

Multiscale analysis and functional validation of the cellular and genetic determinants of skeletal disease

Received: 7 November 2024

Accepted: 15 May 2026

Published online: 10 July 2026

 Check for updates

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

Musculoskeletal diseases are a major health burden. Development of bone-active therapies has been hindered by limited understanding of the cells and genes that regulate the skeleton. We exploited the value of cross-species analysis and developed single-cell methodologies in skeletal tissues to define the critical endosteal compartment that regulates bone turnover. Thirty-four distinct cell types were identified, and disease-relevant cells prioritized using enrichment for rare skeletal disorder genes and bone-mineral-density-associated genes in an extended UK Biobank genome-wide association study. Functional validation was undertaken in over 1,000 genetically modified mouse models. Endothelial cells and vascular smooth muscle cells were identified as new skeletal-disease-relevant cells alongside osteoblast, chondrocyte and osteoclast cell lineages. Hundreds of cell-specific genes with unappreciated roles in skeletal pathophysiology were identified. This comprehensive cellular and molecular framework underpins skeletal physiology and disease and will help prioritize new therapeutic targets to accelerate development of therapies to treat musculoskeletal disease.

Musculoskeletal diseases, such as osteoporosis are a major health burden^{1,2}. Existing treatments prevent skeletal deterioration but rarely restore integrity. New therapies are needed³; however, drug development is hindered by limited understanding of mechanisms that maintain bone mass and strength.

Gene mapping of rare and common diseases has uncovered new mechanisms of skeletal regulation^{4,5}. Genome-wide association studies (GWAS) have identified over 500 loci associated with estimated bone mineral density (eBMD), a measure of skeletal integrity^{6–9}. Yet at most loci, the underlying effector genes, genes with a functional role in regulating the skeleton, remain unknown. Expression quantitative trait locus (eQTL) analysis of bulk tissues can identify effector genes¹⁰ but is constrained by limited knowledge of gene function in specific cells^{11–13}. Application of eQTL methods to single-cell transcriptomics has the potential to address this¹⁴. However, obtaining bone cells from healthy individuals is difficult¹⁵, even more so from the endosteal compartment of bone, which is a key site of bone turnover located at the interface between bone and bone marrow (BM). Consequently,

bone cells are poorly represented in the Human Cell Atlas and Genotype Tissue Expression project^{16,17}. Given that genes in humans and mice are highly conserved^{18,19}, along with recent success in analyzing GWAS with transcriptomic data from mouse tissue^{20–22}, we reasoned that isolating cells from the endosteal compartment of mice could address these challenges.

We hypothesized that integrating gene expression data from a map of cells isolated from the endosteal compartment of murine bone with human gene mapping studies of rare and common skeletal disease traits would identify the effector genes that contribute to skeletal disease. We generated a single-cell RNA sequencing (scRNA-seq) map of cells from the endosteal compartment of healthy mice. We integrated this with analyses of: (1) the nosology of rare skeletal disorders²³; (2) a new and now largest GWAS of eBMD; (3) skeletal phenotyping data from the Mouse Genome Informatics (MGI) database; (4) an in-depth functional analysis of the skeleton in over 1,000 mouse lines with single gene deletions, and (5) validated outcomes in a scRNA-seq dataset from human bone.

✉ e-mail: graham.williams@imperial.ac.uk; d.bassett@imperial.ac.uk; p.croucher@garvan.org.au; j.kemp2@uq.edu.au

Our multiscale approach defined the cellular landscape of the endosteal compartment and identified effector genes in osteoblasts, chondrocytes, osteoclasts, vascular smooth muscle cells (VSMCs) and endothelial cells (ECs) that contribute to skeletal disease. These data were made available via a bespoke web platform (www.musculoskeletal-genomics.org).

Results

Cells in the endosteal compartment

To identify genes that regulate the skeleton and define the cells in which they function, we developed a multiscale framework (Fig. 1a). We isolated cells from the endosteal compartments of metaphysis and diaphysis, and from the BM of femurs of adult male mice for scRNA-seq (Fig. 1b). We then defined a gene program, a set of genes upregulated in a cell cluster relative to all others, for each cell type (Supplementary Table 1).

Analysis of 133,942 cells identified 34 cell clusters (Fig. 1c) including six categories with distinct cell types: nonhematopoietic cells; monocytes and macrophages; neutrophils; hematopoietic progenitors; B cells; and natural killer (NK) and T cells (Supplementary Table 2 and Supplementary Note 1). Nonhematopoietic cells, including chondrocytes, osteoblast lineage cells, ECs and VSMCs originated largely from the endosteal compartments of the diaphysis and metaphysis (Fig. 1d). Chondrocytes were derived exclusively from the metaphysis. Cells of the monocyte and macrophage lineage, including dendritic cells (DCs) and macrophages, which may include osteal macrophages²⁴, were found in both endosteal and marrow compartments. *Txnip*⁺ monocytes were found predominantly in the BM, while *Cd14*⁺ monocytes were more abundant in the endosteal compartments (Fig. 1d), an enrichment confirmed using flow cytometry (Supplementary Fig. 1). This lineage also included cells denoted as osteoclasts, which were found in both endosteal compartments. While multinucleated cells are normally excluded from scRNA-seq, these cells express osteoclast genes and may represent mononuclear osteoclasts or osteomorphs²⁵. Neutrophils at different stages of maturation, hematopoietic progenitor cells (hematopoietic stem cells (HSCs), basophil-eosinophil-mast cell progenitors and granulomyeloid progenitors), B cells, T and NK cells were also identified.

Osteoblast lineage cells mediate bone formation. Therefore, we determined the gene programs that define osteoblasts and other cell types enriched in the endosteal compartment. The osteoblast gene program consisted of 577 genes (Supplementary Table 2). Seventy-nine genes (13.7%), including *Col1a1*, *Bglap2* and *Ibsp*, were highly upregulated with a $\log_2$ (fold change) greater than 2 (Fig. 1e). While most genes (495 genes, 85.8%) were expressed in at least one other cluster, 82 genes (14.2%) had restricted expression (Supplementary Table 1 and Methods), including *Dmp1*, *Vdr* and *Postn* (Fig. 1f); 387 genes (67.1%) in the osteoblast program were neither annotated in the Gene Ontology (GO) database with skeletal terms nor with abnormal skeletal phenotypes from the MGI mammalian phenotype ontology database (Fig. 1g and Supplementary Table 3).

The chondrocyte program included 546 genes, with 71 highly expressed, including *Col2a1*, *Comp* and *Acan*, and 136

genes with restricted expression (Extended Data Fig. 1a,b and Supplementary Table 2). Skeletal vasculature is also important in bone development and remodeling²⁶. The EC program included 610 genes, with 78 highly expressed (Extended Data Fig. 1c,d). The VSMC program included 636 genes, with 112 highly expressed (Extended Data Fig. 1e,f). Importantly, 26.9% and 27.7% of highly expressed genes in ECs and VSMCs, respectively, were annotated in the GO database with a skeletal biological process or skeletal phenotypes in the MGI database (Extended Data Fig. 1d,f). Finally, the osteoclast program included 687 genes: 19 (2.8%) were highly expressed and 49 (7.1%) expressed only in osteoclasts (Extended Data Fig. 1g,h), including *Tnfrsf11a*, *Dcstamp* and *Atp6v0d2*.

As endosteal cells may regulate bone through direct cell–cell communication, we performed interaction analysis. This revealed the greatest interactions among nonhematopoietic cells, including between osteoblast lineage cells and the vasculature (Fig. 1h, Supplementary Table 4 and Extended Data Fig. 2). Osteoblasts interact with these cells via common (for example, transforming growth factor- β) and distinct pathways (for example, WNT with chondrocytes, NOTCH with ECs, insulin-like growth factor 1 receptor with VSMCs; Supplementary Fig. 2).

Our approach identified the cell types and gene programs that define cells in the endosteal compartment. Osteoclasts, osteoblast lineage cells, chondrocytes, ECs and VSMCs originated almost exclusively (> 98%) from this compartment. Of the 1,886 genes defining the gene programs of these cells, 1,374 (72.9%) had not previously been implicated in skeletal regulation.

Cell states in the endosteal compartment

As cells are at different stages of differentiation, we used high-resolution subclustering of nonhematopoietic cells to define these stages. Sixteen clusters were identified (Extended Data Fig. 3a,b and Supplementary Table 2), including six clusters within the osteoblast lineage, three chondrocyte, four EC, two VSMC clusters and a single neuronal cluster.

The osteoblast lineage formed a continuum: mesenchymal stromal cells (MSCs) progressed through fibroblasts and osteoprogenitors to pre-osteoblasts, mature osteoblasts and early osteocytes (Extended Data Fig. 3c and Supplementary Table 2). The pre-osteoblast number was higher in the diaphysis relative to metaphysis (Supplementary Fig. 3). Gene expression changed throughout differentiation (Extended Data Fig. 3b,d). *Cxcl12*⁺/*Adipoq*⁺ MSC and *Apod*⁺/*Igf1bp5*⁺ fibroblast programs were enriched for GO processes in cell proliferation and migration (Extended Data Fig. 3e). *Postn*⁺/*Col3a1*⁺ osteoprogenitors and *Car3*⁺/*Mmp13*⁺ pre-osteoblasts were associated with extracellular matrix organization, *Bglap2*⁺/*Col1a1*⁺ mature osteoblasts with collagen fibril organization and *Mepe*⁺/*Phex*⁺ early osteocytes with ossification. This trajectory coincided with expression of key transcription factors (TFs) and their target genes identified using SCENIC analysis, including *Cebpa* and *Foxc1* in MSCs, which are associated with adipogenic and osteogenic programs^{27,28}, as well as *Runx2* and *Sp7* in committed cells^{29,30} (Extended Data Fig. 4a and Supplementary Table 5). Single-nucleus assay for transposase-accessible chromatin using sequencing

Fig. 1 | Mapping cell types in the endosteal compartment. **a**, Workflow used to identify and functionally validate effector genes at loci associated with the eBMD. Panel **a** created in BioRender; Chai, R. <https://BioRender.com/4m4I20I> (2026). **b**, Anatomical regions of the mouse femur from which endosteal bone cells and BM cells were isolated and sequenced for the individual experiments. The numbers in the table indicate the total number of cells present after filtering. **c**, Uniform manifold approximation and projection (UMAP) of scRNA-seq data from **b**, showing 34 clusters across six broad cell categories. **d**, Distribution of cell types in the six cell categories across the diaphysis (blue), metaphysis (green) and marrow (red). Shades denote individual experiments and are colored as in **b**, **e**, **f**. Density plots showing the $\log_2$ (fold change) of the osteoblast lineage

cluster marker genes (**e**) and the frequency of these genes across other cell clusters (**f**). **g**, Proportion of osteoblast lineage genes stratified according to $\log_2$ (fold change) magnitude (0.5–1, 1–2 and >2). Genes annotated with skeletal process in the GO database (green), the MGI database (purple), in both databases (blue), or not annotated in either (orange) are shown. The Circos plot highlights genes with $\log_2$ (fold change) > 2; the bold text indicates genes unique to the osteoblast lineage program. **h**, Heatmaps of CellPhoneDB-predicted intercellular interactions showing (i) all pairwise interactions across the dataset and (ii) directional interactions between nonhematopoietic clusters. GMP, granulocyte-monocyte progenitor.



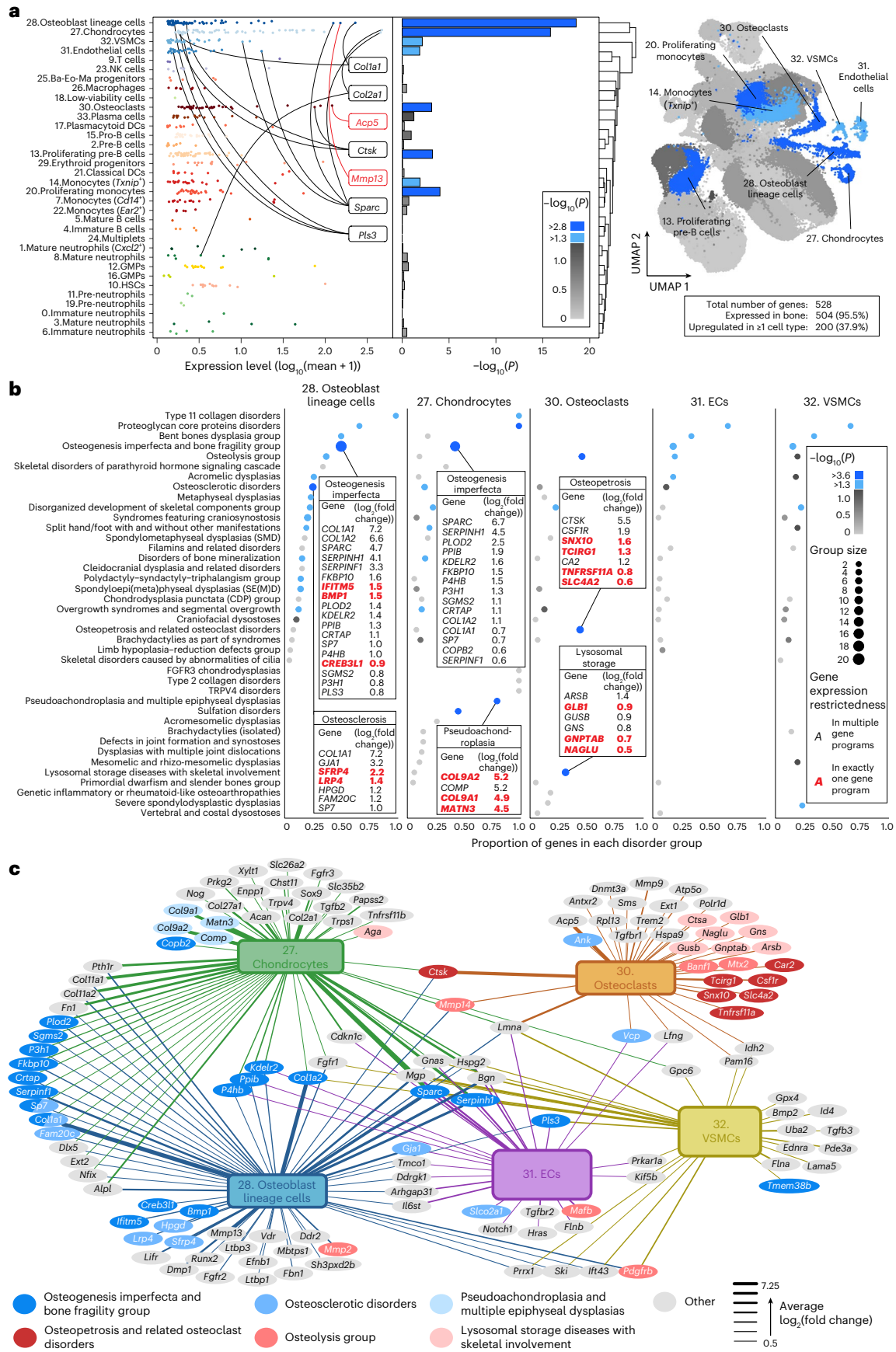


Fig. 2 | Gene programs of specific cell types are enriched for genes involved in human monogenic skeletal disorders. **a**, Dot plot showing mean expression ($\log_{10}(\text{mean} + 1)$) of known monogenic skeletal disorder genes across gene programs. Red highlights genes unique to a single program. The bar and UMAP plots indicate enrichment P value, as determined using a one-tailed Fisher's exact test under the hypergeometric distribution; light blue denotes the nominal significance of $P < 0.05$ ($-\log_{10}(P)$ of > 1.3), while dark blue denotes the Bonferroni-corrected significance of $P < 1.5 \times 10^{-3}$ ($-\log_{10}(P)$ of > 2.8). **b**, Bubble plot showing the number and proportion of genes from 40 monogenic skeletal disorder groups within selected cell-type gene programs. Bold indicates groups enriched in at least one cell type. Circle size represents gene count; color

intensity reflects the P value of enrichment, as determined using a one-tailed Fisher's exact test under the hypergeometric distribution. The light blue circles indicate the nominal significance of $P < 0.05$ ($-\log_{10}(P)$ of > 1.3); dark blue circles indicate Bonferroni-corrected significance of $P < 2.4 \times 10^{-3}$ ($-\log_{10}(P)$ of > 3.6). The boxes identify the causative genes for the indicated disorder group present in a cell-type gene program. Red denotes a gene found only in that gene program and not shared with other gene programs. **c**, Network plot illustrating the expression patterns of monogenic skeletal disorder genes across selected cell types. Nodes are color-coded according to disorder group, with line thickness representing the magnitude of differential expression ($\log_2(\text{fold change})$).

(snATAC-seq) identified unique and overlapping TF-binding motifs in osteoblast lineage cells (Extended Data Fig. 4b–e and Supplementary Table 5). snATAC-seq identified *Tfap2a*, *Ebf2* and *Zic1* as having known roles in the skeleton^{31–33}, as well as new TFs such as *Ebf3*, *Maz* and *Ctcf*. TFs identified using both methods included *Runx1* and *Creb3l1*, as well as *Klf15*, *Etv4* and *Glis2*, which have unappreciated roles in bone (Extended Data Fig. 4e).

Critically, 56.3–72.6% of genes within these osteoblast lineage gene programs were not annotated in GO or the MGI database, indicating that many genes involved in skeletal regulation have yet to be studied (Extended Data Fig. 3f).

Chondrocytes transitioned from *Ucma*⁺/*Serpina1b*⁺ resting chondrocytes to *Ppa1*⁺/*Scrg1*⁺ pre-hypertrophic chondrocytes and *Col10a1*⁺/*Ilhh*⁺ hypertrophic chondrocytes, forming a distinct trajectory (Extended Data Fig. 3a,b,g–i; Supplementary Fig. 3 and Supplementary Table 2). The EC clusters included *Ubd*⁺/*Tfpi*⁺ sinusoidal cells, two clusters of arteriolar cells (*Cldn5*⁺/*Gkn3*⁺ and *Ly6c1*⁺/*Glut*⁺) and *Plvap*⁺/*Aplnr*⁺ type H cells, which have been implicated in skeletal development³⁴ (Extended Data Fig. 3a,b and Supplementary Table 2). VSMC clusters included smooth muscle cells (SMCs) (*Myl9*⁺/*Acta2*⁺) and pericytes expressing *H2-M9* and *Rgs5*. Arteriolar cells (*Cldn5*⁺/*Gkn3*⁺) and SMCs were enriched in the diaphysis, whereas type H ECs were enriched in the metaphysis (Supplementary Fig. 3).

This analysis identified the stages of differentiation in non-hematopoietic cells in the endosteal compartment. Many of the genes in these cells are not known to regulate the skeleton.

Enrichment of rare skeletal disorder genes

We used overrepresentation analyses to identify gene programs of endosteal cells that were enriched for rare monogenic skeletal disorder genes. We used the International Skeletal Dysplasia Society nosology, a set of 528 protein-coding genes with pathogenic mutations causing 719 disorders (Supplementary Table 3)²³. Although multiple cell types expressed causative genes, only some were highly enriched for causative genes, including osteoblast lineage cells, chondrocytes, osteoclasts, proliferating monocytes and proliferating pre-B cells (Fig. 2a and Supplementary Table 6).

Osteoblast lineage cells were enriched, with 61 genes of 553 human orthologs known to cause rare skeletal disorders (Supplementary Table 6). All differentiation states except MSCs were enriched (Extended Data Fig. 5a). Osteoblast lineage cells were primarily enriched for

genes causing osteogenesis imperfecta (OI) and other bone fragility disorders, and osteosclerotic disorders (Fig. 2b,c). Eighteen of 36 known OI and bone fragility genes were present, with causative genes overrepresented in pre-osteoblasts and mature osteoblasts (Extended Data Fig. 5b,d and Supplementary Table 6). Genes causing osteosclerotic and bone mineralization disorders were enriched in pre-osteoblasts and early osteocytes.

Chondrocytes were enriched for causative genes for proteoglycan core protein disorders, sulfation disorders, pseudoachondroplasia, multiple epiphyseal dysplasias and OI and bone fragility (Fig. 2b,c). Enrichment varied among cell states (Extended Data Fig. 5c,d). Osteoclasts were enriched for genes that cause osteopetrosis and osteolysis and lysosomal storage disorders (Fig. 2b,c). Proliferating monocytes and pre-B cells were enriched for genes causing dwarfism, potentially reflecting expression of cell cycle genes³⁵ (Supplementary Table 6). ECs and VSMCs showed nominal evidence of enrichment ($P < 0.05$) for proteoglycan core protein disorders and bent bone disease, possibly because of genes shared with osteoblasts (Fig. 2b,c).

Enrichment of genes associated with eBMD

We then investigated whether gene programs of endosteal cells were enriched for eBMD-associated genes. We conducted (1) GWAS to identify genetic variants associated with eBMD; (2) gene-based tests to identify genes associated with eBMD; and (3) competitive gene set analysis (GSA) to identify gene programs enriched for eBMD-associated genes.

