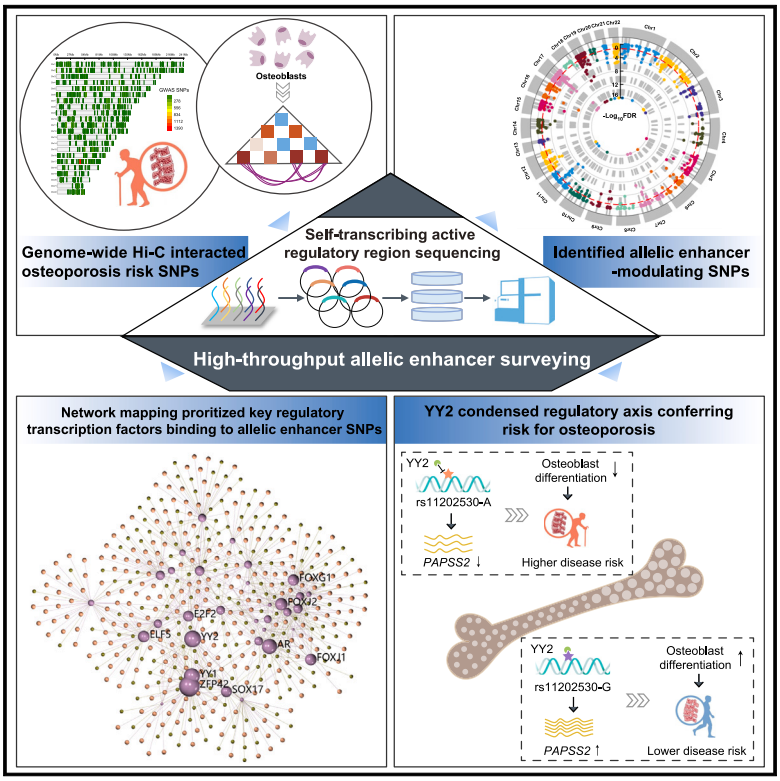
Cell Genomics

Article

Integrative high-throughput enhancer surveying and functional verification divulges a YY2-condensed regulatory axis conferring risk for osteoporosis

Graphical abstract



Highlights

- Massive report assay detected allelic enhancer activities on 319 osteoporosis GWAS SNPs
- 477 chromatin-interaction genes were enriched for roles in osteoporosis pathogenesis
- Network mapping prioritized key TFs binding to enhancer SNPs controlling osteoporosis
- rs11202530-G binds to YY2 to strengthen PAPSS2 expression and osteoblast differentiation

Authors

Xiao-Feng Chen, Yuan-Yuan Duan, Ying-Ying Jia, ..., Rui-Hua Jing, Yan Guo, Tie-Lin Yang

Correspondence

guoyan253@xjtu.edu.cn (Y.G.), yangtielin@xjtu.edu.cn (T.-L.Y.)

In brief

Chen et al. combined chromatin interaction conformation profiling and high-throughput reporter assay to survey osteoporosis GWAS SNPs with allelic enhancer activities. They prioritized key TFs controlling osteoporosis binding to enhancer SNPs and divulged a YY2-condensed regulatory axis conferring risk for osteoporosis through modified *PAPSS2* expression and osteoblast differentiation.

Chen et al., 2024, Cell Genomics *4*, 100501
March 13, 2024
https://doi.org/10.1016/j.xgen.2024.100501

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Integrative high-throughput enhancer surveying and functional verification divulges a YY2-condensed regulatory axis conferring risk for osteoporosis

Xiao-Feng Chen,[1,5] Yuan-Yuan Duan,[1,5] Ying-Ying Jia,[1,5] Qian-Hua Dong,[1] Wei Shi,[1] Yan Zhang,[1] Shan-Shan Dong,[1] Meng Li,[2] Zhongbo Liu,[3] Fei Chen,[1] Xiao-Ting Huang,[1] Ruo-Han Hao,[1] Dong-Li Zhu,[1] Rui-Hua Jing,[4] Yan Guo,[1,*] and Tie-Lin Yang[1,2,6,*]

[1]Key Laboratory of Biomedical Information Engineering of Ministry of Education, Key Laboratory of Biology Multiomics and Diseases in Shaanxi Province Higher Education Institutions, Biomedical Informatics and Genomics Center, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China

[2]Department of Orthopedics, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China

[3]Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an 710004, Shaanxi, China

[4]Department of Ophthalmology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710000, Shaanxi, China

[5]These authors contributed equally

[6]Lead contact

*Correspondence: guoyan253@xjtu.edu.cn (Y.G.), yangtielin@xjtu.edu.cn (T.-L.Y.)

https://doi.org/10.1016/j.xgen.2024.100501

SUMMARY

The precise roles of chromatin organization at osteoporosis risk loci remain largely elusive. Here, we combined chromatin interaction conformation (Hi-C) profiling and self-transcribing active regulatory region sequencing (STARR-seq) to qualify enhancer activities of prioritized osteoporosis-associated single-nucleotide polymorphisms (SNPs). We identified 319 SNPs with biased allelic enhancer activity effect (baaSNPs) that linked to hundreds of candidate target genes through chromatin interactions across 146 loci. Functional characterizations revealed active epigenetic enrichment for baaSNPs and prevailing osteoporosis-relevant regulatory roles for their chromatin interaction genes. Further motif enrichment and network mapping prioritized several putative, key transcription factors (TFs) controlling osteoporosis binding to baaSNPs. Specifically, we selected one top-ranked TF and deciphered that an intronic baaSNP (rs11202530) could allele-preferentially bind to YY2 to augment *PAPSS2* expression through chromatin interactions and promote osteoblast differentiation. Our results underline the roles of TF-mediated enhancer-promoter contacts for osteoporosis, which may help to better understand the intricate molecular regulatory mechanisms underlying osteoporosis risk loci.

INTRODUCTION

Osteoporosis is a common age-related skeletal disease characterized by low bone mineral density (BMD) with increased fragility and fracture risk.[1] Genome-wide association studies (GWASs) have identified over 500 loci associated with fracture[2,3] or BMD at different body positions.[3–6] However, most osteoporosis-associated single-nucleotide polymorphisms (SNPs) reside in noncoding regions, complicating the pinpointing of causal functional SNPs and exact effector genes. Increasing evidences have disclosed that disease-risk variants commonly occurred at enhancer elements to alter gene transcription via spatial enhancer-promoter interaction, and destabilized gene expression could contribute to development and progression of diseases.[7–10] For example, we previously verified that an osteoporosis risk SNP at 1p36.12 could act as an allele-specific enhancer to modulate *LINC00339* expression via long-range loop formation and further influence bone metabolisms through inverse regulation of *LINC00339* and *CDC42*.[11] Therefore, systematic qualification of disease-risk SNPs with enhancer activity impact, as well as ascertainment of their target genes via enhancer-promoter interactions, could strengthen our mechanistic understanding into osteoporosis molecular pathogenesis.

Recent prosperous high-throughput chromatin interaction conformation (Hi-C) assays in diverse disease-relevant cells have provided comprehensive 3D chromatin maps connecting distal genetic variants to their respective target genes,[12–16] which have helped elucidate the long-range regulatory architecture at many diseases risk loci, such as type 2 diabetes,[12] neural diseases,[13] and prostate cancer.[16] However, the precise roles of 3D chromatin organization at osteoporosis risk loci remain largely elusive, perhaps due to the deficiency of chromatin interaction mapping in primary bone-relevant cells. We previously performed high-resolution Hi-C (2-KB) and chromatin immunoprecipitation sequencing (ChIP-seq) in human mesenchymal stem cells (hMSCs) and hMSC-induced osteoblasts

and adipocytes, and we disclosed crucial roles of coordinated change of epigenome and 3D chromatin organization in modulating osteogenesis.[17] The 3D chromatin maps in osteoblasts inspire us to systematically assess roles of chromatin organization at osteoporosis risk loci and to elucidate candidate functional osteoporosis risk variants with distal regulatory effect on target genes through long-range chromatin interactions.

Traditional experimental assays such as dual-luciferase reporter assay or CRISPR could unbiasedly measure enhancer activity of individual candidate SNP fragments.[11] However, they are low throughput and inefficient to systematically uncover regulatory roles at numerous disease risk loci. Recent development of high-throughput reporter assays represented by self-transcribing active regulatory region sequencing (STARR-seq)[18] have demonstrated to be effective in systematically measuring enhancer activity of thousands of DNA sequences simultaneously. STARR-seq relies on the insertion of tested sequences in the 3′ untranslated region of reporter genes and assesses their regulatory activity by directly quantifying their self-transcription using high-through sequencing. We have recently employed an adapted STARR-seq to assess biased enhancer activity of 5,987 noncoding SNP-associated insulin-resistance-relevant phenotypes.[19] Other studies employing STARR-seq have also successfully elucidated regulatory disease risk variants with enhancer activity effect on several other complex diseases, such as prostate cancer,[20] coronary artery disease,[21] and atrial fibrillation.[22] Nevertheless, there is no systematic measurement of enhancer regulatory activities for GWAS SNPs associated with osteoporosis-relevant phenotypes.

Here, we adopted an integrative strategy combing Hi-C profiling,[17] high-throughput enhancer measurement,[19] and multilayer analytical and experimental explorations to systematically dissect enhancer-modulating regulatory SNPs conferring risk for osteoporosis via enhancer-promoter interactions (Figure 1). We firstly fine-mapped 5,642 candidate osteoporosis-associated SNPs with significant chromatin interactions to nearby gene promoters in osteoblasts. Secondly, by utilizing an adapted STARR-seq[19] in human osteoblast-like cell line U2OS, we identified 319 SNPs with allelic enhancer regulatory activities, which showed active epigenetic enrichment in osteoblasts and prevailing genetic regulatory roles for osteoporosis. Thirdly, we identified several candidate regulatory transcription factors (TFs) for osteoporosis etiologies and underscored YY2 as a putative key controlling one through combing allelic motif enrichment and TF-gene regulatory network analyses. Finally, through detailed experimental explorations, we verified one putative YY2-condensed regulatory axis conferring risk of rs11202530 to osteoporosis pathogenesis through regulating *PAPSS2* expression and osteoblast differentiation orchestrated by YY2.

RESULTS

Osteoporosis-associated SNPs with significant promoter chromatin interactions were enriched for active regulatory elements

To explore potential roles of 3D chromatin interactions for osteoporosis-associated SNPs, we induced hMSCs to osteoblast and adipogenic differentiation, and we generated paired significant 3D chromatin interactions (false discovery rate [FDR] < 0.05) and ChIP-seq peak regions (H3K4me1, H3K27ac) in differentiated osteoblasts and adipocytes (Figure S1A; Table S1).[17] Stratified LD score regression (S-LDSC)[23] analysis revealed significant enrichment of heritability in significant 3D chromatin interaction regions (FDR < 0.05) in hMSC differentiated osteoblasts across multiple osteoporosis-relevant traits, including fracture[2,3] and BMD at different body positions (quantitative heel ultrasounds, total body, or total body less head, forearm, and lumbar spine)[3–6] (FDR < 0.05, enrichment ranged from 1.93 to 3.94, Figure S1B). We next prioritized 38,321 candidate SNPs associated with fracture[2,3] or BMD[3–6] through combing conditional GWAS association and fine-mapping analyses (Figure S1C; Tables S2 and S3, STAR Methods), and we found that prioritized SNPs were primarily located within or near significant 3D chromatin interaction regions (FDR < 0.05) in osteoblast cells (Figure S1D). Collectively, these analyses provided evidence for putative critical biological roles for 3D chromatin organization in osteoblasts at osteoporosis risk loci.

To further explore epigenetic characterizations on 5,642 osteoporosis-associated SNPs with significant chromatin interaction (FDR < 0.05) to nearby gene promoters in osteoblasts (Table S4), we compared their overlap with different epigenetic marker peaks in hMSC induced or primary osteoblasts (Table S1) against other osteoporosis-associated SNPs without significant promoter chromatin interactions. We found that promoter chromatin interaction SNPs are significantly enriched (FDR < 0.05, odds ratio [OR] > 1) for active chromatin segments (transcription or enhancer indicative, e.g., 7_Enh), open chromatin region (assay for transposase-accessible chromatin using sequencing [ATAC-seq] or DNase I-hypersensitive site sequencing [DNase-seq]), enhancer-indicative histone modifications (H3K27ac, H3K4me1, H3K4me2, H3K4me3), and CTCF binding while being significantly depleted (FDR < 0.05, OR < 1) for suppressive chromatin segment (e.g., 15_Quies) and H3K9me3 in osteoblasts (Figures S1E and S1F, Fisher's exact test), implying their prevailing endogenous active regulatory activities.

STARR-seq-identified regulatory SNPs with biased allelic enhancer activity effect

To systematically assess putative regulatory activities on 5,642 promoter chromatin interaction osteoporosis-associated SNPs (Table S4), we utilized an adapted STARR-seq using hSTARR-seq_ORI vector with ORI as core promoter based on synthesized oligonucleotide sequences (Figure 2A, STAR Methods). We synthesized 150-bp oligonucleotide sequences centered on reference and alternative allele of each of the 5,642 SNPs to assess their enhancer activities according to our previous report[19] (Figure 2A; Tables S5 and S6). The library containing all SNP oligonucleotides was cloned into hSTARR-seq_ORI vector and transfected into human osteoblast-like cell line U2OS, with plasmid library (input) and transcribed products (output) measured by sequencing (Figure 2A). Specifically, the unique random molecular identifiers (UMIs) were added during reverse transcription or DNA extension to diminish potential bias of PCR duplicates (Figure 2A).

From these sequencing reads (UMIs), we observed significantly high correlation of transcriptional activities among three

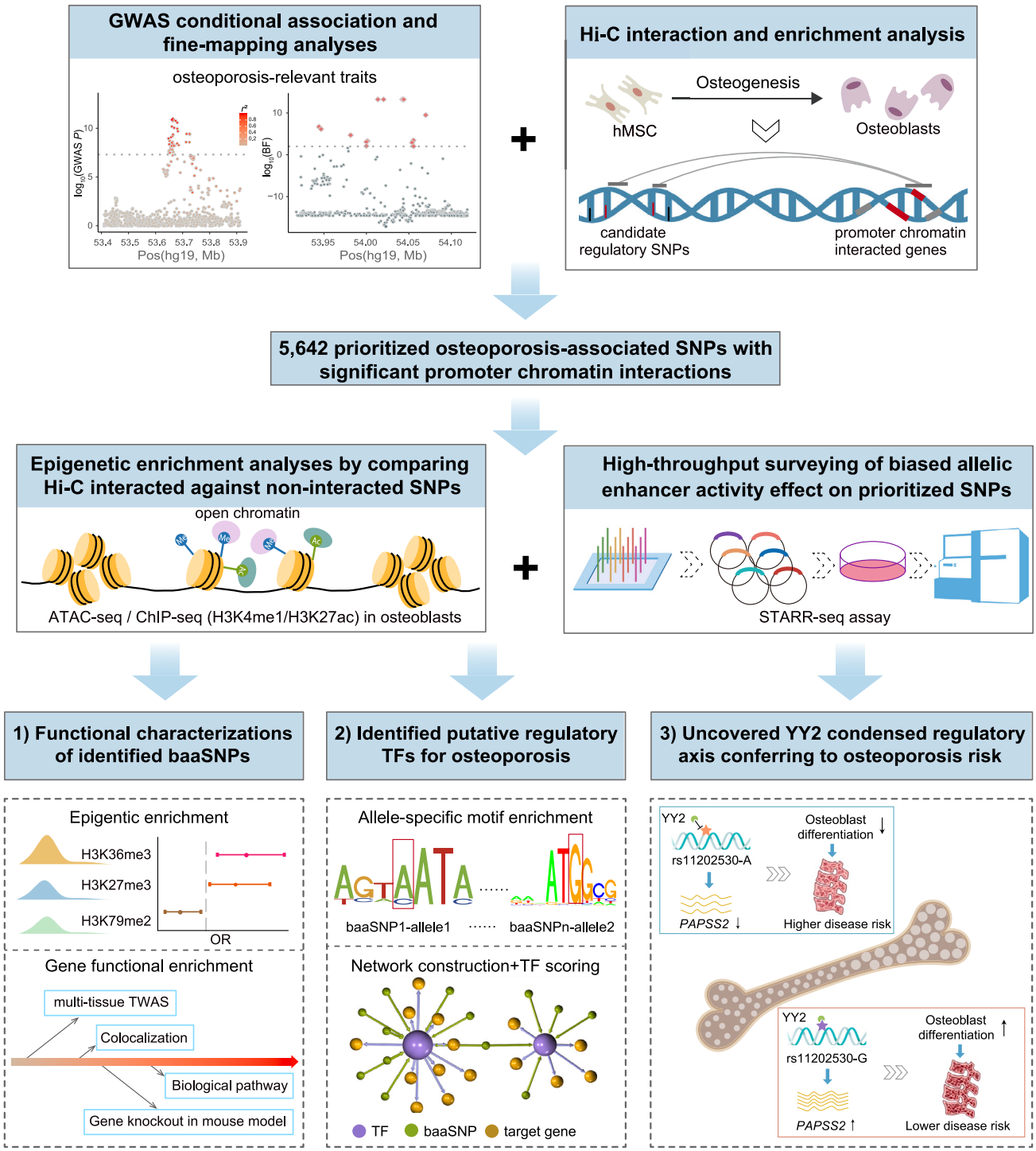


Figure 1. Schematic for the whole study design

Flowchart depicts high-throughput screening for osteoporosis-associated regulatory SNPs with biased allelic enhancer activity effect (baaSNPs), followed by functional characterizations on identified baaSNPs and TF regulatory network analyses, as well as experimental decryption of YY2-condensed regulatory axis conferring risk for osteoporosis. BF: Bayes factor, Hi-C: high-throughput chromatin interaction conformation, hMSC: human mesenchymal stem cell, STARR-seq: self-transcribing active regulatory region sequencing, OR: odds ratio, TF: transcription factor.

