Expression quantitative trait locus study of bone mineral density GWAS variants in human osteoclasts

Benjamin H Mullin[1,2], Kun Zhu[1,2], Jiake Xu[2], Suzanne J Brown[1], Shelby Mullin[1,2], Jennifer Tickner[2], Nathan J Pavlos[2], Frank Dudbridge[4], John P Walsh[1,3], Scott G Wilson[1,2,5]

[1]Department of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA, Australia

[2]School of Biomedical Sciences, University of Western Australia, Crawley, WA, Australia

[3]Medical School, University of Western Australia, Crawley, WA, Australia

[4]Department of Health Sciences, University of Leicester, Leicester, UK

[5]Department of Twin Research & Genetic Epidemiology, King's College London, London, UK

Corresponding author

Benjamin H. Mullin
School of Biomedical Sciences
The University of Western Australia
Crawley, Western Australia 6009
T: +61 8 6457 2466
E-mail: Benjamin.Mullin@uwa.edu.au

A supplemental table has been included with this submission.

Disclosures

All authors state that they have no conflicts of interest.

Abstract

Osteoporosis is a complex disease with a strong genetic component. Genome-wide association studies (GWAS) have been very successful at identifying common genetic variants associated with bone parameters. A recently published study documented the results of the largest GWAS for bone mineral density (BMD) performed to date (n=142,487), identifying 307 conditionally independent single-nucleotide polymorphisms (SNPs) as associated with estimated BMD (eBMD) at the genome-wide significance level. The vast majority of these variants are non-coding SNPs. Expression quantitative trait locus (eQTL) studies using disease-specific cell types have increasingly been integrated with the results from GWAS to identify genes through which the observed GWAS associations are likely mediated. We generated a unique human osteoclast-specific eQTL dataset using cells differentiated *in vitro* from 158 participants. We then used this resource to characterise the 307 recently identified BMD GWAS SNPs for association with nearby genes (+/- 500 kb). After correction for multiple testing, 24 variants were found to be significantly associated with the expression of 32 genes in the osteoclast-like cells. Bioinformatics analysis suggested that these variants and those in strong linkage disequilibrium with them are enriched in regulatory regions. Several of the eQTL associations identified are relevant to genes that present strongly as having a role in bone, particularly *IQGAP1*, *CYP19A1*, *CTNNB1* and *COL6A3*. Supporting evidence for many of the associations was obtained from publicly available eQTL datasets. We have also generated strong evidence for the presence of a regulatory region on chromosome 15q21.2 relevant to both the *GLDN* and *CYP19A1* genes. In conclusion, we have generated a unique osteoclast-specific eQTL resource and have used this to identify 32 eQTL associations for recently identified BMD GWAS loci, which should inform functional studies of osteoclast biology.

Key words: Cell/tissue signalling - paracrine pathways, Osteoclasts, Osteoporosis, Genetic research, Human association studies.

Introduction

Osteoporosis is a common and debilitating disease characterised by low bone mineral density (BMD), deterioration of the bone micro-architecture and increased fracture risk. The disease is particularly prevalent in postmenopausal women due to reduced oestrogen levels. Environmental factors such as dietary calcium intake and exercise have also been shown to have a role in the disease.[1,2] In addition to these effects, it has been demonstrated that there is a strong genetic component to osteoporosis. Twin and family studies have generated BMD heritability estimates of 0.46-0.92 depending on the anatomical site studied,[3,4] and individuals with an affected first-degree relative have an elevated estimated familial relative risk of fragility fracture of 1.31-4.24.[5,6]

Genome-wide association studies (GWAS) have been very successful at identifying common genetic variants associated with bone parameters.[7-11] A recent GWAS performed in 142,487 individuals from the UK Biobank, the largest GWAS performed so far for a bone density phenotype, identified 307 conditionally independent single-nucleotide polymorphisms (SNPs) located in 203 loci as associated with estimated BMD (eBMD) at the genome-wide significance level.[12] Like many complex trait GWAS, the vast majority of these are non-coding variants localised in intergenic or intronic DNA, suggesting that they may influence the eBMD phenotype through regulatory effects on nearby genes. However, for many of these variants the identity of the gene responsible for mediating the observed effect on BMD remains to be determined. Expression quantitative trait locus (eQTL) studies, which characterise associations between genetic variants and gene expression at the cellular level, have increasingly been integrated with the results from disease GWAS to specify the genes involved. However, many of these studies have used complex tissues and, apart from a study using osteoblasts cultured from surgical explants of bone,[13] there is a lack of data that is specific for bone cells.

In an effort to identify some of the regulatory effects caused by BMD GWAS variants, we have generated a unique eQTL dataset using osteoclast-like cells differentiated *in vitro* from peripheral blood mononuclear cells (PBMCs) obtained from 158 subjects. We have integrated this information with that from the recently published eBMD GWAS, selected due to its large sample size and strong power to detect risk loci, to identify genetic regulatory effects relevant to osteoporosis and to gain a better understanding of the biological pathways underlying the disease.

Materials and Methods

Subject recruitment

We recruited 158 female patients aged 30-70 undergoing dual-energy X-ray absorptiometry (DXA) BMD scanning (Hologic, Bedford, MA, USA) in the Bone Density Unit at Sir Charles Gairdner Hospital in Western Australia. All subjects had self-reported European ancestry. Patients suffering from medical conditions or using medications that are likely to affect osteoclastic bone resorption or the ability of their cells to undergo osteoclastogenesis were excluded. This included rheumatoid arthritis, primary hyperparathyroidism, oral glucocorticoid treatment or the medications alendronate, risedronate, zoledronic acid or denosumab. One 4 ml ethylenediaminetetraacetic acid (EDTA) and two 6 ml lithium heparin blood tubes were collected from each patient. All participants gave written informed consent and the study was approved by the Sir Charles Gairdner and Osborne Park Health Care Group Human Research Ethics Committee.

Isolation of peripheral blood mononuclear cells and osteoclastogenesis

PBMCs were isolated from each pair of 6 ml lithium heparin blood tubes by density gradient centrifugation using protocols well established in our laboratory.[14] Briefly, each pair of blood tubes was centrifuged at 2,200 rpm for 13 min at room temperature. The buffy coats were harvested and diluted to a total volume of 4 ml with 1× phosphate buffered saline (PBS) before being gently layered over 3 ml of Ficoll-Paque Premium (GE Healthcare) and centrifuged again at 1,600 rpm for 45 min (no brake). The PBMC layer was collected and washed by re-suspending in 6 ml 1× PBS, centrifuging at 1,000 rpm for 12 min then removing the supernatant. The wash step was repeated, with the resulting cell pellet re-suspended in 1 ml complete α-MEM (minimum essential medium) supplemented with 25 ng/ml macrophage colony stimulating factor (M-CSF). A cell count was then performed, with

each sample used to seed 3 wells (triplicates) of a 24-well cell culture plate with 1.5×10^6 cells each. After two days, the medium was replaced with α-MEM supplemented with 25 ng/ml M-CSF and 100 ng/ml receptor activator of nuclear factor kappa-B ligand (RANKL). The cells were then grown in this medium formulation for another 12 days while osteoclastogenesis occurred.

