Integrated Profiling Identifies Key Molecule Drivers as Diagnostic Biomarkers and Therapeutic Targets in Acute Kidney Injury

Tao Sun
Zhejiang University School of Medicine Second Affiliated Hospital
Ying Cao
Zhejiang University School of Medicine Second Affiliated Hospital
Tiancha Huang
Zhejiang University School of Medicine Second Affiliated Hospital
Yiwen Sang
Zhejiang University School of Medicine Second Affiliated Hospital
Yibei Dai
Zhejiang University School of Medicine Second Affiliated Hospital
Zhihua Tao (zrtzh@zju.edu.cn)
Zhejiang University School of Medicine Second Affiliated Hospital https://orcid.org/0000-0001-7866-8463

Research

Keywords: acute kidney injury, WGCNA, GEO, diagnosis, LASSO, Transcriptional factors, miRNA, immune infiltration.

Posted Date: October 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-970753/v1

Abstract

Background

Acute kidney injury is a common clinical problem with no sensitive and specific diagnostic biomarkers and definitive treatments. The underlying molecular mechanisms of acute kidney injury are unclear. Therefore, it is pivotal and urgent to explore the underlying mechanisms and screen for novel diagnostic biomarkers, as well as therapeutic targets.

Methods

The present study constructed scale-free network using WGCNA analysis. LASSO logistic regression analysis was used to explore the optimal diagnostic model of AKI. In addition, GO and KEGG pathway enrichment analysis were performed and TF-mRNA and miRNA-mRNA network analysis and immune infiltration analysis of hub genes were performed to reveal the underlying mechanisms of AKI.

Results

Fifteen hub genes were uncovered and constructed a diagnostic model by LASSO logistic regression analysis. GO and KEGG analysis revealed that the genes were enriched in oxidation-reduction process, cell adhesion, proliferation, migration, metabolic process, mitochondria and iron ion binding. The enriched TFs were BRD2, EP300, ETS1, MYC, SPI1 and ZNF263. The enriched miRNAs were miR-181c-5p, miR-218-5p, miR-485-5p, miR-532-5p and miR-6884-5p. The immune infiltration analysis showed that Macrophages M2 was decreasing significantly revealing a protective factor for further AKI treatment.

Conclusions

The present study identified fifteen hub genes, a diagnostic model, transcriptional factors, miRNAs, immune infiltration and pathways by analyzing gene expression profiles of AKI, which provides some basis for further experimental studies.

1. Introduction

Acute kidney injury is a common clinical syndromes of acute deterioration or even loss of kidney function due to different causes such as sepsis, cardiac surgery, trauma, contrast medium and nephrotoxic drug[1]. It is a vital complication in patients admitted to hospital where the morbidity is almost 10-15% of all hospitalizations and in patients in the ICU where the prevalence can sometimes be more than 50%[2, 3]. AKI can progress to chronic kidney disease easily and rapidly without effective clinical treatment. AKI has the characteristics of high morbidity and mortality and there is no effective treatment strategy[4]. AKI

will bring high medical expenses if renal dialysis or renal transplant is performed. These issues have contributed AKI to a major public health problem worldwide[5].

Thus, the early, sensitive and rapid diagnosis and intervention of AKI is an important part of the overall management of patients with the various syndromes which cause or are associated with AKI. However, serum creatinine and urinary output are the cornerstone of our current diagnostic approach, which are neither sensitive nor specific for AKI[6]. Although eliminating the short- and long-term outcomes of AKI is urgent, early diagnosis and therapeutic detection biomarkers are still limited[7]. In addition, there are no clinically valid biomarkers for the early diagnosis and progression detection of renal fibrosis after AKI. Therefore, it is urgent and imperative to develop novel and valid therapeutic therapies and find out the potential biomarkers for the treatment of renal fibrosis and delaying the progression[8, 9].

Thanks to the development of technologies on microarray and high-throughput sequencing, an increasing number of biomarkers and therapeutic targets have been discovered and applied in clinical practice[10, 11]. The Gene Expression Omnibus (GEO) database provides an opportunity for bioinformatics mining of the gene expression profiles in different diseases. Selecting the optimal datasets and conducting the interactive analysis on the targeted datasets. Differentially expressed genes (DEGs) and molecular mechanisms relevant to the development of AKI were evaluated by bioinformatics methods. The Weighted gene co-expression network analysis (WGCNA) is widely applied to screen and identify highly collaborative modules and find hub modules of highly correlated genes[12]. In addition, WGCNA has been widely utilized to the omics analysis to mine potential biomarkers and therapeutic targets which provides a useful interpretation tool for various diseases and has brought new insights into making out the molecular pathogenesis[13–15].

In order to explore the underlying mechanisms and to select and identify potential diagnostic biomarkers and therapeutic targets of AKI, in this present study, we performed a well-established bioinformatics methods of WGCNA to analyze the differentially expression genes of expression profile in GEO database (NCBI GEO serial number: GSE139061). We identified key pathways by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, along with constructing Protein-Protein Interaction network to identify hub genes and TF-mRNA network, miRNA-mRNA network. By using these analytical methods, the aim of this present study is providing new insight into the pathogenesis and development of AKI and identifying potential biomarkers and therapeutic targets.

2. Materials And Methods

2.1 Preparation of gene expression profile data

The workflow of this present study was showed in Fig. S1. The gene expression profile data was obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). GSE139061 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139061) was gene expression profile dataset based on the platform of GPL20301 (Illumina HiSeq 4000) which contained 39 native human renal

biopsy samples and 9 reference nephrectomies. GSE30718 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30718) was gene expression profile dataset based on the platform of GPL570 (Affymetrix Human Genome U133 plus 2.0 Array) which contained 28 native human renal biopsy samples and 11 reference nephrectomies. The two gene expression datasets was downloaded with the database were calculated as fragments per kilo base of transcript per million mapped reads. Data standardization was performed by using the multi-array average algorithm in the affy package in Bioconductor (https://www.bioconductor.org) in R 3.6.2 (https://cran.rstudio.com/). GSE139061 was regarded as training dataset and GSE30718 was regarded as verifying dataset for further analysis.

2.2 Identification of differentially expressed genes (DEGs)

Limma R package (Version:3.42.2) was used to screen the differentially expressed genes between AKI cohort and non-AKI cohort with R 3.6.2 (https://cran.rstudio.com/). Log(2)(|FoldChange|) higher than 1 and adjusted P value less than 0.05 were regarded as screening criteria of DEGs. A hierarchical cluster heatmap based on Euclidean distance was generated using the pheatmap R package (Version:1.0.12) and represented the gene expression intensity and direction of differentially expressed genes. A volcano plot displayed the distribution of DEGs which log(2)(|FoldChange|) was calculated and showed in the horizontal axis, -log(10)(P-value) was calculated and showed in the longitudinal axis.

2.3 Construction of Weighted Gene Co-expression Network Analysis (WGCNA)

WGCNA is a systems biology method for constructing scale-free networks using gene expression profile data. Total analyses were performed using the R package WGCNA (Version: 1.68) in R 3.6.2 (https://cran.rstudio.com/). Firstly, the similarity matrix of expression profile was constructed by calculating the Pearson's correlation coefficient between two different genes. Then, the similarity matrix of gene expression was converted into the adjacency matrix and scale-free network was assigned which the optimal soft threshold was $\beta = 4$. The crucial function of this step was strengthening strong correlation and weakening weak correlation at the gene expression level. Next, the adjacency matrix was transformed to the topological matrix (TOM). TOM was used to evaluate the degree association between genes and (1-TOM) was used for hierarchical clustering of genes. The dynamic tree cut algorithm was used to recognize and classify different modules and find the most representative gene in each module which was called module eigengene (ME). The ME represented the first principal component of each module, also meant the overall level of gene expression in this module. The minimum number of genes were 20 in each module, the correlation threshold of hub genes was 0.90 and the unsigned network edge threshold was 0.05. Clinically significant modules were identified by calculating the correlation between ME and clinical trait, and the degree of connection was measured. Gene significance (GS) was used to evaluate this degree and a higher GS indicated the increased significance of genes[16–18].

2.4 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analysis

GO and KEGG pathway enrichment analysis were performed using online database by DAVID 6.8 (https://david. ncifcrf.gov/). GO and KEGG pathway enrichment analysis was used to identify potential biological mechanism of genes. GO involved in three categories: biological process (BP), cellular component (CC) and molecular function (MF). The potential biological features and pathways of differentially expressed genes were further explored and the significance threshold was P value less than 0.05. The brown module was selected as the most representative module and the genes in this module were visualized with Cytoscape 3.5.1 (https:// cytoscape.org/), as this module was optimal related with clinical trait.

2.5 Construction of Protein Protein Interaction (PPI) network

The STRING 11.5 (https://string-db.org/) was used to construct protein protein interaction network. This online database provides a system-wide understanding of cellular function requires knowledge of all function interactions between the expressed proteins[19]. The associations among genes in brown module were obtained by STRING database and were visualized with Cytoscape 3.5.1 (https://cytoscape.org/). The CytoHubba package provides a user-friendly interface to explore hub nodes in biological networks and revealed the association degree among genes by transforming color from dark to light in brown module.

2.6 Identification and validation of hub genes

The correlation of genes were calculated using absolute of Pearson's correlation values by the Cor R package in R 3.6.2 (https://cran.rstudio.com/). Genes that had high correlation with a module which meant the absolute of Pearson's correlation values was more than 0.9 was regarded as hub genes. In addition, the GSE30718 dataset downloaded from GEO database (https://www. ncbi.nlm.nih.gov/geo/)was used to verify the hub genes.

2.7 AKI diagnostic model by LASSO logistic regression

Logistic regression analysis was performed to explore the impact of each gene on the development of AKI. The brown module which was the most related to the development of AKI was selected as the most representative module and fifteen genes was screened as hub genes of this module by bioinformatics methods of WGCNA. The Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis was applied to construct an optimal diagnostic model for AKI samples by using these hub genes. The risk score of each sample was calculated using this formula: Risk Score = Σ exp (genei*βi).

2.8 TF-mRNA and miRNA-mRNA network analysis

Transcription factors (TFs) as crucial regulators can modulate the expression of target genes by binding to specific DNA sequences of their promoters or enhancers[20]. hTFtarget (http://bioinfo.life.hust.edu.cn/hTFtarget#!/) online database provides an opportunity for understanding comprehensively TF-target regulations from large-scale ChIP-seq data of human TFs. The identification of TF-target relationship was a basis for making out the molecular regulatory mechanisms underlying

biological processes containing the development and pathogenesis. TargetScan 7.1 (http://www.targetscan.org/vert_71/) and ENCORI (http://www.starbase.sysu.edu.cn) are online databases integrate biological targets of miRNAs by searching for the presence of conserved sites which matches the seed region of each miRNA. The associations between miRNAs and mRNAs were demonstrated to be related to the molecular regulatory mechanisms and pathogenesis and exploring this regulations would be beneficial for mining novel therapeutic targets of AKI.

