

Full Length Article

Transcriptional profiling of murine osteoblast differentiation based on RNA-seq expression analyses



Layal Abo Khayal^{a,b}, Johannes Grünhagen^a, Ivo Provazník^{b,c}, Stefan Mundlos^{a,d}, Uwe Kornak^{a,d}, Peter N. Robinson^{a,e}, Claus-Eric Ott^{a,d,*}

^a Institute for Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

^b Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

^c International Clinical Research Center, Center of Biomedical Engineering, St. Anne's University Hospital Brno, Brno, Czech Republic

^d Research Group Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany

^e The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032, USA

ARTICLE INFO

Keywords:

Bone cells

RNAseq

Non-coding RNA

Topological domains

Alternative splicing

ABSTRACT

Osteoblastic differentiation is a multistep process characterized by osteogenic induction of mesenchymal stem cells, which then differentiate into proliferative pre-osteoblasts that produce copious amounts of extracellular matrix, followed by stiffening of the extracellular matrix, and matrix mineralization by hydroxylapatite deposition. Although these processes have been well characterized biologically, a detailed transcriptional analysis of murine primary calvaria osteoblast differentiation based on RNA sequencing (RNA-seq) analyses has not previously been reported.

Here, we used RNA-seq to obtain expression values of 29,148 genes at four time points as murine primary calvaria osteoblasts differentiate *in vitro* until onset of mineralization was clearly detectable by microscopic inspection. Expression of marker genes confirmed osteogenic differentiation. We explored differential expression of 1386 protein-coding genes using unsupervised clustering and GO analyses. 100 differentially expressed lncRNAs were investigated by co-expression with protein-coding genes that are localized within the same topologically associated domain. Additionally, we monitored expression of 237 genes that are silent or active at distinct time points and compared differential exon usage.

Our data represent an in-depth profiling of murine primary calvaria osteoblast differentiation by RNA-seq and contribute to our understanding of genetic regulation of this key process in osteoblast biology.

1. Introduction

Osteoblasts are specialized fibroblasts that secrete and mineralize the bone matrix. Osteoblastic differentiation from mesenchymal stem cell (MSCs) is an important step of bone formation. The mineralized extracellular matrix (ECM) is mainly composed of type I collagen and smaller but significant amounts of various proteins including osteocalcin (OC), matrix gla protein (MGP), osteopontin (OPN), bone sialoprotein (BSP), bone morphogenic proteins (BMPs), and transforming growth factor beta (TGF-beta), as well as the inorganic mineral hydroxylapatite [1].

Osteoblast differentiation comprises three major processes: (1) cell proliferation, (2) matrix maturation, and (3) matrix mineralization. Already during proliferation, ECM proteins are secreted by osteoblasts

forming the non-mineralized bone matrix or osteoid. This precedes crosslinking of proteins of the osteoid such as collagen type I fibrils during matrix maturation, thereby forming a rigid structure that acts as a template onto which inorganic minerals are deposited to form mineralized bone. Eventually, osteoblasts transform into inactive osteoblasts, become bone-lining cells or osteocytes, or undergo apoptosis [2].

This process is critical for the fine-tuning of bone tissue structures. During the many molecular events of bone development and osteoblast differentiation, gene expression levels undergo substantial changes. However, gene abundance represents only part of the complexity of the transcriptome. Specific exons are included or excluded from mature mRNAs by alternative splicing, which adds proteomic diversity by generating distinct isoforms from the same gene. It also influences transcripts by changing mRNA-stability, efficiency of translation, the number of microRNA-binding

* Corresponding author at: Institute for Medical Genetics and Human Genetics, Charité - Universitätsmedizin Berlin, Campus Virchow Klinikum, Augustenburgerplatz 1, 13353 Berlin, Germany.

E-mail address: claus-eric.ott@charite.de (C.-E. Ott).

<https://doi.org/10.1016/j.bone.2018.04.006>

Received 21 November 2017; Received in revised form 3 March 2018; Accepted 9 April 2018

Available online 11 April 2018

8756-3282/ © 2018 Elsevier Inc. All rights reserved.

sites, or switching localization signals [3].

Several studies used gene expression microarrays to address osteoblast differentiation. Some focused on osteogenic induction of human adipose-derived stem cells [4] or human bone marrow mesenchymal stem cells [5]. Others used a time series microarray analysis of mineralizing murine primary osteoblast cultures [6]. Here, we have used RNA sequencing (RNA-seq) to develop a high-resolution transcriptome data set of murine osteoblast differentiation. Compared to microarray technologies, RNA-seq is an open system, has better dynamic range in estimates of low-level gene expression and accuracy [7], and allows measurement of alternative splicing. Few studies have used RNA-seq to investigate transcriptional regulation of osteogenic differentiation in other cell types such as ST2 cells in response to BMP2 [8], murine marrow derived MSCs differentiated to day 7 or day 15 with osteogenic or adipogenic media [9], the osteocytogenic cell line IDG-SW3 [10], and hMSC-TERT cells [11] which are human bone marrow stromal cells stably transduced by a retroviral vector containing the *hTERT* gene. To our knowledge, the most comprehensive data set on murine primary calvaria osteoblast differentiation is available in the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/gds>) under the accession GSE54461. The authors obtained pre-osteoblast like cells from neonatal calvaria from transgenic C57BL/6J mice expressing cyan fluorescent protein (CFP) under the control of the Col3.6 promoter. These cells were cultured for 4 days, subjected FACS sorting based on the presence of CFP expression, and then returned to culture. During differentiation RNA was collected at 9 time points. However, the data set is not connected to cellular characterization of the cells at the distinct time points, and the related publication [12] has another focus.

We used murine primary calvaria osteoblasts, i.e. primary cells which already committed to the osteogenic lineage. Thus, transcriptional profiling focused on osteoblast differentiation rather than osteogenic induction. In addition to mRNA expression of protein-coding genes, we assessed alternative splicing and long non-coding RNA expression. We analyzed four different time points and defined microscopically visible mineralization as end point.

2. Material and methods

2.1. Osteoblast differentiation and RNA isolation

Sacrifice of animals followed regulations of the animal welfare organization as approved by the local legal representative (Landesamt für Gesundheit und Soziales Berlin: T 0438/08). Primary calvaria osteoblasts (pCOBs) were harvested from newborn C57BL/6 wild-type mice (P0–P4) [13,14]. Cells were seeded on 6-well plates in Alpha-Mem (Lonza, Basel, Switzerland) containing 10% fetal calf serum (FCS; Gibco, Life Technologies, Carlsbad, California, USA), Pen/Strep (100 U/mL, Lonza), and 2 mM ultra-glutamine (Lonza). At confluence, cells were harvested, and on the other plates medium was supplemented with 50 μ M L-ascorbate-2-phosphate and 10 mM beta-glycerophosphate to promote osteoblast differentiation. Cells were washed with PBS before lysis in RNeasy Lysis Buffer (Qiagen) at day 0 (i.e. confluent cultures without stimulation), day 3, day 6, and day 12. Total RNA was isolated using phenol/chloroform extraction. RNA integrity was confirmed using the Agilent 2100 Bioanalyzer with the Agilent RNA 6000 Nano Kit according to the manufacturer's instructions. This procedure was repeated three times to obtain biological replicates.

2.2. RNA sequencing

We used the TruSeq RNA Sample Prep Kit for library preparation. After cluster generation by the TruSeq PE Cluster Kit, 101 bp paired-end sequencing was performed on a HiSeq 2000 (Illumina, San Diego, USA).

2.3. Alignment of paired end reads

We generated FASTQ files with the standard Illumina pipeline, used FastQC version 0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) to assess quality, and removed poor quality and adapter contamination with Trimmomatic version 0.36 [15]. Reads were aligned to reference genome (GRCm38, Ensembl release 86) using Tophat2 (V. 2.0.13) [16]. A count matrix containing number of reads mapped to each gene in each sample was generated using FeatureCounts [17].

2.4. Statistical analysis of differential gene expression

To compare aligned reads across samples, we used DESeq2, which normalizes raw read counts taking into account the sample size factor and further sources of technical bias such as GC content or gene length [18]. We applied Wald test for testing the significance in gene expression changes between differentiation time points. We adjusted *p*-values for multiple testing using Bonferroni correction. For hierarchical clustering, we calculated the mean per time point of DESeq normalized read counts since PCA indicated high similarity of replicates. Expression values were scaled to a mean of zero and a standard deviation of one. The scaled new value $x_{ij}' = (x_{ij} - \mu)/\sigma$, where x_{ij} is the mean of the normalized count of the replicates at time point *j* and of gene *i*, μ is the mean of x_{ij} , and σ is the standard deviation. Heatmaps were generated using the heatmap.2 function from “gplots” R package, calculating the Euclidean distances of the values x_{ij}' to generate the dendrogram. The cut-off of the dendrogram to define 9 clusters was at the distance 2.23. For protein-coding genes, GO enrichment analyses were performed using Ontologizer [19] with settings Parent-child intersection [20] and Benjamini Hochberg multiple testing correction. GO terms with adjusted *p* < 0.05 were considered significant.

2.5. Analyses of long non-coding RNAs (lncRNAs)

Long non-coding RNAs were identified according to Ensembl gene biotype annotation [21]. We classified lncRNAs with absolute fold change ≥ 1.5 and adjusted *p*-value < 0.01 in at least one pairwise comparison as differentially expressed. Co-expression of lncRNAs and protein-coding genes may suggest potential regulatory mechanisms. We restricted co-expression with protein-coding genes by their localization within the same topologically associated domain (TAD). We derived murine TAD definitions from Jesse R. Dixon et al. [22]. Prior to analysis, we converted given NCBI37/mm9-based TAD coordinates to GRCm38/mm10 using the UCSC Lift Genome Annotations tool [23]. For each significantly expressed lncRNA localized within an annotated TAD, we determined a protein-coding gene within the same TAD with highest positive and highest negative correlation.

