Identification of hub genes associated with postmenopausal osteoporosis by Gibbs sampling method

YA-PENG ZHANG, SHUANG AO, YU LIU, YU WANG, YI-MING JIA, HAO ZHANG and HUI LENG

Department of Orthopedics, Chifeng Municipal Hospital, Chifeng, Inner Mongolia 024000, P.R. China

Received June 4, 2018; Accepted January 3, 2019

DOI: 10.3892/etm.2019.7231

Abstract. Underlying pivotal pathways were identified to reveal potential key genes correlated with postmenopausal osteoporosis (PMOP). The pathways were enriched by Kyoto Encyclopedia of Genes and Genomes (KEGG) with genes intersection greater than 5 based on gene expression profile data, and the acquired pathways were then transformed to Markov chain (MC). Gibbs sampling was conducted to obtain a new MC. Moreover, the average probabilities of each pathway in normal and PMOP were computed via an MC Monte Carlo (MCMC) algorithm, and differential pathways were identified based on probabilities more than 0.7. In addition, frequencies of appearance of pathway genes were counted via MCMC and the hub genes were achieved with the probabilities of gene expression efficiencies in two states. Judging by the gene intersection more than 5, overall 280 pathways were determined. After Gibbs sampling, 2 differential pathways were obtained on the basis of probabilities more than 0.7. Moreover, the hub genes comprising TNNC1, MYL2, and TTN were achieved according to frequencies more than 0.7. The identified pathways and the three hub genes probably are useful for developing approaches for the diagnosis and treatment of PMOP in future preclinical and clinical applications.

Introduction

Osteoporosis termed systemic osteopathy characterized as bone loss and bone microstructure destruction is correlated with aging and menopause, which becomes a primary disturbing issue common in postmenopausal women and the elderly worldwide (1,2). In spite of several marked advances in early diagnosis, work-up and therapeutic managements (3,4), there are still demands for emerging strategies with postmenopausal osteoporosis (PMOP).

Occurrence and progress of PMOP comprise a sophisticated biological process involving a series of genomic changes and diverse molecular pathogenesis. With the development of various bioinformatics, a wide spectrum of disease prediction and investigation of molecular mechanism have come to light. Mounting evidence indicates that several pathways are tightly associated with PMOP, including the Wnt/β-catenin pathway, (RANKL)/RANK/osteoprotegerin (OPG) signaling pathway, and NF-κB signaling pathway (5-7). Previous findings demonstrated that numerous protein biomarkers with PMOP were selected, for instance, SOD, A1AT and TRIM63 (8-10). It is well known that genes contain most of the genetic information that is closely relevant with phenotype of human beings, old, sick, and dead. Thus, genetic expression variations in disease development are more likely to affect a series of biomarker behaviors and signaling transductions. In addition, a large body of previous studies identified that multiple genes relative to PMOP were screened out using gene expression profiling (11,12). Nevertheless, studies on key genes and crucial pathway in PMOP are limited.

Gibbs sampling is one of the Markov chain Monte Carlo (MCMC) algorithms, used to construct a random sample with multivariable probability distribution (13). Notably, on the basis of the probabilities, it is probable that crucial pathways and key genes of importance to uncover various pathogenesis of disorder are identified (14,15). Thus, in our research, Gibbs sampling was applied to investigate the significance of a pathway gene set and their functions in postmenopausal osteoporosis.

In the present study, we selected 280 pathways based on gene intersection greater than 5. The pathways were then transformed to MC, and Gibbs sampling was performed to gain a new MC. Moreover, the mean probability of gene expression in each pathway was calculated using the MCMC algorithm and then differential pathways were identified on the basis of the probability of pathway expression more than 0.7. Furthermore, the emergence times of genes in differential pathway were counted, the average probability of gene expression in pathway was calculated and the hub genes were gained by the probability of gene expression more than 0.7. Our findings are useful in the investigation of potential molecular biomarkers for the diagnosis, therapy and monitoring progression of PMOP.

