# Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture

**Citation for published version:**
UK10K Consortium & AOGC Consortium 2015, 'Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture' Nature. DOI: 10.1038/nature14878

**Digital Object Identifier (DOI):**
10.1038/nature14878

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
Nature

# Whole-genome sequencing identifies non-coding variants near *EN1* with large effects on bone mineral density

## SUPPLEMENTARY INFORMATION

## 1. Cohorts

Cohort-level characteristics are described in **Supplementary Table 18**. **Supplementary Table 19** describes BMD measurements and covariates used in association testing for all cohorts.

## 2. Sequencing

**Supplementary Table 20** summarizes the data generation method for sequencing-based cohorts. A detailed description is provided below for the whole genome and whole exome sequenced cohorts.

### Whole Genome Sequencing

*UK10K*

ALSPAC and TwinsUK cohorts were sequenced to an average read depth of 6.7x through the UK10K program (www.uk10k.org). Whole genome sequencing was completed at the Wellcome Trust Sanger Institute and Beijing Genomics Institute. Briefly, DNA was sheared and subjected to Illumina paired-end DNA library preparation and sequenced using the Illumina HiSeq platform. Reads were aligned to the GRCh37 human reference, using BWA (v0.5.9-r16).[1] BAM files were further processed to realign around known INDELs, base quality score recalibration, addition of BAQ tags using GATK.

SNV calls were completed using samtools/bcftools (version 0.1.18-r579),[2] and were then called to produce a VCF file.[3] The pipeline to create these calls is available from: https://github.com/VertebrateResequencing/vr-codebase/tree/develop.

Sites were called using the Variant Quality Score Recalibration (VQSR)[4] and GATK Unified Genotyper was used to recall the sites and alleles discovered by samtools. The VariantRecalibrator within GATK was used to first model the variants, then GATK ApplyRecalibration was applied to assign VQSLOD scores (see http://www.broadinstitute.org/gsa/wiki/index.php/Variant_quality_score_recalibration for more details).

VQSLOD score threshold was set at -0.6804 which improved site concordance with duplicate samples sequenced using high-depth exome sequence methods. Two further thresholds were applied: failure of HWE ($P<1e-6$) and evidence of batch effects between samples genotyped at Sanger and BGI ($P<1e-2$). In total, 4.2M variants were removed through these QC criteria. The final set contained over 42M SNVs.

In order to assess the validity of the sequence data, 61 samples were also sequenced using high-depth next generation exome sequencing (depth = 70x), described previously.[5] 74,621 sites were shared out of the 86,322 sites that were called in the exome data, representing a sensitivity of 86.4% across the 35Mb bait region. We assessed the non-reference discordance rate and found that 0.2% were discordant, suggesting that the resultant dataset had high concordance with high depth exome sequencing.

### Whole Exome Sequencing

*AOGC*

993 samples underwent exome capture and massive parallel sequencing. Sequencing libraries were constructed from 1.6ug genomic DNA using the Illumina TruSeqDNA sample preparation kit. Libraries were combined in pools of six for target capture using the Illumina TruSeq Exome Enrichment Kit. Libraries were assessed pre- and post-capture for both quality and yield using the Agilent High Sensitivity DNA assay and KAPA Library Quantification Kit. Massively parallel sequencing was performed using the Illumina HiSeq2000 to generate 100bp paired-end reads (2x100PE). Either 6 or 12 samples were run per flow cell lane using version 2 or version 3 SBS reagents respectively.

Illumina Data Analysis Pipeline software (CASAVA 1.8.2) was used for de-multiplexing and initial base calling and sequence data was aligned to the current build of the human genome (hg19, released February 2009) using the Novoalign alignment tool (V2.08.02); sequence alignment files were converted using SAMtools (v0.1.16) and Picard tools (http://picard.sourceforge.net)  (v1.42) SNPs and indels were called using the Genome Analysis Toolkit (GATK v2.2-3)[6] including variant recalibration. Variants were annotated using ANNOVAR and Variant Effect Predictor and GERP++ scores.[7,8]

### *FHS*

DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer's protocol (Illumina Multiplexing_SamplePrep_Guide_1005361_D) with modifications as described in the *BCM-HGSC Illumina Barcoded Paired-End Capture Library Preparation* protocol. Libraries were prepared using Beckman robotic workstations (Biomek NXp and FXp models).  The complete protocol and oligonucleotide sequences are accessible from the HGSC website (https://hgsc.bcm.edu/sites/default/files/documents/Illumina_Barcoded_Paired-End_Capture_Library_Preparation.pdf).

Briefly, 1 µg of DNA in 100ul volume was sheared into fragments of approximately 300-400 base pairs in a Covaris plate with E210 system (Covaris, Inc. Woburn, MA) followed by end-repair, A-tailing and ligation of the Illumina multiplexing PE adaptors.  Pre-capture Ligation Mediated-PCR (LM-PCR) was performed for 6-8 cycles of amplification using the 2X SOLiD Library High Fidelity Amplification Mix (a custom product manufactured by Invitrogen).  Universal primer IMUX-P1.0 and a pre-capture barcoded primer IBC were used in the PCR amplification.  In total, a set of 12 such barcoded primers were used on these samples.  Purification was performed with Agencourt AMPure XP beads after enzymatic reactions. Following the final XP beads purification, quantification and size distribution of the pre-capture LM-PCR product was determined using the LabChip GX electrophoresis system (PerkinElmer).

For the hybridization step, four or six pre-capture libraries were pooled together (~250 ng/sample for a 4-plex and ~166 ng/sample for a 6-plex, totaling 1 µg per pool).  These pooled libraries were then hybridized in solution to the HGSC VCRome 2.1 design (42Mb, NimbleGen) according to the manufacturer's protocol *NimbleGen SeqCap EZ Exome Library SR User's Guide (Version 2.2)* with minor revisions.  Human COT1 DNA and full-length Illumina adaptor-specific blocking oligonucleotides were added into the hybridization to block repetitive genomic sequences and the adaptor sequences.  Post-capture LM-PCR amplification was performed using the 2X SOLiD Library High Fidelity Amplification Mix with 14 cycles of amplification. After the final AMPure XP bead purification, quantity and size of the capture library was analyzed using the Agilent Bioanalyzer 2100 DNA Chip 7500.  The efficiency of the capture was evaluated by performing a qPCR-based quality check on the four standard NimbleGen internal controls.  Successful enrichment of the capture libraries was estimated to range from a 6 to 9 of ΔCt value over the non-enriched samples.

Library templates were prepared for sequencing using Illumina's cBot cluster generation system with TruSeq PE Cluster Generation Kits.  Briefly, these libraries were denatured with sodium hydroxide and diluted to 6-9 pM in hybridization buffer in order to achieve a load density of ~800K clusters/mm$^2$.  Each library pool was loaded in a single lane of a HiSeq flow cell, and each lane was spiked with 1% phiX control library for run quality control.  The sample libraries then underwent bridge amplification to form clonal clusters, followed by hybridization with the sequencing primer.  Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform.  Using the TruSeq SBS Kits, sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional 7 cycles for the index read.

### *RS-I*

Genomic DNA of RS participants were prepared from blood and fragmented into 200-400 bp fragments using Covaris Adaptive Focused Acoustics (AFA) shearing according to the manufacturer¹s instructions (Covaris, Inc., Woburn, MA). Illumina TruSeq DNA Library preparation (Illumina, Inc., San Diego, CA) was performed on a Caliper Sciclone NGS workstation (Caliper Life Sciences, Hopkinton, MA), followed by exome capture using the Nimblegen SeqCap EZ V2 kit (Roche Nimblegen, Inc., Madison, WI). This capture targets 44Mb of exonic regions covering 30,246 coding genes, 329,028 exons and 710 miRNAs.

Paired-end 2 x100 sequencing was performed on Illumina HiSeq2000 sequencer using Illumina TruSeq V3 chemistry.

Reads were mapped to the human genome hg19 with the BWA algorithm and processed with the Genome Analysis Toolkit (GATK) to recalibrate base quality-scores and perform local realignment around known indels.