GWAS involved 448,010 UK Biobank (UKB) participants with eBMD measures (Supplementary Table 7). This included 21,186 additional individuals, analysis of 36% more genetic variants and more detailed analysis of the X chromosome than previously performed³. A total of 169,684 variants were associated at genome-wide significance ($P < 6.6 \times 10^{-9}$)⁷. Twenty-five were high-impact variants and predicted to alter gene function (Fig. 3a and Supplementary Table 7). This included known variants in *MEPE* associated with lower eBMD³⁶, and variants associated with higher eBMD, including a 'stop/lost' variant in the NHERF family PDZ scaffold protein 2 (*NHERF2*), a regulator of sodium transport³⁷ (Fig. 3a,b and Supplementary Table 7).

Fine-mapping identified 1,294 independent lead variants, defined as variants most strongly associated with eBMD that mapped closest to 903 unique protein-coding genes (Extended Data Fig. 6a and Supplementary Table 7). Sixteen lead variants represented previously unreported associations and 13 mapped to the X chromosome

Fig. 3 | Gene programs of specific cell types are enriched for genes associated with eBMD. **a**, Circos plot of high-impact variants that are predicted to alter gene function. The inner ring denotes the type of variant, the middle ring denotes the frequency of the alternate allele and the outer ring denotes the effect of the alternate allele on eBMD. **b**, LocusZoom plots showing eBMD-associated variants mapping closest to *MEPE* and *NHERF2*. High-impact variants are represented by stars. Pairwise linkage disequilibrium (LD) could not be estimated for the lead eBMD-associated variant closest to *MEPE*. **c**, Manhattan plot summarizing the results of the genome-wide gene-based tests of eBMD conducted in the UKB study using MAGMA. The dotted line denotes the threshold of statistical significance ($P < 2.5 \times 10^{-6}$). Orange circles indicate genes that are associated

with eBMD and are monogenic skeletal disorder genes. **d**, Bar and UMAP plots showing gene programs of different cell types that are enriched for eBMD-associated genes. The scale bar in the bar and UMAP plots indicates the P value, as determined using MAGMA GSA with a one-sided test for positive association. The light blue bars in the bar plot and the light blue clusters in the UMAP plot correspond to observations that have nominal evidence of enrichment: $P < 0.05$ ($-\log_{10}(P) > 1.3$). The dark blue bars in the bar plot and the dark blue clusters in the UMAP plot correspond to observations that have robust evidence of enrichment and meet the Bonferroni-corrected significance threshold: $P < 1.5 \times 10^{-3}$ ($-\log_{10}(P) > 2.8$).

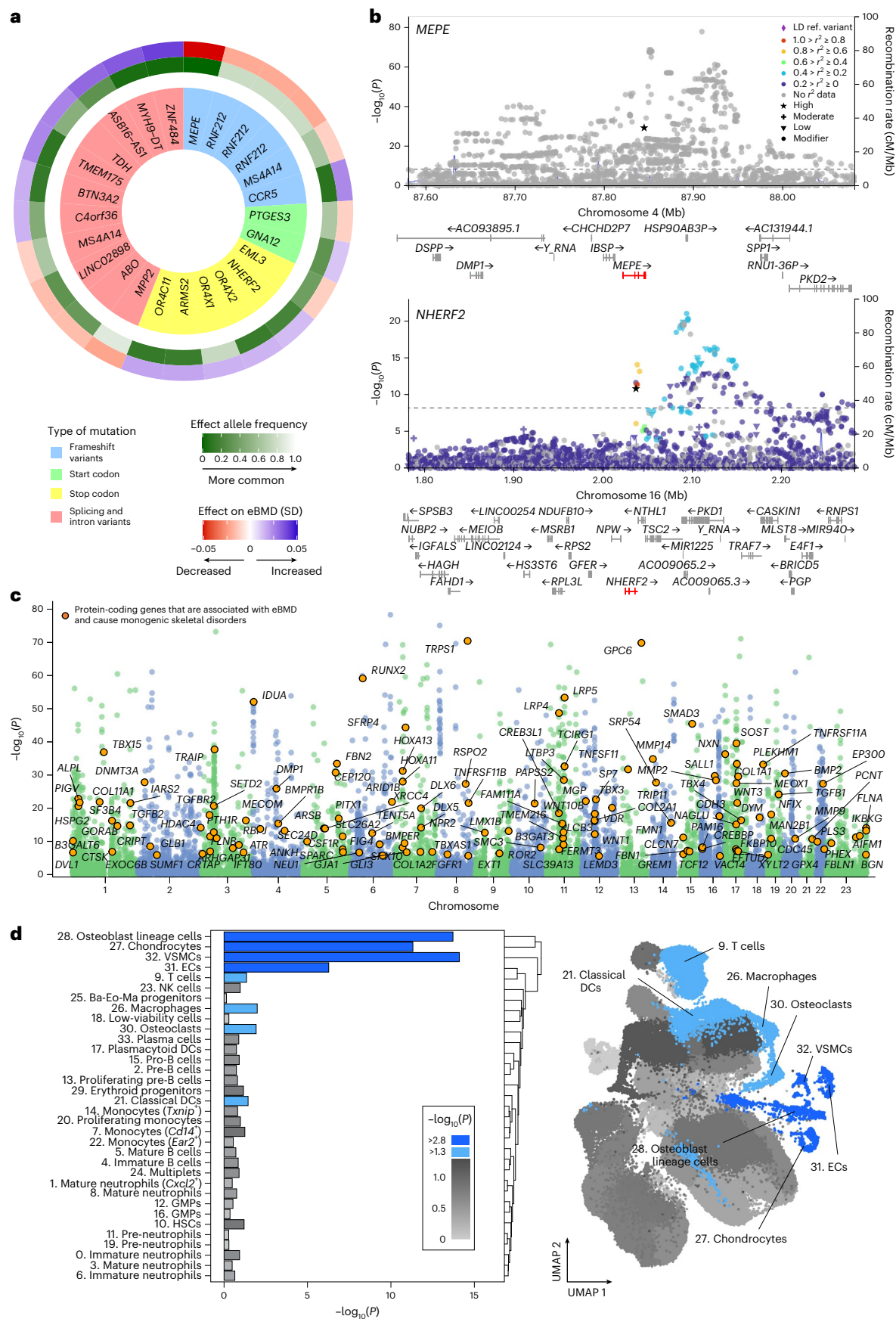




Fig. 4 | Genes that cause abnormal skeletal phenotypes when deleted in mice.

a, The OBCD phenotyping pipeline. Panel **a** created in BioRender; Chai, R. <https://BioRender.com/4m4120I> (2026). **b**, Skeletal effects of deleting genes identified in nonhematopoietic cell and osteoclast gene programs. Circos plots represent individual cell clusters identified in Fig. 1, with total gene counts indicated at the center of each plot. From inner to outer rings: (1) percentage of KO mice with skeletal phenotypes; (2) individual deleted genes (red indicates monogenic disorder gene orthologs; symbols denote homozygous

lethality for heterozygous models); (3) effects on cortical bone; (4) effects on trabecular bone; (5) MAGMA-gene-based eBMD association (blue/gray); (6) cell-type-restricted expression (green); (7) GO skeletal process annotation; and (8) MGI abnormal bone structure annotation. The key summarizes the color coding for each phenotype, association with monogenic disorders and eBMD, gene expression restrictedness and skeleton-related annotations. **a**, Scale bar in Faxitron images of femur and vertebra, 1 mm; scale bars in micro-CT images, 100 μ m. Ct.BMD, cortical BMD; Int.Diam, internal diameter.

(Extended Data Fig. 6a). Three were closest to genes involved in monogenic disorders with abnormal BMD, including *PLS3*, which causes X-linked juvenile osteoporosis³⁸, *IKBKG*, which causes osteopetrosis³⁹, and *PHEX*, which causes abnormal mineralization and X-linked hypophosphatemic rickets (Extended Data Fig. 6b and Supplementary Table 7)⁴⁰. The set of protein-coding genes closest to lead variants was enriched for genes that cause rare skeletal disorders ($P = 2.1 \times 10^{-19}$) (Extended Data Fig. 6c and Supplementary Table 7). These included genes involved in disorders with abnormal BMD, such as OI. Thus, our GWAS extends eBMD-associated variants from 1,103 (518 loci)⁸ to 1,294 (533 loci) and highlights variants and new genes that may regulate skeletal integrity.

As gene-based tests of association offer a powerful approach to estimate gene-trait associations, we conducted a MAGMA analysis⁴¹. We detected eBMD associations for 3,883 of 19,695 protein-coding genes ($P = 2.5 \times 10^{-6}$; Fig. 3c and Supplementary Table 7). A total of 788 eBMD-associated genes were among the 903 genes located closest to lead variants identified using GWAS (Supplementary Table 7). eBMD-associated genes were enriched for rare monogenic skeletal disorder genes (Supplementary Fig. 4a and Supplementary Table 7), including causative genes for high and low BMD disorders. Analysis of pulse rate, an unrelated trait, showed no enrichment for causative genes, suggesting that enrichment was not by chance (Supplementary Fig. 4b).

GSA showed the gene programs of osteoblast lineage cells, chondrocytes, ECs and VSMCs, were enriched for eBMD-associated genes (Fig. 3d and Supplementary Table 7). All stages of osteoblast, chondrocyte, VSMC differentiation, and one of four EC states were enriched for eBMD-associated genes (Extended Data Fig. 7 and Supplementary Table 7). As 57% of genes in nonhematopoietic cell gene programs were present in at least one other program, we performed pairwise conditional GSA to determine whether enrichment was confounded by shared genes (Supplementary Fig. 5a,b and Supplementary Table 7). Osteoblast lineage cells, chondrocytes, ECs and VSMCs remained enriched for genes associated with eBMD ($P_{\text{conditional}} < 0.05$; Supplementary Fig. 5b and Supplementary Table 7). Given that a subset of genes could influence enrichment, we also performed a post-hoc permutation analysis; 93–99% of genes contributed to enrichment, suggesting that small numbers of genes were not unduly influential (Supplementary Fig. 5c).

Together, this demonstrates that gene programs of osteoblast lineage cells, chondrocytes, ECs and VSMCs are enriched for genes associated with eBMD.

Enrichment of eBMD effector genes known to regulate bone

We next investigated whether eBMD-associated genes in nonhematopoietic cell and osteoclast programs influence skeletal structure. We

identified 1,347 protein-coding genes in the MGI mouse phenotype ontology database with human orthologs that when mutated cause ‘abnormal bone structure’ (Supplementary Table 3). Osteoblasts, chondrocytes, osteoclasts and VSMCs, but not ECs, were highly enriched (Extended Data Fig. 8a and Supplementary Table 8). Enrichment increased when the analysis was restricted to eBMD-associated genes (Extended Data Fig. 8a, Supplementary Fig. 6a and Supplementary Table 8). Sensitivity analyses excluding monogenic disorder genes indicated that enrichment persisted for osteoblasts, chondrocytes and VSMCs (Extended Data Fig. 8b and Supplementary Table 8). Enrichment increased for genes with the greatest expression in osteoblast, chondrocyte and osteoclast gene programs and increased further when restricted to genes associated with eBMD (Extended Data Fig. 8c, Supplementary Fig. 6b and Supplementary Table 8).

New eBMD effector genes regulate structure and function

We next determined whether eBMD-associated genes expressed in nonhematopoietic cells and osteoclasts have functional roles in bone and are effector genes. We screened adult female mice from more than 1,000 mouse lines with gene deletions that have undergone structural and functional skeletal phenotyping in our Origin of Bone and Cartilage Disease (OBCD) study (Fig. 4a; see Methods and Supplementary Table 1 for definitions). This extends the original 100 lines⁴² and the 596 lines reported in 2019 (ref. 8).

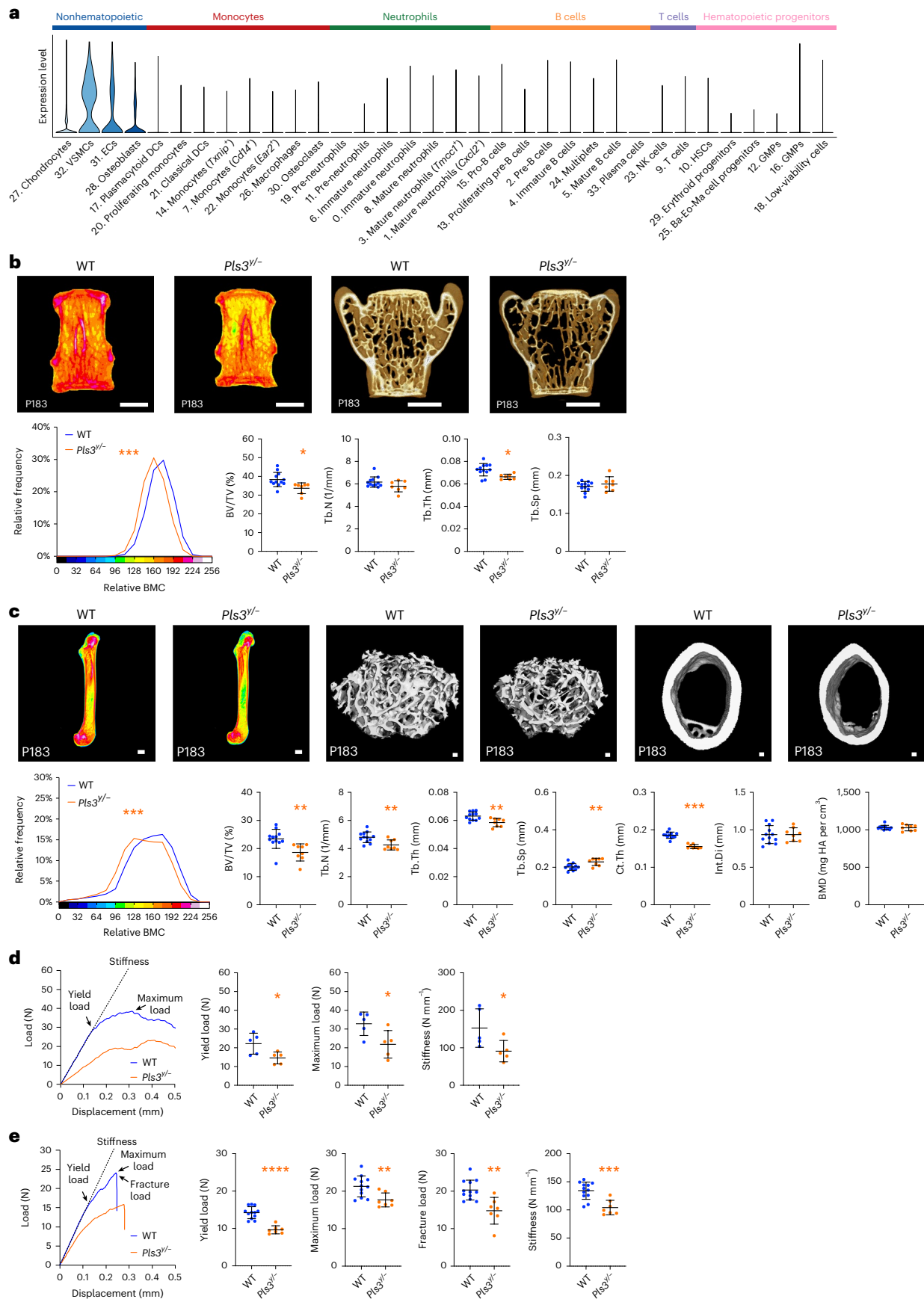
Of the 1,886 genes in the five programs, 101 (5.3%) had deletions in the OBCD study (Fig. 4b); 32 in the osteoblast program; 25 in chondrocytes; 31 in ECs; 26 in VSMCs; and 36 in osteoclasts. Fifty-one (50.5%) exhibited abnormal structural, functional or bone quality phenotypes (Fig. 4b and Supplementary Table 9).

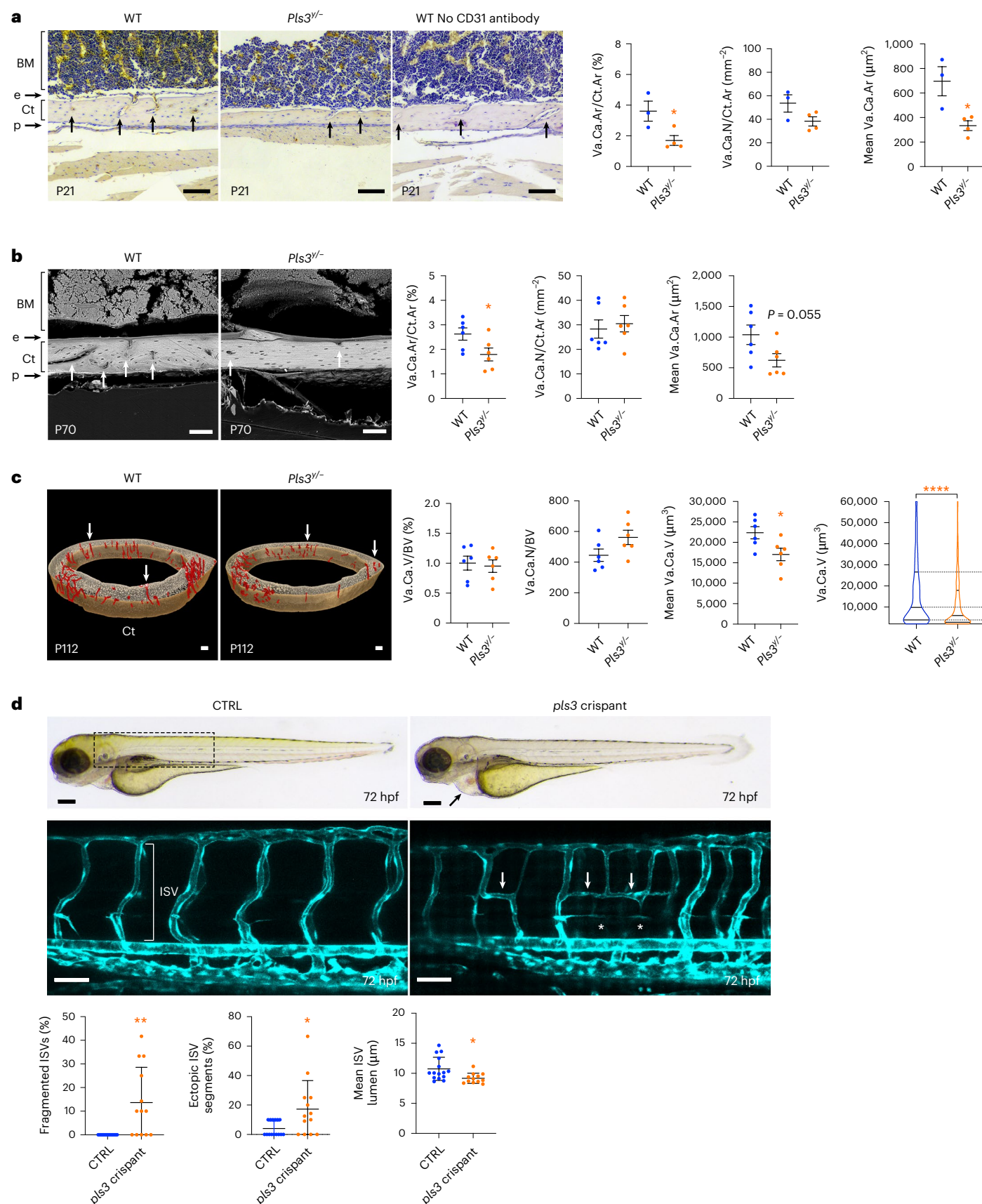
Some genes had established skeletal roles, including *Bgn* (biglycan)⁴³, *Creb3l1* (cAMP responsive element binding protein 3 like 1)⁴⁴, *Gpc6* (glypican 6)⁷ and *Pls3* (plastin 3)³⁸. Many genes identified were not annotated in the GO database or had no or limited skeletal impact in the MGI database, suggesting they are unknown regulators of bone. These included cell-type-restricted genes like *Ptprd*, *Nherf2* (ECs), *Cbx6* (VSMCs) and *Agap1* (osteoclasts) (Fig. 4b and Supplementary Table 9). This suggests that the gene programs defining cells in the endosteal compartment include new eBMD effector genes.

PLS3. To explore a widely expressed effector gene, we performed studies on *Pls3*. While *PLS3* loss-of-function mutations cause OI and X-linked osteoporosis in children³⁸, we also found that the common genetic variation is associated with eBMD in adults (Fig. 3c, Extended Data Fig. 6a and Supplementary Table 7). *Pls3* was expressed in osteoblast lineage cells, ECs and VSMCs in bone (Fig. 5a), and in epithelial, mesenchymal and vascular cells outside bone (Supplementary Fig. 7 and

Fig. 5 | Skeletal phenotype of male *Pls3*^{fl} mice. **a**, Violin plots of *Pls3* expression in the endosteal bone and BM scRNA-seq dataset. **b**, Lumbar vertebra (L5) analysis from P183 mice (WT $n = 12$; *Pls3*^{fl} $n = 8$). Left: Pseudocolored X-ray microradiography and BMC relative frequency histograms (Kolmogorov–Smirnov test, $P < 0.001$). Low BMC is shown in blue/green; high BMC is shown in pink. Right: Mid-coronal micro-CT sections and quantification of bone volume as a proportion of tissue volume (BV/TV), Tb.N, Tb.Th and Tb.Sp. **c**, Femur analysis from P183 mice (WT $n = 12$; *Pls3*^{fl} $n = 8$). Left: Pseudocolored X-ray microradiography and BMC histograms ($P < 0.001$). Low BMC is shown in

blue/green; high BMC is shown in pink. Middle: Distal femur micro-CT and trabecular parameters. Right: Mid-shaft cortical micro-CT and Ct.Th, Int.Di and Ct.BMD. **d**, Compression testing of L5 vertebrae (WT $n = 5$, *Pls3*^{fl} $n = 5$). Representative load displacement curves and quantification of yield load, maximum load and stiffness. **e**, Three-point bend testing of femurs (WT $n = 12$; *Pls3*^{fl} $n = 7$). Representative load displacement curves and quantification of yield load, maximum load, fracture load and stiffness. All dot plots represent the mean $\pm$ s.d.; two-tailed Student's *t*-test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **b**, Scale bar, 1 mm. **c**, Scale bars, 1 mm and 100 μ m (mid-shaft cortical micro-CT).





