Identifying DUSP-1 and FOSB as hub genes in immunoglobulin A nephropathy by WGCNA and DEG screening and validation

Jianping Wu[1#],Xiaona Wei[1#],Jiajia Li[1],Rui Zhang[1], Qianqian Han[1], Qiongqiong Yang[1*]

1.Department of Nephrology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

*Corresponding Author:
Qiongqiong Yang[1]
Yan Jiang Xi Lu 107, Yue Xiu District, Guangzhou, Guangdong, 510120, China
Email address: yangqq@mail.sysu.edu.cn
[#]Jianping Wu and Xiaona Wei contributed equally to this work.

Abstract

Background: The mechanism of immunoglobulin A nephropathy (IgAN) is still unknown. A bioinformatics analysis is a powerful method to identify the biomarkers and possible therapeutic targets of a certain disease from related datasets.

Methods: The GSE93973 dataset, obtained from the Gene Expression Omnibus (GEO) database, was used to construct a weighted gene co-expression network (WGCNA) and filter differentially expressed genes (DEGs). The biological process (BP) enrichment among all the genes in the key modules was analyzed through a Gene Ontology (GO) enrichment analysis. We selected the overlap of hub genes in the WGCNA and PPI network as the final hub genes in IgAN. We verified the final hub genes in other two datasets and in clinical kidney tissue specimens. A receiver operating characteristic (ROC) curve was used to evaluate the diagnostic efficacy of hub genes for IgAN.

Results: The turquoise module, which contained 1,806 genes, was the module with the highest correlation coefficient with IgAN in the GSE93973 dataset. The GO enrichment analysis showed that these 1,806 genes were mainly enriched in inflammation and immune responses. There were 5 hub genes identified by WGCNA and 34 hub genes identified in a DEG analysis in the GSE93973 dataset. DUSP1 and FOSB were identified as the final hub genes in IgAN. The

34 validation results of the final hub genes in two other databases and clinical kidney tissue specimens

35 validated the result that, compared to the control group, FOSB and DUSP1 were expressed at lower

36 levels in the glomerulus of IgAN patients. The ROC curve indicated that DUSP1 and FOSB were

37 good diagnostic indicators for IgAN.

38 **Conclusions:** Our analysis identified two hub genes that might be potential targets for the

39 intervention and treatment of IgAN.

40 **Key words**

41 immunoglobulin A nephropathy, weighted gene co-expression network, hub genes, differentially

42 expressed genes,

Introduction

44 Immunoglobulin A nephropathy (IgAN), the most common primary glomerulonephritis, is

45 characterized by the presence of IgA-containing immune complexes in the mesangial area of the

46 glomerulus. Most IgAN patients will progress to end-stage kidney disease (1, 2). The pathogenesis

47 of IgAN is not very clear and there are not currently any effective pharmacological treatments.

48 There is therefore an urgent need to identify key molecular targets involved in the intervention and

49 treatment of IgAN.

50 A weighted gene co-expression network analysis (WGCNA) is a useful bioinformatics

51 method of identifying and clustering genes related to diseases or clinical manifestations and

52 forming different co-expression networks or co-expression modules (3). WGCNA is also a

53 powerful data analysis tool that can identify potential biomarkers for the intervention and treatment

54 of different diseases, such as advanced kidney disease (4), Sjögren's syndrome (5), acute

55 myocardial infarction (6), and pancreatic cancer (7).

56 In this study, we identified two IgAN hub genes in the GSE93798 dataset using WGCNA and

57 DEG screening. The flowchart of the study was shown in Figure 1. The final hub genes were

58 validated in the GSE26316 and GSE104948 datasets and in kidney tissue from IgAN patients and

59 kidney cancer adjacent sample tissues. This study identified two hub genes, Dual Specificity

60 Phosphatase 1 (DUSP1) and FosB Proto-Oncogene (FOSB), that might be promising targets for

61 the intervention and treatment of IgAN.

Materials & Methods

63

Data Extraction

65   We performed our analysis on August 1, 2021. The GSE93798 dataset

66   (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93798), which includes the gene

67   matrix of the glomerular compartment of renal biopsies from 20 healthy people and 22 IgAN

68   patients, was acquired from the Gene Expression Omnibus (GEO) database. The two groups of

69   people in dataset GSE93798 were with gender and age matched. IgAN (age 43 ± 15 years old,

70   women-to-men ratio of 1:2) and control samples (age 48 ± 13 years old, women-to-men ratio of

71   1:1.27). This dataset was then analyzed using WGCNA and DEG screening to identify key

72   modules and hub genes related to IgAN. The datasets were based on the platform of GPL22945

73   (Affymetrix Human Genome U133 Plus 2.0).

74   **Construction of Co-Expression Network**

75   To select the top 5,000 median absolute deviation (MAD) genes for WGCNA, we sorted all

76   gene expression matrices of the GSE93798 dataset according to median deviation size using the R

77   package "WGCNA" (https://www.rdocumentation.org/packages/WGCNA/versions/1.71) (3). A

78   cluster analysis was performed on the samples before constructing the WGCNA network to test

79   the heterogeneity between the samples included in the GSE93798 dataset. The

80   "PickSoftThreshold" package

81   (https://www.rdocumentation.org/packages/WGCNA/versions/1.71/topics/pickSoftThreshold)

82   was used to determine the best soft threshold and to identify the co-expression module based on

83   the dynamic cut method. The module and sample feature correlation heat map and module and

84   sample feature hierarchical clustering map were constructed to determine which key module had

85   the highest correlation coefficient with IgAN. Genes with a gene significance (GS) value and

86   module membership (MM) that were both greater than 0.9 were determined to be hub genes.

