# Screening of diagnostic markers for osteosarcoma

DAJIANG WU*, KAI CHEN*, YUSHU BAI, XIAODONG ZHU, ZIQIANG CHEN, CHUANFENG WANG, YINGCHUAN ZHAO and MING LI

Department of Orthopedics, Changhai Hospital, Shanghai 200433, P.R. China

Received September 13, 2013; Accepted April 14, 2014

DOI: 10.3892/mmr.2014.2546

**Abstract.** Osteosarcoma, which is the most common type of highly malignant bone tumor in children and adolescents, has poor diagnosis and 2-year survival rates of 15-20% following surgery or radiotherapy, and has therefore generated marked attention. In order to investigate the potential biomarkers for diagnosing osteosarcoma, the expression profiling data from normal and disease tissues were compared, respectively, and the differentially-expressed genes were analyzed by three different statistical tests. Interacting proteins were determined and an interaction network was constructed by Search Tool for the Retrieval of Interacting Genes database. Subsequently, the protein interaction network was decomposed and Gene Otology annotation using Cytoscape, Mcode and Bingo, was conducted on the function modules. Finally, three differentially-expressed genes GJA1, COL1A2 and COL5A2 were identified, and an interaction network was successfully generated with COL1A2 and COL5A2 at the core. From the results, it was observed that COL1A2 and COL5A2 interact with a number of genes of the matrix metalloprotease (MMP) family, including MMP1, MMP2, MMP3 and MMP14, TGFβ and RUNX2. Furthermore, these genes have been confirmed to be important in the tumorigenesis of osteosarcoma. It was hypothesized that the upregulation of the COL gene family may be considered as a diagnostic marker for osteosarcoma and collagen may be administered as a therapy.

## Introduction

Osteosarcoma is a common clinical malignant bone tumor, representing ~35% of cases and constituting ~0.07% of all the neoplasms (1), and it mainly targets the adolescent age group (2). There has been great progress in the treatment of osteosarcoma, with almost 80% of patients being treated with limb-salvage and the 5-year survival rate for patients with osteosarcoma has increased from 20% to about 80%; however, more than half of all patients succumbed to metastasis and recurrence of osteosarcoma (3). With increased knowledge of tumor molecular biology, the concept of gene therapy for tumors is proposed and the experimental results show potential for clinical application. Several cytogenetic and molecular studies have recently been undertaken in osteosarcoma, and a number of genes (tumor suppressors, oncogenes and genes coding for growth factors) have been identified to be abnormally expressed. These genes may be used as diagnostic cytogenetic or molecular markers for osteosarcoma. For example, the retinoblastoma (RB) gene, the association of which has been well recognized between osteosarcoma and RB. It is reported that patients with hereditary RB have up to 1,000 times the incidence of osteosarcoma compared with the general population (4,5), and sporadic osteosarcoma revealed alterations of the RB gene in ~70% of cases (6). p53, a tumor suppressor gene, is also thought to be significant in the development of osteosarcoma. The p53 gene product can induce the transcription of numerous genes that are involved in the cell cycle control and apoptosis (7), and numerous reports have identified abnormalities of the p53 gene in osteosarcoma in up to 50% of cases. Additionally, the MDM2 gene has also been identified to be overexpressed in osteosarcoma, which may provide an alternative mechanism for the disruption of the normal p53 pathway (6,8). In addition, MDM2 is reported to be amplified in 36% of osteosarcomas and 100% of parosteal osteosarcomas (9).

Several alterations of genes and gene pathways associated with osteosarcoma have been identified in the past few years, which enhance the knowledge of the pathological mechanisms of osteosarcomas. However, there remains no specific diagnostic biomarker and further strides are urgently required to identify abnormalities that have prognostic and therapeutic implications, which may be useful in guiding the diagnosis and therapy for osteosarcomas.

In the present study, the expression profiling data between samples from human osteosarcoma and normal individuals were compared in order to identify significant tissue-specific genes. In addition, their function was investigated with the aim of identifying noticeable diagnostic markers for osteosarcoma.

