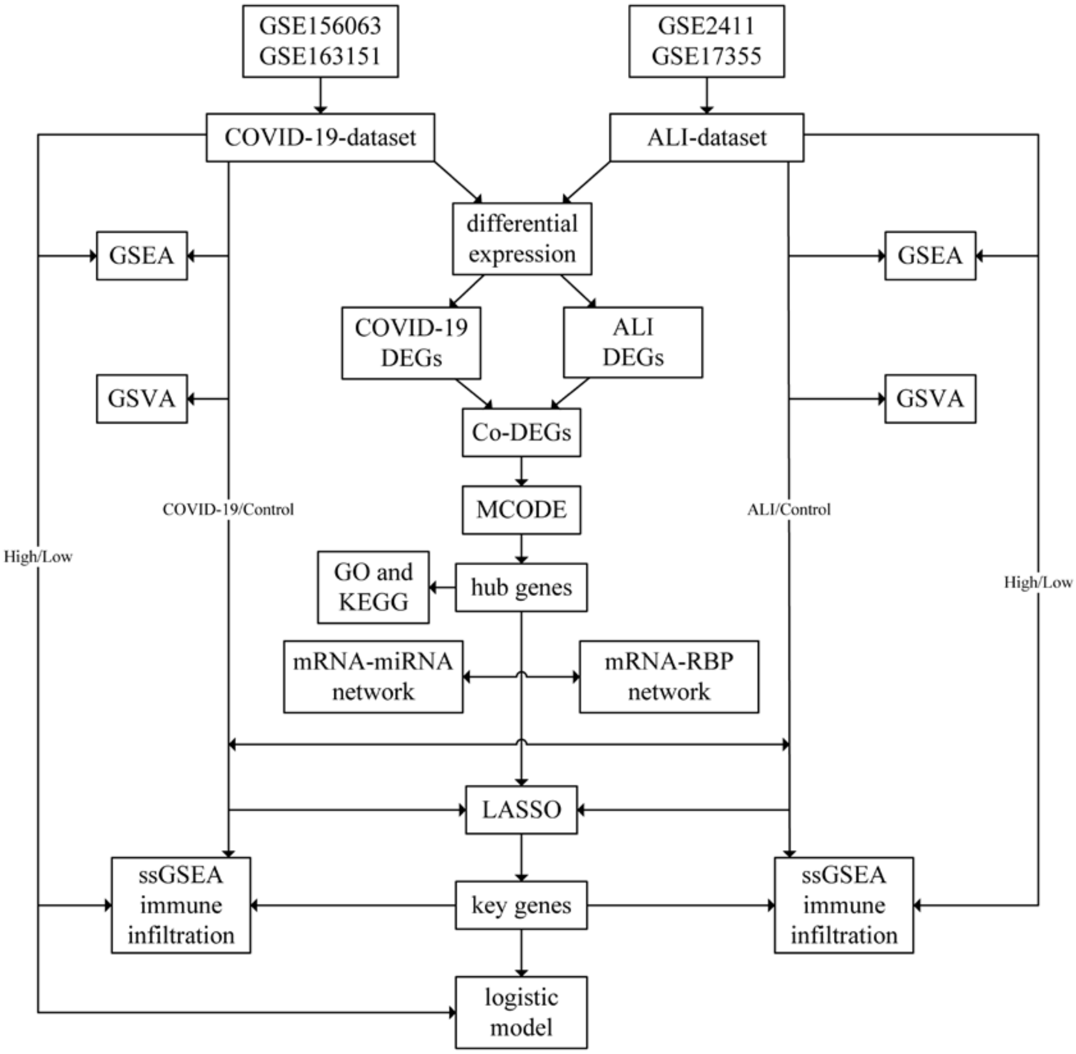
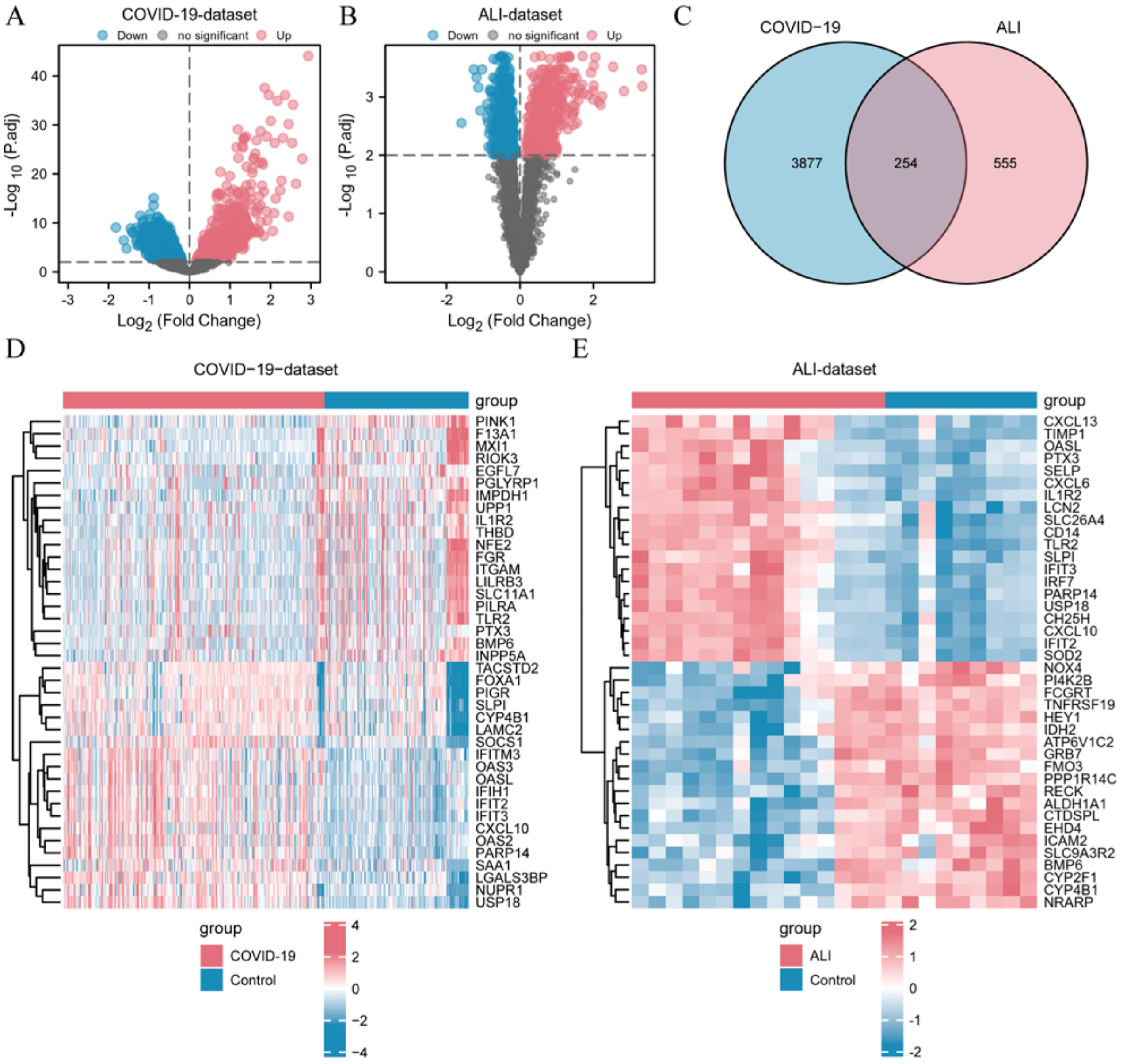


