

RESEARCH ARTICLE

Open Access



Machine learning analysis of gene expression profile reveals a novel diagnostic signature for osteoporosis

Xinlei Chen, Guangping Liu, Shuxiang Wang, Haiyang Zhang and Peng Xue*

Abstract

Background: Osteoporosis (OP) is increasingly prevalent with the aging of the world population. It is urgent to identify efficient diagnostic signatures for the clinical application.

Method: We downloaded the mRNA profile of 90 peripheral blood samples with or without OP from GEO database (Number: GSE152073). Weighted gene co-expression network analysis (WGCNA) was used to reveal the correlation among genes in all samples. GO term and KEGG pathway enrichment analysis was performed via the *clusterProfiler* R package. STRING database was applied to screen the interaction pairs among proteins. Protein–protein interaction (PPI) network was visualized based on Cytoscape, and the key genes were screened using the cytoHubba plug-in. The diagnostic model based on these key genes was constructed, and 5-fold cross validation method was applied to evaluate its reliability.

Results: A gene module consisted of 176 genes predicted to be associated with the occurrence of OP was identified. A total of 16 significantly enriched GO terms and 1 significantly enriched KEGG pathway were obtained based on the 176 genes. The top 50 key genes in the PPI network were identified. Then 22 genes were screened based on stepwise regression analysis from the 50 key genes. Of which, 9 genes were further screened out by multivariate regression analysis with the significant threshold of P value < 0.01 . The diagnostic model was established based on the optimal 9 key genes, which efficiently separated the normal samples and OP samples.

Conclusion: A diagnostic model established based on nine key genes could reliably separate OP patients from healthy subjects, which provided novel lightings on the diagnostic research of OP.

Keywords: Osteoporosis, WGCNA analysis, PPI network, Logistic regression model

Introduction

Osteoporosis (OP) is defined as a systemic skeletal disease which is characterized by low bone mass and micro-architectural deterioration of bone tissue, and even can lead to an increasing risk in bone fracture [1, 2]. According to the Surgeon General's report on bone health and OP, OP has affected more than 8 million women and 2 million men in the USA, as well as

approximately 34 million people undergo low bone mass [3]. Because the early symptoms of OP are not obvious or even asymptomatic, early diagnosis and timely intervention can certainly prevent the disease from progressing to a more serious direction [4, 5]. Although, there are numbers of OP biomarkers and laboratory techniques available, each of which has its own strength and limitation [6]. Therefore, the identification and development of specific and sensitive biomarkers for OP diagnosis or treatment is more urgent.

* Correspondence: xp5879@126.com

Department of Orthopedics, Zibo Central Hospital, Zibo 255000, Shandong, China



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Recently, a series of OP-related genes that have been identified can act as potential diagnostic biomarkers and may potentially be applied for the clinical diagnosis for OP. For example, Zhang et al. analyzed the correlation between VDR gene polymorphisms and OP risk as well as bone mineral density (BMD) in postmenopausal women through a systematic meta-analysis and suggested that VDR is susceptible to OP and might be a potential biomarker [7]. Qian et al. found that PPWD1 may be considered as a potential diagnostic biomarker for the occurrence of postmenopausal OP based on the weighted gene co-expression network analysis (WGCNA) [8]. Mondockova et al. revealed that the estrogen receptor 1 gene can regulate BMD, and the treatment efficiency of OP in Slovak postmenopausal women may more likely be a biomarker [9]. In addition, many non-coding RNAs such as lncRNAs and circRNAs may also play an essential role in OP progression and also can be promising biomarkers [10]. lncRNA-NEF is significantly downregulated in postmenopausal OP and is closely associated with the course of treatment and recurrence [11]. Chen et al. have identified a group of circulating miRNAs as non-invasive biomarkers for the detection of postmenopausal and mechanical unloading OP through a large-scale screening based on microarray [12]. However, model analysis, always in the presence of other covariates, has been widely used to improve the probability of prediction compared with that of single genes [13, 14]. Despite of a large number of single biomarkers identified through different ways, the model analysis based on multiple key genes in OP is still lacking.

In the present study, a reliable diagnostic model was established based on nine key genes which might be involved in the progression of OP, and could efficiently separate normal subjects from OP patients. Our study enriched the research of early diagnosis in OP and exerted certain practical values.

Materials and methods

Data collection

The mRNA profile GSE152073 was downloaded from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) which includes 90 peripheral blood samples of 44 OP patients and 46 healthy subjects. The mRNA profiles of GSE152073 were detected by using Affymetrix Human Transcriptome Array 2.0.

