# Screening and identification of key biomarkers for diabetic kidney disease: based on transcriptome sequencing and bioinformatics analysis

Jing Zhao [Equal first author, 1, 2] , Kaiying He [Equal first author, 1, 2] , Hongxuan Du [1, 2] , Guohua Wei [2] , Yuejia Wen [1, 2] , Jiaqi Wang [1] , Xiaochun Zhou [Corresp., 2] , Jianqin Wang [Corresp. 2]

[1] Lanzhou University, Lanzhou, China

[2] Lanzhou University Second Hospital, Lanzhou, China

Corresponding Authors: Xiaochun Zhou, Jianqin Wang
Email address: ery_zhouxc@lzu.edu.cn, ery_wangjqery@lzu.edu.cn

**Background** Diabetic kidney disease (DKD) is the leading cause of death in people with type 2 diabetes mellitus (T2DM). The main objective of this study is to find the potential biomarkers for diabetic kidney disease. **Materials and Methods** Two datasets (GSE86300 and GSE184836) retrieved from GEO database were used, combined with our RNA sequencing (RNA-seq) results of DKD mice (C57 BLKS-26W db/db) and nondiabetic (db/m) mice for further analysis. After processing the expression matrix of the 3 sets of data using R software "Limma", differential expression analysis was performed. The significantly differentially expressed genes (DEGs) ($|logFC| > 1$, $p$-value $< 0.05$) were visualized by heatmaps and volcano plots respectively. Next, the co-expression genes expressed in the 3 groups of DEGs were obtained by constructing a Venn diagram. In addition, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were further analyzed the related functions and enrichment pathways of these co-expression genes. Then, qRT-PCR was used to verify the expression levels of co-expression genes in the kidney of DKD and control mice. Finally, protein-protein interaction network (PPI), GO, KEGG analysis and Pearson correlation test were performed on the experimentally validated genes, in order to clarify the possible mechanism of them in DKD disease. **Results** Our RNA-seq results identified a total of 124 DEGs, including 58 up-regulated and 66 down-regulated DEGs. At the same time, 181 up-regulated and 151 down-regulated DEGs were obtained in GEO database GSE86300, and 107 up-regulated and 208 down-regulated DEGs were obtained in GSE184836. Venn diagram showed that 10 co-expression DEGs among the 3 groups of DEGs. GO analysis showed that BP was mainly enriched in steroid metabolism, lipid localization, cellular ketone metabolism, fatty acid metabolism, organic hydroxy compounds, response to fatty acid, and lipid transport. KEGG pathway analysis showed that the 3 major enriched pathways were cholesterol

metabolism, steroid hormone biosynthesis, and bile secretion. After qRT-PCR validation, we obtained 8 genes that were significant differentially expressed in the kidney tissues of DKD mice compared with control mice. (The mRNA expression levels of Ugt8a, Apoh, Akr1c14, Lpl, Slc1a4, and Cd36 were declined, whereas Abcc4 and Ltc4s were elevated.)

**Conclusion** Our study, based on RNA-seq results, GEO databases and qRT-PCR, identified 8 significant dysregulated DEGs, which play an important role in lipid metabolism and provide novel targets for diagnosis and treatment of DKD.

**Screening and identification of key biomarkers for diabetic kidney disease: based on transcriptome sequencing and bioinformatics analysis**

4  Jing Zhao[a,b*], Kaiying He[a,b*], Hongxuan Du[a,b], Guohua Wei[b] , Yuejia Wen[a,b] , Jiaqi Wang[a], Xiaochun Zhou[b#],

5  Jianqin Wang[b#]

6  a.Lanzhou University, Lanzhou, Gansu, China

7  b.Department of Nephrology, Lanzhou University Second Hospital, Lanzhou, Gansu,China

9  Short Title: Key biomarkers for diabetic kidney disease from transcriptome sequencing and bioinformatics

10  analysis

11  #Corresponding Author:

12  Xiaochun Zhou

13  Department of Nephrology

14  Lanzhou University Second Hospital

15  No. 82, Cuiyingmen

16  Lanzhou,Gansu, 730030,China

17  Tel:15214057999

18  E-mail: ery_zhouxc@lzu.edu.cn

20  Jianqin Wang

21  Department of Nephrology

22  Lanzhou University Second Hospital

23  No. 82, Cuiyingmen

24  Lanzhou,Gansu, 730030,China

25  Tel:13919038189

26  E-mail: ery_wangjqery@lzu.edu.cn

28  #*Equal study contribution

29

30  Number of Tables:2

31  Number of Figures:5

32  Number of supplementary Tables:2

33

34  Key words: DKD; RNA-seq; bioinformatics analysis; differentially expressed genes (DEGs); biomarkers.

35

36

37

38

39

40

41

42

43

44  **Abstract**

45  **Background** Diabetic kidney disease (DKD) is the leading cause of death in people with type 2 diabetes mellitus

46  (T2DM). The main objective of this study is to find the potential biomarkers for diabetic kidney disease.

47  **Materials and Methods** Two datasets (GSE86300 and GSE184836) retrieved from GEO database were used,

48  combined with our RNA sequencing (RNA-seq) results of DKD mice (C57 BLKS-26W db/db) and nondiabetic

49  (db/m) mice for further analysis. After processing the expression matrix of the 3 sets of data using R software

50  "Limma", differential expression analysis was performed. The significantly differentially expressed genes

51  (DEGs)(|logFC| > 1, $p$-value < 0.05) were visualized by heatmaps and volcano plots respectively. Next, the co-

52  expression genes expressed in the 3 groups of DEGs were obtained by constructing a Venn diagram. In addition,

53  Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis

54  were further analyzed the related functions and enrichment pathways of these co-expression genes. Then, qRT-

55  PCR was used to verify the expression levels of co-expression genes in the kidney of DKD and control mice.

56  Finally, protein-protein interaction network (PPI), GO, KEGG analysis and Pearson correlation test were

57  performed on the experimentally validated genes, in order to clarify the possible mechanism of them in DKD

58  disease.

59  **Results** Our RNA-seq results identified a total of 124 DEGs, including 58 up-regulated and 66 down-regulated

60  DEGs. At the same time, 181 up-regulated and 151 down-regulated DEGs were obtained in GEO database

61  GSE86300, and 107 up-regulated and 208 down-regulated DEGs were obtained in GSE184836. Venn diagram

62  showed that 10 co-expression DEGs among the 3 groups of DEGs. GO analysis showed that BP was mainly

63  enriched in steroid metabolism, lipid localization, cellular ketone metabolism, fatty acid metabolism, organic

64  hydroxy compounds, response to fatty acid, and lipid transport. KEGG pathway analysis showed that the 3 major

65  enriched pathways were cholesterol metabolism, steroid hormone biosynthesis, and bile secretion. After qRT-

66  PCR validation, we obtained 8 genes that were significant differentially expressed in the kidney tissues of DKD

67  mice compared with control mice. (The mRNA expression levels of Ugt8a, Apoh, Akr1c14, Lpl, Slc1a4, and

68  Cd36 were declined, whereas Abcc4 and Ltc4s were elevated.)

69  **Conclusion** Our study, based on RNA-seq results, GEO databases and qRT-PCR, identified 8 significant

70  dysregulated DEGs, which play an important role in lipid metabolism and provide novel targets for diagnosis

71  and treatment of DKD.

