Bioinformatic Analysis Reveals Novel Hub Genes pathways associated with IgA nephropathy

xue Jiang  
Hangzhou Hospital of Traditional Chinese Medicine

Zhijie Xu  
Zhejiang University School of Medicine First Affiliated Hospital

Yuanyuan Du  
Hangzhou Hospital of Traditional Chinese Medicine

Hongyu Chen (✉ chenhy2016@126.com)  
Hangzhou Hospital of Traditional Chinese Medicine  https://orcid.org/0000-0002-8207-1421

Research

Keywords: IgA nephropathy, Gene Expression Profiling, Bioinformatic Analysis

DOI: https://doi.org/10.21203/rs.3.rs-29974/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background

Immunoglobulin A nephropathy (IgAN) is the most common primary glomerulopathy worldwide. However, the molecular events underlying IgAN remain to be fully elucidated. The aim of the study is to identify novel biomarkers of IgAN through bioinformatics analysis and elucidate the possible molecular mechanism.

Methods

Based on the microarray data GSE93798 and GSE37460 were downloaded from the Gene Expression Omnibus database, the differentially expressed genes (DEGs) between IgAN samples and normal controls were identified. With DEGs, we further performed a series of functional enrichment analyses. Protein-protein interaction (PPI) networks of the DEGs were built with the STRING online search tool and visualized by using Cytoscape, then further identified the hub gene and most important module in DEGs, Biological Networks Gene Oncology tool (BiNGO) were then performed to elucidate the molecular mechanism of IgAN.

Results

A total of 148 DEGs were recognized, consisting of 53 upregulated genes and 95 downregulated genes. GO analysis indicates that DEGs for IgAN were mainly enriched in extracellular exosome, region and space, fibroblast growth factor stimulus, inflammatory response, and innate immunity. The modules analysis showed that genes in the top 1 significant modules of PPI network were mainly associated with innate immune response, integrin-mediated signaling pathway and inflammatory response respectively. The top 10 hub genes were constructed in PPI network, which could well distinguish the IgAN and control group in monocytes sample and tissue sample. We finally identified ITGB2 and FCER1G gene may have important roles in the development of IgAN.

Conclusions

We identified a series of key genes along with the pathways that were most closely related with IgAN initiation and progression. Our results provide a more detailed molecular mechanism for the development of IgAN and novel candidate genes targets of IgAN.

Background

IgA nephropathy (IgAN), the most prevalent type of glomerulonephritis in humans, is characterized by mesangial cell proliferation, the expansion of the glomerular mesangial matrix. Nearly 25%–30% of affected patients go on to develop end-stage renal disease. Right now a number of clinical biomarkers have been identified to be associated with IgAN progression, such as proteinuria, serum creatinine, hypertension and advanced histological involvement [1]. In 2011[2], Suzuki et al hypothesized that the
pathogenesis of IgAN was based on four hits, first the occurrence of an abnormal IgA1 glycosylation process leading to galactose-deficient IgA1 (Gd-IgA1); second, the formation of antiglycan antibodies against Gd-IgA1; third, the formation of nephrogenic circulating immune complexes; fourth, the deposition of these complexes in the mesangium of glomeruli leading to renal injury with variable clinical expressions. While the exact pathogenesis is not very clear.

Many studies have also shown a genetic predisposition to IgAN [3]. Serino et al found six miRNAs significantly up-regulated, two of these modulating the O-glycosylation process of IgA1. Specifically, let-7b regulates the gene GALNT2 and miR–148 modulates the gene target C1GALT1. which has been considered to be an underlying biomarker to predict the probability of IgAN [4–5]. Wang et al found low urinary levels of miR–29b and miR–29c that correlated with proteinuria and renal function. High levels of miR–93 were correlated with glomerular scarring. miR–200a, miR–200b, and miR–429 also have been suggested as potential biomarkers for monitoring the progression of the disease at the renal level process in IgAN patients [6]. However, due to lack of large-scale studies, the limitation of animal models and current low throughput genetic studies the crucial genes involved in the development and effective treatment of IgAN have remained elusive.