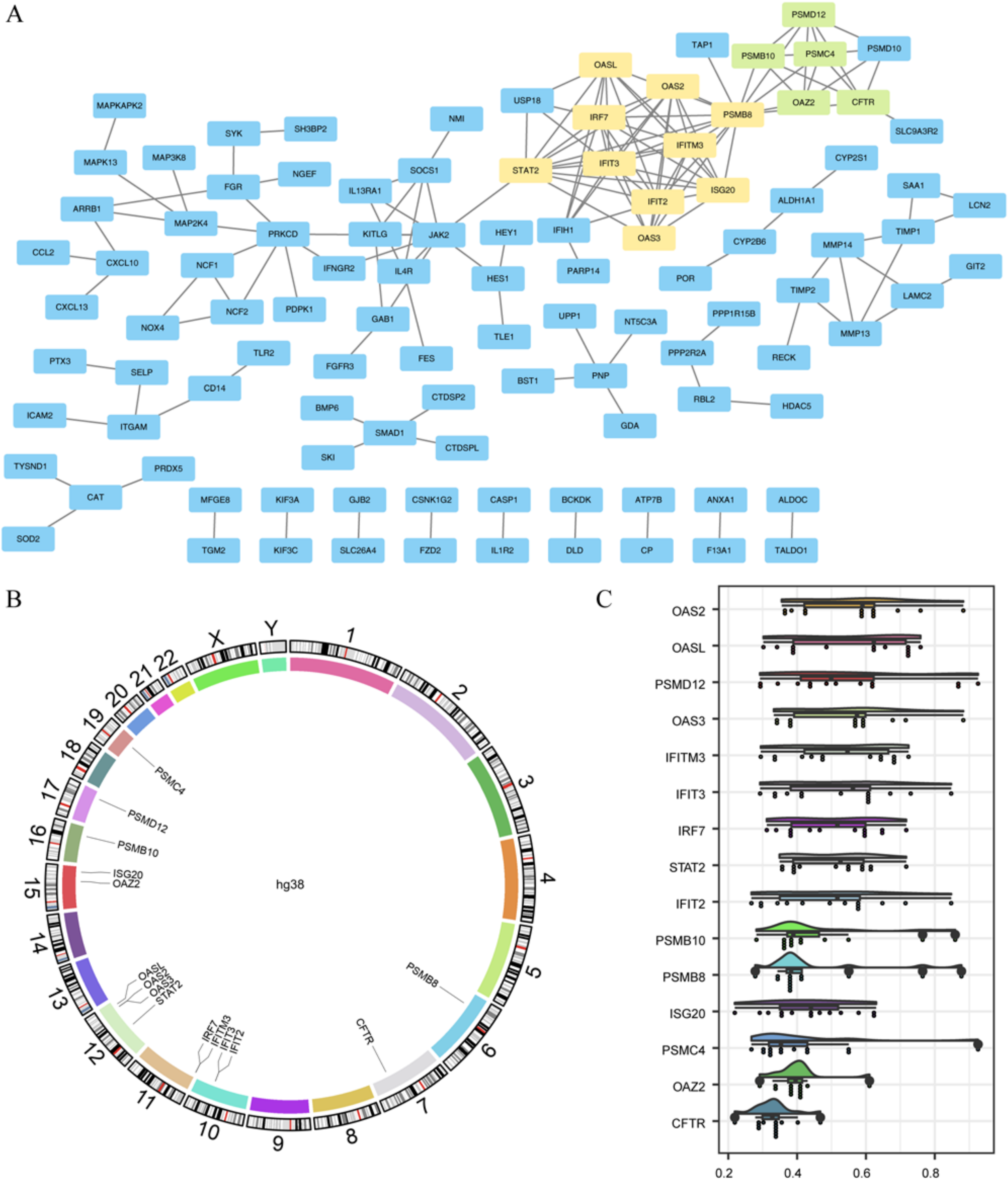
Supplementary Figure S1. The workflow chart

COVID-19 – Coronavirus disease 2019, ALI – acute lung injury, DEGs – differential expression genes; Co-DEGs – common DEGs, GSEA – gene set enrichment analysis, GSVA – gene set variation analysis, GO – gene ontology, KEGG – Kyoto encyclopedia of genes and genomes, MCODE – molecular complex detection, RBP – RNA binding protein, LASSO – least absolute shrinkage and selection operator, ssGSEA – single sample gene set enrichment analysis.



