

Article

Multi-omics Data Integration for Identifying Osteoporosis Biomarkers and Their Biological Interaction and Causal Mechanisms

Chuan Qiu,
Fangtang Yu,
Kuanjui Su, ...,
Yuping Wang,
Hongwen Deng,
Hui Shen

hshen3@tulane.edu

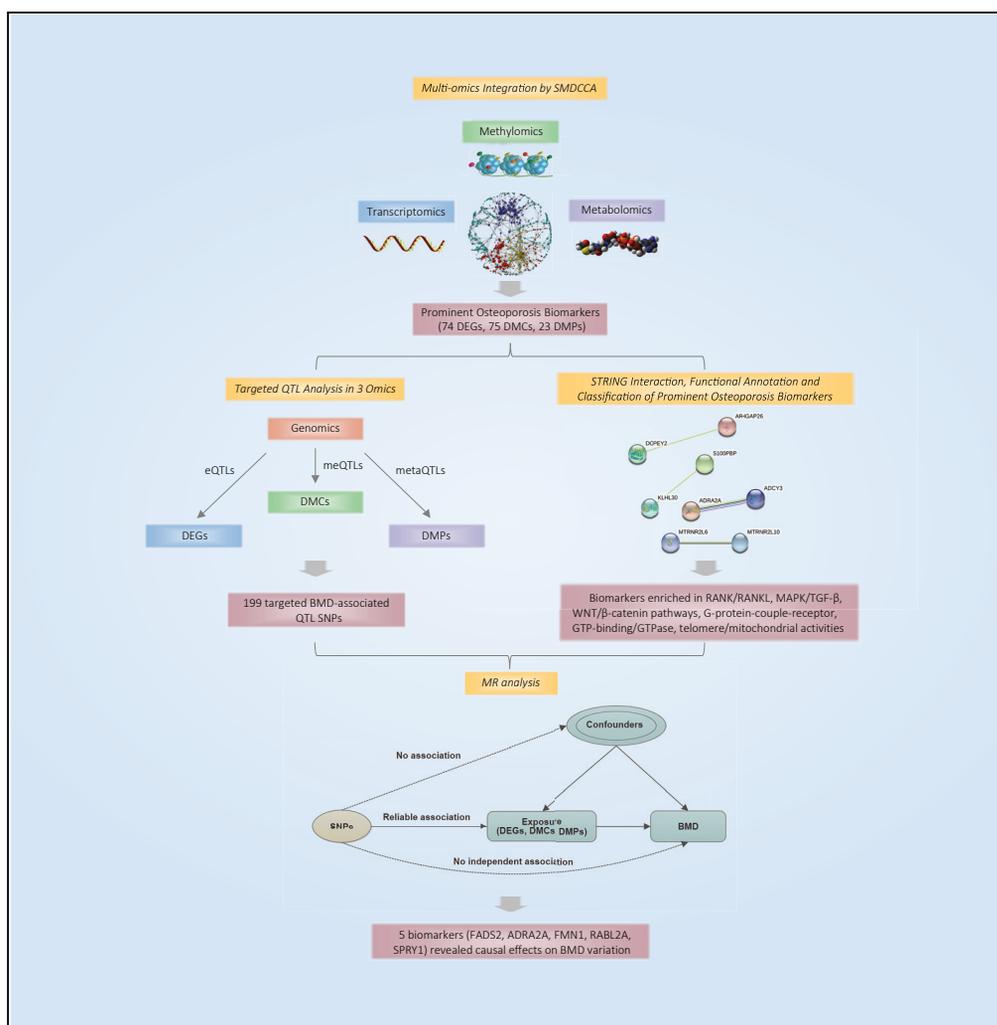
HIGHLIGHTS

Multi-omics integration revealed 172 osteoporosis biomarkers with complex interaction

Genetic variants have multi-level effects on osteoporosis biomarkers

A substantial proportion of biomarkers enriched in bone-related pathways/activities

Several osteoporosis biomarkers have causal effects on the BMD variation



Qiu et al., iScience 23, 100847
February 21, 2020 © 2020 The Author(s).
<https://doi.org/10.1016/j.isci.2020.100847>



Article

Multi-omics Data Integration for Identifying Osteoporosis Biomarkers and Their Biological Interaction and Causal Mechanisms

Chuan Qiu,¹ Fangtang Yu,¹ Kuanjui Su,¹ Qi Zhao,² Lan Zhang,¹ Chao Xu,³ Wenxing Hu,⁴ Zun Wang,^{1,5} Lanjuan Zhao,¹ Qing Tian,¹ Yuping Wang,⁴ Hongwen Deng,^{1,6} and Hui Shen^{1,7,*}

SUMMARY

Osteoporosis is characterized by low bone mineral density (BMD). The advancement of high-throughput technologies and integrative approaches provided an opportunity for deciphering the mechanisms underlying osteoporosis. Here, we generated genomic, transcriptomic, methylomic, and metabolomic datasets from 119 subjects with high (n = 61) and low (n = 58) BMDs. By adopting sparse multiple discriminative canonical correlation analysis, we identified an optimal multi-omics biomarker panel with 74 differentially expressed genes (DEGs), 75 differentially methylated CpG sites (DMCs), and 23 differential metabolic products (DMPs). By linking genetic data, we identified 199 targeted BMD-associated expression/methylation/metabolite quantitative trait loci (eQTLs/meQTLs/metaQTLs). The reconstructed networks/pathways showed extensive biomarker interactions, and a substantial proportion of these biomarkers were enriched in RANK/RANKL, MAPK/TGF- β , and WNT/ β -catenin pathways and G-protein-coupled receptor, GTP-binding/GTPase, telomere/mitochondrial activities that are essential for bone metabolism. Five biomarkers (FADS2, ADRA2A, FMN1, RABL2A, SPRY1) revealed causal effects on BMD variation. Our study provided an innovative framework and insights into the pathogenesis of osteoporosis.

INTRODUCTION

Osteoporosis is a chronic progressive disorder characterized by low bone mineral density (BMD) and deterioration of bone microarchitecture, resulting in increased bone fragility and susceptibility to fracture (Kanis, 2002). The prevalence of osteoporosis in the aging population is over 20% in the United States, and it is becoming an increasingly serious public health problem in the elderly (Wright et al., 2014). Previous genetic studies have indicated that BMD is under strong genetic influence, with estimates of heritability ranging from 0.50 to 0.85 (Ralston and de Crombrughe, 2006; Ralston and Uitterlinden, 2010). In the past decade, researchers have interrogated a wide variety of biological components (e.g., genetic variation, gene expression, and DNA methylation) and uncovered a number of risk biomarkers for BMD. Nonetheless, most of the prior studies have been focused on identification of biomarkers in a single molecular level through univariate statistical methods (e.g., t test, ANOVA, or linear model) and rarely integrated evidences from multiple omics levels. Consequently, so far, the specific functional roles of these identified molecular biomarkers are largely unknown, and their *in vivo* biological interaction and causal mechanisms are not explored.

The advancement of high-throughput technologies, such as whole genome sequencing (WGS), RNA-sequencing (RNA-seq), reduced-representation bisulfite sequencing (RRBS), and liquid chromatography-mass spectrometry (LC-MS), has dramatically increased our ability to comprehensively interrogate diverse molecular features at different omics levels. Meanwhile, several statistical integrative approaches have recently been developed for combining the molecular biomarkers identified from separate analyses of each omics (Liu et al., 2013; Gunther et al., 2012; Rohart et al., 2017), which lead to the discovery of crucial biological insights in a holistic manner and substantially enhance our understanding of molecular networks/pathways underlying the development of human complex diseases. In particular, Le Cao and her colleagues (Singh et al., 2019; Rohart et al., 2017) recently proposed a sparse multiple discriminative canonical correlation method that enables feature selection in multi-omics datasets and answers cutting-edge integrative questions in system biology. This method provided several attractive properties: (1) it allows relax assumptions about data distribution and thus is highly flexible to answer topical questions across various biology-related fields; (2) it is computationally efficient to handle large

¹Center for Bioinformatics and Genomics, Department of Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans 70112, LA, USA

²Department of Preventive Medicine, College of Medicine, University of Tennessee Health Sciences Center, Memphis 38163, TN, USA

³Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City 73104, OK, USA

⁴Department of Biomedical Engineering, Tulane University, New Orleans 70118, LA, USA

⁵Xiangya Nursing School, Central South University, Changsha 410013, China

⁶School of Basic Medical Science, Central South University, Changsha 410013, China

⁷Lead Contact

*Correspondence: hshen3@tulane.edu

<https://doi.org/10.1016/j.isci.2020.100847>



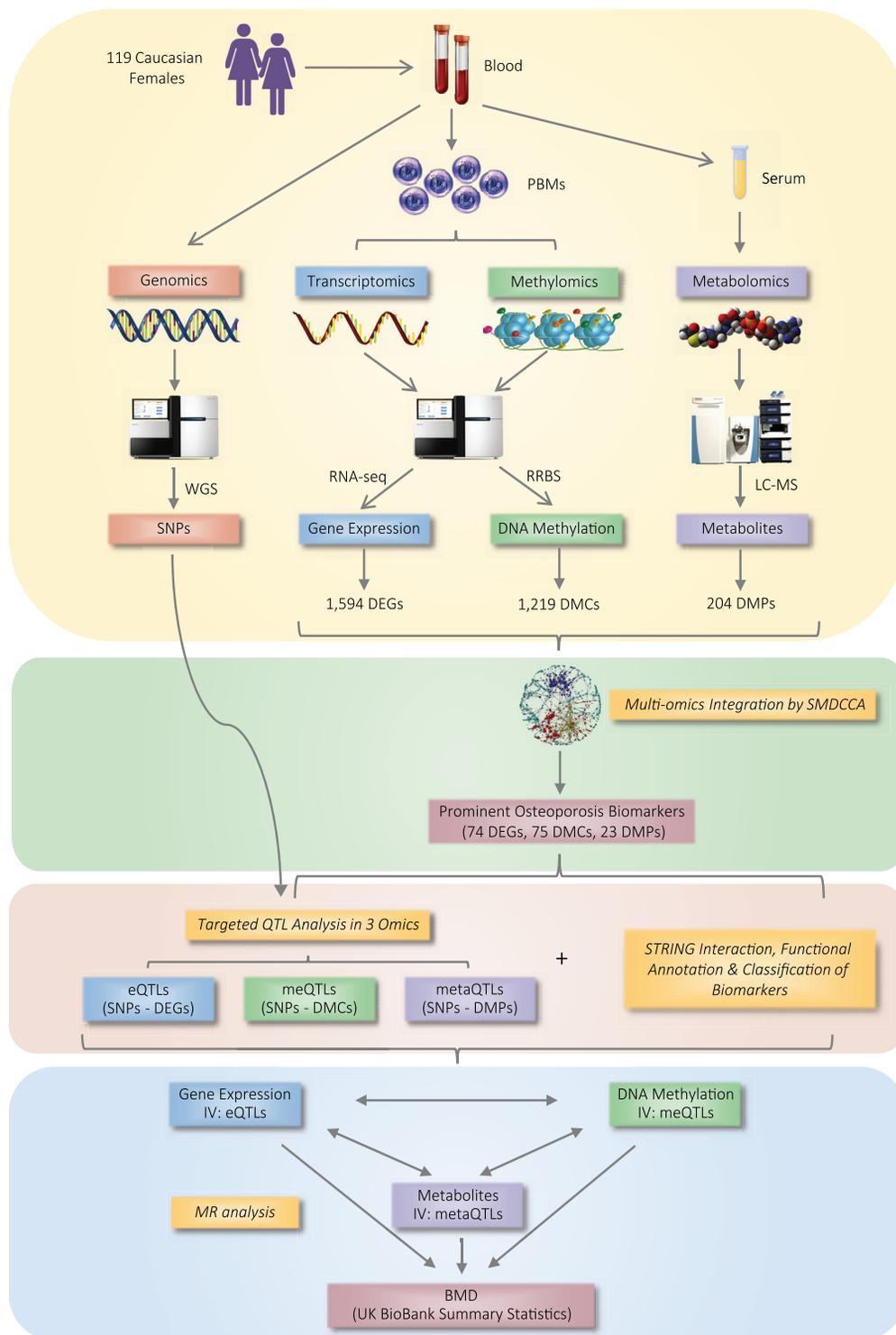


Figure 1. The Overall Workflow for Identifying Osteoporosis Biomarkers and Their Biological Interaction and Causal Mechanisms

The workflow consisted of four phases. Phase 1: individual transcriptomic, methylomic, and metabolomic analyses. Phase 2: SMDCCA integrative analysis of DEGs, DMCs, and DMPs. Phase 3: targeted QTL analyses followed by interaction network analyses, as well as functional annotation and classification analyses. Phase 4: MR analyses. PBMs, peripheral

Figure 1. Continued

blood monocytes; DEGs, differentially expressed genes; DMCs, differentially methylated CpG sites; DMPs, differential metabolic products; SMDCCA, sparse multiple discriminative canonical correlation analysis; IV, instrumental variable; MR, Mendelian randomization.

datasets, where the number of biological markers is much larger than the number of samples; (3) it implements dimension reduction by projecting the data into a smaller subspace while capturing the correlation structure and highlighting the largest sources of variation from the data, resulting in a powerful explanation of the biological system under study.

Genetic variation has a substantial impact on multiple genomic contexts and/or molecular/cellular phenotypes in humans (Pierce et al., 2018; Albert and Kruglyak, 2015; Lemire et al., 2015; McVicker et al., 2013; Kraus et al., 2015), such as transcript abundance (Pierce et al., 2018; Albert and Kruglyak, 2015), DNA methylation (Pierce et al., 2018; Lemire et al., 2015), histone modification (McVicker et al., 2013), and metabolites (Kraus et al., 2015). Whole-genome association scans to detect regions that harbor such variants for gene expression (known as expression quantitative trait locus, eQTL), DNA methylation (meQTL), and metabolite product (metaQTL) have been conducted in multiple human cell/tissue types (GTEx Consortium et al., 2017). Interestingly, previous studies have shown that many QTLs may appear to influence multiple molecular phenotypes. For instance, single nucleotide polymorphisms (SNPs) associated with expression of nearby genes were often associated with methylation of nearby CpG sites (Pierce et al., 2018). These common QTLs in multiple phenotypes may suggest a potential shared biological mechanism by which the common causal variant influences both gene expression and DNA methylation en route to eventually influencing phenotypes. Nonetheless, the precise mechanisms underlying these genetic associations remain poorly understood because of short of the study approaches. Thus, analytical approaches for dissecting the complex biological processes are needed to prioritize the plausible functional variants for further studies. Notably, Mendelian randomization (MR) analysis has been widely used to assess potential causal relationships of genetic/environmental risk factors and diseases (Davey Smith and Hemani, 2014). Recently, MR analysis has been adopted to inspect the causality of biomarkers in disease etiology, utilizing multiple independent SNPs identified by QTL analysis (QTL SNPs) as instrumental variables (Taylor et al., 2019; Yao et al., 2018; Chen et al., 2018). As an example, by applying eQTLs as genetic instruments, Chen et al. recently revealed a causal relationship between LINC00339 gene expression and BMD variation (Chen et al., 2018).

In this work, we performed multi-omics integrative analyses with the largest datasets so far in the bone field to identify osteoporosis biomarkers as well as their biological interaction and causal mechanisms. A simple overview of our workflow is illustrated in Figure 1. Briefly, our approach consisted of four phases. First, we performed individual transcriptomic, methylomic, and metabolomic analyses in 119 Caucasian female subjects with high ($n = 61$) and low ($n = 58$) BMDs to identify potential differentially expressed genes (DEGs), differentially methylated CpG sites (DMCs), and differential metabolic products (DMPs) for osteoporosis risk. The basic characteristics of the study subjects were summarized in Table S1. Second, we integrated the identified DEGs, DMCs, and DMPs via a sparse multiple discriminative canonical correlation analysis (SMDCCA) to retrieve prominent osteoporosis biomarkers that not only reliably distinguish the high-BMD and low-BMD groups, but also are highly correlated spanning different biological layers. Third, we used targeted QTL analyses to test the effects of SNPs on prominent osteoporosis biomarkers in each omics, followed by interaction network analyses, as well as functional annotation and classification analyses, to assess the biological importance of the identified biomarkers. At last, by applying MR analyses using the multiple independent QTL SNPs as instrumental variables, we assessed the causality of the functionally classified biomarkers in BMD variation and inspected whether the identified biomarkers are causally related to one another (e.g., by functional regulation) with the purpose of gaining insights into the *in vivo* molecular functional mechanisms of the etiology of osteoporosis. In aggregate, we identified several osteoporosis biomarkers and reconstructed multi-omics networks/pathways that may mediate variation in risk of osteoporosis *in vivo* in humans. Our study pioneered an innovative integrative approach, and the results illuminated the advantages of multi-omics integrative analysis and provided valuable insights into the pathogenic mechanisms of osteoporosis.

RESULTS**Multi-omics Integration with Supervised SMDCCA**

A total of 25,342 genes, 17,462,566 CpG sites, and 4,209 metabolites were measured in 61 subjects with high BMD and 58 subjects with low BMD, of which 18,774 genes, 763,265 CpG sites, and 2,608 known

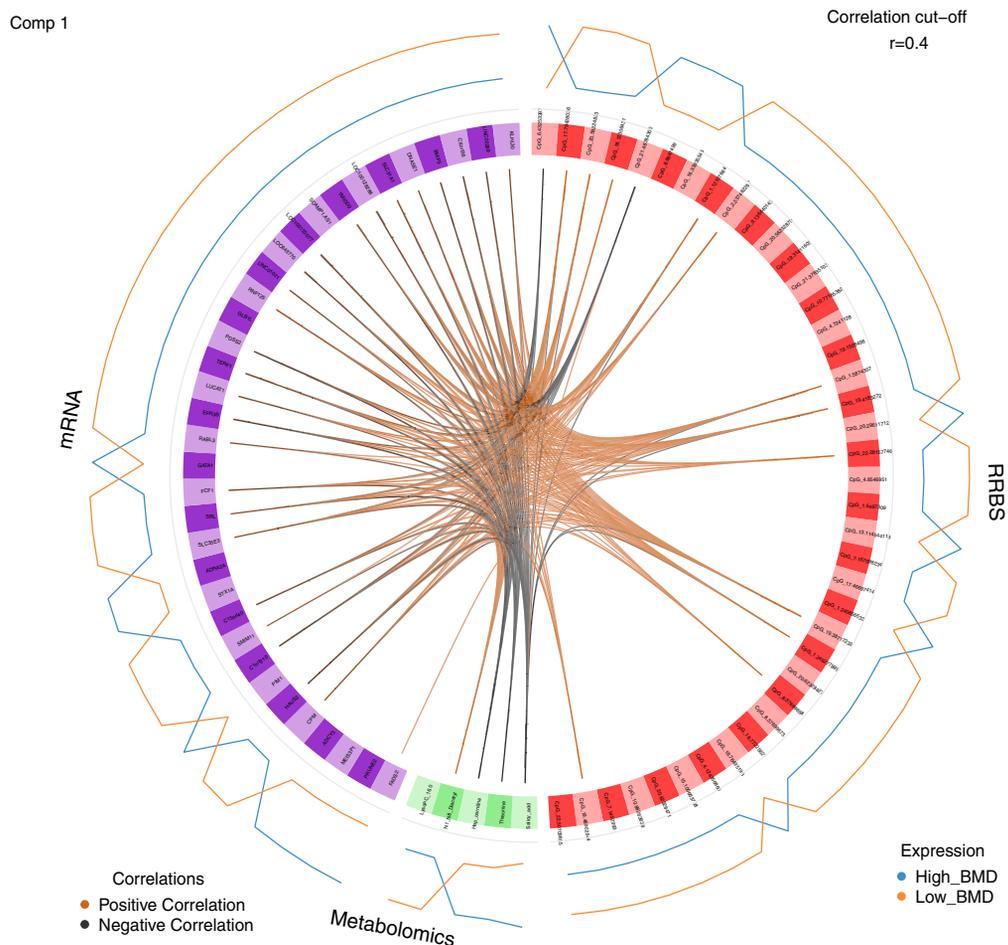


Figure 2. Circos Plot for Prominent Osteoporosis Biomarkers Selected in the Optimal Multi-omics Biomarker Panel

Circos plot displays the different types of osteoporosis biomarkers in the first component on a circle, with links between each omics indicating the positive (brown) or negative (black) correlations with cutoff $r = 0.4$. The blue line and orange line represent the high-BMD and low-BMD groups, respectively.

metabolites were selected (Methods) for subsequent analyses. We identified a total of 1,594 DEGs with false discovery rate (FDR) < 0.05 , 1,219 DMCs ($q < 0.05$ with methylation difference large than 10%), and 204 DMPs ($p < 0.05$) by comparing the high-BMD and low-BMD groups in prevailing single-omics analyses.

