An atlas of genetic influences on osteoporosis in humans and mice

John A. Morris[1,2,33], John P. Kemp[3,4,33], Scott E. Youlten[5], Laetitia Laurent[2], John G. Logan[6], Ryan C. Chai[5], Nicholas A. Vulpescu[7], Vincenzo Forgetta[2], Aaron Kleinman[8], Sindhu T. Mohanty[5], C. Marcelo Sergio[5], Julian Quinn[5], Loan Nguyen-Yamamoto[9], Aimee-Lee Luco[9], Jinchu Vijay[10], Marie-Michelle Simon[10], Albena Pramatarova[10], Carolina Medina-Gomez[11], Katerina Trajanoska[11], Elena J. Ghirardello[6], Natalie C. Butterfield[6], Katharine F. Curry[6], Victoria D. Leitch[6], Penny C. Sparkes[6], Anne-Tounsia Adoum[6], Naila S. Mannan[6], Davide S.K. Komla-Ebri[6], Andrea S. Pollard[6], Hannah F. Dewhurst[6], Thomas A.D. Hassall[3], Michael-John G. Beltejar[12], 23andMe Research Team*, Douglas J. Adams[13], Suzanne M. Vaillancourt[14], Stephen Kaptoge[15], Paul Baldock[5], Cyrus Cooper[16,17,18], Jonathan Reeve[18], Evangelia E. Ntzani[19,20], Evangelos Evangelou[19,21], Claes Ohlsson[22], David Karasik[23], Fernando Rivadeneira[11], Douglas P. Kiel[23,24,25,26], Jonathan H. Tobias[27], Celia L. Gregson[27], Nicholas C. Harvey[16,17], Elin Grundberg[10,28], David Goltzman[9], David J. Adams[29], Christopher J. Lelliott[29], David A. Hinds[8], Cheryl L. Ackert-Bicknell[30], Yi-Hsiang Hsu[23,24,25,26], Matthew T. Maurano[7], Peter I. Croucher[5], Graham R. Williams[6], J. H. Duncan Bassett[6], David M. Evans[3,4,34], J. Brent Richards[1,2,14,31,32,34]

[1] Department of Human Genetics, McGill University, Montréal, Québec, Canada
[2] Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada
[3] University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia
[4] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[5] Garvan Institute of Medical Research, Sydney, New South Wales, Australia
[6] Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, London, UK
[7] Institute for Systems Genetics, New York University Langone Medical Center, New York, New York, USA
[8] Department of Research, 23andMe, Inc., Mountain View, California, USA
[9] Research Institute of the McGill University Health Centre, Montréal, Québec, Canada
[10] McGill University and Genome Quebec Innovation Centre, Montréal, Québec, Canada
[11] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, Netherlands
[12] Department of Biomedical Genetics, University of Rochester, Rochester, New York, USA
[13] Department of Orthopedics, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA
[14] Department of Medicine, McGill University, Montréal, Québec, Canada
[15] Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[16] MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK
[17] NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
[18] NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK
[19] Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece
[20] Center for Evidence Synthesis in Health, Department of Health Services, Policy and Practice, School of Public Health, Brown University, Rhode Island, USA
[21] Department of Epidemiology and Biostatistics, Imperial College London, London, UK
[22] Department of Internal Medicine and Clinical Nutrition, University of Gothenburg, Gothenburg, Sweden
[23] Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts, USA
[24] Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[25] Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

[26] Broad Institute of Harvard and Massachusetts Institute of Technology, Boston, Massachusetts, USA
[27] Musculoskeletal Research Unit, Department of Translational Health Sciences, University of Bristol, Bristol, UK
[28] Children's Mercy Hospitals and Clinics, Kansas City, Missouri, USA
[29] Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
[30] Center for Musculoskeletal Research, Department of Orthopaedics, University of Rochester, Rochester, New York, USA
[31] Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montréal, Québec, Canada
[32] Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
[33] These authors contributed equally to this work
[34] These authors jointly supervised this work

* Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Jennifer C. McCreight, Karen E. Huber, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson.

Correspondence should be addressed to J.B.R. (brent.richards@mcgill.ca) or D.M.E. (d.evans1@uq.edu.au)

Abstract

Osteoporosis is a common aging-related disease diagnosed primarily using bone mineral density (BMD). We assessed genetic determinants of BMD as estimated by heel quantitative ultrasound (eBMD) in 426,824 individuals, identifying 518 genome-wide significant loci (301 novel), explaining 20% of its variance. We identified 13 bone fracture loci, all associated with eBMD, in ~1.2M individuals. We then identified target genes enriched for genes known to influence bone density and strength (maximum odds-ratio=58, $p=10^{-75}$) from cell-specific features, including chromatin conformation and accessible chromatin sites. We next performed rapid-throughput skeletal phenotyping of 126 knockout mice lacking target genes and found an increased abnormal skeletal phenotype frequency compared to 526 unselected lines ($p<0.0001$). In-depth analysis of one gene, *DAAM2*, showed a disproportionate decrease in bone strength relative to mineralization. This genetic atlas provides evidence testing how to link associated-SNPs to causal genes, offers new insights into osteoporosis pathophysiology and highlights opportunities for drug development.

Introduction

Osteoporosis is a common, aging-related disease characterized by decreased bone strength and consequent increased fracture risk.[1] Bone mineral density (BMD), the most clinically relevant risk factor when diagnosing osteoporosis, is highly heritable[2] and is a strong risk factor for fracture.[3] BMD GWAS have demonstrated that it is a highly polygenic trait,[2] and the known genetic determinants of fracture all act through BMD.[cite BMJ paper Trajonoska] Recently, we identified 203 loci associated with estimated BMD (eBMD) by measuring quantitative heel ultrasound, explaining 12% of its variance, demonstrating this polygenicity.[4]

eBMD is predictive of fracture and is highly heritable (50-80%).[5–9] While BMD measured from dual-energy X-ray absorptiometry (DXA)-scanning is most often used in clinical settings, our recent eBMD GWAS identified 84% of all currently known genome-wide significant loci for DXA-BMD[4] and effect sizes were concordant between the two traits (Pearson's r =0.69 for lumbar spine and 0.64 for femoral neck).[4] The largest GWAS to-date for DXA-derived BMD measures contained only 66,628 individuals.[10] Both ultrasound and DXA-derived BMD are strongly associated with fracture risk where a standard deviation decrease in either metric is associated with an approximate 1.5-fold increase in osteoporotic fracture risk.[3,11]

Little is known about how to reliably map associated loci to their causal genes. However, highly polygenic traits such as bone density allow for empirical testing of which methods link associated SNPs to genes enriched for causal proteins. Causal proteins can be identified in human clinical trials when their manipulation by medications leads to changes in BMD.[2] Another source of causal proteins is Mendelian genetic conditions, which may constitute human knockouts and strongly implicate key genes that underlie bone physiology.[12] Given a sufficient number of associated loci, different genomic characteristics that link a SNP to these causal proteins can be tested. These include genomic landscape characteristics such as cell-specific 3-dimensional (3D) contact domains, cell-specific open chromatin states, physical proximity and the presence of coding variation. Furthermore, knockout mice generated by large-scale studies can be used to identify genes whose deletion results in an abnormal murine skeletal phenotype. Rapid-throughput phenotyping data can then be used to determine whether outlier bone phenotypes are enriched in mice harboring deletions of genes identified by GWAS in humans.

Here, we present a comprehensive investigation of genetic influences on eBMD and fracture in humans and mice. We undertook an eBMD GWAS of 426,824 individuals in the UK Biobank, identifying 301 novel loci which explained 20% of its variance, and identified genetic determinants of fracture in up to 1.2 million individuals combining the UK Biobank and 23andMe cohorts. We then assessed SNP-level and genomic landscape characteristics, mapping associated SNPs to genes enriched for known bone density proteins. Identified target genes were enriched up to 58-fold for known causal genes and for genes differentially expressed *in vivo* in osteocytes compared to bone marrow cell models. Finally, we asked whether deletion of GWAS-identified genes resulted in skeletal abnormalities *in vivo* by undertaking rapid-throughput phenotyping of knockout mice, which included 126 target genes. Mice harboring deletions of these 126 genes were enriched for outlier skeletal phenotypes. A convergence of human and mouse genetics, bone-cell expression and cell culture data pointed to a role for *DAAM2* in osteoporosis. We found that mice with a hypomorphic *Daam2* allele had marked decreases in bone strength and increases in cortical bone porosity. Finally, CRISPR/Cas9-mediated edits of *DAAM2* in osteoblast cell lines demonstrated a reduction in mineralization, compared to un-edited cells.

These novel loci will empower future clinical and pharmacological research on osteoporosis, spanning from a better understanding of its genetic susceptibility to, potentially, biomarker discovery and drug targets.

Results

GWAS for eBMD and fracture

We selected 426,824 UK Biobank full release White British individuals (55% female) for an eBMD GWAS (Online Methods, **Supplementary Table 1**, **Supplementary Figure 1**). We analyzed 13,737,936 autosomal and X-chromosomal SNPs for their association with eBMD. Although there was substantial inflation of the test statistics relative to the null for eBMD (λ_{GC}=2.26, **Supplementary Figure 2**), linkage disequilibrium (LD) score regression indicated that most of the inflation was due to polygenicity rather than population stratification [LD score regression intercept =1.06 (0.063), ratio=0.017 (0.018)].

We identified 1,103 conditionally independent signals (423 novel) at genome-wide significance ($p<6.6x10^{-9}$, Online Methods) mapping to 515 loci (301 novel, **Supplementary Table 2**, **Figure 1**). Of the conditionally independent lead SNPs per locus, 4.6% were rare, having a minor allele frequency (MAF) ≤1%, whereas 9.3% were low-frequency (MAF≤5% but >1%) and 86.1% were common (MAF>5%, **Supplementary Figure 3** shows the relationship between MAF and absolute effect size). The average absolute conditional effect sizes for these three categories of SNPs were 0.14, 0.04 and 0.02 standard deviations, respectively. The total variance explained by conditionally independent genome-wide significant eBMD lead SNPs was 20.3%. When partitioning the variance explained by these lead SNPs into three MAF categories, we found that rare variants explained 0.8% of the variance in eBMD, whereas low-frequency and common variants explained 1.7% and 17.8% of the variance, respectively. We found strong correlations between eBMD effect sizes with UK Biobank interim release effect sizes (r=0.93, **Supplementary Figure 4**, **Supplementary Table 3**). In addition, we performed sex heterogeneity analyses to investigate whether the genetic aetiology of eBMD differed between the sexes (**Supplementary Note, Supplementary Figure 6**, **Supplementary Tables 5**, **6** and **7**). The total number of genome-wide significant conditionally independent signals becomes 1,106 (518 loci) when including these analyses, however, we focus on results from the main GWAS for this study.

We identified 416,795 UK Biobank participants [n_{cases}=53,184 (60% female) and $n_{controls}$=373,611 (54%female)] for a GWAS of fracture risk (**Supplementary Table 1**). We assessed 13,977,204 autosomal and X-chromosomal SNPs and identified 14 conditionally independent signals associated with fracture mapping to 13 loci (**Supplementary Table 4**, **Supplementary Figure 5**). Once again, we observed test statistic inflation (λ_{GC}=1.15). However, this was also likely due to polygenicity, rather than population stratification [LD score regression intercept =1.00 (0.008), ratio=0.017 (0.038)]. Conditionally independent genome-wide significant lead SNPs were tested for replication in a cohort of research participants from 23andMe, Inc., a personal genetics company (n_{cases}=367,900 and $n_{controls}$=363, 919). All 14 SNPs showed strong evidence of replication (**Supplementary Table 4**). All genome-wide significant fracture SNPs were also found to be genome-wide significantly associated with eBMD in the expected direction of effect (i.e. alleles lowering eBMD increased fracture risk). Furthermore, there was a highly negative correlation between SNP effect sizes on eBMD and fracture [r=-0.77 (-0.79, -0.74), **Supplementary Figure 4**].

