Genes associated with situs inversus in patients with primary ciliary dyskinesia are identified by the weighted gene correlation network analysis

Pengcheng Xia
Shandong Provincial Hospital, Shandong First Medical University
Jing Chen
Shandong First Medical University
Yingchao Liu
Shandong Provincial Hospital, Shandong First Medical University
Xiaolin Cui
Shandong University
Cuicui Wang
Shandong Provincial Hospital, Shandong First Medical University
Shuai Zong
Shandong Provincial Hospital, Shandong First Medical University
Le Wang
Shandong Provincial Hospital, Shandong First Medical University
Zhiming Lu (luzhiming@sdu.edu.cn)
Shandong Provincial Hospital, Shandong First Medical University

Research Article

Keywords: Primary ciliary dyskinesia, Situs inversus, WGCNA, functional enrichment analysis, protein-protein interaction network, hub gene

Posted Date: April 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1550476/v1

Abstract

Background.

Primary ciliary dyskinesia (PCD) is an autosomal recessive disorder characterized by immotile and dysmotile cilia or absence of cilia. Situs inversus (SI) is one of the phenomena frequently observed in PCD. However, little is known about the association between SI with PCD.

Material and Methods.

The microarray data of PCD were retrieved from the Gene Expression Omnibus (GEO) database to construct the co-expression network by the weighted gene co-expression network analysis (WGCNA). R software package Limma was used to perform the differential analysis. The gene network modules associated with SI screened with the common genes were further annotated based on Gene Ontology (GO) database and enriched based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Gene Set Enrichment Analysis (GSEA) was applied to identify the significant pathways. CIBERSORT was used to perform the correlation analysis of immune cells. The protein-protein interaction (PPI) network was constructed based on STRING database to identify the hub genes in the network.

Results.

Genes involved in PCD were identified related to SI. The results of the correlation analysis of immune cells showed that there were significant differences in the activated contents of dendritic cells. The WGCNA analysis revealed a total of 8 hub genes (i.e., *ATP5A1, NDUFS3, NDUFV2, PSMC3, UBE2M, ALDH18A1, DBT, FOXRED1*). Based on the GO annotations of hub genes, a group of GO terms were revealed to show significant difference, including mitochondrial part, oxidoreductase complex, cytosol, mitochondrial respiratory chain complex I, NADH dehydrogenase complex, respiratory chain complex I, catalytic complex, mitochondrial envelope, respiratory chain complex, and mitochondrial respiratory chain. KEGG enrichment analysis of hub genes identified a group of metabolic pathways showing significant difference, including Propanoate metabolism, beta-Alanine metabolism, Fatty acid degradation, Valine, leucine and isoleucine degradation, Huntington disease, Carbon metabolism, Fatty acid metabolism, Lysine degradation, Butanoate metabolism, and Metabolic pathways.

Conclusion.

Our study identified the variations in immune cells in PCD patients. The hub genes provide novel therapeutic targets for the diagnosis and treatment of PCD and SI.

Introduction

Primary ciliary dyskinesia (PCD) is an autosomal recessive disorder characterized by cilia with either immotility or dysmotility or absence of cilia [1]. The defect in ciliary motion leads to anomalous mucociliary clearance, resulting in clinically recurrent or persistent sinorespiratory infections and infertility [2]. It is estimated that PCD occurs in approximately 1 out of 20,000 to 60,000 individuals in the United States, though this is likely an underestimate of the actual incidence [3]. There is considerable variability in the clinical presentation of PCD, leading frequently to delayed diagnosis. The symptoms of PCD correspond to the organs where ciliary motility is a crucial component of the normal function [4]. For example, the ciliary motility is commonly involved in both the upper and lower respiratory tracts. Neonatal respiratory distress, often attributed to transient tachypnea of the newborn or neonatal pneumonia, is frequently seen retrospectively in patients diagnosed with PCD [5]. Recurrent pneumonia and bronchiectasis can be seen in young children as well. Chronically persistent rhinosinusitis is almost universally present in patients with PCD [6]. Nasal polyps are frequently seen in these cases. Because of the abnormal ciliary functions in the eustachian tubes, children with PCD frequently have either recurrent acute otitis media or chronic serous otitis media both with the risk of conductive hearing loss. After childhood, men with PCD are infertile due to immotile spermatozoa[7]. Adult women with PCD have an approximately 50% risk of infertility due to impaired ciliary function in the fallopian tubes impairing the travel of the ovum. Similarly, women with PCD are also at increased risk for ectopic pregnancy[8].

Clinical studies have shown that PCD is associated with the situs inversus (SI) [9]. SI is a rare anomaly characterized by mirror-image location of the abdominal organs and, in most cases, the cardiac apex relative to situs solitus [10]. Genetic studies have identified extensive genetic heterogeneity underlying PCD in the past decade, when it is sometimes described as a group of disorders rather than a single disease entity. Although the molecular basis of a significant proportion of PCD cases still remains unexplained, a number of genetically stratified subgroups of patients can be identified based on ultrastructural and motility findings in their cilia. Mutations in a total of 27 genes leading to distinct ultrastructural defects have been reported [11–17]. These include mutations in structural components of the axoneme including the outer dynein arms (ODA) and their targeting and docking complexes, causing ODA defects (*DNAH5, DNAI1, DNAI2, DNAL1, NME8/TXNDC3, CCDC114*, and *ARMC4*). Some mutations affecting the components or regulators of the nexin–dynein regulatory complexes cause the microtubule disarrangements and the inner dynein arm (IDA) loss (*CCDC39* and *CCDC40*), while others not greatly disturbing the axoneme organization (*CCDC164* and *CCDC65*), which is a finding also noted for mutations affecting an ODA component gene (*DNAH11*). To date, the molecular mechanism underlying the SI phenomenon in PCD remains unclear.

In order to explore the molecular mechanism regulating the SI metastasis, differentially expressed genes (DEGs) were identified and were further annotated based on Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Database for Annotatio, Visualization and Integrated Discovery (DAVID), and Gene Eet Enrichment Analysis (GSEA) databases. By constructing the protein-protein interaction (PPI) network and using the Search Tool for the Retrieval of Interacting Genes (STRING) database and the Cytoscape software, a key module was screened from the entire PPI network with the hub genes identified based on the key module. This study identified several potentially critical biomarkers involved in the

progress of SI, providing novel insights for exploring the pathogenesis of SI and potential clinical treatment of SI.

Materials & Methods

Data Preparation

Gene expression profiles of PCD were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/; accessed on Feb 24, 2015). The dataset of GSE25186 [18] based on GPL6947 Illumina HumanHT-12 V3.0 expression beadchip contained the microarray data from 9 normal controls and 6 PCD cases (Table S1).

Differential Gene Analysis

Limma[19] is a differential expression screening method based on generalized linear models. We performed the differential analysis based on the R software package Limma (Version 3.40.6) to obtain the differential genes between different comparison groups and control groups. Specifically, we first performed the log2 transformation on the expression spectrum dataset and then used lmFit function to perform multiple linear regression analysis. We further used eBays function to compute the moderated t-statistics, moderated f-statistic, and log-odds of differential expression by empirical Bayes moderation of the standard errors towards a common value, and finally obtained the significant difference of each gene. The adjusted *P*-value was analyzed to correct the false positive results in the GEO datasets. The parameters “Adjusted $P < 0.05$ and Log (Fold Change) > 1 or Log (Fold Change) < − 1” were defined as the thresholds for the screening of differential expression of mRNAs. The box plot and heatmap were generated by the functions ggplot2 and pheatmap of the R software package, respectively.