By multi-omics integrative analysis on DEGs, DMCs, and DMPs via SMDCCA approach, we identified an optimal multi-omics biomarker panel for discriminating high-BMD and low-BMD groups (classification error rate of 0.1, Figure S1). This optimal multi-omics biomarker panel was composed of three components (Table S2) involving a total of 74 DEGs, 75 DMCs, and 23 DMPs (henceforward termed *prominent osteoporosis biomarkers*). We observed moderate correlations between DEGs and DMCs/DMPs and a few weak correlations between DMCs and DMPs (Figures 2 and S2). Notably, a substantial proportion of DEGs were found to be correlated with both DMCs and DMPs. For instance, gene expression of HAUS2 has positive correlation with DNA methylation at CpG sites Chr1:5874307 (NPHP4) but is negatively correlated with metabolite threonine. Interestingly, there is a significant interaction between HAUS2 and NPHP4 with a combined interaction score of 0.905 (Szklarczyk et al., 2019). However, the specific functional roles of threonine and mechanisms underlining the connection between these correlations are unclear.

Next, we examined the regulatory status for regions containing the 75 DMC prominent osteoporosis biomarkers. Our results showed that 51 (68.0%) DMCs were mapped to the DNaseI hypersensitivity cluster; 28

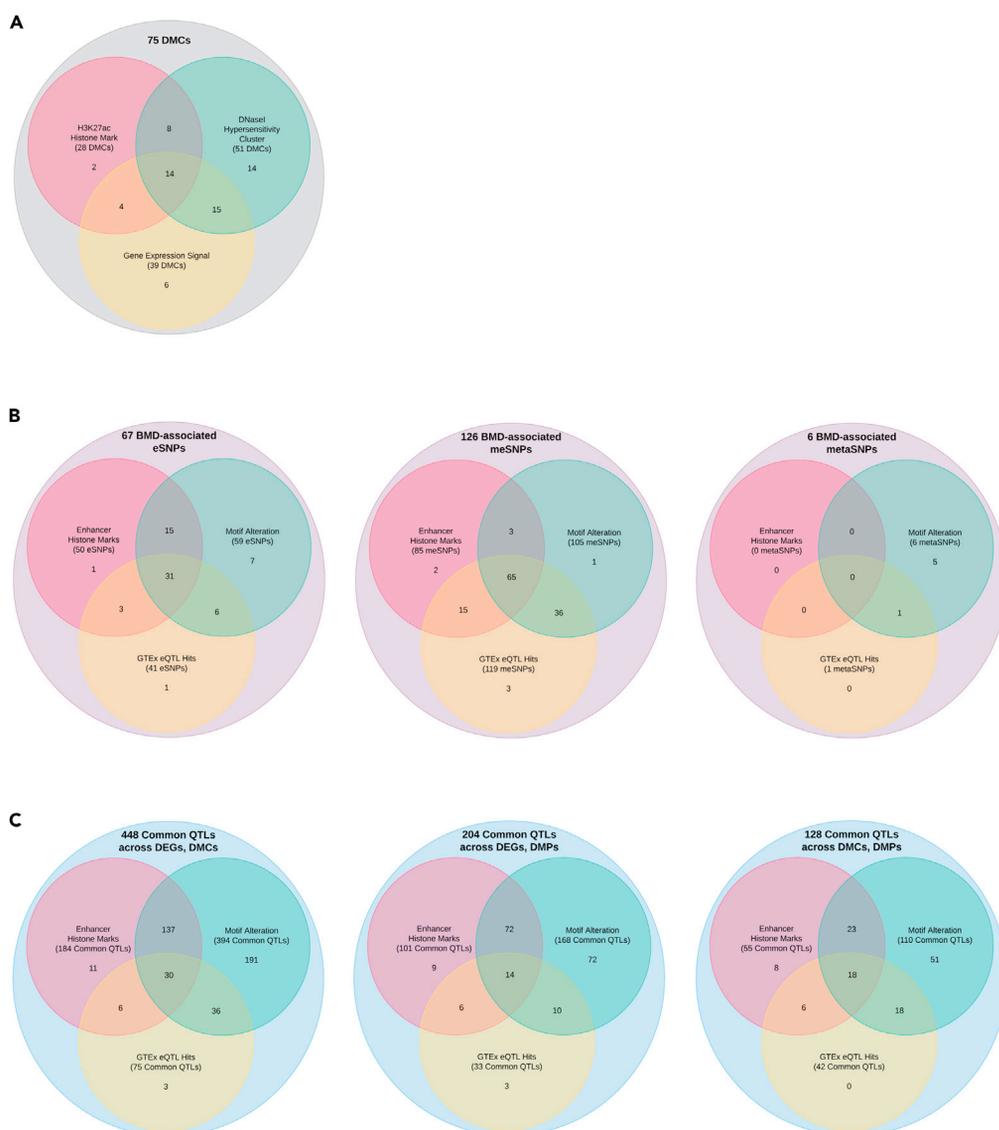


Figure 3. Functional Enrichment of Osteoporosis Biomarkers

Venn diagrams showing the functional enrichment of DMCs (A), BMD-associated QTL SNPs (B), and common QTLs (C) in different potential regulatory elements. The detailed results can be found in [Tables S3, S5, and S8](#).

(37.3%) DMCs were assigned to the H3K27ac chromatin mark, which was often observed near active regulatory elements; and 39 (52.0%) DMCs also showed gene expression signals in multiple tissues from GTEx RNA-seq data (Figure 3A, Table S3). In addition, there were 28 DMCs at sites that were predicted to be fast-evolving (PhyloP score < -1) and 6 DMCs at sites that were predicted to be conserved (PhyloP score > 1) (Table S3). Interestingly, a previous study has indicated that the methylation-PhyloP correlation is dependent on the sequence context (Chuang and Chen, 2014), although the exact mechanisms are still unclear. Therefore, the functional properties of these DMC biomarkers are worth further exploration.

Targeted eQTL, meQTL, and metaQTL Analyses

To assess the genetic effects on the prominent osteoporosis biomarkers (74 DEGs, 75 DMCs, and 23 DMPs), we performed targeted eQTL, meQTL, and metaQTL analyses by utilizing WGS data generated from the same subjects. At significance threshold of FDR = 0.05, we detected 6,778 SNP-DEG pairs, corresponding to 64 DEGs and 4,401 eQTL SNPs (eSNPs). Significant associations were also detected at 2,062

Type of Analysis	Significant SNP-Biomarker Pairs	QTL SNPs	Biomarkers
eQTL	6,778	4,401 (327)	64
meQTL	2,387	2,062 (75)	42
metaQTL	1,559	1,380 (96)	23

Table 1. Patterns of eQTLs, meQTLs, and metaQTLs

Note: The association between SNP and biomarker was tested with a linear regression model in R package *Matrix eQTL*. The significance threshold was defined as FDR < 0.05. QTL SNPs: The SNPs identified by QTL analysis. Biomarkers: DEGs, DMCs, and DMPs. The number in the bracket indicated the independent SNPs after LD pruning ($r^2 < 0.1$).

meQTL SNPs (meSNPs) for 42 DMCs and 1,380 metaQTL SNPs (metaSNPs) for 23 DMPs (Table 1). Notably, a total of 199 QTL SNPs, including 67 eSNPs, 126 meSNPs, and 6 metaSNPs, were associated with BMD ($p < 5.0 \times 10^{-8}$) in previous genome-wide association studies (GWASs) (Kemp et al., 2017; Morris et al., 2019) (Figure 3B, Table S4), supporting their impact on the BMD phenotype. Among the 67 BMD-associated eSNPs, 59 (88.1%) SNPs were predicted to alter transcription factor (TF) binding motifs and 50 (74.6%) SNPs were annotated to enhance histone marks (Figure 3B, Table S5). Similar enrichment for potential regulatory elements were also observed for the 126 meSNPs and 6 metaSNPs associated with BMD (Figure 3B, Table S5). Remarkably, we detected 28 highly correlated (linkage disequilibrium, LD $r^2 > 0.8$) BMD-associated eSNPs for biomarker FADS2. These eSNPs lie in the gene cluster of MYRF, FEN1, FADS1, and FADS2 at the 11q12.2 locus, of which 3 eSNPs were mapped to transcription start sites (TSSs) of FEN1 (rs174538) and FADS2 (rs5792235, rs99780), 1 eSNP rs1535 mapped to FADS2 enhancer, and 1 eSNP rs174562 overlapped with both TSS and enhancer of FADS1/FADS2 in a wide variety of cell/tissue types (Figure 4). Interestingly, these 5 eSNPs also showed the top five highest functionality scores among the 28 eSNPs (Figure S3, Table S6) through 3DSNP prioritization analysis (Lu et al., 2017). This result is supported by a previous functional study, which demonstrated the influence of rs174538 on the expression of FEN1 and its enzyme activity (Yang et al., 2009).

Next, we attempted to identify common QTLs across DEGs, DMCs, and DMPs. A total of 448 potential eSNPs were also associated ($p < 5.0 \times 10^{-5}$) with methylation at 48 DMCs, which corresponded to 883 SNP-DEG-DMC combinations (Table S7). For example, SNP rs2236373 is associated with gene expression of BMP3 ($p = 3.45 \times 10^{-6}$) and DNA methylation at CpG site Chr16:75279661 ($p = 4.42 \times 10^{-5}$) in BCAR1 gene. Interestingly, both biomarkers were reported to be associated with bone-related signaling pathways, such as WNT pathway (Kokabu and Rosen, 2018) and RANK/RANKL pathway (Robinson et al., 2009). Similarly, we identified 451 SNP-DEG-DMP combinations and 148 SNP-DMC-DMP combinations (Table S7). Functional annotation analyses showed that many of these multi-marker QTLs were predicted to alter TF binding motifs and/or mapped to other putative regulatory regions (Figure 3C, Table S8), suggesting that a substantial number of genetic variants may have multi-level effects on functional biomarkers across different omics levels, possibly through regulation via complex functional network/modules across multi-omics layers.

Functional Interaction Network of the Prominent Osteoporosis Biomarkers

Through STRING interaction analysis (Szklarczyk et al., 2019), we revealed functional interactions among the 74 DEGs and/or 75 DMCs (corresponding to 61 DMC-annotated genes) that were identified in the optimal multi-omics biomarker panel (Figures 5 and S4). Notably, by integrating 74 DEGs and 61 DMC-annotated genes together, we revealed a complex interaction network (Figure 5), which may indicate the potential regulatory relationships between gene expression and DNA methylation biomarkers in bone metabolism.

Furthermore, by integrating the STRING interaction results with the known biological functions (through literature review) of each prominent osteoporosis biomarker (74 DEGs, 61 DMC-annotated genes, and 23 DMPs), we assigned these biomarkers into different signaling pathways or functional activities related to bone metabolism. Finally, we identified 29 DEGs, 38 DMCs-annotated genes, and 8 DMPs that may interactively act upon several critical bone-related signaling pathways/activities in osteoblasts and/or osteoclasts (Figure 6), such as RANK/RANKL pathway, MAPK/TGF- β pathway, WNT/ β -catenin pathway, G protein-coupled receptor activity, GTP binding/GTPase activator activity, and telomere/mitochondrial activity.

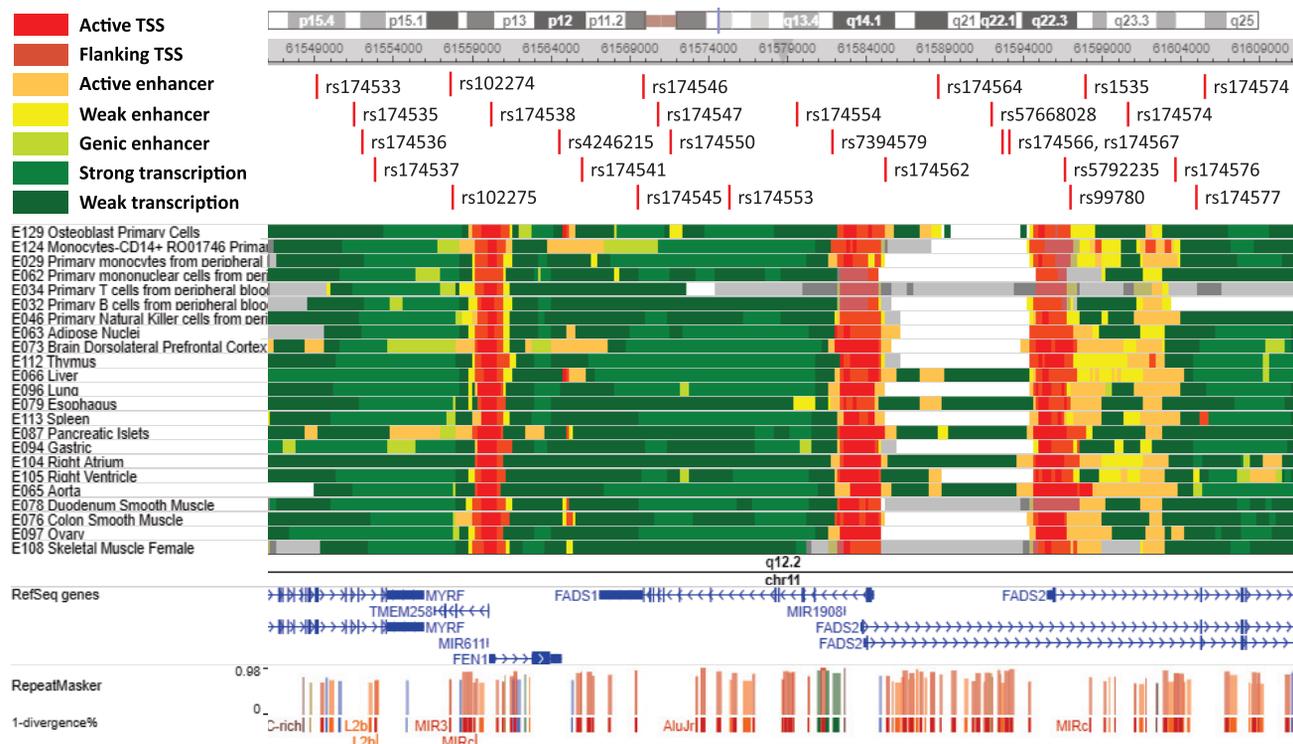


Figure 4. Chromatin-State Annotation of 28 BMD-Associated eSNPs for Biomarker FADS2

The chromatin state annotation tracks were generated by the 18-state ChromHMM model from the Roadmap Epigenomics Project under the human reference genome assembly GRCh37 (hg19) and visualized in the UCSC Genome Browser. eSNP rs174538, rs5792235, rs99780, and rs174562 were mapped to transcription start site (TSS); eSNP rs1535 and rs174562 were mapped to enhancer in a wide variety of cell/tissue types.

MR Analysis

The finding of functionally classified biomarkers prompted us to investigate their causal biological mechanisms in BMD variation. Focusing on the prominent osteoporosis biomarkers assigned to signaling pathways/activities critical for bone metabolism (Figure 6), we conducted MR analysis and identified five biomarkers (DEG biomarkers FADS2, ADRA2A, FMN1, RABL2A, and DMC biomarker CpG_4:124356866 at SPRY1) that may have putative causal effects on BMD variation (Table 2). Interestingly, gene expression of FADS2 has a robust causal effect on BMD variation based on either median-based method or inverse-variance weighted (IVW) method ($p < 0.001$). Previous GWAS meta-analysis has shown that the genetic variants in FADS2 were associated with BMD (Kemp et al., 2017). Furthermore, our QTL analysis revealed that these variants have significant effects on FADS2 expression (FDR < 0.05 , Table S4). Collectively, these results provided convergent and compelling evidence for the significance of genetic regulation of FADS2 expression in bone metabolism.