Materials and methods

Microarray data capturing and data preprocessing. The profile E-MEXP-1618 (16) was downloaded from ArrayExpress...
were selected, as shown in Fig. 4. Moreover, their posterior inference was used to identify probability distributions of pathway expression values for PMOP. During the pathway analysis, the initial transition probability was obtained from expression values of the first state and the second state, the third state was reckoned from the second state.

Gibbs sampling. According to the enrichment condition of the gene expression profile in each pathway, we calculated the average gene expression of each pathway under the first state (normal) and the second state (osteoporosis) of samples, and regarded this average gene expression as the pathway expression value. At this point, the first state was acted as the final state and the second state was acted as a prior state. After having converted all the pathway expression values into MCs, their posterior inference was used to identify probability distributions of pathway expression values for PMOP. During the pathway system, the initial transition probability was obtained from expression values of the first state and the second state, the third state was reckoned from the second state.

Gibbs sampling is MCMC algorithm which aims to gain a sequence of samples approximated from a specified multi-dimensional probability distribution. To perform Gibbs sampling, the above pathway expression value should be converted into Markov chains. Firstly, an empty Gibbs sampling set containing an M-dimensional (M = pathway samples) random vector was defined. Secondly, N samples Markov chain data set including the initial value and prior value were put into the empty Gibbs sampling set. Thirdly, an M-dimensional vector was initialized, M-1 elements of this vector were fixed, the remaining elements were extracted, like this cycled M times which amounts to refresh the whole vectors and generate a new sample. The third state was acquired. Finally, through n cycles of Gibbs sampling, a Markov chain was constructed.

Differential pathway analysis. Based on the posterior value of the pathway generated by the Markov chain and using the probability calculation formula alfa.pi, the probability of each pathway was obtained.

\[ \text{alfa} \cdot \pi = \frac{\sum_{i=0}^{10000} P_i}{10000 - 2000 + 1} \]

where ‘alfa.pi’ is ‘posterior value of a pathway’.

Thereinto, \( P_i \) represented the posterior value of the pathway in subsample i. According to pathway expression values in different states, P-values of pathways were calculated using t-test. Subsequently, P-values were ranked, combining P-values with probability alfa.pi, the correction coefficients (R-value) were calculated and then the adjusted probability of the pathway (alta-adj) was obtained. Pathways of which alfa-adj was >0.7 were regarded as differential pathway. The formula used was:

\[ R\text{value} = 1 - \frac{\text{rank}_i}{n} \]

where \( n \) stands for the number of pathway, and \( \text{rank}_i \) stands for i ranking.

Pathway enrichment analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) is one of the most commonly used biological information databases worldwide, characterized by associating large set of genome information with higher level functional information of cell, species and ecosystems (17). In the present study, gene symbols were enriched to KEGG pathways, from which we chose pathways with gene interaction in pathways >5.

In order to explore the potential key genes relative to PMOP, hub genes in the differential pathway were selected via Gibbs sampling. Based on the alfa-adj >0.7, three hub genes including TNNC1, MYL2, and TTN were selected, as shown in Fig. 4. Moreover, their expression status in the normal and PMOP condition were indicated in a box scatter diagram (Fig. 5).
Discussion

PMOP, defined as a systemic skeletal disorder, is a silent disease without recognition in patients until fractures emerge. Concomitantly, novel diagnostics and therapeutics of PMOP are under investigation. Currently, the diagnostic potential of crucial pathways or key genes for disorders has been studied and several pathways and genes were identified as underlying biomarkers (18,19). In this study, we presented PMOP correlatevially pivotal pathways and key genes, which may be applied to diagnose and monitor PMOP. Gibbs sampling analysis indicated that 2 differential pathways including cardiac muscle contraction and hypertrophic cardiomyopathy were selected. Significantly, three hub genes associated with PMOP were identified containing TNNC1, MYL2, and TTN, which could be good candidates for biomarkers in the diagnosis of PMOP in future clinical applications.

