LETTER

doi:10.1038/nature14878

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

A list of authors and affiliations appears at the end of the paper

The extent to which low-frequency (minor allele frequency (MAF) between 1–5%) and rare (MAF ≤ 1%) variants contribute to complex traits and disease in the general population is mainly unknown. Bone mineral density (BMD) is highly heritable, a major predictor of osteoporotic fractures, and has been previously associated with common genetic variants[1–8], as well as rare, population-specific, coding variants[9]. Here we identify novel non-coding genetic variants with large effects on BMD (n_{total} = 53,236) and fracture (n_{total} = 508,253) in individuals of European ancestry from the general population. Associations for BMD were derived from whole-genome sequencing (n = 2,882 from UK10K (ref. 10); a population-based genome sequencing consortium), whole-exome sequencing (n = 3,549), deep imputation of genotyped samples using a combined UK10K/1000 Genomes reference panel (n = 26,534), and *de novo* replication genotyping (n = 20,271). We identified a low-frequency non-coding variant near a novel locus, *EN1*, with an effect size fourfold larger than the mean of previously reported common variants for lumbar spine BMD[8] (rs11692564(T), MAF = 1.6%, replication effect size = +0.20 s.d., $P_{meta} = 2 \times 10^{-14}$), which was also associated with a decreased risk of fracture (odds ratio = 0.85; $P = 2 \times 10^{-11}$; n_{cases} = 98,742 and $n_{controls}$ = 409,511). Using an $En1^{cre/flox}$ mouse model, we observed that conditional loss of *En1* results in low bone mass, probably as a consequence of high bone turnover. We also identified a novel low-frequency non-coding variant with large effects on BMD near *WNT16* (rs148771817(T), MAF = 1.2%, replication effect size = +0.41 s.d., $P_{meta} = 1 \times 10^{-11}$). In general, there was an excess of association signals arising from deleterious coding and conserved non-coding variants. These findings provide evidence that low-frequency non-coding variants have large effects on BMD and fracture, thereby providing rationale for whole-genome sequencing and improved imputation reference panels to study the genetic architecture of complex traits and disease in the general population.

Recently, genetic discoveries have generally focused on common variants of small effect and rare coding variants identified through genome-wide association studies (GWAS) and whole-exome sequencing initiatives, respectively[11,12]. The effect of low-frequency and rare non-coding variants upon common diseases, and their underlying traits has been recently explored in an isolated population[13,14], but has not been well-studied to date in the general population. The UK10K project has generated a large whole-genome sequence-based resource to address this question in a general European-ancestry population[10], which is tenfold larger than the European subset of the 1000 Genomes project reference[15].

Osteoporosis, diagnosed mainly through measurement of bone mineral density (BMD), is a common systemic skeletal disease characterized by an increased propensity to fracture. The narrow-sense heritability of BMD has been estimated to be ~85%, and GWAS have successfully identified numerous loci associated with BMD which in total explain ~5% of the genetic variance for this trait[16]. However, these studies have been mainly unable to assess the role of low frequency (MAF 1–5%) and rare (MAF ≤ 1%) genetic variation, as these methods rely on testing common variants (MAF ≥ 5%). A recent sequencing-based study identified a rare nonsense variant associated with BMD using 4,931 Icelandic subjects with low BMD and 69,034 population-based controls[9]. This coding variant, which disrupts the function of *LGR4*, appears to be confined to the Icelandic population.

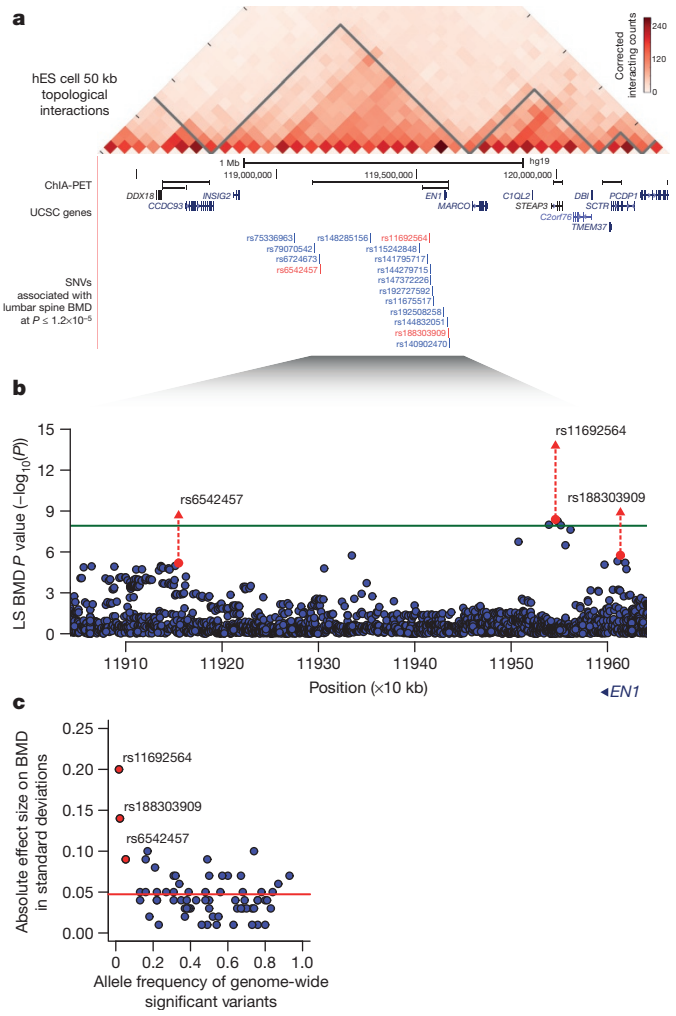


Figure 1 | Association signals near engrailed homeobox-1 for lumbar spine BMD. a, A topological domain includes associated variants and *EN1*, and chromatin interaction analysis with paired-end tag sequencing (ChIA-PET for CTCF in MCF-7 cell line) suggests a smaller interacting region containing *EN1*, and three genome-wide significant variants for lumbar spine BMD (in red). hES cell, human embryonic stem cell. **b**, Association signals at the *EN1* locus (green line at $P = 1.2 \times 10^{-8}$) for lumbar spine BMD. Red circles and triangles represent results from discovery and combined discovery and replication using fixed-effects meta-analysis (see Supplementary Information), respectively. **c**, Allele frequency versus absolute effect size for lumbar spine BMD for previously identified variants (blue)[8] and the three *EN1* novel variants (red). The red line denotes the mean of previously reported effect sizes.

Table 1 | Novel variants from single SNV association tests

			Discovery meta-analysis				Replication meta-analysis				Combined meta-analysis				
BMD phenotype	SNP	Effect allele	*N*	β	*P*	I^2	*N*	β	*P*	I^2	Freq.	*N*	β	*P*	I^2
Lumbar spine	rs11692564	T	25,225	0.24	4.1×10^{-9}	0.37	15,291	0.20	2.8×10^{-6}	0.46	0.016	40,516	0.22	1.7×10^{-14}	0.40
Lumbar spine	rs188303909	T	25,225	0.18	1.7×10^{-6}	0.36	15,228	0.14	3.3×10^{-4}	0.13	0.019	40,453	0.16	1.3×10^{-9}	0.21
Lumbar spine	rs6542457	C	25,225	0.08	6.5×10^{-6}	0.00	15,240	0.09	1.5×10^{-4}	0.00	0.058	40,465	0.09	2.2×10^{-9}	0.00
Femoral neck	rs55983207	C	29,188	0.10	2.5×10^{-7}	0.19	16,248	0.17	9.8×10^{-10}	0.03	0.042	45,436	0.12	7.2×10^{-15}	0.23
Femoral neck	rs11024028	G	29,188	0.06	2.2×10^{-8}	0.00	15,397	0.03	2.6×10^{-2}	0.30	0.198	44,585	0.05	1.3×10^{-9}	0.04
Forearm	rs148771817	T	7,848	0.47	9.3×10^{-9}	0.15	2,539	0.41	5.5×10^{-4}	-	0.012	10,387	0.46	1.1×10^{-11}	0.00

β is the additive effect of the effect allele and is measured in standard deviations of bone mineral density.

To investigate the role of rare and low-frequency genetic variation on BMD in the general population of European descent, we first undertook whole-genome sequencing in 2,882 subjects from two cohorts in the UK10K project and whole-exome sequencing in 3,549 subjects from five cohorts (Supplementary Table 1) with BMD phenotypes. We then used a novel imputation reference panel generated by the UK10K and 1000 Genomes consortia to impute variants that were missing, or poorly captured, from previous GWAS studies in 26,534 subjects (Supplementary Table 1 and Extended Data Fig. 1a). The combined UK10K and 1000 Genomes reference panel, which contained 3,781 and 379 European individuals with whole-genome sequences from UK10K and 1000 Genomes projects, respectively, enabled improved imputation, particularly of low-frequency variants, when compared to the 1000 Genomes reference panel alone[17]. We then undertook *de novo* replication genotyping of lead variants in 13 cohorts for BMD, comprising 20,271 individuals of European descent.

We meta-analyzed association results from all discovery cohorts ($n_{total} = 32{,}965$, Supplementary Table 1) for BMD measured at the forearm, femoral neck and lumbar spine, the sites where osteoporotic fractures are most prevalent. We tested bi-allelic single nucleotide variants (SNVs) with MAF $\geq 0.5\%$ for association, declaring genome-wide statistical significance at $P \leq 1.2 \times 10^{-8}$ (accounting for all independent SNVs above this MAF threshold; Supplementary Methods)[18]. The sequence kernel association test (SKAT) was used to assess association of regions containing SNVs with MAF $\leq 5\%$ and $\leq 1\%$ (Supplementary Methods). All summary-level meta-analytic results are available for unrestricted download (http://www.gefos.org). Novel genome-wide significant loci were then tested for their relationship with fracture in up to 508,253 individuals. Finally, functional genomics as well as cellular and animal models were used to investigate the relevance of these novel genetic associations to bone physiology.

Through meta-analysis of sequenced and imputed single-SNV association tests from the discovery cohorts (Supplementary Table 1), we identified a novel locus at 2q14.2 harbouring variants associated with lumbar spine BMD (lead low-frequency SNV rs11692564(T), MAF = 1.7%, effect size = +0.24 s.d., $P = 4 \times 10^{-9}$, Fig. 1 and Table 1). The direction of effect was consistent across all discovery cohorts (Extended Data Fig. 2) and the mean imputation information score for the imputed cohorts was 0.71 (Supplementary Table 2). This variant is located 53 kilobase pairs (kb) downstream from engrailed homeobox-1 (*EN1*), which, to our knowledge, has not previously been associated with any osteoporosis-related traits in humans. The rs11692564 variant was not present on HapMap imputation panels, nor on genotyping chips, underlining the importance of developing more comprehensive imputation reference panels.

To validate whole-genome sequencing genotypes at rs11692564, we genotyped 1,853 whole-genome sequenced subjects, and found all genotypes to be perfectly concordant (Supplementary Table 3). We validated imputation of rs11692564 in 3,601 imputed subjects through direct genotyping and observed that the association strengthened, and its statistical significance improved, as compared to imputed results (lumbar spine: imputed effect size = 0.22 s.d.; $P = 0.05$, genotyped effect size = 0.31 s.d.; $P = 0.004$) (Supplementary Table 4). We next sought additional evidence for the association at rs11692564 by performing additional *de novo* genotyping in 16,233 independent individuals and found a similarly large effect size in this population (effect size = +0.20 s.d.; $P = 3 \times 10^{-6}$). Meta-analysis of the discovery and replication cohorts provided strong evidence for association ($P_{combined\text{-}meta} = 2 \times 10^{-14}$) (Table 1).

We also identified an additional association signal, arising from rs55983207 (MAF = 4%), 17 kb downstream of rs11692564 ($r^2 = 0.001$) to be associated with femoral neck BMD from the combined meta-analysis ($P_{meta} = 7.2 \times 10^{-15}$, Table 1). A haplotype containing both effect alleles was not observed from within the UK10K whole-genome sequenced cohort (total number of haplotypes = 7,562).

