Identifying hub genes associated with clinical characteristics in IgA nephropathy by Weighted Gene Co-expression Network Analysis

Chengyu Yang
The Affiliated Hospital of Qingdao University

Chenyu Li
The Affiliated Hospital of Qingdao University

Long Zhao
The Affiliated Hospital of Qingdao University

Bin Zhou
The Affiliated Hospital of Qingdao University

Xiaofei Man
The Affiliated Hospital of Qingdao University

Quandong Bu
The Affiliated Hospital of Qingdao University

Congjuan Luo
The Affiliated Hospital of Qingdao University

Xuefei Shen
The Affiliated Hospital of Qingdao University

Wei Jiang
The Affiliated Hospital of Qingdao University

Hong Luan
The Affiliated Hospital of Qingdao University

Lin Che
The Affiliated Hospital of Qingdao University

Yanfei Wang
The Affiliated Hospital of Qingdao University

Yan Xu (✉ xuyanqfy@126.com)
The Affiliated Hospital of Qingdao University

Research article

Keywords: IgA nephropathy; Microarray; Weighted Gene Co-expression Network Analysis; Body Mass Index; Glomerular filtration rate
Abstract

Background: Clinically, IgA nephropathy has a variety of symptoms including paroxysmal gross hematuria, nephritic, and nephrotic syndrome. This study aimed at investigating hub gene and genes modular related to IgA nephropathy clinical characteristics by using weighted gene co-expression network analysis combining clinical, microarray, and network database parameters. Methods: We collected 32 human samples from the European Renal cDNA Bank and used RMA method to preprocess the data and utilize the limma package to obtain differentially expressed gene in renal interstitium and glomeruli. We used the WGCNA package to construct the gene co-expression of differential expression genes and identify hub genes associated with clinical characteristics in renal interstitium and glomeruli, respectively. Gene ontology enrichment analysis and KEGG analysis for hub genes which associated with clinical characteristics were performed by DAVID. PPI information was acquired from STRING. Results: For glomeruli, 1470 genes differentially expressed between IgA nephropathy patients and healthy control, containing 10 hub genes associated with age, 8 hub genes associated with sex, 48 hub genes associated with Bp enriched in ERK1 and ERK2 cascade and Rap1 signaling pathway, 223 hub genes associated with BMI enriched in organic acid catabolic process and fatty acid degradation pathway, 136 hub genes associated GFR enriched in immune response and PI3K-Akt signaling pathway, 82 hub genes associated with proteinuria enriched in extracellular matrix organization and PI3K-Akt signaling pathway. In tubulointerstitium, there were 480 genes differentially expressed between IgA nephropathy patients and healthy control. Among 480 DEGs, 6 hub genes associated with age, 15 hub genes associated with sex, 35 hub genes associated with Bp enriched in positive regulation of apoptotic process, 87 hub genes associated with GFR enriched in negative regulation of macromolecule metabolic process and RNA transport, 33 hub genes associated with proteinuria enriched in regulation of apoptotic process and FoxO signaling pathway. PPI enrichment analysis shown that all hub genes sets are biologically connected cluster. Conclusions: We made a preliminary investigation on molecular mechanisms of relationship between IgA nephropathy and clinical characteristics and identified hub genes and pathways closely related with BMI, GFR and Proteinuria in IgA nephropathy by a series of bioinformatics analysis.

Background

Immunoglobulin A (IgA) nephropathy also known as Berger’s disease initially described in 1968 by Berger and now considered as the most common chronic glomerular disease in the world [1]. Defined by the predominant deposition of IgA in glomerular mesangium by immunofluorescence, IgA nephropathy has light histological changes with glomerular mesangial proliferation, and microscopic hematuria, gross hematuria and even latent onset are the main clinical symptoms of IgA nephropathy. About 30% of patients with IgA nephropathy reached end stage renal disease (ESRD) after 10~20 years of initial diagnosis [2]. Epidemiologically, IgA nephropathy may occur at any age with the peak incidence at 20~30 years old, and males are more susceptible to IgA nephropathy than females with a ratio of 2:1~5:1 for males to females, but the effect of age or sex for prognosis of IgA nephropathy is still uncertain.
In chronic kidney disease (CKD), proteinuria is one of the significant signs for reaching end stage renal disease (ESRD), and the degree of proteinuria associated with IgA nephropathy progression and prognosis. A large number of clinical studies indicated that proteinuria, especially constant large amount of proteinuria (≥3g/24h) is an independent risk factor for poor prognosis of IgA nephropathy: glomerular filtration rate (GFR) of patients with proteinuria>3g/24h was decreased by 20 to 25 times compared to patients with proteinuria<3g/24h; long-term follow-up research indicated that patients with mild proteinuria (proteinuria<0.5g/24h) are more likely to have adverse events with the significant degradation of renal function and significantly elevation of blood pressure [3-5]. Obesity (Body Mass Index, BMI ≥25kg/m²) is now considered as an independent risk factors that may cause or exacerbate chronic kidney disease [6, 7], which can damage the kidneys by affecting renal hemodynamics with glomeruli high blood pressure, high perfusion and high filtration. Meanwhile, hypertension is also a risk factor of progression for IgA nephropathy to end stage renal disease (ESRD). Because of the characteristics of renal vascular structure, early hypertension mainly affect the function of reabsorption in renal tubular and, then, increase the glomerular hyperfiltration and promote glomerular sclerosis leading to the ischemic damage of renal parenchymal, which is the particularly prominent injure in the malignant hypertensive patients. Patients diagnosed with hypertension, especially with proteinuria, generally have worse outcomes, but improved control of blood pressure can effectively reduce the incidence of adverse events [8]. Remarkably, in the progression of IgA nephropathy, the elevation and development of blood pressure is a significant sign for predicting the long-term prognosis of IgA nephropathy [9]. For renal function, many clinical studies have confirmed that the function of glomerular filtration (serum creatinine, GFR) in histology confirmed IgA nephropathy definitely effect on the prognosticate IgA nephropathy. D'Amico retrospective study suggested that serum creatinine levels are the strongest predictors of long-term prognosis of IgA nephropathy: 5 years survival rate of patients with serum creatinine levels <115μmol/L and serum creatinine levels ≥115μmol/L were 98.4% and 41.8% with statistically significant, respectively [10].