As the development of bioinformatics study which has been widely used in various fields to excavate potential information and reveal underlying mechanics and is used in various diseases. Recently bioinformatics analysis has also gradually provided insight into the molecular mechanisms of kidney disease. Such as PSMB8 as a novel hub gene, play a significant role in the occurrence of membrane nephropathy [7]. In lupus, bioinformatic analysis revealed that CD38 and CCL2 were hub macrophage-related genes [8]. EST1 might be a drug target for diabetic nephropathy treatment [9]. While right now, only few bioinformatics analyses have been performed on IgAN. the critical genes and the interaction have not been fully investigated.

In the present study, two original microarray datasets were selected from Gene Expression Omnibus (GEO) database. After identifying the differentially expressed genes (DEGs) in IgAN patients and control group. We employed the Database for Annotation, Visualization and Integration Discovery (DAVID) to identify the functions of the identified DEGs and perform Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The network of protein-protein interaction (PPI) was generated using the STRING database, hub genes and the most significant module in PPI networks were identified using cytoHubba and Molecular Complex Detection (MCODE) plug-in.

The present study aimed to found out potential novel candidate hub gene for the diagnosis and treatment of IgAN.

**Methods**

1. Microarray data
The microarray data was downloaded from GEO database (http://www.ncbi.nlm.nih.gov/geo) using IgAN as search term. GSE93798 was based on Affymetrix Human GeneChip U133 2.0 platform (includes 42 samples, 20 IgAN patients and 22 healthy controls. GSE37460 was based on Human Genome U133A Affymetrix platform (includes 54 samples, 27 IgAN patients and 27 healthy controls)

2. Identification of DEGs

The DEGs were identified based on the series matrix file using Limma package in R software (version 3.5.0), An adj P. Value < 0.01 and a |log FC (fold change) | ≥ 1 were defined as the thresholds for DEGs screening. The DEGs overlapped among the two datasets were identified and then used for further Functional Enrichment analysis. The overlap DEGs were subjected to bidirectional hierarchical clustering analysis using the Pheatmap package in R to recognize and visualize the differences of DEGs between the IgAN and the control.

3. Enrichment analysis of the DEGs

The DAVID(http://david.abcc.Ncifcrf.gov/) [10] tool was used to conduct GO/KEGG (http://www.genome.jp/kegg/pathway.html) [11] pathway enrichment analyses for DEGs. The number of enrichment genes (count number) ≥ 2 and p-value < 0.05 were chosen as cut-off criteria.

4. PPI network construction, module analysis hub gene identification

DEGs identified were subjected to PPI analysis by using the search functionality of STRING (http://string.embl.de/) [12] to explore the association between the DEGs, and a network interaction matrix was built. An interaction with a combined score > 0.4 was set as the cut-off value. Then the Network was visualized using Cytoscape software [13], which is a broadly used tool for the visual exploration of interaction networks among numerous biomolecules including proteins and genes. The MCODE plug-in was used to identify the most significant module in PPI networks with MCODE scores > 5, degree cut-off = 2, node score cut-off = 0.2, Max depth = 100 and k score = 2. CytoHubba [14] is a tool used to identify hub objects and subnetworks from a complex interactome. ‘MCC’ is a topological analysis method in CytoHubba. ‘MCC’ was used to discover featured nodes and identify the hub genes from all DEGs. The biological processes of the hub genes were visualized using the Biological Networks Gene Oncology tool (BiNGO) (version 3.0.3) plugin of Cytoscape [15], with significance threshold 0.01 and Homo sapiens as the selected organism.

Results

1. Identification of DEGs

We identified a total of 96 samples, comprising 47 IgAN samples and 49 normal samples, in the two datasets. Based on the cutoff criteria of |log2 FC| ≥ 1.0 and adj Pvalue ≤ 0.05, 148 overlapping DEGs were showed by a Venn diagram obtained from IgAN group vs. control group, consisted of 53 upregulated and 95 downregulated. The results on the expression level analysis are presented in a volcano plot in Fig. 1A.
As indicated in the clustering heat map (Fig. 1B), these DEGs could well distinguish the IgAN and control group completely.

