Transcriptional Landscape Analysis Identifies Differently Expressed Genes Involved in Follicle-Stimulating Hormone Induced Postmenopausal Osteoporosis

Katre Maasalu¹,², Ott Laius¹, Lidiia Zhytnik¹, Sulev Köks³,⁴, Ele Prans³, Ene Reimann³,⁴, Aare Märtson¹,²

¹Department of Traumatology and Orthopedics, University of Tartu, Puusepa 8, Tartu 51014, Estonia
²Clinic of Traumatology and Orthopedics, Tartu University Hospital, Puusepa 8, Tartu 51014, Estonia
³Department of Pathophysiology, University of Tartu, Ravila 19, Tartu 50411, Estonia
⁴Department of Reproductive Biology, Estonian University of Life Sciences, Kreutzwaldi 1, Tartu 51014, Estonia

Corresponding author:
Katre Maasalu
Clinic of Traumatology and Orthopedics, University of Tartu
Puusepa 8, Tartu 51014, ESTONIA
katre.maasalu@kliinikum.ee
Tel: +3727356490
Fax: +3727356491
ABSTRACT

Osteoporosis is a disorder associated with bone tissue reorganization, bone mass and mineral density. Osteoporosis can severely affect postmenopausal women, causing bone fragility and osteoporotic fractures. The aim of the current study was to compare blood mRNA profiles of postmenopausal women with and without osteoporosis, with the aim of finding different gene expression and thus targets for future osteoporosis biomarker studies.

Our study consisted of transcriptome analysis of whole blood serum from twelve elderly female osteoporotic patients and twelve non-osteoporotic elderly female controls. The transcriptome analysis was performed with RNA sequencing technology. For data analysis, the edgeR package of R Bioconductor was used.

Two hundred and fourteen genes were expressed differently in osteoporotic compared to non-osteoporotic patients. Statistical analysis revealed 20 differently expressed genes with a false discovery rate of less than 1.47×10^{-4} among osteoporotic patients. The expression of 10 genes were up-regulated and 10 down-regulated. Further statistical analysis identified a potential osteoporosis mRNA biomarker pattern consisting of six genes: CACNA1G, ALG13, SBK1, GGT7, MBNL3, and RIOK3. Functional Ingenuity Pathway Analysis identified the strongest candidate genes with regard to potential involvement in a follicle-stimulating hormone activated network of increased osteoclast activity and hypogonadal bone loss. The differentially expressed genes identified in this study may contribute to future research of postmenopausal osteoporosis blood biomarkers.

KEY WORDS: bone, transcriptome, age, musculoskeletal, female, biomarkers
INTRODUCTION

Osteoporosis (OP) is a skeletal fragility disorder characterized by low bone mineral density (BMD), modification of bone tissue microarchitecture quality, and susceptibility to sudden fractures.\(^1\) Osteoporotic fractures, including those of the hip and spine, are often causes of a poor quality of life, disability, and increased risk of mortality among patients.\(^2\) Every year the prevalence of OP increases globally, resulting in new health care and financial concerns.\(^3\)

Postmenopausal women face the largest challenge of bone loss, due to changes in levels of reproductive hormones.\(^4,5\) Postmenopausal osteoporosis (PMOP) is associated with a decrease of estrogen (ESR), and an increase of follicle-stimulating hormone (FSH) and luteinizing hormone (LH).\(^6\) FSH influences bone mass both indirectly and directly, via ESR and an extracellular signal-regulated kinase–mitogen activated protein kinase (Erk/Mek) signaling pathway with Gi2a stimulation of MEK/Erk, the nuclear factor κβ (NF-κβ) and 3-kinase-Akt, respectively.\(^7\)

An understanding of OP mechanisms is crucial for effective disease prevention, diagnosis, and therapy. OP diagnosis and fracture risk estimation is based on a BMD scale (T-score < \(-2.5\) SD), however, OP fractures might occur among those at a moderate risk.\(^8\)\(^–\)\(^10\)