72  **Key words:** DKD; RNA-seq; bioinformatics analysis; differentially expressed genes (DEGs); biomarkers.

73  **1.Introduction**

74  Diabetic mellitus (DM) is a chronic metabolic disease that seriously affects public health. According to the

75  World Health Organization (WHO), about 629 million people will suffer from T2DM by 2045, and the

76  complications may affect various organs throughout the body and have a high mortality and disability rate [1].

77  Diabetic kidney disease (DKD) is a common chronic complication of DM, which has gradually overtaken

78  glomerulonephritis and hypertension and is the leading cause of death in patients with DM and end-stage renal

79  disease (ESRD) [2]. The main features of progression of DKD are hypertension, increased protein in urine, and

80  decreased estimated glomerular filtration rate (eGFR). Early pathological changes mainly include thickening of

81  the glomerular and tubular basement membrane, which gradually develops into glomerular extracellular matrix

82  (ECM) accumulation and tubular interstitial fibrosis as the disease progresses, eventually causing irreversible

83    damage to the renal structure. Most experts believe that the pathogenesis of DKD is mainly due to the interaction
84    of renin-angiotensin-aldosterone system (RAAS), advanced glycation end products (AGEs), transforming
85    growth factor-β1 (TGF-β1), protein kinase C (PKC), mitogen-activated protein kinases (MAPKs) and reactive
86    oxygen species (ROS), which affect renal function through inflammation and oxidative stress. [3]Although
87    various treatments have been used to improve metabolism, hemodynamic disturbances, and fibrosis, the
88    mortality rate of patients with DKD remains high. Therefore, the key to improving the quality of life and survival
89    rate is to find biomarkers for early diagnosis of DKD and to develop new strategies and prevention of
90    deterioration of renal function.

91    In recent years, many biomarker molecules have been found to be associated with changes in renal structure and
92    function in DKD patients, such as urine markers, serum/plasma markers, etc [4]. However, it can only be used
93    for the current diagnosis of DKD, fails to diagnose at early stage. With the development of the new generation
94    of high-throughput sequencing technology and bioinformatics techniques, the ability of humans to understand
95    diseases from the root has greatly improved, and more and more disease-related risk genes have been discovered.
96    This promises revolutionary advances in disease diagnosis and treatment in the future [5]. Many microarray-
97    based studies have shown that some small molecules such as non-coding RNAs and mRNAs play important
98    roles in the pathogenesis of DKD. They influence disease occurrence, progression, and prognosis through their
99    interactions and regulation of signaling pathways. For example, studies have shown that leucine rich-α-2
100   glycoprotein 1 (LRG1) expression is increased in kidneys of diabetic patients and mice, it can be used as a
101   biomarker for early diagnosis of DKD[6]. The expression of SH3YL1 in the serum of DKD patients, in kidney
102   tissues of db/db mice, and in podocytes, is higher than control group, so it also has the potential as a biomarker
103   for the early diagnosis of DKD [7]. By combining clinicopathological data with mRNA analysis of urine from
104   DKD patients, it was found that LYZ, C3, FKBP5, and G6PC are associated with renal tubulointerstitial
105   inflammation and fibrosis, and their expression levels reflect the renal pathological types of DKD patients, and
106   that they can also predict the prognosis of renal function[8]. Some researchers also used bioinformatics methods
107   to analyze multiple datasets in the GEO database and proved that the increased expression of AEBP1 in plasma
108   exosomes of DKD patients may also be a novel biomarker [9].

109   The pathogenesis of DKD is complex, which is related to oxidative stress, inflammation, autophagy, apoptosis,
110   and other mechanisms. However, its pathogenesis still needs further study. Although many mRNAs are currently
111   thought to play a role in DKD, there are few consistent results across studies. We need to obtain more data by
112   biotechnological methods and increase the number of experiments so that the results of the different genes have
113   higher authenticity and reliability.

114   Our analysis was conducted following the procedure presented in Fig 1. Kidney tissues from DKD and control
115   mice were isolated and RNA was extracted for RNA-seq, and DEGs were determined. To reduce the false-
116   positive rate of experimental results, the GSE86300 and GSE184836 RNA-seq datasets of kidneys from DKD
117   mice were extracted from the GEO (Gene Expression Omnibus) database. Finally, 10 co-expression genes were
118   determined by the combined analysis of the 3 datasets, qRT-PCR verified that 8 of them had the same expression
119   trend as the above sequencing results.

120   **2.Materials and Methods**
121   **2.1 Data source**
122   **2.1.1 Transcriptome sequencing of kidney tissues from DKD and Control mice**
123   **1) Animal model**

124    Six C57 BLKS-db/db mice and six db/m mice(6-week-old,male)were selected as the model group (body weight
125    34.35±2.52g),normal control group(body weight 18.94±1.44g) respectively. They were purchased from Nanjing
126    Institute of Model Zoology and reared in the barrier system of Animal Experimental Research Centre of Lanzhou
127    University Second Hospital. The feeding temperature is (20±2) ℃, the humidity is 40%-70%, the light alternates
128    between light and dark every 12 hours, the ordinary Specific Pathogen Free (SPF) food is fed and the drinking
129    water is free. After these mice were fed for 20 weeks, the mice were euthanized by intraperitoneal injection of
130    10% chloral hydrate and the kidney tissues were collected. 50 mg of the right kidney was divided into cryogenic
131    vials, frozen with liquid nitrogen and stored in the refrigerator at -80°C. All the experiments involving animals
132    were performed after Lanzhou University Second Hospital Institutional Ethical Committee's approval and under
133    the strict adherence to the National Institutes of Health Guide for laboratory animals' care and use, our ethical
134    review acceptance number is D2019-198.
135    **2) RNA extraction from kidney tissues**
136    Approximately 50 mg of each kidney tissue sample was placed in grinding tubes, and 1 ml of Trizol was added
137    to each tube and ground completely. Then the samples were allowed to stand for 5 minutes at room temperature,
138    and 0.2 ml of chloroform was added to each tube and shaken vigorously for 15 seconds. After the samples stand
139    at room temperature for 3 minutes, the supernatant was centrifuged at high speed (12000 rpm, 4℃) for 15 minutes
140    and transferred to a Rnase-free centrifuge tube. Next, add 0.5 mL isopropyl alcohol, mix gently and let stand at
141    room temperature for 10 minutes, centrifuge for 10 minutes (12000 rpm, 4℃), wash the precipitate with 1 mL
142    75% ethanol, centrifuge for 5 minutes (7500 rpm, 4℃), remove the liquid, dry at room temperature for 10
143    minutes. 150uL diethyl pyrocarbonate (DEPC) $H_2O$ was added and gently mix. At last, the precipitate was left
144    at 55℃ for 10 minutes to dissolve and stored temporarily at -80℃ for further sequencing.
145    **3) RNA-sequencing(RNA-seq)**
146    In this study, we need to consider and deal with the difference expression caused by the biological variability,
147    by far the most commonly used and most effective way is to set up biological replicates, biological replicates
148    can not only test the relevance of the biology of repeatability, it can also assess the reliability of differentially
149    expressed genes and assist in screening abnormal samples. Pearson's correlation coefficient was taken as the
150    evaluation index of biological repetition correlation (the closer it is to 1, the stronger the correlation between
151    two repeated samples is)(Supplementary fig.1). All biological repeat samples under the same conditions were
152    extracted and built with the same person and batch, sequenced with same Run and Lane, and conducted a detailed
153    analysis of the abnormal samples. Meanwhile, we set the Power value as 0.9, α value as 0.05 and then used the
154    R language RNAseqPower package to calculate the sample size was 2.44, which could achieve 1.5 times of
155    change, proving that the number of mice in each group selected as 3 is validate in our experiment.
156    Illumina's high-throughput sequencing platform was used to perform transcriptome sequencing on kidney tissue
157    samples from DKD mice and Control mice. Clean data were obtained by filtering data from the Illumina high-
158    throughput sequencing platform sequenced with the indicated reference genome. Subsequently, insertion
159    fragment length test, random test, library quality assessment, alternative splicing analysis, novel gene discovery
160    and gene structure optimization were performed. Finally, the number of mapped reads and transcript length were
161    normalized in the sample. Fragments Per Kilobase Million (FPKM) was used as an indicator to measure gene
162    expression levels. The expression matrix of FPKM can be obtained.
163    **2.1.2 GEO database**
164    GEO database was built by the National Center for Biotechnology Information (NCBI) and is a gene expression

