Bioinformatic analysis identifies potential key genes in the pathogenesis of uterine leiomyoma

Yi-Chao Jin1,†, Tong-Hui Ji1,†, Xiong Yuan1, Ying Sun1, Yu-Jie Sun2, Jie Wu1,‡

1 Department of Gynecology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210029 Jiangsu, P. R. China
2 Key laboratory of Human Functional Genomics of Jiangsu Province, Nanjing Medical University, Nanjing, 211126 Jiangsu, P. R. China
*Correspondence: wujiemd@126.com (Jie Wu)
† These authors contributed equally.

Methods: The microarray data of GSE31699, containing 16 uterine leiomyoma tissue samples and 16 matched normal myometrium samples, were downloaded from the Gene Expression Omnibus database (GEO). The “limma” R language package was used to identify differentially-expressed genes (DEGs) between uterine leiomyoma and myometrium. Gene Ontology (GO) and pathway enrichment analyses were performed by using clustering profiler, the DEGs were mostly enriched in post-synapse assembly. Furthermore, we employed Molecular Complex Detection (MCODE) in Cytoscape, then clustered extracellular matrix receptor interaction and coagulation cascades. Subsequently, a protein-protein interaction (PPI) network of DEGs was constructed by Search Tool for the Retrieval of Interacting Genes Database (STRING) and visualized by utilizing Cytoscape software. We screened hub clusters of PPI network by the plug-in WGCNA package in R to conduct co-expression network for all genes in hub clusters. WGCNA used to explore pivotal genes related to different diseases including uterine leiomyomas. For example, Liu and coworkers revealed the potential role of CASP1, ALDH1A1 and PROS1 in uterine leiomyoma diagnosis and treatment by using GEO database [7]. And the location of the uterine leiomyoma may have a different gene expression profile [8]. In this study, cases may present with abnormal uterine bleeding (heavy menstrual bleeding, prolonged menstruation, anemia), pelvic mass and compression symptoms (dysuria, constipation, hydronephrosis), pelvic pain, infertility and pregnancy complications [1]. Although the rate of malignant transformation is low, uterine fibroids are still a tremendous public health burden and the most frequent indication for hysterectomy [2].

Traditionally, it is universally acknowledged that uterine fibroids are clonal smooth muscle cell neoplasms, and gonadal steroids play an important role in their occurrence and development. Some uterine fibroids also have specific molecular genetic background and chromosomal rearrangement [3]. The latest research shows that at least four types of cellular components exist in uterine fibroids: uterine smooth muscle cells, vascular smooth muscle cells, fibroblasts, and fibroid-associated fibroblasts. All of these cells are derived from fibroid progenitor cells which are transformed from myometrial stem cells [4]. In addition, the excessive accumulated extracellular matrix (ECM) including collagens is also an important constituent of uterine fibroids [5]. Compared with normal myometrium, the amount and topological structure of extracellular matrix in uterine fibroids have changed significantly. Besides, the cellular components can secrete growth factors, cytokines, chemokines, proteases, angiogenic and inflammatory response mediators locally in an autocrine or paracrine manner. These molecules may account for the development of uterine fibroids by promoting cell proliferation, angiogenesis, inflammation and ECM accumulation [6]. However, the precise molecular mechanisms and cellular changes of uterine fibroids have not been comprehensively elucidated and effective treatment is limited by the poor understanding of their pathogenesis.

Currently, bioinformatic analysis has been extensively used to explore pivotal genes related to different diseases including uterine leiomyomas. For example, Liu and coworkers revealed the potential role of CASP1, ALDH1A1 and PROS1 in uterine leiomyoma diagnosis and treatment by using GEO database [7]. And the location of the uterine leiomyoma may have a different gene expression profile [8]. In this study,
Weighted gene expression network analysis (WGCNA) was carried out given that only concerning about the genes’ discrepant expression may neglect the potential association among genes. WGCNA is a systematic biological method for describing the correlation patterns among genes across microarray samples. It can be used for identifying modules of highly associated genes, candidate biomarkers or therapeutic targets according to associations of each modules and relationships between modules and external sample phenotypes [9]. In our study, we identified hub genes by both WGCNA analysis and DEG’s PPI network construction. These hub genes and related pathways may contribute to the pathology of uterine leiomyomas or may be potential therapeutic targets in the future.