Weighted gene co-expression network analysis

“WGCNA” function package of R language as previously described [15]. Briefly, the hierarchical clustering of genes was conducted based on the gene expression value of targeted genes, and genes with high expression similarity were placed into the same module by using the dynamic shear tree method. Next, we calculated the Module Eigengene

(ME) of each module and the correlation coefficient between the ME and the phenotype we focused on. In which phenotypes contained disease states (dichotomous phenotypes) (i.e., OP and normal). Besides, age, height, and weight were also included as continuous phenotypes in ME and phenotype correlation analysis. P value < 0.05 was used as the significant threshold of ME and phenotype correlation analysis. The greater the absolute value of ME, the closer the correlation between this module and the concerned phenotype.

Functional enrichment analysis

For these genes in the key modules, the Gene Ontology (GO) term (including biological process, molecular function, and cellular component) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed based on the *clusterProfiler* function package of R language [16] to identify key pathways involved in the progression of OP. $P < 0.05$ was used as the threshold for statistical significance.

The construction of protein–protein interaction (PPI) networks

The STRING (<https://string-db.org/>, version 11.0) is a widely used database to analyze and predict the functional connections and interactions of proteins in human diseases [17]. In the present study, STRING was applied for the functional connections and interactions analysis between proteins during the progression of OP, and the interaction pairs with confidence score ≥ 0.4 . Subsequently, the PPI network was visualized based on Cytoscape software (<https://cytoscape.org/>, version 3.7.2) as previously reported [18], and key genes of the PPI network were screened by using the cytoHubba plug-in of Cytoscape software according to the Maximum neighborhood component (MNC) algorithm.

The construction of logistic regression model

Logistic regression is a widely used method in classification as previously described [19], which is always used to predict a classification, in this study, OP or not. In the present study, all 90 samples were divided into the training set by simple random sampling method and the validation set for the construction of logistic regression model. Here, based on the normal control samples and OP samples, stepwise regression analysis was first used to prioritize genes followed by multivariate logistic regression analysis to determine the final genes contained in the model by using the *glm* function of R language [20]. Of which, the genes' expression values were used as the continuous variable, and the sample type was considered as the binary classification response value. Meanwhile, the 5-fold cross validation method was used to verify the accuracy of the logistic regression model.