2.9 Immune infiltration analysis of hub genes

In order to explore the differences of immune cell subtypes, CIBERSORT R package (Version: 1.04) was used to evaluate the proportions of 22 immune cell subtypes based on gene expression profile. The analysis results were used for further analysis. In addition, we compared differences of immune cell subtypes between AKI cohort and non-AKI cohort by Mann-Whitney U test.

2.10 Statistical analysis

All statistical analyzes were carried out by R 3.6.2 (https://cran.rstudio.com/) and corresponding packages. Statistical significance was set at a probability value of $P < 0.05$. Significant differences were calculated by one-way ANOVA with Dunnett's or Newman-keuls test, or by two-tailed Student's t-test or Mann-Whitney U test.

3. Results

3.1 Identification of DEGs in AKI

The DEGs in 39 AKI samples compared with 9 non-AKI samples were analyzed by limma R package (Version:3.42.2) in R 3.6.2 (https://cran.rstudio.com/). There were 20139 genes were analyzed and 2202 DEGs were obtained, of which 572 DEGs were up regulated significantly and 1630 DEGs were down regulated significantly. The screening criteria for DEGs were log(2)(|FoldChange|) > 1 and the adjusted P value < 0.05. Hierarchical clustering was used to cluster samples and genes according to different gene expression values of genes in various samples and the expression profile of DEGs was visualized using a heat map (Fig. 1A). A volcano plot displayed the distribution of DEGs which log(2)(|FoldChange|) was calculated and showed in the horizontal axis, -log(10)(P-value) was calculated and showed in the longitudinal axis (Fig. 1B). The most significant up regulated ten genes were visualized in Table S1 and the most significant down regulated ten genes were shown in Table S2.

3.2 Construction of WGCNA and identification of clinical crucial module

Cluster analysis was carried out on the samples of GSE139061 using average linkage and Pearson's correlation, and the co-expression network was constructed by co-expression analysis. The soft threshold $\beta = 4$ was identified to ensure a scale-free network for further analysis (Fig. 2A). A total of 7 modules were identified by the average linkage hierarchical clustering, calculating with MEs and combing adjacent

modules with the same module and set the height as 0.25 (Fig. 2B). The Pearson's correlation coefficients among modules were calculated and the associations between modules and inner genes were evaluated (Fig. 2C and D). The brown module (R = 0.58, P = 0.00009) was selected as the target module for the optimal association with the development of AKI from the correlation with clinical trait and the suitable number of genes (Fig. 2E). The black, red and blue modules had high associations with clinical trait, however, these modules were not suitable genes mounts and further evaluation may focus on the correlation between these modules and the development of AKI. TOM was used to display the correlations between hierarchical clustering of genes and modules (Fig. 2F).

3.3 GO and KEGG pathway enrichment analysis

The genes in the crucial clinically significant module were categorized into three function groups, including BP, CC and MF. BP analysis revealed that the genes in brown module were mainly involved in positive regulation GTPase activity, cellular response to amino acid stimulus and transcription (DNA-templated) (Fig. S2A). CC analysis showed that the genes in brown module were mainly involved in nucleus, golgi apparatus, and endoplasmic reticulum lumen (Fig. S2B). MF analysis displayed that the genes in brown module were mainly involved in chromatin binding, zinc ion binding and ploy(A) RNA binding (Fig. S2C). KEGG analysis uncovered that the genes in brown module were mainly involved in Focal adhesion, ECM-receptor interaction and protein digestion and absorption (Fig. S2D). The function and enrichment of total DEGs were analyzed and DEGs were involved in multiple biological processes, cellular components, molecular functions and pathways. BP analysis revealed that DEGs were mainly involved in oxidation-reduction process, positive regulation of GTPase activity, cell adhesion, transport, response to drug, multiple metabolic processes and positive regulation of cell migration (Fig. 3A). CC analysis showed that DEGs were mainly involved in integral component of membrane, plasma membrane, extracellular exosome and mitochondrion (Fig. 3B). MF analysis displayed that DEGs were mainly involved in zinc ion binding, protein homodimerization activity, ligase activity, GTPase activator activity, catalytic activity and hydrolase activity (Fig. 3C). KEGG pathway enrichment analysis uncovered that DEGs were mainly involved in metabolic pathways, biosynthesis of antibiotics, Rap1 signaling pathway, carbon metabolism, protein digestion and absorption, drug metabolism, chemical carcinogenesis and retinol metabolism (Fig. 3D).

3.4 PPI network analysis of the key module

The PPI network of the genes in brown module was constructed by STRING database version: 11.5 (https://string-db.org/) and visualized by Cytoscape 3.5.1 (https://cytoscape.org/) (Fig. 4). There were 203 genes in brown module and 15 genes were identified as hub genes including KMT2B, NOC2L, COL1A1, BAZ1A, PABPN1, HNRNPD, H6PD, SYNE1, DST, RANGAP1, DEK, MACF1, CHD3, CXXC1 and UBTF. The high degree genes calculated by the cytohubba plugin were located in the center of the circle network and the dark color represented the high degree of genes. The expression levels of 15 hub genes between AKI and non-AKI cohort were shown in Fig. 5.

3.5 Validation of hub genes and Construction of AKI diagnostic model

The expression of 15 hub genes was validated by GSE30718 in GEO database. RNA-sequencing expression levels of KMT2B, NOC2L, COL1A1, BAZ1A, PABPN1, HNRNPD, H6PD, RANGAP1, DEK, CHD3, CXXC1 and UBTF were significantly increased in AKI cohort compared with non-AKI cohort. RNA-sequencing expression levels of DST, MACF1 and SYNE1were significantly decreased in AKI cohort compared with non-AKI cohort (Fig. 6). The validated results were consistent with the training dataset of GSE139061. To construct AKI related gene diagnostic model, LASSO regression was used to screen crucial genes from 15 hub genes. Then, five crucial genes were selected to perform binary logistic regression analysis (Fig. 7A and 7B). Finally, a diagnostic model containing five genes (RANGAP1, UBTF, SYNE1, BAZ1A and COL1A1) was established to assess the diagnostic efficiency of each patients as follows: Risk Score = 0.0391* RANGAP1 + 0.5509* UBTF + (-0.4938)* SYNE1 + 0.5081* BAZ1A + 0.0363* COL1A1. In addition, ROC curve analysis was performed to assess the diagnostic power of single gene and this model which was consist of five genes. The area under the ROC curve of RANGAP1 was 0.90, 95% confidence interval was 0.80-0.99, $P < 0.05$. The area under the ROC curve of UBTF was 0.91, 95% confidence interval was 0.83-0.99, $P < 0.05$. The area under the ROC curve of SYNE1was 0.87, 95% confidence interval was 0.76-0.98, $P < 0.05$. The area under the ROC curve of BAZ1A was 0.85, 95% confidence interval was 0.75-0.96, $P < 0.05$. The area under the ROC curve of COL1A1 was 0.77, 95% confidence interval was 0.58-0.96, $P < 0.05$. The area under the ROC curve of combination model was 0.99, 95% confidence interval was 0.98-1.01, $P < 0.05$ (Fig. 7C).

3.6 TF-mRNA and miRNA-mRNA network analysis

Using hTFtarget database, we explored the TF-mRNA interaction in AKI. As presented in Table S3, the fifteen hub genes were used to screen corresponding transcription factors, and each hub gene may be transcriptionally regulated by multiple transcription factors which were combined with transcription factor binding sites in upstream promoter regions. The interaction network was constructed between fifteen hub genes and target TFs by Cytoscape 3.5.1 (Fig. 8). Bromodomain-containing protein 2 (BRD2), Histone acetyltransferase p300 (EP300), Protein C-ets 1 (ETS1), Myc proto-oncogene protein (MYC), Transcription factor PU.1 (SPI1) and Zinc finger protein 263 (ZNF263) were significantly enriched transcription factors for the fifteen hub genes. Furthermore, we explored the miRNA-mRNA interaction in AKI using TargetScan 7.1 and ENCORI databases. As presented in Table S4, the predicted target miRNAs were analyzed and each miRNA was combined with 3’ UTRs of target mRNA. Each mRNA is corresponding to multiple target miRNAs. The interaction network was established between fifteen hub genes and target miRNAs by Cytoscape 3.5.1 (Fig. 9). It was identified that miR-181c-5p, miR-218-5p, miR-485-5p, miR-532-5p and miR-6884-5p were significantly enriched target miRNAs for the fifteen hub genes.

3.7 Immune cell subtypes analysis between AKI and non-AKI cohort

The 22 immune cell proportions of AKI were shown in Fig. 10. T cell CD4 naïve, T cell CD4 memory resting, NK cell resting, Monocytes accounted for a large proportions of AKI immune cell infiltration. AKI and non-AKI cohort displayed different immune cells expression. The expression of Monocytes and Macrophages M0 in AKI cohort was increasing significantly compared with non-AKI cohort. The expression of Macrophages M2 in AKI cohort was decreasing significantly compared with non-AKI cohort (Fig. 11).

4. Discussion

Acute kidney injury is a common and vital diagnostic and therapeutic challenge for clinicians[21] and is typically diagnosed by the accumulation of end products of nitrogen metabolism (urea and creatinine) or decreased urine output, or both[1, 22]. According to the different anatomical structures, AKI was divided into three types of prerenal, renal, postrenal impairment[23]. The pathophysiology of AKI shares common pathogenic denominators including cell death, cell injury, inflammation, and fibrosis, regardless of the initiating insults[24]. AKI can be converted to CKD easily in case of no early diagnosis and effective treatment measures. At present, owing to the lack of sensitive and specific means of AKI prevention and therapy, it is crucial to study the diagnostic biomarkers, novel therapeutic targets and potential pathophysiological mechanism in AKI. Therefore, microarray and high throughout sequencing analysis were used in present study to investigate the function of genes at the whole genome level[25]. WGCNA, a systematic biology method, was used to investigate co-expression in AKI and non-AKI tissues. In addition, the current study explored potential molecular mechanisms, miRNAs, transcription factors and immune cell infiltration and constructed a diagnostic model by LASSO logistic regression.

