

Chapter 13

Clinical Application of Molecular Bioinformatics



Yifei Liu, Furong Yan, Xiaoping Liu, Meili Weng, Xiangdong Wang, Yiming Zeng, and Hongzhi Gao

Abstract With the rapid development of bioinformatics, this subject is more and more closely combined with clinical practice. Clinical bioinformatics not only is used more and more in disease diagnosis and prognosis prediction but also plays an important role in the study of pathogenesis, the search of disease markers, the prediction of drug targets, and other aspects. Here, we highlight some of the techniques of clinical bioinformatics and add examples of using bioinformatics methods to solve clinical problems. It focuses on how molecular networks or protein-protein interaction networks influence diseases and how gene co-expression networks relate to clinical phenotypes. The possibility of assigning chronic obstructive pulmonary disease subtypes based on gene expression was also explored.

Yifei Liu, Furong Yan, Xiaoping Liu and Meili Weng contributed equally with all other contributors.

Y. Liu · F. Yan · X. Wang (✉) · H. Gao

Clinical Center for Molecular Diagnosis and Therapy, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China

e-mail: xdwang@fucb.com

X. Liu

School of Mathematics and Statistics, Shandong University at Weihai, Weihai, Shandong Province, China

M. Weng

Department of Pulmonary and Critical Care Medicine, Respiratory Medicine Center of Fujian Province, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China

Y. Zeng

Clinical Center for Molecular Diagnosis and Therapy, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China

Department of Pulmonary and Critical Care Medicine, Respiratory Medicine Center of Fujian Province, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China

e-mail: zengyiming@fjmu.edu.cn

The application of molecular bioinformatics to re-understand diseases is very important and will provide a broader prospect for the diagnosis and treatment of diseases.

Keywords Clinical bioinformatics · Molecular networks · Subtype · Genomics · Regional medical

13.1 Introduction

Bioinformatics plays an important role in clinical diagnosis and research. At present, clinical bioinformatics has been widely used in the discovery of disease-related genes, determination of new drug molecular targets, disease diagnosis, and prognosis prediction (Wooller et al. 2017; Oliver et al. 2015; Fu et al. 2020). In clinical practice, bioinformatics can effectively predict the prognosis of patients or the occurrence and development of diseases based on the integration of previous diagnosis and treatment data and sequencing data and provide guidance for the diagnosis and treatment of diseases. For diseases whose pathogenesis is not clear, bioinformatics can also provide strong guidance, which can effectively save time and avoid aimless experiments.

The clinical identification of disease subtypes has mainly relied on pathology and symptoms, but the use of molecular bioinformatics to identify molecular subtypes has just begun. The molecular subtypes of diseases can be associated with clinical phenotypes, which may indicate the causes of phenotypic changes and explain the different symptoms of the same disease at the molecular level. This paper introduces how clinical bioinformatics can integrate molecular networks into clinical practice and how bioinformatics can be used to reclassify disease and solve clinical problems in regional medicine.

13.2 Methods Suitable for Clinical Practice

Clinical bioinformatics is widely used, which can not only integrate phenotype and gene expression but also predict phenotype and even find etiology through gene expression. It could also focus on genes, regulatory elements, or microRNAs to find potential ways to treat diseases. The following are some bioinformatics methods that can be popularized in clinical practice.

13.2.1 *Weighted Gene Co-expression Network Analysis (WGCNA)*

Correlation network analysis is becoming more and more widely used in biological research. Weighted correlation network analysis (WGCNA) is a method used to describe the gene association patterns among different samples. It can be used to identify highly covariating gene sets and to identify alternative biomarker genes or therapeutic targets based on the connectivity of the gene sets and the association between the gene sets and phenotypes. Compared with the research method of focusing on differentially expressed genes, WGCNA can study thousands of genes with the greatest variation or all detected genes, form co-expression networks, and then conduct significant association analysis of phenotypes. This can either make full use of the information or obtain the important genes associated with the phenotype by screening the hub genes of the module and also provide reference and inspiration for the diagnosis and treatment of clinical diseases (Yin et al. 2018; Bai et al. 2020). WGCNA mainly include the establishment of gene co-expression network, formation of co-expressed gene modules, correlation of co-expressed gene modules with clinical data, correlations between modules and among genes within modules, and screening of hub genes according to gene significance and module membership (Langfelder and Horvath 2008), of which the WGCNA workflow is shown in Fig. 13.1a.

13.2.2 *Identification of Disease Subtypes*

Clinically, diseases are often classified according to their symptomatic characteristics. Consensus clustering provides a new way to classify molecular subtypes of diseases according to gene expression. Based on consensus clustering results, clinical phenotypes of different molecular subtypes were studied by statistical methods such as chi-square test and T test, or WGCNA was used to construct co-expression network to correlate molecular subtypes with clinical phenotypes, which is beneficial to more efficient and accurate diagnosis and treatment of diseases. The consistent clustering method takes sub-sampling from the gene expression matrix to determine the clusters with a specific cluster count (k). For the consensus value, the two items have the same cluster in the number of occurrences in the same subsample, which is calculated and stored in the symmetric consensus matrix for each k . There are many methods to determine the optimal clustering number K value of consensus clustering. The optimal cluster number can be determined by (principal component analysis) PCA method or by consensus CDF (Fig. 13.1b, c). However, no matter which method is used, the final clustering results need to pass the evaluation of clustering significance.