In addition to rs11692564, we also observed two additional novel genome-wide significant variants for lumbar spine BMD near *EN1*, rs6542457 (MAF = 5.8%) and rs188303909 (MAF = 1.6%), which are 391 kb downstream and 67 kb upstream from rs11692564, respectively (Fig. 1b and Table 1). Variant rs188303909 was in moderate linkage disequilibrium (LD) with rs11692564 ($r^2 = 0.47$), and conditional analysis demonstrated that these two association signals were not independent (Supplementary Table 5). However, rs6542457 was in low LD with rs11692564 ($r^2 = 0.002$), and remained independent in conditional analyses (Supplementary Table 5). Overall, the *EN1* locus harbours multiple non-coding variants associated with lumbar spine

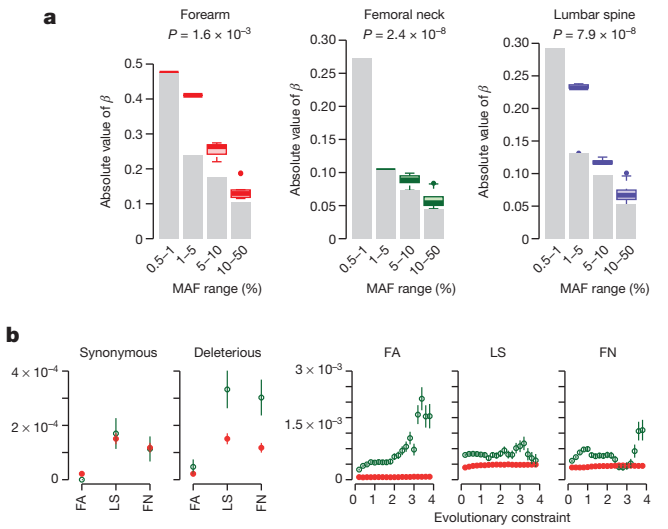


Figure 2 | Genome-wide features of association signals. **a**, Box plots of the effect sizes of genome-wide significant SNVs ($P < 1.2 \times 10^{-8}$), pruned for LD ($r^2 < 0.2$) by MAF bin for discovery cohorts. Grey bars represent the values of β not observed and for which we lack statistical power to observe (at $\alpha \leq 1.2 \times 10^{-8}$ and power ≥ 0.8). *P* values per phenotype are from the non-parametric trend test across MAF bins (see Supplementary Information). **b**, Proportion of single nucleotide variants (SNVs) passing a false discovery rate (FDR) *q*-value of 0.05 across different annotation features in discovery cohorts (green) versus matched control variants (red). The three panels on the right-hand side show enrichment across a range of evolutionary constraint scores (GERP++ score), in which green denotes SNVs above the threshold and red denotes variants below the threshold. Bars represent standard error (for Methods refer to the Supplementary Information). FA, forearm; FN, femoral neck, LS, lumbar spine.

Table 2 | Fracture meta-analysis of *EN1* variants

Locus	SNP	Effect allele	Effect allele freq.	OR (95% CI)	*P*	*N* cases	*N* controls	I^2
EN1	rs11692564	T	0.02	0.85 (0.80–0.89)	2.0×10^{-11}	98,742	409,511	0.00
	rs188303909	T	0.03	0.89 (0.85–0.93)	9.8×10^{-7}	95,669	405,697	0.00
	rs55983207	C	0.05	0.93 (0.90–0.96)	5.4×10^{-6}	97,651	407,487	0.20
	rs6542457	C	0.06	0.98 (0.95–1)	1.2×10^{-1}	95,669	405,697	0.17

and a single variant associated with femoral neck BMD. All three genome-wide significant variants for lumbar spine BMD (Table 1) co-localize solely with *EN1* in a sub-region of high interaction frequency within a single topologically associated domain[19] (Fig. 1a).

The mean effect size of previously reported genome-wide significant single nucleotide polymorphisms (SNPs) (MAF ≥ 5%) from the largest GWAS meta-analysis to date for lumbar spine and femoral BMD was 0.048 s.d. and the largest effect size was 0.1 s.d.[8]. Hence, the observed effect size at rs11692564 is fourfold larger than this mean and twice that of the largest previously reported effect (Fig. 1c)[8]. For all genome-wide significant variants, we observed larger effect sizes across decreasing MAF bins (Fig. 2a).

An increase in BMD is associated with a decrease in risk of bone fracture. We therefore tested the association of rs11692564(T) (the low-frequency allele at *EN1* associated with the largest increase in BMD) in 18 cohorts comprising 508,253 individuals (98,742 cases and 409,511 controls, Supplementary Table 6). rs11692564(T) was strongly associated with a decreased risk of fracture (odds ratio (OR) = 0.85 (95% confidence interval (CI): 0.80–0.89); $P = 2.0 \times 10^{-11}$; $I^2 = 0.00$) (Table 2 and Supplementary Table 7). Table 2 also shows clear associations between other variants near *EN1* and risk of fracture. The fracture association at rs11692564 was 2.9-fold larger than the mean of fracture associations detected in the largest GWAS to date, and 2.0-fold larger than the largest previously identified fracture association[8].

EN1 encodes a homeobox gene central to mouse limb development[20], which has been shown to be involved in Wnt signalling interaction with Dkk1 (ref. 21). Studies of calvarial bone development and fracture healing of long bones in mice have shown that perinatal $En1^{-/-}$ mutants display osteopenia and enhanced skull bone resorption[22], whereas in normal adult mice *En1* is upregulated in the bone callus post-fracture[22]. Investigating the functional role of *EN1*, we detected *En1* expression during osteoblastogenesis in developing and mature cultured murine calvarial osteoblasts, but not in marrow-derived osteoclasts, or in human primary osteoclast cultures (Fig. 3a and Extended Data Fig. 3). To determine where *En1* is active in adult bones, we analysed vertebrae from $En1^{lacZ/+}$ knock-in mice[23] and detected LacZ expression in proliferative and hypertrophic chondrocytes, osteogenic cells in the periosteum and trabecular bone surface, and in osteocytes of cortical and trabecular bone (Fig. 3b and Extended Data Fig. 4).

Using $En1^{cre/+}$; $R26^{lox\text{-}STOP\text{-}lox\text{-}EYFP}$ reporter mice to genetically tag cells for which the *En1* promoter was active at any point within a cell

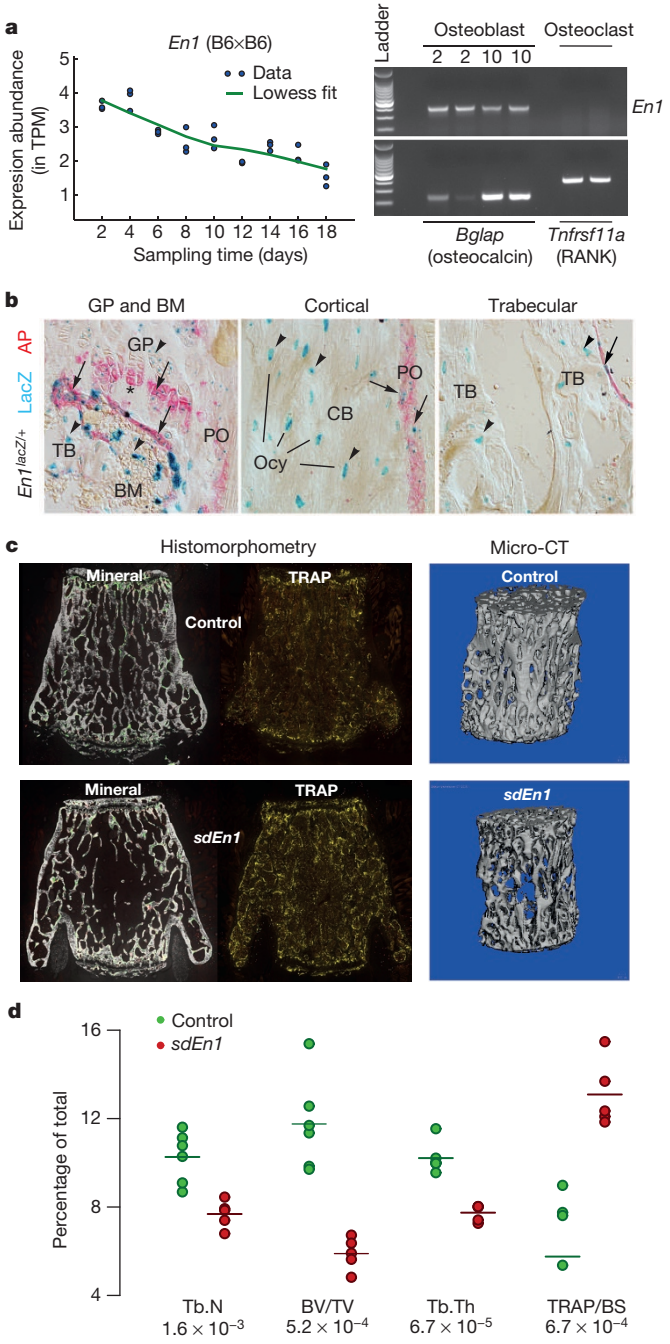


Figure 3 | Mouse *En1* functional experiments. **a**, Left, quantitative expression of *En1* and its temporal pattern (RNA-seq) in cultured calvarial murine osteoblasts ($n = 3$ per time point). Right, confirmation of the expression of *En1* in a separate RT–PCR experiment of cultured calvarial murine osteoblasts and lack of expression in osteoclasts matured from bone-marrow-derived precursor cells (positive controls for osteoblasts (osteocalcin) and osteoclast (RANK) are also shown). TPM, transcripts per million. **b**, Representative sections from lumbar vertebra 2 show the growth plate and bone marrow (GP and BM, left), cortical bone (CB, middle), and trabecular bone (TB, right) at ×40 magnification from $En1^{lacZ/+}$ adult mice ($n = 2$) stained for β-gal activity (LacZ blue, $En1^+$ cells) and alkaline phosphatase (AP, red late chondrocytes and actively calcifying tissues). In the periosteum (PO), all the $LacZ^+$ cells were AP^+; some AP^- BM cells expressed LacZ. Some AP^- proliferative chondrocytes in the GP expressed $LacZ^+$, whereas most AP^+ hypertrophic chondrocytes expressed LacZ. Some AP^- osteocytes (Ocy) in CB and TB were $LacZ^+$. **c**, Left, histomorphometry images of lumbar vertebrae 5 show decreased trabecular bone volume and increased bone surface area occupied by osteoclast cells when comparing $En1^{cre/flox}$ (self-deleted *En1*, *sdEn1*) mutants and $En1^{flox/+}$ control mice. Right, reconstructed micro-CT images show the mineral density in a control and an *sdEn1* animal. **d**, Micro-CT and histomorphometry measures within *sdEn1* ($n = 5$) and controls ($En1^{lox/+}$, $n = 6$). By micro-CT, *sdEn1* mutants exhibit decreased L5 trabecular number (Tb.N) and thickness (Tb.Th), as well as deceased bone volume fraction (BV/TV). Using histomorphometry, *sdEn1* mutants exhibit increased osteoclastic area (TRAP/BS). BS, bone surface; TRAP, tartrate acid staining. Average for each measure denoted by the solid horizontal line. For each group, *P* value between control and *sdEn1* is noted below label and was computed using paired *t*-test.

lineage, we confirmed that *En1* expression was only observed in osteogenic lineages (Extended Data Fig. 4). As most $En1^{-/-}$ animals die soon after birth, we generated $En1^{cre/flox}$ self-deleted *En1* (*sdEn1*) conditional mutants[24] ($n = 5$) and demonstrated by X-ray micro-computed tomography (micro-CT) that mutants have lower trabecular bone volume fraction (BV/TV), trabecular number, and trabecular thickness in both the lumbar L5 vertebrae (Fig. 3c, d and Extended Data Fig. 5) and the femur (Extended Data Fig. 5) as compared to littermate controls ($n = 6$). A decrease in femoral cortical thickness was also observed (Extended Data Fig. 5). By histomorphometry (Fig. 3c), we observed that the *sdEn1* mice had a statistically higher proportion of osteogenic and osteoclastic cells compared to littermate controls (Fig. 3d and Supplementary Table 8). The driving force for the low bone mass would appear to be an increase in osteoclastic activity induced by *En1* null osteogenic cells. This in turn initiates the expected coupled increase in mineralizing bone formation (Fig. 3b, d) mediated by an increased number of osteogenic cells and thus conforms to a high turnover osteoporosis-like phenotype, although dynamic histomorphometry and evidence from bone turn-over markers would be required to confirm an increased rate of bone formation (Extended Data Fig. 4). Genetic evidence from homologous regions in mice also supported a role for *En1* in bone, as the homologous region contained a quantitative trait loci (QTL) peak for femur BMD (Supplementary Table 9)[25]. These findings, together with an earlier study focusing on *En1* function in calvarial bone development[22] implicate this gene as an important mediator in skeletal biology.

Together, these findings suggest that *EN1* plays an important role in bone physiology and that low-frequency non-coding variants mapping near *EN1* have large effects on BMD and risk of fracture in the general European population.

We also identified a novel SNV at 7q31.31 within the intron of *CPED1* (rs148771817(T), MAF = 1.2%, effect size = +0.47 s.d., $P_{discovery} = 9.31 \times 10^{-9}$) associated with forearm BMD (Table 1, Supplementary Table 10 and Extended Data Fig. 6). We replicated the association at rs148771817 in 2,539 independent individuals and found a similar effect size (effect size = +0.41 s.d., $P = 6 \times 10^{-4}$), and combined meta-analysis of the discovery and replication cohorts for further improved statistical evidence for association (+0.46 s.d., $P = 1 \times 10^{-11}$) (Table 1). This variant had an effect size 2.2-fold larger than the mean of previously reported effects for common variants associated with forearm BMD (Extended Data Fig. 6)[26].

We previously identified rs7776725 to be associated with BMD at *WNT16*, a gene neighbouring *CPED1*, (Extended Data Fig. 6) and demonstrated that knockout of *Wnt16* in mice confers a 50% decrease in bone strength ($P = 7 \times 10^{-13}$)[26,27]. We have recently shown that osteoblast-derived Wnt16 represses osteoclastogenesis[28]. As a result, we undertook conditional analysis of rs148771817 upon rs7776725. The rs148771817 variant remained associated after conditioning, albeit with lower statistical significance (effect size = 0.35 s.d.; $P_{meta} = 1 \times 10^{-7}$; Extended Data Fig. 6d). Similarly, conditional analysis of the common variant upon rs148771817 revealed little change in the effect size or the statistical significance (Supplementary Table 5). Although we acknowledge that both variants may be causal, our data does not permit us to distinguish if one or both of these variants have distinct biologic effects.

