Exploration of Dilated Cardiomyopathy for Biomarkers and Immune Microenvironment: Evidence From RNA-seq

Chenggang Fang
Wuhan University
Zhan Lv
Wuhan University
Zhimin Yu
Wuhan University
Kexin Wang
Wuhan University
Chengkai Xu
Wuhan University
Yixuan Li
Wuhan University
Yanggan Wang (ygwang2018@yeah.net)
Zhongnan Hospital of Wuhan University

Research Article

Keywords: Dilated Cardiomyopathy, DCM, Hub genes, Biomarker, Immune microenvironment

Posted Date: January 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1226732/v1

Version of Record: A version of this preprint was published at BMC Cardiovascular Disorders on July 18th, 2022. See the published version at https://doi.org/10.1186/s12872-022-02759-7.

Abstract

Background: The pathogenic mechanism of Dilated Cardiomyopathy (DCM) remain to be defined. This study aimed to identify hub genes and immune cells that could serve as potential therapeutic targets for DCM.

Methods: We downloaded four data sets from the Gene Expression Omnibus (GEO) database, GSE141910, GSE3585, GSE42955 and GSE79962. Weighted gene coexpression network analysis (WGCNA) and differential expression analysis were performed to identified genes panel related DCM. Meanwhile, CIBERSORT algorithm was used to estimate the immune cells in DCM tissues. Multiple machine learning approaches were used to screen the hub genes and immune cells. Finally, the diagnostic value of the hub genes was assessed by receiver operating characteristic (ROC) analysis.

Results: FRZB and EXT1 were identified as hub biomarkers, and the ROC curves suggested an excellent diagnostic ability of above genes for DCM. In addition, B cells naive was up-regulated in DMC tissues, while Eosinophils, Macrophages M2, and T cells CD4 memory were down-regulated in DMC tissues.

Conclusion: These results indicated that FRZB and EXT1 could be used as promising biomarkers, and Eosinophils, Macrophages M2, T cells CD4 memory resting and B cells naive also may affect the occurrence of DCM.

Introduction

Dilated cardiomyopathy(DCM) is defined as left ventricular(LV) dilatation and left ventricular systolic dysfunction in the absence of abnormal loading conditions (hypertension, valve disease) or coronary artery disease sufficient to cause global systolic impairment.[1] DCM is one of the most common cause of heart failure and its prevalence is range from 1:250 to 1:2500 in the general population.[2, 3] If the patients do not get timely treatment ,the 1-year survival rate is 70%-75% and the 5-year survival rate is as low as 50%.[4] Causes of DCM can be classified into two categories, genetic and nongenetic but overlaps do exist in the two categories. The most common pathologies presently known underlie reactive changes are inflammation (viral myocarditis or autoimmune disease), nutritive-toxic influences (alcohol, drugs, chemotoxines), and metabolic disorders. These changes finally lead to a remodeling of the myocardium.[5] But the molecular mechanism underling the remodeling is a complex network of cellular signaling pathway and is still incompletely understood so far. Genetic testing for cardiovascular disease is becoming more and more common in recent years. For hypertrophic cardiomyopathy, the rate of detection is around 60-70%, but the detection rate of the testing for dilated cardiomyopathy is much lower than 60%. It has been shown that most of the mutations occur in genes encoding of the structural proteins of the cytoskeleton, the cardiomyocytic sarcomeres and nuclear envelope proteins.[2, 6, 7] Therefore, it is essential to identify new novel biomarkers significantly correlated with DCM diagnosis to improve the effectiveness of therapeutic approaches.

The combination of myocardial inflammation (myocarditis) and dysfunction is termed inflammatory cardiomyopathy. In patients with recent-onset DCM, identification of myocarditis has important clinical implications due to the high potential for LV recovery. Circulating cardiac autoantibodies are more common in patients with dilated cardiomyopathy and myocarditis than in patients with non-inflammatory heart disease.

Furthermore, in healthy relatives of patients with dilated cardiomyopathy, serum anti-heart autoantibodies are independent predictors of disease progression[8]. Both innate and adaptive immunological aspects may play a role in affecting outcomes in laboratory animals and patients with viral myocarditis.[9] The failing myocardium can provide signals to assist in immune cell infiltration via upregulation or secretion of various cytokines such as P-selectin, e-selectin, intracellular cell adhesion molecule-1, and vascular cell-adhesion molecule-1, which allows transendothelial migration of a range of immune cells into the myocardium, including B cells, T cells, natural killer cells, monocytes, and platelets.[10, 11] If we can find out the most relevant type of immune cells, a novel treatment may be possible by interfering a particular type of immune cell.

In this research, we examined 4 GEO datasets and found 3743 significant differentially expressed genes (DEGs) between normal and DCM samples. A weighted gene coexpression network analysis (WGCNA) was performed to evaluate the key module correlated with DCM. LASSO 10-fold cross-validation was used to further knockout redundant genes, and 38 potential genes were finally screened out. SVM machine learning, random forest tree and logistic analysis were used to conduct in-depth screening and finally FRZB and EXT1 were identified as hub biomarkers. In the screening set, ROC and difference analysis were performed on the above two genes. The results showed that the two genes had good predictive performance in the screening set. Random forest tree analysis and Wilcoxon test were performed to find core immune cells that may affect the occurrence of DCM and finally four types of immune cells were selected.

Materials And Methods

Datasets and data pre-processing

Use the GSE141910 data set based on the GPL16791 platform (166 healthy myocardial tissue samples, 166 DCM samples), which is the original RNA-seq data set downloaded from the GEO database[12], as a screening set.

In addition, GSE3585 data set (GPL96, 5 normal samples, 7 DCM samples), GSE42955 data set (GPL6244, 5 normal samples, 12 DCM samples), and GSE79962 data set (GPL6244, 11 normal samples, 9 DCM samples) are collectively defined as an external verification set, and the sva package[13] is used to perform background correction, normalization and expression calculation on the original data.

Screening of hub biomarkers

Use the limma package [14] to identify differentially expressed genes (DEGs) in the screening set, and select |log2 fold change FC|>0.5 and adj.P.value <0.05 as cut-off criteria. In the WGCNA analysis[15], the entire DEGs are used as input, and the topological calculation is performed with a soft threshold value of 1 to 20. According to the optimal soft threshold value, the relation matrix is converted into an adjacent matrix, then converted into a topological overlap matrix (TOM). Perform average-linkage hierarchical clustering, classify related modules according to TOM, the number of genes in each module is not less than 50, and then merge similar modules. Then use Pearson method to calculate the correlation between the merged module and DCM. Among the core modules screened by WGCNA, perform LASSO 10-fold cross-validation to knockout redundant genes (glmnet package[16]). Subsequently, adopt the support vector machine-recursive feature elimination (SVM-REF)[17, 18] method (take the lowest point feature of RSM), the random forest tree method (take the genes with Top10 weights), and the single-factor logistic regression method (take the Top3 OR values) to screen the de-redundant genes. Ultimately, overlap the genes identified with the above methods to identify the final core gene.

