bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Transcriptome-wide Association Study and eQTL colocalization identify potentially causal genes responsible for bone mineral density GWAS associations

Basel M. Al-Barghouthi[1,2], Will T. Rosenow[1], Kang-Ping Du[3], Jinho Heo[4], Robert Maynard[5], Larry Mesner[1,6], Gina Calabrese[1], Aaron Nakasone[7], Bhavya Senwar[8], Louis Gerstenfeld[9], Virginia Ferguson[8], Cheryl Ackert-Bicknell[5], Elise Morgan[7], David L. Brautigan[4] and Charles R. Farber[1,2,6]

[1] Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA 22908
[2] Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA 22908
[3] Department of Radiation Oncology, University of Virginia, Charlottesville, VA 22908
[4] Department of Microbiology, Immunology, and Cancer Biology, School of Medicine, University of Virginia, Charlottesville, VA 22908
[5] Department of Orthopedics, Anschutz Medical Campus, University of Colorado, Aurora, CO 80045
[6] Department of Public Health Sciences, School of Medicine, University of Virginia, Charlottesville, VA 22908
[7] Department of Mechanical Engineering, Boston University, Boston, MA 02215
[8] Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309
[9] Department of Orthopaedic Surgery, Boston University Medical Center, Boston, MA 02118

Correspondence should be addressed to:
Charles R. Farber
Center for Public Health Genomics
P.O. Box 800717
University of Virginia
Charlottesville, VA 22908
Tel: 434-243-8584
Fax: 434-982-1815
Email: crf2s@virginia.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract:

Genome-wide association studies (GWASs) for bone mineral density (BMD) have identified over 1,100 associations to date. However, identifying causal genes implicated by such studies has been challenging. Recent advances in the development of transcriptome reference datasets and computational approaches such as transcriptome-wide association studies (TWASs) and expression quantitative trait loci (eQTL) colocalization have proven to be informative in identifying putatively causal genes underlying GWAS associations. Here, we used TWAS/eQTL colocalization in conjunction with transcriptomic data from the Genotype-Tissue Expression (GTEx) project to identify potentially causal genes for the largest BMD GWAS performed to date. Using this approach, we identified 512 genes as significant (Bonferroni <= 0.05) using both TWAS and eQTL colocalization. This set of genes was enriched for regulators of BMD and members of bone relevant biological processes. To investigate the significance of our findings, we selected *PPP6R3*, the gene with the strongest support from our analysis which was not previously implicated in the regulation of BMD, for further investigation. We observed that *Ppp6r3* deletion in mice decreased BMD. In this work, we provide an updated resource of putatively causal BMD genes and demonstrate that *PPP6R3* is a putatively causal BMD GWAS gene. These data increase our understanding of the genetics of BMD and provide further evidence for the utility of combined TWAS/colocalization approaches in untangling the genetics of complex traits.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction:

Osteoporosis, a disease characterized by low bone mineral density (BMD), decreased bone strength, and an increased risk of fracture, affects over 10 million individuals in the U.S. [1,2]. BMD is the single strongest predictor of fracture and a highly heritable quantitative trait [3 4 5]. Over the last decade, genome-wide association studies (GWASs) have identified over 1,100 independent associations for BMD [6–8]. However, despite the success of GWAS, few of the underlying causal genes have been identified [9,10].

One of the main difficulties in GWAS gene discovery is that the vast majority (>90%) of associations are driven by non-coding variation [11,12]. Over the last decade, approaches such as transcriptome-wide association studies (TWASs) and expression quantitative trait locus (eQTL) colocalization, have been developed which leverage transcriptomic data in order to inform gene discovery by connecting non-coding disease associated variants to changes in transcript levels [13–17]. These approaches have proven successful for a wide array of diseases and disease-associated quantitative traits [15,18,19]. However, the osteoporosis field has lagged behind such efforts, due to the limited number of large-scale bone-related transcriptomic datasets.

In a TWAS, genetic predictors of gene expression (e.g., local eQTL - sets of genetic variants that influence the expression of a gene in close proximity [20]) identified in a reference population (e.g., the Genotype-Tissue Expression (GTEx) project [21]) are used to impute gene expression in a GWAS cohort. Components of gene expression due to genetic variation are then associated with a disease or disease-associated quantitative trait. Genes identified by TWAS are often located in GWAS associations, suggesting that the genetic regulation of their expression is the mechanism underlying such associations. Several tools (e.g., FUSION, PrediXcan, MultiXcan [13,22,23]) have been developed to perform TWASs. Most of these tools use GWAS summary statistics, making TWAS widely applicable to large GWAS datasets. In contrast, eQTL colocalization is a statistical approach that determines if there is a shared genetic basis for two associations (e.g., a local eQTL and BMD GWAS locus). Recently, it has been demonstrated that prioritizing genes using both TWAS and eQTL colocalization provides a way to identify genes with the strongest support for causality [14,15].

The GTEx project has generated RNA-seq data on over 50 tissues across hundreds of individuals [24]. Even though data on the tissues/cell-types likely to be most relevant to BMD (bone or bone cells) were not included, the project demonstrated that many expression quantitative trait loci (eQTL) were shared across tissues [24,25]. Additionally, it is well known that effects in a wide-array of non-bone cell-types and tissues can impact bone and BMD [26,27]. As a result, we sought to use the GTEx resource in conjunction with TWAS and eQTL colocalization to leverage non-bone gene expression data to identify putatively causal genes underlying BMD GWAS.

Here, we performed TWAS and eQTL colocalization using the GTEx resource and the largest BMD GWAS performed to date to identify potentially causal genes [6]. Using this approach, we identified 512 genes significantly associated via TWAS with a significant colocalizing eQTL. To investigate the significance of our findings we selected Protein Phosphatase 6 Regulatory

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Subunit 3 (*PPP6R3*), the gene with the strongest support not previously implicated in the regulation of BMD, for further investigation. We demonstrate using mutant mice that *Ppp6r3* is a regulator of lumbar spine BMD. These results highlight the power of leveraging GTEx data, even in the absence of data from the most relevant tissue/cell-types, to increase our understanding of the genetic architecture of BMD.

Results:

TWAS and eQTL colocalization identify potentially causal BMD GWAS genes:

To identify potentially causal genes responsible for BMD GWAS associations, we combined TWAS and eQTL colocalization using GTEx data (**Figure 1A**). We began by performing a TWAS using reference gene expression predictions from GTEx (Version 8; 49 tissues) and the largest GWAS performed to date for heel estimated BMD (eBMD) (>1,100 independent associations) [6,24] . The analysis was performed using S-MultiXcan, which allowed us to leverage information across all 49 GTEx tissues [23]. Our analysis focused on protein-coding genes (excluded non-coding genes). A total of 2,156 protein-coding genes were significantly (Bonferroni-adjusted P-value <= 0.05) associated with eBMD (**Supplementary Table 1)**.

Next, we identified colocalizing eQTL from each of the 49 tissues in GTEx using fastENLOC [15,17]. We identified 1,182 colocalizing protein-coding genes with a regional colocalization probability (RCP) of 0.1 or greater (**Supplementary Table 2).** In total, 512 protein-coding genes were significant in both the TWAS and eQTL colocalization analyses (**Table 1 and Supplementary Table 3)**. Among the identified genes were many with well-known roles in the regulation of BMD, such as *RUNX2* (**Figure 1B**)*, IGF1,* and *LRP6*, as well as novel genes such as *RERE* (**Figure 1C**)*.* Overall, the identified genes had significantly colocalizing eQTL in all 49 GTEx tissues, with eQTL from cultured fibroblasts (132 genes), subcutaneous adipose tissue (117 genes), tibial artery (115 genes) and tibial nerve (114 genes) exhibiting the highest number of significant colocalizations **(Supplementary Table 4)**. TWAS predictors were only generated for genes on autosomes and of the 1,103 independent associations identified by Morris, et al.[6] , 1,097 were autosomal. For each of these, we defined a locus as the region consisting of +/- 1 Mbp around each association. Of the 1,097 loci, almost half (542; 49%) of the loci contained at least one of the 512 prioritized genes. Most loci overlapped one gene (mean = 1.7, median = 1); however, 184 loci overlapped multiple genes, including a locus on Chromosome (Chr.) 20 (lead SNP rs6142137) which contained 9 prioritized genes. (**Figure 1D**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Top 10 protein-coding genes significant by colocalization (RCP >=0.1) and TWAS, sorted by TWAS P-value.