Cultures of osteoclast-like cells were stained for tartrate resistant acid phosphatase (TRAP) using a chromogenic TRAP enzyme substrate to confirm production of the TRAP enzyme as an indicator of the osteoclast phenotype. This involved washing the cells with 1× PBS, fixation with 4% (v/v) paraformaldehyde for 10 min, washing twice with 1× PBS before incubation with filtered TRAP stain solution at 37°C for 40 min. The stained cells were then washed with sterile water prior to visualisation using light microscopy.

Nucleic acid extraction

Genomic DNA was extracted from 200 µl EDTA blood for each patient using the QIAamp DNA Blood Mini Kit (QIAGEN) according to the manufacturer's instructions. Each batch of osteoclast-like cell cultures was washed once with 1× PBS prior to nucleic acid extraction at day 14 of culture. Total RNA and genomic DNA were extracted from each set of triplicate cultures using the AllPrep DNA/RNA Mini Kit (QIAGEN) according to the manufacturer's instructions, with on-column DNase digestion for the RNA fraction. Once cell lysis had been performed, the cell lysate from each of the 3 cultures for each sample was combined into a single aliquot. RNA integrity numbers (RINs) were assessed for each RNA sample using the Agilent 2100 Bioanalyzer, with all samples recording RIN values ≥9.7 indicating that the RNA samples used in this study were of very high quality.

Genotyping and imputation

Genotyping of the genomic DNA extracted from each EDTA blood sample was performed by the Australian Genome Research Facility (AGRF) using the Illumina Infinium OmniExpress-24 BeadChip array. QC criteria applied to the genotype data prior to imputation included removal of individuals with a call rate <90%, removal of variants that were monomorphic, unmapped, minor allele frequency (MAF) <5%, Hardy-Weinberg equilibrium $P<5\times10^{-8}$ or call rate <90%, leaving 572,898 variants for imputation. Genotype imputation was performed by the Sanger Imputation Service using the Haplotype Reference Consortium (HRC) release 1.1 reference panel[15] with pre-phasing by Eagle v2.3.2[16] and imputation performed by PBWT (Positional Burrows-Wheeler Transform).[17] Post-imputation QC included removal of variants with an IMPUTE2 info score <0.4.

Relatedness testing and principal components analysis

Relatedness testing and principal components analysis (PCA) were performed on the genotyped data using Plink v1.9.[18] Ten principal components were generated for the cohort and retained for use as covariates in the eQTL analysis to correct for population stratification.

Gene expression data processing

Quantitation of gene expression was performed by AGRF on the RNA samples extracted from the osteoclast-like cells using 50 bp single-end RNA-Seq on an Illumina HiSeq 2500. The primary sequence data was generated using the Illumina bcl2fastq 2.19.0.316 pipeline. Per-base sequence quality for all 158 samples passed QC, with >96% of the bases above Q30 across all samples. Cleaned sequence reads were aligned against the Homo sapiens genome (build version HG38) and the TopHat aligner v2.0.14[19] was used to map reads to genomic sequences. Raw counts were summarised at gene level (GENCODE v25) using the featureCounts v1.4.6-p5 utility of the Subread package.[20] Genes with a read count <1 per million were removed from the dataset, as were those expressed in <10 individuals. Trimmed

mean of M-values (TMM) normalisation and correction of the gene expression data for total read count by conversion to counts per million (CPM) was performed using the edgeR package.[21] Calculation of reads per kilobase per million (RPKM) values for comparison of expression levels between different genes was also performed using the edgeR package.[21]

eQTL analysis

The eQTL analysis was performed on the TMM normalised CPM values and only included variants with a MAF ≥5%. We performed a hypothesis-driven association analysis of quantile-normalised gene expression levels with probabilities of imputed genotypes for 307 genetic variants using FastQTL,[22] which performs linear regressions between the genotypes and gene expression phenotypes. These 307 variants were recently identified as independently associated at the genome-wide significance level with eBMD, a phenotype derived from quantitative ultrasound measurements at the heel, in a GWAS performed in 142,487 individuals from the UK Biobank Study.[12] This is the largest GWAS performed to date for a bone density phenotype, and the 307 genetic variants collectively accounted for ~12% of the phenotype variance in that study. The eQTL analysis was corrected for the covariates patient age, RNA-Seq batch and 10 principal components. Each variant was tested for association with the expression of any gene where the transcription start site (TSS) fell within a 500 kb window on either side of the variant (*cis*-eQTLs). The 500 kb window size was selected as it has been shown that the vast majority of *cis*-eQTL variants are located well within this distance of their gene TSS,[23] and this represents a compromise between capturing as many *cis*-eQTLs as possible without incurring excessive multiple testing penalties. Correction for multiple testing was performed using the Benjamini-Yekutieli procedure[24] using a false discovery rate (FDR) of 5%.

Bioinformatics analysis

Genetic variants identified as significantly associated with expression of nearby genes were analysed using the GREGOR (Genomic Regulatory Elements and Gwas Overlap algorithm) v1.4.0 software,[25] which tests for enrichment of a variant set and its linkage disequilibrium (LD) proxies in Encyclopedia of DNA Elements (ENCODE) regulatory elements relative to MAF, TSS-distance and number of LD neighbours-matched null SNP sets. We used GREGOR to test for enrichment of the eQTL-variants (eSNPs) in the University of California, Santa Cruz (UCSC) ENCODE datasets for DNase Hypersensitivity Clusters (125 cell types, V3) and Transcription Factor ChIP-seq (161 factors, V3). The following parameters were used, as reported in the original GREGOR publication: r^2 threshold = 0.8, LD window size = 1 Mb and minimum neighbour number = 500. Assessment of co-localization of eBMD and eQTL association signals within 500 kb of each eQTL-gene (eGene) TSS was performed using the coloc package in R,[26] which uses a Bayesian framework to calculate posterior probabilities to quantify support for 5 different hypotheses regarding the presence and sharing of causal variants for traits of interest. Bioinformatics analysis to determine potential regulatory effects of individual eSNPs was performed using HaploReg v4.1[27] and RegulomeDB[28], with LD data derived from HaploReg v4.1 (1000G Phase 1 EUR population).

Results

An overview of the steps used in this study to generate the genotype and gene expression data and perform the eQTL analysis is presented in Fig. 1. The demographics of the 158 women aged 30-70 years with self-reported European ancestry recruited into the study are presented in Table 1. No outliers were observed in the PCA. After genotype imputation and filtering, we had genotype data for 5,373,348 variants in the study sample with a MAF ≥5% and an IMPUTE2 info score ≥0.4. After applying QC criteria to the RNA-Seq data, 15,688 expressed gene transcripts were identified.