Supplementary Table 10). It is unclear which cells it acts through to affect bone integrity^{45–48}.

As X-linked osteoporosis is characterized by vertebral fractures in young males, we characterized lumbar vertebrae from *Pls3*-deficient

male mice (*Pls3*^{-/-}). Bone mineral content (BMC) was reduced in vertebrae in *Pls3*^{-/-} mice and micro-computed tomography (CT) analysis demonstrated decreased bone volume (BV/TV) and trabecular thickness (Tb.Th) (Fig. 5b and Supplementary Fig. 8a). In long bones,

Fig. 6 | Vascular phenotype of male *Pls3*^{+/−} mice and *pls3* crispant zebrafish.

a, Anti-CD31 immunohistochemistry of P21 mid-femur sections (WT $n = 3$; *Pls3*^{+/−} $n = 4$). The arrows indicate the cortical vascular canals. Graphs quantify the vascular canal area as a proportion of the cortical area (Va.Ca.Ar/Ct.Ar), vascular canal number per cortical area (Va.Ca.N/Ct.Ar) and mean vascular canal area (Va.Ca.Ar). **b**, Backscattered SEM images of triiodide-stained P70 mid-femur sections (WT $n = 6$, *Pls3*^{+/−} $n = 6$). The arrows indicate the vascular canals. Graphs quantify Va.Ca.Ar/Ct.Ar, Va.Ca.N/Ct.Ar and mean Va.Ca.Ar. **c**, Micro-CT of P112 mid-femur region of interest (ROI) (250 μ m) showing cortical vessels in red (WT $n = 6$, *Pls3*^{+/−} $n = 6$). Graphs quantify vascular canal volume as a proportion of bone volume (Va.Ca.V/BV), vascular canal number per bone volume (Va.Ca.N/BV) and mean vascular canal volume (Va.Ca.Ar). The violin plot shows the Va.Ca.V distribution

(Kolmogorov–Smirnov test, **** $P < 0.0001$). **d**, Top: Morphology of CTRL ($n = 15$) and *pls3* crispant ($n = 13$) zebrafish at 72 h post fertilization (hpf). The arrow indicates pericardial edema; the box denotes the ROI. Bottom: Representative maximum intensity projections of *Tg(kdr-l:eGFP)* expression in the trunk vasculature. The asterisks indicate incomplete ISVs, disconnected from the dorsal aorta or posterior cardinal vein; the white arrows show ectopic ISV segments connecting neighboring ISVs in *pls3* crispants. Graphs quantify the percentage of fragmented ISVs, the percentage of ectopic segments and the mean ISV luminal diameter. Unless otherwise stated, all dot plots represent the mean $\pm$ s.d.; two-tailed Student's *t*-test; * $P < 0.05$, ** $P < 0.01$. **a**, Scale bar, 100 μ m. **b,c**, Scale bar, 100 μ m. **d**, Scale bar, 50 μ m. ct, cortical bone; e, endosteum; p, periosteum.

Pls3^{+/−} mice also showed decreased BMC, reduced BV/TV, trabecular number (Tb.N) and Tb.Th, and increased trabecular spacing (Tb.Sp) (Fig. 5c and Supplementary Fig. 8c,d). Cortical thickness (Ct.Th) was reduced (Fig. 5c and Supplementary Fig. 8e). Biomechanical testing demonstrated that maximum load was impaired in vertebrae at P70; yield load, maximum load and stiffness were all reduced (Fig. 5d and Supplementary Fig. 8b). Long bones also exhibited a reduced yield load, maximum load and stiffness in *Pls3*^{+/−} mice (Fig. 5e and Supplementary Fig. 8f). Female *Pls3*-deficient (*Pls3*^{−/−}) mice demonstrated a similar phenotype (Supplementary Fig. 9).

To investigate the cellular mechanism, we analyzed the major bone cell lineages. Analysis of the postnatal day 1 (P1) skeleton revealed no differences in endochondral ossification (Extended Data Fig. 9a). There were no differences in growth plate reserve and proliferative and hypertrophic zones at P21; linear growth did not differ (Extended Data Fig. 9a). Static and dynamic histomorphometry and serum measures of osteoclastic resorption, osteoblastic bone formation and osteocyte-derived proteins were unaffected in *Pls3*^{+/−} mice (Extended Data Fig. 9b–d). Cortical porosity and osteocyte lacuna size and number were increased in *Pls3*^{+/−} mice (Extended Data Fig. 9e).

As *Pls3* was most highly expressed in VSMCs and ECs (Fig. 5a), we examined the vasculature in bone. *Pls3*^{+/−} mice showed smaller blood vessels in bone compared to controls (CTRLs) (Fig. 6a), which was confirmed by backscattered scanning electron microscopy (SEM) and micro-CT analysis (Fig. 6b,c and Supplementary Fig. 10). Knockdown of *pls3* in zebrafish embryos⁴⁹ resulted in mild cardiac edema and defects in intersegmental vessel (ISV) formation, including disconnections from the dorsal aorta or posterior cardinal vein, ectopic vessel connections and smaller lumens (Fig. 6d). The absence of bone cells at this developmental stage indicates that *pls3* might act directly to control angiogenesis and vascular patterning. These data suggest that X-linked osteoporosis might result from abnormalities in bone vasculature.

Restricted eBMD genes. To understand eBMD-associated genes with restricted expression, we investigated the impact of deleting *Ptprd*, *Nherf2*, *Cbx6* and *Agap1* on bone structure and function. *Ptprd* was restricted to osteoblast lineage cells in bone (Fig. 7a) and expressed in mesenchymal, epithelial and glial cells in lung, brain and ovary (Supplementary Fig. 7 and Supplementary Table 10). Female *Ptprd*^{−/−} mice had normal femoral BMC and length but increased vertebral BMC and decreased vertebral length compared to the wild-type (WT) mice

reference range of 320 female mice from an identical genetic background (Fig. 7b,c). Femurs from *Ptprd*^{−/−} mice showed decreased BV/TV and Tb.N with increased Tb.Sp but normal cortical bone (Fig. 7d,e). Femoral and vertebral bone strength were unchanged (Fig. 7f,g). *Nherf2* expression was restricted to ECs in the endosteal compartment (Fig. 7a) and expressed in ECs in other tissues (Supplementary Fig. 7 and Supplementary Table 10), suggesting that *Nherf2* is important in endothelial function. *Nherf2*^{−/−} mice had decreased vertebral length (Fig. 7c) and reduced femoral BV/TV and Tb.N with increased Tb.Sp (Fig. 7d). Bone strength and stiffness parameters were unchanged (Fig. 7f,g). *Cbx6* expression was restricted to VSMCs in the skeleton (Fig. 7a) but also expressed in the lung, brain, pancreas and stomach (Supplementary Fig. 7 and Supplementary Table 10). *Cbx6*^{−/−} mice exhibited decreased femoral BMC and vertebral length (Fig. 7b,c) and reduced femoral cortical thickness (Ct.Th) (Fig. 7e). Femur yield load and maximum load were decreased compared to WT mice (Fig. 7f,g). Finally, *Agap1* expression was restricted to osteoclasts in bone but also expressed in the heart, brain and liver (Supplementary Fig. 7 and Supplementary Table 10). *Agap1*^{−/−} mice had decreased femur BMC and decreased vertebral BMC and length (Fig. 7b,c). Femurs also had decreased BV/TV and Ct.Th (Fig. 7d,e). Femur maximum load and vertebral yield load, maximum load and stiffness were all decreased (Fig. 7f,g).

Cells and genes involved in skeletal diseases in human bone

To investigate the translational importance of our findings, we isolated and sequenced 125,063 cells from adult human femoral bone (Fig. 8a). Human bone included similar populations of MSCs, pre-osteoblasts, mature osteoblasts, chondrocytes, fibroblasts, ECs and VSMCs to mice (Fig. 8a,b and Supplementary Table 2).

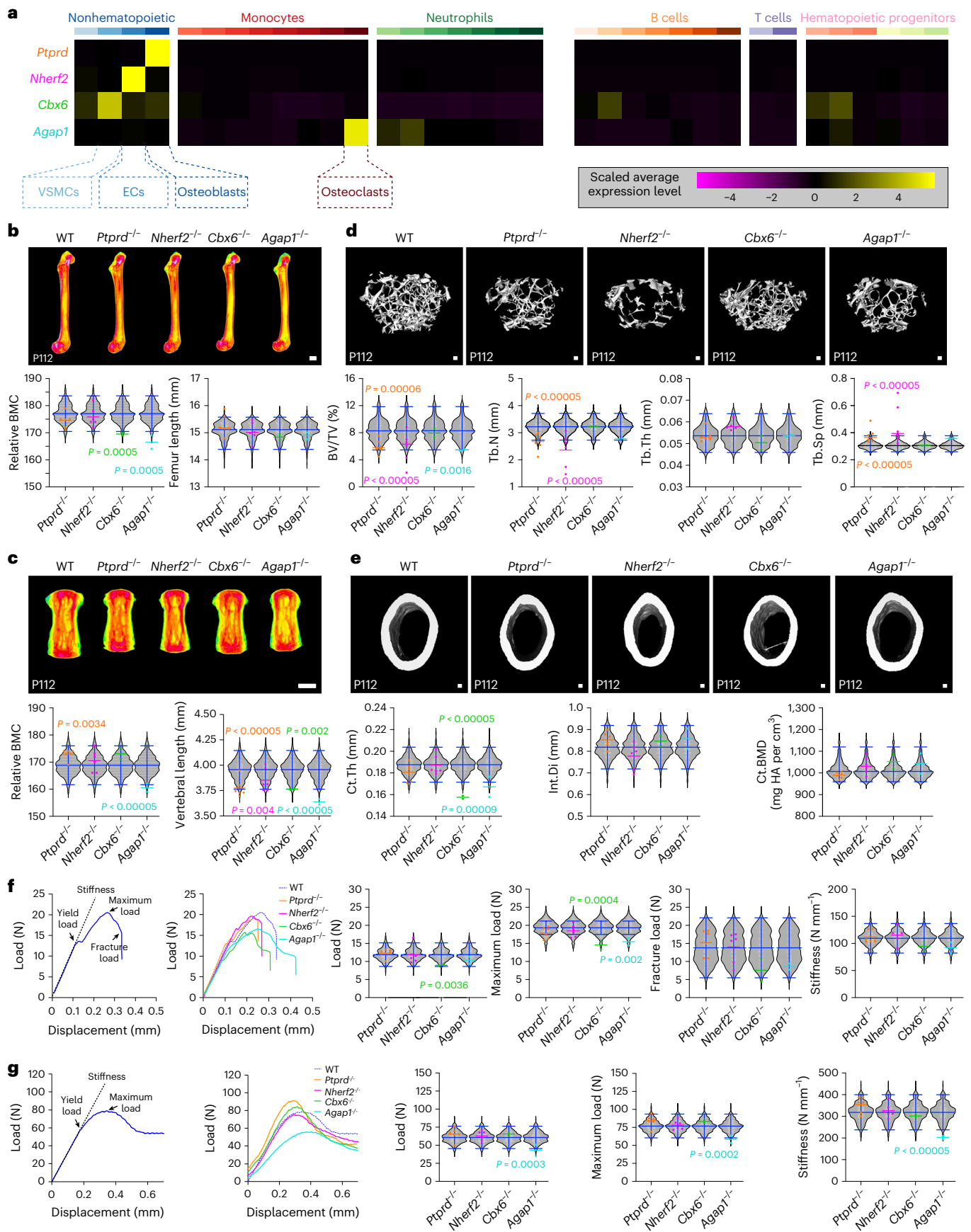
The gene programs that define human MSCs, pre-osteoblasts, mature osteoblasts, chondrocytes and fibroblasts were enriched for genes that cause monogenic skeletal disorders and associated with eBMD (Fig. 8c,d and Supplementary Tables 6 and 7). Human VSMCs and ECs were also enriched for eBMD-associated genes (Fig. 8d); 82–100% of eBMD effector genes that we validated in mice were also found in their cellular counterparts in human bone (Fig. 8e). This included *PLS3* in MSCs, osteoblasts, SMCs, pericytes and ECs; *PTPRD* in MSCs and osteoblasts; *CBX6* in SMCs and pericytes; and *NHERF2* in ECs (Fig. 8e).

To validate our human scRNA-seq data we interrogated a spatial transcriptomic dataset of adult human bone⁵⁰. In addition to

Fig. 7 | Skeletal phenotypes of female *Ptprd*^{−/−}, *Cbx6*^{−/−}, *Nherf2*^{−/−} and *Agap1*^{−/−} mice.

a, Heatmaps of *Ptprd*, *Nherf2*, *Cbx6* and *Agap1* expression in the endosteal bone and BM scRNA-seq dataset. **b,c**, Pseudocolored X-ray microradiography of femurs (**b**) and caudal vertebrae (**c**) from P112 WT and KO mice. Graphs show relative BMC (mean $\pm$ s.d.) and length (femur: median $\pm$ 2.5th and 97.5th percentiles; vertebrae: mean $\pm$ s.d.). WT reference ranges ($n = 320$) are gray violins with individual white dots. Individual values and mean or median for *Ptprd*^{−/−} (orange, $n = 6$), *Nherf2*^{−/−} (pink, $n = 6$), *Cbx6*^{−/−} (green, $n = 2$) and *Agap1*^{−/−} (cyan, $n = 2$) are shown. **d,e**, Micro-CT images and quantification of distal femur trabecular bone (**d**) (BV/TV, Tb.N, Tb.Th: mean $\pm$ s.d.; Tb.Sp: median $\pm$ 2.5th and

97.5th percentiles) and mid-shaft cortical bone (**e**) (Ct.Th, Int.Di: mean $\pm$ s.d.; Ct.BMD: median $\pm$ 2.5th and 97.5th percentiles). **f,g**, Example (left) and representative (right) load displacement curves from femur three-point bending (**f**) and caudal vertebrae compression (**g**) for WT (dotted blue), *Ptprd*^{−/−} (orange), *Nherf2*^{−/−} (pink), *Cbx6*^{−/−} (green) and *Agap1*^{−/−} (cyan). Graphs show yield load (mean $\pm$ s.d.), maximum load (mean $\pm$ s.d.), fracture load (median $\pm$ 2.5th and 97.5th percentiles) and stiffness (mean $\pm$ s.d.). For all graphs, significant *P* values after permutation testing as detailed in the Methods are indicated. **b,c**, Scale bar, 1 mm. **d,e**, Scale bar, 100 μ m.



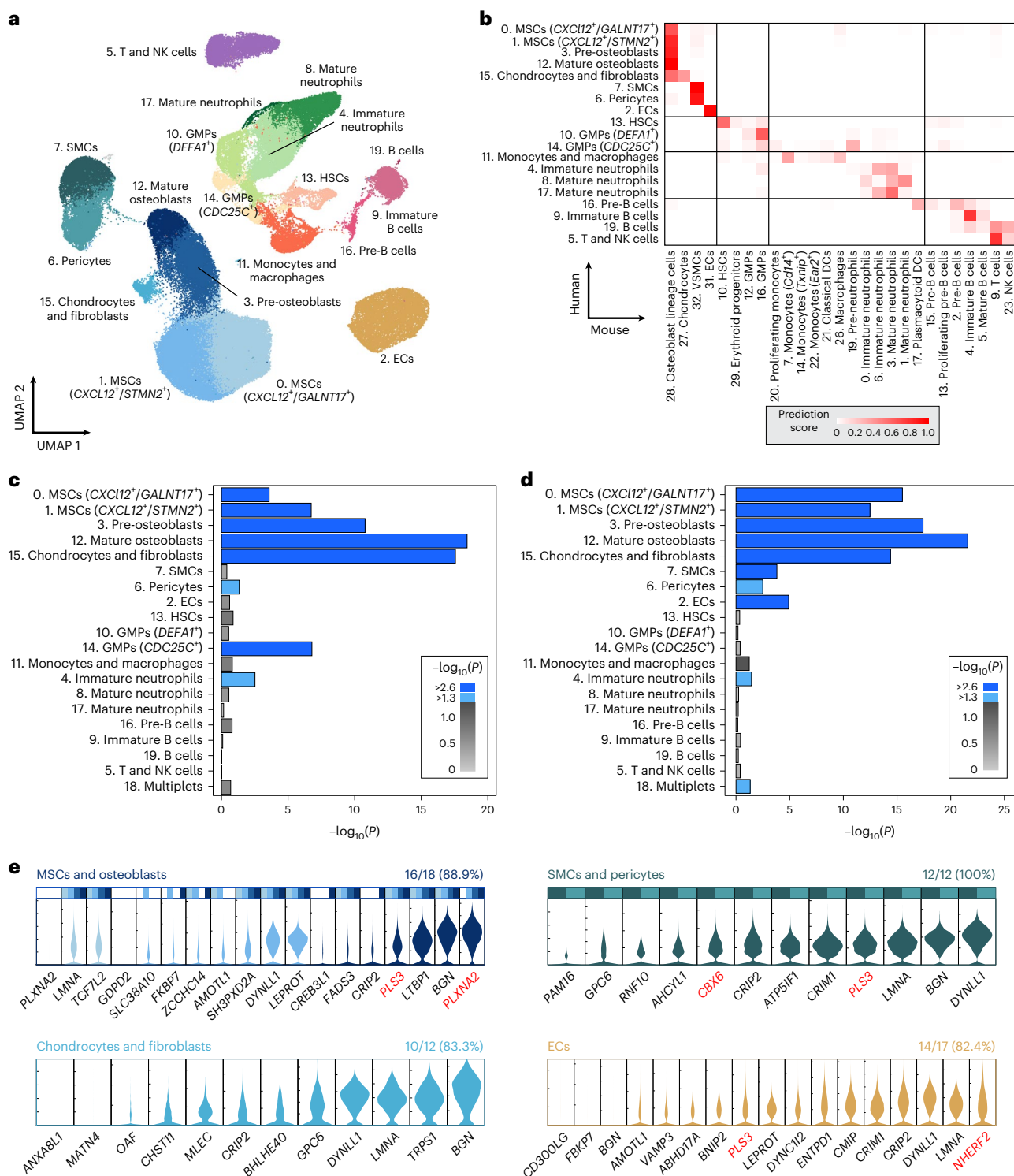


Fig. 8 | Mapping the cell types and genes involved in skeletal diseases in human bone. **a**, UMAP plot derived from scRNA-seq data of cells isolated from human femoral head. **b**, Label transfer analysis using mouse scRNA-seq data from Fig. 1 as a reference. Prediction scores reflect the probability assigned to a transferred label for a given query cell cluster. **c,d**, Bar plots showing enrichment of gene programs for causative genes of monogenic skeletal disorders (**c**) and eBMD-associated genes (**d**). The scale bars in the bar plots indicate the P value. The light blue bars correspond to observations that have nominal evidence of enrichment: P value < 0.05 ($-\log_{10}(P) > 1.3$). The dark blue bars correspond to observations that meet the Bonferroni-corrected significance threshold: $P < 2.5 \times 10^{-3}$ ($-\log_{10}(P) > 2.6$) as determined using a one-tailed Fisher's exact test

under the hypergeometric distribution (**c**) or MAGMA GSA using a one-sided test for positive association (**d**). **e**, Violin plots showing the expression of exemplar genes (cell-type-specific genes that are eBMD-associated and functionally validated in mice) in human bone. Where multiple clusters are present for a lineage, a single violin is shown corresponding to the cluster with the highest expression, with the bars above indicating which clusters pass the threshold for active expression of the gene (Methods), both colored according to the UMAP in **a**. The numbers above the plots for each lineage indicate the proportion of genes that pass the threshold for active expression. Selected exemplars from Figs. 5 and 7 are highlighted in red.

osteoblasts, chondrocytes, cells of the vasculature and osteoclasts, this also identified a population of adipocytes (Extended Data Fig. 10a,b and Supplementary Table 2). The gene programs that define osteoblasts, chondrocytes and osteoclasts were all enriched for genes that caused monogenic skeletal disorders and eBMD-associated genes (Extended Data Fig. 10c,d and Supplementary Tables 6 and 7). In common with the scRNA-seq data, VSMCs were also enriched for eBMD-associated genes (Extended Data Fig. 10d and Supplementary Table 7). Adipocytes were not enriched for monogenic skeletal disorders genes or eBMD-associated genes. Together this demonstrated our approach has important translational potential.