Gene	Tissue with greatest RCP	Max. RCP	TWAS P-value (Bonferroni)
SPTBN1	Cells_Cultured_Fibroblasts	0.9469	$<5 \times 10^{-324}$
CCDC170	Spleen	0.6582	$<5 \times 10^{-324}$
FAM3C	Artery_Tibial	0.4917	$<5 \times 10^{-324}$
SEPT5	Skin_Sun_Exposed	0.4868	2.26×10^{-286}
FGFRL1	Cells_Cultured_Fibroblasts	0.1611	5.31×10^{-272}
GREM2	Cells_Cultured_Fibroblasts	0.9998	4.31×10^{-257}
GPATCH1	Whole_Blood	0.3564	3.44×10^{-226}
RHPN2	Pituitary	0.2181	8.71×10^{-221}
BMP4	Brain_Cortex	0.5468	5.49×10^{-169}
RUNX2	Esophagus_Gastroesophageal_Junction	0.2372	1.99×10^{-146}

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

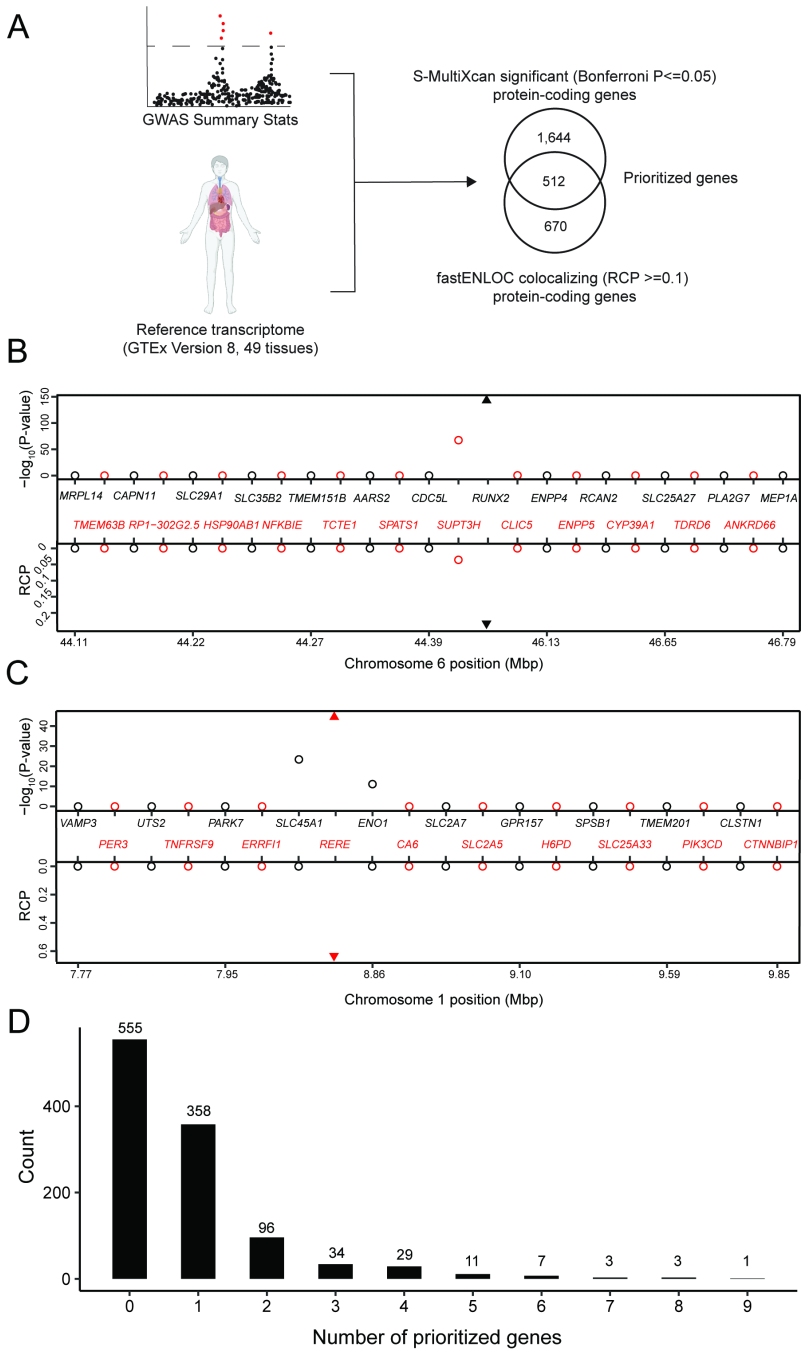

Figure 1. TWAS and eQTL colocalization identify potentially causal BMD GWAS genes. A) Overview of the analysis. The human image was obtained from BioRender.com. TWAS/colocalization plot for genes in the locus around *RUNX2* **(B)** and *RERE* **(C)**. The $-\log_{10}$ Bonferroni-adjusted P-values from the TWAS analysis (top panel) and the maximum RCPs from the colocalization analyses (bottom panel). Genes alternate in color for visual clarity. Triangles represent *RUNX2* **(B)** and *RERE* **(C)**. **D)** Distribution of prioritized genes within eBMD GWAS loci.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Characterization of genes identified by TWAS/eQTL colocalization:
To evaluate the ability of the combined TWAS/colocalization approach to identify genes previously implicated in the regulation of BMD, other bone traits, or the activity of bone cells, we queried the presence of "known bone genes" within the list of the 512 prioritized protein-coding genes. To do so, we created a database-curated set of genes previously implicated in the regulation of bone processes (henceforth referred to as our "known bone genes" list, N=1,399, **Supplementary Table 5)**. Of the 512 genes identified above, 66 (12.9%) were known bone genes, representing a significant enrichment (odds ratio = 1.72; P = 1.0×10^{-4}) over what would be expected by chance (**Supplementary Table 6)**.

We also performed a Gene Ontology (GO) enrichment analysis of the 512 prioritized genes. We observed enrichments in several bone-relevant ontologies, such as "regulation of ossification" (P = 2.6×10^{-5}), "skeletal system development" (P = 2.8×10^{-5}), and "regulation of osteoblast differentiation" (P = 8.7×10^{-5}) (**Figure 2A**, **Supplementary Table 7**).

In order to compare our approach with the approach of prioritizing the closest genes to GWAS associations as potentially causal, we quantified the number of genes that were both the closest genes to eBMD GWAS associations and were members of the "known bone gene" list. Of the 863 genes that were the closest genes to eBMD GWAS associations, 139 were members of the "known bone gene list", representing a more significant enrichment of "known bone genes" than our prioritization approach (OR = 2.56, P = 6.37×10^{-19}). Of our 512 prioritized genes, 206 (40%) were also the closest genes to eBMD GWAS associations, with 27 of the remaining 306 prioritized genes (8.8%) being members of the "known bone gene" list.

The International Mouse Phenotype Consortium (IMPC) has recently measured whole body BMD in hundreds of mouse knockouts [28,29]. We searched the IMPC database for any of the 512 genes identified by TWAS and eQTL colocalization. Of the 512, 142 (27.7%) had been tested by the IMPC and 64 (12.5% of the 512 prioritized genes, 45% of the 142 IMPC-tested genes) had a nominally significant (P<=0.05) alteration of whole-body BMD in knockout/knockdown mice, compared to controls. Of the 64, 49 (76.5%) were not members of the "known bone gene" list.

An example of one of the 64 genes is *GPATCH1*, located within a GWAS association on human chromosome *19q13.11*. Of all the genes in the region, *GPATCH1* had the strongest TWAS association (P=3.44×10^{-226}) (**Figure 2B**) and the strongest eQTL colocalization (whole blood, RCP=0.36) (**Figures 2B-D**). The eQTL and BMD GWAS allele effects for the top SNPs were in the same direction, suggesting that decreasing the expression of *GPATCH1* would lead to decreased BMD. BMD data from the IMPC showed that female mice heterozygous for a *Gpatch1* null allele had decreased BMD (P=2.17×10^{-8}) (**Figure 2E**). Together, these data suggest that many of the genes identified by the combined TWAS/colocalization approach are likely causal BMD GWAS ge

