Single Cell Cortical Bone Transcriptomics Defines Novel Osteolineage Gene Sets Altered in Chronic Kidney Disease

Rafiou Agoro[1], Intawat Nookaew[2], Megan L. Noonan[1], Yamil G. Marambio[1], Sheng Liu[1], Wennan Chang[1,4], Hongyu Gao[1,3], Daniel Horan[5], William R. Thompson[5], Xiaoling Xuei[1], Yunlong Liu[1,3], Chi Zhang[1,4], Alexander G. Robling[5], Lynda F. Bonewald[3,6], Jun Wan[1] and Kenneth E. White[1,7,*]

[1]Department of Medical and Molecular Genetics, Indianapolis, IN 46202, USA
[2]Department of Biomedical Informatics, University of Arkansas for Medical Sciences, AR 72205, USA
[3]Center for Medical Genomics, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[4]Department of Electrical and Computer Engineering, Purdue University, Indianapolis, IN 46202, USA
[5]Department of Anatomy, Cell Biology and Physiology, Indianapolis, IN 46202, USA
[6]Indiana Center for Musculoskeletal Health, Indiana University, Indianapolis, IN 46202, USA
[7]Department of Medicine/Nephrology; Indiana University School of Medicine, Indianapolis, IN 46202

Running title: Osteoblast and osteocyte phenotypes at the single cell level

Keywords: Bone, scRNAseq, Osteoblast, Osteocyte, Chronic Kidney Disease

*Corresponding author information:
Kenneth E. White, Ph.D.
Indiana University School of Medicine
Department of Medical and Molecular Genetics
(Mail) 975 West Walnut St., IB130
(Office) 635 Barnhill Dr., MS5010
Indianapolis, IN 46202
317-278-1775 p
317-274-2293
kenewhit@iu.edu

Abstract

Due to a lack of spatial-temporal resolution at the single cell level, the etiologies of the bone dysfunction caused by diseases such as normal aging, osteoporosis, and the metabolic bone disease associated with chronic kidney disease (CKD) remain largely unknown. To this end, flow cytometry and scRNAseq were performed on long bone cells from Sost-cre/Ai9$^+$ mice, and pure osteolineage transcriptomes were identified, including novel osteocyte-specific gene sets. Clustering analysis isolated osteoblast precursors that expressed *Tnc*, *Mmp13,* and *Spp1,* and a mature osteoblast population defined by *Smpd3*, *Col1a1*, and *Col11a1*. Osteocytes were demarcated by *Cd109*, *Ptprz1, Ramp1, Bambi, Adamts14*, *Spns2, Bmp2*, *Wasl,* and *Phex*. We validated our *in vivo* scRNAseq using integrative *in vitro* promoter occupancy via ATACseq coupled with transcriptomic analyses of a conditional, temporally differentiated MSC cell line. Further, trajectory analyses predicted osteoblast-to-osteocyte transitions via defined pathways associated with a distinct metabolic shift as determined by single-cell flux estimation analysis (scFEA). Using the adenine mouse model of CKD, at a time point prior to major skeletal alterations, we found that gene expression within all stages of the osteolineage was disturbed. In sum, we defined distinct populations of osteoblasts/osteocytes at the single cell level. Using this roadmap of gene assembly, we demonstrated unrealized molecular defects across multiple bone cell populations in a mouse model of CKD, and our collective results suggest a potentially earlier and more broad bone pathology in this disease than previously recognized.

Significance Statement

Understanding the cellular dynamics controlling bone formation and function is necessary for developing new therapeutic approaches. We leveraged flow cytometry and scRNAseq from Sost-cre/Ai9$^+$ mice to identify cortical bone osteoblast precursor, mature osteoblast and osteocyte transcriptomes. Novel osteolineage gene expression was validated in cortical bone RNA as well as using *in vitro* promoter occupancy and transcriptomic analyses. Single-cell flux estimation analysis (scFEA) showed metabolic heterogeneity between osteoblast precursors and osteoblasts/osteocytes. Using a CKD mouse model, gene alterations were found in cells at all stages of the osteolineage, suggesting a simultaneous and wider effect of CKD on bone than previously understood. Thus, our studies underscore the concept that targeting the activity of multiple bone cell types and differentiation stages may be necessary to improve the musculoskeletal disease associated with CKD.

Introduction

Osteocytes are derived from osteoblasts through a dynamic and spatio-temporal process regulated within cortical bone [1]. The exposure of these cells to the bloodstream through dendrites allows osteocytes to play key endocrine functions, including sending and receiving signals to vascularized organs to the kidney, among others. A complete identification of the osteocyte and osteocyte precursor transcriptomes is a critical need towards understanding the molecular pathways of osteolineage differentiation, as well as to specify the contributions of osteoblast/osteocyte genes to musculoskeletal disease.

Common bone pathologies are associated with osteocyte dysfunction, including aging [2], osteoporosis [3], and chronic kidney disease (CKD) [4]. CKD is a worldwide public health problem with an estimated 5-10 million lives lost each year [5]. Endocrine changes due to loss of renal function lead to increased bone production of fibroblast growth factor-23 (FGF23), which acts on the kidney to reduce circulating 1,25$(OH)_2$ vitamin D (1,25D) concentrations. These actions in turn cause hypocalcemia-mediated compensatory elevations in PTH and subsequent bone loss due to increased osteoclastic activity [6]. Indeed, fracture is the most important clinical outcome of the CKD bone disorder, with an estimation that CKD patient hip and spine fracture rates range between 2- to 4- fold greater than the general population [7], markedly increasing patient morbidity and mortality [8].

Several studies have provided initial insight into novel bone cell bioactivity using the analysis of bulk sequencing of cortical bone [9] and single cell analyses of combined cortical bone and calvaria cells [10]. These approaches have led to an increased understanding of transitional states of the skeletal cell populations in diseases typically associated with progressive loss of bone mass, including osteoporosis and aging [11]. Additionally, the previous studies identified novel osteoblast and osteocyte genes, providing key insight into new roles of these loci for a potentially broader set of bone diseases. However, the full spectrum of osteoblast and osteocyte

genes in bone cell populations specifically affected during CKD remain understudied, hampering the ability to target cell lineages towards improved bone health and patient outcomes. Thus, the dynamic and spatial contribution of osteolineage cells affected during CKD as well as the molecular mechanisms and spatial transcriptional reprogramming that contribute to bone loss remain to be determined.

Herein, we developed a successful workflow to enrich cortical bone osteoblast and osteocyte cell populations using flow cytometry followed by single cell RNAseq (scRNAseq). Our transcriptomic data detected two osteoblast populations, as well as novel osteocyte-specific genes, and also revealed cell-specific metabolic profiles. This approach identified previously unrecognized transcriptional reprogramming prior to major bone ultrastructural changes in a mouse model of CKD. Our collective results suggest that the endocrine bone disease arising from CKD affects multiple cell populations in parallel and thus emerging treatments for CKD may be more effective by targeting a broader set of cell lineages.

