

5-1-2019

## Mouse genome-wide association and systems genetics identifies Lhfp as a regulator of bone mass.

Larry D Mesner

Gina M Calabrese

Basel Al-Barghouthi

Daniel M. Gatti

John P Sundberg

*See next page for additional authors*

Follow this and additional works at: <https://mouseion.jax.org/stfb2019>

Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

---

---

**Authors**

Larry D Mesner, Gina M Calabrese, Basel Al-Barghouthi, Daniel M. Gatti, John P Sundberg, Gary A Churchill, Dana A Godfrey, Cheryl L Ackert-Bicknell, and Charles R Farber

---

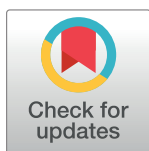
RESEARCH ARTICLE

# Mouse genome-wide association and systems genetics identifies *Lhfp* as a regulator of bone mass

Larry D. Mesner<sup>1,2</sup>, Gina M. Calabrese<sup>1</sup>, Basel Al-Barghouthi<sup>1,3</sup>, Daniel M. Gatti<sup>4</sup>, John P. Sundberg<sup>4</sup>, Gary A. Churchill<sup>4</sup>, Dana. A. Godfrey<sup>5</sup>, Cheryl L. Ackert-Bicknell<sup>5</sup>, Charles R. Farber<sup>1,2,3\*</sup>

**1** Center for Public Health Genomics, University of Virginia, Charlottesville, VA, United States of America, **2** Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States of America, **3** Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, United States of America, **4** The Jackson Laboratory, Bar Harbor, ME, United States of America, **5** Center for Musculoskeletal Research, University of Rochester, Rochester, NY, United States of America

\* [crf2s@virginia.edu](mailto:crf2s@virginia.edu)



**OPEN ACCESS**

**Citation:** Mesner LD, Calabrese GM, Al-Barghouthi B, Gatti DM, Sundberg JP, Churchill GA, et al. (2019) Mouse genome-wide association and systems genetics identifies *Lhfp* as a regulator of bone mass. *PLoS Genet* 15(5): e1008123. <https://doi.org/10.1371/journal.pgen.1008123>

**Editor:** Benedikt Hallgrímsson, University of Calgary, CANADA

**Received:** October 31, 2018

**Accepted:** April 3, 2019

**Published:** May 1, 2019

**Copyright:** © 2019 Mesner et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** BMD data used in this study are available on the "Ackert", "Naggert", and "Tordoff" strain sets from the Mouse Phenome Database (<https://phenome.jax.org>). All GWAS results are provided in supporting information files. Gene expression data sets used are available from the NCBI Gene Expression Omnibus (GEO) database (accession numbers GSE121887, GSE27483, GSE11338, GSE11065, GSE12798, and GSE12795).

## Abstract

Bone mineral density (BMD) is a strong predictor of osteoporotic fracture. It is also one of the most heritable disease-associated quantitative traits. As a result, there has been considerable effort focused on dissecting its genetic basis. Here, we performed a genome-wide association study (GWAS) in a panel of inbred strains to identify associations influencing BMD. This analysis identified a significant ( $P = 3.1 \times 10^{-12}$ ) BMD locus on Chromosome 3@52.5 Mbp that replicated in two separate inbred strain panels and overlapped a BMD quantitative trait locus (QTL) previously identified in a F2 intercross. The association mapped to a 300 Kbp region containing four genes; *Gm2447*, *Gm20750*, *Cog6*, and *Lhfp*. Further analysis found that Lipoma HMGIC Fusion Partner (*Lhfp*) was highly expressed in bone and osteoblasts. Furthermore, its expression was regulated by a local expression QTL (eQTL), which overlapped the BMD association. A co-expression network analysis revealed that *Lhfp* was strongly connected to genes involved in osteoblast differentiation. To directly evaluate its role in bone, *Lhfp* deficient mice (*Lhfp*<sup>-/-</sup>) were created using CRISPR/Cas9. Consistent with genetic and network predictions, bone marrow stromal cells (BMSCs) from *Lhfp*<sup>-/-</sup> mice displayed increased osteogenic differentiation. *Lhfp*<sup>-/-</sup> mice also had elevated BMD due to increased cortical bone mass. Lastly, we identified SNPs in human *LHFP* that were associated ( $P = 1.2 \times 10^{-5}$ ) with heel BMD. In conclusion, we used GWAS and systems genetics to identify *Lhfp* as a regulator of osteoblast activity and bone mass.

## Author summary

Osteoporosis is a common, chronic disease characterized by low bone mineral density (BMD) that puts millions of Americans at high risk of fracture. Variation in BMD in the general population is, in large part, determined by genetic factors. To identify novel genes influencing BMD, we performed a genome-wide association study in a panel of inbred

**Funding:** The authors would like to acknowledge grant support from the National Institute of Arthritis and Musculoskeletal and Skin Disease (R01AR057759, R01AR064790, and R01AR068345 to CRF and R21AR060981 to CLA) and the National Institute of General Medical Sciences (R01GM070683 for GC and DG) of the National Institutes of Health. Partial funding for generation of the *Lhfp* mutant mice was provided by the UVa Cancer Center (P30CA044579) and microCT analysis was partially fund through the Biomechanics, Biomaterials, and Multimodal Tissue Imaging Core in the UR Center for Musculoskeletal Research (P30AR069655). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

mouse strains. We identified a locus on Chromosome 3 strongly associated with BMD. Using a combination of systems genetics approaches, we connected the expression of the *Lhfp* gene with BMD-associated genetic variants and predicted it influenced BMD by altering the activity of bone-forming osteoblasts. Using mice deficient in *Lhfp*, we demonstrated that *Lhfp* negatively regulates bone formation and BMD. These data suggest that inhibiting *Lhfp* may represent a novel therapeutic strategy to increase BMD and decrease the risk of fracture.

## Introduction

It is currently estimated that half of all Americans over the age of 50 already have or are at high risk of developing osteoporosis [1]. Bone mineral density (BMD) is used clinically to diagnose osteoporosis and beyond age, it is the single strongest predictor of the risk of fracture [2]. BMD is also one of the most heritable disease-associated quantitative traits with studies demonstrating that up to 80% of the variance in peak bone mass is heritable [3–6]. Consistent with its high heritability, genome-wide association studies (GWASs) in humans have identified hundreds of loci for BMD [7–9]. However, only a small fraction of the variance in BMD can be collectively explained by these loci, suggesting that BMD is influenced by a large number of small effect size loci [10]. As a result, there remains much to be discovered regarding the genetics of bone mass and genetic mapping efforts using mouse models is a complementary approach to identify novel regulators of bone mass [11–13].

Historically, linkage analyses in intercrosses, backcrosses, and recombinant inbred strain panels were the mainstay of mouse genetics [14]. These approaches were used to identify dozens of quantitative trait loci (QTL) for BMD and other bone traits [15,16]. However, identifying causative genes underlying QTL proved challenging [17]. Over the last decade, gene mapping approaches have transitioned from low-resolution linkage mapping to high-resolution GWASs [11]. The first GWASs in mice used panels of inbred mouse strains [18–21] and by leveraging accumulated recombinations, this approach significantly increased mapping resolution [19]. However, the approach was limited by population structure and low statistical power, due to the complicated breeding histories of inbred mouse strains and the small number of easily accessible and appropriate inbred strains (N typically < 30), respectively. Later studies demonstrated that these issues could be partly addressed by accounting for population structure and leveraging information from linkage-based QTL studies [22,23]. Given the significant amount of existing phenotypic and genotypic data on inbred strain panels [24], this approach is potentially a cost-effective strategy to identify novel regulators of complex traits.

High-resolution mapping approaches have significantly increased our ability to identify narrow regions of the genome harboring trait associated genetic variants. It is still, however, a challenge to identify causal genes and several approaches have been developed that can assist in bridging this gap. Specifically, systems genetics approaches involving the integration of other types of “-omics” data have proven useful [25]. Two systems genetics approaches for informing GWAS are expression quantitative trait loci (eQTL) discovery and co-expression network analysis [26]. eQTL discovery allows one to link variants associated with a trait, such as BMD, to changes in gene expression which leads to the hypothesis that the change in gene expression is causal for the change in phenotype. eQTL studies have been tremendously successful in identifying target genes downstream of genome-wide significant variants (as examples; [27,28]). However, in many cases the identified target genes have no known connection to the phenotype under investigation. It has been shown that co-expressed genes often operate

in the same pathway or are functionally related [29]. Therefore, by using co-expression networks, which cluster genes based on patterns of co-expression across a series of perturbations [30], it is possible to develop hypotheses as to the function of a novel gene. When a locus has been resolved down to a small number of genes using genetic methods, unknown or poorly characterized genes can be ranked as the most likely candidate based on their function predicted from a co-expression network generated in a disease relevant tissue or cell-type [12,31].

Here, we used GWAS in an inbred strain panel to identify two chromosomal regions harboring variants influencing BMD. One of the associations, located on Chromosome (Chr.) 3, affected BMD in both sexes and was replicated in two separate inbred strain panels and an F2 intercross. This locus mapped to a 300 Kbp interval (NCBI37/mm9; Chr3:52.5–52.8 Mbp) encompassing four genes, *Gm2447*, *Gm20750*, *Cog6*, and *Lhfp*. An eQTL analysis and examination of a bone co-expression network suggested that *Lhfp* was a causal gene at this locus. The analysis of BMD, and other bone parameters, in *Lhfp* mutant mice supported this hypothesis. Additionally, SNPs within human *LHFP* were associated with heel BMD. Thus, we have used GWAS and systems genetics to identify *Lhfp* as a novel regulator of bone mass.