2.6. Analysis of ON-OFF genes

To address genes that are turned on and off during osteoblast differentiation we defined a gene to be silent, i.e. “off” at a distinct time point, if the normalized count was 0 in at least two replicates and if the mean of normalized counts of all replicates was < 1.5. Otherwise, a gene was defined to be “on”.

2.7. Analyses of differential exon usage and alternative splicing

To analyze differential exon usage, we used DEXSeq [24]. Analysis focused on protein-coding exons.

2.8. Data access

Data from this work are available from the NCBI BioProject at <http://www.ncbi.nlm.nih.gov/bioproject>. BioProject ID: PRJNA416437; BioSample accessions SAMN07956470 to SAMN07956481.

3. Results and discussion

3.1. Experimental design

We used transcriptome profiling by RNA-seq to identify changes in gene expression as murine primary calvaria osteoblasts (pCOBs) differentiate until onset of mineralization (Supplementary Fig. S1). Harvesting and differentiation of pCOBs is a well-established method in our lab [13,14]. PolyA enriched cDNA libraries were prepared using the Illumina TruSeq protocol and 101-bp paired-end sequences were generated. On average 70 million raw reads per sample were obtained. About 87% of these reads were retained after trimming. 94.2% of the trimmed read pairs were mapped to the murine reference genome (Supplementary Table S1, Supplementary Fig. S2).

3.2. Statistical evaluation of gene coverage and biological replicates

The read coverage was homogeneous along the gene body and similar between samples (Supplementary Fig. S3). To test, whether coverage might vary across transcript features, we separate the first exons in 5' end, from the last exons in 3' end, and the middle exons. This analysis showed a similar distribution of read depth (base mean density) across the 5', middle, and 3' exons (Supplementary Fig. S4).

To visualize the relationship between samples, we used PCA and Euclidean distance analysis (Fig. 1). The first component (PC1) on the x-axis comprises 68% of variance. The proliferative phase samples cluster separately from the other three sample groups in the first principle component. The second component (PC2) on the y-axis explains 20% of the variance. In the second principle component samples obtained at day 12, i.e. onset of visible mineralization, cluster separately from the other three sample groups (Fig. 1A). Likewise, we observed a clear separation between time points when Euclidean distance between samples was computed and used to create a heatmap color image and dendrogram depicting the closeness between samples (Fig. 1B). Both analyses demonstrate that the variance between time points is much greater than the variance between biological replicates.

3.3. Expression of marker genes for lineage commitment and differentiation

To confirm the biological validity of our data, we analyzed differential expression of marker genes that are well known to be differentially expressed during osteoblast differentiation. As shown in Fig. 2A, *Bmp2*, a marker gene for osteogenic induction, displayed highest expression levels at day 0. *Runx2* which encodes one of the key transcription factors of bone formation is highly expressed with slight downregulation over time. *Col1a1* encoding the pro- $\alpha 1$ chains of type I collagen, the most abundant ECM component in bone, shows peak expression at day 6. One of the marker genes for osteoblast matrix remodeling *Mmp13* is highly expressed at day 3 and day 12 with an transitional drop at day 6. Marker genes for osteoblast differentiation such as *Bglap2* showed rising expression levels over time starting at day 3, and *Sost*, a marker for the onset of mineralization, showed no substantial expression until day 6, but expression at day 12.

In contrast, *Pparg*, a marker gene for adipogenic lineage commitment, displayed decreasing expression and *Lep* which is expressed by differentiated fat cells shows constantly low expression. The chondrogenic marker genes *Comp* and *Col9a1* are downregulated over time. Finally, marker genes for myogenic differentiation such as *Myod* and mature muscle cells such as *Ckm*, displayed marginal expression levels (Fig. 2B). Taken together, expression of marker genes confirmed osteogenic differentiation.

3.4. Identification of differentially expressed genes

After trimming, mapping, and filtering genes with > 1 count over all samples, the resulting count table contained expression values of

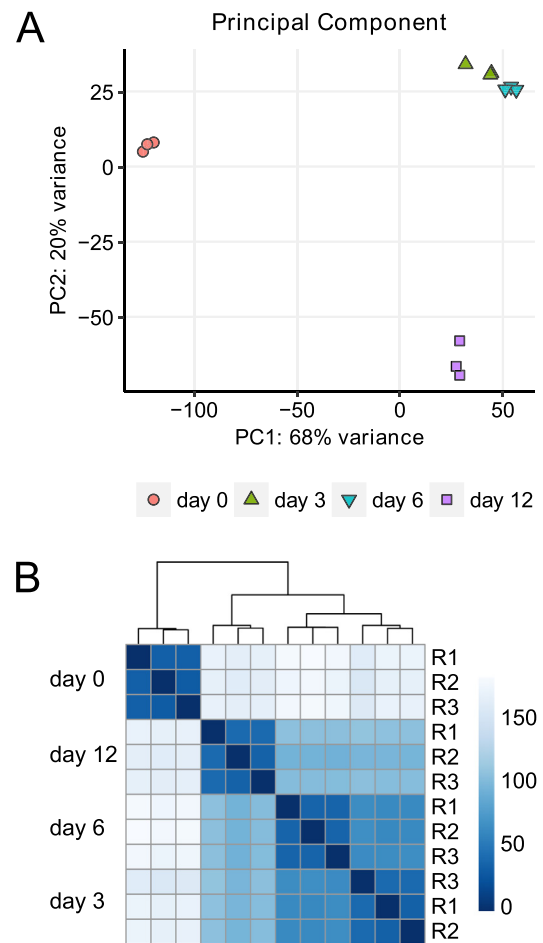


Fig. 1. Global gene expression profiles of osteoblast differentiation. **A.** Principal component analysis performed on regularized logarithmic transformed read counts matrix. Each point represents an experimental sample with point color and shape indicating the time points. The first component explains 68% of variability and the second component explains 20%. **B.** Euclidian distance metric plot. Distance between experimental groups, i.e. time points, is larger than distance between replicates. The dendrogram indicates that samples of day 0 are clearly separated from samples of the other time points. Distance between samples of day 3 and day 6 is much smaller.

29,148 genes. In total, there were 12,932 significantly differentially expressed genes with an adjusted p -value < 0.01 in at least one comparison, i.e. day 12 vs. day 0, day 6 vs. day 0, day 3 vs. day 0, day 12 vs. day 3, day 12 vs. day 6, and day 6 vs. day 3. Of these, 11,773 genes are protein-coding and 287 genes are annotated lncRNAs (Supplementary Table 2).

To describe the overall expression pattern, we visualized the global expression pattern in each comparison using MA-plots (Supplementary Fig. S5). Here, we subset 4012 genes with adjusted p -value < 0.01 and an absolute fold change ≥ 4 (Supplementary Table 3). Of these, 2414 genes were up-regulated and 2124 were down-regulated in at least one comparison. There was a clear predominance of upregulation in the comparisons with day 0, which is most prominent versus day 3 (1466 genes upregulated versus 912 genes downregulated). We speculate that this is an indicator of diversification of transcription, when osteoblasts enter the differentiation program. In contrast, comparison of intermediate time points day 3 and day 6 with day 12 indicate a preponderance of downregulation which is most prominent for the latter two time points (214 genes upregulated versus 412 genes downregulated). This might indicate that terminal differentiation is accompanied with specification of the gene-regulatory program.