Methods

Sost-CreERT2-Cre+/Ai9 mice: Animal protocols and studies were approved by the Indiana University Institutional Animal Care Committee and conform to NIH guidelines [12]. Male and female Sost-CreERT2-Cre+ mice were crossed with the 'Ai9' mouse in which the recombination event causes activation of tdTomato fluorescent protein to generate Sost-CreERT2 Cre+/Ai9 [13]. Both males and females were included in the scRNAseq study. For the sequencing experiments, six mice (eight weeks of age) were sacrificed, and long bones (femurs, tibiae, and humeri) were collected and pooled, followed by a sequential bone digestion (see below).

Preparation of mouse long bones for osteoblast/osteocyte isolation: Isolated long bones were placed in 100 mm petri dishes containing αMEM with 10% penicillin and streptomycin. Any

remaining muscle and connective tissue from the bones were removed and the periosteum was scraped away using a scalpel. The bones were then washed, epiphyses removed, and the marrow flushed using a 27G syringe. Further, the bones were cut in half lengthwise and then into 1 to 2 mm lengths using a scalpel and briefly washed with Hank's Balanced Salt Solution (HBSS; Hyclone).

Bone digestion and isolation of osteoblasts and osteocytes: The long bone pieces were digested as described previously [14]. In brief, bones were immersed in warmed collagenase type IA (2 mg/mL; Sigma) and EDTA (5 mM) alternately and incubated at 37°C for 25 min on a rotating shaker (~200 rpm). A total of 9 digestions was performed. The extracted bone cells were then centrifuged at 600 rpm for 15 minutes at 4°C, washed 3 times in αMEM supplemented with 10% FBS and the pellet resuspended in αMEM supplemented with 10% FBS. TdTomato^{+} cells were then sorted using Fluorescence-activated cell sorting (FACS) Aria Fusion (BD Biosciences) with standard Phycoerythrin (PE) Texas Red gating.

CKD mouse models: Male and female C57BL/6J mice ware purchased from the Jackson Laboratory and housed in the Indiana University School of Medicine Laboratory Animal Resource Center (LARC) for at least 1 week for acclimatization prior to the start of experiments, according to previous protocols [15]. The adenine diet model was used to induce CKD in male and female mice. At 8 weeks of age, mice were fed a control casein-based diet (0.9% phosphate and 0.6% calcium, TD.150303 Envigo) for 2 weeks, or CKD-causing diet with 0.2% added adenine (TD.160020; Envigo) for 2 or 4 weeks. Diets and water were provided ad libitum.

Micro-computed tomography (μCT): Paraformaldehyde fixed femora from CKD and healthy mice were scanned, reconstructed, and analyzed as previously described [16]. Femurs were scanned at 10-μm resolution, 55-kV peak tube potential and 8W. Standard output parameters related to cortical bone mass, geometry, and architecture were measured as reported [17].

Single cell library preparation: We applied a single cell master mix with lysis buffer and reverse transcription reagents according to the Chromium Single Cell 3' Reagent Kits V3 User Guide, CG000183 Rev A (10X Genomics, Inc.). This was followed with cDNA synthesis and library preparation according to standard 10X Genomics methods; all libraries were sequenced on an Illumina NovaSeq6000 platform in paired-end mode (28bp + 91bp).

Data processing/bioinformatic analyses: The 10x Genomics Cellranger (v. 6.0.0) pipeline was used to demultiplex raw base call files to FASTQ files and reads aligned to the mm10 murine genome using STAR [18]. The Cellranger computational output was then analyzed in R (v 4.0.2) using the Seurat package v. 4.1.0 4 [19]. Seurat objects were created, and the top principal components were used to perform unsupervised clustering analysis and visualized using UMAP dimensionality reduction. Using the Seurat package, annotation and grouping of clusters by cell type was performed manually by inspection of differentially expressed genes using the MAST method [20] for each cluster, based on canonical marker genes in the literature. The selected markers were visualized on UMAP coordinates as gene expression density using R package Nebulosa [21]. To perform the pseudotime analysis on the integrated Seurat object, cells were divided into individual gene expression data files organized by previously defined cell types. R package Monocle v3 [22] was used for dataset analysis. Outputs were obtained detailing the pseudotime cell distributions for each cell type. Positional information for the Monocle plot was used to subset and color cells for downstream analyses [22]. Ingenuity Pathway Analysis (IPA) was used to predict the statistically significant canonical pathways regulated in osteoblasts and osteocytes.

scFEA: The python package v1.2 of single cell flux estimation analysis (scFEA) was applied to estimate cell-wise metabolic flux rate against the whole human metabolic map using the generated mouse scRNAseq data [23]. Default parameters were utilized. Statistical significance of the differences in metabolic flux between cell groups was assessed by Mann-Whitney test.

Culture of the MPC cell line: Murine progenitor cells clone 2 (MPC2) [24], a conditionally immortalized mesenchymal stem cell line, was cultured in αMEM (Invitrogen, Thermo-Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; Hyclone), 25 mM L-glutamine, and 25 mM penicillin-streptomycin (Sigma-Aldrich, St. Louis, MO, USA) at 33°C and 5% CO_2 to proliferate. Cells were plated at a density of 1.0×10^5 cells per well in 6-well plates and incubated overnight before being transferred to a 37°C incubator for osteogenic differentiation, by culturing in maintenance media supplemented with 4 mM beta-glycerophosphate and 50 μg/mL ascorbic acid. Cells were differentiated for 0-4 weeks with this 'osteogenic media' which was changed every 2-3 days.

Bulk mRNA sequencing: MPC2 cells were differentiated for 3 weeks in osteogenic media or plated in an undifferentiated state at 33°C (see cell culture methods above). Total RNA was extracted (with a DNase step) using the same methods as described in "RNA preparation" section in Methods (see below). The RNA was evaluated for its quantity and quality using Agilent Bioanalyzer 2100; 100 ng of total RNA was used for the cDNA libraries. Library preparation included mRNA purification/enrichment, RNA fragmentation, cDNA synthesis, ligation of index adaptors, and amplification, following the KAPA mRNA Hyper Prep Kit Technical Data Sheet, KR1352 – v4.17 (Roche Corporate). Each resulting indexed library was quantified and its quality accessed by Qubit and Agilent Bioanalyzer, and multiple libraries pooled in equal molarity. The pooled libraries were denatured and neutralized before loading to NovaSeq 6000 sequencer at 300 pM final concentration for 100b paired-end sequencing (Illumina, Inc.). Approximately 30-40M reads per library were generated. A Phred quality score (Q score) was used to measure the quality of sequencing. More than 90% of the sequencing reads reached Q30 (99.9% base call accuracy). The sequencing data were first assessed using FastQC (Babraham Bioinformatics, Cambridge, UK) for quality control. The reads were mapped to the mouse genome mm10 using STAR (v2.7.2a) [18]. RNAseq aligner with the following parameter: "--outSAMmapqUnique 60"