## Results

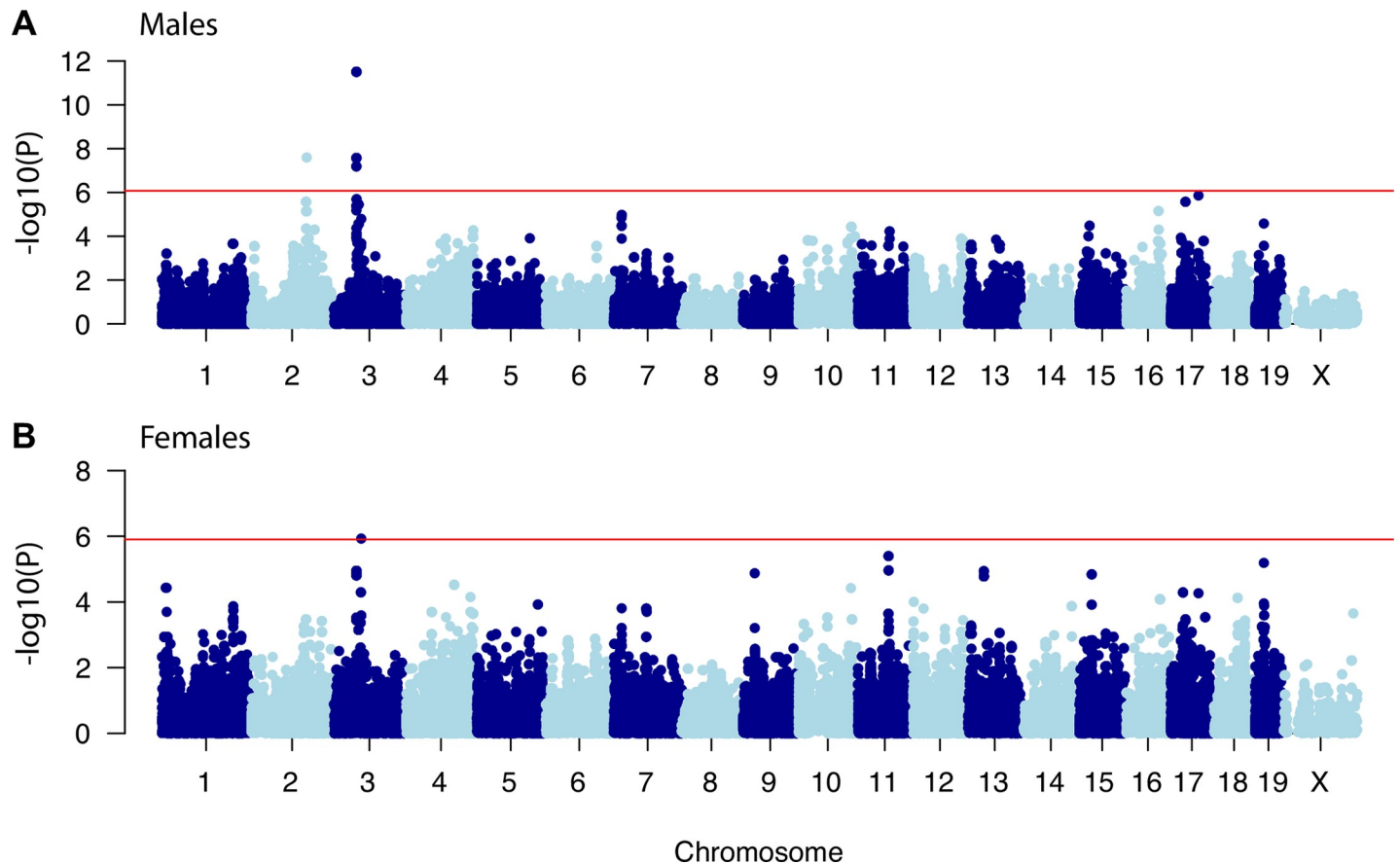
### Identification of genome-wide associations for BMD

We performed a GWAS for total body BMD in 26 classical (non wild-derived) inbred strains at 12 months of age fed a chow diet. Genome scans were performed separately for each sex using the Efficient Mixed Model Algorithm (EMMA) to account for population stratification (S1 and S2 Files). In female mice, a significant (permutation determined threshold of  $-\log_{10}(P) > 6$ ) association was identified on Chromosome (Chr.) 3 and, in males, significant (permutation determined threshold of  $-\log_{10}(P) > 5.9$ ) loci were identified on Chrs. 2 and 3 (Fig 1).

### Replication of the Chr. 3 association

Given our goal of identifying novel genes influencing BMD, we selected the Chr. 3 locus for further investigation. This locus was chosen because it was the most significant and the only one identified in both sexes (Fig 1). However, upon closer inspection, Chr. 3 harbored two associations, with peaks at 52.5 and 63.3 Mbp. In males, the 52.5 Mbp peak was the most significant ( $-\log_{10}(P) = 11.5$ ), whereas in females the 63.3 Mbp peak was the most significant ( $-\log_{10}(P) = 5.9$ ). The lead SNPs at both peaks were in moderate linkage disequilibrium ( $r^2 = 0.46$ ), making it unclear if they represented independent loci. We performed conditional analyses in males and in both cases each peak still exceeded chromosome-wise significance ( $-\log_{10}(P) > 2.9$ ) after controlling for the other, suggesting they represent independent loci.

We next sought to identify independent datasets supporting the validity of the Chr. 3 associations. There were 17 lead SNPs (the B6 reference allele was the minor allele at all SNPs with a frequency of 0.42), with the exact same strain distribution pattern (SDP) (S3 File), at the Chr.3@52.5 Mbp association. All 16 were polymorphic between B6 and C3H. We previously identified a QTL, *Bmd40*, affecting femoral BMD on Chr. 3 in 32 week-old mice from a C57BL/6J (B6) x C3H/HeJ (C3H) (BXH) F2 intercross fed a high-fat diet [16]. The peak of *Bmd40* overlaps both associations (Fig 2A). We also identified two sets of inbred strains with BMD measurements (the “Naggert” and “Tordoff” studies; data available from the Mouse Phenome Database [32] (<https://phenome.jax.org/>)) that were large enough (N strains > 25) to attempt to replicate the associations. In both “Naggert” and “Tordoff” panels, the strains used largely overlapped, but they did represent independent measures of BMD at different ages and conditions (Naggert—15–17 wks old, high-fat diet; Tordoff—14–18 weeks, chow diet). In both strain sets the exact same sets of SNPs at 52.5 Mbp reached chromosome-wide significance



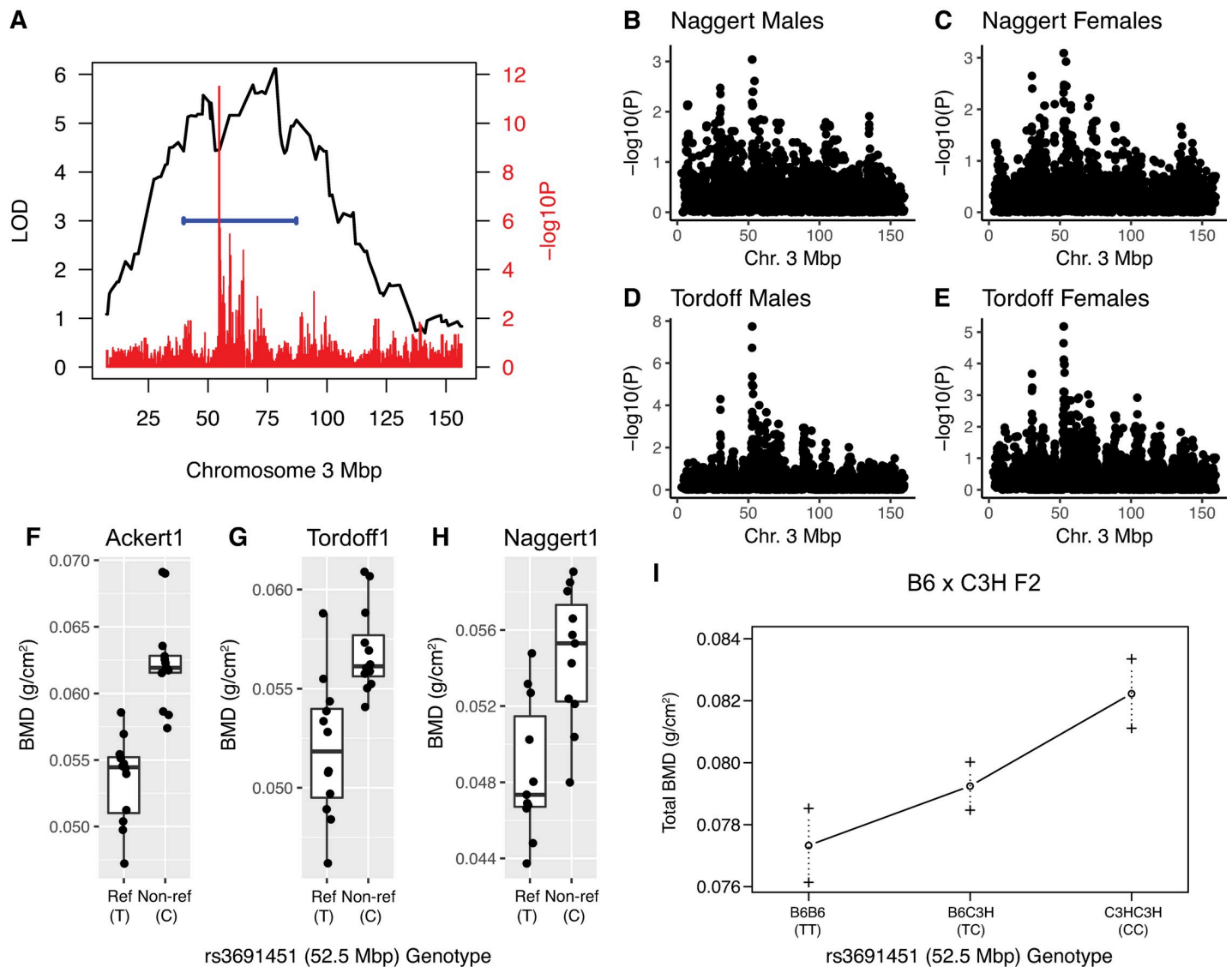
**Fig 1. Manhattan plot for BMD GWAS in the “Ackert” inbred strain panel.** A) GWAS results in male mice. B) GWAS results in female mice. Data from 27 classical inbred strains was used in the GWAS.

<https://doi.org/10.1371/journal.pgen.1008123.g001>

( $-\log_{10}(P) > 2.9$ ) in both sexes (Naggert males  $-\log P = 3.03$ , Naggert females  $-\log P = 3.09$ , Tordoff males  $-\log_{10}P = 7.73$ , and Tordoff females  $-\log_{10}P = 5.17$ ) (Fig 2B–2E). The association at 63.3 Mbp replicated in the Tordoff cohort (male  $-\log_{10}P = 3.18$  and female  $-\log_{10}P = 2.92$ ), but not in the Naggert cohort (male  $-\log_{10}P = 1.15$  and female  $-\log_{10}P = 1.09$ ) (Fig 2B–2E). These data provide additional support for the BMD association at 52.5 Mbp. Importantly, in all three inbred strain panels (“Ackert”, “Naggert” and “Tordoff”) and the BXH F2 intercross, reference (B6) alleles were associated with increased BMD relative to non-reference (C3H) alleles (Fig 2F–2I). Together, these data, from independent sources, are consistent with the hypothesis that a variant(s) in proximity of 52.5 Mbp on Chr. 3 influences BMD.

