

# An RNA-seq Protocol to Identify mRNA Expression Changes in Mouse Diaphyseal Bone: Applications in Mice with Bone Property Altering *Lrp5* Mutations

Ugur M Ayturk,<sup>1,2</sup> Christina M Jacobsen,<sup>1,3</sup> Danos C Christodoulou,<sup>2</sup> Joshua Gorham,<sup>2</sup> Jonathan G Seidman,<sup>2</sup> Christine E Seidman,<sup>2,5</sup> Alexander G Robling,<sup>4</sup> and Matthew L Warman<sup>1,2,5</sup>

<sup>1</sup>Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA, USA

<sup>2</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA

<sup>3</sup>Divisions of Endocrinology and Genetics, Boston Children's Hospital, Boston, MA, USA

<sup>4</sup>Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN, USA

<sup>5</sup>Howard Hughes Medical Institute, Boston, MA, USA

## ABSTRACT

Loss-of-function and certain missense mutations in the Wnt coreceptor low-density lipoprotein receptor-related protein 5 (LRP5) significantly decrease or increase bone mass, respectively. These human skeletal phenotypes have been recapitulated in mice harboring *Lrp5* knockout and knock-in mutations. We hypothesized that measuring mRNA expression in diaphyseal bone from mice with *Lrp5* wild-type (*Lrp5*<sup>+/+</sup>), knockout (*Lrp5*<sup>-/-</sup>), and high bone mass (HBM)-causing (*Lrp5*<sup>p.A214V/+</sup>) knock-in alleles could identify genes and pathways that regulate or are regulated by LRP5 activity. We performed RNA-seq on pairs of tibial diaphyseal bones from four 16-week-old mice with each of the aforementioned genotypes. We then evaluated different methods for controlling for contaminating nonskeletal tissue (ie, blood, bone marrow, and skeletal muscle) in our data. These methods included predigestion of diaphyseal bone with collagenase and separate transcriptional profiling of blood, skeletal muscle, and bone marrow. We found that collagenase digestion reduced contamination, but also altered gene expression in the remaining cells. In contrast, in silico filtering of the diaphyseal bone RNA-seq data for highly expressed blood, skeletal muscle, and bone marrow transcripts significantly increased the correlation between RNA-seq data from an animal's right and left tibias and from animals with the same *Lrp5* genotype. We conclude that reliable and reproducible RNA-seq data can be obtained from mouse diaphyseal bone and that lack of LRP5 has a more pronounced effect on gene expression than the HBM-causing LRP5 missense mutation. We identified 84 differentially expressed protein-coding transcripts between LRP5 "sufficient" (ie, *Lrp5*<sup>+/+</sup> and *Lrp5*<sup>p.A214V/+</sup>) and "insufficient" (*Lrp5*<sup>-/-</sup>) diaphyseal bone, and far fewer differentially expressed genes between *Lrp5*<sup>p.A214V/+</sup> and *Lrp5*<sup>+/+</sup> diaphyseal bone. © 2013 American Society for Bone and Mineral Research.

**KEY WORDS:** MOLECULAR PATHWAY DEVELOPMENT; GENETIC ANIMAL MODELS; WNT/ $\beta$ -CATENIN/LRPS; CELLS OF BONE; DISEASES AND DISORDERS OF/RELATED TO BONE; STATISTICAL METHODS

## Introduction

Low-density lipoprotein receptor-related protein 5 (LRP5) functions as an important coreceptor in the canonical Wnt signaling pathway. *LRP5* loss-of-function and certain missense mutations significantly reduce or increase bone mass, respectively, in humans.<sup>(1-3)</sup> These skeletal phenotypes have been recapitulated in mice harboring knockout and knock-in *Lrp5* mutations.<sup>(4,5)</sup> Studies in these mouse models indicate that LRP5 signaling is involved in the ability of bone to respond to changes in mechanical load.<sup>(6-8)</sup>

Two endogenous inhibitors of canonical Wnt signaling, Sclerostin (SOST) and Dickkopf homolog 1 (DKK1), likely exert their inhibitory effects by binding to LRP5 and/or LRP6.<sup>(9-12)</sup> Missense mutations in *LRP5* that cause increased bone mass appear to do so by impairing the binding between LRP5 and its inhibitors, without impairing the binding between LRP5 and its agonists (ie, Wnt ligands).<sup>(13-15)</sup> The pathways that function upstream and downstream of LRP5-mediated signaling are incompletely understood. We hypothesized that measuring mRNA expression in cortical bone from mice with *Lrp5* wild-type (*Lrp5*<sup>+/+</sup>), knockout (*Lrp5*<sup>-/-</sup>), and high bone mass (HBM)-causing

Received in original form February 21, 2013; revised form March 18, 2013; accepted March 22, 2013. Accepted manuscript online April 2, 2013.

Address correspondence to: Matthew L Warman, MD, Harvard Medical School and Boston Children's Hospital, 320 Longwood Ave, Enders 250, Boston, MA 02115, USA. E-mail: matthew.warman@childrens.harvard.edu

Additional Supporting Information may be found in the online version of this article.

Journal of Bone and Mineral Research, Vol. 28, No. 10, October 2013, pp 2081-2093

DOI: 10.1002/jbmr.1946

© 2013 American Society for Bone and Mineral Research

(*Lrp5<sup>p.A214V/+</sup>*) alleles could identify genes and pathways that regulate or are regulated by LRP5 activity.

Massively parallel sequencing (also referred to as next generation sequencing) of mRNA (RNA-seq) has quickly become a powerful alternative to measuring mRNA expression by microarray and quantitative RT-PCR.<sup>(16,17)</sup> The advantages of RNA-seq are that it can cover a broad range of transcript abundance and can identify mRNA transcripts at a single nucleotide resolution.<sup>(18,19)</sup> To date, the application of RNA-seq in the study of mineralized tissues is limited.<sup>(20–22)</sup> Therefore, despite its potential, RNA-seq has been underutilized to study skeletal biology in a small animal model, such as the mouse. Herein, we describe methods for extracting mRNA from mouse diaphyseal bone, constructing cDNA libraries for use in massively parallel sequencing, and analyzing resultant sequence data in order to identify differentially expressed transcripts among *Lrp5<sup>+/+</sup>*, *Lrp5<sup>-/-</sup>*, and *Lrp5<sup>p.A214V/+</sup>* mice. Additionally, because it is not possible to completely remove nonskeletal tissue (e.g., skeletal muscle, bone marrow, blood) from a freshly excised long bone, we also provide an analytic strategy for robustly identifying and removing nonskeletal “contaminating” transcripts from RNA-seq data. These methods should aid other investigators who use mouse models to study skeletal biology and disease.

## Subjects and Methods

### mRNA library preparation

The Boston Children’s Hospital Institutional Animal Care and Use Committee approved these studies. The mice used in this study have been described.<sup>(4,5)</sup> Male, 16-week-old *Lrp5<sup>+/+</sup>* ( $n = 4$ ), *Lrp5<sup>-/-</sup>* ( $n = 4$ ), and *Lrp5<sup>p.A214V/+</sup>* ( $n = 4$ ) mice were used. One animal at a time was euthanized by a 1-minute exposure to CO<sub>2</sub>. Cervical dislocation was then performed to assure death. Within 10 minutes of the animal’s death, the pair of tibiae was prepared for RNA extraction: Muscle, tendon, and ligament were removed with a scalpel. The distal and proximal epiphyses were excised and the diaphyseal bone marrow was removed by centrifugation at  $>15,000g$  for 1 minute at room temperature. The resultant hollow bone shafts were individually flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until further use.

RNA was recovered from each tibia sample by pulverizing the bone inside 2-mL RNase-free microtubes with metal beads and 1 mL TRIzol (Life Technologies, Grand Island, NY, USA) using a benchtop tissue homogenizer (FastPrep24; MP Biomedicals, Solon, OH, USA). Each tibia was subjected to four cycles of 30-second pulverization in the tissue homogenizer. Total RNA was then separated from bone mineral, DNA, and protein by phenol-chloroform extraction. The RNA containing fraction (400  $\mu\text{L}$  per sample) was then purified (Purelink RNA Mini Kit; Life Technologies, Grand Island, NY, USA) and treated with DNase I (RNase-Free DNase Set; Qiagen, Valencia, CA, USA) for 15 minutes. The amount of recovered total RNA (eluted in 50  $\mu\text{L}$  RNase-free water) was measured with a spectrophotometer (Nanodrop 1000; Thermo Scientific, Wilmington, DE, USA) and the quality of the total RNA was assessed with a Bioanalyzer (Model 2100; Agilent Technologies, Santa Clara, CA, USA). Each

fresh frozen bone RNA sample had an RNA integrity number (RIN)  $> 6$ , indicating they were of sufficient quality to prepare sequencing libraries.<sup>(23)</sup>

An mRNA library for each individual tibia was prepared using the TruSeq RNA Sample Preparation Kit, v2 (Illumina, San Diego, CA, USA). Briefly, mRNA was enriched from the total RNA (starting material  $>250$  ng) by performing polyA<sup>+</sup> purification twice with magnetic beads. The recovered mRNA was then chemically fragmented to yield segments from 120 to 200 bp in length, reverse transcribed using random hexamers, and ligated to bar-coded adapters following the manufacturer’s recommendations. The resultant cDNA fragments were amplified for 15 PCR cycles. Finally, the cDNA libraries were washed with AMPure XP beads to remove primer-dimers (Agencourt AMPure XP; Beckman Coulter, Brea, CA, USA) and 1- $\mu\text{L}$  aliquots were run on a 4–20% TBE gel (Life Technologies, Grand Island, NY) for verification. Equal amounts of DNA ( $\sim 15$  ng per sample) from eight to 10 separately bar-coded cDNA libraries (each generated from a single tibia sample) were pooled and then sent for 50-bp paired-end sequencing on one lane of an Illumina HiSeq 2000 to generate a minimum of 10 million reads/library. For the quantities such as buffer volumes or time and temperature settings not specified here, the exact recommendations of the manufacturers were followed at every step of the protocol.

