Marker Gene Screening for Human Mesenchymal Stem Cells in Early Osteogenic Response to Bone Morphogenetic Protein 6 with DNA Microarray

Shien Zou,¹ Shaofen Zhang,¹ Qiqi Long,¹ Yuankui Cao,¹ and Wei Zhang²

Aims: Microarray data were analyzed using bioinformatic tools to screen marker genes of human mesenchymal stem cells (hMSC) in response to bone morphogenetic protein 6 (BMP6). **Results:** A total of 190 differentially expressed genes were identified. The interaction network was divided into three functional modules. These genes were connected with BMP signaling pathways and regulation of cell processes, while *NOG* and *BMPR2* participated in the transforming growth factor-beta signal pathway. Besides, several related small molecules were acquired. **Conclusion:** Marker genes in osteogenic responses to BMP6 treatment for hMSC were screened with microarray data along with elaborate function analysis by bioinformatics. *NOG* and *BMPR2* showed potential to become indicators to monitor the directed differentiation of hMSC into osteoblasts, which can be used for bone disease treatment. Moreover, small molecules such as W-13 were retrieved and provided directions for future drug design.

Introduction

OSTEOPOROSIS is a common bone disease that leads to an increased risk of fracture. Since osteoblasts play an important role in bone formation, their use is considered an effective means of treatment (Chen *et al.*, 1997). Directed differentiation of human mesenchymal stem cells (hMSCs) is regarded as a prospective way to provide osteoblasts. The hMSCs possess a strong differentiation capacity. Under different inducing conditions, they can differentiate into various types of cells, including mesoderm-type cell lineages such as osteoplasts, chondrocytes and also non-mesoderm-type lineages such as neuronal-like cells (Kassem *et al.*, 2004). The differentiation of hMSCs into osteoblasts is regulated by many factors such as hormones and cytokines (Friedenstein *et al.*, 1987). However, the specific underlying regulation mechanism is not very clear. Markers that can be used to monitor and control the process of differentiation are especially necessary.

Several researchers have explored the expression pattern of known markers (Gonzalez *et al.*, 2007; Riekstina *et al.*, 2009). Gonzalez *et al.* (2007) had demonstrated that general pluripotent markers such as Oct-4, Nanog, Sox-2, and SSEA-4 existed in human second-trimester MSCs. In addition, some novel markers were also discovered, such as W5C5 (Masuda *et al.*, 2012) and Stathmin-like 2 (Chiellini *et al.*, 2008). Bae *et al.* (2011) combined transcriptome data and proteome data to screen new markers and suggested that TM4SF1 (transmem-

brane 4 L6 family member 1) can serve as a surface protein marker to identify hMSCs. Although some advances have been achieved, the knowledge was far from sufficient to support practical application.

High-throughput screening with microarray data is a potent way to discover new markers. Bone morphogenetic proteins (BMPs) can strongly stimulate mesenchymal bone progenitor cell differentiation into mature osteoblasts (Lories *et al.*, 2006) and are, thus, involved in bone formation (Yoon *et al.*, 2005). They are acidic glycoproteins that are widely present in bone matrix and which belong to the transforming growth factor β (TGF- β) superfamily. Therefore, BMP6 stimulation favors identification of markers in the early phase of hMSC differentiation into osteoblasts (Jeon *et al.*, 2006). Based on the earlier considerations, we compared microarray data for hMSC samples treated with BMP6 and that for untreated hMSC samples to obtain differentially expressed genes and then performed function analysis to further characterize them.

Materials and Methods

Microarray data

Dataset GSE13604 (Luo *et al.*, 2009) was downloaded from Gene Expression Omnibus database. The dataset contains a total of four gene chip measurements from duplicate experiments, each with paired measurements of hMSC with or without 8 h BMP6 treatment. The platform used to collect the

¹Department of Gynecology, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China.

²Department of Biostatistics, School of Public Health, Fudan University, Shanghai, China.