Therefore, it is greatly significant to investigate molecular mechanisms of relationship between IgA nephropathy and clinical characteristics, understand the pathogenesis of IgA nephropathy for evaluation of IgA nephropathy disease status and prognosis and guidance diagnosis and treatment of IgA nephropathy.

Weighted Gene Co-expression Network Analysis (WGCNA) is an advanced and comprehensive algorithm for co-expression analysis based on R programming language. At present, gene co-expression networks are increasingly being used to look for functional similar genes, which used a very straightforward concept: the nodes in the network represent the genes, and the significantly co-expressed genes will be clustered together by selecting the appropriate tissue samples. In fact, it is difficult to define whether two nodes can be connected in a network. Traditionally, binary method was used to construct genes co-expression network by defining 1 as connected and 0 as unconnected, but it cannot interpret biological significance between the “hard” threshold 1 or 0. In order to solve this problem, WGCNA used a soft threshold to define a weight value to determine the probability of interaction among a set of genes. Under
this concept, a weighted co-expression network is formed. In order to use soft threshold, the co-expression network was transformed into a weight connection network, and parameters of soft threshold were set by scale-free topology criterion based on biological and statistical significance.

In our study, we used 32 human samples from the European Renal cDNA Bank to construct the gene co-expression of differential expression genes by WGCNA and identify the hub genes associated with clinical characteristics in renal interstitium and glomeruli, respectively. Then, Gene ontology enrichment analysis, KEGG analysis and PPI network analysis were used to further investigate the function of hub genes sets associated with clinical characteristics.

**Methods**

Collecting genes expression data and clinical parameters

A total of 32 human samples from the European Renal cDNA Bank (ERCB) [11], which contains 25 IgA nephropathy tubulointerstitium, 27 IgA nephropathy glomeruli, 6 pretransplant healthy living tubulointerstitial and 5 pretransplant healthy living glomeruli, were collected from GEO [12] database which was used to store gene expression datasets and platform information. For consistent pre-processing, we downloaded total 66 raw data (CEL file) from GSE37463 [13] (the platform is Affymetrix GeneChip Human Genome HG-U133A), a sample dataset which was contributed by Berthier et al. and preformed further analysis. Clinical parameters including age, sex, Bp, BMI, GFR, Scr and proteinuria were collected from nephroseq [14], a database which was used to store microarray datasets and clinical data of kidney disease Clinical characteristics of these patients are provided in Table 1.

Identification of differentially expressed genes

R (version 3.3.4) programming language was applied to data quality assessment, normalization and detection of DEGs. Based on “affyPLM” [15] package, we conducted quality assessment on microarray data by using RNA degradation curve and Normalized Unscaled Standard Errors(NUSE) [16], a simple and sensitive method which is the standard deviation of the PM value of a probe set divided by the median of the standard deviation of PM values over all chip. Then, we used RMA(Robust Multi-array Averaging) [17], a integrative algorithm with fewer false positives, to preprocess the raw data and utilized the Empirical Bayess method to obtain genes differential expression based on limma package, a widely used statistical test in microarray analysis. The expression of ±1.2-fold change and FDR<0.05 was used as the threshold to select DEGs. Finally, genes were annotated by official annotations file (Affymetrix HG-U133A Annotations, Release 35).

Weighted gene co-expression network analysis and identification of hub genes and module
WGCNA \cite{18} is a systematic and robust gene co-expression network algorithm to describe the correlation of gene expression matrix, detect highly correlated genes modules and evaluate the correlation between genes modules and clinical traits. According to the official protocol, WGCNA can be briefly divided into the following steps: 1. construct co-expression network specified by its adjacency matrix that calculated by soft threshold power; 2. transform adjacency into topological overlap matrix (TOM) and using hierarchical clustering and dynamic tree-cutting method to screen module; 3. select module and gene related to external information.