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

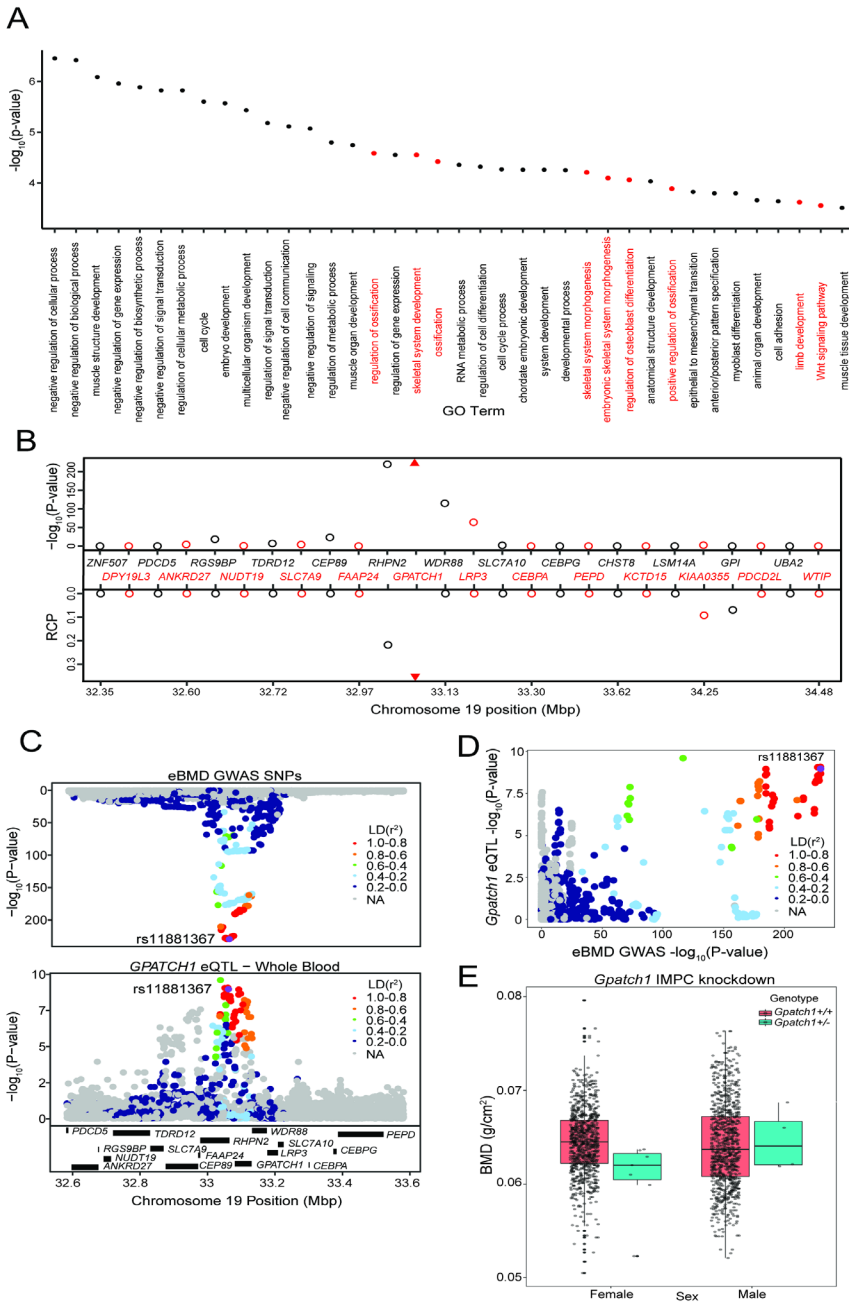

Figure 2. TWAS and eQTL colocalization identify *Gpatch1* a potentially causal BMD GWAS gene. A) The top 40 terms from a Gene Ontology analysis of the 512 potentially causal BMD genes identified by our analysis. Terms with clear relevance to bone are highlighted in red. Only terms from the "Biological Process" (BP) sub-ontology are listed. **B)** TWAS/colocalization plot for genes in the locus around *GPATCH1* (+/- 1.5 Mbp). The $-\log_{10}$ Bonferroni-adjusted P-values from the TWAS analysis (top panel) and the maximum RCPs from the colocalization analyses (bottom panel). Genes alternate in color for visual clarity. Triangles represent *GPATCH1*. **C)** Mirrorplot showing eBMD GWAS locus (top panel) and colocalizing *GPATCH1* eQTL in whole blood (bottom panel). SNPs are colored by their LD with rs11881367 (purple), the most significant GWAS SNP in the locus. **D)** Scatterplot of $-\log_{10}$ P-values for *GPATCH1* eQTL versus eBMD GWAS SNPs. SNPs are colored by their LD with rs11881367 (purple). **E)** Bone mineral density (BMD) in *Gpatch1* knockdown mice. N=7 females and N=4 males for *Gpatch1*$^{+/-}$ mice, N=880 females and N=906 males for *Gpatch1*$^{+/+}$ mice. Boxplots indicate the median (middle line), the 25th and 75th percentiles (box) and the whiskers extend to 1.5 * IQR.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PPP6R3 is a candidate causal gene for a GWAS association on Chr. 11:

To identify novel candidate genes for functional validation, we focused on genes with the strongest evidence of being causal. To do so, we increased the colocalization RCP threshold to 0.5, and then sorted genes based on TWAS Bonferroni-adjusted P-values. Furthermore, we constrained the list of candidates for functional validation to genes which were not members of the "known bone gene" list or genes with a nominal (P<=0.05) alteration in whole-body BMD as determined by the IMPC. This yielded 137 putatively causal BMD genes (**Table 2, Supplementary Table 8**).

Table 2. Top 10 novel protein-coding genes significant by colocalization (RCP >=0.5) and TWAS, sorted by TWAS P-value.

Gene	Tissue with greatest RCP	Max. RCP	TWAS P-value (Bonferroni)
SPTBN1	Cells_Cultured_fibroblasts	0.9469	$<5 \times 10^{-324}$
PPP6R3	Thyroid	0.5291	5.7×10^{-93}
BARX1	Colon_Transverse	0.7764	6.36×10^{-63}
MEOX2	Brain_Nucleus_accumbens_basal_ganglia	0.6286	3.21×10^{-53}
RERE	Adipose_Subcutaneous	0.6431	6.95×10^{-46}
SIPA1	Nerve_Tibial	0.9981	4.26×10^{-41}
CAPZB	Testis	0.6716	3.64×10^{-33}
B4GALNT3	Artery_Aorta	0.9241	2.67×10^{-33}
TRPC4AP	Breast_Mammary_Tissue	0.5577	8.62×10^{-31}
AXL	Minor_Salivary_Gland	0.6205	9.74×10^{-31}

Though it was not on the "known bone gene" list, the first gene ranked by TWAS P-value, *SPTBN1*, has been demonstrated to play a role in the regulation of BMD [30]. The second, *PPP6R3*, has not been previously implicated in the regulation of BMD. *PPP6R3* is located on human Chr. 11 within 1 Mbp of seven independent eBMD GWAS SNPs identified by Morris et al. [6] (subsequently referred to as "eBMD lead SNPs") (**Figure 3A**). Of all the protein-coding genes (N=29) in the ~1.8 Mbp region surrounding *PPP6R3*, its expression was the most significantly associated with eBMD by TWAS (Bonferroni = 5.7×10^{-93}) (**Figure 3B**). Furthermore, *PPP6R3* was the only gene in the region with eQTL (in four GTEx tissues, thyroid, ovary, brain_putamen_basal_ganglia, and stomach with RCPs = 0.53, 0.50, 0.28 and 0.14, respectively) that colocalized with at least one of the eBMD associations (**Figure 3B**). Based on these data, we chose to further investigate *PPP6R3* as a potentially causal BMD gene.

We first determined which of the seven associations colocalized with the *PPP6R3* eQTL (**Figure 3C**). The most significant *PPP6R3* eQTL SNP in thyroid tissue (the tissue with the highest RCP) was rs10047483 (Chr. 11, 68.464237 Mbp) (*PPP6R3* eQTL P = 6.99×10^{-8}, eBMD GWAS P = 1.2×10^{-100}) located in intron 1 of *PPP6R3*. The most significant eBMD lead SNP in the locus was rs11228240 (Chr. 11, 68.450822 Mbp, eBMD GWAS P = 6.6×10^{-101}, *PPP6R3* eQTL P = 3.7×10^{-6}), located upstream of *PPP6R3*. Consistent with the colocalization analysis, these two variants are in high LD (r^2=0.941) and rs10047483 does not exhibit strong LD ($r^2 < 0.104$) with

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

any of the other six eBMD lead SNPs in the locus. The eQTL and BMD GWAS allele effects for rs10047483 were opposing, suggesting that a decrease in the expression of *PPP6R3* would lead to an increase in BMD.

A recent fracture GWAS identified 14 significant associations, one of which was located in the *PPP6R3* region (rs35989399, Chr. 11, 68.622433 Mbp) [6]. We analyzed the fracture GWAS in the same manner as we did above for eBMD. We found that *PPP6R3* expression when analyzed by TWAS was significant for fracture (TWAS Bonferroni-pval = 6.0×10^{-3}) and the same *PPP6R3* eQTL colocalized with the fracture association (RCP = 0.49 in ovary, RCP = 0.36 in thyroid) (**Figure 3D**). Together, these data highlight *PPP6R3* as a strong candidate for one of the seven eBMD/fracture associations in this region.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

