Clinical Trial/Experimental Study

Medicine

OPEN

Identification of hub genes related to Duchenne muscular dystrophy by weighted gene co-expression network analysis

Yanning Wei, BM[a], Qisheng Su, MM[b], Xiaohong Li, MM[b,*]

Abstract

Background: The study was aimed to analyze the potential gene modules and hub genes of Duchenne muscular dystrophy (DMD) by weighted gene co-expression network analysis.

Methods: Based on the muscular dystrophy tissue expression profiling microarray GSE13608 from gene expression omnibus, gene co-expression modules were analyzed using weighted gene co-expression network analysis, gene modules related to DMD were screened, gene ontology and Kyoto encyclopedia of genes and genomes enrichment analyses were performed, and signature genes in the modules were screened. The protein–protein interaction network was constructed through Cytoscape, and hub genes were identified. The expression of hub genes in DMD versus normal muscle tissue was calculated in GSE6011.

Results: 12 co-expressed gene modules were identified, among which black module was significantly related to DMD. The characteristic genes in the module were enriched in the regulation of immune effector processes, immune response mediated by immunoglobulin, immune response mediated by B cells, etc. SERPING1, F13A1, C1S, C1R, and HLA-DPA1 were considered as hub genes in protein–protein interaction network. Analysis of GSE6011 shows that expression of SERPING1, F13A1, C1S, C1R, and HLA-DPA1 in tissues of DMD patients were higher than normal.

Conclusion: SERPING1, F13A1, C1S, C1R, and HLA-DPA1 may participate in the development of DMD by regulating innate immunity and inflammation, and they are expected to be a potential biomarker and novel therapeutic targets for DMD.

Abbreviations: DMD = Duchenne muscular dystrophy, KEGG = Kyoto encyclopedia of genes and genomes, PPI = protein–protein interaction, WGCNA = weighted gene co-expression network analysis.

Keywords: co-expression, Duchenne muscular dystrophy, gene module, immune, WGCNA

1. Introduction

Muscular dystrophy (MD) is a genetic disorder in which genetic abnormalities lead to abnormalities in a group of proteins that maintain stability of skeletal muscle cell structure and function. Chronic and progressive muscle weakness or muscle atrophy is the main clinical manifestation of MD. Common muscular dystrophy are Duchenne muscular dystrophy (DMD), Bekerer muscular dystrophy (BMD), myotonic muscular dystrophy (MD), congenital muscular dystrophy (CMD), limb girdle type Muscular dystrophy (LGMD), facial scapulohumeral muscular dystrophy (FSMD), distal muscular dystrophy, ophthalmopharyngeal muscular dystrophy (OPMD), Emery-Dreifuss muscular dystrophy (EDMD), etc. Among them, DMD is an X-recessive genetic disease caused by mutations in the anti-atropin gene. It is the most common type of progressive muscular dystrophy and the most severe type in childhood.[1] Most of the patients with DMD were found to have disease at the age of 3 to 5 years, often accompanied by delayed motor development, progressive skeletal muscle atrophy, scoliosis, joint spasm, respiratory muscle weakness, and dilated heart disease. Most patients with DMD died of respiratory failure and heart failure around the age of 20.[2,3]

At present, there are a lot of studies on DMD. Gene substitution with adenovirus and CRISPR gene editing are considered as potential therapeutic methods for DMD.[4] Some targets have been found to delay the development of DMD. For example, drugs targeting CAT and VCAM1 can alleviate DMD related cardiomyopathy[5], and the consumption of C3 by cobra venom can increase the survival of muscle cells.[6, p. 5] Although great progress has been made in the treatment of DMD, there is still much room for exploration.[7] We believe that exploring the

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

Data in this study were from open public databases, and how to access them had been explained in manuscript.

[a] School of Public Health, Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, China, [b] Key Laboratory of Clinical Laboratory Medicine of Guangxi Department of Education, Department of Clinical Laboratory, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, China.

** Correspondence: Xiaohong Li, Department of Clinical Laboratory, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region 530021, China (e-mail: 925148420@qq.com).*

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Wei Y, Su Q, Li X. Identification of hub genes related to Duchenne muscular dystrophy by weighted gene co-expression network analysis. Medicine 2022;101:52(e32603).

Received: 20 July 2022 / Received in final form: 17 December 2022 / Accepted: 19 December 2022

http://dx.doi.org/10.1097/MD.0000000000032603

molecular and mechanism involved in the occurrence and development of DMD may help to find new therapeutic targets for DMD, and may provide help for the treatment of DMD.