FIG. 1. Box plots for normalized microarray data. The two gray boxes represent normal human mesenchymal stem cells (hMSCs) samples, and the two black boxes represent bone morphogenetic protein 6 (BMP6)-stimulated hMSCs samples. The black line in the box is the median of each data. The black lines are almost on the same line, indicating good effectiveness of normalization.

microarray data was GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Probe annotations were provided by Affymetrix, including all Affymetrix ATH1 (25K) chip probe. Both raw data and the probe annotations were collected.

Screening of differentially expressed genes

Raw data were first transformed to a recognizable expression profiling format, and then, the missing data were filled in (Troyanskaya *et al.*, 2001). After normalization (Fujita *et al.*, 2006), package limma (Smyth, 2005) in R language was utilized for statistical analysis. p -value < 0.05 and $|\log FC| > 1$ were set as the cutoff criteria to screen the differentially expressed genes.

Establishment of interaction network

Uncovering the interactions between proteins is the basis for understanding their functions and regulation mechanisms. Therefore, software HitPredict was utilized to establish the interaction network for protein products of all the differentially expressed genes.

HitPredict collects protein-protein interactions from IntAct, BIOGRID, and HPRD database by high-throughput or small-scale experiments. It also includes predicted interactions according to the score, integrating information-like sequence, structure, and function using a Bayesian algorithm (Patil *et al.*, 2011). Interactions obtained from experiments or with a score > 1 are considered of high confidence (Patil and Nakamura, 2005), based on which we identified the interaction relationships and applied Cytoscape (Duncan *et al.*, 2010) to visualize the network.

Function annotation

The plugins Mcode of Cytoscape (Goldberg *et al.*, 2007) (cutoff: the degree of each node in the module > 2) and Bingo (Bernstein *et al.*, 2012) were used for module division and annotation based on the hypergeometric distribution (corrected p -value < 0.05). Pathway enrichment analysis was also performed with FuncAssociate (Berriz *et al.*, 2009) for the differentially expressed genes, and a p -value of < 0.05 was set as the cutoff criterion.

Screening of related small molecules

The Connectivity Map (CMap) is a collection of transcriptional expression data and pattern-matching algorithms that together enable the discovery of function associations between drugs, genes, and diseases (Lamb *et al.*, 2006; Lamb, 2007; Wen *et al.*, 2011). Genes were divided into two groups: up-regulated and down-regulated, and then, they were analyzed separately by CMap. Those with a correlation score ($|\text{score}| > 0.8$) were retained.

Results

Differentially expressed genes

Microarray data were preprocessed and normalized (Fig. 1). A total of 190 differentially expressed genes were identified according to the criteria ($p < 0.01$ and $|\log FC| > 1$), including

FIG. 2. Three different modules of interaction network for the differentially expressed genes.

TABLE 1. FUNCTIONS FOR GENES IN MODULE 1

GO-ID	p-value	corr p-value	Description
30509	1.49E-17	1.07E-14	BMP signaling pathway
7178	9.12E-15	3.26E-12	Transmembrane receptor protein serine/threonine kinase signaling pathway
7167	3.57E-11	8.52E-09	Enzyme-linked receptor protein signaling pathway
9653	2.36E-07	1.88E-05	Anatomical structure morphogenesis
7166	3.34E-07	2.39E-05	Cell-surface-receptor-linked signaling pathway
48513	3.41E-06	0.000153	Organ development
23033	1.02E-05	0.000306	Signaling pathway
48731	2.70E-05	0.000603	System development
48856	5.06E-05	0.000823	Anatomical structure development
7275	1.09E-04	0.001215	Multicellular organismal development
23052	1.55E-04	0.001457	Signaling
32502	1.93E-04	0.001587	Developmental process
50896	4.22E-04	0.002769	Response to stimulus
32501	1.46E-03	0.00594	Multicellular organismal process
50794	1.46E-02	0.026024	Regulation of cellular process
50789	2.03E-02	0.033556	Regulation of biological process
65007	2.92E-02	0.043681	Biological regulation

BMP, bone morphogenetic protein.

87 down-regulated genes and 103 up-regulated genes (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/gtmb).