Enrichment analysis

GO enrichment analysis is a commonly used bioinformatics method for searching comprehensive information of large-scale genetic data, including BP, CC, and MF. In addition, KEGG pathway enrichment analysis is widely used to understand biological mechanisms and functions. Furthermore, DO enrichment analysis can explore the diseases in which the relevant genes are mainly involved. GOplot[19] package is used to visualize the GO, KEGG paths, and DO analysis. At the end, use the clusterprofile package[20] and GSVA package[21] to further explore the important signal pathways related to core genes. Download the h.all.v7.4.symbols gene set from MSigDB[22], perform GSEA analysis on the gene set and gene expression matrix to explore the regulatory pathways that may be involved.

Construction of regulatory network

First, use the mirDIP database[23] to predict potential miRNAs targeting hub genes and identify miR regulatory networks. Select the TF-core gene interaction pair with P<0.05 in the TRRUST database[24] to establish an upstream regulatory network. Then search the Comparative Toxicogenomics database[25] for compounds that may be potentially related to core genes. Finally, based on the Network analyst database[26] to realize the visualization of the core gene regulation network.

CIBERSORT algorithm

The CIBERSORT algorithm[27] calculates the proportion of different immune cell types based on the expression levels of immune cell-related genes. Integrate the output results of 22 infiltrated immune cells to generate a matrix of immune cell components for analysis (CIBERSORT package).

Screening of hub immune cells

Use the Wilcoxon test to investigate the differences in the content of immune cells in different tissues. Meanwhile, use the randomForest package to construct a random forest tree of 22 kinds of immune cells,

determine the points with the smallest error, and sort the immune cells according to their importance, then select the immune cells with an importance score greater than 10. Ultimately, overlap the immune cells identified with above method, and screen out the core immune cells that regulate the occurrence of DCM.

Results

Differential expression genes

Differentially expressed genes (DEGs) were analyzed in the screening set, and finally 3743 DEGs were identified (supplementary files.1). The heat map shows the Top20 DEGs (Fig. 1a). The volcano map shows 1861 up-regulated genes and 1882 down-regulated genes (Fig. 1b).

WGCNA analysis

The clinical information and genes were correlated, and WGCNA analysis was performed. The clustering situation of each sample was favorable, with no outlier sample (Fig. 2a), and the optimal soft threshold was determined to be 6 (Fig. 2b). The modules were classified according to the soft threshold and TOM matrix, and the number of genes in each module was not less than 50 (Fig. 2c). The similar gene modules were merged, and 8 modules were finally identified (Fig. 2d). By calculating the correlation between module genes and clinical traits, it was found that the black module containing 1078 genes had the highest positive correlation with the occurrence of DCM (r=0.85), and the red module containing 265 genes had the highest negative correlation with the occurrence of DCM (r=-0.64). Using both as core modules, finally 1343 potential core genes were identified.

Enrichment analysis

In order to explore the potential biological mechanism of DCM, enrichment analysis was performed on 1343 potential core genes. DO analysis revealed the types of diseases that may have common pathogenesis, such as bacterial infectious disease, tuberculosis and sarcoidosis (Fig. 3a). Further GO analysis showed that T cell activation, regulation of immune effector process, positive regulation of leukocyte activation and other processes were significantly enriched (Fig. 3b). In addition, KEGG also described specific pathways, such as Th1 and Th2 cell differentiation, Th17 cell differentiation, Viral protein interaction with cytokine and cytokine receptor, etc. (Fig. 3c). The above results indicate that immune-related factors may affect the occurrence of DCM. Ultimately, GSEA analysis was performed on the gene set and expression matrix, and the results showed that INTERFERON_ALPHA_RESPONSE, INTERFERON_GAMMA_RESPONSE and other pathways were significantly enriched (Fig. 3d). In summary, the strong chain of evidence indicates the important role of immunity in the pathogenesis of DCM.

Exploring of hub biomarkers

First, LASSO 10-fold cross-validation was used to further knockout redundant genes, and 38 potential genes were finally screened out (Fig. 4a). Among the above 38 genes, SVM machine learning method was used to conduct in-depth screening. The results showed that when 19 genes were included, the RMSE

value was the lowest (Fig. 4b). In addition, the random forest tree method was used to rank the weights of 38 genes (Fig. 4c). At the same time, the occurrence of DCM was used as the dependent variable, and logistic analysis was performed. The results of the forest plot showed the OR value and confidence interval corresponding to each gene (Fig. 4d). Finally, the genes identified by the above algorithm were overlapped, and FRZB and EXT1 were identified as hub biomarkers (Fig. 4e).

Validation of hub biomarkers

In the screening set, ROC and difference analysis were performed on the above two genes. The results showed that the two genes had good predictive performance in the screening set: EXT1 (AUC=0.946), which was significantly high expressed in DCM samples; FRZB (AUC=0.985) was also highly expressed in DCM samples (Fig. 5a-b). In the external validation set, the expression of core genes was similar to that in the screening set, which were up-regulated in DCM tissues and also had strong diagnostic performance (EXT1, AUC=0.842; FRZB, AUC=0.954) (Fig. 5c-d). In addition, the regulatory network of the above two core genes was visualized, a TF-mRNA-miRNA network was constructed, and its potential candidate compounds targeting EXT1 and FRZB were predicted to improve the symptoms of DCM patients.

Analysis of differences in immune microenvironment

Considering the important role of immune-related pathways in the occurrence of DCM in gene enrichment analysis (Fig. 2), the CIBERSORT algorithm was used to analyze the content of immune cells in different samples. The histogram shows the overall landscape of immune cell distribution, and the results of heat map show in detail the correlation of 22 types of immune cells. The results of Wilcoxon test analysis showed the difference in the content of immune cells in DCM samples and normal myocardial tissues. In order to identify the core immune cells that change the immune microenvironment in myocardial tissue, random forest tree analysis was performed on 22 immune cells (Fig. 6a-b). Subsequently, the immune cells identified by Wilcoxon test and random forest tree were overlapped, and four core immune cells that may affect the occurrence of DCM were finally identified (Fig. 6c): Eosinophils, Macrophages M2, T cells CD4 memory resting and B cells naive. Among them, only B cells naive was up-regulated in DMC tissues, while Eosinophils, Macrophages M2, and T cells CD4 memory were down-regulated in DCM tissues (Fig. 6d).

