Abstract: Objective: Osteoporosis (OP) can result in low bone mineral density (BMD) and reduced bone strength. This disease has been identified as a major public health problem around the world. Hence, it is necessary to find a proactive method, identifying high-risk OP patients, and to investigate the corresponding pathology. The objective of the current study was to reveal key pathways and hub genes associated with OP, utilizing Gibbs sampling. Methods: Informative pathways (IPs) with genes more than 5 were extracted, based on the KEGG database and microarray profiles. Obtained IPs were then converted into the Markov chain (MC). Afterward, Gibbs sampling was implemented to obtain a new MC. Subsequently, probabilities of IPs were counted through the MC Monte Carlo (MCMC) algorithm, followed by detection of differentially-expressed pathways (DEPs) based on adjusted probabilities of IPs higher than 0.65. Moreover, genes enriched in the DEPs were analyzed using the same sampling strategy. Hub genes were identified based on the threshold of adjusted probabilities greater than 0.80. Results: When the gene set was 5, a total of 278 IPs were extracted. After Gibbs sampling, only 1 DEP, mineral absorption, was identified according to the adjusted alfa.pi > 0.65. Moreover, after the probability of genes in this DEP was evaluated, a total of 8 hub genes were screened out, including VDR, ACP6, FTCD, ALDOB, ATIC, ALDH3A1, MAPK3, and OXCT2. Conclusion: Comprehensive approaches, including Gibbs sampling and Markov chain, might provide good reference for OP treatment in the future. Identified DEPs and hub genes might play pivotal roles in onset and progression of OP. They may be helpful in the development of available therapeutic drugs for treatment of OP in the future.

Keywords: Osteoporosis, Gibbs sampling, hub genes, differentially-expressed pathways

Introduction

Postmenopausal females have a high incidence of osteoporosis (OP). This is due to the co-existence of many independent predisposing factors, including estrogen deficiency, calcium loss, and aging [1-3]. OP, characterized by an imbalance between bone resorption and bone formation [4], results in low bone mineral density (BMD) and reduced bone strength. It also increases the risk of fragility fractures [5]. Significantly, OP is a major public health concern worldwide, particularly in China [6]. Currently, treatment of OP is mainly dependent on drugs, yet they come with a high cost. They are time consuming and have many side effects. Moreover, the curative effects are not ideal.
Differentially-expressed pathways in osteoporosis

Estrogen, excites antibody production by B-cells [14]. Compared to 17ß-Estradiol, bazedoxifene (an estrogen receptor modulator) restrains B lymphopoeisis generation at a later stage of B-cell differentiation [15]. Pineda et al. [16] attempted to reproduce one of the major risk factors (estrogen deficiencies after menopause or bilateral ovariectomy) for OP in women. However, the roles of B-cells in bone metabolism and OP remain largely unknown, especially at the systematic gene expression level in humans in vivo. At present, microarray technology is broadly used to detect key gene biomarkers for OP. Therefore, more investigators are using bioinformatics strategies to investigate the molecular mechanisms of OP, studying the microarray profiles of OP. GSE7429 is one of the microarray profiles of OP and was deposited by Xiao et al. [17]. They detected underexpression of ESR1 and MAPK3 in B-cells regulating the factor secretion, causing increased osteoclastogenesis or reduced osteoblastogenesis. In 2015, using the same data deposited by Xiao et al. [17], Yan et al. [18] extracted 238 differentially-expressed genes (DEGs) which were involved in OP. These included MAPK3, MAP3K10, MAP3K9, COX10, COX15, ATIC, UMPs, and HPRT1. Ma et al. [19] utilized the same gene expression data to identify several crucial genes associated with OP, including CSTA, TUBA1B, and CCNE1. However, thus far, most studies involving OP have paid attention to several important genes. Of note, several of these gene signatures have poor reproducibility and overlapped among different studies, although using the same microarray profiles. Consequently, understanding the complex interaction among genes is very challenging. Generally, detecting pathways participated in a given phenotype is very important. As demonstrated, signaling pathways, instead of single genes, govern the process of diseases [20]. Thus, it is important to identify potential pathways related to OP, further exploring the pathology of OP. Gibbs sampling, a Markov chain Monte Carlo (MCMC) algorithm obtaining a sequence of observations that are approximated from a specified multivariate probability distribution [21], can be used to identify differentially-expressed pathways (DEPs).