A platform for data interrogation

Our analysis provides a systematic approach to identify the cells and genes with functional roles in the skeleton. We developed a platform (www.musculoskeletal-genomics.org) that integrates these data and enables users to identify effector genes and define the cells in which they may function.

Discussion

Advances in understanding the pathophysiology of many diseases has benefited from single-cell technologies. Despite the burden of skeletal disease, the challenges of working with mineralized tissues has meant that skeletal research has lagged behind other fields. We addressed this by developing single-cell methodologies to define the critical endosteal compartment that regulates bone turnover. We then exploited cross-species analysis to identify new cellular and genetic determinants of bone structure and function.

This study identified the cells within the endosteal compartment and defined the gene programs that control them. Cell types enriched for rare monogenic human disorder genes and genes associated with eBMD were considered disease-relevant and investigated for evidence of functional involvement. Our analysis validated the key role of osteoblast, chondrocyte and monocyte/macrophage/osteoclast lineages but also identified a role for underappreciated cell types, including ECs and VSMCs. Many of these cell types were localized to the endosteal compartment illustrating that this is a distinct location within bone.

A role for ECs and VSMCs is exemplified by the finding of a new vascular phenotype in the cortical bone of *Pls3*^{y/-} mice and during *pls3*-deficient zebrafish embryo development. The strength of our approach is highlighted by the fact that mutations in *PLS3* are known to cause X-linked osteoporosis, but the cellular and molecular mechanisms have remained elusive³⁸. Furthermore, most genes defining disease-relevant cells have underappreciated roles in skeletal pathophysiology. We demonstrate causation for exemplar genes by characterizing skeletal phenotypes in knockout (KO) mice. Our results suggest that pathological mechanisms may be lineage-restricted, involve contributions from multiple cell types or may reflect complex systemic effects. To investigate further, we characterized the interactions between cells and showed that ECs and VSMCs can interact directly with osteoblast lineage cells. This provides a potential explanation for how genes expressed in ECs and VSMCs, such as *Pls3*, *Nherf2* and *Cbx6*, regulate bone.

Our approach generated a comprehensive map of cell types across the endosteal compartment and expands the repertoire of cells that regulate bone to include ECs and VSMCs. This complements analyses of specific bone and marrow cell populations^{51–54}.

Our study has limitations. We focused on the endosteal compartment, and this does not include the periosteum. Nevertheless, mesenchymal cells from these locations are similar⁵⁵. The single-cell methodology may not capture fragile cells such as adipocytes, osteoclasts and cells embedded within bone, such as mature osteocytes. Of note, some cells can fragment during disaggregation and remnants can be associated with unrelated cells^{56,57}. Despite this, cells with canonical

osteoclast markers, probably mononuclear osteoclasts or osteomorphs were identified²⁵. Early osteocytes were also isolated, which probably represent cells before entombment. Moreover, the application of spatial transcriptomics validated our approach and enabled identification of adipocytes, although these cells were not enriched for disease-causing genes. We only studied the endosteal compartment in the femur and while different bones have different developmental patterns, osteoblastic cells and osteocytes from different bones express similar gene programs^{22,58}. The eBMD GWAS was restricted to individuals with European ancestry and has not been replicated in an independent cohort. Finally, while we phenotyped more than 1,000 KO mouse lines, this did not include all candidate genes. Moreover, because of the scale, outlier phenotypes were determined by comparison of 2–6 mice from each KO line with data from 320 WT littermates. Outliers were independently identified using permutation testing and all lines selected for further study replicated the phenotype, confirming the strength of this approach. Future studies could explore cell specificity and sexual dimorphism using inducible conditional gene targeting.

In summary, our approach integrates genetic, transcriptional and functional data from humans, mice and zebrafish, and identifies a new cellular and molecular framework through which to understand skeletal physiology and disease. These data have implications for skeletal development, rare bone disorders, aging and skeletal responses to systemic diseases, including malignancy and inflammation. This foundational knowledge will help identify therapeutic targets to treat musculoskeletal diseases.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-026-02640-9>.

References

- Dieleman, J. L. et al. US Health Care Spending by Payer and Health Condition, 1996–2016. *JAMA* **323**, 863–884 (2020).
- Wu, A.-M. et al. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev.* **2**, e580–e592 (2021).
- Ayub, N. et al. The treatment gap in osteoporosis. *J. Clin. Med.* **10**, 3002 (2021).
- Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic evidence on clinical success. *Nature* **629**, 624–629 (2024).
- Rivadeneira, F. & Mäkitie, O. Osteoporosis and bone mass disorders: from gene pathways to treatments. *Trends Endocrinol. Metab.* **27**, 262–281 (2016).
- Zhu, X., Bai, W. & Zheng, H. Twelve years of GWAS discoveries for osteoporosis and related traits: advances, challenges and applications. *Bone Res.* **9**, 23 (2021).
- Kemp, J. P. et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat. Genet.* **49**, 1468–1475 (2017).
- Morris, J. A. et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019).
- Kim, S. K. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PLoS ONE* **13**, e0200785 (2018).
- Aguet, F. et al. Molecular quantitative trait loci. *Nat. Rev. Methods Primers* **3**, 4 (2023).
- Mullin, B. H. et al. Characterisation of genetic regulatory effects for osteoporosis risk variants in human osteoclasts. *Genome Biol.* **21**, 80 (2020).

12. Grundberg, E. et al. Population genomics in a disease targeted primary cell model. *Genome Res.* **19**, 1942–1952 (2009).
13. Al-Barghouthi, B. M. et al. Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for human bone mineral density GWAS associations. *eLife* **11**, e77285 (2022).
14. Yazar, S. et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
15. Bandyopadhyay, S. et al. Mapping the cellular biogeography of human bone marrow niches using single-cell transcriptomics and proteomic imaging. *Cell* **187**, 3120–3140 (2024).
16. Rood, J. E., Maartens, A., Hupalowska, A., Teichmann, S. A. & Regev, A. Impact of the Human Cell Atlas on medicine. *Nat. Med.* **28**, 2486–2496 (2022).
17. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
18. Chinwalla, A. T. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
19. Brommage, R. & Ohlsson, C. High fidelity of mouse models mimicking human genetic skeletal disorders. *Front. Endocrinol.* **10**, 934 (2020).
20. Kaya, S., Schurman, C. A., Dole, N. S., Evans, D. S. & Alliston, T. Prioritization of genes relevant to bone fragility through the unbiased integration of aging mouse bone transcriptomics and human GWAS analyses. *J. Bone Miner. Res.* **37**, 804–817 (2022).
21. Doolittle, M. L., Khosla, S. & Saul, D. Single-cell integration of BMD GWAS results prioritize candidate genes influencing age-related bone loss. *JBM Plus* **7**, e10795 (2023).
22. Youtlen, S. E. et al. Osteocyte transcriptome mapping identifies a molecular landscape controlling skeletal homeostasis and susceptibility to skeletal disease. *Nat. Commun.* **12**, 2444 (2021).
23. Unger, S. et al. Nosology of genetic skeletal disorders: 2023 revision. *Am. J. Med. Genet. A* **191**, 1164–1209 (2023).
24. Batoon, L. et al. CD169⁺ macrophages are critical for osteoblast maintenance and promote intramembranous and endochondral ossification during bone repair. *Biomaterials* **196**, 51–66 (2019).
25. McDonald, M. M. et al. Osteoclasts recycle via osteomorphs during RANKL-stimulated bone resorption. *Cell* **184**, 1330–1347 (2021).
26. Filipowska, J., Tomaszewski, K. A., Niedźwiedzki, Ł., Walocha, J. A. & Niedźwiedzki, T. The role of vasculature in bone development, regeneration and proper systemic functioning. *Angiogenesis* **20**, 291–302 (2017).
27. Lane, M. D., Tang, Q.-Q. & Jiang, M.-S. Role of the CCAAT enhancer binding proteins (C/EBPs) in adipocyte differentiation. *Biochem. Biophys. Res. Commun.* **266**, 677–683 (1999).
28. Ouyang, N. et al. The transcription factor Foxc1 promotes osteogenesis by directly regulating Runx2 in response of intermittent parathyroid hormone (1-34) treatment. *Front. Pharmacol.* **11**, 592 (2020).
29. Rashid, H. et al. Sp7 and Runx2 molecular complex synergistically regulate expression of target genes. *Connect. Tissue Res.* **55**, 83–87 (2014).
30. Wang, S. et al. Regulon active landscape reveals cell development and functional state changes of human primary osteoblasts in vivo. *Hum. Genomics* **17**, 11 (2023).
31. Zhang, J. et al. Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. *Nature* **381**, 238–241 (1996).
32. Kieslinger, M. et al. EBF2 regulates osteoblast-dependent differentiation of osteoclasts. *Dev. Cell* **9**, 757–767 (2005).
33. Thottappillil, N. et al. ZIC1 dictates osteogenesis versus adipogenesis in human mesenchymal progenitor cells via a hedgehog dependent mechanism. *Stem Cells* **41**, 862–876 (2023).
34. Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature* **507**, 323–328 (2014).
35. Klingseisen, A. & Jackson, A. P. Mechanisms and pathways of growth failure in primordial dwarfism. *Genes Dev.* **25**, 2011–2024 (2011).
36. Zhou, S. et al. Converging evidence from exome sequencing and common variants implicates target genes for osteoporosis. *Nat. Genet.* **55**, 1277–1287 (2023).
37. Wang, Q. et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).
38. van Dijk, F. S. et al. PLS3 mutations in X-linked osteoporosis with fractures. *N. Engl. J. Med.* **369**, 1529–1536 (2013).
39. Döffinger, R. et al. X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF-κB signaling. *Nat. Genet.* **27**, 277–285 (2001).
40. Francis, F. et al. A gene (PEX) with homologies to endopeptidases is mutated in patients with X-linked hypophosphatemic rickets. *Nat. Genet.* **11**, 130–136 (1995).
41. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
42. Bassett, J. H. et al. Rapid-throughput skeletal phenotyping of 100 knockout mice identifies 9 new genes that determine bone strength. *PLoS Genet.* **8**, e1002858 (2012).
43. Bianco, P., Fisher, L. W., Young, M. F., Termine, J. D. & Robey, P. G. Expression and localization of the two small proteoglycans biglycan and decorin in developing human skeletal and non-skeletal tissues. *J. Histochem. Cytochem.* **38**, 1549–1563 (1990).
44. Keller, R. B. et al. Monoallelic and biallelic CREB3L1 variant causes mild and severe osteogenesis imperfecta, respectively. *Genet. Med.* **20**, 411–419 (2018).
45. Schwebach, C. L. et al. Osteogenesis imperfecta mutations in platin 3 lead to impaired calcium regulation of actin bundling. *Bone Res.* **8**, 21 (2020).
46. Chin, S. M., Unnold-Cofre, C., Naismith, T. & Jansen, S. The actin-bundling protein, PLS3, is part of the mechanoresponsive machinery that regulates osteoblast mineralization. *Front. Cell Dev. Biol.* **11**, 1141738 (2023).
47. Maus, I. et al. Osteoclast-specific Platin 3 knockout in mice fail to develop osteoporosis despite dramatic increased osteoclast resorption activity. *JBM Plus* **8**, ziad009 (2024).
48. Yorgan, T. A. et al. Mice lacking platin-3 display a specific defect of cortical bone acquisition. *Bone* **130**, 115062 (2020).
49. Wu, R. S. et al. A rapid method for directed gene knockout for screening in G0 zebrafish. *Dev. Cell* **46**, 112–125 (2018).
50. Yip, R. K. H. et al. Profiling the spatial architecture of multiple myeloma in human bone marrow trephine biopsy specimens with spatial transcriptomics. *Blood* **146**, 1837–1849 (2025).
51. Tikhonova, A. N. et al. The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222–228 (2019).
52. Baryawno, N. et al. A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. *Cell* **177**, 1915–1932 (2019).
53. Baccin, C. et al. Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
54. Zhong, L. et al. Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *eLife* **9**, e54695 (2020).
55. Nookaew, I. et al. Refining the identity of mesenchymal cell types associated with murine periosteal and endosteal bone. *J. Biol. Chem.* **300**, 107158 (2024).
56. Millard, S. M. et al. Fragmentation of tissue-resident macrophages during isolation confounds analysis of single-cell preparations from mouse hematopoietic tissues. *Cell Rep.* **37**, 110058 (2021).

57. Hume, D. A., Millard, S. M. & Pettit, A. R. Macrophage heterogeneity in the single-cell era: facts and artifacts. *Blood* **142**, 1339–1347 (2023).
58. Ayturk, U. M. et al. Single-cell RNA sequencing of calvarial and long-bone endocortical cells. *J. Bone Miner. Res.* **35**, 1981–1991 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format,

as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Ryan C. Chai^{1,2,43}, **Mischa Lundberg**^{3,4,5,43}, **Bernard Freudenthal**^{6,43}, **James T. Smith**^{1,2,43}, **Andrew P. Boughton**⁷, **Yuandan Zhang**⁸, **Kaitlyn A. Flynn**⁸, **Monika Frysz**⁹, **Alexander P. Corr**¹, **Weng Hua Khoo**¹, **Davide Komla-Ebri**⁶, **Michael R. G. Dack**⁶, **Siobhan E. Guilfoyle**⁶, **John G. Logan**⁶, **Natalie C. Butterfield**⁶, **Victoria D. Leitch**⁶, **Andrea S. Pollard**⁶, **Riikka E. Mäkitie**^{6,10,11}, **Nathaniel Bradford**¹², **Lorenzo Ramos-Mucci**¹³, **Amaia Vilas-Zornoza**¹³, **Yunshun Chen**¹⁴, **Raymond K. H. Yip**^{14,15}, **Jeremy Er**¹⁴, **Siew Zhuan Tan**¹⁵, **Michelle M. McDonald**¹⁶, **Scott E. Youtlen**¹⁷, **C. Marcelo Sergio**¹, **Ariel Castro-Martinez**¹, **Shelley G. Young**¹, **Elena Skorokhodova**¹, **David M. Evans**^{4,5,18}, **Joseph E. Powell**^{19,20}, **Christiaan A. de Leeuw**²¹, **Adam D. Ewing**⁸, **John A. Eisman**^{1,2,12,22}, **Robert D. Blank**^{1,23}, **Tri Giang Phan**^{2,24}, **International Federation of Musculoskeletal Research Societies (IFMRS) Big Data Working Group***, **Anne K. Lagendijk**^{5,25,26}, **Edwin D. Hawkins**^{14,15}, **Horng Lii Oh**¹², **Rebecca E. McIntyre**¹³, **Edith M. Hessel**¹³, **Jake P. Taylor-King**¹³, **Paul A. Baldock**¹, **Emma L. Duncan**^{27,28}, **Graham R. Williams**^{6,44}✉, **J. H. Duncan Bassett**^{6,44}✉, **Peter I. Croucher**^{1,2,44}✉ & **John P. Kemp**^{8,44}✉

¹Cancer Plasticity and Dormancy Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. ²School of Clinical Medicine, Faculty of Medicine and Health, UNSW Sydney, Sydney, New South Wales, Australia. ³Transformational Bioinformatics, Commonwealth Scientific and Industrial Research Organisation, Sydney, New South Wales, Australia. ⁴UQ Frazer Institute, The University of Queensland, Woolloongabba, Queensland, Australia.

⁵Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia. ⁶Molecular Endocrinology Laboratory, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. ⁷Department of Biostatistics and the Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. ⁸Mater Research Institute, The University of Queensland, Translational Research Institute, Woolloongabba, Queensland, Australia. ⁹Musculoskeletal Research Unit, Translational Health Sciences, Southmead Hospital, University of Bristol, Bristol, UK. ¹⁰Department of Otorhinolaryngology, Head and Neck Surgery, Helsinki University Hospital and University of Helsinki, HUS, Helsinki, Finland. ¹¹Faculty of Medicine, University of Helsinki, Helsinki, Finland. ¹²St. Vincent's Hospital, Darlinghurst, New South Wales, Australia. ¹³Relation Therapeutics, London, UK. ¹⁴Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia. ¹⁵Department of Medical Biology, The University of Melbourne, Parkville, Victoria, Australia. ¹⁶School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia. ¹⁷Department of Genetics, Yale School of Medicine, New Haven, CT, USA. ¹⁸MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. ¹⁹Translational Genomics Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. ²⁰UNSW Cellular Genomics Futures Institute, UNSW Sydney, Kensington, New South Wales, Australia. ²¹Department of Complex Trait Genetics, Center for Neurogenetics and Cognitive Research, VU University Amsterdam, Amsterdam, the Netherlands. ²²School of Medicine Sydney, University of Notre Dame Australia, Darlinghurst, New South Wales, Australia. ²³Division of Endocrinology, Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, USA. ²⁴Precision Immunology Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia. ²⁵School of Biomedical Science, The University of Queensland, St Lucia, Queensland, Australia. ²⁶UQ Centre for Cardiovascular Health and Research (CCVHR), The University of Queensland, St Lucia, Queensland, Australia. ²⁷School of Life Course & Population Sciences, Faculty of Life Sciences and Medicine, King's College London, London, UK. ²⁸Guy's and St Thomas' NHS Foundation Trust, London, UK. ⁴³These authors contributed equally: Ryan C. Chai, Mischa Lundberg, Bernard Freudenthal, James T. Smith. ⁴⁴These authors jointly supervised this work: Graham R. Williams, J. H. Duncan Bassett, Peter I. Croucher & John P. Kemp. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: graham.williams@imperial.ac.uk; d.bassett@imperial.ac.uk; p.croucher@garvan.org.au; j.kemp2@uq.edu.au

International Federation of Musculoskeletal Research Societies (IFMRS) Big Data Working Group

Cheryl L. Ackert-Bicknell²⁹, **Douglas P. Kiel**^{30,31,32}, **Fernando Rivadeneira**³³, **Jennifer J. Westendorf**³⁴, **David Karasik**³⁵, **Yuuki Imai**^{36,37}, **Ralph Müller**³⁸, **Jason Flannick**^{31,39}, **Lynda Bonewald**⁴⁰, **Noël P. Burt**³², **Jonathan H. Tobias**^{9,18}, **Carolina Medina-Gomez**³³, **Qing Wu**⁴¹, **Maria C. Costanzo**³² & **Charles R. Farber**⁴²

²⁹Department of Orthopedics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ³⁰Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. ³¹Harvard Medical School, Boston, MA, USA. ³²Broad Institute of MIT and Harvard, Cambridge, MA, US. ³³Erasmus University Medical Center, Rotterdam, the Netherlands. ³⁴Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, USA. ³⁵The Musculoskeletal Genetics Laboratory, The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel. ³⁶Department of Pathophysiology, Ehime University Graduate School of Medicine, Toon, Japan. ³⁷Division of Integrative Pathophysiology, Proteo-Science Center (PROS), Ehime University, Toon, Japan. ³⁸Institute for Biomechanics, Department of Health Sciences and Technology, ETH Zürich, Zürich, Switzerland. ³⁹Division of Genetics and Genomics at Boston Children's Hospital, Boston, MA, USA. ⁴⁰Indiana Center for Musculoskeletal Health, Indiana University, Indianapolis, IN, USA. ⁴¹Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, USA. ⁴²Department of Genome Sciences, University of Virginia, Charlottesville, VA, USA.

Methods

Full details of methods, reagents and protocols are provided in Supplementary Note 2.

Ethical approval

All studies were approved by the appropriate institutional ethics committees and comply with the ARRIVE⁵⁹ and the 2024 Declaration of Helsinki⁶⁰ guidelines.

Samples for scRNA-seq

Mouse endosteal and BM cells were isolated from femurs of 9–10-week-old WT C57BL/6J male mice obtained from Australian BioResources. Animal experiments were approved by the Garvan Institute of Medical Research Animal Ethics Committee (ARA16/01, ARA19/09 and ARA22/12). Marrow was flushed with PBS; marrow-depleted bone was digested with 2 mg ml⁻¹ collagenase A (cat. no. 10103586001, Merck) and 2.5 mg ml⁻¹ trypsin (cat. no. T1426, Merck) for 30 min at 37 °C to retrieve endosteal cells. Human femoral head samples were obtained from patients undergoing surgery for osteoarthritis at St. Vincent's Hospital Sydney. All participants provided written informed consent; the study was conducted with approval from the St. Vincent's Hospital Human Research Ethics Committee (no. 2022/ETH00475). Bone cores and fragments were excised and digested using 2 mg ml⁻¹ collagenase A or collagenase IV (cat. no. LS004186, Worthington Biochemical) and 2.5 mg ml⁻¹ dispase (cat. no. D4693 Merck) for 30 min at 37 °C to release cells. Patient and sample characteristics are shown in Supplementary Table 11.