87   **Functional Enrichment Analysis**

88   Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/home.jsp/) was

89   used to perform an online Gene Ontology (GO) analysis within the genes in the key module. The

90   "GOplot" R package(https://www.rdocumentation.org/packages/GOplot/versions/1.0.2) was then

91   used to plot and display the results.

92   **DEG screening**

93   Genes with an adjusted P-value <0.05 and | log2(fold change) |> 1 were defined as DEGs

94   between the control group and IgAN patients. The GSE93798 dataset was screened using the

95   "limma" package (https://www.rdocumentation.org/packages/limma/versions/3.28.14) from R

96   (8). All DEGs were drawn as heat maps with R package "pheatmap."
97   (https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12) The Cytoscape software
98   was used to construct a Protein-Protein Interaction (PPI) network based on the protein interaction
99   information obtained from the STRING database (https://string-db.org/cgi/) after screening the
100 DEGs. The Molecular Complex Detection (MOCDE) plugin was then applied to find hub genes
101 in the PPI network.

Validation of final hub genes in other datasets and clinical specimens

103   The hub genes that existed in both the PPI network and WGCNA were determined to be the
104 final hub genes. In order to find appropriate datasets, based on the GEO database, we only included
105 datasets that included: 1) Sequencing data from glomerus of IgAN patients; 2) gene matrix
106 information; 3) and 3 or more samples. Two datasets met the criteria and were included, namely
107 GSE37460 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37460), which contains
108 the renal biopsies of 27 healthy people and 27 IgAN patients, and GSE104948
109 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104948), which contains the renal
110 biopsies of 3 healthy people and 27 IgAN patients. To validate these hub genes, we verified their
111 expressions in 3 cases of adjacent renal cell carcinoma tissues and 3 cases of IgAN renal biopsy
112 tissues with immunohistochemistry. The verified specimens came from Sun Yat-sen Memorial
113 Hospital of Sun Yat-sen University. This study was approved by the Medical Ethics Committee of
114 Sun Yat-sen Memorial Hospital (SYSEC-KY-KS-2018-080). All patients included in the study
115 signed an informed consent form.

Receiver operating characteristic (ROC) curve analysis

117   The ROC curve of the final hub genes was drawn using the sample information in the
118 GSE93798, GSE37460 and GSE104948 datasets with the pROC package
119 (https://www.rdocumentation.org/packages/pROC/versions/1.18.0/topics/pROC-package) of R
120 (3). We evaluated the diagnostic efficacy of the final hub genes with IgAN by calculating the area
121 under the curve (AUC).

Immunohistochemical staining

123   After dewaxing, the paraffin sections with a thickness of about 4 μm were placed in a
124 sodium citrate solution with pH=7.0 heated in a microwaves on high heat for antigen retrieval. A
125 3% hydrogen peroxide solution was then dropped on the sliced tissue to block endogenous
126 peroxidase. We then performed primary antibody incubation (anti-DUSP1, Abcam, ab61201,

127   1:100 dilution; anti-FOSB, Abcam, ab11959, 1:100 dilution), HRP goat anti-rabbit secondary

128   antibody (1:200 dilution) incubation, and nucleus staining incubation. These were followed by

129   DAB coloration, hematoxylin staining, and hydration, at which point we covered the slip with

130   neutral resin and observed the final product under a microscope. We then used the Image J

131   software to perform a semi-quantitative analysis.

132   **Statistics**

133   A statistical analysis was performed using R 4.0.3, SPSS 20.0 and GraphPad Prism 6.0. A

134   Student's t-test was then performed to examine the statistical difference between the group

135   means. All data were presented as Mean ± SEM. $P < 0.05$ was defined as statistically significant.

136   **Results**

137   **Construction of the WGCNA network**

138   All gene expression matrices of the GSE93798 dataset were sorted according to median

139   deviation size and the top 5,000 genes were used for subsequent analyses. A cluster analysis was

140   performed on the samples before we started constructing the WGCNA network and the results

141   showed that there was no obvious heterogeneity between these samples (Figure 2A). The power

142   value (β) was set to 8 and the scale independence value was set to 0.85 (Figure.2B, C). There was

143   a negative correlation between the logarithm log(k) and the logarithm log(P(k)), and the correlation

144   coefficient was greater than 0.8. These results proved that the constructed co-expression network

145   met the requirements of scale-free topology (Figure 2D, E). Therefore, β = 8 was chosen to build

146   a hierarchical clustering tree.

147   **Identifying the key module and hub genes**

148   There were six gene co-expression modules in the WGCNA (Figure 3A). The turquoise

149   module, which contained 1,806 genes, was the one with the highest correlation coefficient (CC)

150   with IgAN (Figure 3B, C). A total of 400 genes were randomly selected for heat map display

151   (Figure 3D). The module and disease characteristic heat map showed that the turquoise module

152   had the most significant correlation with IgAN (Figure 3E). Genes that met the criteria that the

153   value of GS and MM were both greater than 0.9 were determined to be hub genes (Figure 3F).

154   **Functional enrichment analysis of turquoise module genes**

155   A GO enrichment analysis showed that these 1,806 genes in the turquoise module were mainly

156   enriched in the BP of cell adhesion, leukocyte migration, inflammatory response, extracellular

157   matrix organization, immune response, and adaptive immune response (Figure 4A). A cross-check

158   GO analysis showed that a large number of genes related to white blood cell migration,

159   inflammation and inflammatory response were also enriched in extracellular matrix tissue,

160   immune response, adaptive immune response and other BPs. This indicated that these cross-

161   mapped genes were involved in the pathogenesis of IgAN in multiple BPs (Figure 4B).

