Reveal the biomarkers and immune infiltration of dermatomyositis and explore associated diseases as well as drug molecules by bioinformatics analysis

Haonan Zhou
Queen Mary College of Nanchang University

Jie Peng
The Second Clinical Medical College of Nanchang University

Zhen Zong
the Second Affiliated Hospital of Nanchang University

Rui Wu
the First Affiliated Hospital of Nanchang University

Hui Li (lihui0791nc@126.com)
the First Affiliated Hospital of Nanchang University

Research Article

Keywords: Biomarkers, Immune Infiltration, STAT1, IFI44L

Posted Date: May 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1626248/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: Dermatomyositis (DM) is a rare autoimmune disease, which occurs with poorly understood pathogenesis. Unclear pathogenesis and similar symptoms with other inflammatory myopathies make its diagnosis hard. The study was designed to provide more details about the pathogenesis of DM, trying to identify potential biomarkers, biomarkers associated diseases and drugs or DMs diagnosis and treatment.

Method: GSE142807, GSE1551, GSE5370, GSE39454, GSE48280 and GSE46239 datasets were selected in our study from Gene Expression Omnibus (GEO). GEO2R was used to identify differentially expressed genes (DEGs) between DM samples and controls. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and Gene set enrichment analysis (GSEA) were conducted to identify enrichment pathways. We used Cytoscape to build up the protein-protein interaction (PPI) network. Weighted gene co-expression network analysis (WGCNA) was used to verify the preceding analysis. CIBERSORT tool was applied to analyze the immune cell infiltration patterns of DM. The receiver operating characteristic (ROC) analysis was applied to explore the crucial genes. If icaiclsign or icanceot or DMs diagnosis. DisGeNET database was used to identify associated diseases and Drug Signatures database (DSigDB) was used to identify associated molecules.

Results: DEG analysis identified 31 upregulated genes. KEGG analysis indicated that the DEGs were enriched in pathways associated close with viruses. Eight hub genes were obtained through PPI network. STAT1, IFI44L and DDX58 were identified as crucial from eight hub genes after verification of WGCNA. And we found a significantly differential infiltration of regulatory cells and macrophages in DM samples. ROC analysis demonstrated that the 3 genes had excellent clinical values and their combination model had an AUC = 1, potential to be biomarkers for DM’s diagnosis. According to the gene-disease and gene-drugs analysis, Influenza and several cancers including liver carcinoma associated close to the 3 crucial genes and several drugs like prenylamine HL60 UP and suloctidil HL60 UP were identified.

Conclusions: The study identified three potential biomarkers (STAT1, IFI44L and DDX58) and revealed the immune infiltration of DM and explored out several diseases including cancer and potential drugs for DM, which might guide future diagnosis and treatment for DM.

Introduction

Dermatomyositis (DM), characterized by progressive muscle weakness and skin rash, is a rare autoimmune disease belonging to idiopathic inflammatory myopathies [1–3]. Patients with DM might also suffer from edema and joint pain. Its clinical manifestation is similar to other inflammatory myopathies (like polymyositis), which makes it hard to diagnose. It occurs in patients at any age with poorly understood pathogenesis. Additionally, DM has also been connected to several complications including interstitial lung disease (ILD) and various cardiac abnormalities [4, 5]. And it has also been reported that DM might be associated with malignancies such as ovarian, lung, pancreatic, stomach and colorectal cancer [6]. Its occurrence and progress could be attributed to many factors including genetic element and mainly immune element. According to previous studies, some viruses (e.g., infections might also participate) in DM’s pathogenesis [7, 8]. From the perspective of immune elements, it’s believed that DM is mainly mediated by humoral immunity where endomysial capillaries are targeted as antigen wrongly, resulting in recruitment of complement system and membrane attack complex (MAC) [9–12]. And people with specific human leukocyte antigen (HLA)-DR molecules are predisposed to developing DM [13]. Several aspects like cancers, immune system sdis or ders and ≥ ti cific r have been found to be correlated with DM occurrence and progress [6, 9–13], while the pathogenesis is still unclear, leading to the dilemma of DMs diagnosis → somedegree. Therefore or e, it is necessary for us to explore the internal mechanisms and identify some biomarkers for the diagnosis of DM.