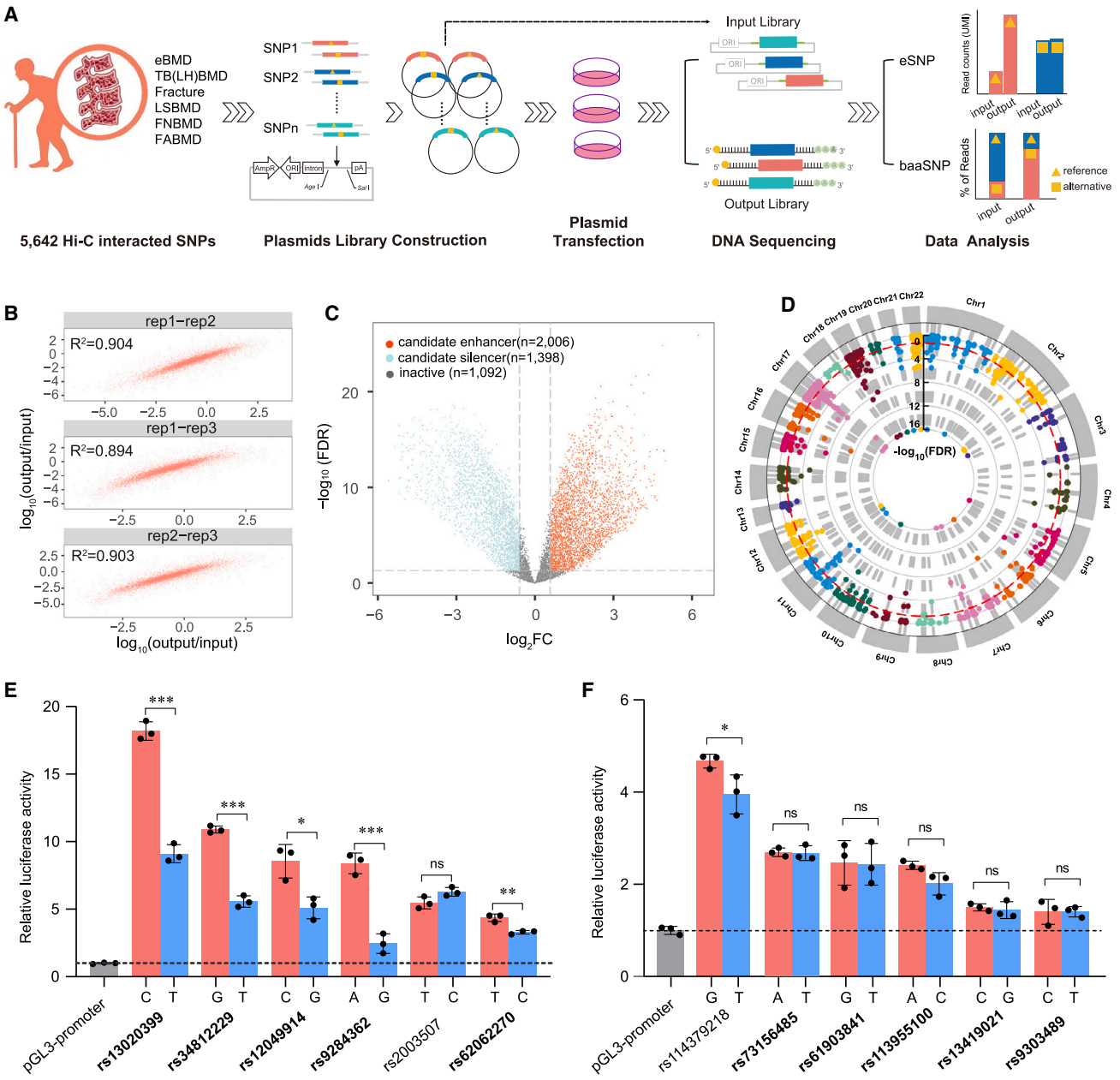


Figure 2. High-throughput identification of osteoporosis-associated regulatory SNPs with allelic-biased enhancer activity effect

(A) Schematization of high-throughput assessment of allelic enhancer activity effect on 5,642 promoter chromatin interaction osteoporosis-associated SNPs by STARR-seq. TB(LH)BMD: total body or total body less head BMD, FABMD: forearm BMD, FNBMD: femoral neck BMD, LSBMD: lumbar spine BMD, eBMD: quantitative heel ultrasound BMD.

(B) Paired Pearson correlation of logarithmic read count ratio of input/output between three biological replications ($p < 2.2 \times 10^{-16}$).

(C) Volcano plot displays regulatory activities for all SNP allele-containing fragments in STARR-seq assay. FC (fold change) represents expression change in output compared with input. SNP fragments with significantly reinforced ($log_2FC > 0.585$, FDR < 0.05) or repressed expression ($log_2FC < -0.585$, FDR < 0.05) analyzed by voom[24] were determined as candidate enhancers (orange circle) or silencers (blue circle), respectively. The rest of the fragments were declared as inactive fragments (gray circle).

(D) Circle diagrams depict significance level ($-log_{10}$FDR) of allelic regulatory activity difference on SNPs showing enhancer activity on at least one allele (eSNP) analyzed by MPRAnalyze.[26]

(E and F) Dual-luciferase reporter assay measuring relative activity on six randomly selected baaSNPs (E) or six randomly selected non-baaSNPs (but eSNPs) (F) in U2OS cells. (*p < 0.05, **p < 0.01, ***p < 0.001, ns: not significant, two-tailed paired Student's t test). Data are presented as mean ± standard deviation. All red bars represent the alleles that had relative higher activities in STARR-seq. SNPs with consistent measured activities in luciferase reporter assay were marked in bold.

replicates (paired Pearson R^2 ranged from 0.894 to 0.904, $p < 2.20 \times 10^{-6}$, Figure 2B). We compared output reads with input reads by voom[24] to evaluate regulatory activity on all SNP-containing fragments, and we detected 2,006 SNPs with potential enhancer activity effect (referred to as eSNPs) for either reference or alternative allele-containing fragments ($log_2FC > 0.585$, FDR < 0.05, Figure 2C; Table S7). Similarly, we detected 1,398 candidate silencer SNPs with potential silencer activity effect ($log_2FC < -0.585$, FDR < 0.05) and referred to other SNPs except for either eSNPs or silencer SNPs as inactive SNPs (n = 1,092, Figure 2C; Table S7). Epigenetic comparison between silencer SNPs and inactive SNPs revealed no enrichment for any active chromatin segments or histone markers or typical repressed histone markers (such as H3K9me3 or H3K27me3) in primary osteoblast cells ($p > 0.05$ by two-sided Fisher's exact test). However, consistent with previous reports,[25] we observed significant enrichment for open chromatin regions in osteoblast cells on candidate silencer SNPs (OR = 1.81, $p = 2.25 \times 10^{-4}$ by two-sided Fisher's exact test), supporting their potential endogenous silencer activities.

To further identify eSNPs with biased allelic enhancer activity effect (referred as baaSNPs), we compared their change of allelic expression ratio between output and input library using MPRAnalyze.[26] We identified 319 candidate baaSNPs from 146 loci (FDR < 0.05, Figure 2D; Table S8). To validate the allelic enhancer activity effect on the identified baaSNPs, we randomly selected 6 baaSNPs and examined their allelic regulatory activity by dual-luciferase reporter assay in U2OS cells (Table S6). The results revealed that 5 of 6 baaSNPs showed significantly different regulatory activity with consistent direction of allelic activity change with STARR-seq ($p < 0.05$, Figure 2E; Table S6). For comparison, we also randomly chose 6 non-baaSNPs by STARR-seq and detected no significant differential luciferase activities between reference and alternative allele for 5 of 6 selected SNPs in U2OS cells (Figure 2F). Taken together, these results supported the credibility of our STARR-seq assays in assessing regulatory activities of baaSNPs.

Functional characterizations unraveled active epigenetic enrichment on baaSNPs

To verify the enhancer regulatory activities on identified baaSNPs, we firstly scrutinized their nearby epigenetic annotations in osteoblast cells (Table S1). The majority of baaSNPs are located on or near open chromatin region or enhancer-indicative markers (H3K4m1, H3K27ac) in primary osteoblast cells (71.2% within 1 kb and 85.0% within 2 kb of nearest peaks, Figure S2A), implying their pervasive endogenous enhancer activities. Further comparison of chromatin segment annotation between baaSNPs and inactive SNPs revealed significantly enrichment for enhancers (7_Enh) while being depleted for weak-repressed Polycomb (14_ReprPCWk) in osteoblast cells on baaSNPs ($p < 0.05$ by Fisher's exact test, Figure 3A). We also detected significant enrichment for several active histone markers (H3K36me3, H3K79me2)[27,28] while being depletion for typical suppressive histone marker (H3K27me3) in osteoblast cells on baaSNPs against inactive SNPs ($p < 0.05$, Figure 3B). Further S-LDSC[23] analysis revealed significant enrichment of heritability across multiple osteoporosis-relevant traits in H3K36me3 and H3K79me2 peaks in osteoblast cells ($p < 0.05$, enrichment ranged from 1.48 to 3.25, Figures 3C and 3D), supporting potential biological roles for these two epigenetic markers at osteoporosis risk loci.

We noted that there was comparable annotation of open chromatin regions or two typical enhancer-indicative epigenetic marker (H3K4me1, H3K27ac) peaks between baaSNPs and inactive SNPs (Figure 3B). However, comparison of allelic effect of baaSNPs (represented by absolute log_2FC) revealed significantly higher regulatory effect on those baaSNPs annotated within H3K27ac peaks ($p = 0.04$ by Wilcoxon rank-sum test, Figure S2B) compared with other non-annotated baaSNPs, as well as distinctly higher regulatory effect on baaSNPs overlaying open chromatin region or H3K4me1 peaks against non-annotated baaSNPs, although not significant (Figure S2B). Moreover, comparison of average reads signals surrounding all eSNPs revealed significantly higher reads on baaSNPs for open chromatin regions ($p = 4.12 \times 10^{-4}$ by Wilcoxon rank-sum test, Figure S2C), as well as discernibly higher reads for H3K4me1 or H3K27ac compared with non-baaSNPs, although not significant (Figure S2C). Collectively, these analyses supported the crucial enhancer regulatory effect on identified baaSNPs.

Multiomics analyses unveiled genetic regulatory roles for osteoporosis on baaSNPs

To verify distal regulatory effect of baaSNPs on their chromatin interaction genes (n = 477, Table S8), we firstly compared significant chromatin interactions between baaSNPs or non-baaSNPs (but eSNPs) and gene promoters. We detected significantly higher interaction intensities on baaSNPs (Wilcoxon rank-sum test, $p < 0.05$, Figure 3E), implying their critical distal enhancer regulatory roles. We further examined *cis*-expression quantitative trait locus (*cis*-eQTL) association from GTEx (V8),[29] and we found that most baaSNPs (65.83%) also had nominally significant *cis*-eQTL association ($p < 0.05$) with at least one promoter chromatin interaction gene (254 total) in a consistent direction of allelic activity effect with STARR-seq (Table S9). To further explore potential causal genetic regulatory roles in osteoporosis on chromatin interaction genes, we employed the fast enrichment estimation aided colocalization analysis (fastENLOC)[30] and identified 89 potential causal osteoporosis-associated chromatin interaction genes (locus-level colocalization probability [LCP] > 0.1) (Table S10). Complementary multi-tissue transcriptome-wide association analysis (TWAS) using S-MulTiXcan[31] further prioritized 202 potential osteoporosis-associated chromatin interaction genes (Bonferroni corrected $p < 0.05$, Table S11). Compared to all nearby genes (1,000 kb surrounding), chromatin interaction genes of baaSNPs were significantly higher enriched for osteoporosis-associated genes predicted by either fastENLOC (OR = 3.19, $p = 1.13 \times 10^{-16}$ by Fisher's exact test, Figure 3F) or S-MulTiXcan (OR = 5.37, $p = 4.05 \times 10^{-44}$ by Fisher's exact test, Figure 3F), implying their potential critical regulatory roles in osteoporosis pathogenesis.

To further interrogate putative biological roles in osteoporosis pathogenesis on chromatin interaction genes of baaSNPs, we performed Gene Ontology (GO) pathway enrichment analysis, and we found that most of the top significantly enriched

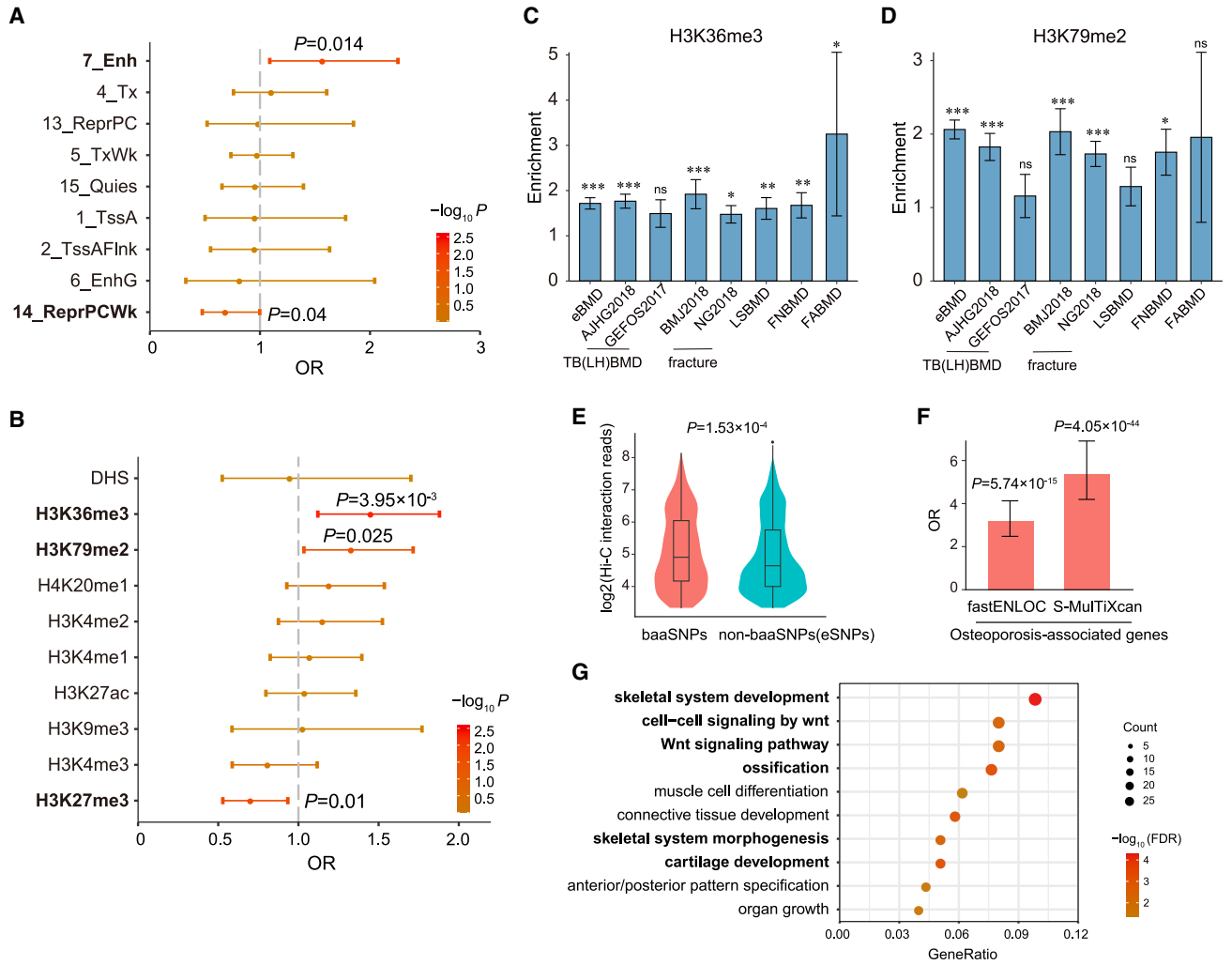


Figure 3. Functional characterizations on baaSNPs and their promoter chromatin interaction genes

(A and B) Comparison of percentage of baaSNPs against inactive SNPs overlying chromatin segments (HMM15) (A) or different histone peak regions (B) in osteoblast cells. Odds ratios (ORs) with 95% confidence interval were shown with significant markers marked in bold (two-sided Fisher's exact test).

(C and D) Estimates of heritability and standard errors by S-LDSC[23] across multiple osteoporosis-relevant traits in peaks of two significantly higher enriched markers (H3K36me3, H3K79me2) in osteoblast cells in (B). Enrichment results are separated by trait (**p < 0.01, ***p < 0.001, ns: not significant).

(E) Violin plot shows comparison of logarithmically significant Hi-C chromatin interaction reads linking gene promoter to baaSNPs against non-baaSNPs (Wilcoxon rank-sum test).

(F) Comparison of percentage of candidate osteoporosis genetically regulatory genes supported by fastENLOC (LCP > 0.1)[30] or S-MulTiXcan (Bonferroni corrected p < 0.05)[31] on promoter chromatin interaction genes or all nearby genes (1,000 kb surrounding) of baaSNPs. Data are presented as mean ± standard deviation. ORs with 95% confidence interval were shown (one-sided Fisher's exact test).

