5 months individual, which clarified the number of DEGs and specifically expressed genes in each sample. DEGs in X11-X12 [(582-up, 303-down) -(595-up, 294-down)], X13-X14 [(595-up, 536-down) -(607-up, 526-down)], X5-X6 [(555-up, 194-down) -(566-up, 189-down)], X7-X8 [(664-up, 417-down) -(678-up, 407-down)], X9-X10 [(711-up, 697-down) -(724-up, 687-down)], were identified respectively. Longitudinal findings of recovered COVID-19 patients might assist to recognize the after-effects of the infection. During 3-months after recovery 44 genes were specifically downregulated, and in the 5th-month 63 genes that were showed specifically high expression in recovered patients.

We also observed inflammatory cytokines and chemokines, like DUSP1/6, IRF1, JUN, IFITM3, and FOS and interferon-stimulated genes (IFRD1, IRF6, and IFI1), were expressed at high levels in recovered patients, which were consistent with the previous study, indicating that these genes were associated with CD14++ inflammatory monocytes and CD4+ T cells. The high expression of some genes IL1B, IL6, CSF1, and CSF2 may be involved with cytokine storm in COVID-19 patients [31]. IL6 was highly expressed in acute COVID-19 patients compared to healthy people, and no significant difference was observed between the acute and recovering patients [32]. In our analysis, IL6 expression was down in recovered patients except for X13-X14, and no expression was observed in X5-X6. In contrast, we observed low expression of JUNB, CCL3, CCL4, CSF2, and no differential expression for CXCR4, and KLF6, which showed that reduced expression of inflammatory genes after the recovery. The dysregulation in the monocyte population balance in patients showed that classical-monocytes CD14++ enhanced circulation to stimulate inflammation in SARS-CoV-2 infection [31]. In this study of recovered patients, we demonstrated that IFITM3 expression was elevated in the X5-X6 group, compared to X13-X14, and X9-X10. IFITM3 was involved in anti-viral response and provides immunity against infection by interrupting the intracellular cholesterol homeostasis and hinders the virus entry into the cytoplasm [33, 34]. Former study at the whole transcriptome level showed some sets of genes that were upregulated during COVID-19 recovery also showed high expression in some patients of our study even after 5-months of recovery like OSM, IL1B, JUN, NR4A3, AREG, DUSP8, and CD69, whereas NR4A3, ZEB1, IL1B, OSM, MAP3K8,

SOSC3, DUSP8 genes presented decreasing trend in some patients after 3 and 5 months of recovery. It disclosed that up-regulation of JUN, AREG, and CD69, and down-regulation of ZEB, MAP3K8, and SOSC3 during the recovery stage exhibited the same trend even after 5-months of recovery. Additionally, according to the previous investigation, some antiviral genes like OAS1, IFIT3, RSAD2 were downregulated at the rehabilitation stage, but current analysis showed these genes were up-regulated after recovery, except group-2 and group-4 where IFIT3 exhibited decreasing trend even after 5-months of recovery [35].

Single-cell RNA-seq exhibited up-regulation of the TNF/IL1B-driven inflammatory response in COVID-19 [36]. Furthermore, previously stated that in COVID-19 highly expressed IL1B gene may be a novel candidate target, however, its expression was decreased in observed recovered patients, except in the X5-X6 group where it had a high expression. There was no differential expression of IL1B in the X13-X14 group, might be its expression uniquely depends on other factors like age, because the X13-X14 individual was 8 years old and X5-X6 was 58 years old. IL1B and CXCL8 were increased during aging and upregulate in COVID-19 patients [37] still showed high expression in aged people after recovery of this infection. Many genes behave differently in both of these groups as compared to others. We observed high expression of CCL2, CXCL10, S100A8, and other B cell activation-related genes TNFSF13, and TNFSF13B in one group (X5-X6), but lessen the expression of TNFSF12-TNFSF13 and TNFSF13B in another group (X13-X14). Based on the previous study, IL18, TNFSF13, TNFSF13B, IL2, and IL4 to elevate the cells proliferation and then produced antibodies into the blood may be helpful for the recovery [31] but we did not observe differential expression of IL2, IL4, and IL18 in our recovered patients. While CCL13, CXCL8 (related with chemoattraction of macrophages or neutrophils (innate-immunity)) [32], CXCL2, and CXCL10, were elevated in COVID-19 patients compared to recovering stage, exhibited altered expression in recovered patients like to up-regulate in X13-X14 and downregulate in others, while no expression in X5-X6. Might be CCL2, CCL13, CXCL2, and CXCL10 were also related to aging. While CCL13 was normal in recovered patients. Meanwhile, a previous study showed that OAS2 and IL16 linked with T cells (adaptive-immunity) were highly expressed in COVID-19 patients with recovery stage and healthy people. In this current analysis, we also observed high expression of OAS2 in some groups of recovered patients even after 5 months of recovery. Some analysis described that the enhanced expression of CXCL10, TNFSF14, S100A8, IL6, and OSM genes in COVID-19 infection is strongly associated with clinical severity [38, 39]. We observed down the expression of TNFSF14 in group-4 and group-5 of recovered patients. Many studies had stated notably higher levels of inflammatory cytokines which were linked with an acuteness of disease in MERS, SARS, and SARS-CoV-2 patients suggested that inflammation was an important component of the immune response during COVID-19 infection [40, 41].

