

Meta-analysis of gene-based genome-wide association studies of bone mineral density in Chinese and European subjects

C.-L. Cheung · P.-C. Sham · S.-M. Xiao · C. H. Bow ·
A. W.-C. Kung

Received: 7 April 2011 / Accepted: 5 July 2011 / Published online: 17 September 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract

Summary Gene-based association approach could be regarded as a complementary analysis to the single SNP association analysis. We meta-analyzed the findings from the gene-based association approach using the genome-wide association studies (GWAS) data from Chinese and European subjects, confirmed several well established bone mineral density (BMD) genes, and suggested several novel BMD genes.

Introduction The introduction of GWAS has greatly increased the number of genes that are known to be associated with common diseases. Nonetheless, such a single SNP GWAS has a lower power to detect genes with multiple causal variants. We aimed to assess the association of each gene with BMD variation at the spine and hip using gene-based GWAS approach.

Methods We studied 778 Hong Kong Southern Chinese (HKSC) women and 5,858 Northern Europeans (dCG); age,

sex, and weight were adjusted in the model. The main outcome measure was BMD at the spine and hip.

Results Nine genes showed suggestive p value in HKSC, while 4 and 17 genes showed significant and suggestive p values respectively in dCG. Meta-analysis using weighted Z-transformed test confirmed several known BMD genes and suggested some novel ones at 1q21.3, 9q22, 9q33.2, 20p13, and 20q12. Top BMD genes were significantly associated with connective tissue, skeletal, and muscular system development and function ($p < 0.05$). Gene network inference revealed that a large number of these genes were significantly connected with each other to form a functional gene network, and several signaling pathways were strongly connected with these gene networks.

Conclusion Our gene-based GWAS confirmed several BMD genes and suggested several novel BMD genes. Genetic contribution to BMD variation may operate through multiple genes identified in this study in functional gene networks. This finding may be useful in identifying and prioritizing candidate genes/loci for further study.

Electronic supplementary material The online version of this article (doi:10.1007/s00198-011-1779-7) contains supplementary material, which is available to authorized users.

C.-L. Cheung (✉) · S.-M. Xiao · C. H. Bow · A. W.-C. Kung
Department of Medicine, The University of Hong Kong,
Queen Mary Hospital,
102 Pokfulam Road,
Hong Kong, China
e-mail: lung1212@hku.hk

P.-C. Sham
Department of Psychiatry, The University of Hong Kong,
Hong Kong, China

S.-M. Xiao · A. W.-C. Kung
Research Centre of Heart, Brain, Hormone & Healthy Aging,
The University of Hong Kong,
Hong Kong, China

Keywords Association study · Bone mineral density · Genetic epidemiology · Meta-analysis · Osteoporosis

Introduction

Genome-wide association studies (GWAS) provide a powerful approach to search for common genetic variants that increase susceptibility to complex diseases or traits. Nonetheless, they do not necessarily lead directly to the gene or genes in a given locus associated with disease, nor typically inform the broader context in which the disease genes operate. They thus provide limited insight into the mechanisms that drive disease. In addition, the amount of

genetic variation explained by GWAS for a given disease is most often significantly less than the heritability estimate for the disease. For example, a number of studies estimate the genetic heritability for spine BMD to be as high as 80%, but the 15 genetic loci identified for spine BMD to date account for only ~2.9% of the variation in spine BMD [1]. This raises the question of whether there are many more common DNA variants with smaller effects that are not being identified in the GWAS because of a lack of power, whether there are many more rare variants with stronger effect that explain the missing variation or whether it is some combination of these two scenarios.

In the current GWAS design, a genetic association is considered only at a single SNP locus level rather than a gene level. The stringent genome-wide significance level may also inflate the false-negative rate and limit its ability to identify disease genes. Different approaches have recently been adopted to ameliorate this situation, including pathway-based and gene-based GWAS. Gene-based analysis is a complementary approach to single-locus analysis. Generally, this type of approach tests whether a set of SNPs in a given gene locus is associated with a trait of interest. Different approaches have been used to identify genes that are associated with trait of interest, such as multiple logistic regression for discrete trait and set-based test for discrete or continuous trait. Nonetheless, the set-based test requires heavy computation and therefore limits its application at a genome-wide level. An efficient genome-wide gene-based association method has recently been developed, based on simulations from the multivariate normal distribution. This approach has provided important biological insight into disease etiology, and a number of disease genes are expected to be identified. These genes may not contain any SNPs that meet the genome-wide significance threshold, but rather a nominal significant p value may be observed in a number of SNPs in each of these genes.

In this study, we performed gene-based GWAS in a Hong Kong Southern Chinese (HKSC) cohort and Icelandic deCODE Study (dCG) [2] and performed meta-analysis of 6,636 adults by combining the results from HKSC and dCG that examined spine and femoral neck BMD. Our findings confirmed several well-known candidate genes and discovered a number of novel candidate genes.

Materials and methods

Study population

The current meta-analysis incorporated 6,643 individuals derived from two GWAS on BMD at the lumbar spine and femoral neck, the HKSC Study ($n=778$), and dCG Study (dCG, $n=5,858$) [2]. In the Hong Kong Osteoporosis Study,

800 unrelated women with extreme high or low BMD were selected from a HKSC cohort with extreme BMD. These subjects were selected from a database (>9,000 Southern Han Chinese volunteers) at the Osteoporosis Centre of the University of Hong Kong. Low-BMD subjects are defined as those with a BMD Z-score ≤ -1.28 at either the lumbar spine (LS) or femoral neck (FN) (the lowest 10% of the total cohort). High-BMD subjects comprised individuals with BMD Z-score $\geq +1.0$ at either site. Subjects who reported diseases or environmental factors that may affect BMD and bone metabolism were excluded. The recruitment procedure and exclusion criteria have been detailed elsewhere [3]. The demographic data of studied population are provided in Supplementary Table 1.

BMD and anthropometric measurements

BMD (grams per square centimeter) at the LS and FN was measured by dual-energy X-ray absorptiometry (Hologic QDR 4500 plus, Hologic Waltham, MA, USA) with standard protocol. The in vivo precision of the machine was 1.2% and 1.5% for LS and FN BMD, respectively.

