

Prediction of osteoporosis candidate genes by computational disease-gene identification strategy

Qing-Yang Huang · Gloria H. Y. Li ·
William M. W. Cheung · You-Qiang Song ·
Annie W. C. Kung

Received: 15 February 2008 / Accepted: 8 April 2008 / Published online: 9 May 2008
© The Japan Society of Human Genetics and Springer 2008

Abstract Osteoporosis is a complex disease with a strong genetic component. To date, more than 20 genome-wide linkage scans across multiple populations have been launched to hunt for osteoporosis susceptibility genes. Some significant or suggestive chromosomal regions of linkage to bone mineral density have been identified and replicated in genome-wide linkage screens. However, identification of key candidate genes within these confirmed regions is challenging. We used five freely available bioinformatics tools (Prioritizer, GeneSeeker, PROSPECTR and SUSPECTS, Disease Gene Prediction, and Endeavor) to analyze the 13 well-replicated osteoporosis susceptibility loci: 1p36, 1q21–25, 2p22–24, 3p14–25, 4q25–34, 6p21, 7p14–21, 11q14–25, 12q23–24, 13q14–34, 20p12, 2q24–32, and 5q12–21. Pathways and regulatory network analyses were performed using the Ingenuity Pathways Analysis (IPA) software. We identified a subset of most likely candidate osteoporosis susceptibility genes that are largely involved in transforming growth factor (TGF)- β signaling, granulocyte-macrophage colony-stimulating factor (GM-CSF) signaling, axonal guidance signaling, peroxisome proliferator-activated receptor (PPAR) signaling, and Wnt/ β -catenin signaling pathway. Six non-overlapping networks were generated by IPA 5.0 from 88 out of the 91 candidate genes. The list of most likely candidate genes and the associated pathway identified will

assist researchers in prioritizing candidate disease genes for further empirical analysis and understanding the pathogenesis of osteoporosis.

Keywords Osteoporosis · Bone mineral density · Genetics · Candidate gene · Bioinformatics

Introduction

Complex diseases such as osteoporosis are influenced by multiple genes with small individual effects, environments, and their interaction. Identification of genetic components of complex diseases is the greatest challenge for human geneticists. For the past two decades, the dominant study design has been linkage analysis in families, which identifies broad intervals of several megabases of DNA (quantitative trait loci) that correlate with the disease status in pedigrees. Linked DNA intervals can encompass dozens to hundreds of candidate genes that might be involved in or causal for the disease. Transition from quantitative trait loci to gene has been proven to be difficult due to the absence of complete functional information for the majority of genes in this susceptibility locus and the limited knowledge of the link between gene function and disease.

Osteoporosis is a complex disease with a strong genetic component. To date, more than 20 genome-wide linkage scans across multiple populations have been launched to hunt for osteoporosis susceptibility genes (Huang and Kung 2006; Ralston et al. 2005; Streeten et al. 2006; Xiao et al. 2006; Hsu et al. 2007; Lee et al. 2006; Ioannidis et al. 2007). Some significant or suggestive chromosomal regions of linkage to bone mineral density (BMD) have been identified and replicated in genome-wide linkage screens. The next daunting task is

Q.-Y. Huang (✉) · G. H. Y. Li · W. M. W. Cheung ·
A. W. C. Kung
Department of Medicine, The University of Hong Kong,
Hong Kong, China
e-mail: huangqy@hkucc.hku.hk; awckung@hkucc.hku.hk

Y.-Q. Song
Department of Biochemistry,
The University of Hong Kong, Hong Kong, China

identifying key candidate genes within these confirmed regions. Exhaustive surveys of all variation in the intervals are needed to determine which genes within these chromosomal regions account for the linkage signals found. Despite the recent drop in genotyping cost, this kind of study is still expensive and in many cases not feasible. Now some promising bioinformatics tools are available for disease-gene identification (Lopez-Bigas and Ouzounis 2004; van Driel et al. 2005; Franke et al. 2006; Adie et al. 2006; Aerts et al. 2006). These tools use information extracted from public online databases, such as sequence data, medical literature, gene ontology, and function annotation, as well as information on biology, function, and gene expression. These methods have been successfully used to prioritize candidate disease genes for type 2 diabetes and obesity (Tiffin et al. 2006; Elbers et al. 2007). Recently, text-mining methods have been applied to the discovery of novel genes related to bone biology (Gajendran et al. 2007). In this study, we used five freely available bioinformatics tools to analyze the 13 most promising osteoporosis susceptibility loci encompassing the 5,492 positional candidate genes. We prioritized a subset of most likely candidate osteoporosis susceptibility genes for further empirical test.

Materials and methods

Susceptibility loci selection

To date, more than 20 genome-wide linkage scans across multiple populations have been performed to identify quantitative trait loci underlying BMD variations (Huang and Kung 2006; Ralston et al. 2005; Streeten et al. 2006; Xiao et al. 2006; Hsu et al. 2007; Lee et al. 2006; Ioannidis et al. 2007). Only those loci that were significant or suggestive [\log of odds (LOD > 2.2)] at least once and replicated (LOD > 1) in at least one independent study were included (Lander and Kruglyak 1995). The regions 1p36, 1q21–25, 2p22–24, 3p14–25, 4q25–34, 6p21, 7p14–21, 11q14–25, 12q23–24, 13q14–34, and 20p12 have been identified. In addition, two significant regions (2q24–32 and 5q12–21) recently identified by a genome-wide linkage scan in a Chinese population (Hsu et al. 2007) are also included for the purpose of our interest. In total, 13 osteoporosis susceptibility loci were selected for analysis (Table 1).