Different inflammatory cells and monocyte populations might play a key role as they were recognized to fuel inflammation [42-44]. We also observed less expression of IL10, IL1A, IL1RN, IL4I1, IL21R, IL31RA, ILDR2, and IL2RA in recovered COVID-19 individuals. Several studies demonstrated that upregulation of IL10 and IL2RA was involved in anti-inflammatory signaling during the infection [45, 46]. The increased level of CXCL10 was correlated with disease severity [47, 48], thus most of the recovered patients showed decreased expression of CXCL10, CCL2, CCL3, and CCL4. This altered expression proposed that these

cytokines might be a helpful candidate target for COVID-19 screening and therapeutics. A previous study showed a correlation between COVID-19 pathogenesis and excessive release of these cytokines [23].

We found some decreasing trends of genes such as CTSL, DDIT4, RRAS, BIRC5, CTSZ, CCNB2, CDK1, and TP53I3 in some recovered patients, which were enriched to P53 signaling pathways and apoptosis according to the previous study, while TNFSF10 showed an increasing trend in recovered patients. The former analysis presented those lymphocytes had been significantly reduced in COVID-19 patients. While, TP53, was a significant gene in the procedure of apoptosis, showed no altered expression in recovered patients [23, 49]. Moreover, we identified decreased expression of VEGFA genes (growth factor family) in few patients. FCGBP encodes IgG-binding protein's Fc fragment, which has been stated with the virus disease and viral vector design. It was highly expressed in some recovered patients (X9, X10, X13, X14) in this study, while, FCGBP was also showed high expression in infected COVID-19 patients [50]. Overall, significantly expressed genes that were enriched in COVID-19 patients (CCL4, CCL3, CXCL9, CXCL16, IL10, CSF2, VEGFA) showed a decreasing trend in recovered patients and these genes might be the hallmark of COVID-19. The presence of inflammatory gene signatures even after 5-months of discharge, suggesting these COVID-19 patients did not return to their normal life yet, was consistent with a previous study of single-cell sequencing of COVID-19 patients in the recovery phase[31]. Furthermore, we performed gene annotation, functional enrichment, and classification analysis in PBMCs to identify the involved disease pathways and biological function of DEGs. The down-regulated cytokines were mostly involved in inflammatory bowel disease, T B+severe combined immunodeficiency syndrome, allergic rhinitis and type 1 diabetes mellitus, and so on. Additionally, we also carried out alternative splicing and gene fusion analysis, which might have a pivotal role in SARS-COV-2 illness precision medicine.

Conclusions

In conclusion, by studying the differential expression of genes after SARS-CoV-2 infection, we provided deep perceptions on the possible usage of transcriptomics data and pointed out the prominent inflammatory genes and cytokines in recovered patients, which might explain immune response alterations after infection, and why some patients feel unwell after being discharged. Moreover, it is beneficial to figure out the disease consequences. Additionally, the cytokine expression profile suggested altered expression of pro-inflammatory cytokine might be a hallmark of recovered patients. Furthermore, a deep investigation is required to verify either these genes could be used as a potential biomarker or have potential medical efficacy in immunotherapies of COVID-19.

Limitations

In this article the limited number of patient samples assessed and uninfected controls are also not included not only due to time limits, but also limited patients.

Abbreviations

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

COVID-19: Coronavirus disease-2019

PBMCs: Peripheral blood mononuclear cell

DEG: Differentially expressed genes

ARDS: Acute respiratory distress syndrome

PCA: Principal component analysis

KEGG: Kyoto encyclopedia of genes and genomes

NK cell: Natural killer cells

Declarations

Ethical Approval and Consent to Participate

This study was reviewed and approved by the Medical Ethical Committee of First Affiliated Hospital of USTC (approval number 2020-XG(H)-019).

Consent For Publication

All patients provided informed consent.

Availability of Supporting Data

The authors declare that the data supporting the findings of this study are available within the article, and its supplementary information files.

Competing Interests

Authors declare no conflict of interest.