## Materials and methods

*Osteosarcoma related gene expression profiles*. The gene expression data GSE16088 were obtained from the Gene

*Correspondence to:* Dr Ming Li, Department of Orthopedics, Changhai Hospital, No. 168 Changhai Road, Yangpu, Shanghai 200433, P.R. China
E-mail: liminglimingdoctor@hotmail.com

*Contributed equally

*Key words:* osteosarcoma, diagnosis, COL1A2, COL5A2, collagen

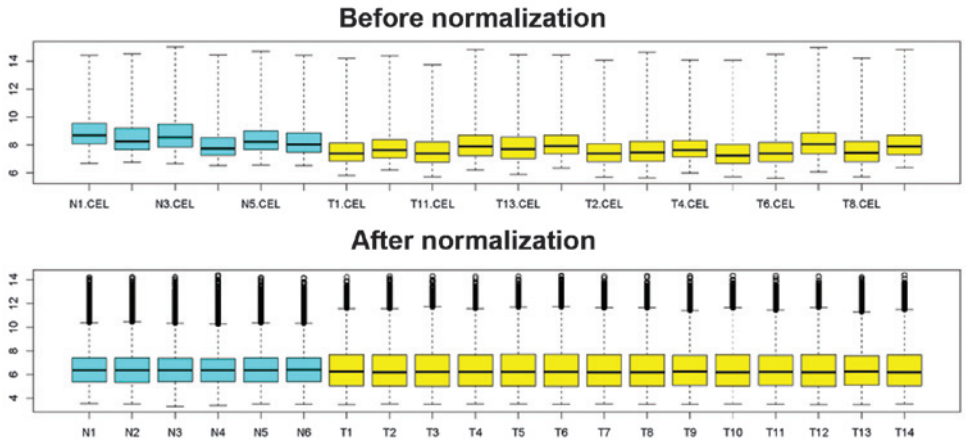


Figure 1. Chip data prior to and following normalization. The blue box represents the normal samples, while yellow represents the disease samples. The black line in the box is approaching on a horizontal line following standardization, which indicates that the standardization degree of data is higher compared with the original data.

Expression Omnibus database (10). A total of 20 samples were examined in the present study, among which 14 patients were diagnosed with osteosarcoma while the 6 normal samples served as negative controls.

*Preprocessing of expression data*. At first, the raw data downloaded were transformed into the expression data format by Affy package in R language (11), and then the missing data was imputed (12). At last, the expression profiling data were standardized using the median standardization method.

*Analysis of differential expression*. Three statistical tests in the multtest package in R language (13), including the t-test and Wilcox and Fisher's exact tests were used to conduct the differential-expression analysis between the osteosarcoma and normal groups. The P-value obtained from each test was corrected by the multiple testing Benjamini and Hochberg method (14), and the log FC (fold change) of every gene expression value was also inspected. The genes that could be screened by a variety of test methods simultaneously were regarded as differentially-expressed genes with high degree of confidence. These differentially-expressed genes were used to identify the abnormal genes between the osteosarcoma and normal control samples according to the threshold of $P<0.05$ and $|logFC|>1$.

*Construction of interaction network*. After the differentially-expressed genes were selected with high confidence through rigorous testing methods, Search Tool for the Retrieval of Interacting Genes (STRING) (15) was utilized to search for the interaction objects of differentially-expressed gene products and a protein interaction network was constructed.

*Analysis of network module*. The module of the whole network obtained previously was decomposed and the function modules with modular nature, including the genes identified through analysis, were found. Gene Otology annotations were performed using Cytoscape (16), Mcode (17) and Bingo (18) software based on the hypergeometric distribution and function enrichment threshold of adjp<0.05.

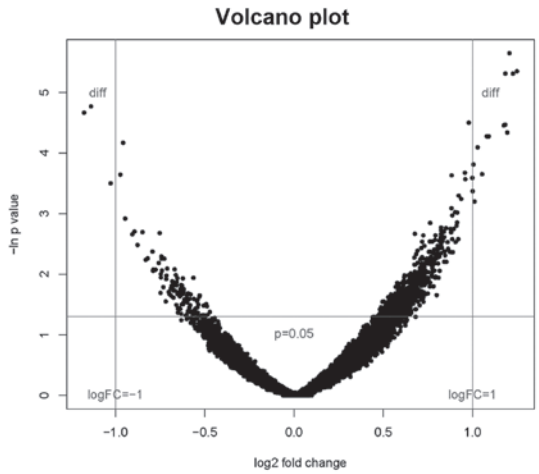


Figure 2. Association between the P-value and log FC. Since the smaller the P-value, the greater the logFC will be, the graph presents a volcano-like shape with 0 as the center, and it is termed a volcano plot. FC, fold change.

