Bioinformatics analysis of inflammation and oncology in pulmonary lymphangioleiomyomatosis

Wanchun Su, PhD[a], Zimin Zhao, PhD[a], Xin Liu, PhD[a], Jianfeng Xin, PhD[a], Song Xia, MD[a], Wenbin Shen, MD[a],*

Abstract

This study investigates the molecular markers and biological pathways of pulmonary lymphangioleiomyomatosis. We analyzed 2 gene expression profiles in the gene expression omnibus Gene Expression Omnibus database for normal lung tissue and lymphangioleiomyomatosis and identified differential expressed genes in pulmonary lymphangioleiomyomatosis. Ninety-one differentially expressed genes were identified, including 36 upregulated genes and 55 downregulated genes. Hub genes and pathogenic pathways associated with disease development were subsequently identified by enrichment analysis and protein–protein interaction network. Analysis showed that differential expressed genes are mainly involved in the biological behavior of tumor cell proliferation and invasion as well as the inflammatory response. We have identified 10 hub genes in the protein–protein interaction network. Hub genes play an important role in the proliferation and inflammatory response involved in tumor cell proliferation. This study deepens the understanding of lymphangioleiomyomatosis disease and provides a biological basis for further clinical diagnosis and treatment.

Abbreviations: DEG = differential expressed gene, GEO = Gene Expression Omnibus, LAM = lymphangioleiomyomatosis, PLAM = pulmonary lymphangioleimyomatosis.

Keywords: bioinformatics analysis, lymphangioleiomyomatosis, pulmonary lymphangioleimyomatosis

1. Background

Lymphangioleiomyomatosis (LAM) is a rare disease that occurs in women of childbearing age and can involve the lungs, mediastinum, and retroperitoneal lymphatics.[1] It is characterized histologically by abnormal proliferation of smooth muscle-like cells that can spread through the lymphatic system, with the lungs being the most commonly involved organ.[2] Pulmonary LAM (PLAM) was once thought to be an interstitial lung disease, but later studies have described PLAM as being intermediate between benign and malignant disease.[3] It is a mutation and loss of heterozygosity in the TSC1 gene or TSC2 gene caused by abnormal smooth muscle-like LAM cell proliferation.[4] The diagnosis and treatment of PLAM is a worldwide challenge. In this study, we first selected GES12027 and GSE16944 from the Gene Expression Omnibus (GEO) database. We applied R software for data analysis and obtained the common differential expressed genes (DEGs) of GSE12027 and GSE160618. Molecular function (MF), cellular composition (CC), biological processes (BP), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were analyzed using R software. We then built the protein–protein interaction (PPI) network and performed some additional analysis with Cytoscape. Our aim is to initially search for pathogenic mechanisms and pathways of PLAM at the genetic level to inform future diagnosis and treatment.

2. Methods

2.1. Microarray data information

National Center for Biotechnology Information-GEO is a free public microarray/gene mapping database. The search strategy ((“lymphangioleiomyomatosis”[MeSH Terms] OR Lymphangioleiomyomatosis[All Fields]) AND “Expression profiling by array”[Filter]) was obtained.

Inclusion criteria are: the sample source was Homo sapiens; and normal tissue and LAM tissue were controls.

We obtained the gene expression profiles of GSE12027 and GSE160618 in LAM and normal lung tissue from the GEO database. Microarray data of GSE12027 were on account of GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array), Microarray data of GSE160618 were on account of GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). Microarray data of GSE12027 contain 14 female LAM nodules and 7 pulmonary artery smooth muscles (CC-2581,

*Correspondence: Wenbin Shen, Department of Lymphatic Surgery, Capital Medical University Affiliated Beijing Shijitan Hospital, No.10, Tieyi St, Beijing 100010, China (e-mail: shenwb@bjsjth.cn).

Lonza). Microarray data of GSE160618 contain 6 LAM samples and 6 normal samples. Data table header descriptions and series matrix files of GSE12027 and GSE160618 were downloaded.

2.2. Data processing of DEGs

Gene annotation was performed using clusterProfiler of R software program, and the data set was normalized by quantile logarithm[5] (Fig. 1).

The limma V3.50.0 (linear models for microarray data) package of the R software program (version 4.1.2) was used to screen the samples for DEGs with a threshold of |log FC| >1, $P < .05$.[6]

The ggplot2 package of the R software program (version 3.5.0) was used to produce volcano plots of the DEGs. The pheatmap package of the R software program (version 3.5.0) was used to produce heat maps of the DEGs and build a Venn diagram.[7]

2.3. Gene Ontology and pathway enrichment analysis

Gene Ontology (GO) is an international standard classification system for gene function, which aims to establish a linguistic vocabulary standard for qualifying and describing gene and protein function that is applicable to all species and can be updated as research progresses. GO is divided into 3 parts: MF, BP, and cellular component (CC).[8]

KEGG is a database on genome decipherment.[9]

In this study, GO terms and KEGG pathway enrichment analysis of DEGs were automatically completed and visualized by the clusterProiler digest, and the Goplot package in the R software statistical analysis platform (significant as $P < .05$ and a q-value < 0.05) and the CluePedia plugin and ClueGO plug-in in Cytoscape software (version 3.7.1, http://www.cytoscape.org/) (with a kappa score ≥0.4).[10–13]

2.4. Gene Set Enrichment Analysis

We used GSESA software for further enrichment analysis of all differential genes. Gene sets were considered significantly enriched with alpha or P values <5% and false discovery rate <25%.[14]