While rs148771817 is intronic in *CPED1*, we found that DNA accessibility at this region, as measured by DNase I hypersensitivity data from ENCODE studies, was moderately correlated with DNA accessibility at the *WNT16* promoter in 305 cell types (maximum $r^2 = 0.4$, $P = 2.2 \times 10^{-15}$, Supplementary Table 11), whereas correlation to the promoter of *CPED1* was lower (maximum $r^2 = 0.1$, $P = 0.06$). Moreover, analysis of chromosome conformation capture Hi-C interaction frequencies from human H1 embryonic stem cells shows elevated interaction frequency between rs148771817 and *WNT16* (Extended Data Fig. 6), though we also observed stronger interactions between these loci and their immediate neighbouring regions.

We assessed whether association signals were enriched for deleterious coding SNVs or SNVs with increased evolutionary constraint (see Supplementary Methods). These two groups of SNVs were matched to control SNVs by MAF and distance to gene (Supplementary Methods and Supplementary Table 12), followed by LD pruning ($r^2 < 0.2$). We observed enrichment of association signal across the spectrum of positive evolutionary constraint thresholds, which was comparable to deleterious coding variants (Fig. 2b).

In total, we have identified multiple variants associated with BMD, including 3 genome-wide significant loci for forearm BMD, 14 for femoral neck and 19 for lumbar spine (Supplementary Tables 10, 13–15, and Extended Data Figs 7 and 8). A common variant not on previous HapMap imputation panels, near the *SOX6* gene was also identified (rs11024028, MAF = 20%) (Table 1), and was found to be an independent signal from a previously reported signal at this locus (rs7108738, $r^2 = 0.002$)[8]. Consistent with recent experiments[29,30], region-based collapsing methods did not identify any convincing novel associations that were not already identified as genome-wide significant through single SNV associations. This included collapsing variants below 1% and 5% MAF thresholds, including all variants, only variants with increased GERP++ scores or those from protein-coding regions (Supplementary Table 16 and Extended Data Figs 9 and 10).

We have identified low-frequency, non-coding genetic variants of large effect that are present in the general population and associate with BMD and fracture. These variants have effect sizes up to fourfold larger than the mean effect described for common variants associated with BMD and approximately threefold larger than those for fracture. Our study illustrates that larger reference panels, covering relevant ethnicities, will facilitate the discovery of low frequency and rare variants. This was enabled here by a large imputation reference panel (UK10K and 1000 Genomes) which offered tenfold more European samples than the 1000 Genomes reference panel available at the time of analysis (phase I version 3). Although we did not identify coding low-frequency or rare variants associated with BMD at a genome-wide significant level, we did observe that deleterious coding variants were enriched for association as a group. This suggests the existence of as yet undiscovered coding variants influencing BMD. Importantly, we have also generated new functional evidence for a central role of the homeobox protein engrailed-1 gene in regulation of BMD and identified EN1 as a critical protein in bone biology. Our findings demonstrate the utility of whole-genome sequencing-based discovery and deep imputation to enable the identification of novel genetic associations. These discoveries provide an improved understanding of the pathophysiology of osteoporosis and suggest that more comprehensive sets of whole-genome sequenced individuals, covering relevant ethnicities, will enable accurate imputation and thus facilitate discovery of low frequency and rare variants influencing complex traits and common disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 April 2014; accepted 30 June 2015.
Published online 14 September 2015.

1. Richards, J. B. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371,** 1505–1512 (2008).
2. Styrkarsdottir, U. *et al.* Multiple genetic loci for bone mineral density and fractures. *N. Engl. J. Med.* **358,** 2355–2365 (2008).
3. Styrkarsdottir, U. *et al.* New sequence variants associated with bone mineral density. *Nature Genet.* **41,** 15–17 (2009).
4. Rivadeneira, F. *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nature Genet.* **41,** 1199–1206 (2009).
5. Duncan, E. L. *et al.* Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet.* **7,** e1001372 (2011).
6. Koller, D. L. *et al.* Genome-wide association study of bone mineral density in premenopausal European-American women and replication in African-American women. *J. Clin. Endocrinol. Metab.* **95,** 1802–1809 (2010).

7. Xiong, D.-H. *et al.* Genome-wide association and follow-up replication studies identified *ADAMTS18* and *TGFBR3* as bone mass candidate genes in different ethnic groups. *Am. J. Hum. Genet.* **84,** 388–398 (2009).
8. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genet.* **44,** 491–501 (2012).
9. Styrkarsdottir, U. *et al.* Nonsense mutation in the *LGR4* gene is associated with several human diseases and other traits. *Nature* **497,** 517–520 (2013).
10. The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* http://dx.doi.org/10.1038/nature14962 (this issue).
11. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
12. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature Genet.* **44,** 623–630 (2012).
13. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nature Genet.* **47,** 435–452 (2015).
14. Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nature Genet.* **47,** 448–444 (2015).
15. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
16. Richards, J. B., Zheng, H.-F. & Spector, T. D. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nature Rev. Genet.* **13,** 576–588 (2012).
17. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature Comm.* **6,** 8111 (2015).
18. Xu, C., Tachmazidou, I., Walter, K., Ciampi, A., Zeggini, E. & Greenwood, C. M. T. Estimating genome-wide significance for whole-genome sequencing studies. *Genet. Epidemiol.* **38,** 281–290 (2014).
19. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
20. Loomis, C. A. *et al.* The mouse *Engrailed-1* gene and ventral limb patterning. *Nature* **382,** 360–363 (1996).
21. Adamska, M., MacDonald, B. T., Sarmast, Z. H., Oliver, E. R. & Meisler, M. H. *En1* and *Wnt7a* interact with *Dkk1* during limb development in the mouse. *Dev. Biol.* **272,** 134–144 (2004).
22. Deckelbaum, R. A., Majithia, A., Booker, T., Henderson, J. E. & Loomis, C. A. The homeoprotein engrailed 1 has pleiotropic functions in calvarial intramembranous bone formation and remodeling. *Development* **133,** 63–74 (2006).
23. Matise, M. P. & Joyner, A. L. Expression patterns of developmental control genes in normal and *Engrailed-1* mutant mouse spinal cord reveal early diversity in developing interneurons. *J. Neurosci.* **17,** 7805–7816 (1997).
24. 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).
25. 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).
26. Zheng, H.-F. *et al. WNT16* influences bone mineral density, cortical bone thickness, bone strength, and osteoporotic fracture risk. *PLoS Genet.* **8,** e1002745 (2012).
27. Medina-Gomez, C. *et al.* Meta-analysis of genome-wide scans for total body BMD in children and adults reveals allelic heterogeneity and age-specific effects at the *WNT16* locus. *PLoS Genet.* **8,** e1002718 (2012).
28. Movérare-Skrtic, S. *et al.* Osteoblast-derived WNT16 represses osteoclastogenesis and prevents cortical bone fragility fractures. *Nature Med.* **20,** 1279–1288 (2014).
29. Ladouceur, M., Dastani, Z., Aulchenko, Y. S., Greenwood, C. M. T. & Richards, J. B. The empirical power of rare variant association methods: results from Sanger sequencing in 1,998 individuals. *PLoS Genet.* **8,** e1002496 (2012).
30. Tang, H. *et al.* A large-scale screen for coding variants predisposing to psoriasis. *Nature Genet.* **46,** 45–50 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements Full acknowledgements are listed in the Supplementary Information.

Author Contributions Principal Investigators: A.H., A.J., A.U., A.X.-A., B.L., C.A.-B., Ch.C., C.L., Cl.C., C.L.D., C.M.v.D., C.O., D.S.E., D.Ga., D.Go., D.Gr., D.H., D.Ki., D.M., E.D., E.O., F.Ri., F.Ro., G.D.S., J.B.R., J.D., J.Re., J.Ri., J.-T.K., J.Tu., K.A., L.A.C., L.L., L.P.G.M.d.G., M.B., M.M.F., N.S., N.T., N.v.d.V., N.vS., P.R., R.D., R.D.J., R.L.P., S.G.W., S.H.R., T.H., T.P., T.S., U.P.-K., V.G., X.N., Y.-H.H. Genotyping: AOGC Consortium, A.U., B.M., B.W., C.L., C.M.v.D., C.N., C.O., C.W., D.C., D.Gr., E.D., E.O., F.Ri., G.G., G.Tr., J.Er., J.Jv.M., J.Re., J.Ri., J.-T.K., J.v.R., M.B., M.C.F., M.J., M.Z., N.A., N.G.-G., N.S., N.T., P.Ar., P.D., P.R., R.K., S.H.R., S.M., S.R., U.P.-K., X.N. and Y.-H.H. Phenotyping: A.E., A.H., A.L., AOGC Consortium, A.P.H., A.U., A.X.-A., B.M., C.G., C.K., C.L., C.L.D., C.M.v.D., C.O., D.Go., D.Ka., D.Ki., D.M., E.D., E.N., E.O., F.E.M., F.K., F.Ri., F.Ro., G.H., J.B., J.C., J.Ei., J.O., J.Re., J.Ri., J.-T.K., J.To., K.E., K.S., K.T., L.O., L.R., L.V., M.B., M.C.F., M.M.F., M.C.Z., M.K., M.Z., N.A., N.S., N.T., O.L., O.S., R.L.P., S.G.W., S.G., S.H.R., S.K., T.N., T.S. and U.P.-K. Functional experiments: A.J., A.R.-D., B.Ge, C.A.-B., C.H., C.L., C.L.D., C.O., C.U., D.Ga., D.P., E.G., H.Y.P.-M., J.D., J.F., K.Ch., Ma.M., M.H., N.S., O.S., S.B., S.C., S.-H.C., St.W., T.K., T.P., U.P.-K., W.C. and X.J. Data analysis: A.E., A.K., A.S., A.V.S., B.M., C.A.-B., Ch.C., C.-H.C., C.K., C.L., C.L.D., C.M.-G., C.M.T.G., C.O., C.T.L., C.W., D.S.E., D.M.E., D.C., D.Ka., D.M., D.P., E.D., E.G., E.N., F.G., F.Ri., G.H., G.Th., H.-F.Z., J.B.R., J.D., J.Er., J.F., J.Ha., J.Hu., J.K., J.v.R., K.Ch., K.E., K.W., L.A.C., L.H., L.M., L.O., L.R., L.V., M.B., M.C., M.H., M.K., N.A., N.S., N.T., O.L., P.Au., P.D., P.L., R.L., S.B., S.C., S.G.W., S.K., U.P., U.P.-K., V.F., W.-C.C., Y.-H.H., Y.M. and Y.Z. Meta-analysis: H.-F.Z., V.F. and Y.-H.H. Lead analysts: H.-F.Z. and V.F. Wrote first draft: J.B.R.

Author Information Source code used in preparation of results is available at https://github.com/richardslab/gefos.seq/. BMD discovery meta-analysis results are available from http://www.gefos.org. Information pertaining to UK10K can be obtained from http://www.uk10k.org. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B.R. (brent.richards@mcgill.ca).