Furthermore, we assessed whether DEGs, DMCs, and DMPs within the same signaling pathway/activities are causally related to one another. As an example, we selected two biomarkers from RANK/RANKL pathway (Figure 6), namely, DEG biomarker ADCY3 and DMC biomarker NFATC1 (CpG_18:77225621). ADCY3, induced by RANKL, encodes a membrane-associated enzyme that catalyzes the formation of secondary messenger cAMP in response to G protein-coupled receptor activity. NFATC1, a ubiquitous TF in many cell types, is a well-known master regulator of both osteoblastogenesis and osteoclastogenesis (Fro-miguet et al., 2010; Kim and Kim, 2014) and has been demonstrated to play crucial roles in regulating bone homeostasis and bone mass (Winslow et al., 2006; Lee et al., 2009). Previous studies have shown that the expression of NFATC1 was markedly elevated in ADCY3-silenced cells since the elevation of intracellular cAMP culminates the PKA-mediated phosphorylation and subsequently inhibits gene expression of NFATC1 (Yoon et al., 2011; Sheridan et al., 2002). Nevertheless, no interaction was reported in STRING interaction network. Notably, in our study, MR analyses revealed a significant causal effect of gene expression of ADCY3 on DNA methylation of CpG_18:77225621 at NFATC1 based on either median-based

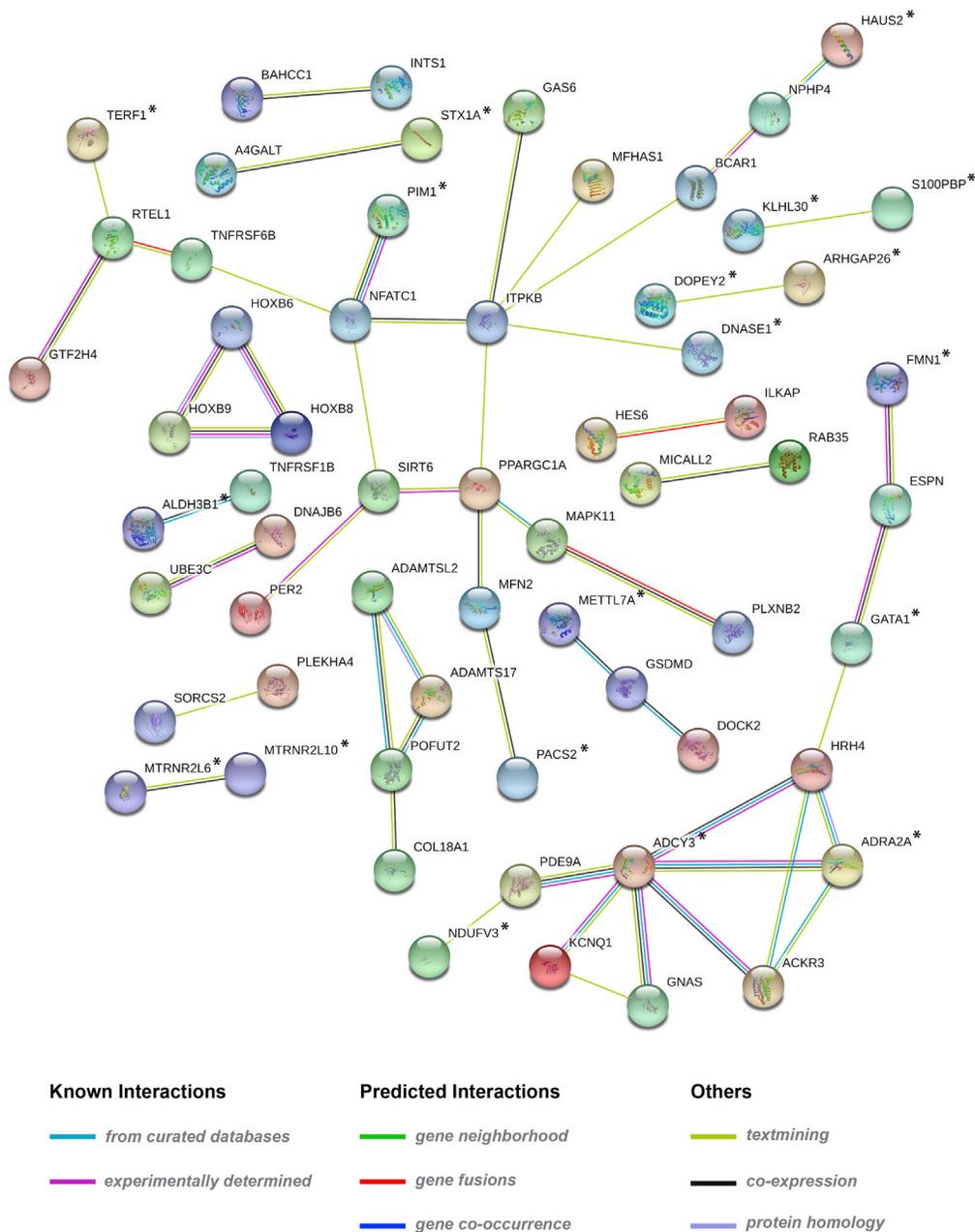


Figure 5. A Functional Interaction Network for DEGs and DMC-Annotated Genes

DMCs were assigned to their nearest gene/gene cluster based on the human reference genome assembly GRCh37 (hg19). Connections are based on co-expression and experimental evidence with a STRING v11.0 summary score above 0.4. Each color-filled node denotes a querying gene; edges between nodes indicate protein-protein interactions between protein products of the corresponding genes. Different edge colors represent the types of evidence for the association (Szkłarczyk et al., 2019). Note: * represents DEGs.

method or IVW method with $p < 0.001$ (Table 3). Interestingly, NFATC1 was identified not only through a DMC (CpG_Ch18:77225621) (Figure 7A), but also as a DEG ($p = 4.37 \times 10^{-2}$), although it did not quite (but close to) reach the adjusted threshold for integrative analysis (Figure 7B). Remarkably, functional annotation analysis revealed that DNA methylation biomarker CpG_Ch18:77225621 was linked to regulatory regions, such as TSSs, transcription regions, enhancer histone marks, and DNaseI hypersensitivity cluster (Figure 7C, Table S3). GTEx RNA-seq data also showed extensive gene expression signals of NFATC1 across multiple human cell/tissue types (Figure 7D). Taken together, these results suggested that DNA

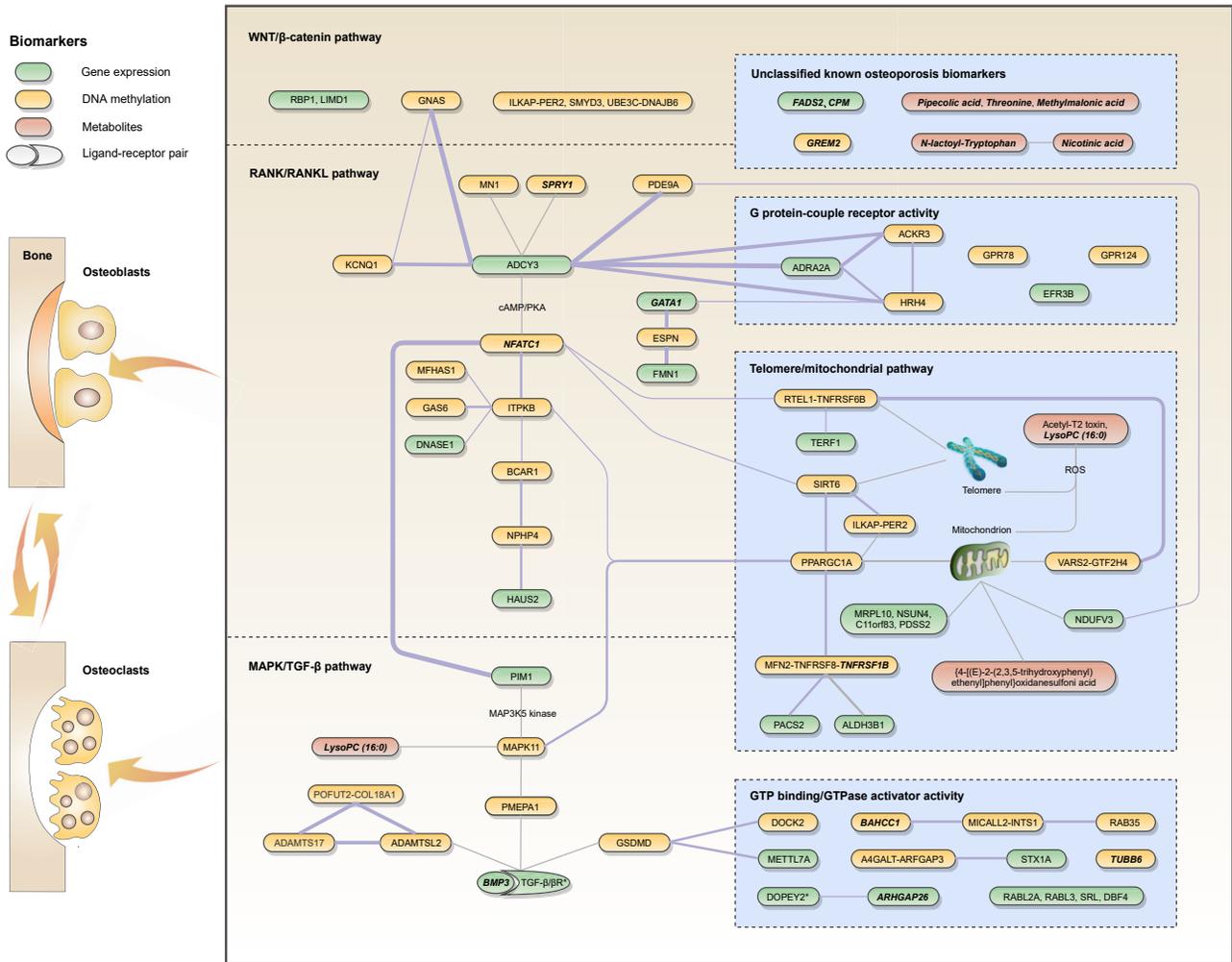


Figure 6. Biomarkers in Bone-Related Signaling Pathways/Activities

The 29 gene expression (DEGs), 38 DNA methylation (DMC-annotated genes), and 8 metabolic (DMPs) biomarkers that may interactively act upon several critical bone-related signaling pathways/activities in osteoblasts and/or osteoclasts. The purple line indicates interaction identified in the STRING interaction network, and the gray line indicates the association reported in the literature. The known osteoporosis biomarkers reported in previous GWASs or functional studies were marked in bold italics. * Several polypeptide members of TGF- β /TGF- β receptor and their coactivators (TGFB111, TGFBR3, and TGFBRAP1) were differentially expressed in single-omics analysis. DOPEY2 (CpG_Ch21:37635103) was also identified as a DNA methylation biomarker.

methylation changes in NFATC1 may regulate its gene expression activity and ultimately regulates bone metabolism.

In addition, DEG biomarker ADRA2A, a key player in G protein-coupled receptor activity, was highly associated with ADCY3 (evidence score 0.919) in STRING interaction network. Interestingly, our MR analysis also showed a significant causal impact of ADRA2A gene expression on BMD variation (Table 2). Therefore, we were interested in testing whether ADRA2A by itself or together with ADCY3 also has a causal effect on NFATC1 DNA methylation. Indeed, we identified a significant causal effect of ADRA2A gene expression or combined effects of ADCY3 and ADRA2A expression on DNA methylation of NFATC1 (Table 3). In contrast, we also conducted reversed causality analysis (e.g., test the causal effect of DNA methylation of NFATC1 on ADCY3 and/or ADRA2A gene expression) and found no significant results (data not shown).

DISCUSSION

Recent development in high-throughput profiling technologies and integrative analysis of multi-omics data offered advanced and powerful approaches to dissect complex biological problems. In this study,

Causal Biomarker	Simple Median	Weighted Median	IVW	MR-Egger	Intercept
FADS2	<0.001	<0.001	<0.001	0.23	0.92
ADRA2A	0.014	0.014	0.044	0.216	0.057
FMN1	0.032	0.023	0.005	0.154	0.626
RABL2A	0.123	0.105	0.185	0.041	0.078
CpG_4:124356866 (SPRY1)	0.043	0.078	0.026	0.21	0.026

Table 2. Significant Causal Biomarkers for BMD Variation

Note: significant results ($p < 0.05$) are marked in bold.

we pioneered an innovative approach by synthesizing the state-of-the-art methods recently developed and performed multi-omics analyses integrating gene expression and DNA methylation data in bone-related cells, as well as serum metabolomics data. We initially identified 74 DEGs, 75 DMCs (in 61 genes), and 23 DMPs for BMD variation (Table S2). There are a set of osteoporosis biomarkers in addition to biomarkers that are known to be associated with BMD. We then investigated the effects of genetic variants on these prominent osteoporosis biomarkers via targeted QTL analysis in each omics and identified hundreds of potential QTL SNPs shared by different omics, which may suggest their common biological mechanisms in pathogenesis of osteoporosis. Furthermore, we reconstructed the integrative networks/pathways in bone metabolism via STRING interaction analysis combined with functional annotation and classification analyses and revealed that a substantial proportion of these biomarkers not only interacted with each other, but also were enriched in several well-known signaling pathways or functional activities (Figure 6) that are crucial for osteoblastogenesis and osteoclastogenesis, such as RANK/RANKL pathway, MAPK/TGF- β pathway, and WNT/ β -catenin pathway, as well as G protein-coupled receptor activity, GTP binding/GTPase activator activity, and telomere/mitochondrial activity. By considering the perplexing relationship between functionally classified biomarkers, we implemented MR analysis to investigate their potential causality. Our MR results further provided supporting evidence that several gene expression and DNA methylation biomarkers have causal effects on the final BMD variation or were causally related to one another. In aggregate, our multi-omics data integration captured the complexity of the prominent interplay among different omics and pointed out a list of candidate biomarkers that may help refine biological hypotheses and propose biological validations for future studies. Furthermore, the integration framework taken here can be adopted to other complex traits/disorders and further extended to incorporate additional types of omics data (e.g., proteomics, lipidomics, and metagenomics) to enhance our understanding of the pathogenesis of human diseases.

In addition to the osteoporosis biomarkers that have been discussed in the results, several other prominent osteoporosis biomarkers that participated in well-known signaling pathways or functional activities of bone metabolism (Figure 6) should also be concerned and highlighted. Briefly, there are 18 genes, including 5 DEGs (ADCY3, DNASE1, HAUS2, GATA1, and FMN1) and 13 DMC-annotated genes (MN1, SPRY1, KCNQ1, NFATC1, ITPKB, BCAR1, NPHP4, MFHAS1, GAS6, PDE9A, RTEL1-TNFRSF6B, MFN2-TNFRSF8-TNFRSF1B, and ESPN), annotated in RANK/RANKL pathway. MN1 acts as a transcriptional activator of the osteoclastogenic cytokine RANKL and plays a crucial role in the formation of the membranous bones in the skull during mammalian development (Zhang et al., 2009). Disruption of MN1 in calvarial osteoblasts resulted in altered morphology, decreased growth rate, impaired motility, and attenuated 1,25(OH) $_2$ D $_3$ /VDR-mediated transcription, as well as reduced alkaline phosphatase activity and mineralized nodule formation (Zhang et al., 2009). SPRY1 encodes a growth factor regulator for marrow progenitor cells and promotes osteoblast differentiation at the expense of adipocytes (Urs et al., 2012). A recent transgenic mouse model revealed that miR-21, a regulator of osteoclastogenesis, can affect RANK/RANKL signaling pathway by targeting SPRY1 (Hu et al., 2017). Genetic variants in SPRY1 have also been associated with osteoporosis in Korean women (Jin et al., 2013).

We identified 2 DEGs (PIM1, BMP3), 6 DMC-annotated genes (MAPK11, PMEPA1, GSDMD, ADAMTSL2, ADAMTSL7, and POFUT2-COL18A1), and 1 metabolite biomarker LysoPC (16:0) in MAPK/TGF- β pathway. PIM1 is a member of the serine/threonine kinase family and can significantly decrease MAP3K5 kinase activity and inhibits MAP3K5-mediated phosphorylation of JNK and JNK/p38MAPK, which subsequently

Causal Biomarker	Simple Median	Weighted Median	IVW	MR-Egger	Intercept
ADCY3	<0.001	<0.001	<0.001	0.525	0.451
ADRA2A	<0.001	<0.001	0.002	0.009	0.068
ADCY3 + ADRA2A	<0.001	<0.001	<0.001	0.002	0.097

Table 3. Significant Causal Effects of Gene Expression of ADCY3 and/or ADRA2A on NFATC1 (CpG_18:77225621) DNA Methylation

Note: significant results ($p < 0.05$) are marked in bold.

reduces caspase-3 activation and cell apoptosis (Gu et al., 2009). Importantly, PIM1 can also regulate RANKL-induced osteoclastogenesis via NF- κ B activation and NFATC1 induction (Kim et al., 2010). BMP3 encodes a secreted ligand of the TGF- β superfamily of proteins. It is one of the most abundant bone morphogenetic proteins in demineralized bone matrix. BMP3 suppresses osteoblastogenesis and negatively regulates bone density by modulating TGF- β receptor availability to other ligands (Wu et al., 2016). Remarkably, we indeed observed the negative regulation of BMP3 gene expression on BMD levels ($p = 2.59 \times 10^{-5}$, Figure S5A). Several polypeptide members of TGF- β /TGF- β receptor and their coactivators (TGFB111, $p = 2.80 \times 10^{-3}$; TGFBR3, $p = 1.26 \times 10^{-5}$; TGFBRAP1, $p = 7.77 \times 10^{-4}$) were also differentially expressed between the high-BMD and low-BMD groups (Figures S5B–S5D). MAPK11, also known as p38- β , is one of the four p38 MAPKs that play a crucial role in osteoblast differentiation and bone development and maintenance (Hu et al., 2003). A recent study showed that MAPK11 can also enhance osteoclastogenesis and bone resorption (He et al., 2014). PMEPA1 encodes a transmembrane protein that contains a Smad interacting motif (SIM). PMEPA1 has a significant role in osteoclastogenesis (Funakubo et al., 2018) and can also act as a TGF- β signaling regulator in osteoblast proliferation (Fournier et al., 2015). ADAMTSL2 is directly involved in TGF- β bioavailability and plays a key role in osteoblast and skeletal development (Le Goff et al., 2008).

In addition, there are 2 DEGs (RBP1, LIMD1) and 4 DMC-annotated genes (GNAS, ILKAP-PER2, SMYD3, and UBE3C-DNAJB6) in WNT/ β -catenin pathway. RBP1 can act as a RUNX2 coactivator and promotes osteoblastic differentiation (Monroe et al., 2010). LIMD1 encodes a scaffold protein that has been implicated in the regulation of osteoclastogenesis through an interaction with the p62/sequestosome protein (Luderer et al., 2008). LIMD1 protein can also influence osteoblast differentiation and function; as such, *Limd1*($-/-$) calvarial osteoblasts displayed increased mineralization and accelerated differentiation (Luderer et al., 2008). Furthermore, there is a significant increase in nuclear beta-catenin staining in differentiating *Limd1*($-/-$) calvarial osteoblasts (Luderer et al., 2008), suggesting that LIMD1 is a negative regulator of canonical WNT signaling in osteoblasts. The GNAS gene is a complex imprinted locus that produces multiple transcripts (such as *Gs α* , *XLAS*, *NESP55*) through the use of alternative promoters and alternative splicing. A recent study by Ramaswamy et al. (Ramaswamy et al., 2017) demonstrated that *Gnas* inactivation in mice negatively affects cortical bone quality and strength, with mutation of the paternal allele causing more severe effects than maternal mutations. These effects of *Gsa* deletion on bone maintenance were exerted through enhanced osteoclast differentiation and increased bone resorption, mediated by *Gsa* signaling via cAMP/PKA and WNT/ β -catenin pathways (Ramaswamy et al., 2017). SMYD3 encodes a histone methyltransferase that functions in RNA polymerase II complexes by an interaction with a specific RNA helicase (Hamamoto et al., 2004) and controls a WNT-responsive epigenetic switch (Wang et al., 2018).