GO Annotation and KEGG Pathway Enrichment Analysis

We performed the enrichment analyses based on KEGG rest API (https://www.kegg.jp/kegg/rest/keggapi.html) identify the metabolic pathways with significant difference. The GO annotations for genes were performed based on the R package org.hs.eg.db (Version 3.1.0) as the background and R software package clusterProfiler (Version 3.14.3) to obtain the enriched gene sets, with the minimum gene set to 5, the maximum gene set to 5000, P value set to < 0.05, and false discovery rate (FDR) set to < 0.25 considered statistically significant.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA; https://www.gseamsigdb.org/gsea/index.jsp) was applied to identify the significant pathways in dataset GSE25186. This method was used to determine the statistically significant differences between two biological states (e.g. phenotypes) in a priori defined set of genes [20]. The coefficients of Spearman correlation between genes and sample labels were defined as the weight of genes [21]. Statistical significance was assessed by comparing the enrichment score with the enrichment results generated from 1,000 random permutations of the gene sets to obtain the nominal

P values. The significant level of pathways was determined by the levels of normalized enrichment score (NES) ≥ 1.0, FDR ≤ 0.25, and $P \leq 0.05$.

Correlation analysis of immune cells

CIBERSORT [21] uses a deconvolution algorithm to estimate the composition and abundance of immune cells in a mixture of cells based on transcriptomic data. CIBERSORT performs deconvolution analysis based on the principle of Linear Support Vector regression among multiple immune infiltration databases with a relatively comprehensive category of immune cells. CIBERSORT method was used to calculate the scores of a total of 22 immune infiltrating cells in each sample.

Weighted Gene Co-expression Network Analysis

Weighted Gene Co-expression Network and co-expression modules were constructed by the Weighted Gene Co-expression Network Analysis (WGCNA) included in the WGCNA package of R [22]. The microarray data of GSE25186 were applied as a primary source of data for the WGCNA. First of all, we calculated the Median Absolute Deviation (MAD) of each gene by using the gene expression spectrum with the first 50% of the smallest MAD genes removed and the outliers of samples removed by the good Samples Genes method in the R software package WGCNA. Furthermore, WGCNA was used to construct the scale-free co-expression network. Specifically, the Pearson's correlation matrices were constructed and the average linkage analyses were performed for all pair-wise genes, Then, a weighted adjacency matrix was constructed using the power function A_mn = | C_mn | ^ β (C_mn = Pearson’s correlation between Gene_m and Gene_n; A_mn = adjacency between Gene M and Gene N), where β was a soft-thresholding parameter protecting strong correlations and penalizing weak correlations between genes. With the power of 16 selected, the adjacency was transformed into a topological overlap matrix (TOM), which was used to measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for network gene ration, while the corresponding dissimilarity (1-TOM) was also calculated. To classify genes with similar expression profiles into gene modules, the average linkage hierarchical clustering was conducted according to the TOM-based dissimilarity measured with a minimum size (gene group) of 30 for the dendrogram of genes with sensitivity to 2. To further analyze the gene modules, we calculated the dissimilarity of module eigen genes to choose the cutoff line for module dendrogram and merged modules. Ee combined the modules with a distance less than 0.35 to finally obtain a total of 24 co-expression modules. It was noted that grey module was considered as a gene set that were not assigned to any module. In order to analyze the correlation between modules and phenotypes, we transformed the classified variables into numerical variables. Specifically, patients with PCD were defined as 100 points, and those without PCD were defined as 0, while patients with SI were defined as 100 points, and those without SI were defined as 0 points. The electron microscopy results showed that lack of both outer and inner dynein arms was defined as 100 points, the presence of single absence was defined as 50 points, and the not determined was defined as 0 points. In this study, a total of 3 phenotypes were analyzed using Spearman method to calculate the correlation.

Protein-Protein Interaction (PPI) Analysis

All common genes from the selected modules were further analyzed by the online Search Tool for the Retrieval of Interacting Genes (STRING) database (Version 11.0; http://string-db.org/) to establish the network through protein-protein interaction (PPI) analysis [23]. A combined score of more than 0.4 was applied to build the PPI network, which was visualized by the Cytoscape software (version 3.8.2, http://cytoscape.org/) [24]. CytoHubba [25]is used to discover key targets and sub-networks of complex networks useful for presenting multiple types of biological data including PPI, gene regulation, cellular pathways and signal transduction. MCODE[26] (Molecular Complex Detection) was used to discover closely linked regions in PPI networks, and these regions may represent molecular complexes. The common genes in the networks were screened by the degree of the gene nodes. The genes with the most interactions were considered as hub genes, which may play important roles in the pathogenesis of the disease.

Statistical Analysis

GraphPad Prism (version 8.0.0) was utilized to perform the statistical analysis. The normality test and homogeneity of variance test were performed on data extracted from GEO datasets. Data that passed these two tests underwent t-testing for comparisons between two groups. Spearman's test was used to investigate the correlation between modules and phenotypes derived from the WGCNA. *P* values less than 0.05 were considered statistically significant.

Results

Screening of Differentially Expressed Genes (DEGs)

Based on the cutoff criteria, a total of 1109 DEGs (663 up-regulated and 446 down-regulated) were identified in the PCD samples derived from the bronchial tissues (Fig. 1A). The top 40 genes with expression levels were displayed in a heatmap (Fig. 1B, Table 1). According to the absolute value of LogFC, the top 5 up-regulated genes included *LOC654163, USP29, MOGAT3, FCGR2C*, and *SIGLEC11*, while the top 5 down-regulated genes included *SOX11, PDHA2, LOC653483, LOC644962*, and *CSF2.*

Table 1
The top 40 genes with significant differences identified in patients with primary ciliary dyskinesia.

Gene	logFC	AveExpr	t	*P*–Value	B	Regulation
SOX11	−6.104	−1.217	−3.043	0.043	−4.580	Down
PDHA2	−6.091	1.532	−5.787	0.005	−4.575	Down
LOC653483	−5.420	0.359	−3.806	0.022	−4.578	Down
LOC644962	−5.016	0.787	−4.518	0.004	−4.563	Down
CSF2	−4.718	0.063	−4.959	0.003	−4.562	Down
LOC342931	−4.513	1.649	−5.886	0.001	−4.502	Down
LOC652513	−4.461	−0.377	−4.008	0.019	−4.577	Down
MLN	−4.374	0.075	−5.505	0.003	−4.567	Down
MBD3L1	−4.324	−0.030	−3.594	0.044	−4.585	Down
C11orf76	−4.318	−1.471	−3.453	0.015	−4.556	Down
LOC642506	−4.297	0.835	−3.184	0.021	−4.570	Down
LOC340268	−4.282	1.223	−2.843	0.039	−4.577	Down
FAM155A	−4.249	1.313	−3.818	0.014	−4.564	Down
OR1L6	−4.238	−0.422	−4.457	0.013	−4.576	Down
TNNT1	−4.226	−0.951	−4.507	0.007	−4.561	Down
FLJ25328	−4.202	−0.527	−3.493	0.047	−4.585	Down
HIST1H2AD	−4.031	0.310	−5.279	0.004	−4.568	Down
ZNF677	−4.025	0.903	−4.327	0.008	−4.562	Down
LOC647147	−3.835	0.114	−3.500	0.019	−4.573	Down
SLC7A13	−3.764	0.788	−3.979	0.008	−4.566	Down
ANKS4B	4.698	0.797	3.905	0.037	−4.585	Up
LOC645689	4.721	0.151	2.699	0.037	−4.574	Up
THY1	4.907	2.806	2.933	0.019	−4.548	Up
TCOF1	4.921	−0.294	7.256	0.004	−4.539	Up
FLJ23356	4.945	0.733	3.174	0.038	−4.580	Up
TRHR	4.977	−0.250	3.921	0.006	−4.541	Up

Gene	logFC	AveExpr	t	*P*–Value	B	Regulation
KCNA7	4.995	1.000	4.152	0.032	−4.585	Up
FLJ44216	5.066	2.267	5.836	0.002	−4.567	Up
ARFIP2	5.179	1.762	6.229	0.002	−4.566	Up
LOC643083	5.240	−1.272	4.356	0.028	−4.584	Up
C22orf31	5.265	0.447	4.376	0.028	−4.584	Up
FAM55D	5.427	−0.377	3.534	0.018	−4.566	Up
LOC158730	5.463	0.287	6.145	0.004	−4.574	Up
MYOG	5.475	0.162	4.551	0.025	−4.584	Up
ASB16	5.630	1.917	3.429	0.021	−4.573	Up
LOC654163	5.653	0.350	4.699	0.023	−4.584	Up
USP29	5.681	0.014	3.511	0.028	−4.579	Up
MOGAT3	5.686	0.461	5.911	0.002	−4.567	Up
FCGR2C	5.708	0.150	4.745	0.023	−4.584	Up
SIGLEC11	6.519	1.146	7.058	0.001	−4.565	Up