2.5. Construction of PPI network and identification of hub genes

STRING is an online tool to identify and predict interactions between genes or proteins. We use STRING to construct PPI networks of DEGs. Visualization of DEG PPI network using Cytoscape software. We use biNGO plug-in and cytoHubba plug-in for analysis.[15]
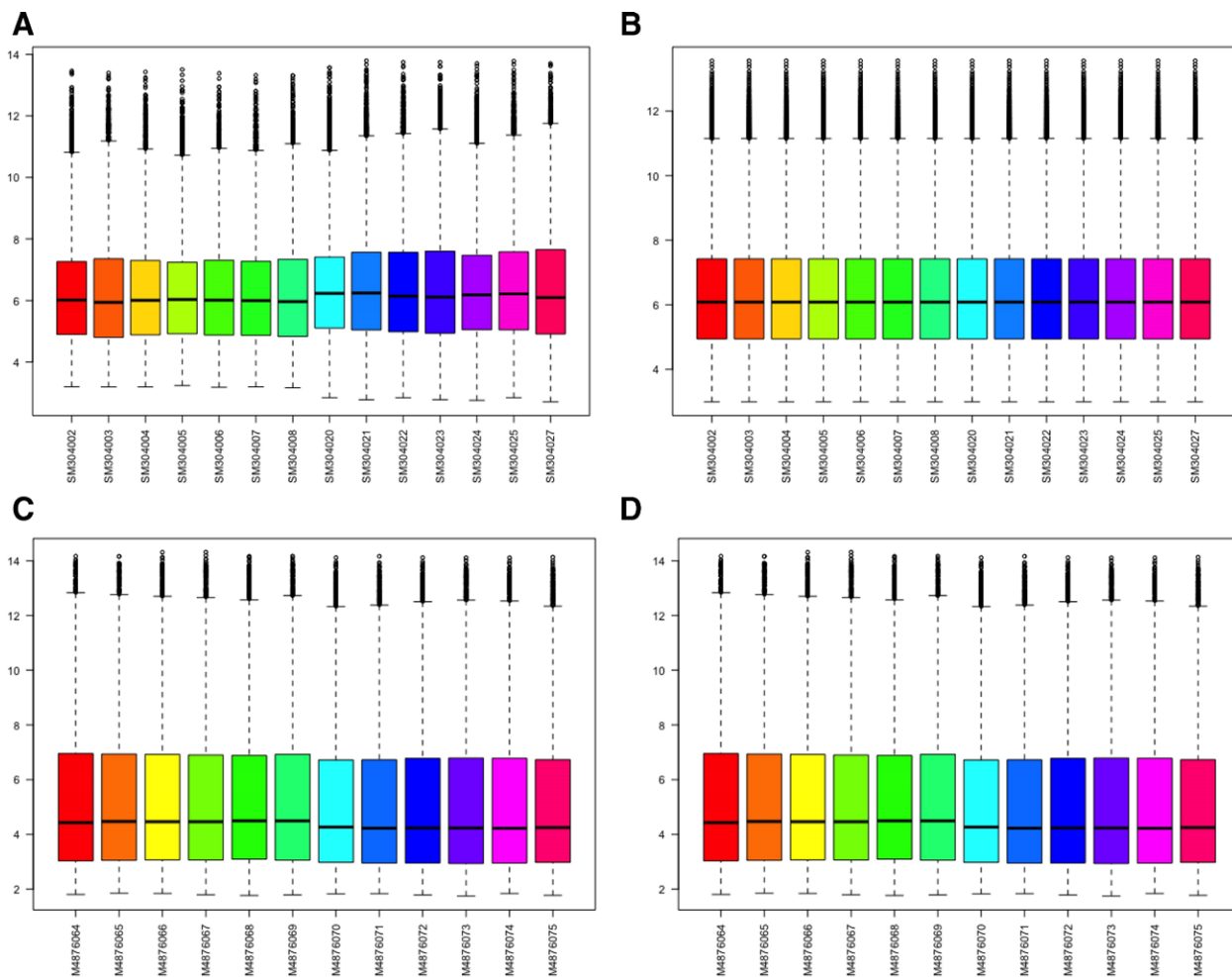

Figure 1. Box plots of gene expression data before and after standardization. (A) unstandardized for GSE12027, (B) standardized for GSE12027, (C) unstandardized for GS E160618, and (D) standardized for GSE160618. x-axis labels represent sample symbols and y-axis labels represent gene expression values. The black line in the box plot represents the median gene expression value.

2.6. DO analysis

Similar to GO database, Disease Ontolog (DO) provides a unified and standardized disease classification system by referring to MeSH, ICD, and other disease classification standards for common human diseases and rare diseases. The DOSE package of the R software program (version 3.5.0) was used to produce plots of the hub genes.

3. Result

3.1. Identification of DEGs

To identify DEGs in LAM and normal tissue samples, we screened 2334 and 496 DEGs from GSE12027 and GSE160618, respectively. After removing the null genes as well as mitochondrial genes, there were 1138 upregulated genes and 1191 downregulated genes in the DEGs of GSE12027. We had 207 upregulated genes and 263 downregulated genes in GSE160168 DEGs. We mapped the differential genes to the volcano map and heat map. There were 36 common upregulated genes and 55 common downregulated genes in both samples. We built the Venn graph with Package (VennDiagram) of R software (version 4.1.2; Figs. 2 and 3; Table 1).