Evaluation of bone quality and fracture risk remains a research area of great interest. Previous studies have concentrated on miRNA signatures and bone turnover biochemical markers of OP.\(^11\)\(^–\)\(^13\) Whole-genome RNA sequencing (RNA-seq) is a powerful tool for investigating the pathological pathways of complex disorders. To the best of our knowledge, there are no previous studies of whole blood mRNA transcriptome analysis among postmenopausal osteoporotic patients. Total blood mRNA shares about 80% of transcriptome with other major tissues.\(^14\) mRNA reflects the functional state of cells, and integrates responses to both genetic
and epigenetic factors of gene regulation, making it a promising way to explore disease progression.\textsuperscript{15–17}

In the present study we report our whole blood RNA-seq transcriptome analysis of 12 elderly postmenopausal osteoporotic and 12 elderly non-osteoporotic females. Differential expression analysis was combined with functional network annotation. As a result, we found a pattern of differently expressed genes (DEGs) that are potentially involved in direct FSH osteoclastogenesis and a bone resorption activation pathway. Our findings might be of interest as new targets for future research of PMOP biomarkers that could result in more effective diagnosis and follow up of OP.

\textbf{METHODS}

\textbf{Patients and controls}

The selection of female individuals for transcriptome analysis was based on bone mineral density. OP patients were selected from the bone densitometry database of the Clinic of Traumatology and Orthopedics, Tartu University Hospital. All selected osteoporotic patients had relatively similar spine BMD T-scores (Table 1b).

We recruited a control group of 12 postmenopausal females with normal BMD from individuals who underwent densitometry testing during regular health care screening. Exclusion criteria for participation in the control group were a history of previous fractures, and disease or medications that can affect bone quality. We performed age and BMI matching among patients and controls in order to reduce an influence of these factors on the transcriptome analysis. Mean age, height, weight, and body mass index (BMI) values were calculated for both the control (KO) and OP groups (Table 1a).
As half of the patients had low BMD only in spine and others in both spine and hip measured regions we created two subgroups for further testing. Group A consisted of patients with osteoporosis only in the spine (T-score spine $-2.92$ SD, hip $0.9$ SD), and Group B with both lower spine ($-2.97$ SD) and hip ($-2.61$ SD) osteoporosis (Table 1b).

The protocols and informed consent form used in this study were approved by the Ethical Review Committee on Human Research of the University of Tartu (permit № 221/M-34). All participants gave written informed consent.

**Sample collection and RNA extraction**

Tempus Blood RNA Tubes (Applied Biosystems, Life Technologies Corp., Carlsbad, CA, USA) were used for collecting the samples of whole blood. Total RNA extraction from whole blood was achieved using a Tempus Spin RNA Isolation Kit (Ambion, Life Technologies Corp., Carlsbad, CA, USA). As total RNA from whole blood consists of up to 70% immunoglobulin mRNA, a GLOBINclear™ Kit (Ambion, Life Technologies Corp., Carlsbad, CA, USA) was applied to purify the samples of globin mRNA. The quality of total RNA was evaluated with an Agilent 2100 Bioanalyzer and RNA 6000 Nano kit (Agilent Technologies Inc., Santa Clara, CA, USA). The average RNA integrity number (RIN) of the samples was at least 7.

**Whole Transcriptome RNAseq library preparation and sequencing**

50ng of each total RNA sample was amplified by applying the Ovation RNA-Seq System V2 (NuGen, Emeryville, CA, USA), after which SOLiD 5500 Wildfire (W) System chemistry (Life Technologies Corp., Carlsbad, CA, USA) was used to prepare the resulting cDNA for the DNA fragment library. Next, the 12 libraries were pooled together in equal amounts to construct two
different library pools. The pooled libraries were converted to SOLiD 5500W libraries, and sequencing was performed using a SOLiD 5500W platform and DNA sequencing chemistry (Life Technologies Corp., Carlsbad, CA, USA). Three-lane sequencing was applied and 12 libraries per lane sequenced. 75bp from a forward direction were sequenced, which altogether gave at least 30 million mappable reads per sample, i.e. sufficient for evaluation of the expression pattern of the transcriptome.