2. Materials and methods

2.1 Ethics statement

This study was a bioinformatical research, and all procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national).

2.2 Microarray data and preprocessing

GEO (http://www.ncbi.nlm.nih.gov/geo/) is a public functional genomics database which provides array- and sequence-based data. Gene expression dataset GSE31699 was performed by Illumina HumanHT-12 V3.0 expression beadchip. We downloaded the dataset from GEO and converted the probes into the matching gene symbol in accordance with the annotation information in the platform. The GSE31699 dataset included 16 uterine leiomyoma tissue samples and 16 matched normal myometrium samples. According to the original study, the patients were all African American premenopausal women (mean age 44 years; range 30-50 years) without any hormonal treatment for at least 6 months prior to surgery, and the tumor sizes ranged from 5 to 17 cm in diameter. And there was no detailed information on the molecular genetic background of the uterine leiomyoma samples.

2.3 Study design & data processing

We designed the study on the basis of the flow diagram (Fig. 1). We evaluated the data quality by sample clustering based on the Pearson’s correlation matrices between different samples. In addition, a heatmap was drawn according to the different expressions of probes (Fig. S1). According to the clustering results and heatmap, we deleted the samples GSM786796 and GSM786789 to eliminate the potential bias in the subsequent analysis.

2.4 Differentially expressed genes

We employed the “limma” R language package to screen the DEGs between uterine leiomyoma and normal myometrium [10]. The adjusted P-value < 0.05 and |log2fold change (FC)| > 1 were considered statistically significant.

2.5 KEGG and GO enrichment analyses of DEGs

Gene Ontology (GO, http://geneontology.org/) provides an ontology of defined terms to represent gene functions (molecular function, cellular component and biological process). Besides, Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) is a database resource for understanding high-level functions and utilities of the biological system. A package in R language called “clusterprofiler” was used to determine the biological significance of DEGs. The package is capable of providing GO and KEGG enrichment analyses and visualization for users to obtain more valuable biological information [11]. P value < 0.05 was considered as a significant enrichment.

2.6 PPI network construction and clusters analyses

Search Tool for the Retrieval of Interacting Genes (http://string-db.org) (version 10.0) online database was utilized to predict PPI network of DEGs [12]. Then we converted the results into visualization by Cytoscape software [13]. Furthermore, we used the plug-in Molecular Complex Detection (MCODE) (version 1.4.2) in Cytoscape to select hub modules of PPI network. The degree cut-off = 2, node score cutoff = 0.2, k-core = 2, and Max depth = 100 were set as significant and the hub clusters were selected [14]. Subsequently, KEGG and GO analyses for genes in the hub clusters were performed by “clusterprofiler” package in R language.

2.7 Weighted co-expression network construction and module enrichment analyses

To investigate the differences between uterine leiomyoma and the matched adjacent normal myometrium, the co-expression network of GSE31699 was created and the module function was analyzed. Firstly, we tested the expression data profile and determined that the samples and genes are suitable. Secondly, the “WGCNA” package in the R language was employed to create the co-expression networks [15]. Thirdly, a power function amn = ((1+cmn)/2)^β (cmn = Pearson’s correlation between gene m and gene n; amn = adjacency between gene m and gene n; β = soft-thresholding power) was used. And we established a weighted adjacency matrix. A soft-thresholding power β was utilized to highlight strong correlations between genes and attenuate weak associations. Fourthly, we switched the adjacency to topological overlap matrix (TOM) and measured network connectivity of each gene. Average linkage hierarchical clustering was created according to the TOM-based dissimilarity, and we set the minimum size of the genes dendrogram as 30. Thereby, genes with strong correlations were grouped into the corresponding gene module. Then, we figured out the dissimilarity of module eigengenes. To determine associated modules that may be involved in the pathogenesis of uterine leiomyoma, GO and KEGG enrichment analyses were employed on these gene modules.