2. Gene Ontology and KEGG Analysis of DEGs

To investigate the biological classification of DEGs, the overall genes in three ontologies were identified using DAVID. The cut-off criterion was set as P < 0.05. The GO function annotation is divided into three functional groups, cell component (CC), molecular function (MF), and biological process (BP). The CC terms of the DEGs were significantly enriched in extracellular exosome, region and space (Fig.2A,3A). The MF terms were mainly enriched in transcriptional activator activity and RNA polymerase II core promoter proximal region sequence-specific binding; heme binding identical protein binding (Fig.2B,3B). The changes in BP were significantly enriched in response to cAMP, cellular response to fibroblast growth factor stimulus and inflammatory response (Fig.2C,3C).

The KEGG pathway enrichment analysis revealed that the DEGs were mainly enriched in Protein digestion and absorption, Pertussis and Staphylococcus aureus infection (Fig.2D,3D).

3. Construction of the PPI network and module analysis

To further investigate the interaction among the 148 DEGs, a PPI network was constructed from STRING(Fig.4A) The most significant module was obtained using Cytoscape MCODE plug-in. The module consisted of 15 nodes and 89 edges, include Fc fragment of IgE receptor Ig (FCER1G), HCK proto-oncogene, Src family tyrosine kinase(HCK), TYRO protein tyrosine kinase binding protein(TYROBP),V-set and immunoglobulin domain containing 4(VSIG4), colony stimulating factor 1 receptor(CSF1R),complement C1q A chain(C1QA), complement C3a receptor 1(C3AR1),cytochrome b–245 beta chain(CYBB), hematopoietic cell-specific Lyn substrate 1(HCLS1), integrin subunit beta 2(ITGB2), interleukin 10 receptor subunit alpha (IL10RA), lysosomal protein transmembrane 5(LAPTM5), neutrophil cytosolic factor 2(NCF2),CD48 and CD53 which exhibited the highest score,12.714 (Fig. 4B). These findings indicated that these genes exhibited higher hub degrees and could play critical roles in the development of IgAN.

GO analysis of the module showed significant enrichment in the CC terms of the DEGs were mostly enriched in integral component of plasma membrane, cell surface and plasma membrane.(Fig.5A,6A) MF terms were mainly enriched in superoxide-generating NADPH oxidase, receptor activity and protein binding (Fig.5B,6B). BP terms: innate immune response, integrin-mediated signaling pathway and inflammatory response. (Fig.5C,6C),

The KEGG pathway enrichment analysis revealed that the DEGs were mainly enriched in Tuberculosis, natural killer cell mediated cytotoxicity, Osteoclast differentiation and Staphylococcus aureus infection (Fig.5D,6D).

4. Identification and analysis of hub genes
We exported the STRING data to Cytoscape to construct and visualize the PPI network by implementing the cytoHubba. Thereafter, we implemented the MCC method to evaluate the significance of the genes in the network. The top ten genes, including IL10RA, ITGB2, HCK, C3AR1, CYBB, LAPTМ5, FCER1G, CD53, C1QA and TYROBP. Hierarchical clustering of the hub genes was performed as indicated in the clustering heat map (Fig. 7A), these Hub gene could well distinguish the IgAN and control group completely. The biological process analysis of hub genes was performed using BiNGO plug-in showed in Fig.7B.

For further analysis, The microarray data GSE58539 was downloaded from GEO, This dataset contained 17 monocytes samples, including 15 monocytes samples isolated from IgAN patients and 2 monocytes samples isolated from health control group. We used these selected hub genes for analysis, The scatter plot showed that each hub gene was significantly different between IgAN and control group (Fig. 8A). Hierarchical clustering of the hub genes was performed. As indicated in the clustering heat map (Fig. 8B), these Hub gene also could well distinguish the IgAN and control group in monocytes sample.