162   **DEGs screening**

163   There were 248 DEGs identified, including 107 up-regulated genes and 141 down-regulated

164   genes, after screening the DEGs based on the GSE93798 dataset (Figure 5A, B). A total of 34 hub

165   genes in three clusters were screened using the MCODE method when the threshold of k- score>

166   2. The three clusters included 15 nodes (Figure 5C), 11 nodes (Figure 5D) and 18 nodes (Figure

167   5E), respectively.

168   **Identification and verification of final hub genes**

169   Hub genes found in both the WGCNA and PPI network, namely FOSB and DUSP1, were

170   identified as the final hub genes (Figure 5F). Compared with the control group, FOSB and DUSP1

171   had low expression levels in the glomerulus of IgAN patients in the GSE93798 dataset and the two

172   validation datasets, GSE104948 (Figure 6A) and GSE37460 (Figure 6B). To validate our results

173   in clinical specimens, we used immunohistochemistry and semi-quantitative analyses, which

174   showed that compared with adjacent tissues of renal cancer, FOSB and DUSP1 were down-

175   regulated in the glomerulus of IgAN patients. This result was consistent with the expressions of

176   these two genes seen in the transcriptome matrix data (Fig. 6C, D, E).

177   **ROC curve analysis of final hub genes**

178   The ROC curve was drawn to evaluate the diagnostic value of FOSB and DUSP1 using the

179   sample information in the GSE93798, GSE37460 and GSE104948 datasets (Figure 7). All the

180   AUC values were greater than 0.75, indicating their potential value as diagnostic markers for

181   IgAN.

182   **Discussion**

183   IgAN is one of the main causes of end-stage renal disease worldwide. The pathogenesis of

184   IgAN is unclear, but the current hypothesis is the multi-strike theory that various pathogenic

185   factors lead to the increase of galactose-deficient O-glycan (Gd-IgA1) in the glomerulus,

186   activating the mesangial cells, and ultimately promoting the occurrence and development of

187   IgAN (1). According to this hypothesis, specific glomerular biological information could better

188   identify the hub genes of IgAN, which supported our decision to use datasets that included

189   glomerulus samples from IgAN patients in this study rather than those that only included

190   peripheral blood samples (10, 11).

191   The co-expression network of diseases and modules constructed by WGCNA can identify

192   biomarkers and possible therapeutic targets for a certain disease. In this study, we used WGCNA

193   to identify the key module with the highest correlation coefficient with IgAN in the GSE93798

194   dataset. A GO enrichment analysis showed that the genes in the key module were mainly

195   enriched in the BP of cell adhesion, leukocyte migration, inflammatory response, extracellular

196   matrix organization, immune response, and adaptive immune response. This result indicated that

197   the pathogenesis of IgAN is related to the activation of immune and inflammation-related

198   pathways, but the inflammatory response plays a more important role in the development of

199   IgAN than the immune response. IgAN is currently considered an autoimmune disease (12), but

200   Sheng et al found that Notch Receptor 1 (NOTCH 1) and NF-κB and other inflammatory

201   pathway-related molecules were increased in the renal tissue of IgAN patients (13). Some animal

202   experiments have also shown that certain drugs can affect the progression of IgAN by inhibiting

203   NF-κB signaling and NLRP3 inflammasome activation (14). In clinical studies, some scholars

204   have found that the neutrophil to lymphocyte ratio is an independent risk factor for poor

205   prognosis in IgAN patients (15). Inflammation thus plays an important role in the occurrence and

206   development of IgAN. This may partly explain why immunosuppressive agents are ineffective in

207   treating IgAN (16).

208   To find the most significant hub gene, we defined GS> 0.9 and ME> 0.9 as target values in

209   the key module of WGCNA and used MCODE to identify the most important gene clusters in

210   the PPI network. There were 5 hub genes in the WGCNA analysis and 34 hub genes in the

211   MCODE analysis. The final hub genes found in both the WGCNA and MCODE analysis were

212   defined as the key hub genes; FOSB and DUSP-1 were identified as the final hub genes of IgAN.

213   An ROC analysis showed that FOSB and DUSP-1 had good diagnostic performance. The

214   validation results of these key genes in two other datasets, as well as in clinical specimens, were

215   consistent with the results that the expression of FOSB and DUSP1 were down-regulated in the

216   renal glomerulus of IgAN patients compared with people that did not have IgAN. Previous

217   studies have shown that FOSB is related to human stress and is a potential therapeutic target for

218   depression. FOS family gene polymorphisms, including FOSB, are related to the clinical

219   phenotypes of IgAN patients (17). In an animal model of renal ischemia/reperfusion, the

220 expression of FOSB was up-regulated in interthalamic core cells (18). Kai et al found that

221 MicroRNA-27A-3P targeted FOSB to regulate cell function involved in the occurrence and

222 development of IgAN (19). These studies show that FOSB is involved in the progression of

223 IgAN, but the specific mechanism of FOSB in the pathogenesis of IgAN unclear. DUSP1 can

224 regulate cell death, tumor growth and plays a role in both inflammation and chronic metabolic

225 diseases. DUSP1 is highly expressed in tissues such as the heart and liver that could activate

226 mitogen-activated protein kinase (MAPKs), extracellular signal-regulated kinase (ERK), c-Jun

227 N-terminal kinase (JNK), and p38 to promote cell apoptosis (20). Jin et al showed that DUSP1

228 inhibited cardiac ischemia-reperfusion (IR) injury by regulating downstream JNK to inhibit

229 mitochondrial fission (21). Low DUSP1 expression is related to glucose metabolism disorder,

230 renal dysfunction and glomerular apoptosis; by affecting mitochondrial homeostasis, DUSP1 is

231 able to protect against diabetic nephropathy progression (22, 23). There is currently no research

232 related to DUSP1 in IgAN and its potential role in the pathogenesis of IgAN needs further study.

