ARTICLE

An Osteoporosis Risk SNP at 1p36.12 Acts as an Allele-Specific Enhancer to Modulate *LINC00339* Expression via Long-Range Loop Formation

Xiao-Feng Chen,[1,2] Dong-Li Zhu,[1,2] Man Yang,[1] Wei-Xin Hu,[1] Yuan-Yuan Duan,[1] Bing-Jie Lu,[1] Yu Rong,[1] Shan-Shan Dong,[1] Ruo-Han Hao,[1] Jia-Bin Chen,[1] Yi-Xiao Chen,[1] Shi Yao,[1] Hlaing Nwe Thynn,[1] Yan Guo,[1,*] and Tie-Lin Yang[1,*]

Genome-wide association studies (GWASs) have reproducibly associated variants within intergenic regions of 1p36.12 locus with osteoporosis, but the functional roles underlying these noncoding variants are unknown. Through an integrative functional genomic and epigenomic analyses, we prioritized rs6426749 as a potential causal SNP for osteoporosis at 1p36.12. Dual-luciferase assay and CRISPR/Cas9 experiments demonstrate that rs6426749 acts as a distal allele-specific enhancer regulating expression of a lncRNA (*LINC00339*) (~360 kb) via long-range chromatin loop formation and that this loop is mediated by CTCF occupied near rs6426749 and *LINC00339* promoter region. Specifically, rs6426749-G allele can bind transcription factor TFAP2A, which efficiently elevates the enhancer activity and increases *LINC00339* expression. Downregulation of *LINC00339* significantly increases the expression of *CDC42* in osteoblast cells, which is a pivotal regulator involved in bone metabolism. Our study provides mechanistic insight into how a noncoding SNP affects osteoporosis by long-range interaction, a finding that could indicate promising therapeutic targets for osteoporosis.

Introduction

Genome-wide association studies (GWASs) have successfully identified numerous genetic variants for human complex diseases or traits. However, many of the identified variants are located in the noncoding regions of human genome.[1] It is particularly challenging to identify the precise gene targets for these noncoding variants and elucidate their functional mechanisms involved in disease pathophysiology. The traditional annotation of GWAS hits usually focuses on the nearest or most biologically plausible gene candidate, which may not be the true target gene and therefore might result in expensive and time-consuming efforts to explore the function of non-causal genes. Strikingly, recent studies have found that some of the noncoding GWAS SNPs are within potential regulatory or functional elements to regulate expressions of distal genes by long-range genome interactions.[2,3] For example, Gupta et al.[3] prioritized a functional variant (rs9349379) at 6p24 associated with five vascular diseases. They further validated that rs9349379 specifically regulates expression of *EDN1* (MIM: 131240), a long-range target gene (>600 kb) with known function on the vasculature.[3] These studies provide us promising insights into deciphering the relationship between noncoding SNPs and diseases. Addressing these knowledge gaps is critical to help translate GWAS findings into clinically useful information.

Osteoporosis (MIM: 166710) is one of the most common metabolic skeletal diseases characterized by low bone mass, poor bone quality, and an increased predisposition to fracture.[4] The clinical diagnosis and assessment of osteoporosis is mainly based on bone mineral density (BMD),[5] which has a high heritability of 0.6–0.8.[6] Previous GWASs have successfully identified more than 60 genetic loci for BMD and osteoporosis.[7,8] Some of these loci are mapped to genes with important function on bone, such as *RANK-RANKL-OPG* (*RANK* [MIM: 603499], *RANKL* [MIM: 602642], *OPG* [MIM: 602643]),[9,10] *ESR1* (MIM: 133430),[10] and *LRP5* (MIM: 603506).[11] However, some loci are localized to genes not known to have a role in bone biology. For example, 1p36.12 was identified by an initial large-scale BMD GWAS[10] and further replicated by multiple GWAS meta-analyses.[9,12–16] The reported SNPs within 1p36.12 are located in the noncoding region, indicating that they may reside within putative regulatory elements. The closest gene is *ZBTB40* (MIM: 612106), which is more than 60 kb away and has unknown function or connection with bone biology. Interestingly, the genes upstream of *ZBTB40*—*WNT4* (392 kb [MIM: 603490]) and *CDC42* (399 kb [MIM: 116952])—both have potential connection with bone or osteoporosis. WNT4 could attenuate bone loss in osteoporosis and skeletal aging mouse models by inhibiting nuclear factor-κB (NF-κB) via noncanonical Wnt signaling.[17] CDC42 is an effector molecule involved in bone metabolism and skeletal development.[18,19] Therefore, it is extremely interesting to find out the true target gene and investigate how the susceptibility SNPs at 1p36.12 affect disease risk.

In this study, we hypothesized that SNPs at 1p36.12 might act as distal regulatory element to influence the

[1]Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, P. R. China
[2]These authors contributed equally to this work
*Correspondence: yangtielin@mail.xjtu.edu.cn (T.-L.Y.), guoyan253@mail.xjtu.edu.cn (Y.G.)
https://doi.org/10.1016/j.ajhg.2018.03.001.

expression of target genes and modulate bone metabolism via long-range interaction. To achieve this aim, we implemented a series of computational analyses using data from sources including Hi-C (high-throughput 3C), expression quantitative trait locus (eQTL), and epigenomic annotation, then followed by various functional validation experiments, with the flowchart shown in Figure 1. We demonstrate that an intergenic SNP (rs6426749) at 1p36.12 acts as strong long-range enhancer to regulate expression of a long noncoding RNA (lncRNA, *LINC00339*) through chromatin loop formation. *LINC00339* could interact with *CDC42*, which is an important regulator involved in bone metabolism. Our findings provide a mechanistic basis for how a noncoding SNP affects osteoporosis by long-range interaction, which would be a potential and promising therapeutic target for osteoporosis.

Material and Methods

Study Subjects

The study was approved by the Institutional Review Board of School of Life Science and Technology at Xi'an Jiaotong University. Signed informed consent documents were obtained from all study participants before recruitment. We enrolled 1,300 unrelated Midwestern Chinese subjects of Han ethnicity from the city of Xi'an and its neighboring areas. The inclusion and exclusion criteria have been detailed in our previous publication.[20] BMD (g/cm^2) values at lumbar spine (LS) and femoral neck (FN) were measured with dual energy X-ray absorptiometry (DXA) using Hologic 4500 W machine (Hologic) that was calibrated daily.

Genotyping and Association Analysis

Genomic DNA was extracted from peripheral blood leukocytes using a commercial DNA isolation kit (Gentra systems) according to the protocol of the kit. For eight BMD-associated SNPs at 1p36.12 collected from the National Human Genome Research Institute (NHGRI) GWAS Catalog,[21] we conducted SNP genotyping using MALDI-TOF mass spectrometry on a MassARRAY system (Sequenom) with iPLEX assay. Genotype calling was performed in real time with MassARRAY RT software v.3.0.0.4 and analyzed using the MassARRAY Typer software v.3.4 (Sequenom). All eight SNPs for the Chinese cohort were successfully genotyped.

Before association analyses, we adjusted raw BMD values using significant covariates, including age, sex, height, and weight. Association analyses with BMD were conducted using linear regression model implemented with PLINK[22] assuming an additive inheritance model. For significant SNPs, we performed conditional association analysis by fitting each SNP genotype as a covariate and testing for secondary association on the remaining ones. We also checked the association signals in the GEFOS (Genetic Factors for Osteoporosis Consortium) dataset. GEFOS is the largest GWAS meta-analysis on BMD so far, including 17 GWASs and 32,965 individuals of 9 European cohorts.[16]

LD and Haplotype Analysis

BMD-associated SNPs at 1p36.12 were collected from the NHGRI GWAS Catalog[21] (October 2016). Linkage disequilibrium (LD) and haplotype analysis were conducted using Haploview v.4.2[23] in different populations (European, East Asian, and African) from the 1000 Genomes V3 genotype data.[24]

Hi-C and TAD Analysis

Hi-C data on IMR90 cells were collected from 4DGenome database.[25,26] Hi-C or capture Hi-C data on GM12878 and CD34 cells were obtained from several studies.[27,28] We also collected capture Hi-C data on 17 human primary blood cell types from a recently published large-scale genome-wide chromatin study.[29] DNase Hi-C data on human embryonic stem cells (H1-hESC) were retrieved from GEO database (GSE56869).[30] TAD data on IMR90 cells were acquired from GEO database (GSE35156).[31] The original ChIA-PET data and newly improved ChIA-PET data on six cell lines (K562, NB4, HCT-116, HeLa-S3, GM12878, and MCF-7)[32] were retrieved from the UCSC ENCODE download portal and GEO database (GSE72816), separately. All data used were summarized in Table S2. Bedtools[33] was used to extract our prioritized SNPs and/or genes within the same pair of Hi-C interaction regions. We only reported SNP-gene pairs within the same TAD region.

cis-eQTL Analysis

We obtained matched SNP genotyping and RNA-seq data for 462 unrelated human lymphoblastoid cell lines (LCLs) samples from ArrayExpress at EBML-EBI (ArrayExpress: E-GEUV-1). Since there is no direct osteoporosis-related human tissues with large sample size, we chose LCLs to explore the eQTL analysis. Previous studies have shown that there is a large overlap in the transcriptomic effects of genetic variation between human osteoblasts and LCLs,[34,35] and LCLs has been widely used as surrogate for eQTL analysis in genetic studies on osteoporosis.[36–39] *cis*-eQTL analysis was conducted between each of BMD SNPs and expression of nearby transcripts in a 1 Mb region using ANOVA test implemented in R software. Eta-squared (η^2) was calculated to measure the effect size [$\eta^2 = SS_{factor}/(SS_{factor} + SS_{error})$].

AseQTL Analysis

Matched SNP genotyping and allele-specific expression (ASE) data were collected from GTEx[40] in dbGap (phs000424.v6.p1). We conducted aseQTL analysis by testing for correlations between heterozygosity of rs6426749 and allelic imbalance at *LINC00339* or *CDC42* expression using Wilcoxon rank sum test. As described by Oldridge et al.,[41] we defined allelic fractions as min (A, B) / (A + B), where A or B is different alleles of a synonymous exonic SNP in the target gene.

Analysis of Shared Causal Genetic Variants

Two complementary methods were used to explore whether eQTL signal and GWAS BMD association were driven by the same causal variants. We downloaded recombination hotspot intervals as defined by McVean et al.[42] from HapMap web site and converted it to hg19 genome assembly using liftOver software. We used the regulatory trait concordance (RTC) test from Nica et al.[43] to distinguish between shared causal effects and coincidental overlaps, which is a rank-based score system testing for association between *cis*-eQTL and GWAS effect. GWAS SNP with the lowest genome-wide association (GWA) meta p value in GEFOS data with LS or FN BMD within corresponding recombination hotspot intervals was extracted to calculate RTC scores for the tested *cis*-eQTL SNP. For the joint likelihood mapping (JLIM) analysis,[44] we ran JLIM (v.1.0.2) with default parameters.

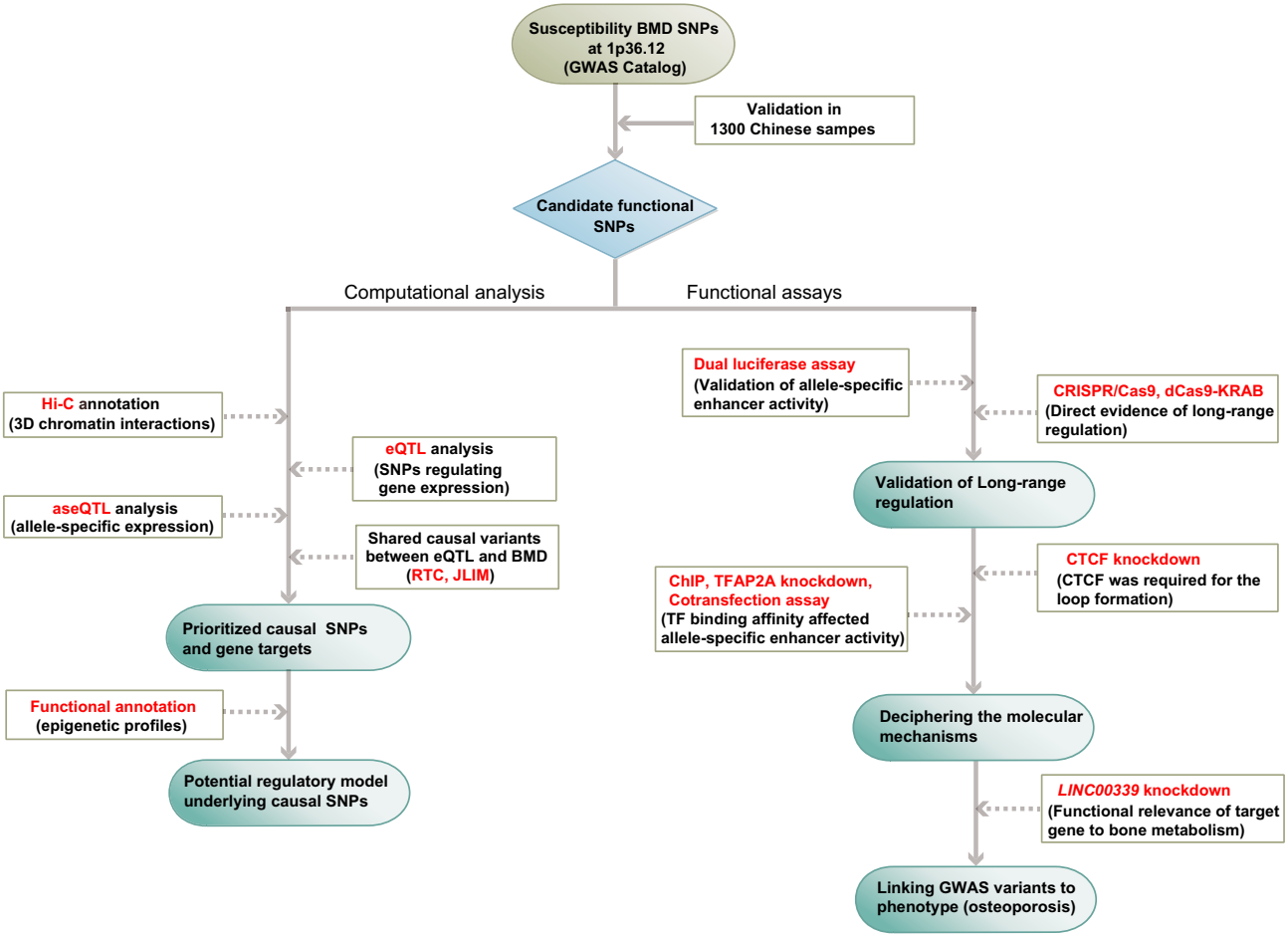

Figure 1. Flowchart for the Integrative Analyses Approach
Flowchart for the identification of functional BMD SNPs at 1p36.12 followed by experimental validation.

Functional Annotation

We annotated epigenetic regulatory features for SNPs and genomic regions of interest using ChIP-seq data from ENCODE, including CTCF insulator marks on six healthy cells (Osteoblast, GM12878, HUVEC, HMEC, H1-hESC, IMR90) and enhancer markers (H3k4me1, H3K4me3, H3K27ac, p300) on osteoblast cells. All data were displayed using WashU EpiGenome Browser.

Motif Analysis

The effect of rs6426749 on transcription factor binding motifs was analyzed using HaploReg v4.1[23] and MEME Suite toolkit[45] with TF motifs available from three public motif databases: JASPAR, HOCOMOCO, and SwissRegulon.[46–48] Motifs with at least three hits by different databases were reported. ChIP-seq data were retrieved from ENCODE and GEO database (GSE44257)[49,50] to validate the motif prediction.

Culture of Cell Lines

The hFOB 1.19 cells were obtained from the Institute of Biochemistry and Cell Biology of Shanghai (Shanghai, China) and cultured in DMEM/F12 supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, and 0.1 mg/mL streptomycin in 5% CO_2 at 37°C. The human embryonic kidney 293T cells (HEK293T) were purchased from the American Type Culture Collection (ATCC) and cultured in DMEM supplemented with 10% FBS, 100 U/mL penicillin, and 0.1 mg/mL streptomycin in 5% CO_2 at 37°C. All cell lines were free of mycoplasma.

Dual-Luciferase Reporter Assays and Site-Directed Mutagenesis

For rs6426749, rs34963268, or rs6684375, we chose a 991-bp, 944-bp, or 995-bp fragment surrounding each SNP as the putative enhancer element, separately. A 1,077-bp fragment upstream of *LINC00339* TSS was selected as the promoter for *LINC00339*. Both enhancer and promoter fragments were PCR amplified from human genomic DNA using the primers listed in Table S7. In order to obtain either the major or minor allele at three SNPs, site-directed mutagenesis was performed with the Quick Change II Site-Directed Mutagenesis Kit (Agilent Technology) according to the manufacturer's instruction, with primers listed in Table S7. All constructs were validated by sequencing and did not contain any other sequence variations. Constructs were co-transfected into hFOB 1.19 or HEK293T cells along with pRL-TK vector containing Renilla luciferase (Promega) using X-treme GENE HP DNA transfection reagent (Roche). After 48 hr of transfection, the cells were harvested and assayed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega), with Renilla luciferase (Rluc) reporter gene as the internal reference. Results were obtained from three independent experiments and each experiment was done in triplicate.

