European Review for Medical and Pharmacological Sciences 2013; 17: 2804-2809

Analysis report for osteosarcoma expression profile

C. LI, C. ZHAN, Y. CHEN, Q. FU, X.D. ZHU, D.W. HE, M. LI, Z.W. WANG

Department of Orthopedics, Changhai Hospital, Second Military Medical University, Shanghai, China
Cheng Li and *Ce Zhan* should be regarded as the co-first Authors

Abstract. – BACKGROUND: Osteosarcoma is a kind of highly malignant primary bone tumor which most common in the teenage, and holds strong aggressive, earlier organs metastases mainly to lung, prone to postoperative recur. Therefore for osteosarcoma, invasion and transfer mechanism and related factors' interaction remains to be a key research subject.

AIM: We aim to find biological molecules marker can be used for osteosarcoma diagnosis through contrast of osteosarcoma sample and normal tissue samples.

MATERIALS AND METHODS: This analysis using human osteosarcoma expression profile data and three lesions normal tissue samples (liver, kidneys, lymph) expression data and compare them, and find significant specifically expressed genes, according to their function.

RESULTS: Research shows that the cancer cell proliferation, invasion, transfer and recurrent process involve many factors interaction, of which angiogenesis is the necessary condition of tumor growth, transfer and the recurrence.

CONCLUSIONS: Now the most important positive regulatory factor of angiogenesis is VEGF (vascular endothelial growth factor) and bFGF (basic fibroblast growth factor). Both of them are with a wide variety and close relationship of tumor angiogenesis and progress.

Key Words:
Osteosarcoma, Expression profile, Angiogenesis, GJA1, COL1A2, COL5A2.

Introduction

Compared with other cancers, the incidence of osteosarcoma is quite low and it is prone to occur among youth, commonly pathogenesis from distal tibia and femoral proximal. It is the most common seen malignant, high malignant degree, rapid developing primary bone tumors. Without formal treatment, six months to a year it would happen tumor metastasis to the lungs, causing patients died.

Because of the fast growth and proliferation as well as the metastasis, 80% of the patients had been diagnosed of tiny lung metastases. Nearly 25 years, the treatment of osteosarcoma had great development, nearly 80% of the patients can keep the suffered limb, and the survival rate of five years has risen from 20% to 55% or 75%[1]. But however, still more than half of the survivors have been dead from metastasis and re-occurrence of osteosarcoma.

So far, the exact mechanism of osteosarcoma pathogenesis is unclear. Many pathogenic factors, such as chemical factors, radioactive substances, virus infection, all participate into the osteosarcoma pathogenesis.

Along with the molecular biology technologies' wide application in the study of tumor etiology, we hold some knowledge on the development of cancer process and the molecule basement of osteosarcoma. Hanahan et al[2] believe that the performance of cancer is based on ten major changes of fine physiology: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis.

This analysis using human osteosarcoma expression profile data and three lesions normal tissue samples (liver, kidneys, lymph) expression data and compare them, and find significant specifically expressed genes, according to their function of, in the hope that it can be a diagnosis of osteosarcoma a biometric marker.

Materials and Methods

Expression Profile of Osteosarcoma Related Genes

Obtain expression data GSE16088 [3] from GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/) database[3]. The

Corresponding Author: Ming Li, MD, Zhiwei Wang, MD; e-mail: mlzhwwang@hotmail.com

experiment contains 20 samples, of which 14 cases are diagnosed to be osteosarcoma patients and the other 6 are normal samples for control (two samples from Liver, kidneys, lymph individually).

Preprocess of Original Data

Firstly we used the affymetrix package of R language to transform the downloaded original data to expression form, and filled up the gap in data[4], normalized them with Median standardization method[5].

Analysis of Differential Expressed Genes

Using the three statistic methods from multtest package[6] in R language, *t*-test, Wilcox test[7] and rigorous examination[8], we performed rigorous expression analysis to case and control group and applied multiple test correction to P values obtained from each method. And we selected genes with $p < 0.05$ and |logFC (FC: fold change)| >1, and through multiple test and selection, we recognized differential expressed genes between osteosarcoma sick samples and the normal ones.