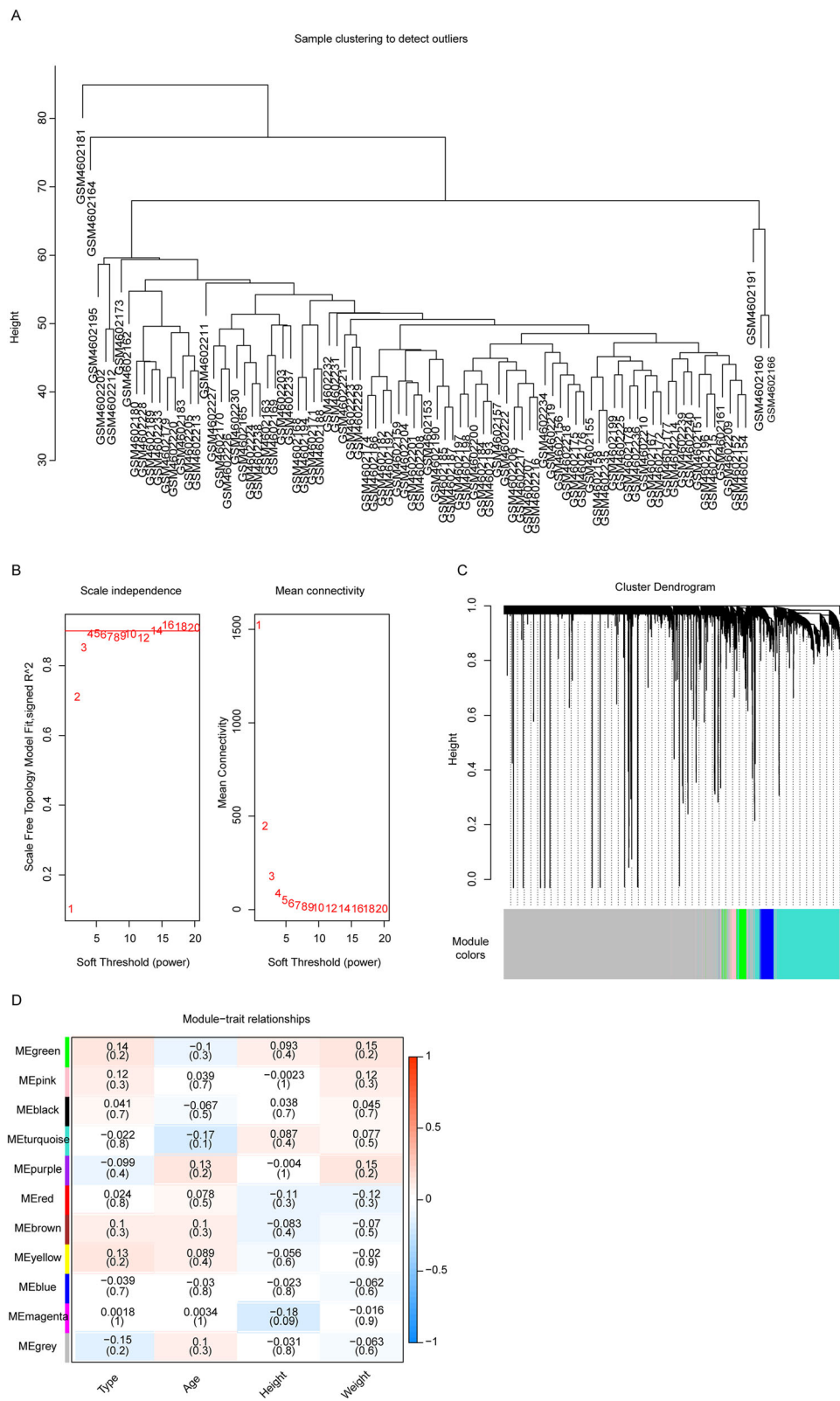


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Weighted gene co-expression network analysis (WGCNA). **a** Clustering of the 90 samples. **b** Selection of the soft threshold in the WGCNA. The red line represents the correlation coefficient, and the first point above the red line is the soft threshold ($\beta = 14$). **c** Clustering of the gene modules. Each designated color represents a gene module, and the genes that could not be grouped into other modules were placed in the grey module. **d** The correlation between gene modules and phenotypes. Darker color indicated greater the correlation between gene modules and phenotypes

Results

The green module was determined to be closely related to the occurrence of OP through WGCNA analysis

Firstly, we standardized the expression data of GSE152073 in order to remove the batch effects, and after standardization, no obvious deviation among all samples was observed. (Fig. S1). Therefore, the standardized data could be adopted for the subsequent analysis. Then the consensus clustering analysis based on the mRNA expression values of 90 samples indicated that there were two outliers (GSM4602181 and GSM4602164), which were excluded in the subsequent WGCNA analysis to ensure the reliability of our analysis (Fig. 1a). Previous studies have demonstrated that the co-expression network conformed to the characteristic of unsigned network [21, 22]. Hence, the soft threshold was selected as 14 to ensure that the co-expression network was unsigned, (Fig. 1b). Then the average-linkage hierarchical clustering method was applied to cluster the genes with setting the minimum cardinality of each gene module to 50 based on the criterion of mixed dynamic shear tree, and Module Eigengene (ME) for each module was then calculated. The cluster analysis on the modules was conducted, and the modules which were close to each other were merged into a new module with setting height 0.25. The results suggested that ten modules were obtained, and the grey module contained the gene set that could not be classified into any module (Fig. 1c). In addition, based on the ME of each module, we also calculated the correlation between each module with sample

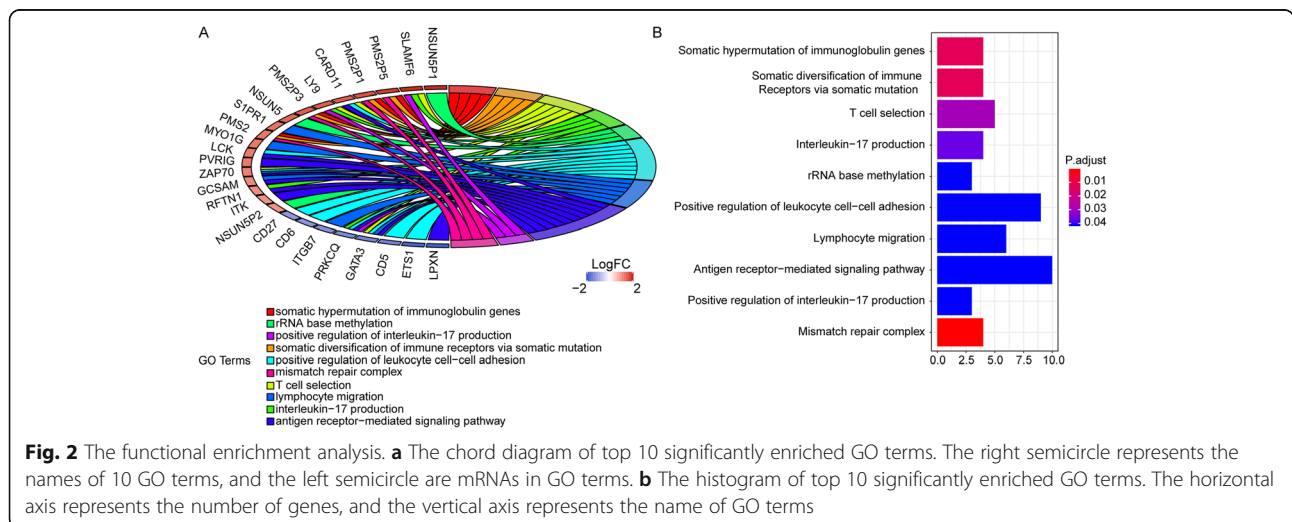
type as well as race information based on the ME of each module, and found that the green module which contained 176 genes had the greatest correlation with the sample type (Fig. 1d, correlation coefficient = 0.14). These results indicated that the genes in the green module were probably related to the initiation of OP.