Fine-mapping associated loci

To map SNPs to potentially causal genes, we first refined associated SNPs at each locus using two statistical fine-mapping methods, GCTA-COJO[13,14] and FINEMAP.[15] These methods identify

SNPs based on their conditional independence and posterior probability for causality, respectively. We generated SNP sets for each genome-wide significant autosomal locus by identifying conditionally independent lead SNPs or SNPs having a high posterior probability of causality, as determined by $\log_{10}$ Bayes factor >3 (**Figure 2a**, we report all SNPs with $\log_{10}$ Bayes factor >2 in **Supplementary Tables 8**, **9** and **10**). Here we refer to the set of "fine-mapped SNPs" as those SNPs achieving either conditional independence or a high posterior probability for causality—on average, we observed two conditionally independent SNPs and five SNPs with a log10 Bayes factor >3 per locus (**Supplementary Note**).

Comparing fine-mapped SNPs for biological activity

Given the large number of associated SNPs per locus, downstream analyses should focus on SNPs most likely to be biologically functional. We used accessible chromatin sites surveyed in relevant cellular contexts as a proxy for biological activity. We generated ATAC-seq maps in the human osteosarcoma cell line SaOS-2—cells that possess osteoblastic features and can be fully differentiated into osteoblast-like cells. We also analyzed DNase I hypersensitive site (DHS) maps from human primary osteoblasts from the ENCODE project.[16] Both ATAC-seq and DHS data were analyzed using a uniform mapping and peak-calling algorithm (Online Methods).

We then analyzed fine-mapped SNPs for enrichment of these functional signatures relative to all SNPs within 1 Mbp of each genome-wide significant association locus. Fine-mapped SNPs, including the set of conditionally independent SNPs and SNPs with $\log_{10}$ Bayes factors >3, were strongly enriched for both missense variants in protein coding regions (**Supplementary Note**, **Supplementary Table 11**) and osteoblast open chromatin sites (**Figure 3a**). As $\log_{10}$ Bayes factor increased, fold-enrichment increased as well (**Figure 3b**). This indicated that fine-mapped SNPs were highly enriched for genomic signatures of function, which can inform the choice of statistical cut-off for SNP selection in follow-up functional studies.

Mapping fine-mapped SNPs to target genes & enrichment for positive control genes

Human genetic associations have rarely been translated to improved clinical care, primarily because causal genes at associated loci have often not been indisputably identified. We therefore sought to test which genomic features linked associated SNPs to genes known to influence bone biology in humans. We identified proteins whose perturbation through pharmacotherapy[2] or Mendelian disease led to changes in bone density or strength. Mendelian disease genes were defined as monogenic disorders characterized with altered bone mass or abnormal skeletal mineralization, osteolysis and/or skeletal fragility or osteogenesis imperfecta (**Supplementary Table 12**) and constitute an informative human knockout resource.[17] We considered such proteins identified through pharmacotherapy or Mendelian disease to be products of "positive control" genes likely critical to bone biology.

Next, we investigated which genomic features linked fine-mapped SNPs to positive control genes. We tested whether positive control genes were enriched among six types of genomic characteristics that can link a SNP to a gene: 1) Genes that were most proximal to fine-mapped SNPs; 2) Genes that contained fine-mapped SNPs overlapping their gene bodies; 3) Genes containing fine-mapped SNPs that were coding variants; 4) Genes identified to be in 3D contact with fine-mapped SNPs in human osteoblasts or osteocytes through high-throughput chromatin conformation capture (Hi-C) experiments; 5) The closest gene to fine-mapped SNPs which also mapped to ATAC-seq peaks in SaOS-2 cells; and 6) Genes within 100 kbp of fine-mapped SNPs (**Figure 2b** emphasizes the target gene selection, **Figure 4** details this entire pipeline). Coding annotations, ATAC-seq peaks and Hi-C interaction peaks were not combined but kept separate to enable different sources of data to provide converging and confirmatory evidence. Distance from a fine-mapped SNP to a gene considered the closer of the 3' and 5' ends, not the

transcription start site. We named identified genes "Target Genes" and tested which of these six methods most enriched Target Genes for positive control genes.

The set of Target Genes most strongly enriched for positive control genes arose from genes targeted by SNPs that were conditionally independent and by SNPs identified to be plausibly causal with a $\log_{10}$ Bayes factor >3 (**Table 1**, **Supplementary Table 13**). This set of Target Genes featured 556 genes total, approximately one per locus. All six methods for linking fine-mapped SNPs to Target Genes yielded strong enrichment for positive control genes. The odds ratios ranged from 5.1 [95% CI: (3.0,8.6), $p=10^{-11}$] for Target Genes within 100 kbp of the fine-mapped SNPs to an odds ratio of 58.5 [(26.4,129.31), $p=10^{-75}$)] for Target Genes closest to fine-mapped SNPs in osteoblast-derived ATAC-seq peaks (**Table 1**). In addition, we used FUMA[18] to assess which pathways from the WikiPathways[19] database were identified by the set of Target Genes most strongly enriched for positive control genes. We observed known pathways such as Wnt signalling, endochondral ossification, osteoclast and osteoblast signalling as well as novel pathways were highlighted by this approach (**Supplementary Figure 7**).

These results suggest that our Target Gene identification method leads to strong enrichment for positive control genes known to be central to bone biology. Such methods may help to prioritize genes at associated loci for functional testing, which are more likely to influence bone biology and therefore, have clinical relevance. The full list of mapped Target Genes and the method through which they were identified is presented in **Supplementary Table 14**.

Mapping fine-mapped SNPs to osteocyte-signature genes

An alternative method to assess the biological plausibility of Target Genes is to test whether their expression is enriched in bone cells. Osteocytes are the most abundant cell type in bone and are key regulators of bone mass, bone formation and bone resorption.[20] We therefore assessed the transcriptome of primary mouse osteocytes derived from three bone types *in vivo*.[21] Genes enriched for expression in osteocytes and expressed in all bone types defined an osteocyte transcriptome signature.[21] We then tested which of the methods used to identify eBMD Target Genes resulted in the greatest enrichment for osteocyte-signature genes.

We found that Target Genes were strongly enriched for osteocyte signature genes, with odds ratios for enrichment ranging from 2.1 [95% CI: (1.7,2.5), $p=2x10^{-17}$)] for Target Genes within 100 kbp of the fine mapped SNPs, to 7.4 [(3.8,14.5), $p=5x10^{-12}$)] for Target Genes mapped through fine-mapped coding SNPs (**Table 2**, **Supplementary Tables 15** and **16**). This again suggested our methods result in enrichment for biologically-relevant genes.

A large-scale high throughput mouse knockout screening program

We investigated whether deletion of Target Genes resulted in enrichment of outlier skeletal phenotypes with the Origins of Bone and Cartilage Disease (OBCD) study ("URLs", **Supplementary Note**). Outlier cortical and trabecular bone phenotypes were more frequent in mice with disruptions of 126 Target Genes compared against 526 unselected knockout lines {**Supplementary Tables 17** and **18**, OR 3.2 [(95% CI: (1.9,5.6), p<0.0001]}. Therefore, enrichment of abnormal skeletal phenotypes in mice with disruption of Target Genes provides clear functional validation that our fine-mapping approach identifies critical and biologically-relevant skeletal genes. Our fine-mapping *in vivo* and *in vitro* data converged to identify *DAAM2* as a highly credible and novel osteoporosis gene, therefore we undertook detailed analyses of mice with a hypomorphic *Daam2* allele to illustrate the potential of this approach.

In-Depth Characterization of DAAM2

Numerous lines of evidence identified *DAAM2* as an important gene for further functional investigation. First, a conditionally independent lead SNP, rs2504101, mapped directly to *DAAM2* ($p_{conditional}=4.3x10^{-10}$) and second, fine-mapping revealed two coding missense variants with high posterior probabilities for causality, rs201229313 in its 19th exon ($log_{10}BF=3.7$), and rs61748650 in its 21st exon ($log_{10}BF=2.5$). Third, a rare variant, rs772843886, near *DAAM2* was suggestively associated with risk of fracture ($p=2x10^{-3}$). Fourth, the *Daam2*$^{tm1a/tm1a}$ mouse was identified to have an outlier skeletal phenotype in our rapid throughput mouse knockout screening program (**Supplementary Table 17**). Fifth, although *DAAM2* has not previously been implicated in osteoporosis, it has been predicted to have a role in canonical Wnt signaling.[22,23]

To investigate the role of *DAAM2* in bone biology, we first tested its expression in bone cells. We performed RNA-seq and ATAC-seq experiments in four different human osteoblast cell lines and found it was expressed in all cell lines (Online Methods, **Supplementary Figure 8**). Staining experiments in the SaOS-2 cell line revealed DAAM2 localized specifically in the cell nuclei (**Supplementary Figures 9** and **10**). This functional evidence from human bone cells also led us to characterize *Daam2* in mouse bone cells. *Daam2* was identified as an osteocyte signature gene (**Supplementary Table 16**) and was expressed in mouse calvarial osteoblasts and bone marrow-derived osteoclasts (**Supplementary Table 19**).

Next using CRISPR/Cas9, we tested the effect on bone mineralization of double-stranded breaks (DSBs) in the second exon of *DAAM2* in SaOS-2 osteoblast cell lines (Online Methods). We found that after 14 days of treatment with osteogenic factors, control cells transfected with the intact plasmid, but not undergoing an DSB of the *DAAM2* gene, had a 9-fold increase in mineralization. After the introduction of a DSB in the second exon of *DAAM2*, induced mineralization was severely impaired (**Figure 5**). These CRISPR/Cas9-based findings suggest that DAAM2 influences mineralization capacity in human osteoblasts.

We next analyzed the skeletal phenotypes of *Daam2*$^{tm1a/tm1a}$, *Daam2*$^{+/tm1a}$ and wild-type littermate mice in detail. Adult male *Daam2*$^{tm1a/tm1a}$ mice had reduced femur and vertebral bone mineral content (BMC), while male *Daam2*$^{+/tm1a}$ and female *Daam2*$^{tm1a/tm1a}$ mice also had reduced vertebral BMC. These changes were accompanied by a small reduction in femur length in *Daam2*$^{tm1a/tm1a}$ mice (males=2.7%, females=3.5%). Despite otherwise normal trabecular and cortical bone structural parameters, cortical porosity was increased in both male and female *Daam2*$^{tm1a/tm1a}$ mice (**Supplementary Figure 11**).

Consistent with their increased cortical porosity, *Daam2*$^{tm1a/tm1a}$ mice had markedly reduced bone strength (**Figure 6**) even though all other cortical bone parameters, including BMD, were normal (**Supplementary Figure 11**). Bone composition and structure were thus investigated in *Daam2*$^{tm1a/tm1a}$ mice by comparing *Daam2*$^{tm1a/tm1a}$ mineralization and biomechanical parameters with values predicted by linear regression analysis of over 300 wild-type age, sex and genetic background matched wild-type controls. Measures of bone composition and structure in *Daam2*$^{tm1a/tm1a}$ mice were reduced compared to wild-type mice, and vertebral stiffness was > 2 standard deviations below that predicted even after accounting for reduced BMC (**Figure 6c**, **Supplementary Table 20**). We observed in additional experiments (**Supplementary Note**) that measures of bone resorption (TRAP) and formation (P1NP) did not differ between wild-type and Daam2 hypomorphic mice (**Supplementary Figure 12**), and that Male Daam2 hypomorphic mice had decreased mineral content per unit matrix protein and increased carbonate substitution (**Supplementary Figure 13**)

Taken together, these data suggest the decreased bone strength in *Daam2*$^{tm1a/tm1a}$ mice is not simply a result of abnormal bone turnover, but also a consequence of increased porosity and

impaired bone composition and structure. If DAAM2 proves to be a tractable drug target, such an agent would represent a complementary therapeutic strategy for prevention and treatment of osteoporosis and fragility fracture.