Several bioinformatic studies have been conducted focused on DM and several potential biomarkers have also been identified like ISG15 and CXCL10[14, 15]. Not limited to biomarkers, we also managed to identify the crucial hub genes associated diseases and mo ≤ cedrugsregat \in gthe \geq \# s expression.

In this study, GEO database and GEO 2R were used to identify differential expressed genes (DEGs). Protein-Protein-Interaction network was built to identify hub genes from DEGs. WGCNA was used to verify three potential biomarkers (STAT1, IFI44L, DDX58). DisGeNET database and Drug Signatures database were used to identify biomarkers associated diseases and mo ≤ cedrugs. And we also conducted im\in f< rationanalisysbyCibers or t. This study was design \# d → provsom s future diagnosis and treatment.

Results

Identification and analysis of DEGs in datasets.

DEGs between DM samples and controls in each dataset were screened out by GEO2R based on adjusted P-value < 0.05 and |logFC|>1. After obtaining the intersection DEGs of six datasets (GSE142807, GSE1551, GSE5370, GSE39454, GSE48280 and GSE46239) by Venn diagram (Figure1, A-B), we obtained 31 DEGs. And they are all upregulated in the DM samples. Then, we conducted GO and KEGG enrichment analysis for them. GO enrichment analysis was used to explore the DEGs’ potential mechanisms from the perspective of molecular function, biological process and cellular component categories. And the results revealed that the NO.1 high-correlated pathway of each module were CXCR chemokine receptor binding; defense response to virus; blood microparticle (only 1 pathway was obtained) respectively (figure1, C). The results of KEGG showed that the top 3 high-correlation pathways are Influenza A, Hepatitis C and Measles (figure1, C). And the circus presented the correlations between the 31 DEGs and the related biological processes at the shared term level (figure1, D).

GSEA analysis.
GSEA was used to analyse enriched KEGG pathways between DM groups and controls in six datasets. And 2 KEGG pathways (cytokine cytokine receptor interaction and cell adhesion molecules cams) that expressed significantly higher in disease groups were exhibited by the figure 2.

Identification and analysis of hub DEGs

STRING database was used to analyze the 31 DEGs and the analysis results were imported into Cytoscape software. MCODE plugins in Cytoscape was used to build up protein-protein interaction network, aimed at searching for the DEGs with high correlation with each other. Therefore, we obtained two modules, characterized by red and blue (figure3, A). According to the results of ClueGo analysis (figure 3, B-D), it suggested that the top 3 high-correlated pathways were "type II interferon signaling way", "negative regulation of viral genome replication" and "regulation of type II interferon production" respectively, implying the immune elements especially type 1 interferon takes up an important position in DM. Subsequently, the Cytohubba of Cytoscape was used to predict hub gene. Shown as the upset Venn map (figure4, A), we obtained 8 hub genes: IFIT3, OAS3, STAT1, MX2, IFI44L, DDX58, MX1, RSAD2. Their expression levels in six datasets were exhibited by heatmap (figure4, B). GeneMANIA database was utilized to operate co-expression analysis, trying to reveal the 8 hub genes correlation (figure4, C). Tounderstand the ernalrelatedmechanismofthehub ≥ s, we applied GO and KEGGenrichmentanalysisaga ∈ . The result turned out that the top 3 high correlation pathways were "Hepatitis", "Influenza" and "Measles". We further investigated the difference of 8 hub genes expression level in each dataset, which was demonstrated by volcano maps. The volcano maps are put in figure S1.

Construction of a ceRNA network

MiRNA and LncRNA play an important role in regulating the protein coding genes. To better understand the internal relation between the expression of hub genes and the regulation of miRNA and LncRNA, we built a mRNA- miRNA- LncRNA ceRNA network by Starbase database and Power BI (https://powerbi.microsoft.com/zh-cn/) (figure5). It turned out that STAT1, DDX58, IFI44L, OAS3 and IFIT3 associated with quit a lot of RNA, which might imply they were crucial regulating genes.

