

Machine Learning and Single-Cell Analysis Identify Molecular Features of IPF-Associated Fibroblast Subtypes and Their Implications on IPF Prognosis

Jiwei Hou *, Yanru Yang and Xin Han *

Department of Biochemistry and Molecular Biology, School of Medicine &
Holistic Integrative Medicine, Jiangsu Collaborative Innovation Center of
Chinese Medicinal Resources Industrialization, Nanjing University of Chinese
Medicine, Nanjing 210023, China

* Correspondence: houjw@njucm.edu.cn (J.H.); xhan0220@njucm.edu.cn (X.H.)

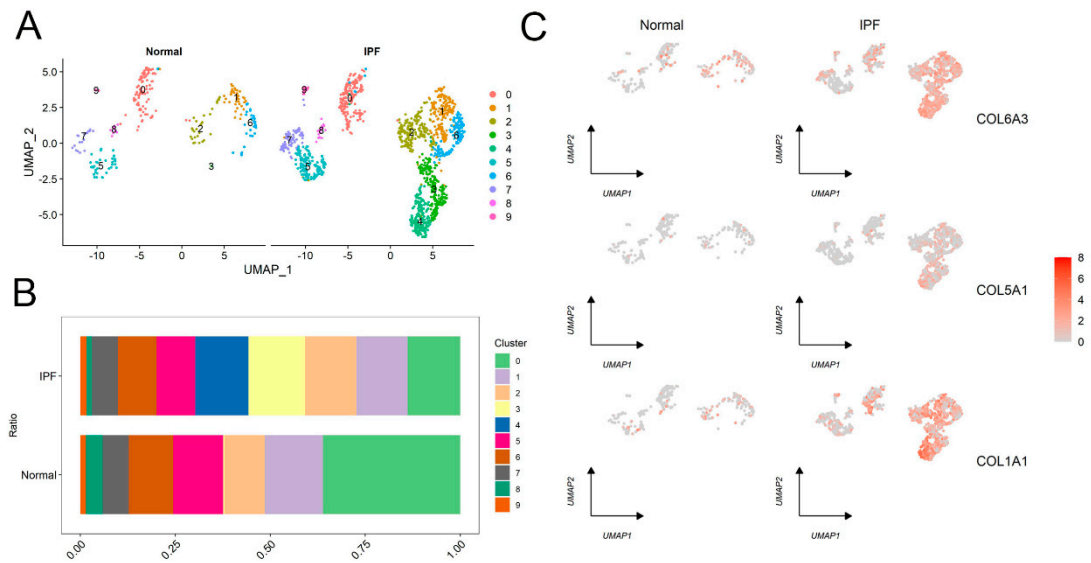


Figure S1: Fibrotic and normal lung fibroblasts subcluster into distinct cell populations. (A) Subclustering of fibrotic and normal lung fibroblasts further identified 10 distinct subtypes. Color-coded UMAP plot is shown and each fibroblast subcluster is defined on the right. (B) Cell proportions of fibroblast subclusters in fibrotic and normal lung tissues. Cells of cluster 4 and cluster 5 were significantly increased in fibrotic samples compared to normal scar samples. (C) Feature plots showing the expression of canonical markers for pulmonary fibrosis, in the fibroblast cell cluster.

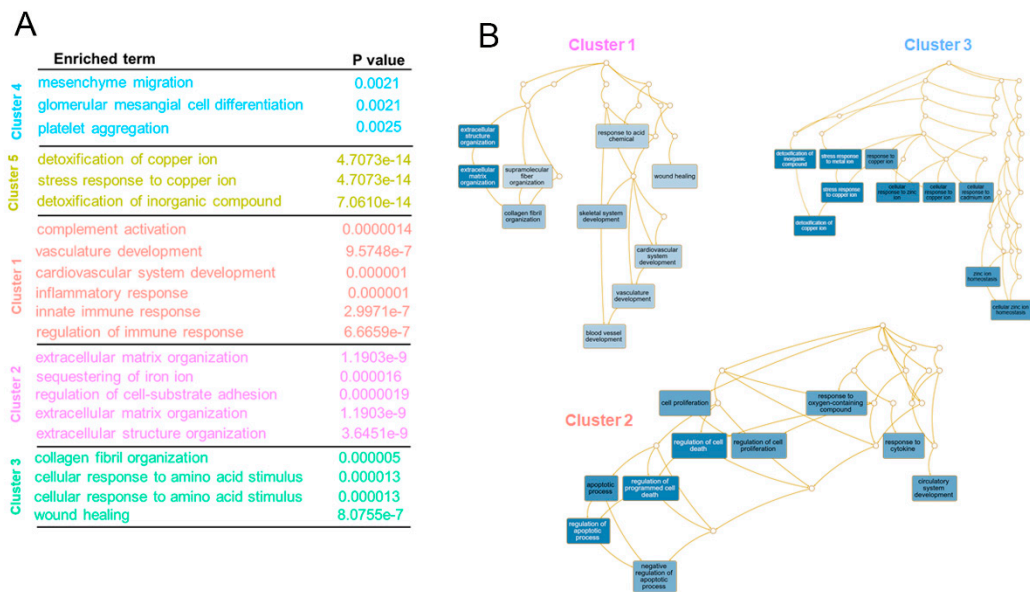


Figure S2: Enrichment analyses of the gene clusters. (A) The representative biological processes related to corresponding DEGs in Figure 3G. (B) GO terms enriched in the three gene clusters in Figure 3H.