With adjusted P value < 0.05 and log(2)(|Foldchange|) > 1 as the cutoff, 2202 DEGs (572 upregulated and 1630 downregulated genes) were identified, which have potential to be novel drivers and may play a role in the pathophysiological mechanism underlying AKI development. Fifteen hub genes (KMT2B, NOC2L, COL1A1, BAZ1A, PABPN1, HNRNPD, H6PD, SYNE1, DST, RANGAP1, DEK, MACF1, CHD3, CXXC1 and UBTF) of brown module were selected using comprehensive analytical method of WGCNA, which were further successfully validated using another dataset of GEO database. To further understand the molecular mechanism, GO and KEGG pathways enrichment analysis of the DEGs were performed. As for GO analysis, it was identified that DEGs were enriched in oxidation-reduction process, cell adhesion, proliferation, migration, metabolic process, mitochondria, iron ion binding, heparin binding, oxygen binding, and so on. As for KEGG pathways enrichment analysis, it was identified that DEGs were enriched in metabolic pathways, biosynthesis of antibiotics, Rap1 signaling pathways, carbon metabolism, drug metabolism, and so on. Recent studies have shown that the primary site of damage during AKI, proximal tubular epithelial cells, are highly metabolically active, relying on fatty acids to meet energy demands, which are rich in mitochondria and peroxisomes. The two organelles mediate fatty acid oxidation[26]. Mitochondria are cytoplasmic organelles with a double phospholipid membrane that generate energy via oxidative phosphorylation[27]. Mitochondria are also associated with calcium homeostasis, intracellular reactive oxygen species (ROS) generation and cell signaling functions[28, 29]. Mitochondrial fatty acid β-oxidation serves as the preferred source of ATP in the kidney and its dysfunction results in ATP depletion

and lipotoxicity to elicit tubular injury and inflammation and subsequent fibrosis progression[30]. The kidney is a highly metabolic organ with high levels of oxidation within cellular mitochondria[31]. Metabolic process includes glucose metabolism, lipid metabolism, drug metabolism and so on which have been discovered in enrichment analysis of DEGs. Lipid metabolism plays a basic role in renal physiology, especially in tubules[32]. Some studies have revealed the emerging association between increased metabolites and AKI pathogenesis and progression from different perspectives, which were consistent with our study[33, 34].

The present study identified that fifteen genes were significantly associated with AKI development as hub genes of brown module. The Mixed Lineage Leukemia 2 (MLL2) protein, also known as KMT2B, belongs to the family of mammalian histone H3 lysine 4 (H3K4) methyltransferases[35]. Moreover, KMT2B plays a key role in development and germ line deletions of MLL2 have been associated with early growth retardation, neural tube defects and apoptosis that leads to embryonic death. The research has revealed that KMT2B acts as a chromatin modifier gene harbors mutations in Renal cell carcinomas through high-throughput sequencing efforts[36]. However, to our knowledge, no experimental studies of KMT2B in acute kidney injury have been reported to date, which is worthy of further study. NOC2L, acts as an inhibitor of histone acetyltransferase activity, prevents acetylation of all core histones by the EP300/p300 histone acetyltransferase at p53/TP53-regulated target promoters in a histone deacetylases-independent manner with chronic kidney disease[37]. COL1A1, acts as type I collagen, is a member of group I collagen (fibrillary forming collagen). The mutations of COL1A1 can cause Osteogenesis imperfecta has been reported extensively[38]. Recent researches have revealed that COL1A1 is highly associated with chronic kidney disease, cardiovascular diseases and bone metabolism disorders[39]. Moreover, the experimental and theory studies of COL1A1 in acute kidney injury are required in further study. Heterogeneous nuclear ribonucleoprotein D (HNRNPD), has been shown to regulate gene expression at the translational and even the transcriptional level and regulate AU rich elements (ARE)-mRNA turnover, primarily functioning to promote rapid ARE-mRNA degradation and various kidney cells express multiple isoforms of HNRNPD[40]. H6PD is a steroid conversion and receptor gene which plays a crucial role in steroid conversion and response in kidney transplantation[41]. The research of SYNE1 revealed that the essential roles in mediating sunitinib cytotoxicity and the loss of function rendered renal cell carcinoma cell resistant to sunitinib in vitro and in vivo[42]. The rest of the hub genes were rarely identified and verified in various kidney disease, especially in acute kidney injury, which demand us to move one step further uncovering the pathophysiologic mechanisms and biological principles. These hub genes will be potential biomarkers and therapeutic targets of acute kidney injury.

The present study identified that BRD2, EP300, ETS1, MYC, SPI1, ZNF263 were significantly enriched transcriptional factors for fifteen hub genes. BRD2 can specially bind acetylated histone H4 and mediate transcription, which belongs to the bromodomain and extraterminal domain (BET) family regulating the expression of many immunity-associated genes and pathways[43]. EP300, a protein with an essential role in controlling cell growth, cell division and prompting cells to differentiate to take on specialized functions, can mediate epigenetic variation of kidney disease as a transcriptional factor[44, 45]. ETS1 is a member of the ETS family and regulates the expression of a variety of genes including growth factors,

chemokines and adhesion molecules and plays a crucial role in the cell cycle progression of renal tubules in acute renal failure (ARF). The ETS1 pathway may regulate the transcription of cyclin D1 and control the regeneration of renal tubules in ARF[46]. MYC acts as a transcriptional factor becomes activated in resident kidney stromal cells early after kidney injury and can regulate metabolic switch in fibrosis initiation and progression[47]. Another study revealed MYC acts as a transcriptional factor participating in the positive feedback loop of MEG3/miR-145-5p/RTKN/ Wnt/β-catenin/c-MYC to promote renal ischemia-reperfusion injury by activating mitophagy and inducing apoptosis[48]. SPI1 can activate H19 which overexpression confers protection against renal injury by stimulating proangiogenic signaling in endothelial cells and tubular epithelial cells of ischemic kidney tissue[49]. ZNF263 acts as a transcriptional factor can regulate a crucial enzyme involved in imparting anticoagulant activity to heparin and also can influence the gene expression of ZRANB2 in human kidney cells[50]. Our study demonstrated that miR-181c-5p, miR-218-5p, miR-485-5p, miR-532-5p and miR-6884-5p were significantly enriched target miRNAs for fifteen hub genes. miR-181c-5p is a member of miR-181c family, and plays a crucial role in regulating extracellular matrix proteins during AKI occurrence and progression[51]. miR-218-5p participates in regulating sepsis-induced acute kidney injury by miR-218-5p/ hemeoxygenase-1 signaling pathway[52]. Wang *et al* reported that miR-218-5p expressed in endothelial progenitor cells contributes to the development and repair of the kidney microvasculature[53]. miR-485-5p regulates the function of TP53 signaling pathway with signal transduction in response to DNA damage and cell cycle regulation of kidney tissues[54]. The research of miR-532-5p uncovered that LINC00052 ameliorates acute kidney injury by sponging miR-532-5p and activating the Wnt signaling pathway[55]. miR-6884-5p regulates the functions of proliferation, invasion, Epithelial-Mesenchymal Transformation in tumors, however, there are rarely researches to reveal the function and mechanism of kidney diseases[56]. Our study results revealed that the expression of Macrophages M2 in AKI cohort was decreasing significantly compared with non-AKI cohort, which meant M2 macrophages were a protective factor for AKI development. The research showed that M2 macrophages can effectively alleviate acute kidney injury by decreasing inflammatory response and promoting primary proximal tubular epithelial cells proliferation, which was consistent with our findings[57].

In conclusion, the present study identified fifteen hub genes in acute kidney injury using WGCNA and constructed a diagnostic model by LASSO logistic regression. Various pathways, Transcription factors, miRNAs and 22 immune cell subtypes which were associated with AKI were analyzed and provided some basis for future experimental studies.

Abbreviations

WGCNA, weighted genes co-expression network analysis; LASSO, least absolute shrinkage and selection operator; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TF, transcriptional factor; AKI, acute kidney injury; DEG, differentially expressed gene; ICU, intensive care unit; TOM, topological overlap measure; ME, module eigengene; GS, gene significance; BP, biological process; CC, cellular component; MF, molecular function; PPI, protein protein interaction; ROC, receiver operating characteristic; CKD, chronic kidney disease.

Declarations

Ethics approval and consent to participate

All patients have written informed contents and this study was approved by the Second Affiliated Hospital of Zhejiang University School of Medicine.

Consent for publication

All authors have agreed to the consents of the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

Not applicable.

Authors' contributions

Tao Sun finished the manuscript and abstract; **Ying Cao** and **Tiancha Huang** completed the figures and tables; **Yibei Dai and Yiwen Sang** consulted relevant literature and completed English revision; **Zhihua Tao** provided constructive feedback and guidance. All authors have read and approved the final manuscript.

Acknowledgments

Not applicable.

Authors' information

[a] Department of Laboratory Medicine, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China.

References

1. Bellomo R, Kellum JA, Ronco C: **Acute kidney injury**. *The Lancet* 2012, **380**(9843):756-766.
2. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA: **Clinical Decision Support for In-Hospital AKI**. *Journal of the American Society of Nephrology : JASN* 2018, **29**(2):654-660.

3. Hoste EA, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, Edipidis K, Forni LG, Gomersall CD, Govil D *et al*: **Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study**. *Intensive care medicine* 2015, **41**(8):1411-1423.
4. Levey AS, James MT: **Acute Kidney Injury**. *Annals of internal medicine* 2017, **167**(9):Itc66-itc80.
5. Chawla LS, Bellomo R, Bihorac A, Goldstein SL, Siew ED, Bagshaw SM, Bittleman D, Cruz D, Endre Z, Fitzgerald RL *et al*: **Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup**. *Nature reviews Nephrology* 2017, **13**(4):241-257.
6. **<Acute Renal Tubular Necrosis.pdf>**.
7. Srisawat N, Kellum JA: **The Role of Biomarkers in Acute Kidney Injury**. *Critical care clinics* 2020, **36**(1):125-140.
8. Venkatachalam MA, Weinberg JM, Kriz W, Bidani AK: **Failed Tubule Recovery, AKI-CKD Transition, and Kidney Disease Progression**. *Journal of the American Society of Nephrology : JASN* 2015, **26**(8):1765-1776.
9. Chen Y, Lin L, Tao X, Song Y, Cui J, Wan J: **The role of podocyte damage in the etiology of ischemia-reperfusion acute kidney injury and post-injury fibrosis**. *BMC nephrology* 2019, **20**(1):106.
10. Zhu Q, Wong AK, Krishnan A, Aure MR, Tadych A, Zhang R, Corney DC, Greene CS, Bongo LA, Kristensen VN *et al*: **Targeted exploration and analysis of large cross-platform human transcriptomic compendia**. *Nature methods* 2015, **12**(3):211-214, 213 p following 214.
11. Alberti C, Manzenreither RA, Sowemimo I, Burkard TR, Wang J, Mahofsky K, Ameres SL, Cochella L: **Cell-type specific sequencing of microRNAs from complex animal tissues**. 2018, **15**(4):283-289.
12. Radulescu E, Jaffe AE: **Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain**. 2020, **25**(4):791-804.
13. Qin S, Long X, Zhao Q, Zhao W: **Co-Expression Network Analysis Identified Genes Associated with Cancer Stem Cell Characteristics in Lung Squamous Cell Carcinoma**. *Cancer investigation* 2020, **38**(1):13-22.
14. Liang JW, Fang ZY, Huang Y, Liuyang ZY, Zhang XL, Wang JL, Wei H, Wang JZ, Wang XC, Zeng J *et al*: **Application of Weighted Gene Co-Expression Network Analysis to Explore the Key Genes in Alzheimer's Disease**. *Journal of Alzheimer's disease : JAD* 2018, **65**(4):1353-1364.
15. Kwon J, Kim YJ, Choi K, Seol S, Kang HJ: **Identification of stress resilience module by weighted gene co-expression network analysis in Fkbp5-deficient mice**. 2019, **12**(1):99.
16. Lin J, Yu M, Xu X, Wang Y, Xing H, An J, Yang J, Tang C, Sun D, Zhu Y: **Identification of biomarkers related to CD8(+) T cell infiltration with gene co-expression network in clear cell renal cell carcinoma**. *Aging* 2020, **12**(4):3694-3712.
17. Sun T, Wang D, Ping Y, Sang Y, Dai Y, Wang Y, Liu Z, Duan X, Tao Z, Liu W: **Integrated profiling identifies SLC5A6 and MFAP2 as novel diagnostic and prognostic biomarkers in gastric cancer patients**. *International journal of oncology* 2020, **56**(2):460-469.