Uniquely mapped sequencing reads were assigned to Gencode M22 gene using featureCounts (v1.6.2) [25] with the following parameters: "–p –Q 10 -O". The genes were kept for further analysis if their read counts > 10 in at least 3 of the samples, followed by the normalization using TMM (trimmed mean of M values) method and subjected to differential expression analysis using edgeR (v3.24.3) [26]. Gene Ontology and KEGG pathway functional enrichment analysis was performed on selected gene sets, e.g., genes undergoing both significantly differential expressions and notable changes of open chromatin accessibilities, with the cut-off of false discovery rate (FDR) < 0.05 using DAVID [27]. Canonical pathways from RNAseq data were generated through the use of IPA (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) [28]. N=3 samples per condition.

Assay for Transposase-Accessible Chromatin sequencing (ATACseq): Cells used for ATACseq were plated, differentiated, and treated identical to samples for RNAseq (see above). Cells were washed twice in 1X PBS, then dissociated with trypsin (Hyclone) for 5 minutes. Cells were resuspended in ice cold 1X PBS, dead cells were removed and processed for ATAC; high-throughput sequencing was performed according to a published protocol [29]. Briefly, cells were collected in cold PBS and cell membranes were disrupted in cold lysis buffer (10 mM Tris–HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2 and 0.1% IGEPAL CA-630). The nuclei were pelleted and resuspended in Tn5 enzyme and transposase buffer (Illumina Nextera® DNA library preparation kit, FC-121-1030). The Nextera libraries were amplified using the Nextera® PCR master mix and KAPA biosystems HiFi hotstart readymix successively. AMPure XP beads (Beckman Coulter) were used to purify the transposed DNA and the amplified PCR products. All libraries were sequenced in a 100 cycle paired-end run on an Illumina NOVAseq instrument. The resulting ATACseq libraries were sequenced on Illumina NovaSeq 6000 at the Center of Medical Genomics (CMG), Indiana University School of Medicine, and paired-end 50 bp reads were generated.

Illumina adapter sequences and low-quality base calls were trimmed off the paired-end reads with Trim Galore v0.4.3. Bowtie2 [30] was used for ATACseq reads alignment on the mouse genome (mm10). Duplicated reads were removed using Picard developed by Broad Institute [https://broadinstitute.github.io/picard/ (Accessed: 2018/02/21; version 2.17.8)]. Low mapping quality reads and mitochondrial reads were discarded in further analysis. Peak calling of mapped ATACseq reads were performed by MACS2 [31] with a Bonferroni adjusted cutoff of p-value less than 0.01. Peaks called from multiple samples were merged, after removing peaks overlapping with ENCODE blacklist regions [32, 33]. Reads locating within merged regions in different samples were counted by pyDNase [34]. The data was filtered using at least 10 cut counts in more than one of the samples, then normalized using TMM (trimmed mean of M values) method and subjected to differential analysis using edgeR (v3.24.3) [26, 35].

RNA isolation and qPCR: Bones were harvested and homogenized in 1 ml of TRIzol reagent (Invitrogen) according to the manufacturer's protocol using a Bullet Blender (Next Advance, Inc.), then further purified using the RNeasy Kit (Qiagen). RNA samples were tested with intron-spanning primers specific for mRNAs expression of gene listed. Mouse *β-actin* was used as an internal control for RT-qPCR. The qPCR primers were purchased as pre-optimized reagents (Applied Biosystems/Life Technologies, Inc.) and the TaqMan One-Step RT-PCR kit was used to perform all reactions. PCR conditions were: 30 minutes 48°C, 10 minutes 95°C, followed by 40 cycles of 15 seconds 95°C and 1 minute 60°C. The data were collected and analyzed by a StepOne Plus system (Applied Biosystems/Life Technologies, Inc.). The expression levels of mRNAs were calculated relative to appropriate controls, and the $2^{-\Delta\Delta CT}$ method described by Livak was used to analyze the data [36]. The references of the primers used in the study are below. Fgf23, Mm00445621_m1; Mmp13, Mm00439491_m1; Tnc, Mm00495662_m1; Gdpd2, Mm00469948_m1; Serping1, Mm00437835_m1; Igfbp5, Mm00516037_m1; Col1a1, Mm00801666_g1; Bglap, Mm03413826_mH; Cthrc1, Mm01163611_m1; Smpd3,

Mm00491359_m1; Dmp1, Mm01208363_m1; Sost, Mm04208528_m1; Pdpn, Mm01348912_g1; Phex, Mm00448119_m1; Ptprz1, Mm00478486_m1; Adamts14, Mm01169225_gH; Pdpn(E11), Mm01348912_g1; Actin, Mm02619580_g1.

Statistical analyses: The most recent updated R software packages with robust affiliated statistics were used to analyze the scRNAseq datasets. The cutoffs used for integration of RNA-seq and ATAC-seq are: peaks withing 10kb upstream of gene, differential peak cutoff FDR < 0.05, differential expression cutoff FDR < 0.05, $\log_2$FC > 1 (up-regulation) or < -1 (down-regulation). Statistical analyses of the *in vivo* data presented were performed by two-way ANOVA to assess the differences between the same gender in response to adenine diet to assess the differences between genders and treatments. Significant changes were considered when at $P<0.05$.

Results

Mouse osteoblast/osteocyte transcriptomic profiling at single cell resolution

The osteocyte transcriptomic profiling at the single cell level solely from long bones remains uncharacterized. To identify cortical bone cells and enrich our cell sample preparation, we used Sclerostin (Sost)-cre/Ai9 'Tomato' reporter mice at 8 weeks of age to isolate fluorescently-labeled osteoblasts/osteocytes. Consistent with previous characterization [13], SOST-cre/Ai9 ('tdTomato' reporter) mice had detectable basal Cre activation in osteocytes as evidenced by the presence of red fluorescence in mouse tail vertebrae (**Figure 1A**). Following cortical bone digestion using collagenase/EDTA, the cells were sorted, and the Ai9/tdTomato$^+$ fraction represented approximately 1.5% of total cells (**Figure 1B**). As determined by qPCR, the Ai9/tdTomato$^+$ cells were highly enriched for osteocyte marker mRNAs *Fgf23* (40-fold), *Phex* (102-fold), *Dmp1* (172-fold), and *Pdpn* (12-fold) compared to the Ai9/tdTomato$^-$ population (**Figure 1B**).