Genotyping

The discovery sample was genotyped via the Infinium assay (Illumina, San Diego, CA) with Human610-quad chip, including 564,214 SNPs. PLINK (version 1.04) was used for data management and quality-control statistics. Seven hundred seventy-eight individuals and 488,853 SNPs were retained for analysis following application of strict quality-control criteria. Subjects were excluded according to the following criteria: (1) genotyping call rate <95% ($n=5$), (2) autosomal heterozygosity <27% or >31% (the same five subjects with low genotyping call rate), (3) being related or identical to other individuals in the sample ($n=7$), and (4) discordance of observed gender and estimated sex ($n=3$). SNPs were excluded if (1) genotyping call rate was $\leq 95\%$ (1,158 SNPs), (2) Hardy–Weinberg equilibrium p value $< 1.0 \times 10^{-4}$ (904 SNPs), (3) minor allele frequency <0.01 (73,589 SNPs). The average genotyping call rate of retained SNPs was 99.91%.

In silico gene-based GWAS of European subjects

All GWA data for spine and hip BMD of dCG were retrieved from the publication [2].

Phenotype modeling

To be comparable and compatible in meta-analysis, both studies used standardized residuals of raw BMDs following adjustment for age, weight, and sex (dCG) in the linear regression model.

Definition of gene locus

We defined the gene locus by its position based on UCSC and ± 50 kb. Fifty kilobytes is the mean distance between the top intergenic SNPs and the nearest genes identified in the latest meta-analysis of GWAS of BMD [1]. *SOX6* and *MEF2C* were excluded from the mean distance calculation as they were the outliers.

Statistical analysis

In the genome-wide association study, the association test of SNPs with standardized residuals of BMD was implemented via PLINK (version 1.04). For each SNP, the asymptotic p value for the relationship between the number of minor alleles and BMD was derived from a two-sided t statistic assuming the minor allele has an additive effect. We identified disease-associated genes with two stages. The first stage was the gene-based test using a simulation approach that took account of LD structure of different populations based on HapMap phase two information. Gene-based analysis in each population was done using VEGAS [4]. In brief, each SNP was assigned to each of 17,787 autosomal genes according to positions on the UCSC Genome Browser hg18 assembly. In order to capture regulatory regions and SNPs in LD, the gene boundaries were defined as 51 kb of 5' and 3' UTRs. Then, for a given gene with n SNPs, association p values were first converted to upper tail chi-squared statistics with 1 degree of freedom (df). The observed gene-based test statistic was then the sum of all (or a pre-defined subset) of the chi-squared 1 df statistics within that gene. In the current study, we summed the statistics of all SNPs located within a gene. Using the Monte Carlo approach, a large number of multivariate normal vectors were simulated, and the empirical gene-based p value was calculated as the proportion of simulated test statistics that exceeded the observed gene-based test statistic. In the second stage, we attempted to meta-analyze the findings from both populations, to increase statistical power and to assess the consistency of evidence in two ethnicities using weighted Z -transformed test as implemented in the R. A weighted Z -transformed test was chosen because it has been suggested that when the number of tests are small, the weighted Z -transformed test performs better than other combination probability methods, such as Fisher's test and generalized binomial test [5, 6].

Gene-based genome-wide significant level and suggestive level

Among 17,640 genes included in the analysis, 14,605 overlapped with either 5' and/or 3' genes with the average overlapping size per gene size (overlapping size with other

gene/gene size) 0.62. We therefore arbitrarily defined the gene-based genome-wide significant level as $0.05/(3,035 \times 1 + 14,605 \times 0.38) = 5.8 \times 10^{-6}$, while the suggestive level was $1/(3,035 \times 1 + 14,605 \times 0.38) = 1.2 \times 10^{-4}$.

Identification of enriched physiological role in genes associated with BMD

The top 35 genes were imported into the Ingenuity Pathways Analysis (IPA) Software (Ingenuity Systems, Redwood City, CA, USA) to obtain networks for further analyses and to determine whether their physiological role was enriched. These top 35 genes were chosen because 35 was the limited number of genes/molecules required to form a functional regulatory gene network in the later gene network inference analysis. The enriched physiological roles were ranked by the p values of the Fisher's Exact Test that indicated the probability of the input gene (from the gene-based GWAS) being associated with genes in the physiological roles by chance.

Gene network inference via knowledge-based data mining

We next analyzed biological interactions among top hits using the IPA tool. The gene annotations from the top hits with suggestive p value were entered into the IPA analysis tool to construct the biological networks of the clustered genes. Networks are generated from the gene set by maximizing the specific connectivity of the input genes, which represents their interconnectedness with each other relative to other molecules to which they are connected in Ingenuity's Knowledge Database. Networks were limited to 35 molecules each to keep them to a functional size. The p value of probability for the genes forming a network was calculated using the right-tailed Fisher's Exact Test based on the hypergeometric distribution.

Results

Genomic control of SNP data before gene-based GWAS

In single SNP GWAS of spine and hip BMD in southern Chinese, an inflation factor of 1 was observed for both sites. An inflation factor of 1.22 and 1.18 for spine and hip BMD was observed in the p value distribution from the dCG GWAS data. After correction of GC, the inflation factor became one in both sites studied.

Identification of BMD genes in each individual study

In southern Chinese, none of the genes reached a genome-wide significant p value (5.8×10^{-6}), whereas seven and two

genes reached a suggestive p value for hip and spine BMD, respectively. The most significant gene for spine BMD was *CCDC55* with an empirical p value of 8.3×10^{-5} (Table 1). The most significant gene for femoral neck BMD was *KPNA4* with an empirical p value of 4.9×10^{-5} (Table 1). The best SNP (rs4470197) in the suggestive genes *EFCAB5* and *CCDC55* for spine BMD was the same. Likewise, the best SNP (rs4680580) in the suggestive genes *SMC4* and *TRIM59* for hip BMD was the same.

In European subjects, three genes (*C6orf97*, *ESPL1*, and *SP7*) were significantly associated with spine BMD (Table 2), and p values of eight genes reached suggestive significance level. Among the three significant genes, rs10876432 was the best SNP in two of them. For femoral neck BMD, two genes (*C6orf97* and *LRP4*) reached a genome-wide significant level (Table 3), and nine genes reached a genome-wide suggestive level. Of the genes significantly associated with femoral neck BMD variation, only *C6orf97* was associated with BMD at both sites in Europeans.

Meta-analysis of gene-based GWAS in southern Chinese and Europeans

In the gene-based GWAS of spine BMD in Europeans, *C6orf97* had an empirical p value of 0. This was because the maximum number of permutations performed was 1,000,000. If there were no simulated test statistics greater than the observed test statistics, the empirical p value would be 0. We replaced the “0” with 1×10^{-6} for the purpose of meta-analysis and applied weighted Z-transformed test to combine the p values of each gene with weighting of sample size of each study. Manhattan plots of the $-\log p$ value of the gene-based GWAS for spine and hip BMD are shown in Supplementary Figures 1a and 1b. The QQ plots of the results are provided in Supplementary Figure 2. The meta-analysis identified five genes from three genomic loci that exceeded the gene-based GWAS significance threshold of association with the BMD traits (Tables 4 and 5). Of these, three loci were associated with LS BMD (Table 4) and two with FN BMD (Table 5).