Gene identification methods

Five bioinformatics tools are freely available for disease-gene identification: Disease Gene Prediction (DGP) (Lopez-Bigas and Ouzounis 2004), GeneSeeker (van Driel

et al. 2005), Prioritizer (Franke et al. 2006), PROSPECTR and SUSPECTS (PandS) (Adie et al. 2006), and Endeavor (Aerts et al. 2006). Prioritizer ranks genes based on its functional interaction with genes on different susceptibility loci, assuming that disease genes in a specific disorder are usually functionally related. GeneSeeker points to genes that are expressed in disease-related tissues. PROSPECTR differentiates between genes that are likely to be involved in diseases and those that are not involved; it uses sequence-based features such as gene length, protein length, and percentage identity of homologues in other species. SUSPECTS scores candidate genes using the PROSPECTR and also assess the similarity between their annotation and that of already known disease genes. DGP assigns probabilities to genes that could indicate involvement in hereditary diseases using parameters based on conservation, phylogenetic extent, protein length, and paralogy. Endeavor is a software application for computational prioritization of test genes based on a training set of genes already known to be involved in the disease of interest. The ranking of a test gene is based on its similarity with training genes. These five tools were combined to analyze the 13 promising osteoporosis susceptibility loci encompassing the 5,492 positional candidate genes. The 1p36, 1q21–25, 2p22–24, 3p14–25, 4q25–34, 6p21, 7p14–21, 11q14–25, 12q23–24, 13q14–34, 20p12, 2q24–32, and 5q12–21 loci were used as inputs in GeneSeeker, DGP, Endeavor, and PandS. The chromosomal start and end locations in the base pair of the corresponding loci were input in Prioritizer. Endeavor and PandS had to be trained with a set of genes. *VDR*, *ESR1*, *ESR2*, *IL6*, *COL1A1*, *LRP5*, *IGF1*, *BMP2*, *SOST*, *CLCN7*, *TGFB1*, and *CYP19A1* are already known to be involved in BMD and/or osteoporosis and were therefore used as training genes. If a training gene was located in a selected susceptibility locus, it was taken out from the training genes while this locus was analyzed. GeneSeeker required disease-related tissue as input. Bone(s), skeletal system, osteoblast, osteoclast, spine, cartilage, marrow, parathyroids, ovary, and testis were used.

Identification of candidate genes

GeneSeeker pinpoints genes expressed in disease-related tissue. Therefore, we took all genes pinpointed by this method into consideration. All other tools produce rankings, and therefore, the top 20 genes from each method were included for comparison. A gene was considered to be interesting as a candidate gene if it was indicated by three or more of the tools. As Endeavor, DGP, and PandS partly use the same input information and show similar outputs, candidate genes were excluded if they were solely identified by these three methods.

Table 1 List of 91 genes in 13 susceptibility loci for osteoporosis pinpointed by the five disease-gene identification tools

Locus	Gene no.	Gene	Tool					Genetic and functional association with osteoporosis	Reference		
			GeneSeeker	Prioritizer	DGP	PandS	Endeavor				
1p36	558	<i>ALPL</i>	•		•	•	•	Bone-specific alkaline phosphatase. It is associated with BMD	Goseki-Sone et al. (2005)		
		<i>HSPG2</i>	•	•	•		•	Essential for cartilage and cephalic development			
		<i>PLOD</i>	•		•	•	•	Well-known candidate gene			
		<i>ENO1</i>	•				•				
		<i>WNT4</i>		•	•	•		A member of the <i>WNT</i> gene family			
		<i>TNFR2</i>	•	•			•	Well-known candidate gene			
		1q21–25	826	<i>LAMC2</i>	•	•	•	•		•	
<i>SELE</i>	•						•				
2p22–24	244	<i>CYP11B1</i>	•	•	•	•	•	Well-known candidate gene	Chapman et al. (2001), van der Weyden et al. (2006)		
		<i>MATN3</i>	•	•	•	•	•	Specifically expressed in cartilaginous tissues and has a role in the development and homeostasis of cartilage and bone. Mutations in this gene result in multiple epiphyseal dysplasia. Functional knockout of the <i>MATN3</i> gene increases BMD in mice			
		<i>POMC</i>	•	•	•	•		Well-known candidate gene			
		<i>HADHA</i>		•	•	•	•				
		<i>GDF7</i>		•	•	•	•				
		<i>ALK</i>	•				•	•		Involvement in regulating normal bone growth	
		<i>EMILIN1</i>		•			•	•			
		<i>SLC5A6</i>		•		•	•				
		<i>APOB</i>		•			•	•			
		<i>SRD5A2</i>	•			•	•				
		3p14–25	625	<i>TDGF1</i>	•	•	•	•		•	
				<i>PPARG</i>		•	•	•		•	Associated with BMD
				<i>FBLN2</i>	•	•		•		•	<i>FBLN2</i> encodes an extracellular matrix protein, which belongs to the fibulin family. This protein binds various extracellular ligands and calcium
3p14–25	625	<i>CCR2</i>	•	•			•	Associated with BMD	Yamada et al. (2002)		
		<i>SCN5A</i>	•		•	•					
		<i>ACVR2B</i>	•	•		•					
		4q25–34	426	<i>EGF</i>	•	•		•		•	EGF stimulates proliferation of osteoblast and inhibits its differentiation
<i>CLGN</i>	•			•			•				
<i>FGB</i>				•	•	•					
<i>FGF2</i>				•		•	•				