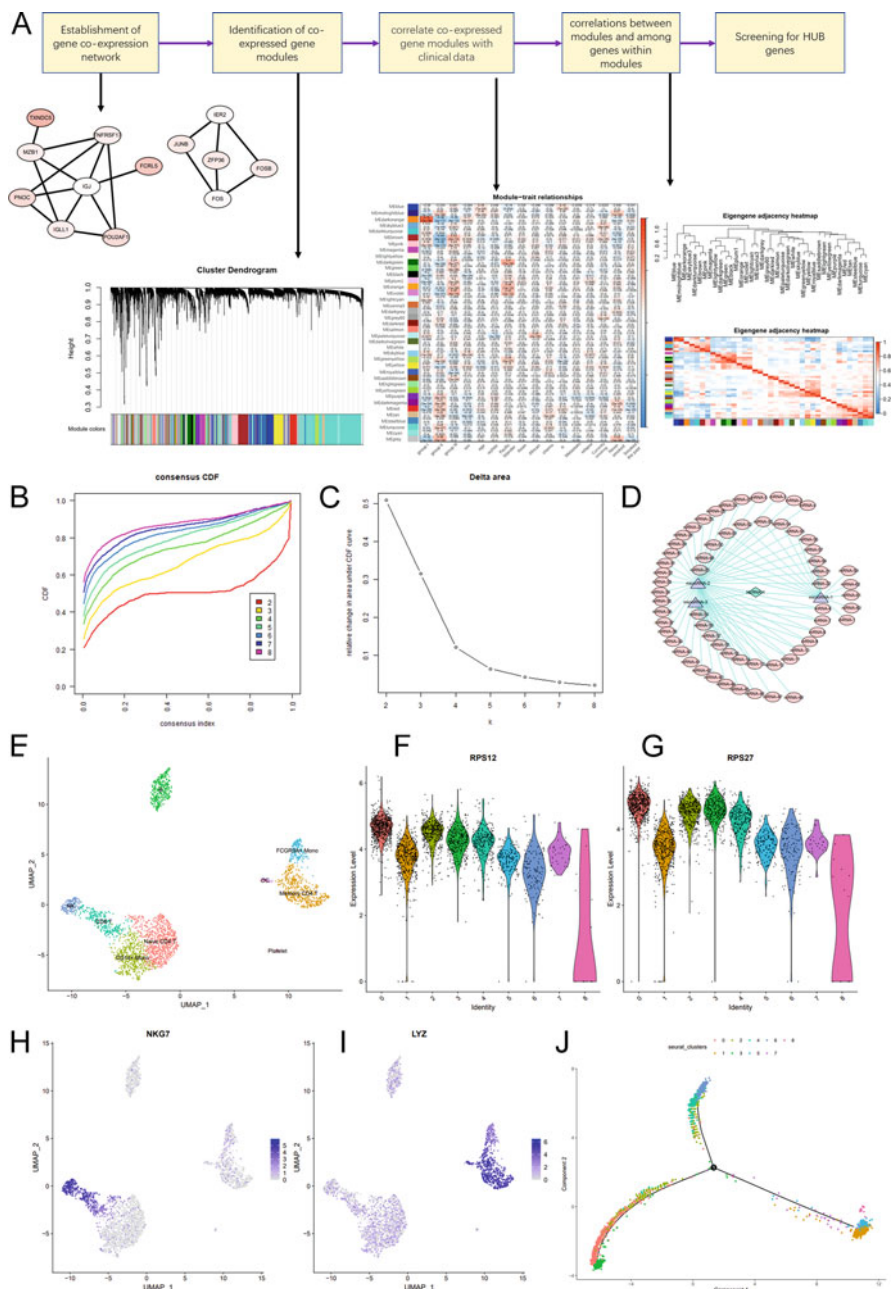


Fig. 13.1 Introduction to several bioinformatics methods. **(a)** The flowchart of WGCNA. First, the co-expression network was constructed, and then gene modules were formed, which were correlated with clinical phenotypes, and HUB genes were selected. **(b, c)** The method of consensus clustering to select K value. **(d)** A schematic diagram of the ceRNA network, which is usually composed of mRNAs, microRNAs, and lncRNAs. **(e, j)** Part of the results of single-cell sequencing analysis, and the analysis data came from the dataset PBMC3K provided by R package Seurat

In addition, before using consensus clustering to classify molecular subtypes of diseases, it is necessary to ensure that no batch effect exists; otherwise, the effect caused by batch effect needs to be eliminated.

13.2.3 The ceRNA Regulatory Network

Competitive endogenous RNA (ceRNA) has attracted much attention in academic circles in recent years. It represents a new regulation mode of gene expression. Compared with the mRNA-miRNA regulation network, the ceRNA regulation network is more sophisticated and complex, involving more RNA molecules, including mRNA, pseudogenes of coding genes, long non-coding RNAs and miRNAs, etc. ceRNA network provides a new way of studying transcriptome and can explain some biological phenomena more deeply. Common ceRNA networks generally contain differentially expressed mRNAs, microRNAs, and lncRNAs or circRNAs. Among them, the expression trend of mRNAs and lncRNAs was consistent, while the expression trend of microRNAs and mRNAs was opposite, and the same was true between microRNAs and lncRNAs. The regulatory relationships among microRNAs, mRNAs, and lncRNAs can be effectively predicted through the construction of the ceRNA regulatory network. It is helpful to excavate gene function and regulation mechanism at a deeper level and facilitate to understand many biological phenomena in a more thorough and comprehensive way (Fig. 13.1d).

13.2.4 Single-Cell Sequencing

Biomarkers are analyzed and mined based on genomics, proteomics, and transcriptomics in a large number of cell or tissue samples, of which the information always ignores the heterogeneity of the sample. In order to fully explore the heterogeneity of cells or tissues and explore the trajectory of cell differentiation, single-cell sequencing is essential (Wang and Song 2017). Techniques such as scRNA-seq and scATAC-seq are gaining popularity in scientific research.

By using Cell Ranger to process single-cell FASTQ files and mapping reads to the reference genome, we can obtain gene expression matrix, annotation information, and cell information. Data are imported into R packages such as Seurat (Satija et al. 2015; Durruthy-Durruthy et al. 2014) and Monocle (<https://cole-trapnell-lab.github.io/monocle3/>) (Trapnell et al. 2014) to create objects, and then principal component analysis (PCA), T-SNE, and other methods can be used to cluster cells, and marker of different clusters can be identified. In addition, cell types can also be identified based on marker identification results. For example, in the clustering results of PBMC samples, we can pick out the cell cluster with CD8a as marker and mark it as CD8+ T cells or pick out the cell clusters with GNLY and

NKG7 as marker and mark them as NK cells (Fig. 13.1e). It is worth noting that different single-cell sequencing methods have different ways of identifying cell types. For example, single-cell ATAC-seq can also identify and cluster similar cell types and states, but it generally uses the open promoter region as a signal of transcriptional activity.

Based on the above analysis results, further pseudotime analysis can be performed. As the cell transitions between states, it undergoes a process of transcriptional recombination, in which some genes are silenced and others are activated. These states are often hard to characterize. Pseudotime analysis of single-cell RNA-seq can view these states without the need to purify the cells (Guerrero-Juarez et al. 2019) (Fig. 13.1j). The single-cell transcriptome analysis data was derived from the PBMC3K dataset provided by R package Seurat.

13.3 Example of Molecular Bioinformatics in Application

Data analysis based on presentation matrices usually requires normalization of the data. Just as the count value in RNA-seq is normalized to obtain the FPKM value, the microarray expression data also needs to be normalized, which can be determined by plotting a boxplot (Fig. 13.2a). If you are using a Series Matrix File on the Geo Dataset for analysis, another problem you may encounter is whether you need to perform log₂ transformations on the data. It can be preliminarily judged from the value of each expression quantity in the expression matrix. The analysis results should not only conform to the set threshold but also be analyzed in combination with the actual situation.

13.3.1 mRNA-MicroRNA Interaction Network

We studied the regulatory networks of mRNA and microRNA in non-specific interstitial pneumonia (NSIP) based on two datasets of GEO dataset GSE110147 (Cecchini et al. 2018) and GSE32538 (Yang et al. 2013) (Table 13.1). The online differential expression analysis tool GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) was used to analyze the differences in the two datasets (GSE110147 and GSE32538), respectively, to obtain the genes and microRNAs differentially expressed in NSIP. Cutoff values were adjusted p -value < 0.01 and $|\log_2FC| > 1.3$ (FC: fold change of expression between NSIP and normal tissue) for DEGs and adjusted p -value < 0.01 for DEMs. GO and KEGG analysis of DEGs was done by DAVID database (<https://david-d.ncifcrf.gov/>) (Kanehisa et al. 2016). The $p < 0.05$ serves as the cutoff value.