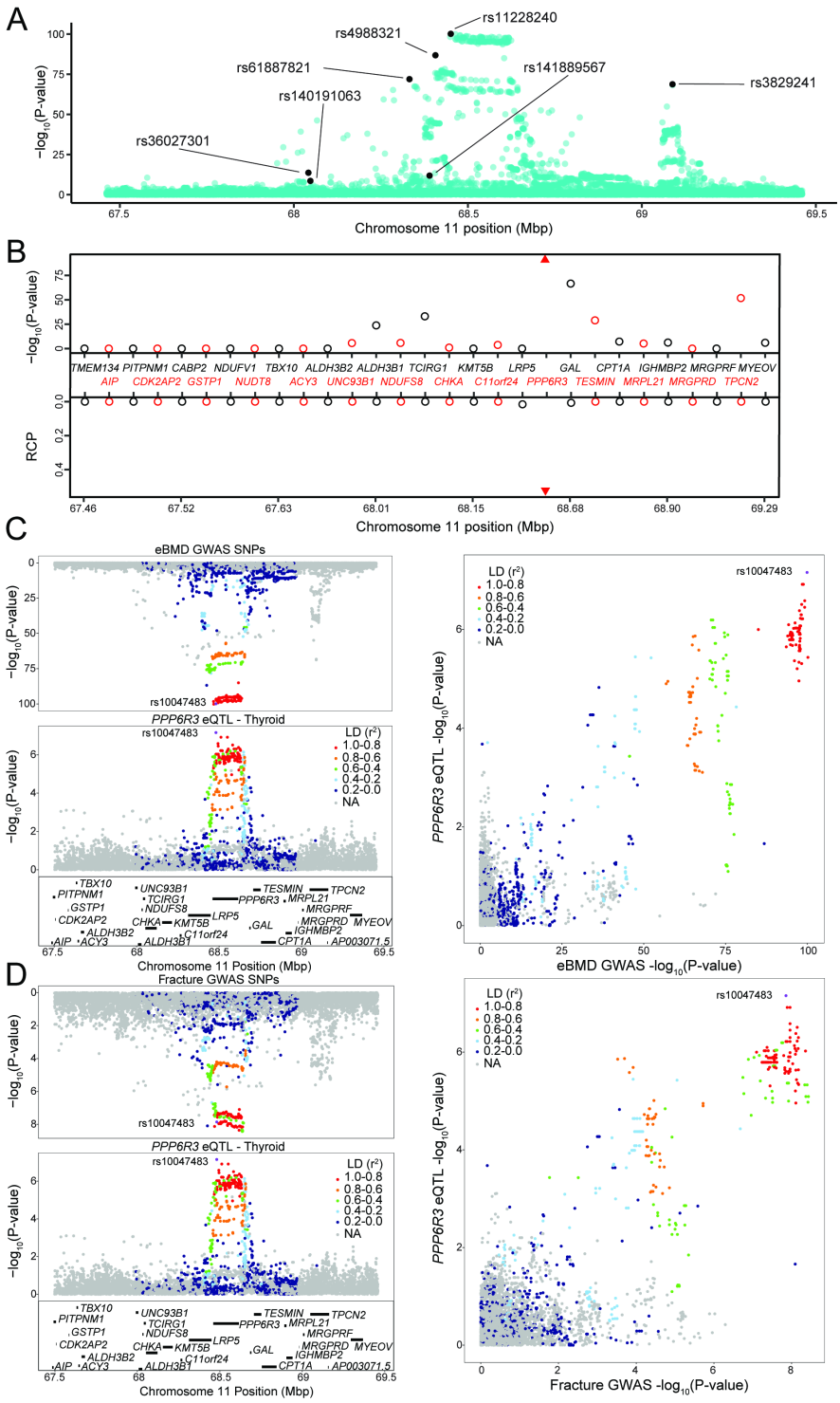

Figure 3. PPP6R3 is a top-10 novel eBMD gene. **A)** eBMD GWAS SNPs around the *PPP6R3* locus (+/- 1 Mbp). The y-axis represents $-\log_{10}$ eBMD GWAS P-values. Highlighted SNPs (black) are the seven lead eBMD GWAS SNPs in the locus. **B)** TWAS/colocalization plot for genes in the locus around *PPP6R3* (+/- 1 Mbp). The $-\log_{10}$ Bonferroni-adjusted P-values from the TWAS analysis (top panel) and the maximum RCPs from the colocalization analyses (bottom panel). Genes alternate in color for visual clarity. Triangles represent *PPP6R3*. Mirrorplot of the eBMD locus **(C)** and *PPP6R3* eQTL in thyroid, and fracture locus and *PPP6R3* eQTL in thyroid **(D)**. The panels on the right are scatterplots of $-\log_{10}$ P-values for *PPP6R3* eQTL and eBMD GWAS SNPs **(C)** and the *PPP6R3* eQTL and fracture GWAS SNPs **(D)**. SNPs are colored by their LD with rs10047483 (purple), the most significant *PPP6R3* eQTL in the locus. Not all genes are shown.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

***PPP6R3* is a regulator of femoral geometry, BMD, and vertebral microarchitecture:**

To assess the effects of *PPP6R3* expression on bone phenotypes, we characterized mice harboring a gene trap allele (*Ppp6r3*$^{tm1a(KOMP)Wtsi}$) (**Figure 4A**). We intercrossed mice heterozygous for the mutant allele to generate mice of all three genotypes (wild-type (WT), heterozygous (HET), and mutant (MUT)). The absence of PPP6R3 protein in MUT mice was confirmed through Western blotting (**Figure 4B**).

The BMD analyses presented above used heel eBMD GWAS data. We used these data because they represent the largest, most well-powered BMD GWAS to date [7]. However, to determine whether perturbation of *Ppp6r3* would be expected to impact femoral or lumbar spine BMD in a similar manner, we turned to a smaller GWAS to look at both of these traits. In a GWAS by Estrada et al. [7], a total of 56 loci were identified for femoral neck (FNBMD) and lumbar spine (LSBMD) BMD. One of the 56 loci corresponded to the same SNPs associated with the *PPP6R3* eQTL. The locus was significant for LSBMD; however, it did not reach genome-wide significance for FNBMD (**Supplemental Figure 1**).

We evaluated BMD at both the femur and the lumbar spine in *Ppp6r3*$^{tm1a(KOMP)Wtsi}$ mice, with the expectation, based on the above data, that perturbation of *Ppp6r3* would have a stronger impact on BMD at the lumbar spine. At approximately 9 weeks of age, we measured areal BMD (aBMD) at the femur and lumbar spine using dual X-ray absorptiometry (DXA). First, we observed no change in body weight at 9 weeks that might impact bone phenotypes (**Supplemental Figure 2A**). As the above analysis predicted, we observed a significant effect of *Ppp6r3* genotype on aBMD at the lumbar spine (WT vs. MUT P=0.01, **Figure 4C**), but not the femur (WT vs. MUT P=0.26, **Figure 4D**). It should also be noted, however, that we observed significantly decreased femoral width, but not length, in *Ppp6r3* mutant mice (anterior-posterior (AP) femoral width, WT vs. MUT P=0.02; medial-lateral (ML) femoral width, WT vs. MUT P=2.2 x 10^{-6}, **Supplemental Figures 2B-D**).

Due to the significant effect of *Ppp6r3* genotype on lumbar spine aBMD, we further characterized the effects of *Ppp6r3* genotype on microarchitectural phenotypes via micro-computed tomography (μCT). We observed significant (P<=0.05) decreases in trabecular bone volume fraction (BV/TV, WT vs. MUT P=0.015, **Figure 4E-F**) and volumetric BMD (vBMD, WT vs. MUT P=0.015, **Figure 4G**) of the lumbar spine as a function of *Ppp6r3* genotype, but found no significant changes in tissue mineral density (TMD, **Supplemental Figure 2E**), trabecular separation (TbSp), trabecular thickness (TbTh) or trabecular number (TbN) (**Figures 4 H-J**).

Finally, in order to assess the effects of *Ppp6r3* genotype on bone matrix composition, we performed periosteal Raman spectroscopy on both the lumbar spines and femurs. We did not observe any significant (P <= 0.05) effects of *Ppp6r3* genotype on bone matrix composition (**Supplemental Figures 3-6**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