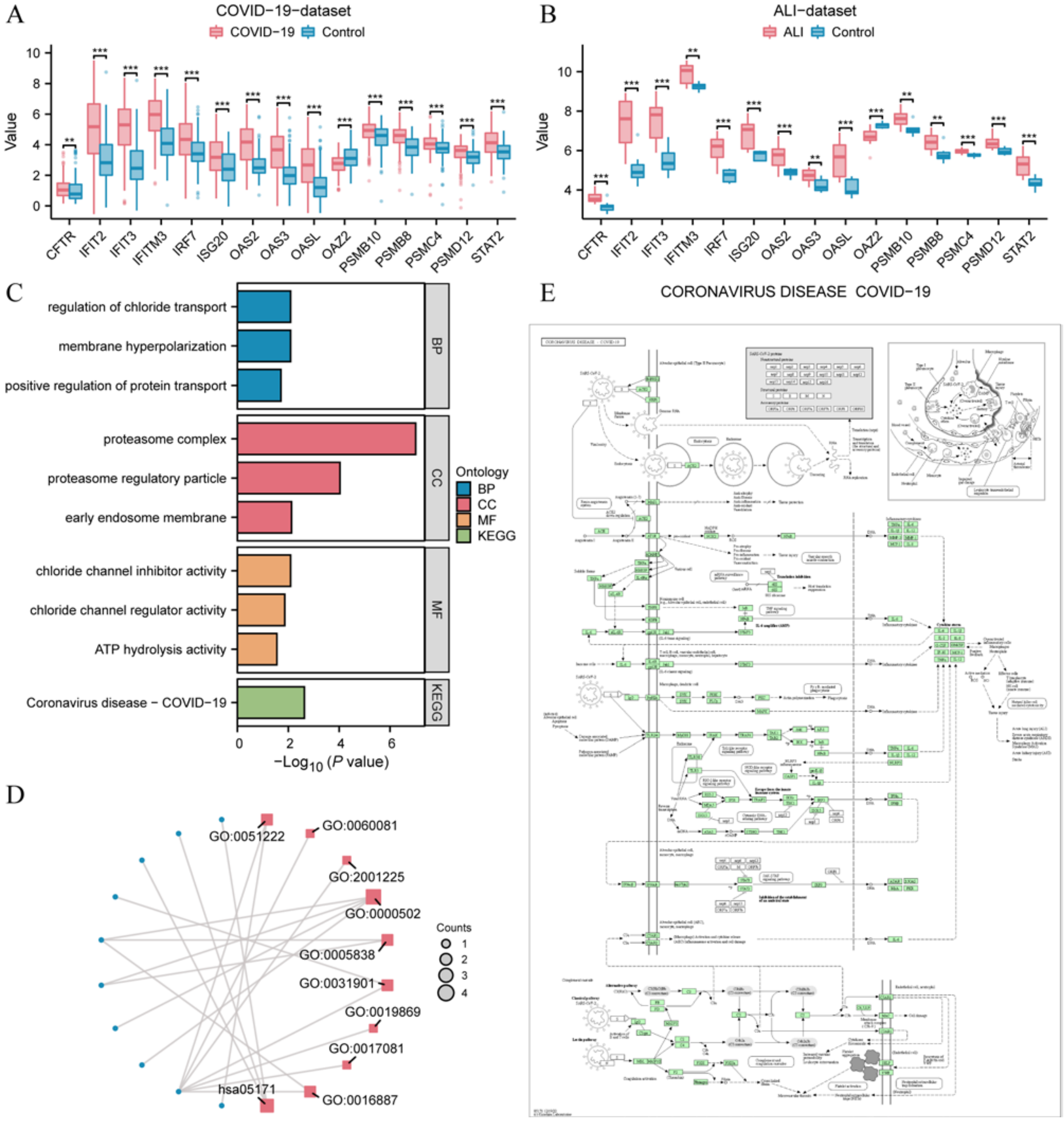
Supplementary Figure S2. Co-DEGs in the COVID-19 and ALI datasets. **A** – Volcano map for DEGs in COVID-19 dataset (healthy control: *n* = 131 and COVID: *n* = 238). **B** – Volcano map for DEGs in ALI dataset (control: *n* = 9 and ALI: *n* = 15). **C** – The Venn diagram of DEGs. **D** – Complex heatmap display of Co-DEGs in the COVID-19 dataset. **E** – Complex heatmap display of Co-DEGs in the ALI dataset

Co-DEGs – common DEGs, ALI – acute lung injury, DEGs – differential expression genes.



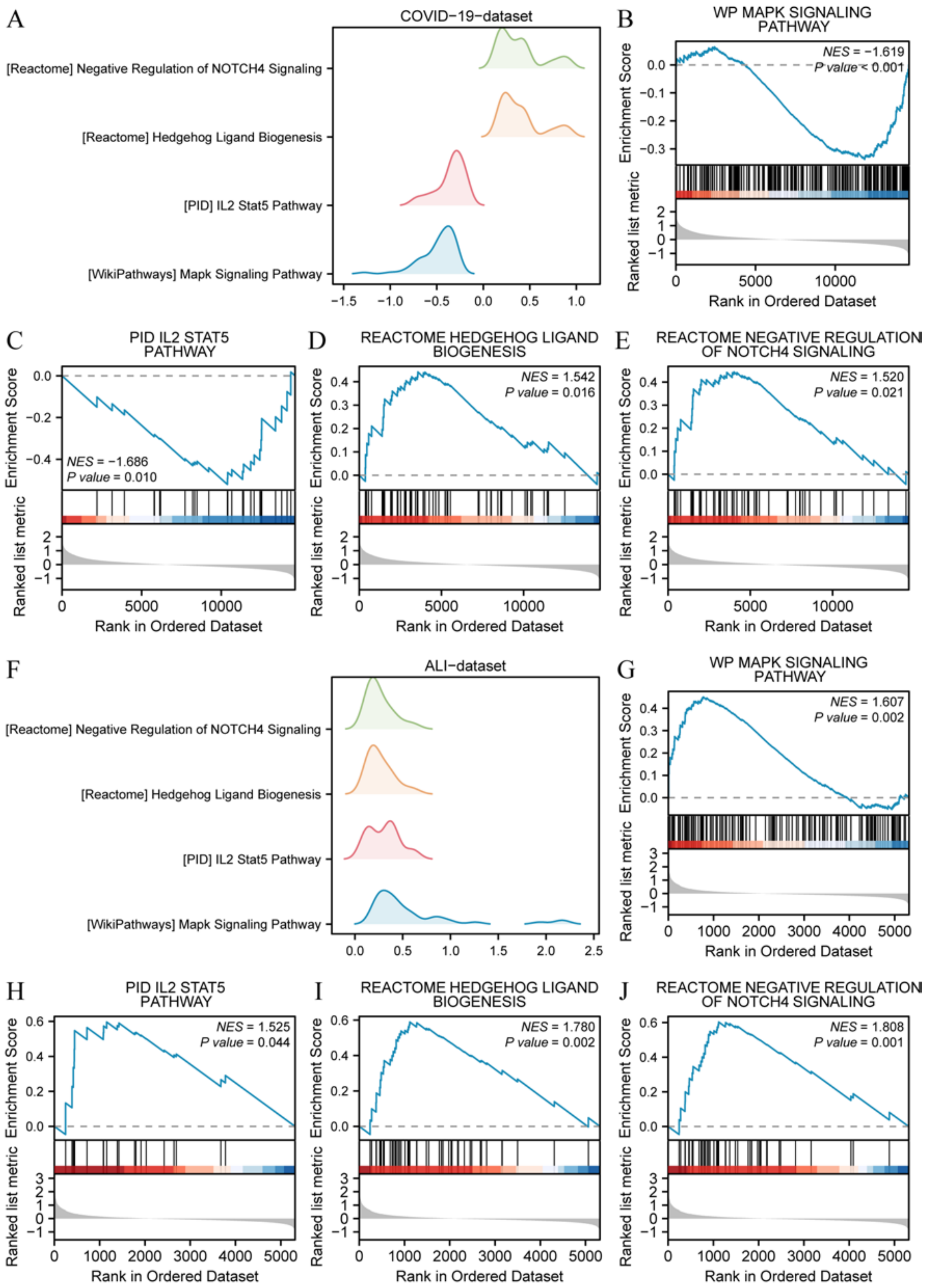
Supplementary Figure S3. MCODE algorithm for computing core gene cluster. **A** – Co-DEGs are displayed in the MCODE plug-in to predict the core gene cluster. **B** – Display the chromosomal localization map of hub genes. **C** – The functional similarity score box diagram of hub genes is displayed