GO Annotation and KEGG Pathway Enrichment Analysis

To further explore the functions of the potential target genes, the DEGs were analyzed by functional enrichment analyses (Fig. 1C, D; Table S2). GO analysis revealed a group of GO terms with significant difference, including cellular response to amyloid-beta, detection of stimulus, sensory perception, sensory perception of chemical stimulus, detection of chemical stimulus, response to amyloid-beta, sensory perception of smell, system process, detection of chemical stimulus involved in sensory perception, and detection of stimulus involved in sensory perception. KEGG analysis revealed the following metabolic pathways with significant difference: Olfactory transduction, Malaria, Alcoholism, Pentose and glucuronate interconversions, Central carbon metabolism in cancer, HIF-1 signaling pathway, Systemic lupus erythematosus, Renin-angiotensin system, Hepatitis B, PI3K-Akt signaling pathway, and EGFR tyrosine kinase inhibitor resistance (Table S2).

Gene Set Enrichment Analysis (GSEA)

The results of the Gene Set Enrichment Analysis'[20, 27] of the microarray data showed that gene set *CHR9P24* was down-regulated, while the gene set *KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG* was up-regulated (Fig. 2).

Correlation analysis of immune cells

The results of CIBERSORT analysis showed that there were significant differences in the activated contents of dendritic cells (Fig. 3A, B). Correlation heat map showed that Plasma cells was positively correlated with Mast cells resting, the T cells CD4 memory resting was negatively correlated with T cells CD8, while the dendritic cells resting was negatively correlated with T cells CD4 memory resting (Fig. 3C).

WGCNA and Key Module Identification

All genes in the microarray dataset were used to conduct the WGCNA (Fig. 4). A total of 24 co-expression modules were identified (Fig. 4C; Table S3). The heatmap displaying the relationships between modules of WGCNA was shown in Fig. 5A. The relationships between modules and genes in the modules are shown in Fig. 5B. The number of genes in each module was provided in Table S4. Based on the heatmap of module-trait correlations, the key module containing a total of 87 genes was the most positively correlated with Situs inversus (Fig. 6A, B). These genes were further annotated based on KEGG and GO databases (Fig. 6C, D). GO analysis of the hub genes revealed the following GO terms with significant difference: mitochondrial part, oxidoreductase complex, cytosol, mitochondrial respiratory chain complex I, NADH dehydrogenase complex, respiratory chain complex I, catalytic complex, mitochondrial envelope, respiratory chain complex, and mitochondrial respiratory chain. The KEGG analysis of hub genes revealed a total of 10 metabolic pathways with significant difference, i.e., Propanoate metabolism, beta-Alanine metabolism, Fatty acid degradation, Valine, leucine and isoleucine degradation, Huntington disease, Carbon metabolism, Fatty acid metabolism, Lysine degradation, Butanoate metabolism, and Metabolic pathways.

PPI Network Construction and Identification of Hub Genes

All the common genes from the selected modules were further analyzed by the STRING database to construct the PPI network (Fig. 7A). Based on the results of PPI analysis, Cytoscape was used for visualization and the plug-in CytoHubba was used to screen the hub genes. The top 10 genes were identified as hub genes (Fig. 7B; Table S4), including *ATP5A1, NDUFS3, NDUFV2, PSMC3, UBE2M, ALDH18A1, DBT, FOXRED1, GNB2L1*, and *NDUFB10*. Then, MCODE was used to identify the key subnetworks and genes. A total of 6 key genes were identified, including *ATP5A1, NDUFV2, FOXRED1, NDUFS3, ATP5I*, and *NDUFB10* (Fig. 7C).

Discussion

PCD is a rare genetic disease characterized by recurrent respiratory tract infections, bronchiectasis, and infertility [18]. SI is often associated with PCD [28]. The first internationally randomized controlled trial in PCD was recently conducted showing azithromycin is effective in reducing exacerbations [29]. Although some cases of SI have been reported [30], the causes of SI remain unclear and the molecular studies of SI are even more sparse. With the rapid advancement of bioinformatics tools, e.g., the WGCNA, we intended to identify the molecules involved in the formation of SI.

We first identified the DEGs involved in PCD by classical analytical methods. Then, these DEGs were annotated and enriched in a group of GO terms and KEGG metabolic pathways, respectively. Studies have shown that many of these enriched pathways are involved in PCD [31–33]. Furthermore, our results of GSEA revealed that the gene set *KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG* showed significant difference, suggesting that PCD might be related to the maturity onset diabetes of the young. These results were consistent with those reported previously, showing that both disorders of SI and PCD occur more frequently in children [34, 35]. Because PCD is related to respiratory infections [36], and the quality of immune function is related to the severity of PCD [37]. Our results of immune cells showed that there were significant differences in the amount of activated contents of dendritic cells in PCD, and the correlation analysis revealed either positive or negative effects of these immune cells, which was important information for the assessment of the immune functions of PCD [38]. Studies on SI have shown that the Bi-allelic mutations in *PKD1L1* are associated with the laterality defects in humans [39], while the *DNAI1* gene defects are associated with PCD and situs solitus [40]. Furthermore, two out of the thirteen novel variants, i.e., (1) c.11839 + 1G > A indynein, axonemal, heavy chain 11 (*DNAH11*) and (2) p.Lys92Trpfs in dynein, axonemal, intermediate chain 1 (*DNAI1*), were associated with dextrocardia with SI, while one variant, i.e., p.Gly21Val in coiled-coil domain-containing protein 40 (*CCDC40*), was involved in the absence of the inner dynein arm. Moreover, the homozygous C1orf127:p.Arg113Ter (rs558323413) was also associated with the laterality defects in two related patients [41]. In addition, it has been reported that *CCDC11* is associated with laterality disorder [42], while a homozygous *NME7* mutation was associated with Situs Inversus Totalis [43]. Although the discovery of these genes partly explains the occurrence of SI, the genes involved in the formation of SI still need to be further explored.

The molecular modules most closely related to SI were identified based on the WGCNA. The functional analysis of these molecules revealed a group of GO terms and KEGG pathways related to situs inversus. Studies have shown that imbalanced mitochondrial function provokes heterotaxy (i.e., the incorrect placement of visceral organs) via aberrant ciliogenesis, while faulty cilia also prevent the development of proper left-right asymmetry to cause heterotaxy [44]. Furthermore, studies have reported that aberrant left-right patterning in the developing human embryo can lead to a broad spectrum of congenital malformations [45]. About 50% of the SI cases are possibly caused by an inability of embryonic cilia to shift the heart to the left side. Based on the observation of electron microscopy, the missing ciliary components can be visualized directly to determine the heterogeneous nature of this disease based on the molecular basis of this congenital defect [46]. To date, our study is the first report with the identification of these KEGG pathways involved in situs inversus, providing a new direction for the study of the causes of SI. Furthermore, a group of 6 core genes (i.e., *ATP5A1, NDUFV2, FOXRED1, NDUFS3, ATP5I*, and *NDUFB10*) closely related to SI were identified based on Cytoscape, indicating that these genes might be the potential targets for the treatment of SI.

Gene regulatory network plays an important role in the pathophysiological process of SI. Our results provide novel insights into the understanding of the pathogenesis of SI and provide potentially novel treatment strategies for this medical disorder. It is noted that our study contains the following limitations: (1) only the top 10 hub genes were comprehensively investigated; (2) it is challenging to verify our

conclusions in this study due to the difficulty in obtaining specimens of SI; (3) the functions of hub genes identified in the network constructed in this study need to be further verified.

Declarations

• Ethics approval and consent to participate

Not necessary.