18. Lin X, Li J, Tan R, Zhong X, Yang J, Wang L: **Identification of Hub Genes Associated with the Development of Acute Kidney Injury by Weighted Gene Co-Expression Network Analysis**. *Kidney & blood pressure research* 2021, **46**(1):63-73.
19. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: **The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible**. 2017, **45**(D1):D362-d368.
20. Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, Guo AY: **hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets**. *Genomics, proteomics & bioinformatics* 2020, **18**(2):120-128.
21. Liangos O, Wald R, O'Bell JW, Price L, Pereira BJ, Jaber BL: **Epidemiology and outcomes of acute renal failure in hospitalized patients: a national survey**. *Clinical journal of the American Society of Nephrology : CJASN* 2006, **1**(1):43-51.
22. Koeze J, Keus F, Dieperink W, van der Horst IC, Zijlstra JG, van Meurs M: **Incidence, timing and outcome of AKI in critically ill patients varies with the definition used and the addition of urine output criteria**. *BMC nephrology* 2017, **18**(1):70.
23. Mercado MG, Smith DK, Guard EL: **Acute Kidney Injury: Diagnosis and Management**. *American family physician* 2019, **100**(11):687-694.
24. Sato Y, Yanagita M: **Immune cells and inflammation in AKI to CKD progression**. *American journal of physiology Renal physiology* 2018, **315**(6):F1501-f1512.
25. Yang JJ, Wu BB, Han F, Chen JH, Yang Y: **Gene expression profiling of sepsis-associated acute kidney injury**. *Experimental and therapeutic medicine* 2020, **20**(5):34.
26. Chiba T, Peasley KD: **Sirtuin 5 Regulates Proximal Tubule Fatty Acid Oxidation to Protect against AKI**. 2019, **30**(12):2384-2398.
27. Green DR, Galluzzi L, Kroemer G: **Mitochondria and the autophagy-inflammation-cell death axis in organismal aging**. *Science (New York, NY)* 2011, **333**(6046):1109-1112.
28. Galley HF: **Oxidative stress and mitochondrial dysfunction in sepsis**. *British journal of anaesthesia* 2011, **107**(1):57-64.
29. Rocha M, Herance R, Rovira S, Hernández-Mijares A, Victor VM: **Mitochondrial dysfunction and antioxidant therapy in sepsis**. *Infectious disorders drug targets* 2012, **12**(2):161-178.
30. Jang HS, Noh MR, Kim J, Padanilam BJ: **Defective Mitochondrial Fatty Acid Oxidation and Lipotoxicity in Kidney Diseases**. *Frontiers in medicine* 2020, **7**:65.
31. Kitada M, Xu J, Ogura Y, Monno I, Koya D: **Manganese Superoxide Dismutase Dysfunction and the Pathogenesis of Kidney Disease**. *Frontiers in physiology* 2020, **11**:755.
32. Li W, Duan A, Xing Y, Xu L, Yang J: **Transcription-Based Multidimensional Regulation of Fatty Acid Metabolism by HIF1α in Renal Tubules**. *Frontiers in cell and developmental biology* 2021, **9**:690079.
33. Bugarski M, Ghazi S, Polesel M, Martins JR, Hall AM: **Changes in NAD and Lipid Metabolism Drive Acidosis-Induced Acute Kidney Injury**. *Journal of the American Society of Nephrology : JASN* 2021,

32(2):342-356.

34. Small DM, Sanchez WY, Roy SF, Morais C, Brooks HL, Coombes JS, Johnson DW, Gobe GC: **N-acetyl-cysteine increases cellular dysfunction in progressive chronic kidney damage after acute kidney injury by dampening endogenous antioxidant responses**. *American journal of physiology Renal physiology* 2018, **314**(5):F956-f968.
35. Klonou A, Chlamydas S: **Structure, Activity and Function of the MLL2 (KMT2B) Protein Lysine Methyltransferase**. 2021, **11**(8).
36. de Cubas AA, Rathmell WK: **Epigenetic modifiers: activities in renal cell carcinoma**. *Nature reviews Urology* 2018, **15**(10):599-614.
37. Hublitz P, Kunowska N, Mayer UP, Müller JM, Heyne K, Yin N, Fritzsche C, Poli C, Miguet L, Schupp IW *et al*: **NIR is a novel INHAT repressor that modulates the transcriptional activity of p53**. *Genes & development* 2005, **19**(23):2912-2924.
38. Marini JC, Dang Do AN: **Osteogenesis Imperfecta**. In: *Endotext.* Edited by Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, Dungan K, Grossman A, Hershman JM, Hofland J *et al*. South Dartmouth (MA): MDText.com, Inc.Copyright © 2000-2021, MDText.com, Inc.; 2000.
39. Perco P, Wilflingseder J, Bernthaler A, Wiesinger M, Rudnicki M, Wimmer B, Mayer B, Oberbauer R: **Biomarker candidates for cardiovascular disease and bone metabolism disorders in chronic kidney disease: a systems biology perspective**. *Journal of cellular and molecular medicine* 2008, **12**(4):1177-1187.
40. Schroeder JM, Ibrahim H, Taylor L, Curthoys NP: **Role of deadenylation and AUF1 binding in the pH-responsive stabilization of glutaminase mRNA**. *American journal of physiology Renal physiology* 2006, **290**(3):F733-740.
41. Christakoudi S, Runglall M, Mobillo P, Rebollo-Mesa I, Tsui TL, Nova-Lamperti E, Norris S, Kamra Y, Hilton R, Bhandari S *et al*: **Steroid regulation: An overlooked aspect of tolerance and chronic rejection in kidney transplantation**. *Molecular and cellular endocrinology* 2018, **473**:205-216.
42. Elgendy M, Fusco JP, Segura V, Lozano MD, Minucci S, Echeveste JI, Gurpide A, Andueza M, Melero I, Sanmamed MF *et al*: **Identification of mutations associated with acquired resistance to sunitinib in renal cell cancer**. 2019, **145**(7):1991-2001.
43. Wang N, Wu R, Tang D, Kang R: **The BET family in immunity and disease**. *Signal transduction and targeted therapy* 2021, **6**(1):23.
44. Liang W, Yamahara K, Hernando-Erhard C, Lagies S, Wanner N, Liang H, Schell C, Kammerer B, Huber TB, Bork T: **A reciprocal regulation of spermidine and autophagy in podocytes maintains the filtration barrier**. *Kidney international* 2020, **98**(6):1434-1448.
45. Chu AY, Tin A, Schlosser P: **Epigenome-wide association studies identify DNA methylation associated with kidney function**. 2017, **8**(1):1286.
46. Feng W, Chumley P, Prieto MC, Miyada K, Seth DM, Fatima H, Hua P, Rezonzew G, Sanders PW, Jaimes EA: **Transcription factor avian erythroblastosis virus E26 oncogen homolog-1 is a novel**

mediator of renal injury in salt-sensitive hypertension. *Hypertension (Dallas, Tex : 1979)* 2015, **65**(4):813-820.

47. Lemos DR, McMurdo M, Karaca G, Wilflingseder J: **Interleukin-1β Activates a MYC-Dependent Metabolic Switch in Kidney Stromal Cells Necessary for Progressive Tubulointerstitial Fibrosis**. 2018, **29**(6):1690-1705.
48. Liu D, Liu Y, Zheng X, Liu N: **c-MYC-induced long noncoding RNA MEG3 aggravates kidney ischemia-reperfusion injury through activating mitophagy by upregulation of RTKN to trigger the Wnt/β-catenin pathway**. *Cell death & disease* 2021, **12**(2):191.
49. Haddad G, Kölling M: **Renal AAV2-Mediated Overexpression of Long Non-Coding RNA H19 Attenuates Ischemic Acute Kidney Injury Through Sponging of microRNA-30a-5p**. 2021, **32**(2):323-341.
50. Yang YH, Markus MA, Mangs AH, Raitskin O, Sperling R, Morris BJ: **ZRANB2 localizes to supraspliceosomes and influences the alternative splicing of multiple genes in the transcriptome**. *Molecular biology reports* 2013, **40**(9):5381-5395.
51. Chen SJ, Wu P, Sun LJ, Zhou B, Niu W, Liu S, Lin FJ, Jiang GR: **miR-204 regulates epithelial-mesenchymal transition by targeting SP1 in the tubular epithelial cells after acute kidney injury induced by ischemia-reperfusion**. *Oncology reports* 2017, **37**(2):1148-1158.
52. Zhang T, Xiang L: **Honokiol alleviates sepsis-induced acute kidney injury in mice by targeting the miR-218-5p/heme oxygenase-1 signaling pathway**. *Cellular & molecular biology letters* 2019, **24**:15.
53. Wang X, Liu J, Yin W, Abdi F, Pang PD, Fucci QA, Abbott M, Chang SL, Steele G, Patel A *et al*: **miR-218 Expressed in Endothelial Progenitor Cells Contributes to the Development and Repair of the Kidney Microvasculature**. *The American journal of pathology* 2020, **190**(3):642-659.
54. He C, Qin H, Tang H, Yang D, Li Y, Huang Z, Zhang D, Lv C: **Comprehensive bioinformatics analysis of the TP53 signaling pathway in Wilms' tumor**. *Annals of translational medicine* 2020, **8**(19):1228.
55. Li X, Zheng P, Ji T, Tang B, Wang Y, Bai S: **LINC00052 ameliorates acute kidney injury by sponging miR-532-3p and activating the Wnt signaling pathway**. *Aging* 2020, **13**(1):340-350.
56. Lv H, Hou H, Lei H, Nie C, Chen B, Bie L, Han L, Chen X: **MicroRNA-6884-5p Regulates the Proliferation, Invasion, and EMT of Gastric Cancer Cells by Directly Targeting S100A16**. *Oncology research* 2020, **28**(3):225-236.
57. Mao R, Wang C: **Peritoneal M2 macrophage transplantation as a potential cell therapy for enhancing renal repair in acute kidney injury**. 2020, **24**(6):3314-3327.

