Abstract. Differential expression pathways and hub genes in circulating monocytes from healthy Chinese women with high peak bone mass (PBM) vs. low PBM were explored using a Markov chain Monte Carlo (MCMC) algorithm. Human circulating monocytes transcription profiling (containing 14 samples with high PBM and 12 samples with low PBM) and KEGG pathways were all downloaded from the public database. Initial state of all the pathways were constructed and Gibbs sampling was performed to obtain a Markov chain and the posterior values of all the pathways were calculated. The probability (α) of occurrence of each pathway was calculated based on the posterior value and it was adjusted by taking gene expression variation into account. Pathways with $\alpha_{adj} > 0.8$ were considered as differentially expressed pathways. Then, these steps were performed again on all the genes in the differentially expressed pathways to find the hub genes in the differential pathways. After Gibbs sampling, neuroactive ligand-receptor interaction (hsa04080) with $\alpha_{adj} = 0.986$ was screened out as the differentially expressed pathway. Analyzing the genes in this pathway, three genes (neurotensin, tachykinin receptor 3 and follicle-stimulating hormone receptor) with $\alpha_{adj} > 0.8$ were identified as hub genes in circulating monocytes which may associate with osteoporosis development. One pathway and three genes which may possess close relationship with osteoporosis development were found in this study. These results provide insights into our understanding of the role of circulating monocytes in osteoporosis development.

Introduction

Peak bone mass (PBM), the highest bone mass obtained during the life of a person, is a critical determinant of osteoporosis (1,2). It has been reported that gaining a high PBM in the first 30 years could effectively prevent osteoporosis (3,4). Circulating monocytes are important cells that serve as early progenitors of osteoclasts and has been reported to be a suitable model for bone-related research (5,6). Deng et al have identified and confirmed five proteins [ras suppressor protein1 (RSU1), gelsolin (GSN), manganese-containing superoxide dismutase (SOD2), glutathione peroxidase 1 (GPX1), and prolyl 4-hydroxylase β subunit (P4HB)] that may involve in differential osteoclastogenesis by a comparative protein expression profiling study and western blot analysis of circulating monocytes in premenopausal females with extremely discordant bone mineral density (7). Lei et al identified three novel risk genes (GBP1, STAT1 and CXCL10) that may be involved in the differentiation of PBM at the monocyte stage (1). Although there have been some studies on the role of circulating monocytes in osteoporosis development, the relationship between circulating monocytes and the pathogenesis of osteoporosis remains poorly understood.

Gibbs sampling, also called Gibbs sampler, is a Markov chain Monte Carlo (MCMC) algorithm for getting a sequence of observations that are approximated from a specified multivariate probability distribution (8). It can be used to identify perturbed pathways and hub genes which may be closely related to occurrence and development of the disease. In this research, we aimed to identify the differential pathways and hub genes in circulating monocytes of healthy Chinese women with high PBM and low PBM using Gibbs sampling, attempting to promote our understanding of the function of circulating monocytes in osteoporosis.

Data collection and preprocessing. Transcription profiling of human circulating monocytes was obtained from the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) and the accession number was E-GEOD-7158. The database consists of 26 samples, including 14 samples from healthy Chinese females with extremely high PBM and 12 samples from healthy Chinese females with extremely low PBM. Followed by preprocessing (9,10), the expression profile was converted from probe level to gene symbol level and the
Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway enrichment. The KEGG database was used in this study (11). A total of 287 pathways, including 6,894 genes were downloaded from the KEGG database. The genes in the transcription profile were enriched to the KEGG pathways. Pathways with gene counts ≥5 were screened out for the next analysis.

Construction of path initial state. The average expression values of the genes in the pathways under high and low PBM conditions were calculated, respectively, and considered as the expression of this pathway. High PBM condition was regarded as the initial state of the pathway and low PBM condition was regarded as a priori value of the pathway.

According to the Markov chain theory, we deduced the initial transition probability from the expression of state 1 and state 2. Then state n is derived from state n-1 and the state n is only associated with state n-1. The original state of the system has no connection with the other Markov processes.