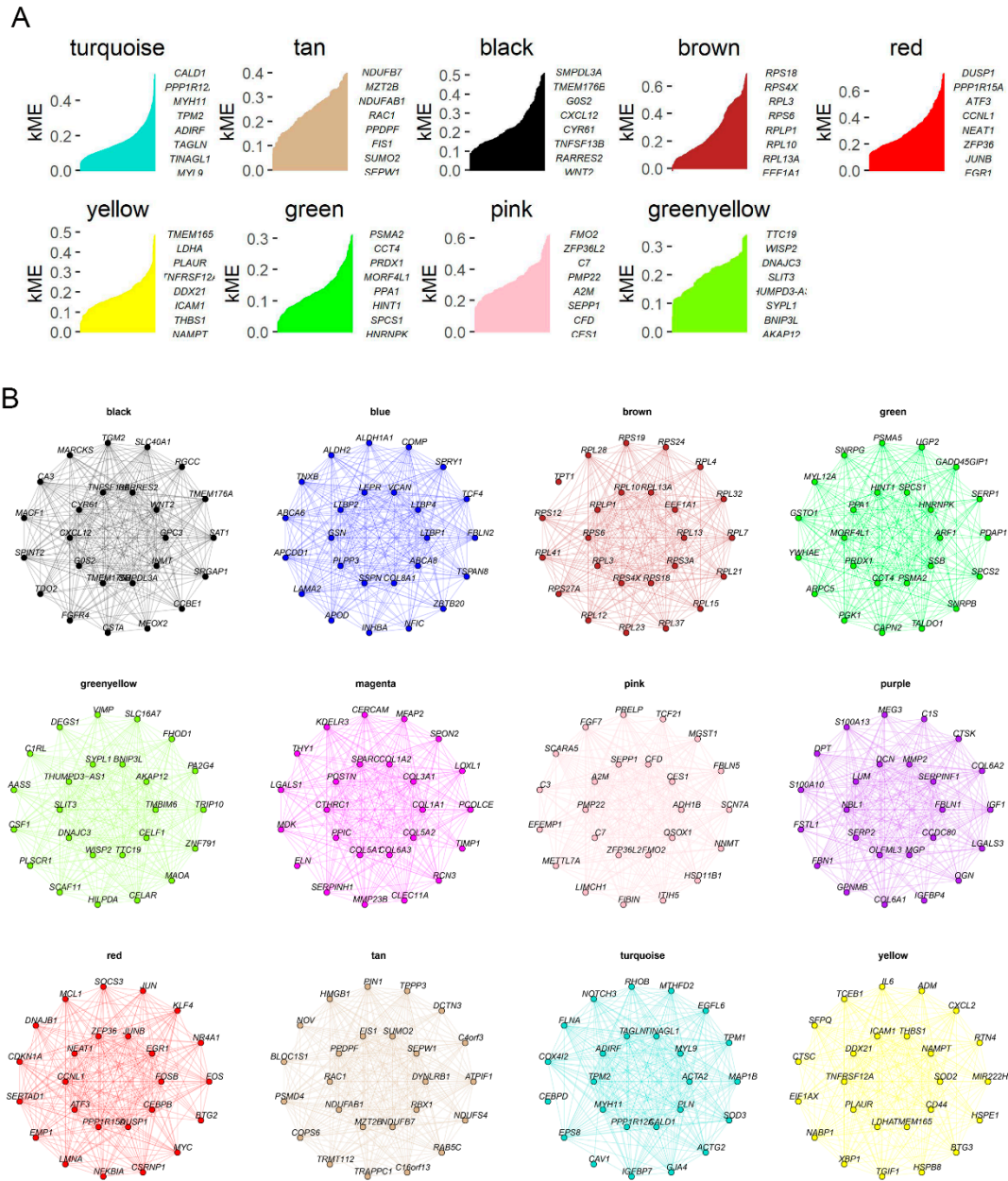


Figure S4:

High dimensional weighted gene co-expression network analysis (hdWGCNA) reveals module-specific hub genes in IPF-related fibroblasts. (A) kME (Eigengene-based connectivity) map showing the top 10 hub genes in each module ranked by kME across the macrophages. Genes with higher connectivity or kME values are considered more central or influential within their respective modules. (B) Protein-protein interaction (PPI) network of the identified hub genes in each module.

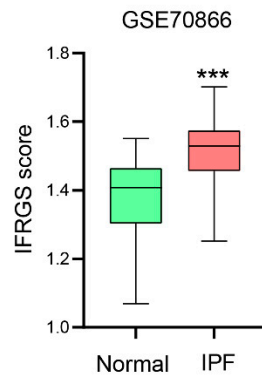


Figure S5

External validation of an IPF-fibroblast-related prognostic signature. (A) Boxplots depict the IPF-fibroblast-related genes (IFRGs) score levels in lung tissue of patients with IPF (n = 212) and normal controls (n = 20). Results are expressed as means \pm SD (***) p < 0.001 vs. Normal).