G protein-coupled receptor activity includes 2 DEGs (ADRA2A, EFR3B) and 4 DMC-annotated genes (HRH4, ACKR3, GPR78, and GPR124). GTP binding/GTPase activator activity includes 7 DEGs (METTL7A, STX1A, ARHGAP26, RABL2A, RABL3, SRL, and DBF4) and 6 DMC-annotated genes (DOCK2, BAHCC1, MICALL2-INTS1, RAB35, A4GALT-ARFGAP3, and TUBB6). For example, ADRA2A, a member of the G protein-coupled receptor superfamily, is involved in neuro-endocrine regulation of bone resorption (Mlakar et al., 2015). ARHGAP26 encodes a GTPase-activating protein, and a mutation in this gene has recently been determined to be associated with BMD (Kemp et al., 2017). Another interesting gene is DBF4, which plays a central role in DNA replication and cell proliferation through nitrogen-containing bisphosphonate-induced cytotoxicity (Bivi et al., 2009). Nitrogen-containing bisphosphonate can potentially inhibit the prenylation and function of GTP-binding proteins required for osteoclast formation and now is firmly established as first-line therapy for osteoporosis (Grey and Reid, 2006). TUBB6 plays a key role in GTP binding and has been associated with BMD variation (Daswani et al., 2015). In addition, there are 8 DEGs (TERF1, MRPL10,

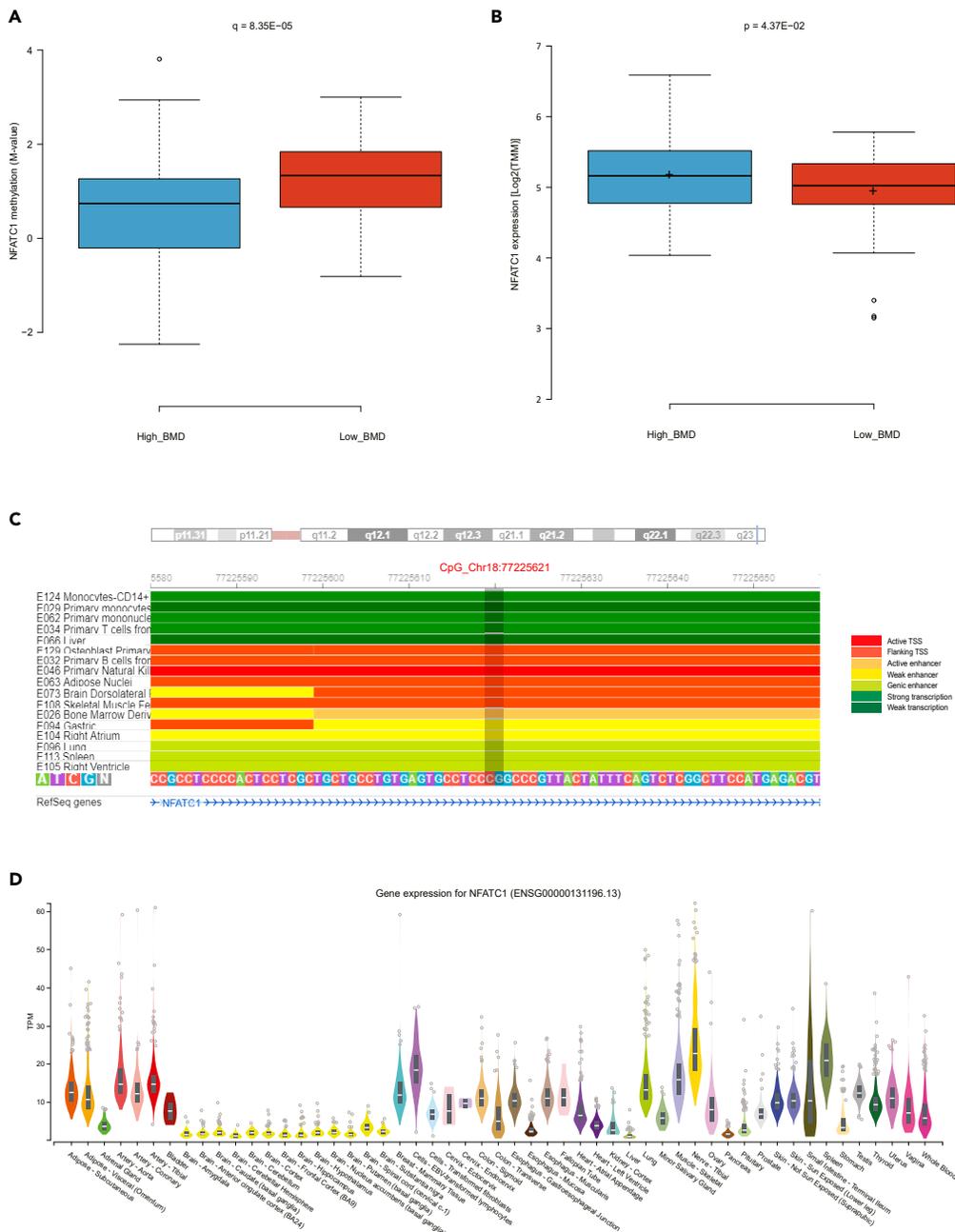


Figure 7. Biological Importance of Biomarker NFATC1

(A and B) Boxplot for NFATC1 DNA methylation (A) and gene expression (B) levels in high-BMD and low-BMD groups. The vertical axis in (A) represents DNA methylation level (M-value); the adjusted q-value was determined by R package *methylKit*. The vertical axis in (B) represents the gene expression level [$\log_2(\text{TMM})$]; p values for moderated statistics were determined by R package *limma*.

(C) The chromatin-state annotation for DNA methylation biomarker CpG_Ch18:77225621 at NFATC1. The chromatin state annotation tracks were generated by the 18-state ChromHMM model from the Roadmap Epigenomics Project under the human reference genome assembly GRCh37 (hg19) and visualized in the UCSC Genome Browser.

(D) The gene expression signals of NFATC1 across multiple human cell/tissue types from GTEx RNA-seq data.

NSUN4, C11orf83, PDSS2, PACS2, NDUFV3, and ALDH3B1), 6 DMC-annotated genes (PPARGC1A, ILKAP-
PER2, SIRT6, RTEL-TNFRSF6B, VARS2-GTF2H4, and MFN2-TNFRSF8-TNFRSF1B), and 3 metabolites
biomarkers ((4-((E)-2-(2,3,5-trihydroxyphenyl)ethenyl)phenyl)oxidanesulfonic acid, Acetyl-T2 toxin, and

LysoPC (16:0)) associated with telomere/mitochondrial activity. TERF1 encodes a component of the telomere nucleoprotein complex, which can regulate telomere elongation and plays a key role in aging-related disease (Blasco, 2005). Previous study has shown that defects in telomere maintenance molecules impair osteoblast differentiation and promote osteoporosis (Pignolo et al., 2008). PPARGC1A encodes a transcriptional coactivator that mediates mitochondrial biogenesis and energy metabolism (Liang and Ward, 2006). This coactivator interacts with PPARG, which permits the interaction of this coactivator with multiple transcription factors (Vega et al., 2000). PPARGC1A can control skeletal stem cell fate and bone-fat balance in osteoporosis and skeletal aging (Yu et al., 2018). Interestingly, WNT signaling can activate PPARGC1A expression and upregulate mitochondrial biogenesis; this upregulation contributes to the osteoblastic differentiation (An et al., 2010). PER2 encodes a transcriptional repressor that forms a core component of the circadian clock. It directly and specifically represses PPARG proadipogenic activity by blocking PPARG recruitment to target promoters and thereby inhibiting transcriptional activation. PER2 is required for fatty acid and lipid metabolism and is involved as well in the regulation of circulating insulin levels (Grimaldi et al., 2010). SIRT6 encodes a member of the sirtuin family of NAD-dependent enzymes that are implicated in cellular stress resistance, genomic stability, aging, and energy homeostasis. A recent study demonstrated that the way SIRT6 regulated osteoclast was predominantly through osteoblast paracrine manner, rather than osteoclast-autonomous behavior, which provided a valuable insight into the pathogenesis of osteoporosis due to SIRT6 mutation (Zhang et al., 2018). Metabolite Acetyl-T2 toxin is the class of organic compounds known as trichothecenes. Trichothecene inhibition of protein synthesis in the mitochondria allows reactive oxygen species (ROS) to build up in the cell, which inevitably leads to oxidative stress and induction of the programmed cell death pathway, apoptosis (Fang et al., 2012). LysoPC (16:0) is a lysophosphatidylcholine. Previous study showed that LysoPC-induced p38 MAPK signaling pathway can control monocyte migration (Tan et al., 2009) and a novel lysophosphatidylcholine derivative (SCOH) can inhibit osteoclast differentiation and bone resorption (Kwak et al., 2004). Moreover, lysophosphatidylcholine can produce mitochondrial ROS generation, increase intracellular free calcium concentration, activate active adenylate cyclase (e.g., gene expression biomarker ADCY3), and enhance glucose-dependent insulin secretion via an orphan G protein-coupled receptor (Watanabe et al., 2006; Chaudhuri et al., 2003; Soga et al., 2005). LysoPC(16:0) isolated from rats plasma was also proved to be related to osteoporosis (Liu et al., 2012).

Notably, there are two adjacent gene clusters RTEL1-TNFRSF6B and MFN2-TNFRSF8-TNFRSF1B involved in multiple signaling pathways or functional activities. RTEL1 encodes a DNA helicase that interacts with proteins in the shelterin complex and plays a key role in the stability, protection, and elongation of telomeres (Deng et al., 2013). Interestingly, telomere deficiency can impair osteoblast differentiation and promote osteoporosis (Pignolo et al., 2008). TNFRSF6B can suppress RANKL-induced osteoclastogenesis via down-regulating NFATC1 and enhancing cell apoptosis (Cheng et al., 2013). MFN2 encodes a mitochondrial membrane protein that participates in the maintenance and operation of the mitochondrial network (Bach et al., 2003), which has been linked to osteoclast activity and bone metabolism via an iron-related fundamental pathway (Ishii et al., 2009). A recent study reported that MFN2 can facilitate osteoclastogenesis by regulating the calcium-calcineurin-NFATC1 axis as well (Szkłarczyk et al., 2019). TNFRSF8 and TNFRSF1B both are members of the TNF-receptor superfamily. Genetic variants in TNFRSF1B gene have been associated with femoral neck BMD (Albagha et al., 2002) and bone structure (Mullin et al., 2008).

We also identified eight unclassified osteoporosis biomarkers, including 2 DEGs (FADS2, CPM), 1 DMC-annotated gene GREM2, and 5 metabolite biomarkers (pipecolic acid, threonine, methylmalonic acid, N-lactoyl-tryptophan, and nicotinic acid), which have been reported in previous association studies. Genetic variants in FADS2, CPM, and GREM2 have been associated with BMD variation in previous GWAS (Kemp et al., 2017). Pipecolic acid is a normal human metabolite present in human blood; it has been associated with both total hip and lumbar spine BMD phenotypes in TwinsUK population (Moayyeri et al., 2018). Interestingly, we also identified two function-related metabolites (N-lactoyl-tryptophan, nicotinic acid). N-lactoyl-tryptophan is lactoyl derivative of tryptophan. Tryptophan acts as the precursor of nicotinic acid (also known as vitamin-B3). Previous study has indicated that tryptophan plays an essential role in osteoblastic differentiation (Pallu et al., 2012). Recently, Michalowska et al. investigated the influence of tryptophan and its metabolites on bone remodeling and observed significant changes of tryptophan levels in bone metabolic diseases (Michalowska et al., 2015). Nicotinic acid occurs naturally in food. Several studies have examined the effect of nicotinic acid on bone metabolism. For example, there is a positive correlation between

dietary intake of nicotinic acid and BMD in premenopausal Japanese women (Sasaki and Yanagibori, 2001). The other study reported a significant inverse association of dietary nicotinic acid intake with hip BMD, but there was no significant association with total body BMD (Carbone et al., 2019). These conflicting findings deserve further serious investigation to better understand the effect of supplementation of nicotinic acid on bone biology. Methylmalonic acid is a dicarboxylic acid that is a c-methylated derivative of malonate. Methylmalonic acid was found to induce osteoclastogenesis in a dose-dependent manner, and vitamin-B12 deficiency may lead to decreased bone mass by increased osteoclast formation due to increased methylmalonic acid level (Vaes et al., 2009). Threonine is an essential amino acid that is used in the biosynthesis of proteins. A prior study has demonstrated that threonine can modulate the growth and the differentiation of osteoblasts cultured *in vitro* and confirmed the relationship between osteoporotic hip fracture and inadequate protein intake (Conconi et al., 2001). Notably, a recent metabolomic study has shown that threonine is associated with BMD and can improve the power for osteoporosis classification in males (Wang et al., 2019).

In summary, we conducted an innovative multi-omics integrative analysis and identified a set of osteoporosis biomarkers as well as biological pathways/networks that may contribute to BMD variation. Our results revealed valuable insights into the pathogenesis of osteoporosis and aided in generating hypotheses for future functional studies.

Limitations of the Study

Several limitations of this study should be noted. First, our sample size is relatively small. However, our study is the largest so far in the bone field for multi-omics analyses and we applied an extreme phenotype sampling strategy with stringent inclusion and exclusion criteria, which is known to provide enhanced statistical power for association analysis compared with studies using comparable numbers of randomly sampled subjects (Bjornland et al., 2018). Moreover, data from different omics levels can provide complementary and inherent validation information with each other, and thus, integrating multi-omics data can partially compensate for the relatively small sample sizes (Hasin et al., 2017). But clearly, our results need to be validated in future studies with large sample size. Second, we used a relatively homogeneous cell type, peripheral blood monocytes (PBMs), as model cells for gene expression and DNA methylation analysis for osteoporosis. PBMs can act as osteoclast precursors, secrete cytokines essential for osteoclast differentiation and function, and represent a major target cell of sex hormones for bone metabolism (Komano et al., 2006). Notably, several transcriptomic and proteomic studies in PBMs have revealed significant insights into the pathogenic mechanisms of osteoporosis (Leung et al., 2011; Kotani et al., 2013; Zhou et al., 2015). On the other hand, we acknowledge that the ideal model cells for osteoporosis study are primary bone cells (e.g., osteoblasts, osteoclasts, and osteocytes). With the continuous development of high-throughput multi-omics profiling technologies, particularly for single-cell sequencing (Macauley et al., 2017), we will be able to apply multi-omics analysis on human primary bone cells in the near future. Third, the identified biomarkers and results of causal analysis exclusively depend on computational modeling; hence, further experimental validation work should be conducted to confirm the biological significance and causality of these osteoporosis biomarkers. Nonetheless, we want to emphasize that traditional validation/further exploration using *in vitro* cells or *in vivo* mice models may be useful in some cases but may not completely reflect human *in vivo* functional mechanisms in other cases.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

DATA AND CODE AVAILABILITY

All whole genome sequencing data, RNA sequencing data, reduced representation bisulfite sequencing data, and metabolite data in this manuscript has been deposited in dbGaP. The accession number for the data reported in this paper is dbGaP: phs001960.v1.p1.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.100847>.

ACKNOWLEDGMENTS

This study was partially supported or benefited by grants from the National Institutes of Health, United States (R01AR069055, R01MH104680, R01AR059781, U19AG055373, and P20GM109036) and the Edward G. Schlieder Endowment and the Drs. W. C. Tsai and P. T. Kung Professorship in Biostatistics from Tulane University.

AUTHOR CONTRIBUTIONS

C.Q. conceived and designed the study. C.Q. and L. Zhang performed all the experiments. C.Q., F.Y., K.S., C.X., Z.W. did the downstream analysis. L. Zhao. and Q.T. recruited the samples. H.S., H.D., W.H., and Y.W. provided additional feedback. C.Q. and H.S. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 23, 2019

Revised: November 22, 2019

Accepted: January 13, 2020

Published: February 21, 2020

REFERENCES

- Albagha, O.M., Tasker, P.N., McGuigan, F.E., Reid, D.M., and Ralston, S.H. (2002). Linkage disequilibrium between polymorphisms in the human TNFRSF1B gene and their association with bone mass in perimenopausal women. *Hum. Mol. Genet.* *11*, 2289–2295.
- Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* *16*, 197–212.
- An, J.H., Yang, J.Y., Ahn, B.Y., Cho, S.W., Jung, J.Y., Cho, H.Y., Cho, Y.M., Kim, S.W., Park, K.S., Kim, S.Y., et al. (2010). Enhanced mitochondrial biogenesis contributes to Wnt induced osteoblastic differentiation of C3H10T1/2 cells. *Bone* *47*, 140–150.
- Bach, D., Pich, S., Soriano, F.X., Vega, N., Baumgartner, B., Oriola, J., Dagaard, J.R., Lloberas, J., Camps, M., Zierath, J.R., et al. (2003). Mitofusin-2 determines mitochondrial network architecture and mitochondrial metabolism. A novel regulatory mechanism altered in obesity. *J. Biol. Chem.* *278*, 17190–17197.
- Bivi, N., Romanello, M., Harrison, R., Clarke, I., Hoyle, D.C., Moro, L., Ortolani, F., Bonetti, A., Quadrioglio, F., Tell, G., et al. (2009). Identification of secondary targets of N-containing bisphosphonates in mammalian cells via parallel competition analysis of the barcoded yeast deletion collection. *Genome Biol.* *10*, R93.
- Bjornland, T., Bye, A., Ryeng, E., Wisloff, U., and Langaas, M. (2018). Powerful extreme phenotype sampling designs and score tests for genetic association studies. *Stat. Med.* *37*, 4234–4251.
- Blasco, M.A. (2005). Telomeres and human disease: ageing, cancer and beyond. *Nat. Rev. Genet.* *6*, 611–622.
- Carbone, L.D., Buzkova, P., Fink, H.A., Raiford, M., Le, B., Isales, C.M., Shikany, J.M., Coughlin, S.S., and Robbins, J.A. (2019). Association of dietary niacin intake with incident hip fracture, BMD, and body composition: the cardiovascular health study. *J. Bone Miner. Res.* *34*, 643–652.
- Chaudhuri, P., Colles, S.M., Damron, D.S., and Graham, L.M. (2003). Lysophosphatidylcholine inhibits endothelial cell migration by increasing intracellular calcium and activating calpain. *Arterioscler. Thromb. Vasc. Biol.* *23*, 218–223.
- Chen, X.F., Zhu, D.L., Yang, M., Hu, W.X., Duan, Y.Y., Lu, B.J., Rong, Y., Dong, S.S., Hao, R.H., Chen, J.B., et al. (2018). An osteoporosis risk SNP at 1p36.12 acts as an allele-specific enhancer to modulate LINC00339 expression via long-range loop formation. *Am. J. Hum. Genet.* *102*, 776–793.
- Cheng, C.P., Sheu, M.J., Sytwu, H.K., and Chang, D.M. (2013). Decoy receptor 3 suppresses RANKL-induced osteoclastogenesis via down-regulating NFATc1 and enhancing cell apoptosis. *Rheumatology (Oxford)* *52*, 609–622.
- Chuang, T.J., and Chen, F.C. (2014). DNA methylation is associated with an increased level of conservation at nondegenerate nucleotides in mammals. *Mol. Biol. Evol.* *31*, 387–396.
- Conconi, M.T., Tommasini, M., Muratori, E., and Parnigotto, P.P. (2001). Essential amino acids increase the growth and alkaline phosphatase activity in osteoblasts cultured in vitro. *Farmacol.* *56*, 755–761.
- Daswani, B., Gupta, M.K., Gavali, S., Desai, M., Sathe, G.J., Patil, A., Parte, P., Sirdeshmukh, R., and Khatkhatay, M.I. (2015). Monocyte proteomics reveals involvement of phosphorylated HSP27 in the pathogenesis of osteoporosis. *Dis. Markers* *2015*, 196589.
- Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* *23*, R89–R98.
- Deng, Z., Glousker, G., Molczan, A., Fox, A.J., Lamm, N., Dheekollu, J., Weizman, O.E., Schertzer, M., Wang, Z., Vladimirova, O., et al. (2013). Inherited mutations in the helicase RTEL1 cause telomere dysfunction and Hoyeraal-Hreidarsson syndrome. *Proc. Natl. Acad. Sci. U S A* *110*, E3408–E3416.
- Fang, H., Wu, Y., Guo, J., Rong, J., Ma, L., Zhao, Z., Zuo, D., and Peng, S. (2012). T-2 toxin induces apoptosis in differentiated murine embryonic stem cells through reactive oxygen species-mediated mitochondrial pathway. *Apoptosis* *17*, 895–907.
- Fournier, P.G., Juarez, P., Jiang, G., Clines, G.A., Niewolna, M., Kim, H.S., Walton, H.W., Peng, X.H., Liu, Y., Mohammad, K.S., et al. (2015). The TGF-beta signaling regulator PMEPA1 suppresses prostate cancer metastases to bone. *Cancer Cell* *27*, 809–821.
- Fromigue, O., Hay, E., Barbara, A., and Marie, P.J. (2010). Essential role of nuclear factor of activated T cells (NFAT)-mediated Wnt signaling in osteoblast differentiation induced by strontium ranelate. *J. Biol. Chem.* *285*, 25251–25258.
- Funakubo, N., Xu, X., Kukita, T., Nakamura, S., Miyamoto, H., and Kukita, A. (2018). Pmepa1 induced by RANKL-p38 MAPK pathway has a novel role in osteoclastogenesis. *J. Cell Physiol.* *233*, 3105–3118.
- Le Goff, C., Morice-Picard, F., Dagoneau, N., Wang, L.W., Perrot, C., Crow, Y.J., Bauer, F., Flori, E., Prost-Squarcioni, C., Krakow, D., et al. (2008). ADAMTSL2 mutations in geleophysic dysplasia demonstrate a role for ADAMTS-like proteins in TGF-beta bioavailability regulation. *Nat. Genet.* *40*, 1119–1123.
- Grey, A., and Reid, I.R. (2006). Differences between the bisphosphonates for the prevention and treatment of osteoporosis. *Ther. Clin. Risk Manag.* *2*, 77–86.
- Grimaldi, B., Bellet, M.M., Katada, S., Astarita, G., Hirayama, J., Amin, R.H., Granneman, J.G., Piomelli, D., Leff, T., and Sassone-Corsi, P. (2010).