(G) Bar diagram depicts top 10 significantly enriched (FDR < 0.05) Gene Ontology (GO) biological processes on promoter chromatin interaction genes of baaSNPs. Pathways related to skeletal development or osteoblast differentiation were marked in bold.

biological pathways are related to skeletal development or osteoblast differentiation, such as skeletal system development, ossification, and Wnt signaling pathway[32] (FDR < 0.05, Figure 3G). Analogous osteoporosis-relevant pathway enrichment was observed when the located/nearest genes of their chromatin interaction baaSNPs were excluded (FDR < 0.05, Figure S3A). We also scrutinized the mouse phenotypic annotation from the International Mouse Phenotyping Consortium (IMPC, release 19)[33] and detected 26 genes whose knockout in mouse model could lead to abnormal skeletal development or bone structure (Table S12). In summary, we found 237 putative osteoporosis-associated chromatin interaction genes (Figure S3B; Table S12), which might provide many mechanistic hypotheses conferring risk of baaSNPs to osteoporosis pathogenesis. For example, our STARR-seq assay suggested biased higher enhancer activity on non-risk allele of one baaSNP at 7p14.1 (rs1721391-G, $p = 1.30 \times 10^{-217}$, and beta = 0.073 on eBMD, Table S8), which might enhance expression of one distal gene (*EPDR1*) through chromatin interactions and promote osteoblast differentiation[34] (Figure S3C). Consistently, knockout of *Epdr1* in

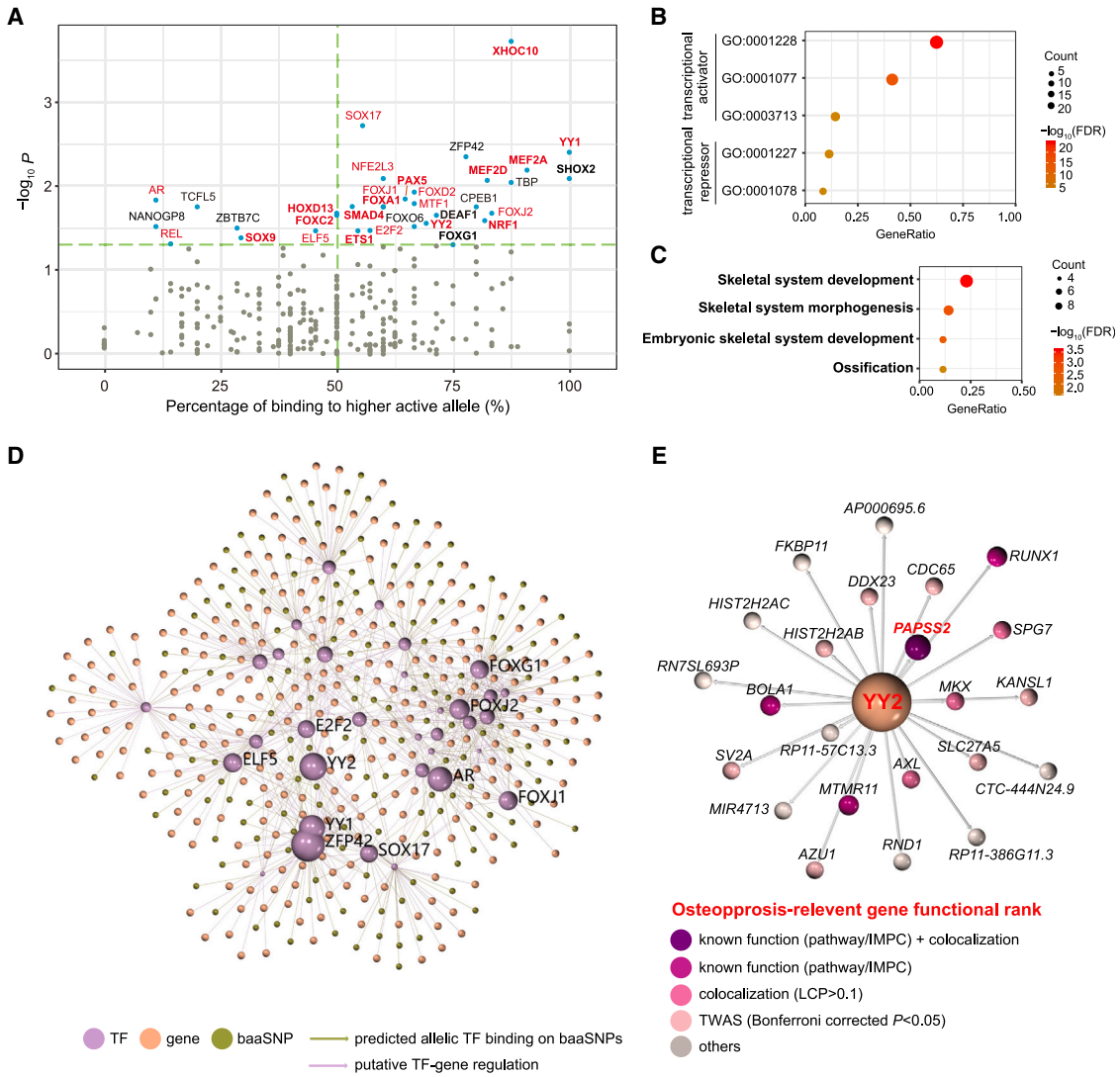


Figure 4. Motif enrichment and regulatory network analyses prioritized YY2 as one putative controlling TF for osteoporosis

(A) Scatterplot depicts enrichment analyses of predicted allele-specific binding motifs on baaSNPs against inactive SNPs by one-sided Fisher's exact test. The green dotted line indicated criteria for retaining nominally significant TFs ($p < 0.05$, OR > 1). All TFs annotated within GO molecular functions related to transcriptional activator/coactivator activity were marked in red. TFs with known roles in osteoblast differentiation or bone metabolism were marked in bold.

(B) Bar diagram depicts comparison of significantly enriched GO molecular function terms related to transcriptional activator/coactivator or transcriptional repressor/corepressor on all enriched TFs in (A).

(C) Bar diagram depicts selected enriched GO biological processes related to skeletal system development (FDR < 0.05) on all enriched TFs in (A).

(D) Network diagram connecting baaSNPs to 33 significantly enriched TFs and their putative regulatory effecter genes. The sizes of TF nodes represent their calculated combined weighting scores (STAR Methods). Network was visualized using Graphia[53] with top 10 highest scored TFs marked.

(E) Extracted TF-gene interaction subnetwork of YY2 from (D). The colors of gene nodes represent different osteoporosis-relevant functional supports. One gene (*PAPSS2*) with all three types of functional support was highlighted in red.

mouse led to short tibia (Figure S3C). In another example, a non-risk allele of one baaSNP at 2p16.3 (rs13035558-T, fine-mapping $Log_{10}FC = 2.38$, $p = 7.80 \times 10^{-8}$, and beta = 0.011 on eBMD, Table S8) displayed biased higher enhancer activity, which may augment expression of *FBXO11* through chromatin interactions and further promote osteoblast differentiation[35] (Figure S3D). Consistently, knockout of *Fbxo11* in mouse could result in decreased BMD,abnormal bone mineralization, and short tibia (Figure S3D).

Motif prediction and network mapping prioritized candidate regulatory TFs for osteoporosis etiologies

To assess potential roles of TFs in mediating allelic enhancer regulatory effect by baaSNPs, we predicted their allelic binding motif using MEME Suite[36] and performed motif enrichment analysis. We detected 33 TFs with nominally significantly higher enrichment for allelic binding to baaSNPs against inactive SNPs ($p < 0.05$ by Fisher's exact test, Figure 4A; Table S13). Analysis of percentage of binding to higher active alleles of

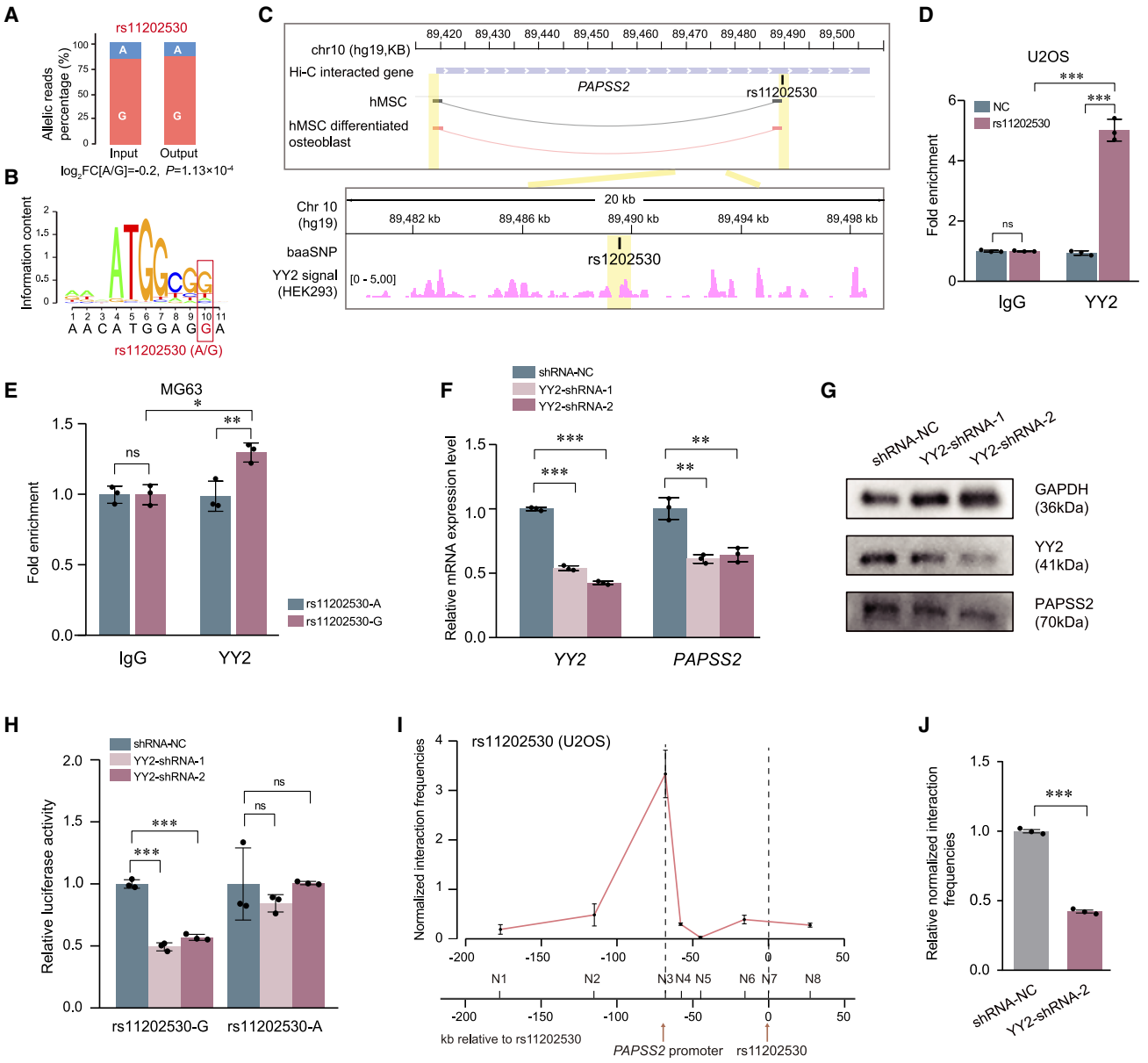


Figure 5. YY2 preferentially binds to rs11202530-G to strengthen its enhancer effect on *PAPSS2* expression

(A) Bar shows comparison of allelic expression ratio of rs11202530 between STARR-seq output and input library samples evaluated by MPRAnalyze.[26]

(B) Graphic representation of rs11202530-G resides within YY2 DNA-binding motif predicted by FIMO from MEME Suite toolkit.[36]

(C) Comparison of significant Hi-C chromatin interactions linking rs11202530 to *PAPSS2* promoter region in hMSCs or hMSC-differentiated osteoblasts or adipose tissue (no significant interaction). Genomic region surrounding rs11202530 was zoomed in with track of ChIP-seq signals on YY2 in HEK293 cell from ENCODE[44] depicted below.

(D) ChIP-qPCR results of YY2 binding at rs11202530 region in U2OS cells. Primers specifically targeting rs11202530 or RPL30 exon region (negative control [NC]) were used. The binding efficiency of YY2 is shown as fold enrichment over IgG.

(E) Allele-specific ChIP-qPCR indicated allele-specific binding of YY2 binding at rs11202530 in MG63. Primers specifically targeting rs11202530-A or G were used.

(F) RT-qPCR assay revealed the effect of *YY2* knockdown on *PAPSS2* expression by two independent shRNAs in U2OS cells.

(G) Western blotting analysis showed the effect of *YY2* knockdown on PAPSS2 expression in U2OS cells. GAPDH was used as loading control.

(H) Dual-luciferase reporter assay comparing regulatory activity on *PAPSS2* promoter inserted by rs11202530-G or rs11202530-A co-transfected with two independent shRNAs of YY2 or shRNA-NC in U2OS cells.

(legend continued on next page)

baaSNPs on enriched TFs revealed that most TFs (25/33) showed preferential binding to more highly active alleles of baaSNPs, implying their potential roles in reinforcing enhancer activity (Figure 4A; Table S13). Consistently, functional annotation from GO revealed that most TFs were involved in molecular function terms related to transcriptional activator or coactivator activity (23/33, FDR < 0.05, Figures 4A and 4B; Table S13). We noted that many enriched TFs had known roles in osteoblast differentiation or bone metabolism (Figure 4A; Table S13), such as HOXC10,[37] MEF2A,[38] or YY2.[39] Consistently, GO biological process enrichment analysis revealed that enriched TFs were significantly enriched for several bone-relevant pathways including skeletal system development and ossification (Figure 4C). To further explore putative roles in osteoblast differentiation for enriched TFs, we compared their expression change in differentiated osteoblast for different time points (0.5–96 h) against non-differentiated cells (Table S1),[40] and we detected 10 TFs (CPEB1, ETS1, FOXC2, FOXD2, HOXC10, MEF2A, NFE2L3, SMAD4, SOX9, and YY1) with significantly changed expression (FDR < 0.05, Table S14). Collectively, these analyses prioritized several regulatory TFs on baaSNPs with putative roles in osteoporosis pathologies.

To further assess putative genetic regulatory roles of enriched TFs for osteoporosis and to prioritize candidate key controlling TFs, we constructed a compounded directed regulatory network connecting baaSNPs to enriched TFs and their corresponding putative regulatory effecter genes (STAR Methods). For 162 baaSNPs with allele-specific motif prediction on 33 enriched TFs, we defined their chromatin interaction genes as putative regulatory genes targets. The network comprised 162 baaSNP nodes, 33 transcription factor nodes, 271 target gene nodes, and their 929 corresponding regulatory connections (324 baaSNP-TF binding and 605 TF-gene regulatory pairs, Figure 4D). To rank putative functionality of TFs in the network, we calculated a combined weighting score for each TF through incorporating biased enhancer effect of baaSNPs, allele-specific motif prediction scores, and intensity of baaSNP-gene promoter chromatin interactions for implicated baaSNP-TF-gene regulatory axes (STAR Methods, Table S15). Notably, all top three highest scored TFs (ZFP42, YY2, YY1) belong to more than 3 adjacent zinc finger factor families (Table S13). The highest scored TF, ZFP42 (score = 6.61, Table S15), was a known regulator of human stem cell pluripotency.[41] However, its role in osteoporosis etiologies remains unknown. In contrast, the second top-ranked TF, YY2 (score = 5.33, Table S15) had been validated to play positive roles in promoting osteoblast differentiation by upregulating transcription activity of Osx.[39] The third top-scored TF, YY1 (score = 5.01, Table S15), was a homolog of YY2 and acts as a known critical structural regulator of active enhancer-promoter loop formation.[42]

Subsequently, we selected YY2 to explore its potential genetic regulatory roles in osteoporosis pathogenesis. The TF-gene subnetwork of YY2 encompassed 23 putative regulatory gene targets (Figure 4E). Functional analyses revealed that the majority of them (60.87%, 14/23) were involved in osteoporosis-related biological pathways or displayed skeletal phenotypic abnormalities in gene knockout mouse models, or they showed putative causal genetic regulatory effect for osteoporosis by colocalization analyses, or TWAS suggested genetic association with osteoporosis (Figure 4E; Table S15), indicating the crucial regulatory roles of YY2 for osteoporosis. In contrast, only 38.89% and 37.04% of regulatory genes on ZFP42 or YY1 showed at least one of above osteoporosis-relevant functionality supports (Table S15). To further verify putative enhancer regulatory roles in osteoblasts for YY2, we predicted its whole-genome active binding sites within footprinting regions in osteoblasts using Wellington.[43] Compared with randomly chosen regions, we detected significant enrichment for several typical enhancer-indicative histone markers (e.g., H3K4me3, H3K27ac, $p < 0.001$) while being depleted for typical suppressive markers (H3K9me3, H3K27me3, $p > 0.05$) in osteoblast cells on YY2 footprints (STAR Methods, Figure S4A). We also observed significant enrichment for significant chromatin interaction regions and CTCF binding sites in osteoblasts on YY2 footprints ($p < 0.001$, Figures S4A and S4B). Additional analyses for chromatin interaction genes of predicted YY2 footprints revealed significant enrichment for osteoblast differentiation-related biological pathways (cell-cell signaling by wnt, Wnt-signaling pathway)[32] ($p < 0.05$, Figure S4C), implying putative roles of YY2 in osteoblast differentiation. Collectively, these analyses coupled risk of baaSNPs to osteoporosis etiologies and prioritized several candidate key regulatory TFs (particularly YY2) via orchestrating the enhancer-promoter regulatory network in osteoblasts.

YY2 could preferentially bind to rs11202530-G to augment *PAPSS2* expression

Considering the highly ranked weighting score of YY2 within the regulatory network as well as the probable functional relevance to osteoporosis on *PAPSS2* (colocalization, TWAS, and osteoporosis-related pathway, Figure 4E; Table S12), we selected rs11202530-YY2-*PAPSS2* regulatory axis (Table S15) for detailed experimentally verification. The rs11202530 existed in the *PAPSS2* intronic region and showed strong linkage disequilibrium ($r^2 > 0.98$) with one conditionally independent eBMD association signal (rs10887745).[3] Rs11202530-G was associated with increased eBMD (beta = 0.016, $p = 1.20 \times 10^{-17}$, Table S3).[3] Our STARR-seq assay suggested biased higher enhancer effect on rs11202530-G compared with A allele (Figure 5A), which also showed G-allele-preferential binding to YY2 according to the motif prediction by MEME Suite[36] (Figure 5B). *PAPSS2* promoter showed osteoblast-specific chromatin interactions with rs11202530 compared with hMSC differentiated adipocytes (Figure 5C), implying potential regulatory roles of YY2 in mediating osteoblast-specific

(I) Chromosome conformation capture (3C) assay in U2OS cells. Normalized chromatin interaction frequencies between rs11202530 region as the anchor point (N7) and *PAPSS2* promoter region (N4) or six other neighboring *Hind* III sites as negative controls (N1, N2, N3, N5, N6, and N8).