Verification of analysis by weighted gene co-expression network analysis (WGCNA) of GSE1551 and GSE142807

GSE1551 and GSE142807 datasets were selected to operate WGCNAs verification. 22 modules were identified in GSE1551 (Figure S2, C) and 6 modules were identified in GSE142807 (Figure S3, C). In the eigengene adjacency heatmap, the red module represents the most positive correlation with occurrence of DM and the green module represents the most negative correlation with occurrence of DM. Analysis results of GSE1551 showed that the 8 hub genes all appeared in the "royalblue" module that is positively associated with the occurrence of DM (figure S2, D). According to the analysis results of GSE142807, the gene DDX58, STAT1 and IFI44L all appeared in the "turquoise" module that is positively related to the occurrence of DM (figure S3, D). The analysis results maps of GSE1551 and GSE142807 were put in Figure S2 and Figure S3 respectively. DDX58, STAT1 and IFI44L appeared in both datasets analysis, which implied that they took up an important position.

Immune cell infiltration

CIBERSORT software was used to operate immune cell infiltration assessment for GSE1551 dataset. According to the analysis result showed by heatmap and box plots (Figure6, A-B), T regulatory cells, Macrophage M0 and eosinophils associated significantly negative with the occurrence and progress of DM. While Macrophage M1 and Macrophage M2 correlated significantly positive with the occurrence and progress of DM. We also conducted Principal Component Analysis (PCA) between controls and DM groups (Figure6, C). The result indicated that DM groups had a different distribution pattern from the normal groups. Not limited to the correlation between immune cell and DM, we also operated the correlation analysis between 22 types of immune cells, which was showed as picture (figure 6, D). It turned out that Macrophages M1 and Macrophages M2 correlated significantly positive with each other. Besides, B cells naive and B cells memory correlated significantly negative with each other. And Macrophages M0 and Macrophages M2 also correlate significantly negative with each other. Additionally, we further operated immunology analysis of each hub gene (STAT1, IFI44L, DDX58) to realize the correlation between each hub gene and the expression level of different types of immune cell. And the result was put in Figure S4.

Diagnosis significance of hub genes.

Aimed at exploring the clinical significance of three hub genes, the ROC analysis was conducted (Figure7, A-E). The result showed that each gene from the six datasets all had an AUC>0.85, which is indicative of a relatively good ability to predict the occurrence of DM. Besides, the STAT1 in five (except GSE46239) datasets all had an AUC>0.95, indicating that STAT1 had very excellent abilities in clinical predicting. And the AUC of the 3 hub genes’ combination model was all equal to 1, indicating that the 3 genes combination model had extremely significant clinical value.

Identification of relative diseases.

The DisGeNET database was applied to identify the diseases associated with the expression of each hub gene. It turned out that Influenza is the disease associating strongest with the 3 hub genes, having the correlation index of 0.33, 0.3 and 0.4 for STAT1, IFI44L and DDX58 respectively. And several cancers (such as liver carcinoma, primary malignant neoplasm, neoplasms and malignant neoplasms) also took up a large part of the predicted diseases, among which the liver carcinoma associated closest with STAT1, having the biggest correlation index of 0.4. Besides, autoimmune diseases and Systemic Lupus Erythematosus also correlated with the 3 genes. The results were showed by the Table 1.

Identification of candidate drugs
We applied DSigDB on Enrichr database to explore the drugs that affect the expression of hub genes. The results revealed the relative drugs of each hub gene based on P value. And we obtained their overlapping drugs, which might affect the DM’s pathogenesis mechanism. We got 5 kinds of overlapping drugs, they are prenylamine HL60 UP, suloctidil HL60 UP, Tetradioxin CTD 00006848, estradiol CTD 00005920 and cyclosporin A CTD 00007121, in an order of increasing P value. And their P values were all smaller than 0.05, indicating their effective statistical significance. The results were demonstrated by the Table 2.