Gene significance\cite{18} measures was used to incorporate clinical characteristics into the gene co-expression network, the higher the absolute GS value, the more biologically significant is the gene. GS is defined as the absolute value of the correlation between the gene profile $x$ and the clinical trait $T$, or defined by minus log of a p-value based on a correlation test or a regression for evaluating the statistical significance between gene profile $x$ and the clinical trait $T$ \cite{18}:

$$GS = |\text{cor}(x, T)|$$

$$GS = -\log p$$

For select hub genes, we used “networkScreening” \cite{18}, a function blends standard and network methods to screening genes (or general variables) highly relevant to a given external trait. On this basis, q.Weight<0.01 (local FDR) \cite{19}, a corrected weighted p-value of association with a clinical characteristic was considered as a threshold to select hub genes.

**Gene ontology analysis of hub genes**

Biological significance of hub genes was explored by Gene ontology \cite{20} (GO) enrichment analysis and KEGG \cite{21} (Kyoto Encyclopedia of Genes and Genomes) analysis based on DAVID \cite{22, 23} (Database for Annotation, Visualization and Integrated Discovery) and Biological Process (BP), the most representative sub-ontologies of GO enrichment analysis, was used to find critical biological function in hub genes closely related to BMI, GFR and proteinuria in tubulointerstitium and glomeruli of IgA nephropathy, respectively. A P<0.01 was considered statically significant and significant enrichment.

**Construct PPI network analysis of hub genes**

Gene “network view” is increasingly being applied to many areas of biology, such as prediction of phenotypes and gene functions, drug discovery and statistical power in genetics. In order to investigate the interactions among hub genes associated with clinical parameters and verify the result of WGCNA, we
look for functional relationships among the gene products and construct PPI network. PPI information was acquired from STRING (Search Tool for the Retrieval of Interacting Genes, http://www.string-db.org/) [24, 25]. Medium confidence (0.4) was chosen be minimum required interaction score to screen the interactions among hub genes. The actived interaction sources including text mining, experiments, databases, co-expression, neighborhood, gene fusion and co-occurrence. Cytoscape [26] (version 3.4.0) software was used for visualization of PPI network. Genes that not display connections with other genes were not shown in the network.

Results

Differentially expressed genes(DEGs) between IgA nephropathy and healthy control

In order to investigate hub genes and genes modules of IgA nephropathy combining clinical parameters, series GSE37463 including 25 IgA nephropathy tubulointerstitium, 27 IgA nephropathy glomeruli, 6 pretransplant healthy living tubulointerstitium and 5 pretransplant healthy living glomeruli was downloaded from GEO [12] database; clinical parameters including age, sex, blood pressure(Bp), BMI, GFR, serum creatinine (Scr) and proteinuria were collected from website database Nephroseq (Figure 1ab). After raw data quality assessment [16] which showed all sample are qualified data for further analysis and data normalization by RMA [17], we used packages limma (Linear Models for Microarray Analysis) [27] for detecting DEGs. The FDR-value for fold-change between IgA nephropathy and healthy control was calculated by moderated t-test, and fold change± 1.2 and FDR<0.05 were considered as the threshold to select DEGs. For glomeruli, there were altogether 1802 differentially expressed RNAs(fold change>1.2, FDR<0.05) containing 1470 unique genes annotated by official annotations, among which consisting significantly 53 down-regulated DEGs and 162 up-regulated DEGs(fold change>2, FDR<0.05); for tubulointerstitium, there were 586 differentially expressed RNAs(fold change>1.2, FDR<0.05) containing 480 unique genes, among which consisting significantly 52 down-regulated DEGs and 3 up-regulated DEGs (fold change>2, FDR<0.05).

Weighted gene co-expression network analysis

Since genes with relationship of function or regulation tend to be a similar expression pattern, we constructed the WGCNA network to sort similar regulatory gene into modules and explore highest correlation module with IgA nephropathy or clinical parameters. After choosing an appropriate soft-thresholding power to fit the scal-free topology index around 0.9, we utilized one-step method “blockwiseModules” which applied hierarchical clustering and dynamic tree cut algorithm to screen modules in both group of tubulointerstitium and glomeruli (Figure 2ac). Subsequently, by incorporating external information into the WGCNA network, Figure 2bd shows correlation coefficients between every module and clinical parameters. Since GFR was a sensitive and reliable index reflecting renal function in IgA nephropathy, we used GFR instead of Scr for further analysis.
Figure 3 and 4 shows Gene Significance (GS) across every module related to IgA nephropathy and clinical parameters. For glomeruli, all modules were high correlation (|r|>0.6, P<0.01) with IgA nephropathy occurrence and brown module (210 genes) have the highest GS(P<0.01) related to IgA nephropathy occurrence; blue (415 genes), yellow (190 genes) and green (79 genes) modules were correlation (|r|>0.5, P<0.01) with BMI in IgA nephropathy and yellow module have the highest GS(P<0.01) related to BMI in IgA nephropathy; turquoise module (418 genes) was correlated (|r|>0.4, P<0.05) to BMI, GFR, Scr and proteinuria with highest GS(P<0.01). For tubulointerstitium, turquoise module were associated with IgA nephropathy occurrence (|r|>0.9, P<0.01) and grey were associated with GRF, and Scr (|r|>0.5, P<0.01) with highest GS(P<0.01). It is worth noting that since the expression pattern of genes is very similar, WGCNA only screened one turquoise module in IgA nephropathy tubulointerstitium and genes that cannot be clustered into one of modules are assigned to the grey module which represents background genes outside of module. Thus, genes possibly related to GFR, Scr in grey module but not belong to WGCNA modules.