(J) 3C assay shows comparison of normalized interaction frequency differences between rs11202530 region and *PAPSS2* promoter region in *YY2*-inhibited (*YY2*-shRNA-2) and control (shRNA-NC) U2OS cells. Data are presented as mean ± standard deviation in (D)–(F) and (H)–(J), *p < 0.05, **p < 0.01, ***p < 0.001, ns: not significant, two-tailed paired Student's t test.

rs11202530-*PAPSS2* chromatin interactions. We also scrutinized publicly available ChIP-seq on YY2 from ENCODE portal,[44] and we observed strong ChIP-seq signals surrounding rs11202530 in HEK293 cells (Figure 5C). To experimentally verify the allelic binding affinity of YY2 on rs11202530, we applied ChIP-qPCR and allele-specific ChIP-qPCR in rs11202530-GG U2OS and rs11202530-GA MG63 cells, respectively (genotyping results showed in Figure S5A). The results indicated that YY2 could bind to rs11202530 region in U2OS cells ($p < 0.001$, Figure 5D). Particularly, YY2 prefer to bind to rs11202530-G allele than A allele in MG63 cells ($p < 0.01$, Figure 5E). Subsequently, to evaluate the role of YY2 in mediating *PAPSS2* expression, we inhibited *YY2* by two independent short hairpin RNAs (shRNAs) in U2OS cells. The results showed that the expression of *PAPSS2* significantly declined both in mRNA and protein levels in *YY2*-knockdown U2OS cells ($p < 0.01$, Figures 5F and 5G). In addition, we transfected pGL3-*PAPSS2* promoter or rs11202530-G or A allele luciferase plasmid into control or *YY2*-knockdown U2OS cells. The results displayed that rs11202530-G allele of *YY2*-knockdown cells could significantly reduce *PAPSS2* expression compared with control cells, while A allele had no obvious effect on *PAPSS2* ($p < 0.001$, Figure 5H).

To further verify the long-range chromatin interactions connecting rs11202530 to *PAPSS2* promoter, we performed chromosome conformation capture (3C) assay in U2OS cells. We found that rs11202530 possessed the strongest chromatin interaction with *PAPSS2* promoter in comparison with other six randomly selected genomic regions (normalized interaction frequency = 3.36, STAR Methods, Figure 5I). We also performed 3C assay in *YY2*-inhibited and control U2OS cells to compare their interaction frequencies differences and found that *YY2* knockdown could significantly attenuate chromatin interactions between rs11202530 and *PAPSS2* promoter region ($p < 0.001$, Figure 5J), indicating the putative roles of YY2 in strengthening long-range chromatin interactions connecting rs11202530 to the *PAPSS2* promoter region. Collectively, these assays provided evidence for potential allele-preferable binding affinity of YY2 to rs11202530-G to augment enhancer activity and to strengthen *PAPSS2* expression through long-range chromatin interactions, which suggested potential genetic regulatory roles of YY2 for osteoporosis.

rs11202530 could act as an allele-biased enhancer to regulate *PAPSS2* expression and to impact osteoblast differentiation

We next performed experimental validation for allelic enhancer regulatory effect of rs11202530 on *PAPSS2* expression. We observed osteoblast-specific H3K27ac signals compared with hMSCs or hMSC-differentiated adipocytes near rs11202530 region (Figure 6A), indicating its osteoblast-specific enhancer effect. *Cis*-eQTL analysis from GTEx[29] revealed significant genetic association with *PAPSS2* expression on rs11202530 in 18 different tissues ($p < 0.05$, Figure 6B), including 15 tissues in consistent direction of allelic activity effect with STARR-seq (Figure 6B; Table S9). To validate allelic enhancer regulation effect of rs11202530 on *PAPSS2*, we firstly performed dual-luciferase reporter assay in U2OS cells. The results exhibited that rs11202530 showed significantly reinforced enhancer activity compared with *PAPSS2* promoter plasmid, and rs11202530-G displayed significant higher enhancer activity on *PAPSS2* promoter compared with the other allele ($p < 0.01$, Figure 6C). Since rs11202530 was homozygous GG in U2OS cells (Figure S5A), we further deleted the genomic region harboring rs11202530 by CRISPR-Cas9, and we found that *PAPSS2* expression was significantly diminished in rs11202530-deleted U2OS cells ($p < 0.001$, Figure 6D). Deletion efficiency of two paired single-guide RNAs (sgRNAs) was confirmed by PCR (Figures 6D and S5B). These results demonstrated that rs11202530-G could act as allelic-biased enhancer to strengthen *PAPSS2* expression in U2OS cells.

Furthermore, we isolated human primary osteoblasts from cancellous bone tissue of human lumbar vertebra (STAR Methods) and deleted the region harboring rs11202530 by CRISPR-Cas9 (heterozygosis GA on rs11202530, Figure S5A). The *PAPSS2* expression was significantly decreased in rs11202530-deleted human primary osteoblasts ($p < 0.001$, Figure 6D). Deletion efficiency of two paired sgRNAs was confirmed by PCR (Figures 6D and S5B). To further explore the potential regulatory effect of rs11202530 in osteoblast differentiation, we measured change of osteoblast differentiation marker gene expression in rs11202530-region-deleted human primary osteoblasts. We found that several marker genes, including *ALP*, *OCN*, *OSX*, *RUNX2*, and *COL1A1* were significantly diminished in rs11202530-region-deleted human primary osteoblasts ($p < 0.05$, Figure 6E). We also detected significantly diminished expression of several osteoblast differentiation marker genes (*ALP*, *OCN*, *OSX*, and *RUNX2*) in rs11202530-region-deleted U2OS cells ($p < 0.05$, Figure S5C). Further alkaline phosphatase (ALP) staining exhibited that deletion of the region harboring rs11202530 could decrease potentiality for osteoblast differentiation in human primary osteoblasts (Figure 6F). Collectively, these experimental results supported the putative genetic regulatory roles of rs11202530 in osteogenesis.

YY2 may play roles in promoting human osteoblast differentiation through positive regulation of *PAPSS2*

The above experimental results suggested one YY2-condensed regulatory axis conferring osteoporosis risk. We next generated ovariectomy (OVX)-induced osteoporosis mice and control mice with parietal ovarian fat tissue removed, and we detected significantly lower expression of both *Yy2* and *Papss2* in osteoporosis mice compared with normal ones ($p < 0.05$, Figure 7A), suggesting their putative protective roles in osteoporosis. To further verify candidate roles of YY2 in osteoblast differentiation, we isolated human primary osteoblasts from cancellous bone tissue of human lumbar vertebra and induced differentiation at different time points (STAR Methods). Differentiation of primary osteoblasts was confirmed by ALP staining ($p < 0.05$, Figure 7B). We observed significantly reinforced *YY2* expression in osteoblasts after differentiation for 7 or 14 days ($p < 0.05$, Figure 7B). We further overexpressed YY2 in primary osteoblasts and detected significantly increased expression of *PAPSS2* and several osteoblast differentiation marker genes (*ALP*, *OCN*, *RUNX2*, and *COL1A1*) ($p < 0.05$, Figure 7C). In contrast, inhibition of *YY2* by two independent shRNAs in primary osteoblasts caused significantly decreased expression of *PAPSS2* and several osteoblast

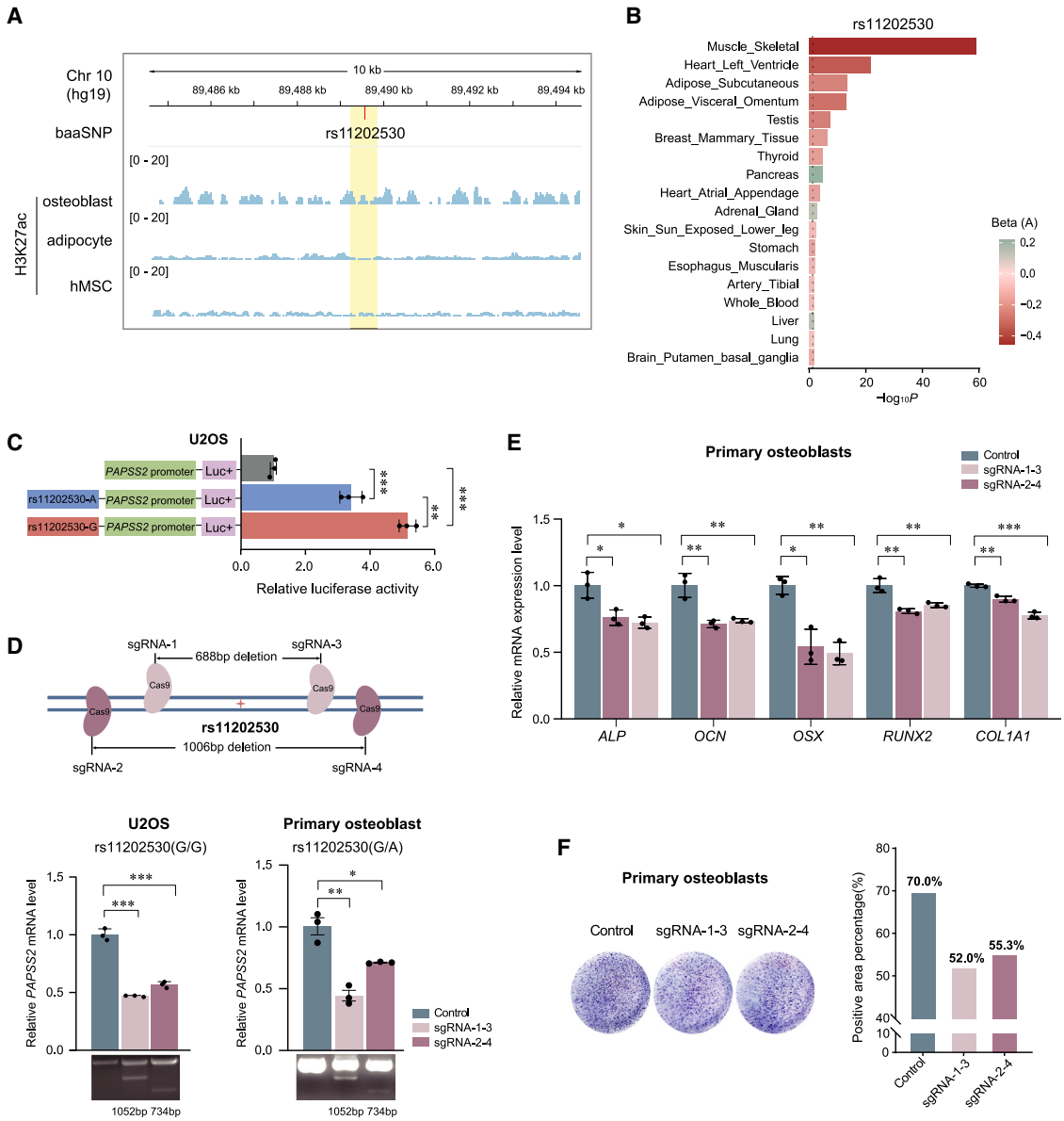


Figure 6. Rs11202530 acts as an allele-biased enhancer to regulate *PAPSS2* expression and osteoblast differentiation

(A) Comparison of H3K27ac histone modification in hMSCs or hMSC-differentiated osteoblasts or adipose tissue surrounding rs11202530.

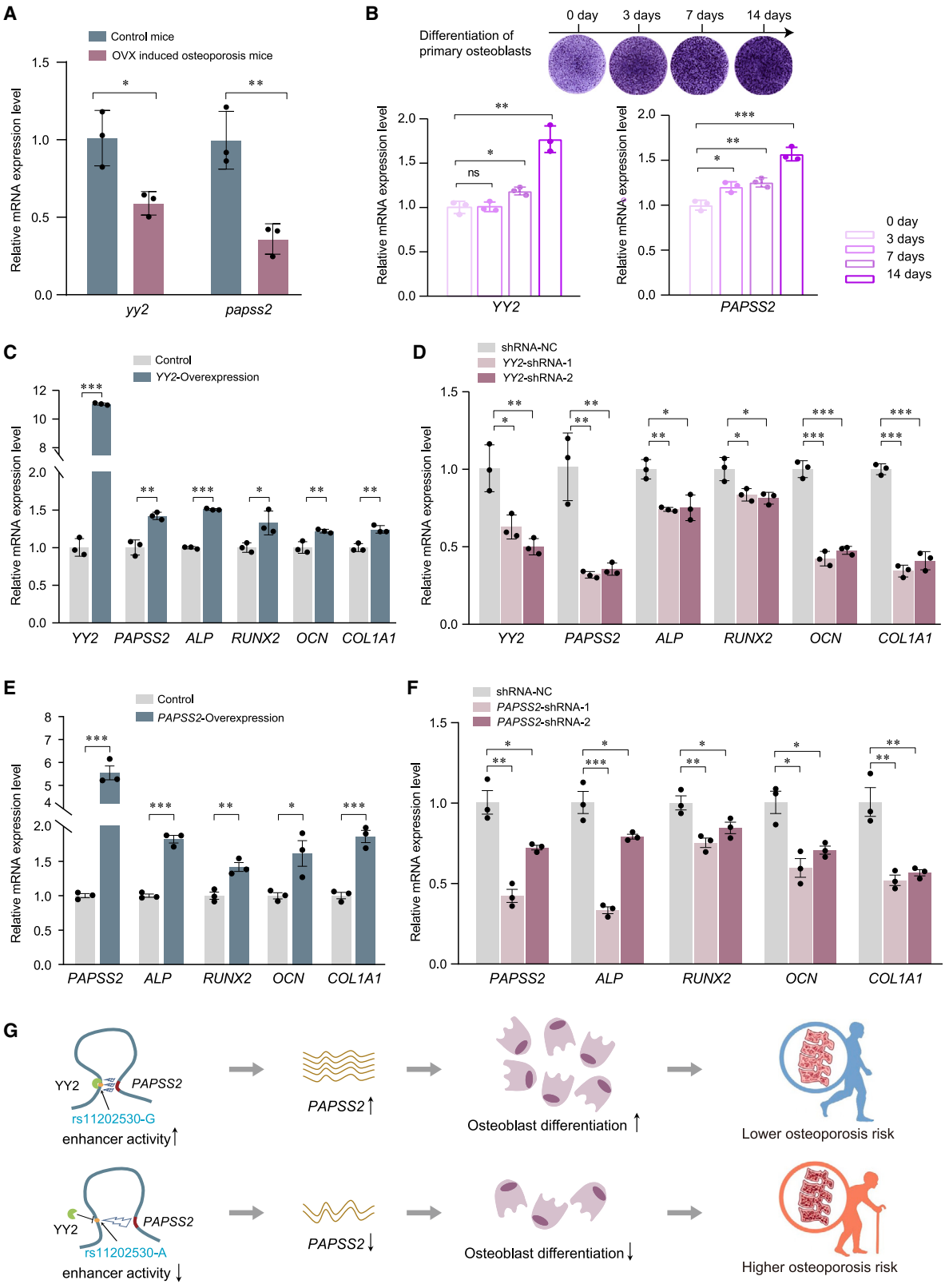
(B) Bar diagrams display all significant *cis*-QTL associations ($p < 0.05$ by linear model controlling for covariates) between genotypes of rs11202530 and *PAPSS2* expression from GTEx (V8),[29] and effect sizes are indicated by shade of colors.

(C) Dual-luciferase reporter assay shows allelic-biased enhancer activity effect on rs11202530 compared with pGL3-*PAPSS2*-promoter in U2OS cells. Luciferase signal was computed as the ratio of firefly luciferase activity to Renilla signal, and relative activity was normalized by pGL3-*PAPSS2*-promoter.

(D) RT-qPCR shows effect of deletion of the genomic region surrounding rs11202530 by CRISPR-Cas9 on *PAPSS2* expression at mRNA level in U2OS or human primary osteoblasts isolated from healthy cancellous bone tissue (STAR Methods). The upper cartoon shows relative position of targeting sites for different pairs of sgRNAs. The bottom gel showed the deletion efficiency of two paired sgRNAs by amplifying a flanking region of 1740 bp (rs11202530) and validated the truncated short regions by PCR.

(E) RT-qPCR displays effect of deletion of the genomic region surrounding rs11202530 by CRISPR-Cas9-decreased expression of osteoblast differentiation marker gene expression (*ALP*, *OCN*, *OSX*, *RUNX2*, and *COL1A1*) at mRNA level in human primary osteoblasts.

(F) Alkaline phosphatase (ALP) staining revealed the effect of deletion of the region surrounding rs11202530 by CRISPR-Cas9 on ALP activity in human primary osteoblasts. The ALP staining results were analyzed using ImageJ software. Data are presented as mean ± standard deviation in (C)–(E), *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns: not significant, two-tailed paired Student's t test.



(legend on next page)

differentiation marker genes (*ALP*, *OCN*, *RUNX2*, and *COL1A1*) ($p < 0.05$, Figure 7D). Considering the positive regulatory effect of YY2 on *PAPSS2* (Figures 5, 7C, and 7D), we further explored potential roles of *PAPSS2* in osteoblast differentiation. Consistently, we detected significant continuously increased *PAPSS2* expression in osteoblasts after differentiation for 3, 7, or 14 days ($p < 0.05$, Figure 7B). Besides, overexpression or inhibition of *PAPSS2* in primary osteoblasts caused significantly enhanced or attenuated expression of several osteoblast differentiation marker genes (*ALP*, *OCN*, *RUNX2*, and *COL1A1*) ($p < 0.05$, Figures 7E and 7F). Collectively, these results supported candidate roles in promoting human osteoblast differentiation for YY2, which provided compelling experimental evidence for one putative YY2-condensed regulatory axis conferring risk of rs11202530 to osteoporosis pathogenesis through modified *PAPSS2* expression and osteoblast differentiation orchestrated by YY2 (Figure 7G).