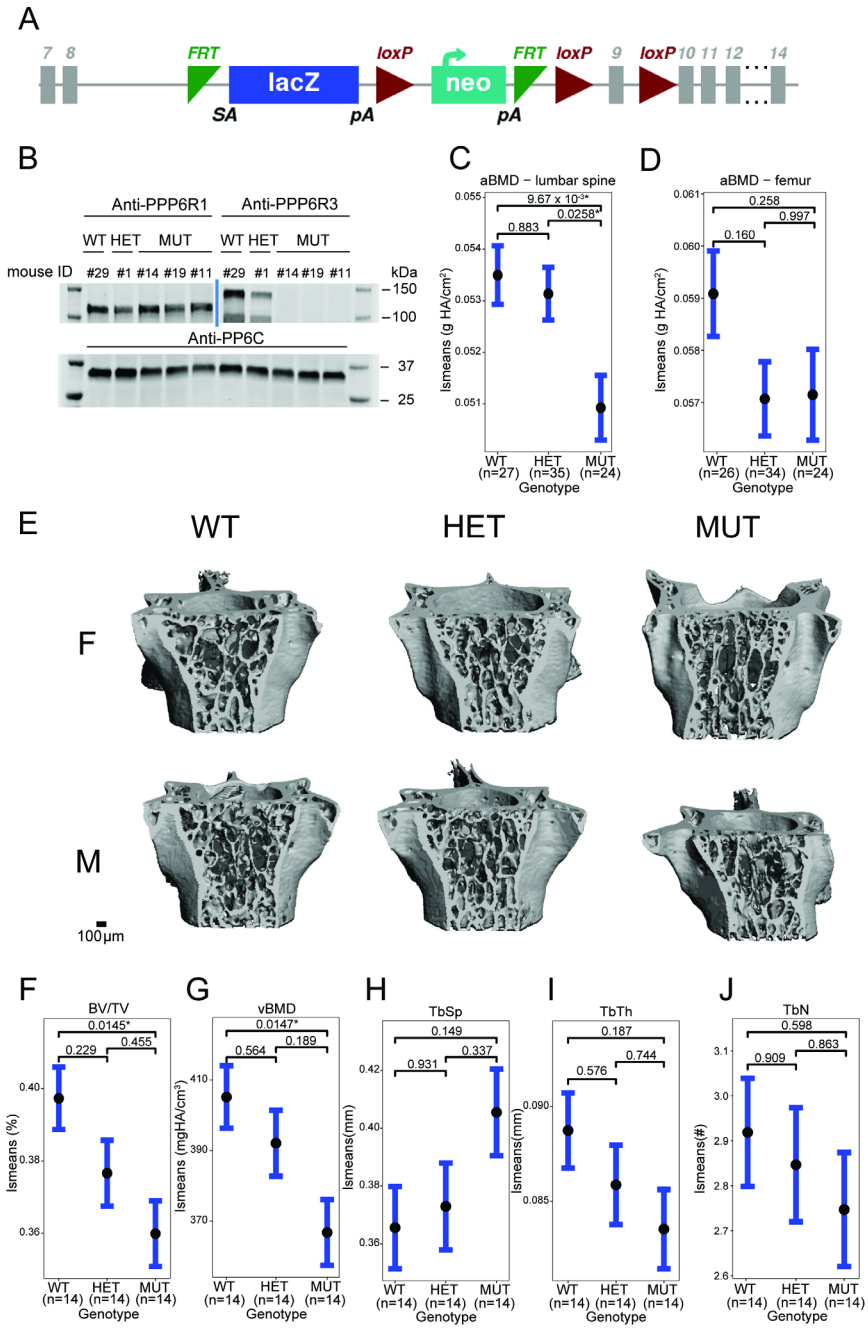

Figure 4. *Ppp6r3* functional validation shows an effect of genotype on bone mass. A) Schematic of the *Ppp6r3* gene-trap allele (*Ppp6r3*$^{tm1a(KOMP)Wtsi}$). Image obtained from the IMPC. **B)** Western blot of the *Ppp6r3* experimental mice. Top left panel shows that PPP6R1 protein (control) levels are not affected by the *Ppp6r3* gene-trap allele. Top right panel shows the effect of the gene-trap allele on PPP6R3 protein levels. The two bands are ostensibly due to different PPP6R3 isoforms. Bottom panel shows that PP6C protein (control) levels are not affected by the *Ppp6r3* gene-trap allele. Least-squares means for spinal (**C**) and femoral (**D**) areal BMD (aBMD) DXA in *Ppp6r3* wild-type (WT), heterozygous (HET), and mutant (MUT) mice. Contrast P-values, adjusted for multiple comparisons are presented. *P<=0.05. **E)** Representative images of vertebrae for the *Ppp6r3* experimental mice. Scale is shown on the bottom right. **F-J)** Least-squares means for µCT measurements in the lumbar spines of *Ppp6r3* WT, HET, and MUT mice. Contrast P-values, adjusted for multiple comparisons are presented. *P<=0.05. Abbreviations: BV/TV - bone volume fraction, vBMD – volumetric bone mineral density, TbSp – trabecular separation, TbTh – trabecular thickness, TbN – trabecular number.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion:

BMD GWASs have identified over 1,100 associations to date. However, identifying causal genes remains a challenge. To aid researchers in further dissecting the genetics of complex traits, reference transcriptomic datasets and computational methods have been developed for the prioritization and identification of causal genes underlying GWAS associations. In this work, our goal was to utilize these data and tools to prioritize putatively causal genes underlying BMD GWAS associations. Specifically, we used the GTEx eQTL reference dataset in 49 tissues to perform TWAS and eQTL colocalization on the largest BMD GWAS. Using this approach, we identified 512 putatively causal protein-coding genes that were significant in both the TWAS and colocalization approaches.

Our approach was inspired by a recent study that used the GTEx resource and a TWAS/eQTL colocalization approach similar to the one we employed. Pividori et al. [15] recently combined TWAS and eQTL colocalization to GTEx and GWAS data on 4,091 traits, including BMD, from the UK Biobank data. A total of 76 protein-coding genes were identified and of the 76, we identified 55 (72.4%) of the same genes in our implementation. There are several reasons for this discrepancy in the number of prioritized genes. First, both studies used a GWAS based on the UK BioBank [31]; however, there were significant differences in sample size. The PhenomeXcan project utilized GWAS data based on the analysis of ~207,000 individuals, whereas we used GWAS data based on the analysis of ~426,000 individuals [6,15]. Second, the two GWAS studies utilized different association models. Finally, due to the breadth of the PhenomeXcan project, they had a higher multiple-testing burden than we did, which led to different Bonferroni-adjusted P-value thresholds ($P<5.49 \times 10^{-10}$ vs. $P<=2.38 \times 10^{-6}$).

One of many novel genes identified in our study was *PPP6R3*, which was also identified in the PhenomeXcan project [15]. *PPP6R3* is a regulatory subunit of protein phosphatase 6 and has been implicated in several cancers ([32,33]). In humans, the PPP6R3 protein shows ubiquitous expression across tissues, and may have an important role in maintaining immune self-tolerance [33]. It is unclear how *PPP6R3* may be influencing BMD. However, Protein Phosphatase 6 has been shown to oppose activation of the nuclear factor kappa-light-chain enhancer of activated B cells (NF-κB) pathway in lymphocytes [34]. Since the NF-κB signaling pathway is highly involved in osteoclastogenesis and bone resorption, it is possible that *PPP6R3* may be involved in the regulation of this pathway in osteoclasts [35]. Further studies that characterize the role of *PPP6R3*, and the effects of its deletion, in bone cells are required to further elucidate its effect on BMD.

The *PPP6R3* locus demonstrated a high level of complexity, containing seven independent GWAS associations, at least one of which was also associated with fracture. Interestingly, just upstream of *PPP6R3* is *LRP5*, a WNT signaling co-receptor [36]. *LRP5* is a well-known regulator of BMD and gain and loss of function mutations lead to high bone mass syndrome and osteoporosis pseudoglioma, respectively [37–40]. *LRP5* expression was not significantly associated with eBMD by TWAS (Bonferroni P= 1), nor did it have a colocalizing eQTL in GTEx tissues

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(most significant RCP=1.6 x 10^{-2} in pancreas). However, another eBMD lead SNP in the region, rs4988321, is a missense mutation in *LRP5* (Val667Met) that has been associated with BMD in multiple studies [41–43]. While this variant represents an association that is independent of the rs10047483 association (r^2 = 0.104), it further highlights the complexity of this locus both in terms of the number of associations as well as target genes.

To determine the effect of *Ppp6r3* expression on bone, we characterized bone phenotypes in mice harboring a gene-trap allele (*Ppp6r3*$^{tm1a(KOMP)Wtsi}$). Consistent with the observation that the *PPP6R3* eQTL SNPs were significantly associated with lumbar spine, but not femoral neck BMD, we observed that *Ppp6r3* deletion had a significant effect on lumbar spine BMD, but not femoral BMD. Using µCT, we further characterized the effect of *Ppp6r3* deletion on lumbar spine microarchitecture. We observed significant decreases in trabecular bone volume fraction (BV/TV) and volumetric BMD of the lumbar spine as a function of *PPP6R3* genotype. While we did not observe significant effects of *Ppp6r3* deletion on trabecular thickness or number, the direction of effects for those phenotypes suggests that the observed decrease in bone volume fraction and BMD may be explained by the cumulative but more subtle effects of *Ppp6r3* deletion on trabecular thickness and number.

Our hypothesis regarding the directions of effect of *Ppp6r3* expression on BMD based on the eQTL and eBMD/lumbar spine BMD GWAS were opposite to what we observed. There are several reasons that may explain this. First, our hypothesis was based on expression data in non-bone tissues and cell-types. Recent studies have shown that the direction of eQTL effects can differ between different cells and tissues within humans [44,45]. Second, our hypothesis was based on human data, while our functional experiments were performed in mice. Third, we globally deleted *Ppp6r3* in mice, as opposed to ablating it in a bone-specific knockout. Future studies of the *PPP6R3* eQTL in bone cells as well as the generation of conditional *Ppp6r3* knockouts will allow us to unravel the precise role of this association and *PPP6R3* in the regulation of bone mass.