Funding

Strategic Priority Research Program of the Chinese Academy of Sciences, XDB29030104, Tengchuan Jin. National Natural Science Foundation of China, 31870731, Tengchuan Jin. Fundamental Research Funds for the Central Universities, Tengchuan Jin. COVID-19 special task grants supported by Chinese Academy of Science Clinical Research Hospital (Hefei), YD2070002017, Tengchuan Jin. Chinese Government Scholarship, Misbah Abbas, PHD.

Authors' Contributions

MA and SD conceived the presented idea, prepared the samples, analyzed the data, wrote the original draft, and formatted the manuscript for submission. DZ, ZK, HH and HM participated the research, reviewed and edited the original version of the manuscript. TJ conceptualized the main idea, provided funding during the whole study, and supervised the whole paper. All the authors read and approved the final version of the manuscript for publication.

Acknowledgments

This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29030104), National Natural Science Fund (31870731 and 31971129), the Fundamental Research Funds for the Central Universities, the new medical science fund of USTC, COVID-19 special task grants supported by Chinese Academy of Science Clinical Research Hospital (Hefei) with Grant No. YD2070002017, and China Postdoctoral Science Foundation (No 2020M670084ZX), respectively. MA is supported by Chinese Government Scholarship.

References

1. Grifoni, A., et al., A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. Cell Host & Microbe, 2020. **27**(4): p. 671-680.e2.

2. Wu, F., et al., A new coronavirus associated with human respiratory disease in China. Nature, 2020. **579**(7798): p. 265-269.

3. Chen, Y., Q. Liu, and D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis. Journal of medical virology, 2020. **92**(4): p. 418-423.

4. Huang, C., et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet, 2020. **395**(10223): p. 497-506.

5. Young, B.E., et al., Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. Jama, 2020. **323**(15): p. 1488-1494.

6. Lima, C.M.A.d.O., Informações sobre o novo coronavírus (COVID-19). Radiologia Brasileira, 2020. **53**(2): p. V-VI.

7. Lee, Y., et al., Prevalence and duration of acute loss of smell or taste in COVID-19 patients. Journal of Korean medical science, 2020. **35**(18).

8. Hornuss, D., et al., Anosmia in COVID-19 patients. Clinical Microbiology and Infection, 2020.

9. Guan, W.-j., et al., Comorbidity and its impact on 1590 patients with Covid-19 in China: A Nationwide Analysis. 2020. **55**(5).

10. Wang, B., et al., Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging (Albany NY), 2020. **12**(7): p. 6049.

11. Zhou, F., et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The lancet, 2020.

12. Corey, L., et al., A strategic approach to COVID-19 vaccine R&D. Science, 2020. **368**(6494): p. 948-950.

13. Yao, X., et al., In vitro antiviral activity and projection of optimized dosing design of hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Clinical infectious diseases, 2020. **71**(15): p. 732-739.

14. Cyranoski, D., This scientist hopes to test coronavirus drugs on animals in locked-down Wuhan. Nature, 2020.

15. Cohen, J., Vaccine wagers on coronavirus surface protein pay off. 2020, American Association for the Advancement of Science.

16. Mahase, E.J.B.B.M.J., Vaccinating the UK: how the covid vaccine was approved, and other questions answered. 2020. **371**.

17. England, P.H., New and emerging respiratory virus threats advisory group (NERVTAG). 2020.

18. Taubenberger, J.K. and D.M. Morens, 1918 Influenza: the mother of all pandemics. Revista Biomedica, 2006. **17**(1): p. 69-79.

19. Trilla, A., G. Trilla, and C. Daer, The 1918 “Spanish Flu” in Spain. Clinical Infectious Diseases, 2008. **47**(5): p. 668-673.

20. Wang, L., A. Zou, and J. Wang, Matrix analysis of clinical characteristics and dynamic observation of immunological features in 90 cases of COVID-19. SSRN 2020, published online April 02.

21. Guo, C., et al., Tocilizumab treatment in severe COVID-19 patients attenuates the inflammatory storm incited by monocyte centric immune interactions revealed by single-cell analysis. BioRxiv, 2020.

22. Voineagu, I., et al., Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature, 2011. **474**(7351): p. 380-384.

23. Xiong, Y., et al., Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerging microbes & infections, 2020. **9**(1): p. 761-770.

24. Cock, P.J.A., et al., The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research, 2010. **38**(6): p. 1767-1771.

25. Kim, D., B. Langmead, and S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. Nature Methods, 2015. **12**(4): p. 357-360.

26. Audic, S. and J.-M. Claverie, The significance of digital gene expression profiles. Genome research, 1997. **7**(10): p. 986-995.

27. Thaker, S.K., J. Ch'ng, and H.R. Christofk, Viral hijacking of cellular metabolism. BMC biology, 2019. **17**(1): p. 1-15.