*Variant calling and Quality Control*

Target coverage or each sample was computed with the GATK. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were called with the Unified Genotyper module of the GATK and filtered to remove SNVs with annotations indicative of technical artifacts (such as strand-bias, low variant call quality, or homopolymer runs). Samples with fewer than 76% of targeted bases covered to 20x, with an abnormally high number of non-reference alleles or heterozygosity, or with an abnormally low concordance with prior SNP array genotypes (based on the distribution across all samples) were excluded from analysis. Any sample genotype at a site with a genotype quality (GQ) < 20 in the sample was ignored (e.g. set as missing). Variants were annotated with the Variant Effect Predictor.[7]

### *ESP*

The processes of library construction, exome capture, sequencing, and mapping were performed as previously described. Sequencing was performed at the University of Washington (UW) and the Broad Institute of MIT/Harvard (Broad). Single Nucleotide Variants (SNVs) were called using the UMAKE pipeline at University of Michigan, which allowed all samples to be analyzed simultaneously, both for variant calling and filtering. Briefly, we used BAM files summarizing Burrows-Wheeler Alignment (BWA) alignments generated at the UW and the Broad as input.[9] These BAM files summarized alignments generated by BWA, refined by duplicate removal, recalibration, and indel re-alignment. We excluded all reads that were not confidently mapped (Phred-scaled mapping quality < 20) from further analysis. To avoid PCR artifacts, we clipped overlapping ends in paired reads. We then computed genotype likelihoods for exome targeted regions and 50 flanking bases, accounting for per base alignment quality (BAQ) using SAMtools.[9] Variable sites and their allele frequencies were identified using a maximum-likelihood model, implemented in glfMultiples.[10] These analyses assumed a uniform prior probability of polymorphism at each site. The final call-set was performed on 6,823 samples.

We used a support vector machine (SVM) classifier to separate likely true positive and false-positive variant sites, applying SNP quality metrics including allelic balance (the proportional representation of each allele in likely heterozygotes), base quality distribution for sites supporting the reference and alternate alleles, and the distribution of supporting evidence between strands and sequencing cycle, amongst others. We used as the positive training set variants identified by dbSNP or 1000 Genomes and we used variants that failed multiple filters as the negative training set. We found this method to be effective at removing sequencing artifacts while preserving good-quality data, as indicated by the transition-transversion (ti-tv) ratio for previously known and newly identified variant sites, the proportion of high frequency variants overlapping with dbSNP, and the ratio of synonymous to non-synonymous variants, as well as attempts at validation of a subset of sites. A total of 1,908,614 SNVs passed the SVM filter. Among these, genotypes with a corresponding read depth less than 10 were set to missing.

### *ERF*

The exomes of 1,336 individual from the ERF population were sequenced at the Center for Biomics of the Cell Biology department of the Erasmus MC, The Netherlands. The individuals were selected random with the regards to ADHD scores. The sequencing was performed using the Agilent version V4 capture kit on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol. The sequence reads were aligned to the human genome build 19 (hg19) using BWA and the NARWHAL pipeline.[11,12] The aligned reads were processed further using the IndelRealigner, MarkDuplicates and TableRecalibration tools from the Genome Analysis Toolkit (GATK) and Picard (http://picard.sourceforge.net). Genetic variants were called using the Unified Genotyper tool of the GATK. About 1.4 million Single Nucleotide Variants (SNVs) were called. After removing the low quality variants (QUAL < 150), variants with a call rate of < 95% and variants that were out of Hardy Weinberg Equilibrium (HWE; $p\text{-value} < 10^{-6}$) and monomorphic variants, and removing individuals with a low call rate (< 90%) we retrieved 540,633 SNVs and 1,301 individuals.

## 3. Whole Genome Genotyping

All genome-wide genotyping has been previously reported,[13] except for MrOs and SOF. Genotyping of MrOS and SOF was performed using the Illumina HumanOmni1_Quad_v1-0 B array at the Broad Institute, Cambridge, MA. Genotypes were called using the Illumina's BeadStudio calling algorithm. The sample quality control exclusion criteria were sample call rate < 97%, excessive autosomal heterozygosity, first and second degree relatives, genotypic sex mismatch using X and Y chromosome probe intensities and gross chromosome abnormalities.

### Imputation of Whole Genome Genotyped Cohorts

All samples were imputed using the same reference panel and pipeline. Before imputation, the phased UK10K haplotypes (3,781 samples) were merged with 1000 Genome haplotypes (1,092 samples) using IMPUTE2[14]. We excluded singletons in both reference panels before merging. To account for the fact that each panel might contain sites that are not present in the other panel, we imputed the missing variants from each study into the other study and then combined the two reference panels as the union set of sites.[15] A total of 41,992,162 variants were finally presented in the combined 9,746 haplotypes.

The GWAS data were pre-phased first without a reference panel, using best-guess haplotypes.[16] Then variants from the combined UK10K/1000Genomes reference panel were imputed into the best-guess haplotypes of genome wide genotyped cohorts. IMPUTE2 reports an information (info) metric as a measurement of post-imputation quality. The info metric typically takes values between 0 and 1. A value near 1 indicates a SNV has been imputed with high certainty, whereas a value of 0 means that there is complete uncertainty about the genotypes. The info metric is used to filter poorly imputed SNVs. We applied a cutoff of an info score of 0.4 to all included SNVs in the association analysis. **Supplementary Table 21** denotes cohort-level imputation statistics. Separating the phasing and imputation steps in this way is beneficial because most of the computational burden of imputation comes from accounting for the unknown phase of the GWAS samples.

## 4. Association Testing

### Single Variant Methods

Single variant association testing was undertaken using an additive model separately for each BMD site. BMD was defined a forearm (distal 1/3 of radius), lumbar spine (L1-4) and femoral neck, as measured by DXA. Since different DXA machines have known systematic differences in BMD measurements, BMD was standardized within each cohort to have a mean of zero and a standard deviation of one. This also assists in interpretation of data since the effect size of each allele can therefore be measured in standard deviations. BMD was adjusted for age, age-squared, sex and weight. In sex-specific analyses, the sex term was removed.

The type of software employed for single variant testing for each cohort is shown in **Supplementary Table 2**. Cohort-specific genomic inflation factors (lambdas) are also shown in **Supplementary Table 2** (the mean lambda was 1.044).

Single variant tests were undertaken for variants with a MAF ≥0.5%, using an additive effect of the minor allele at each variant in each cohort. For single-variant tests, statistical significance was declared after taking into account all of the independent tests among variants with MAF ≥0.05%, as we have described recently.[17] Briefly, to estimate the number of independent SNVs in the UK10K reference panel, we obtained sequence data from chromosome 3 on 2,577,674 genetic variants in 2,432 individuals from the UK10K program. This comprised of 798,175 SNVs with MAF ≥0.005 and the estimated number of SNVs at this threshold, genome-wide was 12,451,530. We resolved these SNVs to 4,268,111 independent tests by calculating the predicted effective number of independent tests and corrected $\alpha = 0.05$ by 4,268,111 to yield a Bonferroni corrected genome-wide significant threshold of 1.2x $10^{-8}$ for single-variant tests at MAF ≥0.5%.

## Single Variant Meta-analysis

Meta-analysis of cohort-level SNV association statistics was undertaken using fixed-effects meta-analysis in GWAMA.[18] Single variant Forest and Locuszoom plots are shown in **Extended Data Fig. 2 and 10**. QQ and Manhattan plots for single variant associations are shown in **Extended Data Figure 1b and 1c**, respectively. QQ and Manhattan plots for variants present in both exome-based and genome-based cohorts are presented in **Extended Data Figure 11a and b**, respectively. Sex-stratified results for novel genome-wide significant variants is provided in **Supplementary Table 22**.

## Rare Variant Analysis and Meta-analysis

To test whether variants with low MAF influenced BMD we undertook a region-based collapsing method, the sequence kernel association test (SKAT), to combine information across low-frequency (MAF ≤5%) and rare (MAF ≤1%) or rare variants alone. Meta-analysis was conducted using skatMeta,[19] where each cohort provided the necessary intermediate files (skatCohort objects and snpInfo files). Analysis with skatMeta requires providing beforehand the genotypes as input to the skatCohort function. For exome-seq data, the analysis can accommodate genotype data from an entire chromosome; however, for whole genome sequence or imputed data the analysis was not possible due to memory constraints. As a result, genotype data was divided into gene regions, defined as the maximal extent of all isoforms of a gene, with an additional 5kb on either end. This approach provided the flexibility to meta-analyze various subgroups of SNVs across an entire gene, such as coding variants or variants under varying evolutionary constraint.

Each cohort performed the following:

1. Converted genotype data to binary IMPUTE2 format using QCTOOL.
2. Using custom scripts, prepared input files for analysis with skatCohort or skatFamCohort (for family-based cohorts):
    a. Fetch all SNVs per gene region, where a region is defined as the maximal extent of all gene isoforms plus a predefined flanking region (i.e. 5kb). Due to computational constraints, gene regions greater than 150kb, are broken into multiple non-overlapping smaller regions of at most 100kb and at least 50kb.
    b. For each gene region, prepare files to support analysis with skatCohort and downstream meta-analysis (SNPInfo files, and SNP statistics).
3. Using custom scripts, execute skatCohort or skatFamCohort per gene-region.