233 In this study, we found two genes closely related to IgAN using bioinformatics and in vitro

234 verification. These genes can be used as potential markers for the diagnosis of IgAN, and

235 perhaps also as potential therapeutic targets.

236   This research has several highlights. First, we used datasets that include glomeruli from the

237 biopsies of IgAN patients for our analysis in order to better identify candidate biomarkers and

238 therapeutic targets, contributing to a more comprehensive understanding of IgAN. Second,

239 WGCNA is a powerful tool for bioinformatics analysis. It can associate gene modules with

240 disease traits. WGCNA plus a DEG analysis can find the most relevant value target genes.

241 However, due to limited sample availability we were unable to verify the mRNA expression

242 levels of FOSB and DUSP1. Instead, we analyzed kidney tissue slices using

243 immunohistochemistry and achieved consistent results with the bioinformatics analysis.

244 **Conclusions**

245   In summary, our study identified FOSB and DUSP1 as final hub genes in IgAN through

246 WGCNA and DEG analysis. These genes were verified to be significantly down-regulated in the

247 renal glomerulus of IgAN patient kidney tissues and other IgAN datasets, compared with non-

248 IgAN subjects. The specific mechanisms of these hub genes involved in IgAN pathogenesis also

249 need to be further studied. Our results provided novel molecular targets as biomarkers and

250 therapeutic targets for future research.

251

Acknowledgements

Thanks to Dr. Anyi Wang for polishing the manuscript.

References

1. Lai KN, Tang SC, Schena FP, Novak J, Tomino Y, Fogo AB, Glassock RJ. IgAN. Nat Rev Dis Primers 2016;2:16001.

2. Woo KT, Chan CM, Mooi CY, H. LC, Tan HK, Foo M, Lee GS, et al. The changing pattern of primary glomerulonephritis in Singapore and other countries over the past 3 decades. Clin Nephrol 2010;74:372-383.

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

4. Xin G, Zhou G, Zhang W, Zhang X. Construction and Validation of Predictive Model to Identify Critical Genes Associated with Advanced Kidney Disease. Int J Genomics 2020;2020:7524057.

5. Yao Q, Song Z, Wang B, Qin Q, Zhang JA. Identifying Key Genes and Functionally Enriched Pathways in Sjogren's Syndrome by Weighted Gene Co-Expression Network Analysis. Front Genet 2019;10:1142.

6. Wu Z, Wen Y, Fan G, He H, Zhou S, Chen L. HEMGN and SLC2A1 might be potential diagnostic biomarkers of steroid-induced osteonecrosis of femoral head: study based on WGCNA and DEGs screening. BMC Musculoskelet Disord 2021;22:85.

7. Zhou YY, Chen LP, Zhang Y, Hu SK, Dong ZJ, Wu M, Chen QX, et al. Integrated transcriptomic analysis reveals hub genes involved in diagnosis and prognosis of pancreatic cancer. Mol Med 2019;25:47.

8. 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 Res 2015;43:e47.

9. Li J, Wang L, Guo M, Zhang R, Dai Q, Liu X, Wang C, et al. Mining disease genes using integrated protein-protein interaction and gene-gene co-regulation information. FEBS Open Bio 2015;5:251-256.

10. Liu H, Liu D, Liu Y, Xia M, Li Y, Li M, Liu H. Comprehensive analysis of circRNA expression profiles and circRNA-associated competing endogenous RNA networks in IgAN. PeerJ

282 2020;8:e10395.

283 11.Jiang X, Xu Z, Du Y, Chen H. Bioinformatics analysis reveals novel hub gene pathways
284 associated with IgAN. Eur J Med Res 2020;25:40.

285 12.Suzuki H, Kiryluk K, Novak J, Moldoveanu Z, Herr AB, Renfrow MB, Wyatt RJ, et al. The
286 pathophysiology of IgAN. J Am Soc Nephrol 2011;22:1795-1803.

287 13.Sheng X, Zuo X, Liu X, Zhou Y, Sun X. Crosstalk between TLR4 and Notch1 signaling in the
288 IgAN during inflammatory response. Int Urol Nephrol 2018;50:779-785.

289 14.Bai L, Li J, Li H, Song J, Zhou Y, Lu R, Liu B, et al. Renoprotective effects of artemisinin and
290 hydroxychloroquine combination therapy on IgAN via suppressing NF-kappaB signaling and
291 NLRP3 inflammasome activation by exosomes in rats. Biochem Pharmacol 2019;169:113619.

292 15.Li Q, Chen P, Shi S, Liu L, Lv J, Zhu L, Zhang H. Neutrophil-to-lymphocyte ratio as an
293 independent inflammatory indicator of poor prognosis in IgAN. Int Immunopharmacol
294 2020;87:106811.

295 16.Rauen T, Wied S, Fitzner C, Eitner F, Sommerer C, Zeier M, Otte B, et al. After ten years of
296 follow-up, no difference between supportive care plus immunosuppression and supportive care
297 alone in IgAN. Kidney Int 2020;98:1044-1052.

298 17.Park HJ, Kim JW, Cho BS, Chung JH. Association of FOS-like antigen 1 promoter
299 polymorphism with podocyte foot process effacement in immunoglobulin A nephropathy patients.
300 J Clin Lab Anal 2014;28:391-397.

301 18.Mutoh J, Ohsawa M, Hisa H. Effect of naloxone on ischemic acute kidney injury in the mouse.
302 Neuropharmacology 2013;71:10-18.

303 19.Liao Y, Wang Z, Wang L, Lin Y, Ye Z, Zeng X, Wei F. MicroRNA-27a-3p directly targets
304 FosB to regulate cell proliferation, apoptosis, and inflammation responses in immunoglobulin a
305 nephropathy. Biochem Biophys Res Commun 2020;529:1124-1130.