### Raw read processing

Paired reads from each library were mapped to the mouse genome (mm9) using RNA-Seq unified mapper (RUM).<sup>(24)</sup> The expression level of each gene was quantified by counting the number of reads that uniquely aligned to the respective set of exonic sequences. Data were then normalized with respect to the total number of mapped reads in each library and the dispersion in distribution of the reads to the genome.<sup>(25)</sup>

### Assessing the reproducibility and reliability of RNA-seq data

We measured the reproducibility of the cDNA sequencing and data analysis platforms by splitting a cDNA library in half, sequencing each half, and determining Pearson’s correlation coefficient ( $R^2$ ) in a comparison of the number of mapped reads to each gene. We measured the reproducibility in preparing bone samples, extracting mRNA, and generating cDNA sequencing libraries by making a separate library for each animal’s right and left tibia and determining  $R^2$  for the resultant RNA-seq data. Due to the logarithmic distribution of reads in all libraries, a small fluctuation in a gene with a very high level of expression (e.g., left tibia versus right tibia fold change  $<1.5$ ) could be much more influential than large fluctuations in a moderately expressed gene (e.g., left tibia versus right tibia fold change  $>5$ ) on  $R^2$ . In order to remove this dispersion bias, we recalculated  $R^2$  using a trimmed mean method; ie, the top and bottom 1% of genes were excluded in the assessment of all correlation values.

We also calculated reads per kilobase of exon model per million mapped reads (RPKM)<sup>(18)</sup> values for each gene in order to determine a detectability threshold: The number of reads mapped to the coding region of each gene was normalized by the total number reads mapped within the respective library and

the length of the associated coding region. The elimination of potential biases due to gene length allowed us to compare the expression levels of different genes to each other, and rank the transcript abundances within a single library. Subsequently, we used RPKM = 5 as a minimum limit for detectability.<sup>(18)</sup> This meant that for a library containing 10 million mapped 50-bp paired-end reads we limited our analyses to genes averaging five or more reads/base of cDNA. However, we did not use RPKM values in our statistical comparisons (ie, when calculating fold-change and *p* values) between different sample groups (see the last two paragraphs in the Methods section for our differential expression analysis protocol).

### Identification and filtering of transcripts representing RNA contamination from nonmineralized tissue

We used three strategies to identify transcripts that likely represented RNA contamination from sources other than bone.

#### Strategy 1

Using the same methods for RNA extraction and cDNA library preparation, we generated separate bar-coded libraries from blood (*n* = 6), skeletal muscle (*n* = 6), bone marrow (*n* = 6), and tibial cortical bone (*n* = 6) from three 12-week old male mice. The source material for all libraries were extracted from the right and left hind limbs of each mouse, except for blood, which was drawn from the heart and split in two. Following RNA-seq we calculated differential expression in order to identify the transcripts with at least twofold greater abundance in blood, muscle, or bone marrow compared to cortical bone (*p* < 0.05).

#### Strategy 2

We removed the femoral bones from these same mice, excised the ends, and centrifuged the marrow from the diaphyseal bone. We then cut the diaphyseal bones into smaller pieces and digested the fragments in a collagenase solution (Collagenase I&I; Worthington Biochemical Corp, Lakewood, NJ, USA) for 45 minutes (3 × 15 minutes at 37°C) in order to enrich for native bone cells.<sup>(26,27)</sup> We then prepared a separate RNA-seq library for each bone. To control for gene expression changes that could result from collagenase digestion and/or from ex vivo tissue culture, we also prepared individual RNA-seq libraries from six diaphyseal femur samples (extracted from 16-week-old *Lrp5*<sup>+/+</sup> male mice) that had been immediately frozen rather than collagenase-digested. Following RNA-seq, we identified the transcripts whose abundances were twofold lower in the collagenase-digested samples compared to the fresh-frozen samples (*p* < 0.05).

#### Strategy 3

We compared RNA-seq data from a single animal's right and left tibia and calculated *R*<sup>2</sup> values. We then employed an algorithm that would calculate the individual effect of removing each individual species of RNA transcript on the *R*<sup>2</sup>, and identified the transcripts whose removal had the greatest effect.

In comparing these three strategies, we noted that most identified transcripts with the largest effect on *R*<sup>2</sup> using strategy 3 were also identified using strategies 1 and 2. Furthermore, transcripts identified as representing the intersection of strategies 1 and 2 accounted for most of the reduction in *R*<sup>2</sup> between paired bones from the same animal. Therefore, we filtered the diaphyseal bone RNA-seq data for transcripts representing the intersection of strategies 1 and 2 before we compared expression across mice with the different *Lrp5* genotypes.

### Identification of differentially expressed genes

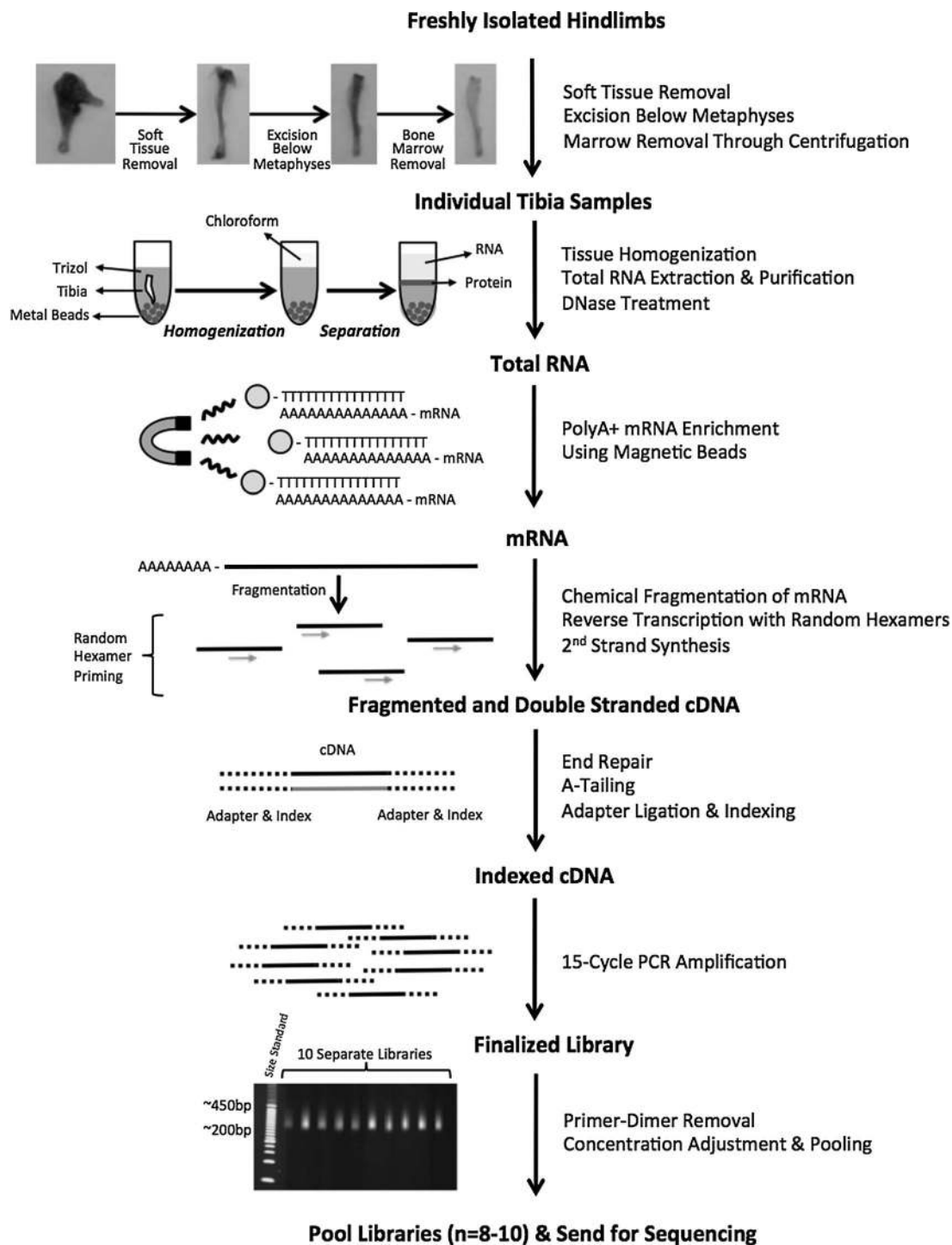
Comparisons of RNA-seq data from different sample groups, including filtered bone libraries with different *Lrp5* genotypes, were made with a Fisher's exact test, and the resultant *p* values were corrected for multiple hypothesis testing (false discovery rate < 0.05).<sup>(28)</sup> All calculations were made in R, using the Rsamtools,<sup>(29)</sup> GenomicFeatures,<sup>(30)</sup> and edgeR<sup>(25)</sup> subroutine packages.

After correcting for multiple hypothesis testing, we performed a series of leave-one-out cross-validation tests: In each comparison between RNA-seq data representing two genotypes (e.g., *Lrp5*<sup>+/+</sup> mice *n* = 8 versus *Lrp5*<sup>-/-</sup> mice *n* = 8), one RNA-seq sample was removed and the test was repeated (*n* = 7 versus *n* = 8) in order to ensure that the excluded library was not solely responsible for the significance of the outcome. This procedure was repeated until the effect of every sample was elucidated. Then, the same protocol was repeated by removing paired tibia libraries from the same mouse at once (*n* = 6 versus *n* = 8), in order to prevent any bias due to a single animal. All genes that lost their significance at any point (*p* > 0.05) were removed from the list of differentially expressed genes.