To isolate the transcriptome of the Ai9/tdTomato$^+$ population, we performed scRNAseq. Approximately 1,200 cells were successfully recovered after sequencing with an estimation of 130,000 reads per cell and 97% valid barcodes. Further stringent bioinformatic analyses were performed to exclude low-quality cells; for downstream analysis, we retained only cells that had less than 3500 genes to exclude doublets and less than 15% of mitochondrial genes to exclude potential dead cells. To aggregate cells based on their similarities, principal component analysis (PCA) was performed on highly variable genes and the result was used as input for clustering using Louvain algorithm with multilevel refinement and Uniform Manifold Approximation and Projection (UMAP) for dimension reduction (**Figure 1C**). To identify gene specific markers in each cluster, we used the function FindMarkersAll from the MAST method algorithm. The quality of clustering was assessed by the dot plot of the most significant marker genes obtained from an individual cluster (**Figure 1D**).

The transcriptomic profiling of Ai9/tdTomato$^+$ cortical bone cells identified a total of six cell populations, including two osteoblast clusters and one osteocyte cluster (**Figures 1C and 1D**). Marker analysis revealed that within the osteoblast population, one cell type was characterized by higher expression of *Smpd3*, *Bglap, Col1a1*, and *Col11a1* mRNAs. The second, referred to as "Tnc/Mmp13 osteoblast," was defined by a high expression of *Tnc*, *Mmp13, Serping1 and Spp1*, which was consistent with previous subpopulation analyses [10]. Osteocytes were associated with a cluster showing the highest expression of *Phex* and *Dmp1*, and other genes that defined osteocytes were *Cd109*, *Dkk1*, and *Ptprz1* (**Figure 1D**). Our analyses also identified two distinct populations of hematopoietic cells. The first population annotated "Hematopoietic I" showed high expression of *Cybb*, *Chil3*, and *Ltf* whereas the "Hematopoietic II" cells were characterized by expression of *Clec4d*, *Il1r2*, and *Cxcl2*. Bone marrow derived endothelial cells ('BMEC') were also identified and distinctly defined by *Cdh5*, *Plvap*, and *Ptprb* gene transcripts (**Figure 1D**). As

expected, *Col1a1* was identified only in osteolineage cells (**Figure 1E**), underscoring the ability to identify transcripts that may be uniquely associated with ossification.

Our UMAP analysis confirmed that the Tnc/Mmp13 osteoblast population clustered farther from osteocytes when compared to the dimensional separation between the defined more mature osteoblasts (high *Bglap* expression (**Figure 1E**)) and osteocytes (high *Phex* expression (**Figure 1E**)). These data suggest that the Tnc-Mmp13 osteoblasts represented a precursor osteoblast population. Importantly, we isolated 25-highly and preferentially expressed genes in osteocytes. Among the genes which were not previously identified as osteocyte-specific based upon the recent proposed bulk RNAseq-based osteocyte transcriptome [9] were *Ramp1, Ptprz1, Ptgis1, Plpp1, Pdgfa, Cd109, Bmp4, Bambi, Wasl, Spns2, Myo1b, Gm41724, Ccn4, Ccdc194, Car12, Bmp2,* and *Adamts14*. The polar plot comparison of recently reported osteocyte transcriptomes highlights the genes in common (*Dkk1*, *Dmp1*, *Irx5, Col24a1, Ackr3*; from our dataset (scRNAseq of long bone), Wang and colleagues (scRNAseq on long bone and calvaria) [10] and Youlten and colleagues (bulk RNAseq on long bone) [9] (**Figure 1F**).

To further test whether the genes identified as highly expressed in osteolineage cells in our scRNAseq analyses were not derived from bone marrow cells, we compared the mRNA expression of these genes in mouse cortical long bone (osteoblast/osteocyte-enriched fraction) *versus* bone marrow. In this regard, *Mmp13*, *Cthrc1*, *Smpd3*, *Ptprz1*, and *Cd109* were highly expressed in cortical bone suggesting that the identified genes were preferentially expressed in osteolineage cells (**Figure 1G**). The erythropoietin receptor (*Epor)* mRNA, tested as a marrow positive control, was not increased in cortical bone when compared to bone marrow expression levels, whereas defined osteocyte genes *Phex, Pdpn, Dmp1, Fgf23,* and *Sost* were all increased in cortical bone samples (**Figure 1G**). To determine the cellular differentiation patterns that govern the maturation of osteoblasts to osteocytes, we performed pseudotime analysis. These analyses

confirmed that the osteocyte cluster of cells identified from our dataset derived from the osteoblast subset, thus identifying the Tnc/Mmp13 osteoblast as a likely precursor population (**Figure 1H**).

scRNAseq identifies pathways involved in the osteoblast to osteocyte transition

To identify the molecular pathways involved in the transition from osteoblasts to osteocytes at single cell resolution, we analyzed differentially expressed osteoblast/osteocyte genes using Ingenuity Pathway Analysis (IPA). Further, a comparison analysis was performed to identify signaling pathways in osteoblasts *versus* osteocytes. We found an enrichment of "*GP6 Signaling*", a major signaling receptor for collagen in osteoblasts, which was almost completely shut down in osteocytes (**Figure 2A**). In contrast, the "*Axonal guidance*", the "*Differentiation via BMP receptors*", "*TGFβ signaling*" and the "*Iron homeostasis signaling*" pathways that were not highly enriched in osteoblasts were significantly increased in osteocytes (**Figure 2A**). Consistent with osteocyte morphology, among the genes predicted to drive "*Axonal guidance*" were *Adamts14*, *Bmp2*, *Bmp4*, and *Wasl*. Furthermore, IPA analysis predicted EGFR among the top upstream regulators of osteoblastic pathways whereas FGF2 was predicted to be a primary upstream regulator associated with the osteocyte maturation pathway (**Figure S1A**). Additionally, our in-silico analysis identified an enriched cell activation in osteoblasts, associated with genes such as *Col11a1*, *Col11a2*, and *Col1a1* which were significantly attenuated in osteocytes (**Figure 2A**). Overall, these data support that osteoblasts and osteocytes are defined by unique pathways associated with their individual homeostatic functions.

To predict the transcriptional profiles of osteocyte genes across the osteolineage cells, we used pseudotime analysis to track the computational-derived kinetics of gene expression during differentiation. We found that "pan-osteolineage" markers *Col1a1*, *Col1a2* and *Runx2* showed a relatively stable expression during osteoblast to osteocyte differentiation (**Figure 2B**). In contrast,

Mmp13, Tnc, and *Spp1* pseudotemporal progression predicted a decrease during differentiation from Tnc/Mmp13 osteoblasts to osteocytes (**Figure 2C**). Further, *Bglap, Col11a2,* and *Smpd3* expression increased from the Tnc/Mmp13 precursor osteoblasts to osteoblasts, but decreased when the osteoblast population transitioned to osteocytes (**Figure 2D**). In osteocytes, the kinetic profiling of *Dmp1, Irx5, Ackr3, Phex, Ptprz1,* and *Cd109* showed progressive positive regulation during differentiation (**Figure 2E**). These data confirmed the dynamic differentiation of canonical genes during the transcriptional reprogramming required for osteoblast transition to osteocytes, as well as identified new genes associated with the cortical bone osteolineage.