**Discussion**

DM, as a rare autoimmune disease, we still lack of relatively comprehensive knowledge of its pathogenesis, which also contributes to the difficulty of accurate diagnosis to some degree. Therefore, it is urgent to explore internally potential mechanisms of the disease to understand its pathogenesis. As what has been mentioned above, here we utilized bioinformatic analysis to identify the hub genes and its relative pathways aiming at discovering an extraordinary method for predicting and diagnosing the complicated disease.

According to GO and KEGG analysis of 31 DEGs (Fig. 1, C-D) and hub genes (Fig. 4, D), pathways including “response to virus”, “defense response to virus”, “type I interferon signaling pathways”, “Influenza A”, “Hepatitis C” and “Measles” took up an important position. Enrichment pathways were related close to response to viruses and regulation of type I interferon. It has been summarized that type I interferon was first found to be crucial in response to viral diseases and later emerged as an important mediator in many autoimmune diseases including DM [16]. Type I IFN could involve in the JAK-STAT signaling pathway to participate in many immunological responses like defending against viral pathogens like HBV (hepatitis B virus) and HCV (hepatitis C virus) [13, 17–20]. Patients with DM have upregulated level of type I interferon in their muscle tissues, skin tissues and peripheral blood, and type I interferons signaling pathway also plays an important role in DMs pathogenesis [16]. Type I IFNs exact mechanisms in DM are still unclear, but it might be correlated to upregulation of Interferon-inducible gene. Several previous studies pointed that increased expression of type I IFN induced genes in DM patients blood cells and muscle tissues usually associated with disease activity [21–23]. Another study suggests that Caucasian JDM patients who are positive for DQA1*0501 express much higher in some IFN-inducible genes compared to controls [21]. DQA1*0501 belongs to HLA class I alleles which are genetic risk factors for DM. People that are positive for this gene are predisposed to develop DM. It has been reported that more than 85% JDM patients tested positive for expression of DQA1*0501, in contrast with controls’ 25% [22]. Moreover, in another study, a large cohort of IM patients was investigated, it demonstrated that 14 genes expressed significantly higher in adult DM patients, of which 12 genes were inducible by type I IFN. And adult DM patients seemed to express higher level of IFN-inducible genes compared to other IM patients [23]. Why do patients with DM have higher level of type I IFN in their bodies? The exact mechanism is still unclear, but it might be associated with the induction of Toll-like receptors (TLR). Previous study demonstrated that signaling through DNA-sensing TLR9 could lead to type I IFN’s potent production [24]. DM’s occurrence has already been correlated with viral infections including parvovirus B19, coxsackie virus, polyomaviruses, Epstein-Barr virus (EBV), viral hepatitis and influenza [25]. Although the exact mechanism is still largely unknown, there are many cases connecting viruses and DM. For example, it has been reported that a 33-year-old European American female developed DM after treatment of doxycycline and levofloxacin and the patient had also been infected by Epstein-Barr viruses [26]. There was also a related report that a 40-year-old man`s DM condition exacerbated into quadriparesis after inoculated with H1N1 vaccine [27]. Interestingly, patients infected by Covid-19 were also predisposed to develop symptoms like myositis [28].

After verification of WGCNA, STAT1, IFI44L and DDX58 were considered as the crucial hub genes in our study. And their combination model had an extremely excellent clinical value.