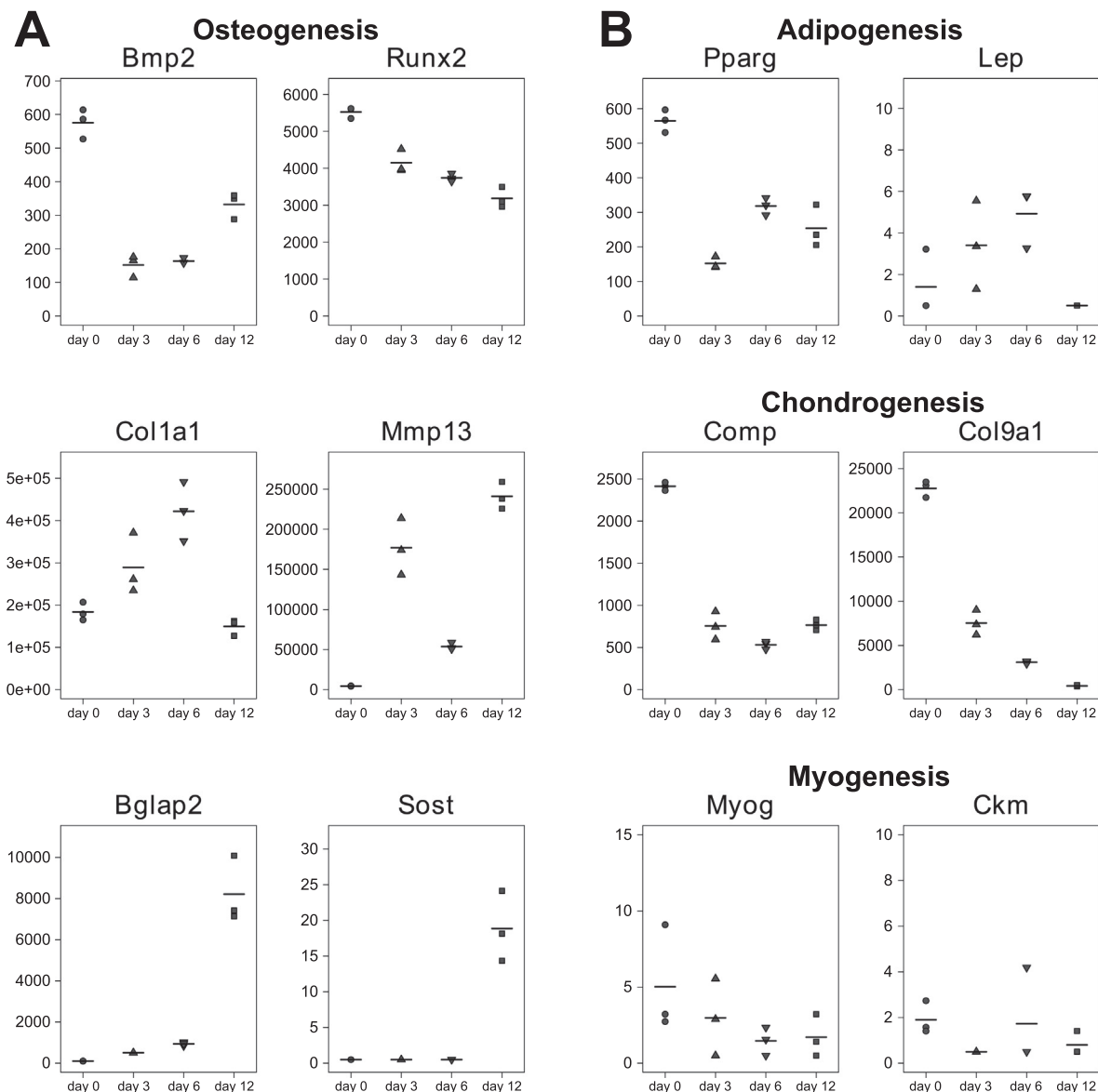


Fig. 2. Differential expression of selected marker genes. **A. Osteogenesis.** *Bmp2* and *Runx2* are known to be involved in osteoblast induction. Highest expression levels are observed at day 0. *Col1a1* encoding the major component of the osseous ECM is highly expressed with peak expression at day 6. *Mmp13*, which is involved in matrix remodeling, displays an intermittent drop of expression at day 6 while it is expressed at high levels at day 3 and day 12. *Bglap2* and *Sost*, marker genes for matrix mineralization, are highly expressed at day 12. **B. Commitment to other lineages.** **Adipogenesis.** *Pparg*, a marker gene for adipogenic lineage commitment, displays decreasing expression. *Lep* (Leptin), which is expressed by differentiated adipocytes, shows marginal expression. **Chondrogenesis.** Chondrocyte marker genes *Comp* and *Col9a1* show decreasing expression levels over time. **Myogenesis.** *Myog* (Myogenin), a marker gene for myogenic lineage commitment, shows low and further decreasing expression. *Ckm* (Creatine Kinase Muscle-type), which is expressed by differentiated myocytes, displays marginal expression. (A,B) Symbols correspond to expression levels of replicates and bars indicate mean expression levels at the indicated time points.

3.5. Exploratory analyses of protein-coding genes

We divided differentially expressed genes into two groups: protein-coding genes and non-coding transcripts. To get a reasonable number of protein-coding genes suitable for clustering and GO analyses, we used volcano plots (Supplementary Fig. S6) to define more stringent cut-offs by visual inspection. A total of 1386 protein-coding genes with an absolute fold change ≥ 5 and adjusted p -value $< 10^{-50}$ in at least one comparison were plotted within a heatmap to identify clusters of genes according to their expression pattern. Nine clusters were identified by visual inspection (Fig. 3).

Cluster A comprises a total of 303 genes that display high expression at day 0 and low expression levels at later time points. 301 of them are annotated to a GO term. 38/301 genes have “nucleic acid binding

transcription factor activity”. Several genes are involved in “response to endoplasmic reticulum stress” including *Atf3*, *Atf4* and *Ddit3*. This cluster contains *Bmpr1b* which is well known to be involved in osteoblast differentiation.

Cluster B contains 247 genes of which 245 genes are annotated to a GO term. Genes in this cluster display a very similar expression pattern in comparison to cluster A. In this cluster we identified again a large number of genes involved in “transcription from RNA polymerase II promoter” and “transcription factor activity, protein binding” including *Actn4* which is involved in regulation of transcription as well as substrate adhesion-dependent cell spreading. Almost 10% of these genes code for proteins localized in the ECM, and nine of them are structural constituents of collagen trimers. About 18% of these genes are involved in “movement of cell or subcellular component”.

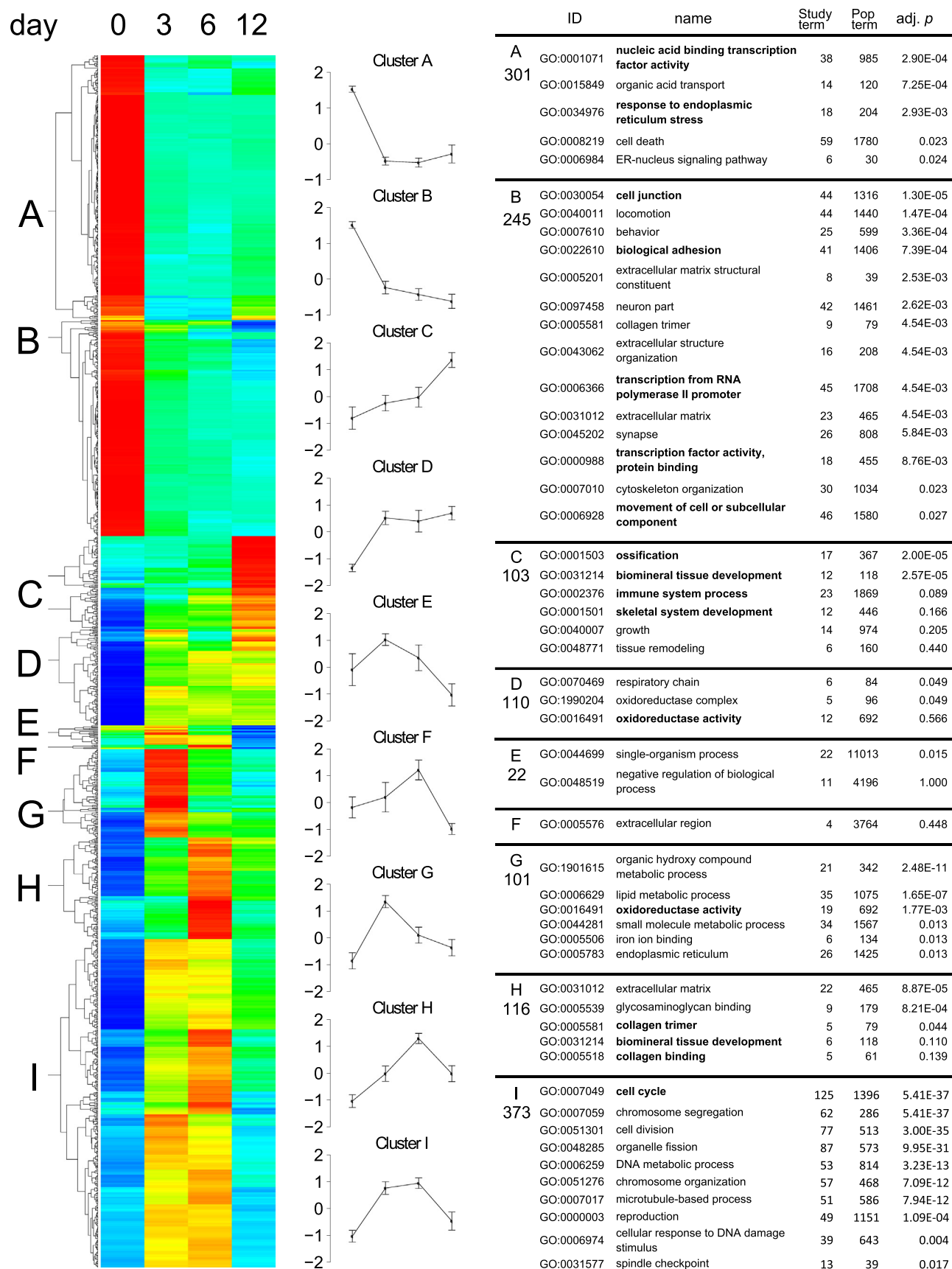


Fig. 3. Expression heatmap. Nine clusters were defined by visual inspection (cluster A-I). Line plots in the middle display expression dynamics of each cluster. For each cluster, GO analysis was performed as outlined in the methods section, and selected GO terms are given in the tables on the right side.

Genes that show increasing expression levels over time reaching highest expression at day 12 are localized in cluster C which contains 106 genes of which 103 genes are annotated to a GO term. 23/103 genes within this cluster are annotated to terms such as “ossification”, “biomineral tissue development”, and “skeletal system development” including well known genes such as *Bglap*, *Bglap2*, *Dmp1*, *Ibsp*, *Mef2c*, *Mepe*, *Ostn*, *Phex*, *Spp1*, *Vdr*. Of note, *Grem1* is also located within this cluster which is a known BMP-antagonist, and its increasing expression is associated with decrease in proliferation and migration of cells. Another 23/103 of the genes in Cluster C are annotated to “immune system process”.

Cluster D contains a total of 110 genes which are all annotated to a GO term. These genes have very low expression levels at day 0, but higher expression levels at the later time points. Although GO analysis did not reveal statistical significance, > 10% of these genes are annotated to “oxidoreductase activity”. Significance was obtained for a subset of these genes annotated to “respiratory chain” and “oxidoreductase complex”. Interestingly, *Tnfrsf11* (aka RANKL) is located within this cluster as well as *Mmp13* and *Arrb1*. The latter has been described to protect against ER stress and its expression is kind of inverse to cluster A, where several genes are annotated to “response to endoplasmic reticulum stress”.

There are 22 genes in cluster E which show highest expression at day 3. They are all annotated to the rather unspecific GO term “single-organism process”. However, this cluster contains *Col9a3* and *Ucma*. Both genes are involved in chondrogenesis, and *Ucma* has been speculated to be involved in negative regulation of osteoblast differentiation [25].

Cluster F contains five genes. These genes display moderate expression levels from day 0 on with peak expression at day 6 and very low expression levels at day 12. Gene products of the four genes *Fbn2*, *Hmnc1*, *Mal2*, and *Serpini1* are localized within the ECM. The fifth gene, *Lrrc75b* has been reported to be a negative regulator of myogenic differentiation [26].