165   database and online genome resource that collects high-throughput gene expression data uploaded from research
166   institutes around the world. Diabetic nephropathy or diabetic kidney disease was entered as search objects. Gene
167   expression microarray datasets GSE86300 and GSE184836 were selected and downloaded. The criteria for
168   selecting the datasets were as follows: DKD and Control mice models with detailed gene expression information.
169   The GSE86300 dataset, based on the GPL7546 platform, includes 10 DKD kidney tissue samples and 10 Control
170   kidney tissue samples. The GSE184836 dataset, based on the GPL21103 platform, includes 6 DKD kidney tissue
171   samples and 6 Control kidney tissue samples. The detailed information on these microarray datasets is listed in
172   Supplemental Table 1.

173   **2.2 Data processing methods**

174   **2.2.1 Differential expression analysis**

175   All differential analyses of the 3 datasets were performed using the Limma packages in R/Bioconductor software.
176   Genes in DKD renal tissue were up- or down-regulated compared with control groups, $p$-value less than 0.05
177   and |log$_2$ fold change (FC)| greater than 1 were considered statistically significant in differential analysis of the
178   3 datasets, logFC＞1 was regarded as up-regulated genes and logFC＜-1 was down-regulated.

179   **2.2.2 Volcano plots and heatmaps**

180   Significant DEGs (|logFC| > 1, $p$-value < 0.05) were visualized using heatmaps and volcano curves. Both
181   visualizations are completed on the "Weishengxin" website (http://www.bioinformatics.com.cn/).

182   **2.2.3 Venn diagram**

183   Determine the common genes of 3 datasets by creating a Venn diagram. The Venn diagram is completed on the
184   "Weishengxin" website.

185   **2.2.4 Functional enrichment analysis.**

186   Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment
187   analysis were performed for DEGs. GO Annotation included analysis of biological processes (BP), cell
188   components (CC), and molecular functions (MF). KEGG is an online database for pathway analysis of a large
189   amount of genetic information. GO functional annotation analysis and KEGG pathway enrichment analysis were
190   performed in the website "Weishengxin". The significant enrichment threshold was set as $p$-value<0.05, ten
191   functions and pathways with the lowest $p$-values were selected as the top 10.

192   **2.2.5 Quantitative real-time polymerase chain reaction(qRT-PCR)**

193   Total RNA was extracted from fresh mice kidneys of DKD and control groups using the TRIZOL method. Total
194   RNA (1 ug) was transcribed into cDNA using the GoScript$^{TM}$ Reverse Transcription System according to the
195   manufacturer's protocol (Promega). qRT-PCR was performed on the ABI7500 system using the GoTaq$^®$ qPCR
196   Master Mix (Promega). All data were normalized to β-actin expression. Relative RNA expression was calculated
197   using the $2^{-\Delta\Delta CT}$ method. Detailed information of primers was listed in Supplemental Table 2.

198   **2.2.6 PPI network analysis and Pearson correlation test.**

199   The protein-protein interaction (PPI) network was created from the "string" database for target genes. The goal
200   is to discover the interaction of target genes and their interactions with other proteins. The expression matrix of
201   the 3 datasets was merged after removing batch effect and the Pearson correlation test was adopted for evaluation
202   of the interactions between DKD related genes at the mRNA level using R.

203   **2.3 Statistical analysis**

204   All values are expressed as the mean±SEM. Statistical analysis was performed using the statistical package SPSS
205   for Windows Version 7.51 (SPSS, Inc, Chicago, IL, USA). Results were analyzed using Student t test for

206   multiple comparisons. Statistical significance was detected at the 0.05 level.

207   **3.Results**

208   **3.1 Differential expression gene**

209   Based on our RNA-seq results, a total of 124 differentially expressed mRNAs were identified, including 58 up-
210   regulated genes and 66 down-regulated genes. After screening the genes with differential expression in
211   GSE86300, 331 genes with differential expression were determined, including 181 up-regulated genes and 151
212   down-regulated genes. In GSE184836 DEGs, there were 315 differentially expressed genes, including 107 up-
213   regulated genes and 208 down-regulated genes. The 3 datasets were analyzed according to the criteria $|\log_2 FC|$
214   ＞1 and $p$-value＜0.05. To visualize the differentially expressed genes, we constructed heatmaps (Fig. 2A,C,E)
215   and volcano curves (Figs. 2B,D,F) and then created a Venn diagram to obtain 10 co-expression genes among the
216   3 datasets(Fig. 2G).