Table 1 continued

Locus	Gene no.	Gene	Tool					Genetic and functional association with osteoporosis	Reference
			GeneSeeker	Prioritizer	DGP	PandS	Endeavor		
6p21	397	<i>RUNX2</i>	•	•	•	•	•	Mutations in this gene have been associated with the bone development disorder cleidocranial dysplasia (CCD). SNPs in the P2 promoter were significantly associated with bone density	Vaughan et al. (2002, 2004), Doecke et al. (2006)
		<i>COL11A2</i>	•	•	•	•	•	Minor collagen of cartilage	
		<i>PPARD</i>	•			•	•	Activation of PPARdelta enhances osteoblast differentiation	
		<i>TAPBP</i>	•	•		•			
		<i>PIMI</i>	•	•			•		
7p14–21	263	<i>INHBA</i>	•	•		•	•		Well-known candidate gene SFRPs act as soluble modulators of Wnt signaling. SFRP4 suppresses the proliferation of osteoblast. SFRP4 is a negative regulator of peak BMD in SAMP6 mice Nakanishi et al. (2006, 2008)
		<i>IL6</i>	•		•	•	•		
		<i>SFRP4</i>		•	•	•	•		
		<i>SOSTDC1</i>	•	•			•	<i>SOSTDC1</i> is a member of the sclerostin family and encodes an <i>N</i> -glycosylated, secreted protein, which functions as a BMP antagonist. Specifically, it directly associates with BMPs, prohibiting them from binding their receptors, thereby regulating BMP signaling during cellular proliferation, differentiation	
		<i>NPY</i>	•		•	•			
		<i>ADCYAP1R1</i>	•		•	•			
		<i>HOXA13</i>	•		•	•			
		<i>GHRHR</i>		•		•	•		
		<i>HOXA1</i>	•	•		•			
		<i>ITGB8</i>		•		•	•		
11q14–25	568	<i>MMP13</i>	•	•		•	•	Matrix metalloproteinase 13	
		<i>MMP3</i>	•	•		•	•	Matrix metalloproteinase 3	
		<i>MMP1</i>	•	•		•		Matrix metalloproteinase 1	
		<i>ETS1</i>	•		•		•	ETS1 may play a role in regulation of osteoblast differentiation	
		<i>SORL1</i>		•		•	•		
		<i>FZD4</i>		•	•	•		An LRP5 coreceptor	
		<i>APOA1</i>	•		•		•		
12q23–24	444	<i>IGF1</i>	•	•	•	•	•	Well-known candidate gene	
		<i>PTPN11</i>	•		•	•	•		
		<i>ACADS</i>		•	•	•	•		
		<i>CIT</i>	•			•	•		
		<i>ABCB9</i>	•			•	•		

Table 1 continued

Locus	Gene no.	Gene	Tool					Genetic and functional association with osteoporosis	Reference
			GeneSeeker	Prioritizer	DGP	PandS	Endeavor		
13q14–34	411	COL4A1	•	•	•	•	•	Collagen of basement membrane	
		ZIC2	•	•	•	•	•		
		<i>EDNRB</i>	•		•	•	•		
		<i>F7</i>		•	•	•	•		
		<i>IRS2</i>		•	•	•	•		
		<i>F10</i>		•	•	•	•		
		<i>PCDH8</i>	•			•	•		
		<i>RB1</i>	•		•		•		
		<i>PROZ</i>		•		•	•		
		<i>CPB2</i>		•		•	•		
		<i>GAS6</i>		•		•	•		
		<i>FGF14</i>		•	•	•			
		20p12	82	BMP2	•	•	•		
<i>JAG1</i>				•	•	•	•		
<i>MCM8</i>				•	•	•	•		
<i>PAK7</i>				•	•	•	•		
<i>PLCB4</i>				•	•	•	•		
<i>CHGB</i>	•					•	•		
<i>HAO1</i>				•	•				
<i>SNAP25</i>				•		•	•		
<i>PCNA</i>				•		•	•		
2q24–32	295			ITGAV	•	•	•	•	•
		COL3A1	•	•	•	•	•	Essential for normal fibrillinogenesis of collagen I	
		<i>ACVR1</i>	•	•		•	•	A recurrent mutation in <i>ACVR1</i> causes inherited and sporadic fibrodysplasia ossificans progressiva	Shore et al. (2006)
		<i>STAT1</i>	•		•	•	•	LRP knockout animals showed bone disease with abnormal urinary excretion of 25(OH) vitamin D ₃	
		<i>LRP2</i>		•	•	•	•		
		<i>NEUROD1</i>	•	•		•	•		
		<i>NR4A2</i>			•	•	•		
		5q12–21	353	<i>THBS4</i>	•	•		•	•
<i>HAPLN1</i>	•			•		•			
<i>CSPG2</i>	•					•	•		
<i>PIK3R1</i>	•					•	•		
<i>CRHBP</i>	•			•		•			
<i>HTR1A</i>	•			•		•			
<i>F2RL1</i>				•		•	•		
<i>NR2F1</i>				•		•	•		

DGP Disease Gene Prediction, *PandS* PROSPECTR and SUSPECTS, *BMD* bone mineral density, *SNP* single nucleotide polymorphisms
The bold signifies the gene pinpointed by all the five tools

Table 2 Canonical pathways pinpointed by candidate genes identified

Pathway	–log <i>P</i>	Genes
TGF- β signaling	4.45	<i>ACVRI, ALK, RUNX2, INHBA, BMP2, ACVR2B</i>
GM-CSF signaling	4.41	<i>PIK3R1, STAT1, PTPN11, ETS1, PIM1</i>
Axonal guidance signaling	3.92	<i>PAK7, PIK3R1, EGF, IGF1, PLCB4, GDF7, WNT4, PTPN11, ALK, BMP2</i>
PPAR signaling	3.64	<i>PTPN11, NR2F1, PPARG, PPARD, TNFRSF1B</i>
Wnt/ β -catenin signaling	3.34	<i>ACVRI, WNT4, FZD4, PPARD, ACVR2B, SFRP4</i>

Pathways and network analyses

Candidate genes identified were imported into the ingenuity pathways analysis (IPA) 5.0 (Ingenuity Systems, Mountain View, CA, USA) to generate putative signaling networks based on the manually curated knowledge database of pathway interactions extracted from the literature. The network was generated by the input genes using both direct and indirect relationships. These networks were ranked by scores that measured the probability that the genes were included in the network by chance alone. Networks with scores > 3 have a 99.9% confidence of not being generated by random chance (Raponi et al. 2004). The overlapping networks were merged to produce the largest possible network such that the number of biological relationships to be examined was maximized. The correctness of the relationships was checked manually based on categorized literature findings provided by the application. Canonical pathways associated with input candidate genes were elucidated with a statistical significance value.

Results

Ninety-one genes were selected as potential osteoporosis susceptibility genes by using five freely available bioinformatics tools for disease-gene identification (Table 1). Some interesting genes indicated by the five disease-gene identification methods are already known to be associated with BMD/osteoporosis. These include genes encoding procollagen lysyl hydroxylase (PLOD), tumor necrosis factor receptor-2 (TNFR2), cytochrome p450 protein (CYP)1B1, proopiomelanocortin (POMC), interleukin (IL)-6, insulin-like growth factor (IGF)-1, and bone morphologic protein (BMP)-2. Of particular interest is the *MATN3* gene. The matrilins are a four-member family of noncollagenous extracellular matrix proteins. Matrilin-3 is specifically expressed in cartilaginous tissue and has a role in the development and homeostasis of cartilage and bone. Mutations in the *MATN3* gene have been reported in a variety of skeletal diseases, including multiple epiphyseal dysplasia (Chapman et al. 2001). A functional knockout of

the *Matn3* gene increases BMD in mice (van der Weyden et al. 2006). The *MATN3* gene has not been previously implicated in osteoporosis (Huang and Kung 2006). Thus, it will be of great interest to see whether polymorphisms in *MATN3* are associated with BMD/osteoporosis in humans.