306 20.Johnson GL, Lapadat R. Mitogen-activated protein kinase pathways mediated by ERK, JNK,
307 and p38 protein kinases. Science 2002;298:1911-1912.

308 21.Jin Q, Li R, Hu N, Xin T, Zhu P, Hu S, Ma S, et al. DUSP1 alleviates cardiac
309 ischemia/reperfusion injury by suppressing the Mff-required mitochondrial fission and Bnip3-
310 related mitophagy via the JNK pathways. Redox Biol 2018;14:576-587.

311 22.Sheng J, Li H, Dai Q, Lu C, Xu M, Zhang J, Feng J. DUSP1 recuses diabetic nephropathy via
312 repressing JNK-Mff-mitochondrial fission pathways. J Cell Physiol 2019;234:3043-3057.

313   23.Ge Y, Wang J, Wu D, Zhou Y, Qiu S, Chen J, Zhu X, et al. lncRNA NR_038323 Suppresses

314   Renal Fibrosis in Diabetic Nephropathy by Targeting the miR-324-3p/DUSP1 Axis. Mol Ther

315   Nucleic Acids 2019;17:741-753.

316
317
318   **Figure lends**

319

320   Fig. 1 Flowchart of this research.

321

322   Fig. 2 Data verification and filter the best soft threshold. A Sample clustering to view

323   heterogeneity. B Soft threshold non-scale fitting index analysis. C Soft threshold average

324   connectivity analysis. D Histogram of the connectivity distribution when the soft threshold value

325   was 8. E Check scale-free topology scale when the soft threshold value was 8.

326

327   Fig. 3 The co-expression network constructed based on the GSE93798 dataset. A Gene

328   dendrogram. B Heat map of module–trait relationships. C Distribution of average gene

329   significance in the 6 modules. D Heat maps visualizing 400 randomly selected genes in the network

330   to depict the TOM. E The combination of eigengene dendrogram and heatmap. F The distribution

331   of GS and MM was presented as a scatter diagram in the turquoise module.

332

333   Fig. 4 GO analysis of the key module genes. A The top 6 components of the GO analysis for

334   biological processes. B Phylogenic tree of the results of the GO analysis for biological processes.

335

336   Fig. 5 DEG screening. A, B Volcano and heatmap of DEGs by screening the GSE93798 dataset.

337   C, D, E Gene cluster filtered using the MCODE plugin. F The final hub genes determined by

338   Venn diagram.

339

340   Fig. 6 Validation of key genes. A, B Expression of FOSB and DUSP1 in the validation datasets

341   of GSE37460 (A) and GSE104948 (B). C Representative image of FOSB and DUSP1

342   immunohistochemistry in renal biopsy tissues of the control group (n=3) and the IgAN group

343   (n=3). D, E The semi-quantitative analyses of C.

344

345   Fig. 7 The ROC curve of FOSB and DUSP1 in datasets. The ROC curve of FOSB and DUSP1 in

346   the GSE37460 (A), GSE93798 (B) and GSE104948 (C) datasets; ROC, receiver operating

347   characteristic; AUC: area under the ROC curve

Figure 1

Flowchart of this research

GEO dataset GSE93973

Expression profiles of 19702 genes

Ranked by MAD

Limma method

Top 5,000 genes to construct WGCNA

Identifying DEGs

Functional enrichment analysis

PPI network and MCODE

Identifying key module genes

Screening hub genes of DEGs

Identification of key genes

ROC curve analysis

Validation of key genes in GSE37460、GSE104948 datasets and clinical Kidney tissues

Figure 2

Data verification and filter the best soft threshold.

A Sample clustering to view heterogeneity. B Soft threshold non-scale fitting index analysis. C Soft threshold average connectivity analysis. D Histogram of the connectivity distribution when the soft threshold value was 8. E Check scale-free topology scale when the soft threshold value was 8.
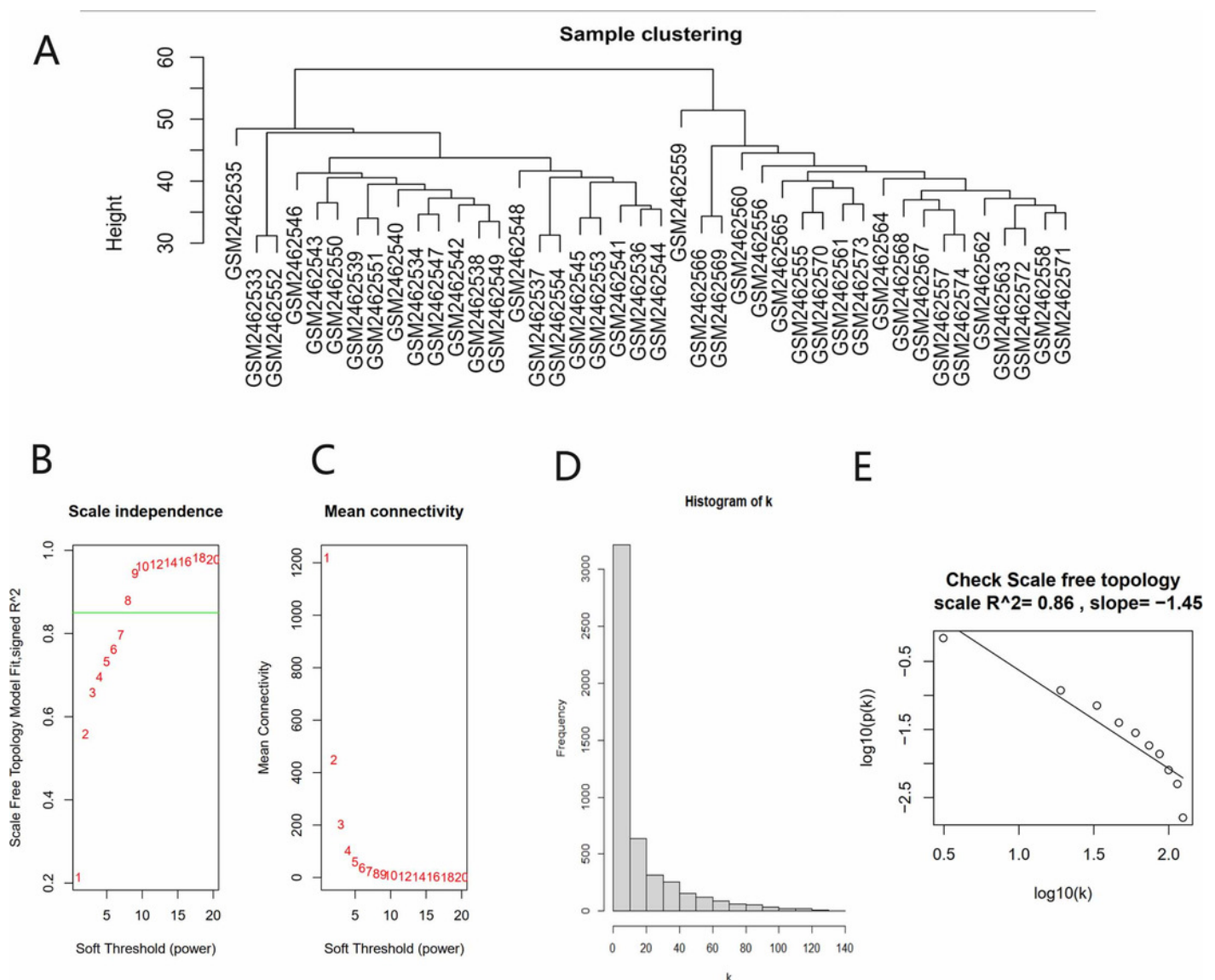
Figure 3