To distinguish the metabolic heterogeneity between osteoblasts and osteocytes, we applied the recently developed computational method single-cell flux estimation analysis (scFEA) [23] that uses a systematically reconstructed human metabolic map to infer the cell-wise cascade from the transcriptome to metabolome. This approach applies multilayer neural networks to capitulate the nonlinear dependency between enzymatic gene expression and reaction rate fluxome from scRNAseq datasets. By reconducting the cell clustering analysis using the original UMAP labeling (see **Figure 1C**), we identified metabolic fluxes in Tnc/Mmp13 osteoblasts, osteoblasts and osteocytes (**Figure S1B**) and found a decrease of glycolytic, TCA, and fatty acid metabolic reactions when pre-osteoblasts transitioned to osteoblasts/osteocytes (**Figure S1C**). Specifically, the genes controlling successive glycolytic reactions from glucose to pyruvate pathways were decreased with maturation of Tnc/Mmp13 osteoblasts to osteoblasts (**Figure 2F** and **Figure S1D**). Further, mRNAs associated with the conversion of fatty acid to acetyl-CoA (**Figure 2F** and **Figure S2E**) and glutamate export also decreased with cell maturation (**Figure S2F**). Importantly, whereas the conversion of serine to cysteine decreased with osteoblast differentiation, the transformation of cysteine to pyruvate increased when Tnc/Mmp13 osteoblasts transitioned to osteoblasts/osteocytes, a potential metabolic route to supply cells with pyruvate since the glycolytic pathway is down regulated during osteoblast differentiation (**Figure 2G**). Based on the

scRNAseq dataset, the top metabolites in osteolineage cells were the inosine monophosphate (IMP), succinyl coenzyme A, and pyrimidine in pre-osteoblasts; the farnesyl pyrophosphate (FPP), cytidine-5'-diphosphate (CDP), and malate in osteoblasts; whereas the xanthosine monophosphate (XMP), acetyl CoA oxaloacetate, and malate were enriched in osteocytes (**Figure 2H**). Collectively, the transcriptional profile across cortical bone cell types predicted differential metabolic states between the osteoblast cell population subsets and osteocytes.

Identification of genes at staged differentiation associated with developmental processes

To validate the osteoblast/osteocyte genes identified from our *in vivo* scRNAseq experiments, the mRNA expression of these genes was tested in mesenchymal progenitor cell (MPC)-2 cells *in vitro*. MPC2 cells were recently characterized as harboring the ability to derive osteoblast- and osteocyte-like cells when cultured in osteogenic media [24]. After 1-4 weeks of culture, the differentiated cells are known to increase alizarin red staining which reflect the ability of the mature cells to mineralize [24]. Our *in vitro* molecular analyses demonstrated that mRNAs encoding *Col1a1* and *Sost* increased during cell differentiation with the highest level observed at 4 weeks (**Figure 3A**). Importantly, the expression of *Ptprz1* and *Sost* were only detected at 3 and 4 weeks of osteocyte differentiation confirming our finding of *Ptprz1* as a likely late osteoblast/osteocyte gene.

To identify the molecular events associated with osteolineage differentiation, we next employed a comprehensive genomic and transcriptomic analyses using RNAseq and ATACseq on MPC2 cells [37]. Based on the kinetic profile performed in **Figure 3A**, a time point of 3 weeks of MSC to osteocyte-like cell differentiation was chosen for RNAseq and ATACseq to capture differentially regulated genes in osteoblasts and osteocytes. Our unbiased approach identified several genes with increased chromatin accessibility and/or gene expression during MPC2 differentiation

including osteocalcin (*Bglap*), osteopontin *(Spp1), Col11a1,* and *Col11a2* (**Figure 3B and Figure S2A-F**). By performing integrative and correlative analyses of the ATACseq and RNAseq, we confirmed our observed changes in mRNA expression and chromatin accessibility in the ATAC-seq dataset. Using this approach, 150 genes were identified with significantly more open chromatin status leading to higher gene expression levels after differentiation (**Figure 3C**), whereas 44 genes were downregulated with notably less DNA accessibility (**Figure 3C**).

The upregulated genes identified from differential expression analysis associated with changes in chromatin accessibility (**Figure 3D**) were then subdivided by gene ontology (GO). Among the GO terms that were increased during osteoblast/osteocyte differentiation were functions related to bone mineralization as well as cellular morphology changes including membrane branching (**Figure 3E**). For example, the chromatin accessibility at the osteocalcin (Bglap) promoter region was also increased after differentiation. Further, the DNA promoter regions of *Col8a1*, *Mgp*, *Igfbp5*, *Sema3a* and *Tnc* genes were more accessible with cell maturation (**Figure 3F**). These genomic changes detected during the cell transitions were robustly and positively associated with their corresponding mRNA expression levels (**Figure 3G**).

We next tested the chromatin accessibility of a set of osteolineage genes identified in the scRNAseq. For instance, among the Tnc/Mmp13 osteoblast signature genes, we detected increased chromatin accessibility at the *Mmp13*, *Serpine2,* and *Lifr* promoter regions in differentiated cells versus MSC control cells (**Figure S3A-C**). We also observed high chromatin accessibility at the promoters of the *Serpinf1* and *Cthrc1* genes (**Figure S3D-E**), that were defined as expressed by fully differentiated osteoblasts based upon our *in vivo* scRNAseq data. Regarding osteocyte genes, we observed an increase chromatin accessibility for *Ackr3, Cd109, Ptgis, Spns2,* and *Bmp2* promoter regions (**Figure S3F-J**). Importantly, non-canonical osteocyte genes identified from the scRNAseq analysis including *Ccn4, Adamts14, Spns2,* and *Bmp2* were significantly upregulated in osteocyte-like cells when compared to undifferentiated cells (**Figure**

3H). *Ptgis* identified in our dataset and the Wang, et al dataset [10] was also significantly increased with differentiation. *Col24a1* found in our dataset and the Youlten, et al dataset [9] was also increased (**Figure 3H**). Taken together, these results pinpoint genomic regions of interest associated with osteoblast differentiation that may play critical roles in cellular transcriptional reprogramming.