Construction of Interaction Network

STRING[9] can construct interaction network of known and predicted proteins. The data is from high throughput genome sequence and experiments, co-expression, published knowledge. Combining these differential expressed genes, we used String to construct protein interaction network.

Network Module Analysis

Splitting the whole network obtained before, and to find the function modules with the target genes included and perform GO (gene ontology) function annotation to them (Cytoscape and its pluigns). The threshold of super geometry distribution is 0.05 (as formulation described below). Cytoscape is a kind of network analysis tool, with Mcode[10] and Bingo[11] are its two plugins for module function analysis and module function annotation. When performing module function annotation, this software combined GO function database, i.e. it will connect to GO database for node search when aroused to run. The algorism is based on super geometry distribution and the result replied is the module function predicted with $p < 0.05$ and number of genes larger than 2. The selected significant function annotation is termed as the function of modules.

Results

Data Preprocess

Because of various problems in original chip data such as background, probe design and so on, there are great differences between chip data. So the normalization of data before analysis is necessary and critical.

Differential Expression Analysis

Using three kinds of statistic examine methods (*t*-test, Wilcox rank-sum test and Fisher rigorous examination), we examined the gene expression data and performed multiple correlation examination (Beyer-Harwick-BH method[12]), and then selected those genes with $p < 0.05$, |logFC (FC: fold change)| > 1 under all the three methods. Finally, we obtained four probes and put them into ID converter for gene names, which is listed in Table I and further analysis was performed.

Construction of Network and the Statistics of Nodes

The obtained three genes COL1A2 (collagen, type 1, alpha 2), COL5A2, GJA1 (gap junction protein alpha 1) as the core, we combined them with the prediction ability of STRING database to identify those proteins who can interact with them and construct the proteins interaction network map (Figure 1). In the map, we contain 103 proteins and 1249 pairs of interaction pairs. Within which, COL1A2 and 29 proteins have interactions, COL5A2 (collagen, tipe V, alpha 2) and 38 proteins have interactions, and GJA1 in-

Table I. List of differentially expressed genes.

ID_REF	*t*-test	adj	Wilcox	adj	Exact_test	adj	LogFC
GJA1	5.46E-14	1.11E-11	5.16E-05	0.000547	4.88E-06	0.027241	1.18444
COL1A2	1.03E-16	6.05E-13	0.000617	0.00138	2.24E-06	0.027241	1.207477
COL5A2	6.25E-16	1.23E-12	0.000617	0.00138	4.47E-06	0.027241	1.248982
COL5A2	2.46E-14	7.02E-12	0.000617	0.00138	4.89E-06	0.027241	1.226827