Figures

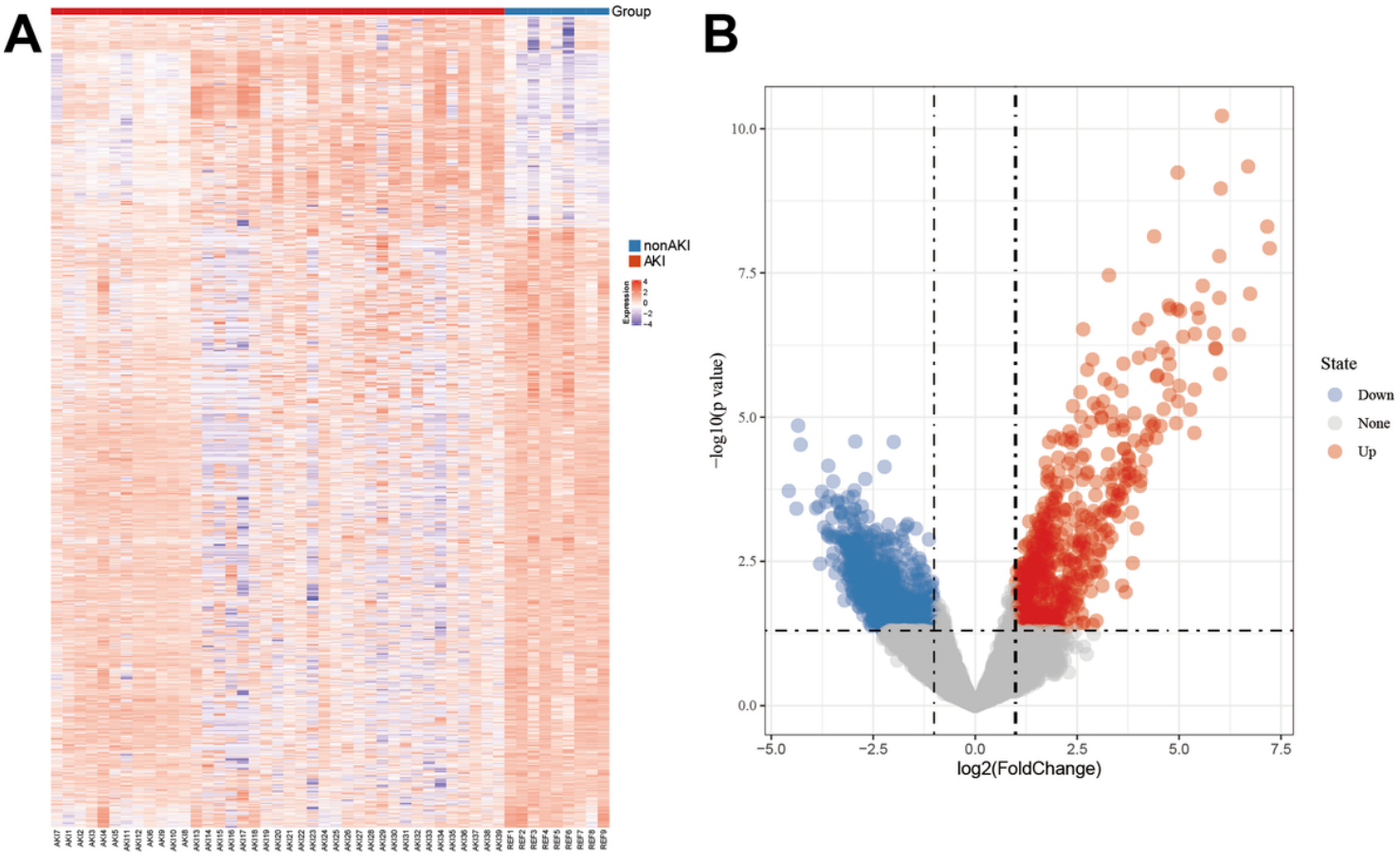

Figure 1

The differentially expressed genes of Acute kidney injury. A: The heatmap of DEGs between AKI and non-AKI cohort. B: The volcano plot of the DEGs. The red dots indicates the high expression and the blue dots indicates the low expression.

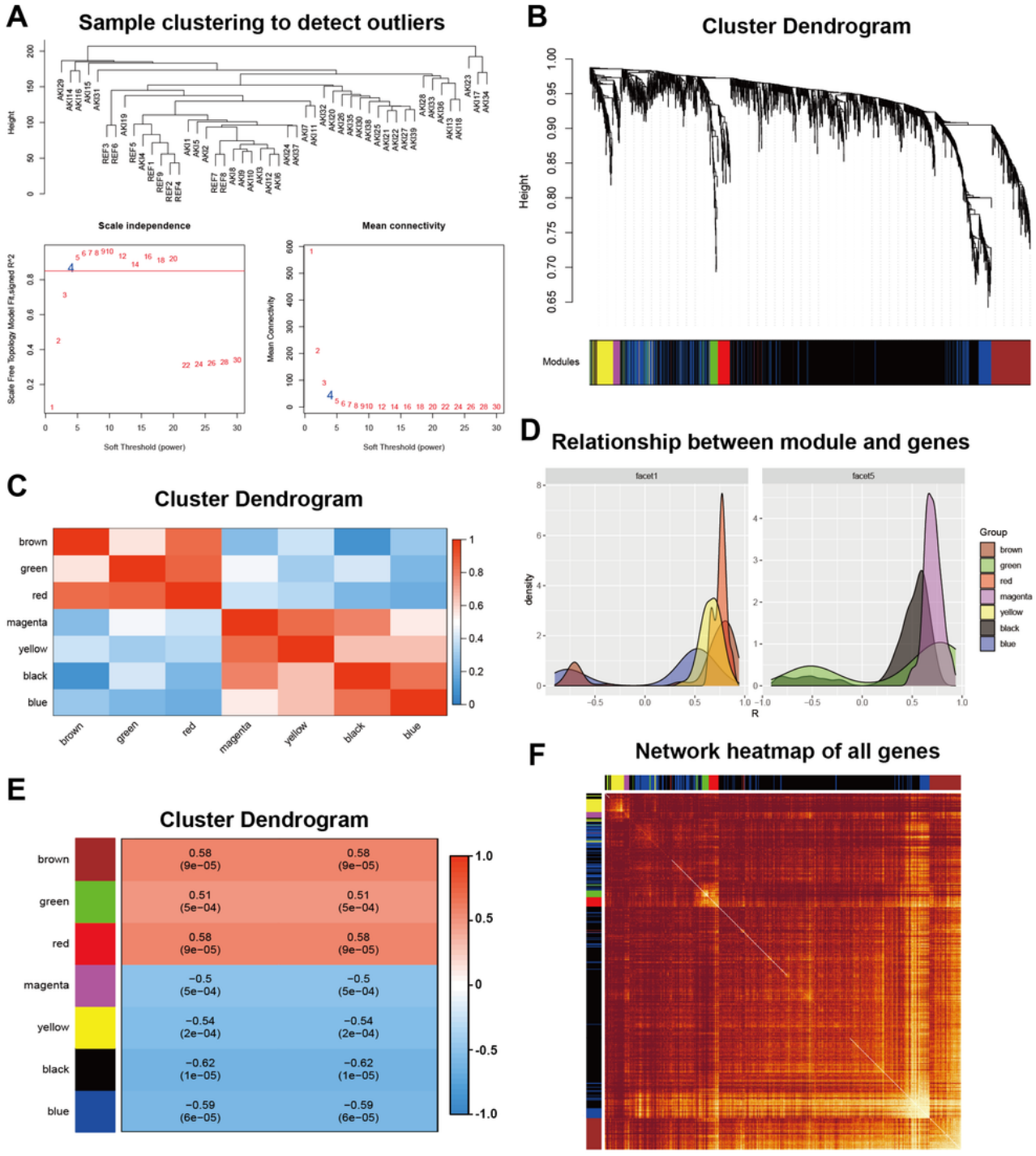

Figure 2

The WGCNA analysis of differentially expressed genes. A: Cluster dendrogram of 48 kidney tissue samples which contain 39 AKI samples and 9 non-AKI samples, and determination of the soft threshold of weighted co-expression network. B: Cluster dendrogram of differentially expressed genes to identify the clinically significant modules associated with the development of AKI. C: Heatmap of the correlations among different modules. D: The relationship between modules and inner genes. E: Heatmap of the

correlations between module eigengenes and clinical traits of AKI. F: The TOM plot of the correlations between hierarchical clustering of genes and modules.

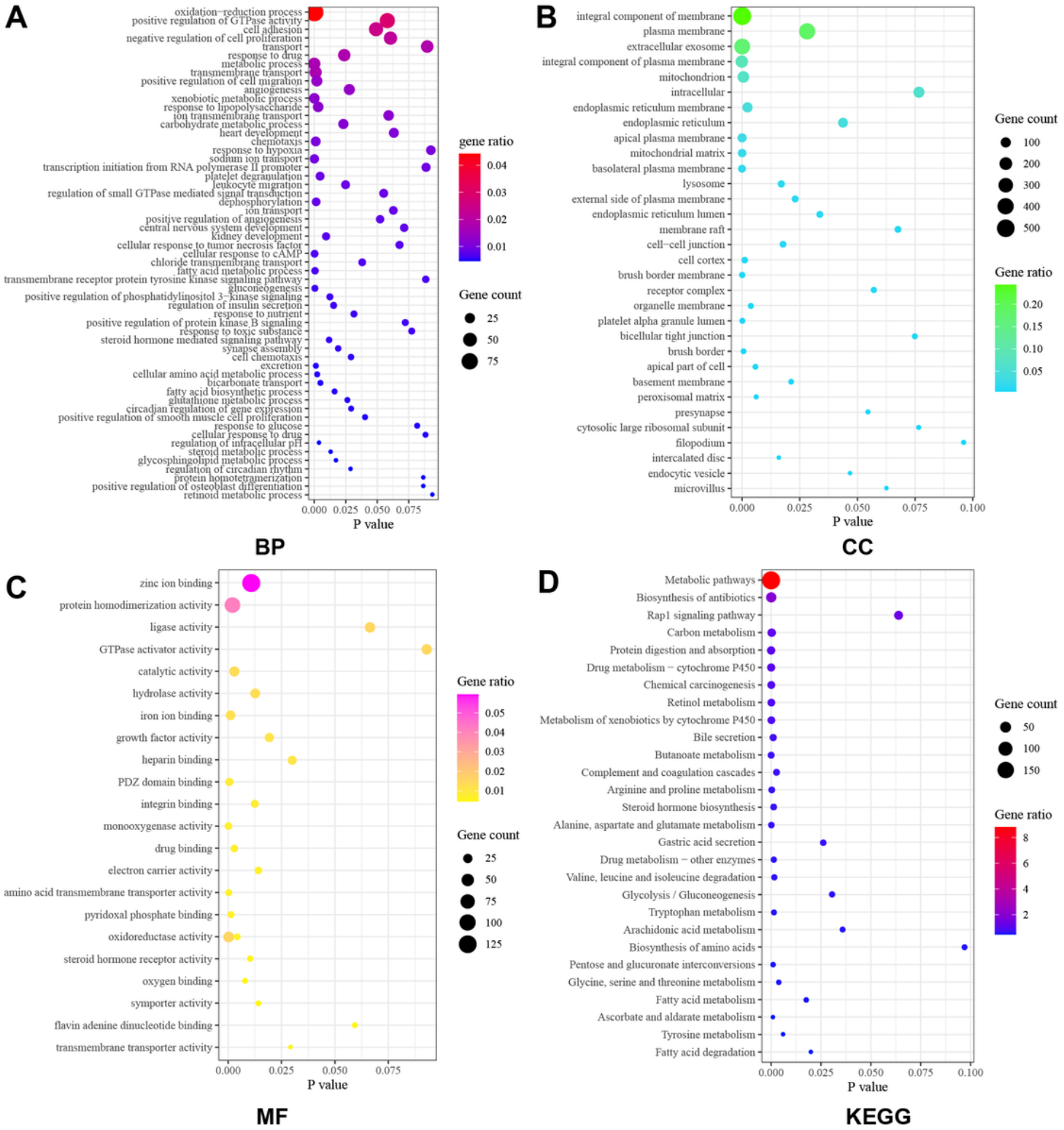

Figure 3

GO and KEGG pathway enrichment analysis of DEGs. A: Biological Process. B: Cellular Component. C: Molecular Function. D: Kyoto Encyclopedia of Genes and Genome pathways.