Correlation analysis of immune cells and hub biomarkers

In the DCM tissue, the correlation analysis between 22 kinds of immune cells and 2 hub biomarkers was performed. Among them, EXT1 was negatively correlated with NK cells resting, and positively correlated with Dendritic cells resting, Mast cells resting, Eosinophils (Fig. 7A). Meanwhile, Fig. 7b specifically shows the scatter plot of the correlation between EXT1 and core immune cells Eosinophils. In addition, FRZB was positively correlated with Monocytes (Fig. 7c-d).

Discussion

As one of the leading causes of heart failure (HF), DCM is the most frequent indication for cardiac transplantation. The DCM is a final common response of myocardium to a quantity of genetic and environmental insults rather than a single disease entity. The contemporary researches using genetic screening show that up to 40% of DCM is genetically determined[3]. More than 50 genes relating to sarcomeric proteins (MYH7, ACTC1, TNNT2, MYH6, MYBPC3), cytoskeleton (TTN, DES, DMD, FLNC, NEXN, LDB3), ion channels (RYR2, SCN5A), nuclear envelop (LMNA, TMPO) and intercellular junctions have been implicated in DCM.

Except for the genetic factors, an important cause of acquired primary cardiomyopathy is myocarditis, which can lead to inflammatory dilated cardiomyopathy (IDC), a subtype of the primary acquired DCM. Idiopathic-inflammatory, viral or autoimmune-mediated cardiomyocyte destructions mediated via several type of immune cells play an important role in this process. [28]

To investigate potential biomarkers for better detection and therapy, we integrated the gene expression profiles of GSE141910, GSE3585, GSE42955 and GSE79962, which contained 194 DCM samples and 187 normal samples. 1861 up-regulated genes and 1882 down-regulated genes were identified. WGCNA, DO, GO, KEGG enrichment analysis and multiple machine learning approaches were performed to find the hub genes and specific immune cells.

We identified two hub genes, FRZB and EXT1. High expression of two genes was significantly associated with DCM. FRZB (Frizzled Related Protein) function as modulators of Wnt signaling through direct interaction with Wnts. They have a role in regulating cell growth and differentiation in specific cell types. It is reported that FRZB serves as a key molecule in abdominal aortic aneurysm progression[29] and can decreases growth and invasiveness of fibrosarcoma cells[30]. It is also a muscle biomarker denervation atrophy in amyotrophic lateral sclerosis[31]. EXT1 is an endoplasmic reticulum-resident type II transmembrane glycosyltransferase involved in the chain elongation step of heparan sulfate biosynthesis. Diseases associated with EXT1 include hereditary multiple exostoses, non-small cell lung carcinoma and Chondrosarcoma[32-34].

Weighted gene coexpression network analysis (WGCNA) and differential expression analysis were performed to identified genes panel related DCM. Meanwhile, CIBERSORT algorithm was used to estimate the immune cells in DCM tissues. Multiple machine learning approaches were used to screen the hub genes and immune cells. Finally, the diagnostic value of the hub genes was assessed by receiver operating characteristic (ROC) analysis.

Conclusion

In our research, 3743 DEGs were identified in DCM. Multiple machine learning approaches were used to screen the hub genes and immune cells. Two hub genes(FRZB and EXT1) could be used as promising biomarkers, and Eosinophils, Macrophages M2, T cells CD4 memory resting and B cells naive also may affect the occurrence of DCM.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declared no potential competing interests with respect to the research, authorship, and/or publication of this article.

Availability of data and materials

The gene expression profiles of GSE141910, GSE3585, GSE42955 and GSE79962 were downloaded from Gene Expression Omnibus (GEO).

Funding

National Natural Science Foundation of China, 8187021424

Authors' contributions

Conceived and designed the experiments: Chenggang Fang

Analyzed the data: Chenggang Fang,Zhan Lv,Zhimin Yu,Kexin Wang,Chengkai Xu,Yixuan Li

Wrote the paper: Chenggang Fang,Yanggan Wang

All authors have read and approved the manuscript

Acknowledgements

We thank all participants for their supports.

References

1. Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, Dubourg O, Kuhl U, Maisch B, McKenna WJ *et al*: **Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases**. *European heart journal* 2008, **29**(2):270-276.
2. Weintraub RG, Semsarian C, Macdonald P: **Dilated cardiomyopathy**. *Lancet* 2017, **390**(10092):400-414.

3. Hershberger RE, Hedges DJ, Morales A: **Dilated cardiomyopathy: the complexity of a diverse genetic architecture**. *Nature reviews Cardiology* 2013, **10**(9):531-547.
4. Dec GW, Fuster V: **Idiopathic dilated cardiomyopathy**. *The New England journal of medicine* 1994, **331**(23):1564-1575.
5. McKenna WJ, Maron BJ, Thiene G: **Classification, Epidemiology, and Global Burden of Cardiomyopathies**. *Circulation research* 2017, **121**(7):722-730.
6. McNally EM, Mestroni L: **Dilated Cardiomyopathy: Genetic Determinants and Mechanisms**. *Circulation research* 2017, **121**(7):731-748.
7. Olson TM, Michels VV, Thibodeau SN, Tai YS, Keating MT: **Actin mutations in dilated cardiomyopathy, a heritable form of heart failure**. *Science* 1998, **280**(5364):750-752.
8. Caforio ALP, Mahon NG, Baig MK, Tona F, Murphy RT, Elliott PM, McKenna WJ: **Prospective familial assessment in dilated cardiomyopathy: cardiac autoantibodies predict disease development in asymptomatic relatives**. *Circulation* 2007, **115**(1):76-83.
9. Huang C-H, Vallejo JG, Kollias G, Mann DL: **Role of the innate immune system in acute viral myocarditis**. *Basic research in cardiology* 2009, **104**(3):228-237.
10. Wilhelmi MH, Leyh RG, Wilhelmi M, Haverich A: **Upregulation of endothelial adhesion molecules in hearts with congestive and ischemic cardiomyopathy: immunohistochemical evaluation of inflammatory endothelial cell activation**. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2005, **27**(1):122-127.
11. Jaakkola K, Jalkanen S, Kaunismäki K, Vänttinen E, Saukko P, Alanen K, Kallajoki M, Voipio-Pulkki LM, Salmi M: **Vascular adhesion protein-1, intercellular adhesion molecule-1 and P-selectin mediate leukocyte binding to ischemic heart in humans**. *Journal of the American College of Cardiology* 2000, **36**(1):122-129.
12. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M *et al*: **NCBI GEO: archive for functional genomics data sets–update**. *Nucleic Acids Res* 2013, **41**(Database issue):D991-D995.
13. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments**. *Bioinformatics (Oxford, England)* 2012, **28**(6):882-883.
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Res* 2015, **43**(7):e47.
15. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC bioinformatics* 2008, **9**:559.
16. Engebretsen S, Bohlin J: **Statistical predictions with glmnet**. *Clin Epigenetics* 2019, **11**(1):123.
17. Yoon S, Kim S: **AdaBoost-based multiple SVM-RFE for classification of mammograms in DDSM**. *BMC Med Inform Decis Mak* 2009, **9 Suppl 1**:S1.