Impairment of osteoblast/osteocyte genes prior to cortical bone deterioration in CKD

We next assessed the dynamic transcriptional profile of identified osteoblast and osteocyte genes within cortical bone prior to the development of cortical porosity in CKD, a severe outcome that leads to bone fracture. We used a mouse model in which CKD is induced by providing an adenine-containing diet, which causes a progressive loss of kidney function [38]. This model reflects the phenotype observed in humans characterized by progressive tubular atrophy, immune cell infiltration, and fibrosis [38]. With a focus on testing osteoblast/osteocyte gene regulation in cortical bone at a timepoint prior to major ultrastructural changes, cohorts of C57BL/6 mice were placed on control diet or a 0.2% adenine-containing diet for 2 or 4 weeks. At this stage of the disease (2 weeks), serum phosphate, alkaline phosphatase, and calcium were not significantly changed compared to controls (**Table 1**). Blood urea nitrogen (BUN) was monitored for decline in renal function, and as expected, the mice receiving the adenine diet had significantly increased BUN (**Table 1**). The cortical porosity and bone volume remained unchanged at 2 or 4 weeks in male or female CKD mice (**Figure 4C**). However, plasma bioactive iFGF23 concentrations were markedly elevated in the mice with CKD, which was associated with significant cortical bone *Fgf23* mRNA increases at 4 weeks (**Figure 4D-E**).

We next determined the effects of CKD on highly expressed osteoblast and osteocyte genes. During CKD, *Mmp13*, identified in the osteoblast precursor population, was increased at 4 weeks

(**Figure 4F**). In contrast, precursor osteoblast genes such as *Tnc, Gdpd2, Serping1,* and *Igfbp5* were decreased during CKD progression. The expression of the osteolineage gene *Col1a1*, as well as osteoblast-specific genes *Bglap, Cthrc1, Smpd3* and osteoblast/osteocyte marker *Dmp1*, were strongly impaired in the mice with CKD (**Figure 4G**). Additionally, osteocyte genes including *Sost*, *Pdpn, Phex, Ptprz1* and *Adamts14* were dramatically downregulated early in CKD (**Figure 4H**). In sum, this work demonstrated that scRNAseq can identify sub-populations of osteoblasts, and together with metabolic profiling, predict differentiation to mature osteoblasts and osteocytes. In addition, we identified new cortical bone genes associated with corresponding changes in their genomic accessibility in *in vitro* differentiation studies. These genes were then shown to be dysregulated in a mouse model of CKD. These collective findings support that the molecular changes observed in cortical bone associated with the severe musculoskeletal phenotypes in CKD rapidly and dramatically occur. Further, these alterations can now be pinpointed to specific cell types within the osteolineage.

Discussion

There is an urgent need to understand bone-forming cell heterogeneity as well as isolate key gene networks that impact skeletal homeostasis. This is especially apparent in diseases such as CKD, where due to the progressive nature and specific calcium- and phosphate-related endocrine disturbances in this disease, there are currently very little treatment options [39, 40]. Our cortical bone single cell RNAseq analysis identified two osteoblast cell populations as well as a distinct osteocyte cluster. The genes demarcating each cell type were confirmed as highly enriched in cortical bone as well as being upregulated during osteoblast differentiation *in vitro*. Interestingly and unexpectedly, at early stage CKD onset in mice, there was a dramatic misregulation of gene expression in the three cortical bone cell types prior to major bone ultrastructural changes.

Recent studies using bulk RNA sequencing developed an osteocyte-defining transcriptome which was comprised of 28 genes [9]. Whereas we observed overlap of genes within the transcriptomes of Wang et al [10] and Youlten et al [9], our scRNAseq analysis identified additional genes which were highly enriched in osteocytes (see **Figure 1F**). It is likely that some of the observed differences in osteocyte gene profiling between our studies and those of others are due to the fact that we exclusively used long bones for cell isolation whereas the previous studies sequenced a mix of long bone and calvariae, or used bulk total cortical bone RNA as a starting material [9, 10]. Future studies will be needed to further sub-set these gene profiles and expand cell-specific targets as more cortical bone datasets become publicly available.

Our bone scRNAseq analyses identified a population of Tnc/Mmp13 mRNA-containing osteoblasts as potential precursors of a mature osteoblast population of cells. Mmp13, the most highly expressed gene in this cell type has been described as a direct target of osteoblast-specific transcription factor osterix (Sp7) in osteoblasts [41] and known to be activated by Runx2, the master transcription factor for osteoblastogenesis [42]. A recent study defined the molecular mechanisms driven by Sp7 to control osteocyte dendrite formation [10]. Whether directly targeting

Mmp13 could influence osteoblast differentiation and osteocyte maturation remains to be investigated. The bone disease arising from CKD has been associated with intrinsic impairment of osteoblast and osteocyte differentiation [43]. Our analyses identified osteolineage genes at all stages of maturation that were dysregulated in modeled CKD prior to major bone structural changes in pre-osteoblasts, osteoblasts and osteocytes. Thus, these genes may serve as early onset markers for CKD bone disease.

Another primary goal of this work was to decipher potential alterations in osteolineage gene regulation during CKD onset. We established a critical link between osteoblasts at early and mature stages, as well as early osteocyte genes that were suppressed prior to the development major bone ultrastructural changes during modeled CKD. These studies imply that targeting one bone cell type in CKD bone disease will likely not be fully effective since damage may be done in early precursor cells, with these cells either having faulty differentiation or potentially carrying these gene changes through differentiation where they may impact mature cells. We also found a decrease in cortical bone Sclerostin expression at an early stage of CKD in mice (**Figure 4H**). In a paracrine signaling manner, rapid decreases of osteocyte genes *Sost* and likely *Dkk1* may represent an adaptive mechanism to maintain effective Wnt signaling as an attempt to delay the initial development of renal osteodystrophy.

The skeleton is an important regulator of systemic glucose homeostasis, with osteocalcin and insulin thought to represent the prime mediators of the interplay between bone and energy metabolism [44, 45]. Using the recently developed computational tool single-cell flux estimation analysis (scFEA) [23] that calculates the cell metabolomic state in scRNAseq datasets, our data predicted *in vivo* transitions during osteoblast differentiation. At the single cell level, we identified the highest glycolysis and glutamine metabolism states in osteoblast precursors (**Figure 2F** and **Figure S1F**). The conversion of arginine to ornithine was similar across all osteolineage cells (**Figure S1G**). Other pathways, including the conversion of citrulline/aspartate to arginosuccinate

or acetyl glucosamine to hyaluronic acid were conserved regardless of cell differentiation or cell type (**Figure S1G**). It will be important to build on these findings to provide genetic evidence that some identified osteoblast precursor genes may be critical in supporting osteoblast differentiation as well as influencing global energy metabolism in normal states and in skeletal disease.

