

Title Page

Comparative transcriptomic analyses of peripheral blood mononuclear cells of patients with non-pneumonia and severe pneumonia at 1 year-Long-COVID-19

Ozgecan Kayalar^{1*}, Pelin Duru Cetinkaya², Vahap Eldem³, Serap Argun Baris⁴, Nurdan Koktürk⁵, Selim Can Kuralay³, Hadi Rajabi¹, Nur Konyalilar¹, Deniz Mortazavi¹, Seval Kubra Korkunc¹, Sinem Erkan¹, Gizem Tuş Aksoy¹, Gul Eyikudamaci¹, Pelin Pinar Deniz², Oya Baydar Toprak², Pinar Yildiz Gulhan⁶, Gulseren Sagcan⁷, Neslihan Kose⁸, Aysegul Tomruk Erdem⁹, Fusun Fakili¹⁰, Onder Ozturk¹¹, Ilknur Basyigit⁴, Hasim Boyaci⁴, Emel Azak¹², Tansu Ulukavak Ciftci⁵, Ipek Kivilcim Oguzulgen⁵, Hasan Selcuk Ozger¹³, Pinar Aysert Yildiz¹³, Ismail Hanta², Ozlem Ataoglu⁶, Merve Ercelik⁶, Caglar Cuhadaroglu⁷, Hacer Kuzu Okur⁷, Muge Meltem Tor⁹, Esra Nurlu Temel¹⁴, Seval Kul¹⁰, Yıldız Tutuncu¹⁵, Oya Itil⁶, Hasan Bayram^{1, 17}

1 Koc University Research Center for Translational Medicine (KUTTAM), Koc University, School of Medicine, Istanbul, Turkey

2 Department of Pulmonary Medicine, Faculty of Medicine, Cukurova University, Adana, Turkey

3 Istanbul University, Science Faculty, Department of Biology, Istanbul, Turkey

4 Department of Pulmonary Medicine, Faculty of Medicine, Kocaeli University, Kocaeli, Turkey

5 Department of Pulmonary Medicine, Faculty of Medicine, Gazi University, Ankara, Turkey

6 Department of Pulmonary Medicine, Faculty of Medicine, Duzce University, Duzce, Turkey

7 Department of Pulmonary Medicine, Altunizade Acibadem Hospital, Istanbul, Turkey

8 Department of Pulmonary Medicine, Bilecik Training and Research Hospital, Bilecik, Turkey

9 Department of Pulmonary Medicine, Faculty of Medicine, Zonguldak Bulent Ecevit University, Zonguldak, Turkey

10 Department of Biostatistics, Faculty of Medicine, Gaziantep University, Gaziantep, Turkey

11 Department of Pulmonary Medicine, Faculty of Medicine, Suleyman Demirel University, Isparta, Turkey

12 Department of Infectious Disease and Clinical Microbiology, Faculty of Medicine, Kocaeli University, Kocaeli Turkey

13 Department of Infectious Diseases and Clinical Microbiology, Faculty of Medicine, Gazi University, Ankara, Turkey

14 Department of Infectious Diseases and Clinical Microbiology, Faculty of Medicine, Suleyman Demirel University, Isparta, Turkey

15 Department of Immunology, Koc University Research Center for Translational Medicine (KUTTAM), Koc University School of Medicine, Istanbul, Turkey

16 Department of Pulmonary Medicine, Faculty of Medicine, Dokuz Eylul University, Izmir, Turkey

17 Department of Pulmonary Medicine, Koc University School of Medicine, Istanbul, Turkey

***Corresponding Author:** Ozgecan Kayalar, Address: Koc University Research Centre for Translational Medicine (KUTTAM), Topkapi, Davutpasa Cad. No:4 Zeytinburnu, Istanbul/Turkey. E-mail: ozgecankayalar@gmail.com; okayalar@ku.edu.tr; Ozgecan Kayalar ORCID: 0000-0001-9107-2381.

Abstract

Long-COVID-19 manifests as a multisystemic condition with varied symptoms lingering beyond three weeks of acute SARS-CoV-2 infection, though its underlying mechanisms remain elusive. Aiming to decipher the long-term molecular impacts of COVID-19, we conducted a transcriptomic analysis on PBMCs from 1-year post-covid patients, including individuals without pneumonia (NP, n=10), those with severe pneumonia (SP, n=11), and healthy controls (C, n=13). Our extensive RNA sequencing revealed 4843 differentially expressed genes (DEGs) and 1056 differentially expressed long non-coding RNAs (DElncRNAs) in “C vs NP,” 1651 DEGs and 577 DElncRNAs in “C vs SP,” 954 DEGs and

148 DElncRNAs in “NP vs SP,” with 291 DEGs and 70 DElncRNAs shared across all groups. We identified 14 hub genes from 291 DEGs, with functional enrichment analysis showing upregulated DEGs mainly linked to inflammation and osteoclast differentiation, and downregulated DEGs to viral infections and immune responses. These hub genes play central roles in inflammatory and immune processes and are significantly associated with pneumonitis and diverse lung diseases. Investigations revealed unique immune cell signatures across DEG categories, associating upregulated DEGs with neutrophils and monocytes, and downregulated DEGs with CD4 memory effector T cells. Analysis of 14 hub genes showed notable upregulation in the no pneumonia group versus healthy controls, displaying complex patterns in the severe pneumonia group. Our study uncovered potential idiopathic pulmonary fibrosis signals in Long-COVID-19 patients’ PBMC transcriptome, highlighting the urgency for thorough monitoring and extended research to understand COVID-19’s lasting effects. This study sheds light on COVID-19’s transcriptomic changes and potential lasting effects, guiding future research and therapeutic approaches for Long-COVID-19.

Abstract: 250 words

Keywords: Lost-COVID-19, Transcriptomic analysis, Peripheral Blood Mononuclear Cells (PBMC), Long-term effects, Inflammatory responses, Pneumonia

Running Title: Long-COVID-19: Transcriptomic Insights from PBMCs

1. Introduction:

SARS-CoV-2, a highly contagious respiratory virus that is the source of the continuing worldwide pandemic, is the cause of the coronavirus disease 2019 (COVID-19). Common symptoms of COVID-19 include fever, cough, headache, lethargy, myalgia, diarrhoea, and anosmia in previously healthy people. COVID-19 often manifests as an asymptomatic or mild to moderate respiratory infection [1, 2]. Additionally, COVID-19 in patients with pre-existing co-morbidities such as obesity, respiratory, cardiovascular, and renal illnesses can swiftly progress into a serious, life-threatening condition needing immediate critical care support [3, 4]. According to the World Health Organization (WHO), there were 6,897,025 fatalities and 762,791,152 confirmed cases of COVID-19 worldwide as of 12 April 2023 (WHO, 2023).

The early acute stages of COVID-19 have been the subject of numerous studies, but studies of the long-term consequences of the disease have received much less attention, and still little is known. In the past two years, studies have reported that 30-70% of recovered individuals who had mild to severe COVID-19 struggling with prolonged symptoms for the time more than a 1 year after infection[5-10], have been commonly referred to as Long-COVID (CDC, 2021) or post-acute sequelae of COVID-19 (PASC)[11]. Long-COVID-19 is defined as a multisystemic condition with severe symptoms for at least 3 weeks or longer after the first negative PCR test following the acute course of SARS-CoV-2 infection [12]. Most cases are resolved within two to four weeks of the initial symptoms' appearance; however, after the initial infection, early studies have showed that these persistent symptoms in some cases may appear four weeks to less than a year later [13, 14]. The most common symptoms observed in the Long-COVID-19 studies include fatigue, headache, attention disorder, memory loss, hair loss, gastrointestinal (GI) distress, dyspnoea, anosmia, shortness of breath, pneumonia, and other symptoms[9, 15, 16].

Recent studies have revealed that Long-COVID-19 is associated with the features of the severity of the disease [7, 17] including the presence of chronic pulmonary fibrosis, pneumonia at initial diagnosis, dyspnoea, and emergency service admission in the Long-COVID-19 period [8]. Long-COVID-19 has been thought to be related to autoimmune mechanisms, unresolved viral fragments, and the presence of high-titre SARS-CoV-2 IgG response induced by the natural infection [7, 18, 19]. Additionally, although some infected people experience long-term radiological abnormalities linked to pulmonary function impairment [20-23], the underlying pathogenic pathways are poorly understood. Besides, it is

unclear why some patients experience complete clinical, physiological, and radiological recovery while others have a more dangerous outcome that includes persisting interstitial lung alterations and related pulmonary function impairment. It is still unknown whether the COVID-19 sequelae that are still present in the lungs are a result of the lingering effects of the initial increase in the inflammatory response or the activation of alternate pathways after the acute illness has subsided. To address these knowledge gaps, we carried out unbiased next-generation RNA sequencing from peripheral blood mononuclear cells (PBMCs) of the patients with the absence of pneumonia and presence of severe pneumonia in 1-year recovery after COVID-19 and the healthy controls without SARS-CoV-2 infection.

2. Material and methods

Ethical Statement:

This study followed the principles of the Declaration of Helsinki and was approved by the institutional ethics committee of Çukurova University School of Medicine (Approval number: 356/22.05.2021). Informed consent was obtained from the patients and the healthy volunteers for the collection of blood samples.

Data Availability:

FASTQ sequences of the PBMC samples have been deposited in NCBI Short Read Archive (SRA) under BioProject PRJNA895325.

Study population:

This study was conducted on patients who participated in the previously published TURCOVID study [8, 24]. During the first wave of the COVID-19 pandemic (between 11 March and 18 July 2020), 1 500 patients over the age of 18, who were monitored and treated because of COVID-19 were included in the targeted trial population in the multi-centre TTS-TURCOVID-19 registry cohort. A total of 831 patients were enrolled in the trial at 13 of the 26 locations (11 university hospitals, 2 sizable tertiary institutions, and 1 private hospital). A standard questionnaire was applied to current patients in the cohort over the phone after receiving written informed consent. Of the cohort of 831 patients, 272 (32.7%) could not be reached, 48 (5.8%) refused to participate in the study, 69 (8.3%) were excluded due to death, and the remaining 442 patients were included (Form-1). Retrospective data entry was performed over the recorded files, and an analysis of the medical records of 442 patients, who could be reached by phone, was applied (Form-2). One year later, 138 patients from 11 centres, who agreed to participate in the study and filled out a signed informed consent form, were called for a follow up check. A routine evaluation (clinical, laboratory and radiological) of the patients was performed (Form-3). From these cases, 27 were randomized into two groups using the computer program at <https://www.medcalc.org/>. Two groups were formed: one consisting of 13 cases (male: 8; female: 5) without pneumonia in their radiology findings after experiencing COVID-19, and the other consisting of 14 cases (male: 10; female: 4) with severe pneumonia. The control group was composed of 13 cases (male: 8; female: 5), who remained uninfected. Age, gender, and smoking status were taken into consideration to

eliminate confounding factors in the randomization process. Thus, the cases were divided into three groups: (i) those with no pneumonia following COVID-19 infection, (ii) those with severe pneumonia during COVID-19 infection, and (iii) the healthy control group with no disease (Table 1). Peripheral blood mononuclear cells (PBMC) were isolated from the blood obtained from the patients and stored at -80°C until used.

Blood Sample Collection and Isolation of Peripheral Blood Mononuclear Cells (PBMCs)

Approximately 20 ml of venous blood was collected from each participant, and PBMCs were isolated using a solution of Lymphoprep™ (Alere Technologies, Norway) by performing density gradient sedimentation at 2000 rpm for 20 minutes.

Transcriptome Library Construction and Next-generation RNA Sequencing:

Following isolation, the purity and integrity of RNA was assessed using a NanoDrop™ spectrophotometer (Thermo Scientific, Nanodrop 2000c) and RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively. Samples with RNA integrity number values >7.5 were retained for further processing. Before library preparation, mRNA and long non-coding RNAs (lncRNAs) were enriched by MGIEasy rRNA Depletion Kit (MGI Tech, China), and DNase I treatment (NEB) was carried out for complete removal of the DNA according to the manufacturer's instruction. RNA libraries were constructed from 500 ng RNA using the MGIEasy RNA Library Prep Kit V3.0 protocol (MGI, Shenzhen, China) according to the manufacturer's instruction. Firstly, enriched RNA samples were fragmented in fragmentation buffer, then short fragments were subjected to reverse transcription and second-strand synthesis. The cDNA fragments were treated with standard library generation steps; end-repair, A-tailing, and adapter ligation. After purification with DNA clean beads, we enriched adapter-ligated fragments using 14 PCR cycles and subjected to the following denaturation and single-strand circularization process to generate a single-stranded circular DNA library. These libraries were then used to generate DNA nanoballs (DNBs) by rolling circle replication (RCR). The resulting DNBs were then loaded into the patterned nanoarrays, and sequencing reaction was performed in a DNBSEQ-G400 sequencer with a pair-end read length of 100 bp.

RNA-Seq Data Analysis and Differential Expression Analysis

The quality of raw sequencing reads was checked with FastQC (Babraham Bioinformatics) before and after sequence trimming. For a comparison of the qualities of all RNA-Seq

libraries, MultiQC [25] software was used to merge the results of FastQC. Raw reads were filtered using fastp v0.23.0 [26] to remove adaptor contamination, ambiguous ($N > 5$) bases, low quality reads (Phred score, $Q < 20$), and fragments < 30 nt. All other options used the default values. Summary statistics for RNA-Seq reads were computed using seqkit v2.0.0 (Shen et al., 2016). Filtered reads were mapped to human reference genome (GRCh38.p13, Ensembl Release 106) using Hisat2 v2.2.1 [27]. The alignment statistics were obtained with Sambamba v0.8.0 [28]. Count matrices and gene-level assignment were generated using featureCounts from Subread package v2.0.0 [29] with annotation version GRCh38.106 (Ensembl “.gtf”). Differential gene expression between groups was performed on raw counts using DESeq2 v1.34.0 [30] after variance-stabilizing transform (vst) normalization. Genes were considered as significantly differentially expressed if the adjusted P -value (Benjamini–Hochberg (BH) multiple test correction method) was less than 0.001 and $\log_2FC > 1.0$. Hierarchical clustering and principal component analysis (PCA) were performed using DEBrowser v1.20.0 [31] to evaluate the correlation between control and disease samples. Volcano plots of most differentially expressed genes (DEGs) among comparison groups were generated using EnhancedVolcano v1.12.0 [32]. The top 34 up-regulated and down-regulated protein-coding genes (sorted by adjusted p -value in increasing order) for each comparison were selected for heatmap generation using the online tool ClustVis[33]. The multiple plots generated from each condition were combined one overall graph using ggarrange() function available in the ggpubr R package (<https://rpkgs.datanovia.com/ggpubr/>).