### The Chr. 3 association implicates a 300 Kbp interval encompassing four transcripts

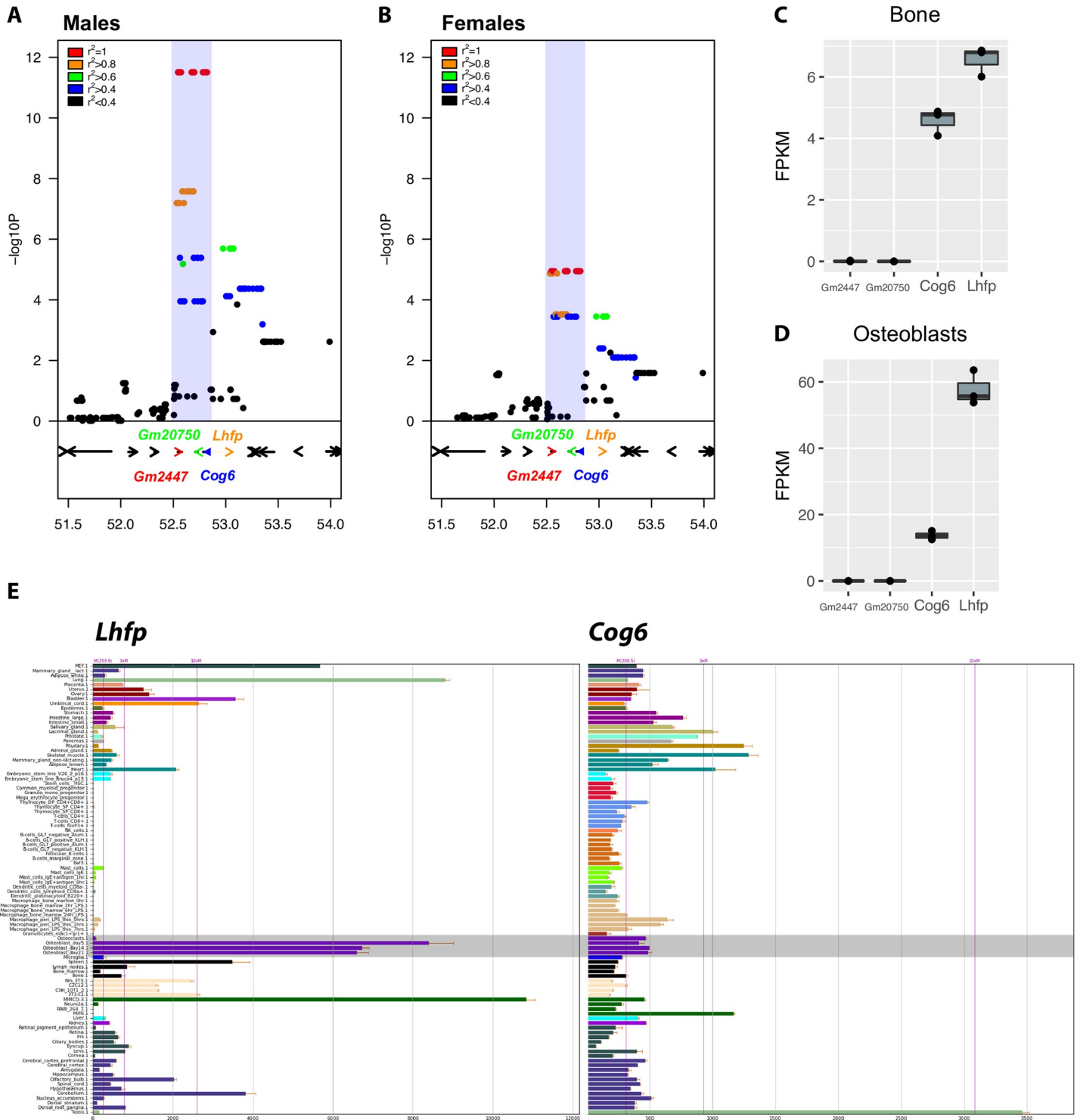
The set of SNPs that were the most significantly associated with BMD spanned a 300 Kbp interval from 52.5 to 52.8 Mbp (Fig 3A and 3B). This region contained four RefSeq transcripts: *Gm2447*, *Gm20750*, *Cog6*, and the 5' end of *Lhfp*. *Gm2447* and *Gm20750* were listed as “predicted” RefSeq transcripts and annotated as long non-coding RNAs (lncRNAs). The evidence for these transcripts was based on prediction models and a small number of expressed sequence tag (EST) sequences. Neither of these transcripts have homologs in humans, rats, or any other mammalian species. To determine if *Gm2447* and *Gm20750* were expressed in



**Fig 2. Replication of BMD association on Chr. 3@52.5 Mbp in both sexes in multiple independent populations.** A) *Bmd40*, a QTL impacting BMD in an F2 cross between C57BL/6J and C3H/HeJ, overlaps the Chr. 3 association at 52.5 Mbp. Replication of the Chr. 3 association at 52.5 Mbp in the “Naggert” (N = 31), males (B) and females (C), and “Tordoff” (N = 30), males (D) and females (E), inbred strain panels. Chromosome-wide significance was  $-\log_{10}P > 2.9$  in both strain sets. Non-ref (C3H/HeJ) alleles of associated SNPs increase BMD in the “Ackert” (F), “Tordoff1” (G), and “Naggert1” (H) strain panels and the BXH F2 intercross (I).

<https://doi.org/10.1371/journal.pgen.1008123.g002>

mouse bone or bone cells, we performed total RNA-seq (poly A+ and poly A-) on three bone and three marrow-derived osteoblasts samples. *Gm2447* and *Gm20750* were not expressed, whereas the other two transcripts, *Cog6* and *Lhfp*, which are well-annotated protein-coding sequences, were highly expressed in both bone (Fig 3C) and osteoblasts (Fig 3D). We also analyzed the expression profiles of *Cog6* and *Lhfp* in 96 mouse tissues and cell lines using data available from BioGPS (<http://biogps.org/>) [33]. *Cog6* was highly expressed in all tissues profiled (Fig 3E). *Lhfp* showed a more restrictive expression profile (Fig 3E). Importantly, *Lhfp* expression in primary calvarial osteoblasts was among the highest of any of the 96 samples surveyed (Fig 3E). *Cog6* is part of the conserved oligomeric Golgi complex required for maintaining normal structure and activity of the Golgi apparatus [34]. *Lhfp* is a member of the lipoma



**Fig 3. Interrogation of the Chr. 3@52.5 Mbp association.** The association implicates four genes (*Gm2447*, *Gm20750*, *Cog6* and *Lhfp*) based on their location with the locus in males (A) and females (B). RNA-seq expression profiles of the four genes in mouse bone (C) and osteoblasts (D) derived from bone marrow stromal cells (N = 3 each sample type). E) Microarray expression profiles for *Cog6* and *Lhfp* in 96 diverse mouse tissues and cell-types (data from BioGPS, <http://biogps.org/>) [33].

<https://doi.org/10.1371/journal.pgen.1008123.g003>



HMGIC fusion partner (*LHFP*) gene family with no known function [35]. All other transcripts on either side of the region were >200 Kbp away.

### Coding polymorphisms in *Cog6* and *Lhfp*

We cannot exclude *Gm2447* and *Gm20750* (or for that matter other genes flanking the association); however, based on the data above we focused on interrogating *Cog6* or *Lhfp* as potential causal genes. First, we evaluated *Cog6* and *Lhfp* for coding polymorphisms among inbred strains. Based on whole genome-sequence data from C57BL/6J and C3H/HeJ (which carry alternative alleles at the association) there are no coding variants between the strains for *Lhfp* [36]. In contrast, there were three non-synonymous SNPs in *Cog6* between B6 and C3H. These SNPs resulted in (rs30302002) I461V, (rs30323949) V620I and (rs30323946) S643N amino acid substitutions. However, using PolyPhen2, SIFT, and PROVEAN all three substitutions were predicted to be benign/tolerated and not impact *Cog6* function [37–39].

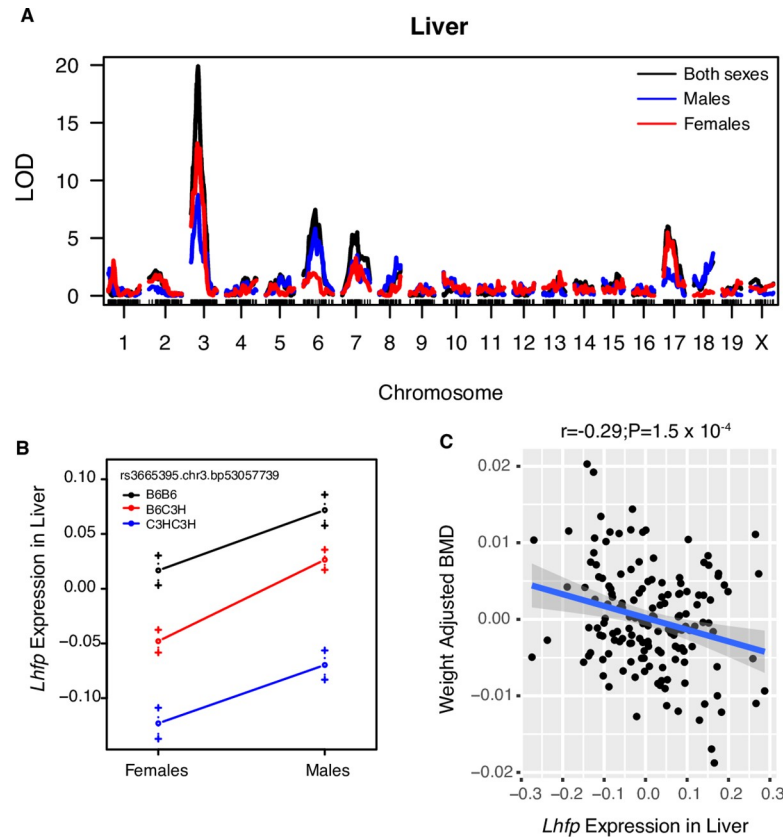
### *Lhfp* is regulated by a local eQTL in liver

We next determined if the same SNPs associated with BMD regulated the expression of *Lhfp* or *Cog6* (or any gene  $\pm$  1Mbp of the association). We searched for local expression quantitative trait loci (eQTL) using expression data in liver, brain, adipose and muscle tissues in the BXH F2 intercross. Although expression data on bone or bone cells would have been ideal for this analysis, these data were not available. This did, however, allow us to identify local eQTL that might also be operative in bone. We observed a highly significant local eQTL for *Lhfp* (LOD = 19.9) in liver (Fig 4A). *Cog6* and all other genes in proximity of the region were not regulated by a local eQTL in any tissue (max *cis* eQTL LOD = 1.8 across all four tissues). The lead *Lhfp* eQTL SNP (rs3665395) was located in the first intron of *Lhfp* and B6 alleles of rs3665395 were associated with increased expression of *Lhfp* relative to C3H alleles (Fig 4B). In liver, we observed a negative correlation ( $r = -0.29$ ,  $P = 1.5 \times 10^{-4}$ ) between *Lhfp* and BMD, as would be expected given that B6 alleles of rs3665395 were associated with decreased BMD and increased *Lhfp* (Fig 4C). We also searched for local eQTL using expression data from bone in the Hybrid Mouse Diversity Panel (HMDP) panel, but did not identify local eQTL for either *Cog6* or *Lhfp*.

### Network analysis predicts a role for *Lhfp* in the regulation of osteoblast activity

Our group and others [12,27,31,40] have shown that co-expression network analysis can identify interactions among genes and knowledge of these interactions can assist in predicting gene function and/or the cell type in which a gene is operative. Therefore, we next used a bone co-expression network to further evaluate *Lhfp* and *Cog6*. For the analysis we used a previously generated whole bone (femur with marrow removed) co-expression network from the Hybrid Mouse Diversity Panel (HMDP) that consisted of 13,759 genes partitioned into 21 co-expression modules [41,42]. In this network, *Lhfp* was a member of module 9 and *Cog6* was a member of module 2. Module 2 was enriched in a large number of gene ontology terms including “mitochondrion”, “oxidative phosphorylation” and “actin cytoskeleton”; all of which are important to bone. However, module 2 did not have a signature of a particular bone cell-type, nor was it enriched for genes known to influence BMD. In contrast, we have previously demonstrated that module 9 is enriched for genes 1) directly involved in osteoblast differentiation, 2) implicated by BMD GWAS, and 3) when knocked-out in mice impact BMD [31,42].