Cohort-level skatCohort objects were meta-analyzed using skatMeta by use of a custom R script. Within each gene region, windows of 30 SNVs (overlapping by 10 SNVs) were analyzed using skatMeta with default parameters, except for the use of the skatOMeta method and the "liu" p-value method. QQ and Manhattan plots of all skatMeta results are shown in **Extended Data Fig. 7a and 7b**, respectively. These results demonstrate that there were few signals that departed from the line of expectation under the null. All significant findings from skatMeta were driven by single variants whose significance was also seen from single variant testing. For example, **Extended Data Fig. 8a and 8c** shows that the skatMeta genome-wide significant signals at *CPED1* were driven by single SNVs, which were present in several cohorts. These SNVs also achieved genome-wide significance in single variant testing. We also identified a region in *DOCK8* and *HEXB* that achieved genome-wide significance, but arose in only one cohort and we feel requires further replication (**Supplementary Table 17** & **Extended Data Fig. 8b**).

## Genome-wide suggestive and significant loci from region-based association tests

**Supplementary Table 17** lists all genome-wide significant ($P < 1.2 \times 10^{-8}$) and suggestive loci ($P < 1.2 \times 10^{-6}$) for region-based association tests.

## Conditional Analyses for Single Variant Associations

Conditional analysis was conducted using GCTAV.0.93.9.[20] This method uses an approximate conditional analysis approach from summary-level statistics from the meta-analysis and LD corrections between SNVs estimated from a reference sample. We used UK10K individuals as the reference sample to calculate the

LD information of SNVs. The associated regions flanking within 400kb of the top SNVs were extracted and the conditional analyses were conducted within these regions. A stepwise model selection procedure was performed to select independently associated SNVs with a threshold of $P \leq 5 \times 10^{-6}$. Then we conditioned on these independently associated SNVs to seek secondary signals. Qualtile-quantile plots for these results are presented in **Extended Data Figure 1**. Conditional analyses of individuals variants presented in **Supplementary Table 6** was conducted using GCTA v 0.93.9 using default parameters.

## Relationship between MAF and Effect Size

For each BMD phenotype, genome-wide significant SNVs ($P \leq 1.2 \times 10^{-8}$) were collected and pruned for linkage disequilibrium by first placing variants into 4 discrete MAF bins: [0.005 - 0.01], [0.01 - 0.05], [0.05 - 0.1], and [0.1 - 0.5]. For each MAF bin, we used the SNP Annotation and Proxy Search (SNAP)[21] to calculate the pair-wise correlation ($r^2$). We retained independent SNVs with the largest effect size by removing all the other SNVs in LD ($r^2 > 0.2$).

Power for single point tests was calculated using standard approaches. Let $N$ be the sample size, $p$ be the minor allele frequency, $\beta$ represent the standardized effect of a SNV on a continuous phenotype (standardized so that $\beta$ is the effect per standard deviation of the phenotype), and let $R^2$ represent the square of the correlation between a true genotype and a genotype measured with error. The non-centrality parameter of the chi-squared distribution for a single SNV has been shown to be $NCP = 2(N-1)p(1-p)\beta^2R^2$.[22] We calculated power from a non-central chi-squared distribution for the genome-wide significance threshold of $1.2\text{x}10^{-8}$ and $r^2 = 1$.[17] This was computed for each BMD phenotype and across 4 MAF bins (0.005-0.01, 0.01-0.05, 0.05-0.1, and 0.1-0.5).

## Fracture Meta-analysis

Fracture was defined as a bone fracture resulting from trauma of any type since even high trauma fractures are strongly associated with risk of osteoporotic fractures. Fractures included were those occurring at any site, except fingers, toes and skull, after age 18. Both incident and prevalent fractures were included. Fractures were verified by either radiographic, casting or clinical reporting.

All 1,482 genome-wide significant SNVs associated with femoral neck, lumbar spine and forearm BMD were tested for their association with risk of fracture in a sample size of 10,459 cases and 27,581 controls (38,040 total), of which 76.2% of the samples overlap the BMD discovery samples (**Supplementary Table 23**).

To obtain association statistics, in each participating study, a logistic regression model (GEE model for family-based studies) adjusted for age, age$^2$, sex, height, weight, estrogen/menopause status (women only), ancestral genetic background (PCs) and cohort-specific covariates (such as clinical centers) was applied. The summary effect estimates for fracture risk were computed using fixed-effects inverse variance meta-analysis unless heterogeneity was detected (as defined by $I^2 > 50\%$), where random effects models were also used. To correct for multiple testing, we estimated the effective number of independent SNVs by principal component analyses in the UK10K sequencing dataset and found that 74 principle components explain 97% of the variance in the number of SNVs tested. We then used a Bonferroni correction to estimate the type 1 error. The multiple testing-corrected significant p-value threshold is $P < 0.000676 = (0.05/74)$.

No novel regions achieved genome-wide significance. Accounting for multiple testing for all BMD genome-wide significant SNVs rs4727923 remained nominally significant ($P = 6.73\text{x}10^{-4}$; **Supplementary Table 4**) which marks the *WNT16* locus, variants at which we have previously described as genome-wide significant for fracture.[23] rs61960954 near the *TNFSF11* locus, was also significant ($P = 3.3 \times 10^{-4}$ **Supplementary Table 4**).

## Rare Variant Meta-Analysis

**Supplementary Table 17** demonstrates that two regions (DOCK8.1.2761-2790 and DOCK8.1.2421-2450) spanning *DOCK8* on chromosome 9 were genome-wide significant in their association with femoral neck BMD ($P = 1.9\text{x}10^{-9}$ and $2.8\text{x}10^{-9}$). This arose from the analysis of all SNVs with MAF ≤1%. These regions have low cumulative MAF (summed across the region) of 0.0055 and 0.006, respectively. To assess whether these regions-based signals were driven by multiple rare variants, or just one single variant, we undertook drop-one analysis, which recomputed the test statistic after sequentially removing one SNV from the collapsed region (**Extended Data Fig. 8b**). These analyses showed that signal at region DOCK8.1.2421-2450 was driven by two variants, while DOCK8.1.2761-2790 was driven by one variant. Neither of these regions contained single SNVs that achieved genome-wide significance on their own.

Next, we tested which cohorts provided this association signal, by sequentially removing one cohort at a time from the test statistic and found that both regions were driven only by signal from the Framingham cohort. While promising, we feel these signals need further follow-up and replication in other cohorts.

Lastly, we tested whether DOCK8 was expressed in mouse calvarial osteoblasts and found that its expression increased importantly over development (**Extended Data Fig. 4a**).

## 5. Replication Genotyping and Combined Meta-analysis

For lumbar spine and femoral neck BMD sites, 17 SNPs were selected to genotype in additional 13 GENOMOS cohorts for replication analysis, the selection criteria of SNPs were listed in **Supplementary Table 24**. Human samples from the 13 GENOMOS cohorts were genotyped on LGC Genomics by KASP genotyping. KASP genotyping assays are based on competitive allele-specific PCR and enable bi-allelic scoring of single nucleotide polymorphisms (SNPs) and insertions and deletions (Indels) at specific loci. ). Replication cohorts were selected only if they contained greater than 1,000 individuals prior to genotyping quality control, given then frequency of the variants in the replication panel. Assays are deemed to be working successfully if clusters are distinct and call rates are consistently high. The data is automatically quality control checked on a per SNP basis. No Template Controls (NTCs) are included on each plate to enable the detection of contamination or non-specific amplification. Two SNPs (rs13046645 and chr3:50906922) were found to be monomorphic in initial test GENOMOS cohorts, thus, only 15 SNPs were genotyped. For replication of rs148771817 for forearm BMD, genotyping was undertaken in a single cohort, since this phenotype is rarely collected in population-based cohorts. Genotyping for this SNP was similarly done at LGC Genomics using KASP genotyping in the AOGC cohort, where the age distribution was matched to the discovery cohorts (age cutoff of <80 years was used). Genotyping quality metrics for all variants are provided in **Supplementary Table 25**.