• Consent for publication

Not applicable.

• Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Gene Expression Omnibus (GEO) repository, [GSE28146].

• Competing interests

The authors declare that they have no competing interests.

• Funding

This research was funded by the Key Research and Development Project of Shandong Province (2022CXGC010507) and Tai 'an City Science and technology Innovation development Project (2021NS261). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

• Authors' contributions

Pengcheng Xia and Jing Chen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

Yingchao Liu, Xiaolin Cui, Cuicui Wang, Shuai Zong and Le Wang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Zhiming Lu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

• Acknowledgements

Not applicable.

References

1. Andjelkovic M, Minic P, Vreca M, Stojiljkovic M, Skakic A, Sovtic A, Rodic M, Skodric-Trifunovic V, Maric N, Visekruna J *et al*: **Genomic profiling supports the diagnosis of primary ciliary dyskinesia and reveals novel candidate genes and genetic variants**. PloS one 2018, **13**(10):e0205422.
2. Djakow J, Kramná L, Dušátková L, Uhlík J, Pursiheimo J, Svobodová T, Pohunek P, Cinek O: **An effective combination of sanger and next generation sequencing in diagnostics of primary ciliary dyskinesia**. Pediatric pulmonology 2016, **51**(5):498–509.
3. Butterfield R: **Primary Ciliary Dyskinesia**. Pediatrics in review 2017, **38**(3):145–146.
4. Postema M, Carrion-Castillo A, Fisher S, Vingerhoets G, Francks C: **The genetics of situs inversus without primary ciliary dyskinesia**. Scientific reports 2020, **10**(1):3677.
5. Horani A, Ferkol T: **Understanding Primary Ciliary Dyskinesia and Other Ciliopathies**. The Journal of pediatrics 2021, **230**:15–22.e11.
6. Guo Z, Chen W, Wang L, Qian L: **Clinical and Genetic Spectrum of Children with Primary Ciliary Dyskinesia in China**. The Journal of pediatrics 2020, **225**:157–165.e155.
7. Sironen A, Shoemark A, Patel M, Loebinger M, Mitchison H: **Sperm defects in primary ciliary dyskinesia and related causes of male infertility**. Cellular and molecular life sciences: CMLS 2020, **77**(11):2029–2048.
8. Guerri G, Maniscalchi T, Barati S, Dhuli K, Busetto G, Del Giudice F, De Berardinis E, De Antoni L, Miertus J, Bertelli M: **Syndromic infertility**. *Acta bio-medica: Atenei Parmensis* 2019, **90**:75–82.
9. Burwick R, Govindappagari S, Sanchez-Lara P: **Situs inversus totalis and prenatal diagnosis of a primary ciliary dyskinesia**. Journal of clinical ultrasound: JCU 2021, **49**(1):71–73.
10. Fulcher A, Turner M: **Abdominal manifestations of situs anomalies in adults**. Radiographics: a review publication of the Radiological Society of North America, Inc 2002, **22**(6):1439–1456.
11. Knowles M, Ostrowski L, Loges N, Hurd T, Leigh M, Huang L, Wolf W, Carson J, Hazucha M, Yin W *et al*: **Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms**. American journal of human genetics 2013, **93**(4):711–720.
12. Zariwala M, Gee H, Kurkowiak M, Al-Mutairi D, Leigh M, Hurd T, Hjeij R, Dell S, Chaki M, Dougherty G *et al*: **ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6**. *American journal of human genetics* 2013, **93**(2):336–345.
13. Tarkar A, Loges N, Slagle C, Francis R, Dougherty G, Tamayo J, Shook B, Cantino M, Schwartz D, Jahnke C *et al*: **DYX1C1 is required for axonemal dynein assembly and ciliary motility**. Nature genetics 2013, **45**(9):995–1003.
14. Moore D, Onoufriadis A, Shoemark A, Simpson M, zur Lage P, de Castro S, Bartoloni L, Gallone G, Petridi S, Woollard W *et al*: **Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia**. American journal of human genetics 2013, **93**(2):346–356.

15. Kott E, Legendre M, Copin B, Papon J, Dastot-Le Moal F, Montantin G, Duquesnoy P, Piterboth W, Amram D, Bassinet L *et al*: **Loss-of-function mutations in RSPH1 cause primary ciliary dyskinesia with central-complex and radial-spoke defects**. American journal of human genetics 2013, **93**(3):561–570.
16. Horani A, Brody S, Ferkol T, Shoseyov D, Wasserman M, Ta-shma A, Wilson K, Bayly P, Amirav I, Cohen-Cymberknoh M *et al*: **CCDC65 mutation causes primary ciliary dyskinesia with normal ultrastructure and hyperkinetic cilia**. PloS one 2013, **8**(8):e72299.
17. Austin-Tse C, Halbritter J, Zariwala M, Gilberti R, Gee H, Hellman N, Pathak N, Liu Y, Panizzi J, Patel-King R *et al*: **Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia**. American journal of human genetics 2013, **93**(4):672–686.
18. Geremek M, Ziętkiewicz E, Bruinenberg M, Franke L, Pogorzelski A, Wijmenga C, Witt M: **Ciliary genes are down-regulated in bronchial tissue of primary ciliary dyskinesia patients**. PloS one 2014, **9**(2):e88216.
19. Ritchie M, Phipson B, Wu D, Hu Y, Law C, Shi W, Smyth G: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. Nucleic acids research 2015, **43**(7):e47.
20. Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, Paulovich A, Pomeroy S, Golub T, Lander E *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. Proceedings of the National Academy of Sciences of the United States of America 2005, **102**(43):15545–15550.
21. Suárez-Fariñas M, Lowes M, Zaba L, Krueger J: **Evaluation of the psoriasis transcriptome across different studies by gene set enrichment analysis (GSEA)**. PloS one 2010, **5**(4):e10247.
22. Luo Y, Coskun V, Liang A, Yu J, Cheng L, Ge W, Shi Z, Zhang K, Li C, Cui Y *et al*: **Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells**. Cell 2015, **161**(5):1175–1186.
23. Szklarczyk D, Morris J, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva N, Roth A, Bork P *et al*: **The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible**. Nucleic acids research 2017, **45**:D362-D368.
24. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. Genome research 2003, **13**(11):2498–2504.
25. Chin C, Chen S, Wu H, Ho C, Ko M, Lin C: **cytoHubba: identifying hub objects and sub-networks from complex interactome**. BMC systems biology 2014:S11.
26. Palukuri M, Marcotte E: **Super.Complex**: **A supervised machine learning pipeline for molecular complex detection in protein-interaction networks**. *bioRxiv: the preprint server for biology* 2021.
27. Mootha V, Lindgren C, Eriksson K, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E *et al*: **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes**. Nature genetics 2003, **34**(3):267–273.