As we and others have shown, the use of both TWAS and eQTL colocalization can prioritize putatively causal genes underlying GWAS associations. Here, we have shown the utility of this approach even in the absence of eQTL data from the most phenotype-relevant tissue. However, it is important to highlight the limitations of our analysis. While studies have shown that many eQTL are shared among tissues, the lack of eQTL data in bone and bone cells means that bone-specific eQTL were missed. For example, a study conducted by Mullin et al performed eQTL colocalization and summary-based Mendelian Randomization (SMR) by utilizing GWAS data and expression data from osteoclast-like cells, and prioritized several eBMD genes [46]. 38% of the colocalizing eQTL and 19% of the SMR genes that they identified overlapped with our 512 prioritized genes, suggesting that we have missed many potential effector genes with eQTL specific to bone cells. In addition, the use of multiple non-bone tissues may have inflated the number of false positives based on coincidence of strong TWAS and eQTL colocalization signals that have no biological impact on bone. Furthermore, the lack of bone transcriptomic data may also explain the observed disparity between our hypothesized and observed direction-of-effect for *PPP6R3.* It is also important to note that due to the reliance of this approach on

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

eQTL data, genes that affect BMD via non-expression related mechanisms were not captured. Another limitation of our approach arises from the definition of loci based on linkage disequilibrium (LD). We used a set of previously-defined approximately independent LD blocks, derived from a cohort of European individuals, in our fastENLOC analysis [47]. The inexact nature of these data may lead to spurious colocalizations due to mismatches in LD structure between the reference LD blocks and the GWAS/eQTL populations. Additionally, because the GWAS and eQTL data have mismatching LD structures, due to their being derived from cohorts with different ancestries, our analyses, particularly the colocalization analyses, may suffer from reduced power [48] . This also raises the related issue of the reduced generalizability of our results in non-European individuals, which brings further attention to the necessity of performing GWASs and providing reference data in diverse and underrepresented populations. Additionally, another issue arises when considering correlations in expression, and predicted expression, between genes in a locus, which may lead to spurious associations in TWAS analyses [49]. Finally, as we show above, our method does not perform as well as prioritizing genes based on their proximity to GWAS associations. However, because our method utilizes systems genetics techniques and data, such as eQTL, we believe that our method prioritizes genes in a more biologically relevant manner. In fact, utilizing the closest gene method alone, *PPP6R3* would not have been prioritized as a bone-relevant gene. We suggest that future studies utilize both prioritization techniques, such as taking the closest genes to GWAS associations and cross referencing them with colocalizing and TWAS-associated genes, in order to provide further evidence for functional validation.

In summary, we applied a combined TWAS/colocalization approach using GTEx and identified 512 putatively causal BMD genes. We further investigated *PPP6R3* and demonstrated that it is a regulator of lumbar spine BMD. We believe this work provides a valuable resource for the bone genetics community and may serve as a framework for prioritizing genes underlying GWAS associations using publicly available tools and data for a wide range of diseases.

Methods:

fastENLOC colocalization:

For each of the eBMD and fracture GWASs, we performed colocalization using fastENLOC, by following the tutorial and guidelines available at https://github.com/xqwen/fastenloc.
Briefly, for each GWAS, we converted variant coordinates to the hg38 human genome assembly, using the UCSC liftOver tool (minimum ratio of bases that must remap = 1) [https://genome.ucsc.edu/cgi-bin/hgLiftOver]. We calculated z-scores by dividing GWAS betas by standard errors. We then defined loci based on European linkage disequilibrium (LD) blocks, as defined based on the results of Berisa and Pickrell, 2015 [47].

Z-scores were then converted to posterior inclusion probabilities (PIPs) using torus [50]. Finally, these data were colocalized with fastENLOC for all 49 GTEx V8 tissues, with the "-total_variants" flag set to 14,000,000. Colocalization was performed using pre-computed GTEx multi-tissue annotations, obtained from https://github.com/xqwen/fastenloc. Finally, to identify

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protein-coding genes in the results, we utilized Ensembl's "hsapiens_gene_ensembl" dataset using biomaRt (version 2.45.8).

S-MultiXcan:

We conducted a transcriptome-wide association study by integrating genome-wide SNP-level association summary statistics from an estimated bone mineral density GWAS [6] with GTEx version 8 gene expression QTL data from 49 tissue types. We used the S-MultiXcan approach for this analysis, to correlate gene expression across tissues to increase power and identify candidate susceptibility genes [23]. Default parameters were used, with the exception of the "--cutoff_condition_number" parameter, which was set to 30. Bonferroni-correction of P-values was performed on the resultant gene set (22,337 genes), using R's p.adjust function. This was followed by the removal of non-protein-coding genes. The analysis was also performed in the same manner using summary statistics from a fracture GWAS [6]. Finally, to identify protein-coding genes in the results, we utilized Ensembl's "hsapiens_gene_ensembl" dataset using biomaRt [51,52].

Creation of the "known bone gene" list:

We generated a "known bone gene" set as follows: First, we downloaded Gene Ontology IDs for the following terms: "osteo*", "bone",and "ossif*" from AmiGO2 (version 2.5.13) [53] . After removal of non-bone related terms, we extracted all mouse and human genes related to the GO terms, using biomaRt. From this list, we retained protein-coding genes.
We also used the "Human-Mouse: Disease Connection" database available at the Mouse Genome Informatics website, to download human and mouse genes annotated with the terms "osteoporosis", "bone mineral density", "osteoblast", "osteoclast" and "osteocyte". We used biomaRt to identify the gene biotypes, and retained protein-coding genes. We then used the MGI human-mouse homology table [http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt] to convert all mouse genes to their human homologs. Finally, we removed genes that weren't interrogated in both the colocalization and the TWAS analyses.

Gene Ontology enrichment analyses:

Gene ontology analysis was performed for the set of protein-coding genes passing the colocalization threshold RCP >= 0.1 and S-MultiXcan Bonferroni P-value <=0.05, using the "topGO" package (version 2.40.0) in R [54]. Enrichment tests were performed for the "Molecular Function", "Biological Process" and "Cellular Component" ontologies, using all protein-coding genes that were subjected to colocalization and multiXcan analysis as background. Enrichment was performed using the "classic" algorithm with Fisher's exact test. P-values were not adjusted for multiple testing.

Linkage disequilibrium calculations:

Linkage disequilibrium between variants was calculated using the LDlinkR (version 1.0.2) R package, using the "EUR" population [55].

***PPP6R3* knockout mouse generation:**

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The study was carried out in strict accordance with NIH's Guide for the Care and Use of Laboratory Animals. Additionally, the University of Virginia Institutional Animal Care and Use Committee approved all animal procedures. *Ppp6r3* gene trap mice were generated using targeted embryonic stem cell clones heterozygous for the *Ppp6r3*$^{tm1a(KOMP)Wtsi}$ gene trap allele obtained from the International Knockout Mouse Project (KOMP; [https://www.komp.org]). KOMP ES clones were karyotyped and injected using a XYClone Laser (Hamilton Thorne, Beverly, MA) into B6N-Tyr^{c-Brd}/BrdCrCrl (Charles River, Wilmington, MA) 8-cell stage embryos to create chimeric mice. Resultant chimeras were bred to B6N-Tyr^{c-Brd}/BrdCrCrl mice to obtain germline transmission of the *Ppp6r3* gene trap allele. From a breeding pair of two heterozygous mice, we generated our experimental population through HET x HET matings. Breeder mice were fed a breeder chow diet (Envigo Teklad S-2335 mouse breeder sterilizable diet, irradiated. Product # 7904), and experimental mice were fed a standard chow diet (Envigo Teklad LM-485 irradiated mouse/rat sterilizable diet. Product #7912).

Genotyping of *PPP6R3* mice:

DNA for genotyping was extracted from tail clips as follows: tail clips were incubated overnight at 55º C in a solution of 200uL digestion/lysis buffer (Viagen Direct PCR (tail), Los Angeles, CA) and 1mg/mL proteinase K (Viagen, Los Angeles, CA). After overnight incubation, tails were heated at 85º C for 45 minutes, and solutions were subsequently stored at 4º C.

For genotyping, PCR reactions were set up as follows. For each reaction, 1 µL of DNA was mixed with 24 µL of a master mix consisting of 19.5 µL nuclease-free H_2O, 2.5 µL 10x PCR reaction buffer (Invitrogen, Waltham, MA), 0.75 µL of mgCl2 (Invitrogen, Waltham, MA), 0.5 µL of 10mMol Quad dNTPs (Roche Diagnostics GmbH, Mannheim, Germany) 0.25 µL of Platinum Taq DNA polymerase (Invitrogen, Waltham, MA), and 0.25 µL of each primer, diluted to 20 µMol.

Primers: PCR primers were obtained from Integrated DNA Technologies, Coralville, IA.
Forward primer: 5'- CAC CTG GGT TGG TTA CAT CC -3'
Reverse primer: 5'- GAC CCT GCC TTA AAA CCA AA -3'

The following PCR settings were used:

- Initialization: 94º C, 120s
- Denaturation: 94º C, 30s (37 cycles)
- Annealing: 54º C, 30s (37 cycles)
- Elongation: 72º C, 35s (37 cycles)
- Final elongation: 72º C, 300s

PCR products were run on a 2% agarose gel for 150 minutes at 60 volts, to distinguish between wild-type, heterozygous and mutant *Ppp6r3* mice.