Weighted gene co-expression network analysis (WGCNA) is 1 way to identify gene modules and key genes related to phenotypic traits.[8] As a commonly used gene modular analysis technology, it has been extensively used to identify and screen molecular markers or drug targets for complex diseases.[9] This research uses WGCNA and other bioinformatics technologies to identify gene modules that are highly related to DMD, and screen out key genes from them. It is intended to provide a basis for a comprehensive understanding of the main biological markers and molecular mechanisms of the occurrence and development of DMD, to distinguish DMD from other muscular dystrophy diseases, and to provide theoretical support for the diagnosis and treatment of DMD.

2. Material and methods

2.1. Data acquisition and analysis

The expression matrix of GSE13608 standardized by Dchip calculated signal intensity is downloaded from gene expression omnibus database (https://www.ncbi.nlm.nih.gov/geo/), which is an open genomics database of NCBI, and the ethical approval involved had obtained by the research of data sources. Our study was based on open-source data, so there are no ethical issues and other conflicts of interest. GSE13608 is based on Affymetrix Human Genome U133 Plus 2.0 Array platform, including samples: 10 cases of DM1, 20 cases of DM2, 16 cases of DMx (DM-like, no DMPK or ZNF9 expansion), 5 cases of BMD (Becker Muscular Dystrophy), 1 cases of MC-AD (Myotonia Congenita-Autosomal Dominant), 3 cases of DMD (Duchenne Muscular Dystrophy, 4) cases of TMD (Tibial Muscular Dystrophy), 3 cases of normal fetal (Normal Fetal), 6 cases of normal adults (Normal individuals).[10] GSE6011 was used to verify the results of GSE13608 analysis, including 22 DMD patients and 14 normal controls.[11] Data analysis was performed by R 4.0.3.

2.2. Construction of weighted gene co expression network

Variance of the expression of each gene in all samples were calculated, and co-expression network was constructed with genes which variance top 25% between samples through the WGCNA package. The appropriate weighting factor β (soft threshold) was determined used the pick soft threshold function. The inter-gene correlation coefficient was further used to build a hierarchical clustering tree. Different gene modules were represented by different colors. In the hierarchical clustering tree, different branches represent different gene modules. Then, on the basis of gene weighted correlation coefficients, genes were classified according to their expression patterns. Genes with similar patterns were grouped into 1 module. After gene expression pattern differences were classified, multiple modules were formed. These modules were used in the next analysis. Finally, under the condition of the coefficient, the correlation matrix was converted into an adjacency matrix, and the adjacency matrix was further converted into a topological overlap matrix (TOM), and 400 genes were randomly selected as TOM heat maps to prove the high independence between modules and the relative independence of gene expression in each module. The minimum module size of 30 and the merged cutting height of 0.25 was set.

2.3. Module–trait correlation analysis

Pearson method was utilized to calculate the correlation coefficient between module characteristic genes and clinical traits. The module with the highest DMD correlation value was identified as the key module for analysis. Gene significance (GS) refers to the correlation between gene expression and each clinical trait, while module membership (MM) refers to the correlation between gene and module characteristics. The genes which absolute value of GS more than 0.5 ($|GS| > 0.5$) and absolute value of MM more than 0.8 ($|MM| > 0.8$) were screened out for further analysis.

2.4. Pathway and function enrichment analysis

In order to evaluate the biological function of the core genes in the key modules, the cluster profiler package was used to analyze gene ontology functional terms and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment of the core genes in the key modules.[12] If the *P* value of go and KEGG pathway enrichment is < .05 and the number of genes is more