To investigate specific network connections for *Lhfp* and *Cog6*, we identified the 150 genes most strongly connected to each gene in their respective module (S4 and S5 Files). The genes



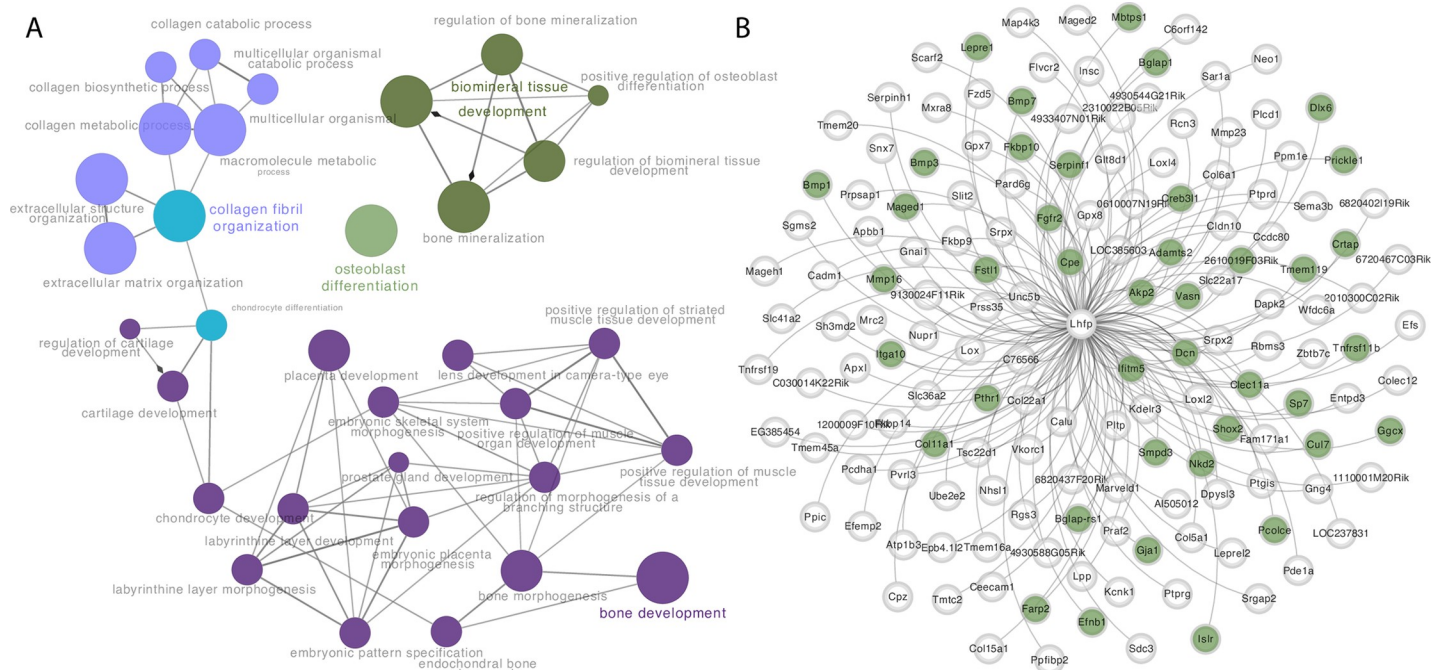
**Fig 4.** *Lhfp* expression is regulated by a local eQTL in liver overlapping the Chr.3@52.5 Mbp BMD association. A) Strong local eQTL for *Lhfp* in liver tissue from BXH F2 mice. B) In BXH livers, C57BL/6J (B6) alleles at the Chr. 3 local eQTL are associated with increased *Lhfp* expression. Data are presented as means  $\pm$  s.e.m. C) *Lhfp* expression in liver is negatively correlated with weight-adjusted femoral BMD in BXH F2 mice.

<https://doi.org/10.1371/journal.pgen.1008123.g004>

with the strongest connections to *Cog6* were enriched for genes involved in “muscle structure development” (FDR =  $2.9 \times 10^{-10}$ ), “muscle cell development” (FDR =  $4.4 \times 10^{-10}$ ), among many other similar muscle-related categories (S6 File). In contrast the genes with the strongest connections to *Lhfp* were, similar to module 9, enriched for genes involved in “ossification” (FDR =  $1.5 \times 10^{-7}$ ), “osteoblast differentiation” (FDR =  $8.0 \times 10^{-4}$ ), “skeletal system development” (FDR =  $6.4 \times 10^{-3}$ ), “bone development” (FDR =  $3.3 \times 10^{-2}$ ), among many other related bone-related functional categories (S7 File and Fig 5A). The *Lhfp*-centric network contained a number of genes with key roles in osteoblast differentiation and activity, including *Sp7*, *Pthrr1*, *Akp2*, *Tmem119*, and *Bmp3* (Fig 5B). Together, these data suggest that *Lhfp* is involved in the activity of osteoblasts, a process of direct relevance to the regulation of bone mass.

### *Lhfp* regulates the number and osteogenic differentiation of bone marrow stromal cells

Bone marrow stromal cells (BMSCs) are adherent marrow cells that contain the mesenchymal progenitors of osteoblasts [43]. To test the role of *Lhfp* in osteoblast function, we quantified the number of BMSCs and their ability to form osteoblasts from mice lacking *Lhfp*. Using CRISPR/Cas9, we created five small deletions (ranging from 4–16 bps) in exon 2 (ATG start codon is in exon 2) of *Lhfp* (Table 1). All five were frameshift mutations resulting in a truncated LHFP protein (S1 Fig). As expected, we observed significantly decreased *Lhfp* transcript



**Fig 5. *Lhfp* in bone is highly connected to genes involved in osteoblast differentiation.** A) Network depiction of gene ontology “biological process” categories containing more of the genes with strong connections to *Lhfp* in a bone co-expression network than would be expected by chance. B) Genes with the strongest connections to *Lhfp* in a bone co-expression network. The genes highlighted in green have been shown to be directly involved in osteoblast differentiation.

<https://doi.org/10.1371/journal.pgen.1008123.g005>

levels in heterozygotes (*Lhfp*<sup>+/−</sup>) and mutants (*Lhfp*<sup>−/−</sup>) from all five lines (Fig 6A). Since all five mutations impacted *Lhfp* expression in the same manner, we grouped littermate mice by genotype from all lines for all downstream experiments.

Next, we performed colony-forming unit-fibroblast (CFU-F) assays, a direct measure of BMSCs, in 16 week-old *Lhfp*<sup>−/−</sup> and littermate *Lhfp*<sup>+/+</sup> mice. We observed similar trends in both sexes; therefore, all data were combined and adjusted for the effects of sex to increase power. In *Lhfp*<sup>−/−</sup> mice, we observed a significant (P = 0.02) increase in CFU-F number (Fig 6B). We next evaluated the ability of BMSCs from *Lhfp*<sup>+/+</sup> and *Lhfp*<sup>−/−</sup> mice to differentiate into mineralizing osteoblasts. Consistent with network predictions, *Lhfp*<sup>−/−</sup> BMSCs exhibited increased mineralization as measured by bound alizarin red (P = 0.02; Fig 6C).

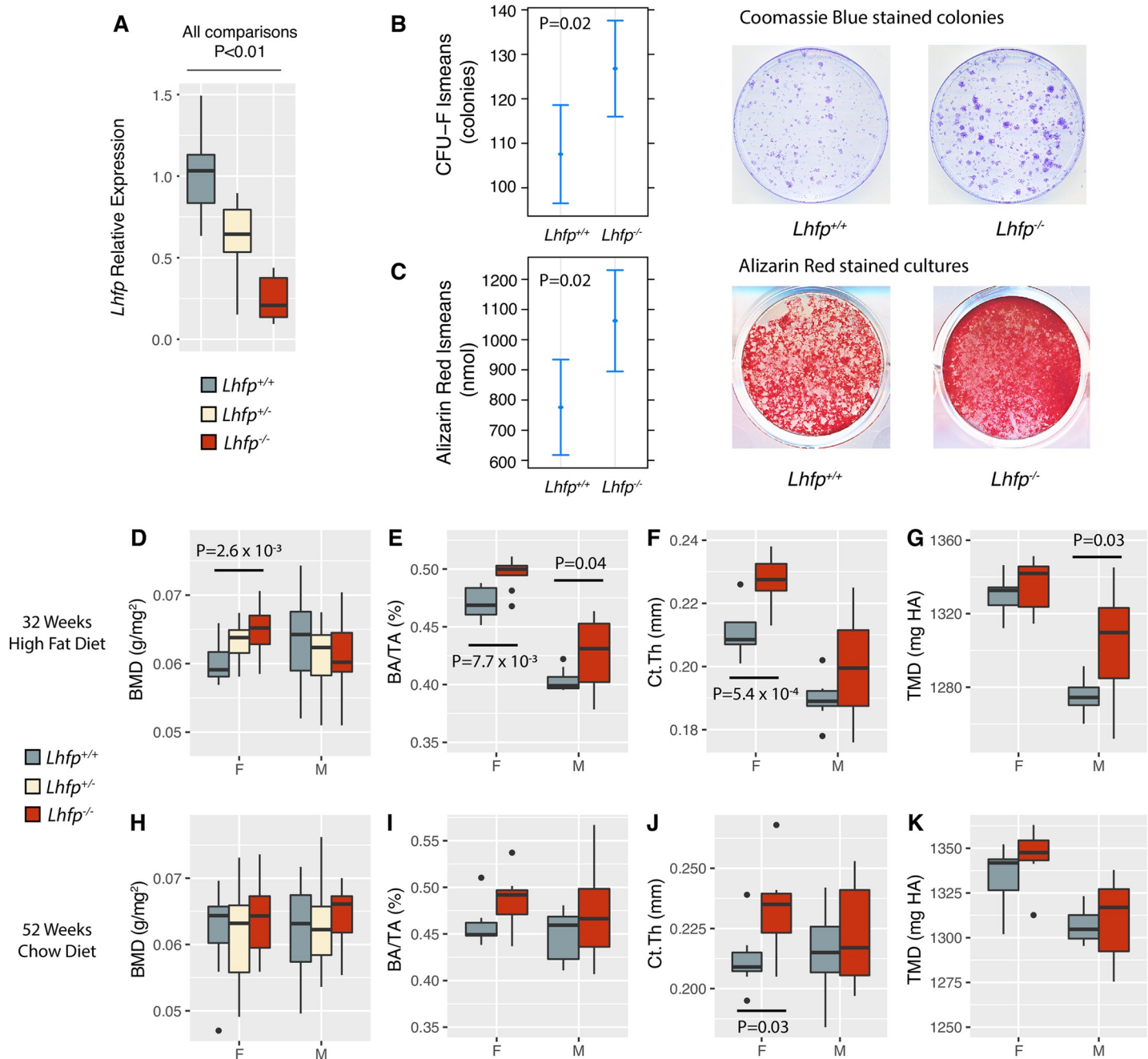
### *Lhfp* regulates cortical bone mass

We next determined if bone mass was altered in *Lhfp* mutant lines. To replicate the conditions of the Ackert inbred strain panel and the BXH F2, we generated two cohorts of mice. The first