Association of eBMD variants with gene expression

Of the 307 eBMD-associated genetic variants included in the analysis, 199 had a MAF ≥5% in our study cohort and an expressed gene located within the +/-500 kb analysis window (mean 7.6 genes per 199 variant). After correction for multiple testing, 24 of these variants were found to be eSNPs significantly associated with the expression of 32 genes (Table 2), including genes with a strong potential role in bone such as *IQGAP1*[29] (Fig. 2A), *CYP19A1*[30,31] (Figs. 2B and 3), *CTNNB1*[7,32] (Fig. 2C) and *COL6A3*[33] (Fig. 2D), as well as many with no known role in bone biology. As expected, a number of the variants demonstrated associations with more than one gene, including rs6680737 (*WARS2* and *RP11-418J17.1*), rs11636403 (*GLDN* and *CYP19A1*) (Figs. 2B and 3), rs58057291 (*MLPH* and *COL6A3*) (Fig. 2D), rs757980 (*CHN2* and *CREB5*), rs72767980 (*FAM129B* and *ZNF79*) and rs2696264 (*CACNA1G*, *EPN3*, *ACSF2*, *PPP1R9B*). Since eSNPs influence gene expression, it is anticipated that many are located in close proximity to the eGene TSS. Out of the 32 eSNP-eGene associations, 23 (72%) were for variants located within 250 kb of the eGene TSS (Table 2). We also found that the associations were stronger for variants located in close proximity to the TSSs and the five strongest eSNP-eGene associations were for variants

located in the 160 kb region immediately 5' of the TSS, suggesting enrichment of these variants in regulatory regions. Bioinformatics analysis using the GREGOR software supported this, with statistically significant enrichment of the 24 eSNPs and their LD proxies ($r^2 \geq 0.8$) observed in DNase hypersensitivity clusters and transcription factor binding sites from ENCODE relative to matched null SNP sets (*P=0.0004* and *0.04* respectively). Co-localization analysis yielded evidence for co-localization of eBMD and eQTL association signals (posterior probability >50%) at 8 loci, including *SF3A3* (63%), *MLPH* (64%), *SERPINB1* (75%), *FADS2* (88%), *GLDN* (94%), *CYP19A1* (94%), *IQGAP1* (57%) and *CACNA1G* (52%).

Replication in other cell/tissue types

We queried each of the 24 eSNPs using the GTEx[23] and Blood eQTL[34] browsers for evidence of replication in other cell/tissue types. For the GTEx portal, eSNPs from our analysis were queried against pre-calculated eQTLs for tissues having more than 70 samples, with a multiple-testing corrected q-value threshold of 0.05 used to identify significant associations. The Blood eQTL dataset was generated using peripheral blood samples from 5,311 individuals, with a multiple-testing corrected FDR threshold of 0.5 used. We found supporting evidence for 14 of the 32 eSNP-eGene associations, including associations in multiple tissue types for rs3790608 – *ST7L* (11 tissues, $P=1.9\times10^{-5}–4.2\times10^{-11}$), rs6680737 – *WARS2* (22 tissues, $P=1.9\times10^{-5}–2\times10^{-35}$), rs6680737 – *RP11-418J17.1* (20 tissues, $P=3.1\times10^{-5}–6.3\times10^{-16}$), rs4360494 – *SF3A3* (27 tissues, $P=2.3\times10^{-5}–2.3\times10^{-37}$), rs58057291 – *MLPH* (3 tissues, $P=3.7\times10^{-5}–1.6\times10^{-13}$), rs6547870 – *TRMT61B* (20 tissues, $P=1.2\times10^{-5}–3\times10^{-13}$), rs6829296 – *DGKQ* (10 tissues, $P=1.1\times10^{-5}–1.5\times10^{-12}$), rs174574 – *FADS2* (19 tissues, $P=6.6\times10^{-5}–1.5\times10^{-23}$), rs11637971 – *IQGAP1* (33 tissues, $P=2.5\times10^{-5}–4.5\times10^{-33}$) and rs2301522 – *TMEM8A* (12 tissues, $P=1.7\times10^{-5}–5.5\times10^{-23}$). Single-tissue associations were seen for rs4589135 – *PIGV* (whole blood, $P=4.4\times10^{-9}$), rs6761129 – *THADA*

(subcutaneous adipose tissue, $P=1.7\times10^{-5}$), rs7578166 – *DYSF* (testis, $P=3.2\times10^{-7}$) and rs757980 – *CREB5* (whole blood, $P=1.3\times10^{-9}$).

Individual variant bioinformatics analysis

Analysis of the 24 eSNPs and their LD proxies ($r^2\geq0.8$) using the HaploReg v4.1[27] and RegulomeDB[28] tools yielded a large number of variants with potential regulatory effects, often with multiple observed for a single locus (including at least 11 for the rs174574 – *FADS2* locus and 20 for the rs6761129 – *THADA* locus) (Supplemental Table 1). Five of the 24 eSNPs were not found to be in strong LD ($r^2\geq0.8$) with any other variants: rs3790608, rs6829296, rs757980, rs72767980 and rs62028332. We confirmed the LD around these variants using the updated 1000G Phase 3 data for the EUR population, and discovered a variant in strong LD with rs6829296 (rs111588353, $r^2=0.97$). The other 4 eSNPs present strongly as having potential regulatory effects and, considering the lack of LD proxies, it is possible that these may be the quantitative trait nucleotides responsible for the associations seen between eBMD and gene expression. The bioinformatics findings for these 4 eSNPs are described below.

rs3790608 (*ST7L*) is predicted to be located in enhancer histone marks in 8 tissues and alter the regulatory motifs DMRT4 and DMRT5. RegulomeDB assigned rs3790608 a strong score of 1f (likely to affect binding and linked to expression of a gene target).

rs757980 (*CHN2* and *CREB5*) presents very strongly as having a potential regulatory role. It is predicted to be located in a GERP (Genomic Evolutionary Rate Profiling) conserved region, in promoter histone marks in 21 tissues, enhancer histone marks in 8 tissues, DNase hypersensitivity sites in 51 tissues, binding sites for 16 transcription factors and to alter the regulatory motifs E2F, Nkx2 and Nkx3.

rs72767980 (*FAM129B* and *ZNF79*) is predicted to lie in promoter histone marks in 4 tissues, enhancer histone marks in 22 tissues, DNase hypersensitivity sites in 13 tissues and alter 13 regulatory motifs.

rs62028332 (*SNX20*) is predicted to lie in a GERP conserved region, promoter histone marks in 2 tissues and alter the regulatory motifs CEBPB and Hoxa7.

Discussion

In this study we have identified a number of significant associations between eBMD GWAS SNPs and gene expression levels in osteoclast-like cells. Although not routinely used clinically in the diagnosis of osteoporosis, the quantitative ultrasound measurements that are used to calculate the eBMD phenotype are moderately correlated with DXA BMD at the hip and spine (*r=0.4–0.6*)[35], and the eBMD GWAS results provided evidence of replication for 54 out of the 64 DXA BMD GWAS loci described by Estrada et al.[7] Our bioinformatics analysis suggests that the eSNPs identified in this study and their LD proxies are enriched in regulatory regions, and we were able to demonstrate strong predicted regulatory effects for many individual variants. Interestingly, we were able to identify significant eQTL associations for several genes that appeared to be expressed at a lower level in this cell type. Although it is tempting to deprioritise the findings for these genes or question if they may be false positives, the validity of many (such as *ST7L*, *CREB5*, *WARS2*, *RP11-418J17.1*) was supported with publicly available data from the GTEx[23] and/or Blood eQTL datasets,[36] suggesting that they may have a legitimate role in bone, acting through mechanisms that are as yet unknown and achieved with less transcription.