The genotyping data of the 16 SNPs were formatted into PLINK, the BMD phenotype were adjusted by sex, age, age$^2$, weight and standardized to have a mean of zero and a standard deviation of one within each cohort. The association analysis were performed by PLINK in each cohort, then combined analysis of the 9 discovery GEFOS cohorts and 13 GENOMOS cohorts of the 16 SNPs were conducted with GWAMA. Cohort-level association results are provided in **Supplementary Table 26**. Meta-analysis results from replication cohorts as well as combined discovery/replication cohorts are provided in **Supplementary Table 27**.

Non-reference discordant rate (NRD) presented in main text and **Supplementary Table 5** was computed as the proportion of non-reference genotypes found to be discordant between compared individuals as compared total measured non-reference genotypes for that variant. For imputed variants, genotype probabilities where converted to hard-calls using a genotype probability of 0.9.

## 6. Functional Genomics

### Functional Class Enrichment

To ascertain if functional classification using GERP++ scores[8] enriches for more significant GWAS associations, we computed, for variants that fall below or above a series of GERP++ thresholds the proportion that surpass FDR q-value of 0.05. Prior to FDR analysis, variants from meta-analysis (UK10K + 1KG reference panel SNVs only) were annotated with GERP++ scores and distance to the nearest gene. Variants were then pruned for LD with PLINK using an $r^2$ of < 0.2, window size of 100kb and step of 20 kb. For each SNV in this LD independent set, we selected the single SNV in LD with the lowest p-value, thus yielding a final dataset of LD independent SNVs with lowest p-values. For each GERP++ threshold considered (from 0 to 4, with step of 0.2), we partitioned SNVs into those above (cases) and below (controls) the threshold. To adjust for bias (such as a correlation of GERP++ score to gene distance or MAF) we matched cases to controls using the R package MatchIt (coarsened exact matching algorithm). Subsequent to matching, FDR analysis was conducted on each set of case and control variants, and proportion of variants surpassing FDR q-value of 0.05 was obtained.

For synonymous and deleterious variants, the above strategy was applied, except only coding variants were partitioned based on whether or not they were synonymous or deleterious. Variant annotations were computed using a local version of Variant Effect Predictor.[7] Deleterious variants were classified as having the following sequence ontology terms (frameshift_variant, inframe_deletion, inframe_insertion, initiator_codon_variant, missense_variant, splice_acceptor_variant, splice_donor_variant, stop_gained, and stop_lost).

### Chromatin Accessibility: DNase I Hypersensitivity Site Correlations

For distal/promoter DHS correlation analyses, we utilized DHS data from ENCODE for 305 cell lines (**Supplementary Table 28**). The distal DHS regions were defined as a peak region from ENCODE DHS data that overlapped the GWAS SNP. For rs148771817 (WNT16 locus), the peak region was obtained from the "Digital DNase I Hypersensitivity Clusters in 125 cell types from ENCODE" track from the UCSC Genome Browser. For rs188303909, (*EN1* locus), the peak region was obtained from the "Osteoblasts DNase I HS Peaks from ENCODE/Duke" track from the UCSC Genome Browser. Promoter regions were defined as 500 nucleotide flanking the TSS of genes within 500 Kb of the GWAS SNP being considered. Within each of these promoter or distal DHS regions, we obtained the DHS measurements for all overlapping 100nt bins across the 305 cell lines. Correlation analyses between the distal and promoter DHS 100 nucleotide bins was conducted using Pearson correlation in the R statistical software package. For each distal nucleotide bin, the maximal $R^2$ to across all bins per gene promoter region was reported. Results for *EN1* and *WNT16* loci are shown in **Supplementary Tables 7 & 11,** and **Extended Data Fig. 3**.

### ENCODE Analysis of *EN1* variant rs188303909

A genome-wide significant SNV 7kb upstream of *EN1* (rs188303909[T], MAF= 2.0%, effect size from replication cohorts = +0.14 SD, $P_{meta}$= 1.3 x $10^{-9}$) overlaps a DNase I hypersensitive site (DHS) in osteoblasts bound by CTCF (**Extended Data Fig. 3a**) as well as histone marks. We found evidence of moderate correlation in chromatin accessibility between rs188303909 and the promoter of *EN1* across 305 cell types[24] (maximum $r^2$ = 0.59, P= 1.5 x $10^{-29}$, **Supplementary Table 7**, **Extended Data Fig. 3a**), suggesting that the DHS overlapping rs188303909 may regulate expression of *EN1*. Hi-C data in human embryonic stem cells further supports an interaction between the rs188303909 DHS and the *EN1* promoter, as both sites lie within the same topologically associated domain (TAD).[25] Similarly, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) data in MCF-7[26] cells also supports an interaction between rs188303909 and *EN1* **(Figure 1).**

### ENCODE Analysis of *CPED1/WNT16* variant rs148771817

A genome-wide significant within an intron of *CPED1* (rs148771817 [T], MAF= 0.9%, replication effect size= +0.41 SD, $P_{meta}$= 1.1 x $10^{-11}$) overlaps a DNase I hypersensitive site (DHS) in 14 cell types (**Extended Data Fig. 3b**). We found evidence of moderate correlation in chromatin accessibility between rs148771817 and the promoter of *WNT16* across 305 cell types[24] (**Supplementary Table 11**, **Extended Data Fig. 3b**), suggesting that the DHS overlapping rs148771817 may regulate expression of *WNT16*. Hi-C data in human embryonic stem cells supports an interaction between the rs148771817 DHS and the *WNT16* promoter, as both sites lie within the same topologically associated domain (TAD).[25] However, interaction frequency to nearby regions is also elevated (**Supplementary Table 11**).

## 7. Functional Experiments

### Murine Osteoblast Gene Expression Profiling

Pre-osteoblast-like cells were obtained from neonatal calvaria collected from C57BL/6J mice expressing cyan florescent protein (CFP) under the control of the Col3.6 promoter (pOBCol3.6CFP), using standard techniques.[27] pOBCol3.6CFP mice used with permission for this study. These mice were made in an identical fashion to the previously described pOBCol3.6GFPtpz transgenic mice.[27] The cells were placed into culture for 4 days in growth media (DMEM containing 10% fetal bovine serum (FBS) and 1X penicillin/streptomycin), removed from culture by trypsin digestions and subjected FACS sorting based on the presence/absence of CFP expression to allow for enrichment of osteoblast lineage cells. Cells expressing CFP were returned to culture, plated at a density of 1x$10^4$ cells per cm$^2$, subjected to an osteoblast differentiation cocktail (αMEM containing 50 μg/ml Ascorbic Acid, 4 mM β-glycerol phosphate, 10% FBS and 1X penicillin/streptomycin) and RNA was collected at 2, 4, 6, 8, 10, 12, 14, 16 and 18 days post differentiation. To examine gene expression across osteoblastogenesis, mRNA profiles for each time point were generated by Next Generation High throughput RNA sequencing (RNAseq), using an Illumina HiSeq 2000. Three technical replicates per sample were sequenced. The alignments for abundance estimation of transcripts was conducted using Bowtie version 0.12.9[28] using the NCBIm37 transcriptome as the reference for alignments. Among all the possible alignments that had fewer than or equal to three mismatches against reference transcriptome, we only accepted those with the minimum number of mismatches for each of 100bp read (using both '--all' and '--best' options). Expression level per gene was calculated using RSEM version 1.2.0 using the following parameters: --fragment-length-mean 280 and --fragment-length-sd 50 and expression level for each sample was normalized relative to the per sample upper quartile.[29,30] This data has been submitted to the gene expression omnibus (Accession Number: GSE54461). Results from these experiments for *Dock8* are shown in **Extended Data Fig. 4a.**