**Table 1. Description of CRISPR/Cas9-induced *Lhfp* mutations.**

Mutant mouse line	Founder mouse	Deletion Size	Base pairs deleted	Chr 3 Map position deleted (GRCm38.p4 C57BL/6j)
1	C	4 bp	TGGG	53043620–530436623
2	B	4 & 3 bp or 8 bp + T	CCTG & TGG CCTGATGG + A inserted	53043615–530436618; 53043620–530436622 53043615–530436622
3	A	8 bp	TGG GTT GC	53043620–530436627
4	B & C	11 bp	CTG ATG GGT TG	53043616–530436626
5	A	16 bp	TCA CTG CCC TGA TGG G	53043608–530436623

<https://doi.org/10.1371/journal.pgen.1008123.t001>



**Fig 6. *Lhfp* is a negative regulator of osteoblast differentiation and cortical bone mass.** A) *Lhfp* transcript levels are decreased due to the five nonsense mutations created using CRISPR/Cas9 (*Lhfp*<sup>+/+</sup>, N = 10; *Lhfp*<sup>+/-</sup>, N = 11; *Lhfp*<sup>-/-</sup>, N = 7; represents at least one mouse from each of the five mutant lines, see Table 1). Data are presented as means ± 1.5 times interquartile range (IQR). The number of CFU-F colonies (B) and osteoblast mineralization (C) are increased in *Lhfp*<sup>-/-</sup> mice (*Lhfp*<sup>+/+</sup>, N = 41; *Lhfp*<sup>-/-</sup>, N = 43). Data are presented as means ± s.e.m. Images are from the sample closest to the mean for each genotype. D-G) Femoral BMD, cortical bone area fraction (BA/TA), cortical thickness (Ct.Th), and tissue mineral density (TMD) in *Lhfp* mutant mice in 32 week old mice fed a high-fat diet. Data are presented as means ± 1.5 times IQR. H-K) Femoral BMD, cortical bone area fraction (BA/TA), cortical thickness (Ct.Th), and tissue mineral density (TMD) in *Lhfp* mutant mice in 52 week old mice fed a chow diet. Data are presented as means ± 1.5 times IQR.

<https://doi.org/10.1371/journal.pgen.1008123.g006>

was fed a chow diet for 12 months, while the second was fed a high-fat diet from 8 to 32 weeks of age. Based on the negative correlation between BMD and *Lhfp* expression, the direction of the genetic effects on expression in liver, and increased osteoblast activity observed above, we

predicted increased BMD in *Lhfp*<sup>-/-</sup> mice. In both cohorts, BMD was measured in mice of all three genotypes and cortical and trabecular microarchitecture was measured by microCT only in *Lhfp*<sup>+/+</sup> and *Lhfp*<sup>-/-</sup> mice. At 32 weeks of age in mice on a high-fat diet we observed significantly ( $P = 2.6 \times 10^{-3}$ ) increased femoral BMD as a function of mutant *Lhfp* alleles in females, but not males (Fig 6D).

BMD is an inherently noisy phenotype; therefore, to generate a more detailed understanding of the effects of *Lhfp* in bone we used microCT to investigate the amount of bone in both the femoral trabecular and cortical compartments. We did not observe effects on trabecular bone mass at the distal femur in either male or female mice. However, *Lhfp*<sup>-/-</sup> mice of both sexes had significantly ( $P < 0.05$ ) increased femoral cortical bone area fraction (BA/TA) and cortical thickness (Ct.Th) as compared to *Lhfp*<sup>+/+</sup> littermates (Fig 6E and 6F). We also observed a significant ( $P = 0.03$ ) increase in tissue mineral density (TMD) in male *Lhfp*<sup>-/-</sup> mice (Fig 6G). In general, we observed the same trends of increased cortical bone mass in *Lhfp*<sup>-/-</sup> mice at 52 weeks of age; however, only Ct.Th in females was significant ( $P = 0.03$ ) (Fig 6H–6K). These data indicate that *Lhfp* is a negative regulator of cortical bone mass in both male and female mice. They are also consistent with *Lhfp* underlying, at least in part, the BMD association on Chr. 3@52.5 Mbp.

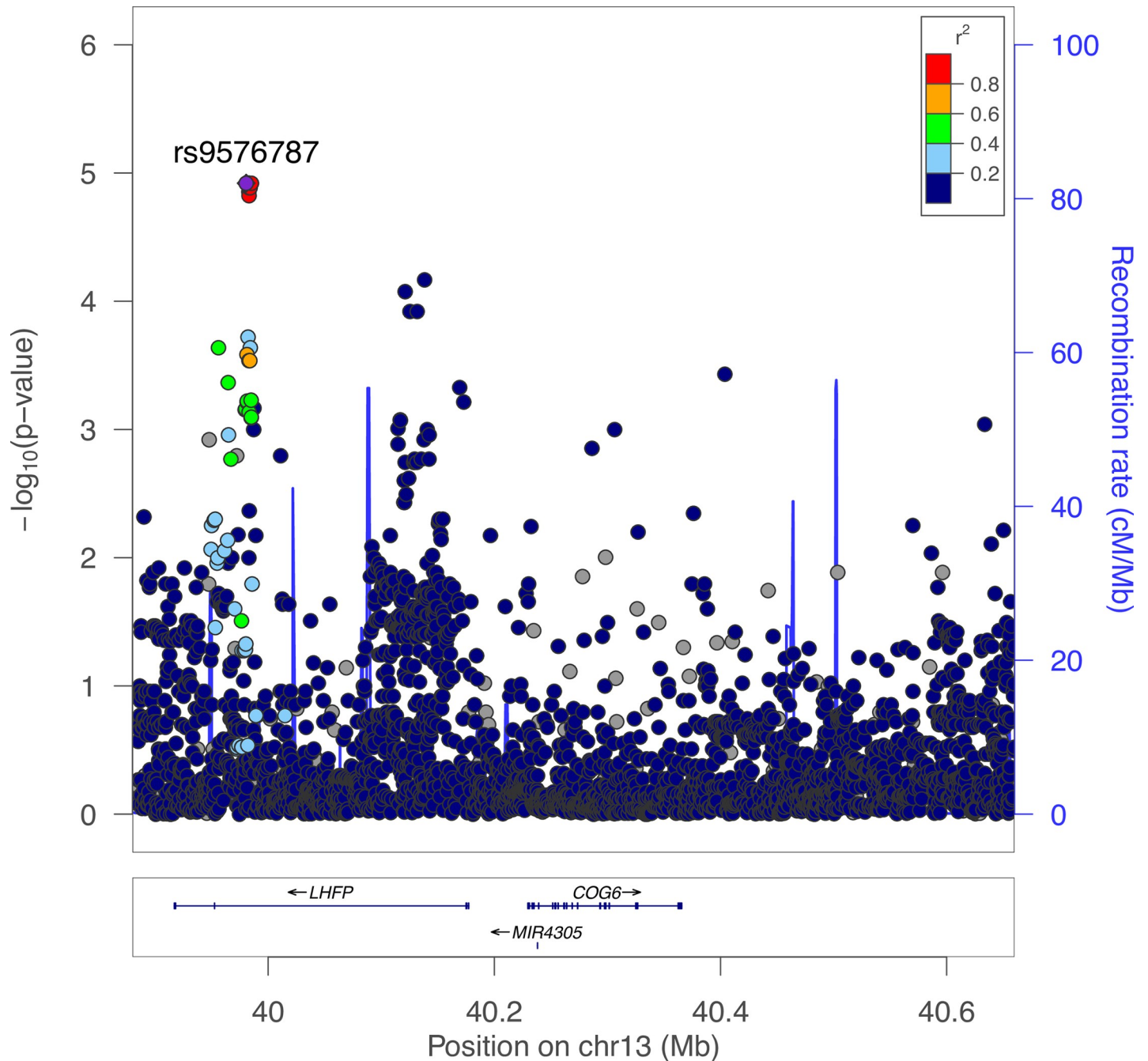
### Variants in human LHFPE are associated with BMD

We next determined if the human region syntenic with the mouse BMD locus harbored SNPs associated with variation in BMD. For this analysis we utilized data from the largest GWAS performed to date ( $N \sim 426K$ ) for heel BMD [9]. Heel BMD has been demonstrated to be highly genetically correlated with BMD at more clinically relevant sites such as the spine and femoral neck [8,9]. The human region syntenic with the mouse Chr. 3@52.5 Mbp spanned from 39.9 to 40.6 Mbp on Chr. 13. This region harbored 4055 SNPs. A set of 14 SNPs were significantly associated ( $P = 1.2 \times 10^{-5}$ ) after adjusting for the total number of SNPs in the region ( $P < 1.23 \times 10^{-5}$ ) (Fig 7). These SNPs were located in intron 3 of *LHFPE*. We queried eQTL for the Gene Tissue Expression (GTEx) project [44], but there were no eQTL for any genes  $\pm 1$  Mbp of the association that colocalized with the heel BMD association. Though these data do not directly implicate *LHFPE*, they do support its potential involvement in the regulation of human BMD.

### Discussion

In this study, we used GWAS in a mouse inbred strain panel and a multifaceted systems genetics approach to identify and validate a high-resolution association for BMD on Chr. 3. The association directly implicated four genes: *Gm2447*, *Gm20750*, *Cog6* and *Lhfp*. Of these, *Lhfp* expression was regulated by a local eQTL in liver and was predicted, based on a bone gene co-expression network, to be involved in osteoblast-mediated bone formation. We demonstrated that mice deficient in *Lhfp* displayed increased BMSC number and increased BMSC osteogenic differentiation. Furthermore, *Lhfp*<sup>-/-</sup> had increased BMD due to increased cortical bone mass. Together these data strongly suggest that *Lhfp* is responsible, at least in part, for the BMD association we identified on Chr. 3@52.5 Mbp. This work defines *Lhfp* as a negative regulator of the pool of osteoprogenitor cells, osteoblast activity, and cortical bone mass.

GWAS in mice has proven to be a powerful approach for the identification of genomic regions harboring trait-associated genetic variation [11]. The earliest applications of GWAS in mice used panels of readily accessible inbred strains [18–20]. However, such approaches were plagued by false positives due to population stratification [22]. Aware of this limitation, we first performed GWAS for BMD after correcting for population structure in inbred strains and then replicated the analysis in two separate strain panels (containing many of the same strains,



**Fig 7. The human region syntenic to the mouse BMD locus harbors SNPs associated with heel BMD.** LocusZoom plot of heel BMD-SNP associations in the human genomic region (Chromosome 13 from 39.8 to 40.7 Mbp) syntenic to mouse Chromosome 3 from 52.5 to 52.8 Mbp.