Function analysis results

A total of 395 pairs of interactions for all the differentially expressed genes were identified from HitPredict (Supplementary Table S2). Three modules were generated containing 8, 13, and 5 genes (Fig. 2), corresponding to 17, 10, and 5 functions (Tables 1, 2 and 3), respectively. As shown in Table 1 to Table 3, the most enriched functions were BMP signaling pathway and regulation of cellular process, respectively.

FuncAssociate is a web-based tool that will return a list of over- or under-represented pathways. Annotations with $p < 0.05$ were regarded as significant in this study. All the differentially expressed genes were analyzed by the software, and the pathway involving *NOG* and *BMPR2* was identified: TGF-beta signaling pathway (Fig. 3).

Related small molecules

A total of 10 small-molecule drugs were recorded as |score| > 0.8 (Table 4), in which W-13 had the highest correlation (0.88723).

TABLE 2. FUNCTIONS FOR GENES IN MODULE 2

GO-ID	p-value	corr p-value	Description
50794	7.59E-04	2.12E-02	Regulation of cellular process
48731	1.42E-03	2.96E-02	System development
32502	1.38E-03	2.96E-02	Developmental process
50789	1.29E-03	2.96E-02	Regulation of biological process
48856	2.49E-03	3.25E-02	Anatomical structure development
65007	2.34E-03	3.25E-02	Biological regulation
7275	4.87E-03	3.71E-02	Multicellular organismal development
9987	6.22E-03	4.06E-02	Cellular process
23052	6.62E-03	4.15E-02	Signaling
32501	1.07E-02	4.61E-02	Multicellular organismal process

Discussion

We compared microarray data for hMSC samples treated with BMP6 versus untreated hMSC samples (as control) and screened out a total of 190 differentially expressed genes. These genes were divided into three modules, while the most enriched functions were BMP signaling pathway and regulation of cellular process. Pathway annotation was also performed, and the TGF-beta signaling pathway was identified, which contained *Noggin* (*NOG*) and *BMP receptor type II* (*BMPR2*).

It is not strange that the BMP signaling pathway is significantly enriched, as samples were treated with BMP6. However, it is worthy of note. The BMP signaling pathway is involved in many processes, including cell growth, cell differentiation, apoptosis, and other functions. Previous researchers have demonstrated the involvement of the BMP signaling pathway and TGF-beta signaling pathway in osteoblast differentiation and bone formation (Marie, 1997; Origuchi *et al.*, 1998; Chen *et al.*, 2012). *NOG* is a BMP antagonist that is induced by BMP (Gazzerro *et al.*, 1998). Okamoto *et al.* (2006) found that over-expression of *NOG* would cause increased bone volume associated with decreased bone formation and decreased osteoclast number. Further, Gerhart *et al.* (2011) demonstrated that *NOG* is a essential regulator for BMP signaling in the early epiblast. However, a recent study pointed out that it was a novel inducer of MSC adipogenesis (Sawant *et al.*, 2012). *BMPR2* is a serine/threonine receptor kinase that can bind to BMPs. Several papers have explored

TABLE 3. FUNCTIONS FOR GENES IN MODULE 3

GO-ID	p-value	corr p-value	Description
48522	5.37E-05	1.09E-03	Positive regulation of cellular process
48518	8.72E-05	1.41E-03	Positive regulation of biological process
50794	1.56E-02	2.99E-02	Regulation of cellular process
50789	2.02E-02	3.51E-02	Regulation of biological process
65007	2.69E-02	4.36E-02	Biological regulation

FIG. 3. KEGG pathway map for genes in the three modules. The dark box indicates the differentially expressed genes.

the association of *BMPR2* mutation with primary pulmonary hypertension (Arbustini *et al.*, 2008; Fu *et al.*, 2008; Rubin, 2008). Recently, one study found that it was necessary for signal transduction and osteogenic differentiation of human adipose-derived MSCs (Zeng *et al.*, 2012). Therefore, we could say that the differentially expressed genes identified in this study are really involved in the regulation of hMSC osteogenic differentiation, especially *NOG* and *BMPR2*.