## Results

*Results of data preprocessing*. There were various factors, including background and probe design that could cause a difference in the original chip data. Therefore, standardization of the data was required prior to the analysis. The difference between data prior to and following standardization was significant (Fig. 1). It was observed that fluctuation of the data following standardization was significantly less compared with that without standardization.

*Analysis of differential expression*. Three statistical testing methods were used to analyze the gene expression data and the data were further corrected by multiple methods. Genes with $P<0.05$ and with $|log FC|>1$ obtained in three different manners were selected (Table I). Three differentially-expressed genes met the strict requirements of high degree confidence, and these genes were analyzed further.

A volcanic plot provides a simultaneous representation of P-values and log2 fold change for the gene expression data.

Table I. List of differentially-expressed genes.

| ID-REF | Gene symbol | T test-adj | Wilcox-adj | Fisher-adj | logFC |
|---|---|---|---|---|---|
| 201667-at | GJA1 | 1.11E-11 | 0.00054674 | 0.02724097 | 1.18443996 |
| 202404-s-at | COL1A2 | 6.05E-13 | 0.00137961 | 0.02724097 | 1.20747741 |
| 221729-at | COL5A2 | 1.23E-12 | 0.00137961 | 0.02724097 | 1.24898200 |

FC, fold change.

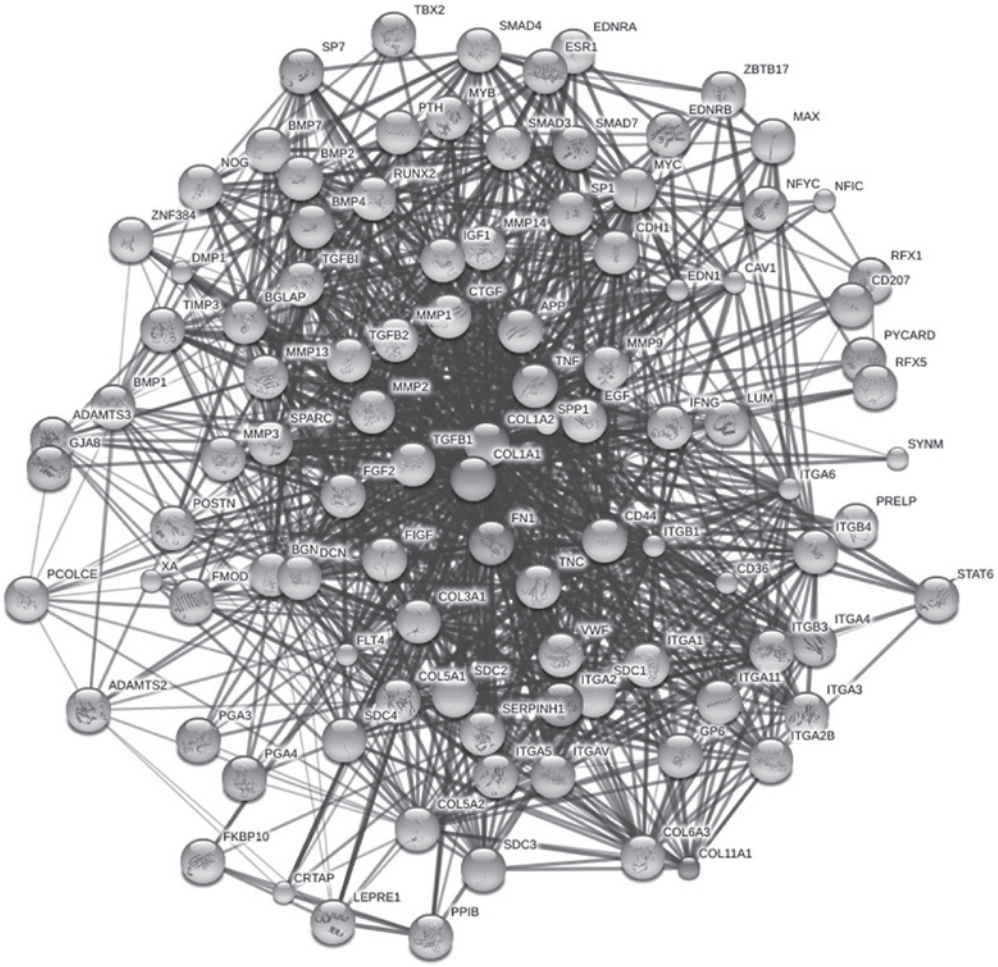


Figure 3. Interaction network diagram. Protein information from multiple databases was integrated by Search Tool for the Retrieval of Interacting Genes and the interaction objects for the target genes were predicted through the characteristics and spatial structure of the protein sequence itself.

The smaller the P-value, the higher the corresponding fold change (Fig. 2)