217   **3.2 Enrichment analysis**

218   GO functional annotation and KEGG pathway enrichment analysis were performed for the DEGs from the 3
219   datasets. The GO analysis results of our RNA-seq showed that the BP analysis was mainly enriched in the steroid
220   metabolic process, lipid localization, cellular ketone metabolic process, fatty acid metabolic process, organic
221   hydroxy compound metabolic process, response to fatty acid and lipid transport, while CC is mainly
222   concentrated in chylomicron, very-low-density lipoprotein particle and triglyceride-rich plasma lipoprotein
223   particle. The top 3 of MF are oxidoreductase activity, monooxygenase activity, alditol: NADP+ 1-
224   oxidoreductase activity respectively (Fig. 3A). KEGG pathway analysis showed that the top 3 enrichment
225   pathways of the DEGs are cholesterol metabolism, steroid hormone biosynthesis and bile secretion (Fig. 3B).
226   GSE86300 analysis showed that the BP of DEGs was mainly in the fatty acid metabolism, steroid metabolism,
227   and regulation of lipoprotein lipase activity (Fig.3C). At the same time, KEGG analysis showed they were mainly
228   in the cholesterol metabolism, complement and coagulation cascades, and legionellosis pathway (Fig.3D).
229   GSE184836 analysis indicated that organic hydroxy compound metabolic process, steroid metabolic process,
230   fatty acid metabolic process as the top 3 BP (Fig.3E). Meanwhile, KEGG analysis manifested that butanoate
231   metabolism, malaria and pyruvate metabolism as the top 3 (Fig.3F).

232   **3.3 Venn diagram**

233   10 overlapping genes, namely, Ltc4s, Cd36, Slc1a4, Slc22a19, Lpl, Akr1c14, Apoh, Fmo5, Abcc4, Ugt8 were
234   identified from the 3 datasets by creating a Venn diagram. The expression levels and trends of the above 10
235   genes in 3 databases were listed in Table1.

236   **3.4 Identification of 10 overlapping genes expression by qRT-PCR**

237   The mRNA levels of the 10 overlapping genes were verified by qRT-PCR in DKD and control mice kidney
238   tissues, meanwhile, the results indicated that the mRNA levels of Ugt8a, Apoh, Akr1c14, Lpl, Slc1a4 and Cd36
239   were down-regulated, while Abcc4 and Ltc4s were up-regulated, these changes were statistically significant ($p$
240   < 0.05). In addition, qRT-PCR results of Fmo5 showed that its expression was up-regulated in DKD kidney
241   tissues, which is contrary to the sequencing results. And the expression of Slc22a19 showed no statistical
242   significance in the DKD group compared with the Control ($p$ >0.05) (Fig 4). In conclusion, the mRNA expression
243   levels of these 8 genes verified by qRT-PCR were consistent with the sequencing results (Ugt8a, Apoh, Akr1c14,
244   Lpl, Slc1a4, Cd36, Abcc4, Ltc4s).

245   **3.5 Protein interaction analysis**

246   The "string" database was used for PPI analysis. The results of PPI analysis of the above verified 8 genes were

247    similar to those of functional annotation of GO and KEGG Pathway Enrichment Analysis (Fig. 5A-B). The 8
248    genes apparently showed mainly correlation with lipid metabolism, and of these 8 genes, 3 had an obvious
249    interaction relationship, namely Cd36, Apoh, and Lpl (Fig. 5C). Similarly, the enrichment analysis of the BP of
250    8 genes revealed that the above 3 genes were all enriched in the triglyceride metabolic process. Additional
251    information of the 8 verified genes related to BP analysis is listed in Table2. In addition, through the Pearson
252    correlation test, there were prominently positive or negative correlations between the DKD related genes at the
253    mRNA levels (Fig. 5D).

254    **4.Discussion**

255    In recent years, the consequences of diabetes complications have gradually attracted people's attention. With the
256    increasing incidence of DKD, it brings a heavy medical burden to society. Nowadays, the clinical diagnosis of
257    DKD mainly depends on the protein content in urine, and the treatment aims to reduce the protein content in
258    urine. With the development of sequencing technology, we now realize that genes play a key role in the
259    diagnosis, occurrence, progression and treatment of DKD. Thus, many mRNAs and non-coding RNAs
260    associated with DKD have been discovered and supported by a large number of studies.

261    In our study, we found 8 genes whose expression levels were significantly altered in the kidney tissues of DKD
262    mice. Cd36, Ugt8a (UDP-galactosyltransferase 8A), Apoh (apolipoprotein H), Akr1c14 (aldoketo reductase
263    family 1, member C14), Lpl (lipoprotein lipase), and Slc1a4 (solute carrier family 1 member 4) were
264    downregulated, whereas the expression levels of Abcc4 and Ltc4s were upregulated. Enrichment analysis of
265    genes with distinct expression differences showed that these genes were involved in the biological processes of
266    KEGG pathways such as steroid metabolism, lipid localization, cellular ketone metabolism, fatty acid
267    metabolism, cholesterol metabolism, steroid hormone biosynthesis and bile secretion. Analysis of the 8 genes
268    verified by qRT-PCR indicated that the function of these genes was mainly related to lipid metabolism.

269    The glucose and lipid metabolism disorder has become a very common metabolic defect, and is also an important
270    factor leading to DKD. Researchers have confirmed that dyslipidemia and the increased free fatty acids (FFA)
271    are risk factors for insulin resistance and glucolipid toxicity is an important cause of T2DM[10].Lipid toxicity is
272    mainly involved in renal damage related to lipid toxicity through the activation of inflammation, oxidative stress,
273    mitochondrial dysfunction and apoptosis [11], and increased lipophagy ameliorates renal damage, lipid
274    deposition, oxidative stress and apoptosis in the kidney[12].

275    There are a number of studies on CD36, it is believed to be involved in fatty acid uptake, apoptosis, angiogenesis,
276    phagocytosis, inflammation and atherosclerosis[13]. Some scholars have proposed that high glucose exacerbates
277    fatty acid uptake and deposition through increased expression of the class B scavenger receptor CD36 via the
278    AKT-PPARγ pathway and that CD36 in plasma may be a biomarker for T2DM[14, 15]. CD36 knockdown prevents
279    renal tubular injury, tubulointerstitial inflammation, and oxidative stress in 16 weeks db/db mice[16]. In addition
280    to these, data suggests that CD36 deficiency aggravates hepatic lipid accumulation and reduced VLDL secretion
281    in ob/ob mice[17]. And the loss of CD36 in mice leads to phenotypes such as lymphatic drainage, visceral fat, and
282    glucose intolerance, which increases the risk for T2DM[18].