Enhancer Deletion and Repression

To delete enhancer fragment containing rs6426749 (749 bp) in hFOB 1.19 cells, the pCas9-dual sgRNA vector containing two sgRNAs was transfected into target cells by using Lipofectamine 2000 transfection reagent (Invitrogen). After selection with puromycin (1 mg/mL) for 1 week, the remaining cells were isolated as clones and verified by PCR sequencing. To repress enhancer activity surrounding rs6426749 in hFOB 1.19 cells, the pdCas9-KRAB vector and hU6 sgRNA vector containing distal sgRNA (sgRNA-1: 315 bp upstream rs6426749) or proximal sgRNA (sgRNA-2: 46 bp upstream rs6426749; sgRNA-3: 67 bp downstream rs6426749) were cotransfected into target cells by using Lipofectamine 2000 transfection reagent (Invitrogen). All sgRNA primers are listed in Table S7.

Genotyping of rs6426749

PCR-RFLP method was used to obtain genotype of rs6426749 in hFOB 1.19 cells and HEK293T cells. A 991-bp sequence centered on rs6426749 was first PCR amplified from human genomic DNA using primers the same with Luciferase Reporter assays (Table S7). The amplified DNAs were digested using the restriction enzyme (Sac I), which were subsequently subjected to the electrophoresis assay.

Chromatin Immunoprecipitation (ChIP) Assay

ChIP assays were performed in HEK293T cells with the Simple-ChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology) according to the manufacturer's instruction. In brief, approximately 3×10^7 cells were cross-linked with 1% formaldehyde for 10 min. After quenching with glycine solution, cells were rinsed, pelleted in cold PBS, and then resuspended and pelleted twice with buffer A and B, respectively. Micrococcal Nuclease (2,000 gel units/μL) was then added for nucleus digesting. After stopping digesting by EDTA, the nuclei fractions were pelleted by centrifugation, with sediment resuspended in ChIP buffer using protease inhibitor cocktail. The lysate was sonicated with the VirTis Virsonic 100 Ultrasonic Homogenizer/Sonicator for 3 pulses, after clarifying lysates by centrifugation, and the supernatant was collected. The supernatant containing sheared chromatin was immunoprecipitated with TFAP2a antibody (ab52222, Abcam) or normal immunoglobulin G (IgG) as a negative control and precleared with agarose beads. DNA protein complex was then precipitated with agarose beads, eluted from the beads, and reversely cross-linked by 5M NaCl and Proteinase K. The DNA fragments enriched in ChIP assays were purified for downstream RT-qPCR analysis, with primers listed in Table S7.

siRNA and shRNA Knockdown

siRNA knockdown experiments for *CTCF* (MIM: 604167) and *TFAP2A* (MIM: 107580) were conducted in hFOB 1.19 cells, separately. The siRNAs targeting *CTCF* or *TFAP2A* with related negative controls were synthesized by GenePharma. All siRNA sequences are listed in Table S7. Transfection of siRNAs was carried out in triplicate using the X-tremeGENE siRNA Transfection Reagent (Roche) according to the manufacturer's instructions. For *LINC00339* knockdown, we constructed the miR30-based short hairpin RNA (shRNA) expression vectors by using two oligonucleotides targeting *LINC00339*. These two oligonucleotides were connected with miR30 backbone and inserted into XhoI and EcoRI site of pcDNA3.1- plasmid (Invitrogen), with shRNA and negative control (NC) sequences shown as follows: shRNA-1: 5′-TGAGATCACTACCCAATGA-3′, shRNA-2: 5′-GACCTGATATC CACACAAA-3′, NC: 5′-GTTCTCCGAACGTGTCACGT-3′. Transfection was performed in triplicate according to the manufacturer's instructions. Briefly, hFOB 1.19 cells were seeded into 6-well plates and cultured to reach the confluence of 80%. Each well was transfected with 3 μg DNA using ViaFect Transfection Reagent (Promega). Cells were collected after 72 hr for further experiments.

Total RNA Isolation and Quantitative Real-Time PCR

Total RNA was isolated from cells using TRIzol reagent (Invitrogen), and complementary DNA (cDNA) was synthesized using the Super Scripts II First-Strand cDNA synthesis kit (Invitrogen). RT-qPCR was performed by BIO-RAD CFX Connect Real-Time System, with primers listed in Table S7. Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) was used as an internal control.

Mendelian Randomization Analysis

Two complementary Mendelian randomization (MR) methods were used to explore the causal relationship between *LINC00339* expression and BMD. The GWAS data on BMD were downloaded from a recent published large-scale GWAS.[51] The eQTL association data on whole blood were extracted from the GTEx database.[40] We first performed a summary-data-based Mendelian Randomization (SMR) analysis[52] (v.0.66) with default parameters which assigned the top *cis*-eQTL as instrumental variable. To exclude potential biased causal effect estimates deriving from invalid instrument variables, we also performed a multi-instrument based MR analysis using R package MendelianRandomization,[53] including inverse-variance weighted method and median-based method for causal test. The *cis*-eQTLs ($p < 0.01$) were first pruned for independence ($r^2 < 0.2$) by PLINK,[22] and the remaining *cis*-eQTLs were used as instrumental variables.

Results

Validation of GWAS SNPs in BMD Locus 1p36.12 in Chinese Population

1p36.12 was identified by an initial large-scale BMD GWAS[10] and replicated by multiple GWAS meta-analyses.[9,10,12–16] According to the GWAS Catalog,[21] there are eight SNPs at 1p36.12 reported by seven different GWASs.[9,10,12–16] Since most of GWAS samples are of European descent, we further examined associations of these eight SNPs with BMD in a Chinese cohort of 1,300 subjects. The basic characteristics of our sample are summarized in Table S1. The detailed association results are summarized in Table 1. These eight SNPs are all located in noncoding regions, which can be classified into three spatial clusters, including intron region of *WNT4* (rs3765350, rs2235529), intergenic region near *WNT4* (rs7521902 and rs3920498, more than 46 kb), and intergenic region near *ZBTB40* (rs7524102, rs34920465, rs6696981, and rs6426749, more than 67 kb) (Figure 2A). Four SNPs near *ZBTB40* were successfully validated for association with both lumbar spine (LS) and femoral neck (FN) BMD ($p < 0.05$, $\beta > 0$, Table 1). However, no significant signals were detected for the other four SNPs

Table 1. Association Results of Eight BMD SNPs at 1p36.12

Closest Gene/ Candidate	SNP	Position[a]	A1/A2[b]	Chinese Cohort					GEFOS					References[c]
				MAF	*P*-LS	β-LS	*P*-FN	β-FN	MAF	*P*-LS	β-LS	*P*-FN	β-FN	
WNT4 intron	rs3765350	22447316	G/A	0.280	0.546	−0.003	0.651	0.003	0.273	0.098	−0.018	0.004	−0.027	Kemp et al.[15]
	rs2235529	22450487	T/C	0.488	0.375	0.001	0.464	0.004	0.179	0.024	−0.028	0.003	−0.032	Kemp et al.[15]
WNT4 proximal	rs7521902	22490724	A/C	0.486	0.142	0.004	0.734	0.006	0.253	0.002	−0.034	6.60E−05	−0.037	Estrada et al.[13]
	rs3920498	22492887	C/G	0.450	0.475	−0.002	0.740	−0.003	0.217	2.91E−06	−0.054	1.07E−05	−0.043	Kemp et al.[15]
ZBTB40 proximal	rs7524102*	22698447	G/A	0.214	0.004*	0.015	0.013*	0.015	0.198	2.41E−14*	0.090	7.36E−17*	0.084	Styrkarsdottir et al.;[9,10] Rivadeneira et al.;[12] Estrada et al.;[13] Zhang et al.;[14] Zheng et al.[16]
	rs34920465*	22700351	G/A	0.213	0.003*	0.015	0.014*	0.015	0.216	1.57E−09*	0.084	1.81E−12*	0.080	Zhang et al.;[14] Zheng et al.[16]
	rs6696981*	22702858	T/G	0.210	0.005*	0.015	0.019*	0.014	0.143	9.02E−09*	0.080	4.78E−08*	0.065	Styrkarsdottir et al.;[9,10] Rivadeneira et al.;[12] Estrada et al.;[13] Zheng et al.[16]
	rs6426749*	22711473	C/G	0.211	0.006*	0.015	0.021*	0.014	0.192	1.15E−13*	0.088	5.90E−16*	0.082	Rivadeneira et al.;[12] Estrada et al.;[13] Zhang et al.;[14] Zheng et al.[16]

Significant SNPs and p values are indicated with an asterisk (*). Abbreviations: MAF, minor allele frequency; LS, lumbar spine BMD; FN, femoral neck BMD; Chinese cohort, 1,300 in-house Chinese cohort; GEFOS, Genetic Factors for Osteoporosis Consortium.

[a]Position is relative to the hg19 version of the human genome

[b]A1 is the minor allele according to 1000 Genomes

[c]Refs included the large-scale GWASs and meta-analysis for BMD and osteoporosis

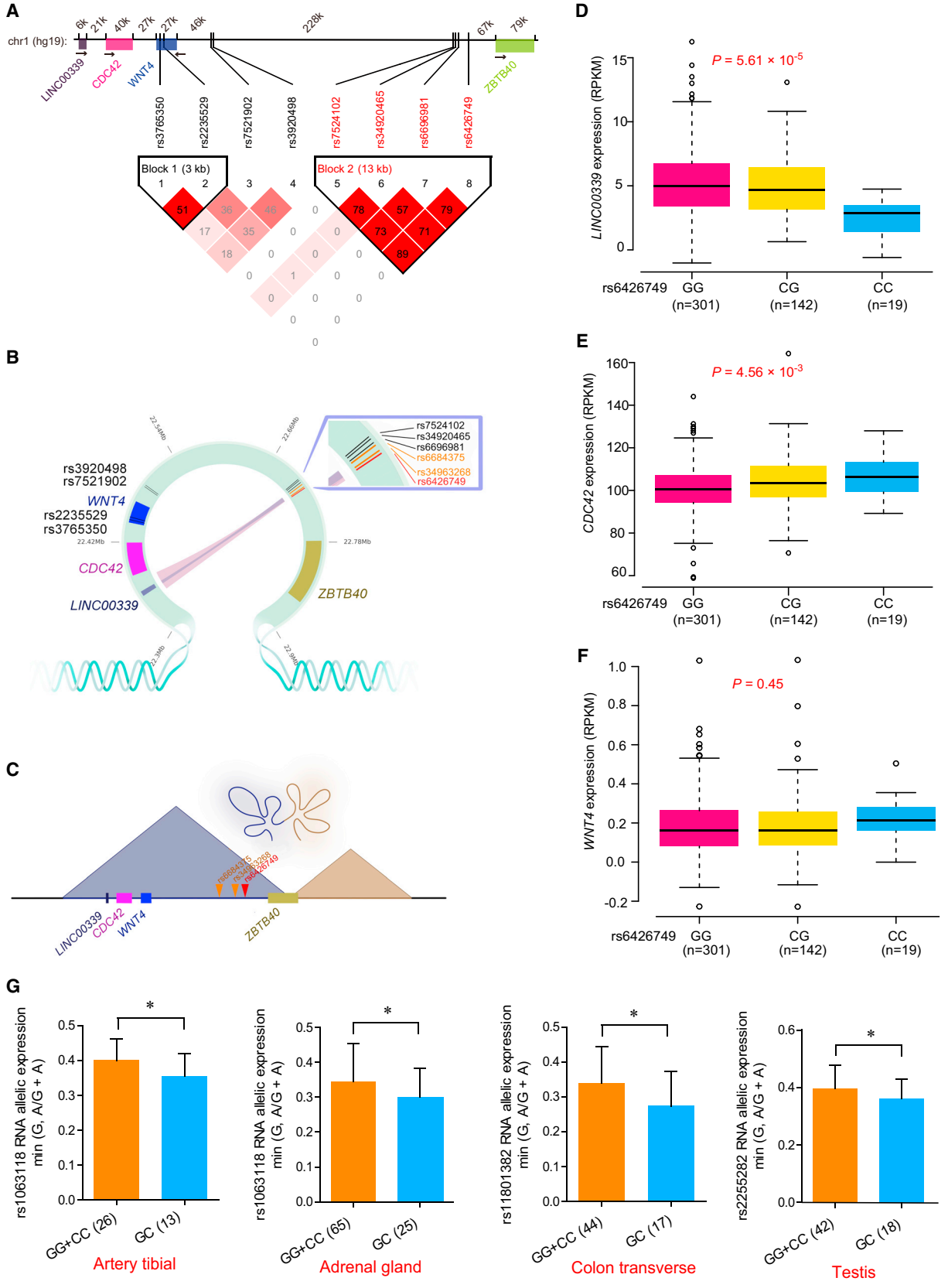

Figure 2. Integrating Analyses Indicate the Long-Range Interaction between rs6426749 and *LINC00339*

(A) LD blocks for eight BMD SNPs. The upper bar shows genomic positions for eight BMD SNPs in 1p36.12 and nearby genes, with distance between genes and (or) SNPs displayed above (kb). The bottom inverted triangle shows the LD blocks for eight BMD SNPs at

(legend continued on next page)

at 1p36.12. We also compared association signals with GEFOS meta-analysis dataset.[16] Consistently, only the four SNPs near *ZBTB40* achieved genome-wide significance level ($p < 5 \times 10^{-8}$) (Table 1). We performed LD and haplotype analyses. As shown in Figure 2A, two blocks in high LD were identified. The four SNPs near *ZBTB40* were in strong LD with each other ($r^2 > 0.7$) and belonged to one block. This block was highly conserved among diverse populations (European, East Asian, and African samples). Conditional analysis using any of these four SNPs as covariate obliterated association signals for the remaining three SNPs (data not shown), suggesting strong correlations of them.

Integrating Hi-C and eQTL Analyses to Identify Regulatory SNPs at 1p36.12 and Their Target Genes

To identify the potential gene targets and evaluate functional role of above noncoding BMD SNPs at 1p36.12, we investigated the long-range chromatin interactions surrounding them using various available Hi-C and ChIA-PET datasets (Table S2).[25–32] We identified seven candidate target genes for six SNPs (Table S3). To validate the predicted gene targets, we conducted *cis*-eQTL analyses using data from 462 unrelated human LCLs samples.[54] Through combining Hi-C and *cis*-eQTL results, we found that only rs6426749 fulfilled both criteria (Table S3), which had long-range chromatin interactions with *LINC00339* promoter in IMR90 cells[26] and H1-hESC cells[30] (Figure 2B), and this loop was located inside a conserved TAD with a size of 600 kb in IMR90 cells (Figure 2C). Moreover, rs6426749-G allele was significantly associated with increased *LINC00339* expression ($p = 5.61 \times 10^{-5}$, $\eta^2 = 0.042$) (Figure 2D and Table S4). Analysis of eQTL from GTEx project[40] further validated the association between rs6426749-G allele and increased *LINC00339* expression in LCLs ($n = 118$, $p = 0.02$, Figure S1A). However, there was no chromatin interaction between the other three SNPs in block 2 and *LINC00339*, implying that rs6426749 might be an independent regulatory SNP for *LINC00339* in this block. We noticed that rs6426749-G allele was also associated with decreased expression of *CDC42* in 462 LCL samples ($p = 4.56 \times 10^{-3}$, $\eta^2 = 0.023$, Figure 2E and Table S4),[54] and no significant association between rs6426749 and *WNT4* expression was found in either 462 LCL samples ($p = 0.45$, Figure 2E and Table S4)[54] or GTEx LCL tissue ($p = 0.76$, Figure S1C).[40] We also found some significant associations between rs6426749 and *CDC42* or *WNT4* expression in several other tissues from GTEx, which are shown in Figure S1.

A different 3D chromatin interaction loop might indicate independent regulatory circuitry affecting expression of target genes.[55] We found that compared with all other SNPs within the same Hi-C interaction region, rs6426749 showed the strongest eQTL association with *LINC00339* and that another two SNPs (rs6684375, rs34963268) in high LD with rs6426749 ($r^2 > 0.8$) had relative weaker eQTL association with *LINC00339* ($p = 4.25 \times 10^{-4}$, $\eta^2 = 0.033$) (Table S4). However, no secondary eQTL signals remained after adjusting residual effect of rs6426749 (Figure S2B), indicating that rs6426749 was the primary eQTL SNP within the Hi-C interaction region. We further performed conditional eQTL analysis for rs6426749 by adjusting the residual effect of each SNP within 1M region surrounding *LINC00339* (Figure S2A) and found that significant eQTL signal for rs6426749 on *LINC00339* was retained ($p < 0.05$, Figure S2C). Together, these data indicated the potential independent long-range regulation on *LINC00339* for rs6426749.

Validation of *cis*-eQTL Regulation on *LINC00339*

To further validate the *cis*-eQTL effect on *LINC00339*, we conducted allele-specific expression (ASE) analysis[56] for rs6426749 on *LINC00339* or *CDC42* expression using matched ASE data and genotype data from GTEx.[40] Individuals with heterozygous genotype for aseQTL should have more imbalanced ASE than those homozygous ones. As expected, we observed significantly higher imbalanced *LINC00339* expression in individuals heterozygous for rs6426749 (GC) than individuals homozygous for rs6426749 (GG + CC) in four different tissues ($p < 0.05$, Figure 2G), which provides strong independent validation of the *cis*-eQTL regulation of rs6426749 on *LINC00339*. However, no aseQTL effects on *CDC42* were detected for rs6426749, indicating that *CDC42* might not be the direct target. We also detected significant aseQTL effect on *LINC00339* instead of *CDC42* for rs6684375 and rs34963268 in six different tissues ($p < 0.05$, Figures S3A and S3B), further supporting the long-range *cis*-regulation on *LINC00339*.