Hub genes were selected by function “networkScreening” and q.Weight<0.01 (local FDR), a corrected weighted p-value of association with a external trait was considered as threshold to select hub genes: In glomeruli, there are 10 hub genes associated with age, 8 hub genes associated with sex, 48 hub genes associated with Bp, 223 hub genes associated with BMI, 82 hub genes associated proteinuria and 136 hub genes associated GFR; In tubulointerstitium, there are 6 hub genes associated with age, 15 hub genes associated with Bp, 7 hub genes associated with BMI, 33 hub genes associated with proteinuria and 87 hub genes associated with GFR. Genesets with more than 40 hub genes were used to perform further PPI network analysis to verify the result of WGCNA by using authenticated genes regulatory relationship.

**Gene ontology and KEGG pathways analysis of hub genes**

Gene ontology enrichment analysis and KEGG analysis for hub genes which associated with BMI, GFR and proteinuria were performed by using DAVID. Firstly, we performed GO enrichment analysis. For glomeruli in IgA nephropathy, genes significantly associated with BMI, GFR and proteinuria were enriched in: ERK1 and ERK2 cascade and cellular response to organic substance (48 hub genes associated with Bp); small molecule catabolic process and organic acid catabolic process (223 hub genes associated with BMI); immune response and regulation of immune response (136 hub genes associated with GFR); and extracellular matrix organization and extracellular structure organization (82 hub genes associated with proteinuria) (Table 1). For tubulointerstitium in IgA nephropathy, 35 hub genes associated with Bp were enriched in positive regulation of apoptotic process and positive regulation of programmed cell death; 87 hub genes associated with GFR were enriched in negative regulation of macromolecule metabolic process and negative regulation of metabolic process; 33 genes significantly associated with proteinuria were enriched in regulation of apoptotic process and regulation of programmed cell death (Table 2).
KEGG pathways were analyzed on the same gene to verify GO enrichment analysis result. For glomeruli in IgA nephropathy, hub genes associated with Bp were enriched in central carbon metabolism in cancer and Rap1 signaling pathway; genes significantly associated with BMI were enriched in type I diabetes mellitus and fatty acid degradation; hub genes associated with GFR were enriched in PI3K-Akt signaling pathway and endocytosis; genes significantly associated with proteinuria were enriched in PI3K-Akt signaling pathway and ECM-receptor interaction (Table 1). For tubulointerstitium in IgA nephropathy, hub genes associated with Bp were enriched in PI3K-Akt signaling pathway metabolic pathways (P > 0.05); genes significantly associated with GFR were enriched in RNA transport and protein processing in endoplasmic reticulum; genes significantly associated with proteinuria were enriched in FoxO signaling pathway and hepatitis B (Table 2).

**Protein-Protein interaction network analysis of hub genes**

PPI analysis was used to obtain further relevant information between hub genes and verify the result of WGCNA by authenticated genes regulatory relationship based on STRING [28] (Search Toll for the Retrieval of Interacting Genes), a public database and tool that construct and display functional protein association networks. Noteworthy, since the hub genes associated with BMI are not duplicated with GFR or Proteinuria in glomeruli of IgA nephropathy, we analyzed these genes separately: the PPI enrichment P-value of genes set associated with BMI, $1.78 \times 10^{-8}$, which means significantly more interaction among these genes than expected from random assignments using proteins of similar size drawn from genome, indicating that the products of those genes are at least partially biologically connected (Figure 5); the network PPI enrichment P-value of genes set associated with GFR and Proteinuria is 0 (Figure 6). For tubulointerstitium in IgA nephropathy, we used genes associated with BMI, GFR and Proteinuria to construct PPI network and the PPI enrichment P-value of this network is $7.33 \times 10^{-5}$ (Figure 7). All three networks have more interactions than expected from random proteins of similar size or drawn from the genome. Moreover, in the PPI network, genes in one module or related to same clinical parameter are more likely to have interaction or be in a cluster, indicating that the WGCNA was successful preformed and the results of gene co-expression analysis were verified by the confirmed PPI experiment.