Fluorescence-activated cell sorting enrichment

Fluorescence-activated cell sorting was performed on mouse and human samples. Viable mouse cells (4',6-diamidino-2-phenylindole (DAPI)⁻; cat. no. D1306, Thermo Fisher Scientific) negative for erythroid marker Ter119 (cat. no. 116208, BioLegend) were sorted. Viable human cells (DAPI⁻) negative for erythroid marker CD235a (cat. no. 563810, BD Biosciences) were sorted into CD45⁺ (hematopoietic; cat. no. 641408, BD Biosciences) and CD45⁻ (nonhematopoietic) populations. Gating is shown in Supplementary Note 3.

scRNA-seq and data analysis

Cells were processed using 10X Chromium and Chromium Single Cell v2 Reagent Kit (target 10,000 cells per channel), and sequenced on the Illumina NovaSeq 6000 sequencing system. Raw data were processed with Cell Ranger (v.2-7) and analyzed using Seurat (v.2-5)⁶¹. Log normalization was performed and low-quality cells were filtered (genes < 300 or mitochondrial unique molecular identifier (UMI) > 10%). Dimensionality reduction used principal component analysis (mouse: 40 principal components; human: 20 principal components) on the top 3,000 variable genes. Human data were batch-corrected using Harmony⁶². Clustering used the Louvain method on shared nearest neighbor graphs and visualized using UMAP. Cell types were annotated using canonical markers and cluster-specific gene programs (Supplementary Note 1).

Defining cell-type gene programs

Differentially expressed genes (gene programs) were identified using Seurat FindAllMarkers, requiring log₂(fold change) > 0.5 and Bonferroni-adjusted *P* < 0.05 (Supplementary Table 2). The restrictedness of gene expression was calculated by determining the number of clusters expressing a given gene. A cluster was deemed to express a gene if one or more UMIs for that gene were detected in at least 25% of cells within the cluster. Gene programs are shown in Supplementary Table 2.

Cell–cell interaction analysis

Cell–cell interactions were identified using CellPhoneDB (v.5.0.1)⁶³ with default settings. Signaling pathways were analyzed using interrogating ReactomeDB⁶⁴ for enriched terms using ReactomePA⁶⁵.

Reconstruction of cell differentiation trajectories

Differentiation trajectories were reconstructed using the Monocle package (v.2) (discriminative dimensionality reduction tree)⁶⁶ on the top 1,000 differentially expressed genes.

Comparing mouse and human scRNA-seq datasets

Label transfer analysis (Seurat v.5.3.0) was performed to compare mouse and human datasets. TransferData with 20 principal components was used; prediction scores > 0.5 indicated high-confidence transfers⁶⁷.

Identifying TFs regulating gene programs

TFs and their target genes (regulons) were identified using pySCENIC (v.0.11.2)⁶⁸. Gene regulatory network inference used GRNBoost2 and cisTarget; regulon activity was scored using AUCell.

snATAC-seq

Endosteal nuclei were isolated from mouse bone and processed using the 10X Genomics Chromium Single Cell ATAC v2 platform. Sequencing was performed on an Illumina NovaSeq 6000 (~25,000–50,000 fragments per nucleus). Raw data were processed with Cell Ranger ATAC (v.2.1.0) and analyzed using Signac (v.1.15.0)⁶⁹ and Seurat. Low-quality nuclei were filtered. Data were normalized (term frequency-inverse document frequency), reduced (singular value decomposition) and clustered (shared nearest neighbor/UMAP). Gene activity scores were computed and integrated with scRNA-seq via canonical correlation analysis. Differential accessibility was assessed using logistic regression and enriched TF-binding motifs identified using FindMotifs.

Flow cytometry validation

Flow cytometry analysis was performed to validate enrichment of endosteal populations. Endosteal cells were isolated using flushing and enzymatic digestion, stained with Zombie NIR Viability Stain (cat. no. 423106, BioLegend), B220-BV510 (cat. no. 103248, BioLegend), TCRβ-BV510 (cat. no. 118131, BioLegend), CD45-BV650 (cat. no. 103151, BioLegend), CD11b-BUV395 (cat. no. 565976, BD Biosciences), Ly6C-BUV737 (cat. no. 755201, BD Biosciences), Ly6G-APC (cat. no. 560599, BD Biosciences) and CD14-PE (cat. no. 569968, BD Biosciences) and acquired on a BD FACSymphony. Analysis was performed using FlowJo v.10.10.0 (FlowJo LLC).

Skeletal gene lists

Two foundational lists of skeletal-relevant genes were curated (Supplementary Table 3). GO Biological Processes⁷⁰: genes associated with GO biological processes related to skeletal function. MGI abnormal skeletal phenotype genes: genes identified from the MGI database⁷¹ (MP:0005390) causing abnormal skeletal phenotypes when mutated in mice.

Hypergeometric overrepresentation testing

Overrepresentation analyses were performed using a Fisher's exact test under the hypergeometric distribution (Supplementary Table 12). *P* values were adjusted via Bonferroni correction. All tests used the RITAN package⁷² and standardized human Ensembl IDs (BioMart).

Gene programs enriched for monogenic skeletal disorder genes

A list of 528 protein-coding genes derived from the International Skeletal Dysplasia Society Nosology database²³ (Supplementary Table 3) was used with Fisher's exact tests. Network plots were generated in Cytoscape (v.3.10.0)⁷³ to visualize cell-gene expression.

Genome-wide association analysis of eBMD

UKB participants with European ancestry were identified; genotypes were projected onto the first 20 principal components of the 1000 Genomes Project phase 3 (1KG) individuals using GCTA (v.1.93.2).

Principal components 1–20 of UKB and 1KG individuals were projected into three UMAP components with the following parameters: $\text{min_dist} = 0.000001$, $\text{n_components} = 3$, $\text{n_neighbors} = 35$, $\text{random_state} = 10293082$. Parameters were defined manually and informed by clustering 1KG individuals with the same ancestry. A total of 461,920 UKB participants clustered with 1KG individuals with European ancestry. Individuals were mapped back to the principal component analysis plots to ensure they co-localized with 1KG individuals with European ancestry.

As described previously⁸, eBMD was derived in 481,380 UKB participants. A total of 448,010 individuals with European ancestry and rank-based inverse normal transformed eBMD were analyzed using a linear mixed noninfinitesimal model implemented in BOLT-LMM v.2.3.4 with age, sex, genotyping array and principal components 1–20 as covariates. Variants with a minor allele frequency > 0.05% and imputation quality score > 0.3 were analyzed. Manhattan plots were generated using ggplot2 (ref. 74).

Population stratification and other latent sources of confounding were investigated using stratified LD score regression implemented in the LDSC Baseline LD model (v.1.0.1) together with the $-\text{chisq-max}$ 5,000 flag. The LDSC attenuation ratio statistic ($\text{LDSC}_{\text{RPS}} = 0.02$ (s.e. = 0.0195)) suggested that polygenicity accounted for genomic inflation ($\lambda_{\text{GC}} = 2.27$). The large single-nucleotide polymorphism (SNP) heritability of eBMD ($h^2_{\text{SNP}} = 0.39$, s.e. = 0.027) and sample size probably resulted in the LDSC intercept exceeding 1 (LD score regression = 1.08, s.e. = 0.073).

SNPTracker (v.1.0) updated reference SNP cluster IDs (rsIDs) and genomic positions from Hg19 to GRCh38 using dbSNP release 151 (ref. 75). SnpEff 5.1d⁷⁶ using the GRCh38.p14 database predicted the functional impact of genetic variants. High-impact variants were validated with the variant effect predictor⁷⁷.

eBMD-associated lead variants were identified using GCTA-COJO⁷⁸ (v.1.94.1)⁷⁹. A genome-wide significance threshold of $P < 6.6 \times 10^{-9}$ was used; variants with high collinearity ($R^2 > 0.8$) were ignored; those more than 10 MB away were assumed to be in complete linkage equilibrium. Approximately 50,000 unrelated individuals of European ancestry from the UKB were used to model LD⁷⁹.

Lead variants were binned to 'loci' using a 1-MB sliding window. Loci were deemed 'new' if located more than 1 MB from any variant in the NHGRI-EBI GWAS Catalog associated with bone density 'EFO_000392' at $P < 5 \times 10^{-8}$. Data were downloaded on 5 March 2024; the Hg38 coordinates of variants were obtained using SnpTracker. Genomic distances were calculated using BEDTools (v.2.29.2)⁸⁰. Each lead variant was annotated to the closest protein-coding gene using BEDTools and the Ensemble Genes 105 (GRCh38) dataset from BioMart⁸¹.

Gene-based tests of association

MAGMA (v.1.10) gene-based tests were conducted using the eBMD GWAS summary statistics of variants with an info score greater than 0.6 and minor allele frequency > 0.5%. LD was modeled with ~50,000 unrelated UKB individuals of European ancestry. Variants were annotated to protein-coding genes using coordinates from Ensemble Genes 105 (GRCh38). Gene annotations included variants ≤ 2 kb upstream of the transcriptional start site and ≤ 1 kb downstream of the stop site. A multi-model approach was used to derive aggregate P values (P_{Multi}) for gene-based tests. A gene-wide significance threshold was determined using Bonferroni correction: the error rate of one test divided by number of genes tested = $0.05/19,695 = P < 2.54 \times 10^{-6}$.

Enrichment analysis involving eBMD-associated genes

Hypergeometric tests were performed as described above to investigate whether rare skeletal disorder genes were overrepresented among genes located closest to the eBMD lead variants, and separately, among genes associated with eBMD (Supplementary Table 7). Hypergeometric tests were also used to determine if enrichment was attributable to genes that caused individual skeletal dysplasia groups.

Gene programs enriched for eBMD-associated genes

MAGMA competitive GSA was used to determine whether gene programs were enriched for eBMD-associated genes. GSA accounted for gene size, gene density and the inverse of the mean minor allele count in the gene, as well as the log value of these three factors. The statistical significance threshold was determined using Bonferroni correction: the error rate of a single test divided by the number of gene programs in each experiment (Supplementary Table 7). A post-hoc permutation analysis was conducted using MAGMA. Pairwise conditional GSA was conducted to account for genes that were shared between gene programs of enriched cell clusters and cell subclusters. A conditional P value < 0.05 was used to reject the null hypothesis of no enrichment after correcting for shared genes.

Pulse rate as a negative control for gene enrichment

We analyzed pulse rate as a negative control in the same individuals from the UKB using the same workflow.

Enrichment in the MGI database

Hypergeometric tests were used to investigate whether genes regulating bone structure were overrepresented among gene programs of cell types that were enriched for monogenic disorder genes or eBMD-associated genes. Genes causing bone structural phenotypes when mutated in mice were identified from the MGI database (MP:0003795, accessed 21 May 2024 (ref. 71)). A total of 1,347 protein-coding genes and genes with human orthologs were extracted using BioMart implemented in R⁸¹ (Supplementary Table 3).

To determine whether genes more highly expressed in each gene program were more strongly enriched for genes that regulate bone structural integrity, we allocated genes to six nested groups based on the magnitude of fold change in gene expression. Enrichment analysis using all genes, and only eBMD-associated genes from each group, was performed.

Genetically modified mice

Skeletal phenotyping was undertaken at Imperial College (PPL70/8785 and PPL1540664) and the Wellcome Trust Sanger Institute (WTSI) Mouse Genetics Project (MGP) as part of the International Mouse Phenotyping Consortium (IMPC) (PPL nos. 80/2485 and P77453634).

Pls3 KO mouse line

C57BL/6N mice carrying a $Pls3^{tm1a(EUCOMM)Wtsi}$ KO first allele (MGI:104807) were obtained from the WTSI EMMA mouse repository and rederived from frozen embryos. The skeletal phenotype was studied in hemizygous male ($Pls3^{y/+}$) mice compared to WT ($Pls3^{y/+}$) male littermates, and in female $Pls3^{-/-}$ mice.

WT and mutant mice generated by the IMPC

Samples from 16-week-old WT and genetically modified mice on C57BL/6 backgrounds were generated by the WTSI MGP. All mice underwent a primary phenotype screen including X-ray skeletal survey and biochemical measures of mineral metabolism.

OBCD bone phenotyping

Female 16-week-old mice from WTSI MGP lines were euthanized and phenotyped to determine 19 parameters of bone mass and strength^{78,42}. Phenotype definitions were based on structural (difference in X-ray or micro-CT parameters), functional (difference in biomechanical strength), structural and functional (difference in both), bone quality (functional outlier not correlating with BMC) or Mahalanobis (significant outlier because of small differences in multiple parameters).

Digital X-ray microradiography

Digital X-ray images were recorded for fixed hindlimbs and caudal vertebrae (Ca6, Ca7) at 10- μm resolution using a Faxitron MX20 variable kV point projection X-ray source and digital imaging system (Qados,

Cross Technologies) to determine relative BMC, a two-dimensional parameter similar to areal BMD⁸².

Micro-CT

The three-dimensional cortical and trabecular structural parameters of femurs were scanned using a SCANCO μ CT50 (SCANCO Medical) at 70 kV, 200 μ A with a 0.5-mm aluminum filter. Cortical parameters were calculated from 10- μ m voxel resolution scans in the mid-shaft (ROI centered at 56% length). Trabecular parameters were calculated from 5- μ m voxel scans of a 1-mm region proximal to the distal femoral growth plate. For male *Pls3*^{W/-} mice, femurs and L5 lumbar vertebrae were imaged. High-resolution 1- μ m scans were performed on the mid-shaft to determine cortical porosity and canal diameter²². Analysis was performed using the SCANCO Medical software suite. Quantification of vascular and osteocyte lacunae was performed on P112 males. Volume distributions were compared using Kolmogorov–Smirnov analysis.

Biomechanical testing

Destructive three-point bend tests were performed on femurs; two-point compression tests were performed on caudal vertebrae (Ca6, Ca7) using an Instron 5543 as described^{7,83}. Load displacement curves were used to calculate yield, maximum and fracture loads, and stiffness.

Histology and histomorphometry

Whole-mount skeletal staining of P1 neonates was performed using Alizarin red (mineralized bone) and Alcian blue (cartilage)⁸⁴.

Growth plate histomorphometry was conducted on Alcian-blue-stained and van Gieson-stained sections of decalcified tibias⁸⁵, measuring the width of the reserve, proliferative and hypertrophic zones using ImageJ.

Osteoclast static histomorphometry was performed on decalcified proximal humeri stained for tartrate-resistant acid phosphatase⁸⁶, and histomorphometry analysis performed using the TrapHisto software (www.liverpool.ac.uk/ageing-and-chronic-disease/bone-hist/trap-hist/) in a 750 × 750 μ m area 250 μ m distal to the proximal humeral growth plate^{86,87}.

Osteoblast dynamic histomorphometry was performed on mice double-labeled with calcein (cat. no. C0875, Sigma-Aldrich) at 6 and 2 days before collection. Calcein fluorescence was imaged using confocal scanning light microscopy on methacrylate-embedded femurs and lumbar vertebrae to quantify total bone surfaces, labeled surfaces and separation between double-labels according to the American Society for Bone and Mineral Research nomenclature^{87,88}.

CD31 immunohistochemistry was performed on decalcified P21 femurs using a primary rabbit anti-CD31 antibody (cat. no. ab182981, Abcam) to quantify vascular canals in cortical bone.

Iodine-contrast-enhanced backscattered electron-SEM

Formalin-fixed P70 male femurs were embedded in polymethyl methacrylate and stained with Lugol's Iodine solution. Blocks were carbon-coated and imaged using SEM (Tescan) at high vacuum using a four-quadrant backscattered electron detector. Cortical bone ROIs were analyzed blind using ImageJ to determine the total cortical area and the number and area of vascular canals.

Serum analysis

Serum markers for bone resorption (CTX) and formation (PINP) were determined using enzyme-linked immunosorbent assay (AC-06F1 and AC-33F1, Immunodiagnostic Systems) at P70. Wnt inhibitors (DKK-1, SOST; cat. nos. MKK100 and MSST00, R&D Systems) were determined at P183.

Statistical analysis of skeletal phenotypes

IMPC lines were compared to WT reference ranges from two different genetic backgrounds (320 C57BL/6N and C57BL/6NTac mice; 80

C57BL/6Brd-Tyrc-Brd and C57BL/6Dnk mice) using reference range analysis and permutation testing. Multivariate analysis was performed using squared robust Mahalanobis distances (MDi2) to identify significant abnormal phenotypes resulting from simultaneous small variances across the 19 skeletal parameters. Bone quality outliers were identified using 95% prediction intervals from BMC versus strength linear regression. *Pls3*^{W/-} data were compared using an unpaired Student's *t*-test (*P* < 0.05). The Kolmogorov–Smirnov test was used to compare cumulative frequency distributions of BMC^{42,89}.

Zebrafish husbandry and genome editing

Zebrafish experimentation adhered to guidelines of the animal ethics committee of the University of Queensland (permit no. 2022/AE000091). The *Tg(kdrl:EGFP)*^{ss43} zebrafish transgenic line was used for all studies⁹⁰. Embryos were obtained through natural paired matings and incubated at 28 °C in a dark-phase incubator. Embryos were maintained in 10-cm Petri dishes containing 1× E3 media at a maximum density of *n* = 60. At 24 hpf, all embryos were changed to E3 medium supplemented with 0.0003% phenylthiourea to prevent pigmentation.

Genome editing of *pls3* mutants in zebrafish

G0 mosaic crispants were generated by injecting ~1 nl of ribonucleoprotein mix containing Alt-R Cas9 Nuclease and three CRISPR RNAs (details in Supplementary Note 2 and Supplementary Table 13) targeting *pls3* at the one-cell stage.

Confocal imaging

Live zebrafish embryos were mounted laterally and imaged on a Zeiss LSM 710 confocal microscope. ISV luminal diameter was quantified using a VasoMetrics macro in ImageJ⁹¹.

Non-osseous tissue gene expression analysis

Gene expression across non-osseous tissues was interrogated using processed scRNA-seq data from the TabulaMuris⁹² and TabulaSapiens⁹³ databases. Cell clusters with fewer than ten cells were excluded. Remaining clusters were manually assigned to a cell group using the Human Protein Atlas classification system. Mean raw expression of selected exemplar genes within each cluster was calculated and visualized.

Statistics and reproducibility

Details for individual experiments are included in each corresponding section in the Methods.

For the scRNA-seq analysis, no statistical method was used to predetermine sample size. Cells were isolated from five independent experiments, each with five biological replicates and combined for subsequent analysis. For snATAC-seq, a single experiment was performed.

For the genetic association analyses, participants were excluded if they had low-quality genomic and outcome data or were missing outcomes and covariates pertinent to the association testing. Obvious outliers were manually filtered by observing the distributions of the data and removing individuals far exceeding the tail ends of the data. Low-quality SNPs were excluded from the analysis in accordance with standard quality control criteria. Multiple-hypothesis correction was performed using Bonferroni correction.

In the OBCD mouse KO studies, sample size was determined as detailed in Supplementary Note 2. The coefficient of variation for each of the OBCD skeletal phenotyping parameters was defined using 400 WT mouse samples from two different genetic backgrounds. Power calculations indicate an 80% power to detect an outlier phenotype of greater or equal to 2 s.d. with a sample size of two for each line in the OBCD screen. Data for all mice were included, including those with and without skeletal phenotypes; no data were excluded. For the analysis of *Pls3*^{W/-} mice, mice with gene deletion were compared to littermate controls. Power calculations using WT data together with the magnitude

of the skeletal phenotype identified in female *Pls3*^{-/-} mice in the OBCD screen ($n = 2$) demonstrated that $n = 7-8$ was required. Investigators were blinded to genotypes during analysis of skeletal phenotypes. Unblinding occurred at the point of statistical analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The scRNA-seq and snATAC-seq raw data are available on the Gene Expression Omnibus (GEO) under accession no. [GSE317069](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE317069). Human spatial transcriptomic data generated by Yip et al.⁵⁰ are available on the GEO under accession no. [GSE299207](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE299207). Genome-wide association summary result statistics are available via the NHGRI-EBIGWAS Catalog (<https://www.ebi.ac.uk/gwas>) under accession no. [GCST90726625](https://www.ebi.ac.uk/gwas/studies/GCST90726625). The interactive web portal can be accessed at www.musculoskeletal-genomics.org. Source data are provided with this paper.

Code availability

All single-cell transcriptomic and genomics analyses were performed using publicly available software as described in the Methods and Supplementary Note 2. Custom scripts are available at GitHub (<https://github.com/ryacha83/Single-cell-RNA-seq-analysis-of-mouse-and-human-bone-and-genomic-analysis-of-eBMD>) and Zenodo (<https://doi.org/10.5281/zenodo.19624990>)⁹⁴.