1p36.12, with each diamond representing the r^2 measure of LD using standard color scheme, where the darker shades of red represent greater values.
(B) Hi-C interactions between eQTLs and promoters of target genes, and different color of lines indicated different Hi-C dataset (pink: Hi-C data on IMR90 cells from 4DGenome;[25] blue: DNase Hi-C data on H1-hESC cells[30]). SNP rs6426749 overlapped with Hi-C regions is labeled in red. Another two SNPs in strong LD with rs6426749 within the same Hi-C interaction regions are labeled in orange.
(C) The loop between rs6426749 and *LINC00339* is located within a 600 kb topologically associated domain (TAD) in IMR90 cells.
(D–F) Boxplot for *LINC00339* (D) or *CDC42* (E) or *WNT4* (F) expression in samples with different genotypes of rs6426749 (GG, CG, CC) taken from 462 LCLs samples.[54] Sample counts are shown.
(G) Allele-specific expression (ASE) analysis between rs6426749 and *LINC00339*, using monoallelic gene expression data from GTEx.[40] Four significant tissues ($p < 0.05$) are shown. The horizontal axis represents individuals with homozygous or homozygous genotypes for rs6426749. The vertical axis represents the exonic SNP chosen as a measurement of allelic expression of *LINC00339*. Error bars, SD; *$p < 0.05$ as determined by Wilcoxon rank sum test.

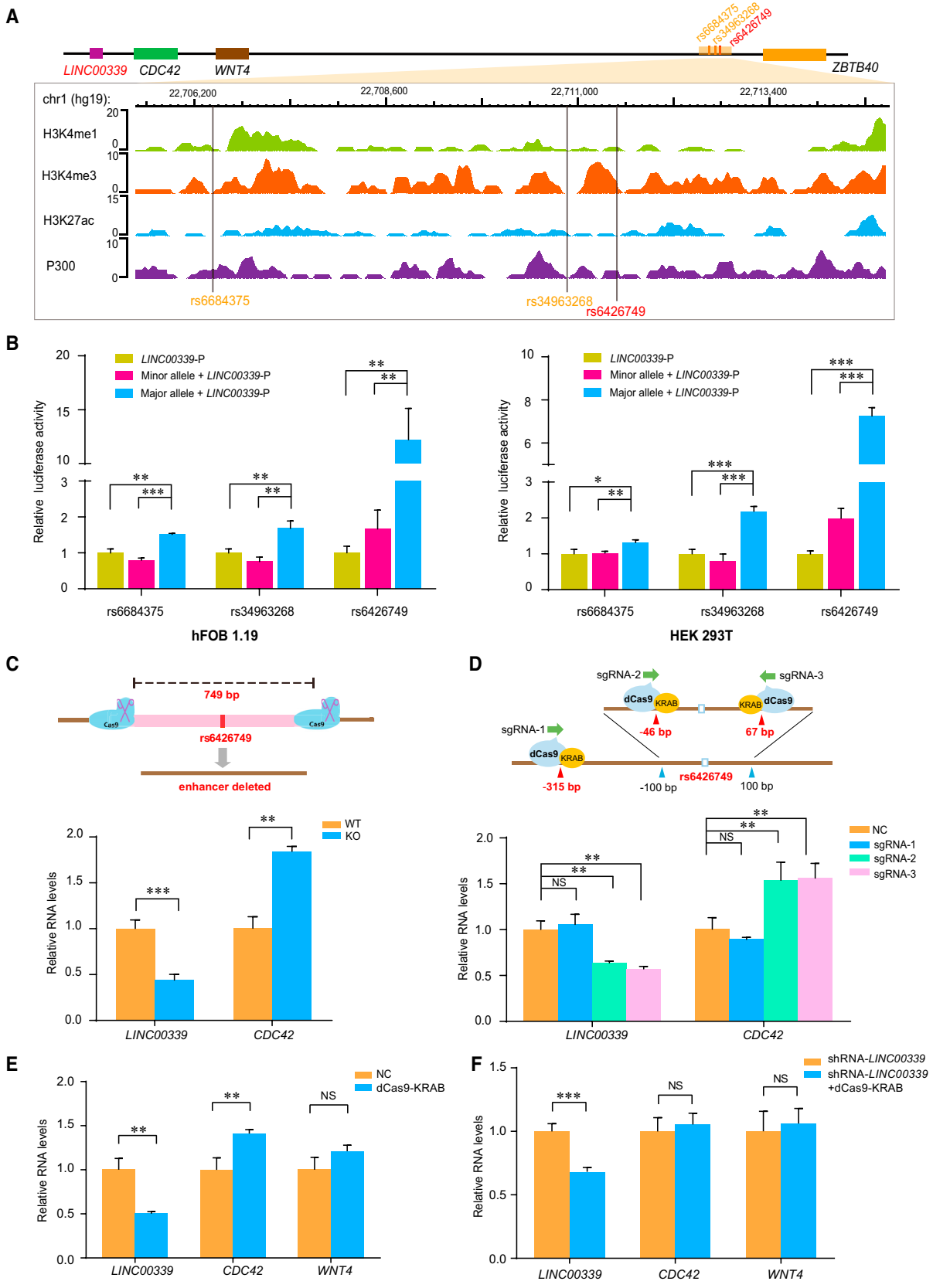

Figure 3. The Region Containing rs6426749 Acts as Strong Allele-Specific Enhancer on the *LINC00339* Promoter

(A) Epigenetic annotation for region surrounding rs6684375, rs34963268, and rs6426749 in osteoblast cells. The data are obtained from ENCODE Project taken from WashU EpiGenome Browser, including active histone modification (H3k4me1, H3K4me3, H3k27ac) as well as acetyltransferase (P300).

(legend continued on next page)

Evidence of Shared Causal Variants for BMD Association and *LINC00339* Expression

It is important to distinguish whether the overlap between GWAS signal and *cis*-eQTLs are coincidental or true shared causal variants. Therefore, we applied two different methods to identify whether associations with BMD and gene expression were driven by the same causal variants. The first is the regulatory trait concordance (RTC) method.[43] We detected high RTC scores for three eQTL SNPs (rs6426749: RTC = 0.97; rs6684375: RTC = 0.98; rs34963268, RTC = 0.98) with *LINC00339* expression, indicating strong evidence of shared causal effects between the eQTLs for *LINC00339* and the BMD GWAS SNPs at 1p36.12. Another method is the joint likelihood mapping (JLIM) analysis,[44] which could assess whether association signals between *cis*-eQTLs and BMD at 1p36.12 were due to the same underlying effect. We detected significant associations between eQTLs for *LINC00339* and BMD at LS ($p = 0.04$) or FN ($p = 0.04$), further supporting the functional relevance of eQTL regulation for *LINC00339* with BMD.

Evaluation of Allele-Specific Enhancer Activity for rs6426749

Noncoding regions of DNA may influence expression of distant genes by acting as enhancers to physically interact with target gene. Enhancers are identifiable by the presence of active epigenetic histone modifications, such as H3K4me1, H3K4me3, and H3K27ac, as well as co-activator and acetyltransferase (CBP/p300).[57] Therefore, we used publicly available ChIP-seq datasets from the ENCODE Project[58] to evaluate the potential regulatory function of regions around rs6426749, rs6684375, and rs34963268. We observed that the regions surrounding these three SNPs overlapped with many enhancer marks, including H3K4me1, H3K4me3, H3K27ac, and p300 in human osteoblast cells (Figure 3A).

To further validate the allele-specific enhancer activity for these three SNPs on target gene *LINC00339*, we cloned rs6684375, rs34963268, and rs6426749 locus with the major or minor allele of corresponding SNP, and inserted into a luciferase reporter vector, upstream of the *LINC00339* promoter, respectively. Upon transfection of these constructs into hFOB 1.19 cells, the major or minor allele of rs6426749 exhibited the greatest different effect on *LINC00339* promoter activity (Figure 3B). The rs6426749-G allele had significant increase in luciferase expression as compared with the rs6426749-C allele ($p < 0.005$, Fold = 7.32) or the *LINC00339* promoter-only construct ($p < 0.005$, Fold = 12.23) (Figure 3B). However, there was no significant difference between the minor allele of rs6426749 and the *LINC00339* promoter constructs. The consistent results were obtained in HEK293T cells (Figure 3B). In contrast, only modest increase in luciferase expression was detected between the major and minor allele of another two LD SNPs in hFOB 1.19 cells ($p < 0.001$, Fold = 1.91 for rs6684375; $p < 0.01$, Fold = 2.22 for rs34963268) or HEK293T cells ($p < 0.01$, Fold = 1.29 for rs6684375; $p < 0.001$, Fold = 2.70 for rs34963268) (Figure 3B). Together, our data demonstrated that rs6426749 could act as strong allele-specific functional enhancer for *LINC00339*.

Validation of Enhancer Activity for rs6426749 via CRISPR/Cas9 and dCas9-KRAB

Genotyping of rs6426749 revealed that HEK293T cells are heterozygous (G/C) and hFOB 1.19 cells are homozygous (G/G). To directly validate the long-range regulation between rs6426749 and *LINC00339*, we deleted a 749-bp enhancer region containing rs6426749 using CRISPR/Cas9 in hFOB 1.19 cells. As shown in Figure 3C, significantly decreased *LINC00339* expression ($p < 0.005$) while increased *CDC42* expression ($p < 0.01$) were detected in enhancer-deleted cells (KO) compared with the normal cells (WT), indicating that *LINC00339* was the direct target gene underlying distal enhancer-promoter regulation. To further validate the central role of rs6426749 in controlling enhancer activity, we designed two proximal sgRNAs (sgRNA-2: 46 bp upstream or sgRNA-3: 67 bp downstream) and one distal sgRNA (sgRNA-1: 315 bp upstream) targeting the rs6426749 locus using dCas9-KRAB in hFOB 1.19 cells, respectively. As shown in Figure 3D, we detected significantly reduced *LINC00339* expression ($p < 0.01$) while elevated *CDC42* expression ($p < 0.01$) on the proximal sgRNAs. However, the expression of *WNT4* was not changed by using the proximal sgRNA-3 in hFOB 1.19 cells ($p > 0.05$, Figure 3E), indicating that *WNT4* might not be the direct target of rs6426749. Moreover, we inhibited *LINC00339* expression using shRNA in hFOB 1.19 cells and then inhibited the enhancer region containing rs6426749 using dCas9-KRAB. As compared with the

(B) The dual-luciferase assay for *LINC00339* promoter (*LINC00339*-P) containing rs6684375, rs34963268, or rs6426749 locus with either the major or minor allele, or individual *LINC00339*-P was measured in hFOB 1.19 cells or HEK293T cells. The individual *LINC00339*-P was used as baseline control. Luciferase signal was normalized to *Renilla* signal. Error bars, SD. $n \geq 3$. $^{**}p < 0.01$, $^{***}p < 0.001$ as determined by an unpaired, two-tailed Student's t test.
(C) Comparison of *LINC00339* and *CDC42* expression between rs6426749 region deleted hFOB 1.19 cells (KO) mediated by CRISPR/Cas9 and normal cells (WT).
(D) Comparison of *LINC00339* and *CDC42* expression between rs6426749-locus repressed hFOB 1.19 cells using dCas9-KRAB and normal cells (NC, negative control). One distal sgRNA (sgRNA-1) and two proximal sgRNAs (sgRNA-2, sgRNA-3) were designed.
(E) Effect of rs6426749-locus repression in hFOB 1.19 cells using dCas9-KRAB (sgRNA-3) on *LINC00339*, *CDC42*, and *WNT4* expression.
(F) RT-qPCR for *LINC00339*, *CDC42*, and *WNT4* expressions in hFOB 1.19 cells after silencing of both *LINC00339* using shRNA and rs6426749-locus using dCas9-KRAB (blue) as compared with *LINC00339* silenced cells (orange), respectively. Error bars, SD. $n \geq 3$. NS: not significant, $^{**}p < 0.01$, $^{***}p < 0.001$ as determined by an unpaired, two-tailed Student's t test.

LINC00339 knockdown cells, we detected significantly decreased expression of *LINC00339* ($p < 0.01$) while no perturbation on either *CDC42* or *WNT4* expression ($p > 0.05$, Figure 3F), suggesting that *LINC00339* was the direct target gene of rs6426749, instead of *CDC42* or *WNT4*.

Analysis of Transcription Factor Binding at rs6426749 Region

We investigated the functional mechanism for rs6426749 underlying the strongest enhancer activity. The allele-specific activity of enhancer might be due to the different binding affinity of transcription factor (TF). We conducted motif analysis using multiple databases and identified two motifs, TFAP2A and TFAP2C of TFAP2 family, as candidate factors specifically binding to rs6426749-G (Figures 4A and S4A). RNA expression analysis revealed that *TFAP2A* is expressed much higher than *TFAP2C* in hFOB 1.19 cells (Figure S4B). ChIP-seq data from ENCODE[49] and GEO[50] identified that TFAP2A could bind to region surrounding rs6426749 in HeLa-S3 cells and MCF7 cells (Figure 4B), and the binding signal was much higher than the background in both HeLa-S3 cells (Fold = 2.9) and MCF7 cells (Fold = 3.1) (Figure S5). To experimentally verify the motif prediction, we performed ChIP-qPCR. Significant enrichment of TFAP2A binding was observed on the rs6426749 region compared with the negative control in HEK293T cells ($p < 0.001$, Figure 4C). We suppressed *TFAP2A* expression by siRNA in both HEK293T cells and hFOB 1.19 cells, which resulted in significant reduction of *LINC00339* expression ($p < 0.05$, Figure 4D). We further provided evidence of allele-specific binding affinity of TFAP2A using cotransfection assays: the TFAP2A knockdown diminished *LINC00339* expression in rs6426749-G allele, while it had no effect on rs6426749-C allele in hFOB 1.19 cells ($p < 0.001$, Figure 4E). Taken together, these data suggest that rs6426749 modulates TFAP2A binding to regulate *LINC00339* transcription.

CTCF Is Involved in Mediating Long-Range Chromatin Interaction between rs6426749 and *LINC00339*

CCCTC-binding factor (CTCF) is the best characterized insulator-binding protein, which is abundant in loop anchors and essential for loop formation and maintenance.[27,32,59,60] Using the annotation data from ENCODE, we found specific enrichment of CTCF binding at *LINC00339* promoter and rs6426749 nearby region (Figure 5A), which suggested that CTCF might play a role in mediating long-range loop interaction between rs6426749 and *LINC00339*. In this case, downregulation of *CTCF* could result in destruction of loop structure and decrease in the expression of target gene. Therefore, to validate the role of CTCF involved in the loop formation, we suppressed the expression of *CTCF* by siRNA in hFOB1.19 cells. As shown in Figure 5B, knockdown of *CTCF* significantly decreased the expression of *LINC00339* ($p < 0.05$) while it increased the expression of both *CDC42* ($p < 0.01$) and *WNT4* ($p < 0.05$), which means that CTCF is required for the loop formation to facilitate the regulatory element approaching and activating the expression of *LINC00339*.

LINC00339 Influences Bone Metabolism by Modulating Expression of *CDC42*

Next, we investigated the potential function for *LINC00339* expression involved in bone metabolism. We estimated the coding probability of *LINC00339* using the Coding Potential Assessment Tool.[61] The score was 0.0079 (with a score >1 indicating a potential coding gene), supporting the non-protein-coding nature of *LINC00339*. Using RNA expression data from GTEx Project[40] and FANTOM5 Project,[62] we found that *LINC00339* was ubiquitously expressed across all 54 various tissues and all 69 primary cells in human comparable levels (Figures S6A and S6B). The FANTOM5 Project defined significant trait-associated genes by systematically annotating susceptibility variants surrounding 59,110 genes.[62] We found strong association for *LINC00339* with bone resorption disease ($p = 3.0 \times 10^{-12}$) and abnormality of bone mineral density diseases ($p = 3.0 \times 10^{-12}$), suggesting the functional relevance of *LINC00339* involved in bone metabolism.

We have demonstrated the direct effect of rs6426749 on *LINC00339*. The above functional assays (including CRISPR/Cas9, dCas9-KRAB, and CTCF knockdown) all imply a negative correlation between *LINC00339* and *CDC42* expressions (Figures 3C–3E and 5B). Given that *CDC42* has been reported to play an important role in bone metabolism[18] and knockout of *Cdc42* in mouse results in severe skeletal abnormalities,[63,64] *LINC003339* might have potential regulatory correlation with *CDC42*. To verify this hypothesis, we conducted co-expression analysis using expression data from GTEx Project[40] and found that *LINC00339* was negatively correlated with *CDC42* in 12 tissues and positively correlated with *CDC42* in another 17 tissues ($p < 0.05$, Table S5). It was notable that more than 70% of positive tissues (12/17) were brain-related tissues, and *LINC00339* was expressed relatively much weaker (mean RPKM < 5) among more than 80% (14/17) of them. We inhibited the expression of *LINC00339* in hFOB 1.19 cells. Similarly, knockdown of *LINC00339* significantly increased the expression of *CDC42* ($p < 0.005$) while it had no effect on *WNT4* expression (Figure 5C), revealing that *CDC42* instead of *WNT4* was negatively regulated by *LINC00339* (Figure 5C). We further analyzed the chromatin interaction between *LINC00339* and *CDC42* using multiple Hi-C and ChIA-PET data (Table S2). Strong long-range interaction was observed between *LINC00339* and *CDC42* in six cells (K562, GM12878, H1-hESC, IMR90, MCF7, and HeLa-S3; Figure 5D and Table S6). Together, our results suggest that *LINC00339* might influence bone metabolism by modulating expression of *CDC42*.