In the current study, integrated approaches of Gibbs sampling and MC were utilized to predict hub genes and pivotal pathways of OP. This study converted 278 IPs based on gene set as

Microarray data

Raw microarray data concerning PMOP (accession number: GSE7429) [17] were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) based on the platform of GPL96 [HG-U133A] Affymetrix Human Genome U133A array. A total of 20 samples were available, including 20 B-cell samples isolated from whole blood, obtained from 10 postmenopausal females with low BMD, aged 56-60 years. Also, there were 10 samples having high BMD, aged 54-60 years. Inclusion criteria for PMOP were: Spine or hip Z-score < -0.84 for the low BMD group; Spine or hip Z-score > 0.84 for the high BMD group. This study was approved by the Institutional Review Board and informed consent was obtained, as mentioned in the article by Xiao et al. [17].

Probe IDs due to concentrated expression levels were transformed into gene symbols. Duplicated genes were then removed in the matrix. Overall, 12,437 genes were obtained for subsequent analysis.

Biological pathways

Kyoto Encyclopedia of Genes and Genomes (KEGG) database (www.genome.jp/kegg/) offers a reference knowledge base for understanding cellular processes. First, this study collected 300 original pathways (6919 human genes) from the KEGG database and named them as OPs. Next, this study mapped the

Materials and methods

In the current study, Gibbs sampling was utilized to explore the significance of pathways, examining their roles in OP. Gibbs sampling, a means of statistical inference, especially Bayesian inference, is an MCMC algorithm used to obtain a sequence of observations. These are approximated from a specified multivariate probability distribution [22-24].
Differentially-expressed pathways in osteoporosis

**Identification of DEPs**

Empirically, the cut-off threshold of posterior probability was set as 0.05, which suggested the reliability of this sampler [28]. However, there was no obvious standard for high frequencies. As reported in a previous study [29], if the probability of a biological process was > 0.6, it was considered to be differentially-expressed. Thus, in the current study, relying on probabilities of IPs > 0.65, DEPs were identified. Moreover, genes in the differentially-expressed MF were believed to be DEGs. These DEGs were then merged. Statistical analysis was implemented on these merged genes that appeared in differentially-expressed MF.

**Selection of hub genes in DEPs**

The gene set enriched in DEPs was obtained, as with DEP analysis using Gibbs sampling. This study also implemented the same analysis for the pathway gene set. When the threshold of adjusted was set to 0.80, hub genes were identified.

**Results**

**Identification of IPs**

Under the criteria of IPs with at least 5 microarray genes, 278 IPs were identified. Of these IPs, 213 pathways displayed a gene number between 1 and 99. A total of 59 categories had more than or equal to 100 genes, but less than 200 genes. Four terms ranged from 200 to 300 genes and 2 terms had more than 300 genes (including hsa05200: Pathways in cancer with 366 genes and hsa04151: PI3K-Akt signaling pathway with 310 genes). Additionally, IPs with possessing gene count > 200 are shown in Table 1.

**Detecting DEPs**

Before assessing the probabilities of IPs based on Gibbs sampling using the MCMC algorithm, this study transformed the 278 IPs in expression data structure to the MC dataset. **Figure 1**

