Potential Mechanism and Therapeutic Targets of Polymyositis

Ye He[1], Junjie Zheng[2], Jing Luo[2], Fei Liao[1], Mingzhi Hong[3] and Xiaochun Zhu[2]

[1]Department of Rheumatology, Taizhou Municipal Hospital, Taizhou, China
[2]Department of Rheumatology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
[3]Department of Burns and Plastic Surgery, Taizhou Municipal Hospital, Taizhou, China

ABSTRACT

Objective: To investigate the variability in the expression profile of genes associated with polymyositis (PM), explore the potential molecular mechanisms underlying PM, and predict novel targets for intervention.

Study Design: Descriptive study.

Place and Duration of the Study: Department of Rheumatology, Taizhou Municipal Hospital, Taizhou, China, from August to November 2023.

Methodology: Three microarray datasets (GSE3112, GSE39454, and GSE128470) were extracted from the gene expression omnibus (GEO). The analysis of this research involved identifying the differentially expressed genes (DEGs) in PM compared to normal samples. Enrichment analysis, gene-microRNA, gene-transcription factor (TF), and protein-protein interaction (PPI) network studies were conducted to identify hub genes and relevant pathways. Additionally, the drug-gene interaction database (DGIdb) was used to predict therapeutic medications.

Results: Eighty-eight DEGs were identified. The enrichment analysis results highlighted the significant involvement of downregulated DEGs in antigen processing and presentation. Based on the PPI networks, seven hub genes with high connectivity degrees were selected including a cluster of differentiation 74 (*CD74*), human leukocyte antigen (*HLA*)-*DPA1*, *HLA-B*, guanylate-binding protein 1 (*GBP1*), recombinant 2', 5'-oligoadenylate synthetase 1 (*OAS1*), *HLA-C*, and *HLA-E*.

Conclusion: This research screened-out core genes, projected prospective therapeutic medications, discovered DEGs between PM and normal samples, and offered fresh perspectives for additional research into the possible mechanism and therapeutic targets of PM.

Key Words: Polymyositis, DEGs, Hub genes, Bioinformatics, Potential therapeutic agents.

INTRODUCTION

Polymyositis (PM) is a type of systemic autoimmune illness that causes skeletal muscle inflammation over time. PM shows bilateral symmetrical weakness of the pharynx, neck muscles, and proximal limb muscles that lasts for weeks to months. The disease may progress from musculoskeletal system involvement to heart, skin, or lung involvement.[1] PM is twice as frequent in women as in men, and its one in 100,000 aetiologies probably combine environmental and genetic risk factors.[1]

Currently, routine clinical medication of PM includes steroids, methotrexate, azathioprine, calcineurin inhibitors, mycophenolate mofetil, cyclophosphamide, rituximab, and Janus kinase (JAK) inhibitors. Although these agents have certain curative effects, there are still some problems, such as poor relief of symptoms in some patients.[1] In order to identify the hub gene and the differentially expressed genes (DEGs), thus it should be planned to extract the original microarray dataset from the gene expression omnibus (GEO), as well as the development of protein-protein interaction (PPI) network. The objective of this study was to investigate the variability in the expression profile of genes associated with PM, explore the potential molecular mechanisms underlying PM, and predict novel targets for intervention.

METHODOLOGY

The gene expression datasets GSE39454, GSE128470, and GSE3112 were obtained from the GEO-NCBI database.[2,3] The GSE39454 dataset was collected using the GPL570 platform (HG-U133_Plus_2; Affymetrix human genome U133 Plus 2.0 array) and consisted of eight samples from PM patients and five

normal samples. The GSE128470 dataset, on the other hand, used the GPL96 platform (HG-U133A; Affymetrix human genome U133A array) and included information from seven PM patients and twelve normal samples. Furthermore, the GSE3112 dataset also used the GPL96 platform, comprised data from six PM samples and eleven normal samples.