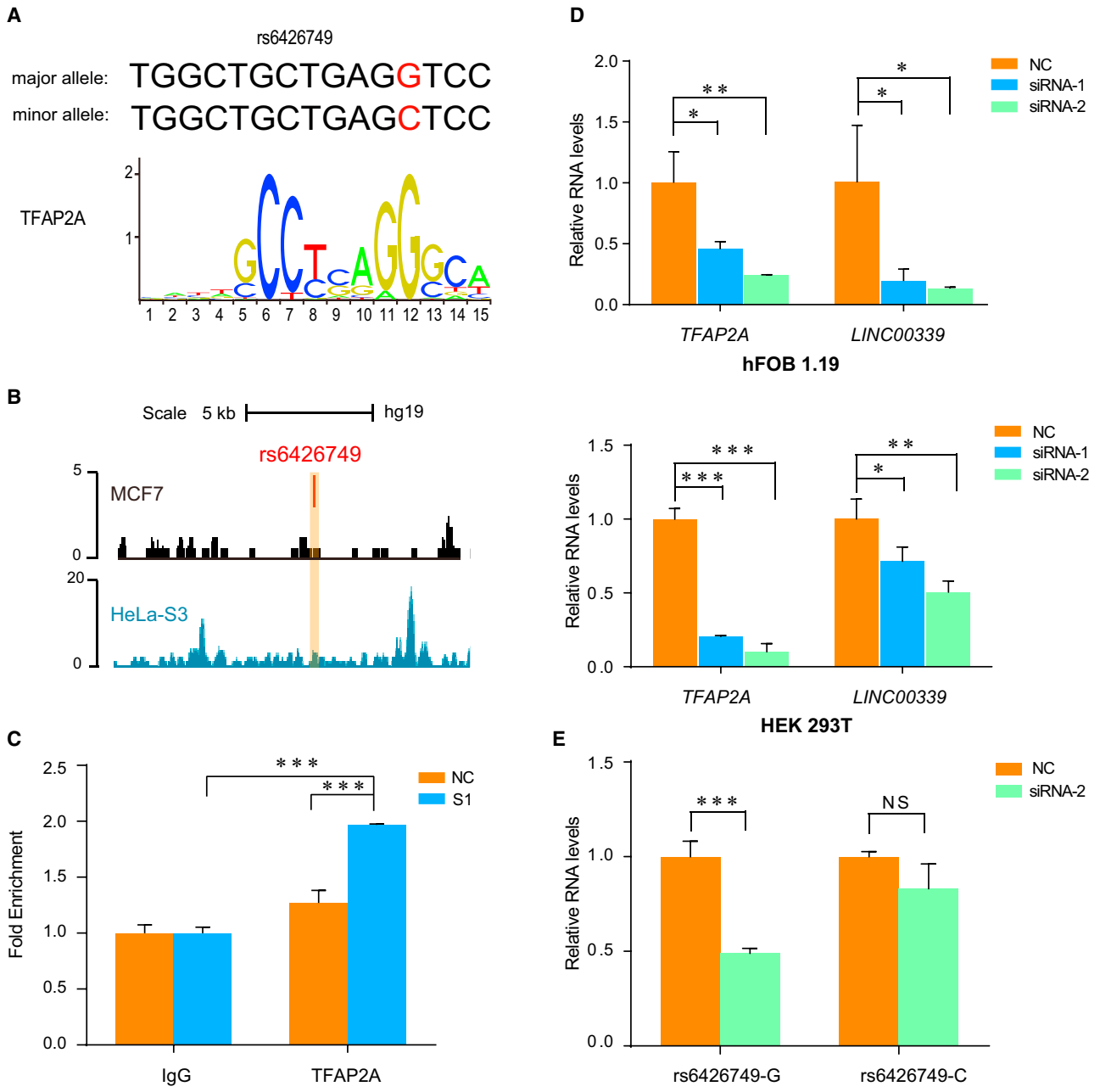

Figure 4. Identification of Transcription Factors Required for the Activity of Enhancer Containing rs6426749
(A) Motif analysis indicated that TFAP2A motif exclusively binds to G allele of rs6426749.
(B) TFAP2A binding surrounding rs6426749 was observed in MCF7 cells (GEO: GSE44257)[50] and HeLa-S3 cells (ENCODE Project, taken from UCSC Genome browser).
(C) ChIP-qPCR of TFAP2A binding at rs6426749 region and negative control region in HEK293T cells. Primers targeting rs6426749 region (S1) or *RPL30* exon (NC) are used. The binding of TFAP2A is shown as fold enrichment over IgG.
(D) The siRNA-mediated depletion of *TFAP2A* diminished *LINC00339* expression. RT-qPCR for *TFAP2A* and *LINC00339* expression in hFOB 1.19 cells or HEK293T cells after knockdown of *TFAP2A* (siRNA-1 and siRNA-2: two different siRNAs, blue and green) compared to NC siRNA-treated cells (NC: negative control, orange), respectively.
(E) The siRNA-mediated depletion of *TFAP2A* specifically diminished activity of enhancer containing rs6426749 on *LINC003339* expression. The pGL3 basic vector containing rs6426749-G (C) allele locus and *LINC00339* promoter (see also Figure 2B), as well as the TFAP2A silencer (siRNA-2) or negative control was cotransfected into the hFOB 1.19 cells. Error bars, SD. $n \geq 3$. NS: not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ as determined by an unpaired, two-tailed Student's t test.

Causal Relationship between *LINC00339* Expression and BMD

We applied Mendelian randomization (MR) analysis to characterize the causal association between *LINC00339* expression and BMD. The SMR analysis using the top *cis*-eQTL on *LINC00339* as instrumental variable (rs2255282) detected significant association between *LINC00339* expression and BMD ($P_{SMR} = 0.02$). To exclude

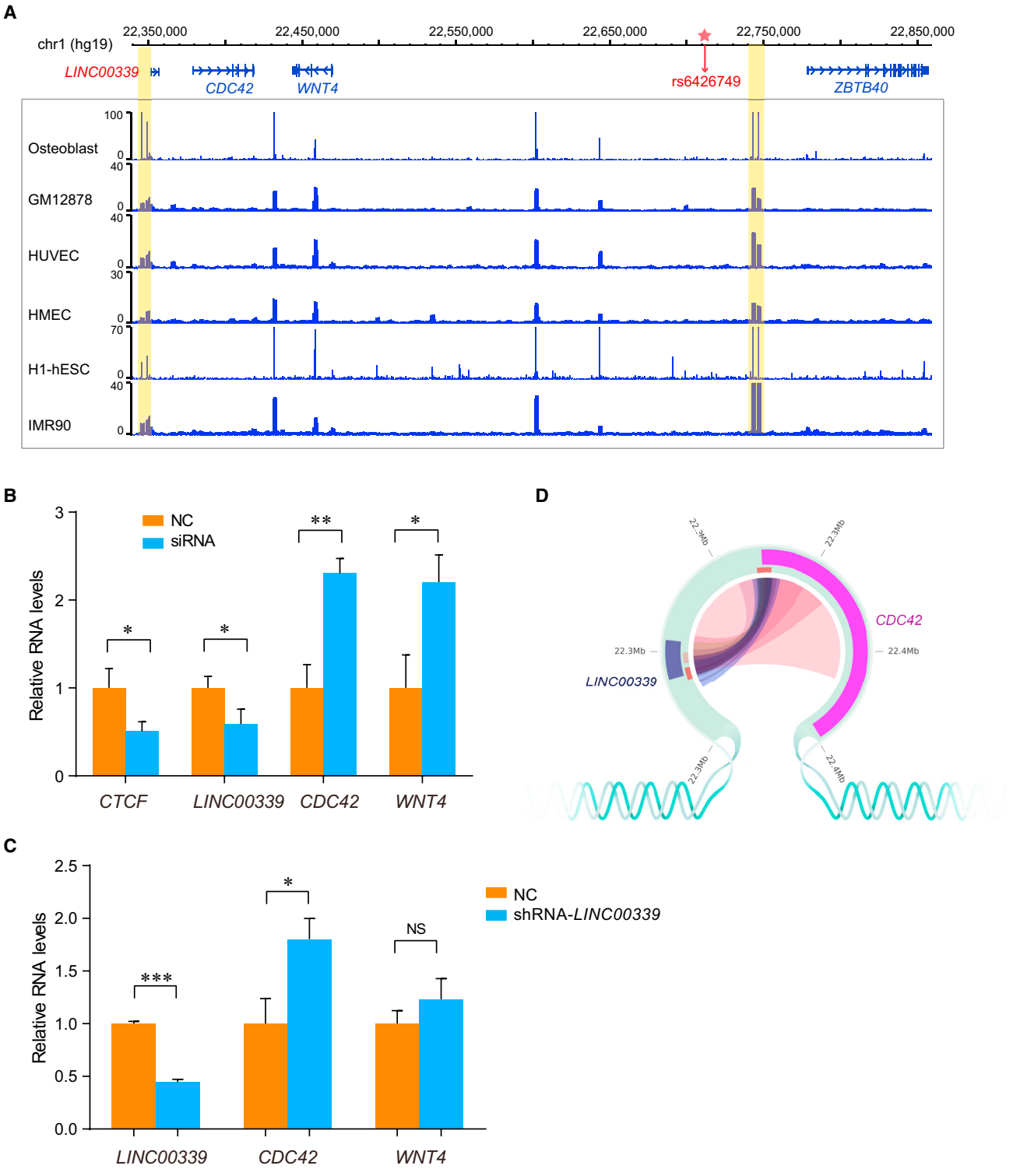

Figure 5. CTCF Modulated Long-Range Loop Formation between *cis*-eQTLs and *LINC00339* and *LINC00339* Negatively Regulated *CDC42*

(A) CTCF binding sites surrounding rs6426749 or *LINC00339* from six different healthy cells taken from WashU EpiGenome Browser are displayed, with focal peak regions highlighted by yellow colors.

(B) The siRNA-mediated depletion of *CTCF* diminished *LINC00339* expression while it elevated both *CDC42* and *WNT4* expression. RT-qPCR for *CTCF*, *LINC00339*, and *CDC42* expression in hFOB 1.19 cells after knockdown of *CTCF* (siRNA, blue) compared to NC siRNA-treated cells (NC: negative control, orange).

(C) RT-qPCR for *LINC00339*, *CDC42*, and *WNT4* expressions in hFOB 1.19 cells after silencing of *LINC00339* using shRNA (blue) compared to NC-treated cells.

(legend continued on next page)

potential biased causal effect estimates deriving from invalid instrument variables, we further performed a multi-instrument-based MR analysis.[53] A total of 44 purified *cis*-eQTL SNPs were selected as instrumental variables with the detailed information summarized in Table S8. We detected robust causal association between *LINC00339* expression and BMD based on either inverse variance-weighted method ($p = 0.009$) or median-based method ($p = 0.0001$). A scatterplot of genetic association with *LINC003339* against association with BMD are shown in Figure S7. Collectively, these data consistently suggested the causal relationship between *LINC00339* expression and BMD.

Discussion

Most of the BMD-associated SNPs identified by GWASs are located in the non-coding regions of genome. The molecular mechanisms underlying the causal actions and biological effects of BMD-associated SNPs are largely unknown. Our study provides extensive evidence that an intergenic SNP (rs6426749) at 1p36.12 acts as a strong long-range allele-specific enhancer to regulate the expression of *LINC00339*. In particular, we demonstrate that the distal enhancer interacts with *LINC00339* via long-range chromatin loop, and CTCF plays a critical role in this loop formation and maintenance. Moreover, the activity of the enhancer containing rs6426749 is mediated by the transcription factor TFAP2A. The rs6426749-G allele robustly recruits TFAP2A, which efficiently elevates the enhancer activity and increases the *LINC00339* expression. The target gene *LINC00339* could negatively modulate the expression of *CDC42*, which is an important gene involved in bone metabolism (Figure 6). Taken together, we elucidate a potential mechanistic basis for the genetic association between rs6426749 and osteoporosis, which highlights the regulatory effect of noncoding SNPs underlying the pathogenesis of diseases.

Our analysis reveals that a distal enhancer could regulate the expression of *LINC00339* via long-range chromatin loop formation. A looped genomic architecture is mediated by some DNA-binding proteins, which facilitate the folding of the 3D genome and bring the distal regulatory elements and promoters into proximity. CTCF is one of the most widely characterized proteins in mediating long-range loop formation.[59,60,65,66] CTCF has been shown to bind to distal enhancer and promoter regions to activate enhancer-promoter interactions.[27,32] Our study also found robust CTCF binding near the boundaries of Hi-C regions involving rs6426749 and *LINC00339*, supporting the potential role of CTCF involved in loop formation. It has been reported that depletion of CTCF could cause global reduction of intradomain interactions.[67] Consistent with this finding, knockdown of *CTCF* efficiently repressed *LINC00339* expression in our functional assays, indicating that CTCF is required for the loop formation. Our finding is comparable to the report by Xiang et al.,[68] in which they found that CTCF was specifically enriched near *MYC* locus, and knockdown of CTCF reduced chromatin interaction frequencies between the MYC promoter and its enhancers as well as *MYC* expression. Therefore, we speculate that loss of CTCF could disrupt the loop structure and restrict the enhancer from approaching the promoter, and therefore inhibit the expression of target gene *LINC00339*.

Here we provide the key mechanistic insight that rs6426749 acts as an allele-dependent enhancer to functionally contribute to differences in allelic gene expression, which demonstrates that genetic variation in regulatory elements can have a strong influence on common human phenotypic traits. It is generally believed that the enhancer regulates target gene transcription via altering TFs occupancy.[69] Our results indicated that TFAP2A could particularly bind to rs6426749-G allele to increase the expression of *LINC00339*. TFAP2A is a transcriptional activator that can bind to enhancer regions to elevate the enhancer activities.[70,71] Our knockdown experiment found that downregulation of *TFAP2A* efficiently repressed *LINC00339* expression in osteoblast cells, which provides functional evidence to support the role of transcriptional activation for TFAP2A.

Our study implicates *LINC00339* as the target for a noncoding susceptibility SNP rs6426749 located at the well-described BMD locus 1p36.12. The nearest gene of this susceptibility SNP is *ZBTB40*, which has unknown function or connection with bone metabolism. Our results reveal that the nearest gene may not be the true target gene for the susceptibility SNPs identified by GWASs, especially for SNPs located in the intergenic region. The number of lncRNAs neighboring those noncoding SNPs far exceeds that of protein-coding genes. The FANTOM5 Project[62] recently elucidated nearly 20,000 potential functional lncRNAs overlapping trait-associated variants or eQTL SNPs, implying the importance of lncRNA in disease development. Another recent study has identified a set of lncRNAs regulated by noncoding SNPs in prostate cancer (MIM: 176807),[72] highlighting the importance of investigating the functional link between the noncoding SNPs and lncRNAs. *LINC00339* is ubiquitously expressed in various tissues and cells with hardly any coding potential. Some variants in *LINC00339* have been identified for

(D) Hi-C annotation revealed interaction between *LINC00339* and *CDC42*. Different shade of colors represents the strength of long-range interactions, and different colors indicated different Hi-C dataset (pink: Hi-C data on IMR90 cells from 4DGenome;[25] blue: DNase Hi-C data on H1-hESC cells;[30] purple: ChIA-PET data taken from ENCODE on three cell lines [HeLa-S3, K562, and MCF-7][49] or CEO database on GM12878 cells[32]). Error bars, SD. $n \geq 3$. NS: not significant, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ as determined by an unpaired, two-tailed Student's t test.