28. Leigh M, Pittman J, Carson J, Ferkol T, Dell S, Davis S, Knowles M, Zariwala M: **Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome**. Genetics in medicine: official journal of the American College of Medical Genetics 2009, **11**(7):473–487.
29. Shoemark A, Harman K: **Primary Ciliary Dyskinesia**. Seminars in respiratory and critical care medicine 2021, **42**(4):537–548.
30. de Castro C, Dos Reis F, de Carvalho G, Fernandes L, Abdalla L, Samano M, Júnior J, Pêgo-Fernandes P: **Technical Challenges in Lung Transplantation of Kartagener Syndrome Recipients**: **A Unique Team Experience With 12 Patients**. *Transplantation proceedings* 2020, **52**(5):1384–1387.
31. Saransh J, Vikas D: **Auditory processing disorders associated with a case of Kartagner's syndrome**. Intractable & rare diseases research 2014, **3**(1):19–24.
32. Kobayashi D, Asano-Hoshino A, Nakakura T, Nishimaki T, Ansai S, Kinoshita M, Ogawa M, Hagiwara H, Yokoyama T: **Loss of zinc finger MYND-type containing 10 (zmynd10) affects cilia integrity and axonemal localization of dynein arms, resulting in ciliary dysmotility, polycystic kidney and scoliosis in medaka (Oryzias latipes)**. Developmental biology 2017, **430**(1):69–79.
33. Laclef C: **[Primary cilia control different steps of brain development]**. Medecine sciences: M/S 2014, **30**(11):980–990.
34. Mirra V, Werner C, Santamaria F: **Primary Ciliary Dyskinesia: An Update on Clinical Aspects, Genetics, Diagnosis, and Future Treatment Strategies**. Frontiers in pediatrics 2017, **5**:135.
35. Anık A, Çatlı G, Abacı A, Böber E: **Maturity-onset diabetes of the young (MODY): an update**. Journal of pediatric endocrinology & metabolism: JPEM 2015, **28**:251–263.
36. Boon M, De Boeck K, Jorissen M, Meyts I: **Primary ciliary dyskinesia and humoral immunodeficiency–is there a missing link?** Respiratory medicine 2014, **108**(6):931–934.
37. Speert D: **Bacterial infections of the lung in normal and immunodeficient patients**. *Novartis Foundation symposium* 2006, **279**:42–51; disussion 51 – 45, 216–219.
38. Cockx M, Gouwy M, Ruytinx P, Lodewijckx I, Van Hout A, Knoops S, Pörtner N, Ronsse I, Vanbrabant L, Godding V *et al*: **Monocytes from patients with Primary Ciliary Dyskinesia show enhanced inflammatory properties and produce higher levels of pro-inflammatory cytokines**. Scientific reports 2017, **7**(1):14657.
39. Vetrini F, D'Alessandro L, Akdemir Z, Braxton A, Azamian M, Eldomery M, Miller K, Kois C, Sack V, Shur N *et al*: **Bi-allelic Mutations in PKD1L1 Are Associated with Laterality Defects in Humans**. American journal of human genetics 2016, **99**(4):886–893.
40. Guichard C, Harricane M, Lafitte J, Godard P, Zaegel M, Tack V, Lalau G, Bouvagnet P: **Axonemal dynein intermediate-chain gene (DNAI1) mutations result in situs inversus and primary ciliary dyskinesia (Kartagener syndrome)**. American journal of human genetics 2001, **68**(4):1030–1035.
41. Alsamri M, Alabdouli A, Iram D, Alkalbani A, Almarzooqi A, Souid A, Vijayan R: **A Study on the Genetics of Primary Ciliary Dyskinesia**. Journal of clinical medicine 2021, **10**(21).
42. Perles Z, Cinnamon Y, Ta-Shma A, Shaag A, Einbinder T, Rein A, Elpeleg O: **A human laterality disorder associated with recessive CCDC11 mutation**. Journal of medical genetics 2012, **49**(6):386–390.

43. Reish O, Aspit L, Zouella A, Roth Y, Polak-Charcon S, Baboushkin T, Benyamini L, Scheetz T, Mussaffi H, Sheffield V *et al*: **A Homozygous Nme7 Mutation Is Associated with Situs Inversus Totalis**. Human mutation 2016, **37**(8):727–731.

44. Burkhalter M, Sridhar A, Sampaio P, Jacinto R, Burczyk M, Donow C, Angenendt M, Hempel M, Walther P, Pennekamp P *et al*: **Imbalanced mitochondrial function provokes heterotaxy via aberrant ciliogenesis**. The Journal of clinical investigation 2019, **129**(7):2841–2855.

45. Li A, Hanchard N, Azamian M, D'Alessandro L, Coban-Akdemir Z, Lopez K, Hall N, Dickerson H, Nicosia A, Fernbach S *et al*: **Genetic architecture of laterality defects revealed by whole exome sequencing**. European journal of human genetics: EJHG 2019, **27**(4):563–573.

46. Afzelius B: **Genetical and ultrastructural aspects of the immotile-cilia syndrome**. American journal of human genetics 1981, **33**(6):852–864.

Figures

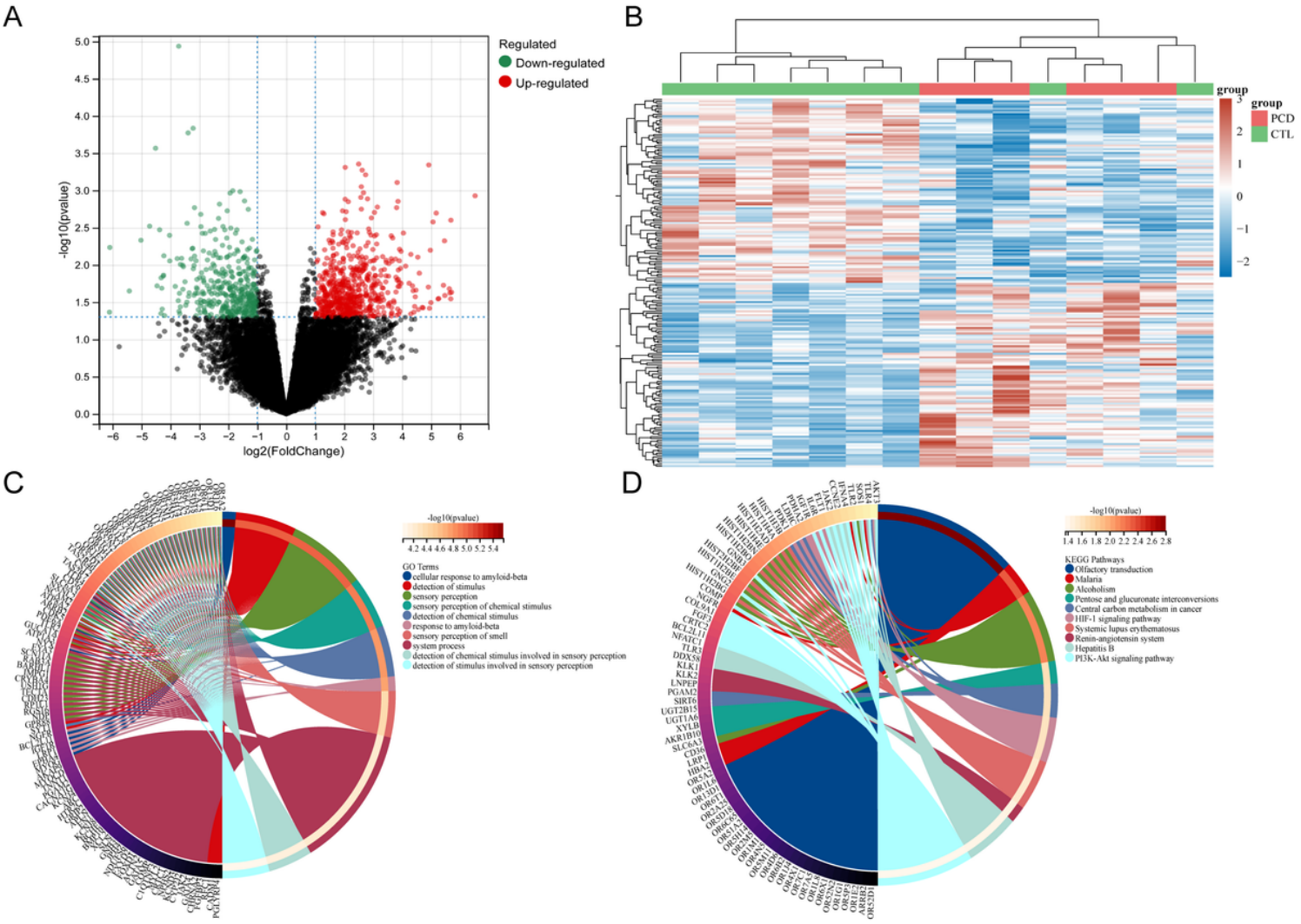

Figure 1

Differential genes and GO/KEGG plots. A. The volcanic map of differential genes. **B.** Heat maps of the top 50 genes with multiple differences. **C.** GO analysis of differential genes shown in a circle diagram. **D.** KEGG analysis of differential genes shown in a circle diagram.