<https://doi.org/10.1371/journal.pgen.1008123.g007>

but representing independent measures of BMD in different environments) and an F2 intercross. Of the multiple loci identified, the association on Chr. 3 at 52.5 Mbp was identified in all datasets, strongly suggesting it represents a bona-fide genetic association.

The Chr. 3 locus, as defined by the interval harboring the most significant SNPs, contained four genes; *Gm2447*, *Gm20750*, *Cog6* and *Lhfp*. *Gm2447* and *Gm20750* were both predicted lncRNAs. This prediction is based on limited data and the fact that we did not observe their

expression in bone tissue or osteoblasts (though we only measured their expression in one inbred strain), suggest they are not likely causal for the locus; though, this alone is not enough to definitely exclude their involvement. For *Cog6* and *Lhfp* we used eQTL data and a bone co-expression network to assist in evaluating their potential causality. Both analyses supported a role for *Lhfp*. Using eQTL data from liver tissue in the BXH F2 intercross, we observed that variants associated with decreased BMD were associated with increased expression of *Lhfp*. We did not observe an association between the BMD-associated variants and *Cog6* expression. Furthermore, *Lhfp* was a member of a well-studied module of co-expressed genes in mouse bone. This module is highly enriched for genes that play a role in osteoblast function, which provides a direct explanation as to how *Lhfp* may be impacting BMD. In contrast, *Cog6* was a member of a module enriched for genes involved in a wide range of “energy-generating” functions. Importantly, all of our experimental results confirmed that *Lhfp* is a negative regulator of osteoblast activity and BMD. While these data support a role for *Lhfp* in the effects of the Chr. 3 locus, they do not exclude any of the other genes in, as well as flanking the locus.

In all four genetic populations used to identify the association on Chr. 3@52.5 Mbp, the strength of the association differed by sex. For example, in the “Ackert” population the association was stronger in males relative to females. In the “Naggert” strain set the strength of the association was similar in both sexes, albeit both were lower than seen in the other three populations. Similar to the “Ackert” strains, the association was stronger in males than females in the “Tordoff” strain set. In the BXH F2, the Chr. 3 QTL was male-specific, with little to no signal in females. The increase in cortical bone mass in *Lhfp*<sup>-/-</sup> mice was also sexually dimorphic. Although *Lhfp* deficiency increased cortical bone mass in both sexes in general, the effects were slightly more pronounced in females than males. This discrepancy could be the result of inaccuracies in estimating genetic effect sizes in the relatively small strain sets, the extent of linkage in the F2 confounding the sex effects, or influences from the different genetic backgrounds of the populations studied (strains sets vs. F2 vs. knockout).

Little is known regarding the molecular function of *Lhfp*. *Lhfp* is a member of the *Lhfp*-like gene family, which is a subset of the superfamily of tetraspan transmembrane protein encoding genes. It was first identified as a translocation partner with the HMGIC gene in benign lipomas [35]. The human *LHFP/COG6* locus was also identified by GWAS as harboring variants associated with hippocampal volume [45]. However, prior to this study *Lhfp* had not been connected to the regulation of osteoblast function or BMD. Based on our experimental results, we hypothesize that *Lhfp* regulates bone mass through a role in cells of the osteoblast lineage. This does fit with prior work implicating *Lhfp* in the mesenchymal differentiation of gliosarcoma [46]. It is possible that *Lhfp* serves as a “brake” regulating the number of osteogenic precursor cells in the bone marrow microenvironment as well as their differentiation potential. However, further work will be required to elucidate its precise molecular role in osteoblasts and bone.

In summary, we have used GWAS in a set of inbred strains to identify an association impacting femoral BMD on Chr. 3 at 52.5 Mbp. We show using a variety of approaches that *Lhfp* is likely responsible for most, if not all, of the effects of this locus. Our results identify *Lhfp* as a novel negative regulator of osteoblast function and BMD and increase our understanding of the genetics of BMD.

## Materials and methods

### Ethics statement

The animal protocol for the generation and characterization of *Lhfp* mutant mice was approved by the Institutional Care and Use Committee (IACUC) at the University of Virginia.

## Association analysis

The “Ackert” strain set contained BMD data on 32 inbred strains at three time points (6, 12 and 18 months). These data were collected by The Jackson Laboratory Nathan Shock Center of Excellence in the Basic Biology of Aging. Cohorts of males and female mice of 32 inbred strains of mice were aged to 6, 12 and 18 months of age. The number of mice per sex and per strain for each age point ranged from 1 mouse to 9 mice per group, with the majority of groups containing 6–7 mice. For the 12 month data set, the focus of this paper, the group size ranged from 3 to 9 mice per strain, per sex. At 6, 12 and 18 months of age, a variety of phenotypes were measured using a cross sectional study design with the hopes of capturing the main definers of Healthspan. This study, and the phenotypes available, is described in detail elsewhere [47]. Whole body BMD, sans the head, was measured by Dual X-ray Absorptiometry as previously described [48]. The complete dataset is available from the Mouse Phenome Database (MPD) (<https://phenome.jax.org/projects/Ackert1>). After removing wild-derived strains, and C57BLKS/J (due to inclusion of this strain producing spurious results) we were left with data on 26 strains. To identify loci influencing BMD, we used the Efficient Mixed Model Association (EMMA) algorithm [23]. For the analysis BMD was rankZ transformed. SNPs were obtained from strains genotyped on the Mouse Diversity Array (<http://churchill-lab.jax.org/website/MDA>) [49]. SNPs with a minor allele frequency < 0.05 were removed, leaving 228,085 SNPs. These SNPs were used to generate a kinship using the ‘emma.kinship’ R script available in the EMMA R package (available at <http://mouse.cs.ucla.edu/emma/>) [23]. The emma.REML.t function of EMMA was used to perform all mapping analyses. The significance of the maximum association peak was assessed by performing 1,000 permutations of the data. In each permutation, the minimum p-value was recorded to produce an empirical distribution of minimum permutation p-values. The quantiles of this distribution were used to assign adjusted p-values. P-values exceeding a genome-wide significant of  $P < 0.05$  were used as thresholds to identify associated loci. GWAS results were visualized using the “qqman” R package [50]. Replication of the association on Chr. 3@52.5 Mbp was performed using “Tordoff” and “Naggert” inbred strains sets. These data are available from MPD (<https://phenome.jax.org/projects/Naggert1> and <https://phenome.jax.org/projects/Tordoff3>). Replication analyses were restricted to Chr. 3 and otherwise performed as described above.

## Generating expression profiles for transcripts in the BMD locus

Femora were isolated from an inbred Collaborative Cross strain (CC016/GeniUnc; Jackson Lab Stock #024684) (N = 3 mice). Marrow was isolated and bone marrow stromal cells (BMSCs) were differentiated as described below. Total RNA was then isolated from bone and BMSC-derived osteoblasts using RNeasy Plus Mini Kit (Qiagen). RNA-Seq libraries were constructed using TruSeq RNA Library Prep Kit v2 sample prep kits (Illumina). Samples were sequenced to an average depth of 24.6 million 2 x 75 bp paired-end reads on an Illumina NextSeq500 sequencer. Fastq files were aligned to the mouse reference (GRCm38) using HISAT2 v 2.0.5 (<https://ccb.jhu.edu/software/hisat2/index.shtml>) with a SNP aware reference index (genome\_snp) [51]. Expression levels in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) were generated using Stringtie [51]. The data are available from GEO (GSE121887). Microarray profiles for *Cog6* and *Lhfp* in 96 tissues/cell-types were downloaded from BioGPS (<http://biogps.org>).

## eQTL analyses

The generation of microarray expression data and eQTL analyses on bone from the 96 strains of the Hybrid Mouse Diversity Panel (HMDP) has been previously described [41,42]. These



data are available from NCBI Gene Expression Omnibus (GEO) (GSE27483). A t-test was used to test for differences in *Cog6* and *Lhfp* expression in strains stratified by genotype at rs3691451 (of the 17 peak BMD SNPs). Liver, brain, muscle and adipose eQTL in the BXH F2 were identified using R/qtl [52]. The expression data is available from GEO (GSE11338, GSE11065, GSE12798, and GSE12795). The genotypes and expression data are also available from GeneNetwork (“BH/HB F2 UCLA”, <http://www.genenetwork.org/webqtl/main.py>).

## Network analysis

The generation of a bone co-expression network and characterization of the module 9 (M9) is described in [42]. We identified genes with the strongest connections to *Cog6* and *Lhfp* based on Topological Overlap Measures (TOMs), calculated as described in [30]. Network depictions were constructed using Cytoscape [53]. Gene Ontology (GO) analysis was performed using the PANTHER database statistical overrepresentation test (<http://www.pantherdb.org/>) [54]. The analysis was restricted to the “GO biological process complete” annotation data set.

## Generation of *Lhfp* mutant mice

The *Lhfp* knockout mice used in this study were generated using the CRISPR/Cas9 genome editing technique. Cas9 mRNA that was injected into C57BL6/N embryos was synthesized exactly as outlined in [55] while the guide RNA (sgRNA) was generated with some modifications. Briefly, the 20 nucleotide (nt) sequence that would be used to generate the sgRNA was chosen using the CRISPR design tool developed by the Zhang lab ([crispr.mit.edu](http://crispr.mit.edu)). The chosen sequence and its genome map position is homologous to a region in Exon 2 that is approximately 300 bp, 3' of the start codon (the 'ATG' is located in Exon 2 of *Lhfp*) (S1 Table). To generate the sgRNA that would be used for injections, oligonucleotides of the chosen sequence, as well as the reverse complement (S1 Table, primer 1 and 2, respectively), were synthesized such that an additional 4 nts (CTTC and AAAC) were added at the 5' ends of the oligonucleotides for cloning purposes. These oligonucleotides were annealed to each other by combining equal molar amounts, heating to 90°C for 5 min. and allowing the mixture to passively cool to room temperature. The annealed oligonucleotides were combined with BbsI digested pX330 plasmid vector (provided by the Zhang lab through Addgene; <https://www.addgene.org/>) and T4 DNA ligase (NEB) and subsequently used to transform Stbl3 competent bacteria (Thermo Fisher) following the manufacturer's protocols. Plasmid DNAs from selected clones were sequenced from primer 4 (S1 Table) and DNA that demonstrated accurate sequence and position of the guide were used for all downstream applications. The DNA template used in the synthesis of the sgRNA was the product of a PCR using the verified plasmid DNA and primers 3 and 5 (S1 Table). The sgRNA was synthesized via in vitro transcription (IVT) by way of the MAXIscript T7 kit (Thermo Fisher) following the manufacturer's protocol. sgRNAs were purified and concentrated using the RNeasy Plus Micro kit (Qiagen) following the manufacturer's protocol.

C57BL/6N female mice (Envigo) were super-ovulated and mated with C57BL/6N males. The females were sacrificed and the fertilized eggs were isolated from the oviducts. The fertilized eggs were co-injected with the purified Cas9 mRNA (100 ng/μl) and sgRNA (30 ng/μl) under a Leica inverted microscope equipped with Leitz micromanipulators (Leica Microsystems). Injected eggs were incubated overnight in KSOM-AA medium (Millipore Sigma). Two-cell stage embryos were implanted on the following day into the oviducts of pseudo pregnant ICR female mice (Taconic or Envigo). Pups were screened by PCR of tail DNA using primers 6 and 7 with subsequent sequencing of the resultant product from primer 8 (S1 Table).