## Results

### Reproducibility of RNA-seq data generation

Our overall strategy for generating RNA-seq libraries is depicted in Fig. 1. On average we obtained 19 million reads/library, with 18% of reads representing PCR duplicates, and 98% of reads successfully mapping (84% uniquely) to the annotated genome (Fig. 2A). The distribution of reads within individual genes was similar across libraries (Fig. 2B). When we split an individual RNA-seq library in half and sequenced each half separately we obtained *R*<sup>2</sup> values > 0.99. When we prepared duplicate separately bar-coded RNA-seq libraries from paired sources within the same animal we obtained *R*<sup>2</sup> values > 0.97 for bone marrow and *R*<sup>2</sup> values ranging between 0.83 and 0.99 for unfiltered paired cortical bone samples (Fig. 2C, D). We assumed that differences in the extent of residual skeletal muscle, blood, and bone marrow within paired bone samples accounted for their reduced *R*<sup>2</sup> values. Therefore, we tested this hypothesis by (1) identifying transcripts that are more highly expressed in skeletal muscle, blood, and bone marrow compared to cortical bone; (2) identifying transcripts that are lower in collagenase-digested bone compared to fresh-frozen bone; and (3) identifying transcripts whose in silico removal has a substantial effect on *R*<sup>2</sup> values when the right and left tibias from the animal with the lowest correlation (*R*<sup>2</sup> = 0.83) are compared (Fig. 2E).

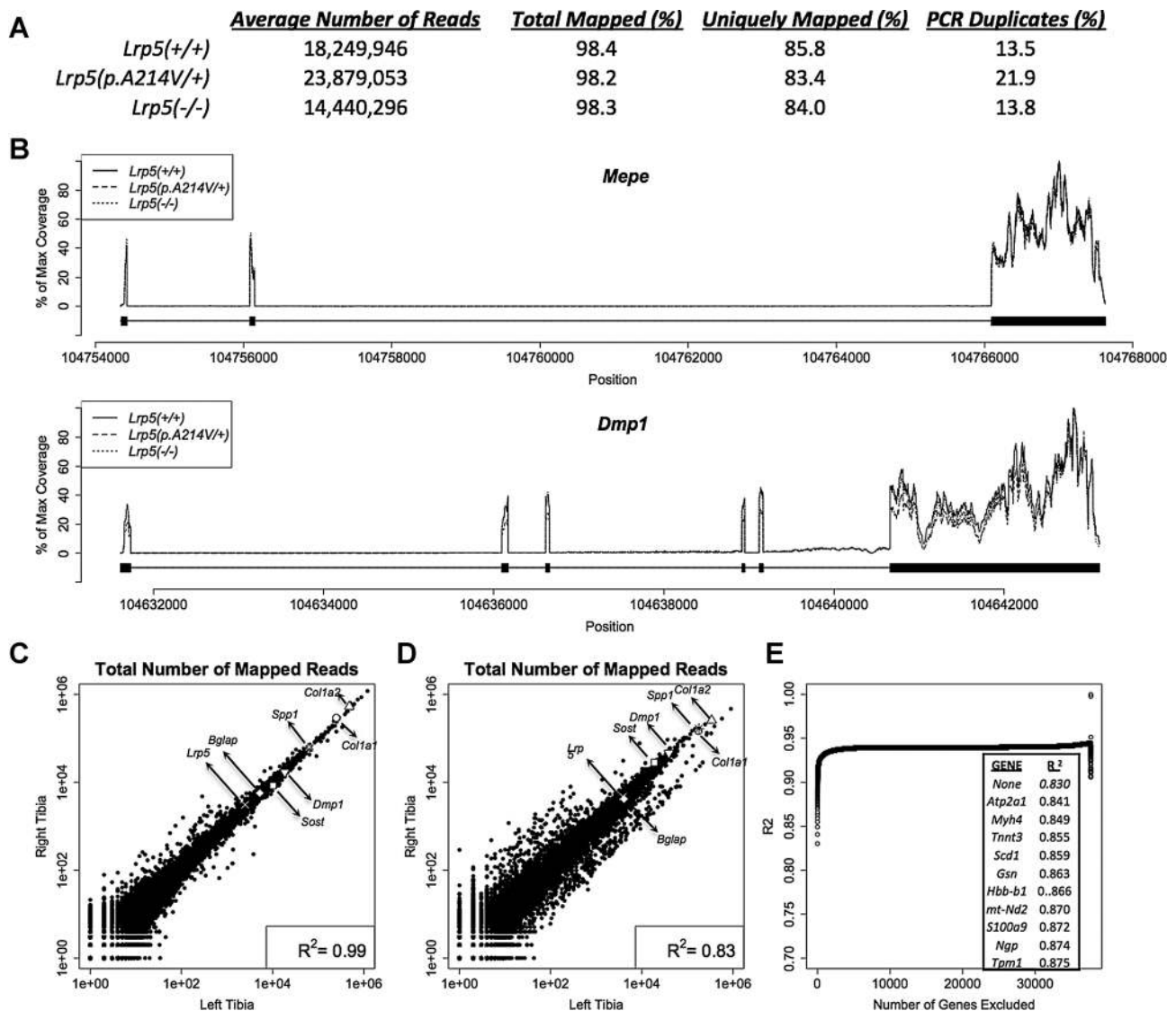


**Fig. 1.** Strategy for generating tibia diaphyseal bone cDNA libraries for RNA-seq. Both tibias were prepared within 10 minutes of euthanasia. Skeletal muscle and the metaphyseal/epiphyseal regions were removed with a scalpel and the marrow was removed by centrifugation. Bone shafts were flash frozen in liquid nitrogen and subjected to pulverization in Trizol for RNA extraction. mRNA was enriched using oligo-dT coated magnetic beads and then chemically fragmented into 120- to 200-nucleotide-long segments prior to random hexamer primed reverse transcription. Bar-coded sequencing adapters were ligated to the cDNA fragments and the indexed cDNA was subjected to 15 cycles of PCR amplification. Each library was analyzed by gel electrophoresis to confirm that the correct fragment size range had been achieved (note the >200-bp fragments include the sequencing adaptors). Libraries constructed from 10 different mouse tibias are depicted here. Eight to 10 separately bar-coded libraries were pooled and loaded onto a single lane of an Illumina HiSeq2000 machine and 50-bp paired-end sequencing was performed.

RNA-seq data obtained from blood, skeletal muscle, bone marrow, and diaphyseal bone

We observed large differences in the transcriptomes between tissues. Using an RPKM value of 5 as the lower boundary for

reliable detection of gene expression, ~5000 genes exceeded this threshold for blood and skeletal muscle cDNA libraries, whereas ~9000 genes exceeded this threshold for bone marrow and diaphyseal bone cDNA libraries (Supplementary Table 1)



**Fig. 2.** RNA-seq metrics and reproducibility. (A) Table containing average values for tibia diaphyseal bone RNA-seq data from mice with the three different *Lrp5* genotypes ( $n = 8$  libraries/genotype). Columns indicate the mouse genotype from which the RNA-seq data were obtained, the number of paired-end reads/library, the percentage of reads that mapped to the mouse genome, the percentage of reads that mapped to unique regions within the mouse genome, and the percentage of reads that likely represent PCR duplicates created during library preparation. (B) Graphs depicting the distribution of RNA-seq data across two representative genes (*Mepe* and *Dmp1*). The read depth relative to the maximum read depth is graphed in alignment to each gene's genomic DNA sequence. Thick bars below the plots indicate the positions of the exons, and the thin lines indicate the positions of the introns. Peaks overlap the exon-containing regions of each gene, consistent with this being RNA-seq data. Note the high degree of similarity with respect to RNA-seq coverage across the genes independent of the animals' *Lrp5* genotypes. (C) Scatter-plot comparing the total number of uniquely mapped reads for RNA-seq data obtained from the right and left tibias of a mouse. Each dot represents an individual gene. Higher-abundance transcripts are closer to the upper right and lower-abundance transcripts are closer to the lower left of this plot. Examples of bone-expressed genes are identified by unique symbols and arrows. The Pearson correlation coefficient ( $R^2$ ) between the right and left tibia RNA-seq data is 0.99 for this sample pair. (D) Scatter-plot comparing total number of uniquely mapped reads for RNA-seq data obtained from the right and left tibias of another mouse. Note that the Pearson correlation coefficient ( $R^2$ ) between the right and left tibia RNA-seq data is 0.83 for this sample pair, although the bone-expressed genes appear to follow the  $y = x$  line. (E) Graph depicting the effect of sequentially filtering transcripts whose removal cause the greatest increases in correlation between the right and left tibia. The inset indicates the top 10 genes whose removal from the RNA-seq data depicted in D had the greatest positive effect on  $R^2$ . Note that these genes are highly expressed in muscle and blood, suggesting the reduced correlation in the tibia pair depicted in D resulted from non-bone tissue contamination.

(Note: Supplementary Tables 1 to 8 are available online for this article; a spreadsheet editor, e.g., Microsoft Excel, would be the ideal tool for visualization and review of these files). The distribution of expressed transcripts within the libraries was least diverse in blood, where 52% of sequence reads corresponded to the four hemoglobin genes (*Hba-a1*, *Hba-a2*, *Hbb-b1*, and *Hbb-*