The co-expression network constructed based on the GSE93798 dataset.

A Gene dendrogram. B Heat map of module–trait relationships. C Distribution of average gene significance in the 6 modules. D Heat maps visualizing 400 randomly selected genes in the network to depict the TOM. E The combination of eigengene dendrogram and heatmap. F The distribution of GS and MM was presented as a scatter diagram in the turquoise module.
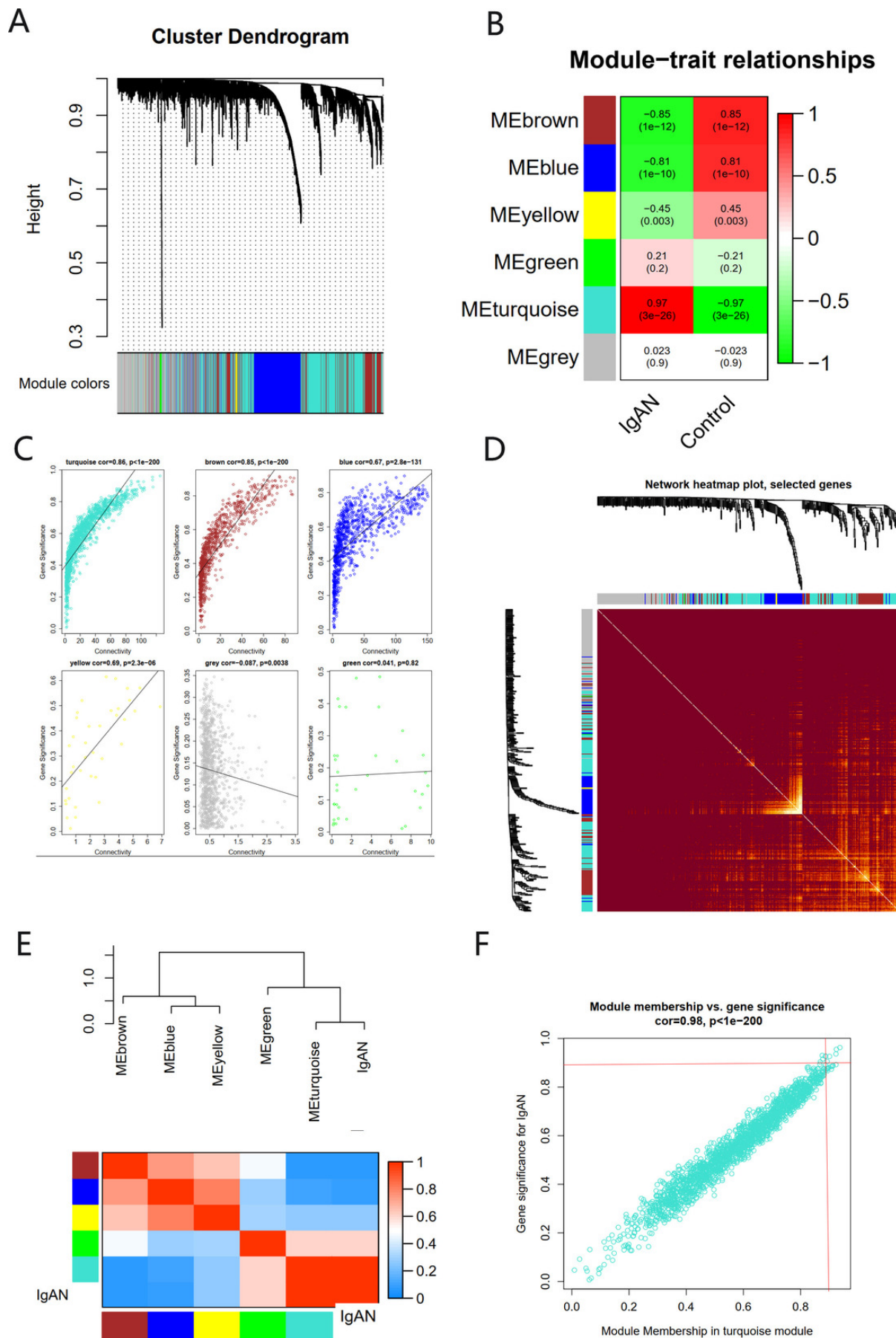
Figure 4

GO analysis of the key module genes.