- PER2 controls lipid metabolism by direct regulation of PPARgamma. *Cell Metab.* 12, 509–520.
- GTEX Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213.
- Gu, J.J., Wang, Z., Reeves, R., and Magnuson, N.S. (2009). PIM1 phosphorylates and negatively regulates ASK1-mediated apoptosis. *Oncogene* 28, 4261–4271.
- Gunther, O.P., Chen, V., Freue, G.C., Balshaw, R.F., Tebbutt, S.J., Hollander, Z., Takhar, M., McMaster, W.R., Mcmanus, B.M., Keown, P.A., et al. (2012). A computational pipeline for the development of multi-marker bio-signature panels and ensemble classifiers. *BMC Bioinformatics* 13, 326.
- Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F.P., Li, M., Yagyu, R., and Nakamura, Y. (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat. Cell Biol.* 6, 731–740.
- Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18, 83.
- He, Z., He, J., Liu, Z., Xu, J., Yi, S.F., Liu, H., and Yang, J. (2014). MAPK11 in breast cancer cells enhances osteoclastogenesis and bone resorption. *Biochimie* 106, 24–32.
- Hu, Y., Chan, E., Wang, S.X., and Li, B. (2003). Activation of p38 mitogen-activated protein kinase is required for osteoblast differentiation. *Endocrinology* 144, 2068–2074.
- Hu, C.H., Sui, B.D., Du, F.Y., Shuai, Y., Zheng, C.X., Zhao, P., Yu, X.R., and Jin, Y. (2017). miR-21 deficiency inhibits osteoclast function and prevents bone loss in mice. *Sci. Rep.* 7, 43191.
- Ishii, K.A., Fumoto, T., Iwai, K., Takeshita, S., Ito, M., Shimohata, N., Aburatani, H., Taketani, S., Lelliott, C.J., Vidal-Puig, A., et al. (2009). Coordination of PGC-1beta and iron uptake in mitochondrial biogenesis and osteoclast activation. *Nat. Med.* 15, 259–266.
- Jin, H.S., Kim, B.Y., Kim, J., Hong, K.W., Jung, S.Y., Lee, Y.S., Huh, D., Oh, B., Chung, Y.S., and Jeong, S.Y. (2013). Association between the SPRY1 gene polymorphism and obesity-related traits and osteoporosis in Korean women. *Mol. Genet. Metab.* 108, 95–101.
- Kanis, J.A. (2002). Diagnosis of osteoporosis and assessment of fracture risk. *Lancet* 359, 1929–1936.
- Kemp, J.P., Morris, J.A., Medina-Gomez, C., Forgetta, V., Warrington, N.M., Youtlen, S.E., Zheng, J., Gregson, C.L., Grundberg, E., Trajanoska, K., et al. (2017). Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat. Genet.* 49, 1468–1475.
- Kim, J.H., and Kim, N. (2014). Regulation of NFATc1 in osteoclast differentiation. *J. Bone Metab.* 21, 233–241.
- Kim, K., Kim, J.H., Youn, B.U., Jin, H.M., and Kim, N. (2010). Pim-1 regulates RANKL-induced osteoclastogenesis via NF-kappaB activation and NFATc1 induction. *J. Immunol.* 185, 7460–7466.
- Kokabu, S., and Rosen, V. (2018). BMP3 expression by osteoblast lineage cells is regulated by canonical Wnt signaling. *FEBS Open Bio* 8, 168–176.
- Komano, Y., Nanki, T., Hayashida, K., Taniguchi, K., and Miyasaka, N. (2006). Identification of a human peripheral blood monocyte subset that differentiates into osteoclasts. *Arthritis Res. Ther.* 8, R152.
- Kotani, M., Kikuta, J., Klauschen, F., Chino, T., Kobayashi, Y., Yasuda, H., Tamai, K., Miyawaki, A., Kanagawa, O., Tomura, M., et al. (2013). Systemic circulation and bone recruitment of osteoclast precursors tracked by using fluorescent imaging techniques. *J. Immunol.* 190, 605–612.
- Kraus, W.E., Muoio, D.M., Stevens, R., Craig, D., Bain, J.R., Grass, E., Haynes, C., Kwee, L., Qin, X., Slentz, D.H., et al. (2015). Metabolomic quantitative trait loci (mQTL) mapping implicates the ubiquitin proteasome system in cardiovascular disease pathogenesis. *PLoS Genet.* 11, e1005553.
- Kwak, H.B., Lee, S.W., Li, Y.J., Kim, Y.A., Han, S.Y., Jhon, G.J., Kim, H.H., and Lee, Z.H. (2004). Inhibition of osteoclast differentiation and bone resorption by a novel lysophosphatidylcholine derivative, SCOH. *Biochem. Pharmacol.* 67, 1239–1248.
- Lee, Y., Ha, J., Kim, H.J., Kim, Y.S., Chang, E.J., Song, W.J., and Kim, H.H. (2009). Negative feedback inhibition of NFATc1 by DYRK1A regulates bone homeostasis. *J. Biol. Chem.* 284, 33343–33351.
- Lemire, M., Zaidi, S.H., Ban, M., Ge, B., Aissi, D., Germain, M., Kassam, I., Wang, M., Zanke, B.W., Gagnon, F., et al. (2015). Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat. Commun.* 6, 6326.
- Leung, R., Cuddy, K., Wang, Y., Rommens, J., and Glogauer, M. (2011). Sbds is required for Rac2-mediated monocyte migration and signaling downstream of RANK during osteoclastogenesis. *Blood* 117, 2044–2053.
- Liang, H., and Ward, W.F. (2006). PGC-1alpha: a key regulator of energy metabolism. *Adv. Physiol. Educ.* 30, 145–151.
- Liu, X., Zhang, S., Lu, X., Zheng, S., Li, F., and Xiong, Z. (2012). Metabonomic study on the anti-osteoporosis effect of *Rhizoma Drynariae* and its action mechanism using ultra-performance liquid chromatography-tandem mass spectrometry. *J. Ethnopharmacol.* 139, 311–317.
- Liu, Y., Devescovi, V., Chen, S., and Nardini, C. (2013). Multilevel omic data integration in cancer cell lines: advanced annotation and emergent properties. *BMC Syst. Biol.* 7, 14.
- Lu, Y., Quan, C., Chen, H., Bo, X., and Zhang, C. (2017). 3DSNP: a database for linking human noncoding SNPs to their three-dimensional interacting genes. *Nucleic Acids Res.* 45, D643–D649.
- Luderer, H.F., Bai, S., and Longmore, G.D. (2008). The LIM protein LIMD1 influences osteoblast differentiation and function. *Exp. Cell Res.* 314, 2884–2894.
- Macaulay, I.C., Ponting, C.P., and Voet, T. (2017). Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33, 155–168.
- McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Wellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science* 342, 747–749.
- Michalowska, M., Znorko, B., Kaminski, T., Oksztulska-Kolanek, E., and Pawlak, D. (2015). New insights into tryptophan and its metabolites in the regulation of bone metabolism. *J. Physiol. Pharmacol.* 66, 779–791.
- Mlakar, V., Jurkovic Mlakar, S., Zupan, J., Komadina, R., Prezelj, J., and Marc, J. (2015). ADRA2A is involved in neuro-endocrine regulation of bone resorption. *J. Cell. Mol. Med.* 19, 1520–1529.
- Moayyeri, A., Cheung, C.L., Tan, K.C., Morris, J.A., Cerani, A., Mohney, R.P., Richards, J.B., Hammond, C., Spector, T.D., and Menni, C. (2018). Metabolomic pathways to osteoporosis in middle-aged women: a genome-metabolome-wide mendelian randomization study. *J. Bone Miner. Res.* 33, 643–650.
- Monroe, D.G., Hawse, J.R., Subramaniam, M., and Spelsberg, T.C. (2010). Retinoblastoma binding protein-1 (RBP1) is a Runx2 coactivator and promotes osteoblastic differentiation. *BMC Musculoskelet. Disord.* 11, 104.
- Morris, J.A., Kemp, J.P., Youtlen, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpesu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* 51, 258–266.
- Mullin, B.H., Prince, R.L., Dick, I.M., Islam, F.M., Hart, D.J., Spector, T.D., Devine, A., Dudbridge, F., and Wilson, S.G. (2008). Bone structural effects of variation in the TNFRSF1B gene encoding the tumor necrosis factor receptor 2. *Osteoporos. Int.* 19, 961–968.
- Pallu, S., Rochefort, G.Y., Jaffre, C., Refregiers, M., Maurel, D.B., Benaitreau, D., Lespessailles, E., Jammé, F., Chappard, C., and Benhamou, C.L. (2012). Synchrotron ultraviolet microspectroscopy on rat cortical bone: involvement of tyrosine and tryptophan in the osteocyte and its environment. *PLoS One* 7, e43930.
- Pierce, B.L., Tong, L., Argos, M., Demanelis, K., Jasmine, F., Rakibuz-Zaman, M., Sarwar, G., Islam, M.T., Shahriar, H., Islam, T., et al. (2018). Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat. Commun.* 9, 804.
- Pignolo, R.J., Suda, R.K., Mcmillan, E.A., Shen, J., Lee, S.H., Choi, Y., Wright, A.C., and Johnson, F.B. (2008). Defects in telomere maintenance

molecules impair osteoblast differentiation and promote osteoporosis. *Aging Cell* 7, 23–31.

Ralston, S.H., and de Crombrughe, B. (2006). Genetic regulation of bone mass and susceptibility to osteoporosis. *Genes Dev.* 20, 2492–2506.

Ralston, S.H., and Uitterlinden, A.G. (2010). Genetics of osteoporosis. *Endocr. Rev.* 31, 629–662.

Ramaswamy, G., Kim, H., Zhang, D., Lounev, V., Wu, J.Y., Choi, Y., Kaplan, F.S., Pignolo, R.J., and Shore, E.M. (2017). Gsalpha controls cortical bone quality by regulating osteoclast differentiation via cAMP/PKA and beta-catenin pathways. *Sci. Rep.* 7, 45140.

Robinson, L.J., Yaroslavskiy, B.B., Griswold, R.D., Zadorozny, E.V., Guo, L., Tourkova, I.L., and Blair, H.C. (2009). Estrogen inhibits RANKL-stimulated osteoclastic differentiation of human monocytes through estrogen and RANKL-regulated interaction of estrogen receptor- α with BCAR1 and Traf6. *Exp. Cell Res.* 315, 1287–1301.

Rohart, F., Gautier, B., Singh, A., and Le Cao, K.A. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13, e1005752.

Sasaki, S., and Yanagibori, R. (2001). Association between current nutrient intakes and bone mineral density at calcaneus in pre- and postmenopausal Japanese women. *J. Nutr. Sci. Vitaminol. (Tokyo)* 47, 289–294.

Sheridan, C.M., Heist, E.K., Beals, C.R., Crabtree, G.R., and Gardner, P. (2002). Protein kinase A negatively modulates the nuclear accumulation of NF-ATc1 by priming for subsequent phosphorylation by glycogen synthase kinase-3. *J. Biol. Chem.* 277, 48664–48676.

Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., and Le Cao, K.A. (2019). DIABLO: an integrative approach for identifying key molecular drivers from multi-omic assays. *Bioinformatics* 35, 3055–3062.

Soga, T., Ohishi, T., Matsui, T., Saito, T., Matsumoto, M., Takasaki, J., Matsumoto, S., Kamohara, M., Hiyama, H., Yoshida, S., et al. (2005). Lysophosphatidylcholine enhances glucose-dependent insulin secretion via an orphan G-protein-coupled receptor. *Biochem. Biophys. Res. Commun.* 326, 744–751.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M.,

Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613.

Tan, M., Hao, F., Xu, X., Chisolm, G.M., and Cui, M.Z. (2009). Lysophosphatidylcholine activates a novel PKD2-mediated signaling pathway that controls monocyte migration. *Arterioscler. Thromb. Vasc. Biol.* 29, 1376–1382.

Taylor, K., Davey Smith, G., Relton, C.L., Gaunt, T.R., and Richardson, T.G. (2019). Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization. *Genome Med.* 11, 6.

Urs, S., Henderson, T., Le, P., Rosen, C.J., and Liaw, L. (2012). Tissue-specific expression of Sprouty1 in mice protects against high-fat diet-induced fat accumulation, bone loss and metabolic dysfunction. *Br. J. Nutr.* 108, 1025–1033.

Vaes, B.L., Lute, C., Blom, H.J., Bravenboer, N., de Vries, T.J., Everts, V., Dhonukshe-Rutten, R.A., Muller, M., de Groot, L.C., and Steegenga, W.T. (2009). Vitamin B(12) deficiency stimulates osteoclastogenesis via increased homocysteine and methylmalonic acid. *Calcif. Tissue Int.* 84, 413–422.

Vega, R.B., Huss, J.M., and Kelly, D.P. (2000). The coactivator PGC-1 cooperates with peroxisome proliferator-activated receptor α in transcriptional control of nuclear genes encoding mitochondrial fatty acid oxidation enzymes. *Mol. Cell. Biol.* 20, 1868–1876.

Wang, T., Wu, H., Liu, S., Lei, Z., Qin, Z., Wen, L., Liu, K., Wang, X., Guo, Y., Liu, Q., et al. (2018). SMYD3 controls a Wnt-responsive epigenetic switch for ASCL2 activation and cancer stem cell maintenance. *Cancer Lett.* 430, 11–24.

Wang, J., Yan, D., and Zhao, A. (2019). Discovery of potential biomarkers for osteoporosis using LC-MS/MS metabolomic methods. *Osteoporos. Int.* 30, 1491–1499.

Watanabe, N., Zmijewski, J.W., Takabe, W., Umezū-Goto, M., Le Goffe, C., Sekine, A., Landar, A., Watanabe, A., Aoki, J., Arai, H., et al. (2006). Activation of mitogen-activated protein kinases by lysophosphatidylcholine-induced mitochondrial reactive oxygen species generation in endothelial cells. *Am. J. Pathol.* 168, 1737–1748.

Winslow, M.M., Pan, M., Starbuck, M., Gallo, E.M., Deng, L., Karsenty, G., and Crabtree, G.R. (2006). Calcineurin/NFAT signaling in osteoblasts regulates bone mass. *Dev. Cell* 10, 771–782.

Wright, N.C., Looker, A.C., Saag, K.G., Curtis, J.R., Delzell, E.S., Randall, S., and Dawson-Hughes, B. (2014). The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine. *J. Bone Miner. Res.* 29, 2520–2526.

Wu, M., Chen, G., and Li, Y.P. (2016). TGF- β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* 4, 16009.

Yang, M., Guo, H., Wu, C., He, Y., Yu, D., Zhou, L., Wang, F., Xu, J., Tan, W., Wang, G., et al. (2009). Functional FEN1 polymorphisms are associated with DNA damage levels and lung cancer risk. *Hum. Mutat.* 30, 1320–1328.

Yao, C., Chen, G., Song, C., Keefe, J., Mendelson, M., Huan, T., Sun, B.B., Laser, A., Maranville, J.C., Wu, H., et al. (2018). Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* 9, 3268.

Yoon, S.H., Ryu, J., Lee, Y., Lee, Z.H., and Kim, H.H. (2011). Adenylate cyclase and calmodulin-dependent kinase have opposite effects on osteoclastogenesis by regulating the PKA-NFATc1 pathway. *J. Bone Miner. Res.* 26, 1217–1229.

Yu, B., Huo, L., Liu, Y., Deng, P., Szymanski, J., Li, J., Luo, X., Hong, C., Lin, J., and Wang, C.Y. (2018). PGC-1 α controls skeletal stem cell fate and bone-fat balance in osteoporosis and skeletal aging by inducing TAZ. *Cell Stem Cell* 23, 615–623.

Zhang, X., Dowd, D.R., Moore, M.C., Kranenburg, T.A., Meester-Smoor, M.A., Zwarthoff, E.C., and Macdonald, P.N. (2009). Meningioma 1 is required for appropriate osteoblast proliferation, motility, differentiation, and function. *J. Biol. Chem.* 284, 18174–18183.

Zhang, D., Jing, J., Lou, F., Li, R., Ping, Y., Yu, F., Wu, F., Yang, X., Xu, R., Li, F., et al. (2018). Evidence for excessive osteoclast activation in SIRT6 null mice. *Sci. Rep.* 8, 10992.