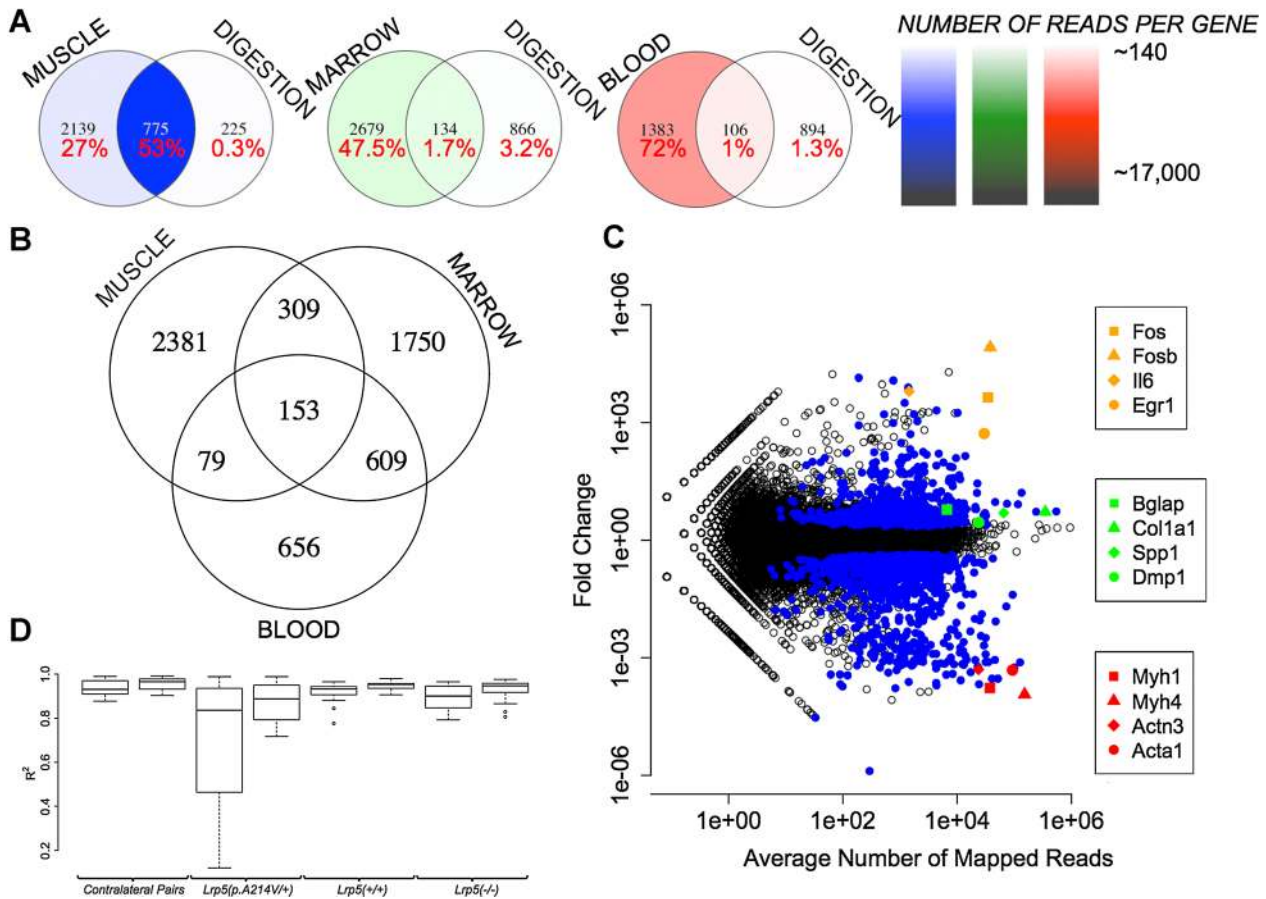
*b2*). These four genes represented only ~0.2% of the RNA-seq reads from diaphyseal bone. Similarly, highly expressed skeletal muscle transcripts *Acta1*, *Tnnc2*, and *Mylpf* together accounted for ~5% of mapped sequence reads in the skeletal muscle libraries, and only ~0.3% of reads in the diaphyseal bone libraries; these data suggest that less than 10% of RNA-seq data

in diaphyseal bone represent tissue contamination. Conversely, transcripts that are highly expressed by osteoblasts and osteocytes, such as *Col1a1*, *Col1a2*, and *Spp1*, represented ~1.4%, ~2.7%, and ~0.9% of the diaphyseal bone RNA-seq data, respectively. When we compared transcript abundance between the libraries, we identified ~6000 genes that were expressed at least twice as abundantly in blood, skeletal muscle, and/or bone marrow compared to diaphyseal bone (Supplementary Table 1). Of these ~6000 genes, 2381 genes were only in the skeletal muscle libraries and 609 genes were in both the blood and bone marrow libraries (Fig. 3A,B). Because we estimated that less than 10% of diaphyseal bone RNA-seq data likely results from tissue contamination, we assumed that filtering all transcripts with

twofold higher expression in the contaminating tissues compared to bone would be too stringent. This is the case, because removing all reads representing these ~6000 genes reduced the diaphyseal bone RNA-seq data by 43%.

### Enzymatic removal of contaminating tissue from diaphyseal bone

There is a tradeoff with regard to processing bone specimens quickly in order to preserve RNA integrity and dissecting the specimens thoroughly in order to remove contaminating tissues, such as muscle, blood, and bone marrow. One approach for removing contaminating tissues relies on ex vivo tissue culture in medium containing collagenase.<sup>(26,27)</sup> We compared RNA-seq



**Fig. 3.** Identification of contaminating transcripts in diaphyseal bone RNA-seq data. (A) Venn diagrams indicating the intersection of genes that have twofold or greater expression ( $p < 0.05$ ) in skeletal muscle, bone marrow, or blood compared to tibia diaphyseal bone and that have demonstrated a significant reduction in transcript abundance in collagenase-digested diaphyseal bone. The percentages of tissue-specific RNA-seq data accounted for by these genes are also indicated. Color-coding represents in heat-map format (see scales on right) the average number of mapped reads/gene in each sector. This is determined by dividing the total number of reads that mapped to genes in this sector by the number of genes in this sector. For example, 2914 (2139 + 775) genes are expressed at least twice as abundantly in muscle compared to bone and 1000 (775 + 225) genes had significant reductions in transcript abundance following collagenase digestion. The intersection of these two data sets comprises 775 genes. These 775 genes account for 53% of all mapped reads in the skeletal muscle RNA-seq data. The average gene in this group of 775 has ~17,000 reads (~13,000,000/775) mapping to it in the muscle RNA-seq libraries. (B) Venn diagram indicating the distribution of ~6000 genes that were twice as abundantly expressed in skeletal muscle, bone marrow, and/or blood compared to tibia diaphyseal bone. (C) Scatter-plot in log<sub>10</sub> scale indicating the fold changes in transcript abundance when RNA-seq data from collagenase-digested bone is compared to RNA-seq data from fresh-frozen bone. Circles represent individual genes and genes having statistically significant changes in abundance ( $p < 0.05$ ) are colored blue. Note genes that are highly expressed in skeletal muscle (red symbols) decrease significantly ( $p < 0.001$ ), whereas genes that are highly expressed in bone (green symbols) have less pronounced increases. The large increases in transcript abundance for other genes (orange symbols) likely represent collagenase-induced changes in gene expression. (D) Box-plots indicating the ranges of trimmed-mean-normalized Pearson's correlation coefficients ( $R^2$ ) between paired tibias from individual animals (contralateral pairs), and between all pairs of tibias representing mice with the same *Lrp5* genotype. Correlations before and after in silico transcript filtering of ~900 genes are shown; note that filtering increases  $R^2$  for all data comparisons ( $p < 0.01$ ).

data obtained from fresh-frozen diaphyseal bone samples to RNA-seq data obtained from samples that had been treated with collagenase. We found that the transcript abundance of ~1000 genes was at least twofold lower in collagenase-digested bone compared to fresh-frozen bone (Fig. 3C; Supplementary Table 2). Most of these genes (~900) were more highly expressed in the skeletal muscle, blood, and/or bone marrow libraries compared to the diaphyseal bone libraries (Supplementary Table 1). As would be expected when contaminating tissues are enzymatically removed, the relative abundance of bone cell (e.g., osteoblast, osteocyte, and osteoclast) transcripts increased (Fig. 3C; Supplementary Table 3). However, more than 700 genes exhibited differences in abundance, ranging from twofold to 80,000-fold, when collagenase-digested bone and fresh-frozen bone were compared. The increased abundance of these transcripts far exceeds the increase expected from simply removing contaminating tissue. For example, transcripts for osteoinductive factors, such as *Egr1*,<sup>(31,32)</sup> *Il-6*,<sup>(33)</sup> and *Cxcl2*<sup>(34)</sup> were 522-fold, 6277-fold, and 11,799-fold more abundant, respectively, and components of the transcription factor Ap-1 complex,<sup>(35,36)</sup> such as *Fos*, *Fosb*, *Jun*, and *Atf3*, were 82,070-fold, 4394-fold, 96-fold, and 1775-fold more abundant, respectively, in collagenase-digested bone compared to fresh-frozen bone (Fig. 3C; Supplementary Table 3). These data suggest that short-term ex vivo culture and/or collagenase digestion in addition to removing contaminating tissues also alters gene expression in the remaining cells.

#### Improved in silico filtering of contaminating transcripts from diaphyseal bone RNA-seq data

All the transcripts whose abundance was lower in collagenase-digested bone are expressed in skeletal muscle, blood, and bone marrow. Therefore, we used these genes to generate one set of transcripts that could be considered “contaminating” and filtered in silico from fresh-frozen diaphyseal bone RNA-seq data. We then intersected this set with another set of transcripts that represent genes whose abundances are at least twofold higher in skeletal muscle, blood, and bone marrow, compared to diaphyseal bone. This intersecting set accounted (Supplementary Tables 1 and 4) for ~53% of the entire set of mapped reads in the skeletal muscle libraries, ~2% of the mapped reads in blood libraries, ~4% of the mapped reads in bone marrow libraries, and ~9% of mapped reads in diaphyseal bone libraries, leading us to conclude that this set is highly enriched for contaminating transcripts in the diaphyseal bone libraries. Further support for this conclusion derives from the significant increases in  $R^2$  between paired tibias from individual mice, and from randomly paired tibias from mice with the same *Lrp5* genotype, when the RNA-seq reads representing this set of genes were removed from the analyses (Fig. 3D). Therefore, this intersecting set of genes (Supplementary Table 4) was computationally filtered from the diaphyseal bone RNA-seq data before data were compared between mice with different *Lrp5* genotypes.