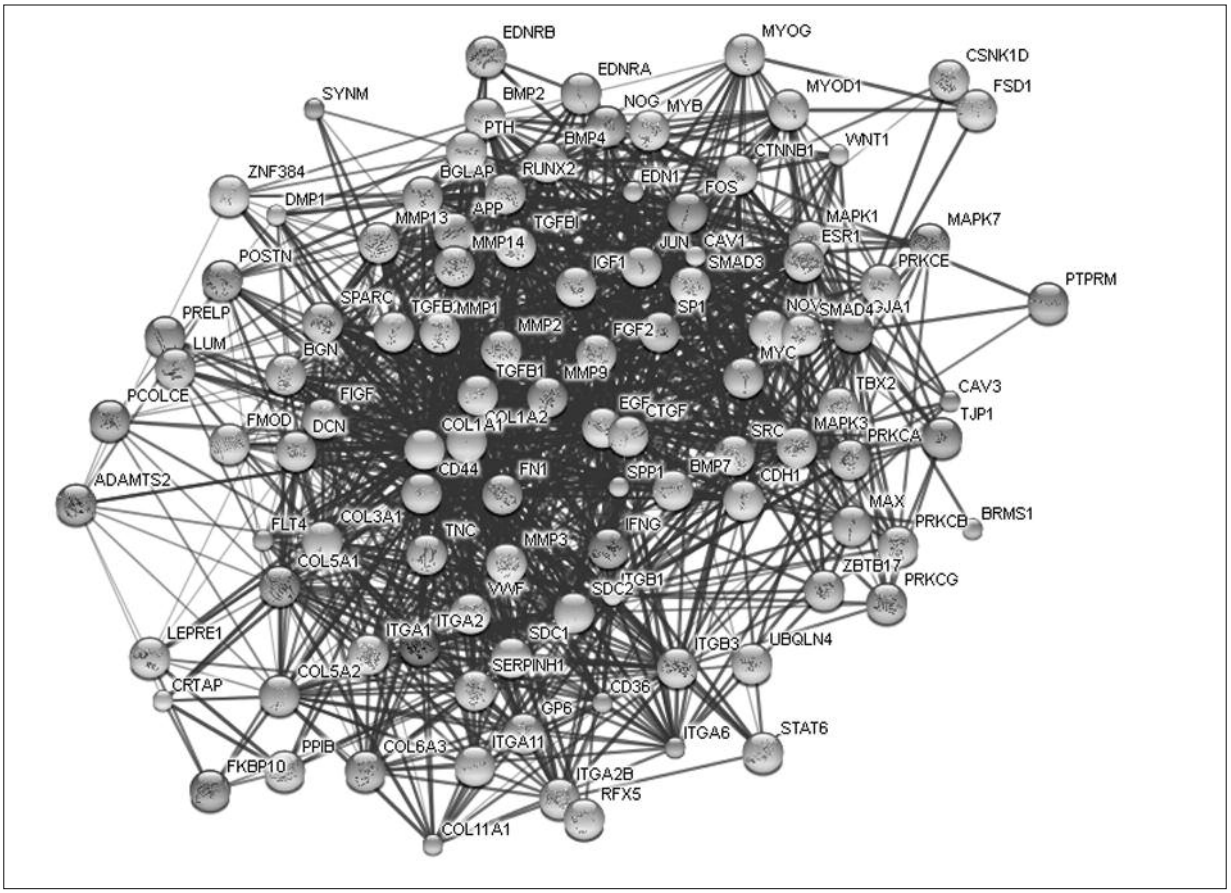

Figure 1. Map of interaction network. String intergrate protein information from several databases and can predict target genes with interactions through sequences of proteins and their space structures.

teracts with 34 proteins. There is some interaction among the proteins which respectively related to COL1A2 and COL5A2.

The interaction pairs in the network obtained are used for module analysis in the next step.

Analysis of Modules in the Network

Interaction network can be divided into many relatively independent son network, which usually has the proteins performing similar functions. This kind of closely relationship is called a function module. Identification of these modules is important for understanding the organization of biosystem structure. So, this analysis applied Cytoscape network analysis software to predict the interaction pairs from String database and carried out modules analysis[13]. With MCODE (molecular complex detection) plugin we unearthed the modules within which the three target genes located[10] (Figure 2), and then annotated the module functions using the plugin Bingo (Table II and Table III).

The parameters list when classifying module partitions are illustrated in Table IV. Parameters of each node in the module are required to be greater than 2 (degree cutoff = 2), each node in the module is scored to be larger than 0.2 (node score cutoff = 0.2), and the number of the adjacent nodes must be greater than 2, and this parameter is essentially as same as the first parameter (K-score = 2), and the requirement of the depth is 100, i.e. the node in division module whose distances to other nodes are all within 100 will be considered.

From the function annotation of these two modules, we can see the significant differentially expressed genes are all related to the signal transduction pathways.

Discussion

For quite some time, the researchers are insisting on the struggle against osteosarcoma, which is a harmful disease to teenagers. Just as Jeon[14]