18. Lin X, Yang F, Zhou L, Yin P, Kong H, Xing W, Lu X, Jia L, Wang Q, Xu G: **A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information**. *J Chromatogr B Analyt Technol Biomed Life Sci* 2012, **910**:149-155.
19. Walter W, Sánchez-Cabo F, Ricote M: **GOplot: an R package for visually combining expression data with functional analysis**. *Bioinformatics (Oxford, England)* 2015, **31**(17):2912-2914.
20. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284-287.
21. Hänzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and RNA-seq data**. *BMC bioinformatics* 2013, **14**:7.
22. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P: **The Molecular Signatures Database (MSigDB) hallmark gene set collection**. *Cell Syst* 2015, **1**(6):417-425.
23. Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild A-C, Tsay M, Lu R, Jurisica I: **mirDIP 4.1-integrative database of human microRNA target predictions**. *Nucleic Acids Res* 2018, **46**(D1):D360-D370.
24. Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E *et al*: **TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions**. *Nucleic Acids Res* 2018, **46**(D1):D380-D386.
25. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ: **The Comparative Toxicogenomics Database: update 2019**. *Nucleic Acids Res* 2019, **47**(D1):D948-D954.
26. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J: **NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis**. *Nucleic Acids Res* 2019, **47**(W1):W234-W241.
27. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA: **Profiling Tumor Infiltrating Immune Cells with CIBERSORT**. *Methods Mol Biol* 2018, **1711**:243-259.
28. Mahrholdt H, Wagner A, Deluigi CC, Kispert E, Hager S, Meinhardt G, Vogelsberg H, Fritz P, Dippon J, Bock CT *et al*: **Presentation, patterns of myocardial damage, and clinical course of viral myocarditis**. *Circulation* 2006, **114**(15):1581-1590.
29. Oh C-K, Ko Y, Park JJ, Heo HJ, Kang J, Kwon EJ, Kang JW, Lee Y, Myung K, Kang JM *et al*: **FRZB as a key molecule in abdominal aortic aneurysm progression affecting vascular integrity**. *Biosci Rep* 2021, **41**(1).
30. Guo Y, Xie J, Rubin E, Tang Y-X, Lin F, Zi X, Hoang BH: **Frzb, a secreted Wnt antagonist, decreases growth and invasiveness of fibrosarcoma cells associated with inhibition of Met signaling**. *Cancer Res* 2008, **68**(9):3350-3360.
31. Kwan T, Kazamel M, Thoenes K, Si Y, Jiang N, King PH: **Wnt antagonist FRZB is a muscle biomarker of denervation atrophy in amyotrophic lateral sclerosis**. *Scientific reports* 2020, **10**(1):16679.
32. Kong W, Chen Y, Zhao Z, Zhang L, Lin X, Luo X, Wang S, Song Z, Lin X, Lai G *et al*: **EXT1 methylation promotes proliferation and migration and predicts the clinical outcome of non-small cell lung**

carcinoma via WNT signalling pathway. *Journal of cellular and molecular medicine* 2021, **25**(5):2609-2620.

33. Guo X, Lin M, Yan W, Chen W, Hong G: **A novel splice mutation induces exon skipping of the EXT1 gene in patients with hereditary multiple exostoses**. *Int J Oncol* 2019, **54**(3):859-868.
34. Delgado MA, Martinez-Domenech G, Sarrión P, Urreizti R, Zecchini L, Robledo HH, Segura F, de Kremer RD, Balcells S, Grinberg D *et al*: **A broad spectrum of genomic changes in latinamerican patients with EXT1/EXT2-CDG**. *Scientific reports* 2014, **4**:6407.