A The top 6 components of the GO analysis for biological processes. B Phylogenic tree of the results of the GO analysis for biological processes.
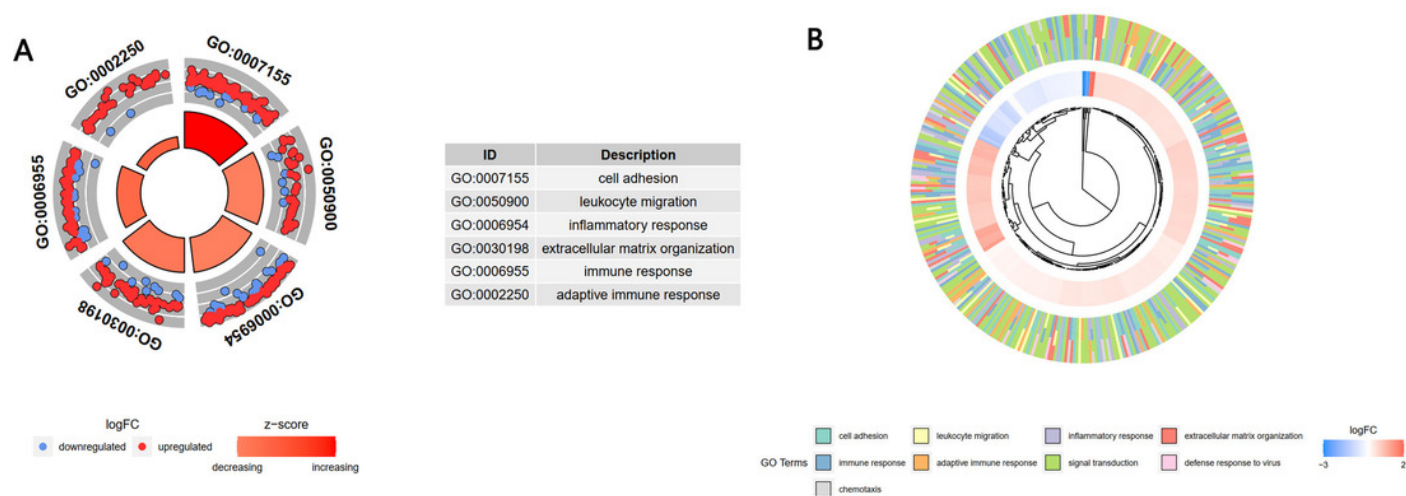
Figure 5

DEG screening.

A, B Volcano and heatmap of DEGs by screening the GSE93798 dataset. C, D, E Gene cluster filtered using the MCODE plugin. F The final hub genes determined by Venn diagram.
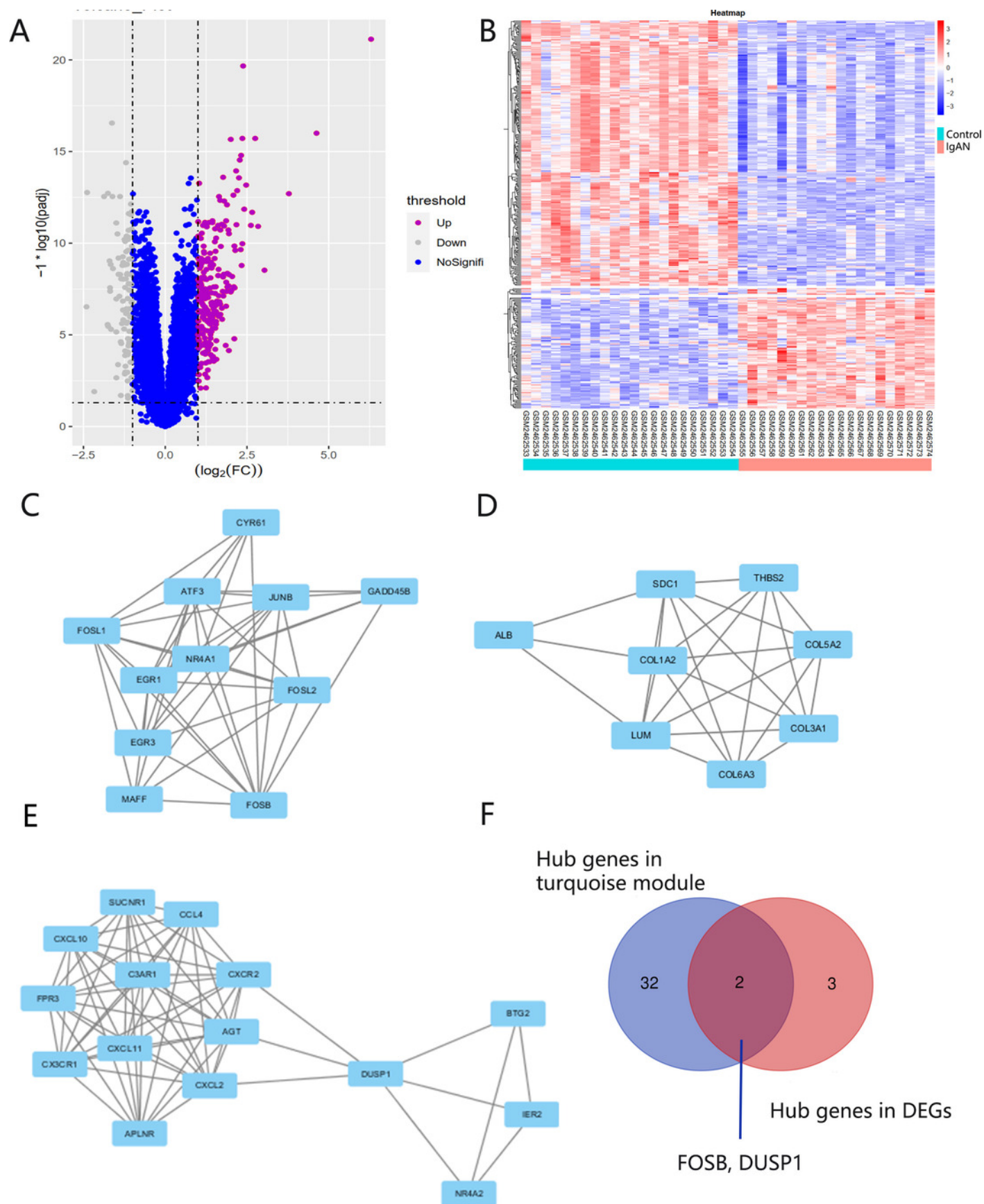
Figure 6