Functional Annotation and Enrichment Analysis

Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEG among groups of control, no pneumonia and severe pneumonia was performed with Metascape using *Homo sapiens* (Ensembl Release 104) as background for enrichment[34]. The KEGG pathway enrichment analysis of DElncRNAs among groups of control, no pneumonia and severe pneumonia was performed with NCPATH using *Homo sapiens* (Ensembl Release 104) as background for enrichment[35]. Significance of enrichment analysis was estimated by Benjamini–Hochberg false discovery rate (FDR) < 0.05 correction. Protein-protein interaction (PIP) network analysis was performed using STRINGdb v12.0 [36]. Normalized expression data were used for discovery co-expressed modules, gene-diseases interaction, gene interactions with transcription factors and their targets, immune cell signatures using Metascape[34] and Enrichr [37, 38]. A variance filter value of 0.01 was used

to ensure the highest level of statistical stringency and Pearson's correlation method was selected for identification of the gene modules.

Identification of 52 Genes Associated with Idiopathic Pulmonary Fibrosis Risk in The PBMCs' Transcriptome

In our PBMC transcriptome data, we compared up and down profiles based on the expression levels of 7 increased genes (PLBD1, TPST1, MCEMP1, IL1R2, HP, FLT3, S100A12) and 45 decreased genes (LCK, CAMK2D, NUP43, SLAMF7, LRRC39, ICOS, CD47, LBH, SH2D1A, CNOT6L, METTL8, ETS1, P2RY10, TRAT1, BTN3A1, LARP4, TC2N, GPR183, MORC4, STAT4, LPAR6, CPED1, DOCK10, ARHGAP5, HLA-DPA1, BIRC3, GPR174, CD28, UTRN, CD2, HLA-DPB1, ARL4C, BTN3A3, CXCR6, DYNC2LI1, BTN3A2, ITK, CD96, GBP4, S1PR1, NAP1L2, KLF12, IL7R, SNHG1, C2orf27A) from a gene signature previously found to be predictive of IPF poor prognosis and COVID-19 outcome [39-41]. The genes (sorted by adjusted p-value in increasing order) for each comparison were selected for heatmap generation using the online tool ClustVis[33].

Statistical Analysis

Data were checked for normality and continuous variables were compared using one-way variance analysis, ANOVA/Dunnett's multiple comparison tests, or Kruskal-Wallis/Dunn's multiple comparison tests in the context of the RNAseq count data-based on gene expression analysis of the hub genes. The findings are presented as median $\square \pm \square$ interquartile (IQ) ranges or mean $\square \pm \square$ SD. P values were considered significant if they were less than 0.05. PRISM version 8 (GraphPad Software Inc, San Diego, CA, United States) was used for the statistical analysis.

3. Results:

3.1. Identification of DEGs and DElncRNAs among Healthy Control, No Pneumonia and Severe Pneumonia Groups

The success of the transcriptome sequencing reaction depends on the quality and quantity of the RNA isolated. Of the 40 PBMC-RNA samples sent for transcriptomic analysis, 36 samples with high quality RNA were analysed, except for three patients with severe pneumonia and one patient without pneumonia. The RNA samples of the other 2 patients without pneumonia were not included in the analysis because it was determined that they could not reach the reading quality with sufficient number and accuracy for bioinformatic analysis. After the bioinformatic analysis of the raw data obtained after the sequence analysis, a total of 34 RNA samples from the control group (N=13), the post-Covid 19 group without pneumonia (N=10) and the post-Covid-19 group with severe pneumonia (N=11) were included in the analysis because the reading quality was of sufficient quality (**Figure 1a**). The relative status of gene expression changes in all three groups is shown by PCA analysis using the transcriptomic gene expression profile from PBMCs from control (n=13), no pneumonia (n=10), and severe pneumonia (n=11) individuals. As a result, we observed that the control samples grouped apart from the other two groups and that the gene expression variations in people with no and severe pneumonia differed more than those in the control group (**Figure 1b**). Moreover, a cluster heat map was created from 34 upregulated and 34 downregulated genes to show differential gene expression in each group comparisons (**Figure 1c**). In the comparison between healthy control (C) and no pneumonia group (NP), we identified 4843 DEGs including 3004 upregulated (up) genes and 1839 downregulated (down) genes, and 1056 DElncRNAs including 694 up and 392 down lncRNAs (**Figure 2a-c; Figure 6a-c**). We identified 1651 DEGs including 1566 up and 85 down genes, and 577 DElncRNAs including 493 up and 84 down lncRNAs between healthy control and severe pneumonia group (C versus SP). In NP vs SP comparison, we identified 954 DEGs including 79 up and 875 down genes, and 148 DElncRNAs including 5 up and 143 down lncRNAs (**Figure 2a-c; Figure 6a-c**). We then overlapped DEGs from all comparisons. Totally, we identified 291 DEGs and 70 DElncRNAs. Additionally, Upregulated and Downregulated DEGs and DElncRNAs were showed in Venn diagrams. Next, we tried to identify the function of common DEGs and DElncRNAs involved during the pneumonia triggered by Long Covid-19 (**Figure 2a-c; Figure 6a-c**).

3.2. Functional Enrichment Analysis of all DEGs

To understand the function and pathways of all DEGs, enrichment analysis was performed. This analysis revealed that upregulated DEGs in “C vs NP” comparison was involved in osteoclast differentiation (KEGG) and positive regulation of inflammatory process (GO:0050727). The DEGs downregulated in the comparison, however, were associated with herpes simplex virus 1 infection (KEGG) and the positive regulation of natural killer cell mediated immunity (GO:0002717). The upregulated DEGs were mainly located in tertiary granules, and the main molecular function of these genes was found as protein serin/threonine kinase activity. The downregulated DEGs were mainly located in sarcoglycan and dystroglycan complexes, and the main molecular function of these down genes was found to bind to DNA via RNA polymerase II transcription regulatory region sequence (**Figure S1a**). Upregulated DEGs in “C vs SP” comparison was involved in osteoclast differentiation (KEGG) and positive regulation of inflammatory process (GO:0050727). The DEGs downregulated in the comparison, however, were associated with primary immunodeficiency (KEGG) and alcohol catabolic process (GO:0046164). The upregulated DEGs were mainly located in secretory granule membrane, and the main molecular function of these genes was found as cytokine receptor activity and G protein coupled receptor activity. The downregulated DEGs were mainly located in keratin and intermediate filaments, and the main molecular function of these down genes was found as oxidoreductase activity acting on NAD or NADP as an acceptor (**Figure S1b**).

Upregulated DEGs in “NP vs SP” comparison was involved in herpes simplex virus 1 infection, viral myocarditis, arrhythmogenic right ventricular cardiomyopathy, sulphur metabolism, hypertrophic cardiomyopathy, and TGF-beta signaling pathway (KEGG) in addition to the regulation of lymphocyte activation (GO:0051249), and some cardiac tissue morphogenesis processes (such as GO:0048738, GO:0061384, GO:0051146). The DEGs downregulated in the comparison, however, were associated in the transcriptional misregulation in cancer and TNF signaling (KEGG), and the positive regulation of transcription by RNA polymerase II (GO:0045944). The upregulated DEGs were mainly located in sarcoglycan and dystroglycan complexes, and the main molecular function of these genes was found as serine-type endopeptidase activity. The downregulated DEGs were mainly located in specific and azurophil granules, and the main molecular function of these genes was found as protein serine/threonine kinase activity. (**Figure S1c**).

3.3. Functional Enrichment Analysis of Common DEGs and DElncRNAs

Enrichment analysis was conducted to gain a deeper understanding of the function and pathways of common DEGs. The results indicated that 291 common DEGs across all DEGs were primarily engaged in inflammatory response processes, including TNF-alpha, NF-κB, and MAPK signalling pathways (**Figure 2a**). Moreover, the extended common 1436 DEGs in upregulated DEGs were involved in neutrophil degranulation, neutrophil extracellular trap formation, and osteoclast differentiation (KEGG), and cellular inflammatory response (GO: 0006954 and GO:0071345) (**Figure 2b**). Furthermore, the extended common 71 DEGs in downregulated DEGs were mainly involved in primary immunodeficiency (KEGG), and some metabolic processes such as acyl-CoA process (GO: 0006637), small molecule catabolic process (GO: 0044282), and protein glycolisation (GO: 0006486) (**Figure 2c**). The results indicated that 70 common lncRNAs across all DElncRNAs were primarily engaged in MAPK and Rap1 signaling pathways (KEGG). Moreover, the extended common 457 lncRNAs in upregulated DElncRNAs were involved in MAPK signaling pathway, focal adhesion, cell cycle, and insulin resistance (KEGG). Furthermore, the extended common 79 lncRNAs in downregulated DElncRNAs were mainly involved in thermogenesis and mTOR signaling pathway (KEGG). Finally, we revealed that 23 common DEGs and 70 DElncRNAs were enriched in MAPK, Rap1, and AMPK signaling pathways (KEGG) (**Table S1**).

3.3. Identification of Hub Genes via Protein–Protein Interaction (PPI) Network

The PPI network analysis revealed an association of DEGs among healthy control, no pneumonia and severe pneumonia groups. We identified 291 common DEGs consisted of 291 nodes with 142 edges (**Figure 3a**). Thirty-six genes were determined from the 291 common genes using Metascape online software. Among 36 genes, the highest scoring (Interactions \geq 3) in 23 central genes, namely ICAM1, TUBB4B, MARCKS, NFKB2, NFKBIA, NFKBIE, HDAC5, ATF3, DDIT3, F3, PRKCD, IL1R1, AREG, CSF1, IL1R2, TOM1, RAB11FIP1, FSCN1, ULK1, RELB, NFKBIB, FOSL1, and JUND were identified from the PPI network in combination with the Metascape (**Table 2**). When we analysed these 23 genes, we obtained a hub gene network with a total of 14 genes in 2 modules (**Figure 3c**), and then, we revealed that these genes were specifically involved in the TNF-alpha/NF-κB signaling complex, osteoclast differentiation, and cytokine signaling in the immune system.

3.3. Identification of Gene-diseases, Gene-transcription factors, and Gene-transcription factor Targets Interactions

Gene-Diseases interaction via DisGeNET interestingly showed that the 291 common genes were related to pneumonitis, middle cerebral artery occlusion, acute pancreatitis, secondary malignant neoplasm of bone, choriocarcinoma, lung diseases, malignant neoplasm of mouth, myocardial ischemia, pancreatic neoplasm, infection and juvenile arthritis. We determined that these genes consisted of genes expressed especially under the control of NFKB1, RELA, JUN and ATF4 transcription factors (**Figure 3b**). Our remarkable discovery was that the 14 hub genes we identified were primarily related to pneumonitis according to DisGeNET analysis. In our analysis, we saw that these genes, especially those carrying NFKB Q6, one of the transcription factor targets, were under the control of JUN (**Figure 3d**).

3.4. Identification of Immune Cell Type Signatures of the Common and Central DEGs

To determine immune cell type signatures in the extended common 1436 upregulated, the extended common 71 downregulated, 291 common, and 23 central DEGs, our Metascape analysis showed that the 1436 DEGs mainly involved in neutrophil cells and monocytes (**Figure 4a**). Downregulated common 71 DEGs mainly involved in CD4 memory effector T cells (**Figure 4b**). The common 291 DEGs involved in monocytes, dendritic cells, basophils, and neutrophils (**Figure 4c**). Interestingly, 23 central genes mainly involved in CCL19 and CCL21 positive cells, monocytes, basophils, and dendritic cells (**Figure 4d**).

3.5. Gene Expression Analysis Based on RNA-seq Count Data of 14 Hub Genes.

We performed a gene expression analysis based on statistical analysis of the normalized counts of the hub genes and GAPDH of each patient and control. According to these results, all genes were significantly upregulated in no pneumonia groups compared to healthy control individuals. Also, the other genes except *FOSL1* and *CSF1* were significantly upregulated in severe pneumonia group compared to the control groups. *FOSL1* ($p < 0.05$) and *CSF1* ($p < 0.01$) genes were downregulated in severe pneumonia group compared to no pneumonia group. In other genes, no significance was found between no pneumonia and severe pneumonia (**Figure 5**).

3.6. Investigation of Idiopathic Pulmonary Fibrosis related 52 Gene Signatures in PBMC Transcriptome of Patients with Non-Pneumonia and Severe Pneumonia

The transcriptome signals of the PBMCs from individual patients in the "NP " and "SP" groups were analysed using a predefined set of 52 PBMC signature genes and visualised in a heatmap to determine whether these groups exhibited IPF signals. Our results shown that the expression levels of the genes PLBD1, TPST1, MCEMP1, IL1R2, HP, FLT3, and S100A12, which were low in the control group, significantly increased in patients, who had COVID-19 but did not develop pneumonia ('NP'). In patients with severe pneumonia due to COVID-19 ('SP'), an increase was observed in PLBD1, S100A12, and MCEMP1 genes, while no significant changes were noted in the other four genes (**Figure S1d**). Also, we demonstrated that the expression levels of the genes LCK, CAMK2D, NUP43, SLAMF7, LRRC39, ICOS, CD47, LBH, SH2D1A, CNOT6L, METTL8, ETS1, P2RY10, TRAT1, BTN3A1, LARP4, TC2N, GPR183, MORC4, STAT4, LPAR6, CPED1, DOCK10, ARHGAP5, HLA-DPA1, BIRC3, GPR174, CD28, UTRN, CD2, HLA-DPB1, ARL4C, BTN3A3, CXCR6, DYNC2LI1, BTN3A2, ITK, CD96, GBP4, S1PR1, NAP1L2, KLF12, IL7R, SNHG1, and C2orf27A, which were high in the control group, significantly decreased in both NP and SP groups (**Figure S1d**).

4. Discussion

Subsequent to the propagation of COVID-19, the observation of protracted health complications among individuals recuperating from the acute phase of the disease has engendered novel and complex challenges for the scientific community and healthcare professionals. Termed as "Long-COVID," which is characterized by persistent symptoms that pervasively affect various organ systems, culminating in a substantial diminution of patients' quality of life. These symptoms encompass fatigue, dyspnoea, pneumonia, pulmonary fibrosis, and pervasive inflammatory responses[8]. The multi-omics analyses of whole blood and PBMC samples have been instrumental in elucidating the cellular signals underpinning the disease's mechanisms and symptoms, contributing significantly to parallel research endeavours [42] [43].