2.8 Hub genes validation

Oncomine database (www.oncomine.org) is an online cancer microarray database for DNA or RNA sequence analysis, which was employed to analyze the transcription levels of selected hub genes in different cancers. The mRNA expressions of the genes in cancers including uterine leiomyoma
Fig. 1. Study design and the flow diagram of study.

were compared with those in normal controls by student’s t test. Cut-off of P value and fold change were as following: P value: 0.01, fold change: 1.5, gene rank: 10%, data type: mRNA. In addition, other GEO datasets were also searched to validate the expression of the real hub genes.

2.9 Gene set enrichment analysis

To explore the potential functions of selected hub genes in uterine leiomyoma, samples of uterine leiomyoma dataset GSE31699 were divided into different groups in accordance with the expression levels of the 5 hub genes respectively. GSEA was utilized to explore whether priority determined biological processes datasets were enriched in these groups [16, 17]. The criteria were set as P value < 0.05 and FDR < 0.25.

3. Result

3.1 Identification of DEGs in uterine leiomyoma

We identified DEGs in GSE31699 by using limma package in R language after deleting the samples GSM786796 and GSM786789 as described previously. The DEGs were displayed in a volcano plot based on |log2FC| (Fig. S2). Then, we draw a heatmap of these DEGs which included 125 up-regulated genes and 148 down-regulated genes in leiomyoma tissues by comparison with normal myometrium tissues (Fig. 2).

3.2 KEGG and GO enrichment analyses of DEGs

The clusterprofiler package was utilized to analyze the biological processes of DEGs. The cutoff of P value was set as < 0.05. In GO analysis, the up-regulated DEGs were mainly enriched in postsynapse assembly, postsynaptic density organization, extracellular structure organization, postsynaptic specialization organization, skeletal system development and regulation of canonical Wnt signaling pathway (Fig. 3A). The down-regulated DEGs were mostly enriched in response to glucocorticoid, toxic substance and corticosteroid, mononuclear cell migration, monocyte chemotaxis and negative regulation of cell migration (Fig. 3B). In the KEGG analysis, the up-regulated DEGs were highly enriched in ECM-receptor interaction, cell adhesion molecules (CAMs) and protein digestion and absorption (Fig. 3C), while the down-regulated ones were significantly enriched in complement and coagulation cascades, staphylococcus aureus infection, fluid shear stress and atherosclerosis, TNF signaling pathway, arachidonic acid metabolism and parathyroid hormone synthesis, secretion and action (Fig. 3D).

3.3 PPI network construction and clusters analyses

The PPI network was constructed via STRING, which included 267 nodes and 555 edges as shown in Fig. 4A. Then, we applied the plug-in MCODE of Cytoscape to find the most significant clusters in the network. With the criteria described above, 7 clusters were identified. The top 4 clusters were used for further analysis with k-core > 3. Cluster1 contains 24 nodes and 69 edges (Fig. 4B); Cluster2 contains 7 nodes and 13 edges (Fig. 4C); Cluster3 contains 4 nodes and 6 edges (Fig. 4D); Cluster4 contains 14 nodes and 24 edges (Fig. 4E). We performed GO analysis and KEGG analysis of these clusters by “clusterprofiler” respectively. Genes in cluster1 were mostly enriched in response to glucocorticoid (Fig. S3A) and Parathyroid hormone synthesis, secretion and action (Fig. S4A). The genes in cluster2 were mainly enriched in platelet degranulation (Fig. S3B) and adipocytokine signaling pathway (Fig. S4B). The genes in cluster3
were significantly enriched in postsynaptic density assembly (Fig. S3C) and nicotine addiction (Fig. S4C). The genes in cluster4 were largely enriched in extracellular matrix organization (Fig. S3D) and protein digestion and absorption (Fig. S4D).

3.4 Weighted co-expression network construction and module enrichment analyses

The “WGCNA” package was conducted in R language. Soft-thresholding power $\beta = 30$ (scale free $R^2 = 0.96$) was chosen to ensure a scale-free network (Fig. S5). A heatmap of module-trait relationships was constructed and 4 modules (blue, turquoise, purple, and tan modules) were extracted according to the $P$ values (Fig. S5A,C) and the cut-off of $P$
Fig. 3. KEGG and GO enrichment of DEGs. (A) GO enrichment of the up-regulated DEGs. (B) GO enrichment of the down-regulated DEGs. (C) KEGG analysis of the up-regulated DEGs. (D) KEGG analysis of the down-regulated DEGs.

value was < 0.01. Among these modules, blue module has the strongest correlation with uterine leiomyoma. Genes except those in grey module were identified for a heatmap (Fig. 5B).