### Temporal Expression of *En1* in Mouse Osteoblasts and Osteoclasts

Pre-osteoblast-like cells were obtained from neonatal calvaria collected from C57BL/6J mice using standard techniques. The cells were placed into culture for 4 days in growth media (DMEM containing 10% fetal bovine serum (FBS) and 1X penicillin/streptomycin) at a density of 2x$10^4$ cells per cm$^2$. The pre-osteoblast cells were subjected to an osteoblast differentiation cocktail (αMEM containing 50 μg/ml Ascorbic Acid, 4 mM β-glycerol phosphate, 10% FBS and 1X penicillin/streptomycin) and RNA was collected at 2 and 18 days post differentiation. These cells were independent from the cells used for the RNA-seq experiments and were not subjected to FACS. Bone marrow derived osteoclast precursor cells were isolated from the hind long bones (femur and tibia) of six week old C57BL/6J female mice. In short, bone marrow was flushed from the long bones and marrow stromal cells were plated at a density of 1x$10^6$ cells per cm$^2$ in αMEM containing 10% FBS, macrophage colony-stimulating factor (25 ηg/ml) and RANKL (100 ηg/ml). Media was changed at day three post plating and RNA was collected three days after that (day six post plating). Each lane for gene expression represents data from one mouse for the osteoclasts cultures and from one well of cells for the osteoblast cultures. For both the osteoclasts and the osteoblast samples, total RNA was isolated using TriZOL (Life Technologies, USA), as described by the manufacturer, and RNA was treated with DNase to remove any contaminating DNA. For each sample, 500 ηg of RNA were then converted to cDNA using standard protocols, with random decamers used as primers for the reaction. Expression of *En1*, *Bglap* (osteocalcin) and *Tnfrsf11a* (RANK) was assessed by PCR. For each PCR reaction, 2 μl of cDNA was added to 48 μl of master mix containing 5 μl of 10X PCR Buffer (Clontech), 0.4mM dNTP's, 2 μM of each of a forward and a reverse primer, 1 μl of Taq (Clontech) with a balance of ddH$_2$O. The PCR was completed using the following cycling conditions: 1 minute (min) hold at 94ºC and 30 cycles of 94ºC for 30 sec and 68ºC for 3 min, followed by a hold at 68ºC for 3 min. The following primers were used: *En1* forward: *TCA AGA CTG ACT CAC AGC AAC CC*, *En1* reverse: *TTG TCC TGA ACC GTG GTG GTA GAG*, *Bglap* forward: *CCA TCT TTC TGC TCA CTC TGC TG*, *Bglap* reverse: *CTT CAA GCC ATA CTG GTC TGA TAG C*, *Tnfrsf11a* forward: *CCA TCA TCT TCG GCG TTT ACT ACA G* and *Tnfrsf11a* reverse: *GGA TTA GGA GCA GTG AAC CAG TCG*. PCR products were visualized by separating them on a standard 2% agarose gel stained with Ethidium Bromide.

### Quantitative Expression of *En1*

#### *iCycler iQ thermal cycler*

For osteoblast marker gene expression, total mRNAs were purified from osteoblast cultures. Real time PCR was performed in triplicate using the iCycler iQ thermal cycler and detection system (Applied Biosystems, Carlsbad, CA) following the manufacturer's protocols. Expression of the tested gene was normalized relative to levels of GAPDH. Primers used are: mALP forward: CAC AAT ATC AAG GAT ATC GAC GTG A, mALP reverse: ACA TCA GTT CTG TTC TTC GGG TAC A, mOSX forward: ATG GCG TCC

TCT CTG CTT GA mOSX reverse: GAA GGG TGG GTA GTC ATT TG, mBglap forward: GGG CAA TAA GGT AGT GAA CAG mBglap reverse: GCA GCA CAG GTC CTA AAT AGT, mRunx2 forward: TAC AAA CCA TAC CCA GTC CCT GTT T, mRunx2 reverse: AGT GCT CTA ACC ACA GTC CAT GCA, mCol1a1: forward ACT GTC CCA ACC CCC AAA G, Ccol1a1 reverse: ACG TAT TCT TCC GGG CAG AA. For results see **Extended Data Fig. 4d**. Expression within sd*En1* is likely due to detection of the non-functional *En1*$^{Cre}$ allele.

### *StepOnePlus Real Time PCR system*

Total RNA from whole vertebral bone extract was prepared using TRIZOLreagent (Sigma) followed by RNeasy Mini Kit (Qiagen). The RNA was reverse transcribed into cDNA using cDNAkit (Appled Biosystems) and real-time PCR analysis was performed using custom designed real-time PCR assays and the StepOnePlus Real Time PCR system (Applied Biosystems). Primers used are: forward ACT CAT GGG TTC GGC TAA CG, reverse GAC GGT CCG AAT AGC GTG T, and probe CGG TGG TCA AGA CTG ACT CA. 18S ribosomal RNA (4310893E Applied Biosystems) was included as an internal standard. For results see **Extended Data Fig. 4e**. Expression within sd*En1* is likely due to detection of the non-functional *En1*$^{Cre}$ allele.

## Murine bone histology

*Mouse bone tissue preparation:* Two-month old *En1*$^{lacZ/+}$ mice[31] were euthanized by $CO_2$ suffocation and the vertebrae from T13 to S1 were collected as a whole block on ice-cold phosphate-buffered saline (PBS). Soft tissue was gently removed with blunt scissors and a size-11 surgical blade, until the intervertebral discs were visible. The vertebral blocks were then fixed by immersion in 0.25% glutaraldehyde in PBS for 90 minutes at room temperature (RT). After several washes with PBS, the vertebrae were decalcified for 4 days with EDTA 0.5M pH 7.4 at 4°C (2 changes per day), followed by further washes with PBS. The tissue was then cryoprotected by incubation with a solution of 30% sucrose in PBS overnight at 4°C, and then embedded in OCT compound (Tissue-Tek) using dry-ice-cold isopentane (Sigma).

*Histochemical methods:* 10-micron sections were collected with a Leica Cryostat on SuperFrost slides (Fisher) and allowed to dry overnight. Sections were stored at –80°C until use. For enzymatic detection of β-galactosidase activity, slides were allowed to reach room temperature (RT) in a closed box, and OCT was washed away for 15 minutes with warm PBS (37°C), followed with several extra PBS washes. The sections were post-fixed 5 minutes with 4% paraformaldehyde (PFA, Electron Microscopy Sciences) in PBS at RT. After PBS washes, the sections were incubated 2x5 minutes with X-gal buffer (2mM $MgCl_2$, 0.02% NP40 and 0.05% deoxycholate in PBS 0.1M pH 7.4) and then overnight at 37°C in X-gal reaction buffer (20 mg/ml X-gal, 5mM $K_4Fe(CN)_6$ and 5mM $K_3Fe(CN)_6$ in X-gal wash buffer). After PBS rinses, the sections were postfixed 10 minutes in 4% PFA and PBS-rinsed again. At this point, if only X-gal staining was required, the sections were then counterstained with Nuclear Fast Red 0.005% for 15 minutes, serially dehydrated, incubated 3x1 min with xylene, and cover-slipped using DPX mountant (Fisher). If Alkaline Phosphatase (AP) activity was also required, the sections were incubated 2x5 minutes with AP buffer (100mM NaCl, 50mM $MgCl_2$, 0.1% Tween-20 in Tris-HCl 0.1M pH 8.2) and AP activity was revealed by incubation with Fast Red (2 h at 37°C) following manufacturer instructions (Roche). After PBS rinses, a gelatin/Tris mounting medium was used to coverslip the slides.

*Imaging:* Z-stacked DIC images were captured of the LacZ and AP double-stained lumbar vertebrae from 2-month old *En1*$^{lacZ/+}$ reporter mice at 400x using a Nikon Eclipse Ni-E upright microscope and NIS$^{Ar}$ software.

## Mouse Micro-CT

*Mice:* The *En1*$^{Cre/+}$ and *En1*$^{flox/+}$ mouse strains[32,33] were maintained in a mixed background and genotyped as described. Male *En1*$^{Cre/+}$ mice were crossed with *En1*$^{flox/flox}$ females to generate experimental and control animals.

*Tissue collection, preparation and imaging:* Lumbar vertebrae from levels L4 to L6 were collected from 4-month old *En1*$^{Cre/flox}$ (self-deleted conditional *En1* mutants, sd*En1*; n=5), and compared to *En1*$^{+/flox}$

littermates (control; n=6). All vertebrae were cleaned of soft tissue, fixed in 4% PFA (paraformaldehyde) overnight, washed three time in 1XPBS for 30 mins each, and stored in 70% ethanol at 4°C until scanning. Before scanning, the bones were allowed to reach room temperature. L5 vertebral body starting 100μm from the growth plates were used for trabecular bone micro-computed tomography (microCT, μCT) analysis on Scanco μCT 35 (Scanco Medical, Brüttisellen, Switzerland) system. 6μm voxel size, 55KVp, 0.36 degrees rotation step (180 degrees angular range) and a 400ms exposure per view were used for the scans, which were performed in 70% ethanol. The Scanco μCT software (HP, DECwindows Motif 1.6) was used for 3D reconstruction and viewing of images. After 3D reconstruction, volumes were segmented using a global threshold of .4g/c. Directly measured bone volume fraction (BV/TV), thickness (Tb.Th), number (Tb.N) and separation (Tb.Sp), TMD and surface to volume ratio (BS/BV) were calculated for the trabecular bone.