**Discussion**

A large number of studies have confirmed that BMI, hypertension, proteinuria and renal function effect on the prognosticate IgA nephropathy, however, the genetic mechanism of these clinical characteristics caused progression of IgA nephropathy are not clear, and the relationship and association between factors (like age sex) and IgA nephropathy are still needed more study and evidence on biological processes at the molecular level. Noteworthy, disease always involved thousands of gene expression changes with a huge and complex gene regulated network, which means are search for a single gene is superficial and unreliable and it is hard to explain the mechanism of a disease. WGCNA is an advanced, successful and comprehensive algorithm for co-expression analysis, and it was not only used to
construct gene co-expression network and screen gene modules, but also a powerful tool to identify hub genes associated external information, especially clinical characteristics, and help researchers to understand the mechanism of the disease and providing a theoretical basis for the diagnosis and treatment of the disease.

In the present study, we collected 32 human samples from the European Renal cDNA Bank and used RMA to preprocession and utilize the limma to obtain DEGs in renal interstitium and glomeruli. For glomeruli, there were altogether 1470 differentially expressed unique genes (fold change>1.2, FDR<0.05). These DEGs were used to construct the gene co-expression and identify the hub genes associated with clinical characteristics by WGCNA: brown module has the highest GS related to IgA nephropathy occurrence; blue, yellow and green modules were correlation with BMI with highest GS of yellow module; turquoise module was correlated to GFR, Scr and proteinuria with highest GS. However, since the genes associated with age or sex are quite few and the genes associated with Bp allocated among many modules, no modules specifically associated with age, sex or Bp. Furthermore, we identify 10 hub genes (HK1, HEY1, MCAM, GPR4, SPRY4, NETO2, DCBLD2, MSX1, GPR124, LPPR2) associated with age, 8 hub genes (CTH, EAF2, LAMTOR5, ZNF331, PRKX, CD99, FECH, DDX3X) associated with sex, 48 hub genes associated with Bp, 223 hub genes associated with BMI, 82 hub genes associated with proteinuria and 136 hub genes associated GFR.

Then, the results of GO, KEGG pathways and PPI network analysis validated the results of WGCNA: 223 hub genes associated with BMI in glomeruli enriched in small molecule catabolic process, organic acid catabolic process and fatty acid degradation pathway; PPI network analysis indicated that 223 hub genes associated with BMI belong to abiological cluster. Above all, these 223 hub genes not only interrelated with IgA nephropathy but also specificity associated with BMI, which provides directions for future research of relationship between BMI and IgA nephropathy.

Meanwhile, 48 hub genes associated with Bp enriched in ERK1 and ERK2 cascade, cellular response to organic substance, central carbon metabolism in cancer and Rap1 signaling pathway; 136 hub genes associated GFR in glomeruli enriched in immune response, cellular response to chemical stimulus and PI3K-Akt signaling pathway, indicated that IgA stimulus and immune response were dominant in impaction of glomerular filtration. It is known that glycosylated IgA has a transferrin receptor (CD71) on the surface of mesangial cells [29] and abnormal glycosylated IgA immune complexes are specifically recognized and deposited in the mesangium causing proinflammatory cytokines and angiotensin released [30], and tumor necrosis factor alpha (TNF-α) derived from IgA nephropathy patients podocytes cells autocrine synthesis caused TNF receptor 1 (TNFR1), TNF receptor 2 (TNFR2) and IL-6 were up-regulated. Elevated expression of TNFR1 leads to podocyte apoptosis and up-regulation of TNFR2 expression leading to chronic inflammation [31]; Phosphatidylinositol-3 kinase (PI3K)-serine/threonine kinase (Akt) is an important pathway in intracellular involve in cell metabolism, apoptosis, proliferation and differentiation [32, 33]. A study by Cox et al. reported that PI3K-Akt signaling pathway was hyperactive in IgA nephropathy patients and played an important role in IgA nephropathy.
Additionally, based on the results of WGCNA, we believed that PI3K-Akt signaling pathway specificity impact the renal function in IgA nephropathy.

For proteinuria, 82 hub genes in glomeruli enriched in extracellular matrix organization, extracellular structure organization and PI3K-Akt signaling pathway, meaning that these genes may played an important role in changing the extracellular matrix and leading to proteinuria. Traditionally, the disruption of glomerular filtration barrier (GFB), a 3-layer structure consisted of endothelial cells, glomerular base membrane (GBM) and podocyte, is the main reason leading to proteinuria. There is plenty of evidence supporting GFB molecular sieve effect and solute with different size are obstructed in varying degrees with the water and small molecules solute free permeability and macromolecular selective permeability, while GFB also is charge barrier preventing anionic molecules such as albumin passing through the GBM [35, 36]. It is now believed that GFB also consists by two additional layers, endothelial surface layer (ESL) and sub podocyte space (SPS). ESL and SPS have solute molecular screening characteristics of glomerular filtration and has a significant impact in renal function [37, 38].