To detect shared DEGs among normal and PM samples, the GEO2R web-based analytical tool offered by NCBI was used. This tool allowed the authors to generate a table of DEGs, which were then ranked by their significance in each gene expression profile. The thresholds were set at $p < 0.05$ and $|log2FC| \geq 1.0$ to ensure the inclusion of highly relevant DEGs. To further analyse the overlapping DEGs across the three datasets, the online tool called Draw Venn diagram available at http://bio informatics.psb.ugent.be/webtools/Venn/ was used. This tool aided the authors in constructing a Venn diagram that visually displayed the shared DEGs among the samples. To examine the biological function of the DEGs between PM and normal samples, various analyses were performed. Firstly, a comprehensive assessment incorporating terms related to biological processes, cell components, and molecular functions were conducted.[4] Additionally, for pathway enrichment analysis, Kyoto Encyclopaedia of Genes and Genomes (KEGG) was used.[5] To further enhance the understanding of the DEGs, the ClueGO plug-in into Cytoscape was introduced, enabling the integration of gene ontology (GO) terms and the creation of back propagation (BP) networks.[6]

The PPI network of the DEGs by using STRING was constructed, the available access is http://www. string-db.org/, considering a typical criterion of combined score $p > 0.4$. Subsequently, the PPI network through Cytoscape software was visualised.[7] To identify significant modules, molecular complex detection V1.5.1 was employed, a plug-in in Cytoscape.[8] Additionally, Cytohubba, which encompasses 12 different approaches was used to explore key nodes and determine the hub genes based on the PPI network.[9] To ascertain the most crucial essential genes, the top 15 genes identified by each topological approach that were common to at least six methods. Finally, the hub genes as those shared by the three methods (PPI network, MCODE, and topological algorithm analysis in cytoHubba) were identified.

An analysis was performed on the connections between target genes at the post-transcriptional level and miRNA or gene-transcription factor (TF) expression in several illness situations.[10,11] The miRNA-mRNA interaction networks were predicted with miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/php/download. php), miRDB (http://www.mirdb.org/), and miRWalk (http://mirwalk.umm.uni-heidelberg.de/). The NCBI and PROMO (http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) were also used to predict TFs. The target gene-miRNA and gene-TF network were visualised using Cytoscape software.

Finding the genes linked to a disease and developing tailored medications for these genes is the ultimate objective of many

illness investigations. DGIdb V3.0.2 with available access http://www.dgidb.org was used to predict potential medication interactions with the DEGs.[12] Cytoscape was used to show the drug-gene interaction network.

RESULTS

To acquire a roster of DEGs linked with PM, a comparison of gene expression profiles in muscle tissues between individuals diagnosed with PM and healthy volunteers was conducted. Based on pre-determined criteria, 1,218 DEGs from GSE39454, 1,517 DEGs from GSE128470, and 2,510 DEGs from the GSE3112 were identified. From the results of the volcano plot, the DEGs were identified from the above datasets using the criteria $p < 0.05$ and $|log2FC| \geq 1.0$ Figure 1(A)-1(C). By analysing the dataset GSE39454, it was found that there were 653 genes that exhibited upregulation and 565 genes that showed downregulation. In the case of GSE128470, there were 127 genes with upregulation and 1,390 genes with downregulation. Similarly, GSE3112 had 66 upregulated genes and 2,444 downregulated genes. To determine the common DEGs, a Venn analysis was performed as shown in (Figure 1D and E). In total, there were 88 DEGs that were common among all three comparisons. Among these, four genes were significantly upregulated, while 84 genes displayed downregulation.

To conduct a more in-depth analysis of the BPs implications related to the 84 downregulated DEGs, the ClueGO plug-in within Cytoscape software was used. This tool allowed the authors to group the genes based on their gene ontology (GO) terms, which provided insights into their involvement in specific biological functions. Moreover, the authors employed it to depict the connections within each cluster and those between different groups, as depicted in (Figure 1G and F).