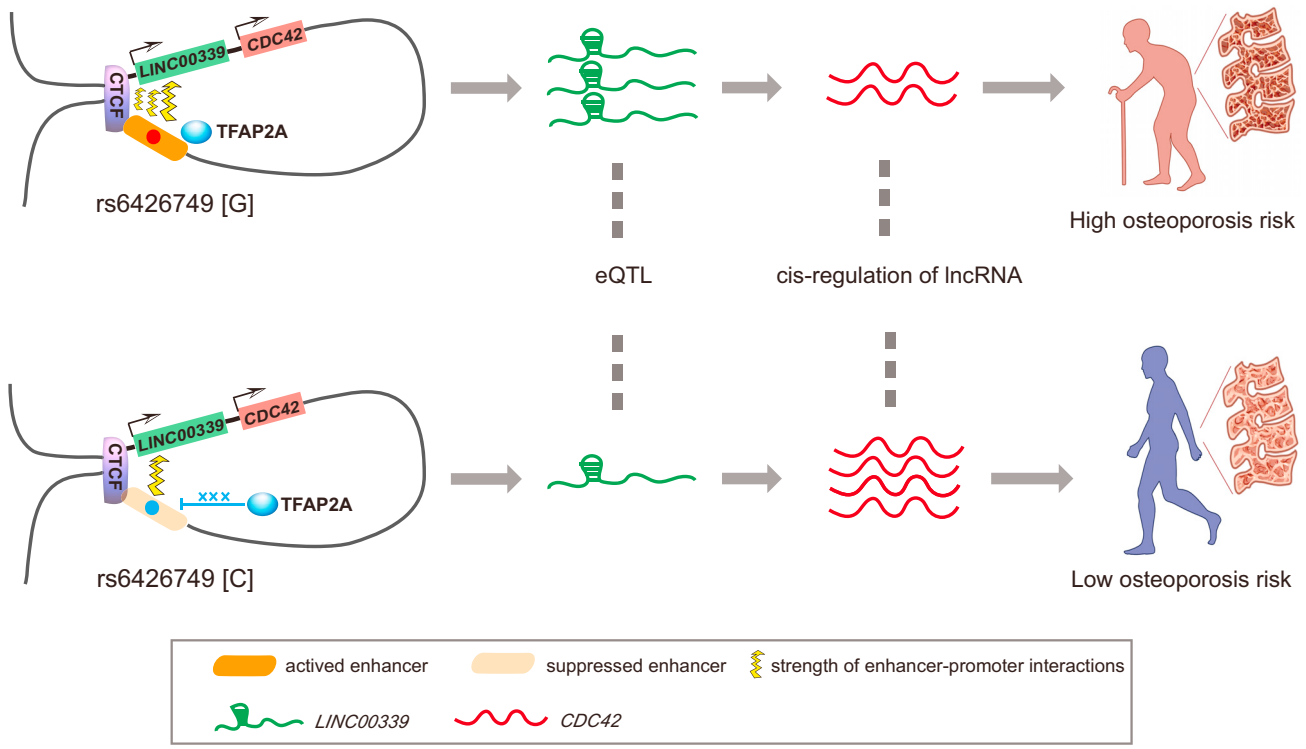

Figure 6. Potential Regulatory Model between rs6426749, *LINC00339*, and *CDC42*
A schematic representation elucidating how genetic variant (rs6426749) affects disease predisposition (osteoporosis). In the top panel, rs6426749-G allele robustly binds to TFAP2A, which elevates the activity of enhancer containing rs6426749 and increases *LINC00339* expression. Overexpressed *LINC00339* acts as a *cis*-regulatory element to suppress *CDC42* expression, whose relative low expression level is a risk factor to decrease BMD and increase osteoporosis incidence. In the bottom panel, in contrast, rs6426749-C allele is absent from TFAP2A binding, which represses the enhancer activity, resulting in relatively lower *LINC00339* expression, which further increases the *CDC42* expression. The relatively high expression level of *CDC42* decreases the risk to osteoporosis incidence.

association with endometriosis (MIM: 131200) and ovarian cancers (MIM: 167000).[73,74] A recent study found that the leading endometriosis risk SNPs within noncoding region at 1p36.12 could act through inverse regulation of *CDC42* and *LINC00339*,[75] supporting the functionality of *LINC00339* and potential negative regulation between *LINC00339* and *CDC42*, which is consistent with our study. The function of *LINC00339* in bone metabolism is unknown. However, considering that lncRNA could regulate expression of target genes in *cis*,[76] we found that downregulation of *LINC00339* can significantly increase the expression of *CDC42* in osteoblast cells. CDC42 (cell division cycle 42) is a small Rho GTPase and key regulator of cytoskeletal components. Moreover, CDC42 is a crucial component of the MAPK (mitogen-activated protein kinase) pathway, which is a pivotal mediator of bone metabolism and plays essential roles in osteoblast differentiation and skeletal development.[18] Previous studies have revealed the important roles of *CDC42* in bone modeling and remodeling.[77] We induced the human umbilical cord mesenchymal stem cells (hUCMSCs) into osteoblast and adipocyte cells, respectively. We found that the expression level of *CDC42* was significantly increased during osteoblast differentiation, but the expression level of *WNT4* was negligible compared with *CDC42* or *LINC00339*, supporting the important role of *CDC42* in bone metabolism (Figure S8). However, no BMD variants in *CDC42* have been reported, indicating that this gene might be regulated by remote BMD susceptibility SNPs. Our study implicates the functional connection between rs6426749 and *CDC42*. Previous GWASs have identified rs6426749-G as the risk allele for BMD. Our data posit that rs6426749-G can enhance the expression of *LINC00339* and therefore suppress the expression of *CDC42*. Deletion of Cdc42 in mice could lead to increased adipocyte differentiation and decreased bone formation,[19] as well as severe skeletal abnormalities,[63] which gives us a strong support that the target gene *CDC42* of rs6426749-G could affect the bone formation and increase the risk of osteoporosis. Future investigations are encouraged to elucidate the precise molecular mechanisms.

Our study also has limitations. First, we leveraged eQTL to prioritize functional GWAS variants. However, due to the smaller sample size and disease or cell type relevance of current eQTL data, there might exist unbalanced signals between eQTLs and GWAS SNPs. We therefore reinforce the need of functional assays to validate the findings indicated by eQTL analysis. Second, it is worth noting that our regulatory model could not exclude the contribution of other genetic variants, but instead highlights the

results of the study at hand, and might be useful in developing hypotheses for future experimentation. Finally, our study highlights the regulatory effect of noncoding SNPs on osteoporosis through *LINC00339*. Future functional experiments are encouraged to investigate the detailed molecular mechanism between *LINC00339* and osteoporosis.

In summary, through an integrative analysis combining various computational analyses and functional assays, we elucidate a potential mechanistic basis for a functional susceptibility SNP (rs6426749) with long-range target genes (*LINC00339*, *CDC42*) at 1p36.12. We anticipate that many other BMD-associated variants in noncoding regions may have similar mechanisms. The integrative approach described in this study can be further used to assign function to more noncoding SNPs in future studies, which is the primary task in our post-GWAS period.

Supplemental Data

Supplemental Data include eight figures and eight tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.03.001.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31771399, 81573241, 31471188, 31701095), China Postdoctoral Science Foundation (2016M602797), Natural Science Basic Research Program Shaanxi Province (2016JQ3026), and the Fundamental Research Funds for the Central Universities. We thank the GTEx Consortium. We obtained GTEx data through dbGaP authorized access at https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login with the accession number of phs000424.v6.p1. We would also like to thank the School of Life Science and Technology at Xi'an Jiaotong University for the sharing platform of laboratory apparatus.

Received: September 25, 2017
Accepted: February 28, 2018
Published: April 26, 2018

Web Resources

1000 Genomes V3 genotype data, ftp://ftp.trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/
4DGenome, https://4dgenome.research.chop.edu/
ArrayExpress, https://www.ebi.ac.uk/arrayexpress/
dbGaP, https://www.ncbi.nlm.nih.gov/gap
FANTOM5, http://fantom.gsc.riken.jp/5/data
GEFOS, http://www.gefos.org/
GEO, https://www.ncbi.nlm.nih.gov/geo/
GTEx Portal, https://www.gtexportal.org/home/
GWAS Catalog, http://www.ebi.ac.uk/gwas/
HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php
International HapMap Project, ftp://ftp.ncbi.nlm.nih.gov/hapmap/
liftOver, http://genome.ucsc.edu/cgi-bin/hgLiftOver
MEME Suite, http://meme-suite.org/
Mouse Genome Informatics, http://www.informatics.jax.org/
OMIM, http://www.omim.org/
R statistical software, https://www.r-project.org/
UCSC ENCODE download portal, https://genome.ucsc.edu/encode/downloads.html
WashU EpiGenome Browser, http://epigenomegateway.wustl.edu/browser/

References

1. Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. Nat. Rev. Genet. *10*, 241–251.
2. Liu, N.Q., Ter Huurne, M., Nguyen, L.N., Peng, T., Wang, S.Y., Studd, J.B., Joshi, O., Ongen, H., Bramsen, J.B., Yan, J., et al. (2017). The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression. Nat. Commun. *8*, 14418.
3. Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C., et al. (2017). A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell *170*, 522–533.e15.
4. Kanis, J.A., Delmas, P., Burckhardt, P., Cooper, C., Torgerson, D.; and The European Foundation for Osteoporosis and Bone Disease (1997). Guidelines for diagnosis and management of osteoporosis. Osteoporos. Int. *7*, 390–406.
5. Johnell, O., Kanis, J.A., Oden, A., Johansson, H., De Laet, C., Delmas, P., Eisman, J.A., Fujiwara, S., Kroger, H., Mellstrom, D., et al. (2005). Predictive value of BMD for hip and other fractures. J. Bone Miner. Res. *20*, 1185–1194.
6. Peacock, M., Turner, C.H., Econs, M.J., and Foroud, T. (2002). Genetics of osteoporosis. Endocr. Rev. *23*, 303–326.
7. Liu, Y.J., Zhang, L., Papasian, C.J., and Deng, H.W. (2014). Genome-wide association studies for osteoporosis: A 2013 update. J. Bone Metab. *21*, 99–116.
8. Richards, J.B., Zheng, H.-F., and Spector, T.D. (2012). Genetics of osteoporosis from genome-wide association studies: advances and challenges. Nat. Rev. Genet. *13*, 576–588.
9. Styrkarsdottir, U., Halldorsson, B.V., Gretarsdottir, S., Gudbjartsson, D.F., Walters, G.B., Ingvarsson, T., Jonsdottir, T., Saemundsdottir, J., Snorradóttir, S., Center, J.R., et al. (2009). New sequence variants associated with bone mineral density. Nat. Genet. *41*, 15–17.
10. Styrkarsdottir, U., Halldorsson, B.V., Gretarsdottir, S., Gudbjartsson, D.F., Walters, G.B., Ingvarsson, T., Jonsdottir, T., Saemundsdottir, J., Center, J.R., Nguyen, T.V., et al. (2008). Multiple genetic loci for bone mineral density and fractures. N. Engl. J. Med. *358*, 2355–2365.
11. Richards, J.B., Rivadeneira, F., Inouye, M., Pastinen, T.M., Soranzo, N., Wilson, S.G., Andrew, T., Falchi, M., Gwilliam, R., Ahmadi, K.R., et al. (2008). Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet *371*, 1505–1512.
12. Rivadeneira, F., Styrkársdottir, U., Estrada, K., Halldórsson, B.V., Hsu, Y.H., Richards, J.B., Zillikens, M.C., Kavvoura, F.K., Amin, N., Aulchenko, Y.S., et al.; Genetic Factors for Osteoporosis (GEFOS) Consortium (2009). Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. Nat. Genet. *41*, 1199–1206.
13. Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y.H., Duncan, E.L., Ntzani, E.E., Oei, L., Albagha, O.M., Amin, N., Kemp, J.P., et al. (2012). Genome-wide meta-analysis identifies

56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat. Genet. *44*, 491–501.
14. Zhang, L., Choi, H.J., Estrada, K., Leo, P.J., Li, J., Pei, Y.-F., Zhang, Y., Lin, Y., Shen, H., Liu, Y.-Z., et al. (2014). Multistage genome-wide association meta-analyses identified two new loci for bone mineral density. Hum. Mol. Genet. *23*, 1923–1933.
15. Kemp, J.P., Medina-Gomez, C., Estrada, K., St Pourcain, B., Heppe, D.H.M., Warrington, N.M., Oei, L., Ring, S.M., Kruithof, C.J., Timpson, N.J., et al. (2014). Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. PLoS Genet. *10*, e1004423.
16. Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L., Park-Min, K.H., Tobias, J.H., Kooperberg, C., et al.; AOGC Consortium; and UK10K Consortium (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature *526*, 112–117.
17. Yu, B., Chang, J., Liu, Y., Li, J., Kevork, K., Al-Hezaimi, K., Graves, D.T., Park, N.H., and Wang, C.Y. (2014). Wnt4 signaling prevents skeletal aging and inflammation by inhibiting nuclear factor-κB. Nat. Med. *20*, 1009–1017.
18. Greenblatt, M.B., Shim, J.-H., and Glimcher, L.H. (2013). Mitogen-activated protein kinase pathways in osteoblasts. Annu. Rev. Cell Dev. Biol. *29*, 63–79.
19. Wuerfel, C., Hoffmann, C., Kawelke, N., and Aszodi, A. (2012). Deletion of cdc42 in osteoblast progenitors leads to increased adipocyte differentiation and decreased bone formation. Bone *50*. https://doi.org/10.1016/j.bone.2012.02.113.
20. Yang, T.L., Guo, Y., Liu, Y.J., Shen, H., Liu, Y.Z., Lei, S.F., Li, J., Tian, Q., and Deng, H.W. (2012). Genetic variants in the SOX6 gene are associated with bone mineral density in both Caucasian and Chinese populations. Osteoporos. Int. *23*, 781–787.
21. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.
22. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
23. Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. *44* (D1), D877–D881.
24. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H., et al.; 1000 Genomes Project Consortium (2015). An integrated map of structural variation in 2,504 human genomes. Nature *526*, 75–81.
25. Teng, L., He, B., Wang, J., and Tan, K. (2015). 4DGenome: a comprehensive database of chromatin interactions. Bioinformatics *31*, 2560–2564.
26. Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.
27. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.
28. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606.
29. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al.; BLUEPRINT Consortium (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell *167*, 1369–1384.e19.
30. Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C.B., Krumm, A., et al. (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat. Methods *12*, 71–78.
31. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
32. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell *163*, 1611–1627.
33. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
34. Kwan, T., Grundberg, E., Koka, V., Ge, B., Lam, K.C.L., Dias, C., Kindmark, A., Mallmin, H., Ljunggren, O., Rivadeneira, F., et al. (2009). Tissue effect on genetic control of transcript isoform variation. PLoS Genet. *5*, e1000608.
35. Hsu, Y.-H., Zillikens, M.C., Wilson, S.G., Farber, C.R., Demissie, S., Soranzo, N., Bianchi, E.N., Grundberg, E., Liang, L., Richards, J.B., et al. (2010). An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. PLoS Genet. *6*, e1000977.
36. Hwang, J.Y., Lee, S.H., Go, M.J., Kim, B.J., Kou, I., Ikegawa, S., Guo, Y., Deng, H.W., Raychaudhuri, S., Kim, Y.J., et al. (2013). Meta-analysis identifies a MECOM gene as a novel predisposing factor of osteoporotic fracture. J. Med. Genet. *50*, 212–219.
37. Yang, T.L., Guo, Y., Zhang, J.G., Xu, C., Tian, Q., and Deng, H.W. (2015). Genome-wide survey of runs of homozygosity identifies recessive loci for bone mineral density in Caucasian and Chinese populations. J. Bone Miner. Res. *30*, 2119–2126.
38. Tang, R., Wei, Y., Li, Z., Chen, H., Miao, Q., Bian, Z., Zhang, H., Wang, Q., Wang, Z., Lian, M., et al. (2016). A common variant in CLDN14 is associated with primary biliary cirrhosis and bone mineral density. Sci. Rep. *6*, 19877.
39. Zhu, D.L., Guo, Y., Zhang, Y., Dong, S.S., Xu, W., Hao, R.H., Chen, X.F., Yan, H., Yang, S.Y., and Yang, T.L. (2017). A functional SNP regulated by miR-196a-3p in the 3′UTR of FGF2 is associated with bone mineral density in the Chinese population. Hum. Mutat. *38*, 725–735.
40. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al.; GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
41. Oldridge, D.A., Wood, A.C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., McDaniel, L.D., Diamond, M., Hart, L.S., Zhu, S., et al. (2015). Genetic predisposition to

neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature *528*, 418–421.
42. McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. Science *304*, 581–584.
43. Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. *6*, e1000895.
44. Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R., and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat. Genet. *49*, 600–605.
45. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202-8.
46. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44* (D1), D110–D115.
47. Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Soboleva, A.V., Kasianov, A.S., Ashoor, H., Ba-Alawi, W., Bajic, V.B., Medvedeva, Y.A., Kolpakov, F.A., and Makeev, V.J. (2016). HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. *44* (D1), D116–D125.
48. Pachkov, M., Balwierz, P.J., Arnold, P., Ozonov, E., and van Nimwegen, E. (2013). SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res. *41*, D214–D220.
49. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.
50. Bogachek, M.V., Chen, Y., Kulak, M.V., Woodfield, G.W., Cyr, A.R., Park, J.M., Spanheimer, P.M., Li, Y., Li, T., and Weigel, R.J. (2014). Sumoylation pathway is required to maintain the basal breast cancer subtype. Cancer Cell *25*, 748–761.
51. 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.
52. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.
53. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int. J. Epidemiol. *46*, 1734–1739.
54. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.
55. Corradin, O., Cohen, A.J., Luppino, J.M., Bayles, I.M., Schumacher, F.R., and Scacheri, P.C. (2016). Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. Nat. Genet. *48*, 1313–1320.
56. Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. *24*, 14–24.
57. Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.
58. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
59. Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. *20*, 2349–2354.
60. Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.
61. Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. *41*, e74–e74.
62. Hon, C.C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. Nature *543*, 199–204.
63. Aizawa, R., Yamada, A., Suzuki, D., Iimura, T., Kassai, H., Harada, T., Tsukasaki, M., Yamamoto, G., Tachikawa, T., Nakao, K., et al. (2012). Cdc42 is required for chondrogenesis and interdigital programmed cell death during limb development. Mech. Dev. *129*, 38–50.
64. Blake, J.A., Eppig, J.T., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J.; and the Mouse Genome Database Group (2017). Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. *45* (D1), D723–D729.
65. Majumder, P., Gomez, J.A., Chadwick, B.P., and Boss, J.M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. J. Exp. Med. *205*, 785–798.
66. Majumder, P., and Boss, J.M. (2010). CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. Mol. Cell. Biol. *30*, 4211–4223.
67. Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. USA *111*, 996–1001.
68. Xiang, J.-F., Yin, Q.-F., Chen, T., Zhang, Y., Zhang, X.-O., Wu, Z., Zhang, S., Wang, H.-B., Ge, J., Lu, X., et al. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. Cell Res. *24*, 513–531.

69. Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.
70. Mitchell, P.J., Wang, C., and Tjian, R. (1987). Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell *50*, 847–861.
71. Williams, T., and Tjian, R. (1991). Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. *5*, 670–682.
72. Guo, H., Ahmed, M., Zhang, F., Yao, C.Q., Li, S., Liang, Y., Hua, J., Soares, F., Sun, Y., Langstein, J., et al. (2016). Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. Nat. Genet. *48*, 1142–1150.
73. Albertsen, H.M., Chettier, R., Farrington, P., and Ward, K. (2013). Genome-wide association study link novel loci to endometriosis. PLoS ONE *8*, e58257.
74. Kuchenbaecker, K.B., Ramus, S.J., Tyrer, J., Lee, A., Shen, H.C., Beesley, J., Lawrenson, K., McGuffog, L., Healey, S., Lee, J.M., et al.; EMBRACE; GEMO Study Collaborators; Breast Cancer Family Registry; HEBON; KConFab Investigators; Australian Cancer Study (Ovarian Cancer Investigators); Australian Ovarian Cancer Study Group; and Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (2015). Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nat. Genet. *47*, 164–171.
75. Powell, J.E., Fung, J.N., Shakhbazov, K., Sapkota, Y., Cloonan, N., Hemani, G., Hillman, K.M., Kaufmann, S., Luong, H.T., Bowdler, L., et al. (2016). Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDC42 and LINC00339. Hum. Mol. Genet. *25*, 5046–5058.
76. Esteller, M. (2011). Non-coding RNAs in human disease. Nat. Rev. Genet. *12*, 861–874.
77. Ito, Y., Teitelbaum, S.L., Zou, W., Zheng, Y., Johnson, J.F., Chappel, J., Ross, F.P., and Zhao, H. (2010). Cdc42 regulates bone modeling and remodeling in mice by modulating RANKL/M-CSF signaling and osteoclast polarization. J. Clin. Invest. *120*, 1981–1993.

The American Journal of Human Genetics, Volume *102*

Supplemental Data

An Osteoporosis Risk SNP at 1p36.12 Acts as an Allele-Specific Enhancer to Modulate *LINC00339* Expression via Long-Range Loop Formation

Xiao-Feng Chen, Dong-Li Zhu, Man Yang, Wei-Xin Hu, Yuan-Yuan Duan, Bing-Jie Lu, Yu Rong, Shan-Shan Dong, Ruo-Han Hao, Jia-Bin Chen, Yi-Xiao Chen, Shi Yao, Hlaing Nwe Thynn, Yan Guo, and Tie-Lin Yang

Supplemental Figures and Legends

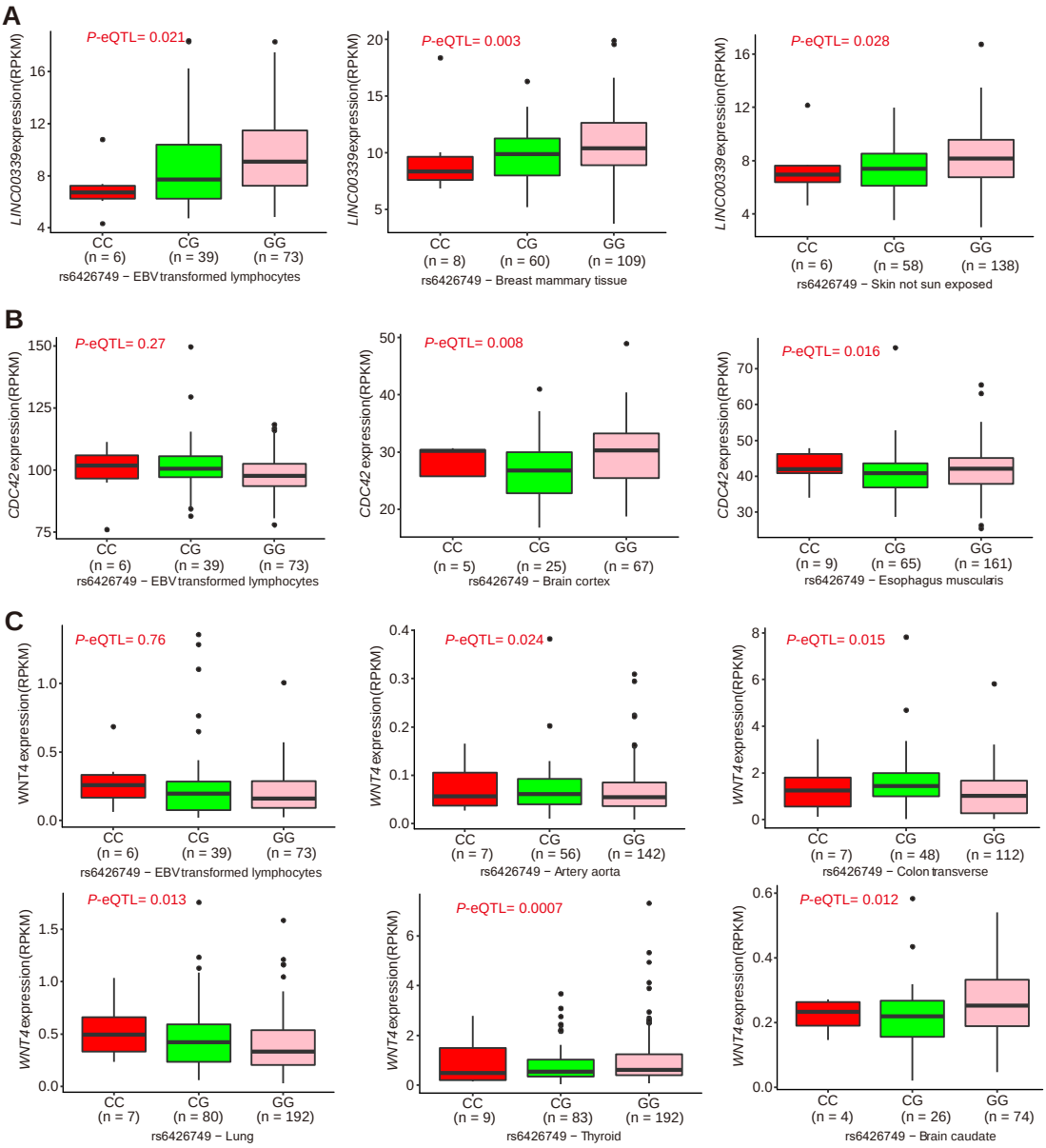

Figure S1. eQTL analyses for rs6426749 with *LINC00339*, *CDC42* and *WNT4* from GTEx

We checked the eQTL association from the Genotype-Tissue Expression (GTEx) database,[1] including 7,051 samples from 449 donors across 44 tissues. The eQTL analysis from GTEx Project[1] between rs6426749 and *LINC00339* (A), *CDC42* (B) or *WNT4* (C) on LCLs and other significant tissues ($P < 0.05$) are shown.

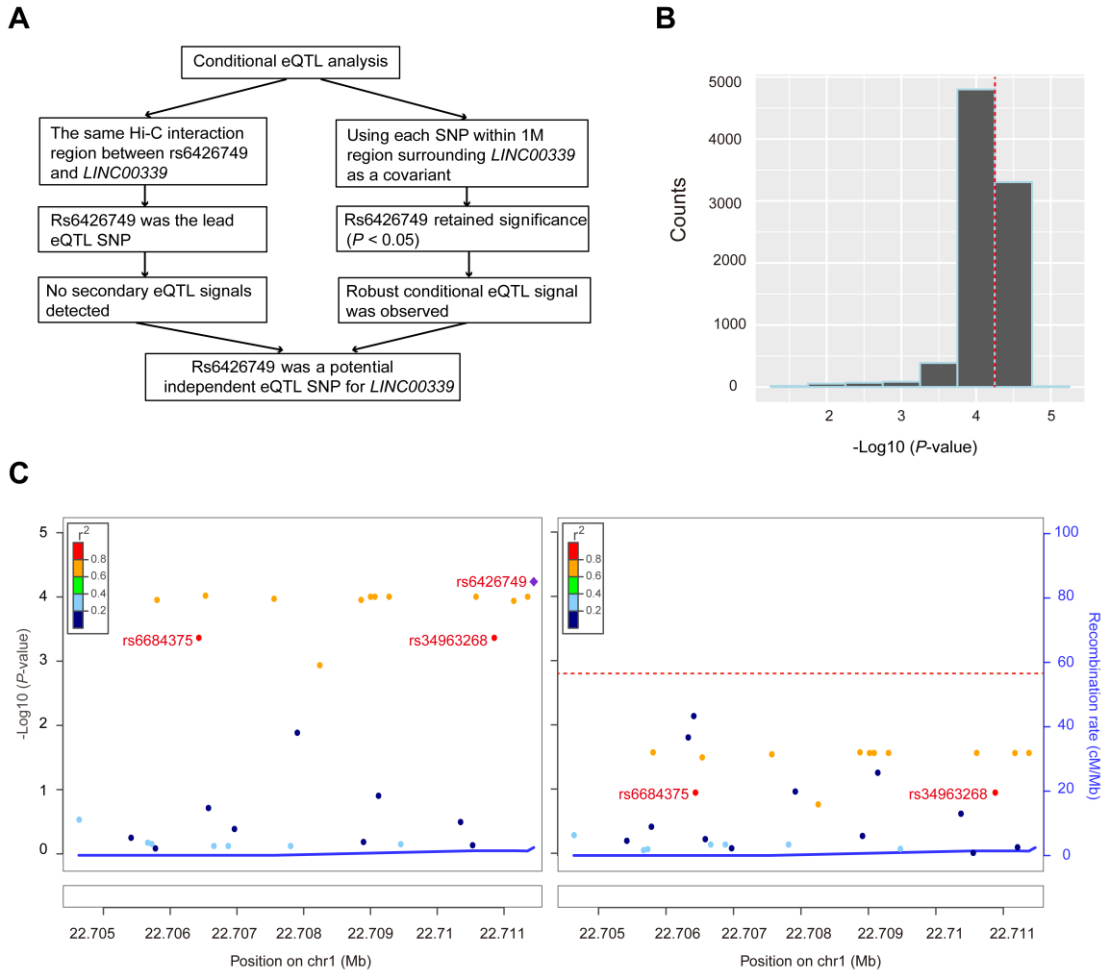

Figure S2. Conditional eQTL analysis for rs6426749

(A) Flowchart of conditional eQTL analysis. (B) Distribution of conditional eQTL signals for rs6426749 on *LINC00339* using each SNP within 1M region surrounding *LINC00339* as a covariant. Raw eQTL signal for rs6426749 with *LINC00339* was marked with dashed red line. (C) Raw (left) and conditional (right) eQTL signal (using rs6426749 as a covariant) for all SNPs within the same Hi-C interaction region with rs6426749. The dashed line represents significant association level of secondary eQTL SNPs (Bonferroni adjusted *P*-value < 0.05). Conditional eQTL analysis was performed by fitting the selected cis-eQTL SNP genotype as a covariate and testing for the secondary association retained using ANNOVA. Bonferroni correction was applied to determine the significance of secondary eQTLs.

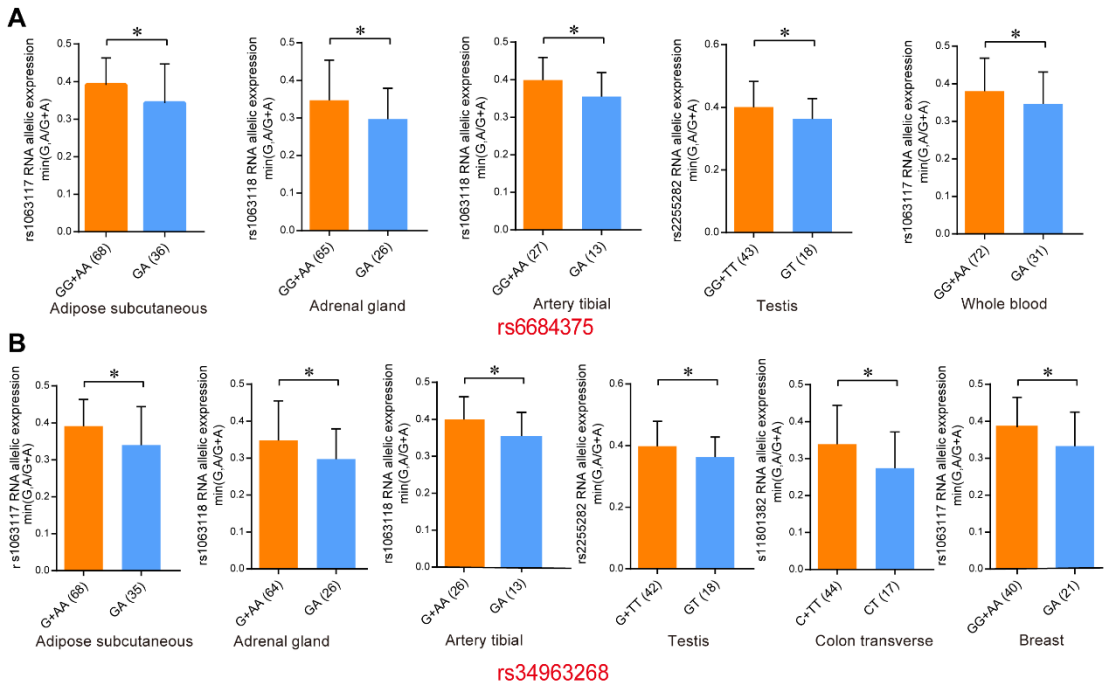

Figure S3. Allele specific expression analysis result on *LINC00339* for rs6684375 and rs34963268

Allele specific expression (ASE) analysis between rs6684375 (A) or rs34963268 (B) and *LINC00339*, using monoallelic gene expression data from GTEx[1]. Only significant tissues ($P < 0.05$) are shown. Error bars, s.d. *$P < 0.05$ as determined by Wilcoxon rank sum test.

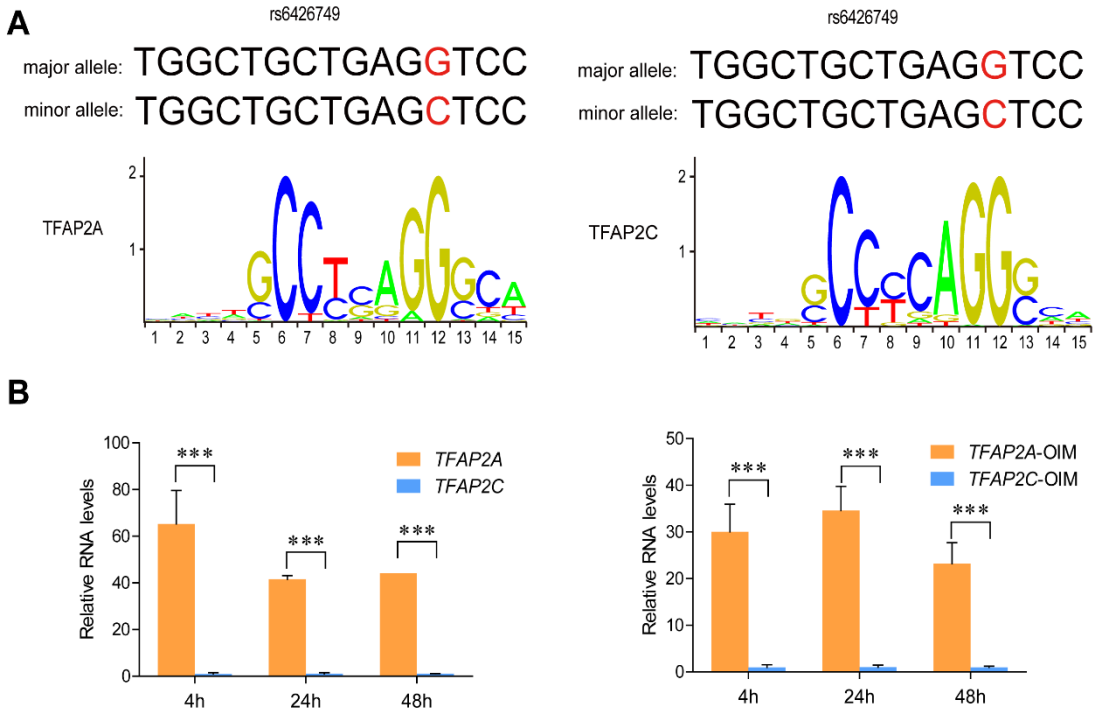

Figure S4. *TFAP2A* is predominantly higher expressed than *TFAP2C* in hFOB1.19 cells

(A) Motif predictions for rs6426749. Two TFAP family motifs (TFAP2A, TFAP2C) were predicted to exclusively bind to major allele (G) of rs6426749. (B) Comparison of mRNA expressions for *TFAP2A* and *TFAP2C* in hFOB1.19 cells undergoing spontaneous differentiation or with the effect of Osteogenic Induction Media (OIM). RNA expression data was extracted from GEO database (GEO: GSE75232).[2] Relative mRNA expression levels were normalized by equalizing *TFAP2C* expression levels to 1. Error bars, s.d. ***$P < 0.001$ as determined by an unpaired, two-tailed Student's t-test.

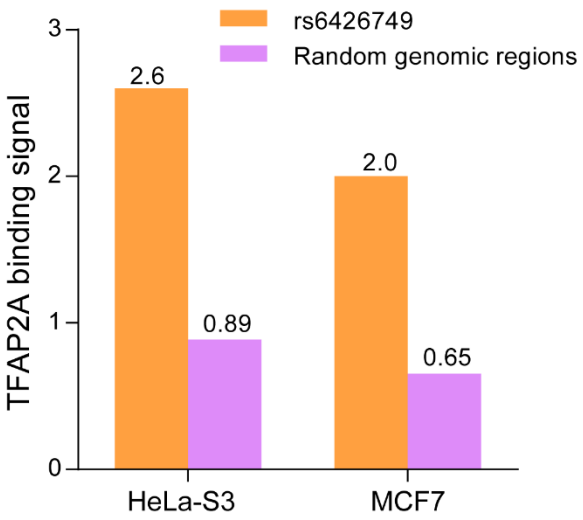

Figure S5. Comparison of TFAP2A binding on rs6426749-locus and random genomic regions

We compared the average TF binding signal surrounding rs6426749 (50-bp) with 1,000,000 randomly chosen genomic regions from 22 autosomes of the same length. Average TFAP2A binding signals for rs6426749-locus and 1,000,000 random genomic regions in HeLa-S3 cells and MCF7 cells were shown, respectively.