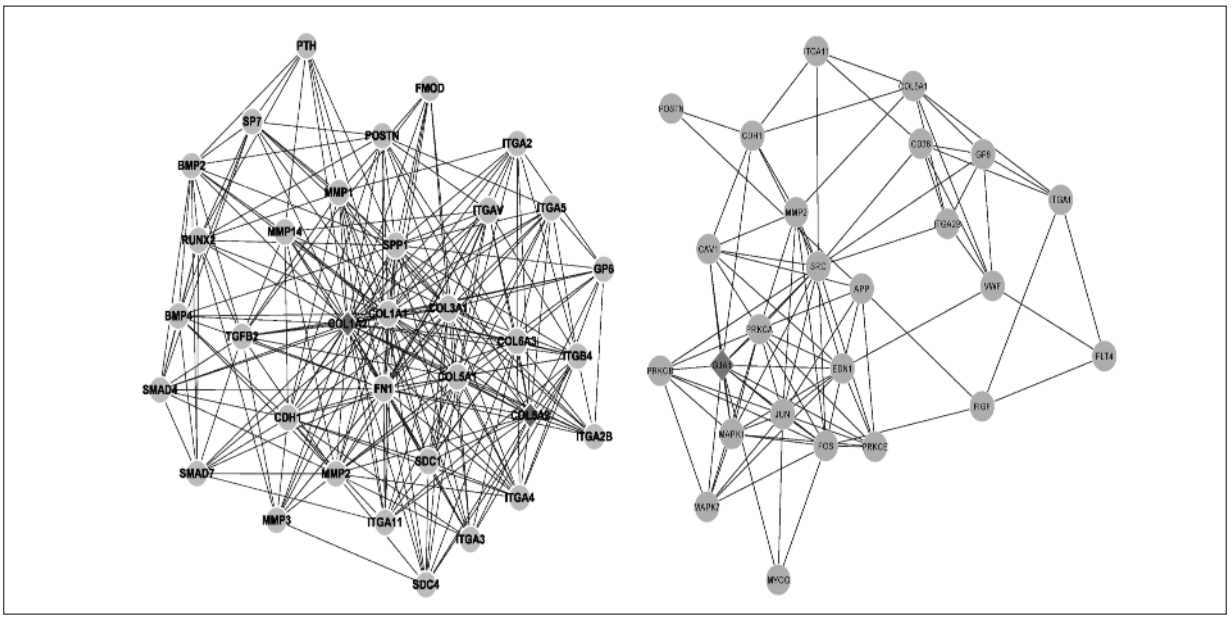

Figure 2. The module analysis using the Mcode. The yellow circle and purple ones represent the interactive genes function with the differentially expressed genes, and the red diamonds represent the target genes found. The left is the modules COL1A2 and COL5A2 belong to, and the right is the module GJA1 belongs to.

reported: 25 cases patients who accepted the primary malignant tumor lesion curettage, hold an increasing local recurrence rate, and the overall tumor-free survival rate was 65%. Because of the local recurrence reduced, the average survival rate of 5 and 10 years of these patients reduced[14]. Bramer et al[15] observed the alkaline phosphatase level of 89 cases adults with osteosarcoma pre- and post-chemotherapy, and predicted its relationship with chemotherapy and survival rate. Finally, they found the results found with poor prognosis when the alkaline phosphatase more than two times of the normal level.

Through the function analysis of the network subsidiary, the function of the modules which COL1A2 and COL5A2 located in are focused on integrin-mediated signaling pathway, biological adhesion (interferon related signal transduction pathways) biological process; at the same time, the main function modules GJA1 located in is focused on the signaling pathway (signal pathway).

These discoveries are matched with the research before. COL1A1, COL1A2, COL3A1 and COL5A2 gene encoded proteins normally combined into different types of collagen fibers[16,17]. These collagen fibers are necessary for the formation of the basic structure of connective tissue[18,19]. When these genes mutated, it will affect the normal connective tissue in the body structure, such as the skin, bones, and present different clinical manifestations, such as Ehlers Danlos syndrome[20]. Its main characteristics is joint loose, skin smooth and soft and easy to extend the spin, wounded, and easily scarred, also associated with muscle bones abnormal phenomenon[21]. For example, B-myb (myeloblastosis), a member of the myb gene family, mediates intracellular signals controlling collagen gene expression in vascular smooth muscle cells (SMCs). Via interaction with a positively-acting matrix regulatory factor, B-Myb indirectly repress the expression of the Col5A2 gene. Also known as

Table II. GO annotation of COL1A2, COL5A2 module.

GO-ID	*p*-value	corr *p*-value	GO-ID description
7229	3.44E-15	1.18E-12	Integrin-mediated signaling pathway
48513	4.40E-15	1.18E-12	Organ development
48731	3.26E-12	3.69E-10	System development
22610	3.42E-12	3.69E-10	Biological adhesion

Table III. GO annotation of GJA1 module.