Hou-Feng Zheng[1,2]*, Vincenzo Forgetta[1,2]*, Yi-Hsiang Hsu[3,4,5]*, Karol Estrada[4,5,6,7]*, Alberto Rosello-Diez[8]*, Paul J. Leo[9]*, Chitra L. Dahia[10,11]*, Kyung Hyun Park-Min[12]*, Jonathan H. Tobias[13,14]*, Charles Kooperberg[15]*, Aaron Kleinman[16], Unnur Styrkarsdottir[17], Ching-Ti Liu[18], Charlotta Uggla[19], Daniel S. Evans[20], Carrie M. Nielson[21,22], Klaudia Walter[23], Ulrika Pettersson-Kymmer[24,25], Shane McCarthy[23], Joel Eriksson[19,26], Tony Kwan[27], Mila Jhamai[6], Katerina Trajanoska[6,28], Yasin Memari[23], Josine Min[14], Jie Huang[23], Petr Danecek[23], Beth Wilmot[29,30], Rui Li[1,2], Wen-Chi Chou[3,4], Lauren E. Mokry[2], Alireza Moayyeri[31,32], Melina Claussnitzer[3,4,5,33], Chia-Ho Cheng[3], Warren Cheung[27,34], Carolina Medina-Gómez[6,28,35], Bing Ge[27], Shu-Huang Chen[27], Kwangbom Choi[36], Ling Oei[6,28,35], James Fraser[37], Robert Kraaij[6,28,35], Matthew A. Hibbs[36,38], Celia L. Gregson[39], Denis Paquette[37], Albert Hofman[28,35], Carl Wibom[40], Gregory J. Tranah[21,22], Mhairi Marshall[9], Brooke B. Gardiner[9], Katie Cremin[9], Paul Auer[41], Li Hsu[15], Sue Ring[42], Joyce Y. Tung[16], Gudmar Thorleifsson[43], Anke W. Enneman[6], Natasja M. van Schoor[44], Lisette C. P. G. M. de Groot[45], Nathalie van der Velde[6,46], Beatrice Melin[40], John P. Kemp[9,14], Claus Christiansen[47], Adrian Sayers[39], Yanhua Zhou[18], Sophie Calderari[48,49], Jeroen van Rooij[6,35], Chris Carlson[15], Ulrike Peters[15], Soizik Berlivet[37], Josée Dostie[37], Andre G. Uitterlinden[6,28,35], Stephen R. Williams[50], Charles Farber[50], Daniel Grinberg[51,52,53], Andrea Z. LaCroix[54], Jeff Haessler[15], Daniel I. Chasman[4,55], Franco Giulianini[55], Lynda M. Rose[55], Paul M. Ridker[4,55], John A. Eisman[56,57,58], Tuan V. Nguyen[56,58], Jacqueline R. Center[56,58], Xavier Nogues[59,60,61], Natalia Garcia-Giralt[59,60], Lenore L. Launer[62], Vilmunder Gudnason[63,64], Dan Mellström[19], Liesbeth Vandenput[19], Najaf Amin[65], Cornelia M. van Duijn[65], Magnus K. Karlsson[66], Östen Ljunggren[67], Olle Svensson[68], Göran Hallmans[25], François Rousseau[69,70], Sylvie Giroux[70], Johanne Bussière[70], Pascal P. Arp[6], Fjorda Koromani[6,28], Richard L. Prince[71,72], Joshua R. Lewis[71,72], Bente L. Langdahl[73], A. Pernille Hermann[74], Jens-Erik B. Jensen[75], Stephen Kaptoge[31], Kay-Tee Khaw[76], Jonathan Reeve[77,78], Melissa M. Formosa[79], Angela Xuereb-Anastasi[79], Kristina Åkesson[66,80], Fiona E. McGuigan[80], Gaurav Garg[80], Jose M. Olmos[81,82], Maria T. Zarrabeitia[83], Jose A. Riancho[81,82], Stuart H. Ralston[84], Nerea Alonso[84], Xi Jiang[85], David Goltzman[86], Tomi Pastinen[27,34], Elin Grundberg[27,34], Dominique Gauguier[48,49], Eric S. Orwoll[22,87], David Karasik[3,88], George Davey-Smith[14], AOGC Consortium†, Albert V. Smith[63,64], Kristin Siggeirsdottir[63], Tamara B. Harris[89], M. Carola Zillikens[6], Joyce B. J. van Meurs[6,28], Unnur Thorsteinsdottir[17,64], Matthew T. Maurano[90], Nicholas J. Timpson[14], Nicole Soranzo[23], Richard Durbin[23], Scott G. Wilson[32,71,91], Evangelia E. Ntzani[92,93], Matthew A. Brown[9], Kari Stefansson[64,94], David A. Hinds[16], Tim Spector[32], L. Adrienne Cupples[18,95], Claes Ohlsson[19], Celia M. T. Greenwood[2,34,96,97], UK10K Consortium†, Rebecca D. Jackson[98]§, David W. Rowe[85]§, Cynthia A. Loomis[99]§, David M. Evans[9,14]§, Cheryl L. Ackert-Bicknell[36]§, Alexandra L. Joyner[8]§, Emma L. Duncan[9,100]§, Douglas P. Kiel[3,4,5,33]§, Fernando Rivadeneira[6,28,35]§ & J. Brent Richards[1,2,32]§

[1]Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Montréal H3A 1A2, Canada. [2]Department of Medicine, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montréal H3T 1E2, Canada. [3]Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts 02131, USA. [4]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Broad Institute of MIT and Harvard, Boston, Massachusetts 02115, USA. [6]Department of Internal Medicine, Erasmus Medical Center, Rotterdam 3015GE, The Netherlands. [7]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [8]Developmental Biology Program, Sloan Kettering Institute, New York, New York 10065, USA. [9]The University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, Brisbane 4102, Australia. [10]Department of Cell and Developmental Biology, Weill Cornell Medical College, New York, New York 10065, USA. [11]Tissue Engineering, Regeneration and Repair Program, Hospital for Special Surgery, New York 10021, USA. [12]Rheumatology Divison, Hospital for Special Surgery New York, New York 10021, USA. [13]School of Clinical Science, University of Bristol, Bristol BS10 5NB, UK. [14]MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK. [15]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [16]Department of Research, 23andMe, Mountain View, California 94041, USA. [17]Department of Population Genomics, deCODE Genetics, Reykjavik IS-101, Iceland. [18]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [19]Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg S-413 45, Sweden. [20]California Pacific Medical Center Research Institute, San Francisco, California 94158, USA. [21]Department of Public Health and Preventive Medicine, Oregon Health & Science University, Portland, Oregon 97239, USA. [22]Bone & Mineral Unit, Oregon Health & Science University, Portland, Oregon 97239, USA. [23]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK. [24]Departments of Pharmacology and Clinical Neurosciences, Umeå University, Umeå S-901 87, Sweden. [25]Department of Public Health and Clinical Medicine, Umeå University, Umeå SE-901 87, Sweden. [26]Centre for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg S-413 45, Sweden. [27]McGill University and Genome Quebec Innovation Centre, Montréal H3A 0G1, Canada. [28]Department of Epidemiology, Erasmus Medical Center, Rotterdam 3015GE, The Netherlands. [29]Oregon Clinical and Translational Research Institute, Oregon Health & Science University, Portland, Oregon

97239, USA. [30]Department of Medical and Clinical Informatics, Oregon Health & Science University, Portland, Oregon 97239, USA. [31]Farr Institute of Health Informatics Research, University College London, London NW1 2DA, UK. [32]Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK. [33]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA. [34]Department of Human Genetics, McGill University, Montréal H3A 1B1, Canada. [35]Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden 2300RC, The Netherlands. [36]Center for Musculoskeletal Research, University of Rochester, Rochester, New York 14642, USA. [37]Department of Biochemistry and Goodman Cancer Research Center, McGill University, Montréal H3G 1Y6, Canada. [38]Department of Computer Science, Trinity University, San Antonio, Texas 78212, USA. [39]Musculoskeletal Research Unit, University of Bristol, Bristol BS10 5NB, UK. [40]Department of Radiation Sciences, Umeå University, Umeå S-901 87, Sweden. [41]School of Public Health, University of Wisconsin, Milwaukee, Wisconsin 53726, USA. [42]School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK. [43]Department of Statistics, deCODE Genetics, Reykjavik IS-101, Iceland. [44]Department of Epidemiology and Biostatistics and the EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam 1007 MB, The Netherlands. [45]Department of Human Nutrition, Wageningen University, Wageningen 6700 EV, The Netherlands. [46]Department of Internal Medicine, Section Geriatrics, Academic Medical Center, Amsterdam 1105, The Netherlands. [47]Nordic Bioscience, Herlev 2730, Denmark. [48]Cordeliers Research Centre, INSERM UMRS 1138, Paris 75006, France. [49]Institute of Cardiometabolism and Nutrition, University Pierre & Marie Curie, Paris 75013, France. [50]Departments of Medicine (Cardiovascular Medicine), Centre for Public Health Genomics, University of Virginia, Charlottesville, Virginia 22908, USA. [51]Department of Genetics, University of Barcelona, Barcelona 08028, Spain. [52]U-720, Centre for Biomedical Network Research on Rare Diseases (CIBERER), Barcelona 28029, Spain. [53]Department of Human Molecular Genetics, The Institute of Biomedicine of the University of Barcelona (IBUB), Barcelona 08028, Spain. [54]Women's Health Center of Excellence Family Medicine and Public Health, University of California – San Diego, San Diego, California 92093, USA. [55]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. [56]Osteoporosis & Bone Biology Program, Garvan Institute of Medical Research, Sydney 2010, Australia. [57]School of Medicine Sydney, University of Notre Dame Australia, Sydney 6959, Australia. [58]St. Vincent's Hospital & Clinical School, NSW University, Sydney 2010, Australia. [59]Musculoskeletal Research Group, Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona 08003, Spain. [60]Cooperative Research Network on Aging and Fragility (RETICEF), Institute of Health Carlos III, 28029, Spain. [61]Department of Internal Medicine, Hospital del Mar, Universitat Autònoma de Barcelona, Barcelona 08193, Spain. [62]Neuroepidemiology Section, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. [63]Icelandic Heart Association, Kopavogur IS-201, Iceland. [64]Faculty of Medicine, University of Iceland, Reykjavik IS-101, Iceland. [65]Genetic epidemiology unit, Department of Epidemiology, Erasmus MC, Rotterdam 3000CA, The Netherlands. [66]Department of Orthopaedics, Skåne University Hospital Malmö 205 02, Sweden. [67]Department of Medical Sciences, University of Uppsala, Uppsala 751 85, Sweden. [68]Department of Surgical and Perioperative Sciences, Umeå Unviersity, Umeå 901 85, Sweden. [69]Department of Molecular Biology, Medical Biochemistry and Pathology, Université Laval, Québec City G1V 0A6, Canada. [70]Axe Santé des Populations et Pratiques Optimales en Santé, Centre de recherche du CHU de Québec, Québec City G1V 4G2, Canada. [71]Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands 6009, Australia. [72]Department of Medicine, University of Western Australia, Perth 6009, Australia. [73]Department of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus C 8000, Denmark. [74]Department of Endocrinology, Odense University Hospital, Odense C 5000, Denmark. [75]Department of Endocrinology, Hvidovre University Hospital, Hvidovre 2650, Denmark. [76]Clinical Gerontology Unit, University of Cambridge, Cambridge CB2 2QQ, UK. [77]Medicine and Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [78]Institute of Musculoskeletal Sciences, The Botnar Research Centre, University of Oxford, Oxford OX3 7LD, UK. [79]Department of Applied Biomedical Science, Faculty of Health Sciences, University of Malta, Msida MSD 2080, Malta. [80]Clinical and Molecular Osteoporosis Research Unit, Department of Clinical Sciences Malmö, Lund University, 205 02, Sweden. [81]Department of Medicine and Psychiatry, University of Cantabria, Santander 39011, Spain. [82]Department of Internal Medicine, Hospital U.M. Valdecilla- IDIVAL, Santander 39008, Spain. [83]Department of Legal Medicine, University of Cantabria, Santander 39011, Spain. [84]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, Western General Hospital, University of Edinburgh, Edinburgh EH4 2XU, UK. [85]Department of Reconstructive Sciences, College of Dental Medicine, University of Connecticut Health Center, Farmington, Connecticut 06030, USA. [86]Department of Medicine and Physiology, McGill University, Montréal H4A 3J1, Canada. [87]Department of Medicine, Oregon Health & Science University, Portland, Oregon 97239, USA. [88]Faculty of Medicine in the Galilee, Bar-Ilan University, Safed 13010, Israel. [89]Laboratory of Epidemiology, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. [90]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [91]School of Medicine and Pharmacology, University of Western Australia, Crawley 6009, Australia. [92]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece. [93]Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, Rhode Island 02903, USA. [94]deCODE Genetics, Reykjavik IS-101, Iceland. [95]Framingham Heart Study, Framingham, Massachusetts 01702, USA. [96]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal H3A 1A2, Canada. [97]Department of Oncology, Gerald Bronfman Centre, McGill University, Montréal H2W 1S6, Canada. [98]Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, The Ohio State University, Columbus, Ohio 43210, USA. [99]The Ronald O. Perelman Department of Dermatology and Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA. [100]Department of Diabetes and Endocrinology, Royal Brisbane and Women's Hospital, Brisbane 4029, Australia.
*These authors contributed equally to this work.
§These authors jointly supervised this work.
†Lists of participants and their affiliations appear in the Supplementary Information.

METHODS

More details for the Methods are in the Supplementary Information. All human studies were approved by their institutional ethics review committees, and all participants provided written informed consent.

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment, except the teams undertaking micro-CT and histomorphometry experiments were blinded to each other's results.

Whole-genome sequencing. ALSPAC and TwinsUK cohorts were sequenced at an average read depth of 6.7× through the UK10K program (http://www.UK10K.org) using the Illumina HiSeq platform, and aligned to the GRCh37 human reference using BWA[31]. SNV calls were completed using samtools/bcftools and VQSR and GATK were used to recall these calls.

Whole-exome sequencing. The AOGC, FHS, RS-I, ESP and ERF cohorts were whole-exome sequenced as described in the Supplementary Information.

Whole-genome genotyping. All remaining discovery cohorts were genome-wide genotyped and imputed to the UK10K/1000 Genomes reference panel, as described in the Supplementary Information.

Association testing for BMD. Single variants with a MAF $>0.5\%$ were tested for an additive effect on lumbar spine, femoral neck and forearm BMD, adjusting for sex, age, age^2, weight and standardized to have a mean of zero and a standard deviation of one. Meta-analysis of cohort-level summary statistics was undertaken using GWAMA[32]. Conditional analyses for significant SNVs was performed using GCTA[33]. Region-based collapsing tests were performed using skatMeta[34], an implementation of the SKAT method[35] that enables the meta-analysis of multiple cohorts. For each cohort, variants with MAF $\leq 5\%$ or $\leq 1\%$ were collected and meta-analysis using skatMeta was conducted for windows of 30 SNVs within each region, overlapping by 10 SNVs.

Replication genotyping. Lead SNVs were selected for replication genotyping, which was performed at LGC Genomics, Erasmus MC and deCODE Genetics using KASP genotyping. Association testing for replication genotyping was undertaken using the same additive model, using the same covariates for BMD, as above.