One of the most convincing eQTL associations identified in this study was between the variant rs11637971 and expression of the gene *IQGAP1*, which encodes the IQ motif containing GTPase activating protein 1. This gene was found to be expressed at a high level in the osteoclast-like cells and supporting evidence for the eQTL association was seen in both the GTEx[23] and Blood eQTL[36] datasets, with consistent allelic effects. We also observed evidence of co-localization of the eBMD and eQTL association signals at this locus, suggesting the presence of a single variant affecting both traits. The product of the *IQGAP1* gene is a ubiquitously expressed scaffolding protein that has been shown to have a role in podosome/sealing zone formation in mouse osteoclasts.[29] Consistent with the potential role

for this gene in osteoclastic bone resorption, the *C* allele at rs11637971 was associated with an increased eBMD[12] and reduced *IQGAP1* expression. Interestingly, the IQGAP1 protein has been found to interact with the RhoA GTPase,[37] and we have previously published evidence suggesting a role for variation in the *RHOA* gene in regulation of BMD.[38] Expression of *IQGAP1* has also been found to strongly positively correlate ($P=9.1\times10^{-5}$) with that of the *RHOA* gene in iliac crest bone biopsies obtained from 84 postmenopausal women.[39]

eQTL associations were identified between the variant rs11636403 and the genes *CYP19A1* and *GLDN*. This variant demonstrated strong associations with expression of both of these genes, which to our knowledge have not been previously identified in any cell/tissue type, and we believe we may have identified a regulatory region located in intron 2 of *CYP19A1* (relative to transcript NM_031226) that is relevant to both genes (Fig. 3). Both of these genes also demonstrated strong evidence for co-localization of eBMD and eQTL association signals. The product of the *GLDN* gene, gliomedin, has a role in the formation of the nodes of Ranvier[40] and does not currently have an established role in bone metabolism. The *CYP19A1* gene on the other hand does present strongly as having a role in bone, and we have previously demonstrated that variation in this gene is associated with postmenopausal osteoporosis.[31] The *CYP19A1* gene encodes the aromatase enzyme, which catalyses the aromatisation of androgens to oestrogens. There is an established relationship between oestrogen deficiency and osteoporosis, thought to be mediated through loss of the restraining effect that oestrogen has on bone resorption, as well as effects on extra-skeletal calcium homeostasis.[41] Aromatase inhibitors, commonly used for treatment of patients with hormone-receptor positive breast carcinoma, are known to increase osteoporosis risk.[42] Consistent with the positive effect of oestrogen on bone, we found that the *T* allele at rs11636403, which was associated with increased expression of *CYP19A1* in this study, was

associated with increased eBMD.[12] The fact that this eQTL association has only been demonstrated in osteoclast-like cells is interesting and could suggest the existence of a local intercellular signalling (i.e. paracrine) mechanism within the basic multicellular unit coupling the activities of bone resorption and formation,[43] and mediated through increased localised production of oestrogen.

The eBMD GWAS variant rs370387 was identified in this study as significantly associated with expression of the gene *CTNNB1*, which encodes the intracellular signal transducer beta-catenin. Beta-catenin has a major role in the canonical Wnt signalling pathway, which is an important regulator of skeletal homeostasis. Loss of beta-catenin has been shown to inhibit osteoblastogenesis[44,45] and reduce production of osteoprotegerin by differentiated osteoblasts,[46] thus increasing osteoclast differentiation and bone resorption. Studies in osteoclasts have suggested that beta-catenin is required for osteoclast precursor proliferation, but that it inhibits differentiation of these precursors into mature osteoclasts.[47] Consistent with this pro-osteogenic role for beta-catenin in bone cells, we found that the *A* allele at rs370387, which is associated with an increased eBMD,[12] was associated with increased expression of *CTNNB1*.

We identified two eQTL associations for the variant rs58057291 in our dataset for the genes *MLPH* and *COL6A3*. The *MLPH* gene encodes melanophilin, which forms a complex with the small Ras-related GTPase Rab27A and the motor protein myosin Va and is involved with melanosome transport. This gene does not have an obvious role in bone metabolism, however the *COL6A3* gene does present strongly as having a role in bone. This gene encodes the alpha 3 chain of type VI collagen, an extracellular matrix protein found in almost all connective tissues including bone. Mutations in this gene have been found to cause Bethlem myopathy[48] and Ullrich congenital muscular dystrophy,[49] both of which are characterised by muscle weakness and joint contractures. Expression of *COL6A3* in articular cartilage and

subchondral bone samples obtained from osteoarthritic knee tissue has been shown to correlate with osteoarthritis severity and bone structural abnormalities.[33] The *A* allele at rs58057291 was associated with increased expression of *MLPH* and *COL6A3*, and a reduced eBMD.[12]

One limitation of our eQTL study is that the cells used were cultured and differentiated *in vitro*, and therefore may not accurately reflect gene expression in osteoclasts *in vivo*. It should also be noted that some of the eSNP-eGene associations identified in this study are for genes that do not present as the most likely bone candidate gene from the chromosomal region. For example, the variant rs2696264 was found to be associated with the expression of *CACNA1G*, *EPN3*, *ACSF2* and *PPP1R9B* in this study. However, the *COL1A1* gene is also located in this locus (around 54 kb from rs2696264) and presents as the most likely BMD candidate gene from the region, having long been associated with osteogenesis imperfecta.[50] This phenomenon could be due to the fact that if rs2696264 is influencing BMD through regulatory effects on the *COL1A1* gene, these effects are likely not mediated through the osteoclast. It is well established that many regulatory regions are cell-type specific.[51] Interestingly, the rs2696264 variant has no demonstrated associations with expression of *COL1A1* in any of the 53 tissues studied by the GTEx project.

In conclusion, we have generated a unique osteoclast-specific eQTL dataset and have identified 32 eSNP-eGene associations for genetic variants recently identified as associated with BMD. The eSNPs are enriched in predicted regulatory regions and several of the genes that appear to be regulated by these variants present strongly as having a role in bone, particularly *IQGAP1*, *CYP19A1*, *CTNNB1* and *COL6A3*. Supporting evidence for many of the associations identified in this study was obtained from publicly available eQTL datasets. We have also generated strong evidence for the presence of a regulatory region on chromosome 15q21.2 relevant to both the *GLDN* and *CYP19A1* genes. These findings

highlight some of the genes that BMD GWAS variants likely influence, and the results from this study will help to target future research in the area.

Acknowledgements

We are very grateful to the women who kindly participated in the study and the staff involved including nursing staff, bone density technologists and clerical staff.