GO-ID	*p*-value	corr *p*-value	Description
7229	3.44E-15	1.18E-12	Integrin-mediated signaling pathway
48513	4.40E-15	1.18E-12	Organ development
48731	3.26E-12	3.69E-10	System development
22610	3.42E-12	3.69E-10	Biological adhesion

proto-oncogene, the mutation of B-Myb will release the expression of Col5A2 and hence make the cell matrix construction disordered and induce the abnormity of cells[22].

The GJA1 gene encodes Gap junction alpha-1 protein (GJA1), also known as connexin 43 (Cx43). This gene is a member of the connexin gene family. The encoded protein is a component of gap junctions, which are composed of arrays of intercellular channels that provide a route for the diffusion of low molecular weight materials from cell to cell[23]. The encoded protein is the major protein of gap junctions in the heart that are thought to have a crucial role in the synchronized contraction of the heart and in embryonic development[24]. Mutations in this gene have been associated with oculodentodigital dysplasia, heart malformations, and Hallermann–Streiff syndrome[25]. And the GJA1 gene also be found associated with bone growth in zebrafish fins[26], mouse oculodentodigital dysplasia[27], mice conditional osteoblast ablation[28] and other bone development diseases[29].

Conclusions

This analysis here is aim at finding biological molecules marker can be used for osteosarcoma diagnosis through contrast of osteosarcoma sample and normal tissue samples. The selected genes, i.e. COL (Collagen) and GJA1 are both significantly up-regulated genes and hence can be used as an important reference gene when performing osteosarcoma diagnosis. And the functions of the genes are performing through interferon signal transduction pathway. As for whether can only use it as a marker gene of osteosarcoma, and what is its function mechanism, more sample data have to be studied and it needs more validation.

Conflict of Interest

The Authors declare that they have no conflict of interests.

References

1) Jaffe N, Bruland, Bielack SS. Pediatric and adolescent osteosarcoma. New York, Springer, 2009.
2) Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell 2011; 144: 646-674.
3) Paoloni M, Davis S, Lana S, Withrow S, Sangiorgi L, Picci P, Hewitt S, Triche T, Meltzer P, Khanna C. Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. BMC Genomics 2009; 10: 625.
4) Fujita A, Sato JR, Rodrigues Lde O, Ferreira CE, Sogayar MC. Evaluating different methods of microarray data normalization. BMC Bioinformatics 2006; 7: 469.
5) Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics 2001; 17: 520-525.
6) Chen BE, Sakoda LC, Hsing AW, Rosenberg PS. Resampling-based multiple hypothesis testing procedures for genetic case-control association studies. Genet Epidemiol 2006; 30: 495-507.
7) Gage TB, Fang F, O'Neill E, Stratton H. Maternal age and infant mortality: A test of the wilcox-russell hypothesis. Am J Epidemiol 2009; 169: 294-303.
8) Condon TJ, Allen GJ. Role of psychoanalytic merging fantasies in systematic desensitization: A rigorous methodological examination. J Abnorm Psychol 1980; 89: 437-443.
9) Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-568.
10) Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.
11) Maere S, Heymans K, Kuiper M. Bingo: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005; 21: 3448-3449.
12) Reiner-Benaim A. FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis. Biom J 2007; 49: 107-126.

13) Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol 2011; 696: 291-303.

14) Ayerza MA, Muscolo DL, Aponte-Tinao LA, Farfalli G. Effect of erroneous surgical procedures on recurrence and survival rates for patients with osteosarcoma. Clin Orthop Relat Res 2006; 452: 231-235.

15) Bramer JA, Abudu AA, Tillman RM, Carter SR, Sumathi VP, Grimer RJ. Pre- and post-chemotherapy alkaline phosphatase levels as prognostic indicators in adults with localised osteosarcoma. Eur J Cancer 2005; 41: 2846-2852.

16) Meienberg J, Rohrbach M, Neuenschwander S, Spanaus K, Giunta C, Alonso S, Arnold E, Henggeler C, Regenass S, Patrignani A, Azzarello-Burri S, Steiner B, Nygren AO, Carrel T, Steinmann B, Matyas G. Hemizygous deletion of COL3A1, COL5A2, and MSTN causes a complex phenotype with aortic dissection: A lesson for and from true haploinsufficiency. Eur J Hum Genet 2010; 18: 1315-1321.