Known genes associated with BMD in previous GWAS meta-analysis

We have previously identified two genes for spine BMD and two genes for femoral neck BMD through a GWAS meta-analysis approach: *SP7* (meta $p=4.4 \times 10^{-6}$) and *C6orf97* (meta $p=7.7 \times 10^{-7}$) for spine BMD, *CKAP5* (meta $p=5.2 \times 10^{-6}$) and *LRP4* (meta $p=1.2 \times 10^{-6}$) for femoral neck BMD. Although *ESPL1* was not identified in the previous study, it is located at the same locus as *SP7* at 12q13–12q13.1.

Novel loci associated with BMD at gene-based GWA suggestive threshold

Four and five genetic loci reached a genome-wide suggestive level for spine and hip BMD respectively: 6q25.1 (*ESR1*), 9q33.2 (*CDK5RAP2*), 12q13 (*C12orf10*, *AAAS*, *SP1*, *PFDN5*, *MFSD5*, and *RARG*), and 20q12 (*EIF6*) for spine BMD; 1q21.3 (*LCE2A*, *KPRP*, *LCE4A*, *LCE2B*, and *LCE2C*), 6q25.1 (*C6orf97*), 9q22 (*FOXE1*), 11p11 (*F2*, *C11orf49*, *ZNF408*, and *ARHGAP1*), and 20p13 (*ADRA1D*) for hip BMD. Of these, 1q21.3, 9q22, 9q33.2, 20p13, and 20q12 were not identified as significant BMD loci in the previous meta-analysis [1].

Enriched physiological role of the top genes

The results of a physiological role analysis (Tables 6 and 7) suggest that genes for spine BMD are involved mainly in connective tissue development (lowest $p=3.7 \times 10^{-6}$) and function and skeletal and muscular system development and function (lowest $p=3.7 \times 10^{-6}$). Genes for hip BMD are involved mainly in cardiovascular system development and function (lowest $p=4.9 \times 10^{-4}$) and tissue morphology (lowest $p=4.9 \times 10^{-4}$). Connective tissue development and function (lowest $p=1.28 \times 10^{-3}$), digestive system development and function (lowest $p=1.28 \times 10^{-3}$), and embryonic development (lowest $p=1.28 \times 10^{-3}$) are also associated with the hip BMD genes.

Novel gene network inference

Gene network inference was performed to evaluate whether the gene set may represent a novel functional gene network that may be involved in bone metabolism. We generated functional gene networks from the BMD genes using IPA. For spine BMD genes, the most significant gene network connected 18 spine BMD genes with 17 connecting genes with a p value of 1×10^{-46} (Fig. 1a). There were several hub genes/molecules in this network, such as *SP1*, *ESR1*, *P38 MAPK*, and *EPK1/2*. This network was significantly associated with connective tissue development and function, skeletal and muscular system development and function, and cell cycle (Fig. 1a). For femoral neck BMD, the most significant gene network connected ten spine BMD genes with 25 connecting genes with a p value of 1×10^{-23} (Fig. 1b). There were several hub genes/molecules in this network, such as *TNF*, *prostaglandin E2*, *NFkB*, and *F2*. This network was significantly associated with cellular development, cellular growth and proliferation, and connective tissue development and function.

Table 1 Genes associated at gene-based genome-wide suggestive level with spine and femoral BMD in HKSC study

Gene information		Lumbar spine BMD				Femoral neck BMD						
Chr	Gene	Number of SNPs	Start position	End position	Test statistic	Gene-based <i>p</i>	Best SNP	SNP <i>p</i>	Test statistic	Gene-based <i>p</i>	Best SNP	SNP <i>p</i>
3	IFT80	15	161457481	161600014	57.5	0.007	rs6798183	0.004	106.3	9.7E-05	rs4679881	4.7E-05
3	SMC4	11	161600123	161635435	56.0	0.003	rs6798183	0.004	93.6	8.2E-05	rs4680580	4.7E-05
3	TRIM59	9	161635984	161650320	47.9	0.003	rs4680588	0.007	80.5	6.2E-05	rs4680580	4.7E-05
3	KPNA4	9	161700655	161766070	56.5	0.001	rs6797357	0.003	85.3	4.9E-05	rs4680588	1.4E-04
4	TBC1D1	118	37569114	37817189	249.5	0.007	rs17425670	6.7E-05	385.9	1.0E-04	rs6845120	3.5E-06
12	OSBPL8	24	75269708	75477720	117.2	0.001	rs10862167	7.0E-04	155.0	9.2E-05	rs2632208	2.3E-05
16	LOC348174-1	8	68542310	68555390	6.4	0.460	rs1052429	0.290	81.2	1.2E-04	rs1052429	1.4E-04
17	EFCAB5	12	25292811	25459596	109.9	1.1E-04	rs4470197	8.1E-06	59.5	0.005	rs4350617	0.004
17	CCDC55	18	25467959	25537612	171.4	8.3E-05	rs4470197	8.1E-06	75.1	0.013	rs4350617	0.004

Table 2 Genes associated at gene-based genome-wide significant and suggestive level with spine BMD in dCG study (*n*=5,858)

Gene information		Lumbar spine BMD				Femoral neck BMD						
Chr	Gene	Number of SNPs	Start position	End position	Test statistic	Gene-based <i>p</i>	Best SNP	SNP <i>p</i>	Test statistic	Gene-based <i>p</i>	Best SNP	SNP <i>p</i>
Significant gene												
6	C6orf97	41	151856919	151984021	248.9	1.0E-06	rs4870044	4.1E-06	270.1	2.0E-06	rs7752591	2.0E-06
12	ESPL1	13	51948349	51973694	140.0	3.0E-06	rs10876432	1.0E-06	47.2	0.013	rs2016266	0.003
12	SP7	6	52006626	52015804	91.6	5.0E-06	rs10876432	1.0E-06	33.3	0.007	rs2016266	0.003
Suggestive gene												
12	C12orf10	8	51979736	51987232	116.3	8.0E-06	rs10876432	1.0E-06	39.7	0.011	rs2016266	0.003
12	SP1	7	52060245	52096493	64.8	8.0E-06	rs10876432	1.0E-06	26.0	0.011	rs2016266	0.003
12	AAAS	8	51987506	52001679	116.3	9.0E-06	rs10876432	1.0E-06	39.7	0.012	rs2016266	0.003
9	CDK5RAP2	16	122190967	122382258	99.0	9.0E-06	rs3780674	7.1E-06	41.2	0.016	rs3780674	0.001
12	PFDN5	8	51975501	51979501	116.3	1.5E-05	rs10876432	1.0E-06	39.7	0.011	rs2016266	0.003
6	ESR1	61	152053323	152466101	234.0	2.7E-05	rs2504063	5.6E-08	176.9	6.0E-04	rs3020331	8.0E-06
12	MFSD5	11	51932146	51934455	73.1	8.8E-05	rs7314769	8.8E-05	26.6	0.046	rs2272313	0.030
12	RARG	12	51890619	51912303	71.7	1.2E-04	rs7314769	8.8E-05	30.6	0.031	rs2272300	0.014