The regulatory relationship between mRNAs and microRNAs was predicted based on the miRWalk database (Dweep et al. 2011; Sticht et al. 2018). Protein-protein interaction (PPI) network was obtained from STRING (<http://string-db.org/>)

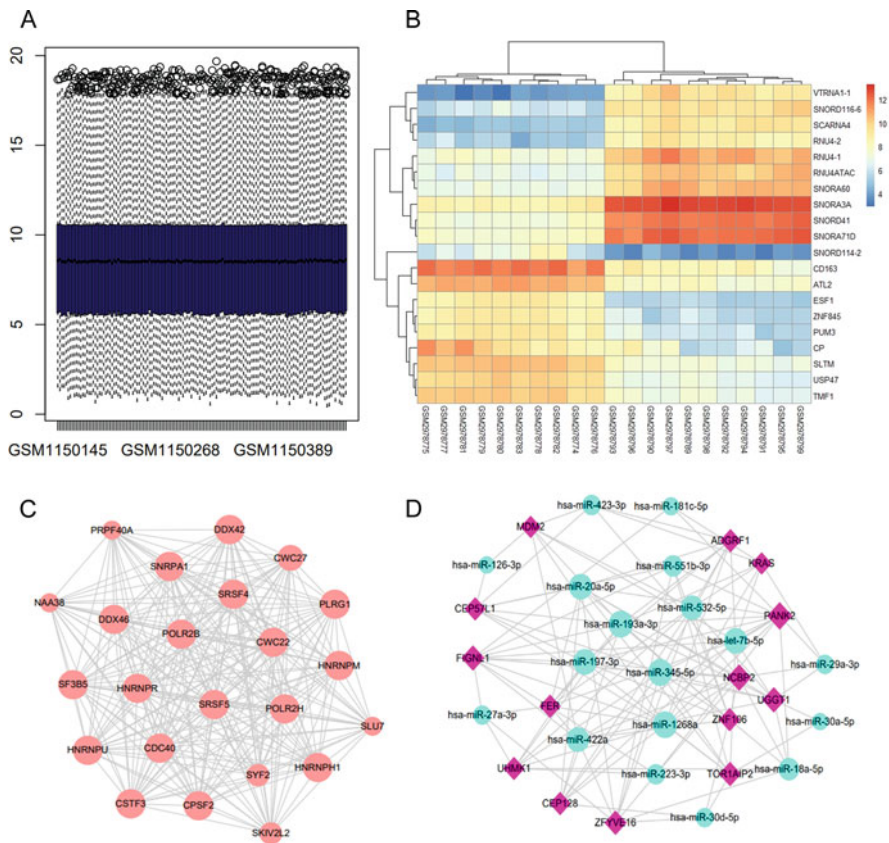


Fig. 13.2 NSIP-related differentially expressed genes and network analysis. **(a)** Boxplot of gene expression data, according to which the normalized data can be roughly judged. **(b)** Heat map of the top 10 upregulated genes and the top 10 downregulated genes. According to the results of cluster analysis, 10 samples on the left were taken from NSIP patients, and 11 samples on the right were normal tissue controls. **(c)** Protein-protein interaction (PPI) network of differentially expressed genes with the highest score. **(d)** Interaction network between differentially expressed genes and differentially expressed microRNAs; 123 interactions between 18 DEMs and 14 DEGs were selected

Table 13.1 The situation of GSE110147 and GSE32538 in GEO database, respectively

GEO number	GSE110147	GSE32538
Experiment type	Expression profiling by array	Non-coding RNA profiling by array
Years	2018	2013
City	London	Aurora
NSIP	10	14
control	11	50

database (Szkarczyk et al. 2015). PPI network was drawn by Cytoscape (Su et al. 2014). The cutoff values were a combined confident score of >0.7 for the PPI network and a node degree of ≥ 10 for screening hub genes. We used the Molecular Complex Detection (MCODE) plug-in for Cytoscape to screen hub genes from the PPI network. As a result, there were 2099 differential expressed genes to be identified between NSIP and normal lung tissue samples, and these genes were potential disease-associated genes for NSIP. 450 genes were upregulated from normal to NSIP, and 1649 genes were downregulated. These genes maybe play key roles in disease onset of NSIP. The heat map of expression quantity of DEGs (the top 10 upregulated genes and the top 10 downregulated genes) was shown in Fig. 13.2b. In addition, we used to adjust p -value < 0.01 as a threshold and identified 21 DEMs between NSIP and normal lung tissue samples.

The functional analysis was performed on GO and KEGG for the 2099 DEGs by DAVID database. In the result analysis, $p < 0.05$ was used as the threshold. The GO analysis revealed that the differential expressed genes were significantly enriched in immune response mechanisms, such as “innate immune response,” “adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains,” “adaptive immune response,” “immunoglobulin-mediated immune response,” “activation of plasma proteins involved in acute inflammatory response,” “immunoglobulin production,” etc. (Table 13.2a). Furthermore, by KEGG pathway analysis, the results indicated that DEGs were significantly enriched in tumor and cell cycle-related pathways, such as “Cell cycle,” “p53 signaling pathway,” or “Pathways in cancer” (Table 13.2b).

The molecular sub-network was identified by mapping the differential expressed genes into the PPI network, choosing the nodes in which the combined score is greater than 0.7 and the degree value is greater than 10. A sub-network with 131 nodes and 1009 edges was obtained from the network. By MCODE, a significant module containing 23 nodes and 246 edges was identified (Fig. 13.2c). We selected the ten genes with the highest degree (degree-value = 22): PLRG1, SRSF4, SNRPA1, HNRNPR, CDC40, DDX42, CWC22, HNRNPU, CPSF2, and CSTF3. Other genes in the significant module network are DDX46, HNRNP, HNRNPH1, SRSF5, POLR2H, POLR2B, SF3B5, CWC27, SKIV2L2, SYF2, SLU7, PRPF40A, and NAA38. Using adjusted p -value < 0.01 as the threshold for DEMs, 21 microRNAs were identified as differential expressed microRNAs between NSIP and normal tissue samples. With miRwalk3.0, a microRNA target gene prediction tool was obtained, and the score >0.95 serves as the cutoff. We predicted the target genes of 21 microRNAs and screened out the overlap between the target genes and the differentially expressed genes. A total of 3687 DEG-DEM interactions are obtained.

In addition, we drew DEG-DEM interaction networks by Cytoscape, calculated degree values of nodes, and further studied sub-networks with degree values ≥ 9 . According to the interaction relationship, we further screened out the pairs of interaction relationship with opposite expression trend, selected a total of 123 interactions between 18 DEMs and 14 DEGs (Fig. 13.2d), and listed them in Table 13.3. In 14 target genes, MDM2, as a target gene of hsa-let-7b-5p, hsa-miR-126-3p,

Table 13.2 Shows the GO and KEGG analysis results of DEGs, in which Table 13.2(a) is the GO analysis results and Table 13.2(b) is the KEGG analysis results with the ten pathways with the lowest p -value