Fracture association testing. Fractures were defined as those occurring at any site, except fingers, toes and skull, after age 18. Both incident and prevalent fractures were included and were verified by either radiographic, casting, physician, or subject reporting. Fractures resulting from any type of trauma were considered. Covariates included in the additive model were age, age^2, sex, height, weight, oestogen/menopause status (when available), ancestral genetic background and cohort-specific covariates (such as clinical centre). Association testing was done in two phases. The first involved all 1,482 genome-wide significant SNVs for BMD. In the second phase of fracture association testing, variants at *EN1* were assessed in 18 cohorts, comprising 98,467 cases and 409,736 controls. Meta-analysis of cohort-level summary statistics was performed using GWAMA[32].

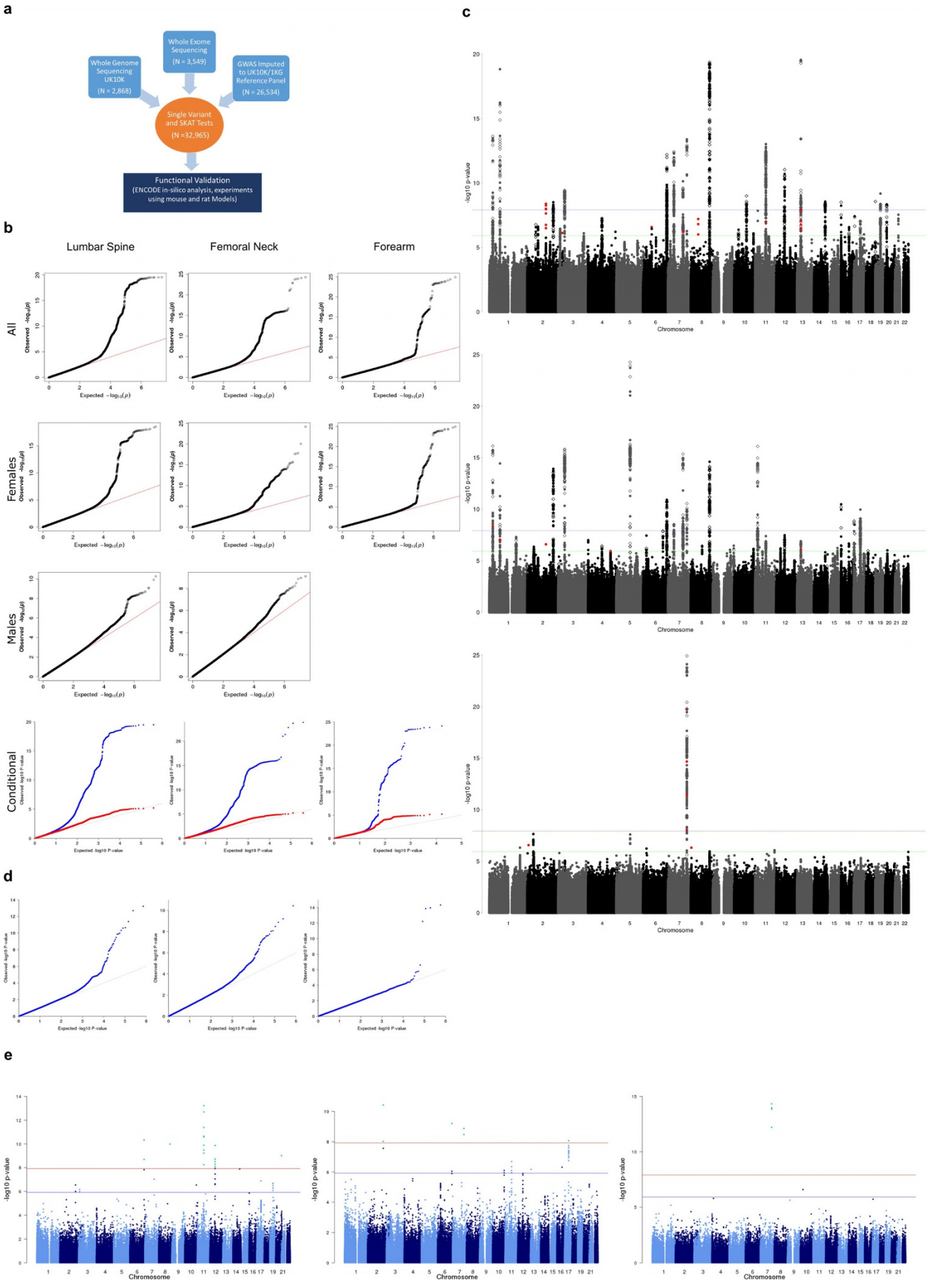
Functional genomics. We tested whether variants with increasing GERP++ scores[36] were more strongly associated with BMD than SNVs matched for distance to gene and MAF, after LD pruning using PLINK[37] at an r^2 of <0.2, using windows of 100 kb and a step of 20 kb. Coding variants were partitioned as deleterious using Variant Effect Predictor[38] LD pruned ($r^2 < 0.2$). The proportion of variants passing an FDR q-value of ≤ 0.05 were reported.

***En1* murine expression experiments.** Pre-osteoblast-like cell were differentiated to osteoblasts from calvaria of C57BL/6J mice and expression levels of each gene was quantified using RNA-seq. The temporal expression of *En1* in cell culture experiments of these osteoblasts and bone-marrow-derived osteoclasts (isolated from long bones of six-week-old mice) was measured by PCR, with *Bglap* (osteocalcin) and *Tnfrsf11a* (RANK), serving as controls. Total mRNA for *En1* in osteoblasts was quantified using real-time PCR.

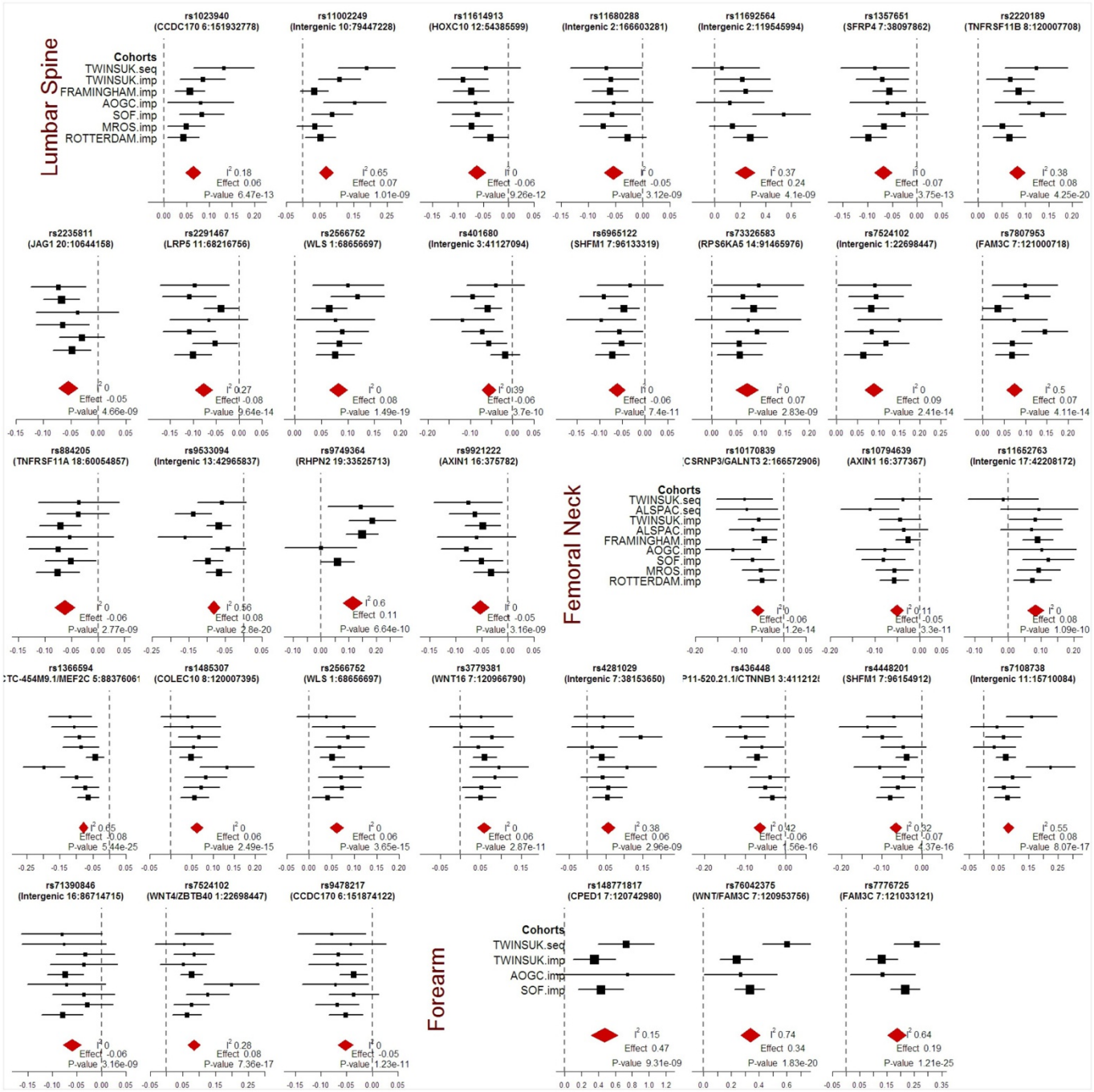
Micro-CT and histomorphometry. Mouse husbandry and all experiments were performed in accordance with Memorial Sloan-Kettering Cancer Center Institutional Animal Care and Use Committee-approved protocols. Bone characteristics of self-deleted conditional *En1*(*sdEn1*) mutants were compared to $En1^{+/flox}$ littermates using micro-CT. The same animals were assessed for histomorphometry (and laboratories performing micro-CT and histomorphometry were blinded to each other's results). After tissue sectioning, samples were stained for calcification (calcein blue), tartrate acid (TRAP) to assess for osteoclasts and alkaline phosphatase to assess for osteoblasts.

Murine histology. Two-month-old $En1^{lacZ/+}$ mice[39] were sectioned at bone sites and stained for X-gal and/or alkaline phosphatase and imaged at ×400.

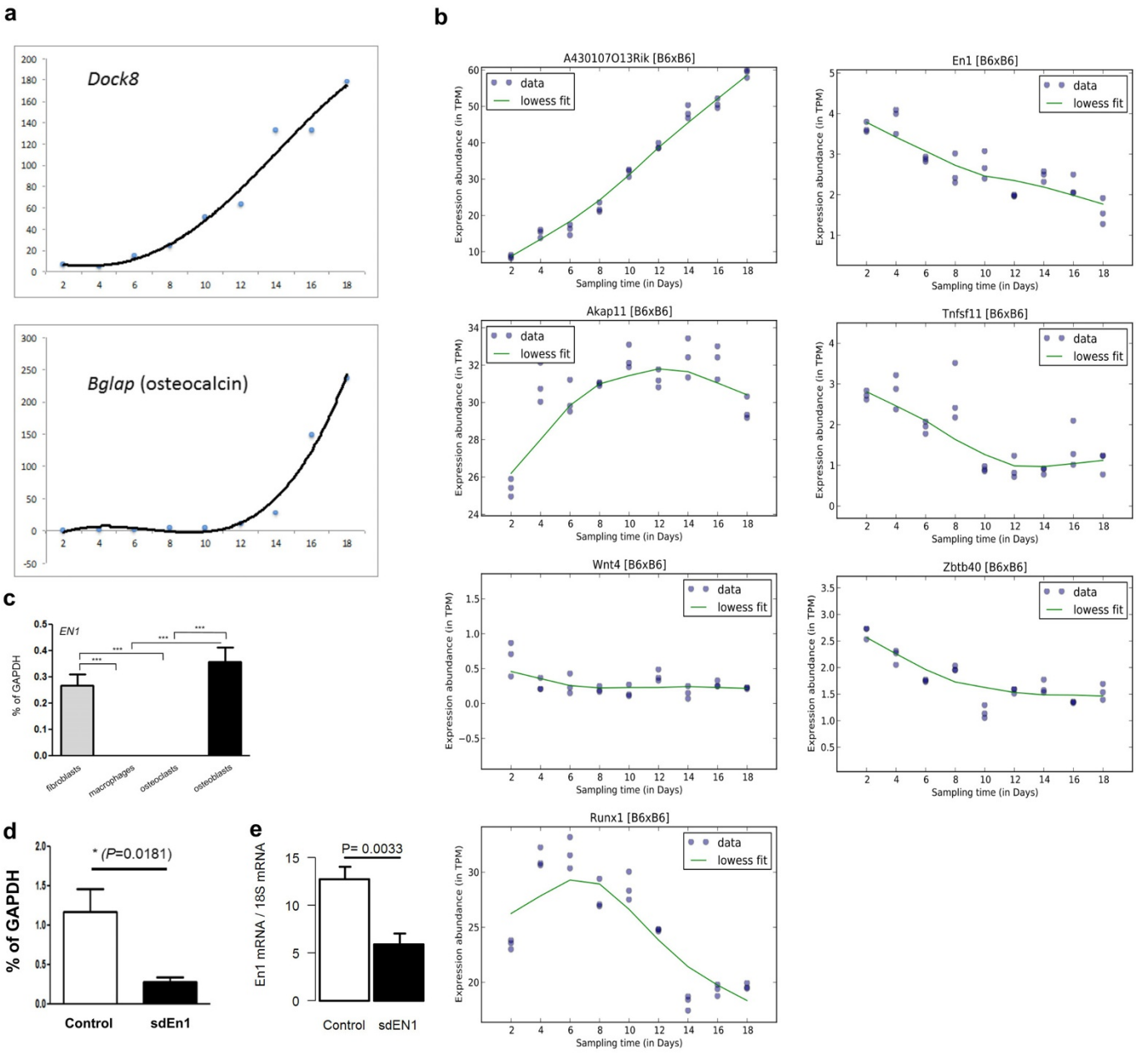
31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
32. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11,** 288 (2010).
33. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genet.* **44,** 369–375 (2012).
34. Voorman, A. A., Brody, J. & Lumley, T. SkatMeta: an R package for meta-analyzing region-based tests of rare DNA variants (http://cran.r-project.org/web/packages/skatMeta) (2013).
35. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89,** 82–93 (2011).
36. 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).
37. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).
38. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* **26,** 2069–2070 (2010).
39. 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).