28. Channappanavar, R., et al., Dysregulated type I interferon and inflammatory monocyte-macrophage responses cause lethal pneumonia in SARS-CoV-infected mice. Cell host & microbe, 2016. **19**(2): p. 181-193.

29. Sanjuán, R. and P. Domingo-Calap, Mechanisms of viral mutation. Cellular and Molecular Life Sciences, 2016. **73**(23): p. 4433-4448.

30. Zhang, B., et al., Clinical characteristics of 82 death cases with COVID-19. MedRxiv, 2020.

31. Wen, W., et al., Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. Cell discovery, 2020. **6**(1): p. 1-18.

32. Sadanandam, A., et al., A blood transcriptome-based analysis of disease progression, immune regulation, and symptoms in coronavirus-infected patients. Cell Death Discovery, 2020. **6**(1): p. 141.

33. Zhang, Y., et al., Interferon-induced transmembrane protein 3 genetic variant rs12252-C associated with disease severity in coronavirus disease 2019. The Journal of infectious diseases, 2020. **222**(1): p. 34-37.

34. Feeley, E.M., et al., IFITM3 inhibits influenza A virus infection by preventing cytosolic entry. PLoS Pathog, 2011. **7**(10): p. e1002337.

35. Zheng, H.-Y., et al., Longitudinal transcriptome analyses show robust T cell immunity during recovery from COVID-19. Signal Transduction and Targeted Therapy, 2020. **5**(1): p. 294.

36. Lee, J.S., et al., Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. Science Immunology, 2020. **5**(49): p. eabd1554.

37. Zheng, Y., et al., A human circulating immune cell landscape in aging and COVID-19. Protein & Cell, 2020. **11**(10): p. 740-770.

38. Arunachalam, P.S., et al., Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science, 2020. **369**(6508): p. 1210.

39. Ren, X., et al., COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. Cell, 2021.

40. Liu, J., et al., Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. Journal of medical virology, 2020. **92**(5): p. 491-494.

41. Channappanavar, R. and S. Perlman. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. in Seminars in immunopathology. 2017. Springer.

42. Gu, J., et al., Multiple organ infection and the pathogenesis of SARS. Journal of Experimental Medicine, 2005. **202**(3): p. 415-424.

43. Nicholls, J.M., et al., Lung pathology of fatal severe acute respiratory syndrome. The Lancet, 2003. **361**(9371): p. 1773-1778.

44. Villani, A.C., et al., Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science, 2017. **356**(6335).

45. Blanco-Melo, D., et al., Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell, 2020. **181**(5): p. 1036-1045.e9.

46. de Melo, C.V.B., et al., Transcriptomic dysregulations associated with SARS-CoV-2 infection in human nasopharyngeal and peripheral blood mononuclear cells. bioRxiv, 2020: p. 2020.09.09.289850.

47. Bailey, M.H., et al., Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell, 2018. **173**(2): p. 371-385.e18.

48. Li, G., et al., Transcriptomic signatures and repurposing drugs for COVID-19 patients: findings of bioinformatics analyses. Computational and Structural Biotechnology Journal, 2021. **19**: p. 1-15.

49. Chen, N., et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. 2020. **395**(10223): p. 507-513.

50. Liu, T., et al., Differential Expression of Viral Transcripts From Single-Cell RNA Sequencing of Moderate and Severe COVID-19 Patients and Its Implications for Case Severity. Frontiers in microbiology, 2020. **11**: p. 603509-603509.