Another interesting gene is *RUNX2*. *RUNX2* is an osteoblast-specific transcription factor 2 that is responsible for controlling osteoblast differentiation and transactivating genes involved in the deposition of bone matrix, such as osteocalcin, osteopontin, and type I collagen. *Runx2*-gene knockout results in animals with no osteoblasts and unmineralized skeleton. Overexpression of a dominant negative form of *Runx2* in mature osteoblasts of transgenic mice leads to less active osteoblasts, reduced bone formation, and short stature. A single nucleotide polymorphism (SNP) in exon 2 of *RUNX2* was associated with higher BMD in an Australian population (Vaughan et al. 2002) and was replicated in a Scottish population (Vaughan et al. 2004). Recently, Doecke et al. (2006) showed that only SNPs in the *RUNX2* P2 promoter were significantly associated with BMD and that greater *RUNX2* P2 promoter activity was associated with higher BMD.

Five canonical pathways are pinpointed by candidate genes identified (Table 2):

1. TGF- β signaling pathway: BMP-2 belongs to the TGF- β superfamily. There are three type II receptors for BMPs, including type II BMP receptor, as well as type II and IIB activin receptors (ActR-II and ACVR2B). BMP-2, as well as cytokines of TGF- β , binds to the type II receptor, which phosphorylates and activates activin-receptor-like kinases (ALKs). Ectopic expression of ACVRI resulted in osteoblast differentiation of chondrocytes (Valcourt et al. 2002). Tissue-specific deletion of ALK2 in neural crest cells resulted in craniofacial defects such as hypotrophic mandibles, thereby suggesting its involvement in regulating normal bone growth (Dudas et al. 2004). BMP signaling is crucial for osteoblastogenesis mediated via Runx2-dependent process (Phimphilai et al. 2006). Osteoblasts regulate mineralization via the coordinated secretion of activin A and follistatin. Whereas activin A also belongs to TGF- β and inhibits mineralization

- in vitro, follistatin increased the process (Eijken et al. 2007). The coordination of activins and inhibins in controlling osteoblastogenesis and osteoclastogenesis has been demonstrated (Gaddy-Kurten et al. 2002).
2. Granulocyte-macrophage colony-stimulating factor (GM-CSF) signaling pathway: Glycoprotein (gp)130-dependent cytokines, which mediate their biological actions via either the signal transducer and activator of transcription (STAT) 1/3 or protein tyrosine phosphatase nonreceptor (PTPN)11/renin angiotensin system (RAS)/mitogen-activated protein kinase (MAPK) pathways, regulate osteoblast and osteoclast formation (Sims et al. 2004). Cytoplasmic Src homology 2-containing phosphotyrosine phosphatase (SHP2) is encoded by *PTPN11*. In bone formation, SHP2 forms a growth factor receptor-bound protein 2 (Grb2)/fibroblast growth factor receptor substrate 2 ((FRS2)/SHP2 complex, which is essential for fibroblast growth factor (FGF) signaling (Mansukhani et al. 2000). FGF induces proliferation in preosteoblasts, whereas in differentiating osteoblasts, FGF inhibits osteoblast differentiation as revealed by its inhibitory effects on alkaline phosphatase activity and mineralization. One important target of FGF is v-ets erythroblastosis virus E26 oncogene homolog 1 (*ETS1*). The latter belongs to a family of transcription factors defined by a conserved ETS DNA binding domain. Treatment with retinoic acid (RA) increases the expression of *ETS1* in preosteoblast cell line MC3T3-E1 cells (Raouf et al. 2000). The conserved sequence for *ETS1* is located in the AJ18 promoter, suggesting its possible role in the regulation of osteoblast differentiation via the transactivation of AJ18 (Jheon et al. 2003).
 3. Axonal guidance signaling pathway: Both premature and mature osteoblasts express and secrete IGF-I. Endogenous transcript expression of IGF-I was detected in osteoblasts and osteocytes in rat tibia (Reijnders et al. 2007). Mechanical loading increases IGF-I expression and is proposed to be a means of translating mechanical loading to osteoblast formation. Epidermal growth factor (EGF), as well as other ligands of the EGF receptor, might regulate bone growth and resorption via the modulation of IGF-I and/or IGF-I-binding protein activities (Xian 2007). As a result, proliferation of osteoblasts is stimulated, but their differentiation is inhibited.
 4. Peroxisome proliferator-activated receptor (PPAR) signaling pathway: Nuclear receptor subfamily 2 (NR2)F1 encodes nuclear receptor chicken ovalbumin upstream promoter-transcription factor-1 (COUP-TFI) that binds to the regulatory response element in BMP-4 promoter and serves as a silencer for the transcription of BMP-4 in osteoblasts (Feng et al. 1995). Tumor necrosis factor (TNF)- α inhibits osteoblast differentiation while upregulating bone resorption (Nanes 2003). Actions of TNF- α require either cell-surface receptor TNFR1 or TNFR2 encoded by *TNFRSF1A* and *TNFRSF1B*, respectively (Chen and Goeddel 2002). The nuclear receptor PPARgamma is an essential key regulator for adipogenesis. Overexpression of PPARgamma2 and its subsequent activation inhibits osteoblast differentiation of mesenchymal stem cells and preosteoblasts (Jeon et al. 2003; Kim et al. 2005). Activation of Wnt/ β -catenin signaling enhances osteoblast differentiation of mesenchymal stem cells (MSCs) via the simultaneous suppression of PPARgamma and CCAAT enhancer binding protein alpha (C/EBPalpha) (Kang et al. 2007). Recent studies hypothesized that activation of PPARdelta enhances osteoblast differentiation (Dang and Lowik 2004).
 5. Wnt/ β -catenin signaling pathway: WNT proteins interact with the receptor frizzled and coreceptor lipoprotein-related protein (LRP)5/6 to activate the stabilization of β -catenin. The cytoplasmic β -catenin translocates to the nucleus and promotes the expression of osteoblast-specific genes. Secreted frizzled-related protein 4 (SFRP4) negatively modulates WNT signaling pathway (Mayr et al. 1997). SFRP4 attenuates bone formation and suppresses the proliferation of osteoblasts, possibly via its antagonistic effect on the WNT signaling pathway (Nakanishi et al. 2006). In addition, six nonoverlapping networks were generated by IPA 5.0 from 88 out of the 91 candidate genes (Fig. 1).