#### Gene expression differences between *Lrp5*<sup>+/+</sup>, *Lrp5*<sup>-/-</sup>, and *Lrp5*<sup>p.A214V/+</sup> mice

We compared in silico-filtered RNA-seq data obtained from pairs of tibia diaphyseal bone mRNA representing *Lrp5*<sup>+/+</sup>, *Lrp5*<sup>-/-</sup>, and

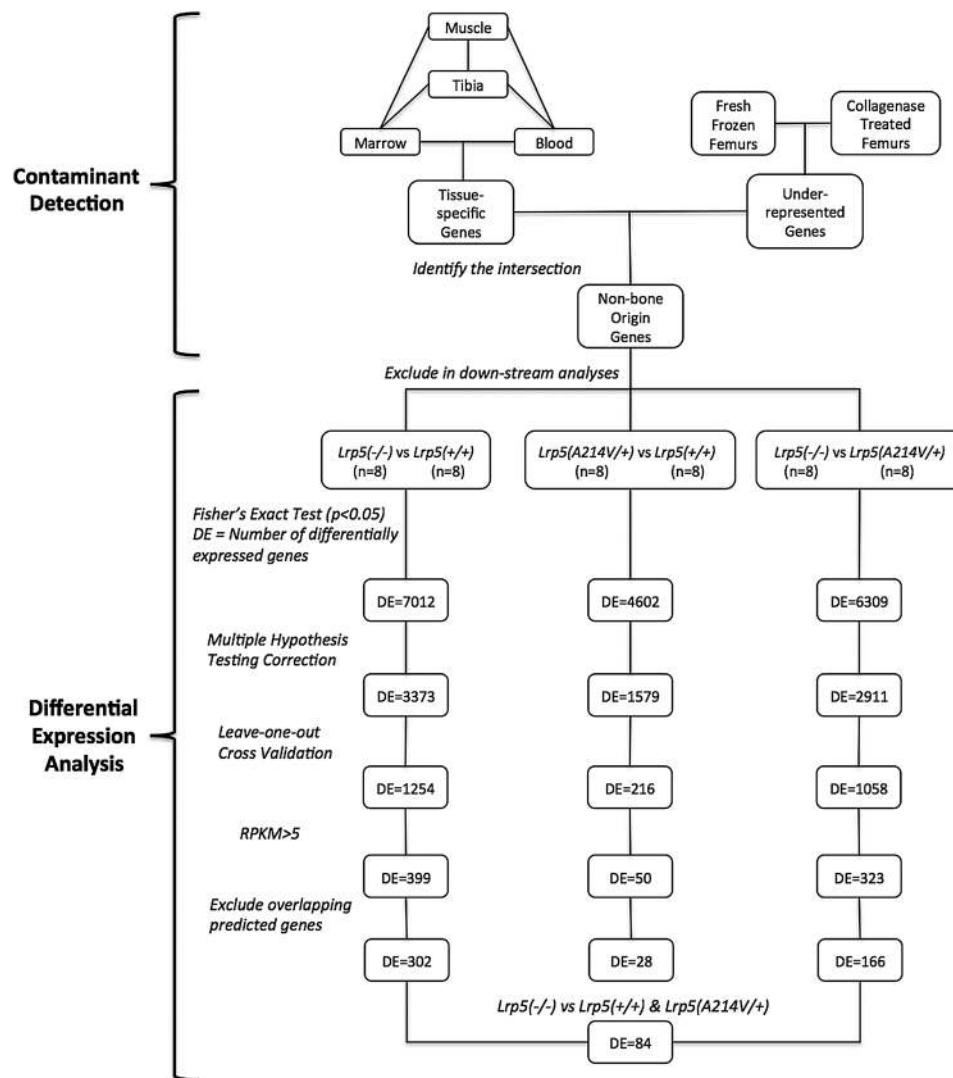
*Lrp5*<sup>p.A214V/+</sup> mice ( $n = 4$  pairs of tibias/genotype). We controlled for differences in gene expression that could result from one outlier bone sample (ie, leave-one-bone out) and one outlier mouse (ie, leave-both-bones-out). We excluded genes whose expression levels would be too low to yield meaningful differences (ie, RPKM values  $< 5$  in both genotypes being compared) and adjusted our significance thresholds to account for multiple hypothesis testing (false discovery rate  $< 0.05$ ) (Fig. 4). Using this strategy we identified 302 differentially expressed genes between *Lrp5*<sup>-/-</sup> and *Lrp5*<sup>+/+</sup> (Supplementary Table 5), 166 differentially expressed genes between *Lrp5*<sup>-/-</sup> and *Lrp5*<sup>p.A214V/+</sup> (Supplementary Table 6), and 28 differentially expressed genes between *Lrp5*<sup>p.A214V/+</sup> and *Lrp5*<sup>+/+</sup> (Supplementary Table 7) mouse diaphyseal bone.

Interestingly, there was substantial overlap between genes that were differentially expressed in *Lrp5*<sup>-/-</sup> mice compared to either *Lrp5*<sup>+/+</sup> or *Lrp5*<sup>p.A214V/+</sup> mice. This overlapping list contains 84 protein-coding genes (Table 1; Supplementary Table 8), which likely represents the consequence of “insufficient” versus “sufficient” LRP5-mediated Wnt signaling. *Lrp5* serves as an important positive control because it is knocked out in the *Lrp5*<sup>-/-</sup> mice. As expected, *Lrp5* was the most significantly reduced transcript when *Lrp5*<sup>-/-</sup> bone was compared to *Lrp5*<sup>+/+</sup> or *Lrp5*<sup>p.A214V/+</sup> bone ( $p < 1 \times 10^{-88}$ ). Also demonstrating significantly reduced expression ( $p < 0.001$ ) in diaphyseal bone from *Lrp5*<sup>-/-</sup> mice were several collagens (*Col1a1*, *Col1a2*, and *Col11a2*), osteocalcin (*Bglap* and *Bglap2*), and the zinc binding proteins (*Mt1* and *Mt2*). Two genes predicted to increase canonical Wnt signaling, *Wnt10b* (a Wnt ligand) and *Fzd4* (a Wnt coreceptor), exhibited significantly increased expression ( $p < 0.001$ ), whereas an inhibitor of the canonical Wnt pathway *Apcdd1* exhibited significantly decreased expression ( $p < 0.001$ ). Interestingly, we detected no significant changes in the expression levels of other Wnt pathway components, including *Sost*, *Dkk1*, *Lrp4*, and *Lrp6*, each of which can significantly affect bone mass in humans and/or mice.<sup>(10,37,38)</sup>

*Lrp5* also serves as an important positive control for the RNA-seq data involving *Lrp5*<sup>p.A214V/+</sup> mice. This mouse strain was intended to contain a true knock-in mutation, so the expression of the mutant allele should be equal to the expression of the wild-type allele in heterozygous mice. Because RNA-seq provides resolution at the single-nucleotide level, we were able show that there was no significant difference in expression between the *Lrp5*<sup>p.A214V</sup> and *Lrp5*<sup>+</sup> allele (83 reads versus 107 reads;  $p = 0.18$ ) in the RNA-seq data from *Lrp5*<sup>p.A214V/+</sup> mouse bone. We found only 28 genes whose expression differed significantly between *Lrp5*<sup>p.A214V/+</sup> and *Lrp5*<sup>+/+</sup> mice (Supplementary Table 7) and would not have predicted any of these genes to have had altered expression a priori.

## Discussion

Massively parallel sequencing of RNA from cells and tissues is increasingly being used to identify the repertoire of expressed mRNAs, their various splice-forms, and their changes in expression following genetic, pharmacologic, and environmental manipulation. RNA-seq has several advantages and disadvantages compared to other expression profiling strategies.



**Fig. 4.** Overview of data filtering and analysis pipeline. A list of contaminating transcripts ( $n \sim 900$ ) was compiled based upon tissue-specific expression profiling and transcript abundances in collagenase-digested versus fresh-frozen bone samples. Transcripts representing the genes on this list were in silico-filtered from the diaphyseal bone RNA-seq data prior to downstream analyses. In each statistical comparison,  $p$  values were computed with Fisher's exact test (with significance set at  $p < 0.05$ ) and corrected for multiple hypothesis testing (false discovery rate  $< 0.05$ ), grouped samples were subjected to leave-one-out cross validation, and genes expressed below the detectability threshold (RPKM  $< 5$ ) were eliminated. The numbers of differentially expressed (DE) genes remaining after each step are noted. Last, predicted genes, noncoding RNAs, and microRNAs that overlapped with the introns of highly expressed protein coding genes were eliminated in order to ensure that the reads mapped to these regions due to incomplete transcription events were not registered in differential expression calculations. A comparison between RNA-seq data from the tibia diaphyseal bones of LRP5 sufficient ( $Lrp5^{p.A214V/+}$  and  $Lrp5^{+/+}$ ) and insufficient ( $Lrp5^{-/-}$ ) mice yielded 84 differentially expressed protein-coding genes.

An advantage of RNA-seq over array-based hybridization technologies, is the former's ability to (1) detect a larger range of transcripts and transcript abundances without the need for a priori target specification, (2) identify alternative transcription start sites and splice-forms, and (3) quantify allele-specific expression at single-nucleotide resolution. For example, in our RNA-seq data from diaphyseal bone, we detected  $\sim 9000$  mRNA species, known alternative transcription start sites and splice-forms for genes such as *Dcn* and *Bglap* (data not shown), and we were able to show that  $Lrp5^{p.A214V/+}$  mice express the knock-in and wild-type alleles equivalently in their diaphyseal bone cells.