Authors' roles: Study conception and design: BM, KZ, JW, and SW. Recruitment and cell culture: BM, KZ, and SM. Data analysis: BM, SB, and FD. Interpretation of data: BM, KZ, JX, SB, JT, NP, FD, JW, and SW. Drafting manuscript: BM, and SW. Revising manuscript content: BM, KZ, JX, SB, SM, JT, NP, FD, JW, and SW. Approving final version of manuscript: BM, and SW. BM takes full responsibility for the integrity of the data analysis.

Funding

This work was supported by the Australian National Health and Medical Research Council (Project Grants 1048216, 1127156), the Sir Charles Gairdner Osborne Park Health Care Group (SCGOPHCG) Research Advisory Committee (Grant 2016-17/017) and the iVEC/Pawsey Supercomputing Centre (with funding from the Australian Government and the Government of Western Australia; Project Grants: Pawsey0162 (S.G.W.), Director2025 (S.G.W.)). The salary of B.H.M. was supported by a Raine Medical Research Foundation Priming Grant.

References

1. Specker BL. Evidence for an interaction between calcium intake and physical activity on changes in bone mineral density. J Bone Miner Res. Oct 1996;11(10):1539-44. Epub 1996/10/01.
2. Uusi-Rasi K, Sievanen H, Vuori I, Pasanen M, Heinonen A, Oja P. Associations of physical activity and calcium intake with bone mass and size in healthy women at different ages. J Bone Miner Res. Jan 1998;13(1):133-42. Epub 1998/01/27.
3. Krall EA, Dawson-Hughes B. Heritable and life-style determinants of bone mineral density. J Bone Miner Res. Jan 1993;8(1):1-9. Epub 1993/01/01.
4. Pocock NA, Eisman JA, Hopper JL, Yeates MG, Sambrook PN, Eberl S. Genetic determinants of bone mass in adults. A twin study. J Clin Invest. Sep 1987;80(3):706-10. Epub 1987/09/01.
5. Deng HW, Chen WM, Recker S, Stegman MR, Li JL, Davies KM, et al. Genetic determination of Colles' fracture and differential bone mass in women with and without Colles' fracture. J Bone Miner Res. Jul 2000;15(7):1243-52. Epub 2000/07/14.
6. Keen RW, Hart DJ, Arden NK, Doyle DV, Spector TD. Family history of appendicular fracture and risk of osteoporosis: a population-based study. Osteoporos Int. 1999;10(2):161-6. Epub 1999/09/29.
7. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, 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. May 2012;44(5):491-501. Epub 2012/04/17.
8. Mullin BH, Walsh JP, Zheng HF, Brown SJ, Surdulescu GL, Curtis C, et al. Genome-wide association study using family-based cohorts identifies the WLS and CCDC170/ESR1 loci as associated with bone mineral density. BMC Genomics. 2016;17(1):136. Epub 2016/02/26.
9. Mullin BH, Zhao JH, Brown SJ, Perry JRB, Luan J, Zheng HF, et al. Genome-wide association study meta-analysis for quantitative ultrasound parameters of bone identifies five novel loci for broadband ultrasound attenuation. Hum Mol Genet. Jul 15 2017;26(14):2791-802. Epub 2017/05/05.
10. Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, et al. Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet. May 3 2008;371(9623):1505-12. Epub 2008/05/06.
11. Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature. Oct 1 2015;526(7571):112-7. Epub 2015/09/15.
12. Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youlten SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet. Sep 04 2017. Epub 2017/09/05.
13. Grundberg E, Kwan T, Ge B, Lam KC, Koka V, Kindmark A, et al. Population genomics in a disease targeted primary cell model. Genome Res. Nov 2009;19(11):1942-52. Epub 2009/08/06.
14. Mullin BH, Mamotte C, Prince RL, Wilson SG. Influence of ARHGEF3 and RHOA knockdown on ACTA2 and other genes in osteoblasts and osteoclasts. PloS one. 2014;9(5):e98116. Epub 2014/05/21.

15. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. Oct 2016;48(10):1279-83. Epub 2016/08/23.
16. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. Nov 2016;48(11):1443-8. Epub 2016/10/28.
17. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. May 01 2014;30(9):1266-72. Epub 2014/01/15.
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. Epub 2015/02/28.
19. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. Apr 25 2013;14(4):R36. Epub 2013/04/27.
20. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. Apr 01 2014;30(7):923-30. Epub 2013/11/15.
21. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. Jan 01 2010;26(1):139-40. Epub 2009/11/17.
22. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. May 15 2016;32(10):1479-85. Epub 2015/12/29.
23. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. May 8 2015;348(6235):648-60. Epub 2015/05/09.
24. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics. 2001;29(4):1165-88.
25. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics. Aug 15 2015;31(16):2601-6. Epub 2015/04/19.
26. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS genetics. May 2014;10(5):e1004383. Epub 2014/05/17.
27. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. Jan 2012;40(Database issue):D930-4. Epub 2011/11/09.
28. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. Sep 2012;22(9):1790-7. Epub 2012/09/08.
29. Steenblock C, Heckel T, Czupalla C, Espirito Santo AI, Niehage C, Sztacho M, et al. The Cdc42 guanine nucleotide exchange factor FGD6 coordinates cell polarity and endosomal membrane recycling in osteoclasts. J Biol Chem. Jun 27 2014;289(26):18347-59. Epub 2014/05/14.
30. Morishima A, Grumbach MM, Simpson ER, Fisher C, Qin K. Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens. J Clin Endocrinol Metab. Dec 1995;80(12):3689-98. Epub 1995/12/01.