The GO analysis showed a significant enrichment of downregulated DEGs in various BPs. These processes included the presentation and processing of peptide antigens (GO:0048002), the cellular response to interferon-gamma (GO:0071346), promoting lymphocyte proliferation (GO:0050671), enhancing mononuclear cell proliferation (GO:0032946), cytokine release (GO:0050671), controlling complement activation (GO:0030449), promoting chemokine production (GO:0032722), and suppressing viral processes (GO:0048525). Moreover, the candidate DEGs were mainly enriched in major histocompatibility complex (MHC) protein complexes (GO:0042611), the luminal side of the endoplasmic reticulum membrane (GO:0098553), ER to Golgi transport vesicle membranes (GO:0012507), blood microparticles (GO:0072562), platelet alpha granule lumens (GO:0031093), and multivesicular bodies (GO:0005771) within cellular components (CC). Importantly, downregulated DEGs exhibited significant enrichment in MHC protein complex binding (GO:0023023), immune receptor activity (GO:0140375), chemokine receptor binding (GO:0042379), and proteoglycan binding (GO:0043394) within molecular function (MF). Additionally, the KEGG pathway analysis revealed the distribution of DEGs in antigen processing and presentation.
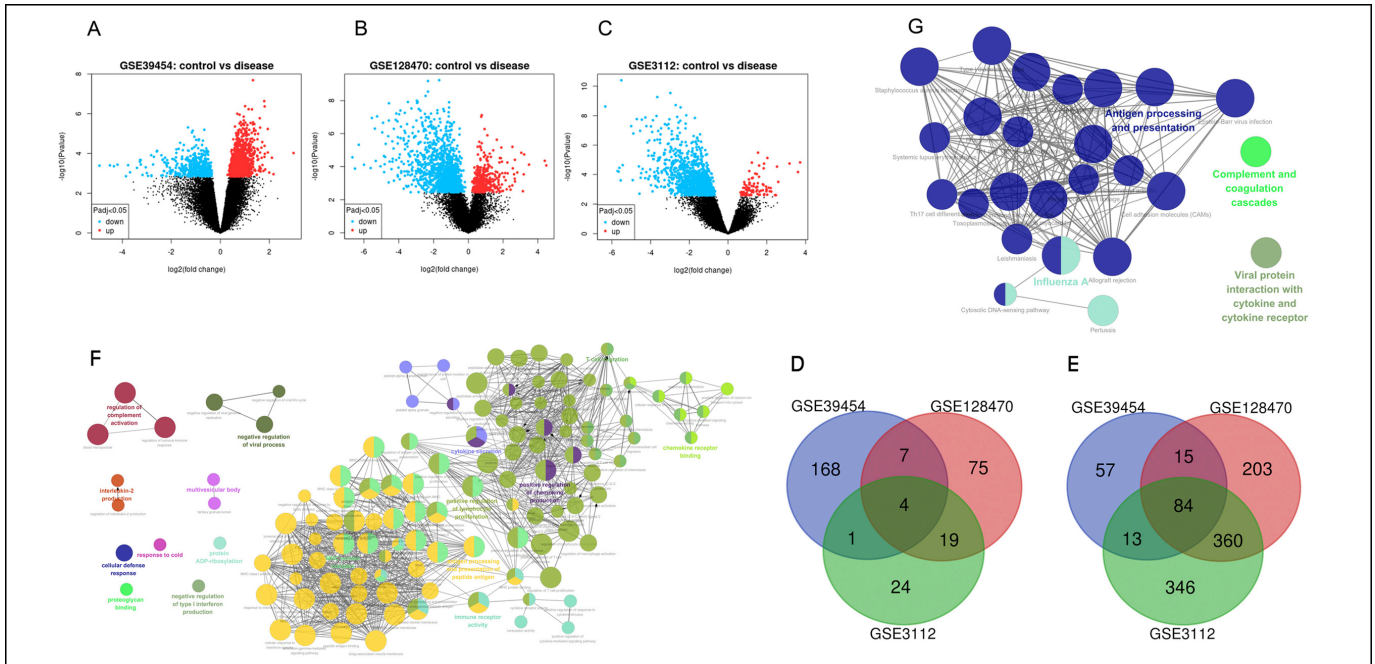
Figure 1: Identification of DEGs and GO term enrichment, KEGG pathway analysis of DEGs. Volcano plot showing the differentially expressed genes identified from the (A) GSE39454, (B) GSE128470, and (C) GSE3112 datasets. Venn diagram of common DEGs identified from the three gene expression profiles: (D) 4 DEGs were upregulated in the three datasets, and (E) 84 DEGs were downregulated in the three datasets. (F) Significantly enriched GO terms among the DEGs in the following three functional groups: MF, BP, and CC. (G) KEGG pathways analysis of DEGs. Nodes were coloured according to grouping of related functions based on statistically significant association of related GO terms. In each group, only the most significant term is labelled, and the node size corresponded to the significance of each GO term.
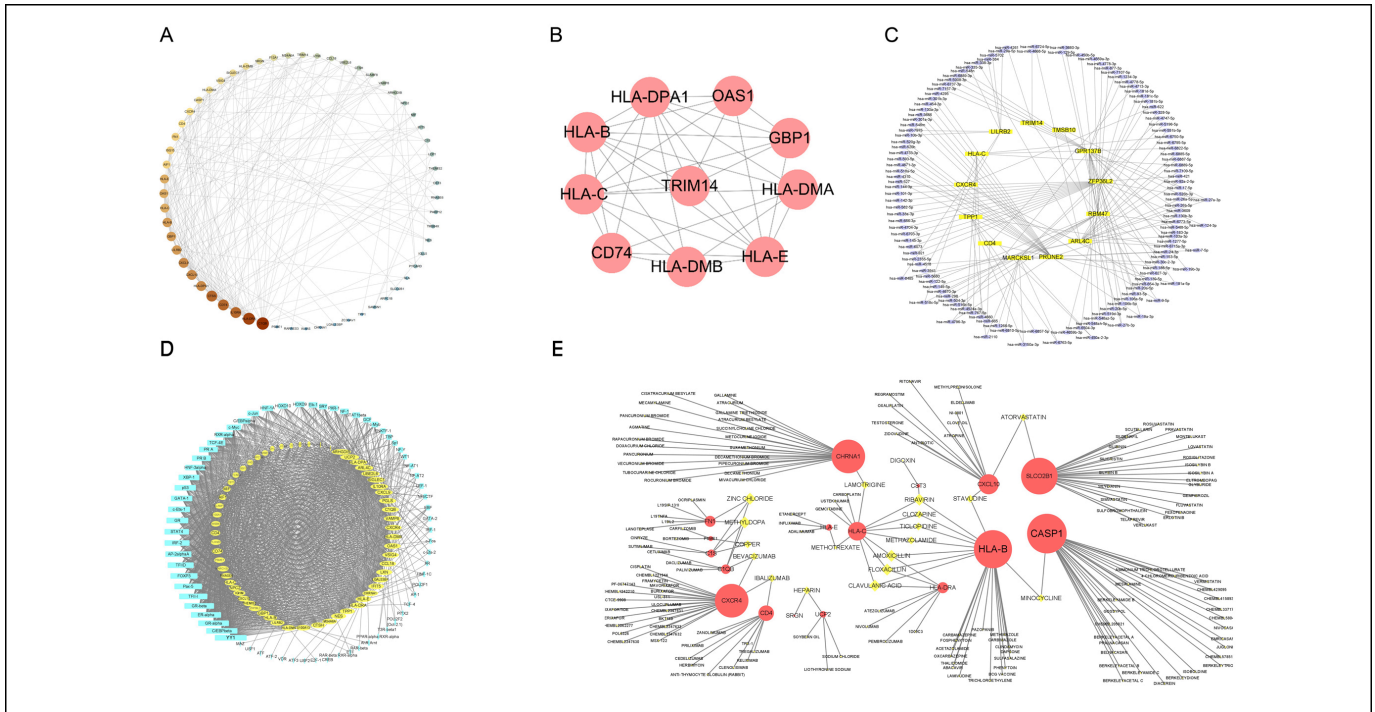