While *DAAM2* represents a detailed validation of a novel Target Gene, we also highlight five additional eBMD Target Genes, with evidence for association with fracture (**Supplementary Table 21**), in the **Supplementary Note**. These five genes had contrasting abnormalities of bone structure and strength when deleted in mice, emphasizing their functional role in skeletal physiology and importance for further study. These genes can be found in **Supplementary Tables 11** and **17** and are *CBX1* (**Supplementary Figure 14**), *WAC* (**Supplementary Figure 15**), *DSCC1* (**Supplementary Figure 16**), *RGCC* (**Supplementary Figure 17**) and *YWHAE* (**Supplementary Figure 18**). Respective bone composition and structure screens are in **Supplementary Figure 19**.

Discussion

In this comprehensive study on the genetic determinants of bone density and fracture in humans and mice, we identified 518 genome-wide significant loci (301 novel) that explained 20% of total eBMD variance. In a meta-analysis of up to 1.2 million individuals, 13 fracture loci were identified, all of which also associated with eBMD. Leveraging the polygenicity of eBMD, we demonstrated strong enrichment for fine-mapped SNPs in bone cell open chromatin. We used fine-mapped SNPs to identify Target Genes strongly enriched for genes with known central roles in bone biology through Mendelian genetics, or as targets for clinically-validated osteoporosis therapies. High-throughput skeletal phenotyping of mice with deletions of 126 Target Genes revealed enrichment for outlier skeletal phenotypes compared to 526 unselected lines. Last, we identified DAAM2 as a protein with critical effects on bone strength, porosity, composition and mineralization. These findings will enable on-going and future studies to better understand genomic characteristics that link fine-mapped SNPs to sets of genes enriched for causal proteins. Furthermore, this comprehensive study of genetic variants associated with osteoporosis will provide opportunities for biomarker and drug development

The polygenicity of eBMD is striking. Few traits and diseases currently have hundreds of loci associated at genome-wide significance.[12,24] This has led to a large proportion of total eBMD variance being explained by now known genetic determinants, which will facilitate bone biology studies and enable osteoporosis drug development.[25] Despite the large number of genetic and biological inputs into eBMD determination, pharmacological perturbation of even only one protein identified in our GWAS can have clinically-relevant effects. For example, RANKL inhibition has been shown to increase bone density by up to 21% after ten years of therapy.[26] Interestingly, the genetic variants near RANKL have small effects on eBMD. Thus, despite small effect sizes for most identified variants, these do not necessarily reflect effect sizes of protein pharmacological manipulation. This is because common genetic variants tend to have small effects on protein function, whereas pharmacotherapies tend to have large effects on protein function. Consequently, dose-response curves describing the effect of small and large genetic perturbations on eBMD are needed to decide which proteins to target for drug development.[12]

Polygenicity improved our statistical power to validate linking associated loci with potentially causal genes. We found that fine-mapped SNPs were able to identify Target Genes strongly enriched for positive control genes—particularly when the approach implemented relatively simple strategies (e.g. nearest gene), or the gene nearest a fine-mapped SNP in cell-relevant open chromatin. We also observed that fine-mapped SNPs were often in 3D contact with Target Genes in human osteoblasts and osteocytes. These data, surveying many genomic landscape

features, provide guidance for investigators attempting to identify causal genes from GWAS-associated SNPs.

The marked reduction in *Daam2*$^{tm1a/tm1a}$ mice's bone strength, despite minimal changes in bone morphology and mineral content, indicated that *Daam2*$^{tm1a/tm1a}$ mice have abnormal bone composition and structure explained in part by increased cortical porosity. Furthermore, CRISPR/Cas9-mediated knockouts of *DAAM2* in osteoblast cells lines resulted in a marked reduction in inducible mineralization. Few such genes have been identified and further investigations will be required to determine whether *DAAM2* represents a tractable drug target. Nevertheless, previous studies have suggested that *DAAM2* indirectly regulates canonical Wnt signalling across several developmental processes.[22,23] Using different sources of data to identify *DAAM2* allowed for greater confidence in results. While each type of data has its own biases, these biases are partially orthogonal, and consequently, concordant evidence from different sources of data increases the quality of the evidence, an approach known as triangulation.[27]

Our fracture GWAS identified 13 loci. These loci also associated with BMD and/or eBMD, highlighting the importance of BMD as a determinant of fracture risk, at least in the age range assessed within the UK Biobank. While BMD-independent loci for fracture likely exist, they were not identified despite this well-powered study. This suggests that screening for fracture drug targets should also include understanding the effect of the protein on BMD.

This study has important limitations. First, we measured eBMD, instead of DXA-derived BMD, which is typically measured in the clinic. Nonetheless, beyond their phenotypic correlation, these two traits have high genetic concordances in terms of their genome-wide significant loci, suggesting that underlying biological properties of these two traits are similar. Importantly, eBMD is a strong predictor of fracture risk in its own right, and contributes to risk assessment over and above DXA-derived BMD at the hip.[28] While our Target Gene approach identified a set of candidate genes enriched for genes with known effects on bone density, it is important to note that there is no gold-standard set of genes known to influence BMD. Our rapid-throughput mouse knockout program is on-going and will investigate many of the Target Genes implicated by our study. Further efforts will be required to functionally validate—or exclude—these genes for effects on bone biology. Our Target Gene approach did not include human gene expression quantitative trait loci (eQTL) data. This is because the largest available eQTL experiments for human osteoblasts involve only 95 individuals,[29] and larger sample sizes with RNA-sequencing data will be required to link fine-mapped SNPs to genes. Finally, this work was limited to individuals of White British genetic ethnicity, leaving the effect of most genome-wide significant SNPs in other populations to be assessed. It is likely that on-going studies in non-British populations will address this question.

In summary, we have generated an atlas of genetic influences on osteoporosis in humans and mice. We have more fully described the genetic architecture of eBMD and fracture and identified Target Genes strongly enriched for known roles in bone biology. We used human and mouse genetics, functional genomics and genome editing to demonstrate the relevance of this approach, formally known as triangulation[27], by identifying *DAAM2*. Disruption of *DAAM2* in mice led to increased cortical porosity and marked bone composition and strength reduction, and in human osteoblasts led to decreased mineralization. We expect these Target Genes to include new drug targets for the treatment of osteoporosis, a common disease for which novel therapeutic options are a health priority.

Accession Codes

Gene Expression Omnibus accession number GSE120755.

URLs

International Mouse Phenotyping Consortium (IMPC), http://www.mousephenotype.org and http://www.sanger.ac.uk/mouseportal; Mouse Genome Informatics (MGI), http://www.informatics.jax.org; the Origins of Bone and Cartilage Disease Study (OBCD), http://www.boneandcartilage.com; UK Biobank, http://www.ukbiobank.ac.uk/; Genetic Factors for Osteoporosis Consortium (GEFOS), http://www.gefos.org/; UK Biobank protocol for measurement of eBMD, https://biobank.ctsu.ox.ac.uk/crystal/docs/Ultrasoundbonedensitometry.pdf; UK Biobank document #155580 on genotyping and quality control, http://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_qc.pdf; Hg19 gene range list, https://www.cog-genomics.org/plink2/; Knockout Mouse Project, https://www.komp.org/; NHS Digital, http://content.digital.nhs.uk/hes; hotspot2, https://github.com/Altius/hotspot2; ENCODE, http://encodeproject.org. liftOver, https://genome.sph.umich.edu/wiki/LiftOver; BGENIX, https://bitbucket.org/gavinband/bgen/wiki/bgenix

Acknowledgments

This research has been conducted using the UK Biobank Resource (accession IDs: 24268, 12703 and 4580). J.B. Richards was supported by the Canadian Institutes of Health Research, the Canadian Foundation for Innovation and the Fonds de Recherche Santé Québec (FRSQ) and a FRQS Clinical Research Scholarship. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. J.A. Morris was funded by the Canadian Institutes of Health Research. D.M. Evans was funded by a National Health and Medical Research Council Senior Research Fellowship (APP1137714) and funded by a Medical Research Council Programme Grant (MC_UU_12013/4). J.P. Kemp was funded by a University of Queensland Development Fellowship (UQFEL1718945). C.L. Gregson was funded by Arthritis Research UK (ref; 20000). G.R. Williams, J.H.D. Bassett and P.I. Croucher were funded by the Wellcome Trust (Strategic Award grant number 101123; project grant 094134) and P.I. Croucher was also funded by the Mrs. Janice Gibson and the Ernest Heine Family Foundation. D. Karasik was supported by Israel Science Foundation grant #1283/14. Y.-H. Hsu was funded by US NIH NIAMS 1R01AR072199. F. Rivadeneira, C. Medina-Gomez, and K. Trajanoska were funded by the Netherlands Organization for Health Research and Development (ZonMw VIDI 016.136.361 grant). C.L. Ackert-Bicknell was funded by NIH/NIAMS AR063702 AR060981. D.P. Kiel was funded by grants from the National Institute of Arthritis Musculoskeletal and Skin Diseases R01 AR041398, R01 AR072199. S. Youlten was funded by the Australian Government Research Training Program Scholarship. J. Reeve and S. Kaptoge were funded by the Genetic Factors of Osteoporosis-GEFOS EU FP7 Integrated Project Grant Reference: 201865 2008-12 and 2007-12 UK NIHR Biomedical Research Centre Grant (Musculoskeletal theme) to Cambridge Clinical School. C. Ohlsson was supported by the Swedish Research Council, Swedish Foundation for Strategic Research, ALF/LUA research grant from the Sahlgrenska University Hospital, Lundberg Foundation, European Calcified Tissue Society, Torsten and Ragnar Söderberg's Foundation, Novo Nordisk Foundation, Knut and Alice Wallenberg Foundation. M.T. Maurano was supported by NIH grant R35 GM119703.

We thank M. Schull for assistance with high-performance computing at the University of Queensland Diamantina Institute, and T. Winkler for invaluable technical support for the EasyStrata Software used in this study. We thank the Sanger Institute's Research Support

Facility, Mouse Pipelines and Mouse Informatics Group who generated the mice and collected materials for this manuscript. We would like to thank the research participants and employees of 23andMe, Inc. for making this work possible.

Author Contributions

J.A.M., J.P.K, A.P., C.L.A.-B., C.L.G., C.O., D.K., D.P.K., E.E., E.G. F.R., G.R.W., J.H.D.B., J.H.T., M.T.M., N.J.H., P.I.C., V.F., Y.-H.H., D.M.E. and J.B.R. conceived of and designed experiments. J.A.M., J.P.K., A.K., A.S.P., A.-T.A., D.G., D.K.-E., E.J.G., H.F.D., J.G.L., K.F.C., M.-J.G.B., N.A.V., N.C.B., N.S.M., P.C.S., R.C., SEY, S.K., T.A.D.H., V.D.L., A.P., C.L.A.-B., C.L.G., D.M.E., E.G. G.R.W., J.H.D.B., M.T.M., N.J.H., V.F., Y.-H.H. and J.B.R. performed data analysis. J.A.M., J.P.K, A.-L.L., A.-T.A., C.M.-G., C.M.S., D.G., Do.J.A., E.J.G., H.F.D., J.G.L., J.V., K.F.C., L.L., L.N.-Y., M.-J.G.B., M.-M.S., N.S.M., P.C.S., R.C., SEY, S.M., A.P., C.L.A.-B., and Y.-H.H conducted experiments. J.A.M., J.P.K., G.R.W., J.H.D.B., D.M.E. and J.B.R. wrote the manuscript. J.A.M. and J.P.K. were the lead analysts. All authors revised and reviewed the paper.