Category	Term	Involved in	p -Value
(a)			
GOTERM_BP_FAT	GO:0002252	Immune effector process	8.23E-04
GOTERM_BP_FAT	GO:0045087	Innate immune response	0.011
GOTERM_BP_FAT	GO:0002449	Lymphocyte-mediated immunity	0.015
GOTERM_BP_FAT	GO:0002250	Adaptive immune response	0.015
GOTERM_BP_FAT	GO:0016064	Immunoglobulin-mediated immune response	0.025
GOTERM_BP_FAT	GO:0002377	Immunoglobulin production	0.035
GOTERM_BP_FAT	GO:0002253	Activation of immune response	0.040
GOTERM_BP_FAT	GO:0002440	Production of molecular mediator of immune response	0.041
(b)			
KEGG_PATHWAY	hsa03040	Spliceosome	3.03E-04
KEGG_PATHWAY	hsa03018	RNA degradation	0.002
KEGG_PATHWAY	hsa04110	Cell cycle	0.003
KEGG_PATHWAY	hsa05222	Small cell lung cancer	0.007
KEGG_PATHWAY	hsa05218	Melanoma	0.007
KEGG_PATHWAY	hsa04510	Focal adhesion	0.008
KEGG_PATHWAY	hsa05016	Huntington's disease	0.009
KEGG_PATHWAY	hsa05200	Pathways in cancer	0.009
KEGG_PATHWAY	hsa05120	Epithelial cell signaling in <i>Helicobacter pylori</i> infection	0.012
KEGG_PATHWAY	hsa00190	Oxidative phosphorylation	0.020

hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-422a, hsa-miR-423-3p, and hsa-miR-532-5p, has been confirmed to be related to NSIP, but the regulatory effects of these four microRNAs on MDM2 in NSIP have not been reported in the literature. Studies have shown that compared with normal lung parenchyma, MDM2 in the epithelial cells of IPF and NSIP patients is significantly upregulated (Nakashima et al. 2005). In addition, CEP128 as a target gene of hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-20a-5p, hsa-miR-30d-5p, hsa-miR-345-5p, hsa-miR-422a, and hsa-miR-532-5p is an autoimmune thyroid diseases' pathogenic factor (Wang et al. 2019).

Based on the above research results, we have identified the interaction relationship between 18 DEMs and 14 DEGs associated with NSIP, which has not been reported yet. Of the 14 NSIP-related DEGs, MDM2 has been shown to be related to NSIP in previous studies (Chen et al. 2017; Wurz and Cee 2019). Therefore, the interaction relationship between 18 DEMs and 14 DEGs selected in this study, especially 4 interaction relationships of MDM2, may provide new ideas for the research of NSIP.

Table 13.3 The list of microRNAs with differential expression and the predicted target genes with an opposing expression trend

Target genes	MicroRNA
ADGRF1	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-345-5p, hsa-miR-423-3p, hsa-miR-532-5p
CEP128	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-20a-5p, hsa-miR-30d-5p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-532-5p
CEP57L1	hsa-miR-126-3p, hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-345-5p, hsa-miR-423-3p, hsa-miR-551b-3p
FER	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-27a-3p, hsa-miR-422a, hsa-miR-532-5p
FIGNL1	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-181c-5p, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-223-3p, hsa-miR-27a-3p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-551b-3p
KRAS	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-181c-5p, hsa-miR-193a-3p, hsa-miR-29a-3p, hsa-miR-423-3p, hsa-miR-532-5p
MDM2	hsa-let-7b-5p, hsa-miR-126-3p, hsa-miR-1268a, hsa-miR-193a-3p, hsa-miR-422a, hsa-miR-423-3p, hsa-miR-532-5p
NCBP2	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-18a-5p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-223-3p, hsa-miR-29a-3p, hsa-miR-30d-5p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-423-3p
PANK2	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-181c-5p, hsa-miR-18a-5p, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-223-3p, hsa-miR-29a-3p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-551b-3p
TOR1AIP2	hsa-let-7b-5p, hsa-miR-1268a, hsa-miR-18a-5p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-30a-5p, hsa-miR-30d-5p, hsa-miR-345-5p, hsa-miR-551b-3p
UGGT1	hsa-let-7b-5p, hsa-miR-18a-5p, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-29a-3p, hsa-miR-30a-5p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-532-5p
UHMK1	hsa-miR-1268a, hsa-miR-18a-5p, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-223-3p, hsa-miR-27a-3p, hsa-miR-345-5p, hsa-miR-532-5p
ZFYVE16	hsa-miR-18a-5p, hsa-miR-20a-5p, hsa-miR-223-3p, hsa-miR-27a-3p, hsa-miR-30d-5p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-532-5p, hsa-miR-551b-3p
ZNF106	hsa-miR-18a-5p, hsa-miR-193a-3p, hsa-miR-197-3p, hsa-miR-20a-5p, hsa-miR-345-5p, hsa-miR-422a, hsa-miR-532-5p, hsa-miR-551b-3p

This table contains 14 genes and 18 microRNAs

13.3.2 Identification of Genes Associated with Open Regions of Chromatin and Super-enhancers in Lung Adenocarcinoma

In addition to the analysis of mRNA-microRNA interaction regulatory network, which can explain the causes of some gene expression changes, the causes of gene expression changes are often explored through the identification of enhancers, super-enhancers, and open regions of chromatin. The presence of super-enhancers and open regions of chromatin generally leads to the upregulation of the corresponding

genes (Buenrostro et al. 2015; Peng and Zhang 2018). Super-enhancers are generally identified by analyzing ChIP-seq processed with H3K27ac (Jiang et al. 2017). The general analysis flow of super-enhancer identification is shown in Fig. 13.3a.

The potential regulatory genes of the super-enhancer can be identified by annotating the genes in the upstream and downstream 50kb range of the super-enhancer. The image shows the results of ChIP-seq analysis of the lung adenocarcinoma cell line A549, Calu-3, and lung fibroblast cell line IMR-90 (Fig. 13.3b–d). Among them, the ChIP-seq data of A549 cell line was derived from the Encyclopedia of DNA Elements (ENCODE) Project (Consortium EP 2012); GEO Accession numbers are GSE91337 and GSM2421889. The ChIP-seq data of the Calu-3 cell line came from the GEO database; GEO Accession numbers are GSM1548075 and GSM1548073 (Fossum et al. 2014). ChIP-seq data for IMR-90 cell line was derived from the Encyclopedia of DNA Elements (ENCODE) Project (Consortium EP 2012); GEO Accession number is GSE16256 (Lister et al. 2009; Hawkins et al. 2010; Bernstein et al. 2010; Lister et al. 2011; Schultz et al. 2015; Micheletti et al. 2017; Rajagopal et al. 2013).

The process of ATAC-seq to identify open regions of chromatin is similar to that of ChIP-seq, but data quality control is required. ATAC-seq data for the A549 cell line came from the Encyclopedia of DNA Elements (ENCODE) Project (Consortium EP 2012); GEO Accession number is GSE114202. Differential expression results of lung adenocarcinoma and normal controls based on TCGA database (Tomczak et al. 2015), we finally screened out five genes: EFNA5, HAVCR1, ATP1B1, DUSP4, and IGF2BP3, among which DUSP4 and IGF2BP3 are associated with prognosis. Prognostic analysis results were obtained from UALCAN (<http://ualcan.path.uab.edu/>) (Chandrashekar et al. 2017) (Fig. 13.3l–p).