STAT1 gene belongs to a big family called signal transducers and activators of transcriptions genes, which are specifically activated by cytokines and growth factors. STAT1 was first mentioned by its participation in response to IFN-α and IFN-γ[32–33]. And STAT1 was also IFN-inducible gene [16], having a strong correlation with IFN. According to a previous study, 57 muscle biopsies (10 DM, 10 PM and 37 controls) were included to perform an immunohistochemical analysis. It turned out that all of the 10 DM patients expressed high level of STAT1 [29]. Besides, in their experiments of cultured muscles in vitro, different treatments including IFN-γ, epidermal growth factor, platelet-derived growth factor and interleukin-2 were used to stimulate the cultured muscles. Finally, STAT1’s activation and transcription in myotubes and rhabdomyosarcoma cells were selectively promoted only by IFN-γ[29]. IFI44L (interferon-induced protein 44-like) gene, like its name, the genes transcription could be induced by IFN, and its also IFN-inducible gene. IFI44L seems to be related to the response to viral infection. It has been reported that the expression of IFI44L would increase when treated with IL-28A and IFN-α to inhibit HCV replication [30]. Additionally, according to another study, increased IFI44L expression level could also promote to inhibiting the propagation of Japanese encephalitis virus (JEV) infection in neurons, exhibited by their receptor interacting serine / threonine - protein kinase 3 (RIPK3) knockout murine model [31]. Especially, the gene connected extremely close with type α interferon signaling pathway. As what’s known, there is just one study connecting DM with IFI44L [32], they selected 5 IFN-stimulated genes including IFI44L as a biomarker to evaluate the DM’s disease activity. While IFI44L has already been noticed because of its important role in the type 1 interferon signaling pathway.

DMs development associated strongly with complicated interactions of immune system. Based on our analysis of immune infiltration cells, T regulatory cells expression level in DM patients was lower than the level in healthy people. T reg cells could alleviate the inflammatory reaction in the body and take part in many immune processes to achieve balance between pro-inflammation and anti-inflammation. It has already been reported that T reg cells associated with DM, and the number of T reg cells in peripheral blood of DM patients was inversely associated with the DM activity [33]. Overmuch inflammation due to reduction or suppression of Treg cells sometimes is the etiology of some disorders. For example, graft versus host disease after haemopoietic stem cells transplantation derived from the grafts immune cells attack against normal host cells, during which the decreased number of Treg cells contributed to increased inflammation [34]. The key point is the alleviated ability to mitigate the inflammation caused by many reasons, which might also occur in the pathogenesis of DM. Not limited to Treg cells, eosinophils level is also lower in DM patients than that of healthy people. According to a previous study of eosinophils in biopsies of “interface dermatitis”, its usually rare in DM patients that eosinophils’ number increases [35]. Therefore, tissue eosinophilia is thought to be unusual in skin biopsies of autoimmune connective disease, traditionally. While there was also study reporting that eosinophil-mediated injury might occur in a wider spectrum (such as idiopathic inflammatory myopathies) not limited to eosinophilic myositis [36], which implied that some DM cases
could occur caused by increased eosinophil infiltration to some degree. What's more, Cibersorts analysis indicated that M2 macrophages infiltrated significantly higher in DM samples and M1 macrophages also infiltrated higher in DM samples. Its reported that macrophages infiltration in the muscle associated closely with the DMs severity [37]. The macrophages are important antigen presenting cells in human's immune system, participating in many immune responses. Moreover, previous study also reported that macrophages contributed to the pathogenesis of DM by presenting T cells antigens, clearing necrotic muscle fibers and generating various type of chemokines and cytokines [38]. Clearly, M1 macrophages are mainly involved in proinflammatory responses and release many proinflammatory cytokines including IL-12, IL-6 and tumor necrosis factors (TNF). While M2 macrophages mainly exert their roles in anti-inflammatory responses and repairing damaged tissues [44–45].