**Statistical and functional analysis**

Raw reads and whole transcriptome analysis workflow were mapped using Lifescope 2.5.1 software (Life Technologies Corp., Carlsbad, CA, USA). This workflow generates a very complex output, including gene and exon counts, alternative splicing, and fusion transcripts. For further analysis, we focused only on gene counts, because our primary question was related to the abundance of gene-targeted transcripts. For differential expression analysis the R Bioconductor package edgeR was used, which implements exact statistical methods and generalized linear models for multi-group and multifactorial experiments. A feature of edgeR is an empirical Bayes method that permits the estimation of gene-specific biological variation, even for experiments with minimal levels of biological replication. EdgeR can be applied to differential expression at gene, exon, or tag level. In our study we used model-based normalization and applied a negative binomial model. Testing for differential expression was done using the exact test. Power analysis was performed with the RNAseqPS web tool. Heatmap clustering analysis was generated with the gplots package in R. Network and pathway analysis was generated with QIAGEN’s Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, CA, USA) software and the iPathwayGuide online tool (Advaita Plymouth, MI, USA).
Additionally, we tested the sensitivity of the DEGs as biomarker candidates of patients with osteoporosis, by comparing the mRNA patterns among OP patients’ subgroups, and controls using edgeR statistical analysis.

RESULTS

Differential expression analysis

We performed whole blood mRNA-seq analysis to investigate the transcriptional profiles of non-OP and PMOP females. Whole transcriptome mRNA sequencing analysis identified 214 DEGs with a confidence level of a false discovery rate (FDR) of less than 0.05. Of the 214 genes, 154 were down-regulated. The twenty strongest DEGs (FDR less than 1.47×10^{-4}) present in the osteoporotic patients and absent from the control group are shown in Table 2. The expression of ten DEGs was up-regulated and ten down-regulated.

The statistical analysis of mRNA expression levels of the all OP patients using the edgeR Bioconductor package, revealed six candidate genes (CACNA1G, ALG13, SBK1, GGT7, MBNL3, and RIOK3) as the strongest candidates for a potential mRNA biomarker pattern indicative of osteoporosis. Subgroup analysis, showed no significant difference in expression of the previously identified 20 candidate osteoporosis biomarkers between the subgroups.

Further analysis revealed a significant correlation between the expression of possible mRNA biomarkers and BMD values for OP patients compared to controls (Figures 1 & 2). The highest correlation of spine BMD was observed with the CACNA1G gene expression ($R^2=0.7842$, $p=5.03\times10^{-7}$). Correlation between lumbar spine BMD and gene expression was more significant than the correlation between hip BMD and gene expression. On scatter plots of spine
BMD, patient and control groups were clearly distinguishable (Figure 1). On scatter plots of hip BMD, three groups formed (controls, only spine OP, and spine and hip OP). Compared to the controls, the expression of genes among group B was more different than among group A. The FDR values of the potential mRNA OP biomarkers of subgroup B were also significantly lower than for group A (Table 3).

DEGs with statistically significant values were clustered with Heatmap analysis (Figure 3). The horizontal axis shows clustering within the two BMD groups (KO with high BMD, OP with low BMD). Differences between gene expression in the control (KO) and OP patient groups were clearly observed (Figures 3 & 4). The statistical power of the performed transcriptome analysis was represented using receiver operating characteristic (ROC) curves for the four strongest candidate genes (GGT7, SBK1, ALG13, and CACNA1G) (Figure 5). All four DEGs had high predictive power of an osteoporosis, with few or no false positives. Accuracy for the strongest candidates ALG13 and CACNA1G was 1.