3.2. GO and KEGG pathway enrichment analyses

We used Package (clusterProiler, Digest, and Goplot) of the R software (version 4.1.2) to perform KEGG and GO analysis on 36 upregulated genes and 55 downregulated genes. GO analysis showed that the significant enrichment in BP was epithelial cell

proliferation, negative regulation of cellular component movement, muscle system process, negative regulation of cell motility, regulation of body fluid levels, negative regulation of locomotion, muscle tissue development, response to peptide hormone, negative regulation of cell migration, regulation of reproductive process. In CC, differential genes were enriched in collagen-containing extracellular matrix, cell–cell junction, focal adhesion, cell–substrate junction, platelet alpha granule, perikaryon, platelet alpha granule lumen, rough endoplasmic reticulum, basement membrane, and T-tubule. For MF, genes were significantly enriched in receptor ligand activity, signaling receptor activator activity, glycosaminoglycan binding, sulfur compound binding, G protein-coupled peptide receptor activity, peptide receptor activity, heparin binding, fibronectin binding, peptide hormone binding, and hormone receptor binding (Fig. 4).

In the KEGG analysis, the differential genes were mainly enriched in hypertrophic cardiomyopathy, Ras signaling pathway, arrhythmogenic right ventricular cardiomyopathy, melanoma, prolactin signaling pathway, Focal adhesion, choline metabolism in cancer, calcium signaling pathway, type II diabetes mellitus, MAPK signaling pathway, cAMP signaling pathway, pathways in cancer, dilated cardiomyopathy, cGMP-PKG signaling pathway, and cytokine–cytokine receptor interaction (Fig. 5).

3.3. Gene Set Enrichment Analysis

We further subjected the LAM samples to Gene Set Enrichment Analysis (GSEA), and we found that these genes were mainly enriched in leukocyte endothelial cell migration, vascular
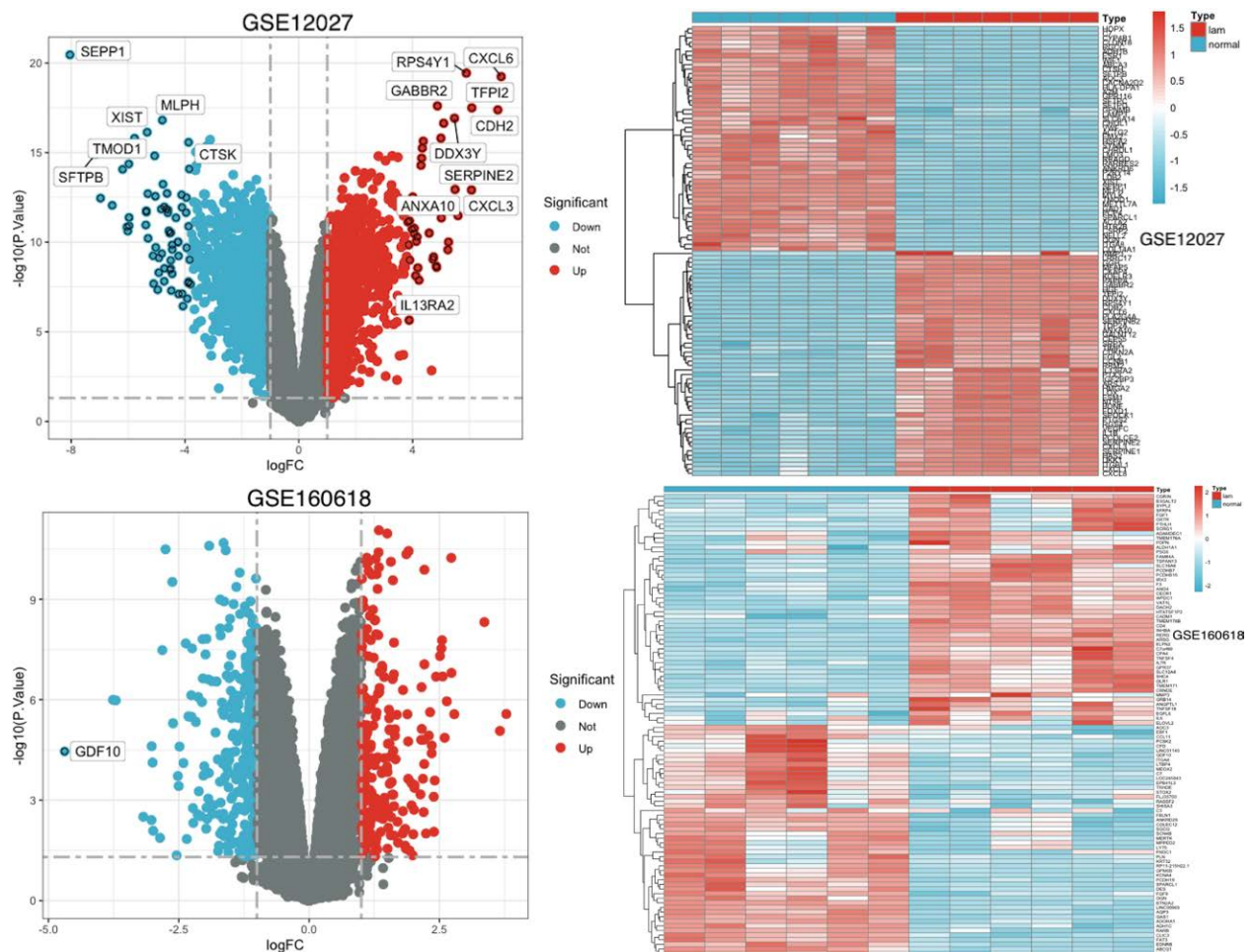