Co-DEGs – common DEGs, MCODE – molecular complex detection.

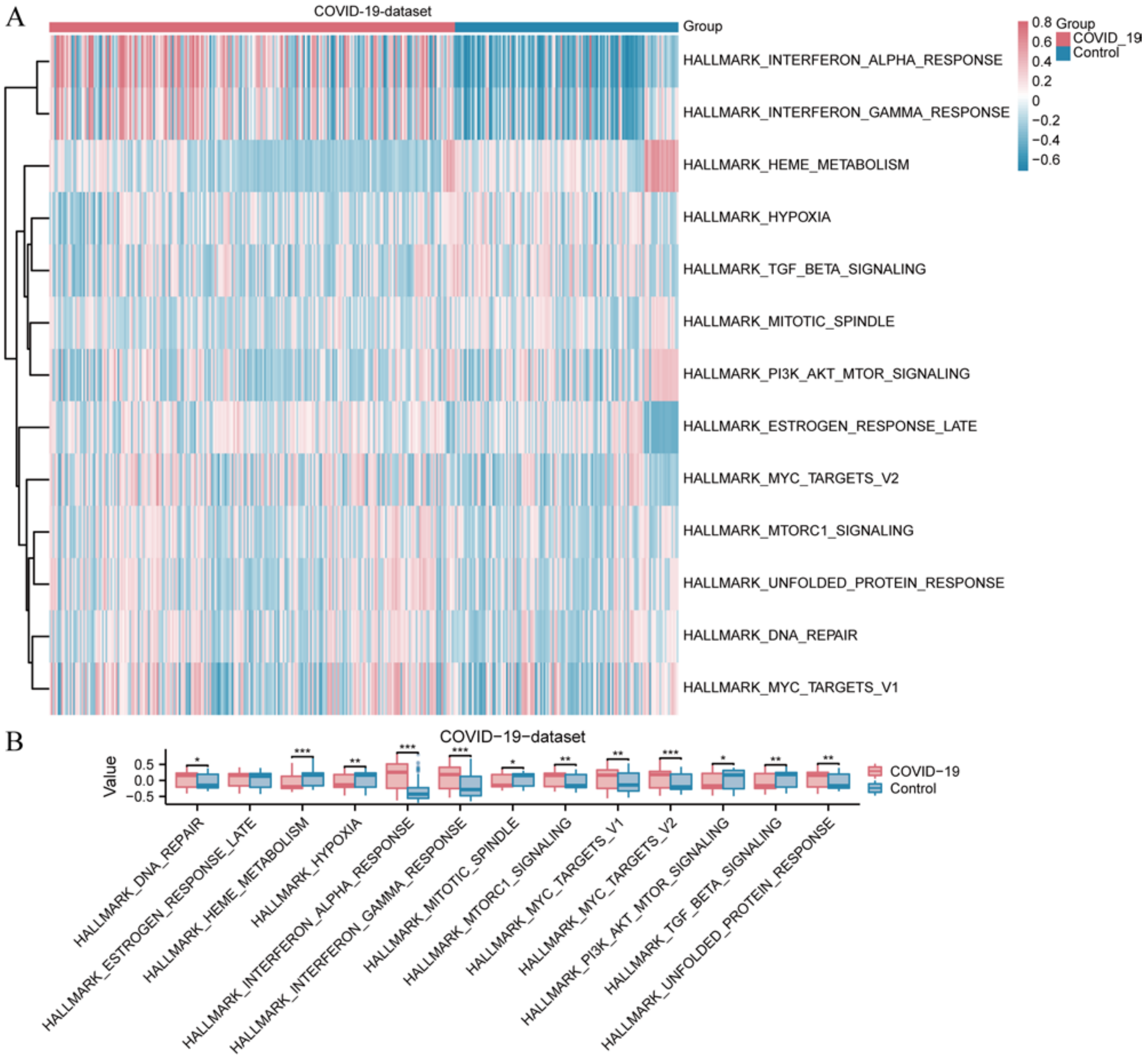


Supplementary Figure S4. GO and KEGG. **A** – Group comparison diagram of hub genes between the disease and control groups in the COVID-19 dataset (healthy control: *n* = 131 and COVID: *n* = 238). **B** –. Group comparison chart in the ALI dataset (control: *n* = 9 and ALI: *n* = 15). **C** – The GO and KEGG enrichment analysis results of hub genes are presented in a bar chart. **D** – In a circular network diagram. **E** – The KEGG pathway diagram of hsa05171. ns – non-significant, **p* < 0.05, ***p* < 0.01 , ****p* < 0.001. The screening criteria for both the analyses are *p* < 0.05 and FDR value < 0.25

GO – gene ontology, KEGG – Kyoto encyclopedia of genes and genomes, ALI – acute lung injury, BP – biological process, CC – cellular component, MF – molecular function.