In our study, we conducted a comprehensive analysis of RNA sequences from PBMCs to interrogate the differential gene and lncRNA expression profiles amongst healthy controls, Lost Covid-19 patients without pneumonia (NP), and Long-COVID-19 patients with severe pneumonia (SP). The investigation's main finding indicates that, regardless of pneumonia manifestation, the TNF-alpha/NFκB signalling complex and MAPK signalling pathways regulate the positive control of inflammatory processes one year after COVID-19. This modulation, together with neutrophil-associated signatures, could be the underlying cause of the reported problems in patients. In one of the pioneering studies, proteomic analysis was conducted with blood plasma collected for two 6-month visits, and multiplex gene expression analysis was performed with RNA obtained from nasal epithelial cells. This study showed the presence of elevated systemic inflammatory signals even 3-6 months post-disease, and reported that the radiological and functional alterations observed in these individuals did not revert to normal within a 12-month period [44]. These findings corroborate the alterations observed in our studies. However, our investigation is distinguished as the inaugural transcriptomic analysis conducted with PBMCs isolated from blood samples obtained at the end of one year, from randomly selected long-Covid-19 patients with no pneumonia, severe pneumonia, and healthy controls, matched for age and gender. The PCA analysis and cluster heat maps provided a clear visual representation of the variations in gene expression, with the control samples distinctly separating from the patient groups. This separation is indicative of a fundamental shift in gene expression dynamics in Long-COVID-19, which could be instrumental in understanding the lingering effects of the disease and potential avenues for therapeutic interventions. The identification of DEGs revealed distinct molecular signatures

between the groups. Based on group comparisons of healthy control vs no pneumonia, healthy control vs severe pneumonia, and no pneumonia vs severe pneumonia, DEGs that were upregulated in both no pneumonia and severe pneumonia patients compared to control individuals were enriched in osteoclast differentiation, HIF1 α , IL-17 and TNF signaling pathways, while downregulated genes were found to be enriched in Herpes simplex virus 1 infection in the "C vs NP" comparison and primary immunodeficiency in the "C vs SP" comparison. This situation suggests the long-lasting effect observed following excessively increased inflammatory signals after virus infection. According to the no pneumonia group, it was determined that the upregulated genes in patients with severe pneumonia observed in their radiology and clinics were mainly clustered in herpes simplex virus 1 infection and cardiac changes secondary to the virus, while the downregulated genes were especially found to be associated with transcriptional mis-regulation in cancer and decrease in TNF- α signaling. The "NP vs SP" comparison revealed a unique set of DEGs involved in cardiac-related processes and viral infections, highlighting the potential cardiovascular risks and lingering susceptibility to viral infections in Long-COVID-19 patients.

Previous investigations have elucidated that TNF- α accelerates osteoclast differentiation independently through the receptor activator of NF- κ B ligand (RANKL), RANK, and TRAF6, while IL-17A facilitates the same process via RANKL-JNK1 signaling pathway[45, 46] [47]. Studies have reported that the differentiation of osteoclasts is intricately coordinated through the binding of macrophage colony-stimulating factor (M-CSF) and RANKL to their respective receptors on the surface of osteoclast precursor cells. It has been documented that M-CSF plays a pivotal role in the proliferation of osteoclasts, whereas RANKL is instrumental in their differentiation[48].

In a murine model study, silica-induced pulmonary fibrosis was observed to stimulate osteoclast-like differentiation, leading to the recruitment of monocytes to the tissue. This process of osteoclast-like differentiation of alveolar macrophages is mediated by the osteoclastogenic cytokine RANKL, which is released from pulmonary lymphocytes and type II alveolar epithelial cells [49]. Collectively, these studies suggest the potential involvement of this mechanism in the persistent effects observed post-viral infection.

The identification of common DEGs and DElncRNAs across all comparisons, as well as the elucidation of hub genes, provided a holistic view of the pivotal molecular players in Long-Covid-19 conditions. In present study, the enrichment of these common DEGs in inflammatory response processes and signaling pathways, such as TNF- α , NF- κ B, and MAPK,

underscores the persistent inflammatory state and potential for chronic complications in Long-Covid-19 patients. This augmentation of signatures was already expressed in an editorial [50]. The PPI network analysis and identification of hub genes further highlighted the interconnectedness of these molecular players, with genes such as *RELB*, *NFKB2*, *NFKBIA*, *NFKBIB*, *NFKBIE*, *HDAC5*, *ATF3*, *DDIT3*, *FOSL1*, *JUND*, *CSF1*, *ICAM1*, *IL1R1* and *IL1R2* emerging as central nodes in the network. These hub genes are known to play crucial roles in immune responses and inflammation, reinforcing the idea that Covid-19 conditions are characterized by prolonged immune activation and potential dysregulation. In a multi-omics analysis performed on the pulmonary tissues of COVID-19 patients [51], the study elucidated that the complications observed in the acute phase of the disease were underpinned by augmented mechanisms of senescence, inflammation, apoptosis, coagulation, and fibrosis. Furthermore, the genes and proteins playing a role in these mechanisms, as well as their associated signaling networks, were elucidated. Intriguingly, within these networks, certain molecular entities (*RELB*, *DDIT3*, and *FOSL1*) were identified as hub genes with increased expression in our patient samples during the 1-year Long-COVID-19 follow-up period. These genes are thought to play a role in ongoing pneumonia and are believed to bridge the networks of apoptosis, inflammation, and fibrosis [51]. The continued upregulation of these genes, initially detected in the acute phase, throughout the 1-year recovery period Long-COVID-19 is noteworthy. This phenomenon could potentially lead to the emergence of pneumonia complications and other inflammation-related outcomes.

Interestingly, our gene-diseases interaction analysis provided intriguing links between the common DEGs and various diseases, including especially pneumonitis, other lung-related conditions, and arthritis. Transcriptomic study of the acute phase PBMCs of COVID-19 patients revealed a number of immune-related illnesses, such as pneumonia, arthritis, and septicemia, in earlier research[43]. As far as is known, the lung is the organ primarily affected by SARS-CoV-2 infection. SARS-CoV-2 spreads in the lower respiratory tract of the severe patients, resulting in hypoxemia, severe pneumonia, and acute respiratory distress syndrome[52]. When single-cell RNA-seq analyses were conducted on pathological samples taken from the lungs of COVID-19 patients during the acute phase, they found strong inflammation and cellular markers associated with apoptosis-dependent fibrotic alterations. They hypothesised that these signals could be connected to inflammation and fibrotic alterations brought on by acute pneumonia [51].

The immune cell type signature analysis revealed distinct profiles between the upregulated and downregulated DEGs, with neutrophils and monocytes dominating the upregulated DEGs and CD4 memory effector T cells being prominent among the downregulated DEGs. Transcriptomic analyses performed on the PBMCs of COVID-19 patients in the acute phase have reported an enrichment of upregulated genes in bone marrow and blood CD33⁺ myeloid cells as well as CD14⁺ monocytes, while downregulated genes were observed to be enriched in CD4 and CD8 T cells, as we have also noted [43]. This shift in immune cell type signatures could have implications for the patient's immune response and susceptibility to secondary infections in Long-COVID-19. The gene expression analysis of the 14 hub genes provided a granular view of their expression patterns across the groups, with significant upregulation observed in the no pneumonia group compared to healthy controls. This upregulation, coupled with the distinct expression patterns in the severe pneumonia group, suggests a complex interplay of molecular factors contributing to the varied clinical outcomes in Long-COVID-19 patients. The investigation of idiopathic pulmonary fibrosis (IPF) related gene signatures revealed a subset of genes with altered expression in the no pneumonia and severe pneumonia groups, providing a potential link between COVID-19 and the risk of developing IPF-like conditions. In our investigation, all seven genes exhibited increased expression in COVID-19 patients without pneumonia, while only certain genes like PLBD1, S100A12, and MCEMP1 showed increased expression in severe pneumonia cases. These seven genes' upregulation correlates with severe IPF and COVID-19 outcomes, as evidenced by prior studies [39-41]. Additionally, all 45 genes displayed decreased expression across both patient groups and have been previously associated with severe IPF and COVID-19 risks. The NP group presented transcriptomic signatures similar to IPF [41]. We identified these high-risk genes predominantly in CD14⁺ monocytes, dendritic cells, and neutrophils, highlighting their pivotal role in mediating high-risk gene-associated responses.

5. Conclusion

In summary, our study provides a comprehensive transcriptomic analysis of PBMCs in post-Covid-19 conditions, uncovering a complex landscape of DEGs, DElncRNAs, and hub genes associated with various clinical outcomes. The persistent inflammatory state, alterations in immune cell type signatures, and potential links to lung-related conditions and IPF highlight the need for ongoing research and monitoring of post-Covid-19 patients. Future studies should focus on validating these findings in larger cohorts, exploring the therapeutic implications of

these molecular alterations, and investigating the long-term health impacts of Covid-19 to enhance our understanding and improve the care of affected individuals.

Author contributions:

OK, HB, VE, SAB, NK, DM, OBT, PDC, FF and OI designed the study; SAB, SCK, OBT, PDC, FF, NK, DM, SK, OK, YT, EA, MK, PAY, PPD, OK, IB, HB, IH, NK, GS, CC, HKO, HSO, BE, MH, AS, ENT, OO, TUC, IKO, VAO, FB, OA, ME, PYG, ATE, MMT, OI and HB collected the data; OK, VE, SCK, HB analysed the data; OK, HR, NK, DM, SKK, SE, GTA, GE and YT searched the literature and OK wrote the manuscript; SAB, OBT, FF, PDC, NK and HB edited and revised manuscript according to journal's instructions; All authors edited and controlled the final version of the manuscript. All the authors approved the final version of the manuscript.

Conflict of interest statement

The authors have no conflict of interest to declare.

Funding

This study was funded by the Research Support Fund of the Turkish Thoracic Society (TTD).

Acknowledgments

We greatly appreciate the patients and healthy volunteers involved in the study. Also, we would like to thank all of the TURCOVID and POST-COVID collaborators, who cooperated for this study. We would like to thank the Turkish Thoracic Society for the funding and support. Furthermore, the authors acknowledge the use of the services and facilities of the Koç University Research Center for Translational Medicine (KUTTAM), funded by the Presidency of Turkey, Head of Strategy and Budget.

References :

1. da Rosa Mesquita, R., et al., Clinical manifestations of COVID-19 in the general population: systematic review. *Wien. Klin. Wochenschr*, 2021. 133(7-8): p. 377-382.
2. Mizrahi, B., et al., Longitudinal symptom dynamics of COVID-19 infection. *Nat. Commun*, 2020. 11(1): p. 6208.
3. Flaherty, G.T., et al., COVID-19 in adult patients with pre-existing chronic cardiac, respiratory and metabolic disease: a critical literature review with clinical recommendations. *Trop. Dis. Travel Med. Vaccines*, 2020. 6: p. 1-13.
4. Williamson, E.J., et al., Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 2020. 584(7821): p. 430-436.
5. Iqbal, F.M., et al., Characteristics and predictors of acute and chronic post-COVID syndrome:

- A systematic review and meta-analysis. *EClinicalMedicine*, 2021. 36.
6. Logue, J.K., et al., Sequelae in adults at 6 months after COVID-19 infection. *JAMA Netw. Open*, 2021. 4(2): p. e210830-e210830.
 7. Peghin, M., et al., Post-COVID-19 syndrome and humoral response association after 1 year in vaccinated and unvaccinated patients. *Clin. Microbiol. Infect*, 2022. 28(8): p. 1140-1148.
 8. Baris, S.A., et al., The predictors of long-COVID in the cohort of Turkish Thoracic Society-TURCOVID multicenter registry: One year follow-up results. *Asian Pac. J. Trop. Med*, 2022. 15(9): p. 400-409..
 9. Imoto, W., et al., A cross-sectional, multicenter survey of the prevalence and risk factors for Long COVID. *Sci. Rep*, 2022. 12(1): p. 22413.
 10. Hastie, C.E., et al., Outcomes among confirmed cases and a matched comparison group in the Long-COVID in Scotland study. *Nat. Commun*, 2022. 13(1): p. 5663.
 11. Groff, D., et al., Short-term and long-term rates of postacute sequelae of SARS-CoV-2 infection: a systematic review. *JAMA Netw. Open*, 2021. 4(10): p. e2128568-e2128568.
 12. Davis, H.E., et al., Long COVID: major findings, mechanisms and recommendations. *Nat. Rev. Microbiol*, 2023. 21(3): p. 133-146.
 13. Huang, C., et al., 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *The Lancet*, 2021. 397(10270): p. 220-232.
 14. Nalbandian, A., et al., Post-acute COVID-19 syndrome. *Nat. Med*, 2021. 27(4): p. 601-615.
 15. Lopez-Leon, S., et al., More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci. Rep*, 2021. 11(1): p. 16144.
 16. Mizrahi, B., et al., Long covid outcomes at one year after mild SARS-CoV-2 infection: nationwide cohort study. *Bmj*, 2023. 380.
 17. Blomberg, B., et al., Long COVID in a prospective cohort of home-isolated patients. *Nat. Med*, 2021. 27(9): p. 1607-1613.
 18. Ramakrishnan, R.K., et al., Unraveling the mystery surrounding post-acute sequelae of COVID-19. *Front. immunol*, 2021. 12: p. 686029.
 19. Su, Y., et al., Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*, 2022. 185(5): p. 881-895. e20.
 20. Guler, S.A., et al., Pulmonary function and radiological features 4 months after COVID-19: first results from the national prospective observational Swiss COVID-19 lung study. *Eur. Respir. J*, 2021. 57(4).
 21. Han, X., et al., Six-month follow-up chest CT findings after severe COVID-19 pneumonia. *Radiology*, 2021. 299(1): p. E177-E186.
 22. Mo, X., et al., Abnormal pulmonary function in COVID-19 patients at time of hospital discharge. *Eur. Respir. J.*, 2020. 55(6).
 23. Xu, Y.-H., et al., Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. *J. Infect*, 2020. 80(4): p. 394-400.
 24. Kokturk, N., et al., The predictors of COVID-19 mortality in a nationwide cohort of Turkish patients. *Respir Med*, 2021. 183: p. 106433.
 25. Ewels, P., et al., MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 2016. 32(19): p. 3047-3048.
 26. Chen, S., et al., Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884-i890 (2018). *Sci. Rep*, 2021. 11: p. 20840.
 27. Kim, D., et al., Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol*, 2019. 37(8): p. 907-915.
 28. Tarasov, A., et al., Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, 2015. 31(12): p. 2032-2034.
 29. Liao, Y., G.K. Smyth, and W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 2013. 30(7): p. 923-930.
 30. Love, M.I., W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 2014. 15(12): p. 1-21.