To explore the connections and interactions among different modules, we completed a cluster analysis and calculated the connectivity of eigengenes. Results showed that 16 modules were divided into two major clusters. Intriguingly, blue, tan, purple, and cyan modules converged into one cluster (cluster1), while turquoise module was in another (cluster2) (Fig. 6A). In addition, there was a positive correlation among four modules in cluster1. However, these four modules were negatively correlated with turquoise module (Fig. 6B). Moreover, we analyzed GS (gene significance, which shows the correlation between the gene and the trait, uterine leiomyoma or normal myometrium) and MM (module membership, which reflects the correlation of the module eigengene and the gene expression profile) of the genes in the 5 modules respectively. The results demonstrated that 4 out of these 5 modules were statistically significant as shown in Fig. 7: blue module, cor = -0.79, P = 9.1 \times 10^{-67} (Fig. 7A);
Fig. 4. PPI network construction and clusters analyses. The red nodes represent the up-regulated genes and the blue nodes represent the down-regulated genes. (A) The PPI network of 273 DEGs was constructed via STRING that contained 267 nodes and 555 edges. (B) Cluster rank 1. This cluster consists of 24 nodes and 69 edges and has the highest score in those clusters. (C) Cluster rank 2. (D) Cluster rank 3. (E) Cluster rank 4.
Fig. 5. Co-expression network creation and hub modules selection. (A) Dendrogram of all genes in GSE31699 clustered based on a dissimilarity measure (1-TOM). (B) A heatmap of selected genes. The intensity of the yellow color indicates the strength of the correlation between pairs of modules on a linear scale. (C) Correlation between modules and traits. The upper number in each cell refers to the correlation coefficient of each module in the trait, and the lower number is the corresponding \( P \)-value. Among them, the blue module was the most relevant modules with uterine leiomyoma traits.
turquoise module, cor = 0.81, $P = 6.4 \times 10^{-161}$ (Fig. 7B);
tan module, cor = -0.52, $P = 0.00025$ (Fig. 7C); cyan module,
cor = -0.34, $P = 0.042$ (Fig. 7D). As the criteria was set as $P < 0.01$, we selected blue, turquoise and tan modules as hub modules.

Functional enrichment analyses of these modules were performed aiming to identify the underlying biological function and pathways those were related to uterine leiomyoma. After GO analysis, we discovered that genes in blue module were mainly enriched in DNA-dependent DNA replication, DNA replication initiation, planar cell polarity pathway involved in neural tube closure, G1/S transition of mitotic cell cycle and ATP-dependent chromatin remodeling (Fig. S6A). The genes in turquoise module were significantly enriched in regulation of vasculature development, regulation of inflammatory response, regulation of angiogenesis, regulation of cell-cell adhesion, cell-substrate adhesion, glomerulus development and response to peptide (Fig. S6C). The genes in tan module were mainly enriched in DNA-dependent DNA replication, DNA replication initiation, planar cell polarity pathway involved in neural tube closure, G1/S transition of mitotic cell cycle and ATP-dependent chromatin remodeling (Fig. S6A). The genes in blue module were mainly enriched in DNA-dependent DNA replication, DNA replication initiation, planar cell polarity pathway involved in neural tube closure, G1/S transition of mitotic cell cycle and ATP-dependent chromatin remodeling (Fig. S6A). The genes in turquoise module were significantly enriched in regulation of vasculature development, regulation of inflammatory response, regulation of angiogenesis, regulation of cell-cell adhesion, cell-substrate adhesion, glomerulus development and response to peptide (Fig. S6C). The genes in tan module were mainly enriched in DNA-dependent DNA replication, DNA replication initiation, planar cell polarity pathway involved in neural tube closure, G1/S transition of mitotic cell cycle and ATP-dependent chromatin remodeling (Fig. S6A). The genes in blue module were mainly enriched in DNA-dependent DNA replication, DNA replication initiation, planar cell polarity pathway involved in neural tube closure, G1/S transition of mitotic cell cycle and ATP-dependent chromatin remodeling (Fig. S6A). The genes in turquoise module were significantly enriched in regulation of vasculature development, regulation of inflammatory response, regulation of angiogenesis, regulation of cell-cell adhesion, cell-substrate adhesion, glomerulus development and response to peptide (Fig. S6C).
module were mostly enriched in extracellular matrix organization, cellular response to amino acid stimulus, protein heterotrimerization, extracellular structure organization, collagen fibril organization, response to amino acid, endodermal cell differentiation and collagen metabolic process (Fig. S6E).