283    In addition, studies have shown that in C57BL/6J mice, Lpl was mainly expressed intracellular and restricted to
284    the proximal tubule[19]. In the DKD group, renal LPL mRNA expression was significantly decreased, and down-
285    regulation of renal LPL increased the level of triglyceride (TG) in renal tissue[20]. Apoh is involved in a variety
286    of physiological processes, including lipoprotein metabolism, coagulation, and antiphospholipid autoantibody
287    production[21]. It is also associated with BMI in diabetics and cardiovascular risk factors and is considered an

288  anti-obesity factor[22].Ugt8a encodes galactosylceramide synthase, and alteration in expression levels can lead to
289  impaired lipid metabolism. The role of Ugt8a in regulating lipid metabolism in the nervous system has been
290  confirmed by a number of studies[23], and some experts have suggested that the elevated Ugt8 levels may be a
291  protective compensatory response in DKD to prevent podocyte degeneration by enhancing their survival[24]. The
292  protein encoded by Ltc4s is an enzyme that catalyzes the synthesis of leukotriene C4 by linking leukotriene A4
293  to reduced glutathione. Leukotrienes are thought to mediate anaphylaxis and inflammatory
294  conditions[25].Unilateral ureteral obstruction (UUO), folate nephropathy and orthologous mouse model were
295  established and Ltc4s synthase was knocked down, it was found that fibrosis was reduced and renal function
296  was improved in the Ltc4s knockout model [26]. Akr1c14, a metabolic enzyme involved in the reduction of
297  aldehydes and ketones and glycoside metabolic process, has been shown to be downregulated in adipocytes in
298  models of adipose tissue inflammation and thus is thought to play a role in the maintenance of normal adipocyte
299  metabolism[27]. Abcc4 was highly expressed in the kidney of mice. Aberrant hypomethylation of Abcc4 was
300  detected in C57BLKS/J db/db mice[28], studies have also shown that the expression of Mrp4/Abcc4 is increased
301  in metabolic tissues such as liver and kidney in obese and diabetic patients[29].Slc1a4 is highly expressed in the
302  central nervous system, and relevant studies are still lacking. Further studies on the mechanism of action of
303  Slc1a4 in DKD are needed in the future.
304  The present study identified novel molecular targets by integrating our RNA-seq and microarray datasets in
305  GEO, we found 8 genes differentially expressed in the kidney tissues of DKD mice and verified by qRT-PCR.
306  Through GO, KEGG and PPI analysis, 3 genes caught our attention: Lpl, Apoh and Cd36, these 3 genes were
307  not only differentially expressed in DKD mice kidney tissue, but also closely related to triglyceride metabolism,
308  the results of the "String" database analysis indicated that these 3 genes may have mutual regulation, and we can
309  see the other proteins that interact with them, for example, they may interact with Angptl4, Apoa1, Apoa2 and
310  Apob. ANGPTL4 has been studied as a potent inhibitor of Lpl that regulates cellular uptake of triglycerides and
311  promotes fatty acid oxidation[30], Binod [31] found that the expression of Cd36 increased in Angptl4$^{-/-}$
312  macrophages. Apo a and Apo b are mainly studied in obesity and cardiovascular diseases, studies suggest that
313  ApoA1, Apo B, and Apo B/ApoA1 ratio have been regarded as the predictors of microvascular and
314  macrovascular complications of diabetes. We have seen less research on kidney disease, but we speculate from
315  our current study that it may play an important role in kidney disease, more experiments will investigate the
316  mechanism in the future.
317  However, the present study has certain limitations. First, we did not perform a joint analysis of biochemical
318  indicators of mice, which is a defect of this study. Second, the DKD mice models are less representative than
319  kidney samples from DKD patients because they may have only moderate renal abnormalities. Third, the array
320  data came from the later stage of DKD, and therefore, the expression levels of certain genes may not be identical
321  to those in the early stage. For instance, the protein levels of CD36 have been reported to be upregulated in this
322  disease, while this gene was downregulated in our study. In addition, the novel potential target genes should be
323  further verified in experimental studies. The present results were obtained using a bioinformatics screening to
324  identify several novel DEGs between DKD and control groups, and which suggested that these 8 genes have
325  vital roles in the pathological process of lipid metabolism in DKD. Importantly, the information provided in this
326  study was not limited to the 8 verified genes, but maybe included certain other typical DEGs. The current results
327  provide a worthy resource for future research on DKD.
328  **5.Conclusion**

329  Our study, based on RNA-seq results, the GEO database and qRT-PCR, identified 8 significant dysregulated
330  DEGs, which play an important role in lipid metabolism and provide novel targets for diagnosis and treatment
331  of DKD.
332

333  **Acknowledgements**
334  We thank Dr. Xiaochun Zhou, Dr. Jianqin Wang (Lanzhou university second hospital) for their helpful
335  suggestions regarding the experiments and their excellent technical assistance.
336  **Author contributions**
337  XCZ, JQW contributed to the conception of the study; JZ, KYH performed the experiment; HXD, GHW
338  contributed significantly to analysis and manuscript preparation; JQW performed the data analyses and wrote
339  the manuscript; YJW helped perform the analysis with constructive discussions.
340  **Funding**
341  This work was supported in part by the National Natural Science Foundation of China (No. 81960142,
342  81560122), Youth Science and Technology Fund Program of Gansu Province (No. 21JR1RA157), and Talent
343  Innovation and Entrepreneurship Project of Lanzhou City, Gansu Province (2021-RC-94). Meanwhile, our
344  experiments are supported by the Clinical Medical Research Center of Gansu Province (21JR7RA436).

345  **Data availability**
346  All data generated or analyzed during this study are included in this article. Further enquiries can be directed to
347  the corresponding author on reasonable request.
348  **Declarations**
349  Conflict of interest the authors declare that they have no competing interests.
350  **Ethical approval**
351  The study protocol was approved by the Institutional Review Board and Ethics Committee of Lanzhou
352  University Second Hospital. All animal experiments were approved by the Committee on Animal
353  Experimentation of Lanzhou University and performed in compliance with the Guidelines for the Care and Use
354  of Laboratory Animals of the university.
355

356  **References**
357  [1]  TANASE D M, GOSAV E M, NECULAE E, et al. Role of Gut Microbiota on Onset and Progression of
358  Microvascular Complications of Type 2 Diabetes (T2DM) [J]. Nutrients, 2020, 12(12):
359  [2]  ALICIC R Z, ROONEY M T, TUTTLE K R. Diabetic Kidney Disease: Challenges, Progress, and Possibilities
360  [J]. Clin J Am Soc Nephrol, 2017, 12(12): 2032-45.
361  [3]  SAMSU N. Diabetic Nephropathy: Challenges in Pathogenesis, Diagnosis, and Treatment [J]. Biomed Res Int,
362  2021, 2021(1497449.
363  [4]  COLHOUN H M, MARCOVECCHIO M L. Biomarkers of diabetic kidney disease [J]. Diabetologia, 2018,
364  61(5): 996-1011.
365  [5]  REGO S M, SNYDER M P. High Throughput Sequencing and Assessing Disease Risk [J]. Cold Spring Harb
366  Perspect Med, 2019, 9(1):
367  [6]  HONG Q, ZHANG L, FU J, et al. LRG1 Promotes Diabetic Kidney Disease Progression by Enhancing TGF-β-
368  Induced Angiogenesis [J]. J Am Soc Nephrol, 2019, 30(4): 546-62.
369  [7]  CHOI G S, MIN H S, CHA J J, et al. SH3YL1 protein as a novel biomarker for diabetic nephropathy in type 2

370     diabetes mellitus [J]. Nutr Metab Cardiovasc Dis, 2021, 31(2): 498-505.