Figures

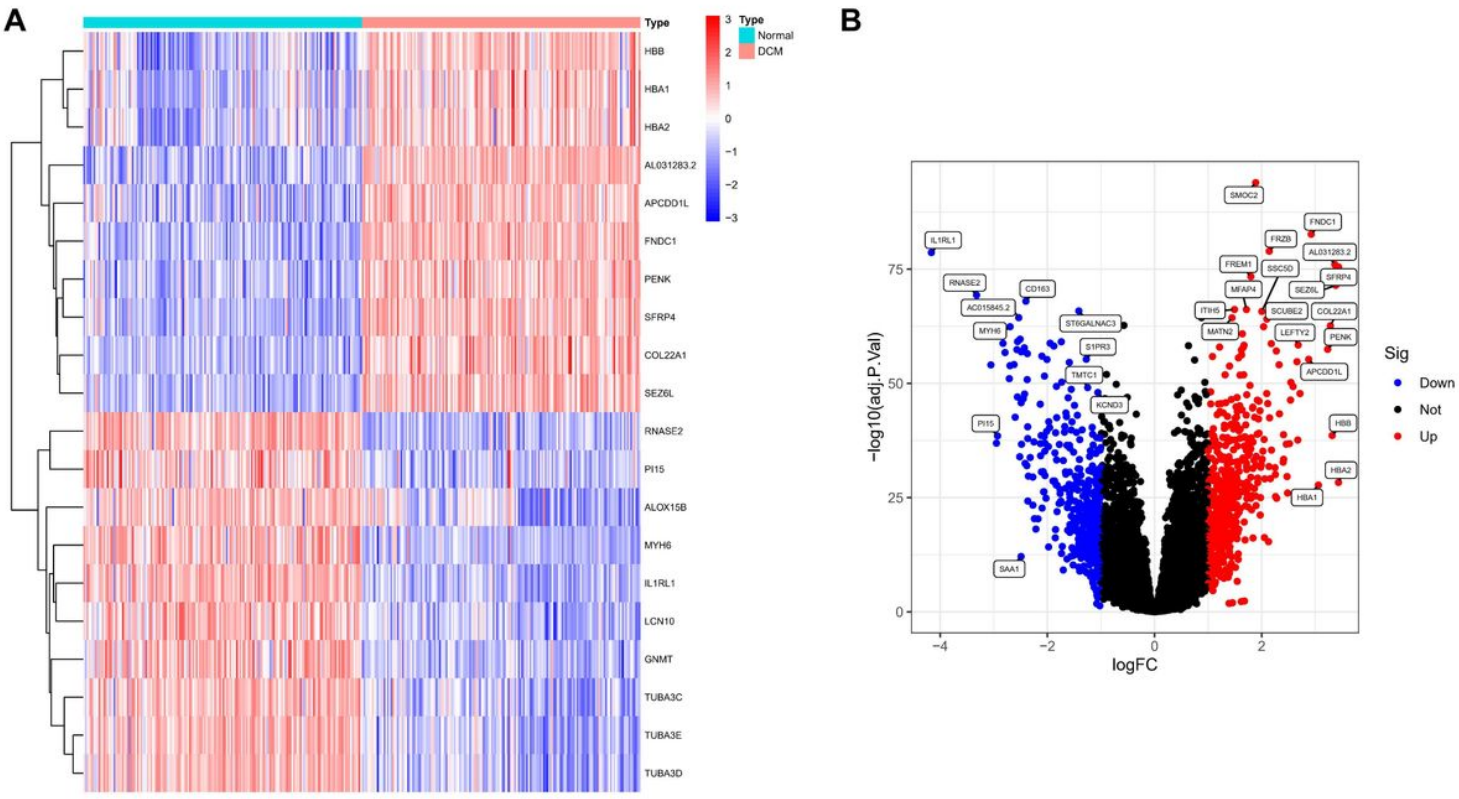

Figure 1

(**A**) heatmap of the 20DEGs between DCM and normal samples. (**B**) Volcano plot visualizing DEGs between normal and DCM samples

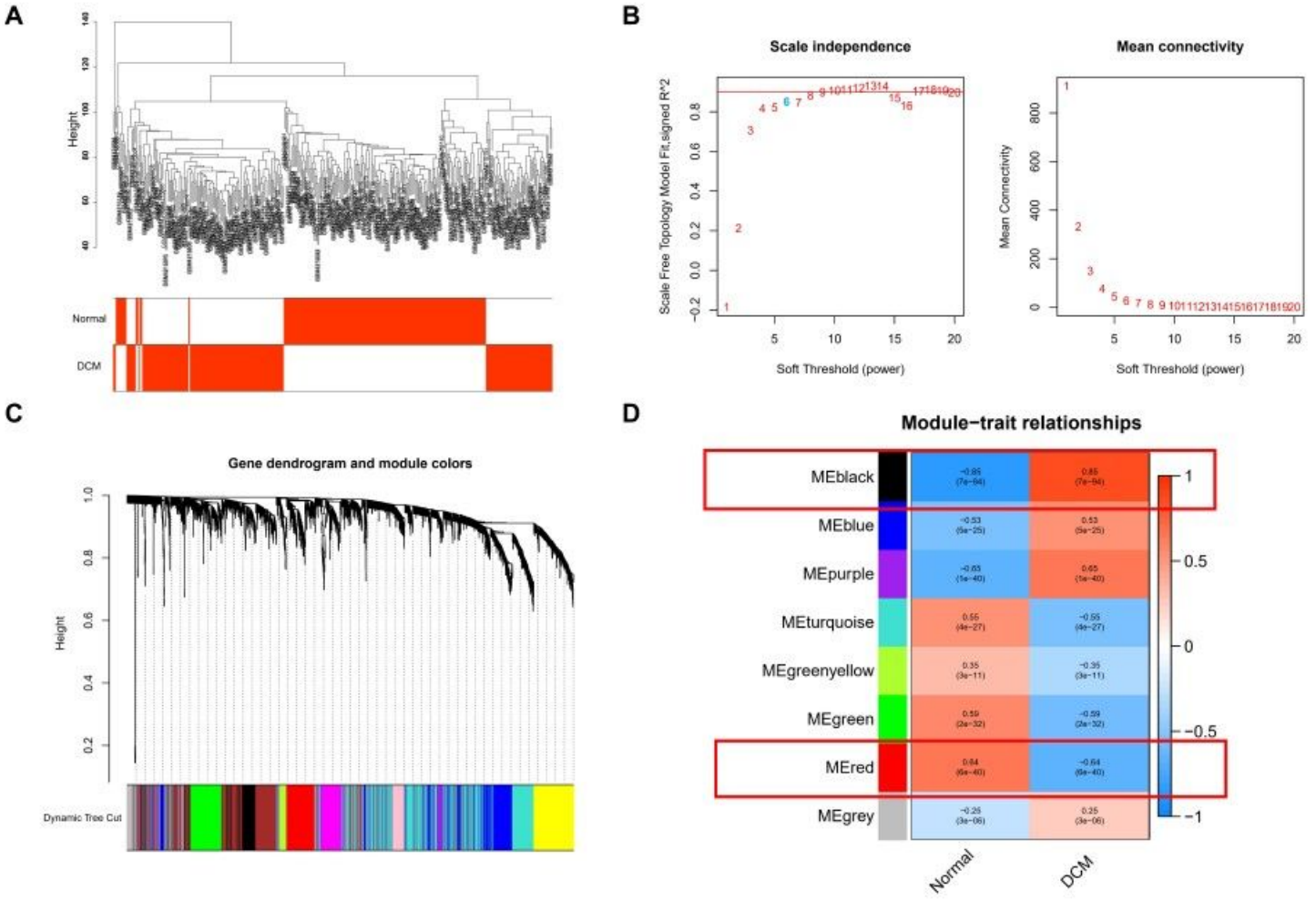

Figure 2

Identification of key modules correlated with clinical traits by WGCNA. (**A**) Clustering dendrograms of samples; (**B**) Analysis of the scale-free fit index and the mean connectivity under various soft-thresholding powers; (**C**) Dendrogram of all DEGs clustered with topological overlap dissimilarity measure; (**D**) Heat map of the correlation between module eigengenes and clinical traits. Each row corresponds to a module eigengene, each column represents a clinical trait and each cell contains the correlation coefficient and p value