References

59. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
60. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. *JAMA* **333**, 71–74 (2025).
61. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
62. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
63. Troulé, K. et al. CellPhoneDB v5: inferring cell–cell communication from single-cell multiomics data. *Nat. Protoc.* **20**, 3412–3440 (2025).
64. Milacic, M. et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res.* **52**, D672–D678 (2024).
65. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
66. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
67. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
68. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
69. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
70. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
71. Smith, C. L., Blake, J. A., Kadin, J. A., Richardson, J. E. & Bult, C. J. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.* **46**, D836–D842 (2018).
72. Zimmerman, M. T., Kabat, B., Grill, D. E., Kennedy, R. B. & Poland, G. A. RITAN: rapid integration of term annotation and network resources. *PeerJ* **7**, e6994 (2019).
73. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
74. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
75. Deng, J. E., Sham, P. C. & Li, M.-X. SNPTracker: a swift tool for comprehensive tracking and unifying dbSNP rs IDs and genomic coordinates of massive sequence variants. *G3* **6**, 205–207 (2015).
76. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
77. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
78. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
79. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
80. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
81. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
82. Butterfield, N. C., Logan, J. G., Waung, J., Williams, G. R. & Bassett, J. H. D. Quantitative X-ray imaging of mouse bone by Faxitron. *Methods Mol. Biol.* **1914**, 559–569 (2019).
83. Esapa, C. T. et al. Bone mineral content and density. *Curr. Protoc. Mouse Biol.* **2**, 365–400 (2012).
84. Rigueur, D. & Lyons, K. M. Whole-mount skeletal staining. *Methods Mol. Biol.* **1130**, 113–121 (2014).
85. Bassett, J. H. et al. Mice lacking the calcineurin inhibitor Rcan2 have an isolated defect of osteoblast function. *Endocrinology* **153**, 3537–3548 (2012).
86. van 't Hof, R. J., Rose, L., Bassonga, E. & Daroszewska, A. Open source software for semi-automated histomorphometry of bone resorption and formation parameters. *Bone* **99**, 69–79 (2017).
87. Dempster, D. W. et al. Standardized nomenclature, symbols, and units for bone histomorphometry: a 2012 update of the report of the ASBMR Histomorphometry Nomenclature Committee. *J. Bone Miner. Res.* **28**, 2–17 (2013).
88. Bassett, J. H. et al. Thyrostimulin regulates osteoblastic bone formation during early skeletal development. *Endocrinology* **156**, 3098–3113 (2015).
89. Bassett, J. H., van der Spek, A., Gogakos, A. & Williams, G. R. Quantitative X-ray imaging of rodent bone by Faxitron. *Methods Mol. Biol.* **816**, 499–506 (2012).
90. Beis, D. et al. Genetic and cellular analyses of zebrafish atrioventricular cushion and valve development. *Development* **132**, 4193–4204 (2005).
91. McDowell, K. P., Berthiaume, A.-A., Tieu, T., Hartmann, D. A. & Shih, A. Y. VasoMetrics: unbiased spatiotemporal analysis of microvascular diameter in multi-photon imaging applications. *Quant. Imaging Med. Surg.* **11**, 969–982 (2021).
92. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
93. The Tabula Sapiens Consortium & Quake, S. R. Tabula Sapiens reveals transcription factor expression, senescence effects, and sex-specific features in cell types from 28 human organs and tissues. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.12.03.626516> (2025).
94. Chai, R. C., Smith, J. T. & Kemp, J. P. Single cell RNA-seq analysis of mouse and human bone and genomic analysis of eBMD. *Zenodo* <https://doi.org/10.5281/zenodo.19624990> (2026).

Acknowledgements

This research has been conducted using the UKB Resource under application no. 53641. We thank E. Lam, A. Swarbrick, S. Wu, C.-L. Chan, N. Farbehi, Y. Husaini, Z. H. Lu, O. Tong, A. Rios Villamil, S. Samarford, C. O'Keefe and the staff of the Garvan Genomics Platform at the Garvan Institute of Medical Research for support with FACS, scRNA-seq and snATAC-seq; and the Australian BioResource Biological Testing Facility. This study was supported by phenotyping expertise provided by A.-T. Adoum, R. Allen, J. Bagchi-Chakraborty, K. F. Curry, H. F. Dewhurst, A. Gogakos, F. Kussy, N. Mannan, J. Miszkiewicz, H. J. Protheroe, P. C. Sparkes and V. Vancollie. We also thank A. Xue, B. Mullin, S. D'Urso, L. Scott, C. Hammond and D. Hume for their intellectual contribution, and L. Hargreaves, O. Powell, L. Elliott and the UQ and IMB ITS support teams for assistance with high-performance computing.

Author contributions

R.C.C., J.H.D.B., G.R.W., P.I.C. and J.P.K. conceived the study. R.C.C., J.H.D.B., G.R.W., P.I.C. and J.P.K. designed the experiments. R.C.C., A.C.-M., S.G.Y. and E.S. isolated cells from mouse and human bone for scRNA-seq and snATAC-seq. R.C.C., J.T.S., A.P.C. and W.H.K. performed the analyses on the scRNA-seq data. M.L. and J.P.K. processed and analyzed the GWAS data. R.C.C., M.L., J.T.S. and J.P.K. performed the integrated analysis of scRNA-seq and GWAS. M.F., Y.Z. and K.A.F. performed the additional analyses on the GWAS data. B.F., D.K.-E., M.R.G.D., S.E.G., J.G.L., N.C.B., V.D.L., A.S.P. and R.E.Mäkitie performed the skeletal phenotyping of mutant mice. J.P.K., A.P.B. and Y.Z. designed and implemented the interactive portal. Y.C., R.K.H.Y., J.E. and E.D.H. generated and analyzed the spatial transcriptomics data. S.Z.T. and A.K.L. generated and performed the vascular phenotyping of mutant zebrafish. N.B. and H.L.O. recruited patients for the human scRNA-seq. L.R.-M., A.V.-Z., H.L.O., R.E.McIntyre, J.P.T.-K., E.M.H., M.M.M., S.E.Y., C.M.S., D.M.E., J.E.P., C.A.L., A.D.E., J.A.E., P.A.B., R.D.B., T.G.P., E.L.D. and the IFMRS Big Data Working Group provided expert guidance and feedback on experimental design, analysis and results. R.C.C., M.L., J.T.S., J.H.D.B., G.R.W., P.I.C. and J.P.K. wrote the manuscript. All authors reviewed and revised the manuscript.

Funding

This work was supported by a Wellcome Trust Strategic Award (101123/Z/13/A) to G.R.W., P.I.C. and J.H.D.B. and by Mrs. Janice Gibson and the Ernest Heine Family Foundation. J.P.K. is funded by a National Health and Medical Research Council (Australia) Investigator Grant (GNT2026272) and a Project Grant (GNT1158758), the Lions Medical Research Foundation (2020 Lions Dunning-Orlich Investigator Award), the American Society of Bone and Mineral Research (2020 ASBMR Rising Star Award) and the Mater Foundation. P.I.C. and T.G.P. are supported by National Health and Medical Research Council Investigator Grants (GNT2009010, GNT1155678) and a National Breast Cancer Foundation Collaborative Research Accelerator grant (2024/

CRA0020). J.H.D.B. and G.R.W. are supported by a Medical Research Council Research Grant (MR/W018950/1), a Wellcome Trust Joint Investigator Award (110141/Z/15/Z and 110140/Z/15/Z) and a European Union H2020 THYRAGE Grant (666869). M.L. was supported by a University of Queensland Research Training Scholarship from The University of Queensland and a Postgraduate Top-Up Scholarship from the Commonwealth Scientific and Industrial Research Organisation. J.T.S. was funded by a UNSW Sydney International Postgraduate Award. A.P.C. and S.E.Y. were funded by the Australian Government Research Training Program Scholarships. R.E.Mäkitie was supported by the Finnish ORL-HNS Foundation, the Orion Research Foundation and the Juhani Aho Foundation. C.A.L. was supported by a European Research Council advanced grant (ERC-2018-ADG 834057). D.M.E. and E.D.H. are supported by fellowships from the National Health and Medical Research Council (GNT2017942, GNT2008652). D.P.K. were supported by a grant from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01 AR041398). A.K.L. is supported by an ARC Future Fellowship (FT250100621). Additional support came from the UNSW Cellular Genomics Futures Institute, Cancer Council NSW, the Leukemia Foundation, the Leukemia and Lymphoma Society (now Blood Cancer United) and Relation Therapeutics.

Competing interests

R.C.C. and P.I.C. received funding from Relation Therapeutics. L.R.-M., A.V.-Z., R.E.McIntyre, E.M.H. and J.P.T.-K. are receiving or have received compensation from Relation Therapeutics. E.M.H. is a trustee of the Kennedy Trust (London). S.E.Y. is founder and director of Compbiosphere, a computational biology analytics and consulting company. R.D.B. is on the advisory board of Amgen, consults for Guidepoint Advisors and receives editorial stipends from Elsevier, The Endocrine Society and Wolters Kluwer. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41588-026-02640-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-026-02640-9>.

Correspondence and requests for materials should be addressed to Graham R. Williams, J. H. Duncan Bassett, Peter I. Croucher or John P. Kemp.

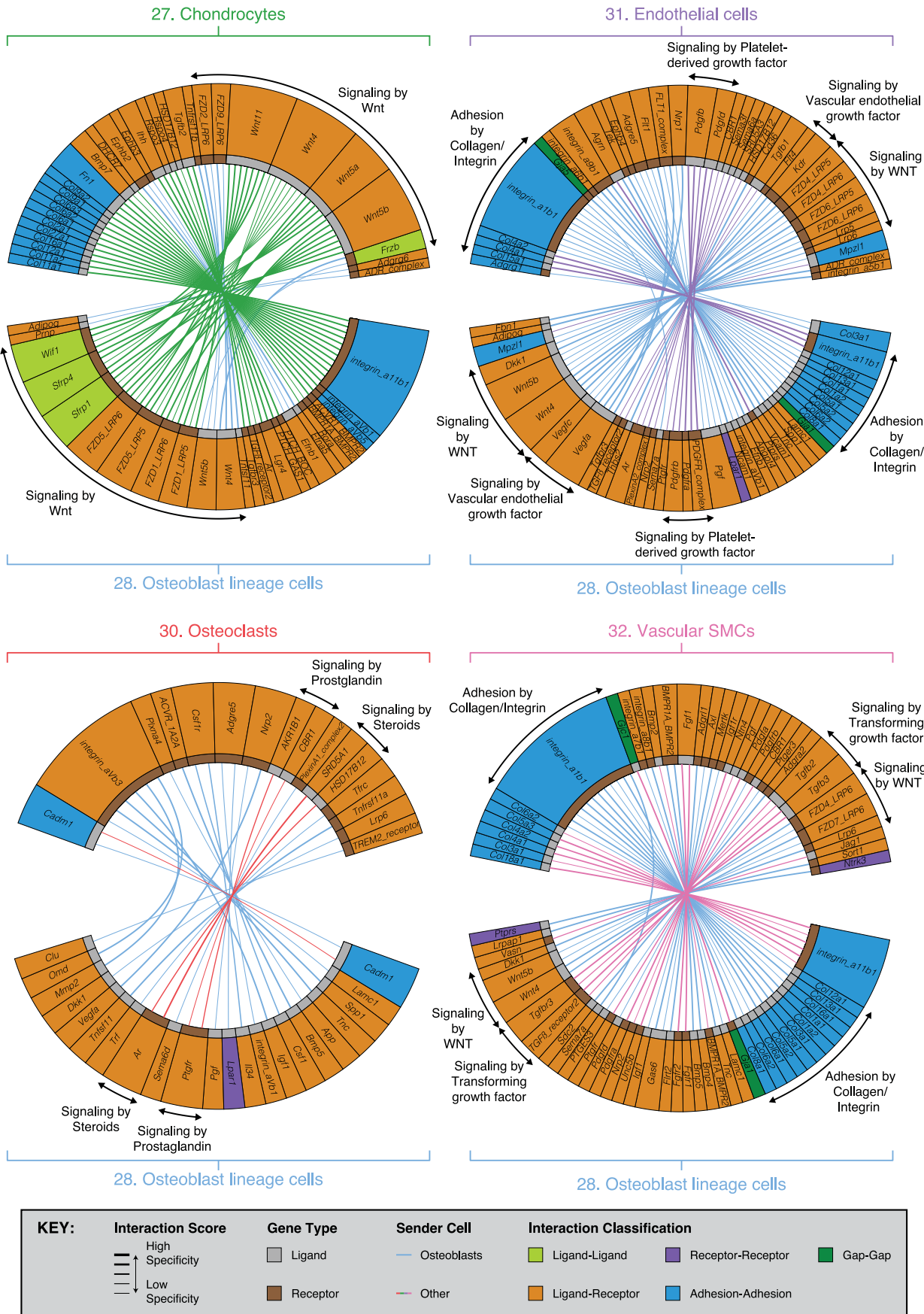
Peer review information *Nature Genetics* thanks for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Genes and gene programs that define murine chondrocytes, endothelial cells, vascular smooth muscle cells and osteoclasts. Analysis of gene programs that define the chondrocyte (cluster 27), endothelial cell (cluster 31), VSMC (cluster 32) and osteoclast (cluster 30) clusters from Fig. 1c. **(a, c, e, g)** Density plots for each cell type (chondrocyte, endothelial cell, VSMC, osteoclast, respectively). The top plot in each panel shows the distribution of the $\log_2$ fold change [$\log_2$ (FC)] in expression of genes defining the respective gene program. The bottom plot shows the total number of other

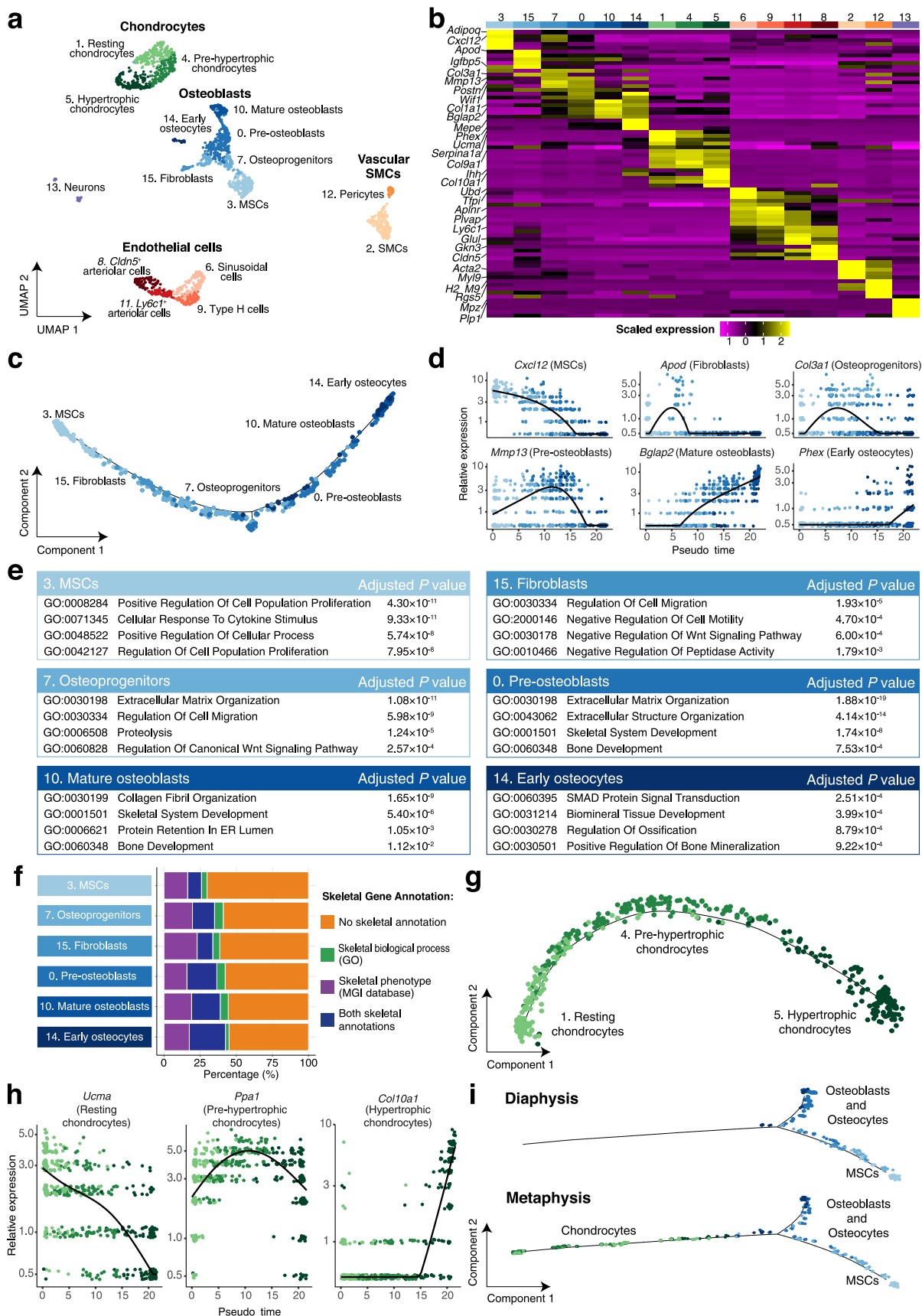
clusters expressing these genes (restrictedness). **(b, d, f, h)** Waffle plots and Circos plots for each cell type (chondrocyte, endothelial cell, VSMC, osteoclast, respectively). Waffle plots display the proportion of genes in the gene program stratified by $\log_2$ (FC) expression bins (0.5 – 1, 1 – 2, >2). Color coding indicates skeletal annotation status: GO database (green), MGI database (purple), both (blue), or unannotated (orange). Circos plots highlight individual genes with $\log_2$ (FC) > 2 in expression, with those found only in that cell type's gene program indicated in bold.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Cell-cell interaction of osteoblast lineage cells with other non-hematopoietic cells and osteoclasts. Circos plots displaying predicted interactions between osteoblast lineage cells and selected cell types. Genes are arranged according to the annotated pathways present within CellPhoneDB and

labelled accordingly in the outer layer. Interaction types are indicated by coloured boxes. Line colour indicates which cell type is the “sender” cell type for a given interaction, whilst thickness reflects interaction specificity. For clarity, only interactions with a specificity score $\geq 50\%$ are included within the plots.

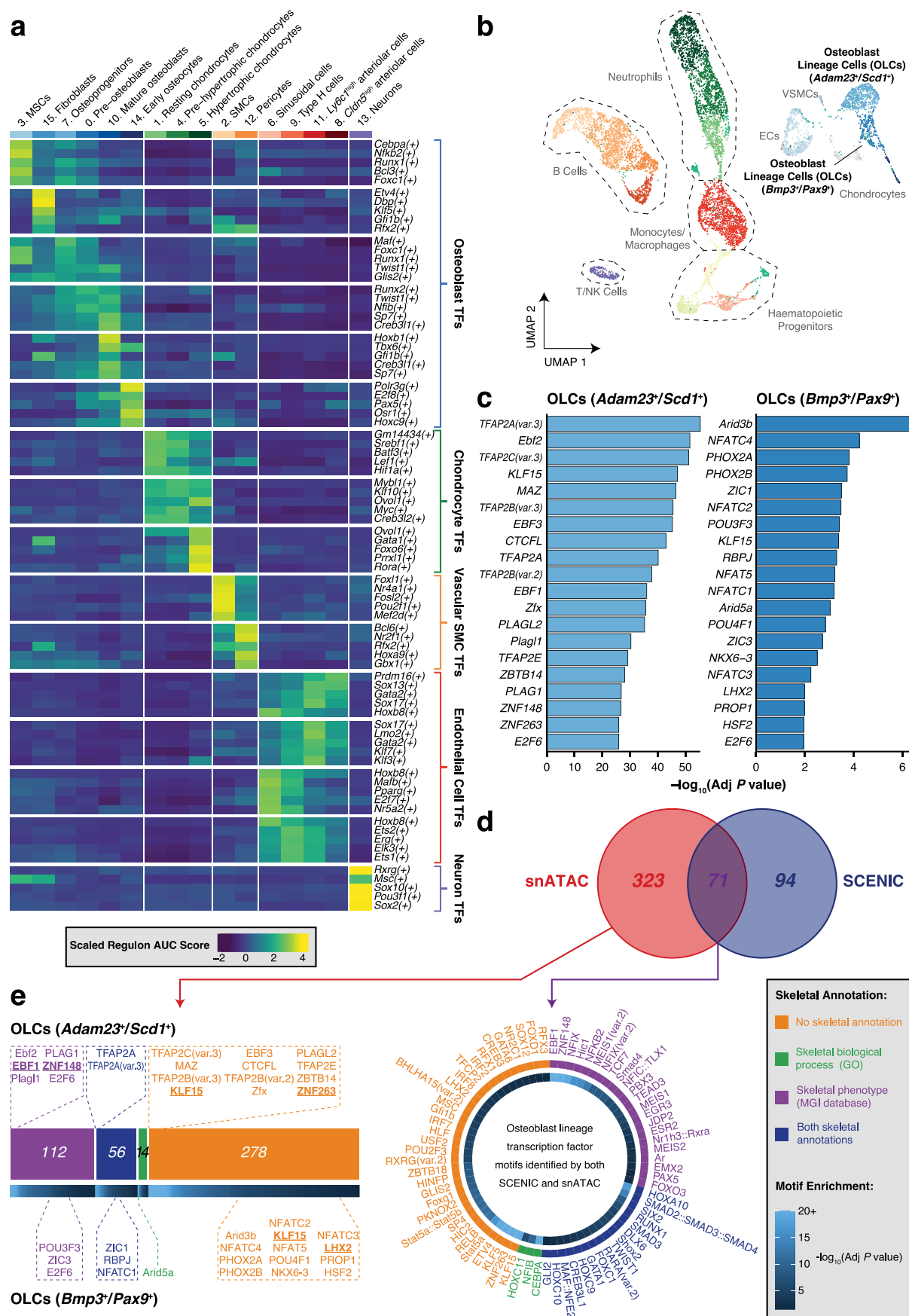


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | High-resolution clustering of non-hematopoietic cells.