13.4 Disease Categories Based on Molecular Networks

In the past, clinical phenotypes have been an important basis for distinguishing disease subtypes. Now, the concept of molecular subtype provides a new research idea for the diagnosis and treatment of diseases. Here, we present a case study of molecular subtypes associated with immune genes in COPD. Chronic obstructive pulmonary disease (COPD) is a form of chronic bronchitis or emphysema characterized by blocked airflow. If not treated, it often develops into pulmonary heart disease or respiratory failure (Blanchette et al. 2014; Kim et al. 2017). With the increase of air pollution, the incidence of COPD is increasing, but its mechanism is still not fully understood. Currently, COPD is still diagnosed and treated based on simple clinical presentation (degree of airflow limitation, symptoms and frequency of exacerbations, etc.). With the popularization of the concept of precision medicine, it has become a general trend to treat patients according to their individual differences (Zhang et al. 2018; Hogg et al. 2004). Reclassification of COPD is essential for developing more effective new treatments or optimizing existing treatments. Therefore, it is necessary for bioinformatics technology and the existing large amount of

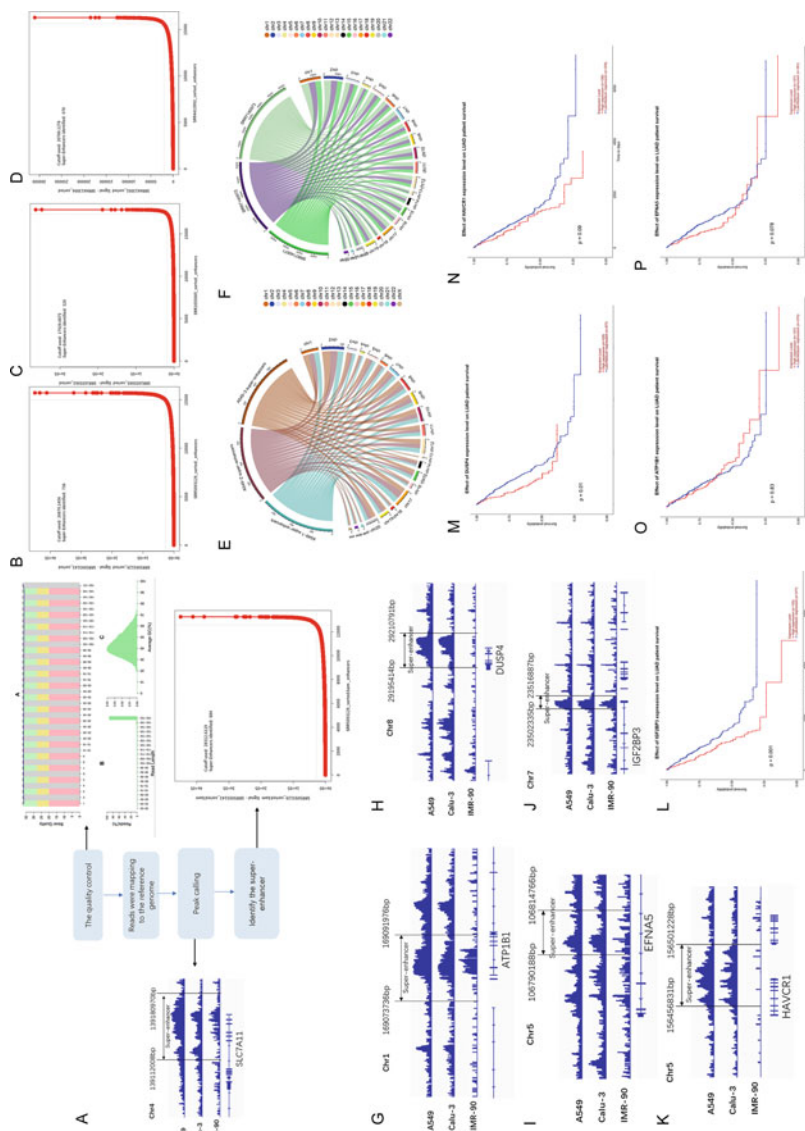


Fig. 13.3 Identification of genes associated with open regions of chromatin and super-enhancers in lung adenocarcinoma. **(a)** The process for identifying super-enhancers. **(b–d)** The identification results of super-enhancers in A549, Calu-3, and IMR-90 cell lines, respectively. **(e, f)** The distribution of super-enhancers and open regions of chromatin on chromosomes, respectively. **(g–k)** The H3K27AC signal near the five genes EFNAS, HAVCR1, ATP1B1, DUSP4, and IGF2BP3 in different cell lines. **(l–p)** Results of survival analysis of five genes. The result suggested that DUSP4 and IGF2BP3 were significantly associated with prognosis

high-throughput data to redefine and interpret large amounts of multi-level information. Two new research strategies (systems biology and network medicine) have the potential to provide new perspectives on the pathology of COPD. Our research has found that the immune-based COPD classification can be used as an auxiliary reference for clinical treatment, which is helpful to the advancement and development of precision medicine.

Common detection items of COPD patients are FEV1 (forced expiratory volume in 1 second) (Chuang and Lin 2019), FVC (forced vital capacity) (Chuang and Lin 2019), emphysema (F-950), DLCO (Hao et al. 2019), etc. DLCO tests the lung's ability to diffuse carbon monoxide. FEV1 is the rapid exhalation of air within 1 s after inspiration to total lung volume. FVC is the maximum amount of breath that can be exhaled as soon as possible after inhaling as much as possible. Emphysema (F-950) is the index involved in quantifying emphysema on CT images by using the density mask method to calculate voxel fraction of the lung (Radder et al. 2017). These indicators play an important role in the clinical diagnosis of COPD.

Although many articles have reported the influence of immune genes and pathways on COPD, the study on the classification of COPD according to the immune gene expression mode of patients' lung tissues has not been reported. COPD is a complex disease driven by a combination of genes; because the gene combinations of different patients are very different, COPD are wildly heterogeneous. Immune-based COPD classification may be used as an auxiliary reference for clinical treatment, which is conducive to the advancement and development of precision medicine.

The expression data of COPD (GSE47460) (Peng et al. 2016; Anathy et al. 2018; Kim et al. 2015; Yu et al. 2018; Tan et al. 2016) were downloaded from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/gds/>). We excluded whole lung homogenate samples with interstitial lung disease and at risk and selected 139 COPD whole lung homogenate samples with different gold stages for analysis. Download the immune gene list from the ImmPort database (<https://www.immport.org/>; SDY1205, DOI: 10.21430/M37N6PJEQT) for research.

Combined with the list of immune genes, all the immune gene expression data in the expression data were taken for consensus clustering by using R package ConsensusClusterPlus (Wilkerson and Hayes 2010). According to the results of the first consensus clustering, we pre-classified the samples into 2 categories, 68 subtype I and 71 subtype II (Fig. 13.4a). By using R package Limma (Ritchie et al. 2015) for differential gene analysis of the 2 subtypes (Fig. 13.4b), 158 different immune genes were obtained. According to the screened 158 immune genes, consensus clustering was carried out for the second time, and the result of 69 subtype I and 70 subtype II was obtained (Fig. 13.4c). 134 immune genes were differentially expressed between the 2 subtypes. Through the third consensus cluster analysis of the 134 different immune genes, the final subtype grouping was obtained, including 70 subtype I and 69 subtype II. A total of 131 immune-related differentially expressed genes were found between the 2 subtypes.