Through the analysis of KEGG, we observed that genes in blue module were enriched in DNA replication, base excision repair, mismatch repair, cell cycle and nucleotide excision repair (Fig. S6B). Genes in turquoise module were enriched in TNF signaling pathway, Malaria, chemokine signaling pathway, fluid shear stress and atherosclerosis, viral protein interaction with cytokine and cytokine receptor, complement and coagulation cascades, Rap1 signaling pathway and leukocyte trans-endothelial migration (Fig. S6D). The genes in tan module were enriched in ECM-receptor interaction, AGERAGE signaling pathway in diabetic complications, amoebiasis, focal adhesion, relaxin signaling pathway and PI3K-Akt signaling pathway (Fig. S6F).

3.5 PPI network construction of hub modules and identification of hub genes

We built a PPI network on the basis of STRING database for genes in the 3 hub modules by Cytoscape software independently. The blue module included 220 nodes and 647 edges (Fig. 8A). The turquoise module contained 592 nodes and 2710 edges (Fig. 8B). The tan module was compromised of 29 nodes and 78 edges (Fig. 8C).

In WGCNA, hub genes were highly connected in a module which may exert important biological functions. So, sorted by intramodular connectivity, the top 10% genes in each hub modules were selected as hub genes (30 out of 307
genes in blue module; 68 out of 687 genes in turquoise module; 4 out of 45 genes in tan module). According to the PPI network of DEGs constructed previously, we took 49 genes in 4 hub clusters as hub genes of DEGs’ PPI network. Eventually, 8 hub genes were identified both in DEGs’ PPI network and co-expression network. These 8 hub genes were regarded as “real” hub genes (Fig. 8D).

3.6 Hub genes validation

In the dataset GSE31699, 3 out of the 8 hub genes were up-regulated while the other 5 genes were down-regulated in uterine leiomyoma compared to normal myometrium. We validated the expression of the selected hub genes in other datasets from GEO: GSE64763 [18], GSE764 [19] and GSE593 [20]. Eventually, we demonstrated that the expressions of COL5A2, ALDH1A1, GNG11, EFEMP1 and ANXA1 were also significantly different in these datasets in accordance with GSE31699 (Fig. 9). The mRNA expressions of these genes were analyzed by Oncomine database as shown in Fig. 10.

3.7 Gene set enrichment analysis

In order to confirm the potential enriched KEGG pathways of the 5 hub genes in uterine leiomyoma, we employed Gene set enrichment analysis of the five hub genes in GSE31699. Tow pathways of “base excision repair” and “complement and coagulation cascades” were mostly enriched (P < 0.05 and FDR < 0.25) (Fig. 11).

4. Discussion

In our study, we identified 273 DEGs related to uterine leiomyoma. GO analysis revealed that the upregulated DEGs were mainly enriched in extracellular structure organization, while the down-regulated DEGs were highly enriched in response to glucocorticoid and mononuclear cell migration. KEGG analysis suggested the upregulated genes were mostly enriched in ECM-receptor interaction. The down-regulated genes were mostly enriched in complement and coagulation cascades. The PPI network was constructed based on DEGs by STRING website and Cytoscape software. We identified 4 clusters when taking k-core > 3 as a criterion.
Fig. 9. The relative expression of hub genes in other datasets from GEO (GSE64763, GSE764 and GSE593, *, $P < 0.01$, **: $P < 0.001$, ***: $P < 0.0001$, ****: $P < 0.00001$).
In WGCNA analysis, we selected 4 hub modules. The turquoise module had the highest correlation which contains 687 genes. GO analysis showed that the genes were mostly enriched in regulation of vasculature development. In KEGG analysis, they were significantly enriched in TNF signaling pathway. We selected 102 genes, which were the top 10% genes according to the intramodular connectivity in each hub module. In view of the PPI network, 8 genes were screened as the hub genes. 5 out of these genes (COL5A2, ALDH1A1, GNG11, EFEMP1 and ANXA1) were validated by other GEO datasets and Oncomine database. The results indicated that the 5 hub genes may be implicated in the pathology of uterine leiomyoma.