In tubulointerstitium, there were 480 DEGs between IgA nephropathy and healthy control. Among 480 DEGs, 6 hub genes (HES1, ACAD10, GEM, RHOB, CREM, FILIP1L) associated with age, 15 hub genes (MFAP1, RPL21, ZMYM4, CTNNA1, EIF5, CCT6A, DNAJC10, MIR7110, TRAPP11, CALCR, TTC37, EPQ7, TUFT1, NUP62, ELL2) associated with sex, 35 hub genes associated with Bp enriched in positive regulation of apoptotic process, cellular response to nutrient levels and regeneration. Moreover, 87 hub genes associated with GFR in tubulointerstitium enriched in negative regulation of macromolecule metabolic process, negative regulation of metabolic process and RNA transport, and 33 hub genes associated with proteinuria enriched in regulation of apoptotic process, regulation of programmed cell death and FoxO signaling pathway. Proteinuria is closely associated with poor cardiovascular outcomes and progression of end-stage renal disease in patients with chronic kidney disease [39, 40]. Our results shown both Bp and proteinuria are related to apoptotic process in tubulointerstitium, indicating hypertension casued damage or apoptosis of cell in tubulointerstitium also leading to Proteinuria. Additionally, in vitro, mesangial cell in IgA nephropathy patients mainly produces humoral factors TNF, TGF-β and angiotensin II, which passed through glomerular filtration barrier or transported by blood to the renal tubules Interstitial, causing localized inflammation cascade amplification and damaging renal tubular epithelial cell [41]. It's remarkable that cysteine-rich angiogenic inducer 61 (Cyr61) [42], a heparin binding activity of secreted protein as matrix related signaling molecules involved in cell proliferation, differentiation, transformation and apoptosis, is one of the hub gene associated with GFR in tubulointerstitium. Moreover, our previous studies have shown protection of renal tubular epithelial cells from apoptosis by Cyr61 expression under hypoxia [43, 44]. Therefore, our results of WGCNA and Gene ontology enrichment analysis were consistent with previous studies results, at least in part, Cyr61 as a matrix molecule related to apoptosis although the underlying mechanisms of proteinuria in IgA nephropathy still need to be explored.

Conclusions
We made a preliminary investigation on molecular mechanisms of relationship between IgA nephropathy and clinical characteristics and identified hub genes and pathways closely related with BMI, GFR and Proteinuria in IgA nephropathy by a series of bioinformatics analysis. Our research gives a direction, but it still need to further explored and validated.

**Abbreviations**

BMI: Body Mass Index; Bp: blood pressure; GFR: glomerular filtration rate; IgA: Immunoglobulin A; ESRD: end-stage renal disease; CKD: In chronic kidney disease; Scr: serum creatinine; WGCNA: Weighted Gene Co-expression Network Analysis; GEO: Gene Expression Omnibus; RMA: Robust Multi-array Averaging; Limma: Linear Models for Microarray Analysis; GO: Gene Ontology; BP: Biological Process; Kyoto Encyclopedia of Genes and Genomes: KEGG; DAVID: Database for Annotation, Visualization and Integrated Discovery; STRING: Search Tool for the Retrieval of Interacting Genes; GS: Gene Significance; GFB: glomerular filtration barrier; GBM: glomerular base membrane; ESL: endothelial surface layer; SPS: subpodocyte space; Cyr61: cysteine-rich angiogenic inducer 61; DEG: differentially expressed gene; IgAN: Immunoglobulin A nephropathy; LD: Healthy Living Donor; FDR: False discovery rate.

**Declarations**

**Ethics approval and consent to participate**

All data were obtained via public database. All the data-devoted researches had acquired informed consent of patients and ethical license in written format according to the data contributors.

**Consent for publication**

Not applicable.

**Availability of data and materials**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This study was supported by the National Natural Science Foundation of China (81770679). The National Natural Science Foundation of China (81770679) was used for the design of the study, the collection of data, the analysis of data and data processing.

**Author’s contributions**
CY, C Li and BZ designed the studies; LZ, XM, XS, WJ, HL, YW and LC performed the study and analyzed the data; YX and QB discussed the results; CY, C Li and C Luo wrote the paper. All authors have read and approved the manuscript.

Acknowledgements

We thank all those colleagues who provided support and help in various forms to this research. We thank all the contributors of the microarrays and clinical data.