**Functional analysis**

Network pathway analysis of the 20 genes with the highest FDR values, revealed the potential involvement of the DEGs in the calcium signaling pathway and ERK/MAPK signaling pathways.

Involvement of the identified PMOP DEGs in these connective tissue disorder pathways was also found in silico using QIAGEN’s Ingenuity software. IPA analysis showed that the DEGs were involved in cell growth and proliferation pathways, and molecular transport. IPA analysis also highlighted the involvement of the DEGs in the calcium signaling pathway and ERK/MAPK signaling pathway, Akt pathway, NF-κβ and FSH network (Figure 6).
DISCUSSION

Osteoporosis alters bone tissue metabolism pathways, which in our study manifested in changes in the mRNA levels of related genes in blood cells, and resulted in a special PMOP gene pattern of differential expression. Six candidate genes (CACNA1G, ALG13, SBK1, GGT7, MBNL3, and RIOK3) were the strongest candidates for a potential mRNA biomarker pattern indicative of osteoporosis in postmenopausal females. Subgroup analysis, showed no significant difference of candidate gene expression between the subgroups, which is expected according to basic knowledge of bone metabolism. Osteoporosis affects entire skeleton and areal BMD differences are connected to differences of bone shape, cortical-trabecular frame and metabolic activity, influenced by physical activity and life style of the individual.

The strongest candidate gene was the alpha 1G subunit of the voltage-dependent calcium channel CACNA1G (FDR 7.75×10^{-69}) and was the most highly up-regulated (logFC 2.502). The CACNA1G gene is involved in the bone morphogenetic protein (BMP) pathway, bone tissue mineralization, intracellular Ca signaling, and the Wnt β-catenin pathway. BMP and Wnt β-catenin pathways are important for osteoblast differentiation and bone formation. Depending on the form of alteration, changes to the BMP and Wnt β-catenin pathways can lead to bone fragility of different severities. We surmise that differential expression of the CACNA1G gene might also reflect alterations in bone tissue metabolism.

In accordance with previous postmenopausal osteoporosis mRNA expression studies in circulating B cells, we found DEGs connected to the ERK/MAPK pathway. The estrogen receptor 1 (ESR1) and mitogen activated protein kinase 3 (MAPK3) network has been proposed as a cause of increased osteoclastogenesis and decreased osteoblastogenesis. However, the IPA analysis of the discovered PMOP profile from our study identified involvement of the DEGs in the FSH–ERK/MEK (MAPK) network, which is non-estrogen dependent. Our
findings support those of a previous study of postmenopausal osteoporosis in haploinsufficient FSH+/−-mice, which showed activation of Gi2a-coupled FSH receptors stimulated MEK/Erk, NF-κβ, and Akt, and resulted in increased osteoclast activity and hypogonadal bone loss. FSH induced Gi2a, MEK/Erk, NF-κβ, and Akt signaling pathways are well-known osteoclast stimulating pathways. Recent investigations have also highlighted an FSH-dependent PMOP mechanism, caused by elevation of FSH and LH levels in elderly females.

Although osteoporosis is connected with the aging process, the similar mean ages of the control (70.2) and osteoporosis patient (70.6) groups would likely exclude the possibility that the discovered mRNAs were a result of “aging” transcriptomes. We are confident that our results reflected a connection between the revealed candidate mRNA biomarkers and bone tissue reorganization. The matched body mass index (BMI) values of the groups also points to body weight being insignificant in terms of differences in the identified gene expression pattern of OP. The average BMD T-score of the control group for total hip (−0.19) and lumbar spine (0.26) showed high bone quality, and allowed for more sensitive distinguishing of the contrasts between the mRNA expression patterns of OP patients and control subjects.