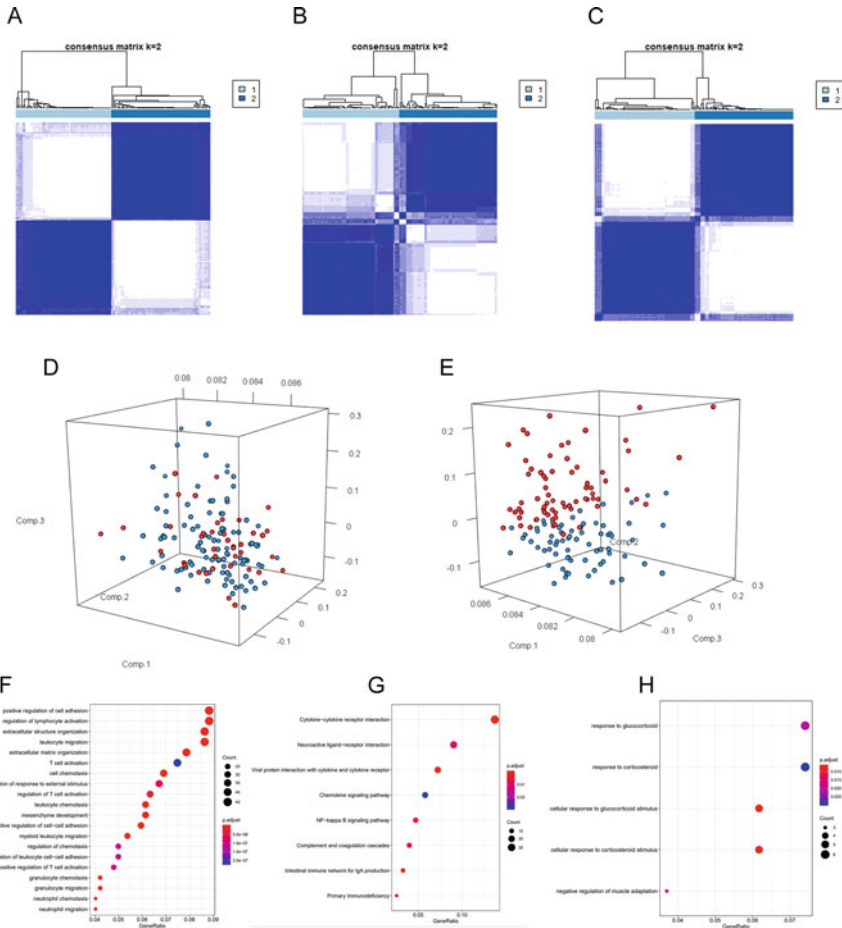


Fig. 13.4 Molecular subtype analysis of chronic obstructive pulmonary disease. (a–c) The result of thrice consensus clustering. (d, e) Principal component analysis (PCA) of immune-related genes. PCA results confirm that the molecular subtypes obtained in this study are not caused by batch effect. (f, g) GO and KEGG enrichment analysis results of differentially expressed genes between subtype I and the control group. (h) GO enrichment analysis result of differentially expressed genes between subtype II and the control group, but there is no significant enrichment KEGG pathway

The R package SigClust (Huang et al. 2012) was used to evaluate the clustering results, and the clustering significance p -values of the two subtypes obtained were shown in Table 13.4. The p -value of the third cluster is the smallest.

In general, the series matrix file of GEO database has preprocessed the data. But we are still trying to verify whether there is a batch effect in the data. To ensure that the intergroup differences we analyzed were not due to batch effect, we queried the sample data one by one from the GEO database and obtained the batch information of 139 COPD samples we used. First, we performed principal component analysis

Table 13.4 shows the clustering significance between the two subtypes of the third consensus clustering

	Subtype I	Subtype II
Subtype I	1	6.175072e-05
Subtype II	6.175072e-05	1

Table 13.5 Chi-square test results between clinical data and subtypes

		Subtype I	Subtype II	df	X-squared	p-value
Gold stages	Gold I and II	38	50	1	4.1916	0.04063
	Gold III and IV	32	19			

The results showed that gold staging is significantly correlated with the two subtypes

(PCA) on the total expression data of 139 COPD samples and on the immune genes differentially expressed between subtypes I and II in the third consensus cluster. Batch or subtype information was labeled in the three-dimensional scatter plot to verify whether the differential gene expression between different subtypes was caused by batch effect. In addition, we did the same for 131 differentially expressed immune genes in the third consensus cluster. Ensure the consistency of clustering results and intergroup differences are independent of the batch effect. As shown in Fig. 13.4d–e, the clustering results of labeled batch information are inconsistent with the clustering results of labeled subtype information, which proves that the subtype we obtained is not caused by batch effect.

To determine the differences between different subtypes and the normal control group, we set $p < 0.01$ and fold change = 0.5 which were used as thresholds to obtain differentially expressed genes between subtypes, and functional enrichment analysis was performed on the results. Functional enrichment analysis was performed using R package clusterProfiler (Yu et al. 2012). As shown in Fig. 13.4g, subtype I was significantly enriched in immune-related pathways, while subtype II was not. This seems to indicate that the immune subtype I identified is more immune-dependent than the immune subtype II.

In combination with clinical data, we investigated the relationship between two immune subtypes and clinical data. We performed chi-square tests on gold stages in the two subtypes. Results showed that the proportion of gold III and gold IV patients in subtype I patients was significantly higher than that in subtype II patients (Table 13.5).

We used R package WGCNA (Langfelder and Horvath 2008) to further investigate the genes that play a key role in the division of molecular subtypes. All the genes in the dataset were included in the analysis so as not to miss out on key information. The results showed that the turquoise module was significantly correlated with the molecular subtypes (Fig. 13.5c). According to gene significance and module membership (Fig. 13.5d), we screened out 11 key genes for subtype classification (Fig. 13.5e–g).

Through bioinformatics and computational analysis, we have determined the possible set of mutations associated with immunity, as well as genes, cell types,