17 whose mutations in this gene are associated with Ehlers-Danlos syndrome. In this study, we found that the expression of COL5A2 in uterine leiomyoma was high in com-
Fig. 11. Gene set enrichment analysis (GSEA) using GSE31699. The most enriched functional gene set in uterine leiomyoma samples with hub genes highly expressed was identified. (A) COL5A2. (B) ALDH1A1. (C) GNG11. (D) EFEMP1. (E) ANXA1.

Comparison to normal myometrium. Accumulation of extracellular matrix is a specific cause of uterine leiomyoma, and collagen is a key structure component of ECM. Many sub-types of collagen were proved to have higher levels such as COL1A1, 4A2, 6A1, 6A2, 7A1 and 16A1 [21], while COL5A2 has not been validated. The up-regulated DEGs of GSE31699...
and tan module including COL5A2 in WGCNA analysis were significantly enriched in extracellular structure organization. Besides, in our GSEA analysis, the functions of COL5A2 were mostly enriched in extra matrix structure constituent and collagen fibril organization. So, we hypothesize that COL5A2 may also be implicated in the pathological process of uterine leiomyoma. In addition, Giri and colleagues found that the region in chromosome 2q31-32 including COL5A2 may mediate interaction between local ancestry and BMI, which were risk factors for uterine leiomyoma [22]. Oncomine database demonstrated overexpression of COL5A2 in several cancers which indicated its potential role in other cancers. Furthermore, some studies showed that COL5A2 was correlated with poor clinical outcomes of bladder and gastric cancer patients while the detailed mechanisms still await clarification [23, 24].

The protein encoded by ALDH1A1 belongs to the aldehyde dehydrogenase family, which is commonly known as a major participant in alcohol metabolism. There have been controversies about the effect of ALDH1A1 on uterine leiomyoma. Zaitseva and coworkers found that αSMA-negative fibroblast-like cells of the connective tissue strongly expressed ALDH1A1 in myometrium when compared with uterine leiomyoma [25], which was consist with the research of Xia and colleagues [26]. ALDH1A1 expression was regulated differently in myometrial and fibroid cells by retinol metabolism pathway which may be important in fibroid pathophysiology [27]. In another study, Shveiky and co-investigators demonstrated an increased level of the ALDH1A1 protein in uterine leiomyoma, while acetaldehyde could inhibit the growth of uterine leiomyoma cells [28]. In our study, the genes positively correlated with ALDH1A1 in GSEA31699 were enriched in fibrinolysis, a process that solubilizes fibrin in the bloodstream. And in WGCNA analysis, the genes in turquoise module containing ALDH1A1 were significantly enriched in regulation of vasculature development. So, we make an assumption that ALDH1A1 may play a role in improving the blood perfusion of uterine leiomyoma. More detailed mechanisms should be explored to validate our hypothesis.

GNG11 is a member of the guanine nucleotide-binding protein (G protein) gamma family and encodes a lipid-anchored, cell membrane protein, which plays a role in transmembrane signaling system [29]. Our results demonstrated GNG11 was down-regulated in uterine leiomyoma. Oncomine database showed that GNG11 had lower level in malignant tumors, especially breast cancer and lung cancer, when compared with normal tissues. Studies have pinpointed that GNG11 may be a biomarker for lung adenocarcinoma and acute myeloid leukemia subtype classification, and lower expression of GNG11 in lung cancer indicated poor prognosis [30, 31]. Yet the knowledge of mechanisms how GNG11 exerts biological functions is still limited. Previous studies showed that GNG11 could suppress cell growth by inducing reactive oxygen species and regulate cellular senescence by activating ERK1/2 of the MAP kinase family [32, 33].