Figure 2: Identification of hub genes and analysis of targeting TFs of DEGs. (A) Protein-protein interaction (PPI) network of 84 downregulated DEGs. (B) Core module (module 1 with an MCODE score of 8) from the PPI network. The colour shadow of each node represents the MCODE score (degree of connection of nodes). (C) Target gene-miRNA network. The yellow rectangle nodes are the genes, and purple ellipse nodes are the miRNAs. (D) Target gene-TF network. The yellow ellipse nodes are the genes, and blue rectangle nodes are the TFs. (E) Drug-gene interaction network. The purple circle nodes are the genes. The diamond nodes are the drugs, which are distinguished by yellow, orange and, red colours according to degree.

As shown in Figure 2(A), the interaction network of DEGs acquired using STRING encompassed 22 genes with reduced expression grouped together in a cluster comprising 75 nodes and 222 edges. Figure 2(B) exhibited the implementation of MCODE to recognise the most notable module, which included 10 nodes, all of which were DEGs with downregulated expression (MCODE score = 8). From each CytoHubba analysis approach, the top 15 pivotal genes were opted and discovered that more than six topological analysis methods identified a total of 15 hub genes. In summary, seven key genes were identified by all three analyses: Cluster of differentiation 74 (*CD74*), human leukocyte antigen (*HLA*)-*DPA1*, *HLA-B*, guanylate-binding protein 1 (*GBP1*), recombinant 2',5'-oligoadenylate synthetase 1 (OAS1), *HLA-C*, and *HLA-E*.

There were six primary DEGs that were subjected to regulation by microRNAs (miRNAs). These DEGs included *ZFP36L2*, which was targeted by a total of 24 miRNAs; *GPR137B*, influenced by 23 miRNAs; *ARL4C*, affected by 19 miRNAs; *PRUNE2*, influenced by 18 miRNAs; *RBM47*, modulated by 16 miRNAs; and *MARCKSL1*, impacted by 13 miRNAs. Among these miRNAs, hsa-miR-27a-3p seemed to possess the highest regulatory control over three DEGs (Figure 2C). On the other hand, the TFs targeted the top five DEGs in significant numbers. *ARHGDIB*, for instance, was regulated by 36 TFs, while *UCP2* was modulated by 35 TFs. Additionally, *ARL4C*, *UBE2L6*, and *HLA-DPA1* were each influenced by 34 TFs (Figure 2D).

To identify prospective therapeutic agents or chemical entities capable of restoring the decreased expression of DEGs in PM, DGIdb was used. Based on the network of drug-gene interactions (Figure 2E), amoxicillin, clavulanic acid, and floxacillin were identified as molecular compounds that were predicted to regulate the expression of *HLA-DRA, HLA-B*, and *HLA-C*. Zinc chloride and methyldopa were predicted to affect the differential expression of *FN1*, *C1S*, and *C1QB*. Additionally, 24 drugs or molecular compounds, such as erlotinib and glyburide, were found to interact with SLCO2B1 and agmatine was also identified.

DISCUSSION

PM is a refractory systemic autoimmune disease. The pathogenesis, phenotype, and progression of PM are believed to be influenced by the intricate interplay of several genetic and environmental factors. There is still a need for further exploration into the pathogenesis and treatment of PM.

During this study's scientific investigation, a combination of bioinformatics strategies was utilised to assist in the examination of crucial alterations in gene expression. Three GEO datasets: GSE39454, GSE128470, and GSE3112 were conducted to unveil potential pathways related to muscle tissue. Through this approach, a total of 88 DEGs were identified successfully. Among these DEGs, 84 were found to be downregulated, while 4 displayed upregulation.

Through functional enrichment analyses, the effect of downregulated DEGs was clarified. The significance of antigen processing and presentation among the DEGs was highlighted by the enriched GO terms identified through a comprehensive analysis of GO enrichment. Consistent with previous evidence, the enrichment analysis also revealed a significant association between HLA alleles and most of the GO terms, indicating the critical role of HLA molecules in PM pathogenesis. The authors' analysis of MF corresponds to the prior research, as it pinpointed MHC protein complex bonding as the foremost significant GO term. Additionally, the enrichment analysis of KEGG revealed the importance of the pathway for complement and coagulation cascades-a pathway vital for a multitude of cellular functions.