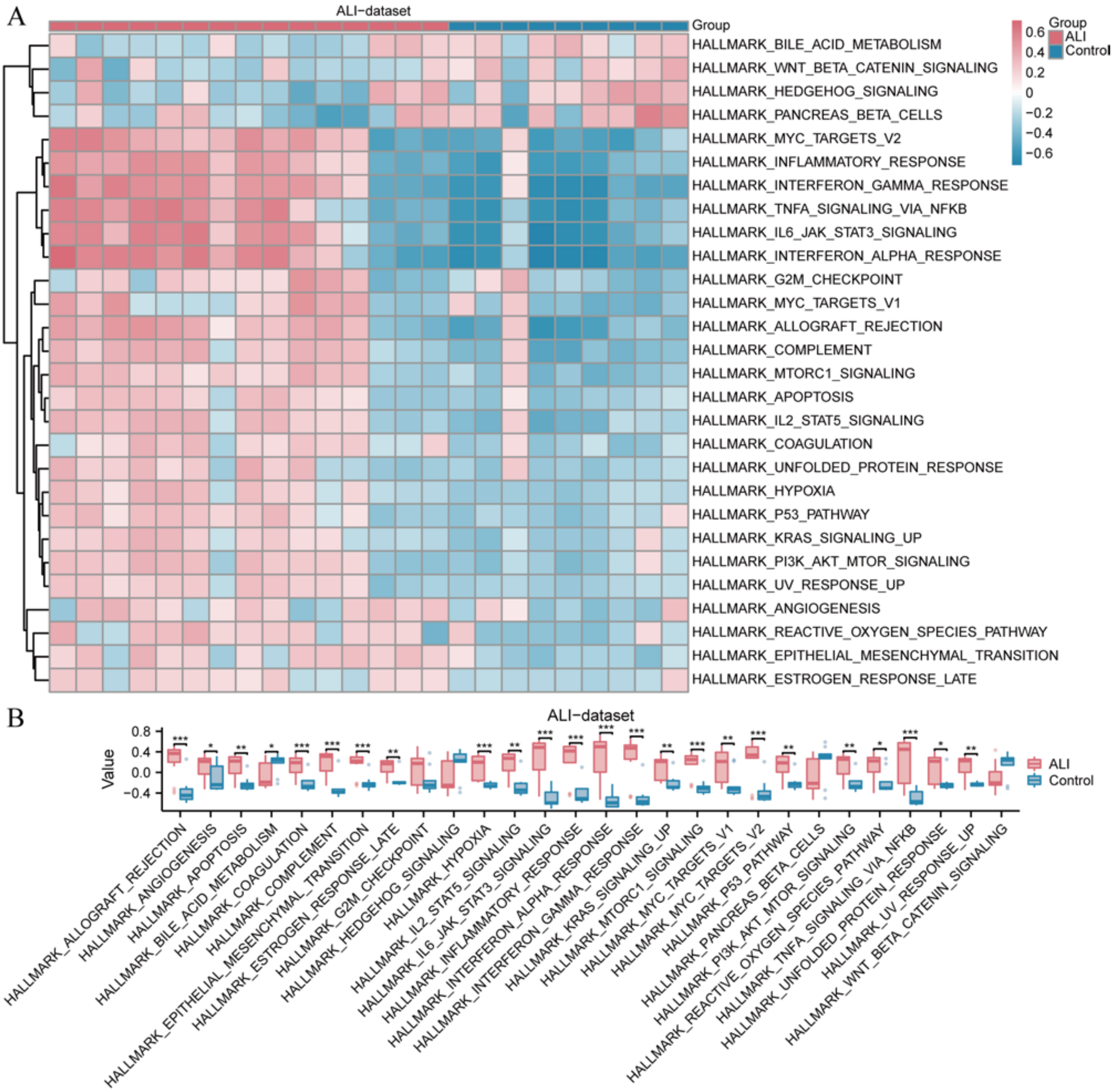


Supplementary Figure S5. GSEA analysis. **A** – Four main biological characteristics in the COVID-19 dataset were found. **B**–**E** – Genes were enriched in pathways as MAPK (**B**), PID NOTCH (**C**), WNT AND PLURIPOTENCY (**D**), and HIPPO (**E**). **F** – Four biological characteristics in the ALI dataset were found. **G**–**J** – Genes were enriched in pathways such as MAPK (**G**), PID NOTCH (**H**), WNT AND PLURIPOTENCY (**I**), and HIPPO (**J**). The criteria for GSEA enrichment analysis were $p < 0.05$ and FDR value (q.value) < 0.25



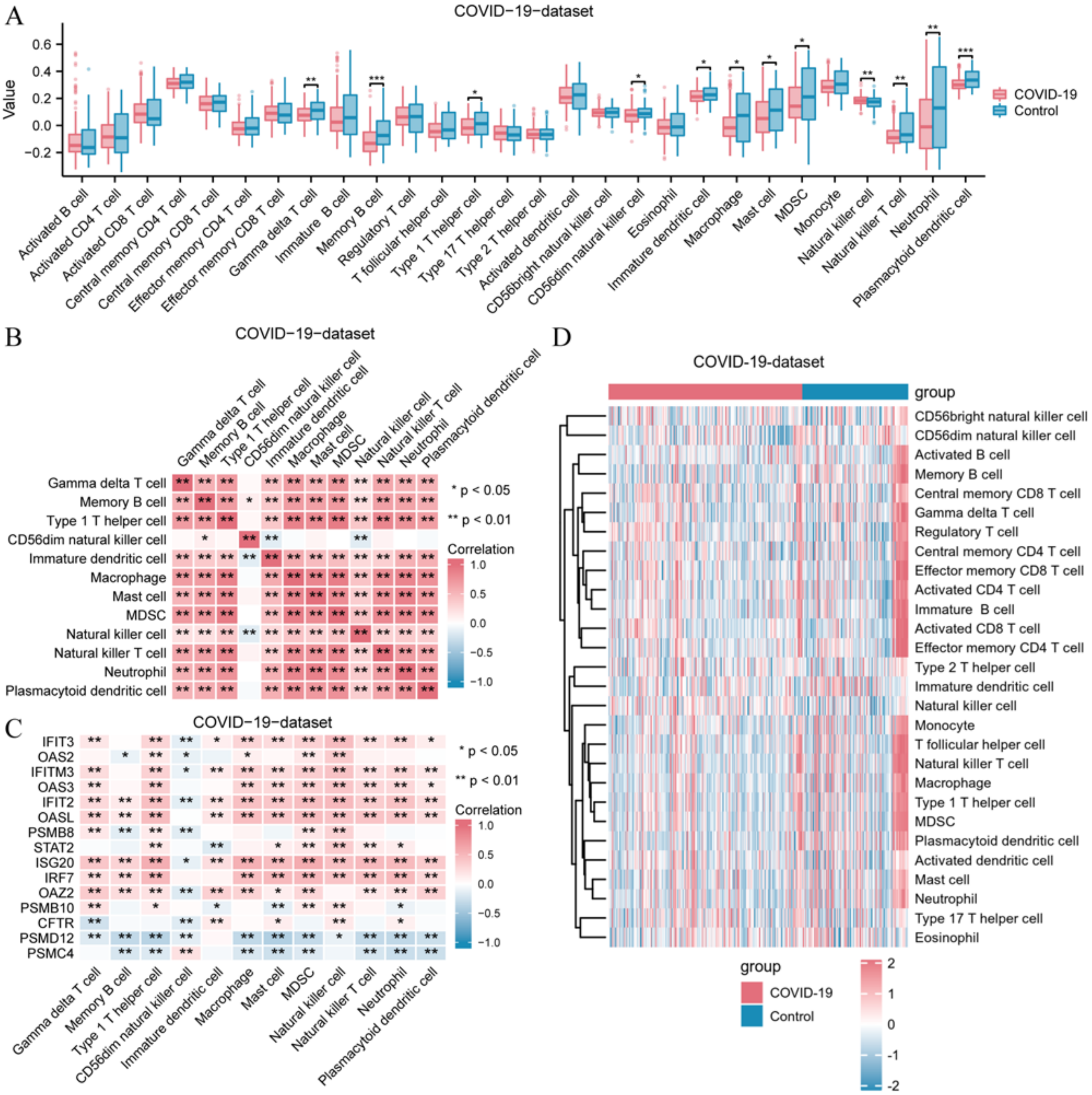
Supplementary Figure S6. GSVA analysis of the COVID-19 dataset. **A** –. Heatmap of the hallmark gene sets. **B** – The group comparison graph of the hallmark gene set between the disease and control groups (healthy control: *n* = 131 and COVID: *n* = 238)