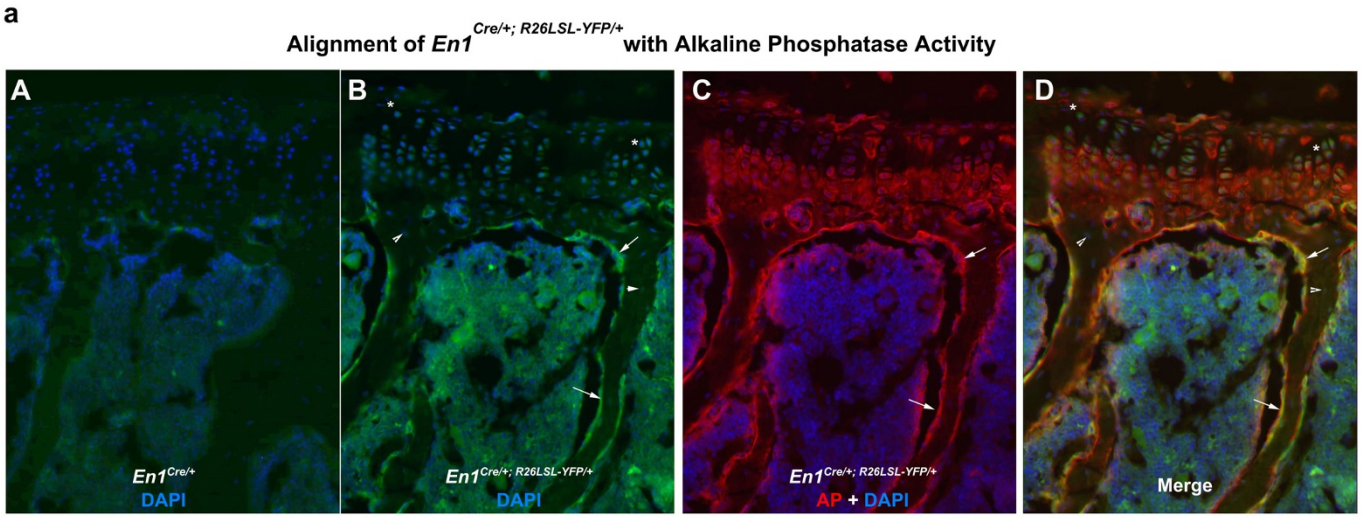
Extended Data Figure 1 | Discovery single variant meta-analysis. **a**, Overall study design. **b**, From top to bottom, quantile–quantile plots for the sex-combined single SNV meta-analysis, sex-stratified single SNV meta-analysis (forearm phenotype consists solely of female-only cohorts), and sex-combined single SNV conditional meta-analysis Plots depicts *P* values prior (blue) and after (red) conditional analysis on genome-wide significant variants (see Supplementary Methods). **c**, From top to bottom, Manhattan plots for sex-combined meta-analysis for lumbar spine BMD, femoral neck BMD, and forearm BMD. Each plot depicts variants from the UK10K/1000G reference panel with MAF $> 0.5\%$ across the 22 autosomes (odd, grey; even, black) against the $-\log_{10} P$ value from the meta-analysis of 7 cohorts (dots). Also depicted are the subset variants from the reference panel that are also present in ref. 8 with *P* value $<5 \times 10^{-6}$ (diamonds). Variants with MAF $< 5\%$ and $P < 1.2 \times 10^{-6}$ are also depicted (red). **d**, Quantile–quantile plots for the sex-combined meta-analysis of lumbar spine, femoral neck, and forearm BMD for SNVs present across both exome-sequenced and genome-sequenced and imputed cohorts, that is, SNV present only in genome-sequenced or imputed cohorts are not shown. **e**, Manhattan plot for the meta-analysis of sex-combined results for lumbar spine BMD for SNVs present in exome-sequenced and genome-sequenced and imputed cohorts, that is, SNV present only in genome-sequenced or imputed cohorts are not shown (from left to right: lumbar spine, forearm and femoral neck BMD).

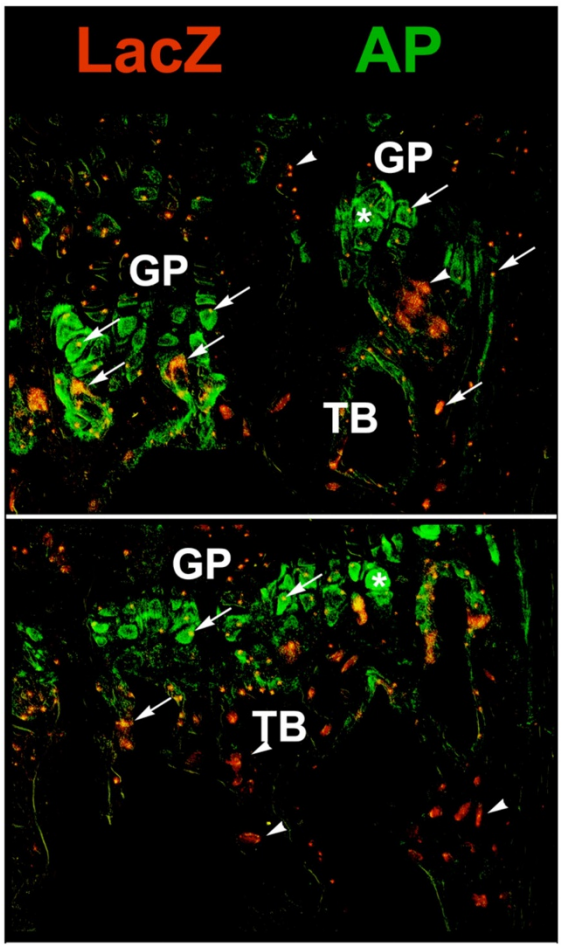


Extended Data Figure 2 | Forest plots by cohort for genome-wide significant loci from discovery meta-analysis. Forest plots for three BMD phenotypes are shown. Title of each plot includes gene overlapping the SNV and its genomic position on build hg19. *P* values are from fixed-effect meta-analysis (see Supplementary Information).

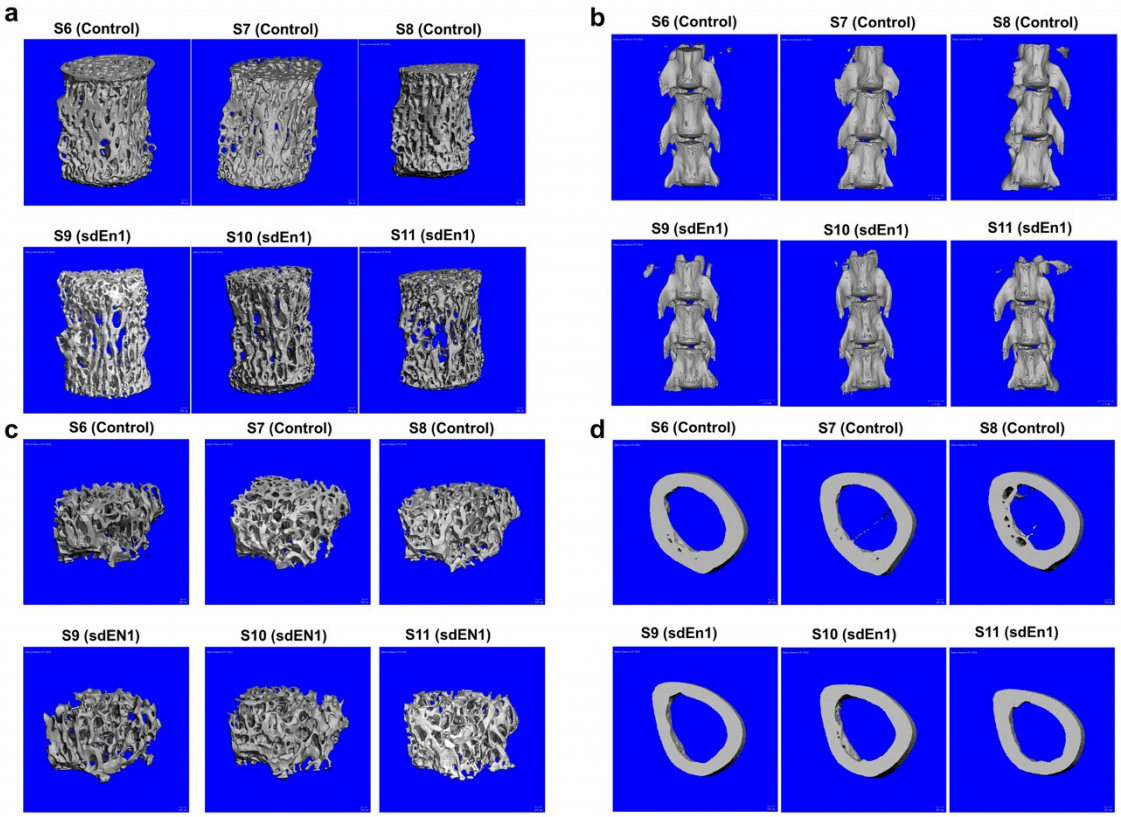


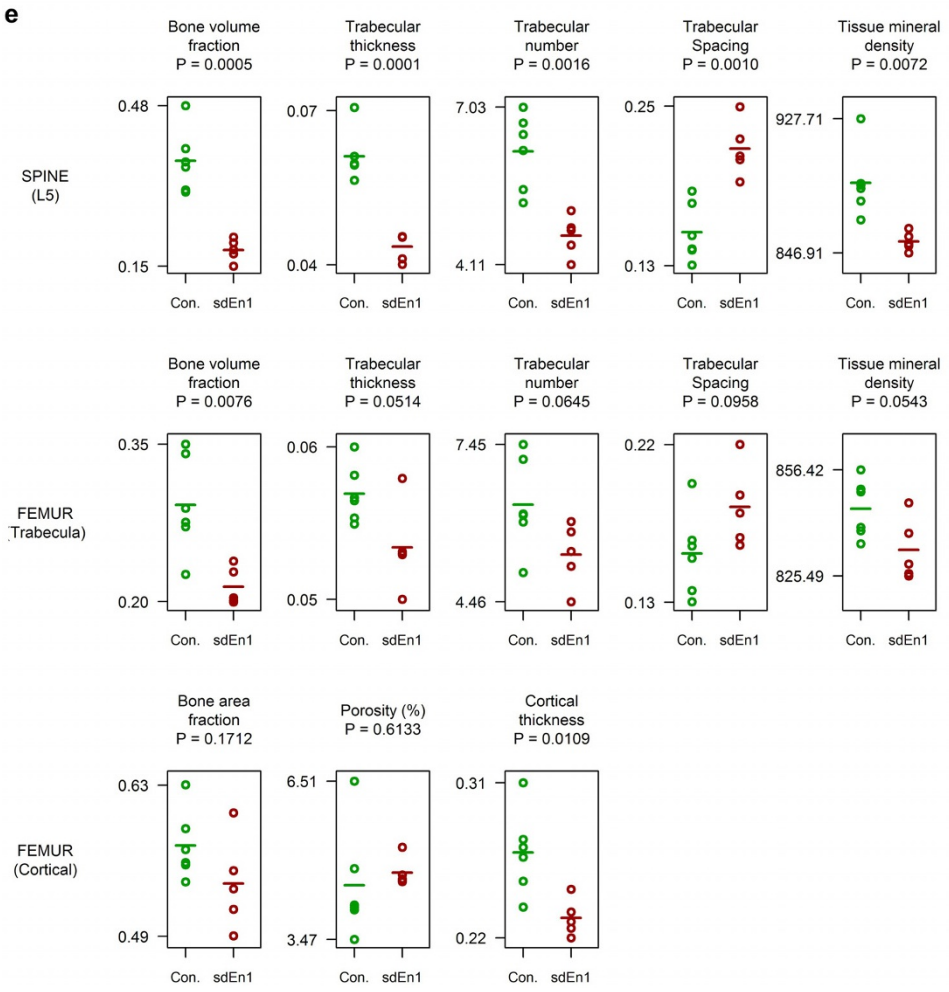
Extended Data Figure 3 | Gene expression in human and mouse. **a**, Quantification of *Dock8* expression and its temporal pattern through RNA-seq in cultured calvarial murine osteoblasts across day 2 through to day 18 of osteoblast development. Shown for comparison is *Bglap*, which encodes osteocalcin, a critical protein in osteoblasts. **b**, Quantification of expression of genome-wide significant genes and their temporal pattern through RNA-seq in cultured calvarial murine osteoblasts across day 2 through to day 18 of osteoblast development. **c**, Expression of *EN1* mRNA in human cells presented as per cent of *GAPDH* mRNA. **d**, Expression of *En1* in control and *sdEn1* mice in purified osteoblast culture. For osteoblast marker gene expression, total mRNAs were purified from osteoblast cultures at day 10 and measured using quantitative real-time PCR. mRNA levels were normalized relative to *GAPDH* mRNA. **e**. Real-time PCR expression of control and *sdEn1* as compared to 18S mRNA in whole vertebral bone extract. All data are shown as mean ± s.e.m. Significance computed by Student's unpaired *t*-test.