31. Kucukural, A., et al., DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC genom*, 2019. **20**(1): p. 1-12.
32. Blighe, K., S. Rana, and M. Lewis, EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling." R package version R package version, 2019. **1**(0).
33. Metsalu, T. and J. Vilo, ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res*, 2015. **43**(W1): p. W566-W570.
34. Zhou, Y., et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun*, 2019. **10**(1): p. 1523.
35. Li, Z., et al., NcPath: a novel platform for visualization and enrichment analysis of human non-coding RNA and KEGG signaling pathways. *Bioinformatics*, 2022. **39**(1).
36. Szklarczyk, D., et al., The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*, 2021. **49**(D1): p. D605-D612.
37. Chen, E.Y., et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 2013. **14**(1): p. 1-14.
38. Xie, Z., et al., Gene set knowledge discovery with Enrichr. *Current protocols*, 2021. **1**(3): p. e90.
39. Herazo-Maya, J.D., et al., Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci. Transl. Med*, 2013. **5**(205): p. 205ra136-205ra136.
40. Herazo-Maya, J.D., et al., Validation of a 52-gene risk profile for outcome prediction in patients with idiopathic pulmonary fibrosis: an international, multicentre, cohort study. *The Lancet Respir Med*, 2017. **5**(11): p. 857-868.
41. Guardela, B.M.J., et al., 50-gene risk profiles in peripheral blood predict COVID-19 outcomes: A retrospective, multicenter cohort study. *EBioMedicine*, 2021. **69**.
42. Li, J., et al., PBMC transcriptomics identifies immune-metabolism disorder during the development of HBV-ACLF. *Gut*, 2022. **71**(1): p. 163-175.
43. Sadanandam, A., et al., A blood transcriptome-based analysis of disease progression, immune regulation, and symptoms in coronavirus-infected patients. *Cell Death Discov.*, 2020. **6**(1): p. 141.
44. George, P.M., et al., A persistent neutrophil-associated immune signature characterizes post–COVID-19 pulmonary sequelae. *Sci. Transl. Med*, 2022. **14**(671): p. eabo5795.
45. Kim, N., et al., Osteoclast differentiation independent of the TRANCE–RANK–TRAF6 axis. *J. Exp. Med*, 2005. **202**(5): p. 589-595.
46. Yao, Z., et al., RANKL cytokine enhances TNF-induced osteoclastogenesis independently of TNF receptor associated factor (TRAF) 6 by degrading TRAF3 in osteoclast precursors. *J. Biol. Chem*, 2017. **292**(24): p. 10169-10179.
47. Ke, D., et al., IL-17A regulates the autophagic activity of osteoclast precursors through RANKL-JNK1 signaling during osteoclastogenesis in vitro. *Biochem. Biophys. Res. Commun.*, 2018. **497**(3): p. 890-896.
48. Lee, K., et al., Selective regulation of MAPK signaling mediates RANKL-dependent osteoclast differentiation. *Int. J. Biol. Sci*, 2016. **12**(2): p. 235.
49. Hasegawa, Y., et al., Pulmonary osteoclast-like cells in silica induced pulmonary fibrosis. *bioRxiv*, 2023.
50. Deng, M.C., Multi-dimensional COVID-19 short-and long-term outcome prediction algorithm. *Expert Rev. Precis. Med. Drug Dev.*, 2020. **5**(4): p. 239-242.
51. Wang, S., et al., A single-cell transcriptomic landscape of the lungs of patients with COVID-19. *Nat. Cell Biol*, 2021. **23**(12): p. 1314-1328.
52. Xu, Z., et al., Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet Respir Med*, 2020. **8**(4): p. 420-422.

Figure Legends:

Figure 1. (a) A graphical abstract of the study. (b) Principal component analysis is shown. Each point represents a sample. Colouring indicates groups. Red circles represent the control samples, green triangles represent the patient samples with COVID-19 disease from a year ago and no pneumonia in the follow-up radiological data, and blue squares represent the samples of patient with COVID-19 disease from a year ago and severe pneumonia confirmed by radiological data in the follow-up period. (c) Heatmap of 34 upregulated and 34 downregulated DEGs in all group comparisons.

Figure 2. (a) All differentially expressed genes (DEGs) comparison among “C vs NP”, “C vs SP” and “NP vs SP”, and common 291 DEGs. Functional enrichment analysis (Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), WikiPathways (WP), and the comprehensive resource of mammalian protein complexes (CORUM)) of common 291 DEGs. (b) Upregulated DEGs among all comparisons, and 1436 extended common upregulated genes. Functional enrichment analysis of common 1436 up-DEGs. (c) Downregulated DEGs among all comparisons, and 71 extended common downregulated genes. Functional enrichment analysis of common 71 down-DEGs. The green circle in Venn diagram represents DEGs in “C vs NP” comparison, and blue circle represents DEGs in “C vs NP” comparison, and red circle represents DEGs in “NP vs SP” comparison.

Figure 3. (a) Protein-protein interaction analysis of 291 common genes determined as a result of comparison of DEGs in all comparable groups. (b) Gene-Diseases (DisGeNET) and Gene Transcription Factor interactions (TRRUST) of these genes. (c) PPI analysis of hub genes from 291 genes. (d) Gene-Diseases (DisGeNET), Gene Transcription Factor interactions (TRRUST), and gene-transcription factor targets (GSEA) of these hub genes. The graph shows the diseases, transcription factors and their targets with which DEGs are highly correlated in these intergroup comparisons. The vertical axis shows the disease names obtained from the DisGenet database, and the horizontal axis shows the $-\log_{10}(P)$ statistical significance level.

Figure 4. Immune cell signatures of (a) the extended common upregulated DEGs, (b) the extended common downregulated DEGs, (c) the 291 common DEGs, and (d) the 14 hub genes. The graph shows the immune cell signatures with which DEGs are highly correlated in these intergroup comparisons. The vertical axis shows the immune cell names obtained from the databases, and the horizontal axis shows the $-\log_{10}(P)$ statistical significance level.

Figure 5. Gene expression analysis based on RNA-seq count data of 14 hub genes. All target genes' counts were normalized the counts of the individual GAPDH. Blue and circles shows control individuals, Orange and squares shows the individuals with no pneumonia, and red and triangles shows the individuals with severe pneumonia. Statistical significance was shown as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

Tables:

Table 1. Demographics of study subjects.

Groups	Healthy controls Before/after RNA quality control	No pneumonia Before/after RNA quality control	Severe pneumonia Before/after RNA quality control
	(n=13)/(n=13)	(n=13)/(n=11)	(n=14)/(n=10)
Male*	8 (46.25±7.85; 41-61)/ 8 (46.25±7.85; 41-61)	8 (49.63±16.20; 25-75)/ 7 (46±13.55; 25-64)	10 (57.5±12.41; 37-77)/ 7 (58.14±11.80; 44-77)
Female *	5 (47.8±6.10; 42-58)/ 5 (47.8±6.10; 42-58)	5 (61±8.06; 53-72)/ 4 (62.75±8.14; 53-72)	4 (51.5+10.25; 40-62) / 3 (53.33±11.72; 40-62)

*Age data were demonstrated as Mean ± SD; Minimum-Maximum

Table 2. Following the comparative analysis of all up-regulated and down-regulated genes across the control, no pneumonia, and severe pneumonia groups, 36 hub genes were identified from a total of 291 common DEGs. Subsequently, a selection of 23 central genes, each interacting with three or more other genes, was made, along with an investigation into the genes, which they interact with.

Gene Symbol	Number of genes interacted	Names of genes interacted
ICAM1	6	AREG, CSF1, IL1R1, IL1R2, PRKCD, TUBB4B
MARCKS	6	F3, FSCN1, PRKCD, RAB11FIP1, TUBB4B, ULK1
TUBB4B	6	F3, GABARAPL1, ICAM1, MARCKS, SNAI1, ULK1
HDAC5	5	ATF3, DDIT3, NFKB2, NFKBIA, NFKBIE
NFKBIE	5	HDAC5, NFKB2, NFKBIA, NFKBIB, RELB
NFKB2	5	HDAC5, NFKBIA, NFKBIB, NFKBIE, RELB
NFKBIA	5	HDAC5, NFKBIB, NFKBIB2, NFKBIE, RELB
ATF3	4	DDIT3, FOSL1, HDAC5, JUND
F3	4	ICAM1, MARCKS, PRKCD, TUBB4B
DDIT3	4	ATF3, FOSL1, HDAC5, JUND
IL1R1	4	CSF1, ICAM1, IL1R2, TOM1
PRKCD	4	F3, FSCN1, ICAM1, MARCKS
AREG	3	ERG, HBEGF, ICAM1
CSF1	3	ICAM1, IL1R1, IL1R2
FOSL1	3	ATF3, DDIT3, JUND
FSCN1	3	MARCKS, MCL1, PRKCD
IL1R2	3	CSF1, ICAM1, IL1R1

JUND	3	ATF3, DDIT3, FOSL1
NFKBIB	3	NFKB2, NFKBIA, NFKBIE
RAB11FIP1	3	IFITM3, MARCKS, TOM1
RELB	3	NFKB2, NFKBIA, NFKBIE
TOM1	3	IL1R1, IFITM3, RAB11FIP1
ULK1	3	GABARAPL1, MARCKS, TUBB4B
ERG	2	AREG HBEGF
GABARAPL1	2	TUBB4B, ULK1
HBEGF	2	AREG, ERG
IFITM3	2	RAB11FIP1, TOM1
MCL1	2	FSCN1, USP2
SNAI1	2	TUBB4B, USP2
USP2	2	MCL1, SNAI1
GADD45B	2	GADD45G, MAP2K3
GADD45G	2	GADD45B, MAP2K3
MAP2K3	2	GADD45B, GADD45G
CSNK1D	2	HSPA5, NOP2
HSPA5	2	CSNK1D, NOP2
NOP2	2	CSNK1D, HSPA5

Supplementary Data:

Supplementary Figure Legends:

Figure S1. (a) KEGG and GO terms of upregulated and downregulated DEGs between Healthy control and No pneumonia groups. (b) KEGG and GO terms of upregulated and downregulated DEGs between Healthy control and Severe pneumonia groups. (c) KEGG and GO terms of upregulated and downregulated DEGs between No pneumonia and Severe pneumonia groups. The top 10 functional terms are sorted from the highest statistical significance score in light color to the lowest score in dark color. (d). Evaluation of 52 validated PBMC genes for disease progression risk assessment in patients with IPF and COVID-19 in post-COVID-19 patients

Figure S2. (a) All DElncRNAs comparison among “C vs NP”, “C vs SP” and “NP vs SP”, and common 70 DElncRNAs. Functional enrichment analysis (KEGG) of common 70 DElncRNAs. (b) Upregulated DEGs among all comparisons, and 457 extended common upregulated lncRNAs. Functional enrichment analysis of common 457 upregulated lncRNAs. (c) Downregulated DEGs among all comparisons, and 79 extended common downregulated lncRNAs. Functional enrichment analysis of common 79 downregulated lncRNAs. The green circle in Venn diagram represents DElncRNAs in “C vs NP” comparison, and blue circle represents DElncRNAs in “C vs NP” comparison, and red circle represents DElncRNAs in “NP vs SP” comparison.

Supplementary Tables:

Table S1: Top 10 functional enrichment (KEGG) analysis of upregulated, downregulated, common, and hub genes. Data were obtained from NCPATH database.