GSVA – gene set variation analysis.



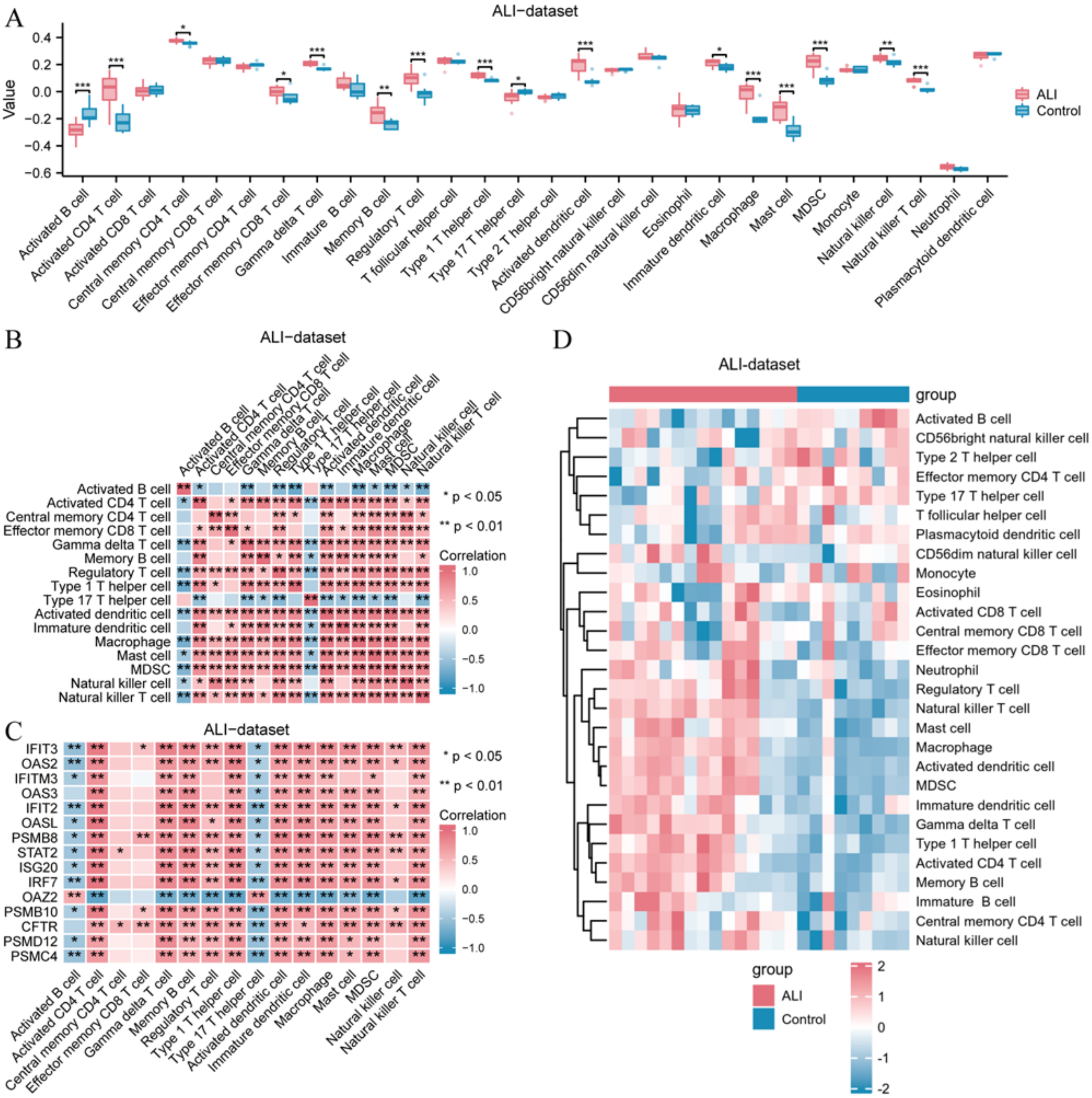
Supplementary Figure S7. GSVA analysis of the ALI dataset. **A** – Heatmap of the hallmark gene sets. **B** – The group comparison graph of the hallmark gene set between the disease and control groups (control: *n* = 9 and ALI: *n* = 15)

GSVA – gene set variation analysis, ALI – acute lung injury.