Previous findings showed that cardiovascular disease is a main reason of death among postmenopausal women (20,21). Many treatments administered are in the form of drug therapy, including, raloxifene (22), drospirenone/17β-estradiol (23). However, few studies focus on the examination and treatment of PMOP associated with cardiac disease or prevention of cardiac disorder in postmenopausal women while there are rare investigations on the underlying molecular mechanisms involved. Concomitantly, research has shown that low bone

Figure 1. The probability distribution of 280 pathways. The X axis represents the pathways, and the Y axis is the posterior value of the pathways. (A) The adjusted posterior value distribution of 280 pathways. (B) The density of posterior value distribution of 280 differential pathways. Pathway was regarded as differential pathway according to adjusted posterior value >0.7.

Figure 2. The differential pathway expression level by box scatter diagram. The pathways expression level of (A) cardiac muscle contraction and (B) hypertrophic cardiomyopathy. The X axis denoted groups, and the Y axis denoted the differential pathways expression level. *P<0.05, indicates statistically significant difference between two groups; maximum value.
mineral density is associated with increased cardiovascular mortality (24) and emerging findings have demonstrated that vascular calcification and bone mineralization share various anatomical and pathophysiological common properties (25). In the current study, by using Gibbs sampling combined with KEGG pathway analysis to investigate gene expression profile of PMOP, we favorably selected crucial pathways including cardiac muscle contraction and hypertrophic cardiomyopathy.

The above facts indicated that cardiac muscle dysfunctions are likely to be associated with the occurrence of PMOP.

Furthermore, we identified key genes TNNC1, MYL2, and TTN by Gibbs sampling analysis of differential pathways. TNNC1, encoded Cardiac troponin C (cTnC), was reported as involved in modulating cardiomyopathy (26). Troponin is composed of three subunits, the troponin C, troponin I and troponin T. Previous research demonstrated that elevated
serum cardiac troponin I was associated with hip fracture in older patients including postmenopausal women (27). Another emerging report studied that highly sensitive cardiac troponin T relates to mortality in perimenopausal women (28). It is common knowledge that \( \text{MYL2} \) is also known as \( \text{MLC-2} \) and encoded myosin light chain-2. It is reported that myosin II is functionally important in bone resorption, besides, myosin activity favors the osteoclast-differentiated activity of bone resorption (29). Recent findings have shown that a decreased expression of Runx2 is accompanied by a lower expression of myosin in ovariectomized rats, simultaneously, have revealed that myosin-dependent nuclear-cytoplasmic shuttling of Runx2 is likely to be crucial for supplying interconnection between myosin and signal transduction cascades, accordingly it commands transcription of downstream factors in osteoblastic cells (30). Titin, encoded by \( \text{TTN} \), is a large sarcomere protein, and an elastic protein. Eccentric exercise, rehabilitation and athletic training, is commonly prescribed for treatment of various types of conditions such as sarcopenia, osteoporosis, and tendinosis, which is mediated by titin-actin and titin-myosin interactions through the study of sports experts (31). In light of the evidence, it is speculated that \( \text{TNNC1}, \text{MYL2}, \text{TTN} \) as hub genes are good for predicting and diagnosis of PMOP. In spite of this, there are limitations of this study. This work is only a slight improvement of probing early-discriminating clues during disease progression as well as this algorithm is anticipated to meliorate or combine with more complicated predicted measures, thus enhancing efficiency and accuracy for predicting disorders with might be the main ones. The effectiveness of the identified differential pathway and hub genes still need further support from animal experiments or clinical investigations. In this work, we
presented a bioinformatics analysis of genetic chip databases based on Gibbs sampling to identify biomarkers including crucial pathways and key genes, which had been shown to be effective by real datasets.

Collectively, the findings of the present study demonstrated that the pathways of cardiac muscle contraction and hypertrophic cardiomyopathy as well as hub genes TNNCI, MYL2, and TTN may exert significant effects in the development of PMOP, which may provide a non-invasive methodology for the prediction, diagnosis, and even personalized treatment of clinical osteoporosis, furthermore it may be helpful for comprehending molecular pathogenesis of PMOP.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

HL conceived the study and drafted the manuscript. YW, YMJ and HZ acquired the data. YPZ, SA and YL analyzed the data and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