*Results of interaction network construction*. The three genes screened as the core were selected and combined with their possible interaction protein predicted from the STRING database. Next, the interaction network was built (Fig. 3).

*Analysis results of network module*. Cytoscape software was used to perform module analysis of the network constructed (Fig. 4). Next, Mcode was applied to identify the common module in the two genes, and the annotation of module function was undertaken by Bingo (Table II).

## Discussion

In the present study, through comparing the expression data from normal and osteosarcoma tissues, three differentially-expressed genes were identified. Among which, the COL1A2 and COL5A2 genes were selected as the core in order to construct a network and to analyze the interaction between genes correlated with osteosarcoma in order to understand the role of the collagen gene (COL) family and proteins in osteosarcoma.

Collagens are the main fibrous proteins of connective tissue and are the most abundant proteins of the extracellular matrix (ECM). Thus, mutations in different COL genes may

Table II. List of GO functional annotation.

| GO-ID | Function | Corr P-value |
|---|---|---|
| 7229 | Integrin-mediated signaling | 1.18E-12 |
| 48513 | Organ development | 1.18E-12 |
| 48731 | System development | 3.69E-10 |
| 22610 | Biological adhesion | 3.69E-10 |
| 7155 | Cell adhesion | 3.69E-10 |
| 48856 | Anatomical structure development | 3.24E-09 |
| 9888 | Tissue development | 5.69E-08 |
| 32501 | Multicellular organismal process | 7.84E-08 |
| 7275 | Multicellular organismal development | 7.84E-08 |
| 7166 | Cell surface receptor linked signal transduction | 1.35E-07 |
| 7167 | Enzyme-linked receptor protein signaling pathway | 1.68E-07 |
| 7178 | Transmembrane receptor protein serine/threonine kinase signaling pathway | 1.94E-07 |

GO, Gene Otology.