Zhou, Y., Deng, H.W., and Shen, H. (2015). Circulating monocytes: an appropriate model for bone-related study. *Osteoporos. Int.* 26, 2561–2572.

iScience, Volume 23

Supplemental Information

Multi-omics Data Integration for Identifying Osteoporosis Biomarkers and Their Biological Interaction and Causal Mechanisms

Chuan Qiu, Fangtang Yu, Kuanjui Su, Qi Zhao, Lan Zhang, Chao Xu, Wenxing Hu, Zun Wang, Lanjuan Zhao, Qing Tian, Yuping Wang, Hongwen Deng, and Hui Shen

Supplemental Information

Multi-omics Data Integration for Identifying Osteoporosis Biomarkers and Their Biological Interaction and Causal Mechanisms

Chuan Qiu, Fangtang Yu, Kuanjui Su, Qi Zhao, Lan Zhang, Chao Xu, Wenxing Hu, Zun Wang,
Lanjuan Zhao, Qing Tian, Yuping Wang, Hongwen Deng, Hui Shen

Supplemental Figures

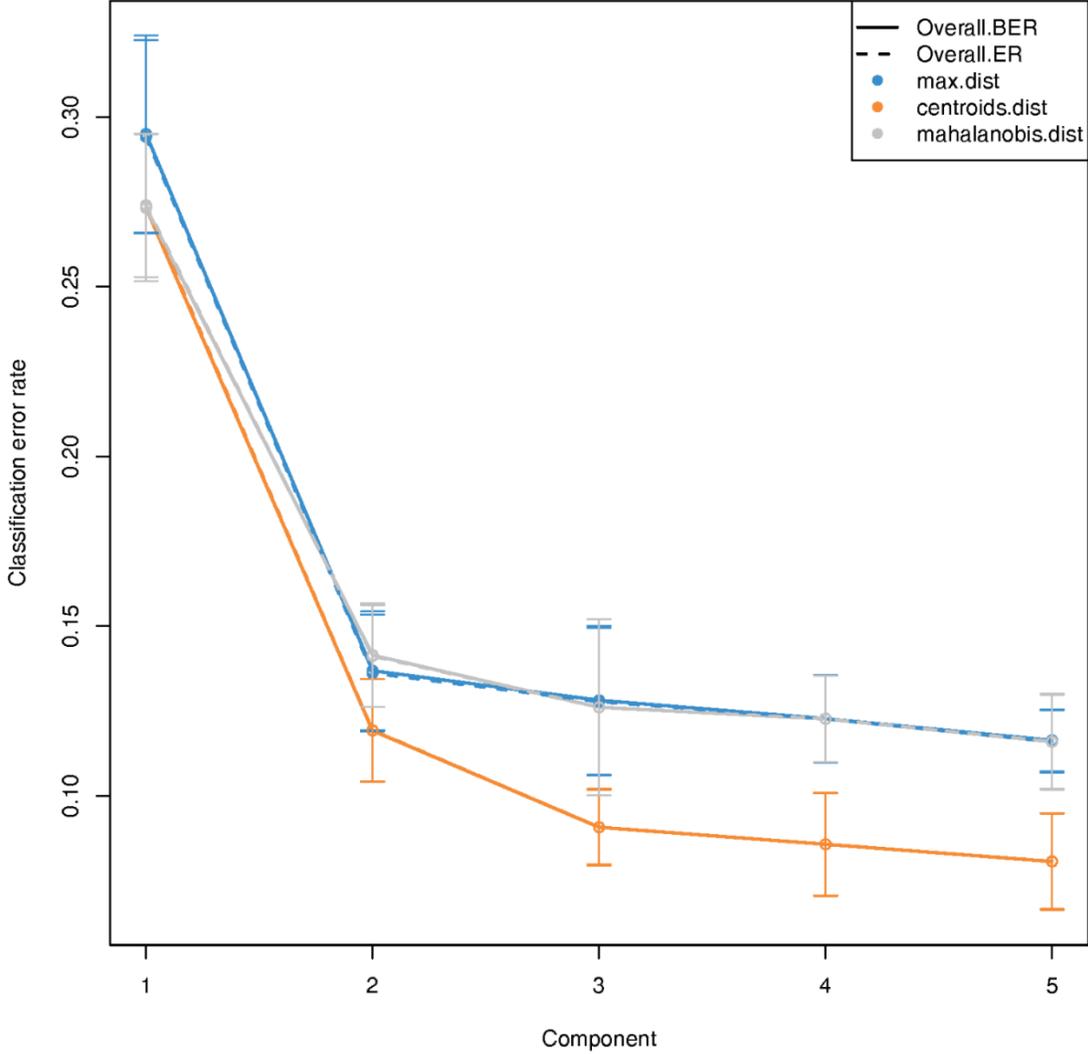


Figure S1. Optimization of multi-omics biomarker panel size. Related to Figure 2. The optimal multi-omics biomarker panel size was assessed by classification error rate using a 5-fold cross-validation repeated 5 times. The horizontal axis represents size of components and the vertical axis represents classification performance.

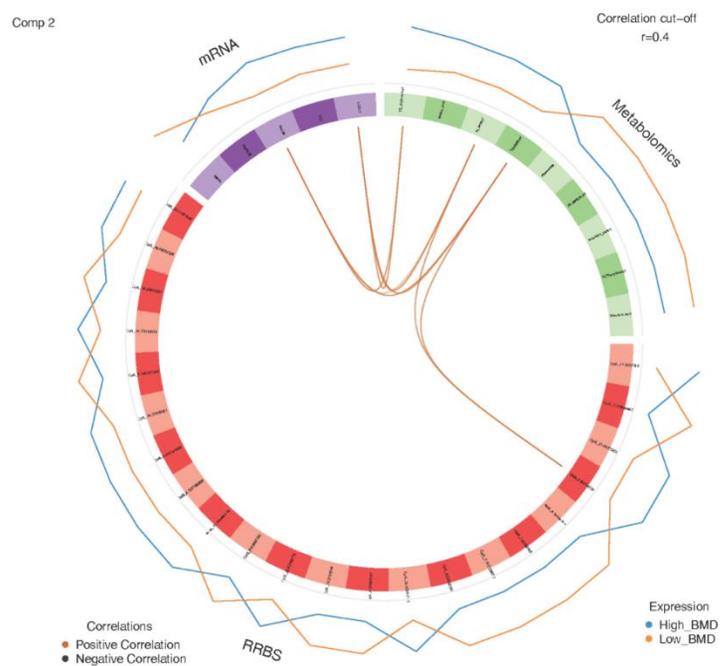
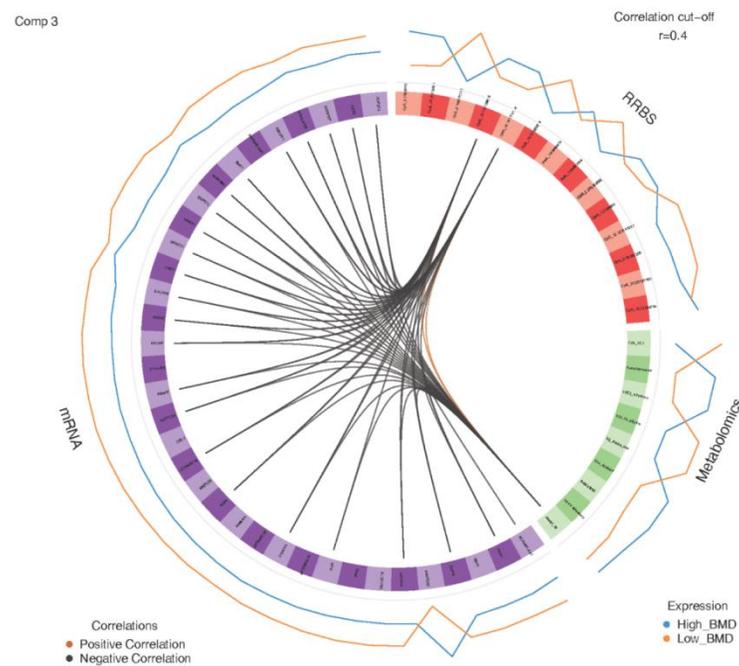
A**B**

Figure S2. Circos plot for multi-omics biomarkers. Related to Figure 2. Circos plot displays the different types of osteoporosis biomarker on a circle generated from the 2nd component (A) and the 3rd component (B), with links between omics indicating the positive (brown) or negative (black) correlations. The blue line and orange line represent high BMD and low BMD group, respectively.

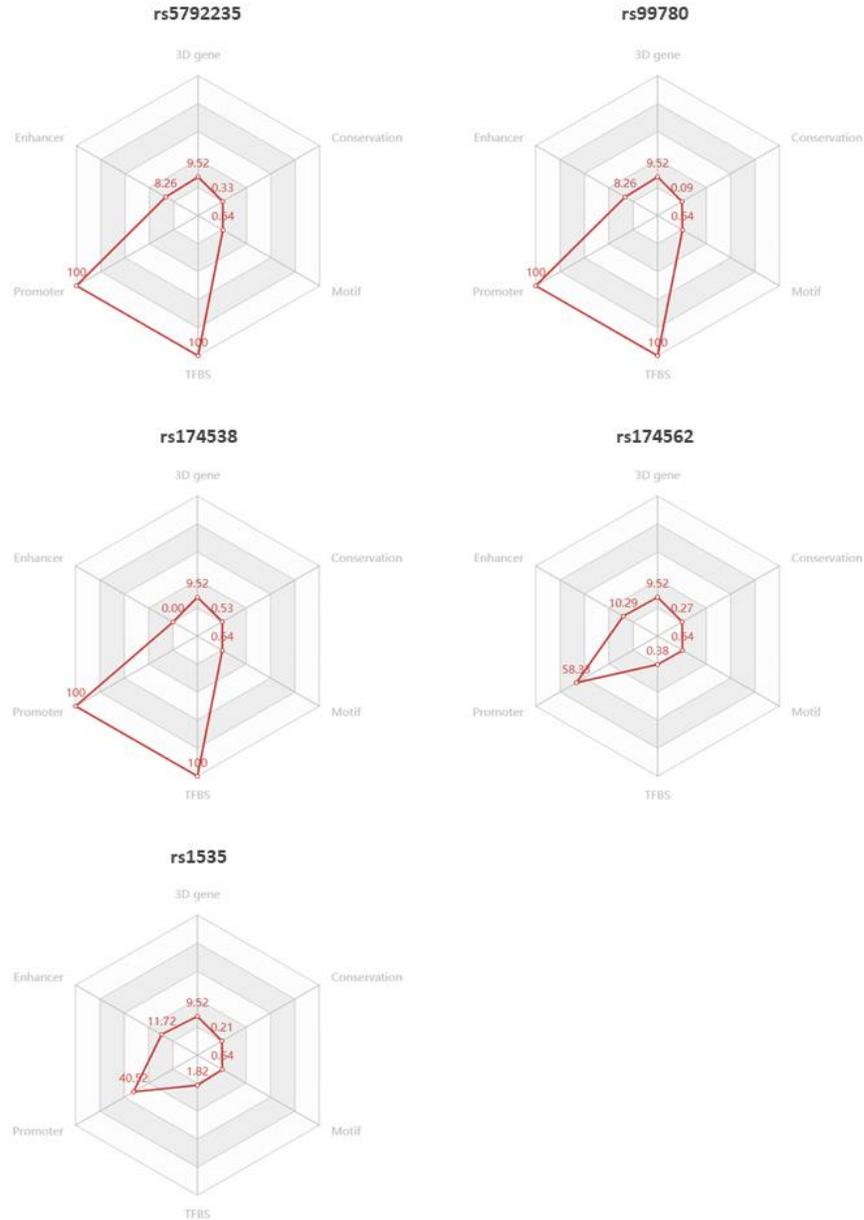
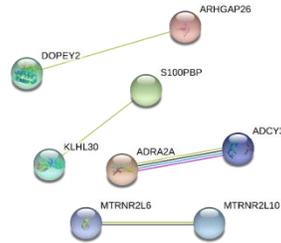


Figure S3. Radar charts of rs5792235, rs99780, rs174538, rs174562, and rs1535. Related to Figure 4. The six axes of the hexagon represent functionality levels (0-100) for enhancer status, promoter status, transcription factor binding site, motifs, evolutionary conservation, and 3D interacting genes, as suggested by 3DSNP.

A



B

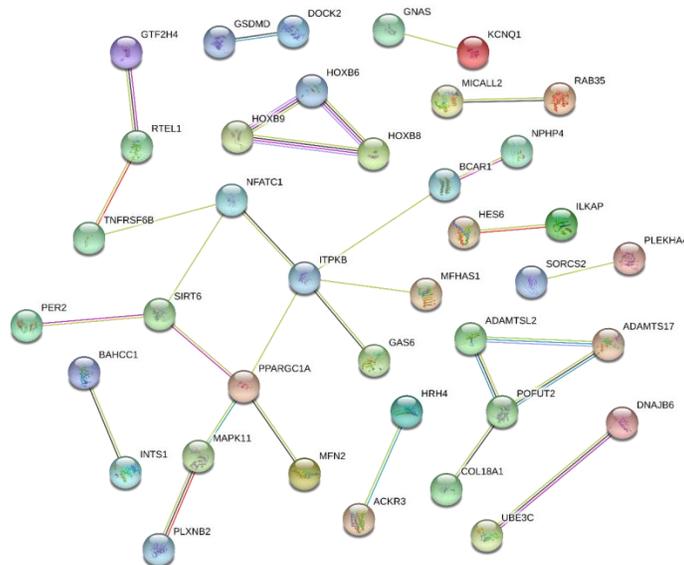


Figure S4. A functional interaction network for DEGs (A) and DMC-annotated genes (B). Related to Figure 5 and Figure 6. DMC was assigned to their nearest gene/gene-cluster based on the human reference genome assembly GRCh37 (hg19) from Genome Reference Consortium. Connections are based on co-expression and experimental evidence with a STRING v10.0 summary score above 0.4. Each filled node denotes a gene; edges between nodes indicate protein-protein interactions between protein products of the corresponding genes. Different edge colors represent the types of evidence for the association.

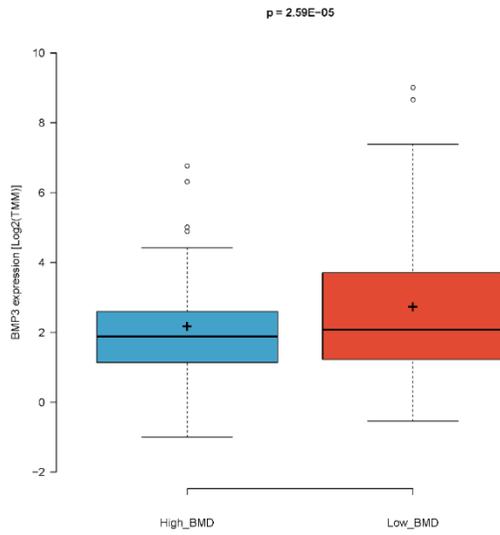
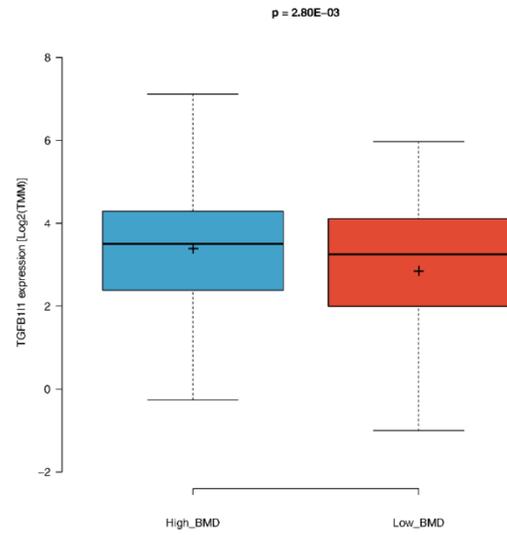
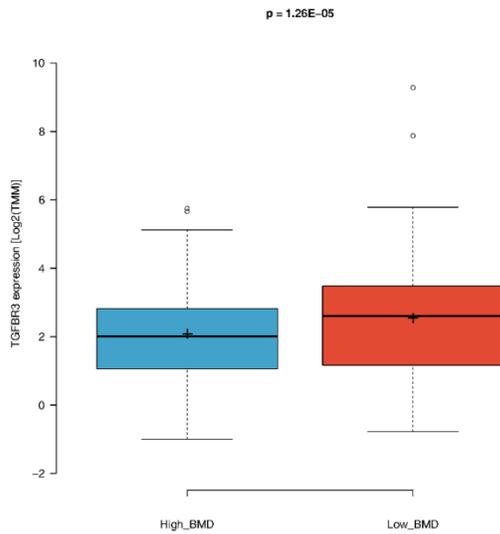
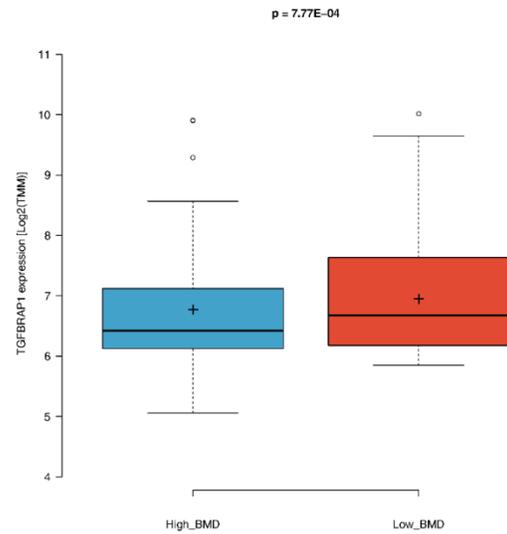
A**B****C****D**

Figure S5. Boxplot for BMP3 (A), TGFB11 (B), TGFBR3 (C), and TGFBRAP1 (D) expression in samples with different BMD status. Related to Figure 6. The horizontal axis represents high BMD and low BMD subgroups. The vertical axis represents the gene expression level [$\log_2(\text{TMM})$], p-values for moderated statistics determined by R package limma.

Supplemental Tables

Table S1. The Basic Characteristics of the Study Subjects. Related to Figure 1.

	Low BMD (n = 58)	High BMD (n = 61)
Age (years) \pm s.d.	31.72 (5.19)	32.23 (5.26)
Height (cm) \pm s.d.	164.37 (6.61)	165.45 (6.42)
Weight (kg) \pm s.d.	59.67 (8.42)	81.75 (25.53)

Note: BMD, bone mineral density; s.d., standard deviation.

Table S2. Multi-omics Osteoporosis Biomarkers Identified by Integration Analysis across the First Three Components. Related to

Figure 2.