Competing Interests Statement

A.K. and D.A.H. are employees of 23andMe, Inc.

References

1. World Health Organization. Consensus development conference: Prophylaxis and treatment of osteoporosis. *Osteoporos. Int.* **1,** 114–117 (1991).
2. Richards, J. B., Zheng, H.-F. & Spector, T. D. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat. Rev. Genet.* **13,** 576–588 (2012).
3. Johnell, O. *et al.* Predictive value of BMD for hip and other fractures. *J. Bone Miner. Res.* **20,** 1185–1194 (2005).
4. 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).
5. Arden, N. K., Baker, J., Hogg, C., Baan, K. & Spector, T. D. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. *J. Bone Miner. Res.* **11,** 530–534 (1996).
6. Hunter, D. J. *et al.* Genetic variation in bone mineral density and calcaneal ultrasound: A study of the influence of menopause using female twins. *Osteoporos. Int.* **12,** 406–411 (2001).
7. Bauer, D. C. Broadband Ultrasound Attenuation Predicts Fractures Strongly and Independently of Densitometry in Older Women. *Arch. Intern. Med.* **157,** 629 (1997).
8. Bauer, D. C. *et al.* Quantitative ultrasound predicts hip and non-spine fracture in men: The MrOS study. *Osteoporos. Int.* **18,** 771–777 (2007).
9. Karasik, D. *et al.* Mapping of quantitative ultrasound of the calcaneus bone to chromosome 1 by genome-wide linkage analysis. *Osteoporos. Int.* **13,** 796–802 (2002).
10. Medina-Gomez, C. *et al.* Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects. *Am. J. Hum. Genet.* **102,** 88–102 (2018).
11. McCloskey, E. V. *et al.* Predictive ability of heel quantitative ultrasound for incident fractures: an individual-level meta-analysis. *Osteoporos. Int.* **26,** 1979–1987 (2015).
12. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19,** 110–124 (2018).
13. 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).
14. 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).
15. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32,** 1493–1501 (2016).
16. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
17. Rivadeneira, F. & Mäkitie, O. Osteoporosis and Bone Mass Disorders: From Gene Pathways to Treatments. *Trends Endocrinol. Metab.* **27,** 262–281 (2016).
18. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8,** 1826 (2017).
19. Kutmon, M. *et al.* WikiPathways: Capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **44,** D488–D494 (2016).
20. Dallas, S. L. & Bonewald, L. F. Dynamics of the transition from osteoblast to osteocyte. *Ann. N. Y. Acad. Sci.* **1192,** 437–443 (2010).
21. Youlten, S. *et al.* Osteocytes express a unique transcriptome that underpins skeletal homeostasis. *J Bone Min. Res 32 (Suppl 1)* (2017).
22. Lee, H. K. & Deneen, B. Daam2 Is Required for Dorsal Patterning via Modulation of Canonical Wnt Signaling in the Developing Spinal Cord. *Dev. Cell* **22,** 183–196 (2012).
23. Lee, H. K. *et al.* Daam2-PIP5K Is a Regulatory Pathway for Wnt Signaling and

Therapeutic Target for Remyelination in the CNS. *Neuron* **85,** 1227–1243 (2015).
24. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101,** 5–22 (2017).
25. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47,** 856–860 (2015).
26. Bone, H. G. *et al.* 10 years of denosumab treatment in postmenopausal women with osteoporosis: results from the phase 3 randomised FREEDOM trial and open-label extension. *Lancet Diabetes Endocrinol.* **5,** 513–523 (2017).
27. Lawlor, D. A., Tilling, K. & Smith, G. D. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45,** 1866–1886 (2016).
28. Moayyeri, A. *et al.* Quantitative ultrasound of the heel and fracture risk assessment: An updated meta-analysis. *Osteoporos. Int.* **23,** 143–153 (2012).
29. Grundberg, E. *et al.* Population genomics in a disease targeted primary cell model. *Genome Res.* **19,** 1942–1952 (2009).

Figure 1. Manhattan plot of genome-wide association results for eBMD in the UK Biobank. The dashed red line denotes the threshold for declaring genome-wide significance ($6.6x10^{-9}$). 1,103 conditionally independent SNPs at 515 loci passed the criteria for genome-wide significance in n=426,824 UK Biobank participants. 301 novel loci (defined as > 1 Mbp from previously reported genome-wide significant BMD variants) reaching genome-wide significance are displayed in blue. Previously reported loci that reached genome-wide significance are displayed in red, and previously reported loci failing to reach genome-wide significance in our study are shown in black.

Figure 2. **Fine-mapping SNPs and target gene selection diagram. a)** For each 500 Mbp region around a conditionally independent lead SNP ($p<6.6x10^{-9}$ after conditional independence testing; n=426,824 UK Biobank participants) we applied statistical fine-mapping to calculate log_{10} Bayes factors for each SNP as a measure of their posterior probability for causality. Conditional independence testing was implemented using GCTA-COJO[13,14] and log_{10} Bayes factors were estimated using FINEMAP.[15] SNPs that were conditionally independent lead SNPs or that had log_{10} Bayes factors > 3 were considered our fine-mapped SNPs that we then used for target gene identification. **b)** Target Genes were identified if: 1) It was the gene closest to a fine-mapped SNP. 2) A fine-mapped SNP was in its gene body. 3) A fine-mapped SNP was coding. 4) The gene mapped closest to a fine-mapped SNP which resided in an SaOS-2 ATAC-seq peak. 5) A fine-mapped SNP was present in a Hi-C osteoblast or osteocyte promoter interaction peak, therefore being closer to a target gene in three-dimensions than linearly on the genome.

Figure 3. SNPs at genome-wide significant loci are enriched for bone-relevant open chromatin sites. Comparison of eBMD-associated SNPs in terms of enrichment for DHSs from primary osteoblasts, and ATAC-seq peaks from SaOS-2 osteosarcoma cells. Odds ratios were computed relative to all SNPs at genome-wide significant regions. Enrichments for missense protein coding SNPs are shown as baselines. **a**) Enrichments for conditionally independent (COJO) or log_{10} Bayes factor >3 (FINEMAP); note the latter set contains nearly twice the number of SNPs. **b**) Ranking SNPs by log_{10} Bayes factor (x-axis) showed increasing enrichment. 95% confidence interval (shaded region) was calculated by a two-sided Fisher's Exact Test.

Figure 4. Target Gene Identification Workflow.

Figure 5. Reduction of DAAM2 protein resulted in reduced mineralization in SaOS-2 cells. Mineralization quantification in control cells and *DAAM2* exon 2 double-stranded break (DSB) induced cells in either the presence of osteogenic factors (treated) or absence (untreated). **a**) Dot plot of n=6 independent experiments ± standard error of the mean (SEM) from Alizarin red staining in (**b**) to quantify mineralization; Bar=5mm. ***$p=1.3x10^{-15}$ compared to untreated control cells and $^{\&\&\&}p=9.3x10^{-15}$ (left) and $8.2x10^{-13}$ (right) compared to treated control cells determined by one-way ANOVA (F=49.7, df=5) and Bonferroni post-hoc tests.

Figure 6. Biomechanical analyses of mice with Daam2 knockdown. a) Femur biomechanical analysis. Destructive 3-point bend testing (Instron 5543 load frame) of femurs from wild-type (WT, $n_{female}=3$, $n_{male}=4$), *Daam2*$^{+/tm1a}$ ($n_{female}=6$, $n_{male}=4$) and *Daam2*$^{tm1a/tm1a}$ ($n_{female}=5$, $n_{male}=9$) mice. Graphs show yield load, maximum load, fracture load, stiffness (gradient of the linear elastic phase) and toughness (energy dissipated prior to fracture). Female data are shown on the left and male data on the right. Data are shown as mean ± standard error of the mean (SEM). Female maximum load analyses for WT versus Daam2$^{tm1a/tm1a}$ (**) and Daam2$^{+/tm1a}$ versus Daam2$^{tm1a/tm1a}$ ($^{\#}$) had statistically significant differences (one-way ANOVA

p=3.0x10^{-3}, F=10.29, df=13, Tukey's post-hoc test **p<0.01 and $^{\#}$p<0.05). Male maximum load analyses for WT versus *Daam2*$^{tm1a/tm1a}$ (***) and *Daam2*$^{+/tm1a}$ versus *Daam2*$^{tm1a/tm1a}$ had statistically significant differences [one-way ANOVA p<1.0x10^{-4} (GraphPad Prism does not report smaller p-values), F=50.11, df=16, Tukey's post-hoc test ***p<1.0x10^{-3} and $^{\#\#\#}$p<1.0x10^{-3}]. Male fracture load analyses for WT vs *Daam2*$^{tm1a/tm1a}$ (***) and *Daam2*$^{+/tm1a}$ vs *Daam2*$^{tm1a/tm1}$ ($^{\#\#}$) had statistically significant differences (one-way ANOVA p=3.0x10^{-4}, F=15.49, df=16, Tukey's post-hoc test ***p<1.0x10^{-3} and $^{\#\#}$p<0.01). **b**) Vertebra biomechanical analyses. Destructive compression testing (Instron 5543 load frame) of caudal vertebrae from WT (n_{female}=3, n_{male}=4), *Daam2*$^{+/tm1a}$ (n_{female}=6, n_{male}=4) and *Daam2*$^{tm1a/tm1a}$ (n_{female}=5, n_{male}=9) mice. Graphs show yield load, maximum load and stiffness. Data are shown as mean ± SEM. Female yield load analysis for WT versus *Daam2*$^{tm1a/tm1a}$ (**) had a statistically significant difference (one-way ANOVA p=6.5x10^{-3}, F=8.26, df=13, Tukey's post-hoc test **p<0.01). Female maximum load analyses for WT versus *Daam2*$^{tm1a/tm1a}$ (**) and WT versus *Daam2*$^{+/tm1a}$ (*) had statistically significant differences (one-way ANOVA p=2.9x10^{-3}, F=10.45, df=13, Tukey's post-hoc test **p<0.01 and *p<0.05). Male maximum load analysis for WT vs *Daam2*$^{tm1a/tm1a}$ (*) had a statistically significant difference (one-way ANOVA p=0.04, F=4.10, df=16, Tukey's post-hoc test *p<0.05). **c**) Bone quality analysis from rapid throughput screening mouse knockouts. The graph demonstrates the physiological relationship between bone mineral content and stiffness in caudal vertebrae from P112 female WT mice (n=320). The blue line shows the linear regression (Pearson's r=0.21, p=1.2x10^{-4}) and the grey box indicates ± 2 standard deviations (SD). The mean value for female *Daam2*$^{tm1a/tm1a}$ [n=2 from initial OBCD screen (**Supplementary Note**)] mice is shown in orange (-2.14 SD).

Tables

Table 1. Target gene identification methods enrichment for 57 positive control genes. Enrichment was calculated with a chi-square test against 19,455 total protein coding genes. No positive control genes were identified by osteocyte Hi-C interactions therefore we did not calculate its enrichment. Distance to gene was determined using 3' and 5' ends, instead of the transcription start site.