Cluster G contains 102 genes of which 101 are annotated to a GO term. Here, genes display low expression levels at day 0 and peak expression at day 3. About one third of these genes are involved in lipid metabolic processes and 19/101 genes are annotated to “oxidoreductase activity”. Genes in this cluster such as *Cyp51*, *Dhcr7*, *Dhcr24*, *Fdps*, *Hmgcs1*, *Idi1*, *Msmo1*, *Mvd*, *Nsdhl*, *Pmvk*, *Sc5d*, *Sqle*, and *Tm7sf2* encode enzymes of the cholesterol biosynthetic pathway, and some of these genes play a known role in bone development as shown by knockout phenotypes. For example, developing *Cyp51* knockout embryos exhibit shortened and bowed limbs and synostosis of femur and tibia [27], *Cyp7b1* knockout animals have increased number of immature or dysfunctional osteoblasts and decreased bone mineral density [28], mutations in *DHCR24* cause severe developmental anomalies including short limbs (Desmoterolosis OMIM 602398), and *Elovl6* plays a role during growth plate development [29]. This cluster also includes *Mgp* encoding matrix Gla protein, a physiological inhibitor of ectopic tissue calcification, *Mmp10* which is involved in matrix remodeling, and *Wisp2* encoding a modulator of bone turnover.

In cluster H, we observed an incremental increase of expression levels with a peak at day 6 and moderate expression at day 12. All 116 genes in this cluster are annotated to a GO term. Products of 22/116 genes are localized within the ECM and several of the genes in this cluster are annotated to “biomineral tissue development”, “collagen binding”, and “collagen trimer”. Five members of the small leucine-rich proteoglycan family *asporin* (*Aspn*), *decorin* (*Dcn*), *epiphygan* (*Epyc*), *lumican* (*Lum*), and *osteoglycin* (*Ogn*) display peak expression at day 6. Although these genes are not directly annotated to “collagen binding”, these genes encode regulators of fibrillogenesis that interact with collagen fibrils and other ECM proteins. Osteoglycin-deficient mice for example have collagen fibril abnormalities. The genes *Mfap2*, *Mfap4*, and *Mfap5* encoding microfibrillar-associated proteins that interact with fibrillin and elastin-associated microfibrils, display a similar

expression pattern. In mice, *Mfap2* deficiency resulted in weakened bone due to increased bone resorption. *Cst3* encoding cystatin C has been suggested to promote osteoblast differentiation via BMP-signaling. However, cluster H contains the BMP-antagonist *Chrdl1*, too. Additionally, this cluster includes *Meox2* which is involved in limb morphogenesis, osteomodulin (*Omd*) where raising expression levels have been reported to mediate the switch from osteoblast proliferation to differentiation, and *Srpf2* which is a negative regulator of Wnt-signaling. Taken together, we observe high expression of genes involved in ECM maturation at day 6. Several of the encoded proteins are incorporated into the matrix and are able to regulate osteoclast activity.

Cluster I is the largest cluster, containing 377 protein-coding genes of which 373 are annotated to a GO term. Expression levels are comparably high at day 3 and day 6, but low at day 0 and day 12. About one third of these genes (125/373) are annotated to “cell cycle” including cyclins, cell division cycle (associated) genes, cyclin dependent kinases and -kinase inhibitors, centromere proteins, E2F transcription factors, and numerous genes encoding kinesin-like proteins. Genes in this cluster with known roles in bone biology include *Birc5* which is a *Comp* target gene, *Clu* (clusterin) which is a marker of zonal articular chondrocytes, *Dkk2*, *Fam111a* where mutations cause Gracile Bone Dysplasia (OMIM 602361), *Gdf10*, *Gdf11*, *Id2*, *Itm2a*, *Kazald1* aka *Bono1*, *Lgas9*, *Mmp9*, *Nell1*, *Postn*, *Ptn*, *Sfrp1*, and *S100a11* which accelerates chondrocyte hypertrophy.

Taken together, these results underscore the plethora of biological and molecular events governing proper ECM formation and osteoblast differentiation.

3.6. Highly abundant genes

To identify genes that are highly expressed during osteoblast differentiation we calculated the mean over all 12 samples. Top twenty of the most abundant genes (Supplementary Fig. S7) included mitochondrially encoded cytochrome c oxidase I (*mt-Co1*), mitochondrially encoded cytochrome b (*mt-Cytb*), and mitochondrially encoded NADH dehydrogenase 1 and 5 (*mt-Nd1*, *mt-Nd5*), which are all part of the respiratory chain. The most abundant nuclear encoded genes were ten components of the ECM such as fibronectin 1 (*Fbn1*), secreted phosphoprotein 1 (osteopontin, *Spp1*), secreted acidic cysteine rich glycoprotein (osteonection, *Sparc*), alpha-1 and alpha-2 subunits of type I collagen (*Col1a1*, *Col1a2*), thrombospondin 1 (*Thbs1*), vimentin (*Vim*), biglycan (*Bgn*), and lumican (*Lum*). Further genes encompass components of the cytoskeleton such as alpha 2 actin (*Acta2*), beta actin (*Actb*), cytoplasmic gamma actin 1 (*Actg1*), and microtubule-binding protein *Tpt1* (tumor protein, translationally-controlled 1), two eukaryotic translation elongation factors (*Eef1a1*, *Eef2*), as well as the non-coding RNA *H19* (H19, imprinted maternally expressed transcript).

3.7. Differential expression of annotated lncRNAs

We identified 285 lncRNAs with absolute fold change ≥ 1.5 and adjusted *p*-value < 0.01 in at least one comparison to be significantly differentially expressed (Supplementary Table S2). Fold change and *p*-value cut-offs were defined based on volcano plots (Supplementary Fig. S8). Within significantly differentially expressed lncRNAs, we first focused on lncRNAs with known function (Supplementary Fig. S9). The braveheart long non-coding RNA (*Bvht*) is highly expressed in the heart and was demonstrated to be a key regulator of cardiac lineage commitment and cardiac gene expression [30]. In our analyses *Bvht* had high expression levels at day 0 and day 3 and was downregulated at day 12. The cardiac mesoderm enhancer-associated non-coding RNA *Carmn* is localized on mouse chromosome 18 in close proximity of *Bvht*. *Carmn* knockdown inhibits cardiac specification and differentiation in cardiac precursor cells [31]. In our analyses *Carmn* was highly expressed at day 0 and downregulated at day 12. The differentiation antagonizing non-protein coding RNA *Dancr* is a suppressor of osteogenic differentiation

[32]. In our data *Dancr* displays low expression at day 0, higher level at day 3, an intermittent drop at day 6, and highest level at day 12. As stated above, we observed highest overall expression of lncRNA *H19*. This lncRNA has been proposed to promote skeletal muscle differentiation and to inhibit proliferation of fetal liver cells and Wnt-signaling [33,34]. During osteoblast differentiation *H19* is strongly expressed with highest levels at day 3 and day 6. The lncRNA downstream of *Cdkn1b* (*Lockd*), which has been proposed to be an enhancer of *Cdkn1b* expression [35], shows a similar expression pattern with highest expression at day 6. *Malat1* (metastasis associated lung adenocarcinoma transcript 1) has been described to accelerated hepatocyte proliferation by stimulating cell cycle progression [36]. It is strongly downregulated between day 0 and day 3 and might thus also play a role in osteoblast proliferation. *Neat1* (nuclear paraspeckle assembly transcript 1) is one of the most upregulated lncRNAs in adipogenesis [37]. During osteoblast differentiation we observed continuous down-regulation between day 0 and day 6. The lncRNA *Pvt1* is a senescence-associated circRNA, that displayed markedly reduced levels in senescent fibroblasts [38]. In our analysis, expression levels dropped more than threefold between day 0 and day 3, stayed rather unaltered at day 6, and then rose again, indicating enforced cell cycle control during matrix formation and maturation.

lncRNAs can modulate developmentally regulated genes in *cis* and in *trans*. Similarly, lncRNAs are themselves regulated by *cis* and *trans* mechanisms. They often exhibit remarkable tissue specificity and may function to fine-tune the expression of their target genes in a tissue-specific manner [39]. To pinpoint lncRNA regulation in *cis*, we determined protein-coding genes within the same TAD with highest positive and highest negative correlation for each significantly expressed lncRNA. For 100/285 lncRNAs we identified at least one correlating protein-coding gene. However, there were seven lncRNAs that had more than ten protein-coding genes with an absolute correlation index > 0.9 and *p*-value < 0.01. Within these pairs, we looked for distance and orientation of the genes. Fig. 4 gives examples of lncRNAs with positively correlating protein-coding genes. In two instances transcriptional start sites are very close to each other and genes are facing in opposite direction. These genes might share same enhancer or promoter elements. In nine instances lncRNAs were localized inside the protein-coding genes, but on opposite strand. Protein-coding gene expression levels were between 15-fold and 8000 fold of lncRNA expression levels. This pattern could be explained by polymerase cycling at DNA-sites with high transcriptional activity and open chromatin. The maximal distance between lncRNA and a single positively correlating protein-coding gene with the same TAD was 1848 kb. Here, co-expression might be explained by long-range regulatory mechanisms, similar transcription factor binding sites or DNA-loop formation.

For overlapping localization of a lncRNA with a negatively correlating protein-coding gene we observed only one example (Fig. 5A). Here, decreasing expression of the lncRNA *Gm26514* is associated with increasing expression of the protein-coding gene *Slc35b3*. The distance between lncRNAs and their best negatively correlating protein-coding gene within the same TAD ranged from 0.1 to 1725 kb. Pairs can be neighbors or be separated by other genes (Fig. 5B–D). One single lncRNA may have both a positively and a negatively correlating gene within the same TAD (Fig. 5E). We observed differential expression of the lncRNA *3222401L13Rik*, which is localized at the protocadherin cluster on chromosome 18 (TAD chr18:37280346–37960346) right between the genes of the β -cluster and the γ -cluster. These genes are regulated by promoter choice and DNA methylation [40,41]. Interestingly, several genes of the β -cluster showed positive correlation while genes of the γ -cluster displayed negative correlation suggesting that the lncRNA *3222401L13Rik* might influence gene cluster regulation. Although these observations are intriguing, functionality of these associations remains to be studied.