	Gene Expression (74 DEGs)	DNA Methylation (75 DMCs)	Metabolites (23 DMPs)
1st Component	<p>35 DEGs: CPM, HAUS2, C1orf210, C15orf40, SRL, FADS2, SMIM11, LINC01021, SLC35E3, PRUNE2, ADCY3, EFR3B, MEIS3P1, FCF1, IM1, TMIGD2, GLB1L, DNASE1, PDSS2, TERF1, RABL3, ADRA2A, STX1A, RNF125, LUCAT1, KLHL30, CXorf36, BMP3, LOC643770, LOC100128288, GATA1, LOC100132077, SLC31A1, LINC00269, SCAMP1.AS1</p>	<p>40 DMCs: CpG_Ch17:79428036, CpG_Ch20:59224823, CpG_Ch6:43253387, CpG_Ch1:12107664, CpG_Ch2:237482267, CpG_Ch20:56292879, CpG_Ch1:5874307, CpG_Ch18:12306431, CpG_Ch16:33935546, CpG_Ch22:28157746, CpG_Ch19:38717239, CpG_Ch21:46784300, CpG_Ch9:136440143, CpG_Ch1:246277899, CpG_Ch4:124356866, CpG_Ch4:8546951, CpG_Ch8:8647499, CpG_Ch19:31411607, CpG_Ch4:7241128, CpG_Ch19:1526496, CpG_Ch1:6497309, CpG_Ch7:157098234, CpG_Ch20:29611712, CpG_Ch17:46697414, CpG_Ch21:37635103, CpG_Ch10:88263638, CpG_Ch18:77225621, CpG_Ch20:62328427, CpG_Ch19:4183272, CpG_Ch8:37699658, CpG_Ch10:77165382, CpG_Ch19:48902944, CpG_Ch15:100663758, CpG_Ch1:240656533, CpG_Ch13:114544118, CpG_Ch7:1452993, CpG_Ch22:50709855, CpG_Ch20:62328471, CpG_Ch8:37699673, CpG_Ch18:76495193</p>	<p>5 DMPs: Salicylic acid, N1,N8-Diacetylspermidine, Threonine, Heptanoylcarnitine, LysoPC(16:0)</p>
2nd Component	<p>5 DEGs: LLGL1, DBF4, PACS2, IYD, PNPLA3</p>	<p>22 DMCs: CpG_Ch7:62574705, CpG_Ch19:49342328, CpG_Ch7:153128979, CpG_Ch20:62601743, CpG_Ch17:36996953, CpG_Ch20:57431283, CpG_Ch1:226823075, CpG_Ch16:33953237, CpG_Ch11:106025516, CpG_Ch22:43166179, CpG_Ch21:44073202, CpG_Ch6:30882761, CpG_Ch8:144641513, CpG_Ch14:70039024, CpG_Ch16:2742246, CpG_Ch4:155703006, CpG_Ch6:30882767, CpG_Ch6:30882780, CpG_Ch11:2537259, CpG_Ch6:168107249, CpG_Ch2:239134803, CpG_Ch14:70039021</p>	<p>9 DMPs: Trans-2-Dodecenoylcarnitine, N-Phenylacetylphenylalanine, Tryptophan, Methylmalonic acid, N-lactoyl-Tryptophan, Isopropyl beta-D-glucoside, Nicotinic acid, 3-Methylcyclohexanethiol, PIPECOLATE</p>
3rd Component	<p>34 DEGs: FMN1, HLCS, METTL7A, NDUFV3, RABL2A, ZNF74, MTRNR2L10, MAN1C1, ANKRD22, MRPL10, UBIAD1, AHR, MTRNR2L6, CES2, RBM47, RBP1, SUPT7L, KRBA2, NSUN4, PPID, CSF2RA, ALDH3B1, ARHGAP26, DOPEY2, LOC646719, LIMD1, SIGLEC16, ARRDC3.AS1, GALNT6, DSC1, EMC1, S100BPB, MBOAT1, C11orf83</p>	<p>13 DMCs: CpG_Ch12:131189685, CpG_Ch12:120541569, CpG_Ch20:26190198, CpG_Ch4:24065208, CpG_Ch1:48021766, CpG_Ch12:131173748, CpG_Ch5:169137783, CpG_Ch5:1659948, CpG_Ch1:5786650, CpG_Ch1:236662741, CpG_Ch18:22098784, CpG_Ch16:75279661, CpG_Ch14:101864613</p>	<p>9 DMPs: Acetyl-T2 Toxin, Acetylcarnosine, CPA(18:1(11Z)/0:0), {4-[(E)-2-(2,3,5-trihydroxyphenyl)ethenyl]phenyl}oxidanesulfonic acid, 2-Amino-4-oxo-6-(1',2',3'-trihydroxypropyl)-diquinoid-7,8-dihydroxypterin, 13, 14-dihydro PGF2a, 5a-Androst-3-en-17-one, Indacaterol, 3,4,5-trihydroxy-6-[[1-(4-methoxyphenyl)pentan-3-yl]oxy]oxane-2-carboxylic acid</p>

Note: DEGs, differentially expressed genes; DMCs, differentially methylated CpG sites; DMPs, differential metabolic products

Table S3. Functional Annotation for DNA Methylation Biomarkers. Related to Figure 3A.

PC	CpG Sites	Closest Gene/Gene Cluster	Histone Mark	Gene Expression Status	DNaseI Status	Transcription Factor Status	Conservation by PhyloP (Mean)
1	CpG_Chrl:12107664	MFN2-TNFRSF8-TNFRSF1B	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	
1	CpG_Chrl:240656533	GREM2		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.93
1	CpG_Chrl:246277899	SMYD3	H3K27ac	Expression signal			-2.64
1	CpG_Chrl:5874307	NPHP4			DNaseI hypersensitivity cluster		-0.85
1	CpG_Chrl:6497309	ESPN		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-2.83
1	CpG_Chrl0:77165382	ZNF503-AS2		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.9
1	CpG_Chrl0:88263638	WAPAL		Expression signal	DNaseI hypersensitivity cluster		-1.03
1	CpG_Chrl3:114544118	GAS6	H3K27ac	Expression signal		Transcription factor ChIP-seq signal	-2.9
1	CpG_Chrl5:100663758	ADAMTS17		Expression signal			-4.43
1	CpG_Chrl6:33935546	NA			DNaseI hypersensitivity cluster		
1	CpG_Chrl7:46697414	HOXB9	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.3
1	CpG_Chrl7:79428036	BAHCC1		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	3.21
1	CpG_Chrl8:12306431	TUBB6	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.42
1	CpG_Chrl8:76495193	NA			DNaseI hypersensitivity cluster		-1.58
1	CpG_Chrl8:77225621	NFATC1	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	
1	CpG_Chrl9:1526496	PLK5		Expression signal		Transcription factor ChIP-seq signal	
1	CpG_Chrl9:31411607	NA					
1	CpG_Chrl9:38717239	DPF1	H3K27ac	Expression signal	DNaseI hypersensitivity cluster		-0.61
1	CpG_Chrl9:4183272	SIRT6			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.19
1	CpG_Chrl9:48902944	GRIN2D	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-0.54
1	CpG_Chrl2:237482267	ACKR3		Expression signal	DNaseI hypersensitivity cluster		-0.44
1	CpG_Chrl20:29611712	FRG1B	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	1.13
1	CpG_Chrl20:56292879	PMEPA1	H3K27ac		DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	1.13
1	CpG_Chrl20:59224823	LOC284757			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.07
1	CpG_Chrl20:62328427	RTEL1-TNFRSF6B	H3K27ac		DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.37
1	CpG_Chrl20:62328471	RTEL1-TNFRSF6B	H3K27ac		DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	
1	CpG_Chrl21:37635103	DOPEY2		Expression signal			0.19
1	CpG_Chrl21:46784300	COL18A1-POFUT2		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.42
1	CpG_Chrl22:28157746	MN1		Expression signal	DNaseI hypersensitivity cluster		-1.57

1	CpG_Ch22:50709855	MAPK11	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-0.93
1	CpG_Ch4:124356866	SPRY1					
1	CpG_Ch4:7241128	SORCS2		Expression signal			-1.82
1	CpG_Ch4:8546951	GPR78			DNaseI hypersensitivity cluster		-1
1	CpG_Ch6:43253387	TTBK1		Expression signal	DNaseI hypersensitivity cluster		
1	CpG_Ch7:1452993	INTS-MICALL2	H3K27ac		DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-0.8
1	CpG_Ch7:157098234	UBE3C-DNAJB6					-2.6
1	CpG_Ch8:37699658	GPR124			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.28
1	CpG_Ch8:37699673	GPR124			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.28
1	CpG_Ch8:8647499	MFHAS1	H3K27ac	Expression signal	DNaseI hypersensitivity cluster		-2.32
1	CpG_Ch9:136440143	ADAMTSL2		Expression signal			
2	CpG_Ch1:226823075	ITPKB	H3K27ac				
2	CpG_Ch11:106025516	AASDHPPT	H3K27ac				-1.9
2	CpG_Ch11:2537259	KCNQ1		Expression signal			-1.15
2	CpG_Ch14:70039021	CCDC177		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.6
2	CpG_Ch14:70039024	CCDC177		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.6
2	CpG_Ch16:2742246	KCTD5	H3K27ac	Expression signal			-2.24
2	CpG_Ch16:33953237	NA	H3K27ac		DNaseI hypersensitivity cluster		
2	CpG_Ch17:36996953	C17orf98					0.2
2	CpG_Ch19:49342328	PLEKHA4		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.2
2	CpG_Ch2:239134803	ILKAP-PER2		Expression signal	DNaseI hypersensitivity cluster		-2.27
2	CpG_Ch20:57431283	GNAS		Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.26
2	CpG_Ch20:62601743	ZNF512B-SAMD10	H3K27ac		DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-2.2
2	CpG_Ch21:44073202	PDE9A			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.86
2	CpG_Ch22:43166179	A4GALT-ARFGAP3	H3K27ac		DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	0.46
2	CpG_Ch4:155703006	RBM46					-5.44
2	CpG_Ch6:168107249	AL832737			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-1.1
2	CpG_Ch6:30882761	VARS2-GTF2H4	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	2.67
2	CpG_Ch6:30882767	VARS2-GTF2H4	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	2.67
2	CpG_Ch6:30882780	VARS2-GTF2H4	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	
2	CpG_Ch7:153128979	NA					
2	CpG_Ch7:62574705	NA			DNaseI hypersensitivity cluster		
2	CpG_Ch8:144641513	GSDMD	H3K27ac	Expression signal	DNaseI hypersensitivity cluster		2.03
3	CpG_Ch1:236662741	EDARADD-LGALS8					

3	CpG_Chrl:48021766	CMPK1					-0.54
3	CpG_Chrl:5786650	NPHP4					-4.55
3	CpG_Chrl2:120541569	RAB35	H3K27ac	Expression signal			-0.79
3	CpG_Chrl2:131173748	RIMBP2					-1.37
3	CpG_Chrl2:131189685	RIMBP2					
3	CpG_Chrl4:101864613	NA			DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	-2.7
3	CpG_Chrl6:75279661	BCAR1	H3K27ac	Expression signal	DNaseI hypersensitivity cluster	Transcription factor ChIP-seq signal	
3	CpG_Chrl8:22098784	HRH4			DNaseI hypersensitivity cluster		
3	CpG_Chrl20:26190198	LOC284801	H3K27ac		DNaseI hypersensitivity cluster		
3	CpG_Chrl4:24065208	PPARGC1A					-0.89
3	CpG_Chrl5:1659948	LOC728613			DNaseI hypersensitivity cluster		-1.39
3	CpG_Chrl5:169137783	DOCK2		Expression signal	DNaseI hypersensitivity cluster		

Table S6. Functional Importance of BMD-associated eQTLs for FADS2. Related to Figure 4.

SNPs	Functionality Score
rs5792235	218.76
rs99780	218.51
rs174538	210.69
rs174562	79.43
rs1535	64.43
rs174574	37.12
rs4246215	22.98
rs174564	20.29
rs174541	20.2
rs102275	16.77
rs102274	16.14
rs174554	14.82
rs7394579	13.38
rs174546	13.2
rs174550	13.17
rs174547	12.86
rs174535	12.47
rs174545	11.62
rs174536	11.15
rs174578	10.83
rs174566	10.75
rs174553	10.58
rs174567	10.4
rs57668028	10.34
rs174576	10.3
rs174533	10.11
rs174577	8.73
rs174537	8.23

TRANSPARENT METHODS

Subjects

A total of 119 unrelated Caucasian females, aged 20-40 years, were recruited through Louisiana Osteoporosis Study (LOS) (Du et al., 2017, Zhao et al., 2018), a repertoire of more than 16,000 subjects (by the end of August 2019) collected for genomic, transcriptomic, methylomic, metabolomic, and metagenomic studies of complex diseases/traits, particularly for osteoporosis. All the subjects were living in New Orleans, Louisiana and its surrounding areas and were self-identified as being of European origin. Hip BMD of each participant was measured by Hologic Discovery-A DXA (dual energy X-ray absorptiometry) machines as the combined BMD of the femoral neck, trochanter, and intertrochanteric region. The measurement precision, as reflected by coefficients of variation for hip BMD, was approximately 1.0%. The 119 subjects included 61 with relatively high hip BMD (Z scores ≥ 0.8) and 58 with relatively low hip BMD (Z scores ≤ -0.8). The BMD Z score was defined as the number of standard deviations a subject's BMD differed from the mean BMD of their age-, gender-, and ethnicity-matched population. Therefore, the high and low BMD subjects were belonging to the top and bottom 20% distribution of hip BMD in the corresponding population, respectively. For each study subject, weight and height were measured using standard procedures, and lifestyle factors (e.g., exercise, alcohol consumption, smoking, etc.) and medical history were assessed by questionnaires. An extensive set of subject exclusion criteria was adopted for this study to exclude subjects with diseases/conditions that may potentially lead to alteration of gene expression, DNA methylation, and metabolite pattern in blood. The basic characteristics of the exquisitely selected study subjects were summarized in **Table S1**. The stringent selection of the subjects was made feasible because of the large base LOS cohort. This study was approved by The Institutional Review Boards for Human Investigation at Tulane University (New Orleans, USA), and the signed informed-consent documents were obtained from all study participants.

Subjects Inclusion and Exclusion Criteria

Individuals must meet the following inclusion criteria to be eligible to participate in the study:

- 1) Caucasian females
- 2) Aged 20-40 years
- 3) With at least one intact ovary
- 4) Hip BMD Z-score ≥ 0.8 or ≤ -0.8

We adopted the following exclusion criteria to minimize non-genetic influence on bone mass variation, so as to empirically enhance the chance to detect individual genetic factors for bone mass.

- 1) Female subjects who are, or could be pregnant;
- 2) Serious residuals from cerebral vascular disease;
- 3) Diabetes mellitus, except for those controlled under medication;
- 4) Chronic renal failure;
- 5) Chronic liver failure;
- 6) Alcohol abuses as defined by those who drink alcohol regularly and cannot control themselves and get drunk at least once a week;
- 7) Chronic obstructive pulmonary disease (COPD);
- 8) Corticosteroid therapy at therapeutic levels for more than 6 months duration;
- 9) Treatment with anticonvulsant therapy for more than 6 months duration;
- 10) Other metabolic or inherited bone diseases including hyper- or hypoparathyroidism, Paget's disease, osteomalacia, osteogenesis imperfecta, and hypochondrogenesis;
- 11) Rheumatoid arthritis, except for minor cases that involve only hand joint and wrist;
- 12) Chronic gastrointestinal disease including celiac disease, postgastrectomy, Crohn's disease, ulcerative colitis, liver transplant, and cirrhosis;

Given that PBMs are essential component of the immune system, we adopted the following additional exclusion criteria in order to minimize the effect of diseases or conditions, which may potentially lead to the transcriptomic and epigenomic profiling changes:

- 1) Autoimmune or autoimmune-related diseases: systemic lupus erythematosus, multiple sclerosis, Graves disease, Hashimoto's thyroiditis, myasthenia gravis, Addison's disease, dermatomyositis, Sjogren's syndrome, Reiter's syndrome.
- 2) Immune-deficiency conditions: AIDS, severe malnutrition, splenectomy, ataxia-telangiectasia, DiGeorge syndrome, Chediak-Higashi syndrome, job syndrome, leukocyte adhesion defects, panhypogammaglobulinemia, selective deficiency of IgA, combined immunodeficiency disease, Wiscott-Aldrich syndrome.
- 3) Haematopoietic and lymphoreticular malignancies: leukaemias, lymphomas (Hodgkin's disease, non-Hodgkin's disease), myeloma, Waldenström's macroglobulinaemia, heavy chain disease, leukemic reticuloendotheliosis, mastocytosis, malignant histiocytosis.
- 4) Other diseases: influenza (within one week of recruitment), active periods of asthma.

WGS Analysis

DNA for WGS was extracted from the whole blood using the Genra Puregene Blood Kit (Qiagen, USA). Concentration and quality of the entire extracted DNA were assessed using Nanodrop 1000 and the samples were kept at -80 °C until further use. Libraries for WGS were prepared with KAPA DNA LTP library preparation kit (KAPA Biosystem, USA) on Biomek FX Laboratory Automation Workstation (Beckman Coulter, USA). Briefly, 300 ng genomic DNA was used as input. The workflow consists of fragmentation of double stranded DNA, end repair to generate blunt ends, A-tailing, adaptor ligation and PCR amplification. Different adaptors were used for multiplexing samples in one lane. Library concentrations and quality were measured using Qubit ds DNA HS Assay kit (Life Technologies,

USA) and Agilent TapeStation (Agilent, USA). WGS was conducted on Illumina HiSeq X-Ten with 150 bp paired-end reads. Data quality check was done on Illumina sequencing analysis viewer (SAV). Sequence reads were trimmed using Cutadapt (version 1.11), aligned to the human reference genome assembly GRCh37 (hg19) using BWA-MEM (version 0.7.12-r1039), duplicates marked with Picard (version 1.129, <http://picard.sourceforge.net>) and coordinates sorted using Samtools (<http://samtools.sourceforge.net>, version 1.3). Single nucleotide variants (SNV) were detected using a dual calling strategy through qSNP (Kassahn et al., 2013) and GATK HaplotypeCaller (McKenna et al., 2010). Variants were annotated with Ensemble v75 gene feature information. Variants were considered “called” and used in subsequent analysis if they passed the following filters: a minimum read depth of 8 reads in each dataset; at least 4 reads containing the variant where the variant was identified on both strands and not within the first or last 5 bases. Variants that did not pass these filters were considered “low evidence” and discarded.

Isolation of Monocytes, Their Genomic DNA, and Total RNA

In the present study, we focused specifically on PBMs, which can act as osteoclast precursors and play important roles in regulating bone metabolism (Fujikawa et al., 1996, Lari et al., 2009). Briefly, peripheral blood mononuclear cells (PBMCs) were firstly separated from 60 ml freshly collected peripheral blood, by a density gradient centrifugation method using Histopaque-1077 (Sigma-Aldrich, USA). The PBMCs were washed repeatedly with 2 mM EDTA in PBS, before being dissolved in 0.5% BSA and 2 mM EDTA in PBS. PBMs were then isolated from the PBMCs with a Monocyte Isolation Kit II (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) according to the manufacturer's protocol. The kit depleted unwanted cells (such as T and B cells) from PBMCs, leaving PBMs free of the surface-bound antibody and beads with minimum disturbance. The isolated PBMs were visually checked for purity and counted under microscope. The genomic DNA used for RRBS and total RNA used for RNA-

seq were extracted from the freshly isolated PBMs with the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, USA) following the manufacturer's protocol and kept at -80 °C until further use.

Transcriptomic Analysis by RNA-seq

Libraries for RNA-seq were prepared with KAPA RNA Hyper kit with RiboErase (KAPA Biosystem, USA) as per the manufacturer's instructions. Briefly, 500 ng RNA was used as input. The workflow consists of rRNA removal, cDNA generation, and end repair to generate blunt ends, A-tailing, adaptor ligation and PCR amplification. Different adaptors were used for multiplexing samples in one sequencing run. Library concentrations and quality were measured using Qubit ds DNA HS Assay kit (Life Technologies, USA) and Agilent TapeStation (Agilent, USA). The libraries were pooled and diluted to 2 nM in EB buffer and then denatured using the Illumina protocol. The denatured libraries were diluted to 10 pM by pre-chilled hybridization buffer and loaded onto Illumina NextSeq 500 and run for 75 cycles using a single-read recipe according to the manufacturer's instructions. Data quality check was done on Illumina SAV. De-multiplexing was implemented with Illumina Bcl2fastq2 v 2.17 program. Sequencing reads were aligned to the human reference genome assembly GRCh37 (hg19) using TopHat (<https://ccb.jhu.edu/software/tophat/index.shtml>, version 2.0.13), allowing only for unique alignments. RefSeq transcript annotations were obtained from the UCSC Genome Browser (<http://genome.ucsc.edu/index.html>), and read fragments aligned to known exons were counted per gene using Htseq (https://htseq.readthedocs.io/en/release_0.11.1/, version 0.6.1p1). All the analyses were conducted at the gene level. The RNA-seq raw counts were normalized by the trimmed mean of M-values (TMM) method through the Bioconductor R package *edgeR* (Robinson et al., 2010). We calculate the counts per million of normalized data and then log₂ transform them into the standard format. The RNA-seq data was pre-processed for differential gene expression analysis. The data was filtered for low-expressed genes by removing genes with less than 1 count per million in more than 20 samples per

group. We applied the Bioconductor R package *limma* (Ritchie et al., 2015) with empirical Bayes moderation to detect DEGs between the high BMD and low BMD groups, with adjusted significance threshold of FDR = 0.05.