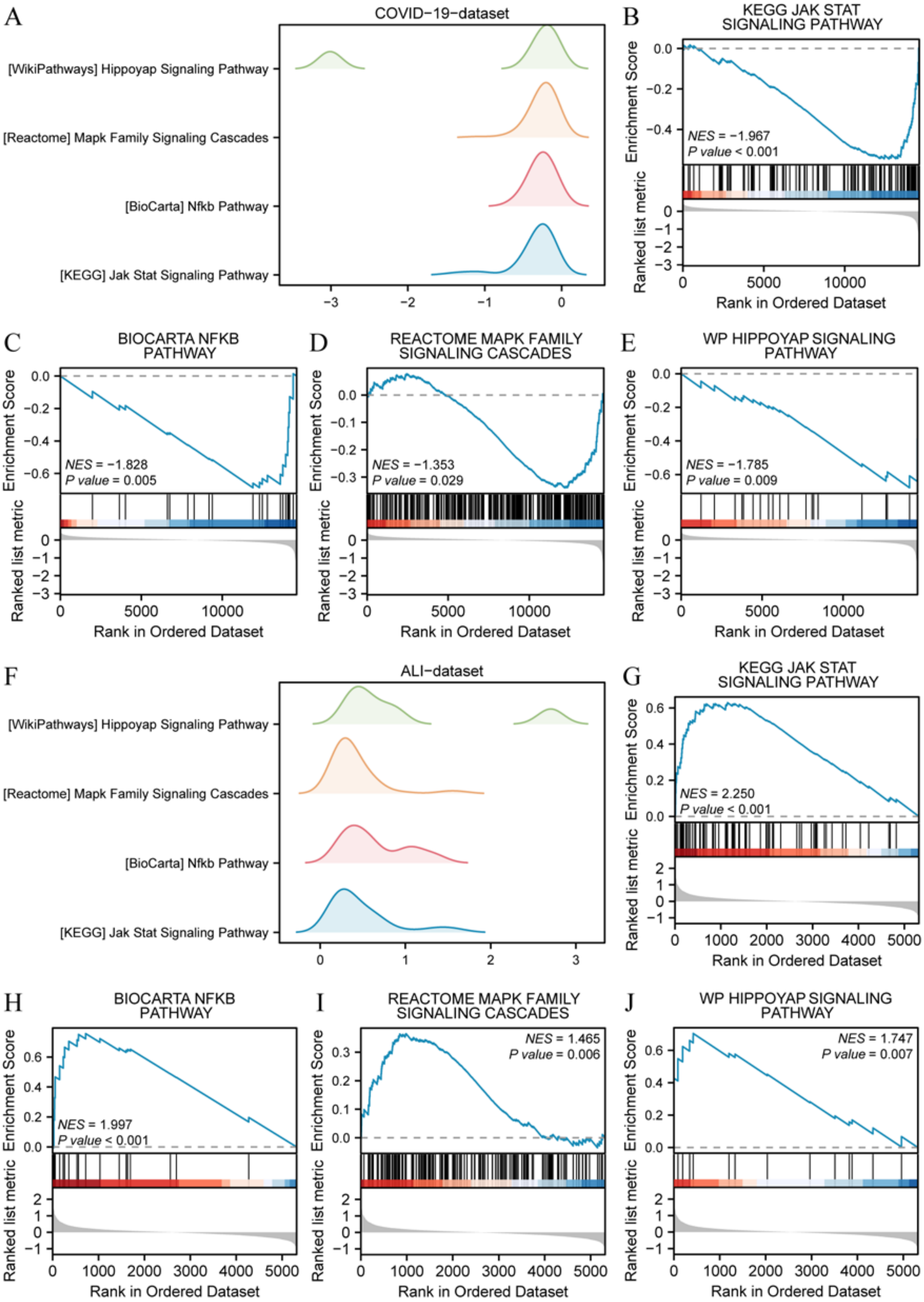
Supplementary Figure S8. ssGSEA immune infiltration analysis in the COVID-19 dataset. **A** – Comparison of the ssGSEA enrichment scores between disease and control groups. **B** – Correlation heatmap between ssGSEA enrichment scores. **C** – Correlation heatmap between ssGSEA enrichment score and hub genes. **D** – The complex heatmap of the ssGSEA enrichment scores between disease and control groups

ssGSEA – enrichment score, infiltration abundance of immune cell, ssGSEA – single sample gene set enrichment analysis.



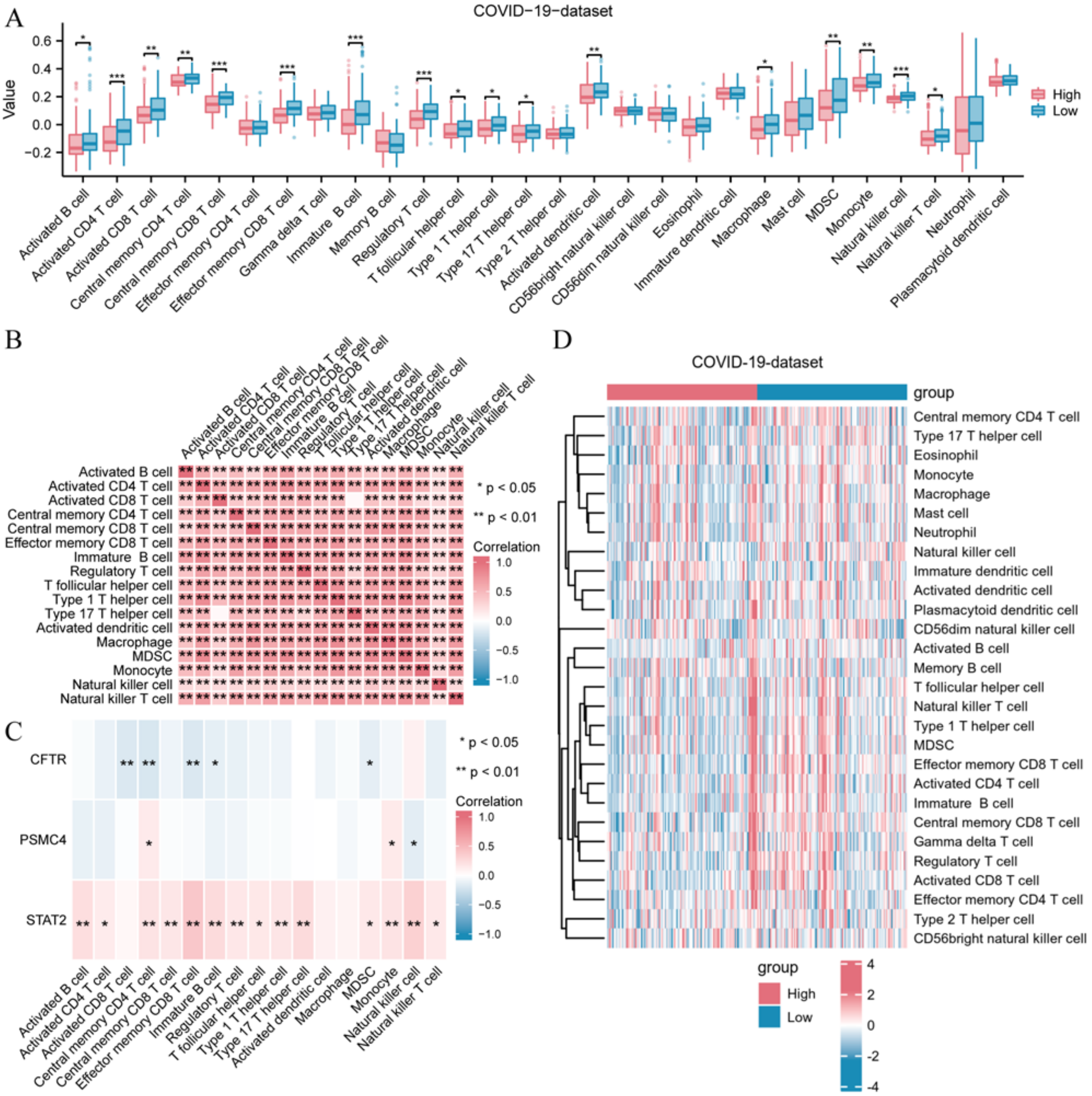
Supplementary Figure S9. ssGSEA immune infiltration analysis in the ALI dataset. **A** – Comparison of the ssGSEA enrichment scores between disease and control groups (control: *n* = 9 and ALI: *n* = 15). **B** – Correlation heatmap between ssGSEA enrichment scores. **C** – Correlation heatmap