3.8. Analyses of ON-OFF genes

As compared to microarray-based expression analyses, RNA-seq data has a better dynamic resolution at low-level expression. An unfiltered overview over all pairwise comparisons showed that fold change distribution is symmetric (Supplementary Fig. S10). Analysis was restricted to significantly differentially expressed genes with adjusted *p*-value < 0.01 and absolute fold change ≥ 4 . We used four classifiers. A gene was defined to be a) “off”, if normalized count was 0 in at least two replicates and if the mean of normalized counts at this time point was < 1.5, to be b) “weak”, i.e. weakly expressed, if the mean of normalized counts at this time point was at least 1.5, but < 5, to be c) “on”, if the mean of normalized counts at this time point was at least 5, and to have d) a “peak”, if the mean of normalized counts was highest at this time point.

Using this approach, we identified 237 genes that were “off” in at least one time point, including 141 protein-coding genes. Of these genes, 68 are annotated to “response to stimulus”, indicating that these genes are indeed expressed in a context dependent manner, 39 are annotated to “cell differentiation”, and 14 to “enzyme regulator activity”. First, genes were divided into two groups according to expression at day 0.

3.8.1. “Peak” or “on” at day 0

In total, 67/237 genes were expressed at day 0 (Supplementary Fig. S11). The majority of them, i.e. 60 genes, had highest expression levels at day 0, and 48 genes were found to be exclusively expressed at this time point with “off” status or “weak” expression at day 3, day 6, and day 12. Products of 13/21 annotated genes are localized to the plasma membrane including three genes involved in G protein-coupled receptor signaling (*Adora2a*, *Gbp2b*, *Gpr132*), GTPase-activating protein (*Rasal1*), delta subunit of the muscle-derived nicotinic acetylcholine receptor (*Chrnd*), a netrin 1 receptor (*Dcc*), and three cation transporter (*Cacna1i*, *Slc22a14*, *Trpc3*). This group also includes *Rbm24* which is a major regulator of muscle-specific alternative splicing. Its down-regulation results in defective exon inclusion impairing the production of muscle-specific isoforms of *Coro6*, *Fxr1* and *NACA* transcripts [42]. Delta like canonical Notch ligand 4 (*Dll4*) is involved in Notch-signaling, which is known to be highly active in MSCs and osteoprogenitor cells. Three genes were expressed at day 0 and day 3 (*Itih5l-ps*, *Kcnk3*, *Lrp8os2*). *Cdr1* and *C230004F18Rik*, which is known to be expressed in limb, were expressed until day 6. The last group contains 13 genes, which are expressed at day 0 and day 12, but “off” or “weak” at day 3 and day 6. Non-coding RNAs *Gm20658* and *RP23-98C9.5*, death domain-containing *Dthd1*, transcription factors *Gbx1* and *Tox2*, *1700001J03Rik*, and the metalloproteinase *Ecel1* had “peak” expression at day 0, whereas the growth factor *Pdgfb*, the bone-derived hormone with metabolic regulatory effects *Lcn2* [43], the host gene of *miR-708* which suppresses osteogenic and adipogenic differentiation of MSCs [44], the brown adipocyte marker *Elovl3*, and non-coding RNAs *Gm27019* and *Unc45bos* had “peak” expression at day 12.

3.8.2. “Off” at day 0

The majority of genes captured by this analysis (170/237) was found to be turned “off” at day 0 (Supplementary Fig. S12). 14 of them are annotated to “immune response” (*Adcyap1*, *Blk*, *C1qc*, *Ccl11*, *Cxcl3*, *Gzmb*, *Gzmd*, *Il12a*, *Il13ra2*, *Nr1h4*, *Ppbp*, *Rag1*, *Reg3g*, *Vip*), and 12 of them to “response to biotic stimulus” (*Casp1*, *Cxcl3*, *Hist2h3c2*, *Ifit1bl2*, *Il12a*, *Nr1h4*, *Ppbp*, *Reg3g*, *Vip*, *Wfdc12*, *Wfdc15b*, *Wfdc6a*).

3.8.3. “Off” at day 0 with “peak” at day 3

Three of seven genes specifically turned “on” at day 3 encode inhibitors of endopeptidase activity (*Kng1*, *Serpina3f*, *Serpina3i*). The transglutaminase *Tgm5* encodes a protein that catalyzes formation of protein cross-links, often resulting in stabilization of protein assemblies. Transglutaminases can crosslink polymeric osteopontin and thereby

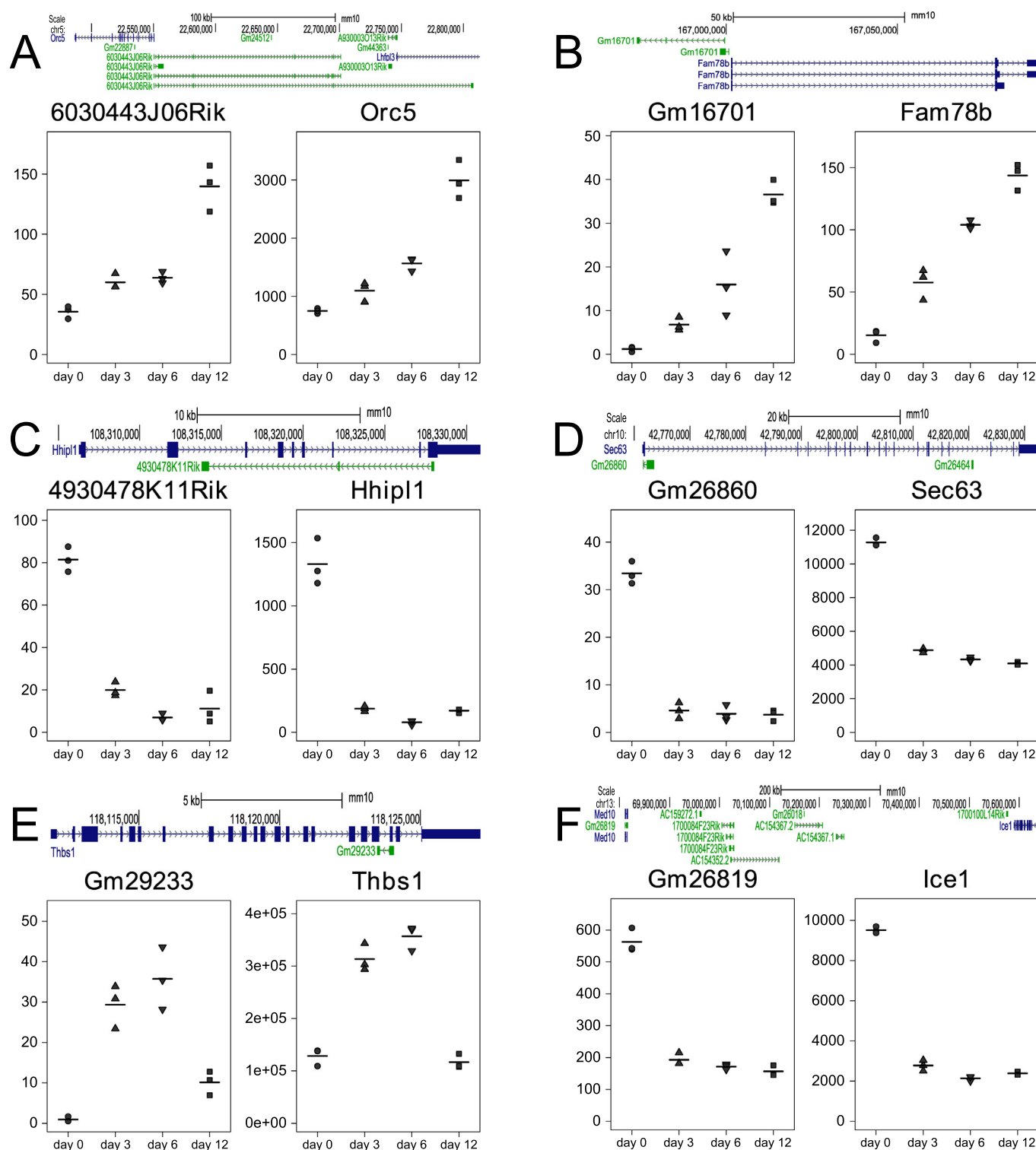


Fig. 4. Pairs of lncRNAs and positively correlating protein-coding genes. (A–F) For the plots on top, we used the UCSC genome browser. Windows were defined to include the protein-coding gene and the lincRNA of interest. Plots on bottom indicate expression levels as outlined in Fig. 2. A,B. LncRNA and best correlating protein-coding gene within the same TAD are on opposite strand and face into opposite direction. Distance of transcriptional start sites is 3 base pairs (A) and 576 base pairs (B). C–E. LncRNA and best correlating protein-coding gene within the same TAD are on opposite strand. The lncRNAs are localized between start and end of the protein-coding gene. Expression levels of protein-coding gene are about 15-fold (C), 300-fold (D), and 8000-fold (E) of lncRNA. F. Although lncRNA and best correlating protein-coding gene are localized within the same TAD, they are 821 kb away from each other.

inhibit mineralization [45]. Six genes that were “off” at day 0 with “peak” at day 3 were still expressed at day 6, but “off” at day 12. The methyltransferase *Mettl11b*, *Spo11*, a subunit of GABA-A receptor *Gabra6*, non-coding RNA *BC064078*, *Dynlt1b* which is involved in

localization of secretory vesicles, and *Slc22a8* which is involved in the sodium-independent transport and excretion of organic anions. 39 genes with “peak” expression at day 3 are “weak” or “on” until day 12. 29 of them display decreasing expression levels over time including the