(a) UMAP of cell clusters identified in osteoblast lineage, chondrocyte, endothelial cell, VSMC and neuronal cell types. (b) Heatmap of the top 10 genes expressed by each of the non-hematopoietic sub-clusters ranked by $\log_2(\text{FC})$. Two example genes that define each sub-cluster are annotated. (c) Predicted relationship between individual osteoblast sub-clusters determined by pseudo-time trajectory analysis. (d) Expression of exemplar genes of each osteoblast sub-cluster in pseudo-time. (e) Gene ontology analysis of the gene programs for each osteoblast

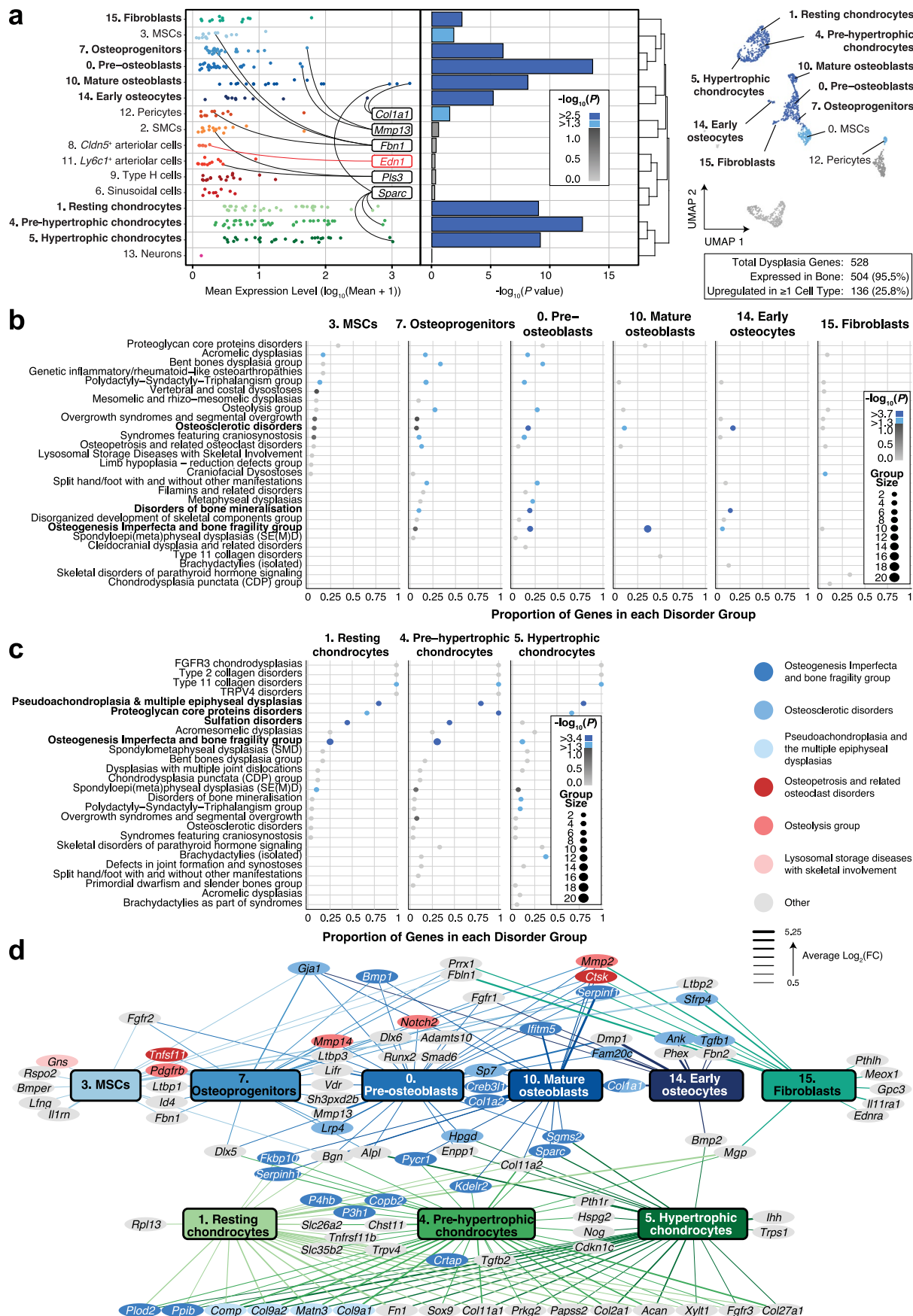
sub-cluster. (f) Bar plot showing the proportion of genes in osteoblast sub-cluster gene programs. Genes annotated with a skeletal process in the GO database (green), MGI database (purple), in both databases (blue) or are unannotated (orange) are shown. (g) Predicted relationship between individual chondrocyte sub-clusters determined by pseudo-time trajectory analysis. (h) Expression of exemplar genes of each chondrocyte sub-cluster in pseudo-time. (i) Predicted relationship between individual osteoblast and chondrocyte sub-clusters in diaphysis and metaphysis determined by pseudo-time trajectory analysis.



Extended Data Fig. 4 | See next page for caption.

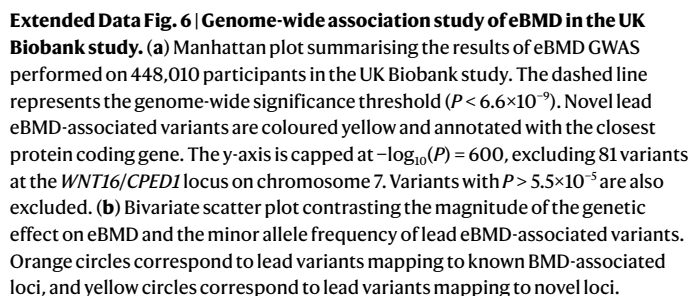
Extended Data Fig. 4 | Transcription factors implicated in regulating non-hematopoietic sub-clusters. (a) Heatmap displaying the mean activity of the target genes (regulons) for the top 5 transcription factors predicted by SCENIC to regulate each non-hematopoietic sub-cluster. Scale bar represents AUCell scores which indicate relative expression levels of the regulon across different cell types. (b) UMAP of cell clusters identified by snATAC-seq. (c) Barplots displaying the top enriched transcription factor binding motifs within differentially accessible peaks identified within the 2 osteoblast lineage cell (OLC) clusters. (d) Venn diagram showing overlap of identified transcription factors from SCENIC and snATAC-seq analyses. (e) Barplot displaying skeletal annotations of all transcription factors

identified from snATAC-seq analysis. Top transcription factors for both osteoblast lineage clusters identified from snATAC-seq analysis in (c) are annotated. Common transcription factors identified from both snATAC-seq and SCENIC analyses are highlighted in bold and underlined. Circos plot displaying known skeletal annotations of all common transcription factors identified from both methods. Motifs involving genes annotated with a skeletal process in the gene ontology (GO) database (green), the mouse genome informatics (MGI) database (purple), in both databases (blue) or are not annotated in either (orange) are shown. Motifs are ordered by strength of enrichment with snATAC-seq analysis, as indicated by the lower bar in barplot and inner circle of circos plot.

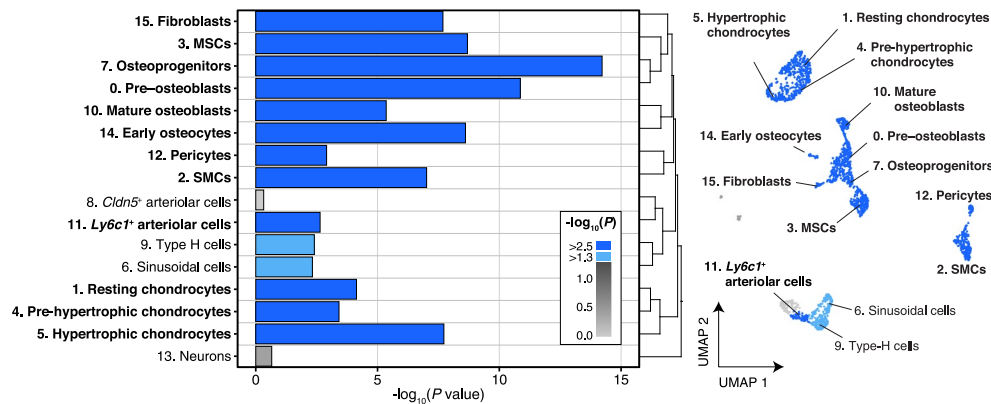


Extended Data Fig. 5 | Gene programs of non-hematopoietic cells and osteoclasts are enriched for monogenic skeletal disorder genes. (a) Dot plot displaying the mean expression level [$\log_{10}(\text{mean}+1)$] of causative genes of rare monogenic skeletal disorders within individual non-hematopoietic sub-cluster gene programs. Boxes highlight exemplar causative genes, with red text denoting genes restricted to a single gene program. Colour shading indicates the P value of enrichment as determined by one-tailed Fisher's exact test under the hypergeometric distribution: Dark blue signifies robust enrichment [$P < 3.1 \times 10^{-3}$, $-\log_{10}(P) > 2.5$] after *Bonferroni* correction, while light blue indicates nominal enrichment [$P < 0.05$, $-\log_{10}(P) > 1.3$]. (b, c) Bubble plots showing the number and proportion of causative genes within rare disorder groups

found in gene programs for (b) osteoblast sub-clusters and (c) chondrocyte sub-clusters. Circle size represents the number of genes from a disorder group present within the gene program. The scale bar and colour shading indicate the P value of enrichment as determined by one-tailed Fisher's exact test under the hypergeometric distribution. Dark blue dots denote robust *Bonferroni*-corrected enrichment [$P < 2 \times 10^{-4}$, $-\log_{10}(P) > 3.7$ for (b); $P < 4.1 \times 10^{-4}$, $-\log_{10}(P) > 3.4$ for (c)], and light blue dots denote nominal enrichment [$P < 0.05$, $-\log_{10}(P) > 1.3$]. (d) Network plot showing the unique and shared causative genes identified in the gene programs of osteoblast lineage and chondrocyte sub-clusters. Genes are colour-coded based on disorder group, and the thickness of the connecting lines indicates the magnitude of differential expression [$\log_2(\text{FC})$].

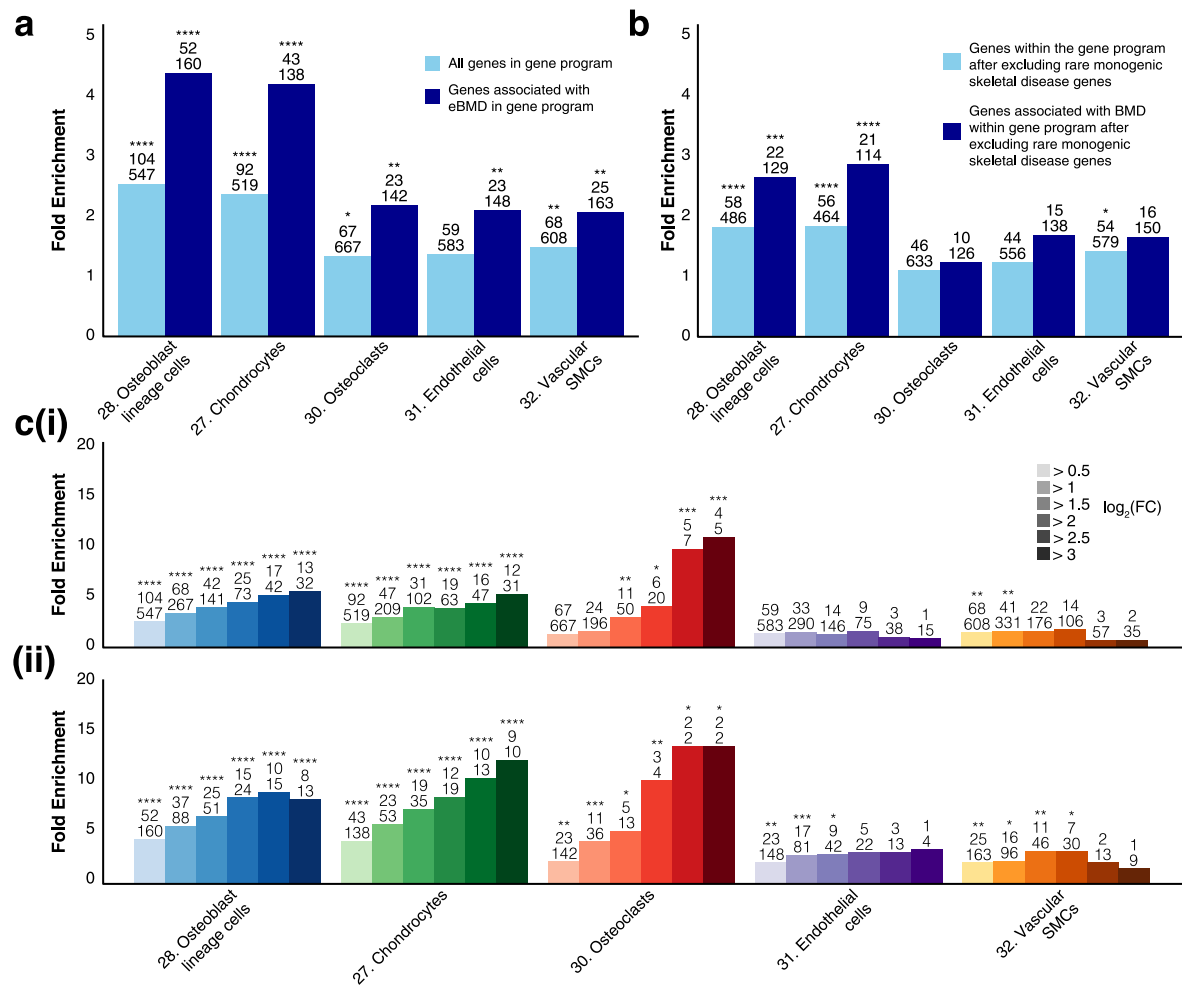


Nature Genetics



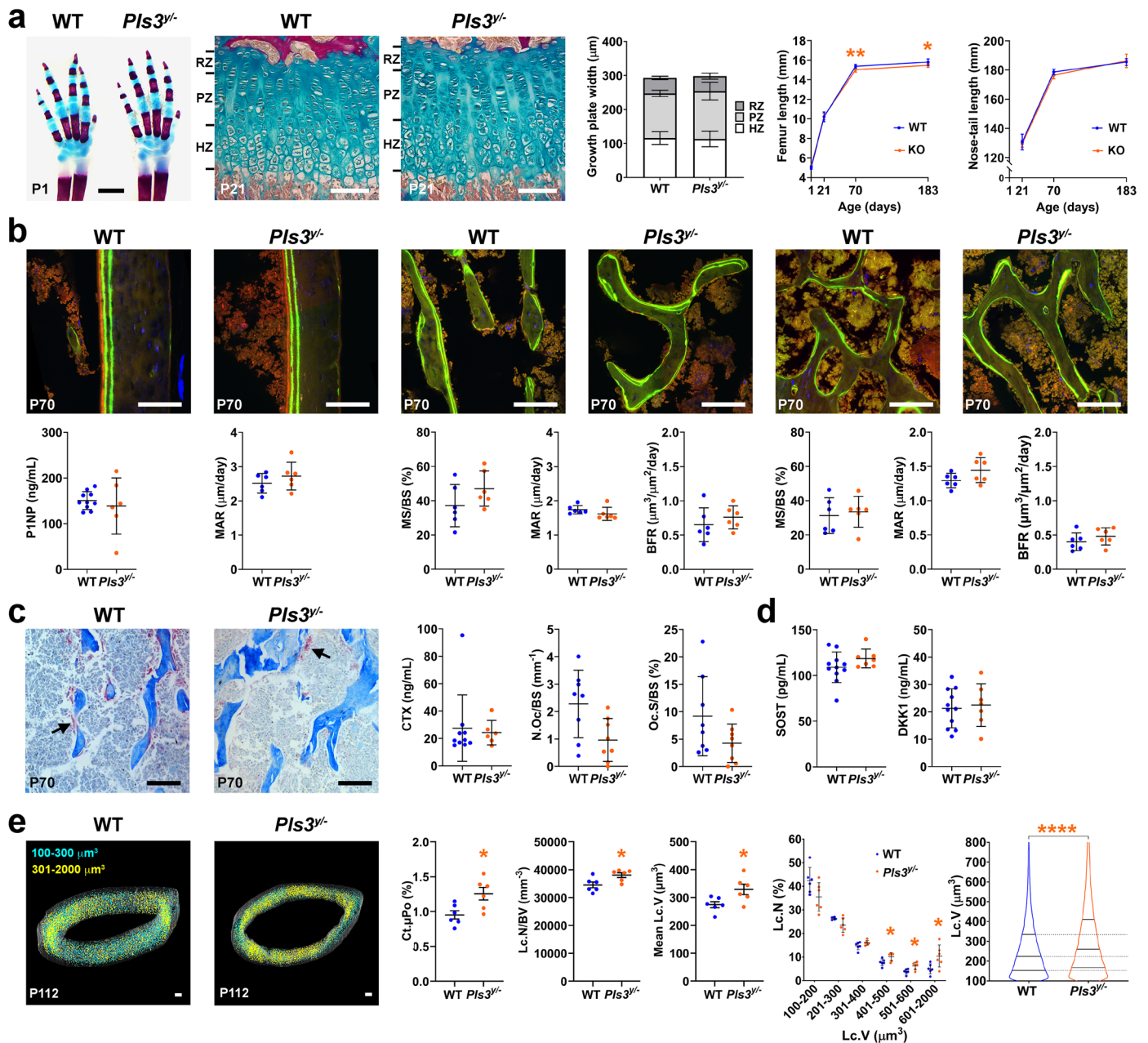
Extended Data Fig. 7 | Gene programs of non-hematopoietic cell sub-clusters are enriched for eBMD-associated genes. Bar plot and UMAP plot showing enrichment of eBMD-associated genes in gene programs for different cell types identified in scRNA-seq as determined by MAGMA competitive gene set analyses using a one-sided test for positive association. Scale bar in bar plot and

UMAP plot indicates the P value. Light blue bars in bar plot and light blue dots in UMAP correspond to observations that have nominal evidence of enrichment: $P < 0.05$ [$-\log_{10}(P)$ of >1.3]. Dark blue bars in bar plot and dark blue dots in UMAP correspond to observations that have robust evidence of enrichment and meet the *Bonferroni*-corrected significance threshold: $P < 3 \times 10^{-3}$ [$-\log_{10}(P) > 2.5$].



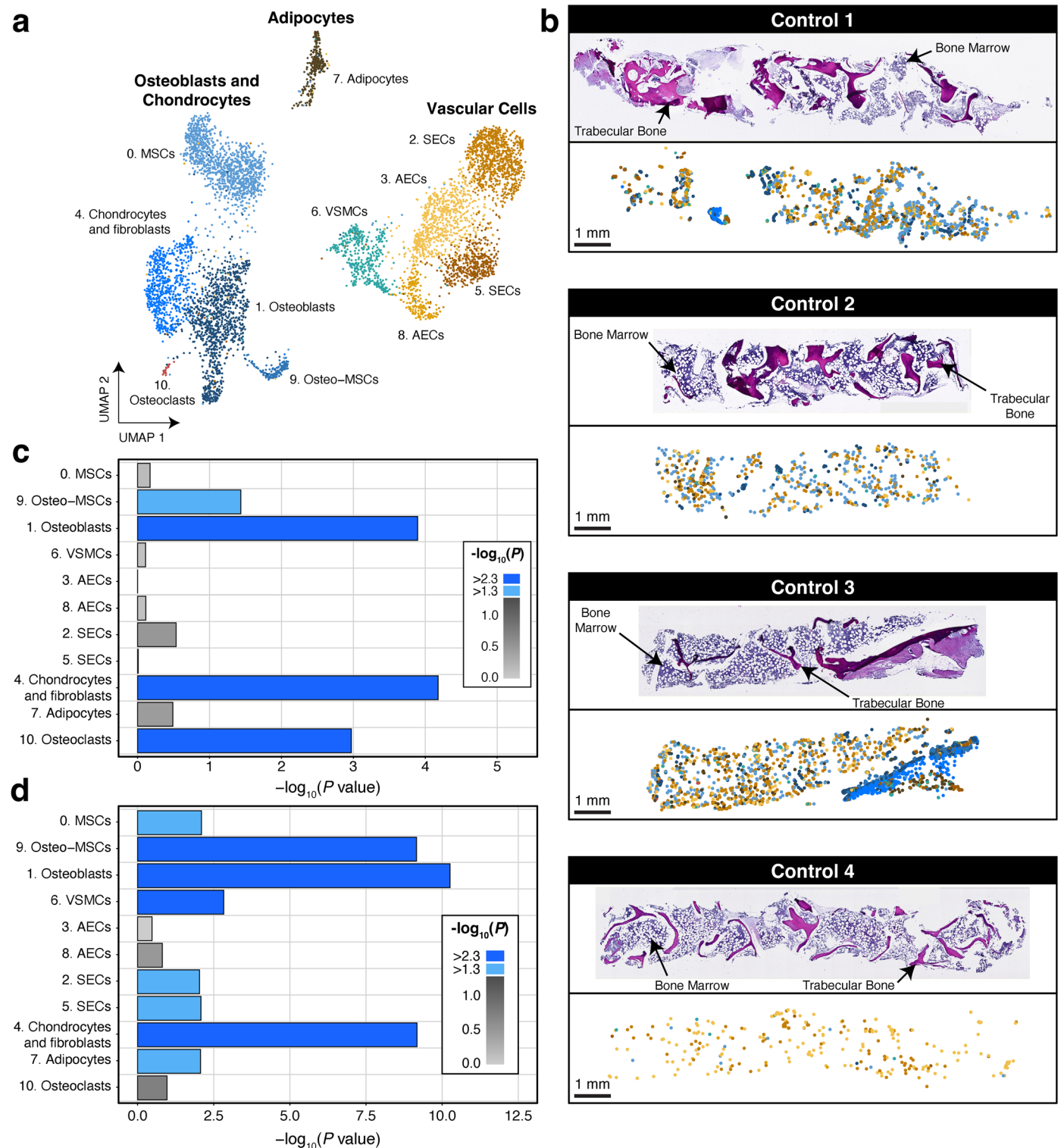
Extended Data Fig. 8 | Gene programs of non-hematopoietic cells and osteoclasts are enriched for genes that cause abnormal bone structure when mutated in mice. This figure details the enrichment of cell type gene programs for genes associated with abnormal bone structure in the MGI database. **(a)** Enrichment of gene programs for MGI abnormal bone structure genes. Light blue bars represent all genes in the gene program; dark blue bars represent the subset of these genes also associated with eBMD. Numbers indicate gene counts (top: MGI genes in gene program; bottom: total genes in gene program). Enrichment was determined by hypergeometric over-representation testing. *Bonferroni*-corrected significance thresholds are indicated. **(b)** Enrichment

analysis excluding genes that cause monogenic skeletal disorders. Light blue bars represent non-causative genes; dark blue bars represent non-causative genes associated with eBMD. Numbers show gene counts. **(c)** Fold-change stratified enrichment for genes causing abnormal BMD in mice. Genes are allocated to six nested groups based on magnitude of differential expression [$\log_2(FC)$, increasing left to right]. Panel (i) includes all genes in gene program, while panel (ii) includes only eBMD-associated genes. *Bonferroni*-corrected significance is indicated. **** $P < 0.001$, *** $P < 0.005$, ** $P < 0.01$, * $P < 0.05$. All statistics are determined by one-tailed Fisher's exact tests under the hypergeometric distribution.