Figures

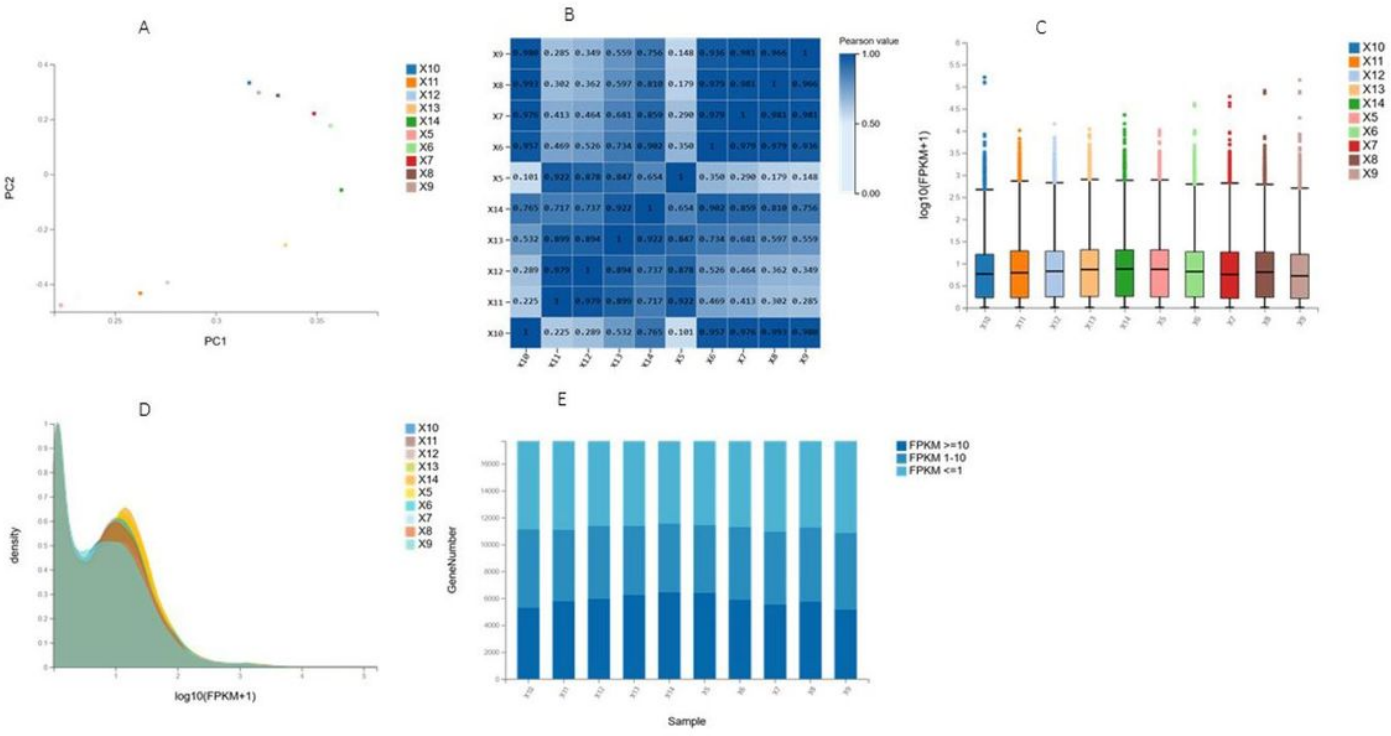

Figure 1

Quantification of PBMCs. A) The X-axes and Y-axes represented new data-sets of the corresponding principal components obtained after the dimension reduction of the sample expression. The dots represented each sample, and the same color represented the same sample group. B) The X-axes and Y-axes signified each sample. Color represented the correlation coefficient (darker colors represented higher correlations, lighter colors represented lower correlations). C) The X-axis was the sample name and the Y-axis was log10. The boxplot of each area corresponds to (Maximum, upper quartile, median, lower quartile, minimum) where outliers were not considered for upper and lower limits. D) The X-axis is log10, and the Y-axis is the density of the genes (the ratio of the number of genes to the total number of genes expression). E) The X-axis indicated the sample name, the Y-axis indicated the number of genes, and the shades of color indicate different gene expression levels: FPKM <= 1 was very low expression level, FPKM between 1-10 with a lower expression level, FPKM Genes with> = 10 were high and moderately expressed.