Seven hub genes, *CD74, HLA-DPA1, HLA-B, GBP1, OAS1, HLA-C*, and *HLA-E*, were identified. All of these hub genes were downregulated in PM. *CD74*, the invariant chain found within the MHC class II complex, is a glycoprotein that traverses the cellular membrane. *CD74* actively participates in B cell maturation by engaging in a pathway that includes the activation of nuclear factor kappa-light-chain-enhancer.[13] *HLA-DPA1*, functions as an *HLA-DP* peptide chain and an MHC class II receptor. It actively participates in immune responses and the presentation of antigenic peptides. *GBP1*, a crucial GTPase belonging to the dynamin superfamily, assumes a significant role for the regulation of membrane dynamics, cell cycle progression, and the organisation of the cytoskeleton. Additionally, it has been observed that *GBP1* acts as a protective factor against IFNγ-induced apoptosis in chronic inflammation.[14] *OAS1* belongs to the family of 2'-5'-oligoadenylate synthetases and is a gene that is induced by interferons. Studies suggest that *OAS1* has the ability to induce apoptosis and enhance IFNαβ signalling. Previous research investigating the connections has established that genetic variations within the HLA complex play a vital role as risk factors for idiopathic inflammatory myopathies. *HLA-B* encodes one of the primary variants of HLA Class I cell surface receptors which present short polypeptides derived from self or foreign sources to cytotoxic T cells for recognition and binding. The *HLA-B**08:01 allele exhibits a slight increase in association with PM.[15] *HLA-C*, on the other hand, is a polymorphic membrane protein that contributes to adaptive immunity. It does so by binding peptides derived from within cells and presenting them to CD8+ cytotoxic T cells. *HLA-E* is a significant molecule in immune surveillance, belonging to the low-polymorphism nonclassical MHC Class I. It is responsible for presenting peptides to T and NK cells. In terms of molecular impact, a total of seven central genes have been identified to play a crucial role in the advancement of PM and its related complexities. Consequently, they offer new prospects for potential therapeutic targets in addressing PM-related issues.

In this investigation, the foremost six targeted DEGs in the miRNA-gene network were *ZFP36L2, GPR137B, ARL4C, PRUNE2, RBM47*, and *MARCKSL1*. According to Makita *et al.*, *ZFP36L2* diminishes the expression of iTregs and hampers their function. Additionally, suppressing *ZFP36L2* in iTregs might offer a potential therapeutic approach for autoimmune diseases.[16] *GPR137B*, on the other hand, is a lysosomal regulatory protein resembling a GPCR that indirectly governs cell proliferation.[17] Furthermore, RNA-binding motif-protein-47 (*RBM47*) enhances the production of IL-10 in B cells while simultaneously preserving the stability of IL-10 messenger RNA. Lastly, myristoylated alanine-rich C kinase substrate like-1 (*MARCKSL1*) is a tethered protein within the cell membrane that assumes a role in tasks such as cell-spreading, activation of integrins, and exocytosis. Further research regarding the relationship of *ARL4C* and *PRUNE2* with PM is clearly needed.

The TF-gene network identified *ARHGDIB, UCP2, ARL4C, HLA-DPA1*, and *UBE2L6* as the top five DEGs. *UCP2*, an anion carrier localised in mitochondria, exerts an influence on inflammatory and metabolic processes.[18] *UBE2L6*, an enzyme involved in the interferon response pathway, assumes a crucial role in mediating the immune response.[19] The roles of *ARL4C* and *HLA-DPA*1 in PM have previously been explored.

There were some limitations to this research. Regrettably, the sample size was not sufficiently extensive, and as a consequence of insufficient experimentation, this study's findings were unable to be corroborated. To surmount this hurdle, the collection of samples would broaden for subsequent investigations and conduct verification experiments.

CONCLUSION

Analysing three microarray datasets, this research picked out seven hub genes of polymyositis, namely *CD74, HLA-DPA1, HLA-B, GBP1, OAS1, HLA-C*, and *HLA-E*. In addition, the top six DEGs targeted by miRNAs were *ZFP36L2, GPR137B, ARL4C, PRUNE2, RBM47*, and *MARCKSL1*. The hsa-miR-27a-3p was indicated as the miRNA responsible for regulating the highest quantity of DEGs. Five primary DEGs that underwent targeting by TFs were A*RHGDIB, UCP2, ARL4C, UBE2L6*, and *HLA-DPA1*. The potential medicines with therapeutic purposes were predicted. Additional investigation is necessary to validate these outcomes and unravel additional mechanisms. The present and future findings may contribute to the discovery of novel therapeutic targets for PM.

ETHICAL APPROVAL:
This study was approved by the Ethics Committee of the authors' institution.

PATIENTS' CONSENT:
There are no patient cases involved in this work.

COMPETING INTEREST:
The authors declared no conflict of interest.