Both RNA-seq and array-based hybridization can be limited in the accuracy with which they quantify changes in gene expression. One reason for inaccuracy derives from the need to

perform a PCR amplification step during the cDNA sample preparation process. A transcript's DNA sequence (e.g., G/C content) and initial abundance can affect the ability of the PCR reaction to increase that transcript's abundance in proportion to all other transcripts in the sample. For example, high G/C content transcripts typically amplify inefficiently whereas highly abundant transcripts tend to amplify more efficiently.<sup>(39-41)</sup> Therefore, PCR amplification-induced inconsistency could lead to false-positive differences in gene expression. Expression profiling technologies that do not rely on amplification steps are available,<sup>(42,43)</sup> but they cannot yet simultaneously quantify expression levels for large numbers of transcripts. Therefore, although imperfect, our RNA-seq strategy does appear capable of quantifying mRNA expression levels with a high degree of



**Table 1.** The Top 30 Genes (Based on *p* Value) That Exhibited Significant Changes in Gene Expression Between LRP5-Insufficient (ie, *Lrp5*<sup>-/-</sup>) and LRP5-Sufficient (ie, *Lrp5*<sup>p.A214V/+</sup> and *Lrp5*<sup>+/+</sup>) Diaphyseal Bone

Gene symbol	Fold change	<i>p</i>	WT RPKM	WT rank	References
<i>Lrp5</i>	0.02	1.24E-89	26.5	1921	(1-4)
<i>Mt2</i>	0.13	4.67E-29	35.9	1340	(46,47)
<i>Zfp445</i>	6.71	7.57E-26	2.8	11878	(48)
<i>Cpz</i>	0.16	1.08E-24	11.6	4755	(49,50)
<i>Slc13a5</i>	0.17	9.76E-24	64.1	634	(51)
<i>Cyp2f2</i>	0.19	3.25E-21	22.3	2334	(52)
<i>Col1a1</i>	0.23	1.48E-17	1245.3	18	(53)
<i>Prss35</i>	0.24	1.16E-15	34.6	1403	(54)
<i>Col11a2</i>	0.25	2.14E-15	94.4	402	(55-57)
<i>Dapk2</i>	3.91	1.83E-14	11.9	4611	(58)
<i>Cfhr2</i>	3.87	4.23E-14	2.6	12272	(59)
<i>1810055G02Rik</i>	3.80	1.47E-13	15.1	3623	
<i>Gm14434</i>	3.76	1.92E-13	4.1	10141	
<i>Mt1</i>	0.28	8.59E-13	23.2	2234	(46,47)
<i>Cfd</i>	0.31	8.85E-12	103.1	353	(60)
<i>Bglap2</i>	0.31	3.94E-11	339.0	89	(61)
<i>Cd5l</i>	0.29	4.33E-11	9.7	5599	(62)
<i>Mdm4</i>	3.31	8.07E-11	5.7	8429	(63,64)
<i>Phf20l1</i>	3.21	1.83E-10	4.1	10084	(65)
<i>Il34</i>	0.32	6.31E-10	11.4	4807	(66,67)
<i>Epor</i>	0.34	2.54E-09	17.1	3148	(68)
<i>Col1a2</i>	0.35	3.18E-09	981.3	25	(53)
<i>Cyr61</i>	0.34	3.51E-09	72.1	538	(69,70)
<i>Angptl4</i>	0.34	4.33E-09	10.2	5323	(71)
<i>Npy</i>	0.36	1.63E-08	127.1	283	(72,73)
<i>Plin1</i>	0.36	1.88E-08	14.5	3774	(74)
<i>Eya4</i>	2.83	3.51E-08	2.4	12598	(75)
<i>Ighg3</i>	0.37	6.30E-08	51.6	836	(76)
<i>Bglap</i>	0.38	8.22E-08	261.7	123	(61)
<i>Bcr</i>	0.38	1.02E-07	8.7	6154	(77)

Each gene's RPKM value and abundance relative to other genes in the *Lrp5*<sup>+/+</sup> diaphyseal bone transcriptome are also indicated.

LRP5 = low-density lipoprotein receptor-related protein 5; RPKM = reads per kilobase of exon model per million mapped reads; WT = wild-type.

accuracy. For example, when we made two independent libraries from paired bones from the same animal, they exhibited high correlations in their rank order of gene expression (range, 0.96-0.97).

Additional challenges exist when RNA-seq is applied to organs and tissues in contrast to its application on homogeneous populations of cells in culture. In diaphyseal bone, osteocytes have a low cell/matrix ratio because they are surrounded by abundant extracellular matrix. Therefore, small amounts of contamination by tissues whose cell/matrix ratios are high, such as muscle and bone marrow, can significantly affect the RNA-seq data. For example, when RNA-seq data are compared between paired bones from the same animal, the transcripts whose expression levels were most discrepant between the specimens predominantly represented genes that are abundantly expressed in muscle (e.g., *Atp2a1* and *Myh4*) and blood and bone marrow (e.g., *Hbb-b1*). Preparing diaphyseal bone samples more carefully would lessen the problem of contamination, but could itself induce changes in gene expression due to the extra time the specimen is kept ex vivo. One proposed strategy for

reducing sample contamination by muscle, blood, and bone marrow, involves a short-term (~1 hour) ex vivo digestion of bone specimens with collagenase.<sup>(26,27)</sup> We confirmed that this approach reduces the abundance of many muscle, blood, and bone marrow-derived transcripts in diaphyseal bone RNA-seq libraries, presumably by enzymatically removing the contaminating tissues. However, we also found that this approach increased the expression of many transcripts, some by as much as 80,000-fold. Consequently, whereas collagenase digestion may remove contaminating tissues it can also alter gene expression in the remaining cells. As an alternative to collagenase treatment, we developed a numerical filtering strategy for eliminating transcripts that likely originate from contaminating tissue. We identified genes (Supplementary Table 4) whose expression is (1) higher in skeletal muscle, blood, and bone marrow, compared to bone; and (2) whose transcript abundance dropped significantly following collagenase digestion. By in silico-filtering diaphyseal bone RNA-seq data for these genes, we significantly improved correlation between paired samples from the same animal (Fig. 3). This method of

numerically controlling for tissue contamination during the specimen preparation process also increased the consistency of correlations between animals with the same genotype. Therefore, this numerical filtering method should be useful to all investigators who perform RNA-seq on freshly processed diaphyseal bone specimens.

RNA-seq can detect a large range of transcript abundances. However, detecting significant differences in transcript abundance depends upon the frequency with which a transcript is detected in RNA-seq data; as a consequence, RNA-seq is less sensitive at detecting differences in low-abundance transcripts. Obtaining larger total numbers of RNA-seq reads will increase sensitivity for low-abundance transcripts and for detecting more subtle fold-changes in gene expression among adequately expressed transcripts. We pooled eight to 10 separately bar-coded RNA-seq libraries in order to obtain >10 million uniquely mapping reads/library/sequencing run. At current pricing, the material cost of generating this RNA-seq data was ~\$400/library. With this amount of data, we did not expect to reliably detect differences in gene expression for transcripts whose RPKM values were less than five. Therefore, for genes like *Axin2* (RPKM < 1), a transcriptional target of Wnt signaling whose expression level is sensitive to mutations in *Lrp5*,<sup>5</sup> we were unable to use RNA-seq to independently confirm this result. In addition, 10 million reads/library is not sensitive enough to accurately detect less than a 1.7-fold change in gene expression. Increasing the number of reads/library would improve the sensitivity of RNA-seq at detecting low-abundance transcripts and lower-fold changes in gene expression across samples, but this would also significantly increase cost.

Signaling via LRP5 clearly affects bone mass.<sup>(4–7,44)</sup> Several missense mutations, including *LRP5*<sup>p.A214V</sup>, cause an HBM phenotype<sup>(1,45)</sup> and *LRP5* loss-of-function mutations cause very low bone mass.<sup>(2)</sup> We performed RNA-seq on diaphyseal bone samples from mice with three different *Lrp5* genotypes (ie, *Lrp5*<sup>+/+</sup>, *Lrp5*<sup>p.A214V/+</sup>, and *Lrp5*<sup>-/-</sup>) to identify changes in gene expression that could inform us about pathways that function upstream and downstream of the LRP5 receptor. Despite the profound effect the *Lrp5*<sup>p.A214V</sup> allele has on mouse bone mass (ie, a fourfold increase in trabecular bone volume/total volume and a 1.5-fold increase in mid-diaphyseal cortical bone area), we observed few greater than twofold differences in gene expression between *Lrp5*<sup>p.A214V/+</sup> and *Lrp5*<sup>+/+</sup> mice (Supplementary Table 7) and no changes in expression that appear to represent “smoking guns.” It is possible that we would have found more differences in gene expression had we studied younger mice when they are more actively increasing bone mass. Alternatively, enhanced LRP5 signaling from the *Lrp5*<sup>p.A214V</sup> allele may only cause changes in gene expression that are below the threshold of detection in our experiment. For example, we cannot detect a ~1.01-fold increase or decrease in gene expression. However, a ~1.01-fold daily increase in bone mass accrual would be sufficient to explain the fourfold increase in trabecular bone volume in *Lrp5*<sup>p.A214V/+</sup> mice compared to wild-type mice by 16 weeks of age.

Interestingly, we observed 84 differentially expressed protein-coding genes when RNA-seq data from either *Lrp5*<sup>p.A214V/+</sup> or *Lrp5*<sup>+/+</sup> mice were compared to data from *Lrp5*<sup>-/-</sup> mice. Because

the *Lrp5*<sup>p.A214V</sup> allele is able to transduce canonical Wnt signaling as efficiently as the *Lrp5*<sup>+</sup> allele, differences in gene expression likely result from deficient LRP5-mediated Wnt signaling. Genes whose protein products contribute to the synthesis of bone’s extracellular matrix (e.g., *Col1a1*, *Col1a2*, *Bglap*, *Bglap2*, *Mt1*, and *Mt2*) had reduced abundance in *Lrp5*<sup>-/-</sup> mice, as did the Wnt pathway inhibitor *Apcdd1*. *Lrp5*<sup>-/-</sup> mice also had increased expression of *Wnt10b*, a Wnt ligand that is anabolic in bone, and the Wnt coreceptor *Fzd4*, which can transduce Wnt signal via LRP5 or LRP6. These data suggest that in the absence of LRP5, bone cells are less able to produce critical matrix components. In addition, the cells alter their expression of specific Wnt pathway components in an attempt to increase Wnt signaling. Also, within the list of genes whose expression levels differ between LRP5-sufficient and LRP5-insufficient mice are genes with unknown function (e.g., *Zfp445*). These genes may now be considered to be targets of LRP5-mediated signaling and candidates for affecting bone properties.

## Disclosures

All authors state that they have no conflicts of interest.

## Acknowledgments

This work was supported by NIAMS/NIH R01AR053237 (to AGR), NIAMS/NIH R21AR062326 (to MLW), the Osteogenesis Imperfecta Foundation (to CMJ), and the Howard Hughes Medical Institute (to MLW and CES). We thank Joshua Levin at the Broad Institute and Victor Johnson at the Boston Children’s Hospital for thoughtful discussions.