**Extended Data Figure 6e** presents the microCT parameters for data in **Supplementary Table 29**. Significance between control and sd*En1* mutants was computed using an unpaired t-test. The thickness map of the 3D image reconstructed microCT image of the LV5 shows the spatial distribution of the mineral density and is presented in **Extended Data Figure 6a**, with morphological characteristics presented in **Extended Data Figure 6b**. **Extended Data Figure 6c and 6d** also presents mictoCT and morphological characteristics for femur trabecula and cortical bone.

## Dynamic and Cellular Histomorphometry

*Mice for skeletal testing and laboratory blinding*: The same animals were used for histomorphometry as MicroCT. Histomorphometry labs (Rower and Ackert) and MicroCT labs (Dahia and Joyner) were blinded to the results of each other's analyses.

*Sample collection*: Prior to sacrifice, mice received intraperitoneal injections of 10 mg/kg calcein (Sigma C-0875) and 30 mg/kg alizarin complexone (Sigma A-3882) at 7 and 2 days, respectively, prior to sacrifice. The dynamic mineralization activity measurements, which are dependent on being able to detect these dyes in the bone, suffered from a loss of 2 animals in each group due to inadequate uptake of one of the two mineralization dyes. The right hind limb was dissected from the hip, and skin and non-adherent muscle and connective tissues were removed from the bone, without scraping periosteal surfaces. The marrow space was exposed by excising the proximal femur and distal tibial bone leaving with knee joint intact. In addition, the vertebra bodies from L1 through L3 were dissected free of adherent muscle. The samples were placed in 10% formalin at 4ºC and shipped to Rowe laboratory. Upon receipt, the samples completed a 3-day exposure to formalin and subsequently were place in 30% sucrose at -80º C until all the samples had been collected.

*Tissue sectioning*: Samples were entered into a modification of the fluorescence-based computer-automated dynamic and cellular bone histomorphometry as described by Hong et al.[34] The workflow and staining protocols are detailed at the bonebase.org, and is briefly outlined here. Three to four distal femur or vertebral bodies are embedded in the sample mold with the aid spacing device and held in place with layering of the cyro-embedding medium (OCT, ThermoScientific). The ice blocks are positioned on cryostat (Leica CM3050 S) and oriented to cut the multiple bones in the same plane of section. Three 5 µm sections are collected at 100 µm levels using an adhesive cryotape (Cryofilm type IIC(10), Section Lab, Japan.) to capture and maintain the morphology of the section. The tape is bonded to a glass slide (Cryojane, Leica Cat# 39475208) using UV fixation with the sample side exposed. Each slide contains two tapes with all the sections from the Control or sd*En1* samples as well as spots of fluorescent beads (5-10µm) for subsequent image registration. Three slides per the Control and sd*En1* samples containing the 3 levels of section are processed in a single batch for the staining and imaging steps.

*Section staining*: In the first step, the samples are incubated for 5 min in a calcein blue solution to identify the accumulated mineral with a fluorescent signal. The slide is cover-slipped with 30% glycerol and imaged for accumulated mineral and the calcein and alizarin complexone (AC) mineralization lines. The second step removes the coverslips and place the slides in a tartrate acid (TRAP) incubation buffer, followed by 5 minutes in the reaction buffer that contains the fluorescent substrate Elf97. The acid conditions remove all the mineralization signals used in the first step. The slides washed in PBS and cover-slipped with 30% glycerol. The TRAP positive cells and registration beads are captured in the second scanning step. The third step removes coverslip and incubates the sections in the alkaline phosphatase (AP) reaction buffer and followed by 10 minutes in the reaction buffer containing fast red substrate. The PBS washed slides are cover-slipped with 30% glycerol/water containing DAPI. The slides are imaged for the red AP signal, DAPI positive nuclei and registration beads. The final step removes the cover slips and stains the slides with toluidine blue to provide visual tissue morphology and identify cartilage proteoglycan of the articular cartilage and growth plate. The slides are cover-slipped again in 50% glycerol/water and imaged from chromogenic signals.

*Section imaging*: The six slides, each containing 5-6 bone sections, are placed in 3 of the 4 positions of the slide holder and loaded into the slide magazine of the Axioscan Z1 microscope. The software recognizes the bar code of the slide and identifies the individual sections on the slide to create an image file for each section. The operator refines the region of interest (ROI) to be scanned and initiates the scanning procedure using preset exposure times and excitation/emission settings. The light source is the Colibri.2 and the image is captured with a high-resolution monochromic digital camera (Zeiss MRm Rev.3). The DAPI, calcein blue and tetracycline sections were excited with the 385nm LED, calcein with the 470nm LED, and the AC and fast red with the 555nm LED. The emission filters used Chroma, #49000ET (LP 400, Emission 460/50) for calcein blue and DAPI, Chroma #49003ET (LP 515, emission 535/30) for calcein, a custom Chroma filter (LP 425, emission 555/30) for tetracycline and Chroma 49005ET (LP 570, emission 600/50) for AC and fast red. The toluidine blue stain section uses a tungsten light source and Hitachi HV-F202 camera. Using a Plan-Apochromat 10x/0.45 WD objective, the entire ROI is captured at 100X using an automated image stack compression algorithm.

*Image Analysis*: The Axioscan generates proprietary multilevel source files that are exported as individual gray scale jpeg files for each filter setting. The files average approximately 65 megapixels per image, which is about 10 pixels per $\mu m^2$ and corresponds to the size of a DAPI positive nucleus. A total of 9 gray scale files and 1 color file are generated. A proprietary algorithm was utilized to threshold each fluorescent signal and map it back to the mineralized bone architecture as defined by the calcein blue stain.[34] The details of the analysis and the production of measured and calculated data are detailed at bonebase.org. Osteoblastic activity is assessed by a fluorescent AP stain, which is strongly positive in active osteoblasts but still detectable by resting (lining) cell of the bone lineage. The histology measures the proportion of trabecular bone surface (BS) that is positive for AP activity (AP/BS). To control for technical or biological variation in staining activity, the distribution of that signal between bone labeling surfaces (AP_L_BS) and non-labeling surfaces (AP_NL_BS) is calculated. Another measurement that is independent of the total AP activity of a tissue, distributes the proportion of AP activity that is adjacent to mineralizing surfaces (%AP_L) or non-labeling surfaces (%AP_NL). Osteoclastic activity is based on a fluorescent TRAP stain and expressed as the proportion of trabecular bone surface that is TRAP positive (TRAP_BS). The analysis does not account for the size of the individual spots of activity nor the number of nuclei per fluorescent spot. Only activity that resides adjacent to the trabecular surface and not within the marrow space is assessed. The distribution of the TRAP activity to sites of active mineralization (TRAP_L_BS) or osteogenesis (AP_TRAP_BS) versus inactive mineralization (TRAP_NL_BS) has proven to be a useful measurement to discriminate areas of high bone turnover versus bone resorption without an osteogenic response.

## Identification of cells belonging to the En1 lineage

*Breeding of mice and sample collection*: Twelve week old *En1*$^{Cre/+}$*;R26*$^{LSL\text{-}EYFP}$ (*En1* lineage tagged) and *En1*$^{+/+}$*;R26*$^{LSL\text{-}EYFP}$ (control) mice were sacrificed, the vertebral 1-3 were excised, placed in chilled 10% formalin, packed in wet ice and sent to the Rowe lab for imaging.

*Tissue process, sectioning and imaging*: The tissues and sectioning protocol was identical to histomorphology including the multiple rounds of imaging and staining. The EYFP signal was captured during the first imaging step using the Colibri.2 470nm LED and the Chroma 49003ET filter (LP 515, Emission 535/30). The EYFP signal from the *R26*$^{LSL\text{-}EYFP}$ allele is very weak in skeletal tissues requiring significant signal enhancement. (**Extended Data Figure 5**)

### Human EN1 Expression Profiling

Human CD14$^+$ cells from healthy donors were incubated with 20 ng/ml of human M-CSF (peprotech) for one day to generate OCPs. Osteoclast precursors were incubated with 20 ng/ml of M-CSF and 40 ng/ml of human soluble RANKL for five additional days in α-MEM supplemented with 10 % FBS. Cytokines were replenished every 3 days. On day 6, cells were harvested. mRNA of human osteoclasts (n=5), human macrophges (n=4), 293T cells (fibroblasts, n=4), and human osteoblasts (a kind gift from Dr. Jae Hycuk Shim, Weill Conell Medical College, n=3) was measured using real-time PCR. EN1 mRNA levels were normalized relative to GAPDH mRNA. Data are shown as mean ± SEM. *** :$P < 0.001$ by one-way ANOVA. All statistical analyses were performed with Graphpad Prism 5.0 software one-way ANOVA for multiple comparisons (more than two conditions) with posthoc Tukey test. $p<0.05$ was taken as statistically significant. Results are presented in **Extended Data Fig. 4c**.