DISCUSSION

Recent studies have employed STARR-seq to elucidate enhancer-modulating regulatory disease-associated GWAS SNPs, such as insulin-resistance-relevant phenotypes,[19] prostate cancer,[20] coronary artery disease,[21] or atrial fibrillation.[22] However, there have been no systematic functional investigations for roles of enhancer elements in orchestrating osteoporosis-associated SNPs. We recently disclosed crucial roles of 3D chromatin organizations in controlling osteogenesis.[17] Nevertheless, whether enhancer-promoter chromatin interactions conferred risk for osteoporosis remains largely elusive. Our preliminary analyses revealed significant enrichment of heritability across multiple osteoporosis-relevant traits[2–6] in significant chromatin interactions in hMSC-induced osteoblasts, implying putatively biological roles of chromatin organizations at osteoporosis risk loci. Consistent with previous functional investigations at several osteoporosis risk loci,[11,45–47] our STARR-seq assays further underscored 319 putative functional baaSNPs with distal enhancer-promoter regulatory effect at 146 osteoporosis risk loci. Particularly, most chromatin interaction genes of baaSNPs displayed known osteoporosis-relevant functions or putative causal genetic associations with osteoporosis, supporting their critical genetic regulatory roles for osteoporosis. These results provided many promising mechanistic hypotheses conferring risk of regulatory GWAS SNPs to osteoporosis pathogenesis via enhancer-promoter interactions, which might accelerate future post-GWAS experimental investigations at osteoporosis risk loci.

Our analyses suggested significant enrichment for several active histone modifications on identified baaSNPs compared with inactive SNPs, including H3K79me2 and H3K36me3. H3K79me2 is associated with active transcription with known roles for modulating enhancer activity to strengthen gene expression through maintenance of enhancer-promoter interactions.[28,48] H3K36me3 is another active chromatin mark with known roles in transcription elongation.[27] H3K36me3 also play roles in antagonizing silencers marked by H3K27me3 by inhibiting PRC2 activity.[49] We also found significant depleted enrichment for H3K27me3, a typical silencer-indicative maker, on identified baaSNPs. Interestingly, previous studies had observed mutually exclusive distribution of H3K36me3 and H3K27me3.[27] Consistently, we found that only 9/144 baaSNPs overlaying with the H3K36me3 peak also located within H3K27me3 peaks, implying potential antagonistic roles between H3K36me3 and H3K27me3 surrounding baaSNPs. In contrast, 81.94% (118/144) of baaSNPs overlaying the H3K36me3 peak also located within peaks of H3K79me2, suggesting their putative cooperative roles in reinforcing enhancer activity and maintenance of enhancer-promoter interactions. Future functional investigations are valued to explore whether H3K36me3 and H3K79me2 play roles in orchestrating enhancer activity and chromatin organization by antagonizing H3K27me3 in osteoblast cells.

Our study also prioritized several candidate controlling TFs for osteoporosis, which exhibited significant enrichment for predicted allele-specific binding on baaSNPs. Interestingly, most enriched TFs (24/33) showed preferential binding to higher enhancer activity alleles of baaSNPs, which was consistent with their significant enrichment for GO molecular terms related to transcriptional activator or coactivator activity and implied their putative enhancer activator effect on baaSNPs. The enriched TFs were also significantly enriched for several bone-relevant pathways (such as skeletal system development and ossification), which provided compelling evidence for their putative regulatory roles on osteoporosis-associated SNPs. Indeed, many of these enriched TFs have known roles in controlling osteogenesis. For example, HOXC10 was proven to regulate osteogenesis of mesenchymal stromal cells and to attenuate bone formation in a mouse model.[37] Another TF, MEF2A, was shown to augment *SOST* expression in osteocytes of adult bone by activating one distal bone enhancer element.[38] For other enriched TFs with uncharacterized roles in osteogenesis, further experimental investigations are worth exploring for their precise roles in controlling osteogenesis and bone development.

Figure 7. YY2 may play roles in promoting human osteoblast differentiation through positive regulation of *PAPSS2*

(A) RT-qPCR shows comparison of *Yy2* and *Papss2* expression between ovariectomy (OVX)-induced osteoporosis mice (females, n = 3) and control normal mice with parietal ovarian fat tissue removed (females, n = 3).

(B) RT-qPCR shows expression change of *YY2* and *PAPSS2* in isolated human primary osteoblasts after differentiation for 3, 7, or 14 days (STAR Methods). The upper ALP staining displays ALP activity change in osteoblasts after differentiation for different days.

(C and D) RT-qPCR shows effect of *YY2* overexpression (C) or knockdown (D) on *PAPSS2* or expression of several osteoblast differentiation marker genes (*ALP*, *OCN*, *RUNX2*, and *COL1A1*) at mRNA level in human primary osteoblasts.

(E and F) RT-qPCR shows effect of *PAPSS2* overexpression (E) or knockdown (F) on expression of several osteoblast differentiation marker genes (*ALP*, *OCN*, *RUNX2*, and *COL1A1*) at mRNA level by two independent shRNAs in human primary osteoblasts.

(G) Speculative regulatory model connecting different genotypes of rs11202530 to allele-preferable binding of YY2 and modified *PAPSS2* expression, osteoblast differentiation, and osteoporosis risk. Data are presented as mean ± standard deviation in (A)–(F), $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns: not significant, two-tailed paired Student's t test.

Our TF-gene regulatory network analyses underlined one putative controlling TF (YY2) for osteoporosis. A recent study has validated that *Yy2* could promote osteoblast differentiation by upregulating transcriptional activity of Osterix in BMP4-induced C2C12 cells.[39] Consistently, we detected significantly lower expression of *Yy2* in osteoporosis mice and experimentally verified putative contributory roles of YY2 in promoting human osteoblast differentiation. The roles of YY2 in chromatin organization are currently unknown. However, a homolog of YY2, YY1, has been validated to act as a structural regulator of active enhancer-promoter loop formation.[42] Consistently, our 3C assay revealed that knockdown of *YY2* could significantly attenuate chromatin interaction frequencies between rs11202530 and *PAPSS2* promoter, which suggested potential involvement of YY2 in mediating long-range chromatin interactions at this locus. Moreover, footprints analyses suggested significant enrichment for both enhancer-like histone peaks, CTCF binding, and significant chromatin interactions on predicted active YY2 binding sites, which implied a putative role of YY2 in orchestrating active enhancer-promoter chromatin interactions analogous to YY1.[42] Future comprehensive experimental investigations are worthy for decrypting the precise roles of YY2 in orchestrating chromatin organizations in osteoblasts.

Our experimental findings also established an enhancer-promoter regulatory relationship between YY2 and *PAPSS2*, which was modified by allelic enhancer activity of one intronic baaSNP (rs11202530) at 10q23.2. A previous study has suggested that *Papss2* could regulate osteoblast ALP activity and cell mineralization in mouse osteoblastic MC3T3-E1 cells.[50] However, roles of *PAPSS2* in human osteogenesis are currently unknown. Mutation in *PAPSS2* has been reported to cause diseases affecting skeletogenesis, such as the spondyloepimetaphyseal dysplasia.[51] By using human primary osteoblast cells, we firstly verified that *PAPSS2* expression was significantly increased during osteoblast differentiation and knockdown of *PAPSS2* expression, or rs11202530 enhancer fragment could suppress expression of osteoblast differentiation marker genes, therefore establishing a new YY2-condensed regulatory axis conferring risk of an osteoporosis risk SNP to osteoporosis pathogenesis. Future experimental investigations are needed to strength this regulatory axis through verifying the skeletal phenotypic effect of rs11202530 enhancer fragment in mouse model.

In summary, our systematical enhancer surveying and experimental verifying divulged a new YY2-condensed regulatory axis conferring risk for osteoporosis. We anticipate that our integrative experimental and regulatory network analyses could be applied to other complex diseases, which may help reduce the gap between traditional genetic findings and mechanistic understanding of disease etiologies and future clinical drug development.

Limitations of the study

Despite the high efficiency in systematically qualifying enhancer activities on whole-genomic GWAS SNPs, the STARR-seq assay lacked *in vivo* contexts. Therefore, some chromatin-dependent enhancers might be neglected. However, we filtered GWAS SNPs with significant promoter chromatin interactions before STARR-seq screening, which showed obvious significant enrichment for enhancer-like epigenetic elements in primary osteoblast cells. Moreover, the high overlap and enrichment of identified enhancer SNPs in active epigenetic features implied their endogenous enhancer activities. We anticipated that future tectological improvements in high-throughput *in vivo* genomic allelic editing coupled with single-cell sequencing[52] could help confirm our results. Besides, future *cis*-eQTL mapping in osteoporosis-related human tissues with large samples would strengthen the enhancer-promoter regulation between baaSNPs and their chromatin interaction genes. Finally, we convincingly divulged a new YY2-condensed regulatory axis conferring risk for osteoporosis. However, our chromatin regulatory network analyses also disclosed putative involvement for osteoporosis pathogenesis on several other TFs, which are worthy of future detailed experimental investigations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - GWAS fine-mapping for refining osteoporosis-associated SNPs
  - Hi-C interaction analysis
  - Partitioned heritability enrichment analysis
  - Epigenetic enrichment analysis on promoter chromatin interacted SNPs
  - Human primary osteoblasts isolation, culture, and differentiation
  - Cell line culture
  - STARR-seq plasmid library
  - Plasmid library transfection
  - STARR-seq sequencing library
  - Dual luciferase reporter assays
  - Genotyping of SNPs
  - ChIP-qPCR
  - Short hairpin RNA knockdown
  - Co-transfection of *YY2* shRNA and rs11202530 luciferase reporter plasmids
  - Chromosome conformation capture assay
  - CRISPR-Cas9 deletion of SNP-harboring regions
  - Alkaline phosphatase (ALP) staining
  - Lentivirus overexpression
  - Osteoporosis mouse model
  - DNA and RNA isolation
  - Reverse transcription and quantitative PCR
  - Western blotting analysis
  - STARR-seq data analysis
  - Epigenetic characterizations on baaSNPs and silencer SNPs

- *Cis*-eQTL association analysis
- GWAS-eQTL colocalization analysis
- Transcriptome-wide association studies analysis
- Pathway enrichment analysis
- Screening genes with skeletal-related physiological effect
- Allele-specific motif prediction
- TF enrichment analyses and functional assessment
- Genetic regulatory network analysis
- Comparative epigenetic and Hi-C interaction analyses for rs11202530
- YY2 footprints prediction and functional enrichment analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100501.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (32000394, 32170616, 82170896, and 31970569), China Postdoctoral Science Foundation (2020TQ0243 and 2020M683453), Science Fund for Distinguished Young Scholars of Shaanxi Province (2021JC-02), Innovation Capability Support Program of Shaanxi Province (2022TD-44), Key Research and Development Project of Shaanxi Province (2022GXLH-01-22), Medical-engineering Interdisciplinary Project of Xi'an Jiaotong University (YGJC202202), and the Fundamental Research Funds for the Central Universities. This study is also supported by the High-Performance Computing Platform and Instrument Analysis Center of Xi'an Jiaotong University. We thank the participants and the organizers of the Genotype-Tissue Expression (GTEx) project and UK Biobank.

AUTHOR CONTRIBUTIONS

T.-L.Y. and Y.G. designed the study. X.-F.C. performed data analysis and wrote the manuscript. Y.-Y.D. and Y.-Y.J. performed STARR-seq screening and experimental assays with assistance by Q.-H.D. and X.-T.H. W.S. performed network construction and analyses. Y.Z. and F.C. contributed to the human primary osteoblast isolation and culture. S.-S.D. and R.-H.H. contributed to Hi-C interaction analysis. L.M., Z.L., D.-L.Z., and R.-H.J. contributed to manuscript preparation. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 16, 2023
Revised: August 23, 2023
Accepted: January 10, 2024
Published: February 8, 2024

REFERENCES

1. Yang, T.L., Shen, H., Liu, A., Dong, S.S., Zhang, L., Deng, F.Y., Zhao, Q., and Deng, H.W. (2020). A road map for understanding molecular and genetic determinants of osteoporosis. Nat. Rev. Endocrinol. *16*, 91–103.
2. Trajanoska, K., Morris, J.A., Oei, L., Zheng, H.F., Evans, D.M., Kiel, D.P., Ohlsson, C., Richards, J.B., and Rivadeneira, F.; GEFOS/GENOMOS consortium and the 23andMe research team (2018). Assessment of the genetic and clinical determinants of fracture risk: genome wide association and mendelian randomisation study. BMJ *362*, k3225.
3. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat. Genet. *51*, 258–266.
4. 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. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature *526*, 112–117.
5. Medina-Gomez, C., Kemp, J.P., Dimou, N.L., Kreiner, E., Chesi, A., Zemel, B.S., Bønnelykke, K., Boer, C.G., Ahluwalia, T.S., Bisgaard, H., et al. (2017). Bivariate genome-wide association meta-analysis of pediatric musculoskeletal traits reveals pleiotropic effects at the SREBF1/TOM1L2 locus. Nat. Commun. *8*, 121.
6. Medina-Gomez, C., Kemp, J.P., Trajanoska, K., Luan, J., Chesi, A., Ahluwalia, T.S., Mook-Kanamori, D.O., Ham, A., Hartwig, F.P., Evans, D.S., et al. (2018). Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects. Am. J. Hum. Genet. *102*, 88–102.
7. Vincentz, J.W., Firulli, B.A., Toolan, K.P., Arking, D.E., Sotoodehnia, N., Wan, J., Chen, P.S., de Gier-de Vries, C., Christoffels, V.M., Rubart-von der Lohe, M., and Firulli, A.B. (2019). Variation in a Left Ventricle-Specific Hand1 Enhancer Impairs GATA Transcription Factor Binding and Disrupts Conduction System Development and Function. Circ. Res. *125*, 575–589.
8. Zhang, Z., Liu, X., Li, L., Yang, Y., Yang, J., Wang, Y., Wu, J., Wu, X., Shan, L., Pei, F., et al. (2021). SNP rs4971059 predisposes to breast carcinogenesis and chemoresistance via TRIM46-mediated HDAC1 degradation. EMBO J. *40*, e107974.
9. Wang, Z., Liang, Q., Qian, X., Hu, B., Zheng, Z., Wang, J., Hu, Y., Bao, Z., Zhao, K., Zhou, Y., et al. (2023). An autoimmune pleiotropic SNP modulates IRF5 alternative promoter usage through ZBTB3-mediated chromatin looping. Nat. Commun. *14*, 1208.
10. Yang, H., Zhang, H., Luan, Y., Liu, T., Yang, W., Roberts, K.G., Qian, M.X., Zhang, B., Yang, W., Perez-Andreu, V., et al. (2022). Noncoding genetic variation in GATA3 increases acute lymphoblastic leukemia risk through local and global changes in chromatin conformation. Nat. Genet. *54*, 170–179.
11. Chen, X.F., Zhu, D.L., Yang, M., Hu, W.X., Duan, Y.Y., Lu, B.J., Rong, Y., Dong, S.S., Hao, R.H., Chen, J.B., et al. (2018). An Osteoporosis Risk SNP at 1p36.12 Acts as an Allele-Specific Enhancer to Modulate LINC00339 Expression via Long-Range Loop Formation. Am. J. Hum. Genet. *102*, 776–793.
12. Su, C., Gao, L., May, C.L., Pippin, J.A., Boehm, K., Lee, M., Liu, C., Pahl, M.C., Golson, M.L., Naji, A., et al. (2022). 3D chromatin maps of the human pancreas reveal lineage-specific regulatory architecture of T2D risk. Cell Metab. *34*, 1394–1409.e4.
13. Lu, L., Liu, X., Huang, W.K., Giusti-Rodríguez, P., Cui, J., Zhang, S., Xu, W., Wen, Z., Ma, S., Rosen, J.D., et al. (2020). Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. Mol. Cell *79*, 521–534.e15.
14. Chandra, V., Bhattacharyya, S., Schmiedel, B.J., Madrigal, A., Gonzalez-Colin, C., Fotsing, S., Crinklaw, A., Seumois, G., Mohammadi, P., Kronenberg, M., et al. (2021). Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. Nat. Genet. *53*, 110–119.
15. Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.
16. Giambartolomei, C., Seo, J.H., Schwarz, T., Freund, M.K., Johnson, R.D., Spisak, S., Baca, S.C., Gusev, A., Mancuso, N., Pasaniuc, B., and Freedman, M.L. (2021). H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. Am. J. Hum. Genet. *108*, 2284–2300.

17. Hao, R.H., Guo, Y., Wang, C., Chen, F., Di, C.X., Dong, S.S., Cao, Q.L., Guo, J., Rong, Y., Yao, S., et al. (2022). Lineage-specific rearrangement of chromatin loops and epigenomic features during adipocytes and osteoblasts commitment. Cell Death Differ. *29*, 2503–2518.

18. Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science *339*, 1074–1077.

19. Duan, Y.Y., Chen, X.F., Zhu, R.J., Jia, Y.Y., Huang, X.T., Zhang, M., Yang, N., Dong, S.S., Zeng, M., Feng, Z., et al. (2023). High-throughput functional dissection of noncoding SNPs with biased allelic enhancer activity for insulin resistance-relevant phenotypes. Am. J. Hum. Genet. *110*, 1266–1288.

20. Zhang, P., Xia, J.H., Zhu, J., Gao, P., Tian, Y.J., Du, M., Guo, Y.C., Suleman, S., Zhang, Q., Kohli, M., et al. (2018). High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. Nat. Commun. *9*, 2022.