Discussion

A number of genomic regions have been identified in previous genome-wide linkage scans for osteoporosis. The challenge now is to find osteoporosis genes in these chromosomal regions. This study generated a list of potential candidate genes in 13 susceptibility loci for osteoporosis using a combination of five computational disease-gene identification methods. These genes are largely involved in TGF- β -, GM-CSF-, axonal-guidance-, PPAR-, and Wnt/ β -catenin-signaling pathways. Some genes have already been genetically or functionally associated with osteoporosis. For example, *BMP2* is the first osteoporosis gene identified by linkage and positional cloning in humans. In this study, *BMP2* was also indicated as such by all the five gene identification methods. Thus, at least some candidate genes presented here might be true candidates. On the other hand, there is no evidence that some candidate genes are

Fig. 1 Six nonoverlapping networks were generated by ingenuity pathways analysis (IPA) 5.0 from 88 out of the 91 candidate genes. *Shaded* genes are the potential candidate genes identified by the bioinformatics approach, and others are those associated with the selected candidate genes based on pathway analysis. **a** Legend for the edges and nodes in the IPA networks. **b** The network with score of 47 and 23 focus genes. **c** The network with score of 41 and 21 focus genes. **d** The network with score of 29 and 16 focus genes. **e** The network with score of 20 and 12 focus genes. **f** The network with score of 18 and 11 focus genes. **g** The network with score of 6 and 5 focus genes