Figure 2. Volcano plots of differential expressed genes. Red data points represent upregulated genes and blue data points represent downregulated genes. Differences are set to | log FC | >1; heat map of differential expressed genes identified in (A) GSE12027, (B) GSE160618.
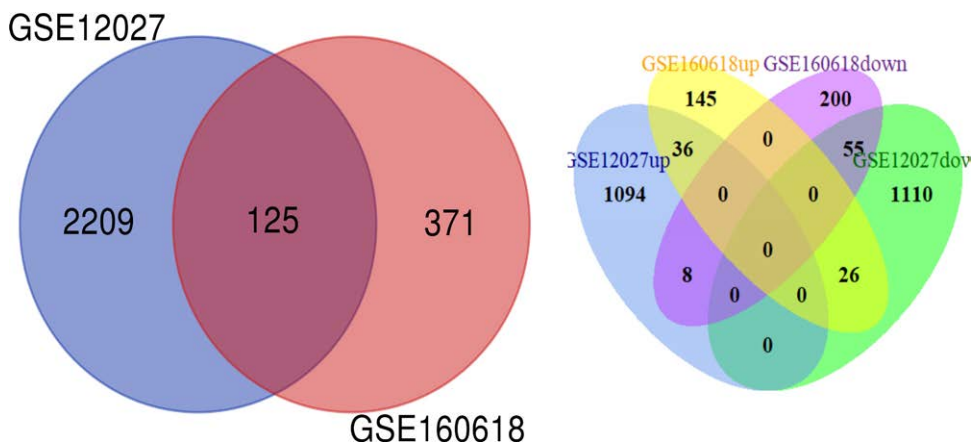
Figure 3. Venn diagrams of commonly differential expressed genes in the 2 datasets.

Table 1

Differential expressed genes of GSE12027 and GSE160618.

| DEGs	Gene name
Both upregulate	MFAP5 HGF PTHLH AADAC BMP2 SULF1 GPR1 FST HEG1 LOX SERPINE1 NFE2L3 IGFBP5 SH2D4A GRB14 DPP4 SLC31A1 ZNF668 CHAC1 LRP8 ANXA3 ZWINT MCM10 SFRP4 TSPAN8 RAB3B FLI1 PLAU OXTR IGFBP3 TNFSF4 SLC12A8 SCG2 CPA4 MTHFD2 MEST
Both downregulate	SPARCL1 AOC3 AQP1 EDNRB SETBP1 OGN C7 TCF21 ZNF204P ITGA8 RASSF2 TNFSF10 PPL TLE2 BZRAP1 MERTK MPPED2 THNSL2 RBM47 IST1 EPCAM SPON1 PPP1R9A APOC1 GPNMB RBM6 ZNF395 TGFBR3 CCL11 CXADR TCEAL2 CFD PLN RAMP1 PDE1A IGF1 RNF128 SCAF4 SERPINF1 PRELP GHR RBM5 SMYD2 CXCL12 MCOLN3 SAT1 FBLN1 FGF9 PPFIA4 SGCG BTN2A2 KIAA0226L ADRB2 NDNF EPB41L3