Table 3 Genes associated at gene-based genome-wide significant and suggestive level with femoral neck BMD in dCG study ($n=5,858$)

Gene information		Lumbar spine BMD				Femoral neck BMD						
Chr	Gene	Number of SNPs	Start position	End position	Test statistic	Gene-based p	Best SNP	SNP p	Test statistic	Gene-based p	Best SNP	SNP p
Significant gene												
6	C6orf97	41	151856919	151984021	248.9	1.0E-06	rs4870044	4.0E-06	270.1	2.0E-06	rs7752591	2.2E-06
11	LRP4	12	46834993	46896652	32.7	0.040	rs7108147	0.004	126.5	4.0E-06	rs1007738	7.1E-06
Suggestive gene												
11	CKAP5	12	46721659	46824419	28.2	0.079	rs7108147	0.004	144.9	1.1E-05	rs1007738	7.1E-06
20	ADRA1D	23	4149277	4177659	50.9	0.025	rs6076639	0.010	108.7	2.9E-05	rs4815683	1.6E-04
11	F2	7	46697318	46717632	8.8	0.282	rs4752926	0.063	80.7	3.4E-05	rs6485690	6.4E-05
1	KPRP	7	150997129	151001153	14.2	0.124	rs1332498	0.063	85.3	3.6E-05	rs1332498	3.3E-05
9	FOXE1	9	99655357	99658818	1.0	0.970	rs6586	0.529	84.7	6.5E-05	rs907580	4.6E-06
1	LCE4A	6	150948146	150948534	13.5	0.119	rs1332498	0.063	79.5	8.9E-05	rs1332498	3.3E-05
1	LCE2A	6	150937463	150938542	12.5	0.129	rs1332498	0.063	70.9	1.0E-04	rs1332498	3.3E-05
10	ERLIN1	6	101899836	101935804	32.3	0.002	rs10883447	2.3E-04	45.8	1.1E-04	rs10883447	1.7E-04
1	LCE2B	8	150925222	150926500	12.9	0.209	rs1332498	0.063	71.0	1.2E-04	rs1332498	3.3E-05

Discussion

GWA is a powerful tool that can identify genes associated with common diseases or traits such as BMD variation. Nonetheless, GWA studies usually focus on the most significant individual variants without considering the global evidence of the gene tested. It should be noted that allelic heterogeneity (i.e., presence of more than one susceptibility allele in a locus or gene) greatly reduces the power for testing of an individual SNP [7, 8]. Therefore, a gene-based test can ameliorate the situation by simply testing the global null hypothesis about the SNPs located per gene. The gene-based test is a direct and powerful means of protecting the overall false-positive rate when a collection of loci are tested, because the p value from the gene-based test has already corrected the number of SNPs included via a simulation approach. Using gene-based analysis of GWA data, our study confirmed several well established candidate genes and suggested several novel genes and loci for BMD variation. Importantly, most of these genes did not contain any SNP that reached genome-wide significance, so the potential importance of these genes would not have been recognized in the absence of gene-based association study.

An ethnic-specific BMD gene may underlie BMD variation in southern Chinese and Europeans. In line with the observations of our recent GWAS, there was no overlap of genes in the significant or suggestive gene list from HKSC and dCG populations. We recently identified a SNP rs2273601 in *JAG1* that was associated with spine BMD (p value = 1.06×10^{-8}) in 1,520 HKSC subjects with extreme BMD; nonetheless, only a modest association of this SNP with spine BMD was observed in three Caucasian cohorts (p range, 0.007–0.045) [3]. In the current study, we observed that top hip BMD genes were more consistent in HKSC and dCG, as reflected by the inflation factor and the results from independent t testing (Supplementary methods, Supplementary Figures 3 to 4, and Supplementary Table 2). The discrepancies of gene-based association results for spine BMD in two populations may be due to a number of factors such as lifestyle, diet, and genetic background. Although these factors may also affect hip BMD, the possibility that spine BMD may be more susceptible than hip BMD to gene and environment interaction cannot be excluded. If this hypothesis is true, identification of gene and environmental interaction will benefit genetic research into osteoporosis and clinical practice.

The study design of HKSC and dCG also differed. In our HKSC cohort, we studied subjects at the extremes of BMD distribution. Studying subjects at the extremes of a quantitative phenotype has proven useful in

Table 4 Top genome-wide significant genes associated with spine BMD in 6,636 adults

Chr	Gene	Start position	End position	Southern Chinese (<i>n</i> =778)			European (<i>n</i> =5,858)			Meta <i>p</i>
				Number of SNPs	Test statistic	Gene-based <i>p</i>	Gene-based <i>p</i>	Number of SNPs	Test statistic	
Significant gene										
6	C6orf97	151856919	151984021	69	46.8	0.734	41	248.9	1.0E-06	1.9E-06
12	ESPL1	51948349	51973694	13	17.2	0.239	13	140.0	3.0E-06	2.3E-06
12	SP7	52006626	52015804	6	6.6	0.309	6	91.6	5.0E-06	4.4E-06
Suggestive gene										
12	C12orf10	51979736	51987232	8	10.4	0.252	8	116.3	8.0E-06	6.4E-06
12	AAAS	51987506	52001679	7	9.5	0.222	8	116.3	9.0E-06	6.7E-06
12	SP1	52060245	52096493	7	5.2	0.414	7	64.8	8.0E-06	8.4E-06
12	PFDN5	51975501	51979501	8	10.4	0.227	8	116.3	1.5E-05	1.1E-05
9	CDK5RAP2	122190967	122382258	35	19.3	0.804	16	99.0	9.0E-06	1.8E-05
6	ESR1	152053323	152466101	132	113.9	0.609	61	234.0	2.7E-05	3.7E-05
12	MFSD5	51932146	51934455	11	14.1	0.271	11	73.1	8.8E-05	7.3E-05
12	RARG	51890619	51912303	12	16.6	0.211	12	71.7	1.2E-04	8.6E-05
20	EIF6	33330138	33336008	14	19.0	0.245	11	66.6	1.6E-04	1.3E-04

identifying functional rare variants [9, 10]. The genes identified in our HKSC cohort may therefore harbor more rare variants than the dCG cohort. Future deep sequencing of the top hits will be required to validate the first hypothesis. The dCG cohort also included both men and women, while our HKSC cohort included only women. Since sex-specific genetic architecture has been

well demonstrated for BMD variation [11–13], this difference likely accounts for some differences in the findings. Although the number of subjects in the HKSC cohort was fewer, the HKSC cohort captured information from the extreme 25% (cases, lowest 10%; super control, highest 15%) of 3,200 subjects. Other heterogeneity in different ethnicities, such as lifestyle, diet,