### Mouse and Rat Homologous Regions

To test possible conservation between human BMD-associated loci and quantitative trait loci (QTLs) linked to bone related traits in experimental mapping panels in rats and mice, comparative genome data were used. For each region of association to BMD in humans, genes flanking SNV markers showing the strongest evidence of association were searched for homologs in the genome assemblies of the mouse (NCBI37/mm9) and rat (RGSC3.4). For the rat QTL, the corresponding locus was tested for co-localization within boundaries of QTLs for bone related phenotypes fine mapped (circa 4Mb) from the heterogeneous stock (HS).[35] For QTL in mice, the peak SNV location was converted to CentiMorgan (cM) distance, as per the Cox-Sex-Average genetic map[36] using the freely available Mouse Map Conversion Tool (http://cgd.jax.org/mousemapconverter/). All identified mouse QTLs[37] mapped within a peak location 5 cM either side of the human peak SNV and are listed, along with the strain pair(s) used to identify the original QTL. Results for these analyses are shown in **Supplementary Table 9**.

## 8. Acknowledgements

### Funding

**ALSPAC**: The UK Medical Research Council (grant 74882), the Wellcome Trust (grant 076467) and the University of Bristol provide core support for ALSPAC. This work was supported by the Medical Research Council MC_UU_12013/4. **Chitra Dahia**: Orthopaedic Research and Education Foundation, grant number# 11-228. **AOGC**: The AOGC was funded by the Australian National Health and Medical Research Council (Australia) (grant references 511132 and 1032571). Funding was also received from the Australian Cancer Research Foundation and Rebecca Cooper Foundation (Australia). MAB is funded by a National Health and Medical Research Council (Australia) Principal Research Fellowship and ELD was previously funded by a National Health and Medical Research Council (Australia) Career Development Award (569807). Ian Reid is supported by the Health Research Council of New Zealand. The Sydney Twin Study and Tasmanian Older Adult Cohort were supported by the National Health and Medical Research Council, Australia. The Dubbo Osteoporosis Epidemiology Study was supported by the Australian National Health and Medical Research Council, MBF Living Well foundation, the Ernst Heine Family Foundation and from untied educational grants from Amgen, Eli Lilly International, GE-Lunar, Merck Australia, Novartis, Sanofi-Aventis Australia and Servier. The OPUS study was supported by Sanofi-Aventis, Eli Lilly, Novartis, Pfizer, Proctor&Gamble Pharmaceuticals and Roche. The Hertfordshire Cohort Study was supported by grants from the Medical research Council UK & Arthritis Research UK. The Oxford Osteoporosis Study was funded by Action Research, UK. The UK HBM study was supported by The Wellcome Trust and the NIHR CRN (portfolio number 5163); supporting CLRNs included Birmingham and the Black Country, North and East Yorkshire and Northern Lincolnshire, South Yorkshire, West Anglia and Western. CLG was funded through a Wellcome Trust Clinical Research Training Fellowship (080280/Z/06/Z) and is now funded by Arthritis Research UK (grant ref 20000). **Bristol HBM Team:** The Wellcome Trust and the NIHR CRN (portfolio number 5163); supporting CLRNs included Birmingham and the Black Country, North and East Yorkshire and Northern Lincolnshire, South Yorkshire, West Anglia and Western. CLG was funded through a Wellcome Trust Clinical Research, Training Fellowship (080280/Z/06/Z) and is now funded by Arthritis Research UK (grant ref 20000). KESP acknowledges the support of Cambridge NIHR Biomedical Research Centre and the MRC Human Nutrition Research unit, Cambridge. **WHI**: Funding for GO ESP was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2

HL-102923 (LungGO), and RC2 HL-102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO). Funding for WHI was provided by NHLBI grants The Women's Health Initiative (WHI) was funded by the National Heart, Lung, and Blood Institute through contracts N01WH22110, N01WH24152, N01WH32100–N01WH32102, N01WH32105, N01WH32106, N01WH32108, N01WH32109, N01WH32111–N01WH32113, N01WH32115, N01WH32118, N01WH32119, N01WH32122, N01WH42107–N01WH42126, N01WH42129–N01WH42132, and N01WH44221. **RSI:** Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), NWO Rubicon Fellowship 825.12.023 ; Research Institute for Diseases in the Elderly (014-93-015; RIDE2); Netherlands Genomics Initiative/Netherlands Consortium for Healthy Aging (050-060-810); German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G. **Framingham**: Tthe project described was supported by grant numbers R01AR061162 and R0141398 from NIAMS/NIH" and is solely the responsibility of the authors and do not necessarily represent the official views of the NIAMS or NIH. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Funding for sequencing was supported by NHLBI grant 5RC2HL102419 through the American Recovery and Reinvestment Act of 2009 (ARRA) for the project, ""Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium (CHARGE-S). **D. Gauguier:** European Commission European FP7 grant EURATRANS (HEALTH-F4-2010-241504). **MrOS-USA:** The Osteoporotic Fractures in Men (MrOS) Study is supported by National Institutes of Health funding. The following institutes provide support: the National Institute on Aging (NIA), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Center for Advancing Translational Sciences (NCATS), and NIH Roadmap for Medical Research under the following grant numbers: U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128. **SOF**: The Study of Osteoporotic Fractures (SOF) is supported by National Institutes of Health funding. The National Institute on Aging (NIA) provides support under the following grant numbers: R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, and R01 AG027576. The National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) provides funding for the MrOS ancillary study 'GWAS in MrOS and SOF' under the grant number RC2ARO58973. **UFO:** The Swedish Research Council, the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg and Umeå, The Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, the Novo Nordisk Foundation, and the European Commission grant HEALTH-F2-2008-201865-GEFOS, BBMRI.se, the Swedish Society of Medicine, the Kempe-Foundation (JCK-1021), the Medical Faculty of Umeå University, the County Council of Västerbotten (Spjutspetsanslag VLL:159:33-2007). **B-Proof**: B-PROOF is supported and funded by The Netherlands Organization for Health Research and Development (ZonMw, Grant 6130.0031), the Hague; unrestricted grant from NZO (Dutch Dairy Association), Zoetermeer; NCHA (Netherlands Consortium Healthy Ageing) Leiden/ Rotterdam; Ministry of Economic Affairs, Agriculture and Innovation (project KB-15-004-003), the Hague; Wageningen University, Wageningen; VU University Medical Center, Amsterdam; Erasmus MC, Rotterdam. **TUK**: The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. TDS is holder of an ERC Advanced Principal Investigator award. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR for TwinsUK. The Australian National Health and Medical Research Council (Project Grants 1010494 and 1048216). **Study Coordinating Centre:** JBR and his team are supported by the Fonds de Recherche Santé Québec, Canadian Institutes of Health Research, the Lady Davis Institute, the Jewish General Hospital and the Canadian Foundation for Innovation. We acknowledge the assistance given by Arend Voorman for meta-analysis of region-based tests using skatMeta.