31. Mullin BH, Carter KW, Lewis JR, Ingley E, Wilson SG, Prince RL. Significant association between common polymorphisms in the aromatase gene CYP19A1 and bone mineral density in postmenopausal women. Calcif Tissue Int. Dec 2011;89(6):464-71. Epub 2011/09/29.
32. Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. Nat Genet. Nov 2009;41(11):1199-206. Epub 2009/10/06.
33. Chou CH, Lee CH, Lu LS, Song IW, Chuang HP, Kuo SY, et al. Direct assessment of articular cartilage and underlying subchondral bone reveals a progressive gene expression change in human osteoarthritic knees. Osteoarthritis Cartilage. Mar 2013;21(3):450-61. Epub 2012/12/12.
34. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. Letter 2013;45(10):1238-43.
35. Gonnelli S, Cepollaro C, Gennari L, Montagnani A, Caffarelli C, Merlotti D, et al. Quantitative ultrasound and dual-energy X-ray absorptiometry in the prediction of fragility fracture in men. Osteoporos Int. Aug 2005;16(8):963-8. Epub 2004/12/16.
36. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. Oct 2013;45(10):1238-43. Epub 2013/09/10.
37. Casteel DE, Turner S, Schwappacher R, Rangaswami H, Su-Yuo J, Zhuang S, et al. Rho isoform-specific interaction with IQGAP1 promotes breast cancer cell proliferation and migration. J Biol Chem. Nov 02 2012;287(45):38367-78. Epub 2012/09/21.
38. Mullin BH, Prince RL, Mamotte C, Spector TD, Hart DJ, Dudbridge F, et al. Further genetic evidence suggesting a role for the RhoGTPase-RhoGEF pathway in osteoporosis. Bone. Aug 2009;45(2):387-91. Epub 2009/05/12.
39. Reppe S, Sachse D, Olstad OK, Gautvik VT, Sanderson P, Datta HK, et al. Identification of transcriptional macromolecular associations in human bone using browser based in silico analysis in a giant correlation matrix. Bone. Mar 2013;53(1):69-78. Epub 2012/12/01.
40. Eshed Y, Feinberg K, Poliak S, Sabanay H, Sarig-Nadir O, Spiegel I, et al. Gliomedin mediates Schwann cell-axon interaction and the molecular assembly of the nodes of Ranvier. Neuron. Jul 21 2005;47(2):215-29. Epub 2005/07/26.
41. Riggs BL, Khosla S, Melton LJ, 3rd. A unitary model for involutional osteoporosis: estrogen deficiency causes both type I and type II osteoporosis in postmenopausal women and contributes to bone loss in aging men. J Bone Miner Res. May 1998;13(5):763-73. Epub 1998/06/04.
42. Choksi P, Williams M, Clark PM, Van Poznak C. Skeletal manifestations of treatment of breast cancer. Current osteoporosis reports. Dec 2013;11(4):319-28. Epub 2013/10/18.
43. Martin TJ. Reflecting on Some Discoveries of 40 Years and Their Outcomes. J Bone Miner Res. Oct 2017;32(10):1971-6. Epub 2017/08/30.
44. Day TF, Guo X, Garrett-Beal L, Yang Y. Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis. Developmental cell. May 2005;8(5):739-50. Epub 2005/05/04.
45. Hill TP, Spater D, Taketo MM, Birchmeier W, Hartmann C. Canonical Wnt/beta-catenin signaling prevents osteoblasts from differentiating into chondrocytes. Developmental cell. May 2005;8(5):727-38. Epub 2005/05/04.

46. Glass DA, 2nd, Bialek P, Ahn JD, Starbuck M, Patel MS, Clevers H, et al. Canonical Wnt signaling in differentiated osteoblasts controls osteoclast differentiation. Developmental cell. May 2005;8(5):751-64. Epub 2005/05/04.
47. Wei W, Zeve D, Suh JM, Wang X, Du Y, Zerwekh JE, et al. Biphasic and dosage-dependent regulation of osteoclastogenesis by beta-catenin. Mol Cell Biol. Dec 2011;31(23):4706-19. Epub 2011/08/31.
48. Pan TC, Zhang RZ, Pericak-Vance MA, Tandan R, Fries T, Stajich JM, et al. Missense mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem myopathy. Hum Mol Genet. May 1998;7(5):807-12. Epub 1998/05/23.
49. Demir E, Sabatelli P, Allamand V, Ferreiro A, Moghadaszadeh B, Makrelouf M, et al. Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. Am J Hum Genet. Jun 2002;70(6):1446-58. Epub 2002/05/07.
50. Sykes B, Ogilvie D, Wordsworth P, Anderson, Jones N. Osteogenesis imperfecta is linked to both type I collagen structural genes. Lancet. Jul 12 1986;2(8498):69-72. Epub 1986/07/12.
51. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. May 5 2011;473(7345):43-9. Epub 2011/03/29.
52. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. Sep 15 2010;26(18):2336-7. Epub 2010/07/17.

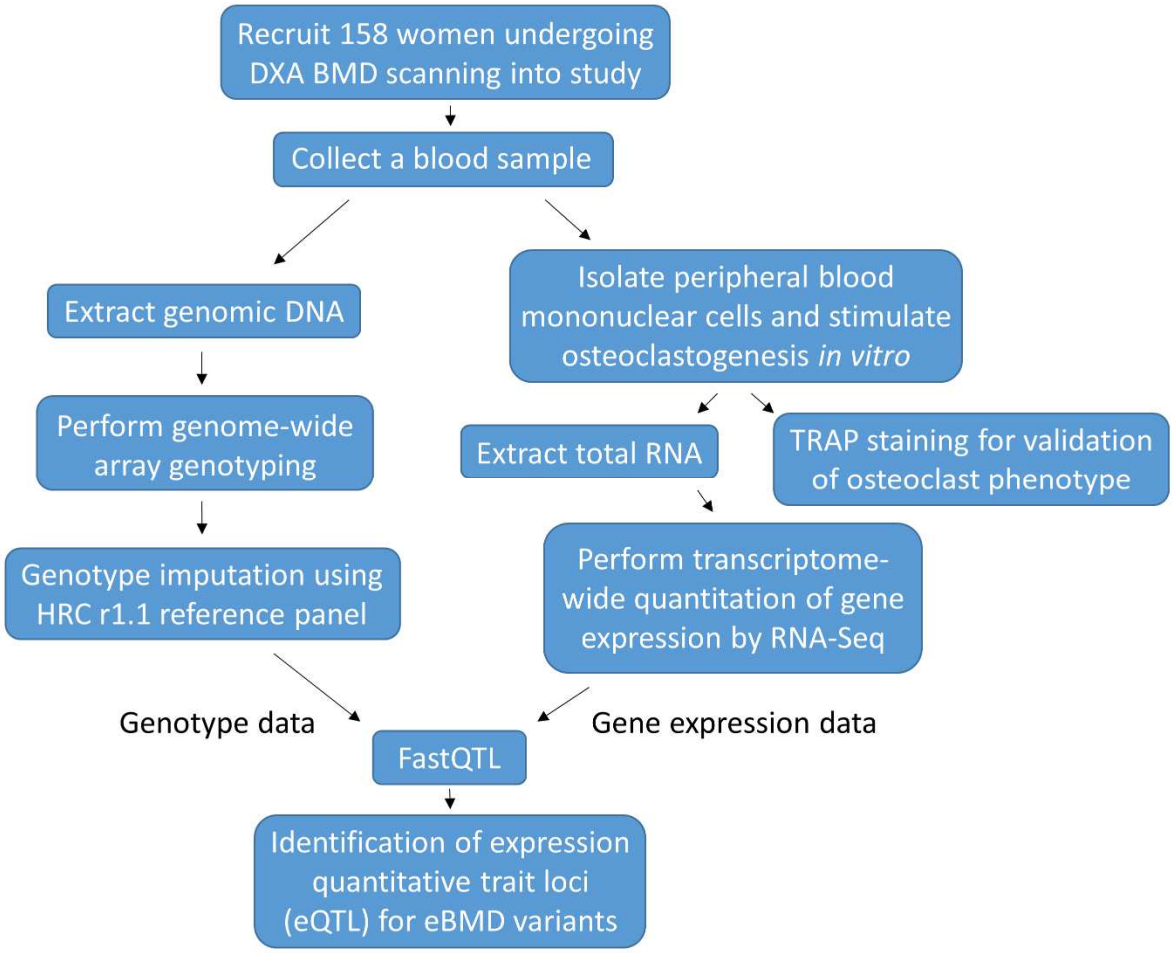

Fig. 1. Flow diagram depicting the steps used to generate the genotype and gene expression data used in the eQTL analysis.