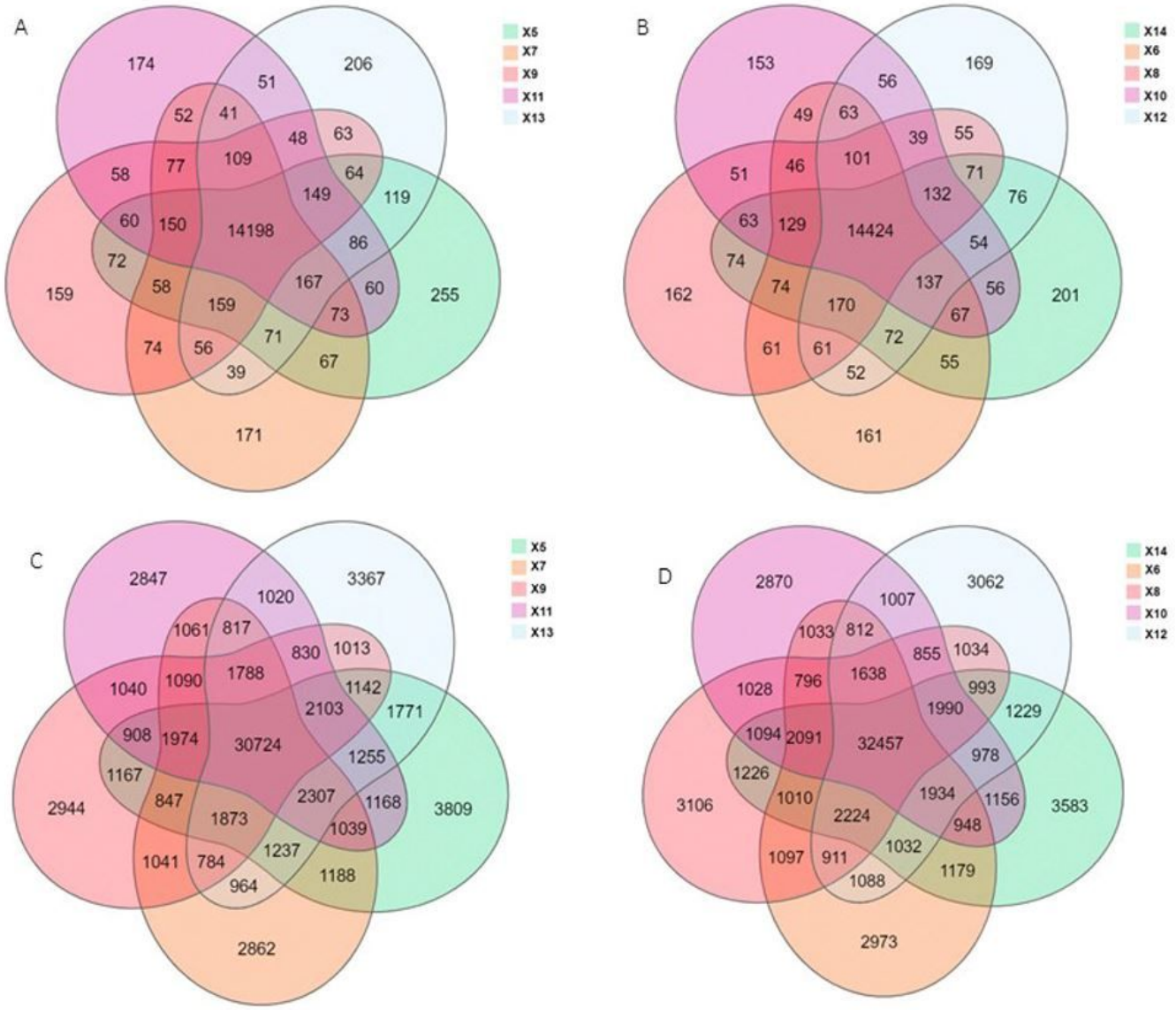

Figure 2

Genes-Transcripts Overview. It is exhibited that number of genes (A, B) and transcripts (C, D) intersection and union in each 3rd to 5th-months sample. Each circle represented group of gene sets, and superimposed areas by circles represented the intersection. Non-overlapping part indicated the uniquely expressed genes, and the numbers on the figure indicated the number of genes in the corresponding area.