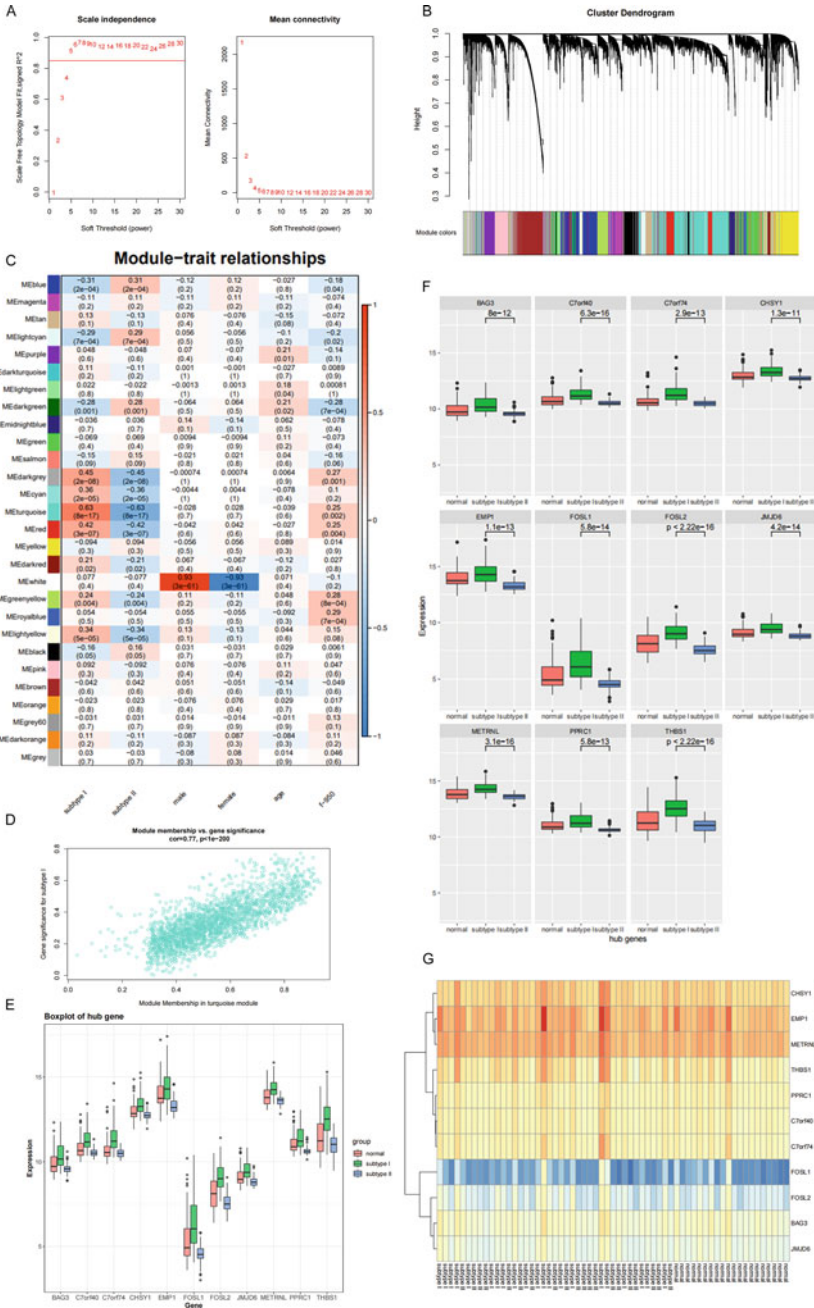


Fig. 13.5 Weighted gene co-expression network analysis and HUB gene screening. (a) The horizontal axis is soft threshold (power), and the vertical axis is the evaluation parameter of scale-free network. The higher the value is, the more the network conforms to the non-scale feature. (b) Gene clustering results. (c) Results of association between gene co-expression modules and clinical phenotypes. Correlation coefficient of threshold setting is greater than 0.5, and p -value is less than 0.05. (d) Correlation between genes and modules and phenotypes in turquoise module.

and biological pathways. Our analysis provides further support for the genetic susceptibility and immune heterogeneity of COPD. We identify the characteristics in each subtype of COPD, which may provide new insights into the biological mechanisms to promote the progress. Studying the use of these endotypes and biomarkers may be helpful for the diagnosis and treatment of COPD and the development of precision medicine.

13.5 Conclusion

Gene sequencing technology helps doctors diagnose patients with symptoms that have no clear cause. But the large amount of data generated is often difficult to get answers quickly. The use of molecular bioinformatics solved this problem. Most diseases are not caused by a single genetic defect but are caused by the interaction of a variety of different genes. Gene expression products such as RNA and proteins interact with other proteins and metabolites in the cell to form a signal regulation network of the disease. Gene mutation did not occur at exactly the same place, but some mutations occur in genes on the same signaling pathway. Gene expression can be changed by the environment, and when changed, specific disease subtypes or endotypes can be formed. Many interventions in the experimental model cannot be completely reproduced on the human body, and therefore molecular bioinformatics provides a way to explore the molecular complexity of a particular disease, to identify disease pathways and modules, and to explore the molecular connections between the different phenotypes. Therefore, molecular bioinformatics has the potential to discover new disease genes, reveal the biological importance of disease-associated mutations, and identify complex diseases, drug targets, and biomarkers (Agusti et al. 2017). The rapid development of molecular bioinformatics provides new ideas for the diagnosis and treatment of diseases. Molecular bioinformatics is defined as a treatment tailored to the individual needs of patients, which distinguishes specific patients from other patients with similar clinical manifestations based on genes, biomarkers, phenotypes, or psychosocial characteristics. Bioinformatics can often reduce research costs and be quick and effective, by computing a large number of sample data, summarizing rules, and associating phenotypes. It helps the precision medicine enter the primary medical system.

For primary hospitals, the simplification of methods is more conducive to the promotion of bioinformatics technology. As bioinformatics tools become more and more accessible, information learning loses some of its complexity and is easier to master quickly through short training. The transition from clinical practice to precision medicine is a more effective and safer way to treat patients than existing



Fig. 13.5 (continued) Genes with GS greater than 0.4 and Mm greater than 0.9 were selected as hub genes. (e–g) The difference and significance of hub genes' expression in the two subtypes and the control group

treatment methods. For the primary medical institutions, it has more development prospects. Most of the training and research related to bioinformatics take place in high-income areas and resource-rich medical institutions, while in primary medical institutions, bioinformatics technology cannot be popularized due to the limited funds and talents. It is becoming more and more urgent to assist primary medical institutions to train professionally talents in the field of bioinformatics. Our article offers an important perspective: molecular bioinformatics can be used in hospitals, and the basic approach we describe is clinically achievable. By learning the methods involved in our research, the personnel of primary medical institutions can use existing resources to re-analyze the published data which helps to re-understand the disease. In addition, the increasing popularity of cloud resources and the availability of online training materials provide excellent opportunities for researchers in primary medical institutions with limited resources. Researchers in primary medical institutions can use cloud resources to analyze large omics datasets, which can reduce the differences caused by equipment shortages to some extent (Mangul et al. 2019). The development of bioinformatics in primary medical institutions is conducive to discovering local related genetic abnormalities.

According to biomedical and life sciences researches, bioinformatics is essential for science to explain treatments and high-throughput omics data meaningful. In the process of disease recognition, diseases are often diagnosed and treated according to phenotypes. Using molecular biological information technology to classify ovarian cancer, it was found that the FGF pathway, a pathway related to tumor proliferation and angiogenesis, plays a significant role in one of the subtypes of ovarian cancer (Hofree et al. 2013). The subtype of liver cancer that overexpress seven hub genes may lead to reduced overall survival in patients (Li et al. 2021). According to the data searched from the public database, bladder cancer is divided into two main molecular subtypes, basic type and differentiated type, and it is found that basic type tumors are associated with a shorter survival period (Volkmer et al. 2012). Cancer involves not only individual mutations but also dysregulation of multiple pathways governing fundamental cell processes such as cell proliferation and apoptosis (Kreeger and Lauffenburger 2010). Increased researches have successfully integrated that database with the molecular to map the signal network of cancer. Through the use of bioinformatics analysis in the molecular signal network, we can subdivide a set of tumor mutation into different subtypes via their biological and clinical information. These subtypes are different from those classified by other clinical markers that are well known to be associated with survival. The subtypes may provide new insight for biological mechanisms driving disease progression.

As molecular bioinformatics become integrated into clinical treatments (Seiler et al. 2017), molecular subtype will become critical for determining the intrinsic feature of many diseases. Heterogeneity is a major challenge to promote precision medicine. If molecular bioinformatics is applied to clinical practice, the treatment and prognosis of diseases will be improved to a new height. We hope that the integration of molecular bioinformatics and multi-omics data will enable patients to receive more accurate, effective, and safe treatments.

Acknowledgments The data for the identification of lung adenocarcinoma super-enhancer and chromatin open region comes from the Encyclopedia of DNA Elements (ENCODE) Project. I sincerely thank the ENCODE Consortium and the ENCODE Production Laboratory(s) for generating the particular dataset(s).