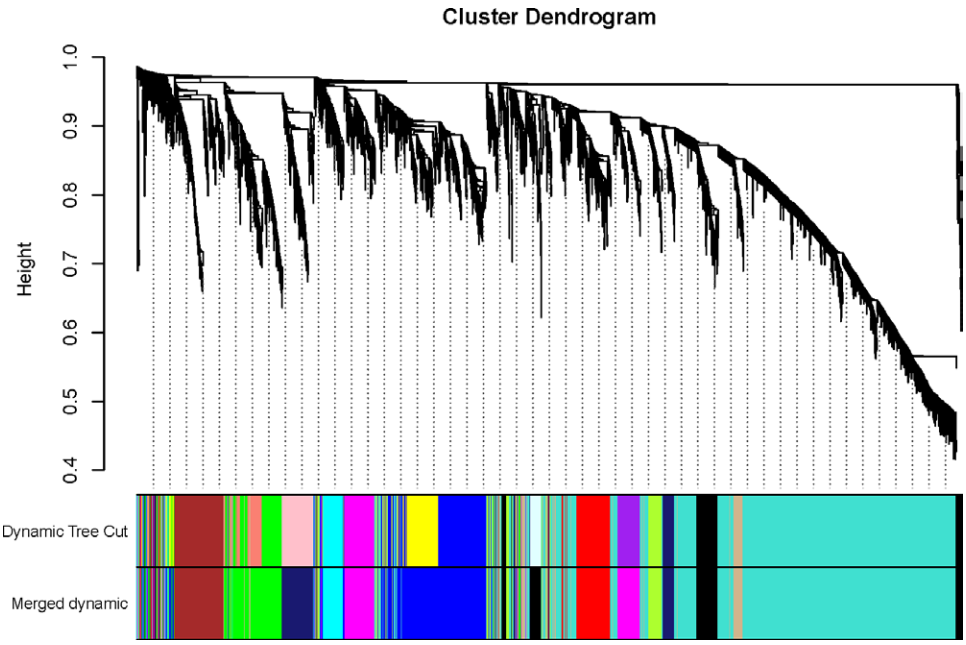

Figure 1. Cluster dendrogram. 12 co-expression modules were identified by dynamic cut tree after modules merged by cutting height which was 0.25. The original and merged modules are shown in the 2 colored bars below the dendrogram.

than 1, it is considered that the gene is significantly correlated with the analysis results.

2.5. Protein-protein interaction network

Genes of | GS | > 0.5 and |MM | > 0.8 were introduced into Cytoscape 3.7.1. Gene mania 3.5.2 was used to construct protein interaction network (PPI), and then Maximal Clique Centrality (MCC) algorithm in cytohubba plug-in was used to calculate the core proteins in the network. Subsequently, the top 5 genes in the network were selected according to the score.

2.6. Expression of hub gene was verified by GSE6011

The top 5 genes were extracted from the expression matrix of GSE6011. Wilcoxon test was used to calculate the difference of these genes expression between normal control and DMD patients. Person correlation was used to calculate the relationship among the expression levels of these 5 genes. $P < .05$ was considered statistically significant.

2.7. Statistics

The statistical methods used in this study have been described above. $P < .05$ was considered statistically significant.