## Other Acknowledgments

**We thank all participants for their cooperation in this study. ALSPAC**: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank 23andMe for funding the generation of the ALSPAC GWAS data. We thank the UK10K Consortium for the sequencing of ALSPAC individuals. This publication is the work of the authors, and they will serve as guarantors for the contents of this paper. JT works within the MRC integrative epidemiology unit at the University of Bristol (MRC grant reference – Medical Research Council MC_II_12013/1-9) **Sloan Kettering Group**: Thanks Daniel Stephen for exceptional technical assistance with analyzing the En1 mutant phenotype. **AOGC**: AOGC Consortium: Geoffrey C Nicholson (Rural Clinical School, The University of Queensland, Toowoomba, Australia), Eugene McCloskey (Academic Unit of Bone Metabolism, Metabolic Bone Centre, University of Sheffield, Sheffield, UK& NIHR Musculoskeletal Biomedical Research Unit, Sheffield Teaching Hospitals Trust, Sheffield, UK), Richard Eastell (Academic Unit of Bone Metabolism, Metabolic Bone Centre, University of Sheffield, Sheffield, UK) Richard L Price (School of Medicine and Pharmacology, University of Western Australia, Perth, Australia & Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Perth, Australia) John A Eisman (Garvan Institute of Medical Research, University of New South Wales, Sydney, Australia) Graeme Jones (Menzies Research Institute, University of Tasmania, Hobart, Australia) Ian R Reid (Department of Medicine, University of Auckland, Auckland, New Zealand) Philip N Sambrook (Kolling Institute, Royal North Shore Hospital, University of Sydney, Sydney, Australia) Elaine M Dennison (Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southhampton, UK) John Wark (Departments of Medicine and Endocrinology, University of Melbourne, Melbourne, Australia) AOGC cohort and Brisbane Sequencing team: We acknowledge gratefully the laboratory assistance given by Jessica Harris, Graeme Clark, Sharon Song and Marina Donskoi. We would like to thank all participants who provided DNA and clinical information for the Anglo-Australasian Osteoporosis Genetics Consortium. The AOGC was funded by the Australian National Health and Medical Research Council (Australia) (grant references 511132 and 1032571). Funding was also received from the Australian Cancer Research Foundation and Rebecca Cooper Foundation (Australia). MAB is funded by a National Health and Medical Research Council (Australia) Principal Research Fellowship and ELD was previously funded by a National Health and Medical Research Council (Australia) Career Development Award (569807). IR is supported by the Health Research Council of New Zealand. The Sydney Twin Study and Tasmanian Older Adult Cohort were supported by the National Health and Medical Research Council, Australia. The help of Dr Joanna Makovey from the Sydney Twin Study is gratefully acknowledged. The OPUS study was supported by Sanofi-Aventis, Eli Lilly, Novartis, Pfizer, Proctor&Gamble Pharmaceuticals and Roche. We thank the co-principal investigators from OPUS (Dr Christian Roux, Dr David Reid, Dr Claus Glüer, and Dr Dieter Felsenber); Ms Judith Finigan and Ms Selina Simpson. The Dubbo Osteoporosis Epidemiology Study was supported by the Australian National Health and Medical Research Council, MBF Living Well foundation, the Ernst Heine Family Foundation and from untied educational grants from Amgen, Eli Lilly International, GE-Lunar, Merck Australia, Novartis, Sanofi-Aventis Australia and Servier. The contribution from A/Prof Jacqueline R Center, A/Prof Tuan V Nguyen, Mr Sing C Nguyen, Ms Denia Mang, and Ms Ruth Toppler is gratefully acknowledged. The assistance of Ms Fatma Gossiel (Sheffield) is gratefully acknowledged. The Hertfordshire Cohort Study was supported by grants from the Medical research Council UK & Arthritis Research UK. We thank the men and women who participated; Prof Cyrus Cooper; and the Herfordshire GPs who supported the study. The Oxford Osteoporosis Study was funded by Action Research, UK. The AOGC thanks the research nurses involved in this study (Ms Barbara Mason and Ms Amanda Horne (Auckland), Ms Linda Bradbury and Ms Kate Lowings (Brisbane), Ms Katherine Kolk and Ms Rumbidzai Tichawangana (Geelong); Ms Helen Steane (Hobart); Ms Jemma Christie (Melbourne); and Ms Janelle Rampellini (Perth)). The AOGC also acknowledges gratefully technical support from Ms Kathryn Addison, Ms Marieke Brugmans, Ms Catherine Cremen, Ms Johanna Hadler, and Ms Karena Pryce. **Bristol HBM Team:** We would like to thank all our study participants and the staff at our collaborating centers particularly the Wellcome Trust Clinical Research Facility in Birmingham with Dr John Ayuk, Cambridge, NIHR Biomedical Research Centre and Addenbrooke's Wellcome Trust Clinical Research Facility, NIHR Bone Biomedical Research Unit in Sheffield and the Centre for Metabolic Bone Disease in Hull with Dr Sue Steel. **RSI**: The GWA study was funded by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Lisbeth Herrera for their help in creating the GWAS database. We also thank Lisette Stolk for her support creating the fracture dataset and members of the GIANT consortium for sharing scripts used in this study. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the participants and staff from the Rotterdam Study, particularly L. Buist and J.H. van den Boogert and also the participating general practitioners and pharmacists. We would like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources. **UFO**: We thank the Bioinformatics Core Facility, at the Sahlgrenska Academy, University of Gothenburg and the Department of Biobank Research at Umeå University for assistance with data analysis and clinical data management. **B-Proof**: The authors gratefully thank all study participants, and all co-workers who helped to succeed this trial, especially P.H. in ‘t Veld, M. Hillen-Tijdink, A. Nicolaas-Merkus, N. Pliester, S. Oliai Araghi, and S. Smits. They also thank Prof. Dr. H.A.P. Pols for obtaining funding. **TUK:** The iVEC supercomputing facilities (EPIC, Murdoch University, Western Australia). We acknowledge the help and support of the twin volunteers. **SOF:** National Institute on Aging (NIA): R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, and R01 AG027576. National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS): RC2ARO58973, K01AR062655. **µCT:** Musculoskeletal imaging Core Facility at Hospital for Special Surgery supported by NIH AR046121.

## 9. References

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
2. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).
3. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).
4. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43,** 491–498 (2011).
5. Williams, F. M. K. *et al.* Genes contributing to pain sensitivity in the normal population: an exome sequencing study. *PLoS Genet.* **8,** e1003095 (2012).
6. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20,** 1297–1303
7. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26,** 2069–2070 (2010).
8. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP. *PLoS Comput Biol* **6,** e1001025 (2010).
9. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
10. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* **18,** 1851–1858 (2008).
11. Brouwer, R. W. W., van den Hout, M. C. G. N., Grosveld, F. G. & van IJcken, W. F. J. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28,** 284–285 (2012).
12. Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. **28,** 2543–2545 (2012).
13. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genetics* 1–13 (2012). doi:10.1038/ng.2249
14. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
15. Huang, J., Howie, B. & Marchini, J. Imputation Performance of ~4,000 genomes from the UK10K Project. in *American Society of Human Genetics Annual Meeting* (2013).

16. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* **44,** 955–959 (2012).

17. Xu, C. *et al.* Estimating genome-wide significance for whole-genome sequencing studies. *Genet. Epidemiol.* **38,** 281–290 (2014).

18. Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11,** 288 (2010).

19. Voorman, A., Brody, J., Lumley, T. & Voorman, M. A. R Package 'skatMeta'. at <http://cran.cc.uoc.gr/web/packages/skatMeta/skatMeta.pdf>

20. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44,** 369–375 (2012).

21. Johnson, A. D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24,** 2938–2939 (2008).

22. Chapman, J. M., Cooper, J. D., Todd, J. A. & Clayton, D. G. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum Hered* **56,** 18–31 (2003).

23. Zheng, H.-F. *et al.* WNT16 Influences Bone Mineral Density, Cortical Bone Thickness, Bone Strength, and Osteoporotic Fracture Risk. *PLoS Genet.* **8,** e1002745 (2012).

24. Project, T. E. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

25. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

26. Li, G. *et al.* Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* **148,** 84–98 (2012).

27. Kalajzic, I. *et al.* Use of Type I Collagen Green Fluorescent Protein Transgenes to Identify Subpopulations of Cells at Different Stages of the Osteoblast Lineage. *J. Bone Miner. Res.* **17,** 15–25 (2002).

28. Ben Langmead, C. T. M. P. S. L. S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10,** R25 (2009).

29. Bo Li, C. N. D. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).

30. James H Bullard, E. P. K. D. H. S. D. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94 (2010).

31. Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A. B. & Joyner, A. L. Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. *Science* **269,** 679–682 (1995).

32. Kimmel, R. A. *et al.* Two lineage boundaries coordinate vertebrate apical ectodermal ridge formation. *genesdev.cshlp.org*

33. Sgaier, S. K. *et al.* Genetic subdivision of the tectum and cerebellum into functionally related regions based on differential sensitivity to engrailed proteins. *Development* **134,** 2325–2335 (2007).

34. Hong, S.-H. *et al.* Computer-Automated Static, Dynamic and Cellular Bone Histomorphometry. *Journal of tissue science & engineering* **Suppl 1,** 004 (2014).

35. Baud, A. *et al.* Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nature Genetics* **45,** 767–775 (2013).

36. Cox, A. *et al.* A New Standard Genetic Map for the Laboratory Mouse. *Genetics* **182,** 1335–1344 (2009).

37. Ackert-Bicknell, C. L. *et al.* Mouse BMD quantitative trait loci show improved concordance with human genome-wide association loci when recalculated on a new, common mouse genetic map. *J. Bone Miner. Res.* **25,** 1808–1820 (2010).