AUTHORS' CONTRIBUTION:
YH: Wrote the manuscript draft.
YH, JZ, JL, FL: Examined the data.
MH, XZ: Manuscript review and experimental design.
All authors approved the final version of the manuscript to be published.

REFERENCES

1. Dalakas MC, Hohlfeld R. Polymyositis and dermatomyositis. *Lancet* 2003; **362(9388)**:971-82. doi: 10.1016/S0140-6736 (03)14368-1.

2. Zhu W, Streicher K, Shen N, Higgs BW, Morehouse C, Greenlees L, *et al*. Genomic signatures characterize leuko-cyte infiltration in myositis muscles. *BMC Med Genomics* 2012; **5**:53. doi: 10.1186/1755-8794-5-53.

3. Greenberg SA, Bradshaw EM, Pinkus JL, Pinkus GS, Burleson T, Due B, *et al*. Plasma cells in muscle in inclusion body myositis and polymyositis. *Neurology* 2005; **65(11)**:1782-7. doi: 10.1212/01.wnl.0000187124.92826.20.

4. Gene ontology consortium. The gene ontology (GO) project in 2006. *Nucleic Acids Res* 2006; **34**(Database issue): D322-6. doi: 10.1093/nar/gkj021.

5. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28(1)**:27-30. doi: 10.1093/nar/28.1.27.

6. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, *et al*. A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009; **25(8)**:1091-3. doi: 10.1093/ bioinformatics/btp101.

7. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al*. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13(11)**:2498-504. doi: 10.1101/gr.1239303.

8. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003; **4**:2. doi: 10.1186/1471-2105-4-2.

9. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 2014; **8(Suppl)**:S11. doi: 10.1186/ 1752-0509-8-S4-S11.

10. Baldwin AS Jr. Series introduction: The transcription factor NF-kappaB and human disease. *J Clin Invest* 2001; **107(1)**: 3-6. doi: 10.1172/JCI11891.

11. Soifer HS, Rossi JJ, Saetrom P. MicroRNAs in disease and potential therapeutic applications. *Mol Ther* 2007; **15(12)**: 2070-9. doi: 10.1038/sj.mt.6300311.

12. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, *et al*. DGIdb 3.0: A redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res* 2018; **46(D1)**: D1068-73. doi: 10.1093/nar/gkx1143.

13. Becker-Herman S, Arie G, Medvedovsky H, Kerem A, Shachar I. *CD74* is a member of the regulated intramem-brane proteolysis-processed protein family. *Mol Biol Cell* 2005; **16(11)**:5061-9. doi: 10.1091/mbc.e05-04-0327.

14. Honkala AT, Tailor D, Malhotra SV. *Guanylate-Binding Protein 1*: An emerging target in inflammation and cancer. *Front Immunol* 2020; **10**:3139. doi: 10.3389/fimmu.2019.03139.

15. Davis DM. Presenting the marvels of immunity. *Nat Rev Immunol* 2014; **14(6)**:351-3. doi: 10.1038/nri3681.

16. Makita S, Takatori H, Iwata A, Tanaka S, Furuta S, Ikeda K, *et al*. RNA-binding protein *ZFP36L2* downregulates helios expression and suppresses the function of regulatory T cells. *Front Immunol* 2020; **11**:1291. doi: 10.3389/fimmu.2020.01291.

17. Gan L, Seki A, Shen K, Iyer H, Han K, Hayer A, *et al*. The lysosomal GPCR-like protein *GPR137B* regulates Rag and mTORC1 localization and activity. *Nat Cell Biol* 2019; **21(5)**:614-26. doi: 10.1038/s41556-019-0321-6.

18. Wei W, Li Y, Li Y, Li D. Adipose-specific knockout of ubiquitin-conjugating enzyme E2L6 (*Ube2l6*) reduces diet-induced obesity, insulin resistance, and hepatic steatosis. *J Pharmacol Sci* 2021; **145(4)**:327-34. doi: 10.1016/j.jphs.2020.12.008.

19. El-Asmi F, McManus FP, Brantis-de-Carvalho CE, Valle-Casuso JC, Thibault P, Chelbi-Alix MK. Cross-talk between SUMOylation and ISGylation in response to interferon. *Cytokine* 2020; **129**:155025. doi: 10.1016/j.cyto.2020.155025.

• • • • • • • • • •