Functional enrichment analysis of potential key genes in the green module

The GO term and KEGG pathway enrichment analysis were then performed to identify important pathways involved in the development of OP based on the 176 genes. The analysis identified 16 significantly enriched GO terms including somatic hypermutation of immunoglobulin genes, interleukin-17 production, and positive regulation of leukocyte cell–cell adhesion as well as one significantly enriched KEGG pathway (T cell receptor signaling pathway). Specifically, the top 10 most significantly enriched GO terms are shown in Fig. 2a and b. All of the significantly enriched GO terms and KEGG pathways are provided in Table S1. These results suggested that these enriched GO terms and KEGG pathways might be associated with OP progression.

Top 50 key genes in the PPI network were identified

The PPI network was visualized by using the 176 genes in the green module and shown in Fig. 3a. The PPI network analysis based on the 176 genes in the green module showed that there were 119 nodes and 224 edges, of



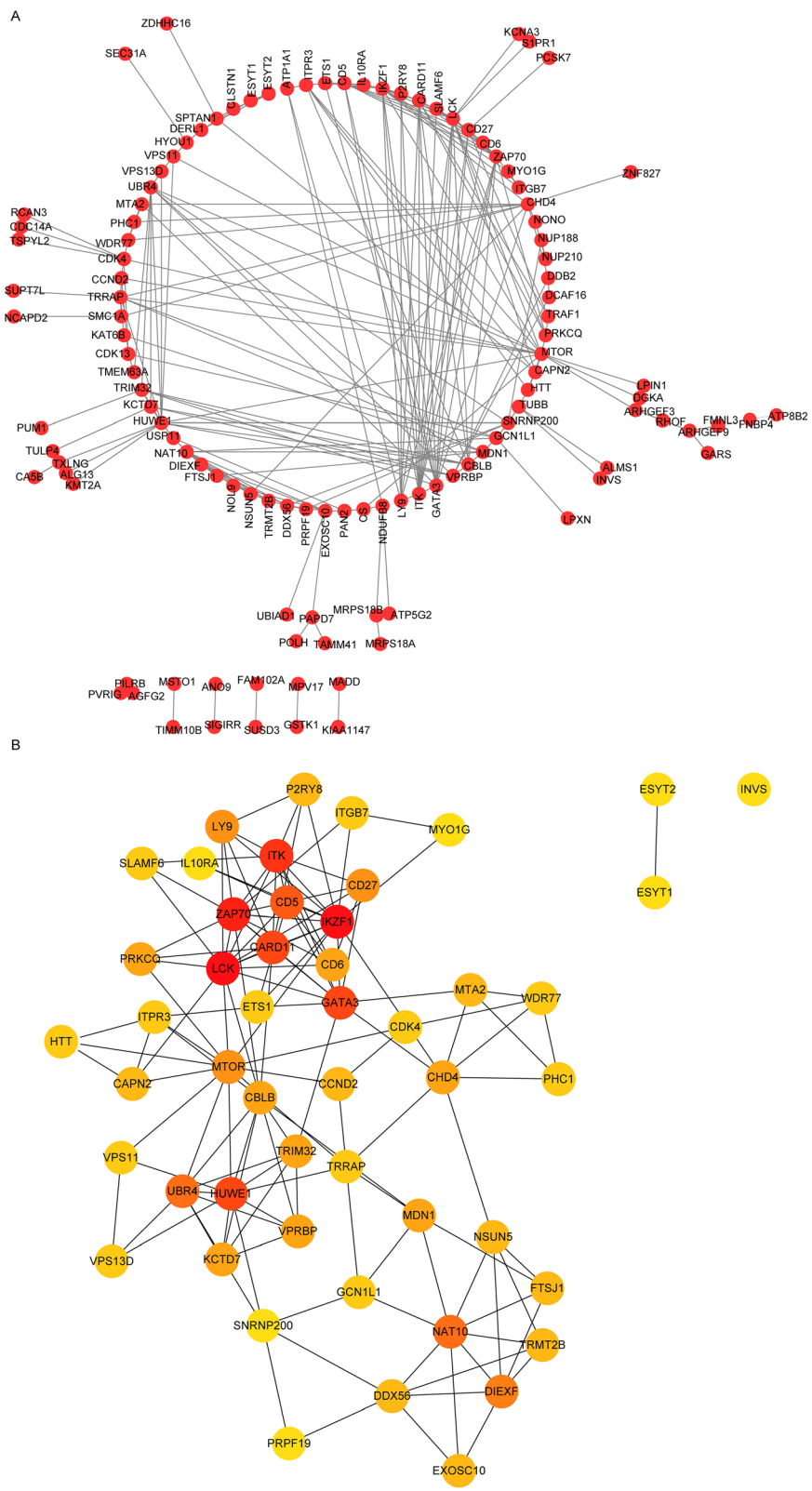


Fig. 3 The construction of PPI network. **a** The PPI network. Each dot represents a node. The more lines connected to the dot, the higher the degree of this node, which means the more important the genes on this node may be in the network. **b** The top 50 genes with high degree in PPI network were screened based on MNC algorithm. The darker the red color was, the higher the degree was