DEG = differentially expressed gene.
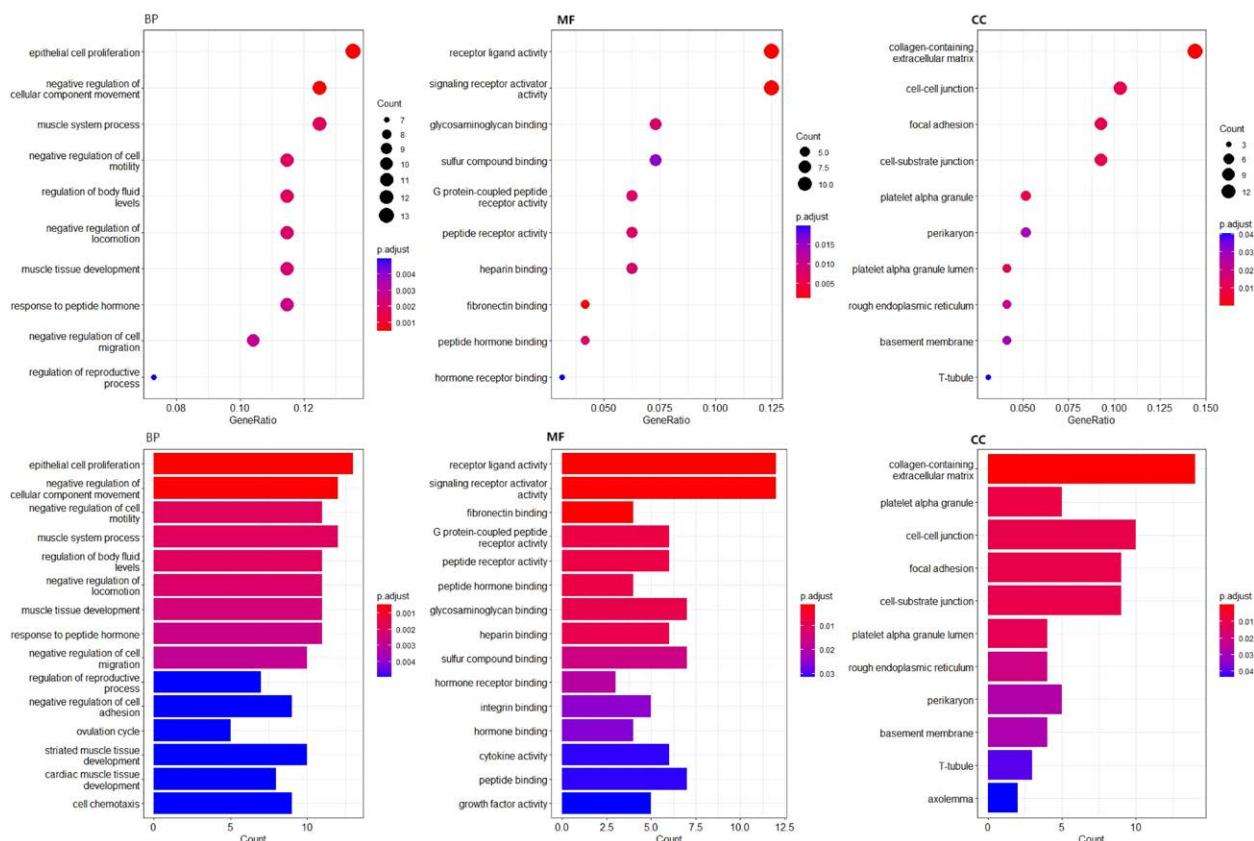
Figure 4. GO enrichment analysis of DEGs. The bubble and bar charts show the GO significance items of DEGs in 3 functional groups: MF, BP, and CC. BP = biological process, CC = cell composition, DEG = differentially expressed gene, GO = Gene Ontology, MF = molecular function.
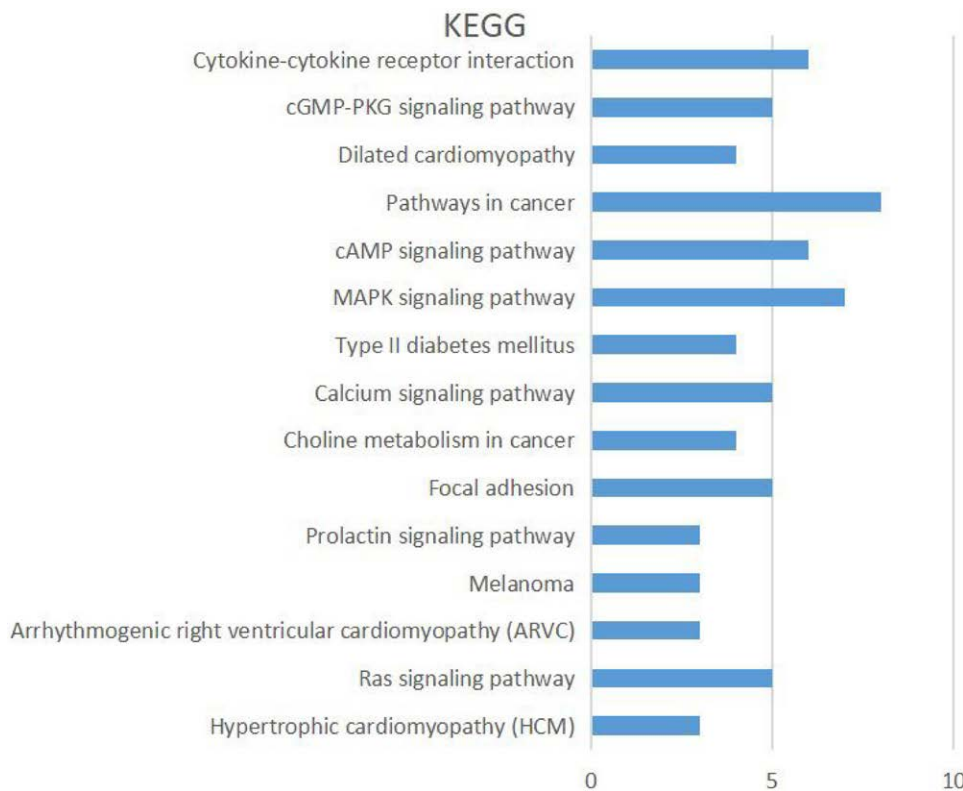
Figure 5. KEGG pathway analysis of DEGs. DEG = differentially expressed gene, KEGG = Kyoto Encyclopaedia of Genes and Genomes.

endothelial growth factor (VEGF) signaling pathway, and T-cell receptor signaling pathway (Fig. 6).

3.4. PPI network analysis and hub gene selection

We used STRING to construct a PPI network with 91 differential genes, of which there were 36 upregulated genes and 55 downregulated genes, which showed 88 nodes and 76 edges. Then we used Cytoscape for further analysis and we found 10 hub nodes, which were IGF1, HGF, SERPINE1, CXCL12, IGFBP3, LOX, PLAU, BMP2, IGFBP5, and DPP4 (Fig. 7).

Among these 10 hub gene, there are 8 upregulated genes and 2 downregulated genes.

3.5. Reanalysis of the hub gene

We continued the enrichment analysis of hub genes and found that hub genes were mainly enriched in receptor ligand activity, signaling receptor activator activity, regulation of insulin-like, regulation of Saloon athay cell migration, muscle cell migration, and receptor signaling pathway (Fig. 8).