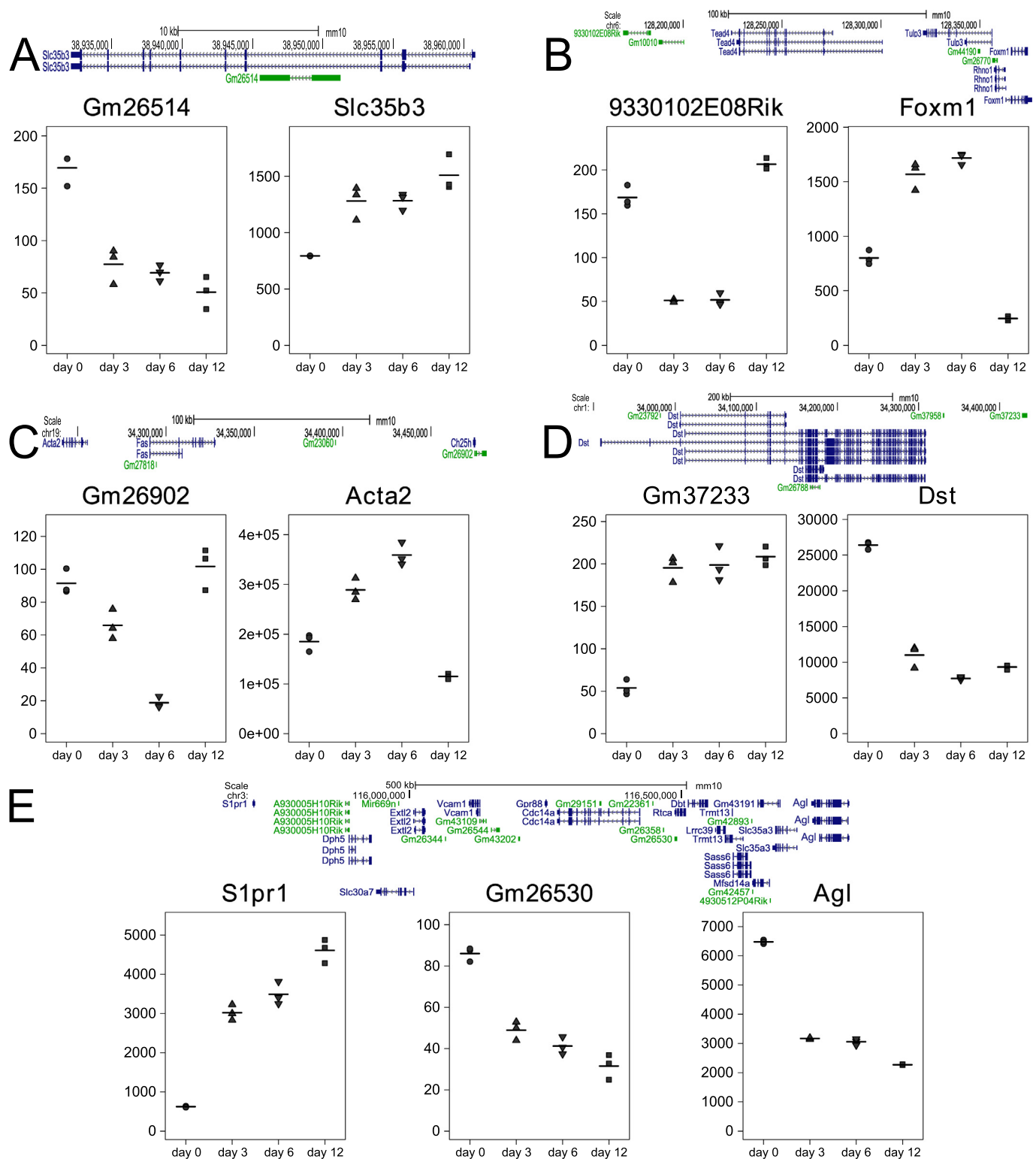


Fig. 5. Pairs of lncRNAs and negatively correlating protein-coding genes. (A–E) Plots are designed as outlined in caption of Fig. 4. **A.** lncRNA and protein-coding gene are on the same strand. Expression is negatively correlated, indicating that either the transcriptional start site of the protein-coding gene or of the lncRNA is used. **B–D.** lncRNA and negatively correlating gene can be separated by other genes. **E.** lncRNAs may have both a positively and a negatively correlating gene within the same TAD.

serine-type endopeptidase inhibitors *Spink2* and *Wfdc12*, matrix metalloproteinase *Mmp27*, as well as genes involved in regulation of proliferation and differentiation such as *Blk*, *Fabp7*, *Gjb5*, and host genes of *miR-450b* and *miR-503*. Ten genes display intermittently lower expression at day 6 including six non-coding RNAs and three regulators of

cell proliferation (*Ccl11*, *Slfn2*, *Vip*). A small group of eight genes (*Ccdc110*, *Gm44029*, *Tph2*, *Ifit1bl2*, *Gm29233*, *Gm19026*, *Gm3550*, *Ccna1*) had comparably high expression levels at day 3 and day 6.

3.8.4. “Off” at day 0 with “peak” at day 6

There were five genes specifically turned “on” at day 6 (*BC147527*, *Slitrk1*, *Rag1*, *Ly6f*, *A830012C17Rik*). Other genes that were “off” at day 0 with “peak” at day 6 were also “on” or “weak” at day 3 and day 12. 17/58 genes with “peak” at day 6 were “off” or “weak” at day 12 including flavin containing monooxygenase *Fmo3*, *Otor*, which may function in cartilage development and maintenance, genes involved in cell differentiation (*Cartpt*, *Sebox*) and eight non-coding RNAs. Further 25/58 genes with “peak” at day 6 were “on” at day 12, but displayed lower expression at that time point in comparison to day 3. These genes include flavin containing monooxygenase *Fmo4*, the transcription factors *Dmrtc1a* and *Nr1h4*, *Dazl*, *Bmp15*, the small serine proteinase inhibitor *Wfdc15b*, and 14 non-coding RNAs. 16/58 displayed higher expression at day 12 in comparison to day 3 such as proteinase inhibitors *Lrrc66*, *Wfdc18*, and *Wfdc6a*, aspartoacylase *Aspa*, *Spr3*, which is the most flexible member of the SPRR family and cross-linked by transglutamination, cadherin *Cdh26*, as well as regulators of T-cell differentiation *H2-Aa* and *Il12a*. A small group of four genes (*Lingo2*, *Atp12a*, *Adcyap1*, *1810019D21Rik*) had comparably high expression levels at day 6 and day 12.

3.8.5. “Off” at day 0 with “peak” at day 12

38 genes display prominent “peak” expression at day 12. Seven of these genes were turned “off” at day 0, day 3, and day 6 and specifically expressed at day 12 including *Sost* (Fig. 2A), which is a marker gene for terminal osteoblast differentiation, the ion transport regulator *Fxyd3*, one of the Iroquois homeobox transcription factors *Irx6*, which might similar to *Irx3* and *Irx5* regulate mineralization, and the protease *Klk4*, which is associated with terminal extracellular degradation of matrix proteins to facilitate lateral expansion of hydroxyapatite crystallites [46].

Twelve genes were “weak” or “on” either at day 3, at day 6 or “weak” at day 3 and day 6. Five of these genes are involved in immune response (*C1qc*, *Gzmb*, *Gzmd*, *Ighd*, and *Reg3g*). *Gzmb* and *Gzmd* encode proteases that process cytokines and degrade ECM proteins. *Dppa2* and *Mepe* play roles in mesenchyme and mineralized tissue development, and leiomodin *Lmod3* is involved in myofibril assembly. In total 19/38 genes were “on” at day 3 and day 6 with “peak” at day 12. These include histone encoding genes *Hist1h1b* and *Hist2h3c2*, two interferon inducible genes *Ifi47* and *Ifi205*, two genes involved in myotube differentiation (*Casp1*, *Neu2*), and the carbohydrate phosphatase *Fbp2*. Finally, there were five genes with high expression at day 3 and “peak” at day 12 or vice versa including *Cxcl3* and two interleukin receptor subunits *Il13ra2* and *Il20ra*.

3.9. Analysis of differential exon usage and alternative splicing

Cellular behavior can not only be influenced by changes in overall transcription of genes, but also by expression of different transcripts. These may result from alternative splicing and/or the use of different transcription start sites. In total, > 120 genes showed alternative expression of transcripts represented by an average exon base mean of at least 10. However, alternative transcripts containing open reading frames were only observed in 22 genes (Table S4). Most of these genes were also globally differentially expressed between any of the time points investigated. One exception, however, is the gene *Wnk1*, whose overall expression levels differ very little, but which produces primarily transcript variant *Wnk1*-223 at day 0 and *Wnk1*-202 at day 12 (Fig. 6A). *Wnk1* regulates ion homeostasis and cell volume, but also plays a role in WNT-signaling [47]. *Col9a1* is a typical cartilage gene, which shows relatively high expression at day 0. Until day 12 expression is 50-fold reduced (Fig. 3 Cluster B; Fig. 2B). If we look at exon usage, we observe that the transcript is almost completely switched to *Col9a1*-202, which lacks the N-terminal laminin G domain due to use of an alternative transcription start site around exon 6 (Fig. 6B). This short variant has been implicated in bone regeneration [48]. In contrast, overall