371     [8]   LEE Y H, SEO J W, KIM M, et al. Urinary mRNA Signatures as Predictors of Renal Function Decline in Patients

372     With Biopsy-Proven Diabetic Kidney Disease [J]. Front Endocrinol (Lausanne), 2021, 12(774436.

373     [9]   TAO Y, WEI X, YUE Y, et al. Extracellular vesicle-derived AEBP1 mRNA as a novel candidate biomarker for

374     diabetic kidney disease [J]. J Transl Med, 2021, 19(1): 326.

375     [10] LYTRIVI M, CASTELL A L, POITOUT V, et al. Recent Insights Into Mechanisms of β-Cell Lipo- and

376     Glucolipotoxicity in Type 2 Diabetes [J]. J Mol Biol, 2020, 432(5): 1514-34.

377     [11] OPAZO-RíOS L, MAS S, MARíN-ROYO G, et al. Lipotoxicity and Diabetic Nephropathy: Novel Mechanistic

378     Insights and Therapeutic Opportunities [J]. Int J Mol Sci, 2020, 21(7):

379     [12] HAN Y, XIONG S, ZHAO H, et al. Lipophagy deficiency exacerbates ectopic lipid accumulation and tubular

380     cells injury in diabetic nephropathy [J]. Cell Death Dis, 2021, 12(11): 1031.

381     [13] PEPINO M Y, KUDA O, SAMOVSKI D, et al. Structure-function of CD36 and importance of fatty acid signal

382     transduction in fat metabolism [J]. Annu Rev Nutr, 2014, 34(281-303.

383     [14] MITROFANOVA A, FONTANELLA A M, MERSCHER S, et al. Lipid deposition and metaflammation in

384     diabetic kidney disease [J]. Curr Opin Pharmacol, 2020, 55(60-72.

385     [15] ALKHATATBEH M J, ENJETI A K, ACHARYA S, et al. The origin of circulating CD36 in type 2 diabetes [J].

386     Nutr Diabetes, 2013, 3(2): e59.

387     [16] HOU Y, WANG Q, HAN B, et al. CD36 promotes NLRP3 inflammasome activation via the mtROS pathway in

388     renal tubular epithelial cells of diabetic kidneys [J]. Cell Death Dis, 2021, 12(6): 523.

389     [17] NASSIR F, ADEWOLE O L, BRUNT E M, et al. CD36 deletion reduces VLDL secretion, modulates liver

390     prostaglandins, and exacerbates hepatic steatosis in ob/ob mice [J]. J Lipid Res, 2013, 54(11): 2988-97.

391     [18] CIFARELLI V, APPAK-BASKOY S, PECHE V S, et al. Visceral obesity and insulin resistance associate with

392     CD36 deletion in lymphatic endothelial cells [J]. Nat Commun, 2021, 12(1): 3350.

393     [19] NYRéN R, MAKOVEICHUK E, MALLA S, et al. Lipoprotein lipase in mouse kidney: effects of nutritional

394     status and high-fat diet [J]. Am J Physiol Renal Physiol, 2019, 316(3): F558-f71.

395     [20] HERMAN-EDELSTEIN M, SCHERZER P, TOBAR A, et al. Altered renal lipid metabolism and renal lipid

396     accumulation in human diabetic nephropathy [J]. J Lipid Res, 2014, 55(3): 561-72.

397     [21] ATHANASIADIS G, SABATER-LLEAL M, BUIL A, et al. Genetic determinants of plasma β$_2$-glycoprotein I

398     levels: a genome-wide association study in extended pedigrees from Spain [J]. J Thromb Haemost, 2013, 11(3): 521-

399     8.

400     [22] HASSTEDT S J, COON H, XIN Y, et al. APOH interacts with FTO to predispose to healthy thinness [J]. Hum

401     Genet, 2016, 135(2): 201-7.

402     [23] HENRIQUES A, CROIXMARIE V, BOUSCARY A, et al. Sphingolipid Metabolism Is Dysregulated at

403     Transcriptomic and Metabolic Levels in the Spinal Cord of an Animal Model of Amyotrophic Lateral Sclerosis [J].

404     Front Mol Neurosci, 2017, 10(433.

405     [24] JAIN S, DE PETRIS L, HOSHI M, et al. Expression profiles of podocytes exposed to high glucose reveal new

406     insights into early diabetic glomerulopathy [J]. Lab Invest, 2011, 91(4): 488-98.

407     [25] FUJIMORI K, UNO S, KURODA K, et al. Leukotriene C(4) synthase is a novel PPARγ target gene, and

408     leukotriene C(4) and D(4) activate adipogenesis through cysteinyl LT1 receptors in adipocytes [J]. Biochim Biophys

409     Acta Mol Cell Res, 2022, 1869(3): 119203.

410     [26] MONTFORD J R, BAUER C, DOBRINSKIKH E, et al. Inhibition of 5-lipoxygenase decreases renal fibrosis

411   and progression of chronic kidney disease [J]. Am J Physiol Renal Physiol, 2019, 316(4): F732-f42.

412   [27]  BECKER J C, GéRARD D, GINOLHAC A, et al. Identification of genes under dynamic post-transcriptional

413   regulation from time-series epigenomic data [J]. Epigenomics, 2019, 11(6): 619-38.

414   [28]  MARUMO T, YAGI S, KAWARAZAKI W, et al. Diabetes Induces Aberrant DNA Methylation in the Proximal

415   Tubules of the Kidney [J]. J Am Soc Nephrol, 2015, 26(10): 2388-97.

416   [29]  DONEPUDI A C, LEE Y, LEE J Y, et al. Multidrug resistance-associated protein 4 (Mrp4) is a novel genetic

417   factor in the pathogenesis of obesity and diabetes [J]. Faseb j, 2021, 35(2): e21304.

418   [30]  ZHOU H, YU B, SUN J, et al. Short-chain fatty acids can improve lipid and glucose metabolism independently

419   of the pig gut microbiota [J]. J Anim Sci Biotechnol, 2021, 12(1): 61.

420   [31]  ARYAL B, ROTLLAN N, ARALDI E, et al. ANGPTL4 deficiency in haematopoietic cells promotes monocyte

421   expansion and atherosclerosis progression [J]. Nat Commun, 2016, 7(12313.