Although this study provided an extensive gene profile analysis of osteolineage cells, some established osteocyte genes such as *Sost* and *Fgf23* were not detected under baseline conditions. It is likely that current single cell sequencing technologies using the short coverage length along each mRNA from the 3'UTR end (which is based on 28 bp of cell barcode and UMI sequences and 91 bp RNA reads generated) have sensitivity limitations to detect low, yet specific gene expression in less prevalent cell populations. Overall, lower expression genes (mainly with a detection at the latest quarter threshold cycle using real time qPCR) were minimally detected at single cell resolution using the current 10X Genomics single-cell gene expression profiling pipeline. However, our approach of combining cell sorting with cell isolation from the Sost-cre/Ai9 mice to develop an enriched cortical bone cell population identified potentially new sets of osteocyte genes which were validated *in vitro* and in RNA from cortical bone *ex vivo*.

In summary, by combining genomic, transcriptomic, and predictive metabolic profiling approaches, our studies identified osteolineage genes associated with distinct cell populations of osteoblast precursors, mature osteoblasts and osteocytes. Genes within these three cortical bone cell populations were misregulated in a mouse model of CKD prior to the development of cortical porosity. Our findings support that CKD osteoblastogenesis is likely impaired prior to the increase of osteoclast activity; thus, it will be necessary to determine the contribution of the newly identified osteolineage genes in CKD-induced bone loss as well as other common musculoskeletal diseases.

Acknowledgements

The authors thank the members of the Indiana University Melvin and Bren Simon Cancer Center Flow Cytometry Resource Facility for their outstanding technical support. This work was supported in part by NIH. The Indiana University Melvin and Bren Simon Comprehensive Cancer Center Flow Cytometry Resource Facility is funded in part by NIH, National Cancer Institute (NCI) grant P30 CA082709 and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant U54 DK106846. The FCRF is supported in part by NIH instrumentation grant 1S10D012270. The authors would like to acknowledge NIH grants K99-DK129705 (RA), R21-AR059278 (KEW), R01-DK112958 (KEW), P20GM125503 (NIGMS to IN), and the David Weaver Professorship (KEW).

Disclosures

KEW receives royalties for licensing FGF23 to Kyowa Hakko Kirin Co., Ltd; had previous funding from Akebia, and current funding from Calico Labs; KEW also owns equity interest in FGF Therapeutics. The other authors have nothing to declare.

Data availability:

scRNAseq, RNAseq and ATACseq datasets will be deposited in the NCBI's Gene Expression Omnibus database.

References

1. Dallas, S.L. and L.F. Bonewald, *Dynamics of the transition from osteoblast to osteocyte.* Ann N Y Acad Sci, 2010. **1192**: p. 437-43.
2. Martin, B., *Aging and strength of bone as a structural material.* Calcif Tissue Int, 1993. **53 Suppl 1**: p. S34-9; discussion S39-40.
3. Seeman, E., *Age- and menopause-related bone loss compromise cortical and trabecular microstructure.* J Gerontol A Biol Sci Med Sci, 2013. **68**(10): p. 1218-25.
4. Nickolas, T.L., et al., *Rapid cortical bone loss in patients with chronic kidney disease.* J Bone Miner Res, 2013. **28**(8): p. 1811-20.
5. Luyckx, V.A., M. Tonelli, and J.W. Stanifer, *The global burden of kidney disease and the sustainable development goals.* Bull World Health Organ, 2018. **96**(6): p. 414-422D.
6. John, G.B., C.Y. Cheng, and M. Kuro-o, *Role of Klotho in aging, phosphate metabolism, and CKD.* Am J Kidney Dis, 2011. **58**(1): p. 127-34.
7. Nitsch, D., et al., *Chronic kidney disease and hip fracture-related mortality in older people in the UK.* Nephrol Dial Transplant, 2009. **24**(5): p. 1539-44.
8. Robertson, L., et al., *Hip fracture incidence and mortality in chronic kidney disease: the GLOMMS-II record linkage cohort study.* BMJ Open, 2018. **8**(4): p. e020312.
9. Youlten, S.E., et al., *Osteocyte transcriptome mapping identifies a molecular landscape controlling skeletal homeostasis and susceptibility to skeletal disease.* Nat Commun, 2021. **12**(1): p. 2444.
10. Wang, J.S., et al., *Control of osteocyte dendrite formation by Sp7 and its target gene osteocrin.* Nat Commun, 2021. **12**(1): p. 6271.
11. Lane, N.E., *Epidemiology, etiology, and diagnosis of osteoporosis.* Am J Obstet Gynecol, 2006. **194**(2 Suppl): p. S3-11.
12. in *Guide for the Care and Use of Laboratory Animals*, th, Editor. 2011: Washington (DC).
13. Maurel, D.B., et al., *Characterization of a novel murine Sost ER(T2) Cre model targeting osteocytes.* Bone Res, 2019. **7**: p. 6.
14. Stern, A.R., et al., *Isolation and culture of primary osteocytes from the long bones of skeletally mature and aged mice.* Biotechniques, 2012. **52**(6): p. 361-73.
15. Clinkenbeard, E.L., et al., *Increased FGF23 protects against detrimental cardio-renal consequences during elevated blood phosphate in CKD.* JCI Insight, 2019. **4**(4).
16. Kedlaya, R., et al., *Sclerostin inhibition reverses skeletal fragility in an Lrp5-deficient mouse model of OPPG syndrome.* Sci Transl Med, 2013. **5**(211): p. 211ra158.
17. Bouxsein, M.L., et al., *Guidelines for assessment of bone microstructure in rodents using micro-computed tomography.* J Bone Miner Res, 2010. **25**(7): p. 1468-86.
18. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
19. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
20. Finak, G., et al., *MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data.* Genome Biol, 2015. **16**: p. 278.
21. Alquicira-Hernandez, J. and J.E. Powell, *Nebulosa recovers single cell gene expression signals by kernel density estimation.* Bioinformatics, 2021.
22. Trapnell, C., et al., *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.* Nat Biotechnol, 2014. **32**(4): p. 381-386.
23. Alghamdi, N., et al., *A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data.* Genome Res, 2021. **31**(10): p. 1867-1884.