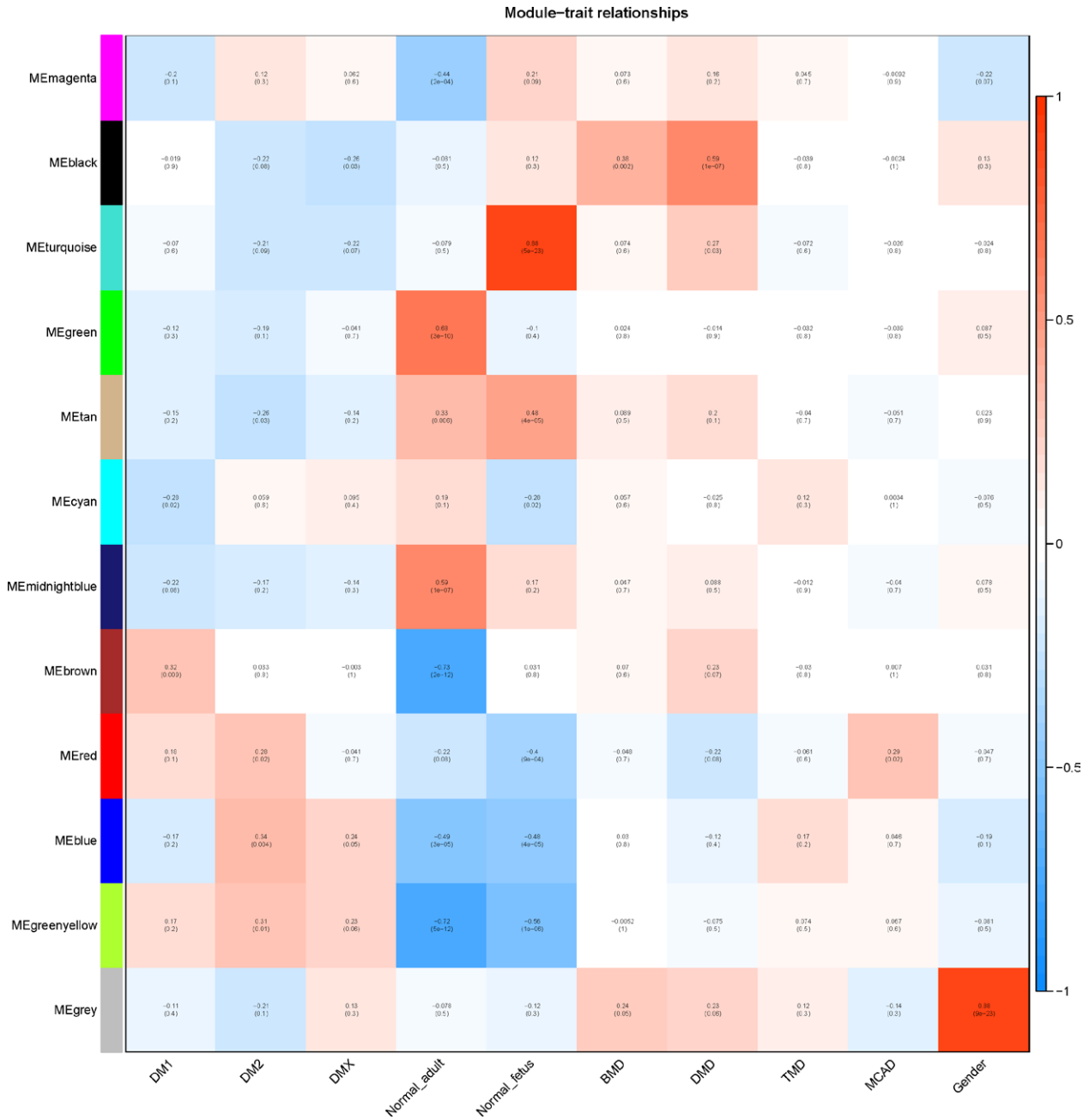

Figure 2. Module–trait relationship heat map. Pearson correlation coefficient for each correlation and *P* value in are shown in each cell.

3. Results

3.1. Construction of weighted gene co expression network

A total of 21,654 genes were obtained from the expression data, of which 5414 genes from the top 25% of the variation genes were used to calculate the co-expression network. In order to ensure the scale-free optimization of the network, pick soft threshold is used to calculate the optimal soft threshold of $\beta = 4$. While the cutting height was 0.25, WGCNA identified 12 gene co-expression modules (Fig. 1).

3.2. Module–trait relationship analysis

The correlation analysis of clinical information of each module and sample showed that DM1 and brown module, DM2 and red module, normal adult and green and midnight blue

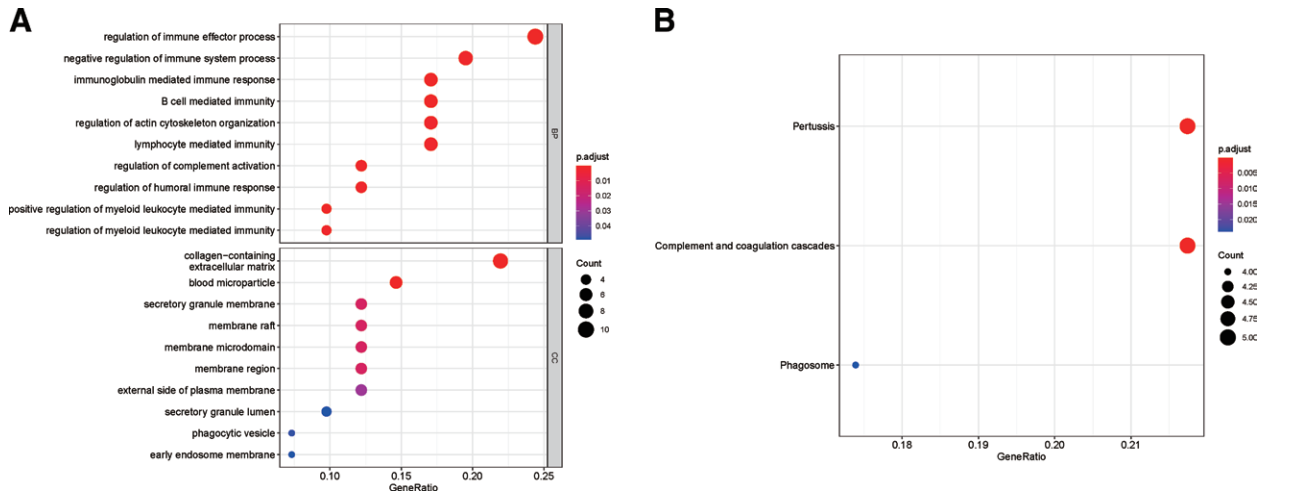

Figure 3. Enrichment of gene ontology (A) and Kyoto encyclopedia of genes and genomes pathway (B). BP, biological process; CC, cellular component. The size of the bubble represented the number of genes that were enriched, and the bubble color indicated adjust *P* value.