422

423

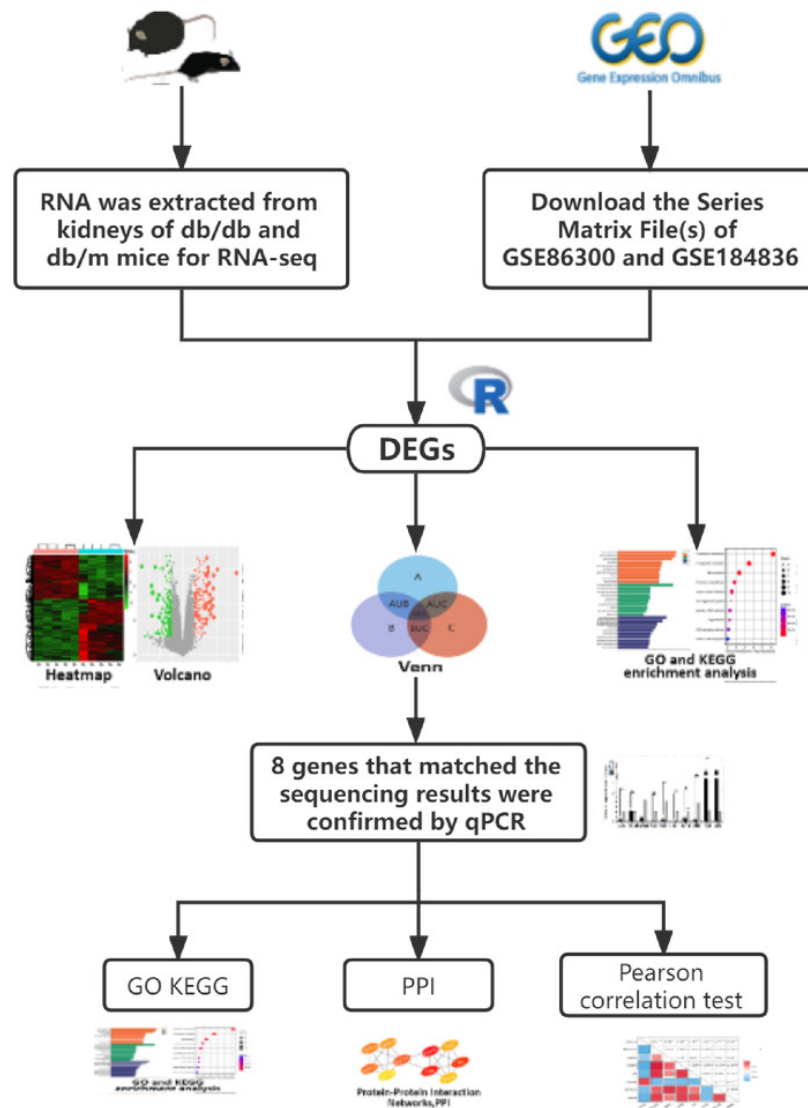
424

# Figure 1

figure 1

Figure 1. The workflow of this study

# Figure 2

figure 2

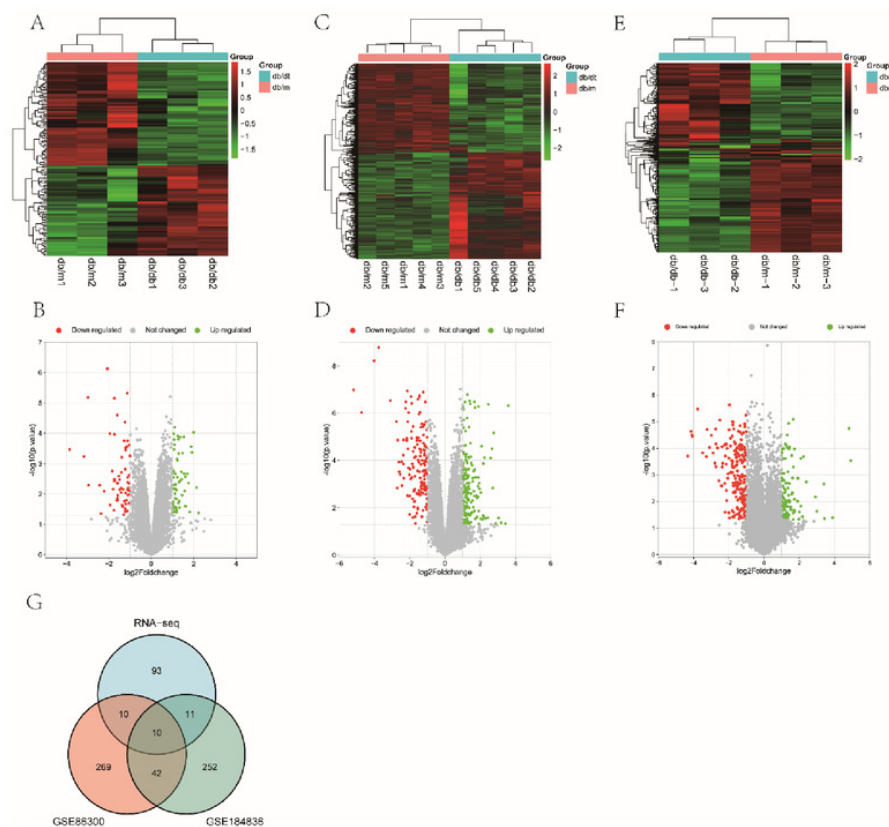



Figure 2 Differential expression analysis of our RNA-seq and two GEO datasets (GSE86300 and GSE184836). (a) Heatmap of DEGs in our RNA-seq. (b) Volcano map of our RNA-seq. A total of 58 upregulated and 66 downregulated DEGs were identified between the DKD and the Control group. (c) Heatmap of DEGs in GSE86300. (d) Volcano map of DEGs in GSE86300. A total of 181 up-regulated and 151 down-regulated DEGs were identified between the DKD and the Control group. (e) Heatmap of DEGs in GSE184836. (f) Volcano map of DEGs in GSE184836. A total of 107 up-regulated and 208 down-regulated DEGs were identified between the DKD and the control group. (g) Venn diagram of three DEGs groups. A total of 10 co-expression genes were obtained. In volcano map, green bubbles mean up-regulated genes, red bubbles mean down-regulated genes, and gray bubbles mean non-significant genes. In heatmap, red mean up-regulated genes, green mean down-regulated genes. ($|\log_2$ FC| > 1 and $p$-value < 0.05)

# Figure 3

figure 3

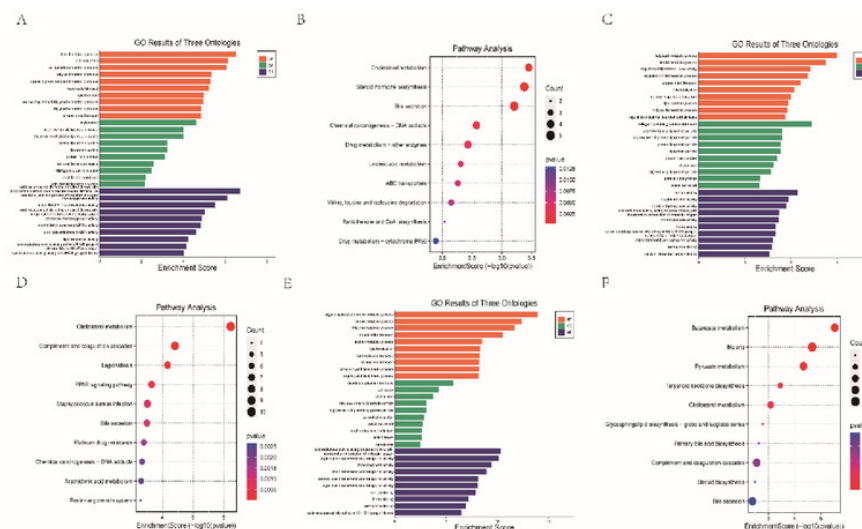



Figure 3. Bar graph of GO Annotation and dot plot of KEGG pathway enrichment analysis of DEGs. (a, b) Our RNA-seq (Top 10). (c, d) GSE86300 (Top 10). (e, f) GSE184836 (Top 10). In the dot plot, the color represents the $p$-value, and the size of the spots represents the gene number.