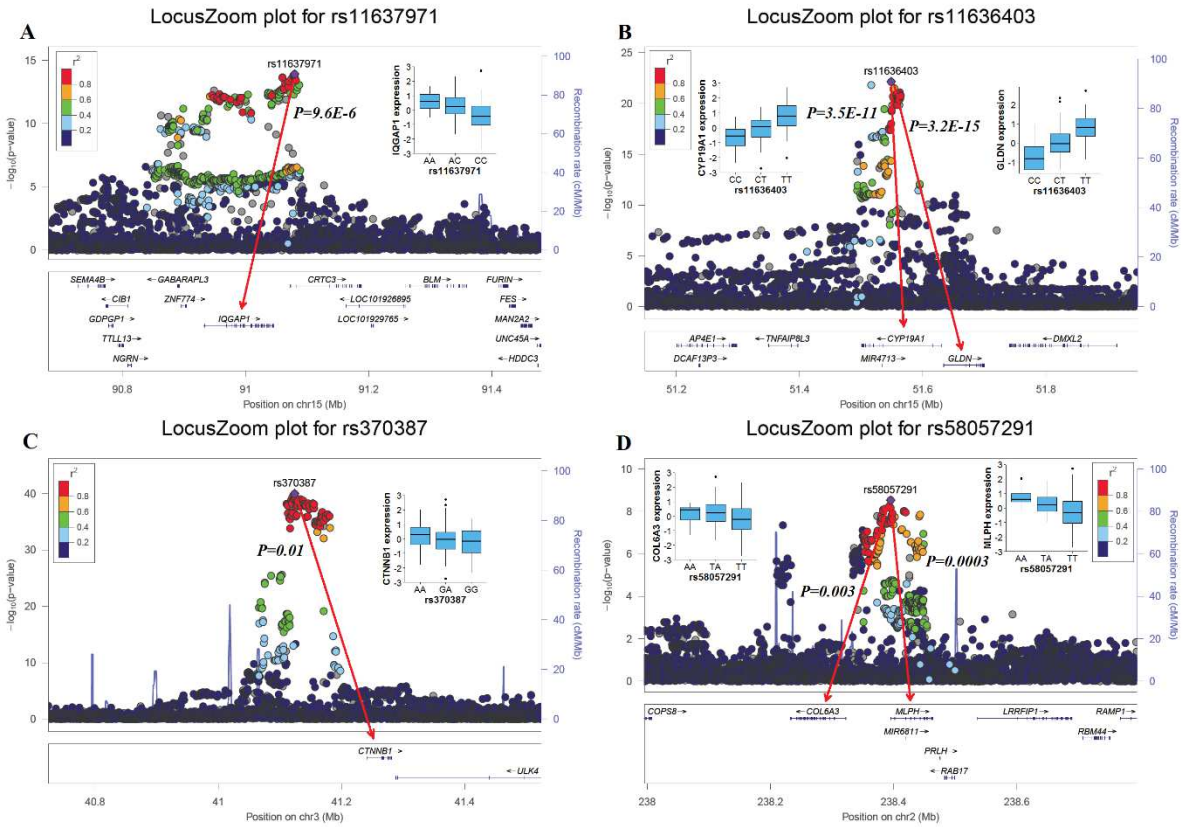

Fig. 2. Regional association plots generated using the eBMD association results from Kemp et al.[12] and combined with the eQTL results for rs11637971 (*IQGAP1*) (A), rs11636403 (*CYP19A1* and *GLDN*) (B), rs370387 (*CTNNB1*) (C) and rs58057291 (*COL6A3* and *MLPH*) (D). Genetic variants within 400kb of the lead variant are depicted (x axis) along with their eBMD P value (–log10). Variants are colour coded according to their LD (r^2) with the lead variant (1000GP Nov 2014 EUR population). The recombination rate (blue line) and position of genes, their exons and direction of transcription is also indicated.[52] Significant association between the lead variants and expression of nearby genes is indicated by the red arrows and P values, with the allelic effects on gene expression presented in the box-and-whisker plots.

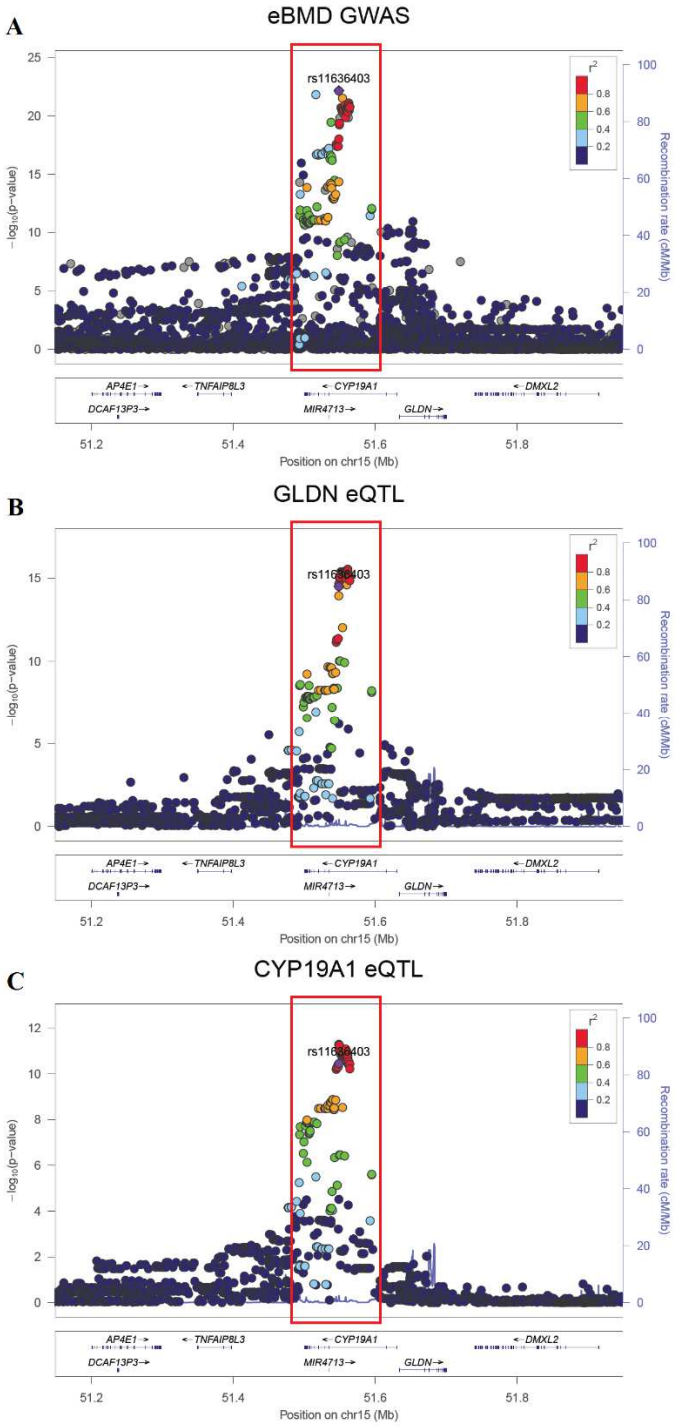

Fig. 3. Regional association plots generated for (A) eBMD GWAS association results from Kemp et al.[12], (B) eQTL association results from this study for the *GLDN* gene and (C) eQTL association results from this study for the *CYP19A1* gene. The location of an association signal for eBMD and expression of both the *GLDN* and *CYP19A1* genes is indicated by the red box.