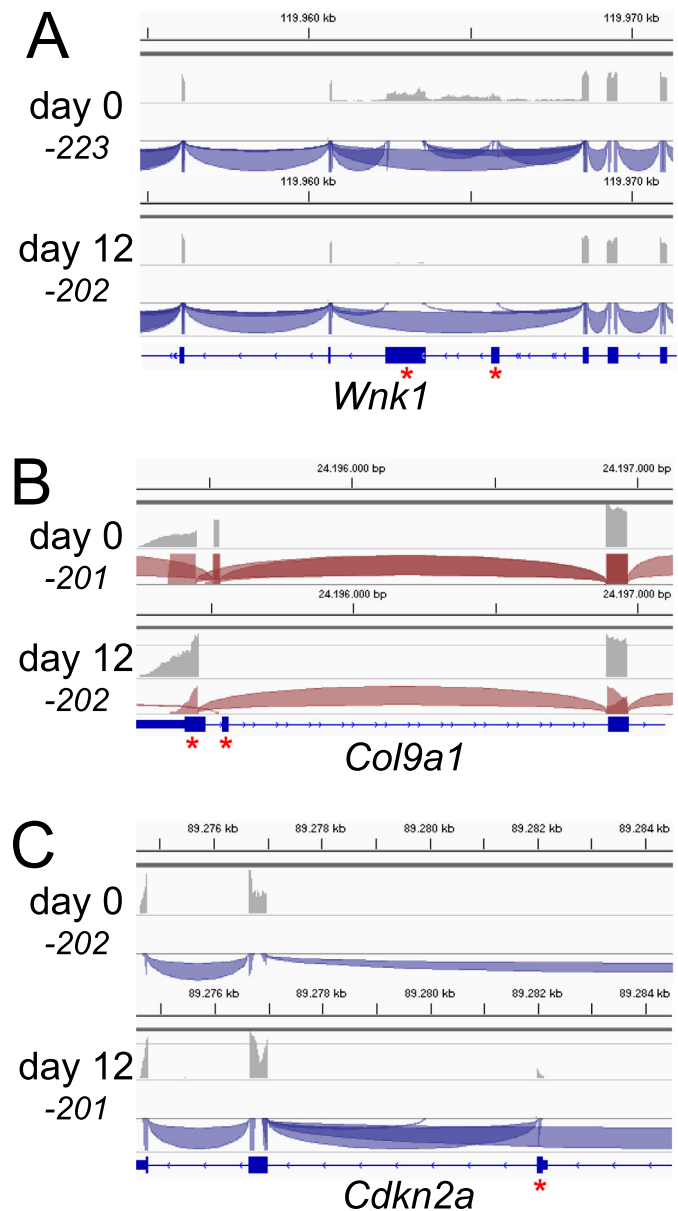


Fig. 6. Three examples of genes showing switching of transcripts with significant expression levels. Alternatively used exons are indicated by asterisks, ENSEMBL transcript names are given in italic. **A.** While at day 0 exons 9 and 10 of the *Wnk1* gene are used they are almost completely skipped at day 12. **B.** Transcript *Col9a1*-201 expressed at day 0 starts at exon 1 (not shown), but at day 12 the main transcript *Col9a1*-202 starts at an alternative exon and skips exon 6. **C.** The only transcript expressed at day 0 is *Cdkn2a*-202, which starts at exon 1 (also called 1 beta; not shown) and encodes p19. But at day 12 a significant portion of the expressed mRNA starts at exon 2 (also called 1 alpha). Transcript *Cdkn2a*-201 encodes p16.

transcription of the gene *Cdkn2a* rises five-fold from day 0 to day 12. *Cdkn2a* encodes two different transcripts that have non-overlapping reading frames resulting in the proteins p19^{ARF} and p16^{INK4a}. Both proteins block cell cycle progression through different ways [49]. This cell cycle arrest is prerequisite for efficient cell differentiation, but is also a hallmark of cellular senescence [50]. Methylation of the human *CDKN2A* locus is implicated in bone mineral density regulation [51]. At day 0 *Cdkn2a* exclusively produces the transcript variant 202, which codes for p19^{ARF}. Until day 12 the variant 201 coding for p16^{INK4a} becomes increasingly expressed in parallel to 202, thus ensuring efficient cell cycle control (Fig. 6C).

4. Conclusion

Here we present an in-depth profiling of murine primary calvaria osteoblast differentiation. We defined confluence of cell cultures as first time point, i.e. day 0, and the onset of mineralization as end point. Expression patterns of well-known marker genes suggest that our experimental procedures resulted in a valid and realistic model system for osteogenic differentiation. In addition to characterizing the expression profiles of 1386 protein-coding genes, we have identified 22 instances of differentially spliced protein-coding genes and showed, that 285 lncRNAs are significantly expressed during osteoblast differentiation. 100 lncRNAs displayed expression patterns that are highly correlated to those of one or more protein-coding genes located in the same TAD, thereby suggesting possible functions of these lncRNAs in osteogenic differentiation. Summarized results for expression profiles of coding genes (Supplementary File 1), ON-OFF genes (Supplementary File 2), lncRNA-TAD associations (Supplementary File 3) and alternative splicing (Supplementary File 4) are available in the online version of this article, and raw RNA-seq data has been made available at NCBI BioProject PRJNA416437.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bone.2018.04.006>.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grant number MU880/11-01 to JG and CEO. LAK received funding from Bundesministerium für Bildung und Forschung (BMBF) grant number 031A427C (SMART). UK and PNR received funding from FP7-EU grant agreement no. 602300 (SYBIL).

Conflict of interest

The authors state that no competing financial interests exist.

Acknowledgments

We thank Marten Jäger for technical support and his great help in developing the overall framework for bioinformatics analyses.

Author contributions

CEO, PNR, and UK conceived the study, JG and CEO performed wet lab experiments, LAK performed bioinformatics analyses and drafted parts of the manuscript, CEO and PNR supervised bioinformatics analyses, IP and SM edited the manuscript, UK analyzed exon-usage data, CEO, UK, and PNR wrote the manuscript. All authors reviewed the manuscript. CEO is the guarantor of this work, had full access to all data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

- [1] L. Polo-Corralles, M. Latorre-Esteves, J.E. Ramirez-Vick, Scaffold design for bone regeneration, *J. Nanosci. Nanotechnol.* 14 (2014) 15–56.
- [2] P. Jayakumar, L. Di Silvio, Osteoblasts in bone tissue engineering, *Proc Inst Mech Eng H* 224 (2010) 1415–1440.
- [3] A. Kalsotra, T.A. Cooper, Functional consequences of developmentally regulated alternative splicing, *Nat Rev Genet* 12 (2011) 715–729.
- [4] G. Huang, Y. Kang, Z. Huang, Z. Zhang, F. Meng, W. Chen, M. Fu, W. Liao, Z. Zhang, Identification and characterization of long non-coding RNAs in osteogenic differentiation of human adipose-derived stem cells, *Cell. Physiol. Biochem.* 42 (2017) 1037–1050.
- [5] W. Zhang, R. Dong, S. Diao, J. Du, Z. Fan, F. Wang, Differential long noncoding RNA/mRNA expression profiling and functional network analysis during osteogenic differentiation of human bone marrow mesenchymal stem cells, *Stem Cell Res Ther* 8 (2017) 30.
- [6] K.A. Staines, D. Zhu, C. Farquharson, V.E. MacRae, Identification of novel regulators of osteoblast matrix mineralization by time series transcriptional profiling, *J. Bone Miner. Metab.* 32 (2014) 240–251.
- [7] P.A. t Hoen, Y. Ariyurek, H.H. Thygesen, E. Vreugdenhil, R.H. Vossen, R.X. de Menezes, J.M. Boer, G.J. van Ommen, J.T. den Dunnen, Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms, *Nucleic Acids Res.* 36 (2008) e141.
- [8] C.M. Karner, S.Y. Lee, F. Long, Bmp induces osteoblast differentiation through both Smad4 and mTORC1 signaling, *Mol. Cell. Biol.* 37 (2017).
- [9] M.B. Meyer, N.A. Benkusky, B. Sen, J. Rubin, J.W. Pike, Epigenetic plasticity drives Adipogenic and osteogenic differentiation of marrow-derived mesenchymal stem cells, *J. Biol. Chem.* 291 (2016) 17829–17847.
- [10] H.C. St John, K.A. Bishop, M.B. Meyer, D.A. Benkusky, N. Leng, C. Kendzioriski, L.F. Bonewald, J.W. Pike, The osteoblast to osteocyte transition: epigenetic changes and response to the vitamin D3 hormone, *Mol. Endocrinol.* 28 (2014) 1150–1165.
- [11] N.A. Twine, L. Chen, C.N. Pang, M.R. Wilkins, M. Kassem, Identification of differentiation-stage specific markers that define the ex vivo osteoblastic phenotype, *Bone* 67 (2014) 23–32.
- [12] J.P. Kemp, C. Medina-Gomez, K. Estrada, B. St Pourcain, D.H. Hepp, N.M. Warrington, L. Oei, S.M. Ring, C.J. Kruitthof, N.J. Timpson, L.E. Wolber, S. Reppe, K. Gautvik, E. Grundberg, B. Ge, B. van der Eerden, J. van de Peppel, M.A. Hibbs, C.L. Ackert-Bicknell, K. Choi, D.L. Koller, M.J. Econs, F.M. Williams, T. Foroud, M.C. Zillikens, C. Ohlsson, A. Hofman, A.G. Uitterlinden, G. Davey Smith, V.W. Jaddoe, J.H. Tobias, F. Rivadeneira, D.M. Evans, Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment, *PLoS Genet.* 10 (2014) e1004423.
- [13] J. Grunhagen, R. Bhushan, E. Degenkolbe, M. Jager, P. Knaus, S. Mundlos, P.N. Robinson, C.E. Ott, MiR-497–195 cluster microRNAs regulate osteoblast differentiation by targeting BMP signaling, *J. Bone Miner. Res.* 30 (2015) 796–808.
- [14] C.E. Ott, S. Bauer, T. Manke, S. Ahrens, C. Rodelsperger, J. Grunhagen, U. Kornak, G. Duda, S. Mundlos, P.N. Robinson, Promiscuous and depolarization-induced immediate-early response genes are induced by mechanical strain of osteoblasts, *J. Bone Miner. Res.* 24 (2009) 1247–1262.
- [15] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114–2120.
- [16] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biol.* 14 (2013) R36.
- [17] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* 30 (2014) 923–930.
- [18] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (2014) 550.
- [19] S. Bauer, S. Grossmann, M. Vingron, P.N. Robinson, Ontologizer 2.0—a multi-functional tool for GO term enrichment analysis and data exploration, *Bioinformatics* 24 (2008) 1650–1651.
- [20] S. Grossmann, S. Bauer, P.N. Robinson, M. Vingron, Improved detection of over-representation of gene-ontology annotations with parent child analysis, *Bioinformatics* 23 (2007) 3024–3031.
- [21] M. Pignatelli, A.J. Vilella, M. Muffato, L. Gordon, S. White, P. Flicek, J. Herrero, ncRNA Orthologies in the Vertebrate Lineage, Database (Oxford), 2016 (2016).
- [22] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions, *Nature* 485 (2012) 376–380.
- [23] D. Karolchik, A.S. Hinrichs, T.S. Furey, K.M. Roskin, C.W. Sugnet, D. Haussler, W.J. Kent, The UCSC table browser data retrieval tool, *Nucleic Acids Res.* 32 (2004) D493–6.
- [24] S. Anders, A. Reyes, W. Huber, Detecting differential usage of exons from RNA-seq data, *Genome Res.* 22 (2012) 2008–2017.
- [25] C. Surmann-Schmitt, U. Dietz, T. Kireva, N. Adam, J. Park, A. Tagariello, P. Onnerfjord, D. Heinegard, U. Schlotzer-Schrehardt, R. Deutzmann, K. von der Mark, M. Stock, Uema, a novel secreted cartilage-specific protein with implications in osteogenesis, *J. Biol. Chem.* 283 (2008) 7082–7093.
- [26] Y. Zhong, L. Zou, Z. Wang, Y. Pan, Z. Dai, X. Liu, L. Cui, C. Zuo, Lrrc75b is a novel negative regulator of C2C12 myogenic differentiation, *Int. J. Mol. Med.* 38 (2016) 1411–1418.
- [27] R. Keber, H. Motain, K.D. Wagner, N. Debeljak, M. Rassoulzadegan, J. Acimovic, D. Rozman, S. Horvat, Mouse knockout of the cholesterologenic cytochrome P450 lanosterol 14alpha-demethylase (Cyp51) resembles Antley-Bixler syndrome, *J. Biol. Chem.* 286 (2011) 29086–29097.
- [28] C.D. DuSelle, E.R. Nelson, X. Wang, J. Abdo, U.I. Modder, M. Umetani, D. Gesty-Palmer, N.B. Javitt, S. Khosla, D.P. McDonnell, The endogenous selective estrogen receptor modulator 27-hydroxycholesterol is a negative regulator of bone homeostasis, *Endocrinology* 151 (2010) 3675–3685.
- [29] M. Kikuchi, M. Shimada, T. Matsuzaka, K. Ishii, Y. Nakagawa, M. Takayanagi, N. Yamada, H. Shimano, Crucial role of Elov16 in chondrocyte growth and differentiation during growth plate development in mice, *PLoS One* 11 (2016) e0159375.
- [30] C.A. Klattenhoff, J.C. Scheuermann, L.E. Surface, R.K. Bradley, P.A. Fields, M.L. Steinhauser, H. Ding, V.L. Butty, L. Torrey, S. Haas, R. Abo, M. Tabebordbar, R.T. Lee, C.B. Burge, L.A. Boyer, Braveheart, a long noncoding RNA required for cardiovascular lineage commitment, *Cell* 152 (2013) 570–583.
- [31] S. Ounzain, R. Micheletti, C. Arnan, I. Plaisance, D. Cecchi, B. Schroen, F. Reverter, M. Alexanian, C. Gonzales, S.Y. Ng, G. Bussotti, I. Pezzuto, C. Notredame, S. Heymans, R. Guigo, R. Johnson, T. Pedrazzini, CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis, *J. Mol. Cell. Cardiol.* 89 (2015) 98–112.
- [32] L. Zhu, P.C. Xu, Downregulated lncRNA-ANCRC promotes osteoblast differentiation by targeting EZH2 and regulating Runx2 expression, *Biochem. Biophys. Res.*