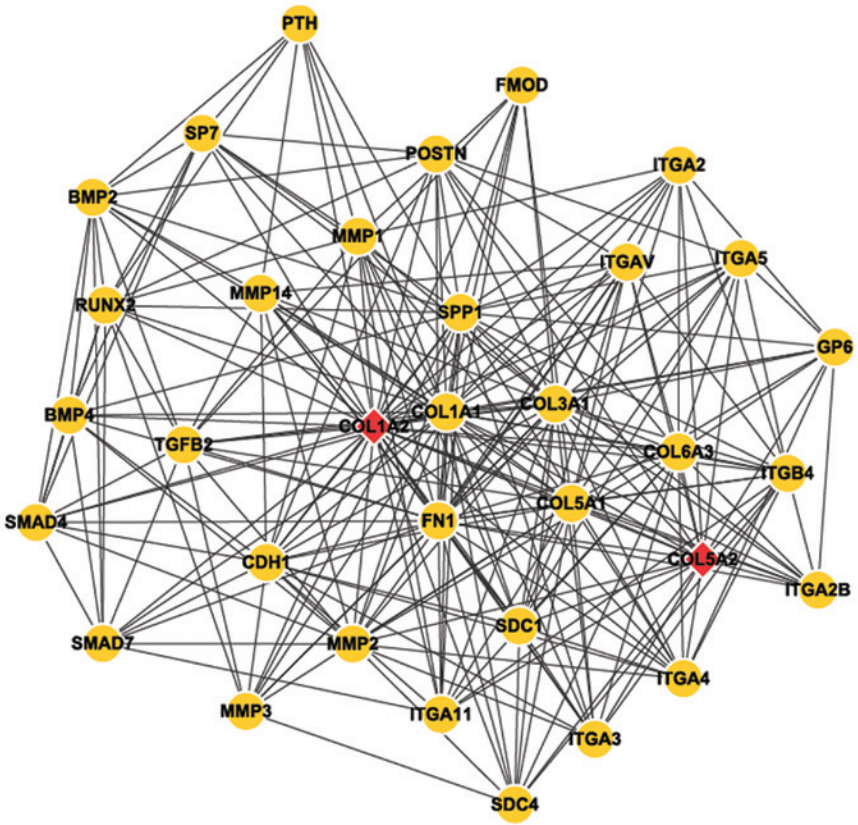


Figure 4. Sub-module diagram excavated by Cytoscape and Mcode. The yellow circle represents the interacting genes, while the red diamond represents the differentially-expressed genes screened.

disrupt the same collagen fibril. In humans, mutations in 13 different COL genes have been associated with disease phenotypes (19,20). Mutations in the same COL gene can also give rise to different human diseases. For example, mutations in the COL1A2 gene result in at least five different forms of chondrodysplasia and cartilage degeneration (21). There are 19 to 20 types of collagen that have been identified to date, and type I COL, the major component of ECM in skin, bone and ligaments is composed of glycine- and proline rich two-α1 (I) and one-α2 (I) chains (22). In the present study, COL was found to be significantly upregulated and the established network module with COL1A2 and COL5A2 as core revealed that numerous other genes participated in this network and interacted with COL1A2 and COL5A2, such as TGFβ, MMP, SMAD and RUNX2. These genes have all been demonstrated to to be important in the pathogenesis of osteosarcoma. A

number of studies have been performed on the correlation between these genes and COL1A2 and COL5A2 and their interaction on the pathogenesis of osteosarcoma. For example, TGFβ2 is a member of the transforming growth factor (TGF)-subfamily, TGFβ induces the synthesis of numerous ECM proteins, such as COL, fibronectin, laminin and tenascin, and inhibits the matrix degrading enzymes. Therefore, TGFβ2 is involved in wound healing, fibrosis, embryogenesis and tumorigenesis (23,24). A previous study provides evidence that TGFβ stimulates human COL1A2 promoter activity through Smad signaling molecules (25). The Smad family contains TGF-β receptor-dependent R-Smads (such as Smad2/3), Co-Smads (such as Smad4) and anti-Smads (such as Smad6/7). It is reported that transient expression of Smad3 or Smad4 in human skin fibroblasts leads to stimulation of COL1A2 promoter activity. In addition, overexpression of anti-Smad and Smad7 represses basal as well as TGFβ-stimulated COL1A2 promoter activity, indicating its antagonistic effect on TGFβ signaling in human dermal fibroblasts (25).

MMP, which is a hallmark of invasive cancers, including osteosarcoma (26), is capable of cleaving type I COL triple helices. Additionally, membrane-bound MMP-14 may be particularly important, as MMP-14 knockout mice exhibit severe COL turnover deficiencies (27).

Collagen fibers, as well as fibers produced by cancer-associated fibroblasts, may form an invasion barrier (28). However, MMPs are essential for maintaining the invasive phenotype and supporting migration and proliferation of cancer cells (29,30). Moreover, the critical function of collagenolytic MMPs in cancer is clearing an invasion path through the barrier of type I collagen fibers in the stroma and blood vessel walls (31,32). Therefore, massive production of MMPs that degrade collagen may solve this problem of invasion barriers. As a consequence, the COL1A2 gene may be used as a diagnostic marker with a higher expression level in cancer cells (33) and restoring type I collagen may be used as a therapeutic target for cancer.

Additionally, the Runt-related transcription factor Runx2, which also interacts with COL1A2, has a well-defined role in mediating the final stages of osteoblast maturation and is required for normal osteogenesis. Runx2 deficiency or mutations affecting the function of Runx2 protein result in severe bone abnormalities in mice and humans. Runx2 is implicated in early osteoblast differentiation, and is involved in the later stages of chondrocyte differentiation, maturation and possibly, endochondral ossification (34). Runx2 expression is essential for the commitment of preosteoblasts to the osteoblast lineage. The cooperative interaction between Runx2 and the RB tumor suppressor protein leads to progressive growth arrest and increases expression of the mature osteoblast phenotype (35), which is reconciled with the molecular pathogenesis of osteosarcoma (36). As the genes mentioned above are all osteosarcoma related and co-interact with COL1A2 and COL5A2, a small mutation in the COL1A2 and COL5A2 genes may affect these other genes, resulting in pathological changes in osteoblasts which lead to osteosarcoma development.