Validation of key genes.

A, B Expression of FOSB and DUSP1 in the validation datasets of GSE37460 (A) and GSE104948 (B). C Representative image of FOSB and DUSP1 immunohistochemistry in renal biopsy tissues of the control group (n=3) and the IgAN group (n=3). D, E The semi-quantitative analyses of C.

A

B

C

D

E
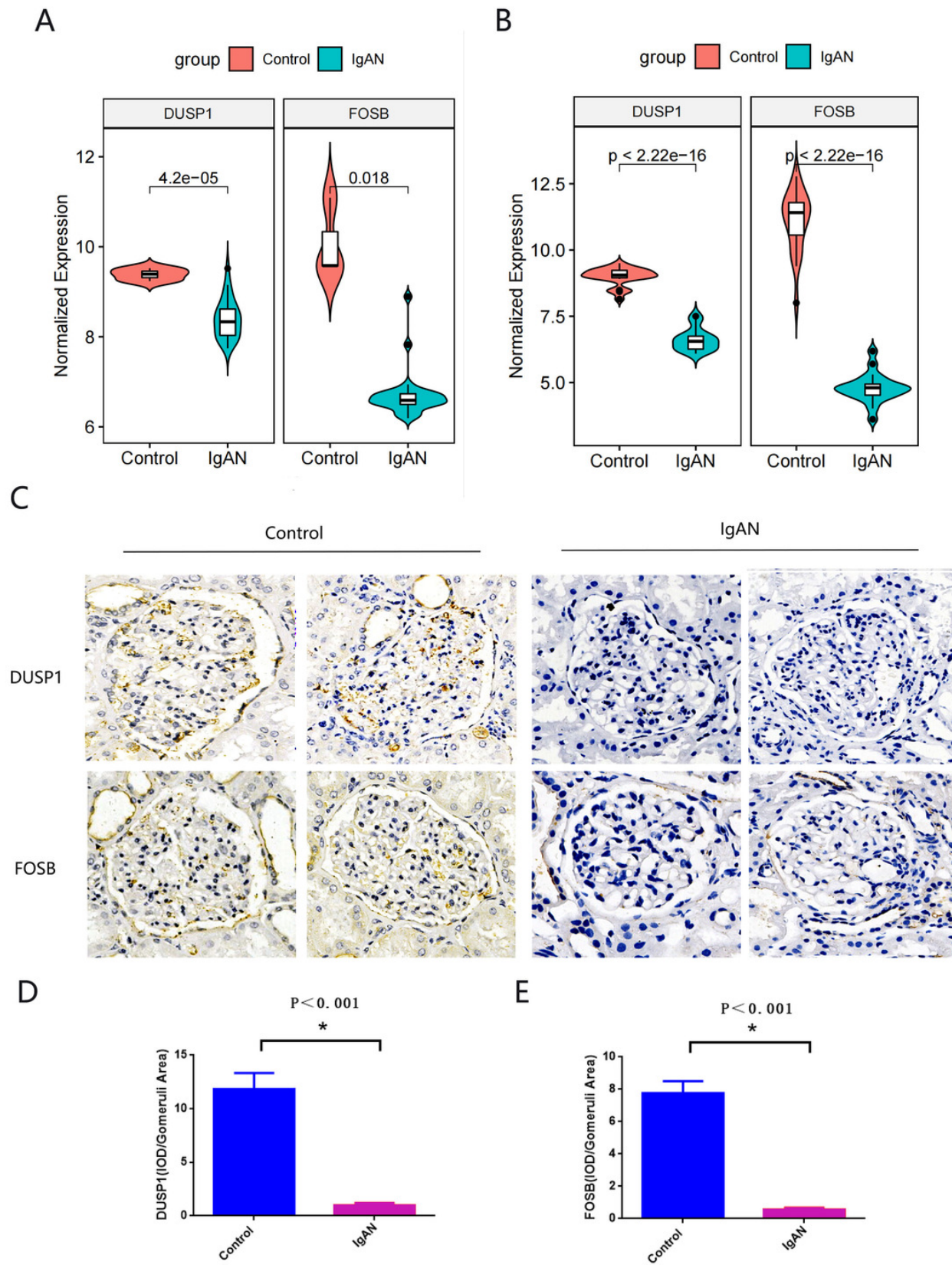
Figure 7

The ROC curve of FOSB and DUSP1 in datasets.

The ROC curve of FOSB and DUSP1 in the GSE37460 (A), GSE93798 (B) and GSE104948 (C) datasets; ROC, receiver operating characteristic; AUC: area under the ROC curve