21. Selvarajan, I., Toropainen, A., Garske, K.M., López Rodríguez, M., Ko, A., Miao, Z., Kaminska, D., Õunap, K., Örd, T., Ravindran, A., et al. (2021). Integrative analysis of liver-specific non-coding regulatory SNPs associated with the risk of coronary artery disease. Am. J. Hum. Genet. *108*, 411–430.

22. van Ouwerkerk, A.F., Bosada, F.M., Liu, J., Zhang, J., van Duijvenboden, K., Chaffin, M., Tucker, N.R., Pijnappels, D., Ellinor, P.T., Barnett, P., et al. (2020). Identification of Functional Variant Enhancers Associated with Atrial Fibrillation. Circ. Res. *127*, 229–243.

23. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

24. Tippens, N.D., Liang, J., Leung, A.K.Y., Wierbowski, S.D., Ozer, A., Booth, J.G., Lis, J.T., and Yu, H. (2020). Transcription imparts architecture, function and logic to enhancer units. Nat. Genet. *52*, 1067–1075.

25. Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. Nat. Commun. *11*, 1061.

26. Ashuach, T., Fischer, D.S., Kreimer, A., Ahituv, N., Theis, F.J., and Yosef, N. (2019). MPRAnalyze: statistical framework for massively parallel reporter assays. Genome Biol. *20*, 183.

27. Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. Biophys. Rep. *4*, 170–177.

28. Godfrey, L., Crump, N.T., O'Byrne, S., Lau, I.J., Rice, S., Harman, J.R., Jackson, T., Elliott, N., Buck, G., Connor, C., et al. (2021). H3K79me2/3 controls enhancer–promoter interactions and activation of the pan-cancer stem cell marker PROM1/CD133 in MLL-AF4 leukemia cells. Leukemia *35*, 90–106.

29. Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A., Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., et al. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

30. Pividori, M., Rajagopal, P.S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y., Consortium, G., Wen, X., and Im, H.K. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci. Adv. *6*, eaba2083.

31. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet. *15*, e1007889.

32. Vlashi, R., Zhang, X., Wu, M., and Chen, G. (2023). Wnt signaling: Essential roles in osteoblast differentiation, bone metabolism and therapeutic implications for bone and skeletal disorders. Genes Dis. *10*, 1291–1317.

33. Smith, C.L., Blake, J.A., Kadin, J.A., Richardson, J.E., and Bult, C.J.; Mouse Genome Database Group (2018). Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. Nucleic Acids Res. *46*, D836–D842.

34. Pippin, J.A., Chesi, A., Wagley, Y., Su, C., Pahl, M.C., Hodge, K.M., Johnson, M.E., Wells, A.D., Hankenson, K.D., and Grant, S.F.A. (2021). CRISPR-Cas9–Mediated Genome Editing Confirms EPDR1 as an Effector Gene at the BMD GWAS-Implicated 'STARD3NL' Locus. JBMR plus *5*, e10531.

35. Huang, H., Lu, J., Aukhil, I., Yu, C., Bhut, B., Marchesan, J., Pirih, F., and Chang, J. (2023). FBXO11 regulates bone development. Bone *170*, 116709.

36. 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–W208.

37. Li, B., Han, H., Song, S., Fan, G., Xu, H., Zhou, W., Qiu, Y., Qian, C., Wang, Y., Yuan, Z., et al. (2019). HOXC10 Regulates Osteogenesis of Mesenchymal Stromal Cells Through Interaction with Its Natural Antisense Transcript lncHOXC-AS3. Stem Cell. *37*, 247–256.

38. Leupin, O., Kramer, I., Collette, N.M., Loots, G.G., Natt, F., Kneissel, M., and Keller, H. (2007). Control of the SOST Bone Enhancer by PTH Using MEF2 Transcription Factors. J. Bone Miner. Res. *22*, 1957–1967.

39. Piao, M., Lee, S.H., Kim, M.J., Kim, H.S., and Lee, K.Y. (2022). YY2 Promotes Osteoblast Differentiation by Upregulating Osterix Transcriptional Activity. Int. J. Mol. Sci. *23*, 4303.

40. van de Peppel, J., Strini, T., Tilburg, J., Westerhoff, H., van Wijnen, A.J., and van Leeuwen, J.P. (2017). Identification of Three Early Phases of Cell-Fate Determination during Osteogenic and Adipogenic Differentiation by Transcription Factor Dynamics. Stem Cell Rep. *8*, 947–960.

41. Son, M.-Y., Choi, H., Han, Y.-M., and Cho, Y.S. (2013). Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. Stem Cell. *31*, 2374–2387.

42. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

43. Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res. *41*, e201.

44. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. *46*, D794–D801.

45. Zhang, Y., Chen, X.F., Li, J., He, F., Li, X., and Guo, Y. (2020). lncRNA Neat1 Stimulates Osteoclastogenesis Via Sponging miR-7. J. Bone Miner. Res. *35*, 1772–1781.

46. Zhu, D.L., Chen, X.F., Zhou, X.R., Hu, S.Y., Tuo, X.M., Hao, R.H., Dong, S.S., Jiang, F., Rong, Y., Yang, T.L., et al. (2022). An Osteoporosis Susceptibility Allele at 11p15 Regulates SOX6 Expression by Modulating TCF4 Chromatin Binding. J. Bone Miner. Res. *37*, 1147–1155.

47. Zhu, D.L., Chen, X.F., Hu, W.X., Dong, S.S., Lu, B.J., Rong, Y., Chen, Y.X., Chen, H., Thynn, H.N., Wang, N.N., et al. (2018). Multiple Functional Variants at 13q14 Risk Locus for Osteoporosis Regulate RANKL Expression Through Long-Range Super-Enhancer. J. Bone Miner. Res. *33*, 1335–1346.

48. Godfrey, L., Crump, N.T., Thorne, R., Lau, I.J., Repapi, E., Dimou, D., Smith, A.L., Harman, J.R., Telenius, J.M., Oudelaar, A.M., et al. (2019). DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. Nat. Commun. *10*, 2803.

49. Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.

50. Wang, W., Li, F., Wang, K., Cheng, B., and Guo, X. (2012). PAPSS2 promotes alkaline phosphates activity and mineralization of osteoblastic

MC3T3-E1 cells by crosstalk and Smads signal pathways. PLoS One *7*, e43475.

51. Faiyaz ul Haque, M., King, L.M., Krakow, D., Cantor, R.M., Rusiniak, M.E., Swank, R.T., Superti-Furga, A., Haque, S., Abbas, H., Ahmad, W., et al. (1998). Mutations in orthologous genes in human spondyloepimetaphyseal dysplasia and the brachymorphic mouse. Nat. Genet. *20*, 157–162.
52. Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell *176*, 377–390.e19.
53. Freeman, T.C., Horsewell, S., Patir, A., Harling-Lee, J., Regan, T., Shih, B.B., Prendergast, J., Hume, D.A., and Angus, T. (2022). Graphia: A platform for the graph-based visualisation and analysis of high dimensional data. PLoS Comput. Biol. *18*, e1010310.
54. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
55. Benner, C., Spencer, C.C.A., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics *32*, 1493–1501.
56. 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.
57. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.Y., et al. (2015). An integrated map of structural variation in 2,504 human genomes. Nature *526*, 75–81.
58. Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522–1536.e22.
59. Thynn, H.N., Chen, X.F., Hu, W.X., Duan, Y.Y., Zhu, D.L., Chen, H., Wang, N.N., Chen, H.H., Rong, Y., Lu, B.J., et al. (2020). An Allele-Specific Functional SNP Associated with Two Systemic Autoimmune Diseases Modulates IRF5 Expression by Long-Range Chromatin Loop Formation. J. Invest. Dermatol. *140*, 348–360.e11.
60. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
61. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
62. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
63. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.
64. Al-Barghouthi, B.M., Rosenow, W.T., Du, K.P., Heo, J., Maynard, R., Mesner, L., Calabrese, G., Nakasone, A., Senwar, B., Gerstenfeld, L., et al. (2022). Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for human bone mineral density GWAS associations. Elife *11*, e77285.
65. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.
66. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al. (2022). JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *50*, D165–D173.
67. Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al. (2018). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. *46*, D252–D259.
68. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.
69. Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547–1548.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-YY2 antibody	Santa Cruz Biotechnology	sc-374455; RRID:AB_10988247
anti-PAPSS2 antibody	Santa Cruz Biotechnology	sc-271429; RRID:AB_10648899
anti-GAPDH antibody	HUABIO	ET1601-4; RRID:AB_3069615
HRP-labeled Goat Anti-Mouse IgG(H + L)	Epizyme Biomedical Technology	LF101; RRID: Pending
HRP-labeled Goat Anti-Rabbit IgG(H + L)	Epizyme Biomedical Technology	LF102; RRID: Pending
Bacterial and virus strains		
Ecoli. DH5α Competent cells	TIANGEN Biotech	CB101
Trans-T1 Competent cells	Transgene	CD501
hSTARR-seq_ORI vector	Addgene	99296
pGL3-basic vector	Promega	E1751
pRL-TK Renilla vector	Promega	E2241
Biological samples		
Human MSCs	Shaanxi Provincial People's Hospital, China	N/A
Human primary osteoblasts	Shaanxi Provincial People's Hospital, China	N/A
Chemicals, peptides, and recombinant proteins		
HyClone DMEM with high glucose	Cytiva	SH30022.01
HyClone RPMI 1640	Cytiva	SH30027.01
HyClone α-MEM	Cytiva	SH30265.01
Fetal Bovine Serum	Biological Industries	04-222-1A
Penicillin-Streptomycin	Solarbio	P1400
Penicillin-Streptomycin-Gentamicin Solution	Beyotime Biotechnology	C0223
Collagenase type I	Solarbio	C8140
Ascorbic acid	Beyotime Biotechnology	ST1434
β-glycerol-phosphate	Solarbio	G8100
Puromycin	Beyotime Biotechnology	ST551
Polybrene	Solarbio	H8761
RIPA lysis buffer	Epizyme Biomedical Technology	PC101
Omni-Easy™ Protein Sample Loading Buffer	Epizyme Biomedical Technology	LT101S
PMSF (Phenylmethanesulfonyl fluoride)	Beyotime Biotechnology	ST506
Critical commercial assays		
NEBNext® Q5 Hot Start HiFi PCR Master Mix	New England Biolabs	M0543S
Age I-HF	New England Biolabs	R3552S
Sal I-HF	New England Biolabs	R3138S
Hind III-HF	New England Biolabs	R3104S
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	E2621
HiSpeed® Plasmid Midi Kit	QIAGEN	12643
Dr. GenTLE™ Precipitation Carrier	Takara	9094
Monarch® Total RNA Miniprep Kit	New England Biolabs	T2010
TURBO DNA-free™ Kit	Thermo Fisher Scientific	AM1907
NEBNext® Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490S
Monarch® RNA Cleanup Kit	New England Biolabs	T2030S
ProtoScript® II First Strand cDNA Synthesis Kit	New England Biolabs	E6560S

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AMPureXP Beads	Beckerman	A63880
Qubit™ dsDNA HS Assay Kit	Invitrogen	Q32851
Qubit™ RNA HS Assay Kit	Invitrogen	Q32852
ViaFect™ Transfection Reagent	Promega	E4981
Dual-Lumi™ II Luciferase Reporter Gene Assay Kit	Beyotime Biotechnology	RG027
Simple ChIP Enzymatic Chromatin IP Kit	Cell Signaling Technology	9003
Alkaline Phosphatase Color Development Kit	Beyotime Biotechnology	C3206
TIANamp Genomic DNA Kit	TIANGEN Biotech	DP304
RNAfast200 Kit	Fastagen	220010
RNAiso Plus	Takara	9109
PrimeScript™ RT reagent Kit	Takara	RR037A
2×Universal SYBR Green Fast qPCR Mix	ABclonal	RK21203
Omni-ECL Femto Light Chemiluminescence Kit	Epizyme Biomedical Technology	SQ201
Deposited data		
Raw data	This paper	GEO: GSE231932
Osteoporosis-relevant GWAS summary data	GEFOS	Table S1
UK Biobank reference genotype panels	Applied from UK Biobank (Application number: 46387)	https://www.ukbiobank.ac.uk/
1000 Genomes genotype data for LD analysis (phase 3)	1000 Genomes[57]	https://www.internationalgenome.org/
Hi-C, ATAC-seq and ChIP-seq data in hMSC induced osteoblasts	Hao et al.[17]	GEO: GSE151319, GSE151315, and GSE151311
GTEx genotype data	Applied from dbGap (dbGap: phs000424.v8.p2)	https://www.ncbi.nlm.nih.gov/gap/
GTEx eQTL association data	GTEx portal (V8)[29]	GTEx Portal: https://www.gtexportal.org/home/datasets
DNase-seq, ChIP-seq peak and chromatin segment (HMM-15) in osteoblasts	ENCODE[44]	http://hgdownload.cse.ucsc.edu/goldenPath/hg19/
YY2 ChIP-seq	ENCODE[44]	ENCODE Project: https://www.encodeproject.org/experiments/ENCSR692HSE/
Pre-computed GTEx-V8 multi-tissue eQTL fine-mapping	GWAS-eQTL colocalization analysis (fastENLOC)[30]	https://drive.google.com/open?id=1rSaHenk8xOFtQo7VuDZevRkjUz6iwuj0
GWAS summary statistics imputation support data	S-MulTiXcan[31]	Zenodo: https://zenodo.org/records/3657902
IMPC (release-17.0) for gene functional analyses	N/A	IMPC Portal: http://www.mousephenotype.org/
gene expression profile for gene expression change analyses in differentiated osteoblasts	Zhang el al.[40]	GEO: GSE80614
Experimental models: Cell lines		
U-2OS	National Collection of Authenticated Cell Cultures	TCHu88
MG-63	National Collection of Authenticated Cell Cultures	TCHu124
HEK293T	National Collection of Authenticated Cell Cultures	GNHu17
Experimental models: Organisms/strains		
Ovariectomy-induced osteoporosis mouse model	This paper	C57BL/6
Oligonucleotides		
Oligonucleotides pool of CustomArray, see Table S5	GenScript	N/A
Primers for STARR-seq plasmid library, see Table S6	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primers for STARR-seq sequencing libraries, see Table S6	This paper	N/A
Primers for dual luciferase reporter assays, see Table S6	This paper	N/A
shRNA targeting sequences: *YY2* or *PAPSS2*, see Table S6	This paper	N/A
Primers for cDNA amplification of *YY2* or *PAPSS2*, see Table S6	This paper	N/A
sgRNA targeting sequences: rs11202530, see Table S6	This paper	N/A
RT-qPCR primers for *YY2* or *PAPSS2* (human), see Table S6	This paper	N/A
RT-qPCR primers for osteoblast marker genes, see Table S6	This paper	N/A
ChIP-qPCR primers for rs11202530, see Table S6	This paper	N/A
Primers for Chromosome conformation Capture assay, see Table S6	This paper	N/A
RT-qPCR primers for *yy2* or *papss2* (mouse), see Table S6	This paper	N/A
Recombinant DNA		
Laboratory-modified pGL3-promoter vector	Duan et al.[19]	N/A
Laboratory-modified lentiCRISPR v2 vector	Chen et al.[11]	N/A
pUC19-hU6-sgRNA vector	Chen et al.[11]	N/A
Laboratory-modified lenti-CMV-MCS-EF1a-PuroR vector	Duan et al.[19]	N/A
Laboratory-modified lenti-shRNA-miR30 vector	Thynn et al.[59]	N/A
Software and algorithms		
ImageJ software	N/A	https://ij.imjoy.io/
GCTA (1.92.1)	Yang et al.[54]	https://yanglab.westlake.edu.cn/software/gcta/#Overview
FINEMAP (1.3.1)	Benner et al.[55]	http://christianbenner.com/
PLINK (1.90)	Purcell et al.[56]	https://www.cog-genomics.org/plink/1.9/
HOMER (v4.10.1)	Heinz et al.[58]	http://homer.ucsd.edu/homer/
bedtools (v2.29.0)	N/A	https://bedtools.readthedocs.io/en/latest/
S-LDSC (v1.0.1)	Finucane et al.[23]	https://github.com/bulik/ldsc
R (v3.4.1)	N/A	https://www.r-project.org/
Python (v2.7.16)	N/A	https://www.python.org/download/releases/2.7/
fastp (v0.20.1)	Chen et al.[60]	https://github.com/OpenGene/fastp
Bowtie 2 (v2.4.1)	Langmead et al.[61]	https://github.com/BenLangmead/bowtie2
R limma package (v3.34.9)	Ritchie et al.[62]	https://bioconductor.org/packages/release/bioc/html/limma.html
MPRAnalyze (v1.16.0)	Ashuach et al.[26]	https://bioconductor.org/packages/release/bioc/html/MPRAnalyze.html
DeepTools (v3.3.0)	Ramírez et al.[63]	https://github.com/deeptools/deepTools
pyDNase (v0.2.4)	Piper et al.[43]	https://pythonhosted.org/pyDNase/index.html
fastENLOC (v2.0)	Pividori et al.[30]	https://github.com/xqwen/fastenloc

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S-MulTiXcan	Barbeira et al.[31]	Zenodo: https://zenodo.org/records/3657902
R clusterProfiler package (v3.6.0)	Yu et al.[65]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
MEME Suite toolkit (v5.4.1)	Bailey et al.[36]	https://meme-suite.org/meme/
R lumi package (v.2.30.0)	Du et al.[69]	https://www.bioconductor.org/packages/release/bioc/html/lumi.html
Graphia	Freeman et al.[53]	https://graphia.app/
Integrative Genomics Viewer	N/A	https://igv.org/
Custom code	This paper	Github: https://github.com/xjtugenetics/2023_Osteoporosis_STARRseq and Zenodo: https://doi.org/10.5281/zenodo.10472555
Other		
Full-length cDNA of *PAPSS2*	NCBI	GenBank: NM_004670.4
Full-length cDNA of *YY2*	NCBI	GenBank: NM_206923.4

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Tie-Lin Yang (yangtielin@xjtu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw STARR-seq sequencing data have been submitted to GEO and are publicly available as of the date of publication (GEO: GSE231932). Hi-C, ATAC-seq and ChIP-seq data in hMSC induced osteoblast are available at GEO (GEO: GSE151319, GSE151315, and GSE151311). Other applied or publicly available data used in this study are listed in the Key resources table and Table S1. Detailed analysis results are listed in supplementary tables. The integrative analytical codes and pipelines are available at Github: https://github.com/xjtugenetics/2023_Osteoporosis_STARRseq and archived under Zenodo: https://doi.org/10.5281/zenodo.10472555. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

For primary cell culture studies, we isolated human primary osteoblasts from cancellous bone tissue of human lumbar vertebra of patients undergoing surgery (vertebral compression fracture or lumbar disc herniation, three males and one female), and cultured with MEM-α (HyClone, USA) at 37°C, 5% CO_2 incubator. The donors were provided by Shaanxi Provincial People's Hospital, China. All protocol was approved by the ethics committee of the involved institutions with written informed consent provided by participants.