Discussion

Bioinformatics analysis play an important role in disease study, and it facilitates the understanding of pathogenesis by integrating data at the genome level with systematic bioinformatic methods. In the present study, 148 DEGs were identified from microarray data by re-analyzing the dataset which could distinguished separate the IgAN and healthy control. Previous studies have shown a significant association of IgAN development and prognosis with the inflammatory reaction, activation of TGFβ signaling is closely related to fibrosis in IgAN [16–17]. In our study the enrichment analysis revealed that the DEGs were significantly enriched in response to cAMP, cellular response to fibroblast growth factor stimulus and inflammatory response which were consistent with the previous study.

We all know that infection play an important role in the onset of IgAN, nearly 30 % patients have a clear history of disease exacerbation after upper respiratory or gastrointestinal infections. Novak et al once reported that viruses (e.g. Epstein–Barr virus) or bacteria (e.g. Streptococcus) that express GalNAc-containing moieties may induce the development of IgG antiglycan autoantibodies, which might subsequently cross-react with glycans on IgA1 resulting in the formation of IgA1-IgG complexes [18]. This ‘molecular mimicry’ could also explain the association of macroscopic hematuria with upper respiratory tract infections. Yamamoto Y reported that antigens of Haemophilus parainuenzae have been detected in renal tissue of patients with IgAN [19]. In our study KEGG analysis also discovered that the DEGs were mainly enriched in pertussis and Staphylococcus aureus infection which coincide with above researches.

Using STRING and MCODE, we selected the most important module consisted of 15 nodes and 89 edges, include CSF1R, IL10RA, ITGB2, HCK, NCF2, C3AR1, CYBB, HCLS1, CD48, C1QA, VSIG4, LAPTМ5, FCER1G, CD53 and TYROBP, further GO analysis revealed that the BP mainly enrichment in the innate immune response, integrin-mediated signaling pathway and inflammatory response. Previous studies have shown a significant association of IgAN development and prognosis with the inflammatory reaction and innate immune response. Toll-like receptors (TLRs) are the key components of the mammalian innate immune
system and mediate immune and inflammatory responses through binding PAMPs and/or DAMPs [20].
Lots of experiments have confirmed the elevation expression of TLR4 mRNA in IgAN rat. In an in vitro coculture system of IgA and mesangial cells, TLR4 mediates MAPK activation and MCP–1 secretion, indicating that TLR4 is engaged in glomerular mesangium damage by inducing inflammatory cytokines in IgAN [21]. TLR4 also involved in the activation of NF-κB, then triggers the transcription of mRNA encoding many inflammatory mediators, such as cytokines, chemokines, fibrinogen, etc., which contribute significantly to the effects of the innate and adaptive immune responses [22].

We further implemented the MCC method and selected 10 hub genes, all of these genes were overlap with the important module selected by MCODE. Hierarchical clustering of the hub genes showed that these Hub gene could well distinguish the IgAN and control group completely. We further introduced these genes into the blood samples for testing, and the results showed that these genes also play a good role in differentiating disease from control in the blood tissue.

Hck is a member of the highly-conserved Src family of cytoplasmic protein tyrosine kinases that transduce a variety of extracellular signals, Hck has been reported significantly upregulated in diabetic nephropathy, IgA nephropathy, and lupus nephritis, Hck is a key mediator of renal fibrosis via its effects on inflammation, fibroblast cell proliferation, and regulation of TGF-β signaling.[23]

LAPTM5, which is preferentially expressed in hematopoietic cells and localized to the lysosome, was initially isolated by a subtractive hybridization strategy between hematopoietic and non-hematopoietic cells. A recent study shows that LAPTM5 is a positive regulator of proinflammatory signaling pathways via facilitating NF-κB and MAPK signaling, and proinflammatory cytokine production in macrophages. [24–25] CYBB is also responsive to a number of inflammatory cytokines such as IFN-γ, LPS, and TNF-α. [26] CD53 codes for cluster of differentiation 53, a leukocyte surface antigen. Many researches indicate that CD53 has a substantial role in cellular stability and in the inflammatory response to adverse conditions [27]. We all know that inflammatory response plays an important role in IgAN. these above hub genes were all identified have the related function to the pathogenesis of IgAN.