Table 1. Informative pathways (IPs) with ≥ 200 genes for osteoporosis

<table>
<thead>
<tr>
<th>IPs</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa05200: Pathways in cancer</td>
<td>366</td>
</tr>
<tr>
<td>hsa04151: PI3K-Akt signaling pathway</td>
<td>310</td>
</tr>
<tr>
<td>hsa04080: Neuroactive ligand-receptor interaction</td>
<td>281</td>
</tr>
<tr>
<td>hsa05166: HTLV-I infection</td>
<td>234</td>
</tr>
<tr>
<td>hsa04060: Cytokine-cytokine receptor interaction</td>
<td>233</td>
</tr>
<tr>
<td>hsa04010: MAPK signaling pathway</td>
<td>231</td>
</tr>
</tbody>
</table>

microarray genes (12,437 genes) to the OPs. As reported, pathways with too few genes may not have sufficient biological information [25]. Thus, a set of pathways was extracted by excluding pathways with less than 5 microarray genes. When removing the pathways having gene sizes < 5, a total of 278 informative pathways (IPs) were identified for further analysis.

In performing Gibbs sampling, the IPs needed to be converted into a data set with functional class expression measurements that were Markov chains (MCs). This was conducted using the Annotation Modified and Faster Global Optimization (MFGO) function of the Bayesian Approach for GeneSet Selection (BAGS) package [26].

**Calculation of probabilities of IPs**

After IPs were converted into MCs, posterior inferences for them were defined to measure probability distributions of IPs from OP [27].

In the current study, an empty set was first defined. Next, the MC dataset was deposited, including IPs with genes > 5 (N = 278) to this empty set. Afterward, Gibbs sampling was performed to establish the 10,000 dimensional-random vectors of N samples. Subsequently, these 10,000 dimensional were initiated into random vectors. Of these, one vector was extracted each time to produce the random number. This process was repeated 10,000 times. A new MC dataset, i.e. 10,000 probability of each IP, was received. Using the following formula, the probability of N IPs was computed:

\[
\text{Alfa.pi} = \text{the average probability of IPs (from 2000 to 10000)} \times 250/(10000 - 2000 + 1)
\]

In this equation, “alfa.pi” denotes “posterior value of an MF”. Afterward, researchers calculated the adjusted “alfa.pi” values for each IP based on the parameters of R values, P values, and rank order. Specifically, Student’s t-test was used to calculate the P-values for each IP. The rank order of each IP was obtained based on the P-values. Next, R-values were computed based on the rank order and “alфа.pi” values.
Differentially-expressed pathways in osteoporosis

Figure 1. Probabilities for 278 informative pathways (IPs).

Figure 2. Expression levels of mineral absorption in high BMD and low BMD groups. From this figure, it can be seen that this pathway was downregulated in low BMD group.

41 genes within the DEP of mineral absorption. Figure 3 illustrates the association of the probability distribution and each gene enriched in the DEP. Using the cut-off criteria of the adjusted alfa.pi > 0.80, 8 hub genes were extracted, including VDR, ACP6, FTCD, ALDOB, ATIC, ALDH3A1, MAPK3, and OXCT2.

Specific information is listed in Table 2.

Discussion

Gibbs sampling has been broadly utilized as way of statistical inference, including Bayesian inference [30]. Of note, Gibbs sampling, a Markov Chain Monte Carlo (MCMC) algorithm, can obtain a sequence of observations, approximated from a specified multivariate probability distribution [22-24]. Remarkably, on the basis of the probabilities, differentially-expressed biological processes and key genes are potentially identified. This might be important in revealing the pathology of disorders. Hence, in the current analysis employed Gibbs sampling to evaluate the significance of pathways, examining their function in OP. Consequently, only 1 DEP, named as mineral absorption, was identified according to the adjusted alfa.pi > 0.65. Moreover, after the probability of genes in this DEP was evaluated, a total of 8 hub genes were screened out, including VDR, ACP6, FTCD, ALDOB, ATIC, ALDH3A1, MAPK3, and OXCT2.