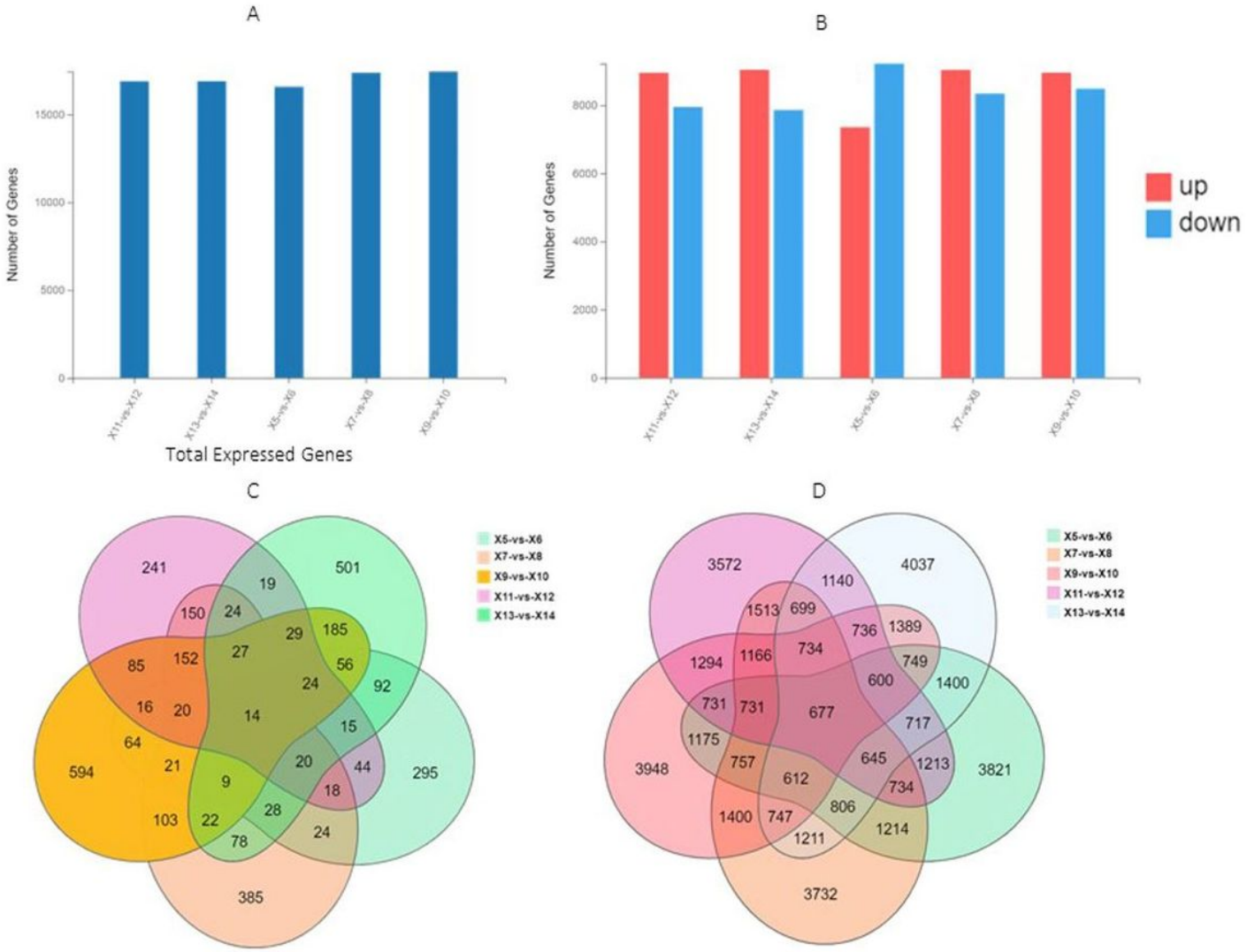

Figure 3

Differential-Expression Analysis. A) Total number of expressed genes B) It represented up/down-regulated genes. The X-axis represented the alignment scheme of DEGs for each group, and the Y-axis represented the corresponding number of DEGs. C) It exhibited number of DEGS, D) represented differentially expressed transcripts. Each circle represented group of gene sets, and the areas superimposed by different circles showed the gene set's intersection. Non-overlapping part indicated the uniquely expressed genes, and the numbers indicated the genes number in the corresponding area.