Gibbs sampling. Gibbs sampling was implemented to form a Markov chain. First, an empty Gibbs sampling set was defined followed by storing the initial state and a priori value of the pathways into the Gibbs sample empty set. Secondly, a k-dimensional vector (k = 280) was initialized randomly and k-1 of these vectors were fixed accompanied by extracting the remaining one vector. The posterior value of this vector was calculated. After repeating this process for k times, a new k-dimensional vector was obtained and defined as state 3. After repeating this whole process 10,000 times, a Markov chain was obtained. The vectors generated after 2,000 times are considered to follow the distribution of the real samples, which is a satisfactory convergence.

Identifying key pathways. After gaining the posterior values of all the pathways, the probability (α) of occurrence of each pathway was calculated using the following formula 1.

\[ α = \frac{\sum_{i=1}^{10000} p_i}{10000 - 2000 + 1} \]

\( p_i \) represents the posterior value of this pathway in the i\textsuperscript{th} sample.

According to the expression level of the pathway in different states, the P-value of the pathway is calculated by Student’s t-test. All the pathways were then ranked according to their P-value. Subsequently, the correction coefficient \( c \) was calculated using the following formula 2.

In formula 2, \( rank_i \) represents the rank of pathway \( i \) and \( n \) represents the number of samples.

\[ c = 1 - \frac{rank_i}{n} \]

Next the adjusted probability of each pathway was calculated as follows: \( α_{adj} = α \times c \). After accomplishing all the adjusted calculation probabilities, the pathways were all ranked according to their \( α_{adj} \) value and pathways with \( α_{adj} ≥0.8 \) were selected as differentially expressed pathways (12).

Identifying hub genes. After identifying the differentially expressed pathways, we proceeded to find the hub genes in these pathways by Gibbs sampling. The process of finding hub genes was the same with the identification of differentially expressed pathways. Firstly, the genes in the identified pathways were transformed into Markov chains. Secondly, Gibbs sampling was performed to obtain a new Markov chain and the posterior values of these genes. Thirdly, the probability (α) of occurrence of each gene was calculated. Then, the P-values were calculated by t-test according to gene expression differences between high group and low group and the genes were ranked basing on the P-value. Next, the correction coefficient \( c \) was calculated and the adjusted probability of each gene was calculated. Finally, the genes with \( α_{adj} ≥0.8 \) were selected as hub genes.

Results

Data preprocessing and pathway enrichment. The transcription profile was downloaded from the ArrayExpress database, including 12 samples from healthy Chinese female with extremely low PBM and 14 samples from healthy Chinese female with extremely high PBM. After pretreatment, we obtained a total of 20,514 genes used for the next analysis.

All human pathways (287) were downloaded from the KEGG database and 280 pathways were selected by screening pathways with a gene intersection ≥5.

Key pathways screened by Gibbs sampling. The posterior probabilities of the 280 pathways were obtained by Gibbs sampling and then the probabilities of the 280 pathways were calculated and adjusted. Fig. 1 shows the probability distribution of the 280 pathways. The blue line indicates the threshold value of probability of 0.80. All 280 pathways were ranked according to their adjusted probability and the top five pathways with the highest \( α_{adj} \) are listed in Table I. Under the criterion of \( α_{adj} ≥0.8 \), neuroactive ligand-receptor interaction (hsa04080) which ranked at first with \( α_{adj} = 0.986 \) was screened out as the key pathway. The expression levels of hsa04080 in the high and low groups are shown in Fig. 2. We can identify that the expression level of hsa04080 was obviously decreased in the high group compared to that in the high group.

Hub genes in the key pathway. By analyzing the key pathway identified previously, we found 309 genes in the neuroactive ligand-receptor interaction. Gibbs sampling was performed again on these pathway genes. The probability distribution of genes are shown in Fig. 3. The blue line represents the threshold value of probability of 0.80. All 309 genes were ranked according to their \( α_{adj} \) and Table II shows the top five genes with the highest \( α_{adj} \) value. Neurotensin (NTS), tachykinin receptor 3 (TACR3), follicle-stimulating hormone receptor (FSHR) with \( α_{adj} \) value ≥0.8 were considered as the hub genes. Of the three genes, TACR3 gained the highest \( α_{adj} \) value (0.930). The expression levels of the three genes were analyzed (Fig. 4). The expression level of NTS was reduced in the low group compared to that in the high group.
The expression level of TACR3 was significantly increased in the low group compared with that in the high group (Fig. 4A). By contrast, the expression level of FSHR was significantly decreased in the low group compared to the high group (Fig. 4C). Heatmap of the three hubgenes is shown in Fig. 5, and the results were consistent with those of Fig. 4.