Table 5 Top genome-wide significant genes associated with femoral neck BMD in 6,636 adults

Chr	Gene	Start position	End position	Southern Chinese (<i>n</i> =778)			European (<i>n</i> =5,858)			Meta <i>p</i>
				Number of SNPs	Test statistic	Gene-based <i>p</i>	Number of SNPs	Test statistic	Gene-based <i>p</i>	
Significant gene										
11	LRP4	46834993	46896652	10	43.6	0.016	12.000	126.5	4.0E-06	1.2E-06
11	CKAP5	46721659	46824419	13	36.9	0.065	12.000	144.9	1.1E-05	5.2E-06
Suggestive gene										
6	C6orf97	151856919	151984021	69	23.9	0.978	41.000	270.1	2.0E-06	8.4E-06
11	F2	46697318	46717632	9	24.8	0.068	7.000	80.7	3.4E-05	1.7E-05
9	FOXE1	99655357	99658818	9	38.0	0.015	9.000	84.7	6.5E-05	2.2E-05
1	LCE2A	150937463	150938542	11	44.4	0.010	6.000	70.9	1.0E-04	3.2E-05
1	KPRP	150997129	151001153	16	18.4	0.329	7.000	85.3	3.6E-05	3.3E-05
1	LCE4A	150948146	150948534	12	37.1	0.023	6.000	79.5	8.9E-05	3.5E-05
20	ADRA1D	4149277	4177659	34	29.8	0.537	23.000	108.7	2.9E-05	3.6E-05
1	LCE2B	150925222	150926500	13	57.9	0.008	8.000	71.0	1.2E-04	3.7E-05
1	LCE2C	150914394	150915673	14	63.8	0.008	8.000	71.0	1.6E-04	5.0E-05
11	C11orf49	46914826	47142507	23	121.2	0.005	20.000	140.1	1.8E-04	5.2E-05
11	ZNF408	46678943	46684037	10	41.8	0.013	9.000	69.9	2.2E-04	7.9E-05
11	ARHGAP1	46655207	46678696	9	37.7	0.012	8.000	57.0	3.1E-04	1.1E-04

Table 6 Bio-function enrichment analysis of spine BMD genes

Physiological role	<i>p</i> value range	Number of molecules
Connective tissue development and function	3.67E–06 to 0.049	4
Skeletal and muscular system development and function	3.67E–06 to 0.046	6
Tissue morphology	6.31E–06 to 0.046	4
Digestive system development and function	1.95E–03 to 0.017	4
Embryonic development	1.95E–03 to 0.029	4

LD structure, might also contribute to the difference in the strength of findings [13].

Interpretation of the gene-based results required extra attention. For example, two spine suggestive genes (*CCDC55* and *EFCAB5*) identified in HKSC harbored the SNP rs4470197 which showed a strong association signal with spine BMD ($p=8.1 \times 10^{-6}$). This SNP was located between these two genes, and the gene-based *p* value was partly contributed by the *p* value of rs4470197. Nonetheless, it is unknown whether rs4470197 is associated with *CCDC55* or *EFCAB5* or both. *CCDC55* (coiled-coil domain containing 55) and *EFCAB5* (EF-hand calcium binding domain 5) are newly annotated genes with no known function; both are conserved in a number of animals such as the chimpanzee, cow, mouse, rat, and chicken. A future functional study is required to validate their role in bone metabolism. The most significant hip BMD gene identified in HKSC was *KPNA4* (karyopherin alpha 4 (importin alpha 3)). The primary function of karyopherins is to recognize nuclear localization signals (NLSs) and dock NLS-containing proteins to the nuclear pore complex. A number of bone genes contain NLS, such as *RUNX2* and *PTHrP*. A recent study [14] demonstrated that NLS of *PTHrP* regulates skeletal development, including bone mass and osteoblast development. Therefore, defective recognition of NLS may affect bone metabolism.