Osteosarcoma, has been a key focus of research as it is a disease that effects adolescents. Jeon *et al* (37) investigated 25 cases of patients with primary malignant curettage, the local recurrence rate was 18% and the overall survival rate was 65%. The average 5- and 10-year survival rates of these patients were reduced due to local recurrence (38). Bramer *et al* (39) observed 89 cases of adult osteosarcoma following chemotherapy and observed that their alkaline phosphatase (AP) levels correlated with response to chemotherapy and survival. The results revealed that the prognosis was poor when the AP level was elevated 2-fold compared with the normal level.

In conclusion, in the present study, it was observed that the selected COL gene was significantly upregulated, and this gene was involved in the interleukin conducted signaling pathway. Thus, the upregulation of the COL gene may be considered as a diagnostic marker for osteosarcoma. However, further studies are required to understand the role of collagen in osteosarcoma development in detail.

## References


1. Fletcher CD, Unni KK and Mertens F; WHO; IARC: WHO classification of tumours: Pathology and genetics of tumours of soft tissue and bone. 4th edition. IARC Press, Lyon, France, 2002.
2. Bielack SS, Kempf-Bielack B, Delling G, *et al*: Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. J Clin Oncol 20: 776-790, 2002.
3. Ferrari S, Smeland S, Mercuri M, *et al*: Neoadjuvant chemotherapy with high-dose Ifosfamide, high-dose methotrexate, cisplatin, and doxorubicin for patients with localized osteosarcoma of the extremity: a joint study by the Italian and Scandinavian Sarcoma Groups. J Clin Oncol 23: 8845-8852, 2005.
4. Abramson DH, Ellsworth RM, Kitchin FD and Tung G: Second nonocular tumors in retinoblastoma survivors. Are they radiation-induced? Ophthalmology 91: 1351-1355, 1984.
5. Kitchin FD and Ellsworth RM: Pleiotropic effects of the gene for retinoblastoma. J Med Genet 11: 244-246, 1974.
6. Ladanyi M and Gorlick R: Molecular pathology and molecular pharmacology of osteosarcoma. Pediatr Pathol Mol Med 19: 391-413, 2000.
7. Hung J and Anderson R: p53: functions, mutations and sarcomas. Acta Orthop Scand Suppl 273: 68-73, 1997.
8. Miller CW, Aslo A, Won A, Tan M, Lampkin B and Koefflar HP: Alterations of the p53, Rb and MDM2 genes in osteosarcoma. J Cancer Res Clin Oncol 122: 559-565, 1996.
9. Noble-Topham SE, Burrow SR, Kandel RA, *et al*: SAS is amplified predominantly in surface osteosarcoma. J Orthop Res 14: 700-705, 1996.
10. Paoloni M, Davis S, Lana S, *et al*: Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. BMC Genomics 10: 625, 2009.
11. Fujita A, Sato JR, Rodrigues Lde O, Ferreira CE and Sogayar MC: Evaluating different methods of microarray data normalization. BMC Bioinformatics 7: 469, 2006.
12. Troyanskaya O, Cantor M, Sherlock G, *et al*: Missing value estimation methods for DNA microarrays. Bioinformatics 17: 520-525, 2001.
13. Pollard KS, Dudoit S and van der Laan MJ: Multiple testing procedures: R multtest package and applications to genomics. Bioinformatics and Computaional Biology Solutions using R and Bioconductor. Statistics for Biology and Health 2005. pp 249-271, 2005.
14. Benjamini Y and Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc. B: 289-300, 1995.
15. Szklarczyk D, Franceschini A, Kuhn M, *et al*: The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39: D561-D568, 2011.
16. Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431-432, 2011.
17. Rivera CG, Vakil R and Bader JS: NeMo: network module identification in Cytoscape. BMC Bioinformatics 11: S61, 2010.
18. Maere S, Heymans K and Kuiper M: BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448-3449, 2005.