# Figure 4

figure 4

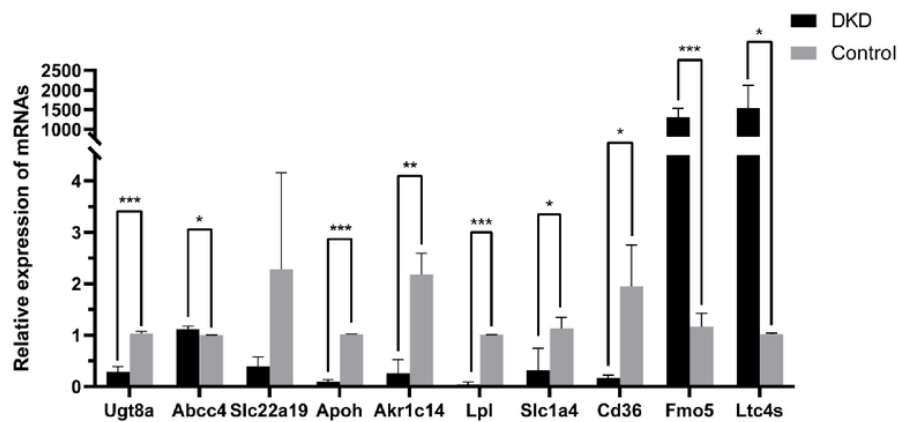



Figure 4. The relative mRNA expression of 10 overlapping genes in DKD and control group determined by qRT-PCR. $^{e}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

# Figure 5

figure 5

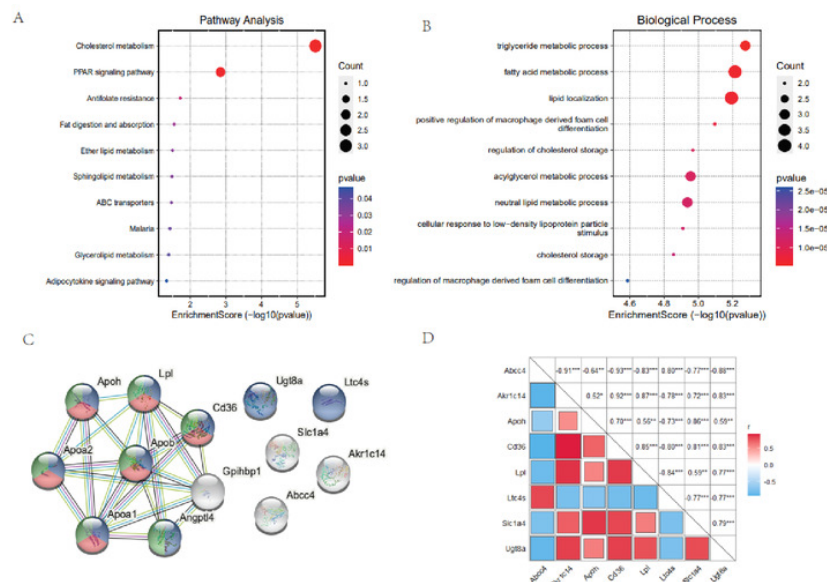



Figure 5 (a) Dot plot of KEGG pathway enrichment analysis of verified 8 genes, in the dot plot, the color represents the p-value, and the size of the spots represents the gene number. (b) Bar plot of GO enrichment analysis of verified 8 genes. (c) PPI network of verified 8 DEGs, which also showed that several other genes were involved. In the PPI analysis, green represents the cholesterol metabolic pathway, red represents the triglyceride metabolic process and blue represents the lipid metabolic process, white has no meaning. (d) Pearson correlation test between 8 DKD related genes, blue represents negative correlation, while red represents positive correlation.

table 1

1  **Table 1 Expression of 10 overlapping genes in three datasets**

| group | RNA-seq | | GSE86300 | | GSE184836 | | |
|---|---|---|---|---|---|---|---|
| Gene name | logFC | p.Value | logFC | p.Value | logFC | p.Value | up/down |
| Ltc4s | 1.99 | 0.0001 | 1.70 | 5.9E-07 | 1.12 | 0.033 | Up |
| Cd36 | -1.10 | 0.0002 | -2.11 | 2.3E-06 | -2.13 | 7.1E-05 | Down |
| Slc1a4 | -1.20 | 0.0004 | -1.28 | 1.0E-03 | -1.53 | 0.004 | Down |
| Slc22a19 | -3.18 | 0.0006 | -2.54 | 3.9E-04 | -1.64 | 0.004 | Down |
| Lpl | -1.05 | 0.0023 | -1.01 | 1.5E-03 | -1.34 | 0.009 | Down |
| Akr1c14 | -2.43 | 0.0052 | -1.86 | 1.4E-04 | -2.98 | 3.3E-05 | Down |
| Apoh | -1.42 | 0.0089 | -1.54 | 1.1E-02 | -1.09 | 0.0002 | Down |
| Fmo5 | -1.58 | 0.0111 | -2.36 | 1.4E-05 | -1.18 | 0.001 | Down |
| Abcc4 | 1.19 | 0.0149 | 1.42 | 4.0E-07 | 1.31 | 3.2E-05 | Up |
| Ugt8a | -1.22 | 0.0380 | -1.12 | 1.0E-04 | -1.13 | 0.029 | Down |

table 2

1 **Table 2 The BP analysis of 8 verified genes**

| ID | Description | *P*.value | Gene |
|---|---|---|---|
| GO:0006641 | triglyceride metabolic process | 5.31354E-06 | Cd36/Lpl/Apoh |
| GO:0006631 | fatty acid metabolic process | 6.10541E-06 | Ltc4s/Cd36/Lpl/Akr1c14 |
| GO:0010876 | lipid localization | 6.40694E-06 | Cd36/Lpl/Apoh/Abcc4 |
| GO:0010744 | positive regulation of macrophage derived foam cell differentiation | 8.01174E-06 | Cd36/Lpl |
| GO:0010885 | regulation of cholesterol storage | 1.07813E-05 | Cd36/Lpl |
| GO:0006639 | acylglycerol metabolic process | 1.10996E-05 | Cd36/Lpl/Apoh |
| GO:0006638 | neutral lipid metabolic process | 1.15891E-05 | Cd36/Lpl/Apoh |
| GO:0071404 | cellular response to low-density lipoprotein particle stimulus | 1.23194E-05 | Cd36/Lpl |
| GO:0010878 | cholesterol storage | 1.39596E-05 | Cd36/Lpl |