Two sets of injections (of ~100 eggs each) were performed resulting in 2 mice possessing mutations from each set of injections (A,B and C,D, respectively; Table 1). All 4 mice

possessed out of frame bi-allelic deletions ranging from 1–16 bp; progeny from only 3 of the founders (mice A, B, C, [Table 1](#)) were used in this study. Note that an identical 11bp deletion was found in two mice from two separate injections. qPCR with primers 9 and 10 ([S1 Table](#)) was used to assess *Lhfp* expression as outlined in [\[31\]](#).

### CFU-F and osteogenic differentiation assays

Isolation [\[56\]](#) and differentiation of mesenchymal stromal cells (MSC) from the bone marrow of mouse femurs was performed as described for osteoblasts [\[57\]](#) with minor modifications. Briefly, one or both femurs from a given mouse were aseptically isolated, denuded of soft tissue and the marrow extracted by removing the proximal end of each bone and centrifuging at 2000 xg for 30 s such that the marrow collects into 25  $\mu$ l of fetal bovine serum (FBS). Exudates from a single femur were dispersed, via trituration, in 5ml complete media (MEM-alpha, 10% FBS, 100U penicillin/100ug Streptomycin per ml, 2 mM glutamine). Cells were manually counted where upon 4 million were used to seed a 10 cm dish for CFU-F determination and the remainder were applied to a 60mm dish. Media on the 10 cm dishes was changed on days 2, 4 and 8; on day 14, cells were fixed (NBF), stained (Coomassie, BioRad #161–0436) and the number of CFU-Fs determined via Image J analysis and the manual counting of colonies. Media on the 60 mm dishes was changed on day 2 with cells removed via trypsin/EDTA (Gibco) digestion on day 4. Detached cells were trituated in 5ml complete media, pelleted at 1000 xg for 5 min., re-suspended in 1 ml complete media and counted where upon 150,000 cells were used to seed a well of a 12 well dish. A minimum of 2 wells per sample were obtained for all samples reported here in. Osteoblast differentiation was initiated 3 days after plating (7 days after bone marrow isolation) by replacing the media with complete media supplemented with 50  $\mu$ g/ml ascorbic acid, 10 mM beta-glycerophosphate and 10 nM dexamethasone. Media was changed every other day for 8 days at which time cells were either used as a source for RNA (mirVana, Thermo Fisher) or used to determine the amount of hydroxyapatite formed during differentiation [\[31\]](#). Briefly, cells were washed with PBS, fixed with neutral buffer formalin (NBF) for 15 min. and subsequently stained with 40 mM Alizarin Red (AR), pH 5.6 for 20 min and washed extensively with H<sub>2</sub>O. The amount of AR bound to mineral was quantitated by Image J analysis of scanned images as well as the 5% Perchloric Acid eluate absorbance at 405 nm.

### Measurement of BMD and microarchitecture (microCT)

Femoral BMD was measured ex vivo using a Lunar PIXImus II Mouse Densitometer (GE Medical Systems Model 51045; Madison, WI, USA). Morphologies of the trabecular bone of the distal femur and cortical bone of the femoral midshaft were measured using micro-focus X-ray computed tomography (vivaCT 40, Scanco Medical AG, Bassersdorf, Switzerland) following guidelines for assessment of bone microstructure [\[58\]](#). Tomographic volumes were acquired at 55 kV and 145  $\mu$ A, collecting 2000 projections per rotation at 300 millisecond integration time. Three-dimensional 16-bit grayscale images were reconstructed using Scanco Evaluation software, Version 6.5–3. Threshold values were adjusted to best match the silhouette of features of interest in the threshold-subtracted image compared to the grey-scale image. The resulting threshold for hydroxyapatite-equivalent density was 370 mg/cm<sup>3</sup> for compact or cortical bone and 270 mg/cm<sup>3</sup> for the trabecular bone region; these values were applied to subsequent samples. Volumetric analysis was confined to the trabecular region for the distal femur by manual exclusion of the cortical bone. A 1.03 mm high region of interest was analyzed beginning at 1 mm proximal to the growth plate. For the cortical bone, a 0.3 mm high region was analyzed at the mid-diaphysis. Measures analyzing the distal femur trabecular site

included total volume, bone volume, trabecular bone volume fraction (BV/TV), thickness, number, connectivity density and structure model index (SMI). Cross-sectional measurements of the cortical bone included bone volume, total volume, marrow area and polar moment of inertia.

### Statistical analysis of *Lhfp* mutant data

All statistical analyses were conducted using the R language and environment for statistical computing [59]. The *Lhfp* qPCR data was analyzed using a t-test. Data are presented as means  $\pm$  1.5 times the interquartile range. CFU-F and osteogenic differentiation data was analyzed using the “lsmeans” R package [60]. The data were fit to a linear model including the effects of genotype and sex. P-values were adjusted using the “tukey” method. Data are presented as lsmeans  $\pm$  s.e.m. BMD and microarchitectural bone data were analyzed using ANOVA with a linear model including the effects of genotype, body weight, and any other phenotype-specific covariates. Data are presented as means  $\pm$  1.5 times the interquartile range.

### Human BMD GWAS data

Data from the largest heel BMD GWAS performed to date were downloaded from <http://www.gefos.org/?q=content/data-release-2018> [9]. LocusZoom was used to create a regional association plot [61]. The GTEx database V7 was queried for colocalizing eQTL(<https://gtexportal.org/home/>) [44].

### Supporting information

**S1 Table. List of primer sequences.**

(DOCX)

**S1 Fig. Impact of each CRISPR/Cas9 induced mutation on *Lhfp* protein sequence.**

(DOCX)

**S1 File. Summary statistics for BMD GWAS in males.** Contains three columns, 1) Chromosomal location of SNP (CHR), 2) bp position (BP), and 3) P-value.

(XLSX)

**S2 File. Summary statistics for BMD GWAS in females.** Contains three columns, 1) Chromosomal location of SNP (CHR), 2) bp position (BP), and 3) P-value.

(XLSX)

**S3 File. Strain distribution patterns of rs3691451 (one of the lead BMD SNPs) in the Ackert, Tordoff, and Naggert cohorts.**

(XLSX)

**S4 File. Top 150 genes with the strongest network to *Cog6*.**

(XLSX)

**S5 File. Top 150 genes with the strongest network to *Lhfp*.**

(XLSX)

**S6 File. GO enrichments for genes strongly connected to *Cog6*.**

(TXT)

**S7 File. GO enrichments for genes strongly connected to *Lhfp*.**

(TXT)

## Acknowledgments

We thank Dr. Wenhao Xu in the Genetically Engineered Murine Model (GEMM) Core at the University of Virginia (UVa) for assistance in generating the *Lhfp* CRISPR/Cas9 mutants.

## Author Contributions

**Conceptualization:** Larry D. Mesner, Cheryl L. Ackert-Bicknell, Charles R. Farber.

**Data curation:** John P. Sundberg, Gary A. Churchill, Dana. A. Godfrey, Cheryl L. Ackert-Bicknell, Charles R. Farber.

**Formal analysis:** Basel Al-Barghouthi, Daniel M. Gatti, Cheryl L. Ackert-Bicknell, Charles R. Farber.

**Funding acquisition:** Charles R. Farber.

**Investigation:** Larry D. Mesner, Charles R. Farber.

**Methodology:** Larry D. Mesner, Gina M. Calabrese, Cheryl L. Ackert-Bicknell, Charles R. Farber.

**Project administration:** Cheryl L. Ackert-Bicknell, Charles R. Farber.

**Validation:** Larry D. Mesner, Charles R. Farber.

**Writing – original draft:** Charles R. Farber.

**Writing – review & editing:** Larry D. Mesner, Cheryl L. Ackert-Bicknell, Charles R. Farber.

## References

1. Black DM, Rosen CJ. Clinical Practice. Postmenopausal Osteoporosis. *N Engl J Med*. 2016; 374: 254–262. <https://doi.org/10.1056/NEJMcp1513724> PMID: 26789873
2. Kanis JA, Borgstrom F, De Laet C, Johansson H, Johnell O, Jonsson B, et al. Assessment of fracture risk. *Osteoporos Int*. Springer-Verlag; 2005; 16: 581–589. <https://doi.org/10.1007/s00198-004-1780-5> PMID: 15616758
3. Ralston SH, Uitterlinden AG. Genetics of osteoporosis. *Endocr Rev*. 2010; 31: 629–662. <https://doi.org/10.1210/er.2009-0044> PMID: 20431112
4. Farber CR, Rosen C. Genetics of Osteoporosis. *Translational Endocrinology & Metabolism*. 2010; 87.
5. Ralston SH, de Crombrughe B. Genetic regulation of bone mass and susceptibility to osteoporosis. *Genes Dev*. Cold Spring Harbor Lab; 2006; 20: 2492–2506. <https://doi.org/10.1101/gad.1449506> PMID: 16980579
6. Peacock M, Turner CH, Econs MJ, Foroud T. Genetics of osteoporosis. *Endocr Rev*. 2002; 23: 303–326. <https://doi.org/10.1210/edrv.23.3.0464> PMID: 12050122
7. Estrada K, Styrkarsdottir U, Evangelou E, Hsu Y-H, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet*. 2012; 44: 491–501. <https://doi.org/10.1038/ng.2249> PMID: 22504420
8. Kemp JP, Morris JA, Medina-Gómez C, Forgetta V, Warrington NM, Youtlen SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet*. 2017; 49: 1468–1475. <https://doi.org/10.1038/ng.3949> PMID: 28869591
9. Morris JA, Kemp JP, Youtlen SE, Laurent L, Logan JG, Chai RC, et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat Genet*. Nature Publishing Group; 2019; 51: 258–266. <https://doi.org/10.1038/s41588-018-0302-x> PMID: 30598549
10. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*. Nature Publishing Group; 2015; 47: 1114–1120. <https://doi.org/10.1038/ng.3390> PMID: 26323059
11. Flint J, Eskin E. Genome-wide association studies in mice. *Nat Rev Genet*. 2012; 13: 807–817. <https://doi.org/10.1038/nrg3335> PMID: 23044826