According to the results of gene-diseases analysis, the 3 hub genes associated with several kinds of diseases like malignancy, autoimmune disease and several kinds of virus diseases. For example, the 3 genes all have relatively stronger association with influenza and virus diseases. Some infectious agents have already been associated with the induction or exacerbation of DM in genetically predisposed people, leading to a secondary "infectious" vasculitis [39]. Not limited to viral Influenza, several kinds of malignancy diseases also take up an important part. It's been a long time since the association between DM and malignancy was described. In particular, DM has already been correlated to the occurrence of tumors in ovaries, lungs, pancreas, and stomach as well as non-Hodgkin lymphoma and colorectal cancer [40]. Besides, patients diagnosed as DM have relative higher risk of suffering from malignancies within a year from the onset of myopathy. And the extent of DM symptoms were also associated with the progression of metastasis [41]. Some report also demonstrated that the exacerbation of DM's condition related to the recurrence of cancers [42]. Moreover, the 3 hub genes also associated with autoimmune diseases and Systemic Lupus Erythematosus (SLE). Actually, DM was a kind ofautoimmune disease and DM patients always have the presence of auto-antibodies within their bodies. SLE was a chronic inflammatory disease, classified into autoimmune connective tissue diseases and affected multiple organs and systems [43]. Both DM and SLE were caused by similar genetic, hormonal and environmental factors [44].The two diseases had some subtle associations. They both had several common features in the skins, manifested as erythematous eruptions, photosensitivity and the Koebner phenomenon [45, 46]. Recent studies also reported that DM patients and SLE patients had extraordinarily similar gene expression profiles. For example, it's been reported that interferon (IFN)-related genes are over-expressed in the skin, muscle and peripheral blood of both DM patients and SLE patients [47, 48].

As so far, several chemical agents and drugs have been used as therapeutic medicine for DM. For example, cyclophosphamide, azathioprine, mycophenolate mofetil and rituximab have been used in combination with corticosteroid for the treatment of DM [49–52]. Additionally, several other drugs were also identified, which could regulate the expression of the 3 genes. Maybe they are potential to be the alternative therapeutic drugs for DM or the genes' related diseases.

**Conclusion**

Based on integrated bioinformatical analysis, we identified 8 hub genes that associated strongly with the occurrence of DM, of which three hub genes (STAT1, IFI44L and DDX58) were potential to be the biomarkers of DM. And they correlated closely to viral infection and pathways about type I interferon. According to the immune infiltration analysis, T regulatory cells and Macrophages may be critical in remission of DM. Besides, we also explored the diseases associated with three hub genes and some drugs regulating the 3 hub genes' expression which might be potential to be the therapeutic drugs for DM.

**Methods**

**Access to GEO datasets**

Utilize GEO (http://www.ncbi.nlm.nih.gov/geo) database to obtain genes data. GSE142807, GSE1551, GSE5370, GSE39454, GSE48280 and GSE46239 datasets were utilized. The GSE1551, GSE5370 series on the GPL96 platform (Affymetrix Human Genome U133A Array). The GSE142807 series on the GPL17692 platform (Affymetrix Human Gene 2.1 ST Array), the GSE39454, GSE46239 series on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) and the GSE48280 series on the GPL626244 platform (Affymetrix Human Gene 1.0 ST Array).

**Identify DEGs by GEO2R**

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) was applied to screen out the DEGs between DM cases and healthy controls. GEO2R is a powerful online analysis tool which identified DEGs through the comparative analysis of the differential classifications. As for the genes without a corresponding symbol, we chose to abandon them. Besides, the genes having more than one probe set also have the same fate. To identify significant DEGs, we utilized Venn online tool (http://bioinformatics. psb.ugent.be/webtools/Venn/) to draw a Venn map so that overlapping DEGs among six datasets could be obtained for subsequent analysis.

**Enrichment analysis through gene set enrichment analysis (GSEA)**

We used GSEA version 4.1.0 software (http://software. broadinstitute.org/gsea/msigdb) to analyze genes' function in means of the GSEA website MSIGDB database. And enrichment analysis was carried out using the default weighted method. Set the random combination as 1000 times. KEGG pathway enrichment analysis and GSEA analysis were performed. FDR<0.25, NOM p-value<0.05 and |NES|>1 could be considered significant enrichment.

**Enrichment analysis, construction of Protein-Protein interaction network and analysis of DEGs.**

To annotate functions of the DEGs identified by above comparative analysis, we conducted KEGG and Go enrichment analysis via R language. STRING website was used to analyze DEGs, after which the analysis result was imported into Cytoscape website. MCODE plugin in Cytoscape website was responsible for the visualization of the analysis result to build up a protein-protein interaction network. And Clue GO plugin in Cytoscape was used to conduct pathway analysis behind protein-protein network. Cytohubba plugin in Cytoscape was applied to predict hub genes. Besides, we drew volcano maps with the assistance...
of volcano plotting tool. Expression heatmap of hub genes in different series was drawn using drawing tool (http://soft.sangerbox.com). It is the correlation analysis between genes-genes and series-series that was applied in the tool. Besides, GENEMANIA (http://genemania.org/search/) was used to construct a gene-gene interaction network.