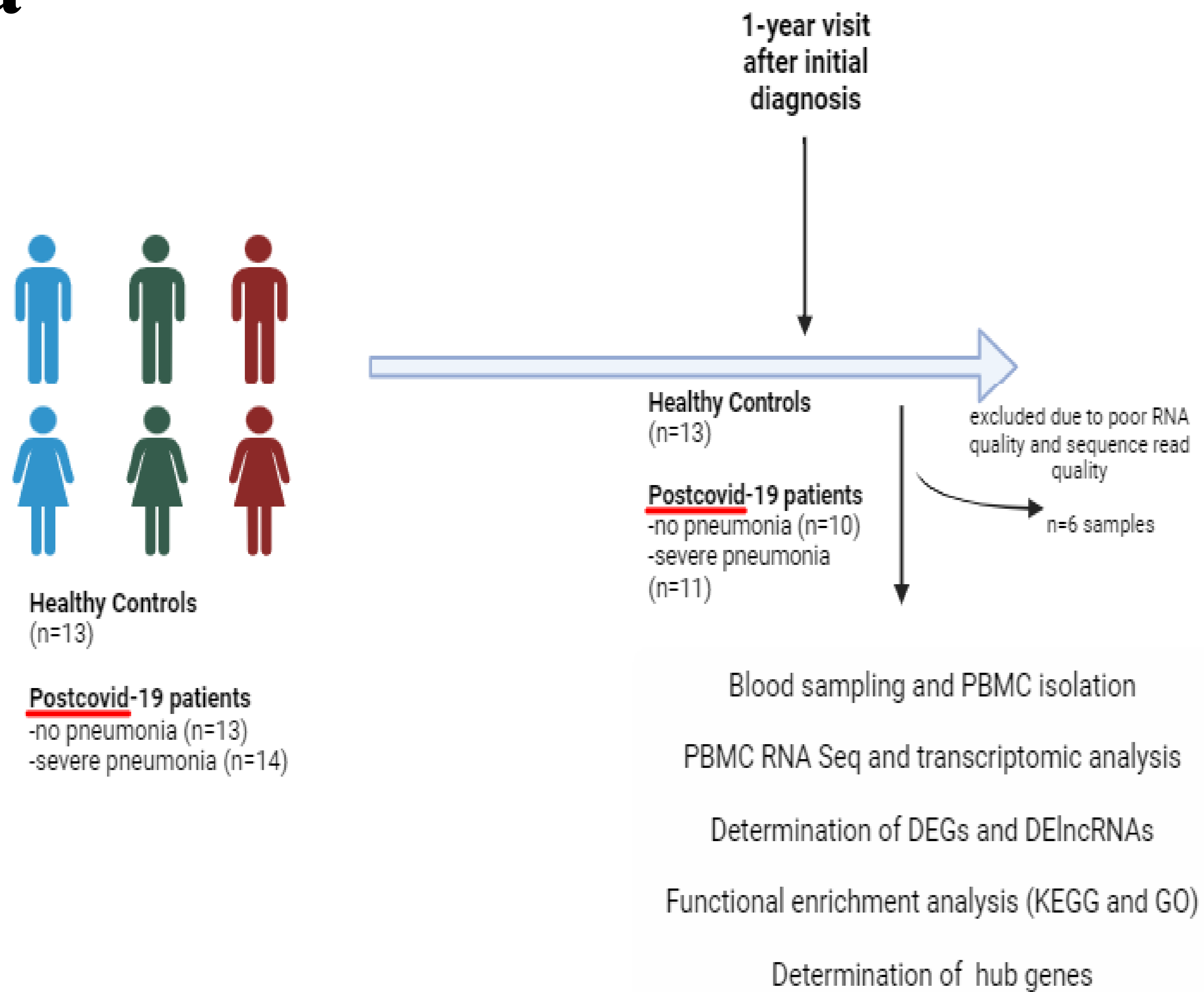
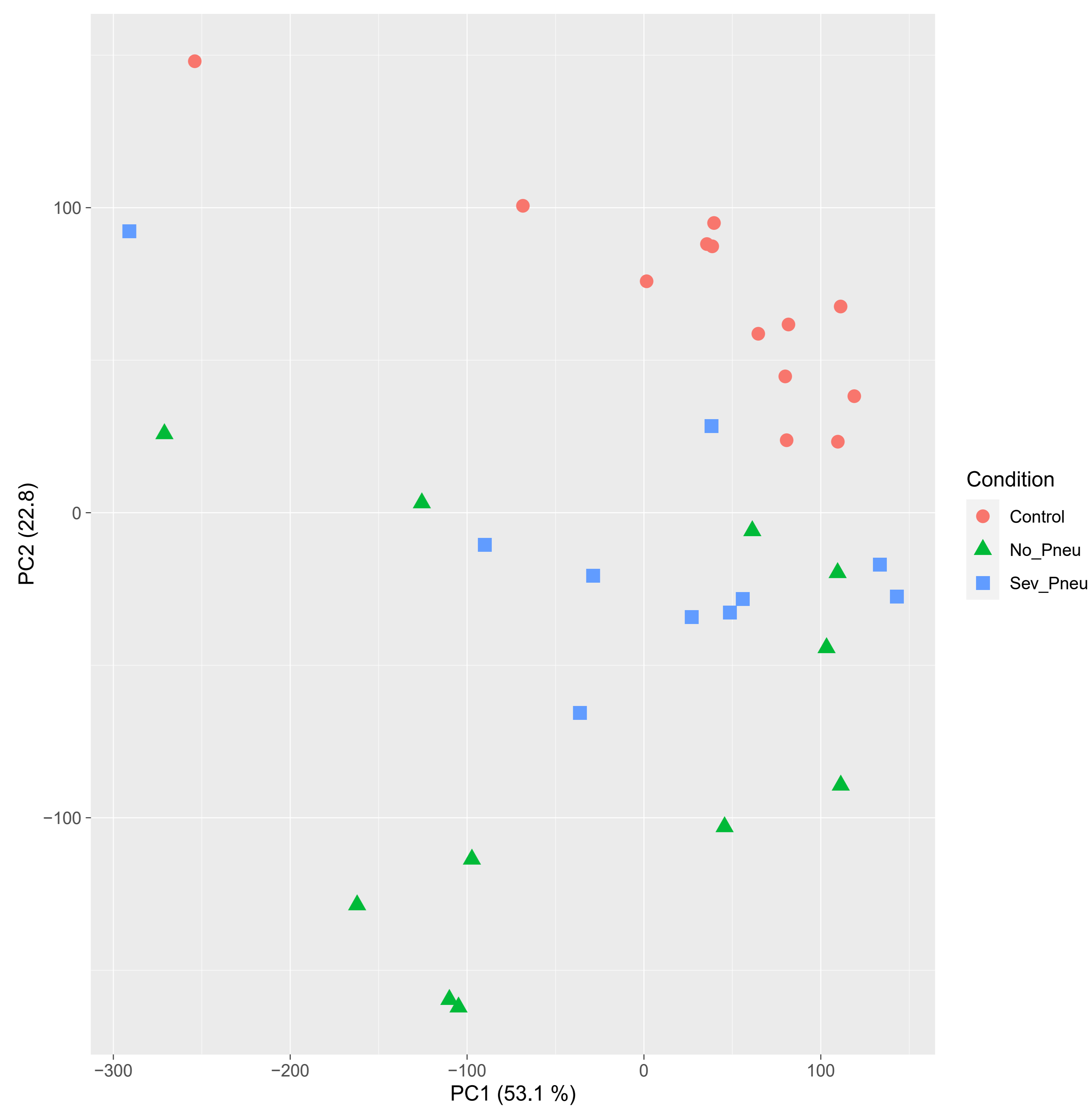
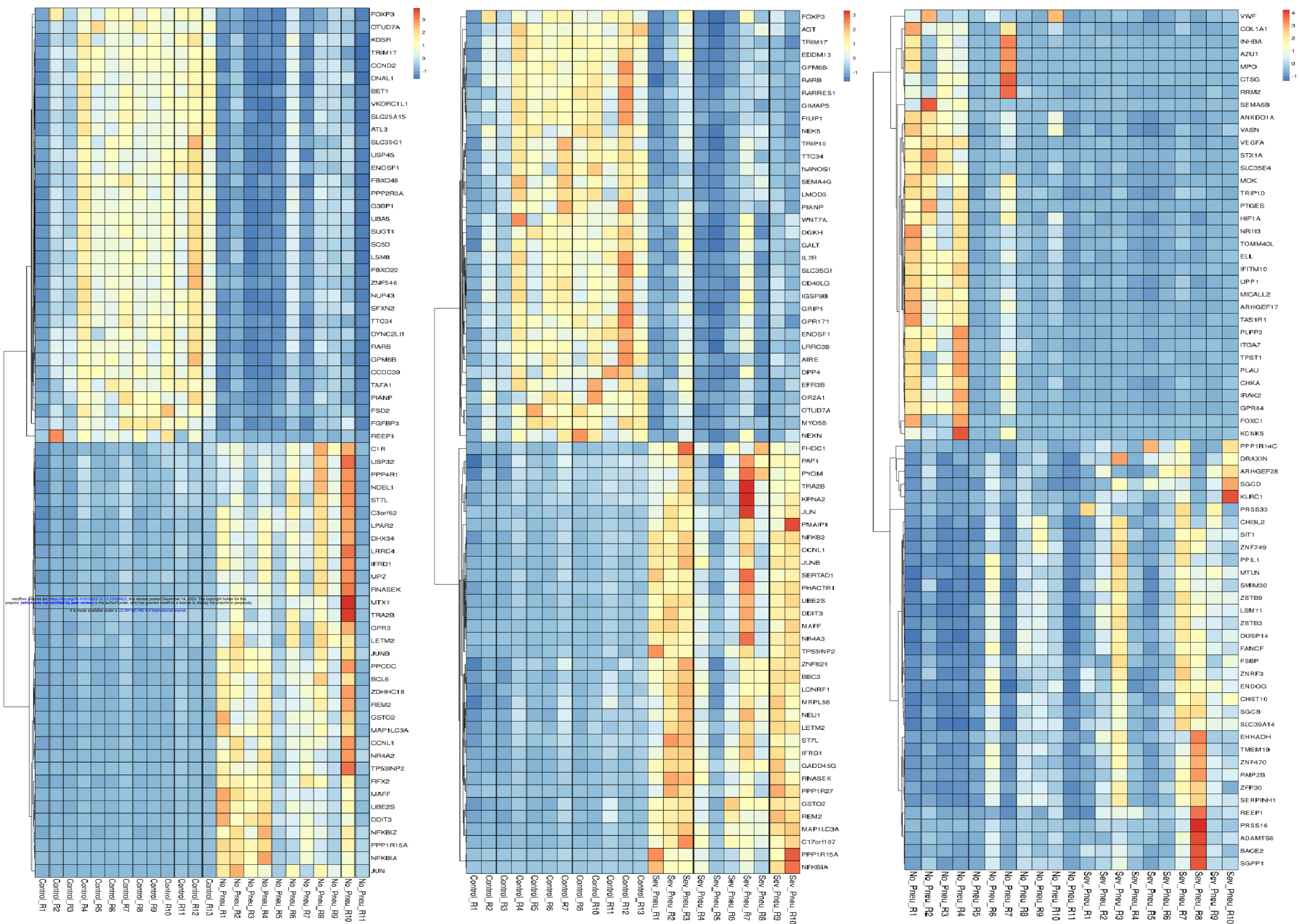
Common 70 DElncRNAs-KEGG Functional Enrichment Analysis				
hsaid	pathway	p-adjust	p-value	genes
hsa04010	MAPK signaling pathway	9.53933919587235e-17	1.1776961970212778e-18	223
hsa04015	Rap1 signaling pathway	2.444343597210432e-12	9.053124434112712e-14	165
hsa05215	Prostate cancer	2.444343597210432e-12	8.804392672203224e-14	84
hsa04152	AMPK signaling pathway	2.6608061331489008e-12	1.3139783373574818e-13	101
hsa04722	Neurotrophin signaling pathway	9.495107472233918e-12	5.861177451996246e-13	98
hsa04520	Adherens junction	1.0647390579295397e-11	7.886955984663257e-13	65
hsa05165	Human papillomavirus infection	1.7895532145452728e-11	1.5465274693601124e-12	242
hsa04218	Cellular senescence	2.324204209971124e-11	2.295510330835678e-12	124
hsa01522	Endocrine resistance	2.6330695483170815e-11	2.925632831463424e-12	82
hsa04510	Focal adhesion	2.6692401663290662e-11	3.2953582300358844e-12	151
Upregulated common 457 DElncRNAs-KEGG Functional Enrichment Analysis				
hsaid	pathway	p-adjust	p-value	genes
hsa04010	MAPK signaling pathway	3.805507722682358e-15	5.0740102969098097e-17	250
hsa04510	Focal adhesion	2.892857382144406e-11	7.714286352385083e-12	173