As demonstrated, calcium is the dominant mineral in bones. It has been regarded as a shortfall nutrient reported in the Dietary Guidelines
Differentially-expressed pathways in osteoporosis

Table 2. List of hub genes based on the adjusted alfa.pi > 0.80

<table>
<thead>
<tr>
<th>Hub gene</th>
<th>P value</th>
<th>Rank p</th>
<th>R value</th>
<th>alfa</th>
<th>Alfa-adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>VDR</td>
<td>0.016282</td>
<td>8</td>
<td>0.972318</td>
<td>0.905137</td>
<td>0.880081</td>
</tr>
<tr>
<td>ACP6</td>
<td>0.011281</td>
<td>4</td>
<td>0.986159</td>
<td>0.886014</td>
<td>0.873751</td>
</tr>
<tr>
<td>FTCD</td>
<td>0.019950</td>
<td>12</td>
<td>0.958478</td>
<td>0.911511</td>
<td>0.873663</td>
</tr>
<tr>
<td>ALDOB</td>
<td>0.012046</td>
<td>5</td>
<td>0.982699</td>
<td>0.873266</td>
<td>0.858157</td>
</tr>
<tr>
<td>ATIC</td>
<td>0.017116</td>
<td>10</td>
<td>0.965398</td>
<td>0.879647</td>
<td>0.849203</td>
</tr>
<tr>
<td>ALDH3A1</td>
<td>0.047386</td>
<td>29</td>
<td>0.899654</td>
<td>0.905137</td>
<td>0.814310</td>
</tr>
<tr>
<td>MAPK3</td>
<td>4.00E-05</td>
<td>1</td>
<td>0.99654</td>
<td>0.815898</td>
<td>0.810375</td>
</tr>
<tr>
<td>OXCT2</td>
<td>0.034067</td>
<td>20</td>
<td>0.930796</td>
<td>0.873266</td>
<td>0.812832</td>
</tr>
</tbody>
</table>

for Americans [31]. Enhanced calcium intake has been suggested to be related to increased bone accrual [32]. Significantly, promoting calcium absorption, as well as other bone-related minerals, is an attractive strategy in reducing risks of OP [33]. Legette et al. [34] implicated that, during middle-age, mineral absorption decreases and bone resorption rates increase. Results suggested that this led to a higher risk for OP. Moreover, current results demonstrated that mineral absorption is very important for OP development.

VDR, a nuclear transcription factor, affects calcium absorption, mineralization, and bone remodeling [35]. Some studies have confirmed the correlation between VDR polymorphisms and occurrence of fractures. For example, VDR FocI and TaqI polymorphisms have been reported to be associated with low BMD at the lumbar spine and femoral neck, according to several studies [36-38]. Stathopoulou et al. [39] implicated that, under lower calcium intake (<680 mg/d), the presence of the B-allele of VDR Bsml polymorphisms enhanced the risk of OP by 118%. Thus, Horst-Sikorska et al. [40] concluded that adequate calcium intake “masked” the VDR genetic influence on bones. In the current study, VDR was the hub gene with the highest probability. Therefore, current results were in line with the above reports, suggesting an important role for VDR in OP development.

In general, the current study provides a comprehensive bioinformatics analysis (using the Gibbs sampling method), identifying DEPs and hub genes that might be involved in the development of OP. Current findings revealed that the pathway of mineral absorption might play key roles in the development of OP. In addition, present results might provide a better understanding for pathologies of OP, indicating potential targets for development of therapeutic drugs for OP in future. However, the current study had several disadvantages. The main drawback was the relatively small population. Another limitation is that additional experiments are necessary to confirm the results obtained above using bioinformatic approaches. Moreover, additional animal or tissue experiments should be conducted, validating current results using more samples based on Western blotting or PCR.

Disclosure of conflict of interest

None.

Address correspondence to: Yong-Kun Wei, Department of Orthopedics, 3201 Hospital, No. 783 Tianhan Road, Hanzhong 723000, Shaanxi Province, China. Tel: 86-0916-2383527; Fax: 86-0916-2383527; E-mail: weiyongkun2005@sohu.com

References

Differentially-expressed pathways in osteoporosis


[20] Sahakyan AB and Balasubramanian S. Long genes and genes with multiple splice variants are enriched in pathways linked to cancer and other multigenic diseases. BMC Genomics 2016; 17: 225.


Differentially-expressed pathways in osteoporosis