Table 1. Demographics of the study cohort

Demographic variable	Mean (SD)
Subjects (n)	158
Age (years)	57.2 (9.7)
Age (range)	30.5 – 69.8
Height (cm)	162.7 (6.2)
Weight (kg)	67.7 (13.0)
BMI (kg/m^2)	25.6 (5.1)
Spine BMD (g/cm^2)	0.94 (0.15)
Spine BMD T-score	-0.96 (1.34)
Total hip BMD (g/cm^2)	0.84 (0.12)
Total hip BMD T-score	-0.80 (0.99)
Femoral neck BMD (g/cm^2)	0.72 (0.11)
Femoral neck BMD T-score	-1.17 (0.96)

SD = standard deviation; BMD = bone mineral density.

Table 2. eBMD variants significantly associated with expression of nearby genes in the osteoclast-like cells

Variant	Location	EA	OA	EAF	eBMD β	Gene	Expression	Distance to TSS	P	Slope*
rs4589135	chr1:26715223	C	T	0.42	-0.02	*PIGV*	4.22 ± 0.57	-72249	3.40E-05	-0.447
rs4360494	chr1:37990219	C	G	0.57	0.02	*SF3A3*	7.45 ± 1.01	-703	3.76E-04	0.408
rs3790608	chr1:112512401	A	G	0.13	0.04	*ST7L*	1.35 ± 0.36	-108425	9.94E-07	0.850
rs6680737	chr1:118979008	A	G	0.45	0.03	*WARS2*	1.88 ± 0.33	-161664	5.88E-04	-0.379
rs6680737	chr1:118979008	A	G	0.45	0.03	*RP11-418J17.1*	0.38 ± 0.11	-161388	3.97E-03	-0.348
rs7516171	chr1:210293587	T	C	0.20	0.03	*IRF6*	0.65 ± 0.38	487411	4.96E-03	-0.409
rs6547870	chr2:28714735	C	G	0.60	-0.02	*TRMT61B*	4.67 ± 0.77	-155567	1.91E-06	-0.565
rs6761129	chr2:43508574	T	C	0.10	0.04	*THADA*	5.16 ± 1.74	-87473	1.86E-03	-0.598
rs7578166	chr2:71402911	C	A	0.64	0.02	*DYSF*	4.95 ± 3.71	-50811	2.61E-08	-0.680
rs58057291	chr2:237486836	A	T	0.24	-0.02	*MLPH*	0.95 ± 0.87	1408	3.38E-04	0.450
rs58057291	chr2:237486836	A	T	0.24	-0.02	*COL6A3*	1.16 ± 0.86	72460	3.35E-03	0.380
rs370387	chr3:41082493	A	G	0.56	0.04	*CTNNB1*	38.1 ± 5.32	-112344	1.48E-02	0.298
rs6829296	chr4:1006350	G	C	0.39	-0.02	*DGKQ*	4.44 ± 1.72	19454	5.81E-04	-0.358
rs6870556	chr5:31134730	A	G	0.58	-0.02	*C5orf22*	7.72 ± 1.29	-397536	5.78E-03	-0.338
rs4959677	chr6:2500586	C	G	0.51	-0.03	*SERPINB1*	36.3 ± 6.40	-341421	2.49E-04	0.441
rs757980	chr7:28685919	A	G	0.73	-0.04	*CHN2*	0.65 ± 0.32	-436354	1.46E-02	0.308
rs757980	chr7:28685919	A	G	0.73	-0.04	*CREB5*	0.41 ± 0.39	386596	3.57E-02	0.266
rs283324	chr8:69268016	A	G	0.25	-0.02	*GS1-44D20.1*	20.3 ± 6.84	138418	3.69E-02	0.281
rs72767980	chr9:127558278	T	C	0.15	0.03	*FAM129B*	99.2 ± 19.33	-20712	7.92E-04	-0.523
rs72767980	chr9:127558278	T	C	0.15	0.03	*ZNF79*	3.56 ± 1.03	133904	9.62E-04	-0.542
rs174574	chr11:61832870	C	A	0.62	-0.02	*FADS2*	21.1 ± 6.96	39890	1.17E-04	-0.438
rs7959604	chr12:1527963	G	C	0.06	-0.06	*RPS4XP14*	1.27 ± 0.58	-230049	1.01E-03	-0.807
rs11636403	chr15:51256547	T	C	0.49	0.03	*GLDN*	0.05 ± 0.07	-85082	3.17E-15	0.756
rs11636403	chr15:51256547	T	C	0.49	0.03	*CYP19A1*	5.61 ± 5.31	-82064	3.53E-11	0.673
rs11637971	chr15:90536631	C	A	0.70	0.03	*IQGAP1*	66.3 ± 9.26	148413	9.60E-06	-0.540
rs2301522	chr16:309953	G	A	0.71	-0.03	*TMEM8A*	28.4 ± 4.18	-77161	1.59E-04	-0.514
rs62028332	chr16:50991557	A	G	0.13	0.04	*SNX20*	12.7 ± 4.09	310203	7.52E-04	0.573
rs4888151	chr16:81525404	C	A	0.71	-0.03	*RP11-303E16.2*	4.06 ± 1.29	494634	1.83E-03	-0.381
rs2696264	chr17:50255988	A	G	0.18	0.03	*CACNA1G*	0.04 ± 0.04	-305080	1.43E-03	0.455

rs2696264	chr17:50255988	A	G	0.18	0.03	*EPN3*	0.18 ± 0.53	-276555	1.54E-03	-0.438
rs2696264	chr17:50255988	A	G	0.18	0.03	*ACSF2*	7.84 ± 1.97	-170170	1.87E-03	0.458
rs2696264	chr17:50255988	A	G	0.18	0.03	*PPP1R9B*	32.4 ± 4.36	105357	2.03E-03	0.435

EA = effect allele; OA = other allele; EAF = effect allele frequency; eBMD = estimated BMD; TSS = transcription start site; variant locations derived from dbSNP build 150 (GRCh38/hg38); eBMD β values obtained from Kemp et al.;[12] expression levels are stated as mean RPKM ± standard deviation. eQTL associations are significant using multiple testing corrected FDR of 5%.

*Slope of the linear regression between the variant and gene expression level (effect size).