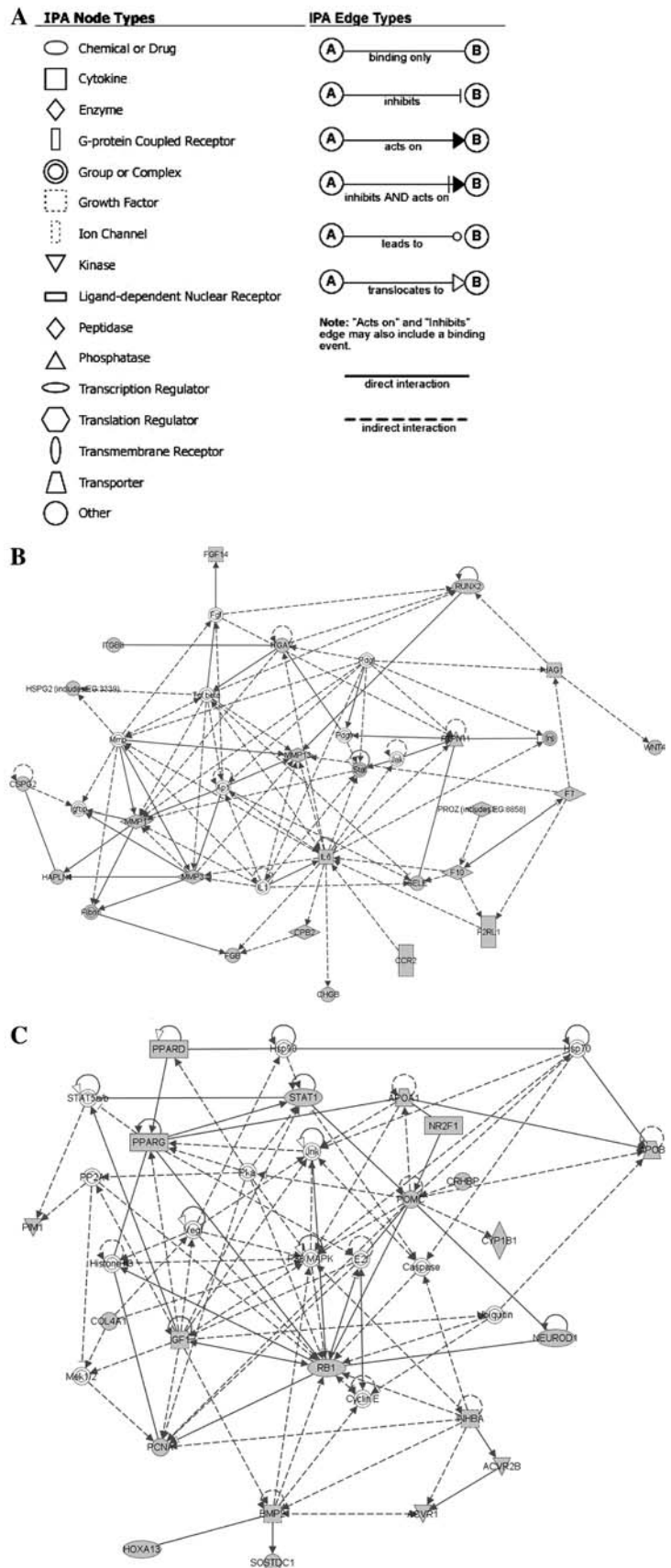
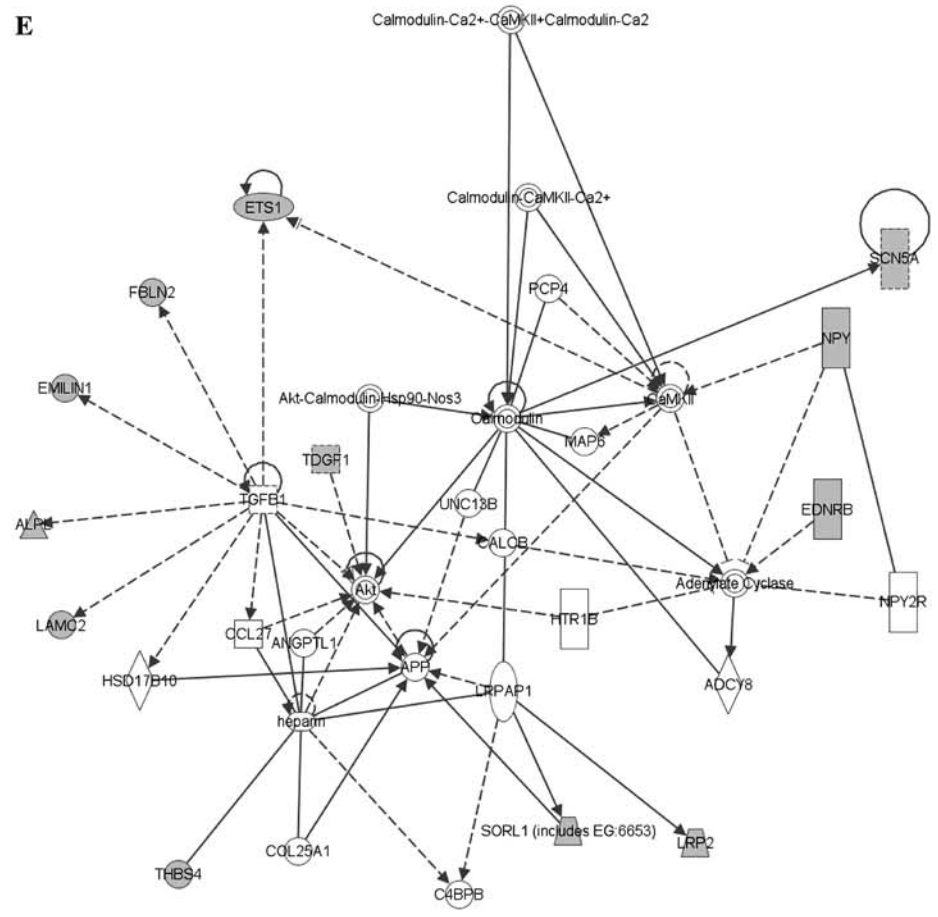
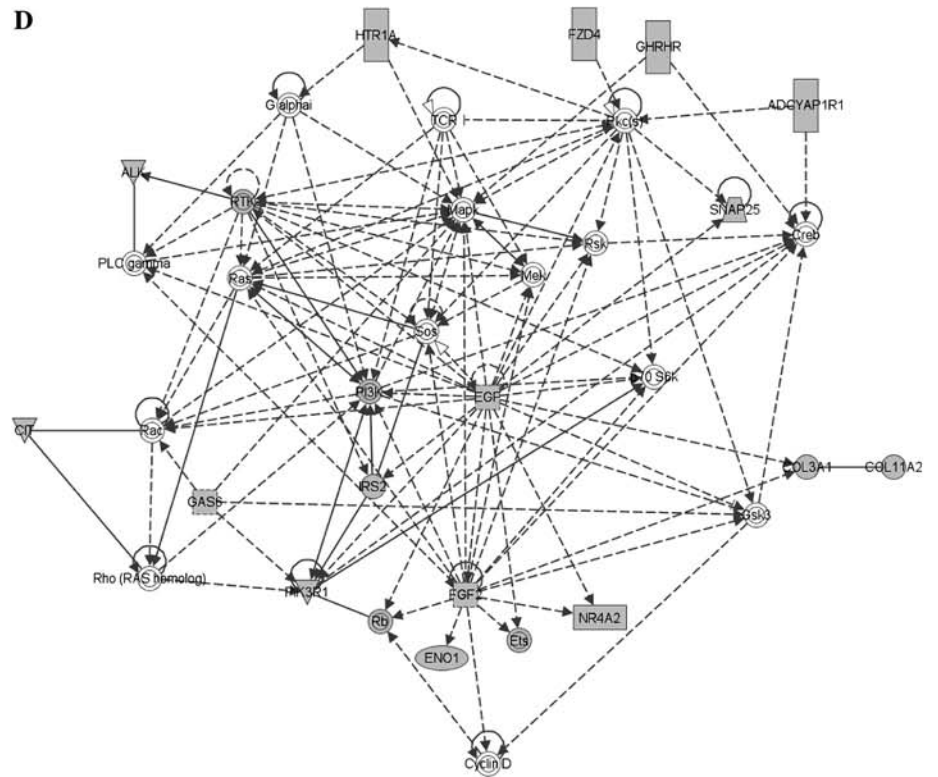


Fig. 1 continued



genetically or functionally involved in osteoporosis. An appropriate follow-up to our study would be to determine the roles played in osteoporosis by the novel genes identified through computational disease-gene-identification strategy. This two-step process of identifying candidate genes using bioinformatics tools followed by conventional experimentation (e.g., gene-wide and tag-SNP-based association analyses in a large population and/or family samples) will greatly expedite the process of gene discovery in complex diseases such as osteoporosis.

The bioinformatics approaches employ information extracted from public online databases and are based on the assumption that novel disease genes may share some characteristics with identified disease genes. Some potential bias to disease-gene selection might be introduced. First, the extent of gene annotation and available data for genes vary greatly. The greater the extent of annotated data for a gene, the more likely it is to be selected. A gene that has been extensively studied for a long time will have a large amount of associated literature and has a better chance of being selected. Moreover, it tends to point to function/features already associated with the disease. Second, some methods have a significant overlap with other methods in their input data. For example, Endeavor, DGP, and PandS partly use the same input information and show similar outputs. The selection of a candidate gene by several methods using the same input data may be less valuable than the selection of a candidate by several methods using disparate data sources. Some of these limitations can be overcome by using multiple methods and data sources in a complementary fashion. In this study, we excluded candidate genes that were solely identified by Endeavor, DGP, and PandS.

Although these methods yielded a list of most likely candidates rather than a single gene, computational approaches reduced the 5,492 positional candidate genes from genome-wide studies to just 92 most likely candidates in 13 susceptibility loci linked with osteoporosis/BMD in multiple studies. Our results showed that computational approaches are helpful in the hunt for complex disease genes and the identification of pathways and regulatory networks involved in complex disorders. We believe the list of most likely candidate genes and the associated pathways identified will assist researchers in prioritizing candidate disease genes for further empirical analysis and understanding the pathogenesis of osteoporosis.