which, a node is a gene and the edge represents the interaction between them. More importantly, the top 50 genes were screened after sorting by score obtained using MNC algorithm (Fig. 3b), and these genes were considered as the potential key genes for OP (Table S2).

The diagnostic model could reliably separate OP patients from healthy subjects

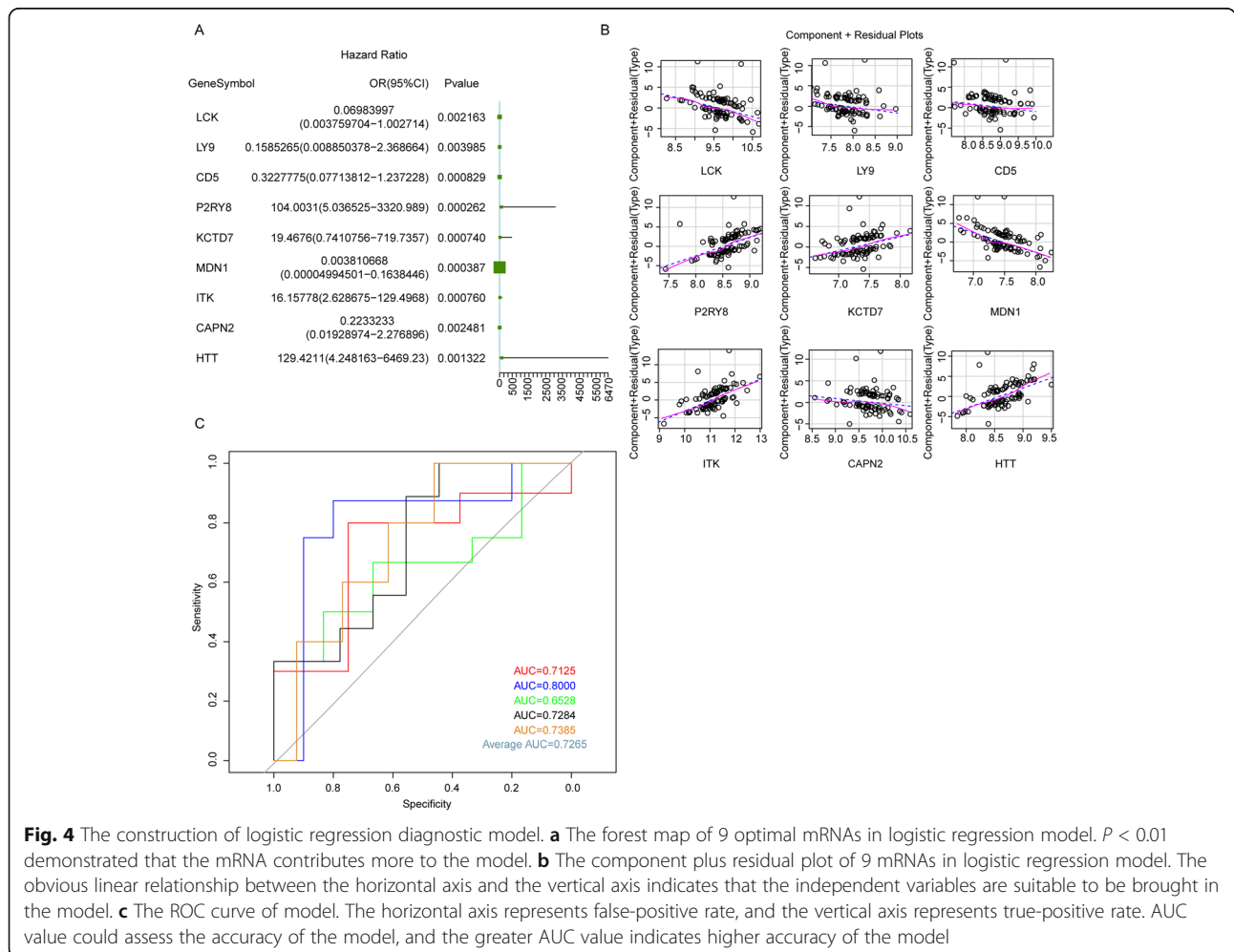
The stepwise regression analysis was first performed, which screened out 22 genes from the 50 key genes. Multivariate logistic regression analysis further identified 9 genes, including LCK, LY9, CD5, P2RY8, KCTD7, MDN1, ITK, CAPN2, and HTT, that are still significant ($P < 0.01$) in multivariate logistic regression analysis (Fig. 4a), suggesting that these 9 genes might be significantly related to the occurrence of OP. Finally, the logistic regression model was constructed based on the 9 key genes, and we found that the model obeys the normal distribution (Fig. S2A). Meanwhile, we found that the independent variables in the model all had significant linear correlation with the response variables (Fig. 4b), and no extreme points that could