**Construct weighted gene co-expression network analysis (WGCNA)**

We applied WGCNA package in R to construct a co-expression network targeting DEGs. The weighted adjacency matrix was built, defining a correlation power (soft threshold parameter) which could reflect strong relations between genes and penalize the weak correlation. Next, the adjacency relationship was transformed into a topological overlap matrix (TOM) to measure each others network connectivity of genes. TOM summed up adjacent genes to obtain the network gene ratio and calculate the corresponding dissimilarity degree. The average linkage hierarchical clustering based on TOM dissimilarity measurement was used to classify gene modules with similar expression profiles, which were represented by clustering trees branches and different colors. So, it was better to construct modules’ relationship, calculate the correlation degree between gene modules and phenotypes and recognize the gene modules associated with clinical traits.

**Immune related research**

We uploaded the gene expression matrix data into CIBERSORT (https://cibersort.stanford.edu/) obtaining the immune cell infiltration matrix aimed at evaluating the abundance of immune infiltrates. Subsequently, the “corrlot” package was applied to draw a related heatmap, trying to visualize the correlation of 22 types of immune cell infiltration. And the “ggplot” package was utilized to operate PCA clustering analysis in which we drew a two-dimensional clustering map and box plots to visualize the difference of immune cell infiltration.

**ROC analysis**

The multivariate modelling with combined hub genes was used to demonstrate the capacity of selected hub genes serving as biomarkers with high specificity and sensitivity for DM diagnosis through visualization tool. One data was selected as training (GSE1551) and others (GSE142807, GSE5370, GSE39454, GSE48280 and GSE46239) were selected as validation samples, the process was conducted repeatedly. Then, the receiver operating characteristic curves were plotted, and area under curves (AUC) was calculated for each hub gene model to evaluate the performance of each model using the R packages “pROC”. AUC >0.9 is indicative of the good effect of models.

**Association analysis of drugs and diseases**

Drug molecules that regulate the three hub genes expression level was identified based on the p values on Drug Signatures database (DSigDB) via Enrichr database. And we obtained the overlapping drugs which affect all the three hub genes expression. We used DisGeNET database (https://doi.org/10.1016/j.csbj.2021.05.015.) to analyse the association between each hub gene and the potential diseases. And the relevance index was used to assess their association.

**Abbreviations**


**Declarations**

**Acknowledgements**

No applicable.

**Funding**

This work was supported by "Science and Technology plan of Jiangxi Provincial Health and Family Planning Commission" (Grant Number: 20182930) and “TCM scientific research project of Jiangxi Provincial Health and Family Planning Commission” (Grant number:2018B038).

**Availability of data and material**

The datasets applied in the study are available in the GEO database. GEO database is a publicly free database which stores many gene functions and expressions. The working links are as follow:


**Ethics approval and consent to participate**

No applicable.
Competing interests

No potential conflicts exist between authors.

Consent for publication

All the authors agree the publication.

Authors contributions

Haonan Zhou was responsible for the writing of the whole article. Jie Peng was responsible for all the bioinformatic analysis and generation of all the pictures. And both authors could prove that they have read the article manuscript. Rui Wu, Hui Li and Zhen Zong provided leading suggestions and lots of supports for the study.

References


### Tables

Tables 1 to 2 are available in the Supplementary Files section

### Figures
Figure 1

DEGs between DM samples and controls and their enrichment analysis.

A: Venn Diagram of down-regulated DEGs between DM and controls. It turned out that there was no common intersection of 6 datasets.

B: Venn Diagram of up-regulated DEGs between DM and controls. 31 DEGs were in the intersection of 6 datasets.