Considering that there would be practical value to develop molecules directing the differentiation process, we retrieved several small molecules from the CMap database, among which W-13 showed the highest correlation with the gene-expression changes in hMSC treated with BMP6. W-13 is a

calmodulin antagonist that inhibits the growth of tumor cells (Wei *et al.*, 1983; Strobl and Peterson, 1992). Although the connections between these molecules and hMSC's directed differentiation was not discovered, it provides certain guidelines for future developments.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (No. 81200415) and a grant from the PhD Programs Foundation of the Ministry of Education of China (No. 20090071120020).

Author Disclosure Statement

The authors state that there are no conflicts of interest, including specific financial interests, relationships, and affiliations, which are relevant to the subject of this article.

References

Arbustini E, Marziliano N, Porcu E, *et al.* (2008) Gene symbol: *BMPR2*. Disease: pulmonary hypertension, primary. Hum Genet 123:112–113.
Bae S, Shim SH, Park CW, *et al.* (2011) Combined omics analysis identifies transmembrane 4 L6 family member 1 as a surface protein marker specific to human mesenchymal stem cells. Stem Cells Dev 20:197–203.
Bernstein BE, Birney E, Dunham I, *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.

TABLE 4. RELATED SMALL-MOLECULE DRUGS	
CMap name	Correlation
W-13	0.88723
Praziquantel	0.84861
Telenzepine	0.84774
Paroxetine	0.83978
Flucloxacillin	0.83523
Urapidil	0.83481
Betahistine	0.82679
Metitepine	0.82126
Dorzolamide	0.81307
Dextromethorphan	0.80077

CMap, Connectivity Map.