The reliability of our results might seem limited by the small number of study individuals, but the high power analysis value of 0.9 of the data should give cause for confidence in the soundness of our results. Furthermore, the transcriptome analysis power of 12 samples is 0.7 and 0.8 for both weak and strong compounds respectively, which is sufficient to reveal differentially expressed genes.

Blood cells do not express all bone cell proteins, thus some protein translation changes may have gone unnoticed during our study. Nonetheless, the present study revealed a connection between whole blood mRNAs and FSH, and postmenopausal bone loss in humans.
CONCLUSION

In our study we investigated whole transcriptome RNA-sequencing of the blood serum of postmenopausal osteoporotic Estonian females, with the aim of revealing a candidate mRNA biomarker pattern for osteoporosis. We discovered a pattern of differently expressed mRNAs of OP that consisted of six genes: *CACNA1G, ALG13, SBK1, GGT7, MBNL3*, and *RIOK3*. This transcriptional landscape was connected to FSH induced Gi2a, MEK/Erk, NF-κβ, and Akt signaling pathways, which are known to directly activate osteoclastogenesis and stimulate postmenopausal bone loss. The current findings may be useful for the development of a blood mRNA PMOP biomarker set which is a promising method of PMOP diagnosis and follow up. Further studies with larger numbers of independent cohorts of PMOP patients and controls are required.

ACKNOWLEDGMENTS

This work was supported by institutional research funding (IUT 20–46) of the Estonian Ministry of Education and Research, and the Centre of Excellence in Health Promotion and Rehabilitation (Project EU 41384 of the European Regional Development Fund).

CONFLICT OF INTEREST

Katre Maasalu, Ott Laius, Lidiia Zhytnik, Sulev Kõks, Ele Prans, Ene Reimann and Aare Märtson declare that they have no conflicts of interest.

AUTHORS’ ROLES

Study design: KM, SK, AM, OL. Study conduct: KM, OL, ER, EP, LZ. Data collection: KM, OL. Data analysis: KM, OL, LZ, SK. Data interpretation: KM, LZ, SK. Drafting the
manuscript: KM, OL, LZ, SK, AM. Revising the manuscript content: KM, OL, LZ, EP, ER, SK, AM. Approving the final version of the manuscript: KM, OL, LZ, EP, ER, SK, AM. KM, SK, OL take responsibility for the integrity of the data analysis.

REFERENCES


Table 1. Patients and controls. (a) Characteristics of healthy control group individuals and OP patients. (b) Mean values for subgroups of OP patients: group A (OP only in lumbar spine), B (lumbar spine and total hip OP).

Table 2. Candidate genes for osteoporosis mRNA biomarkers. The False Discovery Rate (FDR), log Fold Change (logFC) and $p$-values for the candidate genes are listed.

Table 3. False Discovery Rate (FDR), log Fold Change (logFC) and $p$-values of mRNA biomarkers present in both A and B patient’s subgroups compared to healthy controls

Figure 1. Correlation of candidate OP biomarkers gene expressions with spine BMD (T-score). Circles indicate OP patients; diamonds indicate controls.

Figure 2. Correlation of candidate OP biomarkers gene expressions with total hip BMD (T-score). Circles indicate OP patients; diamonds indicate controls.

Figure 3. Heatmap analysis of differently expressed genes. The clustering according to gene expression between control samples (KO) and osteoporotic samples (OP) is observed.

Figure 4. Differences in expressions of the strongest candidate OP biomarker genes between osteoporotic patients (turquoise) and healthy controls (red).

Figure 5. Mean ROC curves for the strongest candidate OP biomarkers gene expressions

Figure 6. Genetic networks revealed among potential OP mRNA biomarkers with Ingenuity pathway Analysis software. The functional analysis identified involvement of the differently expressed genes into connective tissue disorders and the “RNA Post-Transcriptional Modification, Molecular Transport, RNA Trafficking” network. Down-regulated and up-regulated genes are highlighted in green and red colors respectively.