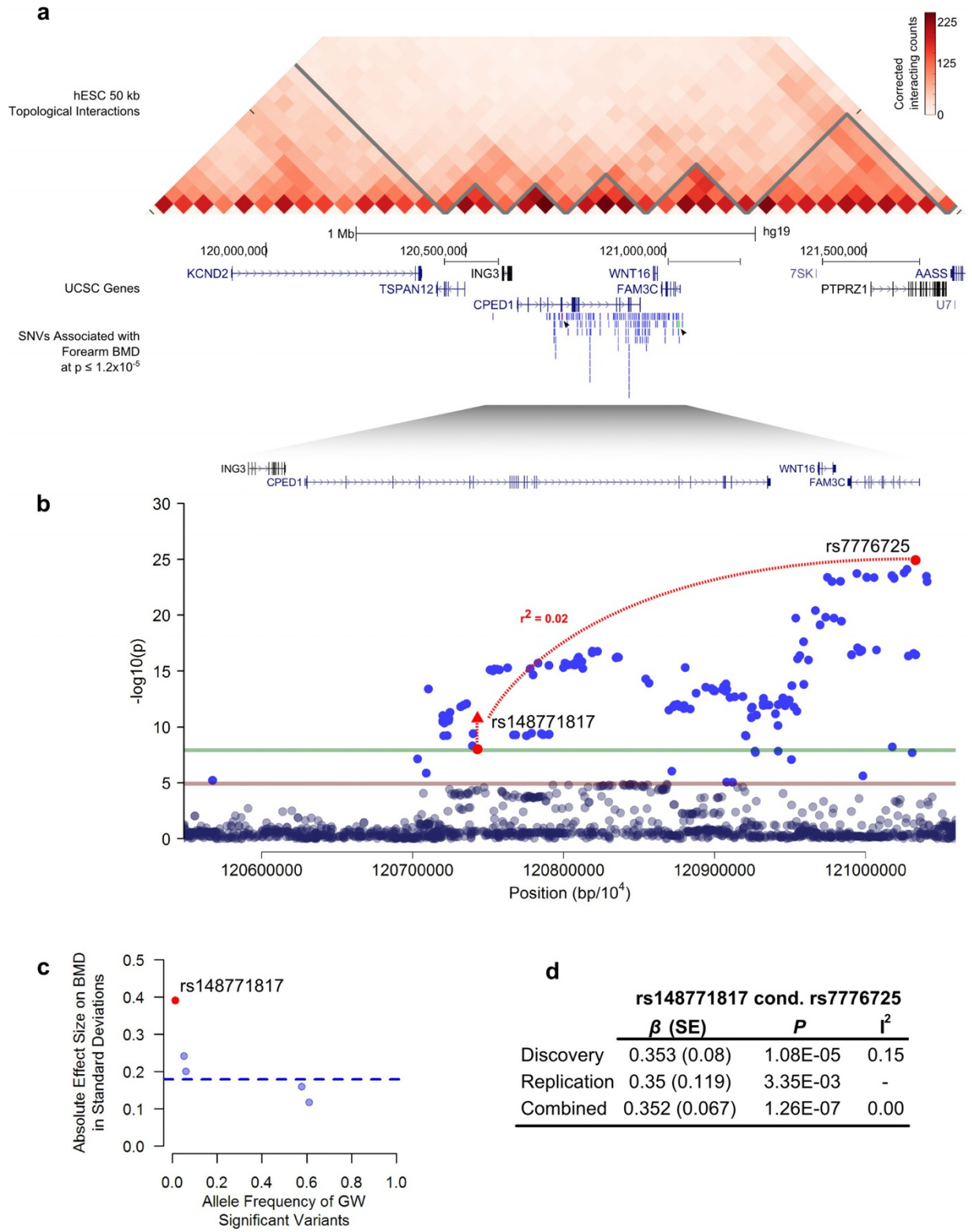


Extended Data Figure 4 | Histological assessment of *En1*^{cre}-expressing cells in skeletal cells of the vertebra. **a**, Lineage history of *En1*^{cre}-expressing cells in skeletal cells of the vertebra. The *En1*^{cre} allele was combined with the *R26*$^{LSL\text{-}YFP}$ reporter allele and examined using frozen fluorescent immunohistochemistry and alkaline phosphatase (AP) staining. Cell nuclei were detected with DAPI. YFP-expressing cells have expressed Cre (*En1*) at some time in their history. In subpanel A, control animals lacking the *R26*$^{LSL\text{-}YFP}$ reporter show low background YFP signal (green). In subpanel B, *En1*$^{cre/+}$; *R26*$^{LSL\text{-}YFP/+}$ mice YFP-expressing cells are detected in the growth plate chondrocytes of the vertebra (asterisk), trabecular bone lining cells (arrow) and osteocytes (arrowhead). Note, high fluorescent background staining in the marrow space. In subpanel C, the same section is shown stained for AP activity using the Fast Red substrate. Strong activity is present in the hypertrophic chondrocytes of the growth plate and trabecular bone lining cells (arrow). In subpanel D, alignment of the AP and YFP images shows that the trabecular lining cells co-express AP and YFP. **b**, Co-localization of *En1* and alkaline phosphatase expression. Images of lumbar vertebrae sections (growth plate and trabecular bone regions, ×40 magnification) from two-month old *En1*$^{lacZ/+}$ mice (see Fig. 3b), stained for LacZ and alkaline phosphatase (AP), false-coloured as indicated. Double-positive cells are indicated by arrows, single-positive cells are indicated by arrowheads (LacZ^{+}) or asterisks (AP^{+}). Except for some chondrocytes, most AP^{+} cells are also LacZ^{+}, that is, express *En1*. The bone marrow was digitally removed, as it contains no AP^{+} cells.



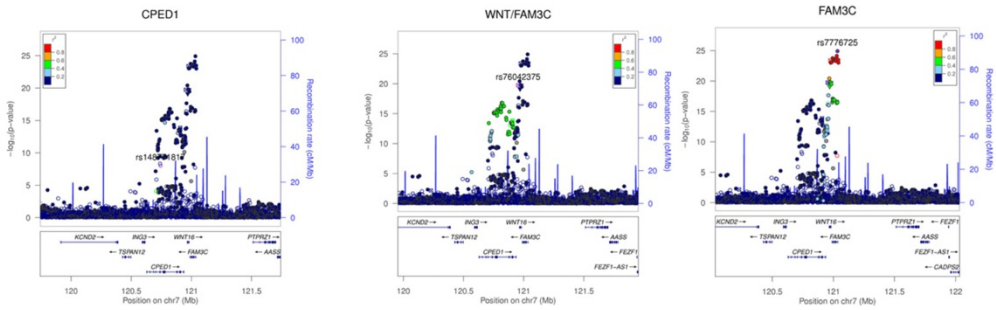


Extended Data Figure 5 | Micro-CT results for control (*En1*$^{flox/+}$) and self-deleting *En1* knockout (*sdEn1*, *En1*$^{cre/flox}$) animals. **a**, Trabecular bone micro-CT images from lumbar vertebra 5. **b**, Morphological characteristics at lumbar vertebra 4, 5, and 6 (from bottom to top). **c**, **d**, Morphological characteristics of left femur trabecular bone (**c**) and left femur cortical bone (**d**). **e**, Micro-CT parameter results for the comparison of control and *sdEn1* animals at lumbar vertebra 5, femur trabecula, and femur cortical bone. Horizontal lines denote mean of observations. Significance between control and *sdEn1* is calculated using an unpaired *t*-test.

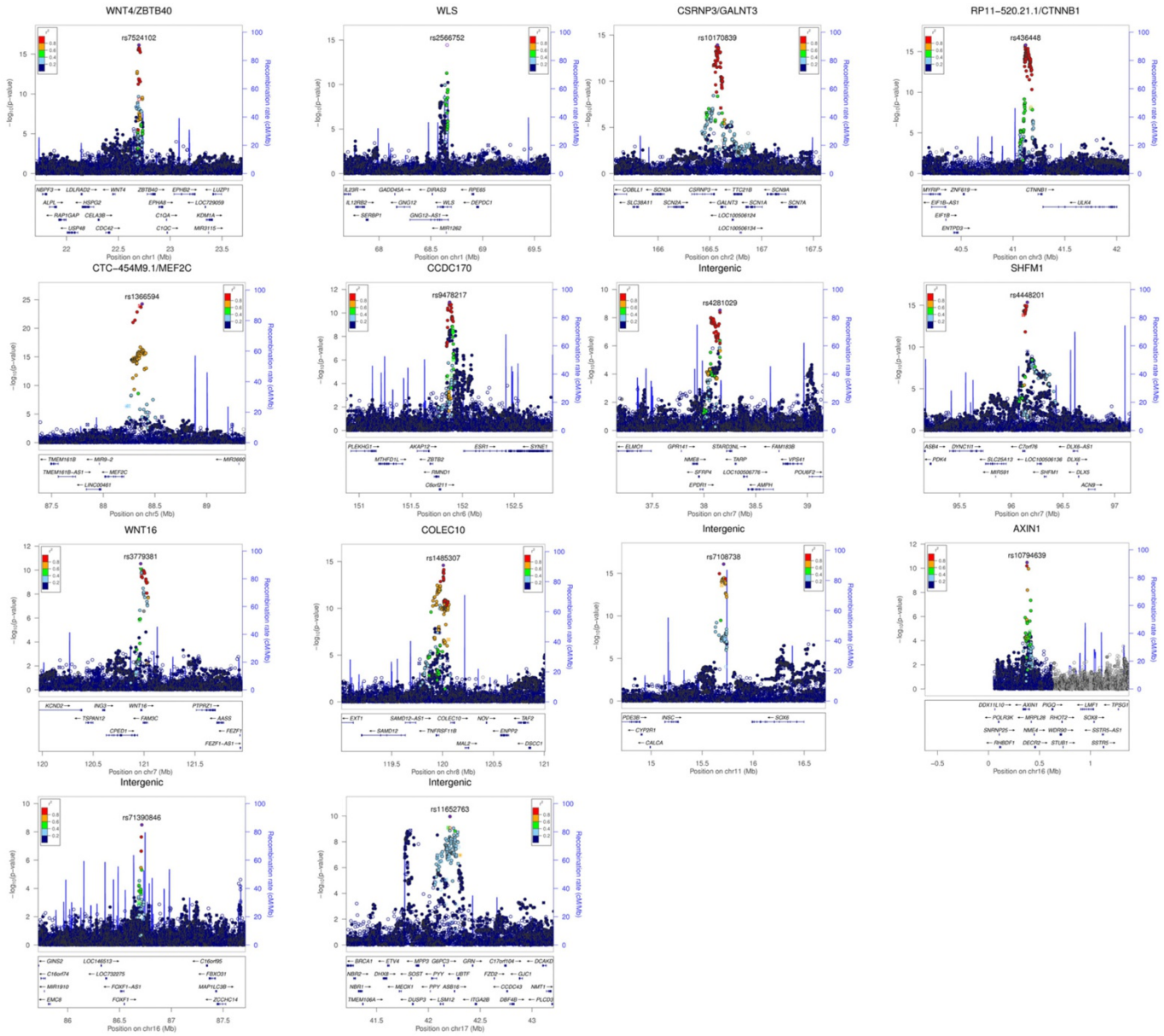


Extended Data Figure 6 | Novel association from 7q31.3. **a**, Chromatin interaction data from Hi-C performed in H1 embryonic stem cells[23] of a 2 Mb region encompassing rs148771817 (red and identified by arrow) and *WNT16*. **b**, The left axis denotes the association *P* value (red and green lines at $P = 1.2 \times 10^{-5}$ and 1.2×10^{-8}, respectively). The novel genome-wide significant SNV, rs148771817, within an intron of *CPED1*, and the lead genome-wide significant SNV rs7776725 upstream to *WNT16* (within *FAM3C*) are in low LD with each other. **c**, Allele frequency versus absolute effect size (in standard deviations) for forearm BMD of all previously identified genome-wide significant variants (blue)[8] and the novel variant within *CPED1* (red), rs148771817 from replication meta-analysis. The blue line denotes the mean of effect sizes for previously reported forearm BMD variants. **d**, Meta-analysis summary statistics of rs148771817 conditioned on rs7776725.