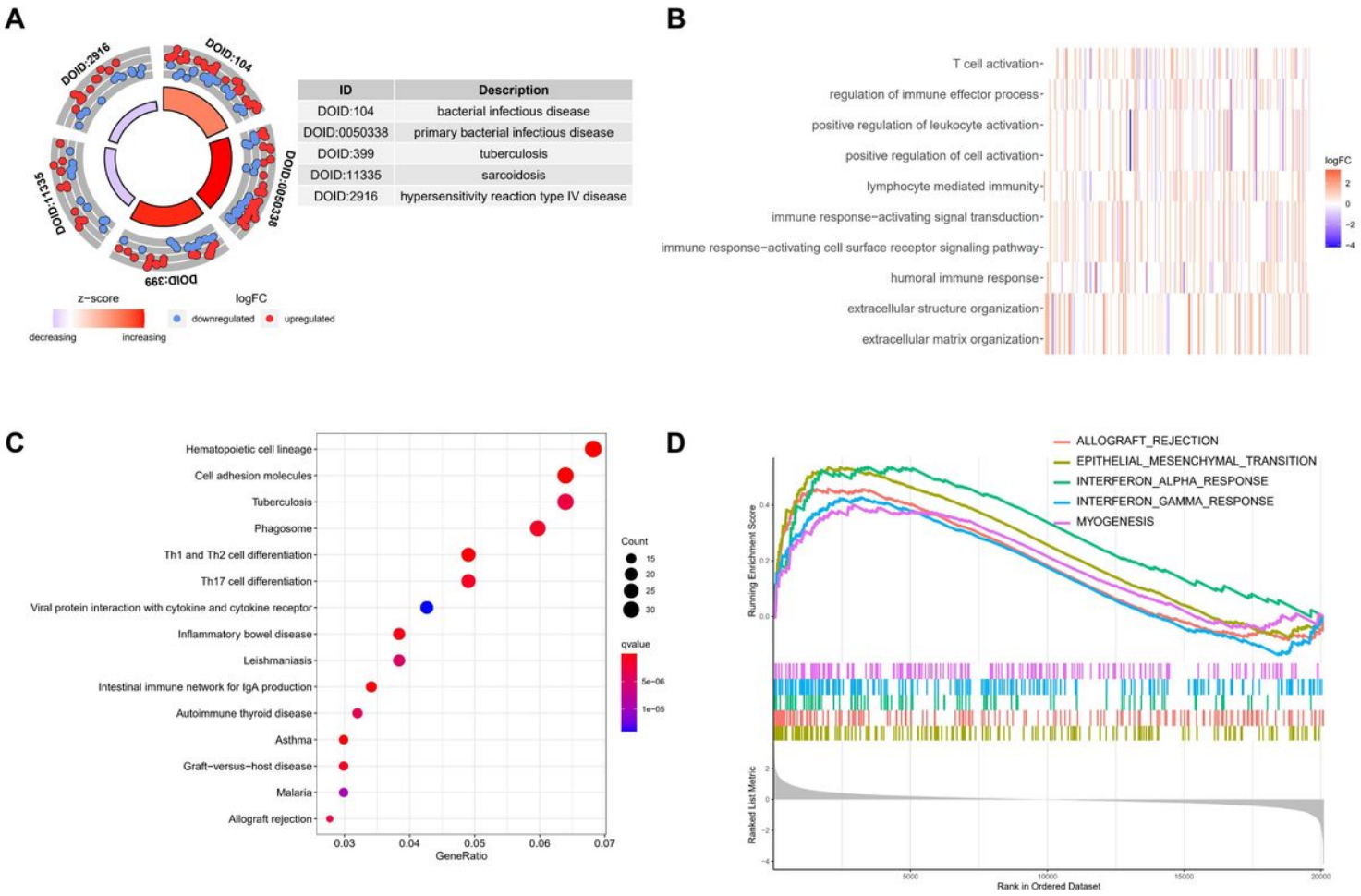

Figure 3

Enrichment analysis of 1343 potential core genes (**A**) DO enrichment analysis (**B**) GO enrichment analysis (**C**) KEGG enrichment analysis (**D**) GSEA enrichment analysis