References

1. Wyatt RJ, Julian BA: IgA nephropathy. The New England journal of medicine 2013, 368(25):2402-2414.
2. Le W, Liang S, Hu Y, Deng K, Bao H, Zeng C, Liu Z: Long-term renal survival and related risk factors in patients with IgA nephropathy: results from a cohort of 1155 cases in a Chinese adult population. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association 2012, 27(4):1479-1485.
3. Reich HN, Troyanov S, Scholey JW, Cattran DC, Toronto Glomerulonephritis R: Remission of proteinuria improves prognosis in IgA nephropathy. Journal of the American Society of Nephrology : JASN 2007, 18(12):3177-3183.
4. Moriyama T, Tanaka K, Iwasaki C, Oshima Y, Ochi A, Kataoka H, Itabashi M, Takei T, Uchida K, Nitta K: Prognosis in IgA nephropathy: 30-year analysis of 1,012 patients at a single center in Japan. PloS one 2014, 9(3):e91756.
5. Lee SM, Rao VM, Franklin WA, Schiffer MS, Aronson AJ, Spargo BH, Katz AI: IgA nephropathy: morphologic predictors of progressive renal disease. Human pathology 1982, 13(4):314-322.
6. Ross WR, McGill JB: Epidemiology of obesity and chronic kidney disease. Advances in chronic kidney disease 2006, 13(4):325-335.
7. Berthoux F, Mariat C, Maillard N: Overweight/obesity revisited as a predictive risk factor in primary IgA nephropathy. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association 2013, 28 Suppl 4:iv160-166.
8. Kamei K, Nakanishi K, Ito S, Ishikura K, Hataya H, Honda M, Nozu K, Iijima K, Shima Y, Yoshikawa N et al: Risk factors for persistent proteinuria after a 2-year combination therapy for severe childhood IgA nephropathy. Pediatric nephrology 2015, 30(6):961-967.
9. Bartosik LP, Lajoie G, Sugar L, Cattran DC: Predicting progression in IgA nephropathy. American journal of kidney diseases : the official journal of the National Kidney Foundation 2001, 38(4):728-735.
10. D'Amico G: Natural history of idiopathic IgA nephropathy and factors predictive of disease outcome. Seminars in nephrology 2004, 24(3):179-196.
11. Cohen CD, Kretzler M: [Gene expression analyses of kidney biopsies: the European renal cDNA bank--Kroner-Fresenius biopsy bank]. Der Pathologe 2009, 30(2):101-104.

12. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM et al: NCBI GEO: archive for functional genomics data sets--10 years on. Nucleic acids research 2011, 39(Database issue):D1005-1010.

13. Gene Expression Omnibus,https://www.ncbi.nlm.nih.gov/gds/?term=GSE37463. Accessed 10 November 2018.

14. Nephroseq, www.nephroseq.org. Accessed 30 November 2018.

15. Rajagopalan D: A comparison of statistical methods for analysis of high density oligonucleotide array data. Bioinformatics (Oxford, England) 2003, 19(12):1469-1476.

16. Heber S, Sick B: Quality assessment of Affymetrix GeneChip data. Omics : a journal of integrative biology 2006, 10(3):358-368.

17. Giorgi FM, Bolger AM, Lohse M, Usadel B: Algorithm-driven artifacts in median polish summarization of microarray data. BMC bioinformatics 2010, 11:553.

18. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 2008, 9:559.

19. Storey JD, Tibshirani R: Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America 2003, 100(16):9440-9445.

20. Gene Ontology C: Gene Ontology Consortium: going forward. Nucleic acids research 2015, 43(Database issue):D1049-1056.

21. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research 2017, 45(D1):D353-D361.

22. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 2009, 4(1):44-57.

23. Database for Annotation, Visualzation and Integrated Discovery, https://david.ncifcrf.gov/. Accessed 20 December 2018.

24. 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. Nucleic acids research 2017, 45(D1):D362-D368.

25. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Mingeau P, Doerks T, Stark M, Muller J, Bork P et al: The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research 2011, 39(Database issue):D561-568.

26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, 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.

27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 2015, 43(7):e47.
28. STRING, https://string-db.org/. Accessed 3 February 2019.

29. Haddad E, Moura IC, Arcos-Fajardo M, Macher MA, Baudouin V, Alberti C, Loirat C, Monteiro RC, Peuchmaur M: Enhanced expression of the CD71 mesangial IgA1 receptor in Berger disease and Henoch-Schönlein nephritis: association between CD71 expression and IgA deposits. Journal of the American Society of Nephrology : JASN 2003, 14(2):327-337.

30. Lai KN, Tang SC, Guh JY, Chuang TD, Lam MF, Chan LY, Tsang AW, Leung JC: Polymeric IgA1 from patients with IgA nephropathy upregulates transforming growth factor-beta synthesis and signal transduction in human mesangial cells via the renin-angiotensin system. Journal of the American Society of Nephrology : JASN 2003, 14(12):3127-3137.

31. Lai KN, Leung JC, Chan LY, Saleem MA, Mathieson PW, Tam KY, Xiao J, Lai FM, Tang SC: Podocyte injury induced by mesangial-derived cytokines in IgA nephropathy. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association 2009, 24(1):62-72.

32. Wiza C, Nascimento EB, Ouwens DM: Role of PRAS40 in Akt and mTOR signaling in health and disease. American journal of physiology Endocrinology and metabolism 2012, 302(12):E1453-1460.

33. Xu N, Lao Y, Zhang Y, Gillespie DA: Akt: a double-edged sword in cell proliferation and genome stability. Journal of oncology 2012, 2012:951724.

34. Cox SN, Sallustio F, Serino G, Pontrelli P, Verrienti R, Pesce F, Torres DD, Ancona N, Stifanelli P, Zaza G et al: Altered modulation of WNT-beta-catenin and PI3K/Akt pathways in IgA nephropathy. Kidney international 2010, 78(4):396-407.