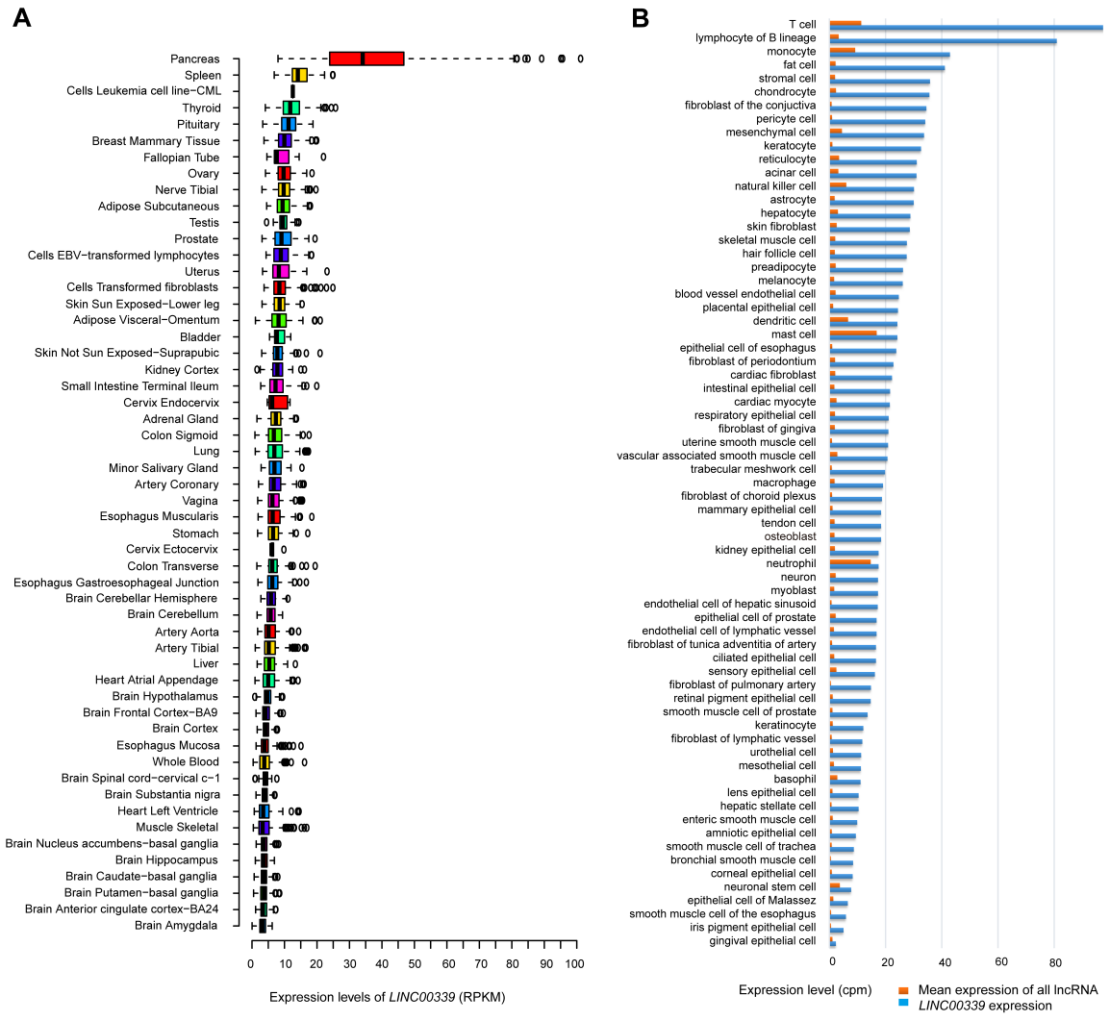

Figure S6. *LINC00339* is extensively expressed in both tissue and cell levels

LINC00339 expression levels were analyzed in 54 diverse tissues from GTEx Project[1] (A) and 69 diverse primary cells from FANMOT5[3] (B), respectively. cpm, counts per million.

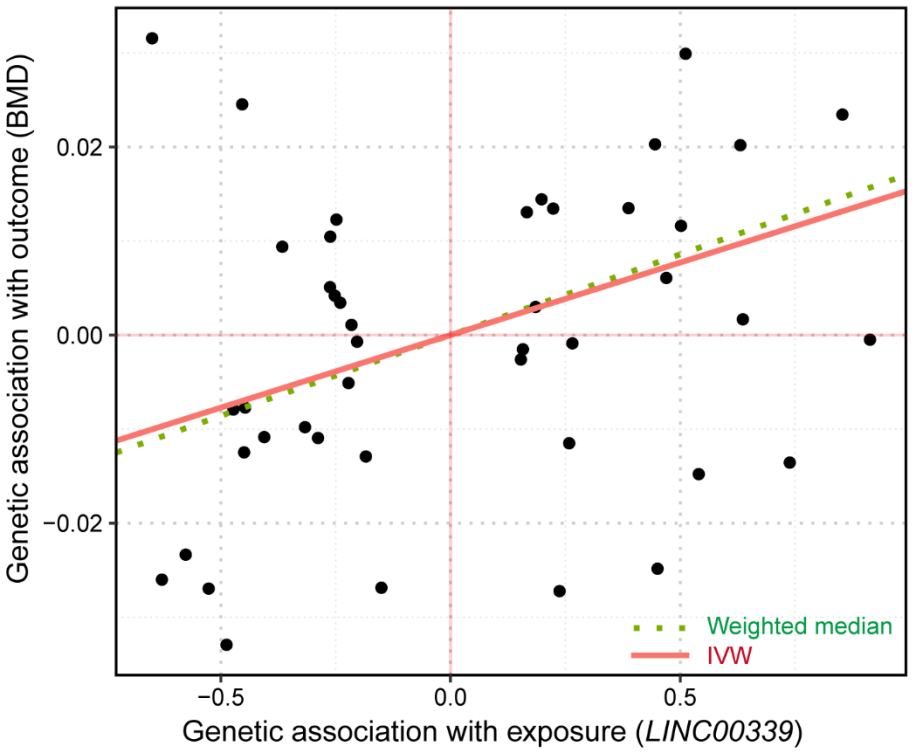

Figure S7. Scatter plot of genetic association with *LINC003339* against association with BMD

Each dot indicated one of 44 total genetic variants used as instrumental variables for multi-instrument based Median randomization analysis. The x and y axis represented coefficients of genetic association with *LINC00339* (eQTL) from GTEx whole blood tissue[1] or genetic association with BMD from UK Biobank,[4] respectively. The red line and green dashed line corresponded to slope from IVW or weighted median, respectively. IVW, inverse-variance weighted.

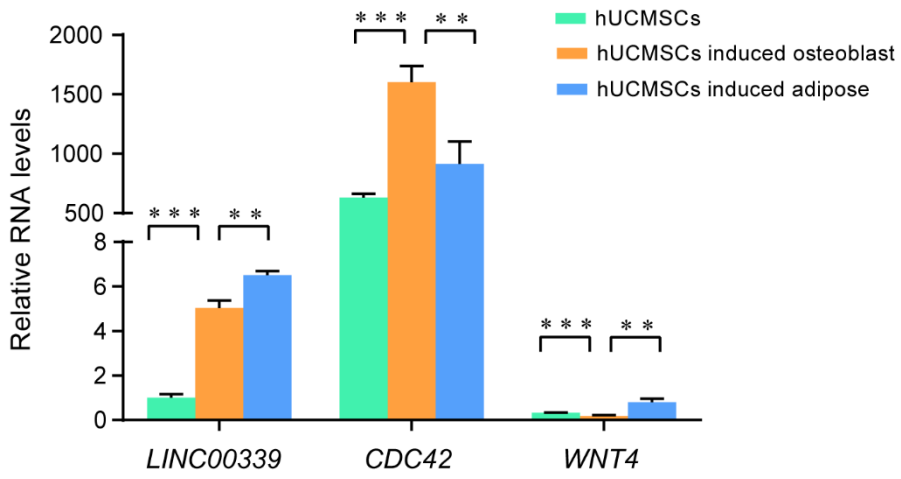

Figure S8. Comparison of gene expression in osteoblast and adipocyte cells induced from the same human mesenchymal stem cells

We obtained umbilical cords from donors with signed informed consent in local hospital and isolated human umbilical cord mesenchymal stem cells (hUCMSCs) as described previously.[5] The hUCMSCs cells were cultured with α-MEM supplemented with 10% FBS, 1% penicillin-streptomycin, and 0.2% cytokine CK2, CK4, CK8, CK9, and maintained at 37°C, 5% CO_2. The osteogenic and adipogenic differentiation was performed by using the OriCell™ hUCMSCs osteogenic differentiation medium kit (HUXUC-90021, Cyagen, China) and the OriCell™ hUCMSCs adipogenic differentiation medium kit (HUXUC-90031, Cyagen, China) according to the manufacturer's instruction, respectively. Cells were maintained in differentiation medium and the medium was changed every 3 days. Two weeks later, cells were harvested for RNA extraction and RT-qPCR. Relative mRNA expression levels were normalized by equalizing *LINC00339* expression levels in human umbilical cord mesenchymal stem cells (hUCMSCs) to 1. Error bars, s.d. $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by an unpaired, two-tailed Student's t-test.

Supplemental Tables

Table S1. Basic characteristics of the Chinese cohort for genetic association analysis

Traits	Chinese cohort (n=1300)
Male/Female	600/700
Age (years)	33.42 (11.32)
Weight (kg)	59.63 (10.41)
Height (cm)	163.94 (8.11)
Lumbar spine BMD (g/cm^2)	0.921 (0.131)
Femoral neck BMD (g/cm^2)	0.815 (0.131)

Note: data are shown as mean (standard deviation, SD).

Table S2. Summary of Hi-C or ChIA-PET data used in this study

Dataset[a]	Data type	Cell[b]	Reference
4DGenome	Hi-C	IMR90	Jin, F. et al.[6]
Cell2014	Hi-C	GM12878	Rao, Suhas S.P. et al.[7]
ChIA-PET^{GEO}	ChIA-PET	GM12878	Tang, Z. et al. [8]
ChIA-PET^{ENCODE}	ChIA-PET	K562, NB4, HCT-116, HeLa-S3, MCF7	Harrow, J. et al.[9]
Cell2016	Capture Hi-C	17 human primary blood cell types	Javierre, B.M. et al.[10]
Mm2015	DNase Hi-C	H1-hESC	Ma, W. et al.[11]
NG2015	Capture Hi-C	GM12878; CD34	Mifsud, B. et al.[12]
TAD	--	IMR90	Dixon, J.R. et al.[13]

Note: [a]Dataset, Hi-C or ChIA-PET data used; [b]Cell, Hi-C data on human healthy cells or ChIA-PET data on all cells were collected; ChIA-PET^{GEO}, ChIA-PET data retrieved from GEO database (GEO: GSE72816); ChIA-PET^{ENCODE}, ChIA-PET data retrieved from UCSC ENCODE download portal.

Table S3. Integrating Hi-C and cis-expression quantitative trait locus (eQTL) analysis for 8 BMD SNPs at 1p36.12

SNP	Target gene predicted from Hi-C	Distance(kb)[a]	Dataset (cell)	Locus1	Locus2	P_{eQTL}	η^2
rs6426749	***LINC00339***	**-359.4**	**4DGenome; IMR90**	**chr1:22341459-22371546**	**chr1:22704394-22711600**	**5.61×10^{-5}**	**0.04**
rs6426749	***LINC00339***	**-359.4**	**Nm2015; H1-hESC**	**chr1:22351460-22356461**	**chr1:22704000-22711000**	**5.61×10^{-5}**	**0.04**
rs6426749	*WNT4*	-242	4DGenome; IMR90	chr1:22434529-22463352	chr1:22704394-22711600	0.45	3.50×10^{-3}
rs6426749	*WNT4*	-242	NG2015; GM12878	chr1:22466747-22478213	chr1:22705031-22711598	0.45	3.50×10^{-3}
rs6426749	*WNT4*	-242	4DGenome; IMR90	chr1:22466749-22528553	chr1:22704394-22711600	0.45	3.50×10^{-3}
rs6426749	*WNT4*	-242	4DGenome; IMR90	chr1:22466749-22478215	chr1:22703397-22711600	0.45	3.50×10^{-3}
rs6426749	*RP1-224A6.3*	-360	4DGenome; IMR90	chr1:22341459-22371546	chr1:22704394-22711600	0.43	6.25×10^{-3}
rs6426749	*ZBTB40*	67	4DGenome; IMR90	chr1:22769969-22795701	chr1:22704394-22711600	0.93	2.96×10^{-4}
rs6426749	*RP1-224A6.9*	-284.4	4DGenome; IMR90	chr1:22425025-22427905	chr1:22703397-22721228	NA	NA
rs34920465	*WNT4*	-229.9	NG2015; GM12878	chr1:22466747-22478213	chr1:22697090-22702342	0.53	2.79×10^{-3}
rs6696981	*WNT4*	-232.4	4DGenome; IMR90	chr1:22440795-22463352	chr1:22702345-22704393	0.63	2.03×10^{-3}
rs6696981	*WNT4*	-232.4	NG2015; CD34	chr1:22466747-22478213	chr1:22702343-22703394	0.63	2.03×10^{-3}
rs7524102	*WNT4*	-228.0	NG2015; GM12878	chr1:22466747-22478213	chr1:22697090-22702342	0.57	2.45×10^{-3}
rs2235529	*HSPG2*	-186.7	ChIA-PET^{ENCODE}; MCF7	chr1:22262651-22265288	chr1:22448523-22450702	0.92	3.63×10^{-4}
rs2235529	*WNT4*	20.0	ChIA-PET^{ENCODE}; K562	chr1:22450461-22453077	chr1:22468006-22470655	0.99	5.56×10^{-5}
rs2235529	*WNT4*	20.0	ChIA-PET^{ENCODE}; MCF7	chr1:22446379-22450951	chr1:22454408-22458465	0.99	5.56×10^{-5}
rs2235529	*WNT4*	20.0	ChIA-PET^{ENCODE}; K562	chr1:22448838-22451961	chr1:22466949-22469506	0.99	5.56×10^{-5}
rs2235529	*RP1-224A6.9*	-2.4	4DGenome; IMR90	chr1:22450003-22488812	chr1:22425025-22427905	NA	NA
rs3765350	*HSPG2*	-183.5	NG2015; GM12878	chr1:22263134-22276213	chr1:22440793-22450000	0.88	5.50×10^{-4}
rs3765350	*WNT4*	23.1	NG2015; GM12878	chr1:22466747-22478213	chr1:22440793-22450000	0.62	2.06×10^{-3}
rs3765350	*WNT4*	23.1	ChIA-PET^{ENCODE}; K562	chr1:22446637-22449453	chr1:22469771-22471584	0.62	2.06×10^{-3}
rs3765350	*WNT4*	23.1	ChIA-PET^{ENCODE}; MCF7	chr1:22446379-22450951	chr1:22454408-22458465	0.62	2.06×10^{-3}
rs3765350	*WNT4*	23.1	NG2015; CD34	chr1:22466747-22478213	chr1:22440793-22450000	0.62	2.06×10^{-3}

Note: Hi-C, Capture Hi-C, DNase Hi-C and ChIA-PET data on over 20 cells summarized in Table S1 were used, with chromatin interaction regions showed in Locus1 and Locus2 (hg19); NA, not available; [a]Distance is the distance between SNP and transcription start site of target gene.