Target Gene Set	Odds Ratio (95% Confidence Interval)	p-value
SaOS-2 ATAC-seq Peak Gene	58.5 (26.4 – 129.3)	$1.3x10^{-75}$
Coding SNP Gene	41.8 (14.3 – 121.6)	$1.0x10^{-30}$
Osteoblast Hi-C Interaction Gene	21.1 (6.4 – 69.6)	$7.8x10^{-13}$
Closest Gene	12.9 (7.1 – 23.4)	$1.8x10^{-27}$
Overlapping Gene Body	11.2 (5.2 – 23.8)	$3.4x10^{-15}$
All Genes Within 100 kbp	6.8 (3.9 – 11.7)	$2.1x10^{-15}$
Osteocyte Hi-C Interaction Gene	NA	NA

Table 2. **Target gene identification methods enrichment for 1,240 osteocyte signature genes.** Enrichment was calculated with a chi-square test against 19,455 total protein coding genes. Distance to gene was determined using 3' and 5' ends, instead of the transcription start site.

Target Gene Set	Odds Ratio (95% Confidence Interval)	p-value
Coding SNP Gene	7.4 (3.8 – 14.5)	$5.2x10^{-12}$
SaOS-2 ATAC-seq Peak Gene	6.1 (3.5 – 10.6)	$2.6x10^{-13}$
Overlapping Gene Body	5.1 (3.8 – 6.7)	$1.1x10^{-37}$
Closest Gene	4.6 (3.7 – 5.6)	$4.1x10^{-53}$
Osteoblast Hi-C Interaction Gene	3.8 (1.9 – 7.4)	$2.5x10^{-5}$
Osteocyte Hi-C Interaction Gene	2.9 (1.0 – 8.6)	$4.0.x10^{-2}$
All Genes Within 100 kbp	2.1 (1.7 – 2.5)	$1.8x10^{-17}$

Online Methods

Curating osteoporosis associated outcomes in the UK Biobank study

During the period from 2006 to 2010, half a million British adults were recruited by the UK Biobank ("URLs").[30] Subjects provided biological samples, consented to physical measurements and answered questionnaires relating to general health and lifestyle. Ethical approval was granted by the Northwest Multi-Centre Research Ethics Committee, and informed consent was obtained from all participants prior to participation. Heel bone quality was evaluated in 487,428 subjects by quantitative ultrasound speed of sound (SOS) and broadband ultrasound attenuation (BUA) using a Sahara Clinical Bone Sonometer (Hologic Corporation, Bedford, Massachusetts, USA). Further information regarding the assessment protocols are publicly available on the UK Biobank website ("URLs"). For in-depth details on participant selection, see the **Supplementary Note**. The R script used to curate the raw data is available on request, together with all supporting summary data and plots. Descriptive statistics of the cohort, after quality control, are detailed in **Supplementary Table 1.**

Fracture cases were identified using two mutually non-exclusive methods: Hospital Episodes Statistics linked through NHS Digital ("URLs") with a hospital-based fracture diagnosis irrespective of mechanism within the primary (n=392,292) or secondary (n=320,448) diagnosis field, and questionnaire-based self-reported fracture within the past five years (n=501,694). We defined a set of International Classification of Diseases codes, 10th revision (ICD10), to separate fracture cases from controls with the Hospital Episodes Statistics data. We excluded fractures of the skull, face, hands and feet, pathological fractures due to malignancy, atypical femoral fractures, periprosthetic and healed fracture codes. A full list of ICD10 codes used can be found in **Supplementary Table 22**. We did not exclude any self-reported fracture cases by fracture site, since participants were only asked if they sustained a fracture at ankle, leg, hip, spine, write, arm, other or unknown. We identified 20,122 fractures using ICD10 codes and 48,818 using questionnaire-based self-reported data. Descriptive statistics of the cohort, after quality control and ancestry selection, are detailed in **Supplementary Table 1**.

For details on ancestry assignment of UK Biobank participants to White British and the identification of unrelated samples for LD reference estimation and X chromosome analyses, see the **Supplementary Note** and **Supplementary Figures 20**, **21** and **22**.

Genome-wide association analysis

A maximum of 426,824 White British individuals (233,185 females and 193,639 males) with genotype and valid QUS measures were analyzed (**Supplementary Table 1**). For fracture, a maximum of 426,795 White British individuals, comprising 53,184 fracture cases (60% female) and 373,611 controls (54% female) were analyzed. We note that the sample sizes between the two assessed traits are similar but different, due to not all fracture cases and controls having eBMD measured, and vice-versa. We tested autosomal genetic variants for association with eBMD and fracture, separately, assuming an additive allelic effect, using a linear mixed non-infinitesimal model implemented in the BOLT-LMM v2 software package[31] to account for population structure and cryptic relatedness. The following covariates were included as fixed effects in all models: age, sex, genotyping array, assessment center and ancestry informative principal components 1 to 20. Autosomal analysis was restricted to up to 13,977,204 high quality HRC imputed variants with a MAF >0.05%, minor allele count >5, info score >0.3, genotype hard call rate >0.95, and Hardy-Weinberg $p>1x10^{-6}$. We also analyzed the association between eBMD and fracture and directly genotyped SNPs on the X chromosome, adjusting for the same covariates, using the Plink2 (October 2017) software package[32] and a nested sample of unrelated participants (n=362,926 for eBMD and n=45,087 cases and 317,775 controls for fracture). As the analyses for the X chromosome data were based upon observed genotypes,

we excluded SNPs with evidence of deviation from Hardy-Weinberg Equilibrium ($p<1\times10^{-6}$), MAF <0.05%, minor allele count ≤5, and overall missing rate >5%, resulting in up to 15,466 X chromosome SNPs for analysis. Heterogeneity in effect size coefficients between sexes was tested in EasyStrata[33], using Cochran's test of heterogeneity[34]

$$X_{het} = \sum_i [(\beta_i - \beta_{Overall})^2 w_i] \sim \chi^2(m-1)$$

β_i *effect size estimates of stratum i*

SE_i *standard error of stratum i*

$$w_i = 1/SE_i^2$$

i = 1..m

Manhattan plots of our genome-wide association scans were generated using the same software. We have previously estimated the genome-wide significance threshold $\alpha=6.6\text{x}10^{-9}$ for analyzing data from the UK Biobank using the above critera.[4]

Fracture replication meta-analysis

14 genome-wide significant conditionally independent lead SNPs identified from our fracture GWAS were tested for replication in the 23andMe cohort. Genetic associations were tested against the fracture phenotype on a set of unrelated individuals of European ancestry. Analyses were adjusted for age, sex, principal components 1 to 5, and the genotyping platform. There were 367,900 cases and 363,919 controls. Meta-analysis of UK Biobank discovery and 23andMe replication data was performed using METAL.[35] In order to compare the effect estimates and standard errors of the UK Biobank discovery and 23andMe replication data, we transformed the UK Biobank discovery effect estimates and standard errors as per the manual specifications in the BOLT-LMM[31] documentation, specifically:

$$\log \text{OR} = \frac{\beta}{\mu * (1-\mu)}$$

where μ = case fraction and standard errors of SNP effect estimates should also be divided by $(\mu * (1-\mu))$.

Approximate conditional association analysis

To detect multiple independent association signals at each of the genome-wide significant eBMD and fracture loci, we applied approximate conditional and joint genome-wide association analysis using the software package GCTA v1.91.[14] Variants with high collinearity (multiple regression R^2 >0.9) were ignored and those situated more than 20 Mbp away were assumed to be independent. A reference sample of 50,000 unrelated White British individuals randomly selected from the UK Biobank was used to model patterns of linkage disequilibrium (LD) between variants. The reference genotyping dataset consisted of the same variants assessed in our GWAS. Conditionally independent variants reaching genome-wide significance were annotated to the physically closest gene using Bedtools v2.26.0[36] and the hg19 gene range list ("URLs").

Estimation of variance explained by significant variants and SNP heritability

We estimated the proportion of eBMD phenotypic variance tagged by all SNPs on the genotyping array (i.e. the SNP heritability) using BOLT-REML[31] and Linkage Disequilibrium Score Regression (LDSC)[37]. To calculate the variance explained by independent genome-wide significant SNPs (i.e. all 1,103 genome-wide significant conditionally independent lead SNPs) we summed the variance explained per SNP using the formula: $2p(1-p)\beta^2$, where p is the

effect allele frequency and β is the effect of the allele on a standardized phenotype (mean=0, variance=1).[38–40]

Estimating genomic inflation with LD score regression (LDSC)

To estimate the amount of genomic inflation present in the data that was due to residual population stratification, cryptic relatedness, and other latent sources of bias, we used stratified LDSC[41] in conjunction with partitioned LD scores that were calculated for high quality HM3 SNPs derived from a sample of unrelated 1000G EUR individuals.

Fine-mapping SNPs

Fine-mapped SNPs were defined as those being conditionally independent, as identified by GCTA-COJO or exceeding our threshold for posterior probability of causality, as defined by FINEMAP. Here we describe the generation of this set of fine-mapped SNPs.

First, SNPs were defined as being conditionally independent using GCTA-COJO.[13,14] We next calculated the posterior probability of causality. To do so, we defined each conditionally independent lead SNP as a signal around which we would undertake posterior probability testing. We used all imputed SNPs within 500 kbp of a conditionally independent lead SNP and treated each signal independently. For details on our application of FINEMAP for statistical fine-mapping to calculate $\log_{10}$ Bayes factors per SNP, see the **Supplementary Note**. We used a $\log_{10}$ Bayes factor >3 threshold to only consider SNPs with the strongest posterior probabilities for causality, and those SNPs that were identified as genome-wide significant conditionally independent lead SNPs, as being fine-mapped SNPs.

RNA sequencing for mouse osteocytes

We performed an analysis of whole transcriptome sequencing data of three distinct bone types from the mouse skeleton to measure osteocyte expression.[4] The three sites were the tibia, femur and humerus, and in each, the bone marrow was removed (n=8 per site). The distribution of normalized gene expression for each sample was used to calculate a threshold of gene expression[42], with genes above this threshold for 8 out of 8 replicates in any bone type deemed to be expressed. Osteocyte enriched genes were determined by comparing the transcriptomes of matched bone sample controls, one with the marrow removed and the other with the marrow left intact (n=5 per site). Genes significantly enriched in osteocytes and expressed in all bone types were defined as osteocyte transcriptome signature genes.

Mapping accessible chromatin

ATAC-seq libraries were generated by the McGill University and Genome Quebec Innovation Centre on 100,000 SaOS-2 cells, using a modified protocol to that previously described.[43] The modifications included: reducing the transposase reaction volume from 50 µl to 25 µl, increasing the transposase concentration from 1x to 40x, and using 12 cycles of PCR to enrich each library. Libraries were quantified by Q-PCR, Picogreen and LabChip, then were sequenced on the Illumina HiSeq 4000 (pair-ended 125 bp sequences), using the Nextera sequencing primers. DNase-seq data from primary osteoblast samples[16] were obtained from ENCODE (“URLs”) under accessions ENCLB776DWN and ENCLB906BCL.

Reads were processed using a uniform pipeline to produce both ATAC-seq and DNase-seq peaks. Illumina adapters were trimmed using Trimmomatic v. 0.36.[44] Reads were aligned to the hg38 human reference using BWA v.0.7.15.[45] Peak calling was performed using hotspot2 (“URLs”) with a cutoff of 1% FDR and converted to hg19 reference coordinates using UCSC liftOver (“URLs”).