influence the accuracy of the model were observed (Fig. S2B). All those data should illustrate the robustness of the model in OP occurrence prediction.

Further, the reliability and validity of this model were evaluated by 5-fold cross validation method. The AUC value of the 5 logistic models in the 5 validation sets was 0.7125, 0.8000, 0.6528, 0.7284, and 0.7385, respectively, with the average AUC 0.7265 (Fig. 4c). Besides, C-index of the model was 0.781, indicating that the model we established had a good consistency with the actual situation. These data suggested that the logistic regression model established by bringing into the 9 optimal genes could reliably distinguish the type of samples (OP or not).

Discussion

According to the World Health Organization (WHO), OP has become one of the most common diseases, and 30–50% of all women worldwide will undergo fractures because of OP throughout their lives [23]. On the other hand, since OP is clinically symptomless until the first



fracture occurs, early diagnosis is more critical and contributes to timely intervene OP to relieve the patient's pain [24]. In this study, we identified 176 genes that might be related to the occurrence and development of OP through WGCNA analysis. Meanwhile, the functional enrichment analysis was performed based on the 176 genes, and the results indicated that several GO terms such as somatic hypermutation of immunoglobulin, interleukin-17 production, and positive regulation of leukocyte cell–cell adhesion as well as only one KEGG pathway T cell receptor signaling pathway were significantly enriched. Previous studies have revealed that both the bone and immune systems work as a close-knit functional unit (osteimmune system) due to their common developmental niche [25]. McInnes et al. demonstrated that dysregulation of the immune system has already been involved in the initiation of various inflammatory autoimmune diseases resulting in adverse effects on bone integrity including OP [26]. In addition, T cells are identified to participate in the pathological bone loss associated with a series of conditions including ovariectomy-induced bone loss [27], rheumatoid arthritis [28], and bisphosphonate-associated osteonecrosis of the jaw [29]. Our analysis indicated that a total of 176 genes might be closely associated with the development of OP, which consists of previous studies that immune response occurs in the progression of OP. These results further provided a certain theoretical basis for our predictive model.

Then we identified the top 50 genes that were predicted to be closely related to the development of OP according to the score based on the PPI network. To establish a more explanatory model, the logistic regression analysis was performed and screened 9 optimal genes (LCK, LY9, CD5, P2RY8, KCTD7, MDN1, ITK, CAPN2, and HTT) ($P < 0.01$) which contributed more to the model. Lewis et al. describe a disorder in bone homeostasis in transgenic mice that anomalously express the cytokine interleukin 4 (IL-4) under the direction of the lymphocyte-specific proximal promoter for the LCK gene, and found that bone disease in LCK-defective mice appeared to be because of markedly decreased bone formation [30]. Sundberg et al. performed a flow cytometric analysis of peripheral blood lymphocytes in OVX rats revealed that both CD5 and CD4 positive lymphocytes or the mean fluorescence per cell are all increased [31]. One previous study reported that the potentiation of 5-HT signaling through the inhibition of the 5-HT transporter (5-HTT) has significant skeletal effects [32]. Although the effect of the other genes in OP has not been reported, their functions in several diseases have been better understood. For example, it has been reported that LY9 antibody targeting depletes marginal zone and germinal center B cells in lymphoid tissues and reduces