PPP6R3 Western blotting:

Mouse spleens 20-40 mg in weight were suspended in 1% NP40 buffer (50 mM Tris (pH 8) 100 mM NaCl,

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 % NP40, 1 mM EGTA, 1 mM EDTA, Protease inhibitor cocktail (04-693-116-001, Roche), 1 mM PMSF, 50 mM NaF, 0.2 mM sodium vanadate). The tissue was homogenized by RNase-free disposable pestles (ThermoFisher #12-141-364) and incubated for 10 min on ice. After brief sonication, the sample was centrifuged for 10 min at 13,000 x rpm at 4C. The protein concentration in the extract was measured by Bradford assay. 100ug of sample protein was boiled 5 min in SDS sample buffer, loaded in each lane, resolved by gradient SDS– PAGE (Bio-Rad #456-1085) and immunoblotted as described in Guergnon et al [56]. Primary antibodies were diluted 1:1000. (SAPS1 Ab: ThermoFisher #PA5-44275, SAPS3 Ab: ThermoFisher #PA5-58405, PP6C Ab: Sigma #HPA050940)

PPP6R3 functional validation:

Experimental mice were sacrificed at approximately 9 weeks of age (mean age = 61 days). At sacrifice, the right femurs were isolated, and femoral morphology (length and widths in AP and ML orientations) was measured with digital calipers (Mitoyuto American, Aurora, IL).
Femurs were then wrapped in PBS-soaked gauze and stored at -20º C, until analysis. Lumbar vertebrae L3-L5 were also dissected at sacrifice and were wrapped in PBS-soaked gauze and frozen at -20º C.

Dual X-ray absorptiometry:

Individual right femurs and the lumbar spine (L5 vertebrae) were isolated from surrounding soft tissues and frozen at -20º C in PBS. Dual X-ray absorptiometry (DXA) was performed on the femurs and lumbar vertebrae using the Lunar Piximus II (GE Healthcare) as described previously by Beamer et al [57] . In short, 10 isolated bones were placed in the detector field at a time and the samples were analyzed one by one, such that the region of interest (ROI) was set for one specimen at a time for data collection. The ROI for the femurs was on the entire isolated femur. For the spine, was on the entire isolated L5. Care was taken to ensure that the sample orientation was identical for all samples.

Micro-computed tomography and image analysis:

All µCT analyses were carried out at the µCT Imaging Core Facility at Boston University using a Scanco Medical µCT 40 instrument (Brütisellen, Switzerland). The power, current, and integration time used for all scans were 70 kVp, 113 µA, and 200 msec respectively. The L5 vertebrae were scanned at a resolution of 12 microns/voxel. Two volumes of interest (VOIs) were selected for analysis: 1) the entire portion of the L5 vertebra extending from 60 microns caudal to the cranial growth plate in the vertebral body to 60 microns cranial to the caudal growth plate; and 2) only the trabecular centrum contained in the first VOI. Semi-automated-edge detection (Scanco Medical) was used to define the boundary between the trabecular centrum and cortical shell to produce the second VOI. Gaussian filtering (sigma= 0.8, support=1) was used for partial background noise suppression. A scan of a potassium hydroxyapatite phantom allowed conversion of grayvalues to mineral density. For segmentation of bone tissue, the threshold was set at a 16-bit gray value of 7143 (521 mgHA/ccm), and this

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

global threshold was applied to all of the samples. For each VOI, the following were calculated: total volume (TV), bone volume (BV), bone volume fraction (BV/TV), bone mineral density (BMD), and tissue mineral density (TMD). BMD was defined as the average density of all voxels in the VOI, whereas TMD was defined as the average density of all voxels in the VOI above the threshold [58]. For the second VOI, the following additional parameters were calculated: trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), trabecular number (Tb.N), connectivity density (Conn.D), and structure model index (SMI) [58].

Raman spectroscopy:

Raman spectroscopy was performed using a Renishaw inVia Raman Microscope (Gloucestershire, UK) on each bone sample using a 785 nm edge red incident laser. A rectangular filled map was created with 3 points in the x-axis and 20 points in the y-axis, for a total of 60 collected points. Each point was exposed 10 times for 6 seconds per exposure. A custom MATLAB script was used to evaluate the peak position, maximum intensity, peak width, full width at half maximum (FWHM) and the area under each peak. Peak area ratios were calculated for mineral:matrix, carbonate:phosphate and crystallinity. Furthermore, the standard deviations of peak area ratios were calculated for each mouse, and were further used to evaluate the material heterogeneity in groups.

Statistical analyses:

To calculate the enrichment of bone genes in prioritized genes, we performed Fisher's exact test, using R's "fisher.test" function, with the alternative hypothesis set as "greater".
For the statistical analysis of the phenotyping results, we calculated least-squares means (lsmeans) using the "emmeans" R package (version 1.5.2.1) [59]. Input for the lsmeans function was a linear model including terms for genotype, weight and age in days. For sex-combined data, we also added a term for sex. For DXA phenotypes, we included a term for "CenterRectX" and "CenterRectY". For the Raman spectroscopy data, weight and age were not included as terms in the linear model.
We used Tukey's HSD test to test for significant differences in lsmeans, for each pair of genotype levels. Tukey's HSD also controls the family-wise error rate.

Analyses involving data from the International Mouse Phenotyping Consortium:

For the IMPC data, we obtained data using their "statistical-result" SOLR database, using the "solrium" R package (version 1.1.4) [60]. We obtained experimental results using the "Bone*Mineral*Density" parameter. We then pruned the resulting data to only include "Successful" analyses, and removed experiments that included the skull. To generate the *Gpatch1* boxplot, we obtained raw data using from IMPC's "statistical-raw-data" SOLR database for *Gpatch1*, and analyzed the data in the same manner as IMPC, using the "OpenStats" R package (version 1.0.2), using the method="MM" and MM_BodyWeightIncluded = TRUE arguments [61]. Finally, mouse genes were converted to their human syntenic counterparts using Ensembl's "hsapiens_gene_ensembl" and "mmusculus_gene_ensembl" datasets through biomaRt.

PhenomeXcan data analysis:

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We obtained all significant PhenomeXcan gene-trait associations from their paper [https://advances.sciencemag.org/content/6/37/eaba2083], and used data for the "3148_raw-Heel_bone_mineral_density_BMD" phenotype [15]. Furthermore, we constrained our search to only include genes that were annotated by the authors as "protein_coding".

LSBMD/FNBMD GWAS analysis:
We obtained sex-combined LSBMD and FNBMD GWAS summary statistics from GEFOS [http://www.gefos.org/?q=content/data-release-2012], and then used a custom script that utilized the biomaRt R package to convert variants to their GRCh38 coordinates.

Data availability:
eBMD and fracture GWAS summary statistics were obtained from GEFOS, as were the LSBMD and FNBMD GWAS summary statistics. GTEx eQTL data were obtained from the GTEx web portal. Data from the PhenomeXcan project were obtained from Pividori et al [15]. Statistical data from the IMPC were obtained using an R interface to their SOLR database. *Ppp6r3* experimental data are provided on our GitHub [https://github.com/basel-maher/BMD_TWAS_colocalization]. Mouse-Human homologs were obtained from MGI [http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt]. We also obtained data from the MGI Human-Mouse:Disease Connection database [http://www.informatics.jax.org/diseasePortal]. Gene Ontologies were obtained from AmiGO2 [http://amigo.geneontology.org/amigo].

Code availability:
Analysis code and the raw data for our *Ppp6r3* functional validation analyses are available on GitHub [https://github.com/basel-maher/BMD_TWAS_colocalization].

Acknowledgements:
Research reported in this publication was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under Award Number AR071657 to C.R.F., L.C.G and E.F.M., and by the National Center for Research Resources of the National Institutes of Health under Award Number S10RR021072 to E.F.M. B.M.A-B was supported in part by a National Institutes of Health, Biomedical Data Sciences Training Grant (5T32LM012416). The authors acknowledge Wenhao Xu (University of Virginia) and the Genetically Engineered Mouse Models (GEMM) core for their technical assistance in generating the *Ppp6r3* gene-trap mice. We thank the IMPC for accessibility to BMD data in knockout mice (www.mousephenotype.org). The data used for the analyses described in this manuscript were obtained from the IMPC SOLR database on 3/8/21. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 6/30/20.

References:

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Black, D.M., and Rosen, C.J. (2016). Clinical practice. Postmenopausal osteoporosis. N. Engl. J. Med. *374*, 254–262.

2. Burge, R., Dawson-Hughes, B., Solomon, D.H., Wong, J.B., King, A., and Tosteson, A. (2007). Incidence and economic burden of osteoporosis-related fractures in the United States, 2005-2025. J. Bone Miner. Res. *22*, 465–475.

3. Miller, P.D., Zapalowski, C., Kulak, C.A., and Bilezikian, J.P. (1999). Bone densitometry: the best way to detect osteoporosis and to monitor therapy. J. Clin. Endocrinol. Metab. *84*, 1867–1871.

4. Ralston, S.H., and Uitterlinden, A.G. (2010). Genetics of osteoporosis. Endocr. Rev. *31*, 629–662.

5. Peacock, M. (2002). Genetics of Osteoporosis. Endocr. Rev. *23*, 303–326.

6. Morris, J.A., Kemp, J.P., Youlten, S.E., Laurent, L., Logan, J.G., Chai, R.C., Vulpescu, N.A., Forgetta, V., Kleinman, A., Mohanty, S.T., et al. (2019). An atlas of genetic influences on osteoporosis in humans and mice. Nat. Genet. *51*, 258–266.

7. Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y.-H., Duncan, E.L., Ntzani, E.E., Oei, L., Albagha, O.M.E., Amin, N., Kemp, J.P., et al. (2012). Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat. Genet. *44*, 491–501.

8. Kemp, J.P., Morris, J.A., Medina-Gomez, C., Forgetta, V., Warrington, N.M., Youlten, S.E., Zheng, J., Gregson, C.L., Grundberg, E., Trajanoska, K., et al. (2017). Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat. Genet. *49*, 1468–1475.

9. Rocha-Braz, M.G.M., and Ferraz-de-Souza, B. (2016). Genetics of osteoporosis: searching for candidate genes for bone fragility. Arch. Endocrinol. Metab. *60*, 391–401.

10. Sabik, O.L., and Farber, C.R. (2017). Using GWAS to identify novel therapeutic targets for osteoporosis. Transl. Res. *181*, 15–26.

11. Giral, H., Landmesser, U., and Kratzer, A. (2018). Into the wild: GWAS exploration of non-coding RNAs. Front. Cardiovasc. Med. *5*, 181.

12. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.

13. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.

14. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. *9*, 1825.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

15. Pividori, M., Rajagopal, P.S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y., Consortium, G., Wen, X., and Im, H.K. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci. Adv. *6*, eaba2083.

16. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genetics *10*, e1004383.

17. Wen, X., Pique-Regi, R., and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet. *13*, e1006646.

18. Bhattacharya, A., García-Closas, M., Olshan, A.F., Perou, C.M., Troester, M.A., and Love, M.I. (2020). A framework for transcriptome-wide association studies in breast cancer in diverse study populations. Genome Biol. *21*, 42.

19. Thom, C.S., and Voight, B.F. (2020). Genetic colocalization atlas points to common regulatory sites and genes for hematopoietic traits and hematopoietic contributions to disease phenotypes. BMC Med. Genomics *13*, 89.

20. Nica, A.C., and Dermitzakis, E.T. (2013). Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120362.

21. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

22. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., GTEx Consortium, Nicolae, D.L., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.

23. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet. *15*, e1007889.

24. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

25. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

26. Fitzpatrick, L.A. (2002). Secondary causes of osteoporosis. Mayo Clin. Proc. *77*, 453–468.

27. Mirza, F., and Canalis, E. (2015). Management of endocrine disease: Secondary osteoporosis: pathophysiology and management. Eur. J. Endocrinol. *173*, R131-51.

28. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29. Swan, A.L., Schütt, C., Rozman, J., Del Mar Muñiz Moreno, M., Brandmaier, S., Simon, M., Leuchtenberger, S., Griffiths, M., Brommage, R., Keskivali-Bond, P., et al. (2020). Mouse mutant phenotyping at scale reveals novel genes controlling bone mineral density. PLoS Genet. *16*, e1009190.

30. Calabrese, G.M., Mesner, L.D., Stains, J.P., Tommasini, S.M., Horowitz, M.C., Rosen, C.J., and Farber, C.R. (2017). Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. Cell Syst *4*, 46-59.e4.

31. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

32. Stefansson, B., and Brautigan, D.L. (2006). Protein phosphatase 6 subunit with conserved Sit4-associated protein domain targets IkappaBepsilon. J. Biol. Chem. *281*, 22624–22634.

33. Cristiano, L. (2020). PPP6R3 (protein phosphatase 6 regulatory subunit 3). Atlas Genet. Cytogenet. Oncol. Haematol.

34. Ziembik, M.A., Bender, T.P., Larner, J.M., and Brautigan, D.L. (2017). Functions of protein phosphatase-6 in NF-κB signaling and in lymphocytes. Biochem. Soc. Trans. *45*, 693–701.

35. Abu-Amer, Y. (2013). NF-κB signaling and bone resorption. Osteoporos. Int. *24*, 2377–2386.

36. Mao, J., Wang, J., Liu, B., Pan, W., Farr, G.H., 3rd, Flynn, C., Yuan, H., Takada, S., Kimelman, D., Li, L., et al. (2001). Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. Mol. Cell *7*, 801–809.

37. Mizuguchi, T., Furuta, I., Watanabe, Y., Tsukamoto, K., Tomita, H., Tsujihata, M., Ohta, T., Kishino, T., Matsumoto, N., Minakami, H., et al. (2004). LRP5, low-density-lipoprotein-receptor-related protein 5, is a determinant for bone mineral density. J. Hum. Genet. *49*, 80–86.

38. Boyden, L.M., Mao, J., Belsky, J., Mitzner, L., Farhi, A., Mitnick, M.A., Wu, D., Insogna, K., and Lifton, R.P. (2002). High bone density due to a mutation in LDL-receptor-related protein 5. N. Engl. J. Med. *346*, 1513–1521.

39. Marques-Pinheiro, A., Levasseur, R., Cormier, C., Bonneau, J., Boileau, C., Varret, M., Abifadel, M., and Allanore, Y. (2010). Novel LRP5 gene mutation in a patient with osteoporosis-pseudoglioma syndrome. Joint Bone Spine *77*, 151–153.

40. Gong, Y., Slee, R.B., Fukai, N., Rawadi, G., Roman-Roman, S., Reginato, A.M., Wang, H., Cundy, T., Glorieux, F.H., Lev, D., et al. (2001). LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. Cell *107*, 513–523.

41. van Meurs, J.B.J., Trikalinos, T.A., Ralston, S.H., Balcells, S., Brandi, M.L., Brixen, K., Kiel, D.P., Langdahl, B.L., Lips, P., Ljunggren, O., et al. (2008). Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis. JAMA *299*, 1277–1290.

42. Brixen, K., Beckers, S., Peeters, A., Piters, E., Balemans, W., Nielsen, T.L., Wraae, K., Bathum, L., Brasen, C., Hagen, C., et al. (2007). Polymorphisms in the low-density lipoprotein

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

receptor-related protein 5 (LRP5) gene are associated with peak bone mass in non-sedentary men: results from the Odense androgen study. Calcif. Tissue Int. *81*, 421–429.

43. Giroux, S., Elfassihi, L., Cardinal, G., Laflamme, N., and Rousseau, F. (2007). LRP5 coding polymorphisms influence the variation of peak bone mass in a normal population of French-Canadian women. Bone *40*, 1299–1307.

44. Peters, J.E., Lyons, P.A., Lee, J.C., Richard, A.C., Fortune, M.D., Newcombe, P.J., Richardson, S., and Smith, K.G.C. (2016). Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease. PLoS Genet. *12*, e1005908.

45. Mizuno, A., and Okada, Y. (2019). Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects. Eur. J. Hum. Genet. *27*, 1745–1756.

46. Mullin, B.H., Tickner, J., Zhu, K., Kenny, J., Mullin, S., Brown, S.J., Dudbridge, F., Pavlos, N.J., Mocarski, E.S., Walsh, J.P., et al. (2020). Characterisation of genetic regulatory effects for osteoporosis risk variants in human osteoclasts. Genome Biol. *21*, 80.

47. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics *32*, 283–285.

48. Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H.K., and Wen, X. (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. Am. J. Hum. Genet. *108*, 25–35.

49. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. *51*, 592–599.

50. Wen, X. (2015). Effective QTL Discovery Incorporating Genomic Annotations (bioRxiv).

51. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

52. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

53. Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., AmiGO Hub, and Web Presence Working Group (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289.

54. Alexa, A., and Rahnenfuhrer, J. (2020). topGO: Enrichment Analysis for Gene Ontology. R package version 2.40.0.

55. Myers, T.A., Chanock, S.J., and Machiela, M.J. (2020). LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations. Front. Genet. *11*, 157.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.12.464046; this version posted January 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

56. Guergnon, J., Derewenda, U., Edelson, J.R., and Brautigan, D.L. (2009). Mapping of protein phosphatase-6 association with its SAPS domain regulatory subunit using a model of helical repeats. BMC Biochem. *10*, 24.

57. Beamer, W.G., Shultz, K.L., Coombs, H.F., 3rd, DeMambro, V.E., Reinholdt, L.G., Ackert-Bicknell, C.L., Canalis, E., Rosen, C.J., and Donahue, L.R. (2011). BMD regulation on mouse distal chromosome 1, candidate genes, and response to ovariectomy or dietary fat. J. Bone Miner. Res. *26*, 88–99.

58. Bouxsein, M.L., Boyd, S.K., Christiansen, B.A., Guldberg, R.E., Jepsen, K.J., and Müller, R. (2010). Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. J. Bone Miner. Res. *25*, 1468–1486.

59. Lenth, R. (2020). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.2.1.

60. Chamberlain, S. (2019). solrium: General Purpose R Interface to “Solr”. R package version 1.1.4.

61. Mashhadi, H.H. (2020). OpenStats: A Robust and Scalable Software Package for Reproducible Analysis of High-Throughput genotype-phenotype association. R package version 1.0.2.