RNA sequencing for human osteoblast cell lines

RNA library preparations were carried out on 500 ng of RNA from SaOS-2, U2OS, MG63 and HOS cells with RNA integrity number (RIN) >7 using the Illumina TruSeq Stranded Total RNA Sample preparation kit, according to manufacturer's protocol. Final libraries were analyzed on a Bioanalyzer and sequenced on the Illumina HiSeq 4000 (pair-ended 100 bp sequences). Raw reads were trimmed for quality (phred33 ≥30) and length (n ≥32), and Illumina adapters were clipped off using Trimmomatic v. 0.35.[44] Filtered reads were aligned to the GRCh37 human reference using STAR v. 2.5.1b.[46] Raw read counts of genes were obtained using HTseq-count v.0.6.1.[47]

High-throughput chromosome conformation capture

High-throughput chromosome conformation capture (Hi-C) was performed on primary human osteoblasts and osteocytes from human bone biopsies of non-fracture subjects. Hi-C libraries were prepared as described previously.[48] Instead of using HindIII restriction enzyme, we used DpnII[49] which increased coverage and insensitivity of CpG methylation.[50] The Hi-C libraries were sequenced on Illumina HiSeq 4000 instruments to 2 billion pair-end reads. Replicates of osteoblasts and osteocytes were independently generated and sequenced. HiC-Pro was used to process the HiC-Pro pipeline[51] beginning with aligning each read end to hg38 reference genomes. The Chimeric read ends were filtered to keep only 5′ alignments with MAPQ >10, and then read-ends were paired and de-duplicated. Contact matrices were constructed, and significant interactions were estimated with Homer[52], GOTHiC[53] and Juicer.[54] We defined significant interactions as $p<10^{-15}$ (comparing observed interactions to estimated expected interactions and taking into account DNA fragment size, GC content, and other genomic features). Only interaction pairs that were significant ($p<10^{-15}$) from all three tools were considered significant. The resolution of Hi-C interactions was from 1.5 to 2 kbp with average 1.8 kbp. ATAC-seq experiments were also performed in primary osteoblasts and osteocytes that were used for HI-C experiments. We only considered and reported chromatin interactions that mapped to open chromatin.

Target Gene identification

We identified Target Genes for the autosomal fine-mapped sets by annotating fine-mapped sets of SNPs to the closest protein-coding gene, making additional note if the SNP mapped directly to the gene's introns or exons, or was coding. We identified Target Genes on the X chromosome by the closest gene to a conditionally independent lead SNP, as we did not calculate $\log_{10}$ Bayes factors for SNPs on the X chromosome. Additionally, we annotated Target Genes that may be functional in bone cells by marking which fine-mapped SNPs mapped to open chromatin in human bone cells, identified by SaOS-2 ATAC-seq peaks, and we mapped chromosomal positions of fine-mapped SNPs to significant Hi-C interactions of primary osteoblast and osteocytes. When the interaction chromatin mapped to multiple isoforms of protein coding genes, we selected the one with the most significant interaction (usually with highest interaction counts). When the interaction chromatin mapped to multiple bins, we selected the one(s) with looping domains. We further annotated Target Genes using the osteocyte signature gene set where genes within this set are enriched for osteocyte activity.[4]

Target Gene enrichment analyses

We performed a series of enrichment analyses by calculating the odds of Target Genes being either positive control genes or osteocyte signature genes. We identified a set of 57 proteins whose perturbation through pharmacotherapy[2], or Mendelian disease leads to changes in bone density, monogenic disorders presenting with abnormal skeletal mineralization or low bone mass, osteolysis and/or skeletal fragility and osteogenesis imperfecta and abnormal skeletal mineralization (**Supplementary Table 12**).[17] For all protein-coding genes in the genome, which

were identified using refGene[55] (n=19,455), we annotated whether they were found to be Target Genes and/or positive control genes. These annotations allowed us to construct contingency tables and calculate an odds ratio for enrichment of Target Genes amongst positive control genes. We then used chi-square tests to calculate p-values. We used multiple genomic features to test which methods of identifying Target Genes enriched for positive control genes. To do so, we tested if positive control genes were enriched amongst targeted genes identified by four different methods: 1) Genes that were most proximal to the fine-mapped set SNPs; 2) Genes that contained fine-mapped SNPs overlapping their gene bodies; 3) Genes containing fine-mapped SNPs that are coding variants; 4) Genes identified to be in 3D-contact with fine-mapped sets in human osteoblasts or osteocytes through Hi-C experiments; 5) The closest gene to fine-mapped SNPs, which also mapped to ATAC-seq peaks in human osteoblast SaOS-2 cell lines; and 6) Those genes within 100 kbp of fine-mapped SNPs (**Figures 2 and 4**). We then repeated this analysis using the osteocyte signature gene set (n=1,240) instead of the positive control set, to calculate the odds of Target Genes being active in the osteocyte. For details on the Target Gene pathway analyses using FUMA[18], see the **Supplementary Note**.

CRISPR/Cas9 Methods

SaOS-2 cells were obtained from ATCC (#ATCC HTB-85) and cultured in McCoy5A medium (ATCC) supplemented with 15% of FBS (Wisent inc) and 1% of penicillin and streptomycin (Wisent Inc.) according to the manufacturer. Three different guide RNAs (gRNA) targeting the second exon of *DAAM2* were cloned in the plasmid pSpCas9(BB)-2A-GFP (PX458), which was a gift from Feng Zhang (Addgene plasmid #48138)[56]. For gRNA sequences, see **Supplementary Note**. We observed the cutting frequency determination (CFD) scores[57] for each gRNA was < 0.1, therefore we did not consider off-target effects to merit testing[58]. The construct plasmids were purified using the QIAGEN filter midi prep kit (QIAGEN #12243) according to manufacturer instructions. SaOS-2 cells were cultured to 80% confluence in a 100-mm^2 petri dish. Cells were then transfected with one of the three different plasmids generated, or with the intact plasmid as a control, using TransIT LT1 transfection reagent (Mirus #MIR2304) with a reagent-to-DNA ratio of 3:1. 48 hours post-transfection, GFP positive cells were sorted by FACS in a single cell model. The remaining colonies were expanded and then assessed for the presence of DAAM2 protein using immunofluorescence technique (Anti-DAAM2 antibody, Sigma-Aldrich #HPA051300). For PCR primers designed against regions of *DAAM2* flanking the three gRNA target sequences to generate 355 bp amplicons, see the **Supplementary Note**. PCR products of the identified clones were sequenced using MiSeq (Genome Quebec). For DAAM2 Western blots that show DAAM2 protein expression reduced to 17.5% and 33.5% in the gRNA1 and gRNA2 edited clones (**Supplementary Figure 23**), respectively, see the **Supplementary Note**.

To induce mineralization (**Figure 5**), cells were then cultured to 90% confluence in a 6-well plate and then treated, or left untreated for a control, with osteogenic factors (Ascorbic acid 50 µg/ml and ß-Gycerophosphate 10 mM). Fresh media containing osteogenic factors was added every 2-3 days over 13 days. At day 14, mineralization was quantified using the osteogenesis assay kit according to manufacturer instructions (Millipore #ECM815). The Alizarin red concentration (µM) was normalized with the protein content assessed in the media in each culture (Pierce BCA Protein assay kit; Thermo Fisher #23227).

Rapid throughput mouse knockout program

For specifics on the Origins of Bone and Cartilage Disease (OBCD) high-throughput phenotyping, see the **Supplementary Note** and **Supplementary Table 18**.

***Daam2* knockout mice**

Mouse studies undertaken at the Garvan Institute of Medical Research (Darlinghurst, NSW, Australia) were approved by the Garvan Institute / St Vincent's Hospital Animal Ethics Committee in accordance with New South Wales (Australia) State Government legislation. *Daam2*$^{tm1a(KOMP)Wtsi}$ mice (designated *Daam2*$^{tm1a/tm1a}$) were obtained from the Wellcome Trust/Sanger Institute (Cambridge, UK) where the mice were generated as part of the International Mouse Phenotyping Consortium ("URLs"), using ES cells produced by the Knockout Mouse Project ("URLs"). The *Daam2* gene in these mice was disrupted by a cassette containing an insertion with an additional splice acceptor site between exons 5 and 6 ("URLs"). The success of this strategy was confirmed with an 80% knockdown of *Daam2* in *Daam2*$^{tm1a/tm1a}$ and 50% knockdown in *Daam2*$^{+/tm1a}$. Age and sex matched 16-week old mice were used for detailed skeletal phenotyping, as described above.

For details on RNA sequencing for mouse calvarial osteoblasts, *in vitro* osteoblast mineralization, *in vitro* assays of osteoclast formation, the detection of serum markers of bone resorption and formation and for Fourier-transform infrared spectroscopy analyses see the **Supplementary Note.**

Data availability

Human genotype and phenotype data on which the results of this study were based are available upon application from the UK Biobank ("URLs"). GWAS summary statistics for eBMD and fracture can be downloaded from the GEFOS website ("URLs"). RNA-seq and ATAC-seq data generated for human osteoblast cell lines, including re-called DHS peaks from human primary osteoblasts, can be downloaded from the Gene Expression Omnibus (accession number GSE120755). Mouse phenotype data are available online from the IMPC ("URLs") and OBCD ("URLs").

Code availability

Analysis scripts available by request from the authors.