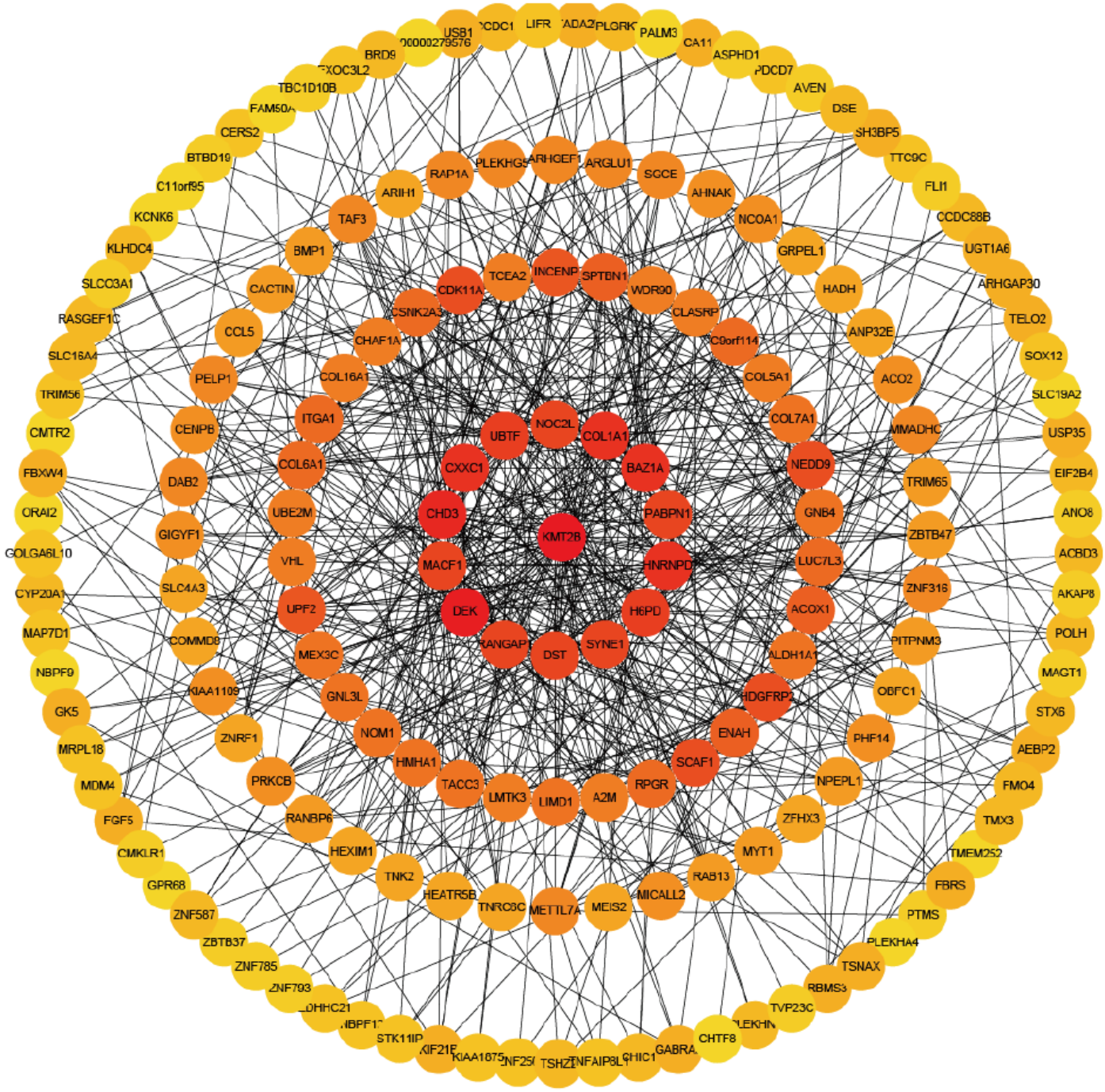

Figure 4

PPI network analysis of DEGs. The high degree genes calculated by the cytohubba plugin were located in the center of the circle network and the dark color represented the high degree of genes.

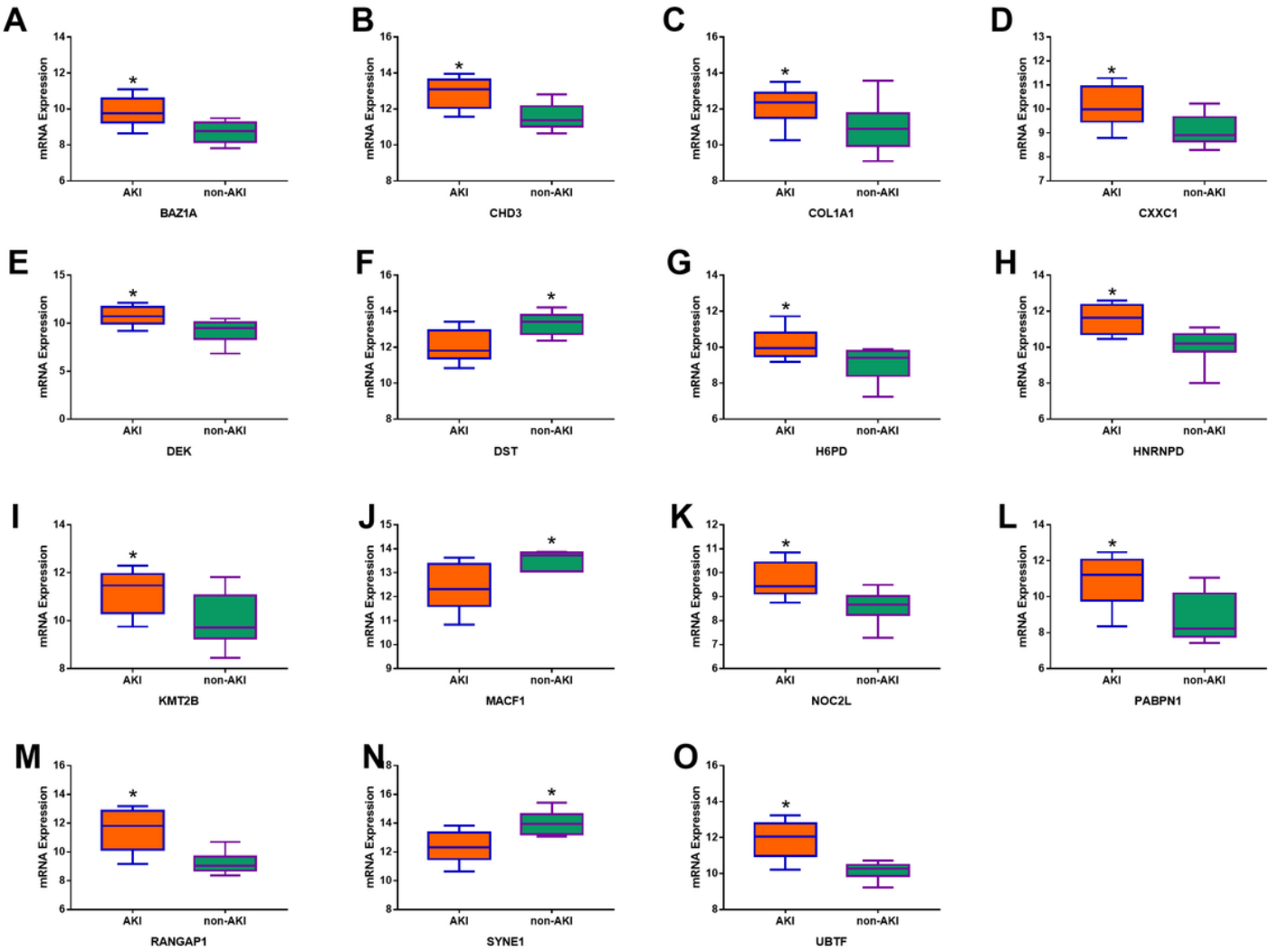

Figure 5

The expression levels of fifteen hub genes in brown module. A: BAZ1A. B: CHD3. C: COL1A1. D: CXXC1. E: DEK. F: DST. G: H6PD. H: HNRNPD. I: KMT2B. J: MACF1. K: NOC2L. L: PABPN1. M: RANGAP1. N: SYNE1. O: UBTF.