3.6. DO analysis

We subjected hub gene to DO analysis and found that these genes were associated with prostate cancer, male reproductive organ cancer, renal cell carcinoma, non–small cell lung carcinoma, kidney cancer, retinal vascular disease, diabetic retinopathy, and pneumonia (Fig. 9).
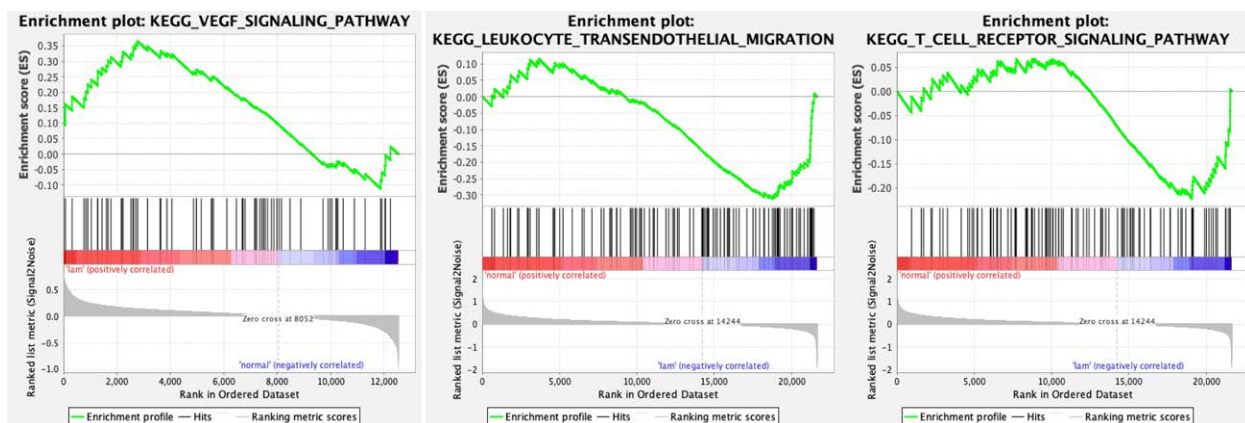

Figure 6. GSEA plots showing the most abundant set of all LAM genes in the GSE12027 and GSE160618 datasets. (1) VEGF signaling pathway, (2) leukocyte endothelial cell migration, (3) T-cell receptor signaling pathway. GSEA = Gene Set Enrichment Analysis, KEGG = Kyoto Encyclopaedia of Genes and Genomes, LAM = lymphangioleiomyomatosis, VEGF = vascular endothelial growth factor.
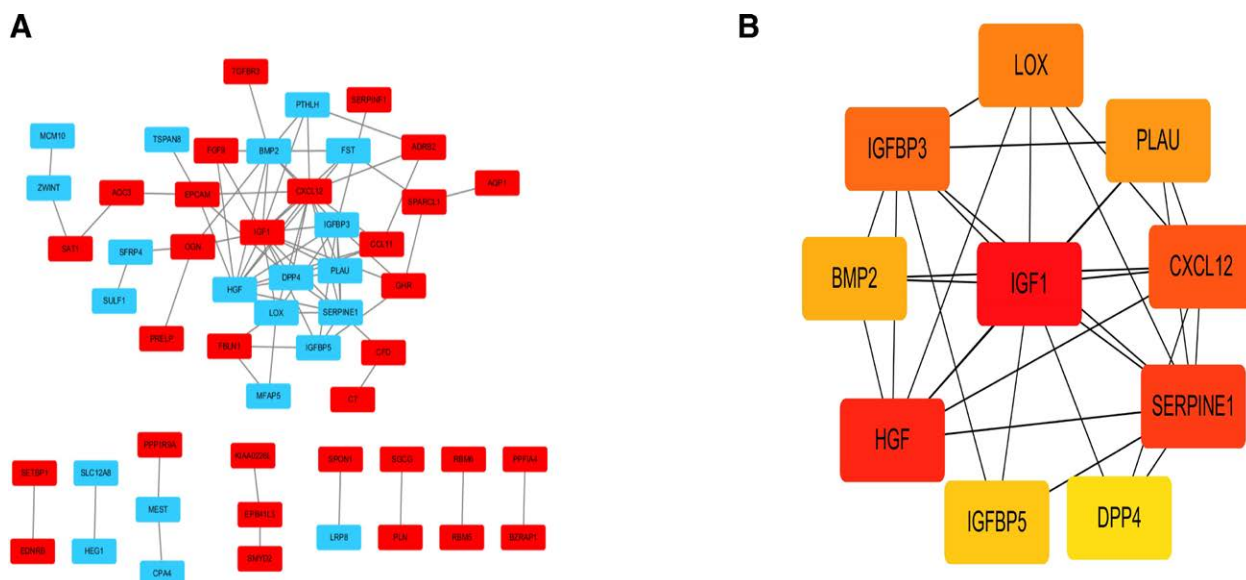
Figure 7. PPI networks. (A) PPI network of DEG, red represents upregulated genes and blue represents downregulated genes (B) subnetwork of the top 10 central genes from the PPI network. Red indicates a high degree of connectivity, yellow indicates a low degree of connectivity. DEG = differentially expressed gene, PPI = protein–protein interaction.
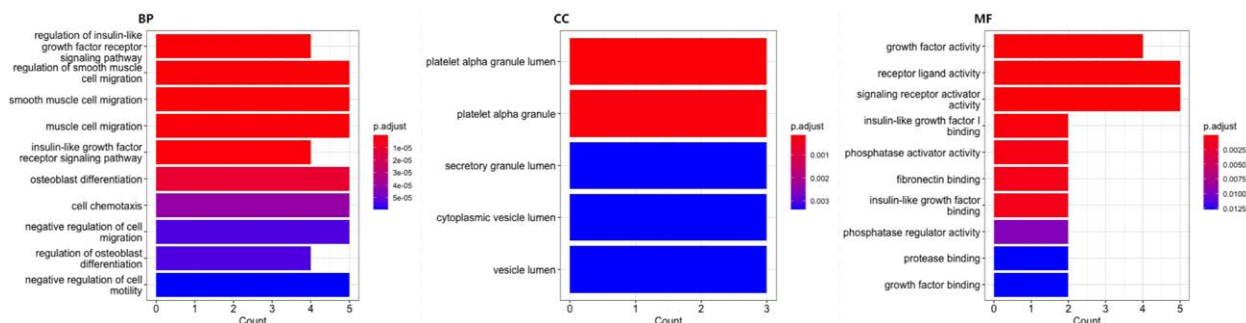