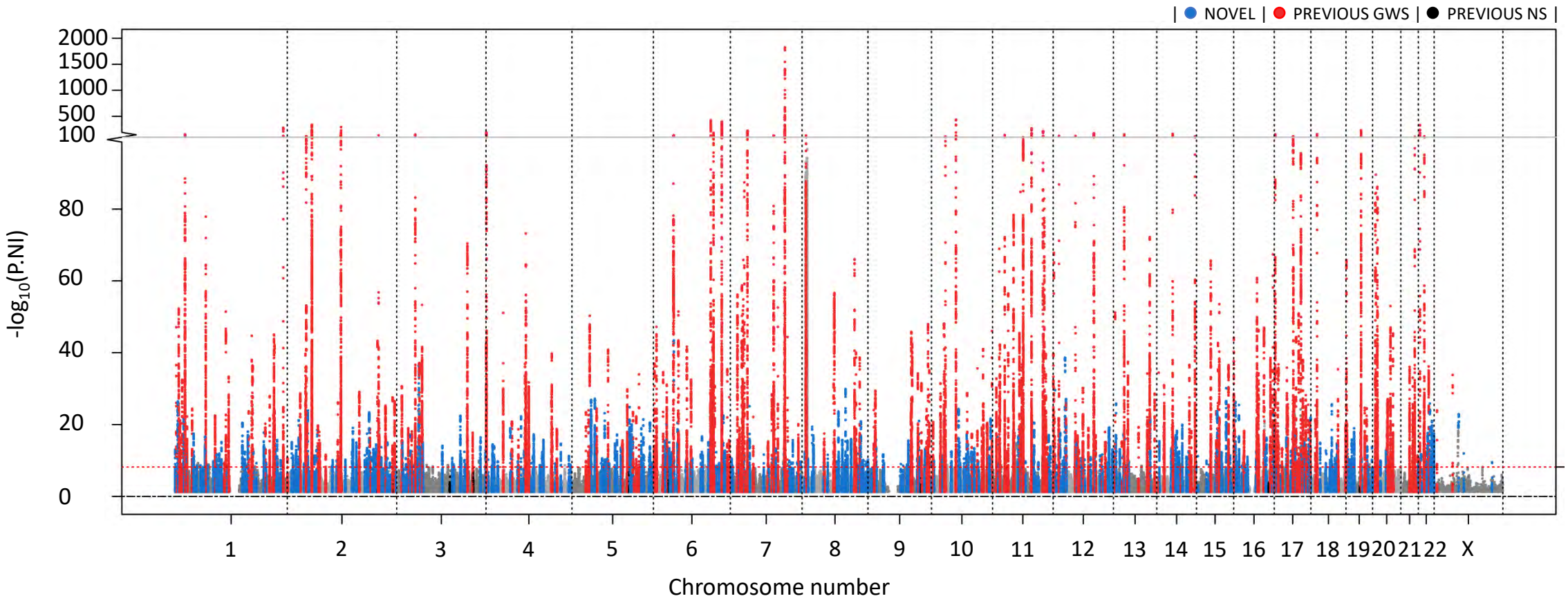
Ethical compliance

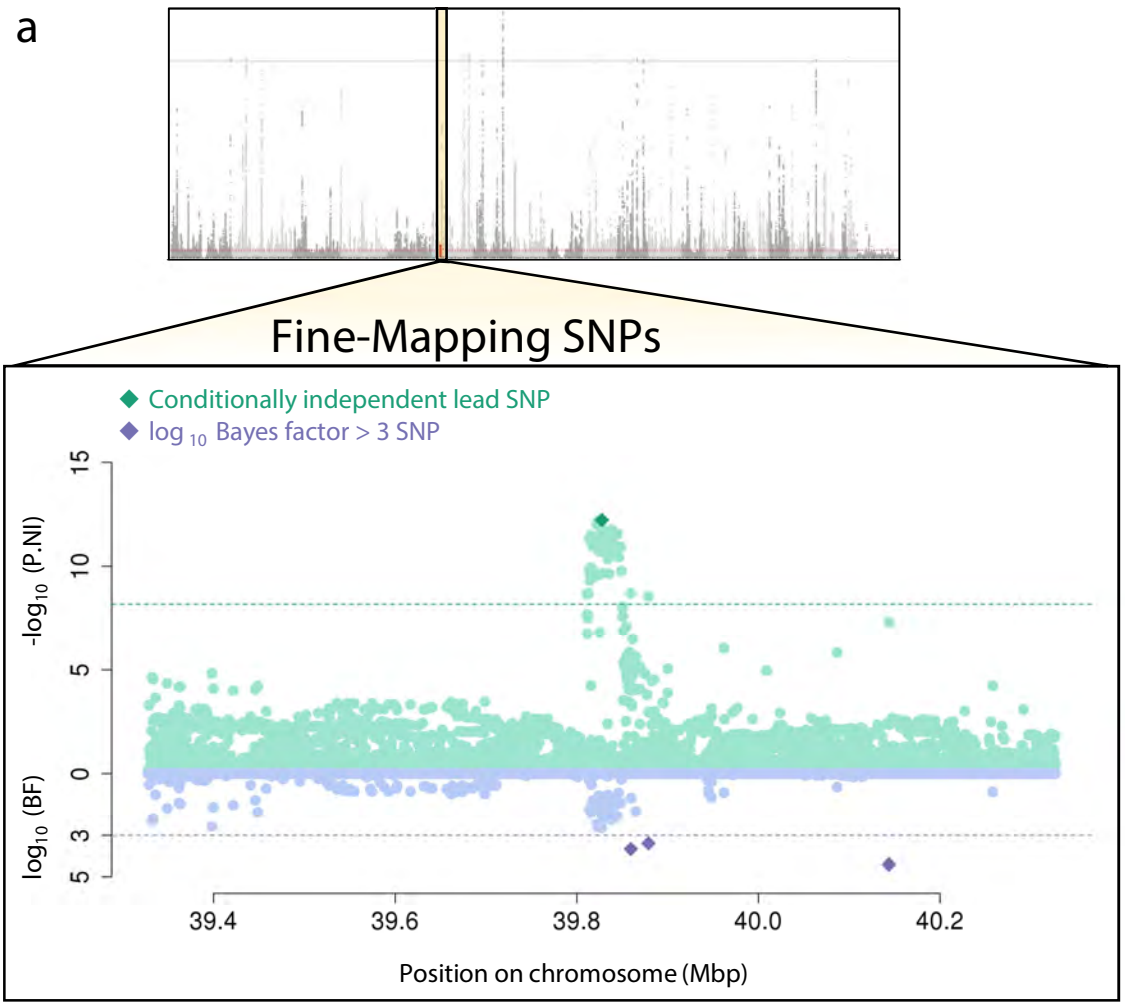
All relevant ethical regulations were complied with for human- and mouse-based research.

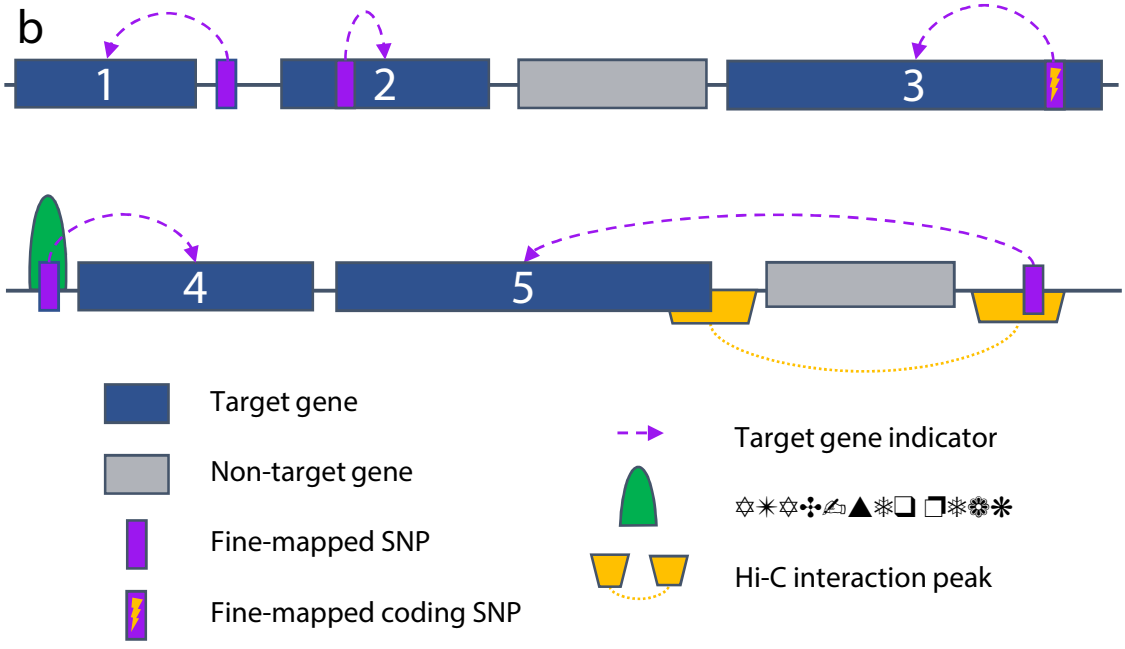
A **Life Sciences Reporting Summary** for this paper is available.

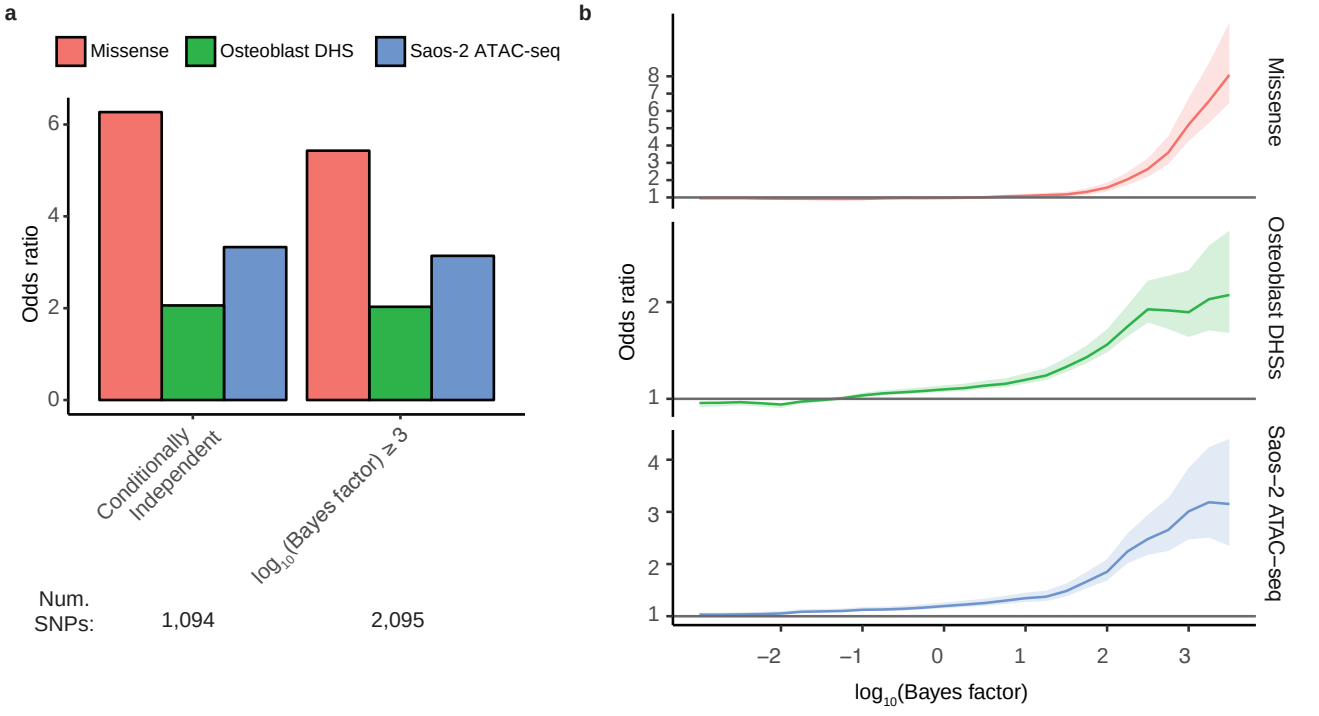
30. Sudlow, C. et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 12, e1001779 (2015).
31. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. 47, 284–290 (2015).
32. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
33. Winkler, T. W. et al. EasyStrata: Evaluation and visualization of stratified genome-wide association meta-Analysis data. Bioinformatics 31, 259–261 (2015).
34. Cochran, W. G. The Combination of Estimates from Different Experiments. Biometrics 10, 101 (1954).
35. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
36. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
37. Bulik-Sullivan, B. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
38. Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. Nat. Rev. Genet. 15, 765–776 (2014).
39. Chapman, J. M., Cooper, J. D., Todd, J. A. & Clayton, D. G. Detecting disease associations due to linkage disequilibrium using haplotype tags: A class of tests and the determinants of statistical power. Hum. Hered. 56, 18–31 (2003).
40. Spencer, C. C. A., Su, Z., Donnelly, P. & Marchini, J. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 5, e1000477 (2009).
41. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).
42. Hart, T., Komori, H. K., LaMere, S., Podshivalova, K. & Salomon, D. R. Finding the active genes in deep RNA-seq gene expression studies. BMC Genomics 14, 778 (2013).
43. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
44. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
46. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
47. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
48. Schmitt, A. D. et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. 17, 2042–2059 (2016).
49. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).
50. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods 123, 56–65 (2017).
51. Servant, N. et al. HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259 (2015).
52. Heinz, S. et al. Simple Combinations of Lineage-Determining Transcription Factors