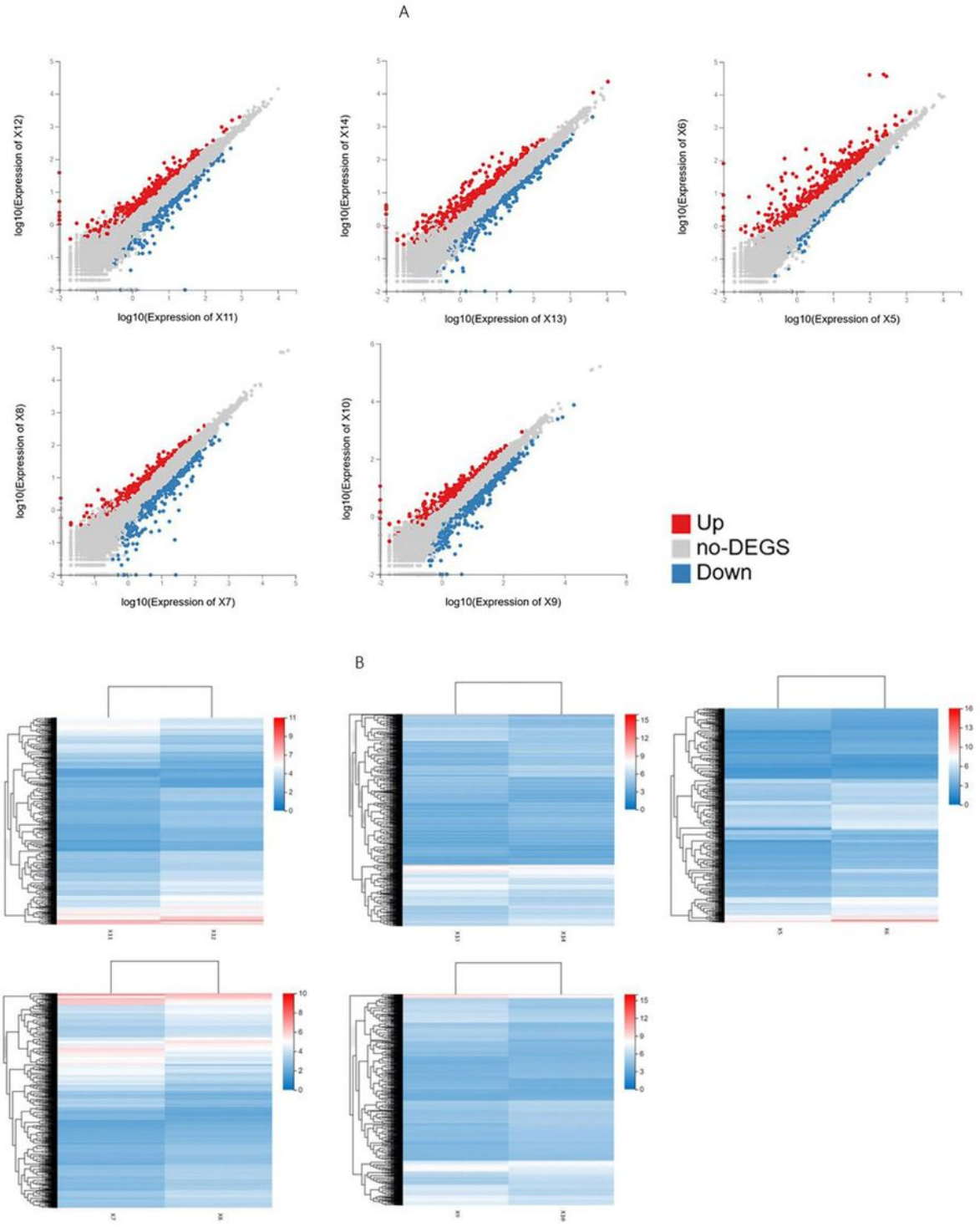

Figure 4

Differential Expression Pattern. A) X-axes and Y-axes characterized the logarithmic value of gene expression in each group. B) The horizontal axis is the log2 of sample (expression-value +1), and the vertical axis is the gene. Under the default color matching, the warmer the color block is, the higher the expression level is, and the colder the color block is, the lower the expression level.

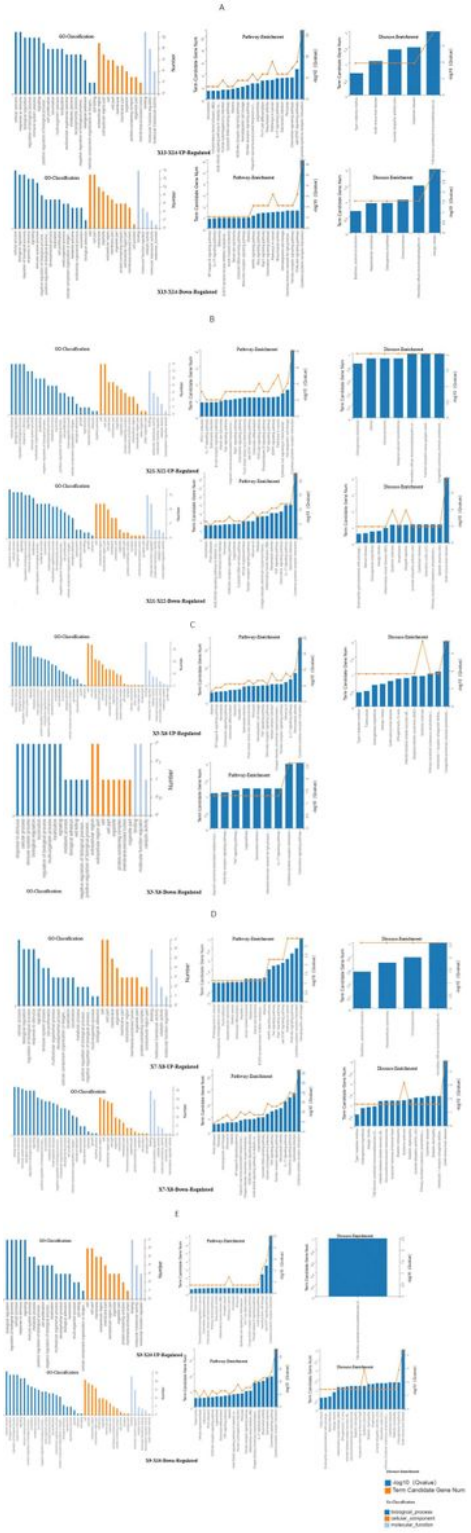

Figure 5

Pathway Enrichment of Cytokines. In Go-CPF analysis the X-axis represented the number of genes annotated to the GO entry, and the Y-axis showed the GO functional classification. While in pathway enrichment X-axis is the enrichment ratio (the ratio of the number of genes annotated to an entry in the selected gene set to the total number of genes annotated to the entry in the species), Y-axis is KEGG

Pathway. In kegg disease enrichment analysis X-axis is the disease enrichment ratio and Y-axis is cytokines number.

Supplementary Files

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

- SUPPLEMENTARYFIGURES.docx.pdf
- SUPPLEMENTARYTABLES.docx.doc