Figure 8. GO enrichment analysis of hub genes. BP = biological process, CC = cell composition, GO = Gene Ontology, MF = molecular function.

4. Discussion

LAM can involve the lungs, mediastinum, and retroperitoneal lymphatics, with the lungs being the most predominantly involved organ. PLAM occurs mostly in women of childbearing age, with studies showing 3 to 8 cases per million women. The current gold standard for the diagnosis of PLAM is lung tissue biopsy, but lung tissue biopsy is invasive and not suitable for large-scale implementation.

In this study, we mined 2 datasets, GSE12027 and GSE160618, from the GEO database, and bioinformatically analyzed the sequencing results of the whole RNA of lung LAM tissue and normal lung tissue. Ninety-one DEGs were identified, including 36 upregulated genes and 55 downregulated genes.

Enrichment analysis of these DEGs revealed that DEGs were mainly enriched in tumor-related signaling pathways and inflammatory responses.

Through the PPI network, we filtered out Hub gene, IGF1, HGF, SERPINE1, CXCL12, IGFBP3, LOX, PLAU, BMP2, IGFBP5, and DPP4.

4.1. Low-grade destructive metastatic tumors

In 2012, it was suggested that PLAM is a low-grade destructive metastatic tumor. 2015 World Health Organization lung tumor classification recognized this concept and classified it as a perivascular epithelioid cell tumor.[16]

In our study, GO enrichment analysis showed that differential genes are mainly associated with epithelial cell proliferation, microenvironment of tumors, produce extracellular matrix. In MF, these differential genes play an important role in tumor signaling. This also proves that LAM is a tumor disease. According to the KEGG results, the RAS signaling pathway and the enrichment of choline metabolism in cancer also suggest that LAM also has a malignant character. GSEA showed that the gene was enriched in the VEGF pathway. Currently, there is clear evidence that VEGF-D can be used as diagnostic evidence for LAM.[17] It was found that serum VEGF-D levels were significantly higher in PLAM patients compared to healthy controls.[2]. The American Thoracic Society/Japanese Respiratory Society recommends VEGF-D ≥ 800 pg/mL as as the threshold value for diagnosis.[18] In the hub gene, HGF promotes tumor cell motility, invasion, and metastasis and mediates tumor–mesenchymal interactions. Elevated expression of IGF1 and its receptors showed a correlation with differentiation, proliferation, and prognosis of malignant tumors. IGFBP3 levels are negatively correlated with cancer development. This reciprocal suppression of genes may be responsible for LAM being low-grade destructive metastatic tumors. SERPINE1 is a key regulator of urokinase and tissue-type plasminogen activator. The urokinase-type plasminogen activator system plays a major role in tumor growth,
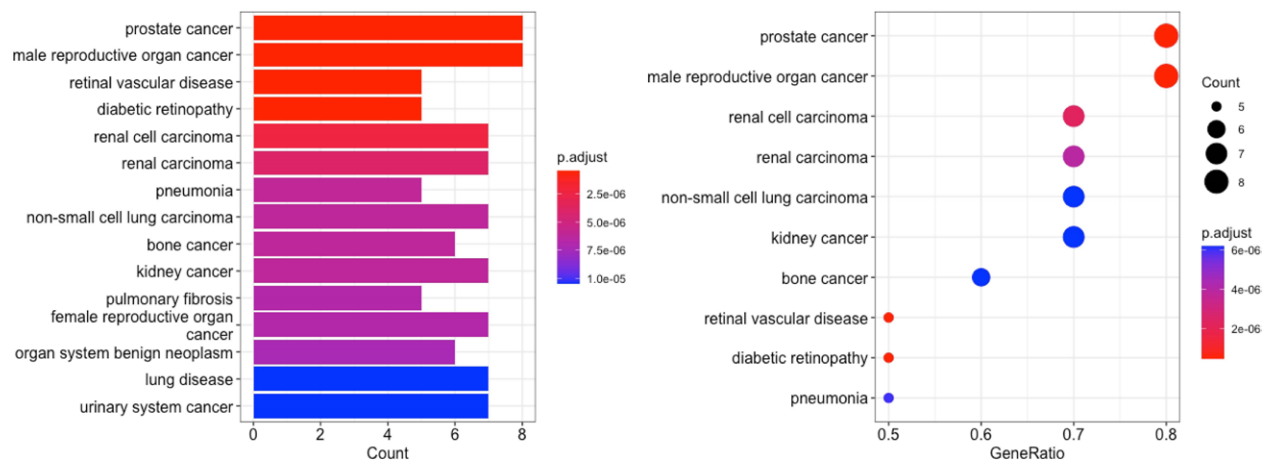
Figure 9. DO analysis of hub genes. Bubble plots and bar graphs showing the expression of hub gene in the different diseases involved. DO = Disease Ontology.

angiogenesis, tumor cell invasion, migration, and metastasis.[19] High expression of SERPINE1 increases the risk of metastasis.