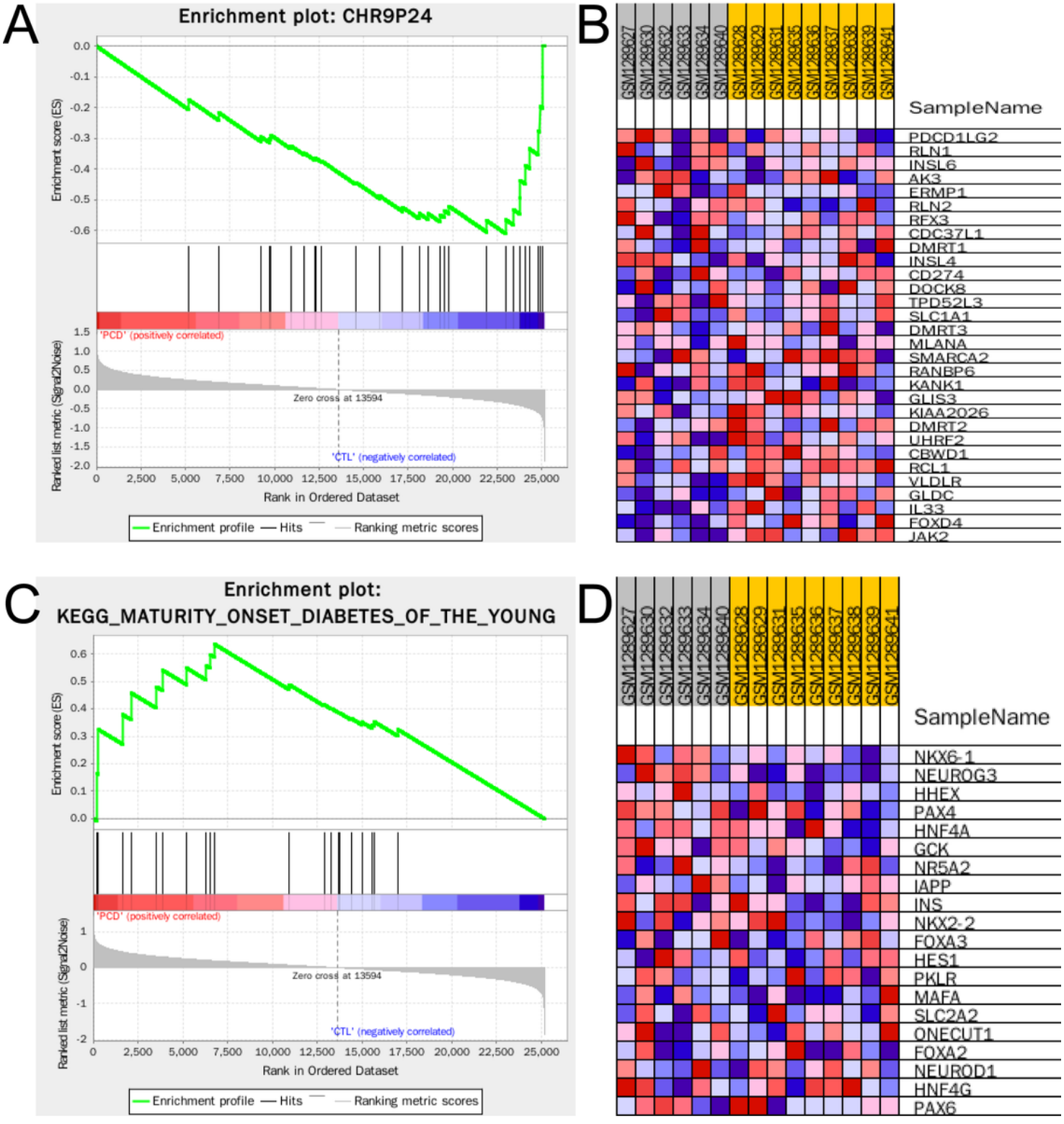

Figure 2

Results of GSEA analysis. A. Enrichment plot of gene set *CHR9P24*. **B.** Heatmap of gene set *CHR9P24*. **C.** Enrichment plot of gene set *KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG*. **D.** Heatmap of gene set *KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG*.

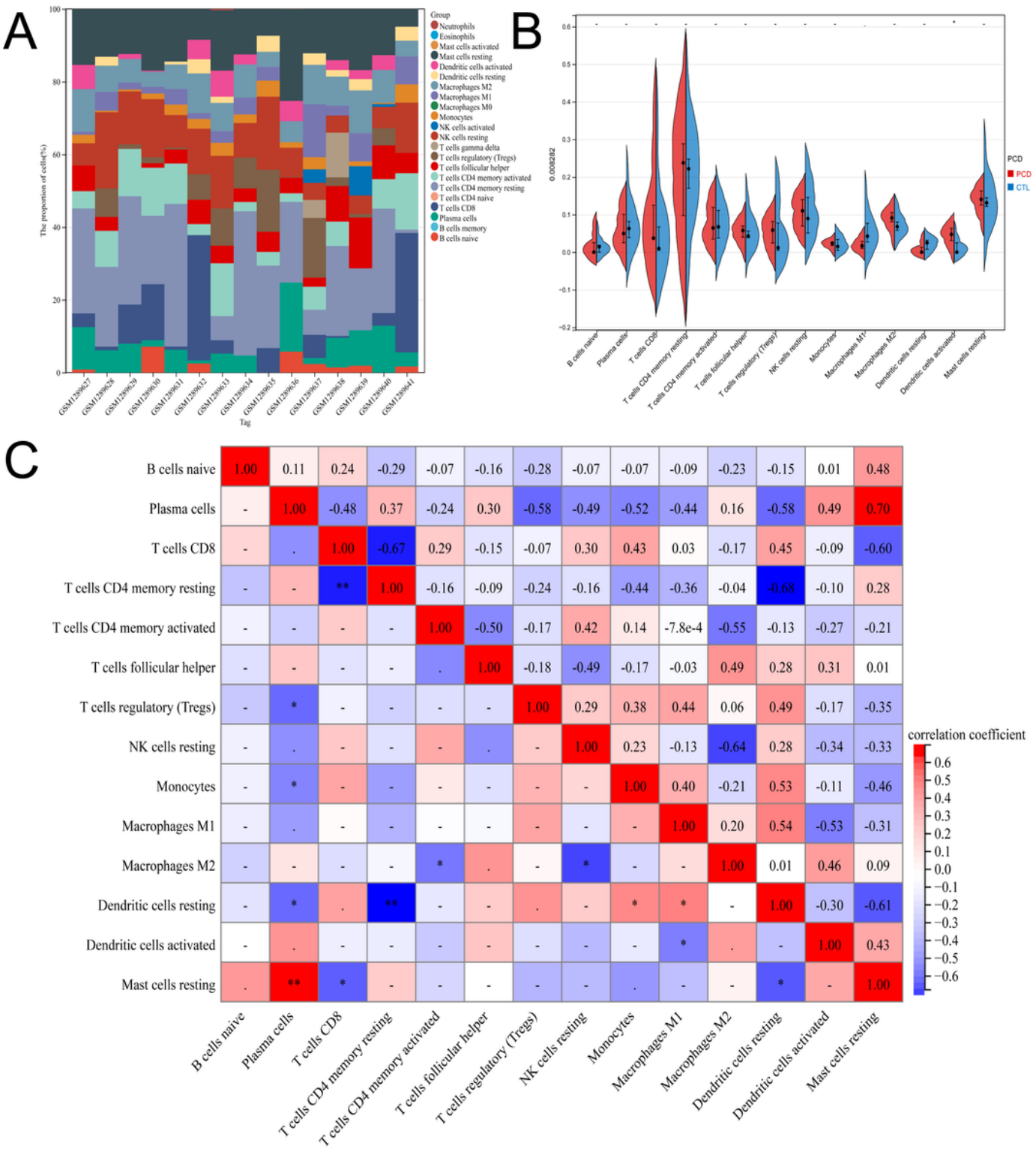

Figure 3

Correlation analysis of immunity. A. Diagram of the composition of 22 immune cells. **B.** Comparison of various types of immune cells. **C.** Heat map of correlations of various immune cells.