Authors’ roles: UMA contributed to experimental design, data collection, data analysis, and writing the first draft of the manuscript. CMJ contributed to data collection. DCC, JG, JGS, CES, and AGR contributed to experimental design. MLW contributed to experimental design, data analysis, and writing the first draft of the manuscript. All authors reviewed and approved the final version of the manuscript. MLW accepts responsibility for the integrity of the data analysis.

## References

1. Boyden LM, Mao J, Belsky J, Mitzner L, Farhi A, Mitnick MA, Wu D, Insogna K, Lifton RP. High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med*. 2002;346(20):1513–21.
2. Gong Y, Slee RB, Fukai N, Rawadi G, Roman-Roman S, Reginato AM, Wang H, Cundy T, Glorieux FH, Lev D, Zacharin M, Oexle K, Marcelino J, Suwairi W, Heeger S, Sabatakos G, Apte S, Adkins WN, Allgrove J, Arslan-Kirchner M, Batch JA, Beighton P, Black GC, Boles RG, Boon LM, Borrone C, Brunner HG, Carle GF, Dallapiccola B, De Paepe A, Floege B, Halfhide ML, Hall B, Hennekam RC, Hirose T, Jans A, Jüppner H, Kim CA, Keppeler-Noreuil K, Kohlschütter A, LaCombe D, Lambert M, Lemyre E, Letteboer T, Peltonen L, Ramasar RS, Romanengo M, Somer H, Steichen-Gersdorf E, Steinmann B, Sullivan B, Superti-Furga A, Swoboda W, van den Boogaard MJ, Van Hul W, Vikkula M, Votruba M, Zabel B, Garcia T, Baron R, Olsen BR, Warman ML. Osteoporosis-Pseudoglioma Syndrome Collaborative Group. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell*. 2001;107(4):513–23.
3. Little RD, Carulli JP, Del Mastro RG, Dupuis J, Osborne M, Folz C, Manning SP, Swain PM, Zhao SC, Eustace B, Lappe MM, Spitzer L,

- Zweier S, Braunschweiger K, Benchekroun Y, Hu X, Adair R, Chee L, FitzGerald MG, Tulig C, Caruso A, Tzellas N, Bawa A, Franklin B, McGuire S, Nogues X, Gong G, Allen KM, Anisowicz A, Morales AJ, Lomedico PT, Recker SM, Van Eerdewegh P, Recker RR, Johnson ML. A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am J Hum Genet.* 2002;70(1):11–9.
4. Clement-Lacroix P, Ai M, Morvan F, Roman-Roman S, Vayssières B, Belleville C, Estrera K, Warman ML, Baron R, Rawadi G. Lrp5-independent activation of Wnt signaling by lithium chloride increases bone formation bone mass in mice. *Proc Natl Acad Sci U.S.A.* 2005;102(48):17406–11.
  5. Cui Y, Niziolek PJ, MacDonald BT, Zylstra CR, Alenina N, Robinson DR, Zhong Z, Matthes S, Jacobsen CM, Conlon RA, Brommage R, Liu Q, Mseeh F, Powell DR, Yang QM, Zambrowicz B, Gerrits H, Gossen JA, He X, Bader M, Williams BO, Warman ML, Robling AG. Lrp5 functions in bone to regulate bone mass. *Nat Med.* 2011;17(6):684–91.
  6. Niziolek PJ, Warman ML, Robling AG. Mechanotransduction in bone tissue: The A214V and G171V mutations in Lrp5 enhance load-induced osteogenesis in a surface-selective manner. *Bone.* 2012;51(3):459–65.
  7. Sawakami K, Robling AG, Ai M, Pitner ND, Liu D, Warden SJ, Li J, Maye P, Rowe DW, Duncan RL, Warman ML, Turner CH. The Wnt co-receptor LRP5 is essential for skeletal mechanotransduction but not for the anabolic bone response to parathyroid hormone treatment. *J Biol Chem.* 2006;281(33):23698–711.
  8. Saxon LK, Jackson BF, Sugiyama T, Lanyon LE, Price JS. Analysis of multiple bone responses to graded strains above functional levels, and to disuse, in mice in vivo show that the human Lrp5 G171V High Bone Mass mutation increases the osteogenic response to loading but that lack of Lrp5 activity reduces it. *Bone.* 2011;49(2):184–93.
  9. Bafico A, Liu G, Yaniv A, Gazit A, Aaronson SA. Novel mechanism of Wnt signalling inhibition mediated by Dickkopf-1 interaction with LRP6/Arrow. *Nat Cell Biol.* 2001;3(7):683–6.
  10. Li X, Zhang Y, Kang H, Liu W, Liu P, Zhang J, Harris SE, Wu D. Sclerostin binds to LRP5/6 and antagonizes canonical Wnt signaling. *J Biol Chem.* 2005;280(20):19883–7.
  11. Semenov M, Tamai K, He X. SOST is a ligand for LRP5/LRP6 and a Wnt signaling inhibitor. *J Biol Chem.* 2005;280(29):26770–5.
  12. Semenov MV, Tamai K, Brott BK, Kühl M, Sokol S, He X. Head inducer Dickkopf-1 is a ligand for Wnt coreceptor LRP6. *Curr Biol.* 2001;11(12):951–61.
  13. Ai M, Holmen SL, Van Hul W, Williams BO, Warman ML. Reduced affinity to and inhibition by DKK1 form a common mechanism by which high bone mass-associated missense mutations in LRP5 affect canonical Wnt signaling. *Mol Cell Biol.* 2005;25(12):4946–55.
  14. Bourhis E, Wang W, Tam C, Hwang J, Zhang Y, Spittler D, Huang OW, Gong Y, Estevez A, Zilberleyb I, Rouge L, Chiu C, Wu Y, Costa M, Hannoush RN, Franke Y, Cochran AG. Wnt antagonists bind through a short peptide to the first beta-propeller domain of LRP5/6. *Structure.* 2011;19(10):1433–42.
  15. Semenov MV, He X. LRP5 mutations linked to high bone mass diseases cause reduced LRP5 binding and inhibition by SOST. *J Biol Chem.* 2006;281(50):38276–84.
  16. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 2011;9:34.
  17. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* 2011;12(2):87–98.
  18. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621–8.
  19. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science.* 2008;320(5881):1344–9.
  20. Davanian H, Stranneheim H, Båge T, Lagervall M, Jansson L, Lundeberg J, Yucel-Lindberg T. Gene expression profiles in paired gingival biopsies from periodontitis-affected and healthy tissues revealed by massively parallel sequencing. *PLoS One.* 2012;7(9):e46440.
  21. Jäger M, Ott CE, Grünhagen J, Hecht J, Schell H, Mundlos S, Duda GN, Robinson PN, Lienau J. Composite transcriptome assembly of RNA-seq data in a sheep model for delayed bone healing. *BMC Genomics.* 2011;12:158.
  22. Yao B, Zhao Y, Zhang H, Zhang M, Liu M, Liu H, Li J. Sequencing and de novo analysis of the Chinese Sika deer antler-tip transcriptome during the ossification stage using Illumina RNA-Seq technology. *Biotechnol Lett.* 2012;34(5):813–22.
  23. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol.* 2006;7:3.
  24. Grant GR, Farkas MH, Pizarro AD, Lahens NF, Schug J, Brunk BP, Stoekert CJ, Hogenesch JB, Pierce EA. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics.* 2011;27(18):2518–28.
  25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
  26. Wasserman E, Webster D, Kuhn G, Attar-Namdar M, Muller R, Bab I. Differential load-regulated global gene expression in mouse trabecular osteocytes. *Bone.* 2013;53(1):14–23.
  27. Xiong J, Onal M, Jilka RL, Weinstein RS, Manolagas SC, O'Brien CA. Matrix-embedded cells control osteoclast formation. *Nat Med.* 2011;17(10):1235–41.
  28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57(1):289–300.
  29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
  30. Carlson M, Pages H, Aboyoun P, Falcon S, Morgan M, Sarkar D, Lawrence M. 2009; Genomic features: tools for making and manipulating transcript centric annotations [Internet]. Bioconductor: Open Source Software for Bioinformatics; 2009 [cited 2013 Apr 6]. Available from: <http://www.bioconductor.org/packages/2.6/bioc/html/GenomicFeatures.html>.
  31. McMahon AP, Champion JE, McMahon JA, Sukhatme VP. Developmental expression of the putative transcription factor Egr-1 suggests that Egr-1 and c-fos are coregulated in some tissues. *Development.* 1990;108(2):281–7.
  32. Srivastava S, Weitzmann MN, Kimble RB, Rizzo M, Zahner M, Milbrandt J, Ross FP, Pacifici R. Estrogen blocks M-CSF gene expression and osteoclast formation by regulating phosphorylation of Egr-1 and its interaction with Sp-1. *J Clin Invest.* 1998;102(10):1850–9.
  33. Ishimi Y, Miyaura C, Jin CH, Akatsu T, Abe E, Nakamura Y, Yamaguchi A, Yoshiki S, Matsuda T, Hirano T. IL-6 is produced by osteoblasts and induces bone resorption. *J Immunol.* 1990;145(10):3297–303.
  34. Ha J, Choi HS, Lee Y, Kwon HJ, Song YW, Kim HH. CXC chemokine ligand 2 induced by receptor activator of NF-kappa B ligand enhances osteoclastogenesis. *J Immunol.* 2010;184(9):4717–24.
  35. Asagiri M, Takayanagi H. The molecular understanding of osteoclast differentiation. *Bone.* 2007;40(2):251–64.
  36. Wagner EF. Bone development and inflammatory disease is regulated by AP-1 (Fos/Jun). *Ann Rheum Dis.* 2010; 69(Suppl 1):i86–8.
  37. Choi HY, Dieckmann M, Herz J, Niemeier A. Lrp4, a novel receptor for Dickkopf 1 and sclerostin, is expressed by osteoblasts and