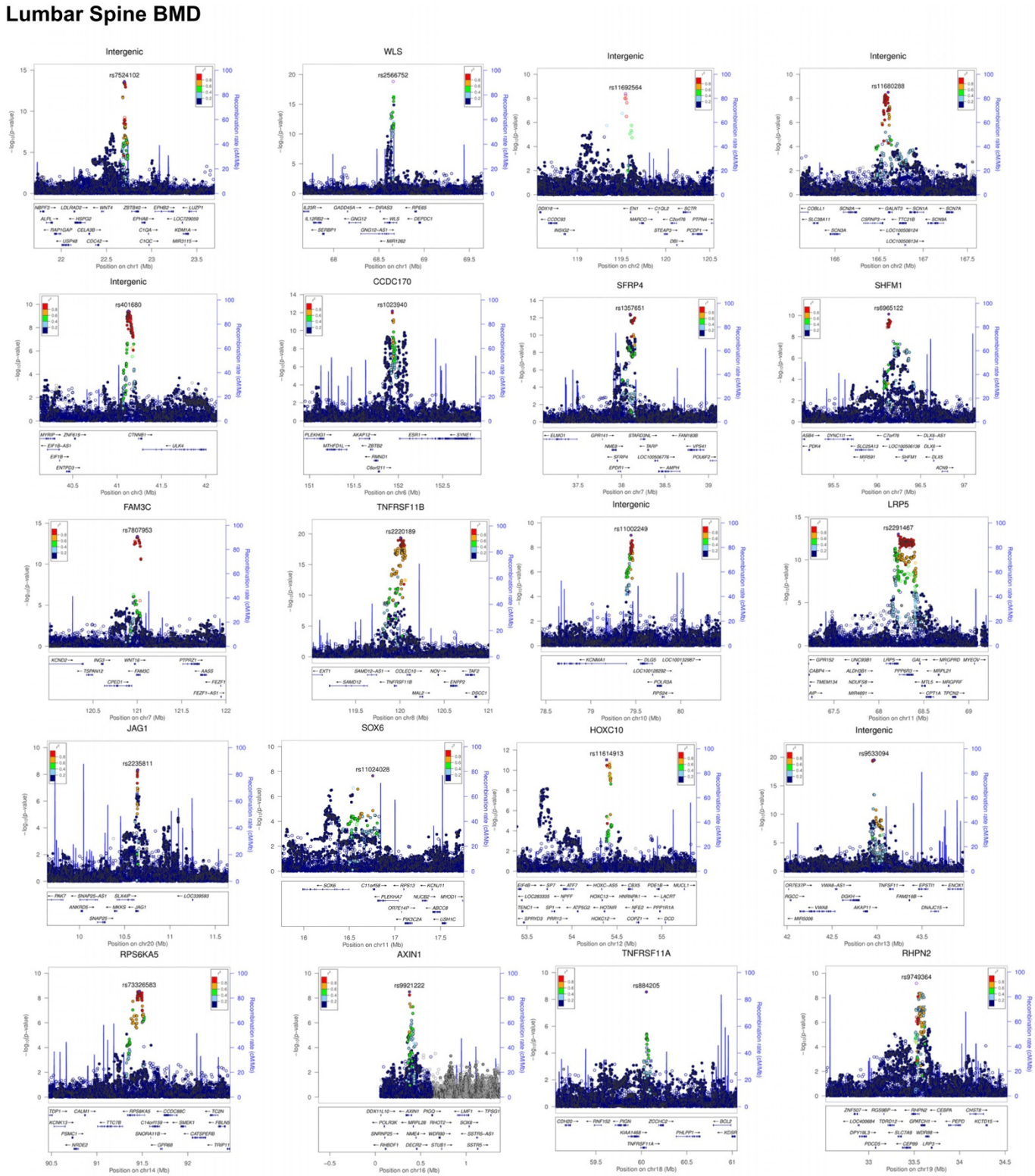
Forearm BMD



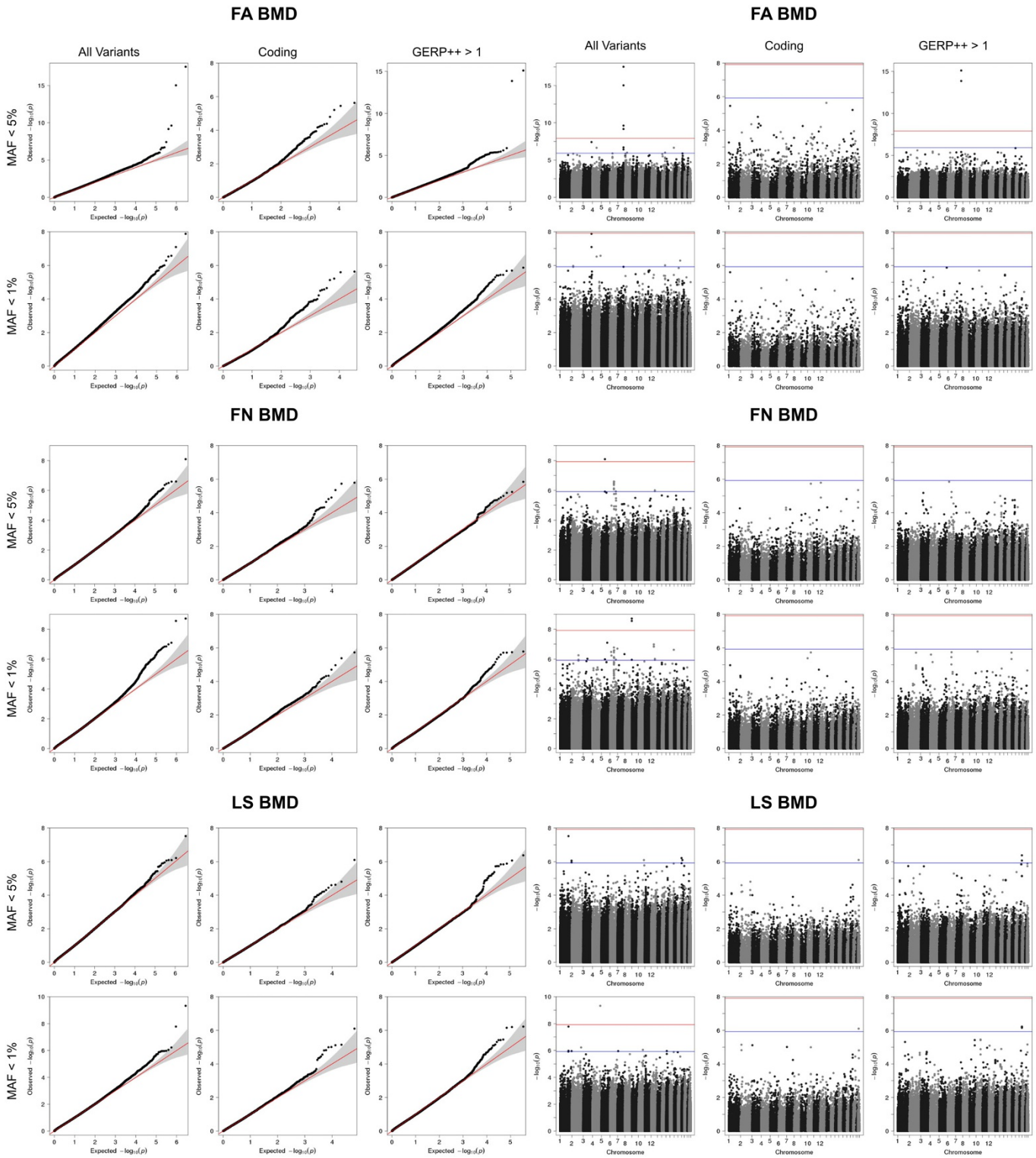
Femoral Neck BMD



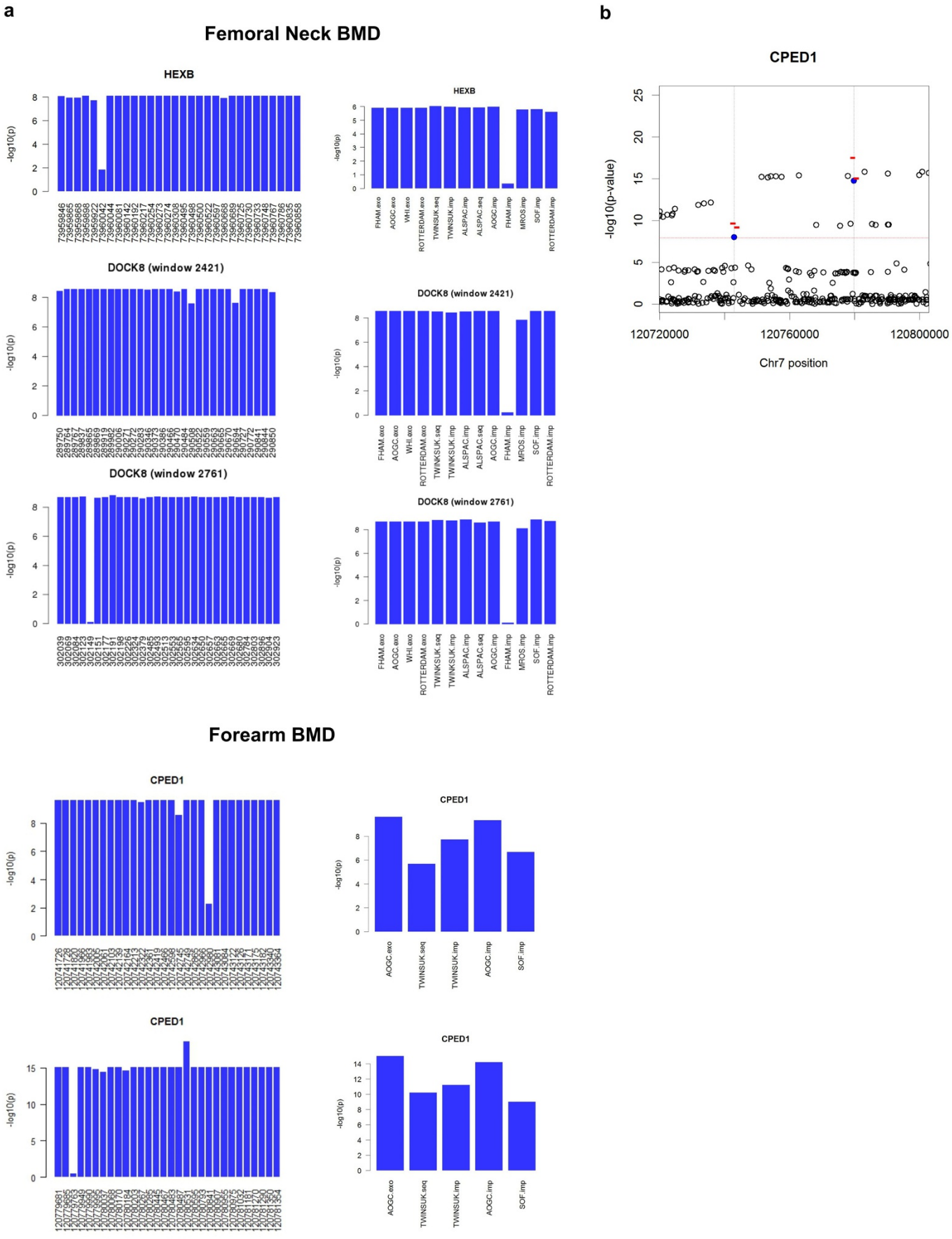
Extended Data Figure 7 | Regional plots of genome-wide significant loci from single-SNV association tests for forearm and femoral neck BMD. Each regional plot depicts SNVs within 1 Mb of a locus' lead SNV (*x* axis) and their associated meta-analysis *P* value ($-\log_{10}$). SNVs are colour-coded according to r^2 with the lead SNV (labelled, r^2 calculated from UK10K whole-genome sequencing data set). Recombination rate (blue line), and the position of genes, their exons and the direction of transcription are also displayed (below plot).



Extended Data Figure 8 | Regional plots of genome-wide significant loci from single-SNV association tests from lumbar spine BMD. Each regional plot depicts SNVs within 1 Mb of a locus' lead SNV (x axis) and their associated meta-analysis P value ($-\log_{10}$). SNVs are colour coded according to r^2 with the lead SNV (labelled, r^2 calculated from UK10K whole genome sequencing data set). Recombination rate (blue line), and the position of genes, their exons and the direction of transcription are also displayed (below plot).



Extended Data Figure 9 | Region-based association tests using skatMeta for windows of 30 SNVs and window step of 20 SNVs. **a**, Left, quantile–quantile plots for forearm (FA) BMD, femoral neck (FN) BMD, and lumbar spine (LS) BMD. For each MAF range considered (<5% or < 1%), analysis was conducted across all variants, variant overlapping coding exons, and variants with GERP++ score >1. **b**, Right, Manhattan plots forearm BMD, femoral neck BMD, and lumbar spine BMD. For each MAF range considered (<5% or < 1%), analysis was conducted across all variants, variant overlapping coding exons, and variants with GERP++ score >1. Blue lines indicate genome-wide suggestive ($P = 1.2 \times 10^{-6}$) thresholds and red lines indicate genome-wide significant ($P = 1.2 \times 10^{-8}$) thresholds.



Extended Data Figure 10 | Single variant analysis of signals from region-based tests. **a**, Drop-one SNV (left) and drop-one cohort (right) for genome-wide significant 30 SNV windows for femoral neck and forearm BMD from skatMeta analysis. On left, for a given 30 SNV window, the $-\log_{10}P$ of skatMeta test for 29 SNVs, excluding (that is, dropping) the SNV at position labelled on the *x* axis. On right, for given 30 SNV window on left, the $-\log_{10}P$ of skatMeta test for all cohorts, excluding (that is, dropping) cohort labelled on *x* axis. **b**, Regional view of CPED1/WNT16 locus for forearm BMD. Significant SNVs from single variant meta-analysis (rs148771817 and rs79162867, in blue) overlap significant regions found using region-based test (red bars).