salivary gland inflammation in a mouse model of Sjögren's Syndrome [33]. Schmah et al. found that the expression of P2RY8/CRLF2 was significantly upregulated in childhood with acute lymphoblastic leukemia, and suggested that it might be related to the development of acute lymphoblastic leukemia in childhood [34]. Previous studies have revealed that a compound heterozygous missense mutation and a large deletion in the KCTD7 gene presenting as an opsoclonus-myoclonus ataxia-like syndrome [35]. In addition, Liu et al. demonstrated that the inhibition of ITK can induce anti-tumor activity by downregulating TCR signaling pathway in malignant T cell lymphoma both *in vitro* and *in vivo* [36]. All these researches indicated that these 9 optimal genes might be involved in the occurrence and development of human diseases including OP. Therefore, a diagnostic model was established based on these 9 genes and verified that it could efficiently separate patients with or without OP. Moreover, the 5-fold cross validation method was conducted to evaluate the reliability of this model and indicated that the model had certain practical values.

However, more samples are needed to be collected to evaluate the reliability of our predictive model. In addition, although the role of several key genes has been well studied in OP progression, the others should be explored in detail in the future.

Conclusion

In summary, a reliable diagnostic model was established based on 9 potential key genes including LCK, LY9, CD5, P2RY8, KCTD7, MDN1, ITK, CAPN2, and HTT, and we verified it could reliably separate patients with or without OP. Our results suggested that the diagnostic model might be potentially applied for the clinical diagnosis of OP.

Abbreviations

OP: Osteoporosis; WGCNA: Weighted gene co-expression network analysis; PPI: Protein–protein interaction; BMD: Bone mineral density; GEO: Gene Expression Omnibus; ME: Module Eigengene; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MNC: Maximum neighborhood component; WHO: World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13018-021-02329-1>.

Additional file 1. Fig. S1 The distribution of mRNA expression values in each sample after GSE152073 standardization. The horizontal axis represents the samples and the vertical axis represent the relative expression of mRNA.

Additional file 2. Fig. S2 The diagnosis diagram of logistic regression model. (A) The normal Q-Q diagram of logistic regression model. The points should fall on a line at an angle of 45 degrees. If the deviation is too large, the model violates the normal assumption. (B) The diagram of Residuals vs. Leverage. The red dotted line indicates the COOK distance. Generally, a point with the COOK greater than 0.5 is a very "influential" point, which affects the reliability of the model.

Additional file 3. Table S1. The full list of significantly enriched GO terms and KEGG pathways based on the 176 genes of the green module.

Additional file 4. Table S2. The top 50 genes screened from PPI network and respective scores.

Acknowledgements

Not applicable

Authors' contributions

XC and PX put forward the ideas of this article, written this article, and analyzed the data. XC, GL, SW, HZ, and PX helped for the acquisition of data and analysis and interpretation of data. GL, SW, and HZ helped for revising the manuscript. All authors read and approved the final manuscript.

Funding

Not applicable

Availability of data and materials

The datasets generated and analyzed during the current study are available in the [GEO] repository (<https://www.ncbi.nlm.nih.gov/geo/>).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Received: 23 November 2020 Accepted: 1 March 2021