References

- Agusti A, Celli B, Faner R. What does endotyping mean for treatment in chronic obstructive pulmonary disease? *Lancet*. 2017;390:980–7.
- Anathy V, Lahue KG, Chapman DG, Chia SB, Casey DT, Aboushousha R, et al. Reducing protein oxidation reverses lung fibrosis. *Nat Med*. 2018;24:1128–35.
- Bai KH, He SY, Shu LL, Wang WD, Lin SY, Zhang QY, et al. Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Med*. 2020;9:4290–8.
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol*. 2010;28:1045–8.
- Blanchette CM, Gross NJ, Altman P. Rising costs of COPD and the potential for maintenance therapy to slow the trend. *Am Health Drug Benefits*. 2014;7:98–106.
- Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109(1):21–9.
- Cecchini MJ, Hosein K, Howlett CJ, Joseph M, Mura M. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. *Respir Res*. 2018;19:153.
- Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19:649–58.
- Chen Y, Wang DD, Wu YP, Su D, Zhou TY, Gai RH, et al. MDM2 promotes epithelial-mesenchymal transition and metastasis of ovarian cancer SKOV3 cells. *Br J Cancer*. 2017;117:1192–201.
- Chuang ML, Lin IF. Investigating the relationships among lung function variables in chronic obstructive pulmonary disease in men. *PeerJ*. 2019;7:e7829.
- Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
- Durruthy-Durruthy R, Gottlieb A, Hartman BH, Waldhaus J, Laske RD, Altman R, et al. Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell*. 2014;157:964–78.
- Dweep H, Sticht C, Pandey P, Gretz N. miRWalk--database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J Biomed Inform*. 2011;44:839–47.
- Fossum SL, Mutolo MJ, Yang R, Dang H, O’Neal WK, Knowles MR, et al. Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells. *Nucleic Acids Res*. 2014;42:13588–98.
- Fu Y, Ling Z, Arabnia H, Deng Y. Current trend and development in bioinformatics research. *BMC Bioinformatics*. 2020;21:538.
- Guerrero-Juarez CF, Dedhia PH, Jin S, Ruiz-Vega R, Ma D, Liu Y, et al. Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat Commun*. 2019;10:650.
- Hao W, Li M, Zhang C, Zhang Y, Du W. Increased levels of inflammatory biomarker CX3CL1 in patients with chronic obstructive pulmonary disease. *Cytokine*. 2019;126:154881.
- Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*. 2010;6:479–91.

- Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nat Methods*. 2013;10:1108–15.
- Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350:2645–53.
- Huang H, Liu Y, Marron J. sigclust: Statistical Significance of Clustering. R package version: 1.0.0. In 2012.
- Jiang YY, Lin DC, Mayakonda A, Hazawa M, Ding LW, Chien WW, et al. Targeting super-enhancer-associated oncogenes in oesophageal squamous cell carcinoma. *Gut*. 2017;66:1358–68.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44:D457–62.
- Kim S, Herazo-Maya JD, Kang DD, Juan-Guardela BM, Tedrow J, Martinez FJ, et al. Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics*. 2015;16:924.
- Kim SW, Rhee CK, Kim KU, Lee SH, Hwang HG, Kim YI, et al. Factors associated with plasma IL-33 levels in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2017;12:395–402.
- Kreeger PK, Lauffenburger DA. Cancer systems biology: a network modeling perspective. *Carcinogenesis*. 2010;31:2–8.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Li Z, Lin Y, Cheng B, Zhang Q, Cai Y. Identification and analysis of potential key genes associated with hepatocellular carcinoma based on integrated bioinformatics methods. *Front Genet*. 2021;12:571231.
- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Toni-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462:315–22.
- Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*. 2011;471:68–73.
- Mangul S, Martin LS, Langmead B, Sanchez-Galan JE, Toma I, Hormozdiari F, et al. How bioinformatics and open data can boost basic science in countries and universities with limited resources. *Nat Biotechnol*. 2019;37:324–6.
- Micheletti R, Plaisance I, Abraham BJ, Sarre A, Ting CC, Alexanian M, et al. The long noncoding RNA Wisper controls cardiac fibrosis and remodeling. *Sci Transl Med*. 2017;9(395)
- Nakashima N, Kuwano K, Maeyama T, Hagimoto N, Yoshimi M, Hamada N, et al. The p53-Mdm2 association in epithelial cells in idiopathic pulmonary fibrosis and non-specific interstitial pneumonia. *J Clin Pathol*. 2005;58:583–9.
- Oliver GR, Hart SN, Klee EW. Bioinformatics for clinical next generation sequencing. *Clin Chem*. 2015;61:124–35.
- Peng Y, Zhang Y. Enhancer and super-enhancer: positive regulators in gene transcription. *Animal Model Exp Med*. 2018;1:169–79.
- Peng X, Moore M, Mathur A, Zhou Y, Sun H, Gan Y, et al. Plexin C1 deficiency permits synaptotagmin 7-mediated macrophage migration and enhances mammalian lung fibrosis. *FASEB J*. 2016;30:4056–70.
- Radder JE, Zhang Y, Gregory AD, Yu S, Kelly NJ, Leader JK, et al. Extreme trait whole-genome sequencing identifies PTPRO as a novel candidate gene in emphysema with severe airflow obstruction. *Am J Respir Crit Care Med*. 2017;196:159–71.
- Rajagopal N, Xie W, Li Y, Wagner U, Wang W, Stamatoyannopoulos J, et al. RFECs: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput Biol*. 2013;9:e1002968.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.

- Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol.* 2015;33:495–502.
- Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature.* 2015;523:212–6.
- Seiler R, Ashab HAD, Erho N, van Rhijn BWG, Winters B, Douglas J, et al. Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. *Eur Urol.* 2017;72:544–54.
- Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One.* 2018;13:e0206239.
- Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics.* 2014;47(1):8–13. 11–24
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447–52.
- Tan J, Tedrow JR, Dutta JA, Juan-Guardela B, Nouraie M, Chu Y, et al. Expression of RXFP1 Is decreased in idiopathic pulmonary fibrosis. Implications for relaxin-based therapies. *Am J Respir Crit Care Med.* 2016;194:1392–402.
- Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn).* 2015;19:A68–77.
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.* 2014;32:381–6.
- Volkmer JP, Sahoo D, Chin RK, Ho PL, Tang C, Kurtova AV, et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc Natl Acad Sci U S A.* 2012;109:2078–83.
- Wang J, Song Y. Single cell sequencing: a distinct new field. *Clin Transl Med.* 2017;6:10.
- Wang B, Jia X, Yao Q, Li Q, He W, Li L, et al. CEP128 is a crucial risk locus for autoimmune thyroid diseases. *Mol Cell Endocrinol.* 2019;480:97–106.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010;26:1572–3.
- Wooller SK, Benstead-Hume G, Chen X, Ali Y, Pearl FMG. Bioinformatics in translational drug discovery. *Biosci Rep.* 2017;37
- Wurz RP, Cee VJ. Targeted degradation of MDM2 as a new approach to improve the efficacy of MDM2-p53 inhibitors. *J Med Chem.* 2019;62:445–7.
- Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, et al. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax.* 2013;68:1114–21.
- Yin L, Cai Z, Zhu B, Xu C. Identification of key pathways and genes in the dynamic progression of HCC based on WGCNA. *Genes (Basel).* 2018;9(2):92.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284–7.
- Yu G, Tzouveleki A, Wang R, Herazo-Maya JD, Ibarra GH, Srivastava A, et al. Thyroid hormone inhibits lung fibrosis in mice by improving epithelial mitochondrial function. *Nat Med.* 2018;24:39–49.
- Zhang Z, Cheng X, Yue L, Cui W, Zhou W, Gao J, et al. Molecular pathogenesis in chronic obstructive pulmonary disease and therapeutic potential by targeting AMP-activated protein kinase. *J Cell Physiol.* 2018;233:1999–2006.