For *in vivo* mouse studies, we purchased 2-month C57BL/6 female healthy mice from the Laboratory Animal Center of Air Force Medical University (Xi'an, Shanxi, China). The mice were housed at permanent temperature (25°C) and alternating light-dark (12 h/12 h) cycle under Specific Pathogen Free (SPF) condition, with *ad libitum* food and water provided. All mouse experiments were in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the Laboratory Animal Ethics & Welfare Committee of Xi'an Jiaotong University.

For other *in vitro* cell culture studies, the human embryonic kidney 293T cells (HEK293T) and human bone osteosarcoma epithelial cells (U2OS and MG63) were purchased from the National Collection of Authenticated Cell Cultures (Shanghai, China). HEK293T and MG63 were cultured in DMEM (Cytiva, USA), U2OS was cultured in RPMI-1640 (Cytiva, USA) with 10% fetal bovine serum (Biological Industries, Israel), 100U/mL penicillin and 0.1 mg/mL streptomycin (Solarbio, USA). All cells were culture at 37°C, 5% CO_2 incubator and authenticated with STR profiling.

METHOD DETAILS

GWAS fine-mapping for refining osteoporosis-associated SNPs

GWAS summary statistics across multiple osteoporosis-relevant traits, including fracture, total body or total body less head bone mineral density (TB[LH]BMD), forearm bone mineral density (FABMD), femoral neck bone mineral density (FNBMD), lumbar spine bone mineral density (LSBMD) and quantitative heel ultrasounds bone mineral density (eBMD) were collected from the Genetic Factors for Osteoporosis Consortium (GEFOS) portal (http://www.gefos.org/) (Table S1). We downloaded UK Biobank genotype data (Application number: 46387) and randomly selected 50,000 White British non-relative samples and removed SNPs with MAF<0.001 or information score (imputation) < 0.3 as reference panels. Conditional GWAS association followed by fine-mapping analysis were conducted as reported by Morris et al..[3] All conditionally independently associated SNPs (indSNP, adjusted $p < 5 \times 10^{-8}$) and fine-mapped potential causal SNPs at base-10 logarithm of Bayes Factor ($Log_{10}BF$) $\geq$ 2 on eBMD were collected from original report by Morris et al.[3] For other traits, we firstly performed a stepwise model selection procedure to identify indSNPs on each GWAS summary data using GCTA –cojo-slct –cojo-p 5e−8 (v1.92.1).[54] We then implemented FINEMAP[55] (v1.3.1) to identify potential causal SNPs at 500-KB region surrounding each indSNP, which was evaluated by $Log_{10}BF$ (Bayes Factor) with higher value corresponding higher probability of being causal. Linkage disequilibrium (LD) analysis for all identified inSNPs were implemented using PLINK (v.1.90)[56] in European samples from 1000 Genomes phase 3,[57] with maximum distance for r^2 calculation set as 1000-kb. All inSNPs and LD expanded ($r^2 > 0.8$) SNPs as well as fine-mapped potential causal SNPs ($Log_{10}BF \geq 2$) were merged as candidate osteoporosis-associated SNPs. In summary, we identified 1,241 indSNPs and 15,765 candidate causal associated SNPs ($log_{10}BF \geq 2$, Table S2), as well as another 27,918 LD expanded osteoporosis-associated SNPs ($r^2 > 0.8$, Table S3).

Hi-C interaction analysis

We previously induced osteoblast and adipogenic differentiation from human mesenchymal stem cells (hMSC) and performed 2-kb resolution Hi-C experiments in hMSC and hMSC differentiated osteoblast or adipocyte.[17] Hi-C data processing have been described before.[17] We called significant chromatin interactions at 2-kb resolution (FDR<0.05) using HOMER's Hi-C analysis tool analyzeHiC (v4.10.1, -res 2000 -superRes 10000 -maxDist 2000000).[58] All osteoporosis-associated SNPs with significant chromatin interactions in osteoblast to any nearby (1000-KB) gene promoter region (1-KB surrounding gene transcription start site [TSS]) analyzed by bedtools (v2.29.0) were prioritized as functional SNPs candidates.

Partitioned heritability enrichment analysis

We used Stratified LD Score Regression (S-LDSC, v.1.0.1)[23] to examine whether heritability of osteoporosis-associated SNPs was enriched in significant Hi-C chromatin interaction regions (FDR<0.05) in osteoblast[17] or two selected epigenetic markers (H3K36me3, H3K79me2). ChIP-seq peak data of epigenetic markers in primary osteoblasts were collected from ENCODE portal[44] (Table S1). GWAS summary statistics across multiple osteoporosis-relevant traits (fracture, TB[LH]BMD, FABMD, FNBMD, LSBMD and eBMD) were used for heritability enrichment analysis. Heritability enrichment is defined as the proportion of heritability assigned to one annotation divided by the proportion of SNPs in that same annotation. For S-LDSC analysis, annot file across 22 chromosomes including all above cell-specific annotation was prepared based on recommended baselineLD v2.2 model using custom python script for annotation-specific LD scores computing. The precomputed genotype, frequency and weight files on HapMap 3 SNPs from 1000 Genomes Europeans were downloaded at https://storage.googleapis.com/broad-alkesgroup-public/LDSCORE/.

Epigenetic enrichment analysis on promoter chromatin interacted SNPs

We collected multiple ChIP-seq (histone modifications and CTCF), chromatin segment (HMM15), ATAC-seq and DNase-seq peak data in hMSC induced or primary osteoblast cells from our previous report (GEO: GSE151311, GSE151315)[17] and ENCODE portal[44] (Table S1). For selected SNPs, overlapping with above annotation were analyzed by bedtools (v2.29.0). We performed enrichment analysis by comparing overlap with each epigenetic annotation type between prioritized promoter chromatin interacted osteoporosis-associated SNPs against all other non-interacted osteoporosis-associated SNPs using two-sided Fisher's exact test, and defined any epigenetic features significantly enriched at FDR<0.05.

Human primary osteoblasts isolation, culture, and differentiation

Human primary osteoblasts were isolated from cancellous bone tissue of human lumbar vertebra of patients undergoing surgery (vertebral compression fracture or lumbar disc herniation). The donor was provided by Shaanxi Provincial People's Hospital, China. In our study, the sample acquired was performed under the ethical approval and patient informed consent. Briefly, the sample of cancellous bone tissue was washed twice with Wash buffer (PBS +1% penicillin/streptomycin/gentamicin) by vigorously shaking. Next, the sample was transferred to the new tube, and added 4mL trypsin for 30min in a shaker at 37°C. Subsequently, the cancellous bone tissue was treated with collagenase type I (Solarbio, USA) for 2 h in a shaker at 37°C after washing twice with Wash buffer. Finally, the sample was placed in culture flasks with α-MEM (Cytiva, USA) supplemented with 10% fetal bovine serum, 1% Penicillin-Streptomycin-Gentamicin Solution (Beyotime, China), and incubated at 37°C and 5% CO_2. We changed the medium twice every week until cells reached confluence. For osteoblast differentiation, primary osteoblasts were induced with osteoblast medium containing 50 μg/mL ascorbic acid (Beyotime, Cat#ST1434, China) and 10 mM β-glycerol-phosphate (Solarbio, Cat#G8100, USA).

The protocol was approved by the ethics committee of the involved institutions with written informed consent provided by participants.

Cell line culture

Human embryonic kidney 293T cells (HEK293T) and human bone osteosarcoma epithelial cells (U2OS) were purchased from National Collection of Authenticated Cell Cultures (Shanghai, China). HEK293T and U2OS were cultured in DMEM (Cytiva, USA) or RPMI (Cytiva, USA) with 10% fetal bovine serum (Biological Industries, Israel), 100U/mL penicillin and 0.1 mg/mL streptomycin (Solarbio, USA), respectively. All cells were culture at 37°C, 5% CO_2 incubator.

STARR-seq plasmid library

For each SNP, we designed two 120-bp sequences centered on the reference or alternative allele. In addition, a 15-bp adapter sequence was added on the 5′ and 3′ region of 120-bp sequence, respectively. A total of 11,284 oligonucleotides (5,642 SNPs interacted to nearby gene promoters) of 150-bp length (Table S5) were synthesized in CustomArray (GenScript, USA). The synthesized oligonucleotides were amplified using adapter primer (Table S6) (NEB, M0543S, USA) and purified using AMPure XP beads (Beckman, A63880, USA). Products were inserted into linearized hSTARR-seq_ORI vector (Plasmid#99296, Addgene, USA) with *Age* I-HF (NEB, Cat#R3552S, USA) and *Sal* I-HF (NEB, Cat#R3138S, USA) by DNA Assembly (NEB, Cat#E2621, USA). The ligation products were precipitated (Takara, Cat#9094, Japan) and transformed into Trans1-T1 Competent cell. Finally, the plasmid library was extracted for cell transfection (QIAGEN, Cat#12643, USA).

Plasmid library transfection

We transfected 12 μg plasmid library into 3.6 × 10^6 U2OS cells using ViaFect transfection reagent (Promega, Cat#E4981, USA) according to the manufacturer's recommendations. Three biological replicates were performed.

STARR-seq sequencing library

Before cell transfection, plasmid library was used for input sequencing library construction. The 100ng plasmid library was applied for one-step PCR using RT-UMI primer (Table S6) (98°C for 30s, 63°C for 30s and 72°C for 2min). Purified products were used for the first round of amplification using Vector-forward and Read1-reverse primers (Table S6). The second round of PCR was performed using sample barcode primers (Table S6) and products were purified for paired-end Illumina sequencing (Novaseq 6000 platform).

After 24 h transfection, total RNA was firstly extracted (NEB, Cat#T2010, USA) and treated with DNase I (Thermo Fisher, AM1907, USA). Next, poly(A) RNA was isolated and precipitated for reverse transcription (NEB, Cat#E7490S, Cat#T2030S, USA). Subsequently, first strand cDNA synthesis was performed using RT-UMI (Table S6) (NEB, Cat#E6560S, USA). Next, cDNA was treated with RNase A (Takara, Cat#2158, Japan) and purified with AMPure XP beads. Next, cDNA was amplified using junction primer[18] and Read1-reverse (Table S6). The second round of PCR was applied using barcode primers (Table S6) and purified with AMPure XP beads. Three biological replicates were performed for input and output sequencing libraries. Sequencing libraries were generated (NEB, Cat#E7645S, USA) and paired-end sequenced on Illumina Novaseq 6000 platform (PE150).

Dual luciferase reporter assays

Luciferase constructs were generated by cloning 120-bp candidate fragments centered on different allele of six randomly selected baaSNPs and six non-baaSNPs (but eSNPs) into the laboratory-modified pGL3-promoter vector,[19] separately (Table S6). In addition, the 120-bp putative enhancer fragments centered on different allele of rs11202530, the 1434-bp *PAPSS2* promoter (1275-bp upstream to 158-bp downstream of TSS) were amplified from healthy human genomic DNA (Table S6). The putative enhancer and promoter fragments were cloned into pGL3-basic vector (Promega, USA). Using transfection reagent (Promega, Cat#E4981, USA), the constructs were co-transfected with renilla plasmid to enable normalization of luciferase signal. Three biological replicates were performed. After 48 h transfection, luciferase and renilla activity were measured (Beyotime, Cat# RG027, China). Luciferase signals was computed as the ratio of firefly luciferase activity to Renilla signals and relative activity was normalized by pGL3-promoter.

Genotyping of SNPs

The genomic DNA was extracted from U2OS, MG63 or human primary osteoblasts. We amplified the DNA fragment surrounding rs11202530 (Table S6) and purified the DNA for Sanger sequencing.

ChIP-qPCR

We performed chromatin immunoprecipitation assay (ChIP) of YY2 with Simple ChIP Enzymatic Chromatin IP Kit (CST, Cat#9003, USA) in U2OS (homozygous GG of rs11202530) and MG63 (heterozygous AG of rs11202530) according to the manufacturer's protocol, which was described in detail as previously.[59] The cross-linked chromatin was immunoprecipitated with YY2 antibody (Santa Cruz Biotechnology, sc-374455, USA) or normal immunoglobulin G (IgG) as a negative control. Following quantification was undertaken by qPCR (U2OS) or allele-specific qPCR (MG63) with primers listed in Table S6. The allele-specific ChIP-qPCR in MG63 was performed using primers specifically targeting rs11202530-G or rs11202530-A to compare TF allelic binding affinity. For

normalization, the control primers were supplied in the kit, which are specific for the human RPL30 gene. Data analysis was performed according to our previous study.[59]

Short hairpin RNA knockdown

Two independent shRNA sequences were designed targeting *PAPSS2* or *YY2*, respectively (Table S6). The oligonucleotides targeting *PAPSS2* or *YY2* were inserted into linearized Laboratory-modified lenti-shRNA-miR30 backbone. The shRNA plasmid (shRNA-1, shRNA-2 or shRNA-NC) and two helper plasmids (pCMV-VSV-G and psPAX2) were co-transfected into HEK293T. The lentiviral supernatant was collected for human primary osteoblasts or U2OS infection. Total RNA and protein were extracted for RT-qPCR and Western blot, respectively (Table S6).

Co-transfection of *YY2* shRNA and rs11202530 luciferase reporter plasmids

YY2 shRNA plasmids (*YY2*-shRNA1, *YY2*-shRNA2, shRNA-NC plasmid) were independently co-transfected with the expression plasmids including rs11202530-G allele or rs11202530-A allele with *PAPSS2* promoter used in the luciferase reporter assay by transfection reagent (Promega, Cat#E4981, USA). Three biological replicates were performed. The measurement of luciferase activity is the same as described in dual-luciferase reporter assay section.

Chromosome conformation capture assay

Chromosome conformation capture (3C) assay was performed in U2OS according to our previous study.[59] Briefly, the *Hind III* (NEB, R3104S, USA) was chosen for 3C because they are near the tested SNP (rs11202530) and target gene (*PAPSS2*) promoter region, allowing fragments containing these regions to be evaluated separately. To correct for potential different primer efficiencies for 3C assay, we constructed a control library that contains rs11202530 regions (N7), *PAPSS2* promoter regions (N3) and six randomly selected neighboring *Hind III* site regions (N1, N2, N4, N5, N6 and N8). We amplified these eight genomic fragments (N1∼N8) using PCR from U2OS genomic DNA with primers listed in Table S6. The equimolar amounts of purified PCR products for each fragment were mixed and digested using *Hind III*. The random ligations were therefore allowed to generate a pool of interaction products for control library, which was further purified by phenol-chloroform extraction and ethanol acetate precipitation. We estimated the percentage of interaction frequency for both libraries (3C and control library) by dividing the amount of distal PCR products between rs11202530 and seven distal target sites (N1, N2, N4, N5, N6 and N8) by the amount of local PCR products for rs11202530-enhancer spanning fragment (N7). The 3C interaction frequency was then normalized by dividing the percentage of interaction frequency in 3C library by the percentage of interaction frequency in control library. We also performed comparative 3C assays in YY2 inhabited (by shRNA) and control (by shRNA-NC) U2OS. Relative normalized 3C interaction frequency in two groups of cells were compared using paired Student's T-test.

CRISPR-Cas9 deletion of SNP-harboring regions

Dual-sgRNA CRISPR-Cas9 was implemented to delete the region surrounding SNP- rs11202530 harboring region, respectively. Briefly, we designed sgRNAs surrounding each SNP using CHOPCHOP (http://chopchop.cbu.uib.no) and CRISPOR (http://crispor.tefor.net) (Table S6). The dual-sgRNA was cloned into pUC19-hU6-sgRNA vector and then ligated into Laboratory-modified lentiCRISPR v2 plasmid. The target plasmid with two helper plasmids (pCMV-VSV-G and psPAX2) were co-transfected into HEK293T. The lentiviral supernatant was collected for target cell infection (U2OS or human primary osteoblasts). Puromycin (2 mg/mL) was used to select sgRNA-positive cells. Genomic DNA were extracted and a flanking region of 1740-bp (rs11202530) was amplified by PCR to verify deletion efficiency with primers listed in Table S6. Total RNA was extracted for RT-qPCR.

Alkaline phosphatase (ALP) staining

The human primary osteoblasts were washed with PBS three times for 5 min, fixed with 4% paraformaldehyde solution for 20 min, and washed with ddH_2O twice for 5 min. The ALP staining (Beyotime, Cat#C3206, China) was added and incubated at 37°C overnight in the dark. The cells were rinsed with ddH_2O three times for 5 min and observed under an inverted microscope and photographed for analysis. ALP staining area ratio of different experimental groups was analyzed using ImageJ software (https://ij.imjoy.io/).

Lentivirus overexpression

Full-length cDNA of *YY2* (GenBank: NM_206923.4) or *PAPSS2* (GenBank: NM_004670.4) was amplified and inserted into Laboratory-modified lenti-CMV-MCS-EF1a-PuroR vector. Empty vector was used as negative control. The YY2-overexpression and PAPSS2-overexpression plasmids and two helper plasmids (pCMV-VSV-G and psPAX2, Addgene, USA) were co-transfected into HEK293T. We collected the lentiviral supernatant from HEK293T and infected human primary osteoblasts. Total RNA was extracted for quantitative real-time PCR (RT-qPCR) (Table S6).