Acknowledgments This project is supported by Hong Kong Research Grant Council (HKU7514/06M), seed funding for basic research, The University of Hong Kong, The KC Wong Education Foundation, and Osteoporosis Research Fund.

Competing interests The authors have no conflicts of interest.

References

- Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS (2006) SUSPECTS: enabling fast and effective prioritization of positional candidates. *Bioinformatics* 22:773–774
- Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, Tranchevent LC, De Moor B, Marynen P, Hassan B, Carmeliet P, Moreau Y (2006) Gene prioritization through genomic data fusion. *Nat Biotechnol* 24:537–544
- Chapman KL, Mortier GR, Chapman K, Loughlin J, Grant ME, Briggs MD (2001) Mutations in the region encoding the von Willebrand factor A domain of matrilin-3 are associated with multiple epiphyseal dysplasia. *Nat Genet* 28:393–396
- Chen G, Goeddel DV (2002) TNF-R1 signaling: a beautiful pathway. *Science* 296:1634–1635
- Dang Z, Lowik CW (2004) The balance between concurrent activation of ERs and PPARs determines daidzein-induced osteogenesis and adipogenesis. *J Bone Miner Res* 19:853–861
- Doecke JD, Day CJ, Stephens AS, Carter SL, van Daal A, Kotowicz MA, Nicholson GC, Morrison NA (2006) Association of functionally different RUNX2 P2 promoter alleles with BMD. *J Bone Miner Res* 21:265–273
- Dudas M, Sridurongrit S, Nagy A, Okazaki K, Kaartinen V (2004) Craniofacial defects in mice lacking BMP type I receptor Alk2 in neural crest cells. *Mech Dev* 121:173–182
- Eijken M, Swagemakers S, Koedam M, Steenbergen C, Derkx P, Uitterlinden AG, van der Spek PJ, Visser JA, de Jong FH, Pols HA, van Leeuwen JP (2007) The activin A-follistatin system: potent regulator of human extracellular matrix mineralization. *FASEB J* 21:2949–2960
- Elbers CC, Onland-Moret NC, Franke L, Niehoff AG, van der Schouw YT, Wijmenga C (2007) A strategy to search for common obesity and type 2 diabetes genes. *Trends Endocrinol Metab* 18:19–26
- Feng JQ, Chen D, Cooney AJ, Tsai MJ, Harris MA, Tsai SY, Feng M, Mundy GR, Harris SE (1995) The mouse bone morphogenetic protein-4 gene: analysis of promoter utilization in fetal rat calvarial osteoblasts and regulation by COUP-TFI orphan receptor. *J Biol Chem* 270:28364–28373
- Franke L, Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am J Hum Genet* 78:1011–1025
- Gaddy-Kurten D, Coker JK, Abe E, Jilka RL, Manolagas SC (2002) Inhibin suppresses and activin stimulates osteoblastogenesis and osteoclastogenesis in murine bone marrow cultures. *Endocrinology* 143:74–83
- Gajendran VK, Lin JR, Fyhrie DP (2007) An application of bioinformatics and text mining to the discovery of novel genes related to bone biology. *Bone* 40:1378–1388
- Goseki-Sone M, Sogabe N, Fukushi-Irie M, Mizoi L, Orimo H, Suzuki T, Nakamura H, Orimo H, Hosoi T (2005) Functional analysis of the single nucleotide polymorphism (787T > C) in the tissue-nonspecific alkaline phosphatase gene associated with BMD. *J Bone Miner Res* 20:773–782
- Hsu YH, Xu X, Terwedow HA, Niu T, Hong X, Wu D, Wang L, Brain JD, Bouxsein ML, Cummings SR, Rosen CJ, Xu X (2007) Large-scale genome-wide linkage analysis for loci linked to BMD at different skeletal sites in extreme selected sibships. *J Bone Miner Res* 22:184–194
- Huang QY, Kung AWC (2006) Genetics of osteoporosis. *Mol Genet Metab* 88:295–306
- Ioannidis JPA, Ng MY, Sham PC, Zintzaras E, Lewis CM, Deng HW, Econs MJ, Karasik D, Devoto M, Kammerer CM, Spector T,