Table S4. Cis-expression quantitative trait locus (eQTL) analysis results for rs6426749, rs6684375, and rs34963268

Gene	rs6426749			rs6684375			rs34963268		
	Distance	*P*-value	η^2	Distance	*P*-value	η^2	Distance	*P*-value	η^2
C1orf213	984.3	0.62	2.07×10^{-3}	989.3	0.72	1.44×10^{-3}	984.9	0.72	1.44×10^{-3}
HNRNPR	959.3	0.49	3.15×10^{-3}	964.4	0.73	1.35×10^{-3}	959.9	0.73	1.35×10^{-3}
RP5-1057J7.1	859.8	0.33	4.81×10^{-3}	864.8	0.26	5.87×10^{-3}	860.4	0.26	5.87×10^{-3}
LUZP1	724.2	0.52	2.84×10^{-3}	729.2	0.40	3.97×10^{-3}	724.7	0.40	3.97×10^{-3}
KDM1A	634.5	0.29	5.36×10^{-3}	639.5	0.47	3.33×10^{-3}	635.1	0.47	3.33×10^{-3}
EPHB2	326.0	0.53	2.72×10^{-3}	331.0	0.44	3.60×10^{-3}	326.6	0.44	3.60×10^{-3}
ZBTB40	67.0	0.93	2.96×10^{-4}	72.0	0.86	6.59×10^{-4}	67.6	0.86	6.59×10^{-4}
WNT4	-242.0	0.45	3.50×10^{-3}	-237.0	0.62	2.08×10^{-3}	-241.4	0.62	2.08×10^{-3}
CDC42	**-332.4**	**4.56×10^{-3}**	**0.023**	**-327.3**	**5.70×10^{-3}**	**0.022**	**-331.8**	**5.70×10^{-3}**	**0.022**
LINC00339	**-359.4**	**5.61×10^{-5}**	**0.042**	**-354.4**	**4.25×10^{-4}**	**0.033**	**-358.9**	**4.25×10^{-4}**	**0.033**
RP1-224A6.3	-360.0	0.43	6.25×10^{-3}	-355.0	0.27	5.75×10^{-3}	-359.4	0.27	5.75×10^{-3}
HSPG2	-447.7	0.26	5.84×10^{-3}	-442.6	0.37	4.26×10^{-3}	-447.1	0.37	4.26×10^{-3}
RP11-26H16.1	-476.9	0.24	3.64×10^{-3}	-471.8	0.38	4.15×10^{-3}	-476.3	0.38	4.15×10^{-3}
LDLRAD2	-572.7	0.40	4.02×10^{-3}	-567.7	0.59	2.26×10^{-3}	-572.1	0.59	2.26×10^{-3}
USP48	-658.4	0.91	4.15×10^{-4}	-653.4	0.69	1.64×10^{-3}	-657.8	0.69	1.64×10^{-3}
NBPF3	-944.9	0.27	5.65×10^{-3}	-939.8	0.43	3.63×10^{-3}	-944.3	0.43	3.63×10^{-3}
HS6ST1P1	-956.7	0.91	4.12×10^{-4}	-951.6	0.85	7.26×10^{-4}	-956.1	0.85	7.26×10^{-4}
NBPF2P	-957.0	0.81	9.39×10^{-4}	-952.0	0.75	1.26×10^{-3}	-956.4	0.75	1.26×10^{-3}
PPP1R11P1	-987.0	0.39	4.04×10^{-3}	-982.0	0.41	3.87×10^{-3}	-986.4	0.41	3.87×10^{-3}

Note: Distance is the distance between SNP and transcription start site of target gene (kb).

Table S5. Co-expression analysis between *LINC00339* and *CDC42*

Tissues	Samples[a]	*CDC42 expression*	*LINC00339* expression	*P*-value	R^{2b}
Thyroid	355	42.34(6.88)	12.22(3.55)	6.15E-15	-0.398
Vagina	97	53.61(9.50)	6.96(2.88)	2.66E-04	-0.362
Ovary	108	42.90(6.86)	10.08(2.81)	2.72E-04	-0.344
Colon transverse	204	53.93(9.69)	6.61(2.43)	2.14E-05	-0.293
Stomach	204	43.82(9.21)	6.71(2.45)	1.36E-03	-0.223
Spleen	118	55.72(8.46)	14.49(3.58)	2.23E-02	-0.21
Small intestine terminal ileum	104	53.27(7.22)	7.90(3.32)	3.78E-02	-0.204
Prostate	119	40.57(6.24)	9.60(3.36)	4.37E-02	-0.185
Liver	137	23.75(7.51)	5.44(2.19)	3.24E-02	-0.183
Colon sigmoid	173	45.29(6.50)	7.32(3.03)	2.40E-02	-0.172
Esophagus mucosa	331	55.43(7.16)	4.22(1.87)	6.68E-03	-0.149
Nerve tibial	335	46.76(6.78)	9.98(2.68)	4.91E-02	-0.108
Brain cortex	128	28.58(5.81)	4.39(1.32)	7.47E-03	0.235
Brain cerebellar hemisphere	115	40.26(9.50)	6.09(1.86)	2.65E-03	0.278
Brain spinal cord (cervical c-1)	76	45.03(13.42)	4.00(1.13)	6.25E-04	0.384
Brain nucleus accumbens (basal ganglia)	123	24.74(8.58)	3.81(1.39)	1.03E-07	0.458
Brain caudate (basal ganglia)	134	25.62(8.06)	3.77(1.32)	5.64E-09	0.477
Brain putamen (basal ganglia)	103	23.44(6.98)	3.71(1.50)	3.21E-08	0.512
Brain frontal cortex (BA9)	117	37.36(10.21)	4.43(1.68)	1.67E-09	0.521
Brain anterior cingulate cortex (BA24)	99	34.32(11.84)	3.65(1.38)	2.10E-10	0.585
Brain hippocampus	103	30.66(10.36)	3.77(1.17)	3.69E-12	0.618
Brain substantia nigra	71	34.63(11.66)	3.96(1.37)	2.75E-10	0.664
Brain amygdala	81	27.83(9.94)	3.26(1.25)	4.44E-16	0.754

Brain hypothalamus	104	42.04(14.33)	4.72(1.69)	$P < 2.20E-12$	0.759
Whole blood	445	83.71(37.45)	4.10(2.24)	$P < 2.20E-12$	0.586
Testis	199	29.88(8.00)	9.68(1.67)	7.93E-12	0.46
Adrenal gland	159	44.93(6.44)	7.55(2.23)	2.63E-02	0.176
Muscle skeletal	469	20.52(6.10)	3.91(2.47)	8.37E-10	0.279
Heart left ventricle	267	25.92(9.09)	3.98(2.17)	3.77E-05	0.249

Note: Co-expression analysis was conducted by Pearson correlation using GTEx RNA expression data[1] in 50 tissues (4 tissues with sample counts less than 20 were excluded). Only significantly correlated tissues ($P < 0.05$) were showed. Expression data was shown as mean (standard deviation, SD); [a]Samples were sample counts without missing *CDC42* or *LINC00339* expression data; [b]R^2 was Pearson Correlation Coefficient.

Table S6. Chromatin interactions between *LINC00339* and *CDC42*

Cell	Validation	Locus1	Gene1	Locus2	Gene2	Score
K562	ChIA-PET^{ENCODE}	chr1:22348117-22354021	*LINC00339*	chr1:22376823-22382698	*CDC42*	15
K562	ChIA-PET^{ENCODE}	chr1:22348152-22354988	*LINC00339*	chr1:22377027-22382698	*CDC42*	18
MCF7	ChIA-PET^{ENCODE}	chr1:22350975-22355164	*LINC00339*	chr1:22378266-22380971	*CDC42*	7
MCF7	ChIA-PET^{ENCODE}	chr1:22351119-22355075	*LINC00339*	chr1:22377954-22381978	*CDC42*	5
HeLa-S3	ChIA-PET^{ENCODE}	chr1:22351928-22352455	*LINC00339*	chr1:22379030-22380021	*CDC42*	3
K562	ChIA-PET^{ENCODE}	chr1:22354998-22357992	*LINC00339*	chr1:22377814-22380762	*CDC42*	2
GM12878	ChIA-PET^{GEO}	chr1:22349358-22349912	*LINC00339*	chr1:22379575-22381183	*CDC42*	5
GM12878	ChIA-PET^{GEO}	chr1:22350646-22354253	*LINC00339*	chr1:22377722-22381071	*CDC42*	61
H1-hESC	Nm2015	chr1:22351460-22356461	*LINC00339*	chr1:22374000-22388000	*CDC42*	NA
IMR90	4DGenome	chr1:22341459-22375876	*LINC00339*	chr1:22375877-22377917	*CDC42*	4.17E-06
IMR90	4DGenome	chr1:22351108-22359290	*LINC00339*	chr1:22379987-22393227	*CDC42*	5.70E-04
IMR90	4DGenome	chr1:22351108-22376206	*LINC00339*	chr1:22377918-22379986	*CDC42*	3.52E-12
IMR90	4DGenome	chr1:22359539-22409777	*LINC00339*	chr1:22351108-22359290	*CDC42*	3.47E-14

Note: Hi-C, DNase Hi-C and ChIA-PET data summarized in Table S1 were used, with chromatin interaction regions showed in Locus1 and Locus2 (hg19); NA, not available; [a]Score: Confidence *P*-value for Hi-C or confidence scores for ChIA-PET chromatin interactions.

Table S7. Summary of primers or siRNA sequences used

Assays	Target	Primers (5'-3')
Luciferase Report-Fusion PCR	rs6426749-F1	GGGGTACCTTTTAGGGAGTTTGAATTGGGCTC (Kpn I)
	rs6426749-R1	AGGCCAGAGGACTATTGTATTTGA
	LINC00339-F1	AATAGTCCTCTGGCCTTGGTTAGCATCTCTGCTTCCTCTA
	LINC00339-R1	CGACGCGTGGACGAGGAAAGATCAGGATAAGA (Mlu I)
Luciferase Report PCR	rs34963268-F1	GGGGTACCAGGCATCTGATAAAGACTCCG (Kpn I)
	rs34963268-R1	CGACGCGTTAAAAGGCCCCAGTAACCC (Mlu I)
	rs6684375-F1	GGGGTACCCCTCATGCCAATGACTCTGGT (Kpn I)
	rs6684375-R1	CGACGCGTATAGCCTGTCCTCATCCTTCCG (Mlu I)
	LINC00339-F2	CGACGCGTTGGTTAGCATCTCTGCTTCCTCTA (Mlu I)
	LINC00339-R2	GAAGATCTGGACGAGGAAAGATCAGGATAAGA (Bgl II)
Luciferase Report-Promoter PCR	*LINC00339*-F	GGGGTACCTGGTTAGCATCTCTGCTTCCTCTA (Kpn I)
	LINC00339-R	CGACGCGTGGACGAGGAAAGATCAGGATAAGA (Mlu I)
Site-directed mutagenesis	rs6426749-F (G-C)	CATACTGGCTGCTGAGCTCCAGGCCAATGGAC
	rs6426749-R (G-C)	GTCCATTGGCCTGGAGCTCAGCAGCCAGTATG
	rs34963268-F (C-G)	CTGGATCGTTGACGTCATTTGAGTGCCTGGAT
	rs34963268-R (C-G)	TGACGTCAACGATCCAGGCACTCAAATGACGT
	rs6684375-F (C-T)	TGGGAATCTGCTCCTCTTCTCTTTTGGGTTGG
	rs6684375-R (C-T)	AGAGGAGCAGATTCCCAGGGGCCCTCCGGCTAAGC
siRNAs (sense)	*CTCF*	UCACCCUCCUGAGGAAUCACCUUAA
	TFAP2A (siRNA-1)	CCGUCUCCGCCAUCCCUAUUAACAA
	TFAP2A (siRNA-2)	AACAUCCCAGAUCAAACUGUA
CRISPR/Cas9	sgRNA 1-F	ACCGTCCTTTCTTCTTTGGACAC
	sgRNA 1-R	AAACGTGTCCAAAGAAGAAAGGA
	sgRNA 2-F	ACCGGCCGCACATTGACATCACC

	sgRNA 2-R	AAACGGTGATGTCAATGTGCGGC
dCas9-KRAB	sgRNA-1-F	ACCGGGGAGCCCTTCCATTCTCG
	sgRNA-1-R	AAACCGAGAATGGAAGGGCTCCC
	sgRNA-2-F	ACCGGCTGATATTAGCAGTGTAC
	sgRNA-2-R	AAACGTACACTGCTAATATCAGC
	sgRNA-3-F	ACCGGCCAATGGGGCATGAGTTG
	sgRNA-3-R	AAACCAACTCATGCCCCATTGGC
RT-qPCR	*LINC00339*-F	GTCCAGATTCCACGAGAGCCTT
	LINC00339-R	GTCTCAGCCACCACCGTCCA
	CDC42-F	GATGGTGCTGTTGGTAAA
	CDC42-R	TAACTCAGCGGTCGTAAT
	CTCF-F	GTGTTCCATGTGCGATTACG
	CTCF-R	TCATGTGCCTTTTCAGCTTG
	TFAP2A-F	GTTCACGCCGATCCATGAAAA
	TFAP2A-R	AGATTGACCTACAGTGCCCAG
ChIP-qPCR	rs6426749-F	ATGTGAAATGCTTACACTGGAGTTC
	rs6426749-R	ATGTGAAATGCTTACACTGGAGTTC

Note: F, forward primer; R, reverse primer; Restriction enzyme site sequences were underlined; For rs6426749, we used fusion PCR[14] to effectively get the long fragment containing both enhancer and *LINC00339* promoter, which was further inserted into the pGL3-basic vector. For rs34963268 and rs6684375, we appended the same restriction enzyme sites to both enhancer and *LINC00339* promoter, which were further inserted into the pGL3-basic vector sequentially.

Table S8. Genetic association with *LINC00339* and BMD for 44 selected SNPs used for multi-instrument based Mendelian randomization analysis

SNP	Chr	Position	eQTL[a]		GWAS[b]	
			P	Beta	*P*	Beta
rs471359	1	21656500	0.010	-0.248	0.006	0.012
rs78885464	1	21807864	0.001	-0.240	0.440	0.003
rs61778393	1	21902436	0.006	0.512	0.003	0.030
rs1130564	1	21952884	0.005	-0.366	0.380	0.009
rs12128206	1	21980091	4.231×10^{-4}	0.631	0.036	0.020
rs60765766	1	22017013	0.010	0.258	0.043	-0.012
rs2315928	1	22189447	0.009	-0.527	0.680	-0.027
rs114537356	1	22214279	0.003	0.739	0.240	-0.014
rs114568494	1	22241660	0.004	-0.650	0.009	0.032
rs6684979	1	22261395	0.008	-0.406	0.240	-0.011
rs35601247	1	22272915	0.004	0.540	0.350	-0.015
rs145444626	1	22287577	2.310×10^{-4}	0.913	0.680	0.000
rs6661287	1	22298481	0.005	0.265	0.740	-0.001
rs12059804	1	22304585	0.002	0.185	0.500	0.003
rs61777960	1	22311348	0.002	-0.184	3.100×10^{-4}	-0.013
rs10917101	1	22314475	1.416×10^{-4}	-0.262	0.036	0.010
rs2865210	1	22342050	0.002	-0.203	0.960	-0.001
rs2255282	1	22352040	1.012×10^{-20}	-0.473	0.011	-0.008
rs116674939	1	22354237	2.753×10^{-7}	0.853	0.044	0.023
rs150153349	1	22355890	0.005	-0.449	0.330	-0.012
rs2473277	1	22361845	1.618×10^{-18}	-0.447	0.012	-0.008
rs2473317	1	22395251	8.522×10^{-4}	-0.252	0.490	0.004
rs16826588	1	22424113	2.482×10^{-4}	0.445	0.074	0.020
rs1046310	1	22443887	2.005×10^{-10}	-0.317	8.900×10^{-4}	-0.010
rs10917161	1	22460208	0.006	-0.454	0.004	0.025
rs113155445	1	22472435	9.761×10^{-4}	-0.262	0.330	0.005
rs4655026	1	22473658	2.811×10^{-8}	-0.288	2.100×10^{-4}	-0.011
rs735475	1	22482230	0.004	0.470	0.680	0.006
rs2807352	1	22495261	3.293×10^{-5}	0.223	6.700×10^{-5}	0.013
rs2982286	1	22506729	0.003	-0.151	5.300×10^{-15}	-0.027
rs140767127	1	22512667	0.008	-0.628	0.170	-0.026
rs115963111	1	22534928	0.001	0.636	0.830	0.002
rs2807331	1	22565967	1.798×10^{-4}	0.198	2.900×10^{-5}	0.014
rs75868741	1	22594676	0.008	0.237	3.300×10^{-8}	-0.027
rs1007243	1	22614839	0.006	0.166	0.001	0.013
rs74816778	1	22641134	0.008	0.451	0.009	-0.025
rs11585537	1	22656868	0.004	-0.222	0.250	-0.005
rs61769163	1	22678805	7.692×10^{-4}	0.388	0.057	0.013
rs4654807	1	22949552	0.009	-0.216	0.730	0.001
rs7549888	1	23004019	0.004	0.153	0.250	-0.003

rs11811882	1	23019404	0.006	0.158	0.570	-0.002
rs75858988	1	23035195	0.006	-0.488	0.011	-0.033
rs111727123	1	23063461	0.007	-0.576	0.300	-0.023
rs76603191	1	23133152	0.003	0.502	0.055	0.012

Note: eQTL[a]: Genetic association with *LINC00339* expression extracted from GTEx whole blood tissue;[1] GWAS[b]: Genetic association with BMD collected from UK Biobank.[4]

Supplemental References

1. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-585.
2. Thompson, B., Varticovski, L., Baek, S., and Hager, G.L. (2016). Genome-Wide Chromatin Landscape Transitions Identify Novel Pathways in Early Commitment to Osteoblast Differentiation. PLOS ONE 11, e0148619.
3. Hon, C.C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. Nature 543, 199-204.
4. 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.
5. Lu, L.L., Liu, Y.J., Yang, S.G., Zhao, Q.J., Wang, X., Gong, W., Han, Z.B., Xu, Z.S., Lu, Y.X., Liu, D., et al. (2006). Isolation and characterization of human umbilical cord mesenchymal stem cells with hematopoiesis-supportive function and other potentials. Haematologica 91, 1017-1026.
6. Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290-294.
7. Rao, Suhas S.P., Huntley, Miriam H., Durand, Neva C., Stamenova, Elena K., Bochkov, Ivan D., Robinson, James T., Sanborn, Adrian L., Machol, I., Omer, Arina D., Lander, Eric S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665-1680.
8. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 163, 1611-1627.
9. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22, 1760-1774.
10. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell 167, 1369-1384.e1319.
11. Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C.B., Krumm, A., et al. (2015). Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nat Meth 12, 71-78.
12. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47, 598-606.
13. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.
14. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989). Engineering hybrid genes

without the use of restriction enzymes: gene splicing by overlap extension. Gene 77, 61-68.