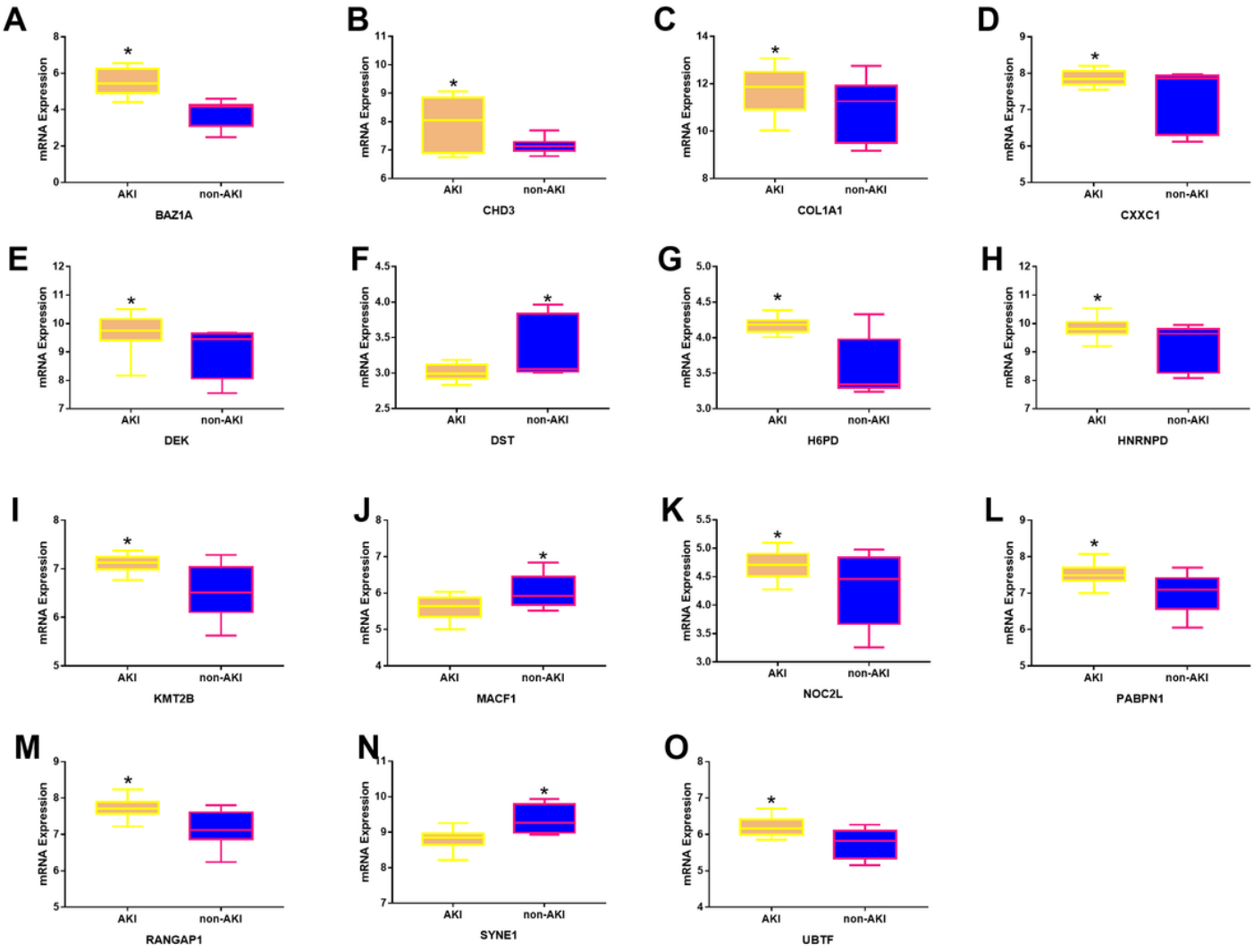

Figure 6

The validation of the expression levels of fifteen hub genes by GSE30718 in GEO database. A: BAZ1A. B: CHD3. C: COL1A1. D: CXXC1. E: DEK. F: DST. G: H6PD. H: HNRNPD. I: KMT2B. J: MACF1. K: NOC2L. L: PABPN1. M: RANGAP1. N: SYNE1. O: UBTF.

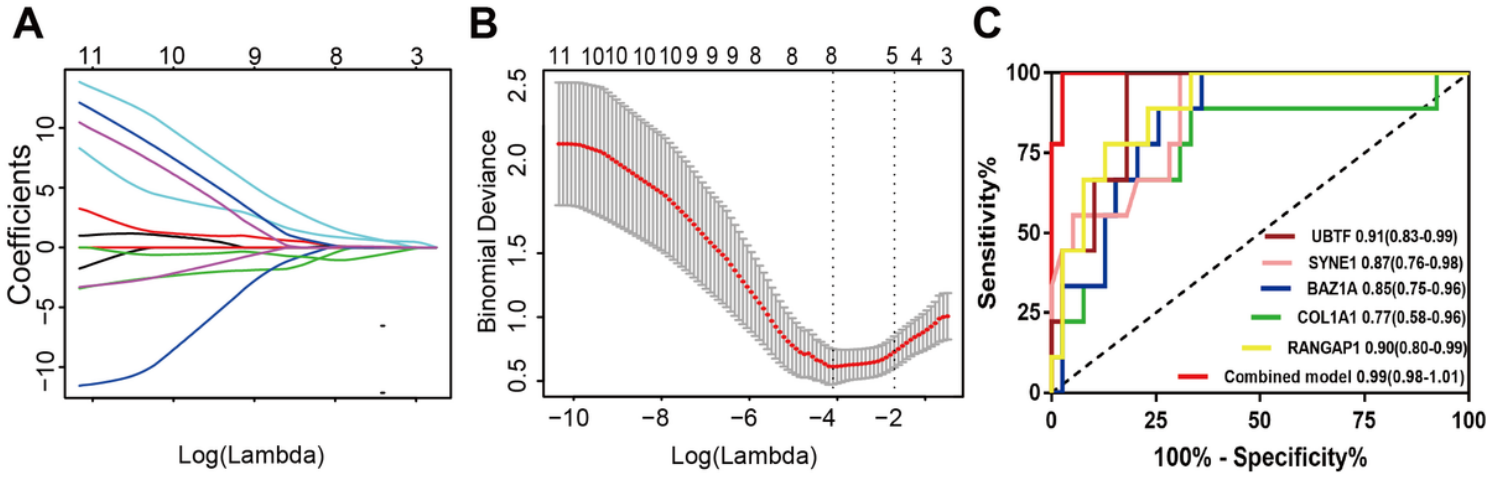

Figure 7

Construction of AKI diagnostic model by LASSO logistic regression analysis. A and B: Screening five crucial genes from fifteen hub genes by LASSO regression for constructing diagnostic model. C: ROC curve analysis for evaluating the diagnostic efficiency of single gene and the model.

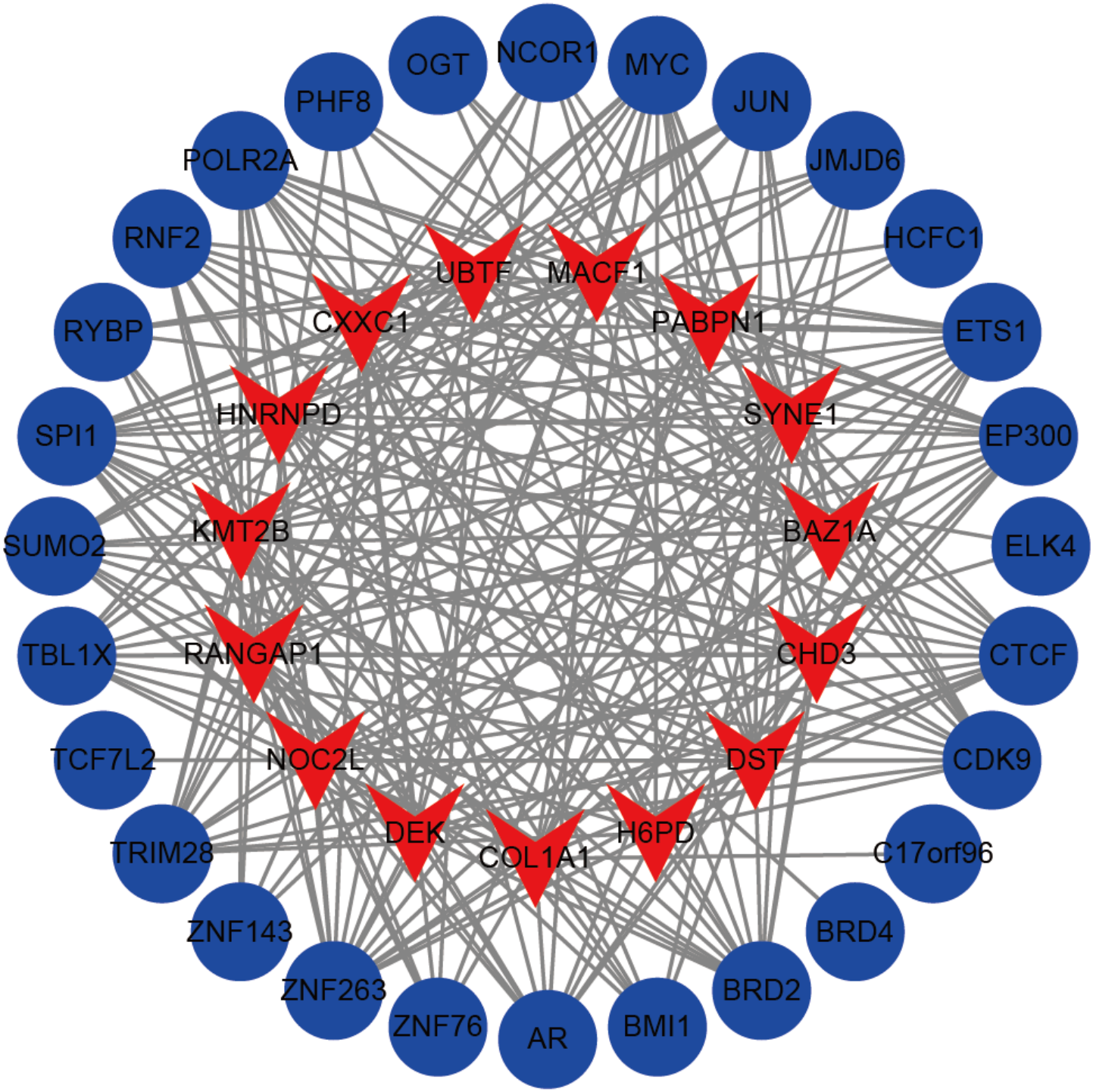

Figure 8

The interaction network between fifteen hub genes and target transcriptional factors. The red nodes represent fifteen hub genes and the blue nodes represent target transcriptional factors.