GSEA analysis in this study also showed that the genes positively correlated with GNG11 mostly enriched in regulation of necrotic cell death, which indicated that lower expression of GNG11 in uterine leiomyoma may contribute to cell proliferation.

EFEMP1 (EGF containing fibuln extracellular matrix protein 1) contains tandemly repeated epidermal growth factor-like repeats followed by a C-terminus fibulin-type domain. EFEMP1 encoded the protein Fibulin-3, which may act as an antagonist of angiogenesis or may promote tumor growth. We found that EFEMP1 was down-regulated in uterine leiomyoma based on the datasets. Marsh et al. confirmed the significant down-regulation in leiomyoma both in vivo and in vitro, which was consistent with other studies on EFEMP1 expression in other solid tumors [34]. In this study, we utilized Oncomine database and discovered that EFEMP1 was down regulated in several kinds of solid cancers. Hu et al. demonstrated that EFEMP1 had the capacity to suppressed the growth of hepatocellular carcinoma (HCC) cells, while its decreased expression was related to the extent of spread to the lymph nodes and the prognosis of patients with HCC [35]. Further study suggested that lower level of EFEMP1 expression may be caused by the hypermethylation in the promoter region, while additional researches are warranted to explore the role of EFEMP1 in uterine leiomyoma pathology.

ANXA1 encodes a membrane-localized protein. This protein binds phospholipids and inhibits phospholipase A2 to serve a role in anti-inflammatory. According to Oncomine database, loss of function or expression of this gene has been detected in multiple tumors including uterine leiomyoma, while overexpression was also found in many cancers of the gut [36, 37]. ANXA1 may play protective antitumorigenic roles or promote metastasis and invasion depending on the type of cancer and cancer grade [38]. Anjum and coworkers used Next-generation RNA Sequencing to discover that ANXA1 has lower level in uterine leiomyoma than in normal tissues, being confirmed by Real-Time PCR further [39]. ANXA1 is involved in several cellular signal transduction pathways, especially those related to inflammation, cell differentiation, proliferation and migration [40]. In the present study, the GSEA analysis revealed that the genes positively correlated with ANXA1 significantly enriched in vasculogenesis and blood vessel endothelial cell migration, indicating that ANXA1 may be involved in the angiogenesis during the development of uterine leiomyoma. However, more detailed mechanisms of ANXA1 for the growth of uterine fibroids still await elucidation.

There were some limitations in our study. Firstly, all the data analyzed in our study was obtained from the online databases, further studies including larger quantity of clinical specimens are required to validate our findings. Secondly, as uterine leiomyomas are commonly benign tumors, there were not enough clinical data in GEO or TCGA database, so survival analysis could not be conducted in this study. The
clinical importance of these genes should be assessed in the future. Finally, we did not clarify the roles of the hub genes in uterine leiomyomas. Future basic researches are worth being undertaken to elucidate the underlying mechanisms.

In conclusion, by using the analysis of bioinformatics, we found that COL5A2, ALDH1A1, GNG11, EFEMP1 and ANXA1 might be the pivotal genes for the pathogenesis of uterine leiomyomas by participating in the pathways of ECM accumulation, vasculature development, angiogenesis and cell proliferation. These genes may be related to the progression of uterine leiomyomas or be the therapeutic targets. The present findings probably provide us with a new perspective on diagnosis and therapy of uterine leiomyomas. However, the screened genes and pathways still need to be confirmed.

Author contributions
Jie Wu and Yu-Jie Sun conceived and designed the study; Yi-Chao Jin, Tong-Hui Ji, Xiong Yuan and Ying Sun performed the data analysis, Yi-Chao Jin and Tong-Hui Ji wrote the paper.

Acknowledgment
We would like to express our gratitude to all those who gave us suggestions and instructions during the writing of this manuscript. And thanks all the peer reviewers and editors for their opinions and suggestions.

Conflict of interest
The authors declare no competing interests.

Supplementary material
Supplementary material associated with this article can be found, in the online version, at https://ejgo.imrpress.com/EN/10.31083/j.ejgo.2021.01.2151.

References