FCER1G is a Protein Coding gene. It has been reported that FCER1G interacts with other factors and participates in various nuclear pathways [28]. Specifically, FCER1G is a constitutive component of the high-affinity immunoglobulin E receptor and interleukin–3 receptor complex. It is mainly involved in mediating the allergic inflammatory signaling of mast cells, selectively mediating the production of interleukin 4 by basophils, and initiating the transfer from T-cells to the effector T-helper 2 subset [29]. It is associated with the progression of clear cell renal cell carcinoma and may improve prognosis by affecting the immune-related pathways. Moreover, FCER1G is a critical molecule in signaling pathways that are widely involved in a variety of immune responses and cell types [30]. By now there is no study reported FCER1G was related to IgAN. In our research, FCER1G was the first-ranked hub gene, Bingo analysis results confirmed that FCER1G was directly involved in innate immune response, so the novel hub gene FCER1G was indicated to have an important role in IgAN.
ITGB2 is a Protein Coding gene, encodes an integrin beta chain, which combines with multiple different alpha chains to form different integrin heterodimers. Integrins are integral cell-surface proteins that participate in cell adhesion as well as cell-surface mediated signaling. The encoded protein plays an important role in immune response and defects in this gene cause leukocyte adhesion deficiency. ITGB2 once reported that involved in cellular adhesion and ECM remodeling in patients with renal cancer [31]. Furthermore, ITGB2 was identified to be closely associated with apoptosis in patients with Alzheimer's disease [32]. A bioinformatics analysis in CKD patients showed that ITGB2, CTSS and CCL5 connected negatively to the eGFR of CKD patients [33]. In the present study, ITGB2 was the second-ranked hub gene, right now, limited research has been reported the association between ITGB2 and IgAN, so we think, this gene deserve further research.

Conclusions

To summarize, at present study through bioinformatics analysis we identified some hub genes involved in the pathological changes of IgAN, these gene not only can be used in tissue sample, but also play important roles in blood sample too. In conclusion, the present study was the first to apply an integrated bioinformatics analysis to investigate novel candidate genes and mechanisms involved in the pathogenesis of IgAN. Among them, ITGB2 and FCER1G gene may have important roles in the development of IgAN and act as potential candidate molecular targets deserve our further research.

List Of Abbreviations

List of abbreviation
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgAN</td>
<td>Immunoglobulin A nephropathy</td>
</tr>
<tr>
<td>DEGs</td>
<td>differentially expressed genes</td>
</tr>
<tr>
<td>PPI</td>
<td>Protein-protein interaction</td>
</tr>
<tr>
<td>BiNGO</td>
<td>Biological Networks Gene Oncology</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Oncology</td>
</tr>
<tr>
<td>Gd-IgA1</td>
<td>galactose-deficient IgA1</td>
</tr>
<tr>
<td>GEO</td>
<td>Gene Expression Omnibus</td>
</tr>
<tr>
<td>DAVID</td>
<td>Database for Annotation, Visualization and Integration Discovery</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>MCODE</td>
<td>Molecular Complex Detection</td>
</tr>
<tr>
<td>CC</td>
<td>cell component</td>
</tr>
<tr>
<td>MF</td>
<td>molecular function</td>
</tr>
<tr>
<td>BP</td>
<td>biological process</td>
</tr>
<tr>
<td>TLRs</td>
<td>Toll-like receptors</td>
</tr>
<tr>
<td>LAPTM5</td>
<td>Lysosomal-associated multispansing membrane protein 5</td>
</tr>
<tr>
<td>FCER1G</td>
<td>Fc fragment of IgE receptor Ig</td>
</tr>
<tr>
<td>ITGB2</td>
<td>Integrin Subunit Beta 2</td>
</tr>
<tr>
<td>TYROBP</td>
<td>TYRO protein tyrosine kinase binding protein</td>
</tr>
<tr>
<td>C1QA</td>
<td>complement C1q A chain</td>
</tr>
<tr>
<td>C3AR1</td>
<td>complement C3a receptor 1</td>
</tr>
<tr>
<td>CYBB</td>
<td>cytochrome b-245 beta chain</td>
</tr>
<tr>
<td>HCK</td>
<td>HCK proto-oncogene, Src family tyrosine kinase</td>
</tr>
<tr>
<td>IL10RA</td>
<td>interleukin 10 receptor subunit alpha</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>VSIG4</td>
<td>V-set and immunoglobulin domain containing 4</td>
</tr>
<tr>
<td>CSF1R</td>
<td>colony stimulating factor 1 receptor</td>
</tr>
<tr>
<td>HCLS1</td>
<td>hematopoietic cell-specific Lyn substrate 1</td>
</tr>
<tr>
<td>NCF2</td>
<td>neutrophil cytosolic factor 2</td>
</tr>
</tbody>
</table>