12. Farber CR, Bennett BJ, Orozco L, Zou W, Lira A, Kostem E, et al. Mouse genome-wide association and systems genetics identify *Asxl2* as a regulator of bone mineral density and osteoclastogenesis. *PLoS Genet*. 2011; 7: e1002038. <https://doi.org/10.1371/journal.pgen.1002038> PMID: 21490954
13. Ackert-Bicknell CL, Demissie S, Tsaih S-W, Beamer WG, Cupples LA, Paigen BJ, et al. Genetic variation in *TRPS1* may regulate hip geometry as well as bone mineral density. *Bone*. 2012; 50: 1188–1195. <https://doi.org/10.1016/j.bone.2012.01.011> PMID: 22306695
14. Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet*. 2005; 6: 271–286. <https://doi.org/10.1038/nrg1576> PMID: 15803197
15. Ackert-Bicknell CL, Karasik D, Li Q, Smith RV, Hsu Y-H, Churchill GA, et al. Mouse BMD quantitative trait loci show improved concordance with human genome-wide association loci when recalculated on a new, common mouse genetic map. *J Bone Miner Res*. 2010; 25: 1808–1820. <https://doi.org/10.1002/jbmr.72> PMID: 20200990
16. Farber CR, van Nas A, Ghazalpour A, Aten JE, Doss S, Sos B, et al. An integrative genetics approach to identify candidate genes regulating BMD: combining linkage, gene expression, and association. *J Bone Miner Res*. 2009; 24: 105–116. <https://doi.org/10.1359/jbmr.080908> PMID: 18767929
17. Abiola O, Angel JM, Avner P, Bachmanov AA, Belknap JK, Bennett B, et al. The nature and identification of quantitative trait loci: a community's view. *Nat Rev Genet*. Nature Publishing Group; 2003; 4: 911–916. <https://doi.org/10.1038/nrg1206> PMID: 14634638
18. Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, et al. In silico mapping of complex disease-related traits in mice. *Science*. American Association for the Advancement of Science; 2001; 292: 1915–1918. <https://doi.org/10.1126/science.1058889> PMID: 11397946
19. Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, Lagler E, et al. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *Plos Biol*. Public Library of Science; 2004; 2: e393. <https://doi.org/10.1371/journal.pbio.0020393> PMID: 15534693
20. Cervino AC, Li G, Edwards S, Laurie C, Tokiwa G, Lum PY, et al. Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics*. 2005; 86: 505–517. <https://doi.org/10.1016/j.ygeno.2005.07.010> PMID: 16126366
21. McClurg P, B J, Wu C, Delano DL, Walker JR, Batalov S, et al. Genomewide association analysis in diverse inbred mice: power and population structure. *Genetics*. *Genetics*; 2007; 176: 675–683. <https://doi.org/10.1534/genetics.106.066241> PMID: 17409088
22. Payseur BA, Place M. Prospects for association mapping in classical inbred mouse strains. *Genetics*. *Genetics*; 2007; 175: 1999–2008. <https://doi.org/10.1534/genetics.106.067868> PMID: 17277361
23. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. *Genetics*. 2008; 178: 1709–1723. <https://doi.org/10.1534/genetics.107.080101> PMID: 18385116
24. Bogue MA, Grubb SC, Walton DO, Philip VM, Kolishovski G, Stearns T, et al. Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice. *Nucleic Acids Res*. 2018; 46: D843–D850. <https://doi.org/10.1093/nar/gkx1082> PMID: 29136208
25. Civelek M, Lusk AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. Nature Publishing Group; 2014; 15: 34–48. <https://doi.org/10.1038/nrg3575> PMID: 24296534
26. Farber CR, Lusk AJ. Integrating global gene expression analysis and genetics. *Adv Genet*. Elsevier; 2008; 60: 571–601. [https://doi.org/10.1016/S0065-2660\(07\)00420-8](https://doi.org/10.1016/S0065-2660(07)00420-8) PMID: 18358333
27. Mesner LD, Ray B, Hsu Y-H, Manichaikul A, Lum E, Bryda EC, et al. *Bicc1* is a genetic determinant of osteoblastogenesis and bone mineral density. *J Clin Invest*. 2014; 124: 2736–2749. <https://doi.org/10.1172/JCI73072> PMID: 24789909
28. Civelek M, Wu Y, Pan C, Raulerson CK, Ko A, He A, et al. Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am J Hum Genet*. 2017; 100: 428–443. <https://doi.org/10.1016/j.ajhg.2017.01.027> PMID: 28257690
29. Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. Learning from Co-expression Networks: Possibilities and Challenges. *Front Plant Sci*. *Frontiers*; 2016; 7: 444. <https://doi.org/10.3389/fpls.2016.00444> PMID: 27092161
30. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005; 4: Article17. <https://doi.org/10.2202/1544-6115.1128> PMID: 16646834
31. Calabrese GM, Mesner LD, Stains JP, Tommasini SM, Horowitz MC, Rosen CJ, et al. Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes *SPTBN1* and *MARK3* and an Osteoblast Functional Module. *Cell Systems*. 2017; 4: 46–59.e4. <https://doi.org/10.1016/j.cels.2016.10.014> PMID: 27866947

32. Bogue MA, Grubb SC, Walton DO, Philip VM, Kolishovski G, Stearns T, et al. Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice. *Nucleic Acids Res.* 2018; 46: D843–D850. <https://doi.org/10.1093/nar/gkx1082> PMID: 29136208
33. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA. National Academy of Sciences*; 2004; 101: 6062–6067. <https://doi.org/10.1073/pnas.0400782101> PMID: 15075390
34. Laufman O, Hong W, Lev S. The COG complex interacts directly with Syntaxin 6 and positively regulates endosome-to-TGN retrograde transport. *J Cell Biol.* 2011; 194: 459–472. <https://doi.org/10.1083/jcb.201102045> PMID: 21807881
35. Petit MM, Schoenmakers EF, Huysmans C, Geurts JM, Mandahl N, Van de Ven WJ. LHFP, a novel translocation partner gene of HMGIC in a lipoma, is a member of a new family of LHFP-like genes. *Genomics.* 1999; 57: 438–441. <https://doi.org/10.1006/geno.1999.5778> PMID: 10329012
36. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature.* 2011; 477: 289–294. <https://doi.org/10.1038/nature10413> PMID: 21921910
37. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature Methods.* 2010; 7: 248–249. <https://doi.org/10.1038/nmeth0410-248> PMID: 20354512
38. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015; 31: 2745–2747. <https://doi.org/10.1093/bioinformatics/btv195> PMID: 25851949
39. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 2012; 40: W452–7. <https://doi.org/10.1093/nar/gks539> PMID: 22689647
40. Leiserson MDM, Eldridge JV, Ramachandran S, Raphael BJ. Network analysis of GWAS data. *Curr Opin Genet Dev.* 2013; 23: 602–610. <https://doi.org/10.1016/j.gde.2013.09.003> PMID: 24287332
41. Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* 2010; 20: 281–290. <https://doi.org/10.1101/gr.099234.109> PMID: 20054062
42. Calabrese G, Bennett BJ, Orozco L, Kang HM, Eskin E, Dombret C, et al. Systems genetic analysis of osteoblast-lineage cells. *PLoS Genet.* 2012; 8: e1003150. <https://doi.org/10.1371/journal.pgen.1003150> PMID: 23300464
43. Karsenty G, Kronenberg HM, Settembre C. Genetic control of bone formation. *Annu Rev Cell Dev Biol.* 2009; 25: 629–648. <https://doi.org/10.1146/annurev.cellbio.042308.113308> PMID: 19575648
44. 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, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017; 550: 204–213. <https://doi.org/10.1038/nature24277> PMID: 29022597
45. Melville SA, Buros J, Parrado AR, Vardarajan B, Logue MW, Shen L, et al. Multiple loci influencing hippocampal degeneration identified by genome scan. *Ann Neurol. Wiley Subscription Services, Inc., A Wiley Company*; 2012; 72: 65–75. <https://doi.org/10.1002/ana.23644> PMID: 22745009
46. Nagaishi M, Kim Y-H, Mittelbronn M, Giangaspero F, Paulus W, Brokinkel B, et al. Amplification of the STOML3, FREM2, and LHFP genes is associated with mesenchymal differentiation in gliosarcoma. *Am J Pathol.* 2012; 180: 1816–1823. <https://doi.org/10.1016/j.ajpath.2012.01.027> PMID: 22538188
47. Bogue MA, Peters LL, Paigen B, Korstanje R, Yuan R, Ackert-Bicknell C, et al. Accessing Data Resources in the Mouse Phenome Database for Genetic Analysis of Murine Life Span and Health Span. *J Gerontol A Biol Sci Med Sci.* 2016; 71: 170–177. <https://doi.org/10.1093/gerona/glu223> PMID: 25533306
48. Ackert-Bicknell CL, Shockley KR, Horton LG, Lecka-Czernik B, Churchill GA, Rosen CJ. Strain-specific effects of rosiglitazone on bone mass, body composition, and serum insulin-like growth factor-I. *Endocrinology.* 2009; 150: 1330–1340. <https://doi.org/10.1210/en.2008-0936> PMID: 18948404
49. Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, et al. A customized and versatile high-density genotyping array for the mouse. *Nature Methods.* 2009; 6: 663–666. <https://doi.org/10.1038/nmeth.1359> PMID: 19668205
50. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv. Cold Spring Harbor Laboratory*; 2014;: 005165. <https://doi.org/10.1101/005165>
51. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 2016; 11: 1650–1667. <https://doi.org/10.1038/nprot.2016.095> PMID: 27560171

52. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 2003; 19: 889–890. PMID: [12724300](https://pubmed.ncbi.nlm.nih.gov/12724300/)
53. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. 2003; 13: 2498–2504. <https://doi.org/10.1101/gr.1239303> PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
54. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019; 47: D419–D426. <https://doi.org/10.1093/nar/gky1038> PMID: [30407594](https://pubmed.ncbi.nlm.nih.gov/30407594/)
55. Francis SP, Krey JF, Krystofiak ES, Cui R, Nanda S, Xu W, et al. A short splice form of Xin-actin binding repeat containing 2 (XIRP2) lacking the Xin repeats is required for maintenance of stereocilia morphology and hearing function. *J Neurosci. Society for Neuroscience*; 2015; 35: 1999–2014. <https://doi.org/10.1523/JNEUROSCI.3449-14.2015> PMID: [25653358](https://pubmed.ncbi.nlm.nih.gov/25653358/)
56. Dobson KR, Reading L, Haberey M, Marine X, Scutt A. Centrifugal isolation of bone marrow from bone: an improved method for the recovery and quantitation of bone marrow osteoprogenitor cells from rat tibiae and femuræ. *Calcif Tissue Int*. 1999; 65: 411–413. PMID: [10541770](https://pubmed.ncbi.nlm.nih.gov/10541770/)
57. Phinney DG, Kopen G, Isaacson RL, Prockop DJ. Plastic adherent stromal cells from the bone marrow of commonly used strains of inbred mice: variations in yield, growth, and differentiation. *J Cell Biochem*. 1999; 72: 570–585. PMID: [10022616](https://pubmed.ncbi.nlm.nih.gov/10022616/)
58. Bouxsein ML, Boyd SK, Christiansen BA, Guldberg RE, Jepsen KJ, Muller R. Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. 2010; 25: 1468–1486. <https://doi.org/10.1002/jbmr.141> PMID: [20533309](https://pubmed.ncbi.nlm.nih.gov/20533309/)
59. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2013 [cited 26 Aug 2016]. Available: <http://www.R-project.org/>
60. Lenth RV. Least-squares means: the R Package lsmeans. *J Stat Softw*. 2016. <https://doi.org/10.18637/jss.v069.i01>
61. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26: 2336–2337. <https://doi.org/10.1093/bioinformatics/btq419> PMID: [20634204](https://pubmed.ncbi.nlm.nih.gov/20634204/)