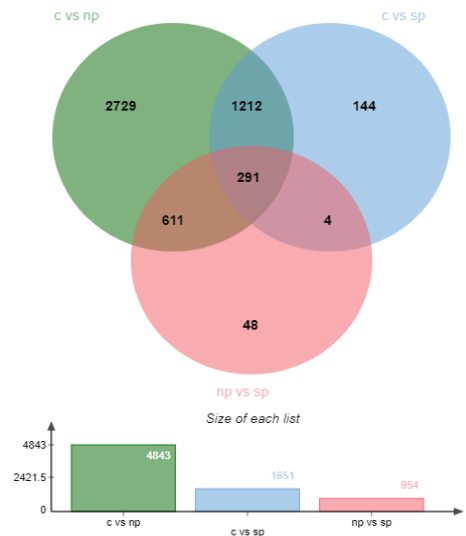
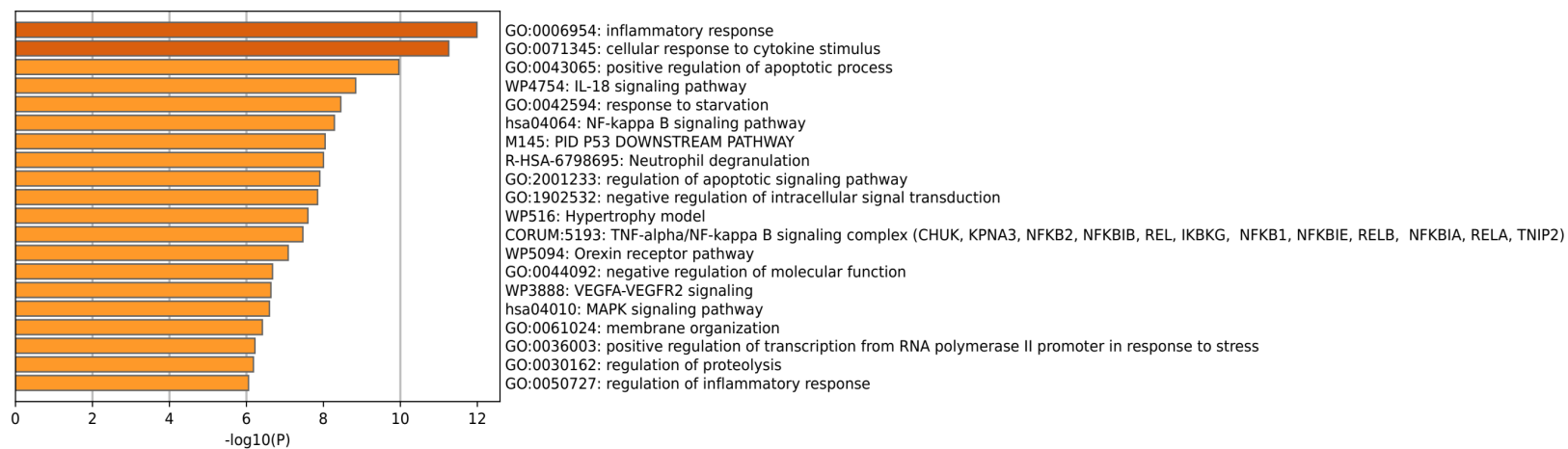
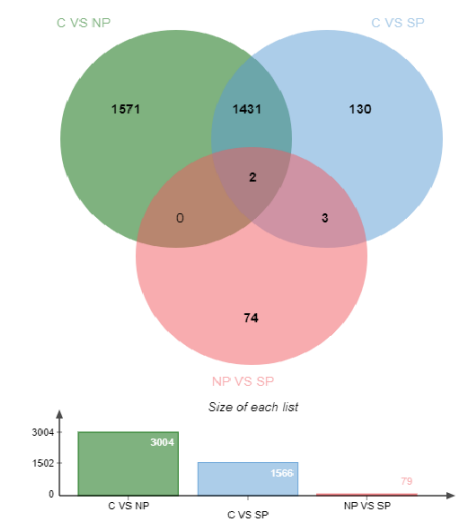
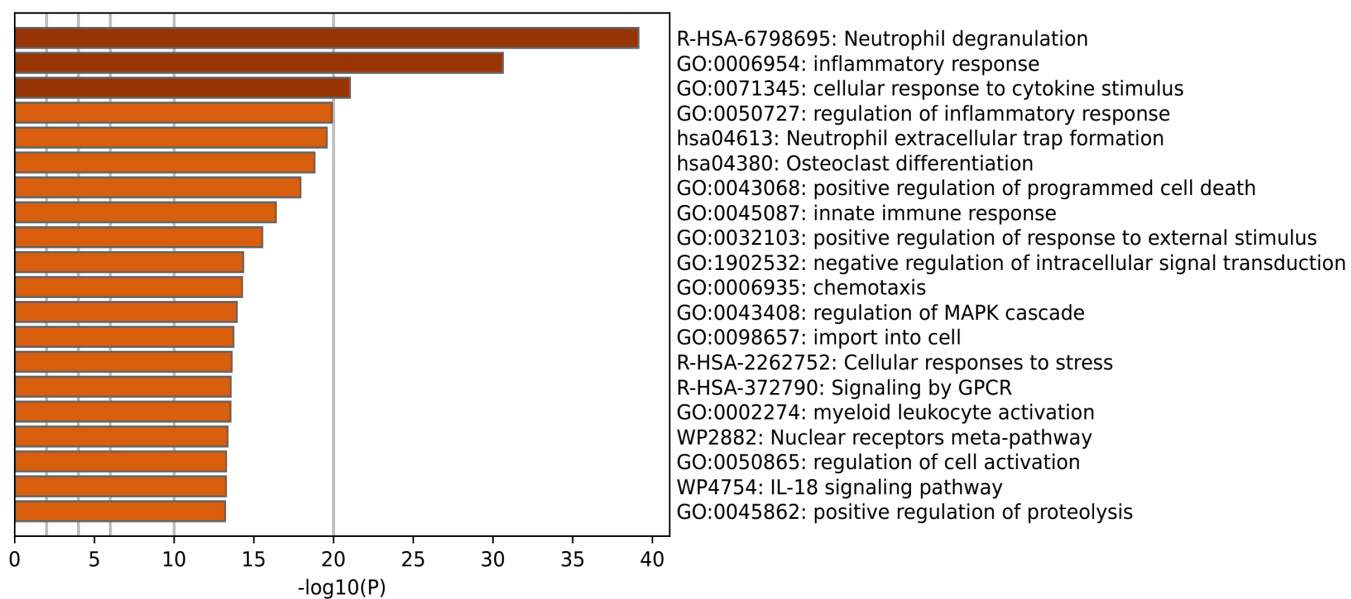
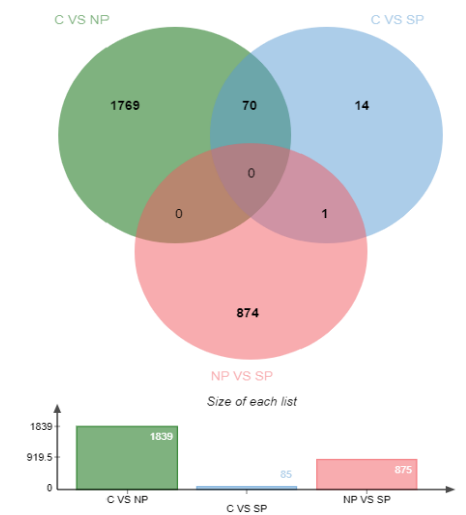
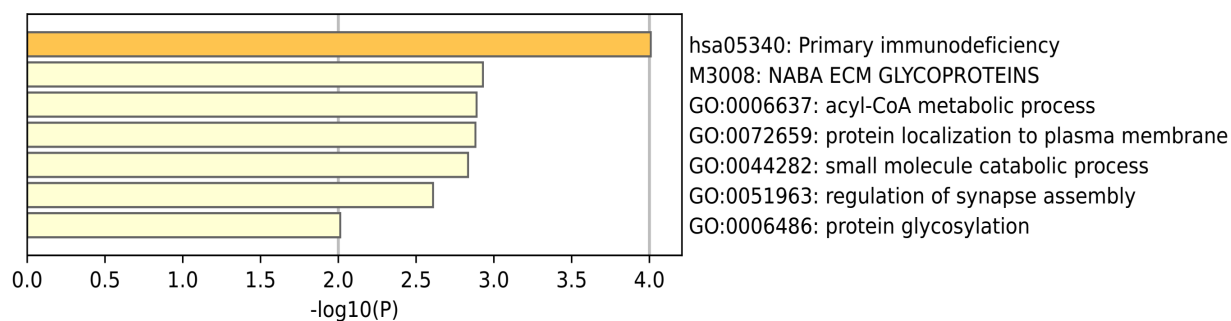
0		-11	13	
hsa0411	Cell cycle	1.5296857804112035e-10	6.118743121644814e-12	113
hsa0493	Insulin resistance	3.8773516016758865e-10	2.0679208542271393e-11	99
hsa0415	AMPK signaling pathway	3.937591631844922e-10	2.6250610878966148e-11	109
hsa0521	Prostate cancer	6.763741254546563e-10	5.41099300363725e-11	89
hsa0472	Neurotrophin signaling pathway	8.295350701461146e-10	7.742327321363737e-11	106
hsa0401	Rap1 signaling pathway	1.3129177372101546e-09	1.4004455863574982e-10	181
hsa0415	mTOR signaling pathway	1.5725927540822539e-09	1.8871113048987045e-10	132
hsa0421	Cellular senescence	2.2339661496262503e-09	3.490933500542371e-10	136
Downregulated common 79 DElncRNAs-KEGG Functional Enrichment Analysis				
hsaid	pathway	p-adjust	p-value	genes
hsa0152	Platinum drug resistance	0.00024072411175947426	1.2269490024761625e-05	35
hsa0471	Thermogenesis	0.00024072411175947426	1.4589340106634803e-05	88
hsa0415	mTOR signaling pathway	0.0006114137926066184	6.154770701913108e-05	59
hsa0522	Hepatocellular carcinoma	0.0006114137926066184	7.411076274019616e-05	63
hsa0521	Glioma	0.0008561020059955173	0.00012971242515083596	33
hsa0522	Chronic myeloid leukemia	0.00097164552531481	0.00017666282278451092	33

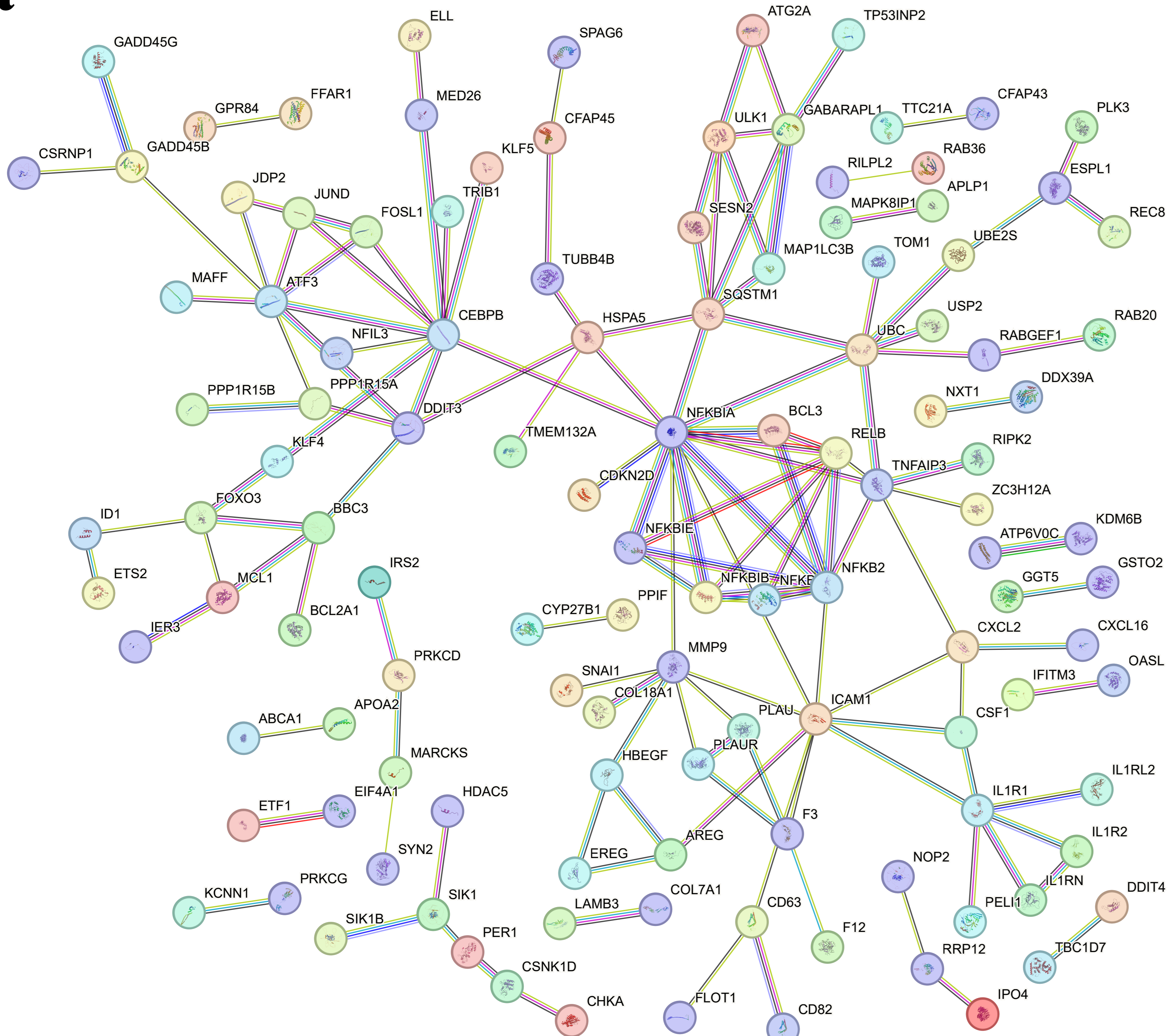
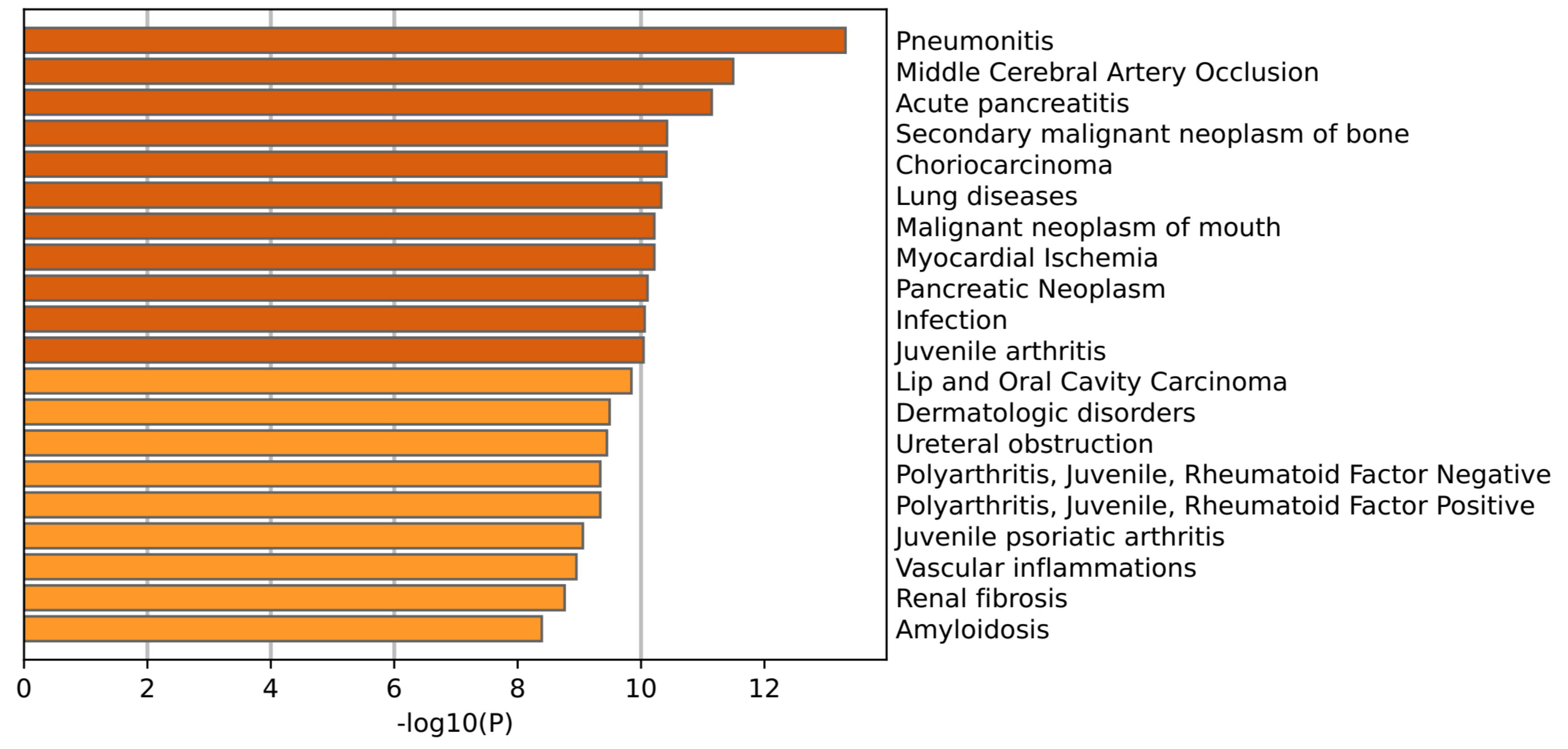
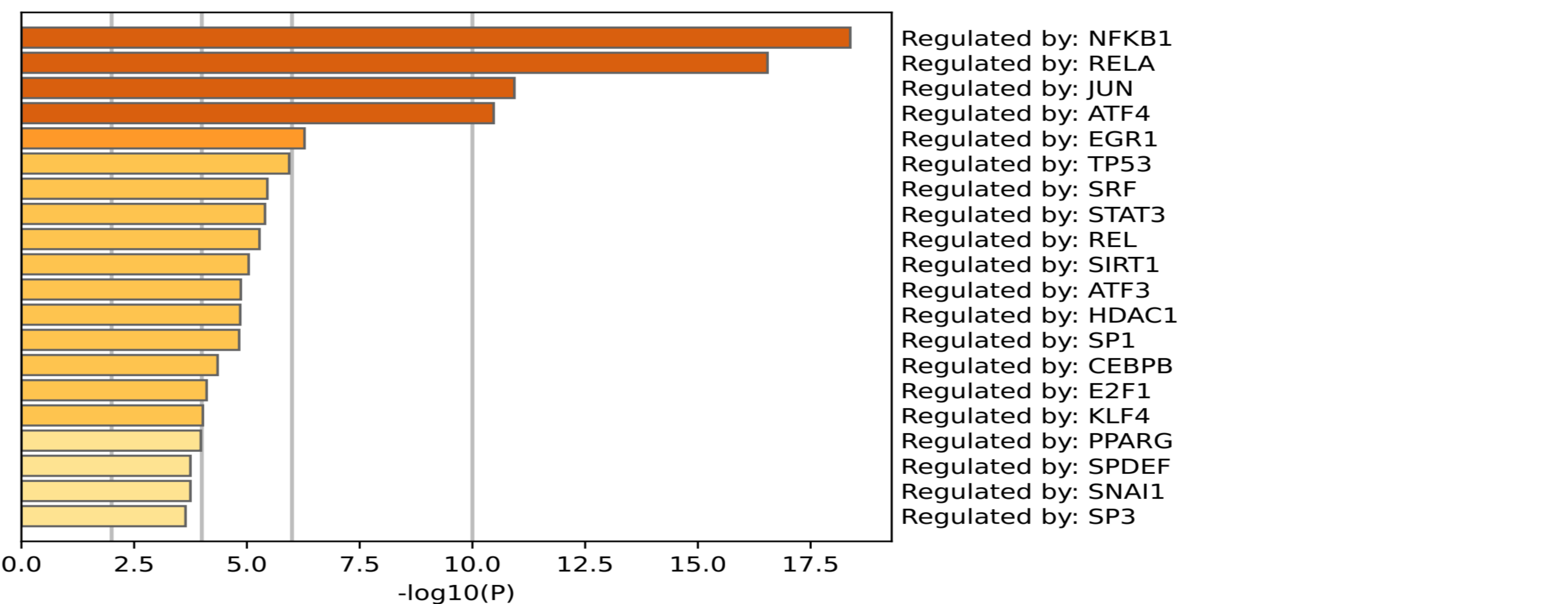
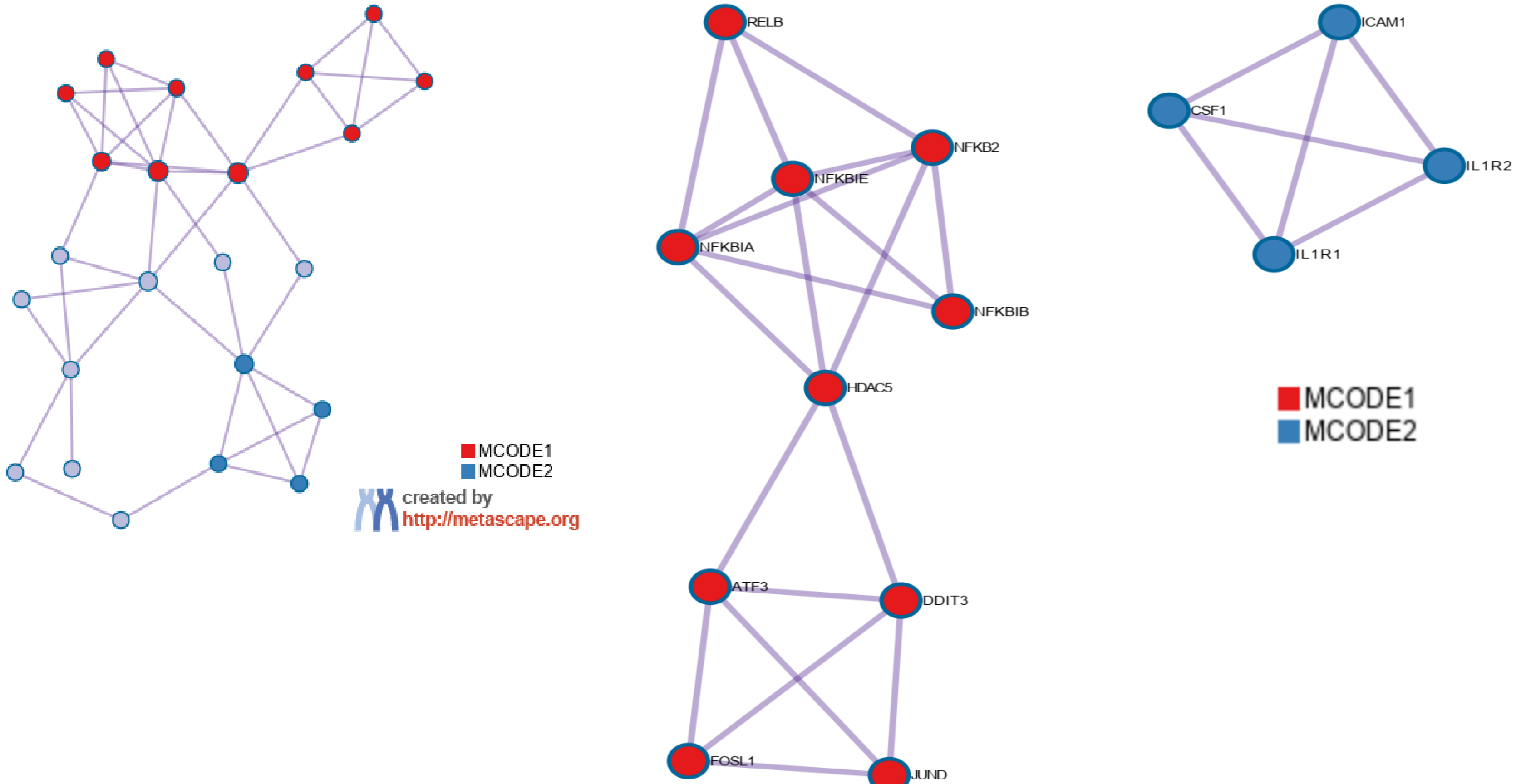
hsa0521 2	Pancreatic cancer	0.0018137416627057 09	0.0004327559007987 3394	32
hsa0522 2	Small cell lung cancer	0.0018137416627057 09	0.0004543111882996 7885	38
hsa0421 1	Longevity regulating pathway	0.0018137416627057 09	0.0004946568171015 57	36
hsa0152 1	EGFR tyrosine kinase inhibitor resistance	0.0031958699234137 06	0.0009684454313374 866	32