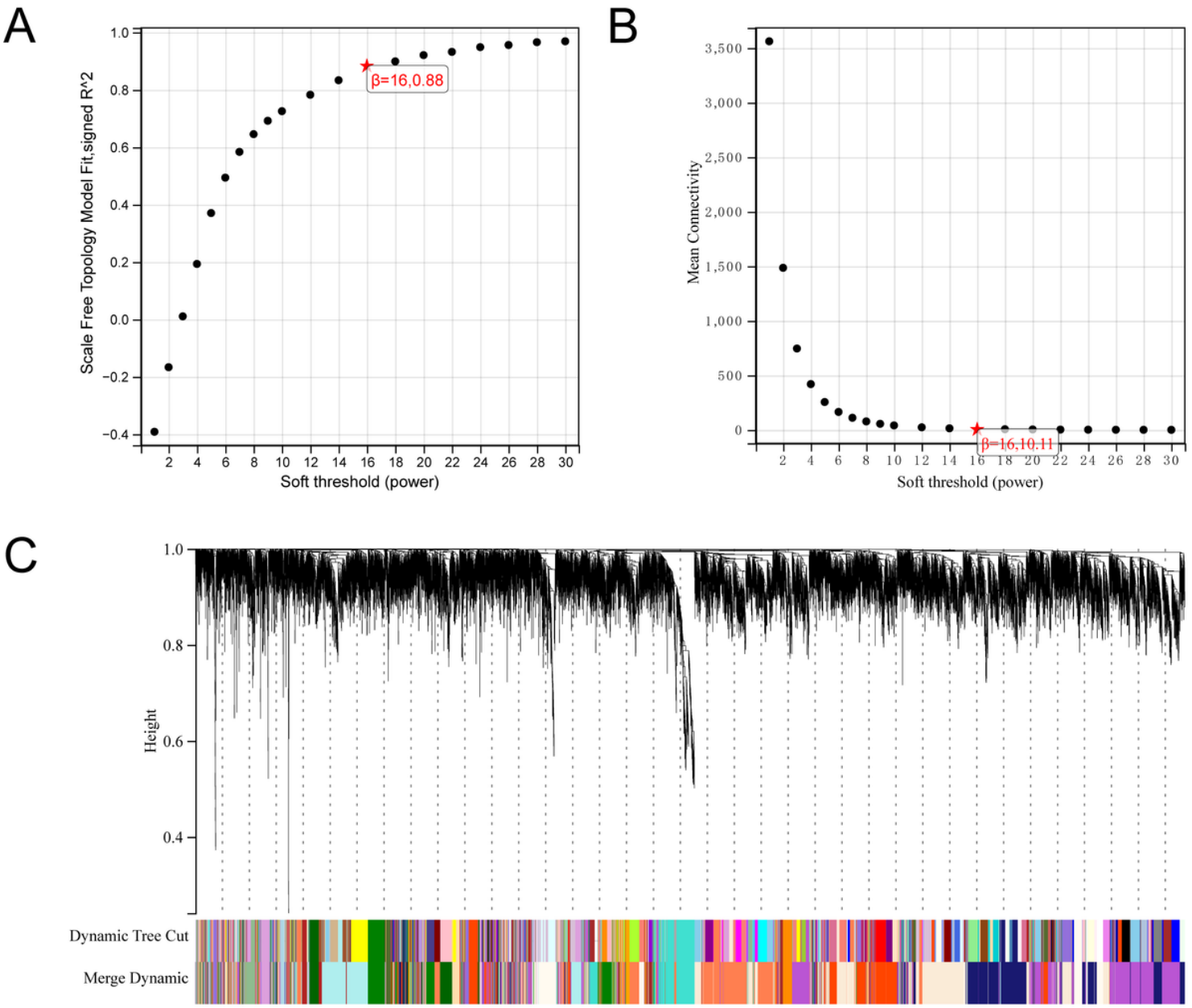

Figure 4

WGCNA analysis of genes. A. Soft threshold (power) of the network. **B.** Mean Connectivity of the network. **C.** Gene cluster diagram of the network.

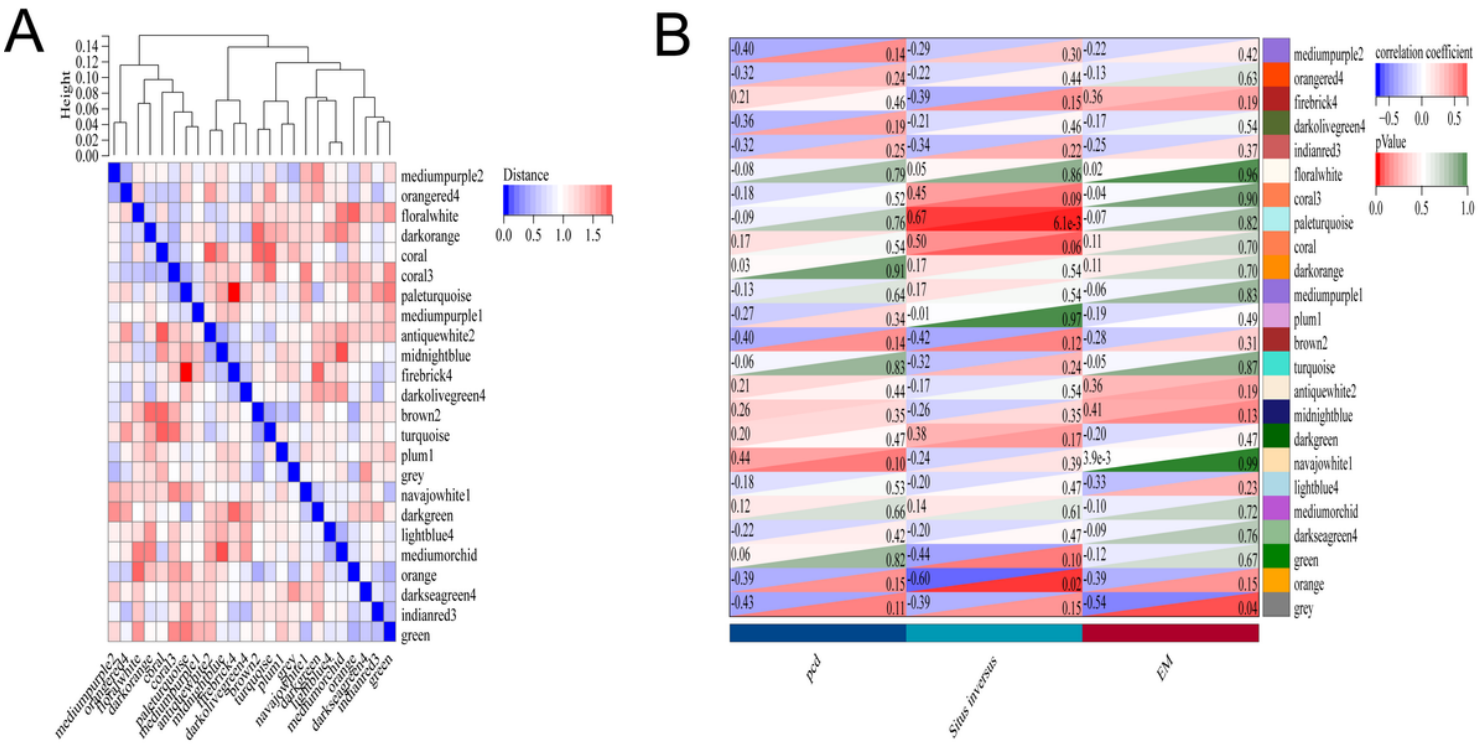

Figure 5

Module cluster heat map and clinical relationship heat map.