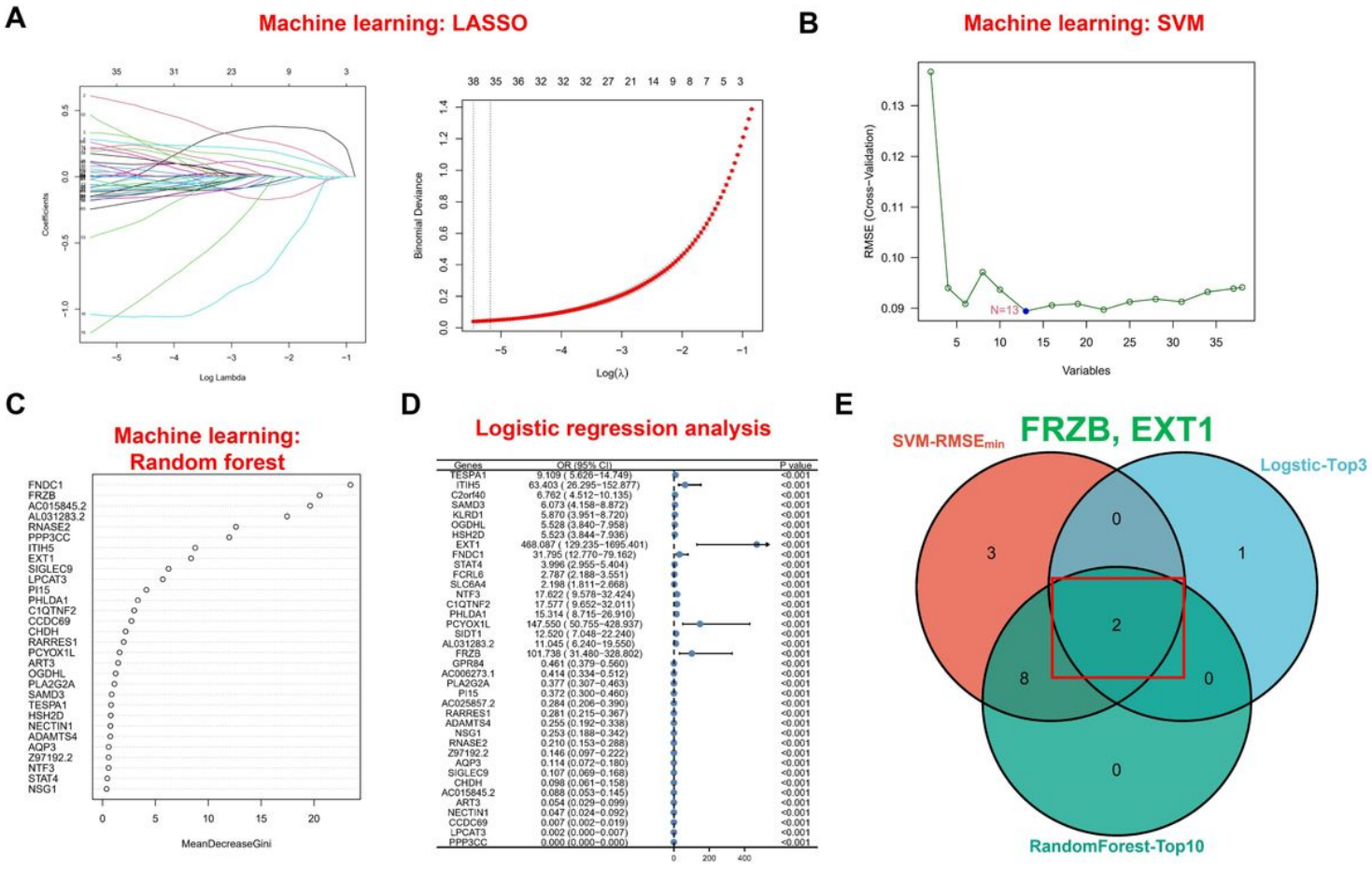

Figure 4

Outcome of multiple machine learning approaches (**A**) LASSON regression (**B**) SVM (**C**) Random forest (**D**) Logistic regression analysis and Venn diagram(**E**) showing the overlapping genes of three machine learning approaches

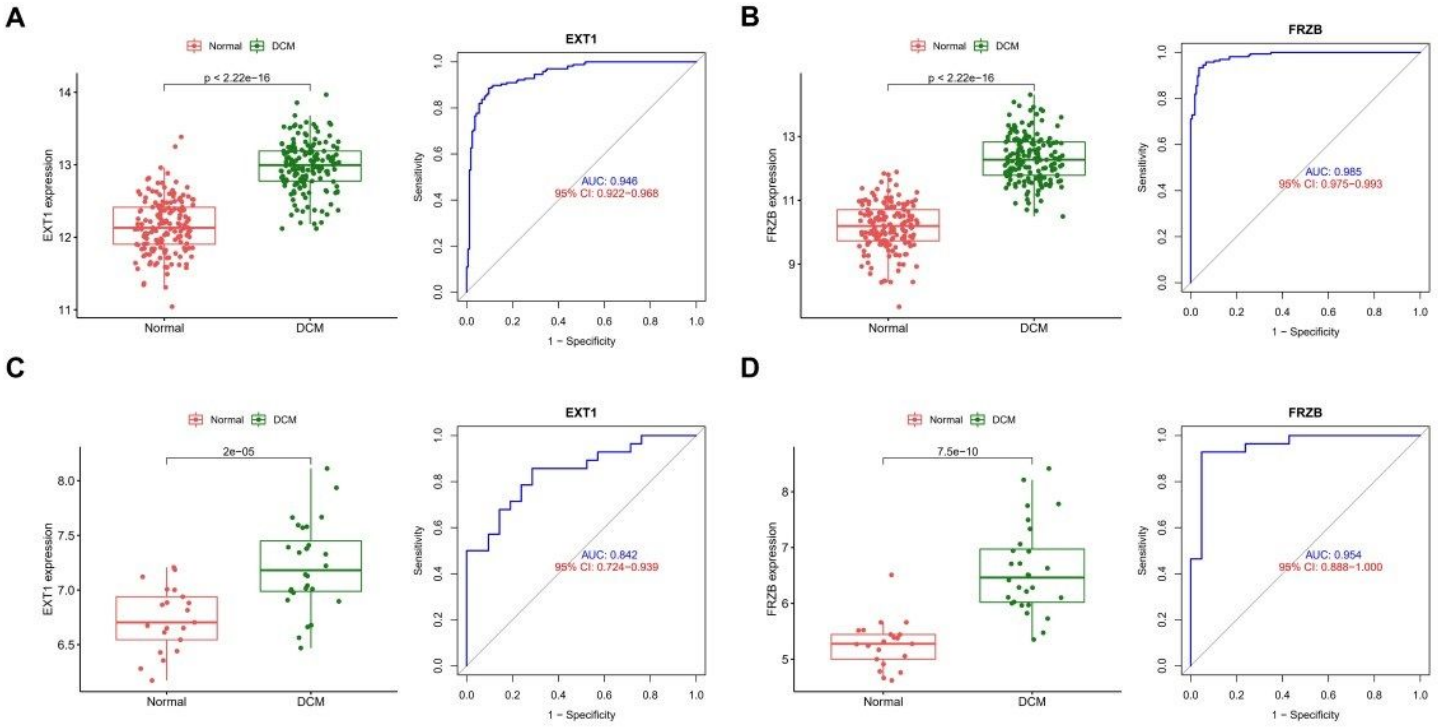

Figure 5

The diagnostic values of hub genes in DCM. ROC curves and AUC statistics to evaluate the diagnostic efficiency of hub genes on the incidence of DCM in the screening set(**A**,**B**) and in the external validation set(**C**,**D**)