4.2. Inflammatory response in LAM

We went on to perform GSEA of the samples and found that in LAM disease, the disease is mainly enriched in leukocyte endothelial cell migration, T-cell receptor signaling pathway, also confirming that inflammatory responses play an important role in LAM disease.

Existing studies have shown that the process of leukocyte migration involves: leucocyte margination, capture, rolling, activation, adhesion, and migration.[20] The adhesion molecules and other chemokines involved in this process act in a sequential manner.[21]

Adhesion molecules are important molecules that mediate communication between extracellular matrix and cells,and are glycoproteins that regulate cell–cell and cell–extracellular matrix interactions and act as adhesions. In the hub gene, CXCL12 has critical role in the inflammatory response. IL-17A promotes the activation and conversion of lung fibroblasts into myofibroblasts, which secrete significantly increased type I collagen and promote the deposition of extracellular matrix, leading to the development and progression of pulmonary fibrosis, and the chemokine CXCL12 secreted by activated fibroblasts may be involved in this process.[22]

The process of leukocyte migration exacerbates the inflammatory response in the lungs, which may contribute to the large amount of pleural fluid produced by PLAM.

5. Conclusion

PLAMis a systemic disease with a severe inflammatory response as well as borderline tumor. IGF1, IGFBP3, SERPINE1, and CXCL12 are potential key genes for inflammation and tumor invasion in PLAM, which point the way to further research on PLAM.

Acknowledgments

We would like to thank all the staff in Department of Lymphatic Surgery, Beijing Shijitan Hospital of Capital Medical University for their contribution on our research.

Author contributions

Conceptualization: Wenbin Shen
Data curation: Zimin Zhao, Wanchun Su, Jianfeng Xin, Xin Liu

Formal analysis: Zimin Zhao, Wanchun Su
Methodology: Zimin Zhao
Supervision: Wenbin Shen
Writing—original draft: Zimin Zhao
Writing—review & editing: Zimin Zhao

References

[1] Meraj R, Wikenheiser-Brokamp KA, Young LR, McCormack FX. Lymphangioleiomyomatosis: new concepts in pathogenesis, diagnosis, and treatment. Semin Respir Crit Care Med. 2012;33:486–97.

[2] Harari S, Torre O, Moss J. Lymphangioleiomyomatosis: what do we know and what are we looking for? Eur Respir Rev. 2011;20:34–44.

[3] Gupta R, Kitaichi M, Inoue Y, Kotloff R, McCormack FX. Lymphatic manifestations of lymphangioleiomyomatosis. Lymphology. 2014;47:106–17.

[4] Steagall WK, Pacheco-Rodriguez G, Darling TN, Torre O, Harari S, Moss J. The lymphangioleiomyomatosis lung cell and its human cell models. Am J Respir Cell Mol Biol. 2018;58:678–83.

[5] Khomtchouk BB, Van Booven DJ, Wahlestedt C. HeatmapGenerator: high performance RNAseq and microarray visualization software suite to examine differential gene expression levels using an R and C++ hybrid computational pipeline. Source Code Biol Med. 2014;9:30.

[6] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

[7] Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinf. 2011;12:35.

[8] Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 2006;34:D322–6.

[9] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.

[10] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16:284–7.

[11] Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics. 2015;31:2912–14.

[12] Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics. 2013;29:661–3.

[13] Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25:1091–3.

[14] Subramanian A, Tamayo P, Mootha VK, et al. Gene Set Enrichment Analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.

[15] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

[16] Nijmeh J, El-Chemaly S, Henske EP. Emerging biomarkers of lymphangioleiomyomatosis. Expert Rev Respir Med. 2018;12:95–102.

[17] Amaral AF, de Oliveira MR, Dias OM, et al. Concentration of serum vascular endothelial growth factor (VEGF-D) and its correlation with

functional and clinical parameters in patients with lymphangioleiomyomatosis from a Brazilian Reference Center. Lung. 2019;197:139–46.

[18] McCormack FX, Gupta N, Finlay GR, et al. Official American Thoracic Society/Japanese Respiratory Society clinical practice guidelines: lymphangioleiomyomatosis diagnosis and management. Am J Respir Crit Care Med. 2016;194:748–61.

[19] Geng JG. Directional migration of leukocytes: their pathological roles in inflammation and strategies for development of anti-inflammatory therapies. Cell Res. 2001;11:85–8.

[20] Takano-Ishikawa Y, Goto M, Yamaki K. Analysis of leukocyte rolling and migration—using inhibitors in the undisturbed microcirculation of the rat mesentery—on inflammatory stimulation. Mediators Inflamm. 2004;13:33–7.

[21] Ma JY, Zhu YF, Bai G. Expression of SERPINE1 in colorectal cancer and its prognostic significance. J Hebei Med Univ. 2020;41:1421–7.

[22] Wang HY, Lv JP, Chen LP, Yu WJ. Chemokine CXCL12 may be involved in the activation of mouse lung fibroblasts by interleukin 17A. J Zhejiang Univ. 2020;49:758–64.