**Discussion**

Circulating monocytes, which possess the potential to differentiate into osteoclasts and to produce potent pro-inflammatory and anti-inflammatory cytokines, are likely to play vital roles in bone metabolism (5). Previously, the analysis of key genes that may be associated with osteoporosis in circulating monocytes has been conducted by a traditional method based on Package linear models for microarray data of R (13). Herein, an MCMC algorithm based on Bayesian theory (Gibbs sampling) was performed for the first time to identify differentially expressed pathways and hub genes which probably associated with osteoclasts in circulating monocytes.

After Gibbs sampling, neuroactive ligand-receptor interaction (hsa04080) with the highest $\alpha_{adj}$ (0.986) was identified as a differentially expressed pathway in circulating monocytes that may relate to osteoporosis development. A total of
309 genes were found in this pathway and Gibbs sampling was performed again to identify the key genes among these 309 genes. Finally, three genes (NTS, TACR3 and FSHR) with $\alpha_{adj}>0.8$ were identified as hub genes that may be associated with osteoporosis. NTS is a neuropeptide composed of 13 amino acids (14) found in the brain and secreted mainly by neurons and astrocytes (15). Later, researchers found that NT was also expressed in the intestine (16). NTS in the central nervous system was associated with Parkinson's disease, eating disorders, schizophrenia, substance dependence and neurodegeneration, sexually transmitted diseases and other neuropsychiatric disorders (17,18). In the intestine, it involved in regulating intestinal function. In the present study, NTS was differentially expressed in circulating monocytes of high group and low group and was identified as a hub gene by Gibbs sampling. However, only few reports have demonstrate the relationship between NTS, circulating monocytes and osteoporosis.

**TACR3**, a member of genes which function as receptors for TACs, are mainly detected in the central nervous system and play vital roles in physiological development and human reproductive system (19,20). Recently, it has been reported that the expression of TACR3 was significantly evaluated in Human Oral Squamous Cell Carcinoma compared with normal epithelium (19). However, the roles of TACR3 in circulating monocytes remain unknown. Further studies of the role of NTS and TACR3 in circulating monocytes may uncover new mechanism and provide new target for osteoporosis.

FSH is a heterodimeric glycoprotein which plays a role in controlling follicle development and steroidogenesis by binding to the FSHR located in ovarian granulosa cells (21). Allan et al reported that FSH could increase bone mass in female mice (21). By contrast, Wang et al demonstrated FSH serum concentrations were increased in postmenopausal women with osteoporosis suggesting that FSH may increase the risk of postmenopausal osteoporosis (22). Additionally, FSH receptors were found in osteoclasts cultured in vitro and may promote osteoporosis in postmenopausal women (23). Furthermore, it has been reported that lacking either FSH β or FSHR in female mice could protect from bone loss despite hypogonadism (24). In the present study, FSHR was identified as a hub gene in circulating monocytes and its expression level in low group is significantly lower than that in high group, indicating that this gene may closely relate to the development of osteoporosis and it has the potential to be a therapeutic target for osteoporosis.
In conclusion, we identified a differentially expressed pathway (neuroactive ligand-receptor interaction) and three hub genes (NTS, TACR3 and FSHR) by analyzing the expression profiling of human circulating monocytes using Gibbs sampling. These findings provide insights into understanding the role of circulating monocytes in osteoporosis development. However, further experiments are needed to confirm these findings.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

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

Authors' contributions

YX conceived the study, acquired and analyzed the data, and drafted the manuscript. XPZ helped to explain the results and revised the manuscript. Both authors approved the final study.

Ethics approval and consent to participate

Not applicable.
References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.