Extended Data Fig. 9 | Cellular mechanism of skeletal phenotype in *Pls3* deficient mice. (a) Left panel: Alcian blue and Alizarin red-stained upper limb paws from P1 mice; scale bar = 1 mm. Second panel: P21 proximal tibia growth plate sections stained with Alcian blue and van Gieson; scale bars = 100 μm. Bar plot shows widths of the growth plate zones (resting zone, RZ; proliferating zone, PZ; hypertrophic zone, HZ). Line plots show femur and nose-to-tail length measured across various ages. Mean ± s.d.; two-tailed Student's *t*-test; **P* < 0.05; ***P* < 0.01. (b) Confocal images of P70 mid-femur cortical, distal femur trabecular, and lumbar vertebral trabecular bone (L1-L3) following calcein double-labelling; scale bars = 100 μm. Graphs show serum PINP levels (P70) and dynamic histomorphometry parameters: MAR (mineral apposition rate), MS/BS (mineralising surface per bone surface), and BFR (bone formation rate) for

both femur and lumbar vertebrae (P70). Mean ± s.d. (c) Decalcified proximal tibia sections from P70 mice stained for TRAP and aniline blue. Black arrows indicate osteoclasts. Scale bars = 100 μm. Graphs show serum CTX levels (P70) and osteoclast parameters: osteoclast number (OcN/BS) and osteoclast surface (OcS/BS) (P70). Mean ± s.d. (d) Graphs show serum levels (mean ± s.d.) of Wnt inhibitors sclerostin (SOST) and dickkopf-related protein 1 (DKK1) in P183 mice. (e) μCT images (P112) of a 250 μm mid-femur ROI, showing small (blue) and large (yellow) osteocyte lacunae. Scale bars = 100 μm. Graphs quantify cortical microporosity (Ct.μPo), lacuna number (Lc.N/BV), mean lacuna volume (Mean Lc.V), and lacuna number by volume (mean ± s.d.; two-tailed Student's *t*-test; **P* < 0.05). Violin plot shows the distribution of Lc.V (Kolmogorov-Smirnov test; *****P* < 0.0001).



Extended Data Fig. 10 | Spatial transcriptomics of human bone samples.

(a) UMAP plot of non-hematopoietic cell types and osteoclasts identified in trephine biopsies of 4 control patients as identified by Yip and Er *et al.*⁵⁰. MSCs = mesenchymal stromal cells; VSMCs = vascular smooth muscle cells; SECs = sinusoidal endothelial cells; AECs = arterial endothelial cells. (b) Haematoxylin and eosin (H&E) images and spatial plots of individual samples used for spatial transcriptomics. Scale bars = 1 mm. Spatial plots display only non-hematopoietic cells and osteoclasts for clarity and are coloured according to (a). Bar plots

showing enrichment of gene programs as determined by one-tailed Fisher's exact test under the hypergeometric distribution. for (c) causative genes of monogenic skeletal disorders and (d) eBMD-associated genes. Scale bars in bar plots indicate the P value. Light blue bars in bar plot correspond to observations that have nominal evidence of enrichment: $P < 0.05$, $-\log_{10}(P)$ of >1.3 . Dark blue bars in bar plot correspond to observations that have robust evidence of enrichment and meet the Bonferroni-corrected significance threshold: $P < 1.6 \times 10^{-3}$, $-\log_{10}(P) > 2.3$.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

1. Collection of cells for single cell RNAseq (scRNA-seq) and single nucleus ATACseq (snATAC-seq)
Processing of scRNA-seq raw data using Cell Ranger versions 2 to 7
Processing of snATAC-seq raw data using Cell Ranger version 2
2. Data from the UK Biobank Study were downloaded via their FTP protocols
3. Human protein coding gene annotations ENSEMBLE Gene v105 (GRCh38)
<https://asia.ensembl.org/biomart/>
4. Genomic co-ordinates and rsids for GRCh38 dbSNP b151
https://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh38p7/database/organism_data/
5. Published BMD SNP associations from NGRI GWAS Catalog
https://www.ebi.ac.uk/gwas/efotraits/EFO_0003923
6. Nosology of genetic skeletal disorders: 2023 Revision
DOI: 10.1002/ajmg.a.63132
7. Mouse Genome Informatics Data
Phenotype ontology terms (http://www.informatics.jax.org/vocab/mp_ontology/)
Genes associated with ontology terms (<http://www.informatics.jax.org/marker/>)

8. For mouse functional phenotyping experiments in the Origins of Bone and Cartilage Diseases (OBCD) study:
 NRecon (Bruker, <http://bruker-microct.com/products/downloads.html>)
 CTAn (Bruker, <http://bruker-microct.com/products/downloads.html>)
 Scanco (v6.4-2)
 Xming 6.9.0.31 (Colin Harrison; <http://www.straightrunning.com/XmingNotes/>)
 PuTTY 0.62 (Simon Tatham; <https://www.putty.org>)
 Instron Bluehill Universal (v4.23) (<https://www.instron.com/en/products/materials-testing-software/bluehill-universal>)
 Drishti-2 (v2.6.1, <https://github.com/nci/drishti>, RRID:SCR_017999)
 Aperio Imagescope (Leica, <https://leicabiosystems.com/digital-pathology/manage/aperio-imagescope>)
 ImageJ (<https://imagej.net>, RRID:SCR_003070)
 TrapHisto (<https://www.liverpool.ac.uk/ageing-and-chronic-disease/bone-hist/trap-hist/>)

Data analysis

1. scRNA-seq analysis was performed using Seurat (versions 2-5), Monocle 2 (version 2.2.0), pySCENIC (version 0.11.2) and CellPhoneDB (version 5.0.1)
 Graphs for scRNA-seq analysis were plotted using Seurat and ggplot2 (version 3.4.2)
 Human orthologs of mouse genes were annotated using bioMaRt (version 2.56.1)

2. snATAC-seq analysis was performed using Signac (version 1.15.0), Seurat (version 5.3.0), JASPAR2020 (version 0.99.10) and presto (version 1.0.0)
 Graphs for snATAC-seq analysis were plotted using Seurat and ggplot2 (version 3.4.2)

3. Genetic association and enrichment analyses
 BOLT-LMM (version 2.3.4), GCTA (version 1.93.2), LDSC (version 1.0.1), BGENIX, R, RStudio, SNPTracker (version 1.0), SnpEff (version 5.1d), VEP, Plink (version 1.9), Plink (version 2.0), Bedtools (version 2.29.2), Bash, Gawk, MAGMA (version 1.10), GraphPad Prism 10, RITAN (version 1.24.0), Cytoscape (version 3.10.0), KING (implemented in Plink version 1.9).

4. Statistical analyses and plots for the wild type baselines in the OBCD phenotyping study were performed and generated using GraphPad Prism 9 and 10 (GraphPad Software Inc; RRID:SCR_002798)

All single-cell, spatial transcriptomic and MAGMA analyses were performed using standard, publicly available packages. The custom R scripts are available at (doi.org/10.5281/zenodo.18529624) and upon reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Source data is available in Supplementary Tables, Source Data section and in online repositories. The raw scRNAseq and snATAC-seq data has been deposited on the Gene Expression Omnibus (GEO) database under the accession number GSE317069.

Genome-wide association summary result statistics is available via the NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) with the accession number GCST90726625.

Spatial transcriptomics data of human bone from Yip et al (2025) were accessed from the Gene Expression Omnibus under accession number GSE299207.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Analyses of human genetic data include both males and females. Genetically inferred sex was utilized as a covariable to adjust for differences in estimated bone mineral density

Analysis of human scRNA-seq data include both males and females.

Reporting on race, ethnicity, or other socially relevant groupings

Our study did not report on socially relevant groupings. Instead genetically inferred ancestry was used to control for population stratification.

Population characteristics

Descriptive statistics of UK-Biobank Study participants are presented in Supplementary Table 7.

Description of human research participants for scRNA-seq analysis are summarised in Supplementary Table 11 to include age, gender and primary diagnosis (reason for elective surgery).

Recruitment

Details of subject recruitment to the UK-Biobank Study are described in the methods, and available from the UK-Biobank website (<https://www.ukbiobank.ac.uk/>)

Femoral head bone samples were obtained from patients with osteoarthritis undergoing total hip arthroplasty at St. Vincent's Hospital Sydney. All participants provided written informed consent, and the study was conducted with approval from the St. Vincent's Hospital Human Research Ethics Committee (2022/ETH00475).

Ethics oversight

Our study has been approved by an appropriate Institutional Review Boards. The UK-Biobank Study obtained appropriate ethics approval as described in the methods and available from the UK-Biobank website (<https://www.ukbiobank.ac.uk/>).

Human bone samples for scRNA-seq were collected with approval from the St. Vincent's Hospital Sydney Human Research Ethics Committee (2022/ETH00475).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

1. For scRNA-seq endosteal bone cells and marrow cells were isolated and captured using 10x Genomics from femurs of 5 groups (endosteal) and 4 groups (marrow) of mice. Each group consisted of 5 mice. No power calculation was performed to pre-determine the sample size. Cells from each of the 4-5 groups of mice were captured with a 10x Genomics run, yielding between 5636-14881 cells each. In total 133,942 cells were included in the analysis. This number was adequate to detect rare populations, such as early osteocytes with 35 cells detected or 0.026% of all cells analysed.

2. For genetic association analyses. Data from 448,010 participants of the UK Biobank Study were selected based on stringent quality control criteria. Participants were selected if they had high-quality quantitative heel ultrasound and genotyping data, and if their genetically inferred ancestry was predominantly European.

3. The coefficient of variation for each of the OBCD skeletal phenotyping parameters were precisely defined using 400 WT mouse samples from two different genetic backgrounds. Power calculations indicate an 80% power to detect outlier phenotype of greater or equal to 2 s.d. with a sample size of n=2 for lines in the OBCD screen.

Power calculations using WT data together with the magnitude of the skeletal phenotype identified female Pls3^{-/-} mice in the OBCD screen (n=2) demonstrated that n=7-8 was required for the comparison of Pls3y^{-/-} mice and WT littermates.

10 other KO lines have been selected from the OBCD screen for further validation by rederivation into independent colonies. Analysis of mice from these colonies have successfully replicated the skeletal phenotypes from the screen. This includes four lines that have been published (Gpc6, Daam2, Cadm1, Sparc).

Data exclusions

1. For scRNA-seq all cells with less than 300 distinct genes observed, or cells with more than 10% of genes mitochondrial genes were removed from analysis as described in methods.

2. For genetic association analyses, we excluded participants if they were missing outcomes and covariates pertinent to the association testing. This is because participants lacking covariates for association testing cannot be fit properly to our mixed-model approach. We also manually filtered for obvious outliers by observing the distributions of the data and removing individuals far exceeding the tail ends of the data. SNPs were excluded from analysis in accordance with our QC criteria which are described in methods.

3. In the OBCD mouse knockout studies, data for all mice were included, including those with and without skeletal phenotypes

Replication

Experiments were conducted using multiple independent biological replicates as outlined in the 'Sample size' section above, and in results and methods section of the manuscript.

1. For scRNA-seq, the captures were conducted using 4 (marrow) or 5 (endosteal bone) independent biological replicates as outlined in results and methods. All annotated cell types and sub-types were replicated and identified in these independent biological replicates.

2. For genetic association analyses, associations with estimated bone mineral density (eBMD) could not be replicated as independent cohorts with eBMD and genetics data are too small to conduct well powered replication studies. To accommodate this our reporting criteria was adjusted accordingly, and we used a more stringent GWAS significance threshold of 6.6×10^{-9} .

3. In the mouse and zebrafish knockout studies, the exact number of biological replicates examined for each knockout line is specified in results.

The OBCD skeletal phenotyping screen has been developed, validated using skeletal samples from 400 WT mice of two different genetic backgrounds with continuous WT sample calibration over >10 years. X-ray microradiography and micro-CT equipment was also calibrated weekly over this period. Skeletal parameters were determined once for each skeletal sample and not replicated. Abnormalities identified in female *Pls3*^{-/-} mice in the OBCD screen (n=2) were replicated in male and female *Pls3* KO mice rederived at Imperial College.

Abnormalities of the skeletal vasculature identified in *Pls3y*^{-/-} mice by CD31 immunohistochemistry were replicated and validated using two different analytical methods Iodine contrast enhanced BSE-SEM and high-resolution micro-computerised tomography (μCT)

10 other KO lines have been selected from OBCD screen for further validation by rederivation into independent colonies. Analysis of mice from these colonies have successfully replicated the skeletal phenotypes. This includes four lines that have been published (*Gpc6*, *Daam2*, *Cadm1*, *Sparc*).

Randomization

1. For scRNAseq samples were collected in 4-5 groups of 5 mice (one 10x Genomics capture for each group of mice). No significant batch effect was detected between the 4-5 independent runs.
2. Genetic association analyses: Participants were recruited at various sites through the UK without any selection criteria. We do adjust for genotyping chip in our association studies, as this assignment was not random.
3. For the OBCD rapid throughput phenotyping of knockout mouse lines - Following the principles of the 3Rs mice were assigned to age groups on the basis of their date of birth. Oldest cohorts first youngest cohorts being the final litters. There were no experimental groups in the skeletal studies.

Blinding

1. Not relevant in the scRNA-seq analysis.
2. Not relevant to our genetic association analyses as the participants represent the general population of the UK.
3. In all skeletal phenotyping analyses investigators were blinded to genotype until statistical comparisons were undertaken

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| ☐ | ☒ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☐ | ☒ Animals and other organisms |
| ☐ | ☒ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

Methods

- | | |
|-------------------------------------|--|
| n/a | Involved in the study |
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

Antibodies

Antibodies used

Fc block; clone 93; catalogue number 101302; Biolegend; 1ul per sample
 TER119-PE; clone TER-119; catalogue number 116208; Biolegend; 1:100 dilution
 CD235-BUV395; clone GA-R2; catalogue number 563810; BD Biosciences; 1:100 dilution
 CD45-APC-H7; clone 2D1; catalogue number 641408; BD Biosciences; 1:100 dilution
 ZombieNIR Viability stain; catalogue number 423106; Biolegend; 1:500 dilution
 B220-BV510; clone RA3-6B2; catalogue number 103248; Biolegend; 1:200 dilution
 TCRb-BV510; clone GL3; catalogue number 118131; Biolegend; 1:200 dilution
 CD45-BV650; clone 30-F11; catalogue number 103151; Biolegend; 1:200 dilution
 CD11b-BUV395; clone M1/70; catalogue number 565976; BD Biosciences; 1:200 dilution
 Ly6C-BUV737; clone HK1.4.rMAb; catalogue number 755201; BD Biosciences; 1:200 dilution
 Ly6G-APC; clone 1A8; catalogue number 560599; BD Biosciences; 1:200 dilution
 CD14-PE; clone Sa14-2; catalogue number 569968; BD Biosciences; 1:200 dilution
 CD31 primary antibody for IHC; clone EPR17259; catalogue number ab182981; 1:200 dilution

Validation

All antibodies used were commercially available and were validated by the manufacturers. No additional validation experiments were performed. Details on validation experiments can be found on the website for each of the catalogue numbers listed.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

1. For scRNA-seq and snATAC-seq experiments, cells were isolated from 9-10-week-old male C57BL/6J mice obtained from Australian Bioresources. Animal holding areas were maintained within a constant temperature range of 21±2 °C and 55±10% humidity. Animals were kept in individually ventilated cages in groups of 2-5 per cage, provided with environmental enrichment and had ad libitum access to sterilised water and a standard irradiated rodent chow.

2. Not relevant to our genetic association analyses

3. OBCD phenotyping study

Knockout mouse lines screened in the OBCD phenotyping pipeline were produced on a background of C57BL/6Brd-Tyrc-Brd, C57BL/6Dnk, and C57BL/6N mice, as part of the Wellcome Trust Sanger Institute's (WTSI) Mouse Genetics Project (MGP), part of the International Mouse Phenotyping Consortium (IMPC; <http://www.mousephenotype.org>). Phenotype analyses were performed on skeletal samples from 16-week-old female mice.

C57BL/6N mice carrying a Pls3 tm1a(EUCOMM)Wtsi knockout first allele (MGI:104807) were obtained from the Wellcome Trust Sanger Institute EMMA mouse repository and re-derived from frozen embryos. Male hemizygous mice (Pls3y/-) were phenotyped at ages P21, P70 and P183.

Animal holding areas were maintained within a constant temperature range of 21±2 °C and 55±10% humidity. Animals were kept in individually ventilated cages in groups of 2-5 per cage, provided with environmental enrichment and had ad libitum access to sterilised water and a standard irradiated rodent chow.

Wild animals

n/a

Reporting on sex

The sex of the animals is detailed in results and methods sections.

Field-collected samples

n/a

Ethics oversight

1. All experiments and procedures involving mice for scRNA-seq and snATAC-seq were conducted in accordance with procedures and protocols approved by the Garvan Institute of Medical Research Animal Ethics Committee (ARA16/01, ARA19/09 and ARA22/12).

2. Not relevant to our genetic association analyses

3. All OBCD mouse skeletal phenotyping studies were undertaken under licence at Imperial College (project licence PPL70/8785 and PP1540664) and the Wellcome Trust Sanger Institute Mouse Genetics Project as part of the International Mouse Phenotyping Consortium and licenced by the UK Home Office (PPLs 80/2485 and P77453634) in accordance with the 1986 Animals (Scientific Procedures) Act and the recommendations of the Weatherall report. Animal experiments were approved by the Sanger or Imperial College Hammersmith Campus Animal Welfare Ethical Review Bodies (AWERB) as appropriate.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

n/a

Study protocol

n/a

Data collection

n/a

Outcomes

n/a

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To obtain endosteal compartment and bone marrow cells for scRNA-seq, mice were sacrificed via CO2 asphyxia. Femurs were harvested from 5 groups of 5 mice. Soft tissue and epiphysis were removed from the femurs before being separated into diaphysis and metaphysis. Marrow cells were collected by flushing the diaphysis with PBS. Marrow-depleted diaphyseal and metaphyseal bone were crushed and cut up gently, then cells that are adherent to the endosteal surface were removed by digestion using 2mg/ml of collagenase A and 2.5mg/ml of trypsin for 30 mins at 37°C. After digestion, bone fragments were vortexed for 10s and the supernatant containing digested endosteal cells was filtered through a 100µm filter into collection tubes containing 10% fetal calf serum (FCS; Bovogen Biologicals). Marrow cells and endosteal cells were collected by centrifugation at 400x g for 5 mins and resuspended in 200µl PBS supplemented with 2% FCS prior to staining for FACS sorting. Cells were stained for relevant antibodies at 4°C for 30 mins and rinsed with PBS supplemented with 2% FCS. Dead cells and debris were excluded by FSC, SSC and DAPI (ThermoFisher Scientific) or ZombieNIR.
Instrument	Cell sorting was done using BD Aria II and Aria III sorters. Flow validation experiments were performed using BD FACSymphony machine
Software	FlowJo (Version 10.10.0)
Cell population abundance	We were able to collect or sort more than enough (>100,000) live non-erythroid cells for each experiment.
Gating strategy	Gating strategy for scRNA-seq experiments are shown in Supplementary Note 3 Gating strategy for validation experiments are shown in Supplementary Figure 1

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.