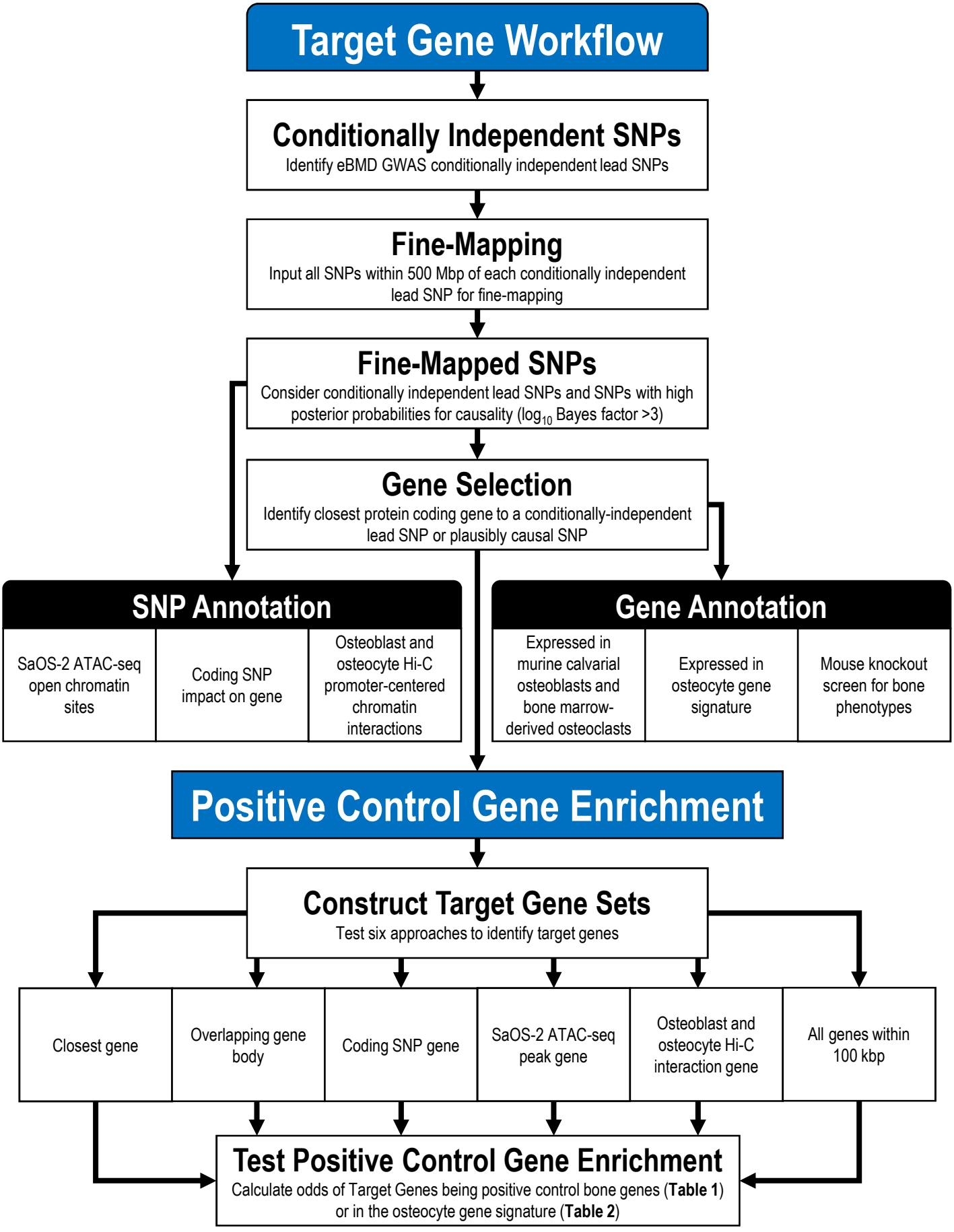
Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell 38, 576–589 (2010).
53. Mifsud, B. et al. GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. PLoS One 12, e0174744 (2017).
54. Durand, N. C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 3, 95–98 (2016).
55. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–D745 (2016).
56. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).
57. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34, 184–191 (2016).
58. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 17, 148 (2016).

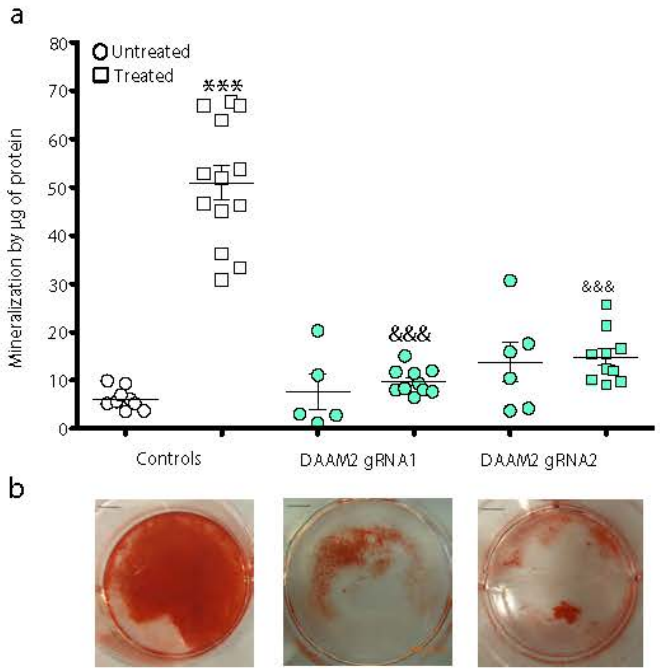

| NOVEL | PREVIOUS GWS | PREVIOUS NS |
2000
1500
1000
500
100
80
60
40
20
0
-log10(P.NI)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome number

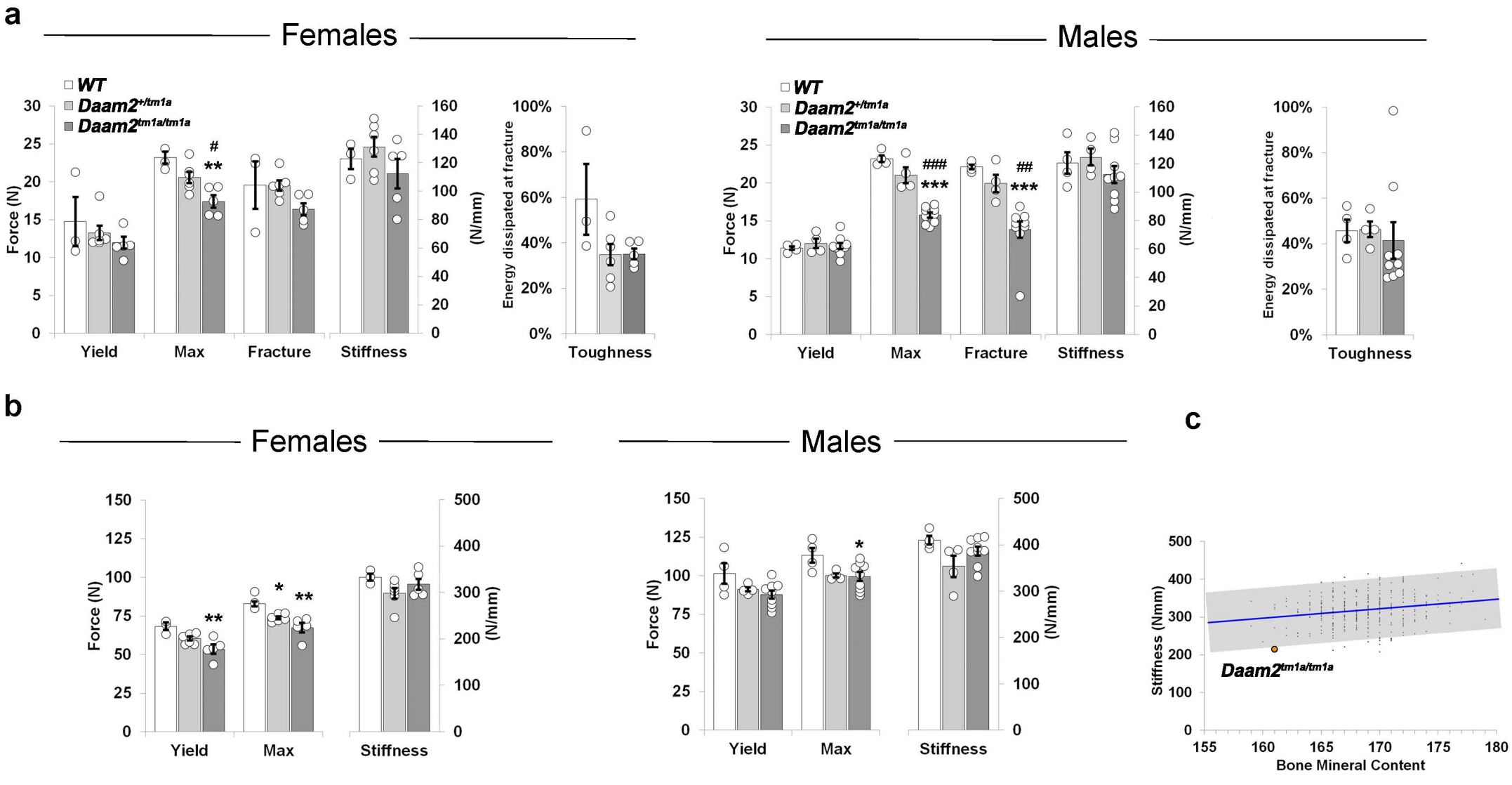

a
Fine-Mapping SNPs
Conditionally independent lead SNP
log 10 Bayes factor > 3 SNP
-log10 (P.NI)
log10 (BF)
15
10
5
0
3
5
39.4
39.6
39.8
40.0
40.2
Position on chromosome (Mbp)


b
1
2
3
4
5
Target gene
Non-target gene
Fine-mapped SNP
Fine-mapped coding SNP
Target gene indicator
Hi-C interaction peak


a
Missense
Osteoblast DHS
Saos-2 ATAC-seq
Odds ratio
0
2
4
6
Conditionally Independent
log10(Bayes factor) ≥ 3
Num. SNPs:
1,094
2,095
b
Odds ratio
Missense
Osteoblast DHSs
Saos−2 ATAC−seq
log10(Bayes factor)
−2
−1
0
1
2
3


Target Gene Workflow
Conditionally Independent SNPs
Identify eBMD GWAS conditionally independent lead SNPs
Fine-Mapping
Input all SNPs within 500 Mbp of each conditionally independent lead SNP for fine-mapping
Fine-Mapped SNPs
Consider conditionally independent lead SNPs and SNPs with high posterior probabilities for causality (log_{10} Bayes factor >3)
Gene Selection
Identify closest protein coding gene to a conditionally-independent lead SNP or plausibly causal SNP
SNP Annotation
SaOS-2 ATAC-seq open chromatin sites
Coding SNP impact on gene
Osteoblast and osteocyte Hi-C promoter-centered chromatin interactions
Gene Annotation
Expressed in murine calvarial osteoblasts and bone marrow-derived osteoclasts
Expressed in osteocyte gene signature
Mouse knockout screen for bone phenotypes
Positive Control Gene Enrichment
Construct Target Gene Sets
Test six approaches to identify target genes
Closest gene
Overlapping gene body
Coding SNP gene
SaOS-2 ATAC-seq peak gene
Osteoblast and osteocyte Hi-C interaction gene
All genes within 100 kbp
Test Positive Control Gene Enrichment
Calculate odds of Target Genes being positive control bone genes (Table 1) or in the osteocyte gene signature (Table 2)


a
Untreated
Treated
Mineralization by µg of protein
80
70
60
50
40
30
20
10
0
***
&&&
&&&
Controls
DAAM2 gRNA1
DAAM2 gRNA2
b


a
Females
WT
Daam2+/tm1a
Daam2tm1a/tm1a
Force (N)
(N/mm)
Yield
Max
Fracture
Stiffness
Energy dissipated at fracture
Toughness
#
**
Males
###
***
##
b
Females
Males
*
c
Stiffness (Nmm)
Bone Mineral Content
Daam2tm1a/tm1a