ssGSEA – single sample gene set enrichment analysis, ALI – acute lung injury.



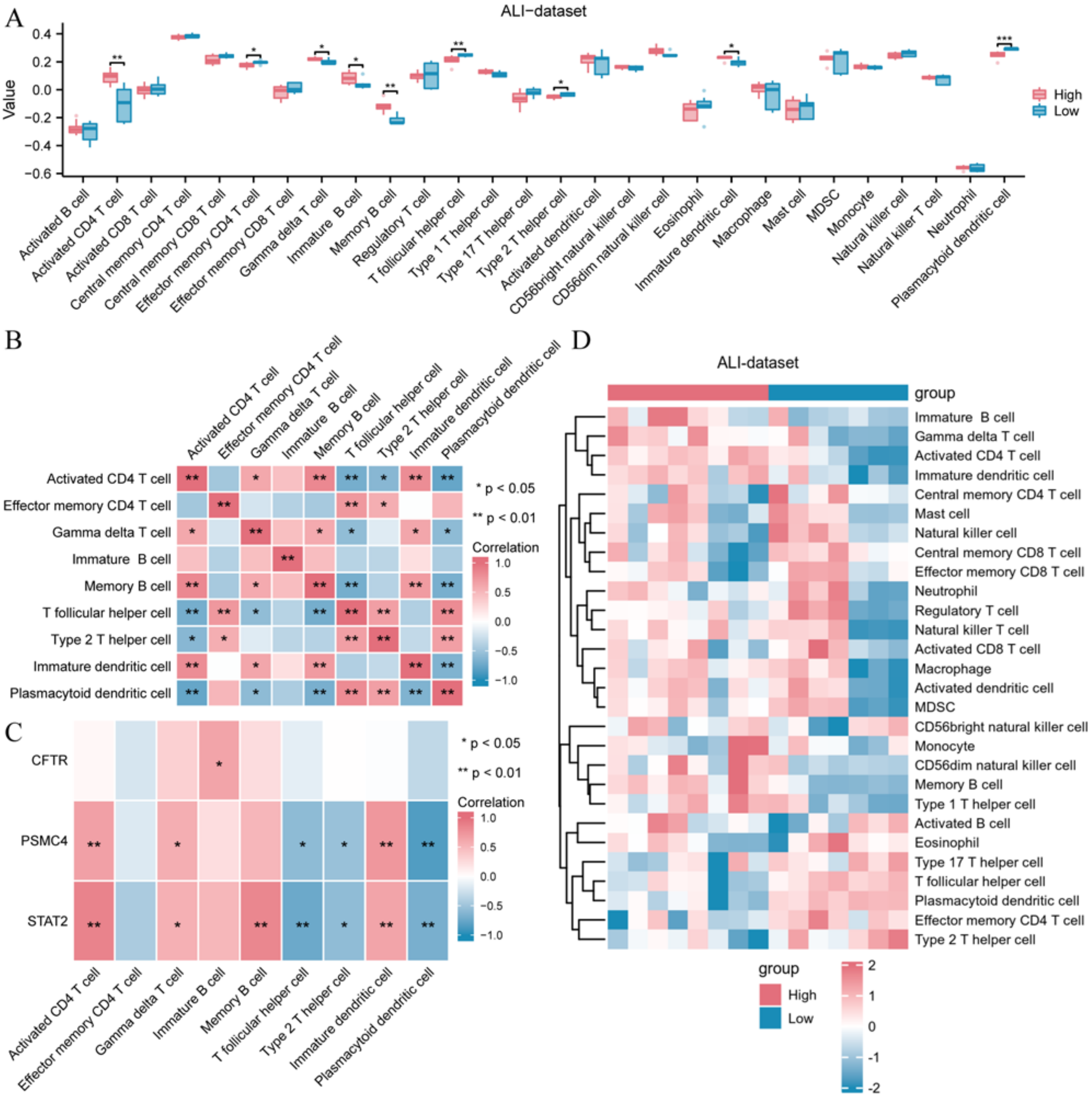
Supplementary Figure S10. GSEA analysis. **A** – The GSEA in the COVID-19 dataset mainly includes four biological characteristics. **B**–**E** – Genes were significantly enriched in pathways such as JAK STAT (**B**), BIOCARTA NFKB (**C**), MAPK (**D**), and HIPPOYAP (**E**). **F** – The GSEA in the ALI dataset mainly includes four biological characteristics. **G**–**J** – Genes were enriched in pathways such as JAK STAT (**G**), BIOCARTA NFKB (**H**), MAPK (**I**), and HIPPOYAP (**J**)

GSEA – gene set enrichment analysis, ALI – acute lung injury.



Supplementary Figure S11. ssGSEA immune infiltration analysis in the COVID-19 dataset between high- and low-score groups. **A** – Comparison of the ssGSEA enrichment scores between the groups (healthy control: *n* = 131 and COVID: *n* = 238). **B** – Correlation heatmap between ssGSEA enrichment scores. **C** – Correlation heatmap between ssGSEA enrichment score and hub genes. **D** – The complex heatmap of the ssGSEA enrichment scores between disease and control groups

ssGSEA – single sample gene set enrichment analysis.



Supplementary Figure S12. ssGSEA immune infiltration analysis in the ALI dataset between high- and low-score groups. **A** – Comparison of the ssGSEA enrichment scores between the groups (control: *n* = 9 and ALI: *n* = 15). **B** – Correlation heatmap between ssGSEA enrichment scores. **C** – Correlation heatmap between ssGSEA enrichment score and hub genes. **D** – The complex heatmap of the ssGSEA enrichment scores between disease and control groups

ssGSEA – single sample gene set enrichment analysis.