- regulates bone growth and turnover in vivo. *PLoS One*. 2009;4(11):e7930.
38. Holmen SL, Giambernardi TA, Zylstra CR, Buckner-Berghuis BD, Resau JH, Hess JF, Glatt V, Boussein ML, Ai M, Warman ML, Williams BO. Decreased BMD and limb deformities in mice carrying mutations in both *Lrp5* and *Lrp6*. *J Bone Miner Res*. 2004;19(12):2033–40.
  39. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*. 2011;12(2):R18.
  40. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Catenazzi MCE, Chang S, Cooley RN, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtkova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, LingNg B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Sohna Sohna JE, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53–9.
  41. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res*. 2008;36(16):e105.
  42. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26(3):317–25.
  43. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods*. 2009;6(4):291–5.
  44. Robinson JA, Chatterjee-Kishore M, Yaworsky PJ, Cullen DM, Zhao W, Li C, Kharode Y, Sauter L, Babji P, Brown EL, Hill AA, Akhter MP, Johnson ML, Recker RR, Komm BS, Bex FJ. Wnt/beta-catenin signaling is a normal physiological response to mechanical loading in bone. *J Biol Chem*. 2006;281(42):31720–8.
  45. Van Wesenbeeck L, Cleiren E, Gram J, Beals RK, Bénichou O, Scopelliti D, Key L, Renton T, Bartels C, Gong Y, Warman ML, De Vernejoul MC, Bollerslev J, Van Hul W. Six novel missense mutations in the LDL receptor-related protein 5 (*LRP5*) gene in different conditions with an increased bone density. *Am J Hum Genet*. 2003;72(3):763–71.
  46. Fong L, Tan K, Tran C, Cool J, Scherer MA, Elovaris R, Coyle P, Foster BK, Rofe AM, Xian CJ. Interaction of dietary zinc and intracellular binding protein metallothionein in postnatal bone growth. *Bone*. 2009;44(6):1151–62.
  47. Nagata M, Lonnerdal B. Role of zinc in cellular zinc trafficking and mineralization in a murine osteoblast-like cell line. *J Nutr Biochem*. 2011;22(2):172–8.
  48. Purcell P, Joo BW, Hu JK, Tran PV, Calicchio ML, O'Connell DJ, Maas RL, Tabin CJ. Temporomandibular joint formation requires two distinct hedgehog-dependent steps. *Proc Natl Acad Sci U S A*. 2009;106(43):18297–302.
  49. Moeller C, Swindell EC, Kispert A, Eichele G. Carboxypeptidase Z (CPZ) modulates Wnt signaling and regulates the development of skeletal elements in the chicken. *Development*. 2003;130(21):5103–11.
  50. Wang L, Shao YY, Ballock RT. Carboxypeptidase Z (CPZ) links thyroid hormone and Wnt signaling pathways in growth plate chondrocytes. *J Bone Miner Res*. 2009;24(2):265–73.
  51. Liu S, Tang W, Fang J, Ren J, Li H, Xiao Z, Quarles LD. Novel regulators of *Fgf23* expression and mineralization in Hyp bone. *Mol Endocrinol*. 2009;23(9):1505–18.
  52. Cruzan G, Bus J, Banton M, Gingell R, Carlson G. Mouse specific lung tumors from CYP2F2-mediated cytotoxic metabolism: an endpoint/toxic response where data from multiple chemicals converge to support a mode of action. *Regul Toxicol Pharmacol*. 2009;55(2):205–18.
  53. Karsenty G, Park RW. Regulation of type I collagen genes expression. *Int Rev Immunol*. 1995;12(2–4):177–85.
  54. Letra A, Menezes R, Fonseca RF, Govil M, McHenry T, Murphy MJ, Hennebold JD, Granjeiro JM, Castilla EE, Orioli IM, Martin R, Marazita ML, Bjork BC, Vieira AR. Novel cleft susceptibility genes in chromosome 6q. *J Dent Res*. 2010;89(9):927–32.
  55. Eames BF, Amores A, Yan YL, Postlethwait JH. Evolution of the osteoblast: skeletogenesis in gar and zebrafish. *BMC Evol Biol*. 2012;12:27.
  56. Goto T, Matsui Y, Fernandes RJ, Hanson DA, Kubo T, Yukata K, Michigami T, Komori T, Fujita T, Yang L, Eyre DR, Yasui N. Sp1 family of transcription factors regulates the human alpha2 (X1) collagen gene (*COL11A2*) in Saos-2 osteoblastic cells. *J Bone Miner Res*. 2006;21(5):661–73.
  57. Li SW, Arita M, Kopen GC, Phinney DG, Prockop DJ. A 1,064 bp fragment from the promoter region of the *Col11a2* gene drives lacZ expression not only in cartilage but also in osteoblasts adjacent to regions undergoing both endochondral and intramembranous ossification in mouse embryos. *Matrix Biol*. 1998;17(3):213–21.
  58. Kawai T, Nomura F, Hoshino K, Copeland NG, Gilbert DJ, Jenkins NA, Akira S. Death-associated protein kinase 2 is a new calcium/calmodulin-dependent protein kinase that signals apoptosis through its catalytic activity. *Oncogene*. 1999;18(23):3471–80.
  59. Rodriguez deCordoba, Esparza-Gordillo S, Goicoechea J, de Jorge E, Lopez-Trascasa M, Sanchez-Corral P. The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol*. 2004;41(4):355–67.
  60. Okazaki R, Toriumi M, Fukumoto S, Miyamoto M, Fujita T, Tanaka K, Takeuchi Y. Thiazolidinediones inhibit osteoclast-like cell formation and bone resorption in vitro. *Endocrinology*. 1999;140(11):5060–5.
  61. Raymond MH, Schutte BC, Torner JC, Burns TL, Willing MC. Osteocalcin: genetic and physical mapping of the human gene *BGLAP* and its potential role in postmenopausal osteoporosis. *Genomics*. 1999;60(2):210–7.
  62. Sosnoski DM, Gay CV. Evaluation of bone-derived and marrow-derived vascular endothelial cells by microarray analysis. *J Cell Biochem*. 2007;102(2):463–72.

63. Marine JC, Jochemsen AG. Mdmx as an essential regulator of p53 activity. *Biochem Biophys Res Commun.* 2005;331(3):750–60.
64. Xu N, Wang Y, Li D, Chen G, Sun R, Zhu R, Sun S, Liu H, Yang G, Dong T. MDM4 overexpression contributes to synoviocyte proliferation in patients with rheumatoid arthritis. *Biochem Biophys Res Commun.* 2010;401(3):417–21.
65. Wrzeszczynski KO, Varadan V, Byrnes J, Lum E, Kamalakaran S, Levine DA, Dimitrova N, Zhang MQ, Lucito R. Identification of tumor suppressors and oncogenes from genomic and epigenetic features in ovarian cancer. *PLoS One.* 2011;6(12):e28503.
66. Baud'huin M, Renault R, Charrier C, Riet A, Moreau A, Brion R, Gouin F, Duplomb L, Heymann D. Interleukin-34 is expressed by giant cell tumours of bone and plays a key role in RANKL-induced osteoclastogenesis. *J Pathol.* 2010;221(1):77–86.
67. Chemel M, Le Goff B, Brion R, Cozic C, Berreur M, Amiaud J, Bougras G, Touchais S, Blanchard F, Heymann MF, Berthelot JM, Verrecchia F, Heymann D. Interleukin 34 expression is associated with synovitis severity in rheumatoid arthritis patients. *Ann Rheum Dis.* 2012;71(1):150–4.
68. Kim J, Jung Y, Sun H, Joseph J, Mishra A, Shiozawa Y, Wang J, Krebsbach PH, Taichman RS. Erythropoietin mediated bone formation is regulated by mTOR signaling. *J Cell Biochem.* 2012; 113(1):220–8.
69. Crockett JC, Schütze N, Tosh D, Jatzke S, Duthie A, Jakob F, Rogers MJ. The matricellular protein CYR61 inhibits osteoclastogenesis by a mechanism independent of  $\alpha$ v $\beta$ 3 and  $\alpha$ v $\beta$ 5. *Endocrinology.* 2007;148(12):5761–8.
70. Si W, Kang Q, Luu HH, Park JK, Luo Q, Song WX, Jiang W, Luo X, Li X, Yin H, Montag AG, Haydon RC, He TC. CCN1/Cyr61 is regulated by the canonical Wnt signal and plays an important role in Wnt3A-induced osteoblast differentiation of mesenchymal stem cells. *Mol Cell Biol.* 2006;26(8):2955–64.
71. Knowles HJ, Cleton-Jansen AM, Korsching E, Athanasou NA. Hypoxia-inducible factor regulates osteoclast-mediated bone resorption: role of angiopoietin-like 4. *FASEB J.* 2010;24(12):4648–59.
72. Igwe JC, Jiang X, Paic F, Ma L, Adams DJ, Baldock PA, Pilbeam CC, Kalajzic I. Neuropeptide Y is expressed by osteocytes and can inhibit osteoblastic activity. *J Cell Biochem.* 2009;108(3):621–30.
73. Matic I, Matthews BG, Kizivat T, Igwe JC, Marijanovic I, Ruohonen ST, Savontaus E, Adams DJ, Kalajzic I. Bone-specific overexpression of NPY modulates osteogenesis. *J Musculoskelet Neuronal Interact.* 2012;12(4):209–18.
74. Yamada Y, Ando F, Shimokata H. Association of polymorphisms in forkhead box C2 and perilipin genes with bone mineral density in community-dwelling Japanese individuals. *Int J Mol Med.* 2006;18(1):119–27.
75. Okabe Y, Sano T, Nagata S. Regulation of the innate immune response by threonine-phosphatase of Eyes absent. *Nature.* 2009;460(7254):520–4.
76. Quintana FJ, Solomon A, Cohen IR, Nussbaum G. Induction of IgG3 to LPS via Toll-like receptor 4 co-stimulation. *PLoS One.* 2008;3(10): e3509.
77. Rens A, Moelling K. Bcr is a negative regulator of the Wnt signalling pathway. *EMBO Rep.* 2005;6(11):1095–100.