17) Valkkila M, Melkoniemi M, Kvist L, Kuivaniemi H, Tromp G, Ala-Kokko L. Genomic organization of the human COL3A1 and COL5A2 genes: COL5A2 has evolved differently than the other minor fibrillar collagen genes. Matrix Biol 2001; 20: 357-366.

18) Christner PJ, Ayitey S. Extracellular matrix containing mutated fibrillin-1 (Fbn1) down regulates COL1A1, COL1A2, COL3A1, COL5A1, and COL5A2 mRNA levels in Tsk/+ and Tsk/Tsk embryonic fibroblasts. Amino Acids 2006; 30: 445-451.

19) Andrikopoulos K, Liu X, Keene DR, Jaenisch R, Ramirez F. Targeted mutation in the COL5A2 gene reveals a regulatory role for type v collagen during matrix assembly. Nat Genet 1995; 9: 31-36.

20) Richards AJ, Martin S, Nicholls AC, Harrison JB, Pope FM, Burrows NP. A single base mutation in COL5A2 causes ehlers-danlos syndrome type II. J Med Genet 1998; 35: 846-848.

21) Grond-Ginsbach C, Wigger F, Morcher M, von Pein F, Grau A, Hausser I, Brandt T. Sequence analysis of the COL5A2 gene in patients with spontaneous cervical artery dissections. Neurology 2002; 58: 1103-1105.

22) Kypreos KE, Marhamati DJ, Sonenshein GE. B-myb represses trans-activation of the COL5A2 collagen promoter indirectly via inhibition of binding of factors interacting with positive elements within the first exon. Matrix Biol 1999; 18: 275-285.

23) Hoptak-Solga AD, Nielsen S, Jain I, Thummel R, Hyde DR, Iovine MK. Connexin43 (GJA1) is required in the population of dividing cells during fin regeneration. Dev Biol 2008; 317: 541-548.

24) Yuan D, Wang Q, Wu D, Yu M, Zhang S, Li L, Tao L, Harris AL. Monocyte-endothelial adhesion is modulated by Cx43-stimulated ATP release from monocytes. Biochem Biophys Res Commun 2012; 420: 536-541.

25) Pizzuti A, Flex E, Mingarelli R, Salpietro C, Zelante L, Dallapiccola B. A homozygous GJA1 gene mutation causes a Hallermann-Streiff/ODDD spectrum phenotype. Hum Mutat 2004; 23: 286.

26) Iovine MK, Higgins EP, Hindes A, Coblitz B, Johnson SL. Mutations in connexin43 (GJA1) perturb bone growth in zebrafish fins. Dev Biol 2005; 278: 208-219.

27) Flenniken AM, Osborne LR, Anderson N, Ciliberti N, Fleming C, Gittens JE, Gong XQ, Kelsey LB, Lounsbury C, Moreno L, Nieman BJ, Peterson K, Qu D, Roscoe W, Shao Q, Tong D, Veitch GI, Voronina I, Vukobradovic I, Wood GA, Zhu Y, Zirngibl RA, Aubin JE, Bai D, Bruneau BG, Grynpas M, Henderson JE, Henkelman RM, McKerlie C, Sled JG, Stanford WL, Laird DW, Kidder GM, Adamson SL, Rossant J. A GJA1 missense mutation in a mouse model of oculodentodigital dysplasia. Development 2005; 132: 4375-4386.

28) Grimston SK, Brodt MD, Silva MJ, Civitelli R. Attenuated response to *in vivo* mechanical loading in mice with conditional osteoblast ablation of the connexin43 gene (GJA1). J Bone Miner Res 2008; 23: 879-886.

29) Cella W, de Vasconcellos JP, de Melo MB, Kneipp B, Costa FF, Longui CA, Costa VP. Structural assessment of PITX2, FOXC1, CYP1B1, and GJA1 genes in patients with Axenfeld-Rieger syndrome with developmental glaucoma. Invest Ophthalmol Vis Sci 2006; 47: 1803-1809.