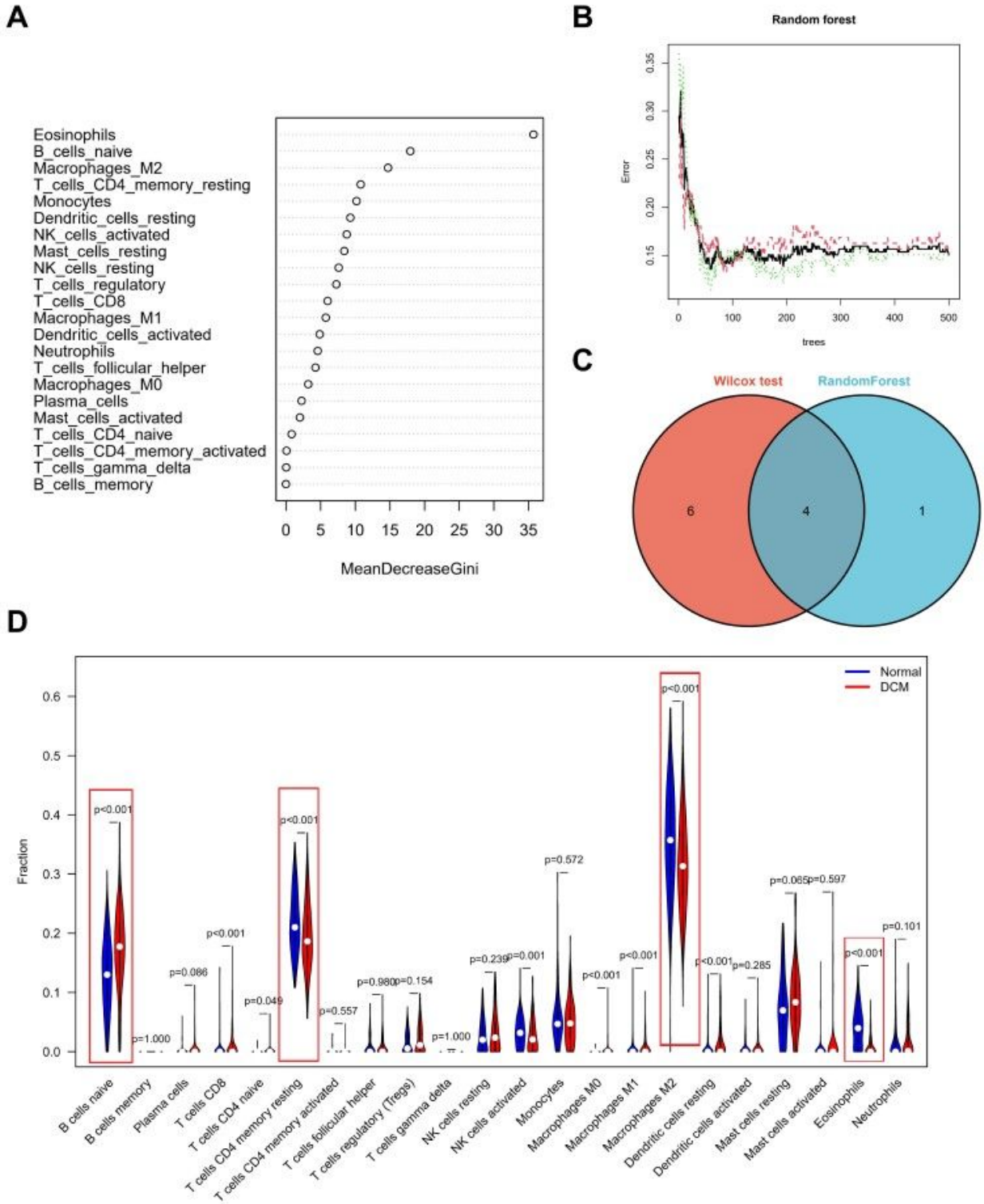

Figure 6

Wilcoxon test(**A**) and random forest tree(**B**) were performed to identify the core immune cells and Venn diagram(**C**) showing the overlapping immune cells and the fraction of immune cells were shown(**D**)

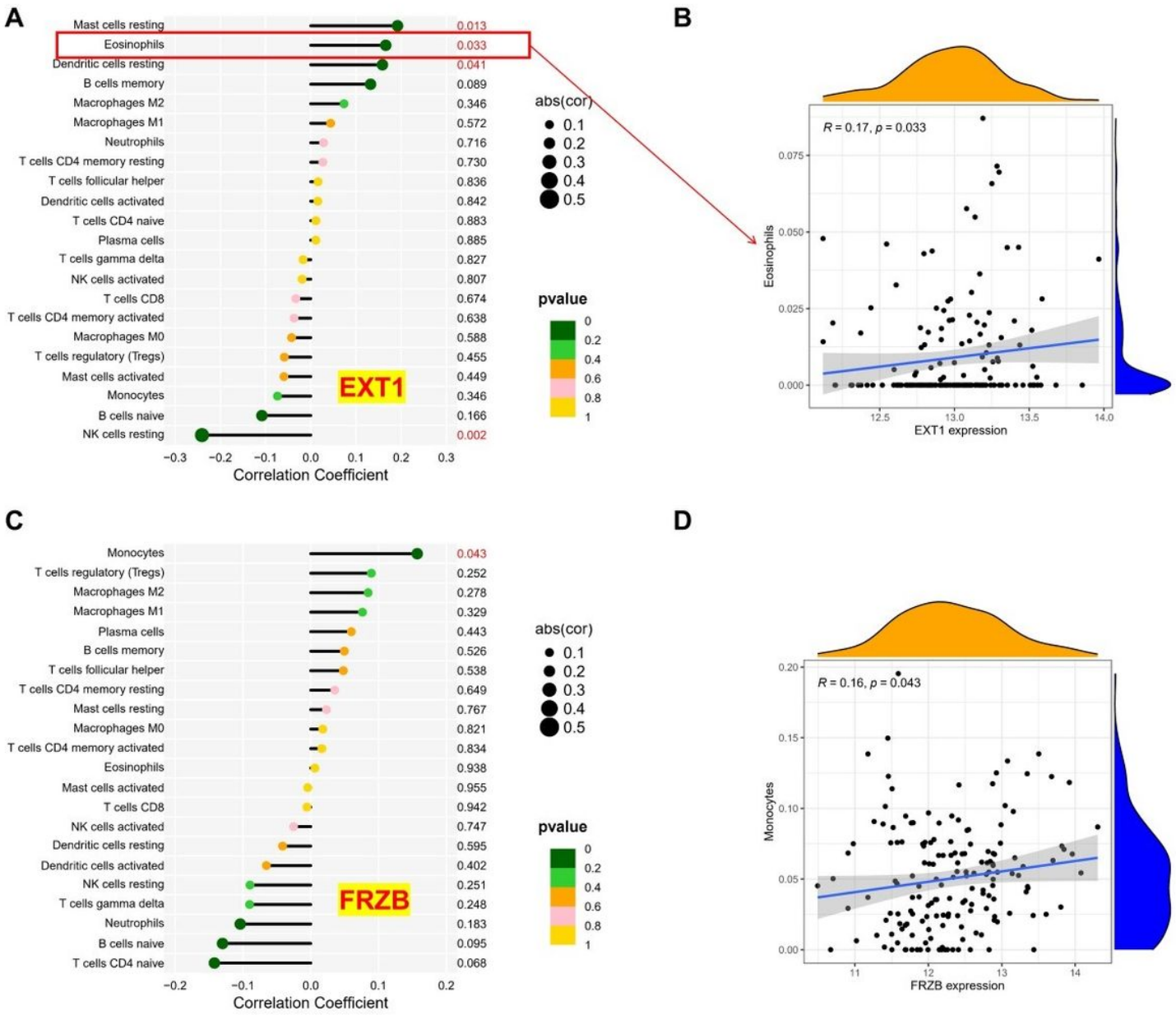

Figure 7

the correlation analysis between 22 kinds of immune cells and two hub genes. (**A**,**B**)EXT1 (**C**,**D**)FRZB

Supplementary Files

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

- Supplementarymaterial.docx
- figureS1.pdf
- figureS2.pdf
- figureS3.pdf