Epigenome-wide DNA Methylation Analysis

Epigenome-wide DNA methylation profiles were determined by RRBS according to previously published protocols (Chatterjee et al., 2012). Briefly, 100 ng genomic DNA isolated from PBMs was first digested overnight with MspI restriction enzyme (Thermo Scientific, USA), followed by end repair, adenylation and adapter ligation using NEXTflex Bisulfite-Seq Library Prep Kit and NEXTflex Bisulfite-Seq Barcodes (BioO Scientific, USA) with a modification of bead size selection to capture MspI fragments of 40-220 bp size. The resulting libraries were bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research Corp, USA), followed by 20 cycles of PCR amplification using the NEXTflex Bisulfite-Seq U+PCR Master Mix and NEXTflex Primer Mix (BioO Scientific, USA). Different adaptors were used for multiplexing samples in one lane. Library concentrations and quality were measured using Qubit ds DNA HS Assay kit (Life Technologies, USA) and Agilent Bioanalyzer (Agilent, USA). Purified and quantified libraries were pooled at 6 samples per sequencing lane and read by 1×50 bp on Illumina HiSeq3000. Data quality check was done on Illumina SAV. De-multiplexing was performed with Illumina Bcl2fastq2 v2.17 program and standard fastq files were trimmed with cutadapt v1.3. Trimmed reads were aligned to the human reference genome assembly GRCh37 (hg19) and methylation calling were obtained using Bismark v0.10.0 (Krueger and Andrews, 2011). Output files were reformatted using a custom script. DMCs between the high BMD and low BMD groups were identified using the Bioconductor R package *methylKit v1.8.1* (Akalin et al., 2012) for CpGs with $\geq 10X$ coverage in at least 40 samples per group. The significance threshold for DMCs was defined as adjusted $q < 0.05$ with more than 10% methylation difference.

Metabolomic Analysis

The LC-MS based metabolomics platform developed by Dr. Garrett's lab in the Southeast Center for Integrated Metabolomics at University of Florida was used to perform the metabolomic analysis. The detailed experimental procedures have been previously described (Liu et al., 2017, Zhao et al., 2018). Briefly, each frozen serum sample (100 μ L) after thawing at room temperature from -80 $^{\circ}$ C storage was mixed with 20 μ L internal standard mix followed by vortex mixing for 20 s. In our protein precipitation procedure, the extracted serum sample (25 μ L) was added 800 μ L of acetonitrile:acetone:methanol (8:1:1, v: v: v) and centrifuged at 20,000 xg for 10 min at < 10 $^{\circ}$ C to remove proteins. The supernatant (250 μ L) was then removed and placed into a new 1 mL Eppendorf tube and dried under a gentle stream of nitrogen (Organomation Associates, USA). For the removal of debris, the dried sample was reconstituted by mixing 100 μ L of 0.1% formic acid and four injection standards T-Boc amino acids in water, and placed in an ice bath for 10-15 min followed by centrifugation at 20,000 xg for 5 min at < 10 $^{\circ}$ C. The untargeted metabolomics profiling was conducted on a Thermo Q-Exactive Orbitrap mass spectrometer with Dionex UHPLC and autosampler. All the pre-processed samples were analyzed in positive and negative heated electrospray ionization with a mass resolution of 35,000 at m/z 200 as separate injections. Separation was achieved on an ACE 18-pfp 100 \times 2.1 mm, 2 μ m column with mobile phase A as 0.1% formic acid in water and mobile phase B as acetonitrile. The flow rate was 350 μ L/min with a column temperature of 25 $^{\circ}$ C. 4 μ L was injected for negative ions and 2 μ L for positive ions. A metabolomic data processing program, MZmine, was used for automatic peak detection, mass spectral deconvolution, peak alignment, filtering, baseline correction (Pluskal et al., 2010), and the metabolite identification was searched against the Human Metabolite Database (HMDB) 4.0 (Wishart et al., 2018) and an internal retention time library of over 600 compounds developed by Dr. Garrett's lab (Liu et al., 2017). We removed metabolites with missing rates $> 20\%$ or coefficients of variation $> 20\%$ from further analyses. Metabolites with missing rates $< 20\%$ were imputed by the R package *missForest*

(Stekhoven and Buhlmann, 2012). The log transformation and autoscale for each metabolite were implemented by using the R package *specmine* (Costa et al., 2016). The two samples t-test was applied to individual metabolite at the time to determine the DMPs for BMD with cutoff threshold of $p = 0.05$.

Multi-omics Integration with SMDCCA

Canonical correlation analysis (CCA) is a multivariate method for cross-data association detection. Given two omics data $X_1 \in \mathbb{R}^{n \times p}, X_2 \in \mathbb{R}^{n \times q}$, where n denotes the sample size and p, q denote the feature/variable sizes of X_1, X_2 respectively, CCA searches the optimal linear combination (or loading vectors $u_1 \in \mathbb{R}^{p \times 1}, u_2 \in \mathbb{R}^{q \times 1}$) of features in X_1, X_2 with the highest Pearson correlation. The formulation of CCA is given in Eq. (1).

$$(u_1^*, u_2^*) = \arg \max_{u_1, u_2} \frac{u_1^T X_1^T X_2 u_2}{\sqrt{u_1^T X_1^T X_1 u_1 u_2^T X_2^T X_2 u_2}} \quad (1)$$

subject to $\|u_1\|_2 = 1, \|u_2\|_2 = 1$.

Sparse CCA (Parkhomenko et al., 2009, Witten and Tibshirani, 2009) was proposed to overcome overfitting when data is of relatively small sample size but large feature size, in which case the algorithm of conventional CCA fails due to the singular covariance matrix. Sparse CCA solves the overfitting problem by enforcing certain regularization constraints, e.g., L_1 norm regularization, on loading vectors u_1, u_2 , as shown in Eq. (2).

$$(u_1^*, u_2^*) = \arg \max_{u_1, u_2} u_1^T X_1^T X_2 u_2 \quad (2)$$

subject to $\|u_1\|_2 \leq 1, \|u_2\|_2 \leq 1, \|u_1\|_1 \leq \lambda_1, \|u_2\|_1 \leq \lambda_2$, where λ_1, λ_2 are tuning parameters controlling the sparsity of u_1, u_2 .

As a result of the sparsity of u_1, u_2 , sparse CCA can be used as a method for feature selection (for biomarkers). Sparse CCA can also be extended to sparse multiple CCA, or generalized CCA, to combine three or more omics data. The formulation of generalized sparse CCA is given as follows

$$\begin{aligned} (u_1, u_2, \dots, u_m) &= \arg \max_{u_1, u_2, \dots, u_m} \sum_{i < j} u_i^t X_i^t X_j u_j \\ \text{s.t. } &u_i^t u_i \leq 1, \|u_i\|_1 \leq \lambda_i, \forall i. \end{aligned} \quad (3)$$

In this study, we used a supervised CCA method, namely sparse multiple discriminative CCA (SMDCCA), both to incorporate the label information (such as high or low BMD group) and to exploit the complimentary discriminative information of multi-omics data. SMDCCA in the work by Le Cao and her colleagues (Singh et al., 2019, Rohart et al., 2017), seeks the correlations between multi-omics and require the correlation to be phenotype related by incorporating label (such as group/disease) information into the model, as given by Eq. (4).

$$\begin{aligned} (u_1, u_2, \dots, u_m) &= \arg \max_{u_1, u_2, \dots, u_m} \sum_{i < j} u_i^t X_i^t X_j u_j + \sum_{i=1}^m u_i^t X_i^t Y \\ \text{s.t. } &u_i^t u_i \leq 1, \|u_i\|_1 \leq \lambda_i, \forall i. \end{aligned} \quad (4)$$

Where $Y \in \mathbb{R}^{n \times 1}$ is the phenotype data.

Here, multi-omics integration with SMDCCA were performed via the Bioconductor R package *mixOmics* (Rohart et al., 2017) by utilizing potential osteoporosis biomarkers (DEGs, DMCs, and DMPs) identified from single-omics analysis. Contrary to the previous omics integrative methods, such as concatenation and the ensemble approaches (Gunther et al., 2012, Liu et al., 2013), supervised SMDCCA analysis can identify biomarkers composed of highly correlated features across the different types of omics, by modeling relationships between the omics datasets. The optimal multi-omics biomarker panel was identified by utilizing a grid search approach where, for any given combination of variables, we assessed the classification performance using a 5-fold cross-validation repeated 5 times. A

circos plot was used to visualize the osteoporosis biomarkers in different omics types on a circle and their correlations between each omics. Those correlations were estimated by applying the latent components as a proxy (Gonzalez et al., 2012).

Targeted QTL Analysis

Targeted QTL analysis was conducted by the R package *Matrix eQTL* (Shabalín, 2012) for the identified DEGs, DMCs, and DMPs, respectively. Eligible variants were 9,265,832 autosomal SNPs with a minor allele frequency (MAF) > 0.01, genotype hard call rate > 0.95, and Hardy-Weinberg $p > 1.0 \times 10^{-6}$. Age, height, and weight were included as covariates. In addition, the first principal component of SNPs, viewed as continuous axes of variation that reflect genetic variation due to ancestry in the sample, was used to control for potential population stratification (Price et al., 2006). *Matrix eQTL* performs a separate test for each SNP and biomarker pair in each omics. SNPs showing a significant association (FDR < 0.05) with DEGs, DMCs, and DMPs were defined as eQTLs, meQTLs, and metaQTLs, respectively. To test whether genetic variant has effects on multiple biomarkers from different omics, we set the cutoff for statistical significance and included all potential eQTLs, meQTLs, and metaQTLs with $p < 5.0 \times 10^{-5}$.

Interaction Network, Functional Annotation and Classification Analysis

In an effort to explore the potential functional importance of the identified DNA methylation biomarkers, we evaluated the chromatin states, expression status, and DNaseI hypersensitivity cluster for regions containing the DMCs across a wide variety of cell line derived from ENCODE (V3) (Roadmap Epigenomics et al., 2015). In addition, to inspect the sequence evolutionary conservation (relating to functional importance) of DNA methylation biomarkers, we obtained PhyloP score for each DMC sites from the UCSC Genome Browser. As functional annotation and classification analyses were conducted on the gene level, we assigned each DMC to their nearest gene/gene-cluster, termed DMC-annotated

gene, based on the human reference genome assembly GRCh37 (hg19). To construct functional interaction network of the identified gene expression and/or DNA methylation biomarkers, we implemented the STRING (version 11.0) interaction analysis with default settings (Szklarczyk et al., 2019). STRING interaction analysis quantifies the genetic interaction based on physical and functional association, which was derived from five main data sources, e.g. genomic context predictions, high-throughput lab experiments, co-expression, automated textmining, and previous knowledge in databases. To further investigate the functional importance of the identified osteoporosis biomarkers, we conducted a comprehensive scientific literature review in PubMed using the names of the biomarkers combined with keywords of “osteoporosis”, “BMD”, “bone”, “PBMs”, “osteoblast”, “osteoclast”, and “pathway”, and assigned these biomarkers into different signaling pathways or functional activities that are critical in bone metabolism. Finally, we reconstructed the biological networks/pathways in gene expression, DNA methylation, and metabolite levels by incorporating STRING interaction network.

The QTL SNPs were annotated to candidate target genes using SNPnexus (Chelala et al., 2009) under NCBI RefSeq gene annotation system. We applied HaploReg (Ward and Kellis, 2012) to explore the potential functional importance of the identified QTL SNPs on regulatory chromatin states across diverse cell types, predicted effects on TF binding, and the effects on gene expression (eQTL hits). We also evaluated the sequence evolutionary conservation (PhyloP score) for each QTL SNP. Analysis of the LD around susceptibility QTL SNPs was performed by SNiPA (Arnold et al., 2015) with 1000 Genomes reference panel (Phase 3 v5 under European population). 3DSNP prioritization analysis (Lu et al., 2017) was used to retrieve the total functionality score of candidate SNPs.

MR Analysis

To further investigate the potential causal effects of the significant biomarkers on BMD variation, we first selected candidate biomarkers via functional annotation and classification analysis, then we

extracted QTL SNPs ($p < 5.0 \times 10^{-5}$) for each osteoporosis biomarkers and downloaded BMD summary statistics released from UK Biobank (Kemp et al., 2017), followed by LD assessment and data harmonization. By applying the multiple independent SNPs as instrumental variables, we conducted MR analysis using simple median, weighted median, IVW and MR Egger methods. In addition, we also assessed the causal effects of these functionally classified osteoporosis biomarkers among different omics. MR analysis was implemented through the R package *MendelianRandomization v0.3.0* (Yavorska and Burgess, 2017).

Data and Code Availability

All whole genome sequencing data, RNA sequencing data, reduced representation bisulfite sequencing data, and metabolite data in this manuscript has been deposited in dbGaP. The accession number for the data reported in this paper is dbGaP: phs001960.v1.p1.

Supplemental References

- AKALIN, A., KORMAKSSON, M., LI, S., GARRETT-BAKELMAN, F. E., FIGUEROA, M. E., MELNICK, A. & MASON, C. E. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*, 13, R87.
- ARNOLD, M., RAFFLER, J., PFEUFER, A., SUHRE, K. & KASTENMULLER, G. 2015. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*, 31, 1334-6.
- CHATTERJEE, A., RODGER, E. J., STOCKWELL, P. A., WEEKS, R. J. & MORISON, I. M. 2012. Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. *J Biomed Biotechnol*, 2012, 741542.
- CHELALA, C., KHAN, A. & LEMOINE, N. R. 2009. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*, 25, 655-61.
- COSTA, C., MARASCHIN, M. & ROCHA, M. 2016. An R package for the integrated analysis of metabolomics and spectral data. *Comput Methods Programs Biomed*, 129, 117-24.
- DU, Y., ZHAO, L. J., XU, Q., WU, K. H. & DENG, H. W. 2017. Socioeconomic status and bone mineral density in adults by race/ethnicity and gender: the Louisiana osteoporosis study. *Osteoporos Int*, 28, 1699-1709.
- FUJIKAWA, Y., QUINN, J. M., SABOKBAR, A., MCGEE, J. O. & ATHANASOU, N. A. 1996. The human osteoclast precursor circulates in the monocyte fraction. *Endocrinology*, 137, 4058-60.
- GONZALEZ, I., CAO, K. A., DAVIS, M. J. & DEJEAN, S. 2012. Visualising associations between paired 'omics' data sets. *BioData Min*, 5, 19.
- GUNTHER, O. P., CHEN, V., FREUE, G. C., BALSHAW, R. F., TEBBUTT, S. J., HOLLANDER, Z., TAKHAR, M., MCMASTER, W. R., MCMANUS, B. M., KEOWN, P. A., et al. 2012. A computational pipeline for the development of multi-marker bio-signature panels and ensemble classifiers. *BMC Bioinformatics*, 13, 326.
- KASSAHN, K. S., HOLMES, O., NONES, K., PATCH, A. M., MILLER, D. K., CHRIST, A. N., HARLIWONG, I., BRUXNER, T. J., XU, Q., ANDERSON, M., et al. 2013. Somatic point mutation calling in low cellularity tumors. *PLoS One*, 8, e74380.
- KEMP, J. P., MORRIS, J. A., MEDINA-GOMEZ, C., FORGETTA, V., WARRINGTON, N. M., YOULTEN, S. E., ZHENG, J., GREGSON, C. L., GRUNDBERG, E., TRAJANOSKA, K., et al. 2017. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet*, 49, 1468-1475.
- KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27, 1571-2.
- LARI, R., KITCHENER, P. D. & HAMILTON, J. A. 2009. The proliferative human monocyte subpopulation contains osteoclast precursors. *Arthritis Res Ther*, 11, R23.
- LIU, H., GARRETT, T. J., SU, Z., KHOO, C. & GU, L. 2017. UHPLC-Q-Orbitrap-HRMS-based global metabolomics reveal metabolome modifications in plasma of young women after cranberry juice consumption. *J Nutr Biochem*, 45, 67-76.
- LIU, Y., DEVESCOVI, V., CHEN, S. & NARDINI, C. 2013. Multilevel omic data integration in cancer cell lines: advanced annotation and emergent properties. *BMC Syst Biol*, 7, 14.
- LU, Y., QUAN, C., CHEN, H., BO, X. & ZHANG, C. 2017. 3DSNP: a database for linking human noncoding SNPs to their three-dimensional interacting genes. *Nucleic Acids Res*, 45, D643-D649.
- MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M., et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 20, 1297-303.
- PARKHOMENKO, E., TRITCHLER, D. & BEYENE, J. 2009. Sparse canonical correlation analysis with application to genomic data integration. *Stat Appl Genet Mol Biol*, 8, Article 1.
- PLUSKAL, T., CASTILLO, S., VILLAR-BRIONES, A. & ORESIC, M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 11, 395.

- PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. & REICH, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, 38, 904-9.
- RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. & SMYTH, G. K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, 43, e47.
- ROADMAP EPIGENOMICS, C., KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature*, 518, 317-30.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-40.
- ROHART, F., GAUTIER, B., SINGH, A. & LE CAO, K. A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*, 13, e1005752.
- SHABALIN, A. A. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, 28, 1353-8.
- SINGH, A., SHANNON, C. P., GAUTIER, B., ROHART, F., VACHER, M., TEBBUTT, S. J. & LE CAO, K. A. 2019. DIABLO: an integrative approach for identifying key molecular drivers from multi-omic assays. *Bioinformatics*.
- STEKHOVEN, D. J. & BUHLMANN, P. 2012. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28, 112-8.
- SZKLARCZYK, D., GABLE, A. L., LYON, D., JUNGE, A., WYDER, S., HUERTA-CEPAS, J., SIMONOVIC, M., DONCHEVA, N. T., MORRIS, J. H., BORK, P., et al. 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47, D607-D613.
- WARD, L. D. & KELLIS, M. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*, 40, D930-4.
- WISHART, D. S., FEUNANG, Y. D., MARCU, A., GUO, A. C., LIANG, K., VAZQUEZ-FRESNO, R., SAJED, T., JOHNSON, D., LI, C., KARU, N., et al. 2018. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*, 46, D608-D617.
- WITTEN, D. M. & TIBSHIRANI, R. J. 2009. Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat Appl Genet Mol Biol*, 8, Article28.
- YAVORSKA, O. O. & BURGESS, S. 2017. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*, 46, 1734-1739.
- ZHAO, Q., SHEN, H., SU, K. J., TIAN, Q., ZHAO, L. J., QIU, C., GARRETT, T. J., LIU, J., KAKHNIASHVILI, D. & DENG, H. W. 2018. A joint analysis of metabolomic profiles associated with muscle mass and strength in Caucasian women. *Aging (Albany NY)*, 10, 2624-2635.