35. Scott RP, Quagggin SE: Review series: The cell biology of renal filtration. The Journal of cell biology 2015, 209(2):199-210.

36. Patrakka J, Tryggvason K: Molecular make-up of the glomerular filtration barrier. Biochemical and biophysical research communications 2010, 396(1):164-169.

37. Daniels BS: The role of the glomerular epithelial cell in the maintenance of the glomerular filtration barrier. American journal of nephrology 1993, 13(5):318-323.

38. Neal CR, Muston PR, Njegovand D, Verrill R, Harper SJ, Deen WM, Bates DO: Glomerular filtration into the subpodocyte space is highly restricted under physiological perfusion conditions. American journal of physiology Renal physiology 2007, 293(6):F1787-1798.

39. Chronic Kidney Disease Prognosis C, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, Coresh J, Gansevoort RT: Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet 2010, 375(9731):2073-2081.

40. Hemmelgarn BR, Manns BJ, Lloyd A, James MT, Klarenbach S, Quinn RR, Wiebe N, Tonelli M, Alberta Kidney Disease N: Relation between kidney function, proteinuria, and adverse outcomes. Jama 2010, 303(5):423-429.

41. Chan LY, Leung JC, Tsang AW, Tang SC, Lai KN: Activation of tubular epithelial cells by mesangial-derived TNF-alpha: glomerulotubular communication in IgA nephropathy. Kidney international 2005,
42. Chen Y, Du XY: Functional properties and intracellular signaling of CCN1/Cyr61. *Journal of cellular biochemistry* 2007, **100**(6):1337-1345.

43. Ma R, Zhang J, Liu X, Yue S, Zhao Q, Xu Y: 7,8-DHF Treatment Induces Cyr61 Expression to Suppress Hypoxia Induced ER Stress in HK-2 Cells. *BioMed research international* 2016, **2016**:5029797.

44. Xu Y, Shen XF, Ma RX, Jiang W, Zhang W.: Protection of renal tubular epithelial cells from apoptosis by Cyr61 expression under hypoxia. *Cell Biology International Reports* 2014: 1-4.

**Figures**

**Figure 1**

Sample of dendrogram and trait heatmap a. Sample of tubulointerstitium dendrogram and trait heatmap. In heatmap, each row corresponds to a clinical characteristic with color-coded: for age, Bp, BMI, GFR, creatinine and Proteinuria, white means low, red means high and grey means missing data; for sex, red represents male and white represents female; for last row, red means IgA nephropathy patient, white means Healthy living donor(LD). a. Sample of glomeruli dendrogram and trait heatmap.
Figure 2

Weighted gene co-expression network analysis in IgA nephropathy glomeruli and tubulointerstitium. a. Clustering dendrogram of genes in glomeruli, with dissimilarity based on topological overlap, and assigned module colors. b. associations of module and traits in glomeruli. Each column corresponds to a trait, row to a module eigengene. Each cell contains the corresponding correlation, with color-coded according to the color legend, and P-value. c. Clustering dendrogram of genes and assigned module in tubulointerstitium. d. associations of module and traits in tubulointerstitium.
Module significance related to clinical characteristics in IgA nephropathy glomeruli a. Module significance related to IgA nephropathy. b. Module significance related to BMI b. Module significance related to GFR. d. Module significance related to proteinuria.

Gene significance across modules related to IgAN, GFR and Proteinuria in tubulointerstitium

Figure 3

Figure 4
Module significance related to clinical characteristics in IgA nephropathy glomeruli.

a. Module significance related to IgA nephropathy.
b. Module significance related to GFR.
c. Module significance related to proteinuria.

Figure 5

PPI enrichment analysis of hub genes associated with BMI in glomeruli. Colors of nodes represent the module that genes belong to. Genes that do not display connections with other genes were not shown in the network. PPI enrichment analysis parameters: number of nodes: 218; number of edges: 227; average node degree: 2.08; avg. local clustering coefficient: 0.346; expected number of edges: 153; PPI enrichment p-value: 1.78e-08.
Figure 6

PPI enrichment analysis of hub genes associated with GFR and Proteinuria in glomeruli. Colors of nodes represent the module that genes belong to and the Colors of circle represents gene relevant clinical characteristics. Genes that not display connections with other genes were not shown in the network. PPI enrichment analysis parameters: number of nodes: 168; number of edges: 390; average node degree: 4.46; avg. local clustering coefficient: 0.375; expected number of edges: 174; PPI enrichment p-value: 0.
Figure 7

PPI enrichment analysis of hub genes associated with BMI in tubulointerstitium. Colors of nodes represent the module that genes belong to and the colors of circle represent gene relevant clinical characteristics. Genes that not display connections with other genes were not shown in the network. PPI enrichment analysis parameters: number of nodes: 111; number of edges: 94; average node degree: 1.69; avg. local clustering coefficient: 0.361; expected number of edges: 62; PPI enrichment p-value: 7.33e-05.