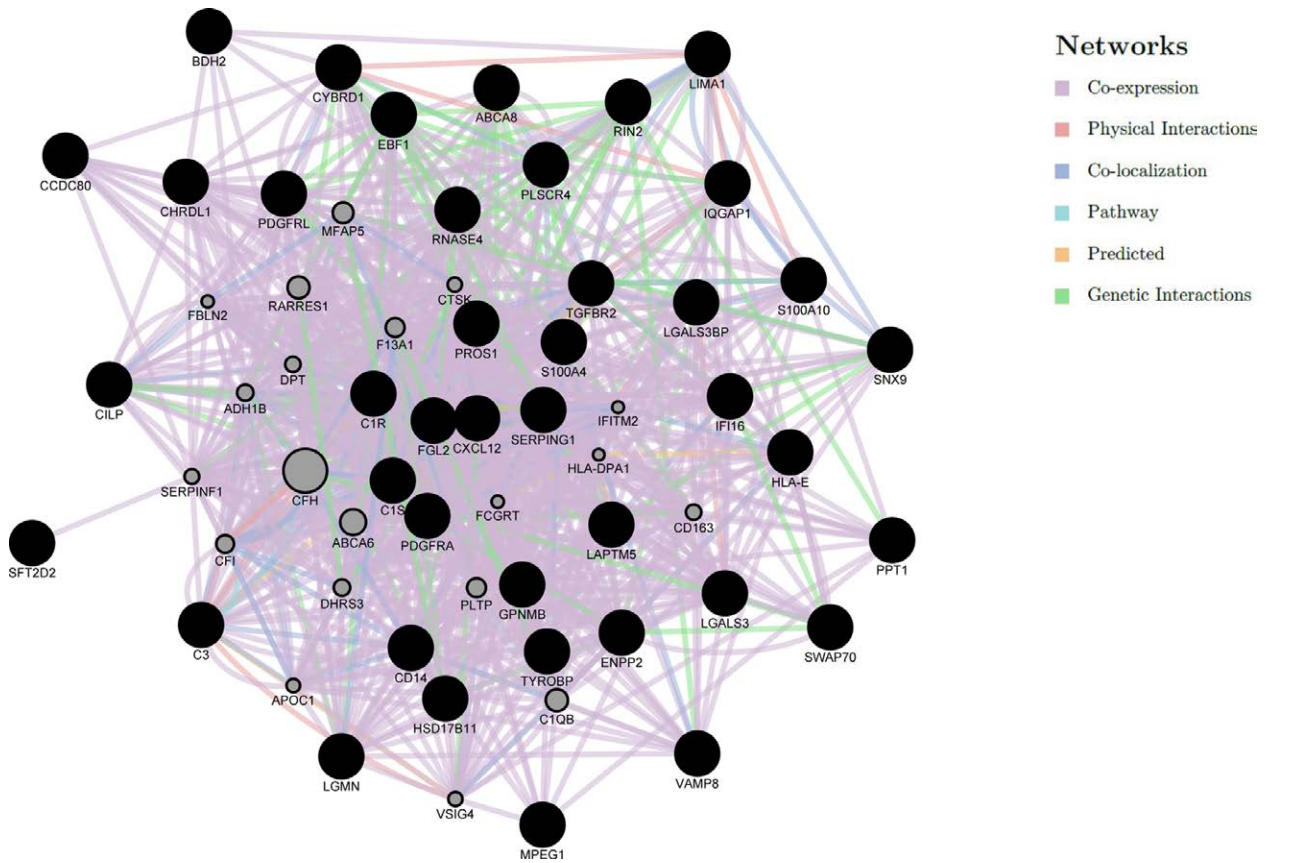

Figure 4. Protein–protein interactions network. A dot represents a protein, the size of dot represents the weight of the protein in the network, line shows the type of association between proteins.

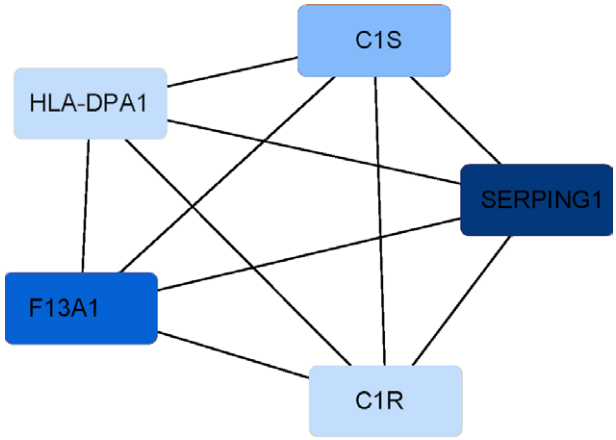

Figure 5. Top 5 genes in PPI network were distinguished by MCC algorithm of cytohubba. The depth of the color indicates the MCC calculated score. PPI = protein–protein interaction.

module, normal fetus and turquoise and Tan module, BMD and black module, gender and green module were highly correlated ($P < .05$). The black module had the highest correlation with DMD (correlation coefficient was 0.59, $P = 1e{-}07$, Fig. 2).