### Declarations

#### Ethics approval and consent to participate

Ethics approval not applicable. The data does not compromise anonymity or confidentiality or breach local data protection law.

#### Consent for publication

Not applicable.

#### Availability of data and materials

All data of the databases are available.

#### Acknowledgements

No applicable.

#### Funding:

This work was supported by the National Science Foundation of Zhejiang province (Grant No. LQ20H29003), Zhejiang Traditional Chinese Medicine Administration (Grant No. 2020ZB154)

#### Contributions

Chen Hongyu designed the study, Du Yuanyuan and Xu Zhijie analyzed the data; Jiang Xue wrote the article. All authors read and approved the final version of the article.

#### Competing interests:

The authors declare to have no competing interests.

### References


Figures
Figure 1

Identification and hierarchical clustering of DEGs. A. Volcano plot of DEGs in IgAN. The cut-off criteria were $|\log_{2}Fc|>1$ and $P$ value $<0.05$. The orange dots represented the up-regulated genes, and the blue dots denoted the down-regulated genes. The grey dots indicated the genes with $|\log_{2}Fc|<1$ and/or $P$ value $>0.05$. B. Heat map of the DEGs. Horizontal band with the cluster tree at the top: blue, normal samples; orange, IgAN. Each row represented a single gene. blue, downregulated DEGs; orange, upregulated DEGs. The depth of the color denoted the change degree.
Figure 2

GO and KEGG pathway enrichment analysis of DEGs. The color depth of nodes refers to the P-value. The size of nodes refers to the numbers of genes. A: GO CC terms. B: GO MF terms. C: GO BP terms. D: KEGG Pathway of DEGs.
**Figure 3**

Figure 4

PPI network, and the most significant module of DEGs. A: The PPI network of DEGs was constructed using Cytoscape. green, downregulated DEGs; red, upregulated DEGs. The size of nodes refers to the gene degree. B: The most significant module was obtained from PPI network. The depth of the color denoted the change degree.
Figure 5

GO pathway enrichment analysis of DEGs in the most significant module. GO pathway enrichment analysis of DEGs in the most significant module which was obtained from PPI network. The color depth of nodes refers to the P-value. The size of nodes refers to the numbers of genes. A: GO CC terms. B: GO MF terms. C: GO BP terms. D: KEGG Pathway of DEGs.
Figure 6

Figure 7

Interaction network and biological process analysis of the hub genes. A. Heat map of the hub genes. Horizontal band with the cluster tree at the top: blue, normal samples; orange, IgAN. Each row represented a single gene. blue, downregulated DEGs; orange, upregulated DEGs. The depth of the color denoted the change degree. B: GO enrichment of hub genes was analyzed using BiNGO.
Figure 8

Hub genes analysis in blood monocytes sample. A. scatter plot of hub genes. B. Heat map of the hub genes in blood monocytes sample. Horizontal band with the cluster tree at the top: blue, normal samples; orange, IgAN. Each row represented a single gene. blue, downregulated DEGs; orange, upregulated DEGs. The depth of the color denoted the change degree.