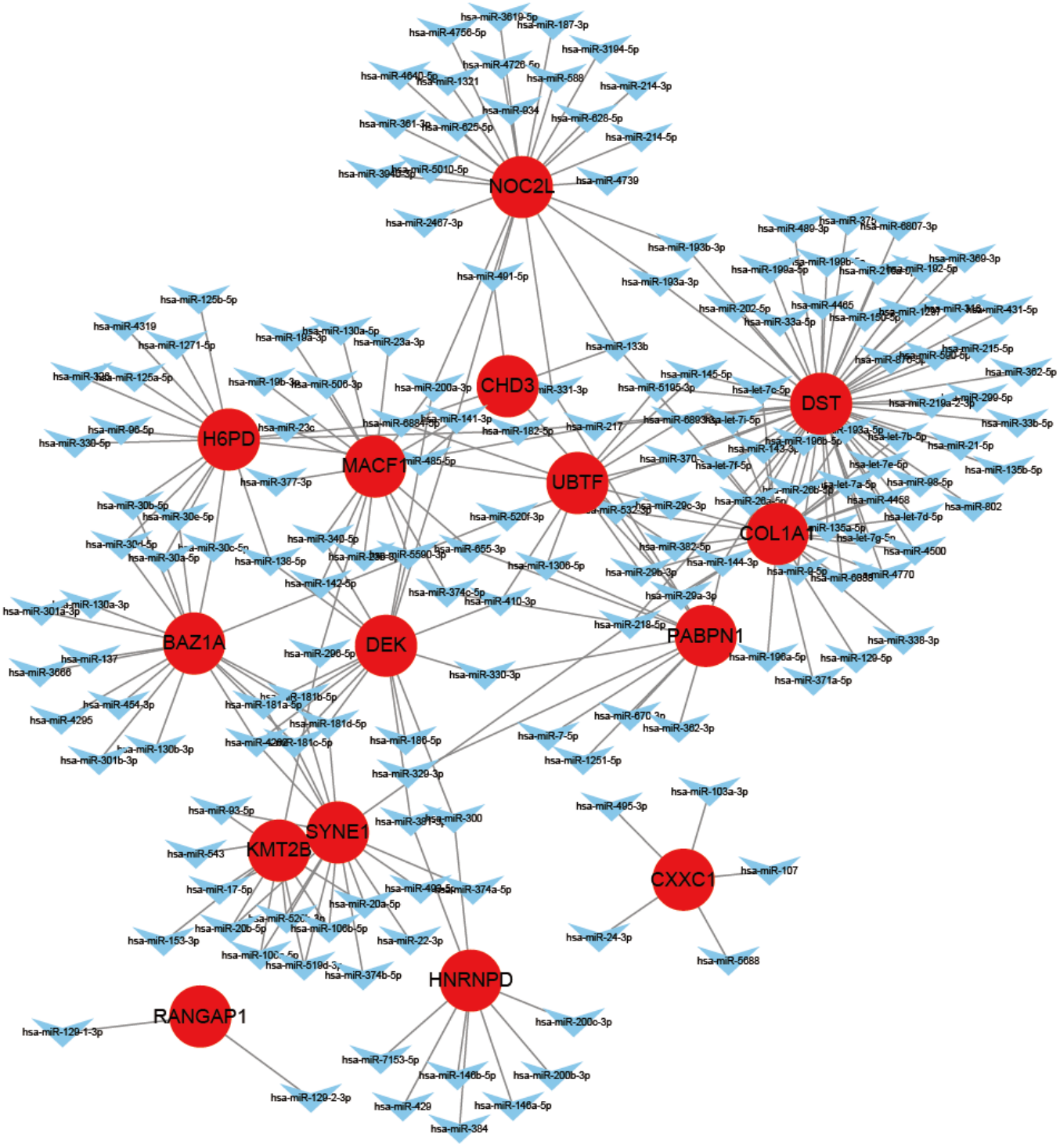

Figure 9

The interaction network between fifteen hub genes and target miRNAs. The red nodes represent fifteen hub genes and the blue nodes represent target miRNAs.

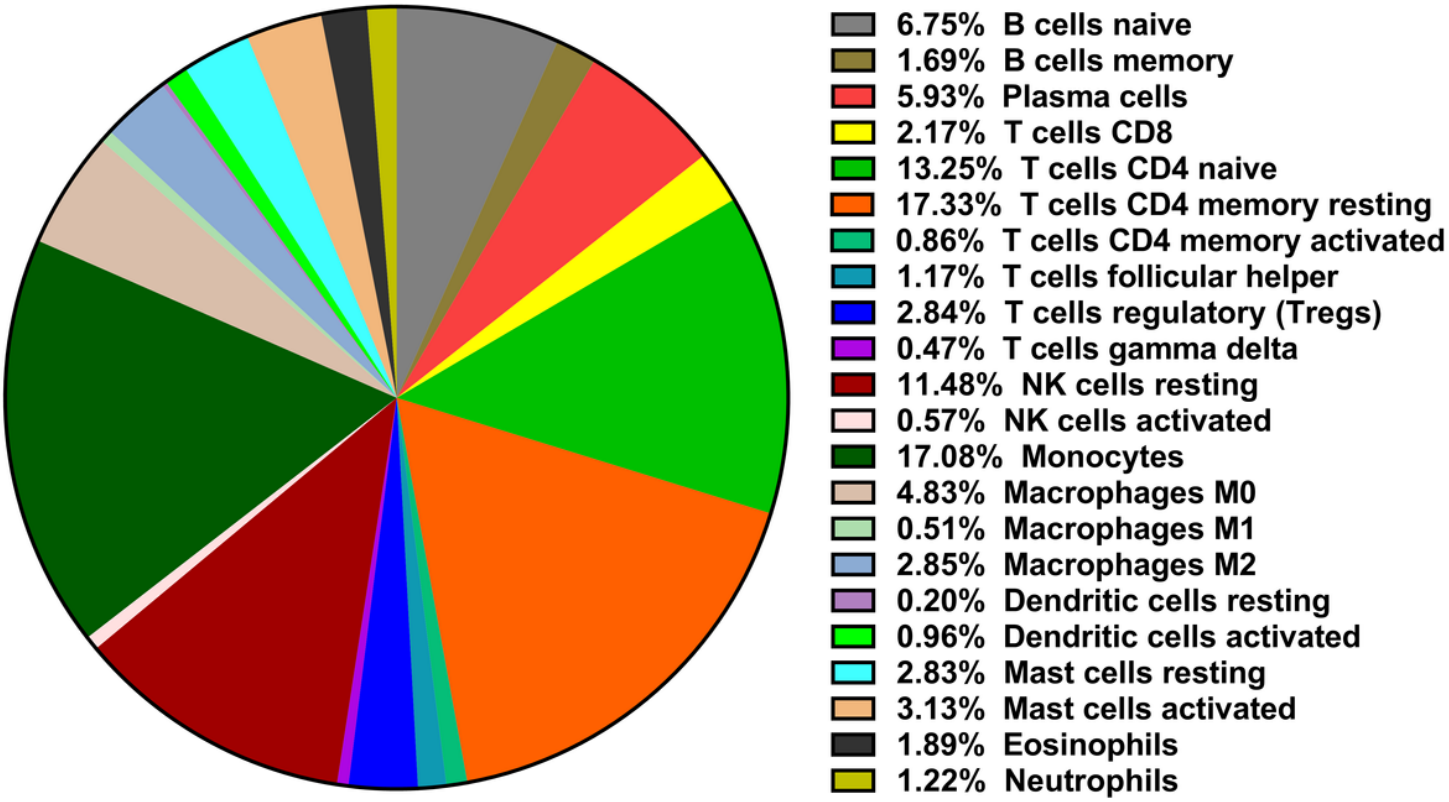

Figure 10

The proportion of 22 immune cells in AKI. T cell CD4 naïve, T cell CD4 memory resting, NK cell resting, Monocytes accounted for a large proportions of AKI immune cell infiltration.

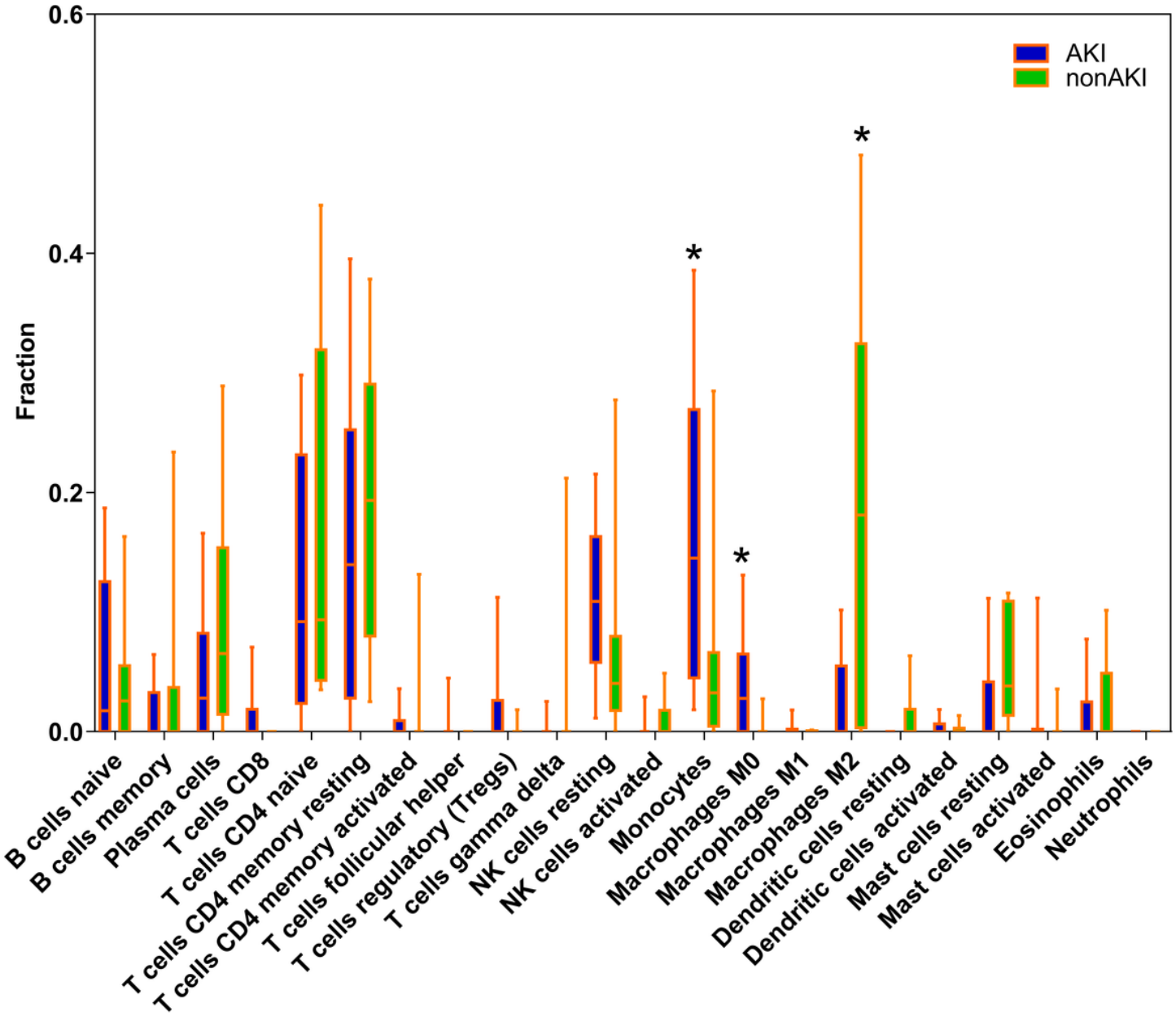

Figure 11

AKI and non-AKI cohort displayed different immune cells expression. The blue color represents AKI cohort and the green color represents non-AKI cohort. Different immune cell type expression was observed between AKI and non-AKI cohort.

Supplementary Files

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

- Tables.docx
- SupplementFigure1.tif

- SupplementFigure2.tif