Common 23 DEGs and 70 DElncRNAs-KEGG Functional Enrichment Analysis

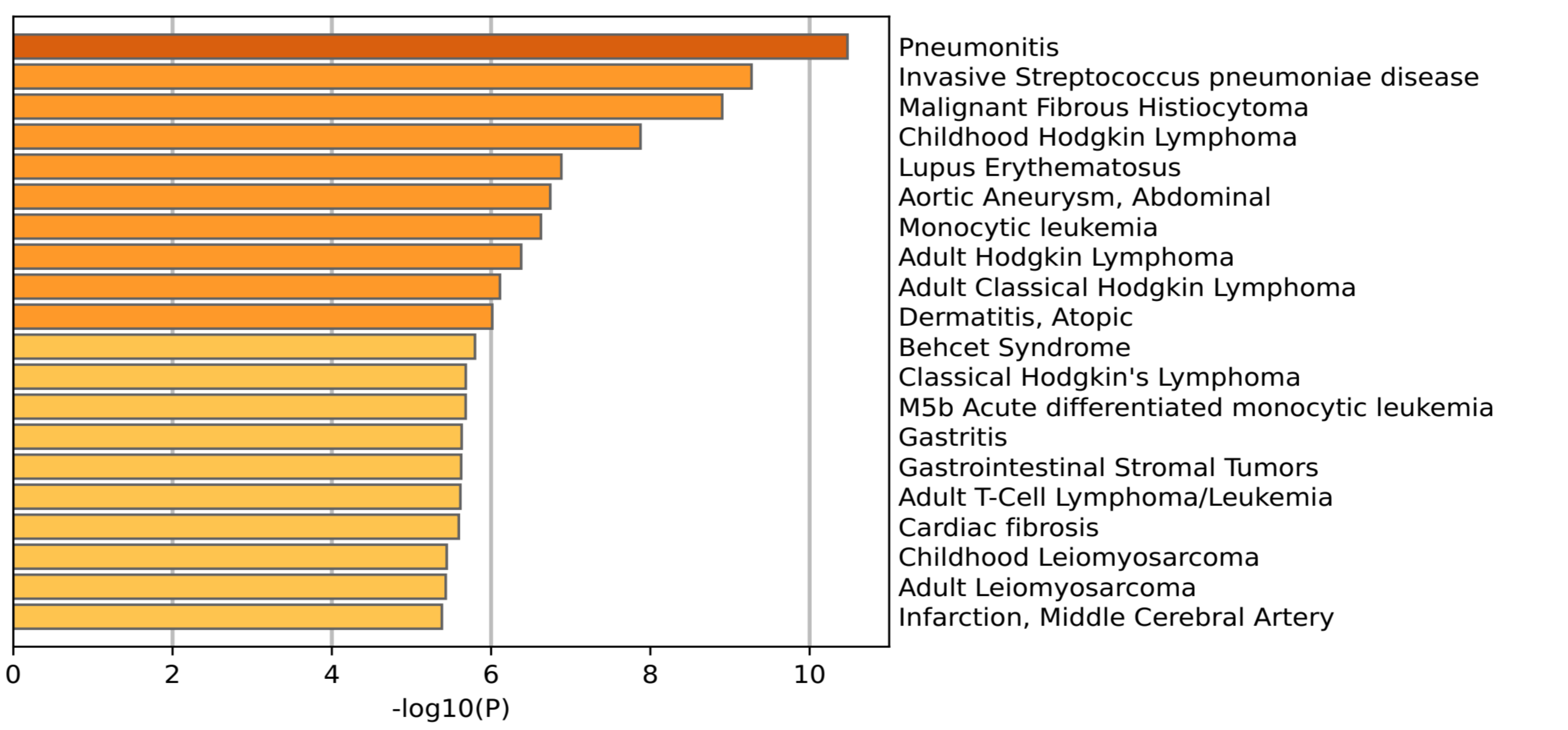
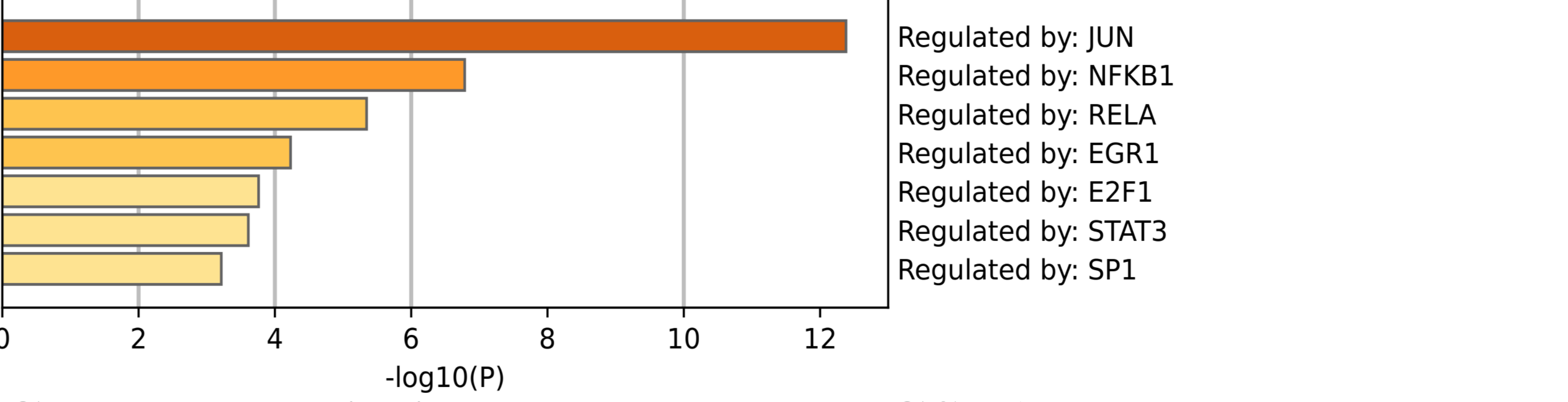
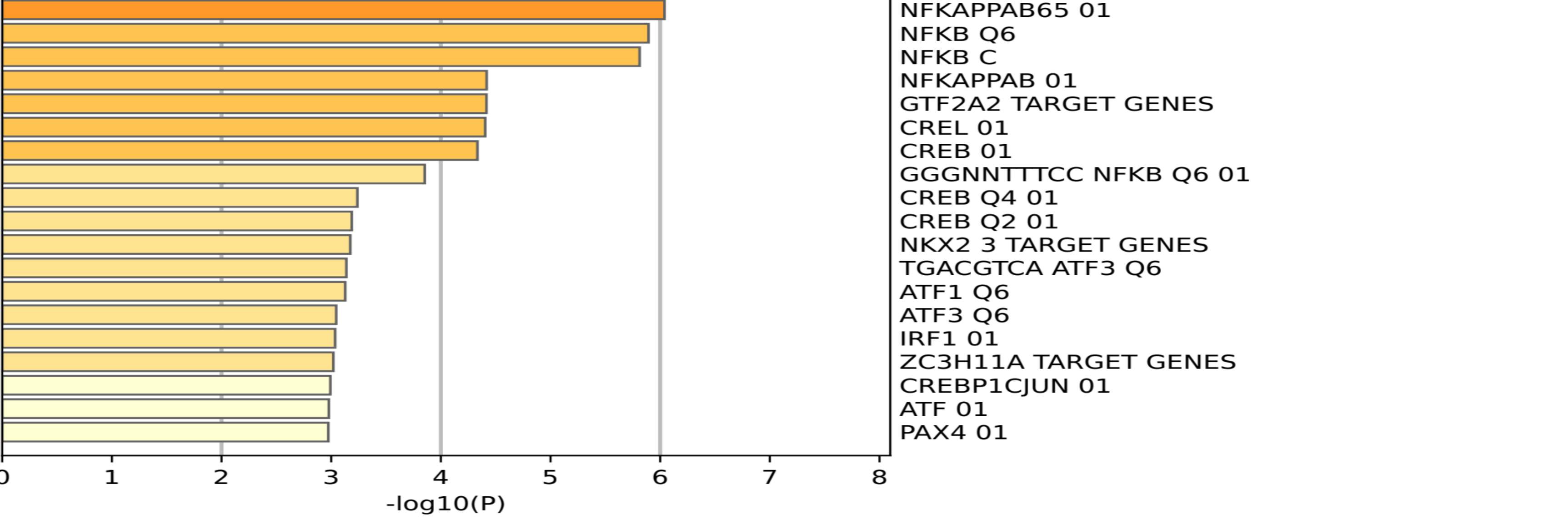
hsaid	pathway	p-adjust	p-value	Gene s
hsa0401 0	MAPK signaling pathway	1.22923251613022e- 17	1.499064044061244e- 19	225
hsa0521 5	Prostate cancer	8.612608644679657e- 14	2.1006362547999165e- 15	86
hsa0472 2	Neurotrophin signaling pathway	7.647697967280332e- 13	2.797938280712317e- 14	100
hsa0401 5	Rap1 signaling pathway	2.1920773103682596e- 12	1.0693060050576875e- 13	165
hsa0415 2	AMPK signaling pathway	2.4298777304097346e- 12	1.4816327624449603e- 13	101
hsa0522 0	Chronic myeloid leukemia	9.72383680842217e- 12	7.11500254274793e- 13	68
hsa0452 0	Adherens junction	1.0075525944866993e- 11	8.601058733423044e- 13	65
hsa0516 5	Human papillomavirus infection	1.9429917103537356e- 11	1.8956016686377907e- 12	242
hsa0421 8	Cellular senescence	2.3876880558830318e- 11	2.6206332320667423e- 12	124
hsa0152 2	Endocrine resistance	2.6522564008199485e- 11	3.2344590253901814e- 12	82

a**b****c**

a**All DEGs****Functional Enrichment****b****Up DEGs****Functional Enrichment****c****Down DEGs****Functional Enrichment**

a**b****Gene-Diseases Interaction-DisGeNET****Gene-Transcription Factor Interaction-TRRUST****c**