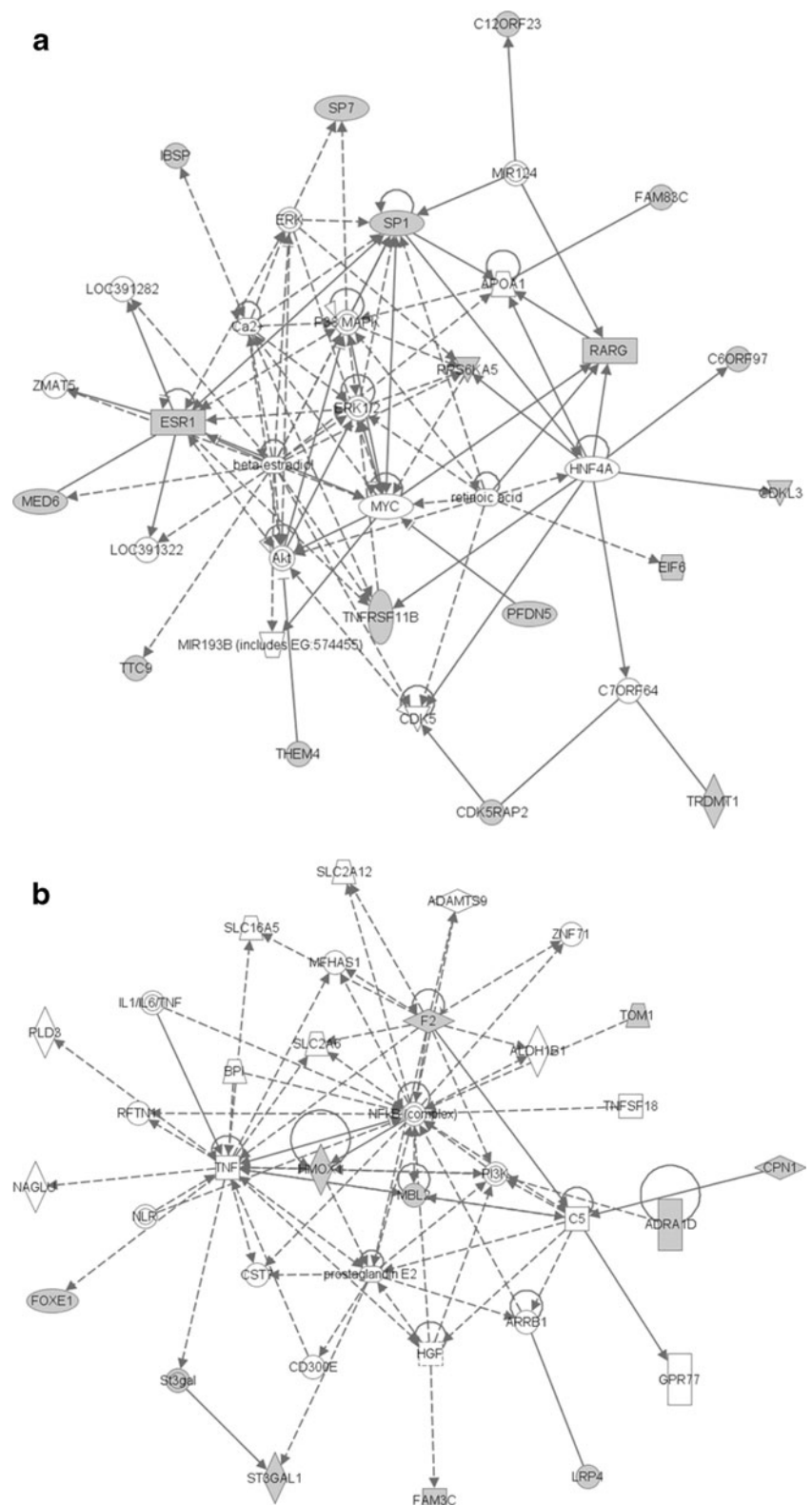
The findings in the dCG cohort were similar to the findings in meta-analysis, despite the fact that *CKAP* became significant and *C6orf97* became insignificant in the meta-analysis for hip BMD. In the meta-analysis, we identified a number of gene loci that have been implicated in bone metabolism in the latest meta-analysis of GWAS in 19,195 subjects [1], such as 6q25 and 12q13 for spine BMD and 11p11.2 for hip BMD. We also identified a

number of novel suggestive loci associated with BMD. 1q21.3 encompasses late cornified envelope protein (LCE) gene cluster and keratinocyte proline-rich protein (*KPRP*) and is known as the epidermal differentiation complex [15]. Both *LCE2A* and *LCE4A* were induced and responsive to the extracellular calcium level and UV irradiation. Though thought to be mainly involved in skin conditions (such as psoriasis [16]), deletion of LCEs was also associated with rheumatoid arthritis [17], thus offering an insight into the role of LCEs in the autoimmune system. The mammalian Forkhead Box (Fox) family of transcription factor is involved in a number of biological processes such as tissue-specific transcription, cell fate determination during embryogenesis, and cell survival. *FOXE1* at 9q22 was identified as a BMD candidate gene in the current study. *FOXE1* is involved in thyroid organogenesis and development of cleft palate [18, 19]. A recent study has shown that this gene is also associated with skeletogenesis in zebrafish. Knocking down of *FOXE1* in zebrafish using morpholino resulted in severe reduction in the expression of *sox9a*, *colla1*, and *runx2*. In addition, this gene and another candidate gene in the same gene family identified in the recent meta-analysis [1], *FOXL1*, are downstream targets of Hedgehog-Gli signaling pathway [20, 21]. The Hedgehog and Gli signaling pathway is important in bone development [22] and osteoblast differentiation [23]. *CDK5RAP2* (CDK5 regulatory subunit associated protein 2) at 9q33.2 is involved in the regulation of neuronal differentiation and associated with microcephaly [24]. Microcephaly is a disease in which head size is smaller than average and is often associated with osteoporosis [25, 26]. Adrenergic, alpha-1D-receptor (*ADRA1D*) at 20p13 is a G-protein coupled receptor that mediates actions in the sympathetic nervous system through a

Table 7 Bio-function enrichment analysis of hip BMD genes

Physiological role	<i>p</i> value range	Number of molecules
Cardiovascular system development and function	4.93E–04 to 0.050	4
Tissue morphology	4.93E–04 to 0.043	6
Connective tissue development and function	1.28E–03 to 0.034	3
Digestive system development and function	1.28E–03 to 0.017	3
Embryonic development	1.28E–03 to 0.036	2

Fig. 1 a Inference of gene regulatory networks using spine BMD genes. **b** Inference of gene regulatory networks using hip BMD genes



number of neurotransmitters, such as catecholamines, epinephrine, and norepinephrine. The sympathetic nervous system is important in bone mass regulation [27, 28]; male mice without beta1/beta2 adrenergic receptor

have increased cortical bone mass [29]. The role of *ADRA1D* in bone metabolism has been demonstrated in MC3T3E-1 osteoblast-like cells, in which *ADRA1D* is expressed in MC3T3E-1 cells, and *RANKL* expression is

regulated via alpha-adrenergic receptor stimulation in osteoblasts [30]. Eukaryotic translation initiation factor 6 (eIF6) at 20q12 is a gene that controls translation at the rate-limiting step of initiation. Recently, Gandin et al. demonstrated that heterozygous mice of eIF6 had fewer hepatic and adipose cells due to impaired G1/S cell cycle progression [31]. They found that the reduction of adipose tissue was due to a decreased proliferation of pre-adipocytes derived from mesenchymal stem cells. Although bone phenotype was not investigated in their study, we believe that eIF6 could affect bone metabolism by regulating the cell number of osteoblasts, since both adipocytes and osteoblasts are derived from the same progenitor–mesenchymal stem cell; eIF6 also regulates Wnt/beta-catenin signaling via regulation of beta-catenin synthesis [32].