Osteoporosis mouse model

We purchased 2-month C57BL/6 female healthy mice from the Laboratory Animal Center of Air Force Medical University (Xi'an, Shanxi, China). All mice were housed in the School of Life Sciences and Technology of Xi'an Jiaotong University at permanent temperature (25°C) and alternating light-dark (12h/12h) cycle under Specific Pathogen Free (SPF) condition, with *ad libitum* food and

water provided. We then selected three healthy mice after routine feeding for one week for surgery of removing ovaries and ligating the fallopian tubes to generate osteoporosis mice model. The control mice group (n = 3) was operated in the same way, but only the parietal ovarian fat was removed after opening the abdominal cavity. Four months later, mice right tibia was isolated and stored in RNAiso Plus (Takara, Cat#9109, Japan) for RNA extraction. The ovariectomized mouse model for osteoporosis was validated by decreased BMD and deteriorated microarchitecture measured by micro-CT scanning.

DNA and RNA isolation

Genomic DNA was extracted from target cells (TIANGEN Biotech, Cat#DP304, China) according to the manufacturer's protocol. For human primary cells and cell lines, total RNA was isolated using RNAfast200 Kit (Fastagen, Cat#220010, China). For mice tissue, total RNA was extracted using RNAiso Plus (Takara, Cat#9109, Japan). Briefly, mouse osseous tissue was homogenized by Tissue Homogenizer (Servicebio, China) with 600μL RNAiso Plus. Then the homogenate was transferred to a 1.5mL tube with additional 400μL RNAiso Plus. After 5 min incubation at room temperature, they were centrifuged at 12,000×g, 4°C for 5 min. The supernatants were then mixed with 0.2mL of chloroform and centrifuged at 12,000×g, 4°C for 15min. Isopropanol was used to precipitate RNA and washed with 75% ethanol afterward. Finally, ethanol was discarded and the precipitate was dissolved by RNase free water. Total RNA was stored at −80°C.

Reverse transcription and quantitative PCR

Total RNA was reverse-transcribed to cDNA with PrimeScript RT reagent Kit (Takara, Cat#RR037A, Japan). Real time quantitative PCR was performed using 2× Universal SYBR Green Fast qPCR Mix (ABclonal, Cat#RK21203, China) by CFX Connect Real-Time PCR Detection System (Bio-Rad, USA). The *GAPDH* (human) or *β-actin* (mouse) was used as an endogenous control to normalize the differences between samples (Table S6).

Western blotting analysis

Cells were lysed with RIPA lysis buffer (Epizyme Biomedical Technology, Cat# PC101, Cat#LT101S, China), each protein sample was subjected to SDS/PAGE and transferred onto PVDF membranes. After sealed with 5% nonfat milk for 2 h, PVDF membranes were incubated with corresponding primary antibodies at 4°C overnight. Primary antibodies included anti-YY2, anti-PAPSS2 (Santa Cruz Biotechnology, sc-374455, sc-271429, USA), and anti-GAPDH (HUABIO, ET1601-4, USA). Then corresponding HRP-conjugated secondary antibodies (Epizyme Biomedical Technology, Cat#LF101, Cat#LF102, China) were subsequently incubated for 1 h at room temperature. The results were visualized using Omni-ECL Femto Light Chemiluminescence Kit (Epizyme Biomedical Technology, Cat#SQ201, China) in MiniChemi610 (Beijing Sage, China).

STARR-seq data analysis

From FASTQ files of input or output library sequencing, we firstly separated each replicated sample paired reads according to 8-bp barcode sequences with no more than two mismatches/deletions of nucleotides permitted (barcode sequences in Table S6). Percentage of 100% barcode matched reads ranged from 96.93% to 98.82% in all separated samples. Trimmed reads with low quality reads (Q-score<30) were then filtered using fastp 0.20.1 with -q 30 -u 50.[60] Next, custom python scripts was used to trim and extract 120-bp SNP-containing sequences. Before sequences trimming, a 13-bp random unique molecular identifiers (UMI) sequences with no more than two deletions of nucleotides permitted were extracted to remove PCR duplications. Final merged reads were then aligned to our selected SNP sequences library using Bowtie 2.4.1 with default parameters.[61] All 100% matched reads were filtered for counting unique UMI of each SNP reference/alternative allele. Any SNP allele with no expressions in any one sample or with pooled input/output expression counts less than 10 were excluded for downstream analysis.

SNP fragment enhancer activity analyses were implemented as described by D. Tippens et al.[24] Briefly, raw reads were processed using voom from R limma package (v3.34.9).[62] All replicated input or output samples were treated as different experimental conditions. Output SNP fragments were divided into three categories (candidate enhancer, silencer and inactive fragments). Candidate enhancers were defined as showing significantly higher expression in output compared with input at FDR adjusted $p < 0.05$ and at least 1.5-fold increased activities ($\log_2FC > 0.585$). In contrast, candidate silencers were determined as showing significantly lower expression in output compared with input (FDR<0.05, $\log_2FC < -0.585$). All other fragments showing comparable expression between output and input were considered as inactive ones.

For SNPs showing significant enhancer activity effect on at least one allele-containing fragment (eSNP), we further employed MPRAnalyze (v1.16.0)[26] for allelic reads comparison between two alleles to identify SNPs with biased allelic enhancer activity effect (baaSNP). We calculated p-value using likelihood ratio test with MPRAnalyze[26] and determined eSNPs with FDR<0.05 as potential baaSNPs.

Epigenetic characterizations on baaSNPs and silencer SNPs

SNPs showing silencer activity effect on at least one allele while without enhancer activity effect on both two alleles were defined as silencer SNPs. All other SNPs except for silencer SNPs or eSNPs were classified as inactive SNPs. We analyzed whether identified baaSNPs are located or near DNase-seq peak or two typical enhancer-indicative markers (H3K4me1, H3K27ac) peaks in osteoblast cells collected from ENCODE portal[44] (Table S1) using bedtools (v2.29.0). To further explore epigenetic characterizations on

identified or baaSNPs or silencer SNPs, we compared percentage of baaSNPs/silencer SNPs overlapping with different epigenetic marker peaks or chromatin segment (HMM15) or DNase-seq peaks in osteoblast cells (Table S1) against inactive SNPs using Fisher's exact test. We also compared average signals on DNase-seq or two typical enhancer-indicative markers (H3K4me1, H3K27ac) surrounding baaSNPs against non-baaSNPs (but eSNPs) using Wilcoxon rank-sum test. Normalized signal reads for selected epigenetic markers were obtained from UCSC Genome Browsers (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/, Table S1). Comparison of average surrounding signals between baaSNPs and non-baaSNPs was visualized using computeMatrix (reference-point –referencePoint center -b 20000 -a 20000 –binSize 10) and plotProfile from DeepTools (v3.3.0).[63] For DNase-seq and two enhancer-indicative (H3K4me1, H3K27ac) epigenetic markers, we also evaluated their relevance to biased enhancer activity effect of baaSNPs by comparing their allelic activity effect (absolute value of $\log_2$FC) between annotated against non-annotated baaSNPs per marker using Wilcoxon rank-sum test.

Cis-eQTL association analysis

Since there were no *cis*-eQTL data in direct osteoporosis-related human tissues with large sample available, we examined potential genetic association between baaSNPs and nearby gene expression from GTEx V8 across 49 tissues,[29] which had been widely used for identifying potentially causal osteoporosis regulatory genes before.[11,64] We applied GTEx genotype data across 8,383 samples from dbGap (dbGap: phs000424.v8.p2), and downloaded precomputed lead SNPs of conditionally independent *cis*-eQTLs for each gene and gene expression and covariates files in each tissue from GTEx portal (https://gtexportal.org/home/). To investigate additional *cis*-QTL association signals, we implemented plink2 (–glm) for conditional *cis*-QTL association analyses between baaSNPs and nearby genes using linear model controlling for all covariates and all other conditionally independent *cis*-eQTL signals. For gene with multiple conditionally independent *cis*-eQTL association, the most significant one was retained. Any genes with accordant direction of *cis*-eQTL association and STARR-seq allelic enhancer activity on baaSNPs at nominal significance ($p < 0.05$) were extracted.

GWAS-eQTL colocalization analysis

To explore whether Hi-C interacted genes on baaSNPs shared the same causal variants between osteoporosis-relevant GWAS association and *cis*-eQTL association, we employed fast enrichment estimation aided colocalization analysis (fastENLOC)[30] by integrating osteoporosis-relevant GWAS summary statistics (fracture, TB[LH]BMD: FABMD, FNBMD, LSBMD and eBMD) and GTEx V8 *cis*-QTL association across 49 tissues.[29] All data resources used are summarized in Table S1. We ran fastENLOC (v2.0) using default parameters and used the recommended locus-level colocalization probability (LCP) > 0.1[30] to determine potential causal gene-trait associations. We performed enrichment analysis by comparing percentage of potential osteoporosis causal genes (LCP>0.1) between Hi-C chromatin interacted genes and all surrounding genes on baaSNPs (1000-KB) using Fisher's exact test.

Transcriptome-wide association studies analysis

Transcriptome-wide association studies (TWAS) was conducted using the S-MulTiXcan software[31] by integrating GWAS summary statistics on multiple osteoporosis-relevant traits (fracture, TB[LH]BMD: FABMD, FNBMD, LSBMD and eBMD) and SNP-expression correlation from GTEx (V8 version).[29] S-MulTiXcan combines gene-disease association results across tissues from S-PrediXcan to increase power. Preprocess of each GWAS data included harmonization and imputation as recommend by software manual. We ran S-PrediXcan on all processed GWAS summary statistics using pre-computed MASHR-M models across 49 GTEx tissues.[31] All imputation support data and PrediXcan MASHR model were downloaded at Zenodo: https://zenodo.org/record/3657902/. We considered genes at Bonferroni-adjusted $p < 0.05$ as osteoporosis-relevant. Enrichment analysis was performed using Fisher's exact test by comparing percentage of osteoporosis-associated genes (Bonferroni-adjusted $p < 0.05$) between Hi-C chromatin interacted genes and all surrounding genes on baaSNPs.

Pathway enrichment analysis

To gain an overview of biological pathways involved for chromatin interacted genes on baaSNPs, we used clusterProfiler R package (v3.6.0) with default parameter[65] to analyze their enrichment of GO biological process terms. We included GO terms with annotated genes ≥ 5, and used an FDR of <0.05 to select significant enriched ones. We also manually selected all annotated GO biological process terms related to skeletal development or osteoblast processes (e.g., Wnt signaling pathway) and extracted involved genes as putative known osteoporosis-associated genes.

Screening genes with skeletal-related physiological effect

We queried skeletal relevant physiological effect of gene knock out in mouse model for chromatin interacted genes of baaSNPs based on the International Mouse Phenotyping Consortium (IMPC) portal (http://www.mousephenotype.org/, release-17.0). Any genes with abnormal skeletal development or bone structure effect in gene knockdown mouse model were retained as putative known osteoporosis-associated genes.

Allele-specific motif prediction

To predict potential allelic occupying transcription factor binding motifs on baaSNPs or inactive SNPs, we extracted genomic sequences encompassing different allele of each SNP and employed FIMO from MEME Suite toolkit (v5.4.1)[36] with default parameters

and TF motifs from multiple public motif databases, including JASPAR (2022 version),[66] HOCOMOCO (v11),[67] Transfac and Jolma2013.[68] Any motifs with allele-specific prediction at default $p < 1 \times 10^{-4}$ were retained.

TF enrichment analyses and functional assessment

For all predicted motif TFs showing allele-specific binding to at least 5 baaSNPs, we firstly compared percentage of predicted binding baaSNPs against inactive SNPs using Fisher's exact test, and determined any TFs significantly higher enriched at $p < 0.05$ and $FC > 1$. We also compared counts of binding baaSNPs on higher or lower enhancer activity allele for predicted TFs. The calculated percentage of higher enhancer activity allele binding might act as an empirical indicator for TF function, with higher value implying potential enhancer activation effect while lower value indicating potential enhancer repressor function. To confirm the enhancer activation or repressor function on enriched TFs, we performed functional enrichment analysis for Gene Ontology (GO) molecular function terms related to transcriptional activator/coactivator activity or transcriptional repressor/corepressor activity using clusterProfiler R package with default parameter.[65] Similarly, we scrutinized enrichment for skeletal-relevant GO biological process terms on enriched TF genes using clusterProfiler R package.[65] To explore putative roles of enriched TFs in early-stage osteoblast differentiation, we analyzed their expression change in osteoblast differentiated for different timepoint (0-3h [Phase I, differentiation initiation], 6-24h [Phase II, lineage-acquisition], 48-96h [Phase III, lineage-progression]) compared with non-differentiated cells based on gene expression profile from GEO (GEO: GSE80614) (Table S1).[40] Raw probe expression was converted and normalized using lumiExpresso function form R package lumi (v.2.30.0)[69] with any probes detected at least once in all biological replicates (detection $p < 0.01$) retained. We then implemented the moderated T-statistics with default parameters from R package limma (v.3.34.9) for gene expression comparison. For genes corresponding to multiple probes, the highest expressed ones were retained. We declared TF genes at FDR of <0.05 as significantly differentially expressed.

Genetic regulatory network analysis

To prioritize putative key controlling TFs for osteoporosis, we constructed a compounded directed regulatory network encompassing baaSNP-TF and TF-gene interactions. Above identified TFs significantly enriched for baaSNPs ($p < 0.05$) were selected for network construction. For any baaSNPs with allele-specific motif prediction on selected TF, we defined their promoter chromatin interacted genes as putative regulatory genes on predicted TF. We speculated that candidate genes were regulated by TFs thorough enhancer-promoter chromatin interactions and the enhancers activities were dependent on different allele of binding baaSNPs. The network was visualized by Graphia (https://graphia.app/).[53] The weighting of TF-gene interaction was defined by the incorporation of three pieces of information: the intensity of biased enhancer effect on baaSNPs with predicted TF binding measured by STARR-seq($\log_2$FC), TF motif prediction score on baaSNP from MEME Suite,[36] and intensity of significant Hi-C chromatin interactions between baaSNP with predicted TF binding and candidate target gene ($\log_2$[interaction reads]). For baaSNP-gene pair with multiple Hi-C chromatin interactions, the strongest one was selected for scoring. Each of these three pieces of information was ranked and then normalized, and their products are multiplied to obtain the final weight for each TF-target gene connection as follows (Equation 1):

$$W_{TF-gene} = rank_{baaSNP\ starr-seq} * rank_{baaSNP-gene\ HiC} * rank_{TF\ motif\ score} \quad \text{(Equation 1)}$$

where $W_{TF\text{-}gene}$ denotes weighting for each TF-gene pair normalized via dividing it by total pair counts ($0 < W_{TF\text{-}gene} \leq 1$). Subsequently, for each TF, the cumulative weighting of all involved TF-gene pairs was calculated to derive its final functionality ranking score, which was defined as follows (Equation 2):

$$Score_{TF} = \sum_{j=1}^{n} W_{(TF-gene)j} \quad \text{(Equation 2)}$$

where $W_{TF\text{-}gene}$ denotes calculated weighting by Equation 1, and j denotes involved TF-gene pair ($1 \leq j \leq n$) for calculated TF (assuming n total TF-gene pairs). For selected top ranked TF, we extracted its putative regulatory genes and analyzed their functional relevance to osteoporosis, including whether they were involved in osteoporosis-relevant biological pathways, whether they showed skeletal phenotypic abnormalities in gene knockout mouse models, or showed putative causal genetic regulatory effect for osteoporosis by colocalization analyses, or potential genetic association with osteoporosis by TWAS analyses. Normalized ChIP-seq signals on YY2 in HEK293 cell surrounding rs11202530 region was downloaded from ENCODE portal[44] and visualized using Integrative Genomics Viewer (https://igv.org/).

Comparative epigenetic and Hi-C interaction analyses for rs11202530

We compared significant chromatin interactions (FDR<0.05) between rs11202530 and *PAPSS2* promoters (1-KB surrounding TSS) in hMSC and hMSC differenced osteoblast or adipocyte.[17] Normalized ChIP-seq signals on H3K27ac surrounding rs11202530 were compared and visualized using Integrative Genomics Viewer (https://igv.org).

YY2 footprints prediction and functional enrichment analyses

To identify potential active YY2 binding sites, we called digital footprint-like genomic regions based on DNase-seq in human primary osteoblast cells from ENCODE[44] by Wellington from pyDNase (v0.2.4),[43] followed by YY2 motif prediction by FIMO from MEME Suite

toolkit (v5.4.1)[36] (Table S1). For Wellington analysis, default parameters were used with genomic regions at FDR adjusted $p < 0.01$ filtered as potential footprints. Predicted footprints regions were then extracted for potential YY2 motif prediction to refine potential active YY2 binding sites. To explore whether predicted YY2 binding sites are significantly enriched in peak region of different epigenetic markers (inlcuding CTCF binding site) in osteoblasts or significant Hi-C chromatin interaction regions in hMSC differentiated osteoblast cells (Table S1), we randomly permuted genomic locations of predicted YY2 binding sites from DNase-seq peak regions in osteblasts using bedtools shuffle (-chrom -excl <YY2 motif> -noOverlapping) 1000 times. We then calculated counts of overlapped random locations against true locations for each feature using bedtools (v2.29.0). Fold enrichment and standard deviation per feature was calculated by dividing overlap of true locations by random locations, and empirical *p* value was calculated counting how many times euqual or higher overlap in true YY2 motif locations was observbed compare to randomly permuted locations. For predicted YY2 binding sites, we also extracted their chromatin interacted gene promoters using bedtools and performed GO biological process enrichment analyses using clusterProfiler R package (v3.6.0).[65]