19. Olsen BR: Mutations in collagen genes resulting in metaphyseal and epiphyseal dysplasias. Bone 17: S45-S49, 1995.
20. Prockop DJ and Kivirikko KI: Collagens: molecular biology, diseases, and potentials for therapy. Ann Rev Biochem 64: 403-434, 1995.
21. Vikkula M, Metsäranta M and Ala-Kokko L: Type II collagen mutations in rare and common cartilage diseases. Ann Med 26: 107-114, 1994.
22. Ghosh AK: Factors involved in the regulation of type I collagen gene expression: implication in fibrosis. Exp Biol Med 227: 301-314, 2002.
23. Branton MH and Kopp JB: TGF-β and fibrosis. Microbes Infect 1: 1349-1365, 1999.
24. Blobe GC, Schiemann WP and Lodish HF: Role of transforming growth factor β in human disease. N Engl J Med 342: 1350-1358, 2000.
25. Chen SJ, Yuan W, Mori Y, Levenson A, Trojanowska M and Varga J: Stimulation of type I collagen transcription in human skin fibroblasts by TGF-β: involvement of Smad 3. J Invest Dermatol 112: 49-57, 1999.
26. Mignatti P and Rifkin DB: Biology and biochemistry of proteinases in tumor invasion. Physiol Rev 73: 161-195, 1993.
27. Holmbeck K, Bianco P, Caterina J, *et al*: MT1-MMP-deficient mice develop dwarfism, osteopenia, arthritis, and connective tissue disease due to inadequate collagen turnover. Cell 99: 81-92, 1999.
28. Makareeva E, Han S, Vera JC, *et al*: Carcinomas contain a matrix metalloproteinase-resistant isoform of type I collagen exerting selective support to invasion. Cancer Res 70: 4366-4374, 2010.
29. Yong HY and Moon A: Roles of calcium-binding proteins, S100A8 and S100A9, in invasive phenotype of human gastric cancer cells. Arch Pharm Res 30: 75-81, 2007.
30. Chen PN, Kuo WH, Chiang CL *et al*: Black rice anthocyanins inhibit cancer cells invasion via repressions of MMPs and u-PA expression. Chem Biol Interact 163: 218-229, 2006.
31. Itoh Y and Seiki M: MT1-MMP: a potent modifier of pericellular microenvironment. J Cell Physiol 206: 1-8, 2006.
32. Nabha SM, dos Santos EB, Yamamoto HA, *et al*: Bone marrow stromal cells enhance prostate cancer cell invasion through type I collagen in an MMP-12 dependent manner. Int J Cancer 122: 2482-2490, 2008..
33. Mori K, Enokida H, Kagara I, *et al*: CpG hypermethylation of collagen type I alpha 2 contributes to proliferation and migration activity of human bladder cancer. Int J Oncol 34: 1593-1602, 2009.
34. Pratap J, Galindo M, Zaidi SK, *et al*: Cell growth regulatory role of Runx2 during proliferative expansion of preosteoblasts. Cancer Res 63: 5357-5362, 2003.
35. Thomas DM, *et al*: Terminal osteoblast differentiation, mediated by runx2 and p27KIP1, is disrupted in osteosarcoma. J Cell Biology 167: 925-934, 2004.
36. Thomas DM, Carty SA, Piscopo DM, *et al*: The retinoblastoma protein acts as a transcriptional coactivator required for osteogenic differentiation. Mol Cell 8: 303-316, 2001.
37. Jeon DG, Lee SY and Kim JW: Bone primary sarcomas undergone unplanned intralesional procedures-the possibility of limb salvage and their oncologic results. J Surg Oncol 94: 592-598, 2006.
38. Ayerza MA, Muscolo DL, Aponte-Tinao LA and Farfalli G: Effect of erroneous surgical procedures on recurrence and survival rates for patients with osteosarcoma. Clin Orthop Relat Res 452: 231-235, 2006.
39. Bramer JA, Abudu AA, Tillman RM, Carter SR, Sumathi VP and Grimer RJ: Pre-and post-chemotherapy alkaline phosphatase levels as prognostic indicators in adults with localised osteosarcoma. Eur J Cancer 41: 2846-2852, 2005.