A. Module feature vector clustering diagram. **B.** Heat map of correlations between modules and phenotypes.

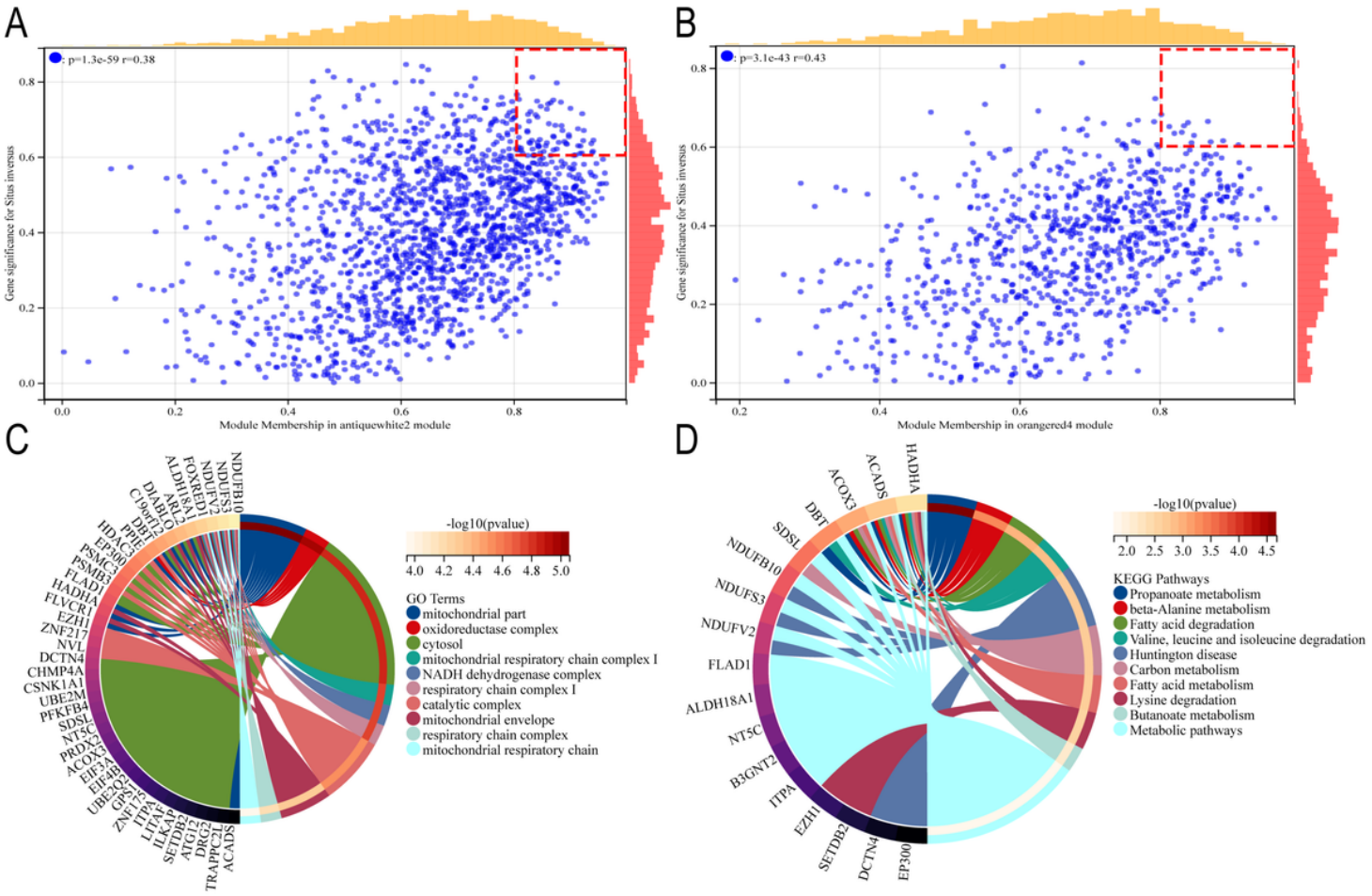

Figure 6

Screening and GO and KEGG analysis of hub genes. A. The hub genes of antiquewhite2 module. **B.** The hub genes of orangered4 module. **C.** The GO analysis of hub genes. **D.** The KEGG analysis of hub genes.

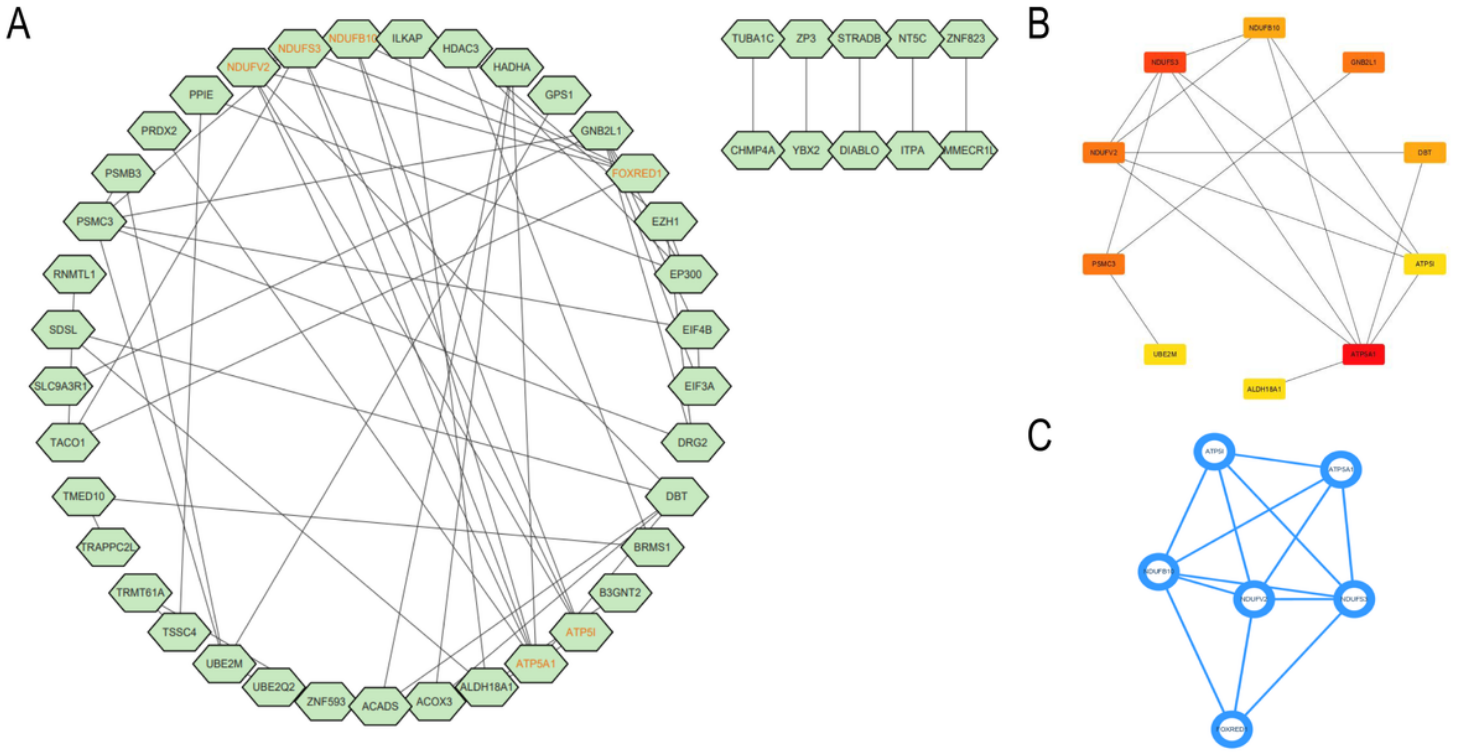

Figure 7

Hub genes of PPI network. A. PPI network diagram of hub genes. **B.** CytoHubba analysis showing 10 hub genes. **C.** MCODE analysis showing 6 hub genes.

Supplementary Files

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

- TableS1.GSE25186.matrix.xlsx
- TableS2.KEGGandKEGGanalysisdata.xlsx
- TableS3.wgcnamodule.xlsx
- TableS4.HUBgene.xlsx
- TableS5.ElectronmicroscopyNOmeasurementsandsitusstatusinthePCDpatientsstudiedrevised.docx