3.3. Pathway and function enrichment

According to the thresholds of | GS | > 0.5 and | mm | > 0.8, 43 genes were selected. Enrichment analysis showed that the 43 genes in black module were enriched in biological processes related to immune molecular regulation, such as regulation of immune effector process, immune response mediated by immunoglobulin, immune response mediated by B cells, positive regulation mediated by bone marrow leukocytes, and regulation of complement activation (Fig. 3A). The 43 genes were also enriched into 3 KEGG pathways: pertussis pathway, coagulation and complement cascade system pathway and phagosome pathway (Fig. 3B).

3.4. Protein–protein interactions network

Protein–protein interactions are often needed to achieve biological functions and metabolic reactions. Gene mania identified 41 genes of the 43 genes in black modules and PPI network was constructed. From the network, it was obvious that the interaction relationship came from co-expression (83.71%), physical interactions (8.91%), co-localization (5.24%), pathway (1.35%), predicted (0.45%), and genetic interactions (0.33%) (Fig. 4). MCC algorithm in cytohubba plug-in was used to calculate the core proteins in the network. The top 5 genes in the network were SERPING1, F13A1, C1S, C1R, and HLA-DPA1 (Fig. 5).

3.5. Expression of hub gene was significantly higher than that in normal

In GSE6011, expression of SERPING1, F13A1, C1R, C1R, and HLA-DPA1 in the muscle of DMD patients were significantly higher than that of normal muscle tissue (Fig. 6A), and the expression levels of 5 genes in GSE6011 samples showed significant correlation (Fig. 6B).

4. Discussion

DMD is the most common neuromuscular disease in the world, affecting as many as 1/ 3600 male births in the world.[13] It is a progressive muscular dystrophy due to the deletion or reduction of myofibrin caused by gene mutation on X chromosome, which is X-linked recessive inheritance.[14] Women are generally carriers, and may have mild cardiac dysfunction, but rarely have the disease.[15,16] The pathological degree of DMD is generally related to the decrease of muscle function and the deficiency of dystrophin on the plasma membrane of muscle fiber. However, these characteristics do not fully explain the formation of DMD muscle injury, which changes were not reported until the onset of the apparent disease. The immune and inflammatory processes of muscle tissue are the other important causes of muscle pathology.[17,18] In the early stage of DMD, the rising of inflammation factors (e.g., cytokines, chemokines) and the consequent activation of the immune system (e.g., CD4+ and CD8+ T cells, Tregs, neutrophils, macrophages) are hallmarks of DMD.[17,18] Many components of the innate immune system are strongly activated shortly after birth in patients with DMD before clinical symptoms appear. With the increase of age, the interaction between chronic activation of innate immunity and non-adjacent and adjacent tissue degeneration and regeneration is combined to promote the further development of DMD.[19]

Through WGCNA, black module was found to be significantly associated with DMD. Go and KEGG enrichment