GO	Description	Log10(P)	Color	MCODE	GO	Description	Log10(P)
CORUM:5193	TNF-alpha/NF-kappa B signaling complex (CHUK, KPNA3, NFKB2, NFKB1B, REL, IKKKG, NFKB1, NFKB1A, NFKB1E, NFKB1F, NFKB1G, NFKB1H, NFKB1I, NFKB1J, NFKB1K, NFKB1L, NFKB1M, NFKB1N, NFKB1O, NFKB1P, NFKB1Q, NFKB1R, NFKB1S, NFKB1T, NFKB1U, NFKB1V, NFKB1W, NFKB1X, NFKB1Y, NFKB1Z)	-13.1	Red	MCODE_1	CORUM:5193	TNF-alpha/NF-kappa B signaling complex (CHUK, KPNA3, NFKB2, NFKB1B, REL, IKKKG, NFKB1, NFKB1A, NFKB1E, NFKB1F, NFKB1G, NFKB1H, NFKB1I, NFKB1J, NFKB1K, NFKB1L, NFKB1M, NFKB1N, NFKB1O, NFKB1P, NFKB1Q, NFKB1R, NFKB1S, NFKB1T, NFKB1U, NFKB1V, NFKB1W, NFKB1X, NFKB1Y, NFKB1Z)	-15.3
hsa04380	Osteoclast differentiation	-11.2	Red	MCODE_1			
R-HSA-1280215	Cytokine Signaling in Immune system	-10.3	Red	MCODE_1	CORUM:5232	TNF-alpha/NF-kappa B signaling complex (RPL6, RPL30, RPS13, CHUK, DDX3X, NFKB2, NFKB1B, REL, IKKKG, NFKB1, MAP3K8, RELB, GLG1, NFKB1A, REL, TNIP2, GTF2I)	-10.8
			Red	MCODE_1	M12705	SIG CD40PATHWAYMAP	-9.6
			Blue	MCODE_2	R-HSA-6783783	Interleukin-10 signaling	-11.3
			Blue	MCODE_2	R-HSA-449147	Signaling by Interleukins	-7.2
			Blue	MCODE_2	hsa04640	Hematopoietic cell lineage	-6.9

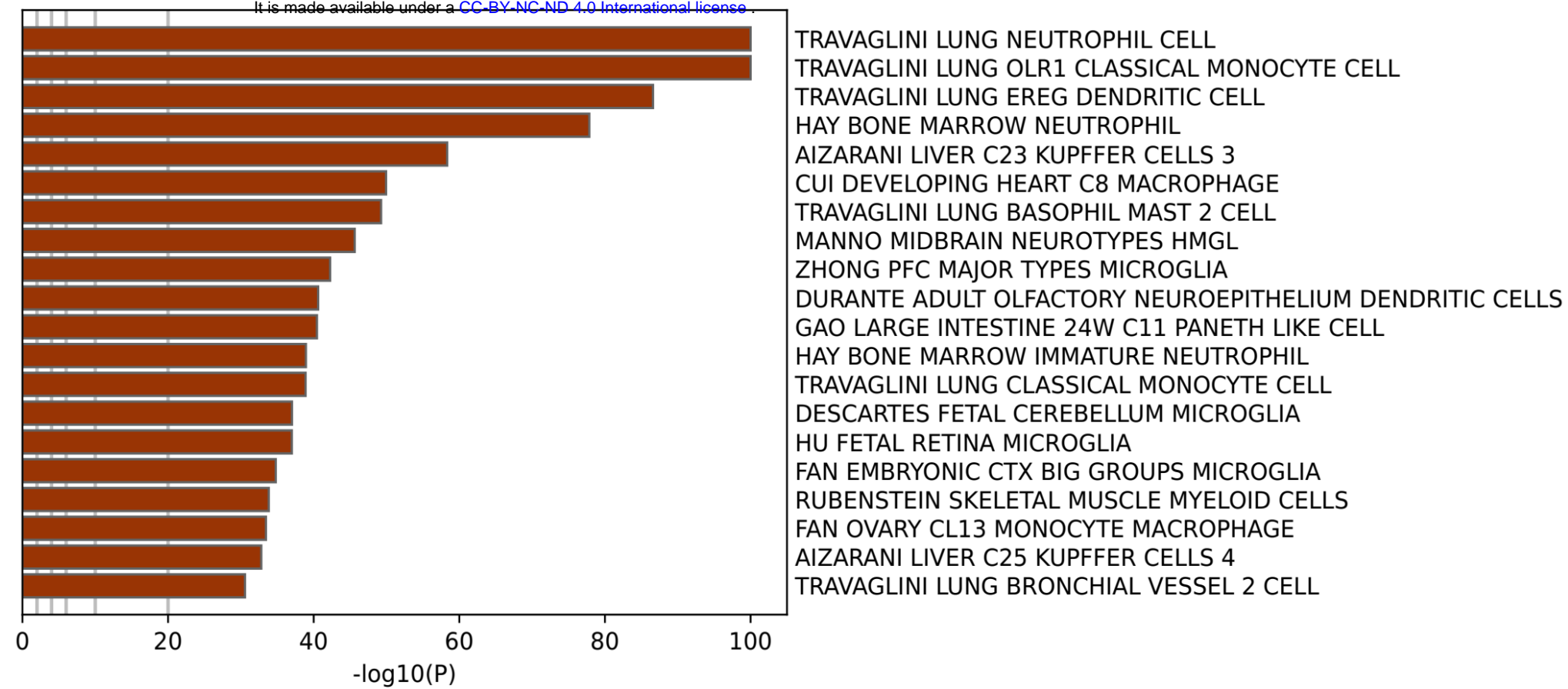
d**Gene-Diseases Interaction-DisGeNET****Gene-Transcription Factor Interaction-TRRUST****Gene-Transcription Factor Targets-GSEA**

Immun cell type signatures-GSEA

a

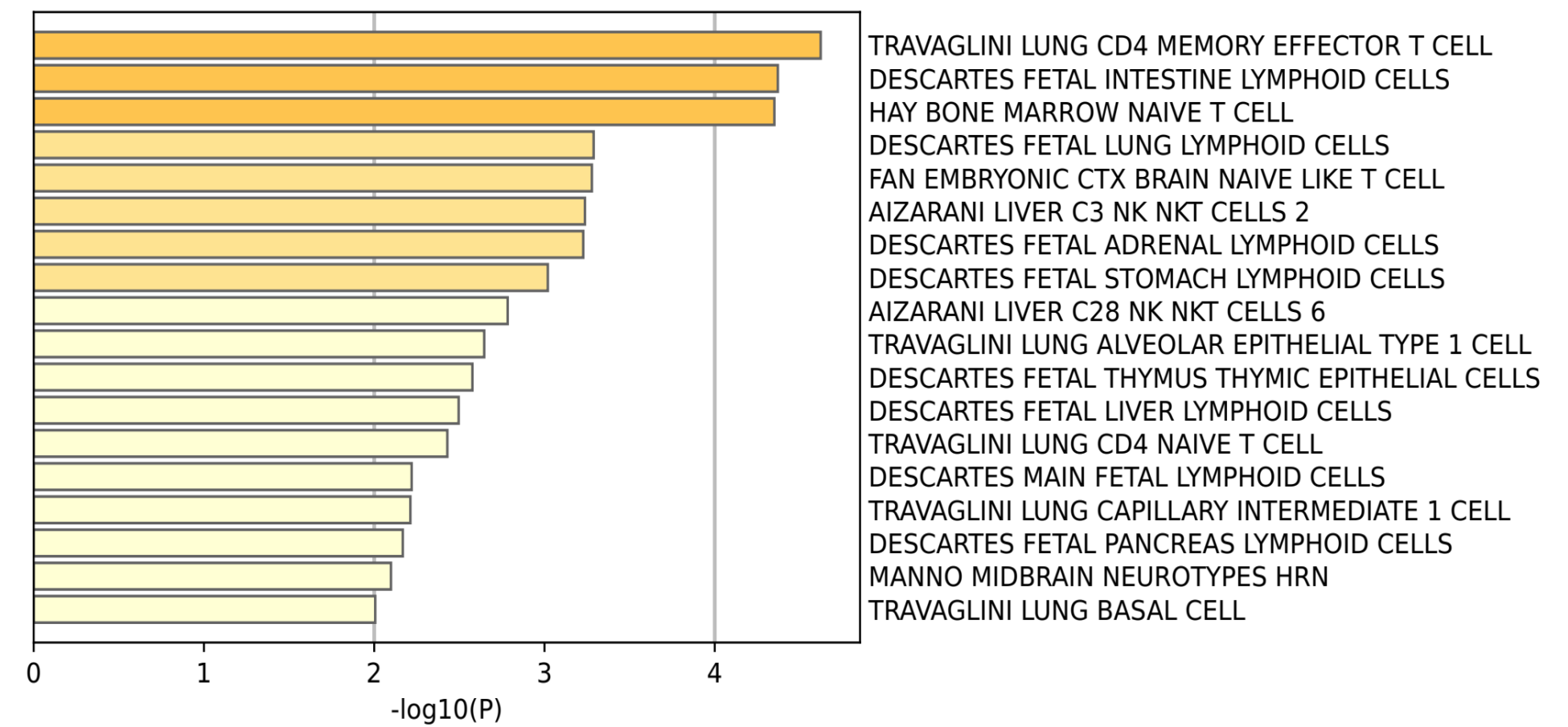
medRxiv preprint doi: <https://doi.org/10.1101/2023.12.12.23299222>; this version posted December 14, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Up DEGs



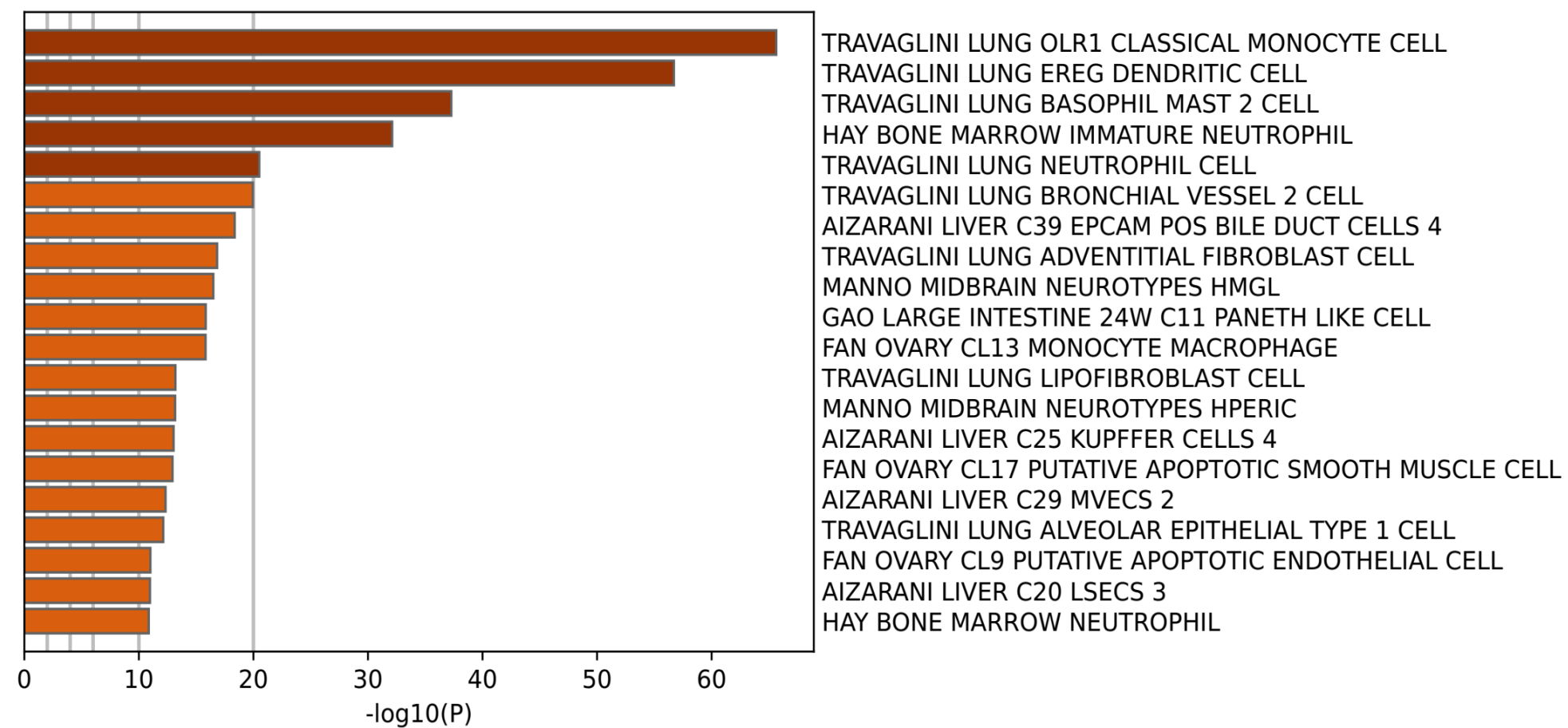
b

Down DEGs



c

All DEGs-291 Common Genes



d

Central Genes (23)

