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

Citation for published version:

Digital Object Identifier (DOI):
10.1038/nature14878

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Extended Data Figure 1 | Discovery single variant meta-analysis. a. Overall study design. 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 depict p-values prior (blue) and after conditional analysis (red). 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/1KG reference panel with MAF > 0.5% across the 22 autosomes (odd=grey, even=black) against the –log10 p-value from the meta-analysis of 7 cohorts (dots). Also depicted is the subset variants from the reference panel that are also present in Estrada et al. (2012) with p value < 5e-6 (diamonds). Variants with MAF < 5% and p < 1.2e-6 are also depicted (red).
Extended Data Figure 2 | Forest Plots by Cohort for Genome-wide Significant Loci from Discovery Meta-analysis.
Extended Data Figure 3 | Distal/promoter DHS correlations for SNV rs18830909 (a) and rs148771817 (b). Distal/promoter DHS correlations for the DHS region overlapping rs18830909 (1). The distal DHS region was defined from the “Osteoblasts DNaseI HS Peaks from ENCODE/Duke” track (rs18830909) or “Digital DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE” track (rs148771817) (2), and was used to obtain all overlapping 100 nucleotides DHS bin measurements. These measurements were correlated to DHS measurements obtained from promoter regions of all genes within 500 kb, where promoter regions were defined as 500 nucleotides flanking the transcription start site (top-most tracks) (3).
Extended Data Figure 4 | 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. Bglap is shown for comparison, 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 percent 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±SEM. Significance computed by student unpaired t-test.
Extended Data Figure 5 | 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 R26R-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. A. Control animals lacking the R26R-YFP reporter show low background YFP signal (green). B. In En1-Cre+/+; R26R-YFP+/+ mice YFP-expressing cells are detected in the growth plate chondrocytes of the vertebra (*), trabecular bone lining cells (arrow) and osteocytes (arrow head). Note, high fluorescent background staining in the marrow space. 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). D. Alignment of the AP and YFP images shows that the trabecular lining cells co-express AP and YFP.

b. Colocalization of En1 and Alkaline Phosphatase expression. Images of lumbar vertebrae sections (growth plate and trabecular bone regions, 40x) from two-month old En1-lacZ+ mice. (see Figure 3b), stained for LacZ and Alkaline phosphatase (AP), false-coloured as indicated. Double-positive cells are indicated by arrows, while single-positive cells are indicated by arrowheads (LacZ+) or asterisks (AP+). Except for some chondrocytes, most AP+ cells are also LacZ+, i.e. express En1. The bone marrow was digitally removed, as it contains no AP+ cells.
Extended Data Figure 6 | MicroCT Results for control (\(En1^{+/+}\)) and self-deleting \(En1\) knockout (sdEn1, \(En1^{Cre/flox}\)) animals

**a.** Trabecular Bone MicroCT images from Lumbar Vertebra 5. **b.** Morphological characteristics at lumbar vertebra 4, 5, and 6 (from bottom to top). **c.** Morphological characteristics of left femur trabecular bone and **d.** left femur cortical bone. **e.** MicroCT parameter results for the comparison of control type 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 7 | Region-based association tests using skatMeta for windows of 30 SNVs and window step of 20 SNVs. 

**a.** From top to bottom, quantile-quantile plots for forearm BMD (FA), femoral neck BMD (FN), 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.** From top to bottom, 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 and red lines at genome-wide suggestive ($P = 1.2 \times 10^{-6}$) and genome-wide significant ($P = 1.2 \times 10^{-8}$) thresholds, respectively.
**Extended Data Figure 8 | Single Variant Analysis of Signals from Region-based Tests.**

**a.** Drop-one SNV and drop-one cohort for genome-wide significant 30 SNV windows for forearm BMD from skatMeta analysis. (A, B) For given 30 SNV window, the $-\log_{10}(p)$ of skatMeta test for 29 SNVs, excluding (i.e. dropping) the SNV at position labeled on x-axis. (C, D) For given 30 SNV window, the $-\log_{10}(p)$ of skatMeta test for 4 cohorts, excluding (i.e. dropping) cohort labeled on x-axis. **b.** Drop-one SNV and drop-one cohort for genome-wide significant 30 SNV windows for femoral neck BMD for skatMeta analysis. (A) For given 30 SNV window, the $-\log_{10}(p)$ of skatMeta test for 29 SNVs, excluding (i.e. dropping) the SNV at position labeled on x-axis. (B) For given 30 SNV window, the $-\log_{10}(p)$ of skatMeta test for 4 cohorts, excluding (i.e. dropping) cohort labeled on x-axis. **c.** Regional view of CPED1/WNT16 locus for forearm BMD. In top panel, significant SNVs from single variant meta-analysis (rs148771817 and rs79162867, in blue) overlap significant regions found using region-based test (red bars).
Forearm BMD

Extended Data Figure 9 | 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 (-log10). SNVs are color coded according to $r^2$ with the lead SNV (labelled, $r^2$ calculated from UK10K whole genome sequencing dataset). Recombination rate (blue line), and the position of genes, their exons and the direction of transcription are also displayed (below plot).
Lumbar Spine

**Extended Data Figure 10 | 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 (-log10). SNVs are color coded according to $r^2$ with the lead SNV (labelled, $r^2$ calculated from UK10K whole genome sequencing dataset). Recombination rate (blue line), and the position of genes, their exons and the direction of transcription are also displayed (below plot).
Extended Data Figure 11 | Discovery Sex-combined Meta-analysis for SNVs present across both exome-sequenced and genome-wide cohorts. **a.** 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-wide cohorts i.e. SNV absent from all exome-sequenced cohort are not shown. **b.** Manhattan plot for the Meta-Analysis of Sex-Combined results for Lumbar Spine BMD for SNVs present in exome-sequenced and genome-wide cohorts i.e. SNV absent from all exome-sequenced cohort are not shown (from top to bottom: lumbar spine, forearm and femoral neck BMD).