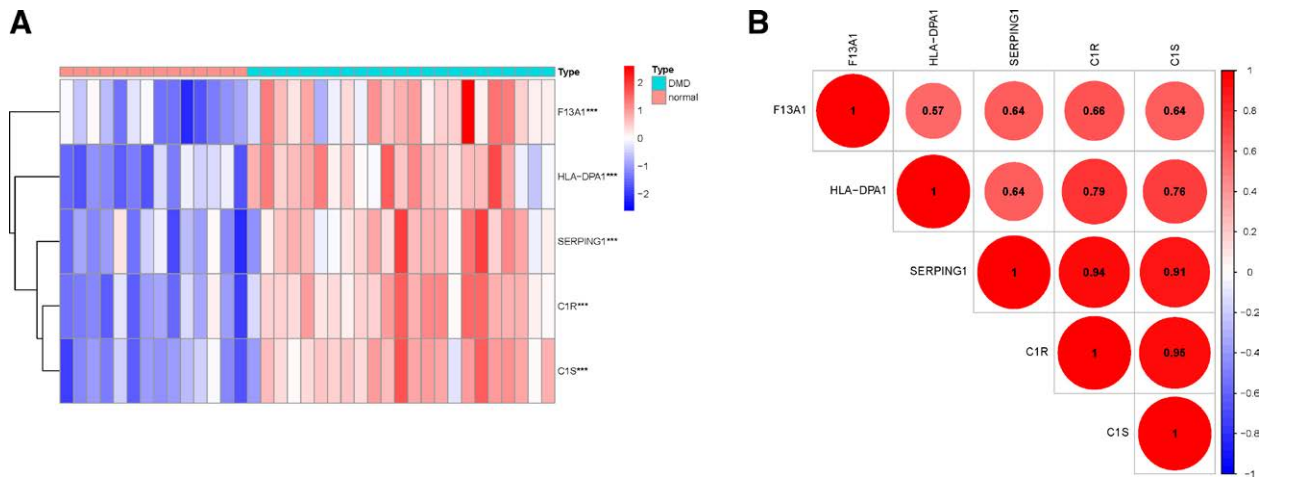

Figure 6. Top 5 genes expression in GSE6011 and their correlation. SERPING1, F13A1, C1R, C1R, and HLA-DPA1 in DMD were significantly higher than that in normal (A, $P < .05$), The correlation heat map showed that the 5 genes had significant correlation (B, $P < .05$, the number in the circles were the correlation coefficient). PPI = protein–protein interaction.

analysis of the 43 genes in black showed that these genes were significantly enriched in the regulation of immune effector process, immunoglobulin mediated immune response, B cell mediated immunity and other immune and inflammation related gene ontology entries, as well as the 3 KEGG pathways of pertussis, complex and coagulation cascades, and phagosome. Genes in the black module may participate in the development of DMD through immune and inflammation related processes.

GSE6011 was used to verify the differential expression of SERPING1, F13A1, C1S, C1R, and HLA-DPA1. Expression of 5 genes in DMD were significantly higher than that in normal muscle tissues, and they had significant correlation among samples. These 5 genes were also linked with the type of co-expression in PPI network, which is consistent with the results of WGCNA. These results indicate that these 5 genes are likely to participate in the common biological process in DMD. However, there was also study show that level of immunoreactive activity of C1 inactivator in the plasma DMD was lower than that in normal.[20] This seems to suggest that C1 has a potential role in DMD, and in our study, C1S and C1R also reflect an association with DMD. In our knowledge, SERPING1, F13A1 and HLA-DPA1 have not been reported in DMD.

It is noteworthy that SERPING1, also known as C1-INH, is a regulator of early classical and lectin pathways. Its coded C1 inhibitor inhibits activated C1r and C1, thus regulating the process of complement activation.[21] C1s/C1-INH complex increases in other diseases involving classical pathway activation, such as systemic lupus erythematosus (SLE), glomerulonephritis and rheumatoid arthritis.[22] F13A1 is a gene encoding the subunit A of coagulation factor XIII. The combination of C1-INH and factor XIII can reduce the adhesion of endothelial leukocytes and IL-1β in plasma, and increase IL-6.[23] Therefore, regulation of coordination and completion cascades may be an important way for black module genes to play a role in the occurrence and development of DMD.

Multiple molecules that in same co-expression module (black module) also show association with the occurrence and development of DMD in previous studies and were considered as targets for the treatment of DMD, such as: $C3^{6(p5)}$, CXCL12,[24] LGALS3.[25] We believe that 5 genes are potentially important molecules in DMD, they may participate in the occurrence and development of DMD by regulating innate immunity and inflammation, but the specific molecular mechanism needs further study.

5. Conclusion

This study was aimed to find genes and gene modules related to DMD, and to provide help for the diagnosis and clinical treatment of DMD. Through WGCNA, gene modules were divided, in which, black module was significantly correlated with DMD. After screening the genes in the black module, we found that SERPING1, F13A1, C1S, C1R, and HLA-DPA1 may participate in the development of DMD by regulating innate immunity and inflammation, and they are expected to be a potential biomarker and novel therapeutic targets for DMD. In our knowledge, SERPING1, F13A1 and HLA-DPA1 have not been reported in DMD and their functions and molecular mechanisms need to be confirmed by experiments. However, there are still some limitations in this study. This study was based on bioinformatics analysis, which has not been verified by experiments, and the clinical samples was relatively small. Thus, further study is required in the future.

Acknowledgments

The principal authors are very grateful for the data support provided by Gene Expression Omnibus databases.

Author contributions

Conceptualization: Xiaohong Li.
Investigation: Yanning Wei.
Methodology: Yanning Wei, Qisheng Su.
Software: Qisheng Su.
Validation: Qisheng Su.
Visualization: Xiaohong Li.
Writing – original draft: Yanning Wei.
Writing – review & editing: Xiaohong Li.