- Commun. 432 (2013) 612–617.
- [33] B.K. Dey, K. Pfeifer, A. Dutta, The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration, *Genes Dev.* 28 (2014) 491–501.
- [34] S. Wang, X. Wu, Y. Liu, J. Yuan, F. Yang, J. Huang, Q. Meng, C. Zhou, F. Liu, J. Ma, S. Sun, J. Zheng, F. Wang, Long noncoding RNA H19 inhibits the proliferation of fetal liver cells and the Wnt signaling pathway, *FEBS Lett.* 590 (2016) 559–570.
- [35] V.R. Paralkar, C.C. Taborda, P. Huang, Y. Yao, A.V. Kossenkova, R. Prasad, J. Luan, J.O. Davies, J.R. Hughes, R.C. Hardison, G.A. Blobel, M.J. Weiss, Unlinking an lncRNA from its associated cis element, *Mol. Cell* 62 (2016) 104–110.
- [36] C. Li, L. Chang, Z. Chen, Z. Liu, Y. Wang, Q. Ye, The role of lncRNA MALAT1 in the regulation of hepatocyte proliferation during liver regeneration, *Int. J. Mol. Med.* 39 (2017) 347–356.
- [37] L. Sun, L.A. Goff, C. Trapnell, R. Alexander, K.A. Lo, E. Hacisuleyman, M. Sauvageau, B. Tazon-Vega, D.R. Kelley, D.G. Hendrickson, B. Yuan, M. Kellis, H.F. Lodish, J.L. Rinn, Long noncoding RNAs regulate adipogenesis, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 3387–3392.
- [38] A.C. Panda, I. Grammatikakis, K.M. Kim, S. De, J.L. Martindale, R. Munk, X. Yang, K. Abdelmohsen, M. Gorospe, Identification of senescence-associated circular RNAs (SAC-RNAs) reveals senescence suppressor CircPVT1, *Nucleic Acids Res.* 45 (2017) 4021–4035.
- [39] J.D. Ransohoff, Y. Wei, P.A. Khavari, The functions and unique features of long intergenic non-coding RNA, *Nat. Rev. Mol. Cell Biol.* 19 (3) (2018) 143–157.
- [40] M. Kawaguchi, T. Toyama, R. Kaneko, T. Hirayama, Y. Kawamura, T. Yagi, Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin- α gene cluster, *J. Biol. Chem.* 283 (2008) 12064–12075.
- [41] B. Tasic, C.E. Nabholz, K.K. Baldwin, Y. Kim, E.H. Rueckert, S.A. Ribich, P. Cramer, Q. Wu, R. Axel, T. Maniatis, Promoter choice determines splice site selection in protocadherin α and γ pre-mRNA splicing, *Mol. Cell* 10 (2002) 21–33.
- [42] B. Cardinali, M. Cappella, C. Provenzano, J.M. Garcia-Manteiga, D. Lazarevic, D. Cittaro, F. Martelli, G. Falcone, MicroRNA-222 regulates muscle alternative splicing through Rbm24 during differentiation of skeletal muscle cells, *Cell Death Dis.* 7 (2016) e2086.
- [43] I. Mosialou, S. Shikhel, J.M. Liu, A. Maurizi, N. Luo, Z. He, Y. Huang, H. Zong, R.A. Friedman, J. Barasch, P. Lanzano, L. Deng, R.L. Leibel, M. Rubin, T. Nickolas, W. Chung, L.M. Zeltser, K.W. Williams, J.E. Pessin, S. Kousteni, MC4R-dependent suppression of appetite by bone-derived lipocalin 2, *Nature* 543 (2017) 385–390.
- [44] C. Hao, S. Yang, W. Xu, J.K. Shen, S. Ye, X. Liu, Z. Dong, B. Xiao, Y. Feng, MiR-708 promotes steroid-induced osteonecrosis of femoral head, suppresses osteogenic differentiation by targeting SMAD3, *Sci. Rep.* 6 (2016) 22599.
- [45] B. Hoac, V. Nelea, W. Jiang, M.T. Kaartinen, M.D. McKee, Mineralization-inhibiting effects of transglutaminase-crosslinked polymeric osteopontin, *Bone* 101 (2017) 37–48.
- [46] C.E. Smith, A.S. Richardson, Y. Hu, J.D. Bartlett, J.C. Hu, J.P. Simmer, Effect of kallikrein 4 loss on enamel mineralization: comparison with mice lacking matrix metalloproteinase 20, *J. Biol. Chem.* 286 (2011) 18149–18160.
- [47] A.R. Rodan, A. Jenny, WNK kinases in development and disease, *Curr. Top. Dev. Biol.* 123 (2017) 1–47.
- [48] K. Ting, H. Ramachandran, K.S. Chung, N. Shah-Hosseini, B.R. Olsen, I. Nishimura, A short isoform of Col9a1 supports alveolar bone repair, *Am. J. Pathol.* 155 (1999) 1993–1999.
- [49] L. Chin, J. Pomerantz, R.A. DePinho, The INK4a/ARF tumor suppressor: one gene—two products—two pathways, *Trends Biochem. Sci.* 23 (1998) 291–296.
- [50] D. Munoz-Espin, M. Serrano, Cellular senescence: from physiology to pathology, *Nat. Rev. Mol. Cell Biol.* 15 (2014) 482–496.
- [51] E.M. Curtis, R. Murray, P. Titcombe, E. Cook, R. Clarke-Harris, P. Costello, E. Garratt, J.D. Holbrook, S. Barton, H. Inskip, K.M. Godfrey, C.G. Bell, C. Cooper, K.A. Lillycrop, N.C. Harvey, Perinatal DNA methylation at CDKN2A is associated with offspring bone mass: findings from the Southampton Women's survey, *J. Bone Miner. Res.* 32 (2017) 2030–2040.