Published online: 15 March 2021

References

- Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med.* 1993;94:646–50.
- Kanis JA, Cooper C, Rizzoli R, et al. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int.* 2019;30:3–44.
- Iwamoto J, Sato Y, Takeda T, Matsumoto H. Whole body vibration exercise improves body balance and walking velocity in postmenopausal osteoporotic women treated with alendronate: Galileo and Alendronate Intervention Trail (GAIT). *J Musculoskelet Neuronal Interact.* 2012;12:136–43.
- Eastell R, Szulc P. Use of bone turnover markers in postmenopausal osteoporosis. *Lancet Diabetes Endocrinol.* 2017;5:908–23.
- Ensrud KE, Crandall CJ. Osteoporosis. *Ann Intern Med.* 2017;167:ITC17–32.
- Dalle-Donne I, Rossi R, Colombo R, Giustarini D, Milzani A. Biomarkers of oxidative damage in human disease. *Clin Chem.* 2006;52:601–23.
- Zhang L, Yin X, Wang J, et al. Associations between VDR gene polymorphisms and osteoporosis risk and bone mineral density in postmenopausal women: a systematic review and meta-analysis. *Sci Rep.* 2018;8:981.
- Qian GF, Yuan LS, Chen M, et al. PPWD1 is associated with the occurrence of postmenopausal osteoporosis as determined by weighted gene coexpression network analysis. *Mol Med Rep.* 2019;20:3202–14.
- Mondockova V, Adamkovicova M, Lukacova M, et al. The estrogen receptor 1 gene affects bone mineral density and osteoporosis treatment efficiency in Slovak postmenopausal women. *BMC Med Genet.* 2018;19:174.
- Jin D, Wu X, Yu H, et al. Systematic analysis of lncRNAs, mRNAs, circRNAs and miRNAs in patients with postmenopausal osteoporosis. *Am J Transl Res.* 2018;10:1498–510.
- Ma X, Guo Z, Gao W, et al. lncRNA-NEF is downregulated in postmenopausal osteoporosis and is related to course of treatment and recurrence. *J Int Med Res.* 2019;47:3299–306.
- Chen J, Li K, Pang Q, et al. Identification of suitable reference gene and biomarkers of serum miRNAs for osteoporosis. *Sci Rep.* 2016;6:36347.
- Garcia TP, Marder K, Wang Y. Statistical modeling of Huntington disease onset. *Handb Clin Neurol.* 2017;144:47–61.
- Bonte C, Vercauteren F. Privacy-preserving logistic regression training. *BMC Med Genet.* 2018;11:86.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284–7.
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47:D607–13.
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
- Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med.* 2011;18:1099–104.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33:1–22.
- Du J, Wang S, He C, Zhou B, Ruan YL, Shou H. Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J Exp Bot.* 2017;68:1955–72.
- Greenham K, Guadagno CR, Gehan MA, et al. Temporal network analysis identifies early physiological and transcriptomic indicators of mild drought in *Brassica rapa*. *Elife.* 6:2017.
- Rachner TD, Khosla S, Hofbauer LC. Osteoporosis: now and the future. *Lancet.* 2011;377:1276–87.
- Dera AA, Ranganath L, Barraclough R, et al. Altered levels of mRNAs for calcium-binding/associated proteins, Annexin A1, S100A4, and TMEM64, in peripheral blood mononuclear cells are associated with osteoporosis. *Dis Markers.* 2019;2019:3189520.
- Geusens P, Lems WF. Osteoimmunology and osteoporosis. *Arthritis Res Ther.* 2011;13:242.
- McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol.* 2007;7:429–42.
- Onal M, Xiong J, Chen X, et al. Receptor activator of nuclear factor kappaB ligand (RANKL) protein expression by B lymphocytes contributes to ovariectomy-induced bone loss. *J Biol Chem.* 2012;287:29851–60.
- Hartgring SA, Willis CR, Bijlsma JW, Lafeber FP and van Roon JA: Interleukin-7-aggravated joint inflammation and tissue destruction in collagen-induced arthritis is associated with T-cell and B-cell activation. *Arthritis Res Ther* 14: R137, 2012.
- Kalyan S, Quabius ES, Wiltfang J, Monig H, Kabelitz D. Can peripheral blood gamma delta T cells predict osteonecrosis of the jaw? An immunological perspective on the adverse drug effects of aminobisphosphonate therapy. *J Bone Miner Res.* 2013;28:728–35.
- Lewis DB, Liggitt HD, Effmann EL, et al. Osteoporosis induced in mice by overproduction of interleukin 4. *Proc Natl Acad Sci U S A.* 1993;90:11618–22.
- Sundberg I, Rasmusson AJ, Ramklint M, Just D, Ekselius L, Cunningham JL. Daytime melatonin levels in saliva are associated with inflammatory markers and anxiety disorders. *Psychoneuroendocrinology.* 2020;112:104514.
- Warden SJ, Robling AG, Haney EM, Turner CH, Biliotes MM. The emerging role of serotonin (5-hydroxytryptamine) in the skeleton and its mediation of the skeletal effects of low-density lipoprotein receptor-related protein 5 (LRP5). *Bone.* 2010;46:4–12.
- Punet-Ortiz J, Saez Moya M, Cuenca M, Caleiras E, Lazaro A, Engel P. Ly9 (CD229) antibody targeting depletes marginal zone and germinal center B cells in lymphoid tissues and reduces salivary gland inflammation in a mouse model of Sjogren's syndrome. *Front Immunol.* 2018;9:2661.
- Schmah J, Fedders B, Panzer-Grumayer R, et al. Molecular characterization of acute lymphoblastic leukemia with high CRLF2 gene expression in childhood. *Pediatr Blood Cancer.* 2017;64.
- Blumkin L, Kivity S, Lev D, et al. A compound heterozygous missense mutation and a large deletion in the KCTD7 gene presenting as an opsoclonus-myoclonus ataxia-like syndrome. *J Neurol.* 2012;259:2590–8.
- Liu Y, Wang X, Deng L, et al. ITK inhibition induced in vitro and in vivo anti-tumor activity through downregulating TCR signaling pathway in malignant T cell lymphoma. *Cancer Cell Int.* 2019;19:32.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.