References

[1] Heydemann A. Skeletal muscle metabolism in Duchenne and Becker muscular dystrophy-implications for therapies. Nutrients. 2018;10:796.
[2] Verhaart IEC, Aartsma-Rus A. Therapeutic developments for Duchenne muscular dystrophy. Nat Rev Neurol. 2019;15:373–86.
[3] Mah J. An overview of recent therapeutics advances for Duchenne muscular dystrophy. Methods Mol Biol. 2018;1687:3–17.
[4] Mollanoori H, Rahmati Y, Hassani B, et al. Promising therapeutic approaches using CRISPR/Cas9 genome editing technology in the treatment of Duchenne muscular dystrophy. Genes Dis. 2020;8:146–56.
[5] Li B, Xiong W, Liang W-M, et al. Targeting of CAT and VCAM1 as novel therapeutic targets for DMD cardiomyopathy. Front Cell Dev Biol. 2021;9:659177.
[6] Hodgetts SI, Grounds MD. Complement and myoblast transfer therapy: donor myoblast survival is enhanced following depletion of host complement C3 using cobra venom factor, but not in the absence of C5. Immunol Cell Biol. 2001;79:231–9.
[7] Arora H. Duchenne muscular dystrophy: still an incurable disease. Neurol India. 2019;67:717–23.
[8] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008;9:559.
[9] Jia R, Zhao H, Jia M. Identification of co-expression modules and potential biomarkers of breast cancer by WGCNA. Gene. 2020;750:144757.
[10] Bachinski LL, Sirito M, Böhme M, et al. Altered MEF2 isoforms in myotonic dystrophy and other neuromuscular disorders. Muscle Nerve. 2010;42:856–63.
[11] Pescatori M, Broccolini A, Minetti C, et al. Gene expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression. FASEB J. 2007;21:1210–26.
[12] Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.
[13] Chung J, Smith AL, Hughes SC, et al. Twenty-year follow-up of newborn screening for patients with muscular dystrophy. Muscle Nerve. 2016;53:570–8.
[14] Ryder S, Leadley RM, Armstrong N, et al. The burden, epidemiology, costs and treatment for Duchenne muscular dystrophy: an evidence review. Orphanet J Rare Dis. 2017;12:79.
[15] Park J, Jang W, Han JY. Differing disease phenotypes of Duchenne muscular dystrophy and Moyamoya disease in female siblings of a Korean family. Mol Genet Genomic Med. 2019;7:e862.
[16] Silva T, Anequini IP, Fávero FM, et al. Functional performance and muscular strength in symptomatic female carriers of Duchenne muscular dystrophy. Arq Neuropsiquiatr. 2020;78:143–8.
[17] Tulangekar A, Sztal TE. Inflammation in Duchenne muscular dystrophy-exploring the role of neutrophils in muscle damage and regeneration. Biomedicines. 2021;9:13661366.
[18] Tripodi L, Villa C, Molinaro D, et al. The immune system in Duchenne muscular dystrophy pathogenesis. Biomedicines. 2021;9:14471447.
[19] Evans NP, Misyak SA, Robertson JL, et al. Immune-mediated mechanisms potentially regulate the disease time-course of Duchenne muscular dystrophy and provide targets for therapeutic intervention. PM R. 2009;1:755–68.
[20] Nagao N, Shomura I, Sawada Y. Serum activity of C1 inactivator in Duchenne-type progressive muscular dystrophy. Biochem Med Metab Biol. 1987;37:385–8.
[21] Wouters D, Wagenaar-Bos I, van Ham M, et al. C1 inhibitor: just a serine protease inhibitor? New and old considerations on therapeutic applications of C1 inhibitor. Expert Opin Biol Ther. 2008;8:1225–40.
[22] Hurler L, Toonen EJM, Kajdácsi E, et al. Distinction of early complement classical and lectin pathway activation via quantification of C1s/

C1-INH and MASP-1/C1-INH complexes using novel ELISAs. Front Immunol. 2022;13:1039765.
[23] Birnbaum J, Klotz E, Spies CD, et al. Impact of combined C1 esterase inhibitor/coagulation factor XIII or N-acetylcysteine/tirilazad mesylate administration on leucocyte adherence and cytokine release in experimental endotoxaemia. J Int Med Res. 2008;36:748–59.
[24] Maeda Y, Yonemochi Y, Nakajyo Y, et al. CXCL12 and osteopontin from bone marrow-derived mesenchymal stromal cells improve muscle regeneration. Sci Rep. 2017;7:3305.
[25] Rancourt A, Dufresne SS, St-Pierre G, et al. Galectin-3 and N-acetylglucosamine promote myogenesis and improve skeletal muscle function in the mdx model of Duchenne muscular dystrophy. FASEB J. 2018;32:6445–55.