- Andrew T, Cupples LA, Duncan EL, Foroud T, Kiel DP, Koller D, Langdahl B, Mitchell BD, Peacock M, Recker R, Shen H, Sol-Church K, Spotila LD, Uitterlinden AG, Wilson SG, Kung AW, Ralston SH (2007) Meta-analysis of genome wide scans provides evidence for gender and site specific regulation of bone mass. *J Bone Miner Res* 27:173–183
- Jeon MJ, Kim JA, Kwon SH, Kim SW, Park KS, Park SW, Kim SY, Shin CS (2003) Activation of peroxisome proliferator-activated receptor-gamma inhibits the Runx2-mediated transcription of osteocalcin in osteoblasts. *J Biol Chem* 278:23270–23277
- Jheon AH, Suzuki N, Nishiyama T, Cheifetz S, Sodek J, Ganss B (2003) Characterization of the 5'-flanking region of the rat AJ18 gene. *Gene* 310:203–213
- Kang S, Bennett CN, Gerin I, Rapp LA, Hankenson KD, Hankenson KD, Macdougald OA (2007) Wnt signaling stimulates osteoblastogenesis of mesenchymal precursors by suppressing CCAAT/enhancer-binding protein alpha and peroxisome proliferator-activated receptor gamma. *J Biol Chem* 282:14515–14524
- Kiel DP, Ferrari S, Cupples LA, Karasik D, Dupuis J, Rosen CJ, Imamovic A, Demissie S (2005) Polymorphisms in the PPAR γ gene influence bone density in humans. *J Bone Miner Res* 20:S234
- Kim SW, Her SJ, Kim SY, Shin CS (2005) Ectopic overexpression of adipogenic transcription factors induces transdifferentiation of MC3T3-E1 osteoblasts. *Biochem Biophys Res Commun* 327:811–819
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting the reporting results. *Nat Genet* 11:241–247
- Lee YH, Rho YH, Choi SJ, Ji JD, Song GG (2006) Meta-analysis of genome-wide linkage studies for bone mineral density. *J Hum Genet* 51:480–486
- Lopez-Bigas N, Ouzounis CA (2004) Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res* 32:3108–3114
- Mansukhani A, Bellosta P, Sahni M, Basilico C (2000) Signaling by fibroblast growth factors (FGF) and fibroblast growth factor receptor 2 (FGFR2)-activating mutations blocks mineralization and induces apoptosis in osteoblasts. *J Cell Biol* 149:1297–1308
- Mayr T, Deutsch U, Kühl M, Drexler HC, Lottspeich F, Deutzmann R, Wedlich D, Risau W (1997) Frizzled-3: a secreted frizzled-related protein that inhibits Wnt activity. *Mech Dev* 63:109–125
- Nakanishi R, Akiyama H, Kimura H, Otsuki B, Shimizu M, Tsuboyama T, Nakamura T (2008) Osteoblast-targeted expression of *sfrp4* in mice results in low bone mass. *J Bone Miner Res* 23:271–277
- Nakanishi R, Shimizu M, Mori M, Akiyama H, Okudaira S, Otsuki B, Hashimoto M, Higuchi K, Hosokawa M, Tsuboyama T, Nakamura T (2006) Secreted frizzled-related protein 4 is a negative regulator of peak BMD in SAMP6 mice. *J Bone Miner Res* 21:1713–1721
- Nanes MS (2003) Tumor necrosis factor-alpha: molecular and cellular mechanisms in skeletal pathology. *Gene* 321:1–15
- Phimphilai M, Zhao Z, Boules H, Roca H, Franceschi RT (2006) BMP signaling is required for RUNX2-dependent induction of the osteoblast phenotype. *J Bone Miner Res* 21:637–646
- Ralston SH, Galwey N, MacKay I, Albagha OME, Cardon L, Compston JE, Cooper C, Duncan E, Keen R, Langdahl B, McLellan A, O'Riordan J, Pols HA, Reid DM, Uitterlinden AG, Wass J, Bennett ST (2005) Loci for regulation of bone mineral density in men and women identified by genome wide linkage scan: the FAMOS study. *Hum Mol Genet* 14:943–951
- Raouf A, Li V, Kola I, Watson DK, Seth A (2000) The Ets1 proto-oncogene is upregulated by retinoic acid: characterization of a functional retinoic acid response element in the Ets1 promoter. *Oncogene* 19:1969–1974
- Raponi M, Belly RT, Karp JE, Lancet JE, Atkins D, Wang Y (2004) Microarray analysis reveals genetic pathways modulated by tipifarnib in acute myeloid leukemia. *BMC Cancer* 4:56
- Reijnders CM, Bravenboer N, Tromp AM, Blankenstein MA, Lips P (2007) Effect of mechanical loading on insulin-like growth factor-I gene expression in rat tibia. *J Endocrinol* 192:131–140
- Shore EM, Xu M, Feldman GJ, Fenstermacher DA, Cho TJ, Choi IH, Connor JM, Delai P, Glaser DL, LeMerrer M, Morhart R, Rogers JG, Smith R, Triffitt JT, Urtizberea JA, Zasloff M, Brown MA, Kaplan FS (2006) A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat Genet* 38:525–527
- Sims NA, Jenkins BJ, Quinn JM, Nakamura A, Glatt M, Gillespie MT, Ernst M, Martin TJ (2004) Glycoprotein 130 regulates bone turnover and bone size by distinct downstream signaling pathways. *J Clin Invest* 113:379–389
- Streeten EA, McBride DJ, Pollin TI, Ryan K, Shapiro J, Ott S, Mitchell BD, Shuldiner AR, O'Connell JR (2006) Quantitative trait loci for BMD identified by autosome-wide linkage scan to chromosomes 7q and 21q in men from the Amish Family Osteoporosis Study. *J Bone Miner Res* 21:1433–1442
- Tiffin N, Adie E, Turner F, Brunner HG, van Driel MA, Oti M, Lopez-Bigas N, Ouzounis C, Perez-Iratxeta C, Andrade-Navarro MA, Adeyemo A, Patti ME, Semple CA, Hide W (2006) Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res* 34:3067–3081
- Valcourt U, Gouttenoire J, Moustakas A, Herbage D, Mallein-Gerin F (2002) Functions of transforming growth factor-beta family type I receptors and Smad proteins in the hypertrophic maturation and osteoblastic differentiation of chondrocytes. *J Biol Chem* 277:33545–33558
- van der Weyden L, Wei L, Luo J, Yang X, Birk DE, Adams DJ, Bradley A, Chen Q (2006) Functional knockout of the matrilin-3 gene causes premature chondrocyte maturation to hypertrophy and increases bone mineral density and osteoarthritis. *Am J Pathol* 169:515–527
- van Driel MA, Cuelenaere K, Kemmeren PP, Leunissen JA, Brunner HG, Vriend G (2005) GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases. *Nucleic Acids Res* 33:W758–W761
- Vaughan T, Pasco J, Kotowicz M, Nicholson G, Morrison NA (2002) Alleles of the RUNX2/CBFA1 gene are associated with differences in bone mineral density and risk of fracture. *J Bone Miner Res* 17:1527–1534
- Vaughan T, Reid D, Morrison NA, Ralston S (2004) RUNX2 alleles associated with BMD in Scottish women; interaction of RUNX2 alleles with menopausal status and body mass index. *Bone* 34:1029–1036
- Xian CJ (2007) Roles of epidermal growth factor family in the regulation of postnatal somatic growth. *Endocr Rev* 28:284–296
- Xiao P, Shen H, Guo YF, Xiong DH, Liu YZ, Liu YJ, Zhao LJ, Long JR, Guo Y, Recker RR, Deng HW (2006) Genomic regions identified for BMD in a large sample including epistatic interactions and gender-specific effects. *J Bone Miner Res* 21:1536–1544
- Yamada Y, Ando F, Niino N, Shimokata H (2002) Association of a polymorphism of the CC chemokine receptor-2 gene with bone mineral density. *Genomics* 80:8–12