- Berriz GF, Beaver JE, Cenik C, *et al.* (2009) Next generation software for functional trend analysis. *Bioinformatics* 25:3043–3044.
- Chen G, Deng C, Li YP (2012) TGF-beta and BMP signaling in osteoblast differentiation and bone formation. *Int J Biol Sci* 8:272–288.
- Chen JL, Hunt P, McElvain M, *et al.* (1997) Osteoblast precursor cells are found in CD34+ cells from human bone marrow. *Stem Cells* 15:368–377.
- Chiellini C, Grenningloh G, Cochet O, *et al.* (2008) Stathmin-like 2, a developmentally-associated neuronal marker, is expressed and modulated during osteogenesis of human mesenchymal stem cells. *Biochem Biophys Res Commun* 374:64–68.
- Duncan D, Prodduturi N, Zhang B (2010) WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics* 11:P10.
- Friedenstein AJ, Chailakhyan RK, Gerasimov UV (1987) Bone marrow osteogenic stem cells: *in vitro* cultivation and transplantation in diffusion chambers. *Cell Tissue Kinet* 20:263–272.
- Fu LJ, Zhou AQ, Huang MR, *et al.* (2008) A novel mutation in the BMPR2 gene in familial pulmonary arterial hypertension. *Chin Med J (Engl)* 121:399–404.
- Fujita A, Sato JR, Rodrigues Lde O, *et al.* (2006) Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7:469.
- Gazzero E, Gangji V, Canalis E (1998) Bone morphogenetic proteins induce the expression of noggin, which limits their activity in cultured rat osteoblasts. *J Clin Invest* 102:2106–2114.
- Gerhart J, Scheinfeld VL, Milito T, *et al.* (2011) Myo/Nog cell regulation of bone morphogenetic protein signaling in the blastocyst is essential for normal morphogenesis and striated muscle lineage specification. *Dev Biol* 359:12–25.
- Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell* 128:635–638.
- Gonzalez R, Maki CB, Pacchiarotti J, *et al.* (2007) Pluripotent marker expression and differentiation of human second trimester Mesenchymal Stem Cells. *Biochem Biophys Res Commun* 362:491–497.
- Jeon EJ, Lee KY, Choi NS, *et al.* (2006) Bone morphogenetic protein-2 stimulates Runx2 acetylation. *J Biol Chem* 281:16502–16511.
- Kassem M, Kristiansen M, Abdallah BM (2004) Mesenchymal stem cells: cell biology and potential use in therapy. *Basic Clin Pharmacol Toxicol* 95:209–214.
- Lamb J (2007) The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer* 7:54–60.
- Lamb J, Crawford ED, Peck D, *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935.
- Lories RJ, Daans M, Derese I, *et al.* (2006) Noggin haploinsufficiency differentially affects tissue responses in destructive and remodeling arthritis. *Arthritis Rheum* 54:1736–1746.
- Luo W, Friedman MS, Shedden K, *et al.* (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* 10:161.
- Marie P (1997) Growth factors and bone formation in osteoporosis: roles for IGF-I and TGF-beta. *Rev Rhum Engl Ed* 64:44–53.
- Masuda H, Anwar SS, Buhning HJ, *et al.* (2012) A novel marker of human endometrial mesenchymal stem-like cells. *Cell Transplant* 21:2201–2214.
- Okamoto M, Murai J, Yoshikawa H, *et al.* (2006) Bone morphogenetic proteins in bone stimulate osteoclasts and osteoblasts during bone development. *J Bone Miner Res* 21:1022–1033.
- Origuchi N, Ishidou Y, Nagamine T, *et al.* (1998) The spatial and temporal immunolocalization of TGF-beta 1 and bone morphogenetic protein-2/-4 in phallic bone formation in inbred Sprague Dawley male rats. *In Vivo* 12:473–480.
- Patil A, Nakai K, Nakamura H (2011) HitPredict: a database of quality assessed protein-protein interactions in nine species. *Nucleic Acids Res* 39:D744–D749.
- Patil A, Nakamura H (2005) Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics* 6:100.
- Riekstina U, Cakstina I, Parfejevs V, *et al.* (2009) Embryonic stem cell marker expression pattern in human mesenchymal stem cells derived from bone marrow, adipose tissue, heart and dermis. *Stem Cell Rev* 5: 378–386.
- Rubin LJ (2008) BMPR2 mutation and outcome in pulmonary arterial hypertension: clinical relevance to physicians and patients. *Am J Respir Crit Care Med* 177:1300–1301.
- Sawant A, Chanda D, Isayeva T, *et al.* (2012) Noggin is novel inducer of mesenchymal stem cell adipogenesis: implications for bone health and obesity. *J Biol Chem* 287:12241–12249.
- Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, *et al.* (eds) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, pp 397–420.
- Strobl JS, Peterson VA (1992) Tamoxifen-resistant human breast cancer cell growth: inhibition by thioridazine, pimozide and the calmodulin antagonist, W-13. *J Pharmacol Exp Ther* 263: 186–193.
- Troyanskaya O, Cantor M, Sherlock G, *et al.* (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17:520–525.
- Wei JW, Hickie RA, Klaassen DJ (1983) Inhibition of human breast cancer colony formation by anticalmodulin agents: trifluoperazine, W-7, and W-13. *Cancer Chemother Pharmacol* 11:86–90.
- Wen Z, Wang Z, Wang S, *et al.* (2011) Discovery of molecular mechanisms of traditional Chinese medicinal formula Si-Wu-Tang using gene expression microarray and connectivity map. *PLoS One* 6:e18278.
- Yoon BS, Ovchinnikov DA, Yoshii I, *et al.* (2005) Bmpr1a and Bmpr1b have overlapping functions and are essential for chondrogenesis *in vivo*. *Proc Natl Acad Sci U S A* 102:5062–5067.
- Zeng Y, Qu X, Li H, *et al.* (2012) MicroRNA-100 regulates osteogenic differentiation of human adipose-derived mesenchymal stem cells by targeting BMPR2. *FEBS Lett* 586:2375–2381.

Address correspondence to:

Wei Zhang, PhD

Department of Biostatistics

School of Public Health

Fudan University

No. 130 Dongan Road

Shanghai 200032

China

E-mail: weizhang@shmu.edu.cn

Shaofen Zhang, MD

Department of Gynecology

Obstetrics and Gynecology Hospital of Fudan University

No. 128 Shenyang Road

Shanghai 200011

China

E-mail: zhangshaofen@163.com