C, D: GO and KEGG pathway enrichment analysis of 31 DEGs. It turned out that enrichment pathways associated with virus closely, including “Influenza A”, “Hepatitis C”, “Measles” and “response to virus”.

CXCL10, IFI44, IFI44L, IFIT3, STAT1, MX1, DDX58, HERC5, IFI16, OAS3, TNFSF10, PARP12, USP18, CXCL11, RSAD2, IFI27, XAF1, LGALS3BP.
Figure 2

Gene set enrichment analysis (GSEA) of DEGs among six datasets.

Gene set enrichment analysis (GSEA) was used to analyze KEGG pathways enriched in different datasets. And two pathways (Cytokine cytokine receptor interaction and Cell adhesion molecules CAMs) were obtained to show. A. GSE1551 B. GSE142807 C. GSE5370 D. GSE39454 E. GSE48280 F. GSE46239.
Figure 3

The Protein-Protein interaction network and GO terms showed interaction network using Cytoscape’s plug-in ClueGO.

A: Protein-Protein interaction network of 31 DEGs. Red and blue modules represented different modules with high intra-module correlation.

B, C, D: ClueGo enrichment analysis conducted by ClueGo plugin of Cytoscape. It indicated that pathways associated with type II interferon including "type II interferon production" and "regulation of type II interferon production" and response to virus including "negative regulation of viral genome replication", "negative regulation of viral process" and "defense response to virus"
Cytohubba plugin of Cytoscape predicted hub genes and analysis of hub genes.

A: Upset Venn diagram applied to demonstrate the predicted hub genes

B: Heatmap demonstrated the different expression levels of 8 hub genes in 6 datasets.

C: The gene-gene interaction network for 8 hub genes was analysed using the GeneMANIA database. The 20 most frequently changed neighboring genes are shown as nodes and colors represent the possible functions of the respective gene. D: GO and KEGG pathway enrichment analysis of 8 hub genes demonstrates major important pathways associated with hub genes. It showed that "defense response to virus", "response to virus", "type I interferon signaling pathway", "Hepatitis C", "Influenza" and "Measles" correlated strongly with the 8 hub genes.

E: GO and KEGG pathway enrichment analysis of 8 hub genes showed the correlation between two hub genes.
Figure 5

IncRNA-mRNA-miRNA interaction network about 8 hub genes. It turned out that STAT1, DDX58, IFI44L, OAS3 and IFIT3 associated with quit a lot of RNA, which might imply they were crucial regulating genes.
Figure 6

Immune infiltration analysis of GSE1551 via Cibersort software. It exhibited that T regulatory cells and Macrophages M2 infiltrated significantly differential between DM samples and controls.

A: Heatmap showing different expression levels of 22 types of immune cells between normal and DM groups.

B: Box plots presented different expression levels of 22 types of immune cells between normal and DM groups. Horizontal axis represented immune cell types. Vertical axis represented correlation index. And blue color was indicative of normal groups, yellow color is referred to DM groups.

C: Principal Component Analysis. It showed that there was existing obvious difference between DM samples and controls.

D: Correlation matrix of all 22 types of immune cells. Both horizontal and vertical axes represented immune cell types. Besides, higher, lower, and same correlation levels were displayed in red, blue, and white respectively.
Figure 7

Diagnostic performance of 3 hub genes and their combination model. The diagnostic performance was conducted based on selecting one dataset (GSE1551) as training sample and 5 other datasets as validation samples. An AUC>0.9 indicated that the model has a good fitting effect. Here are the diagnostic values of DDX58, STAT1 and IFI44L as well as the combination model in dataset GSE1551 (A), GSE39454 (B), GSE5370 (C), GSE142807 (D), GSE46239 (E), GSE48280 (F).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1Thediseasesassociatedwiththe3genes1.pdf
- Table2Thedrugsaffectingtheexpressionofthe3genes.pdf
- figureS1.jpg
- figureS2.jpg
- figureS3.jpg
- figureS4.jpg