Collectively, our data showed that the BMD genes identified in our meta-analysis play an important role in bone metabolism. Although additional studies will be necessary to validate their function, our current findings indicate that these BMD genes are involved in connective tissue development and function and skeletal and muscular system development and function using bio-function analysis implemented in IPA ($p < 0.05$) (Tables 6 and 7). Due to the limitation of genechip, gene-based GWAS may not be able to identify BMD genes if the SNP coverage of the BMD genes is low. One method to remedy this is to perform more genotyping with denser SNP; another method is to perform gene network inference to identify genes that are connected with other BMD genes. Using the gene network inference approach, several bone-related hub genes or complexes have been identified, such as ERK1/2 [33, 34], P38 MAPK [35, 36] (Fig. 1a), prostaglandin E2 [37], and TNF [38] (Fig. 1b). Overlaying the gene network with known canonical signaling pathways revealed that aryl hydrocarbon receptor signaling; role of osteoblasts, osteoclasts, and chondrocytes in rheumatoid arthritis; and role of macrophages, fibroblasts, and endothelial cells in rheumatoid arthritis (7 genes out of 35 genes in each signaling pathway) were the predominant themes of the spine BMD gene network (Supplementary Table 3a), whereas acute phase response signaling (8 genes out of 35 genes) was the predominant theme of the hip BMD gene network (Supplementary Table 3b). Interestingly, acute phase response was one of the underlying mechanisms of action of bisphosphonate in the treatment of osteoporosis [39]. Our findings suggest that hip BMD genes *F2*, *MBL2*, and *HMOX1* may be the genes involved in bisphosphonate treatment and may be used to monitor treatment response.

There are a number of limitations in the current gene-based GWAS. First, our definition of gene-based

GWAS significance level may not be accurate. The most accurate way would be to use simulation; however, this would require extremely heavy computations, as the number of SNPs included in each study and the number of independent genes will vary from study to study. The LD structure also varies in different ethnicities. Nonetheless, our gene-based GWAS significant level 5.8×10^{-6} was not much different to the conservative Bonferroni-corrected GWAS significance level of 2.8×10^{-6} ($=0.05/17,640$, assuming each gene is independent to each other). Second, our definition of the gene locus (± 50 kb 5' upstream and 3' downstream of the coding region) might strongly affect the test statistics and hence the gene-based p value. Noting that large boundaries lead to a longer overlapping region with the neighbor genes, hence some markers are included in multiple genes. Thus, we justified how long the boundaries should be included by averaging the distance between the top intergenic SNPs identified in the recent meta-analysis of GWAS to the nearest coding genes [1]. Notably, the highly significant SNP may also inflate the test statistics in a number of nearby genes, which poses interpretation difficulty. Thirdly, although a gene-based approach can identify genes with multiple causal SNPs with small effect size, it cannot identify genes with only one very significant SNP, while other SNPs in the gene do not show any significant p value. Fourthly, we did not account for the heterogeneity in the meta-analysis although there is obvious heterogeneity between HKSC and dCG studies (as reflected by the Q statistics, data not shown). Moreover, since the sample size of the dCG cohort was much larger than the HKSC cohort, many significant p values of the top findings were driven primarily by the dCG study. Caution should therefore be exercised in interpreting meta-analysis findings, especially when our current data suggested that there was a large genetic heterogeneity for spine BMD present between Chinese and European. Lastly, correction for stratification or any inflation has not been established in gene-based GWAS study; therefore, all QC should be done in the single-locus GWAS before performing the gene-based GWAS.

In conclusion, our results demonstrate the potential applicability of a gene-based approach to the interpretation and further mining of GWAS data. The importance of a gene-based approach is that single-locus GWAS mainly focuses on the association between a single marker and disease trait. It may not be able to identify a disease gene that harbors several causal variants with small effect size (allelic heterogeneity). Testing the overall effect of all SNPs in a gene, thus leveraging this information, may provide significant power to identify disease genes. In this study, we identified and/or confirmed a number of BMD genes. These BMD genes were

significantly enriched in connective tissue development and function and skeletal and muscular system development and function. Using a gene network inference approach, we observed that a large number of BMD genes were connected with each other and contributed to a significant physiological function related to bone metabolism. Our approach suggests a concept of how variation in multiple genes linked in a functional gene network contributes to BMD variation and provides a useful tool to reveal the hidden information of GWAS that would be missed in single SNP analysis.

Acknowledgments This work was supported by the Research Grant Council of the Hong Kong Government, The Osteoporosis Research Fund, and Matching Grant of the University of Hong Kong

Conflicts of interest None.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB, Zillikens MC, Kavvoura FK, Amin N, Aulchenko YS et al (2009) Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet* 41(11):1199–1206
- Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, Gudbjartsson DF, Walters GB, Ingvarsson T, Jonsdottir T, Saemundsdottir J, Center JR, Nguyen TV et al (2008) Multiple genetic loci for bone mineral density and fractures. *N Engl J Med* 358(22):2355–2365
- Kung AW, Xiao SM, Cherny S, Li GH, Gao Y, Tso G, Lau KS, Luk KD, Liu JM, Cui B et al (2010) Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am J Hum Genet* 86(2):229–239
- Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG et al (2010) A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 87(1):139–145
- De Meeus T, Guegan JF, Teriokhin AT (2009) MultiTest V. 1.2, a program to binomially combine independent tests and performance comparison with other related methods on proportional data. *BMC Bioinforma* 10:443
- Whitlock MC (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* 18(5):1368–1373
- Slager SL, Huang J, Vieland VJ (2000) Effect of allelic heterogeneity on the power of the transmission disequilibrium test. *Genet Epidemiol* 18(2):143–156
- Cheung CL, Sham PC, Chan V, Paterson AD, Luk KD, Kung AW (2008) Identification of LTBP2 on chromosome 14q as a novel candidate gene for bone mineral density variation and fracture risk association. *J Clin Endocrinol Metab* 93(11):4448–4455
- Plomin R, Haworth CM, Davis OS (2009) Common disorders are quantitative traits. *Nat Rev Genet* 10(12):872–878
- Nejentsev S, Walker N, Riches D, Egholm M, Todd JA (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324(5925):387–389
- Ioannidis JP, Ng MY, Sham PC, Zintzaras E, Lewis CM, Deng HW, Econs MJ, Karasik D, Devoto M, Kammerer CM et al (2007) Meta-analysis of genome-wide scans provides evidence for sex- and site-specific regulation of bone mass. *J Bone Miner Res* 22(2):173–183
- Huang QY, Ng MY, Cheung CL, Chan V, Sham PC, Kung AW (2006) Identification of two sex-specific quantitative trait loci in chromosome 11q for hip bone mineral density in Chinese. *Hum Hered* 61(4):237–243
- Cheung CL, Xiao SM, Kung AW (2010) Genetic epidemiology of age-related osteoporosis and its clinical applications. *Nat Rev Rheumatol* 6:507–17
- Toribio RE, Brown HA, Novince CM, Marlow B, Hemon K, Lanigan LG, Hildreth BE 3rd, Werbeck JL, Shu ST, Lorch G et al (2010) The midregion, nuclear localization sequence, and C terminus of PTHrP regulate skeletal development, hematopoiesis, and survival in mice. *FASEB J* 24(6):1947–1957
- Jackson B, Tilli CM, Hardman MJ, Avilion AA, MacLeod MC, Ashcroft GS, Byrne C (2005) Late cornified envelope family in differentiating epithelia—response to calcium and ultraviolet irradiation. *J Invest Dermatol* 124(5):1062–1070
- de Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, Giardina E, Stuart PE, Nair R, Helms C et al (2009) Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* 41(2):211–215
- Docampo E, Rabionet R, Riveira-Munoz E, Escaramis G, Julia A, Marsal S, Martin JE, Gonzalez-Gay MA, Balsa A, Raya E et al (2010) Deletion of the late cornified envelope genes, LCE3C and LCE3B, is associated with rheumatoid arthritis. *Arthritis Rheum* 62(5):1246–1251
- Moreno LM, Mansilla MA, Bullard SA, Cooper ME, Busch TD, Machida J, Johnson MK, Brauer D, Krahn K, Daack-Hirsch S et al (2009) FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum Mol Genet* 18(24):4879–4896
- De Felice M, Ovitt C, Biffali E, Rodriguez-Mallon A, Arra C, Anastassiadis K, Macchia PE, Mattei MG, Mariano A, Scholer H et al (1998) A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat Genet* 19(4):395–398
- Brancaccio A, Minichiello A, Grachtchouk M, Antonini D, Sheng H, Parlato R, Dathan N, Dlugosz AA, Missero C (2004) Requirement of the forkhead gene Foxe1, a target of sonic hedgehog signaling, in hair follicle morphogenesis. *Hum Mol Genet* 13(21):2595–2606
- Madison BB, McKenna LB, Dolson D, Epstein DJ, Kaestner KH (2009) FoxF1 and FoxL1 link hedgehog signaling and the control of epithelial proliferation in the developing stomach and intestine. *J Biol Chem* 284(9):5936–5944
- Kimura H, Ng JM, Curran T (2008) Transient inhibition of the Hedgehog pathway in young mice causes permanent defects in bone structure. *Cancer Cell* 13(3):249–260
- Kesper DA, Didt-Kozziel L, Vortkamp A (2010) Gli2 activator function in preosteoblasts is sufficient to mediate Ihh-dependent osteoblast differentiation, whereas the repressor function of Gli2 is dispensable for endochondral ossification. *Dev Dyn* 239(6):1818–1826
- Bond J, Roberts E, Springell K, Lizarraga SB, Scott S, Higgins J, Hampshire DJ, Morrison EE, Leal GF, Silva EO et al (2005) A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat Genet* 37(4):353–355
- de Wit MC, de Coo IF, Julier C, Delepine M, Lequin MH, van de Laar I, Sibbles BJ, Bruining GJ, Mancini GM (2006) Microcephaly and simplified gyral pattern of the brain associated with early

- onset insulin-dependent diabetes mellitus. *Neurogenetics* 7 (4):259–263
26. Zackai EH, Sly WS, McAlister WG (1972) Microcephaly, mild mental retardation, short stature, and skeletal anomalies in siblings. *Am J Dis Child* 124(1):111–115
 27. Burt-Pichat B, Lafage-Proust MH, Duboeuf F, Laroche N, Itzstein C, Vico L, Delmas PD, Chenu C (2005) Dramatic decrease of innervation density in bone after ovariectomy. *Endocrinology* 146(1):503–510
 28. Takeda S, Karsenty G (2008) Molecular bases of the sympathetic regulation of bone mass. *Bone* 42(5):837–840
 29. Bouxsein ML, Devlin MJ, Glatt V, Dhillon H, Pierroz DD, Ferrari SL (2009) Mice lacking beta-adrenergic receptors have increased bone mass but are not protected from deleterious skeletal effects of ovariectomy. *Endocrinology* 150(1):144–152
 30. Nishiura T, Abe K (2007) Alpha1-adrenergic receptor stimulation induces the expression of receptor activator of nuclear factor kappaB ligand gene via protein kinase C and extracellular signal-regulated kinase pathways in MC3T3-E1 osteoblast-like cells. *Arch Oral Biol* 52(8):778–785
 31. Gandin V, Miluzio A, Barbieri AM, Beugnet A, Kiyokawa H, Marchisio PC, Biffo S (2008) Eukaryotic initiation factor 6 is rate-limiting in translation, growth and transformation. *Nature* 455 (7213):684–688
 32. Ji Y, Shah S, Soanes K, Islam MN, Hoxter B, Biffo S, Heslip T, Byers S (2008) Eukaryotic initiation factor 6 selectively regulates Wnt signaling and beta-catenin protein synthesis. *Oncogene* 27 (6):755–762
 33. Li JL, Cui B, Qi L, Li XY, Deng LF, Ning G, Liu JM (2008) NMDA enhances stretching-induced differentiation of osteoblasts through the ERK1/2 signaling pathway. *Bone* 43(3):469–475
 34. Binder NB, Niederreiter B, Hoffmann O, Stange R, Pap T, Stulnig TM, Mack M, Erben RG, Smolen JS, Redlich K (2009) Estrogen-dependent and C–C chemokine receptor-2-dependent pathways determine osteoclast behavior in osteoporosis. *Nat Med* 15(4):417–424
 35. Chen JR, Lazarenko OP, Wu X, Kang J, Blackburn ML, Shankar K, Badger TM, Ronis MJ (2010) Dietary-induced serum phenolic acids promote bone growth via p38 MAPK/beta-catenin canonical wnt signaling. *J Bone Miner Res* 25:2399–411
 36. Whitehouse CA, Waters S, Marchbank K, Horner A, McGowan NW, Jovanovic JV, Xavier GM, Kashima TG, Cobourne MT, Richards GO et al (2010) Neighbor of Brca1 gene (Nbr1) functions as a negative regulator of postnatal osteoblastic bone formation and p38 MAPK activity. *Proc Natl Acad Sci USA* 107 (29):12913–12918
 37. Naik AA, Xie C, Zuscik MJ, Kingsley P, Schwarz EM, Awad H, Guldberg R, Drissi H, Puzas JE, Boyce B et al (2009) Reduced COX-2 expression in aged mice is associated with impaired fracture healing. *J Bone Miner Res* 24(2):251–264
 38. Bar-Shavit Z (2008) Taking a toll on the bones: regulation of bone metabolism by innate immune regulators. *Autoimmunity* 41 (3):195–203
 39. Coxon FP, Thompson K, Rogers MJ (2006) Recent advances in understanding the mechanism of action of bisphosphonates. *Curr Opin Pharmacol* 6(3):307–312