24. Prideaux, M., et al., *Generation of two multipotent mesenchymal progenitor cell lines capable of osteogenic, mature osteocyte, adipogenic, and chondrogenic differentiation.* Sci Rep, 2021. **11**(1): p. 22593.
25. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
26. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
27. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.
28. Kramer, A., et al., *Causal analysis approaches in Ingenuity Pathway Analysis.* Bioinformatics, 2014. **30**(4): p. 523-30.
29. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.
30. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
31. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
32. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
33. Amemiya, H.M., A. Kundaje, and A.P. Boyle, *The ENCODE Blacklist: Identification of Problematic Regions of the Genome.* Sci Rep, 2019. **9**(1): p. 9354.
34. Piper, J., et al., *Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data.* Nucleic Acids Res, 2013. **41**(21): p. e201.
35. McCarthy, D.J., Y. Chen, and G.K. Smyth, *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.* Nucleic Acids Res, 2012. **40**(10): p. 4288-97.
36. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.
37. Shashikant, T. and C.A. Ettensohn, *Genome-wide analysis of chromatin accessibility using ATAC-seq.* Methods Cell Biol, 2019. **151**: p. 219-235.
38. Diwan, V., L. Brown, and G.C. Gobe, *Adenine-induced chronic kidney disease in rats.* Nephrology (Carlton), 2018. **23**(1): p. 5-11.
39. Nickolas, T.L., et al., *Discriminants of prevalent fractures in chronic kidney disease.* J Am Soc Nephrol, 2011. **22**(8): p. 1560-72.
40. Maravic, M., et al., *Incidence and risk factors for hip fractures in dialysis patients.* Osteoporos Int, 2014. **25**(1): p. 159-65.
41. Zhang, C., W. Tang, and Y. Li, *Matrix metalloproteinase 13 (MMP13) is a direct target of osteoblast-specific transcription factor osterix (Osx) in osteoblasts.* PLoS One, 2012. **7**(11): p. e50525.
42. Hecht, J., et al., *Detection of novel skeletogenesis target genes by comprehensive analysis of a Runx2(-/-) mouse model.* Gene Expr Patterns, 2007. **7**(1-2): p. 102-12.
43. Pereira, R.C., et al., *Impaired osteocyte maturation in the pathogenesis of renal osteodystrophy.* Kidney Int, 2018. **94**(5): p. 1002-1012.
44. Karsenty, G. and M. Ferron, *The contribution of bone to whole-organism physiology.* Nature, 2012. **481**(7381): p. 314-20.
45. Ducy, P., *The role of osteocalcin in the endocrine cross-talk between bone remodelling and energy metabolism.* Diabetologia, 2011. **54**(6): p. 1291-7.

Figure 1. Single-cell transcriptomic profiling of bone cells identified osteolineage cells heterogeneity. A-B. Long bones (femur, tibia, and humeri) were subjected to serial collagenase/EDTA digestions (see Methods), and cells from fractions 4-9 were collected for sorting to isolate tdTomato-positive cells, followed by single-cell RNAseq library construction. Parallel analysis confirmed high mRNA expression of osteocyte genes *Fgf23*, *Dmp1*, *Phex*, and *Pdpn* in Ai9 positive cells *versus* Ai9 negative cells by qPCR. **C.** Bioinformatic analysis was performed to cluster the cells by UMAP. Each dot represents a single cell, and cells sharing the same color code indicate discrete populations of transcriptionally similar cells. **D.** Defining cortical bone cell types. Tnc/Mmp13 osteoblast cells were distinctly marked by *Tnc* and *Mmp13*. Osteoblast cells showed high expression of *Smpd3* and *Bglap*. Osteocytes had the highest expression of *Dmp1* and *Phex*. **E.** Expression density plots indicated cells with high transcription of *Col1a1*, *Tnc*, *Bglap*, and *Phex*. **F.** The polar figure highlights the osteocyte markers isolated from our dataset (long bone cell scRNAseq), Wang et al. (long bone and calvaria cell scRNAseq) and Youlten et al (long bone bulk RNAseq)**. G.** Canonical and non-canonical osteolineage genes were validated as being highly expressed in cortical bone when compared to the expression detected in bone marrow. **H.** Pseudotime analysis revealed Tnc/Mmp13 Osteoblasts as an osteoblast precursor (P-OB) and shows their differentiation to osteoblasts (OB) and then osteocytes (OC).

Figure 2. Differential signaling pathways in osteoblasts versus osteocytes. A. Ingenuity Pathway Analysis (IPA) software was used to predict the statistically significant canonical pathways upregulated in Tnc/Mmp13 osteoblasts (P-OB), osteoblasts (OB) and osteocytes (OC) from the scRNAseq data based upon a calculated probability score of ≥2 and a p-value of <0.05. Further, a comparative analysis between osteolineage cells was performed to generate the heatmap. The increase of blue intensity reflects the more significant pathways within the cell types. The gray dots indicate non-significant pathways. **B-E.** Monocle analysis shows pseudotime trajectory mapping of osteolineage gene sets *Col1a1, Col1a2, Runx2;* pre-osteoblast gene sets *Mmp13, Tnc, Spp1;* osteoblast gene sets *Bglap, Col11a2, Smpd3;* and osteocyte gene sets *Dmp1, Irx5, Ackr3, Phex, Ptprz1,* and *Cd109*. **F.** The computational method scFEA was used to infer cell-wise fluxome from the scRNAseq data to predict the metabolic profiling as well as the differential metabolite conversion rate in cells that correspond to a metabolic flux value. Ridgeline plots indicate the distributions values of metabolic flux in Tnc/Mmp13 osteoblasts (P-OB), osteoblasts (OB), and osteocytes (OC). Each ridgeline represents the flux between two metabolites, shown on the x-axis, for different osteolineage cells, shown on the y-axis. **G.** The boxplots indicate the predictive value of the conversion of methionine to cysteine and to pyruvate in P-OB, OB, and OC. **H.** The bar plots show the top 10 different metabolites enriched in P-OB, OB, and OC.

Figure 3. Molecular markers associated with osteoblasts/osteocytes morphogenesis change. A-top. Schematic representing cell differentiation from MSCs to osteoblasts/osteocytes. **A-lower.** *Col1a1*, *Sost,* and *Ptprz1* mRNAs were analyzed in MPC2 cells during osteogenic differentiation over the course of 1-4 weeks and normalized to β-actin. **B.** Volcano plots derived from the bulk RNAseq (left) and ATACseq datasets (right) show the significant alterations of gene expression and chromatin accessibility detected in differentiated *versus* undifferentiated cells. The x-axis shows the fold-change (in $\log_2$ scale). The y-axis represents the significance. **C.** Venn diagrams highlight the number of genes with differentially accessible regions (DARs), or differentially expressed genes (DEGs), or both differentiated *versus* undifferentiated cells. **D.** Integration of ATACseq and RNAseq data exhibited a higher correlation between genome-wide chromatin accessibility changes (x-axis) and gene expression alterations (y-axis). **E.** Functional enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) on upregulated DEGs with more open DARs identified biological pathways that induced with osteogenic differentiation (numbers in parenthesis are gene counts). **F.** Representative ATACseq peaks of undifferentiated cells (top track in gray) compared to differentiated cells (osteoblasts/osteocytes bottom track in black) are shown. **G.** Representative mRNA expression of upregulated genes in differentiated versus undifferentiated cells enriched in cell morphogenesis and ossification. All the genes shown were significantly upregulated (false discovery rate, FDR < 0.05). **H.** The osteocyte genes identified from the scRNAseq dataset in **Figure 1J** were confirmed to be significantly upregulated in osteoblasts/osteocytes cells *versus* control undifferentiated MPC2 cells.

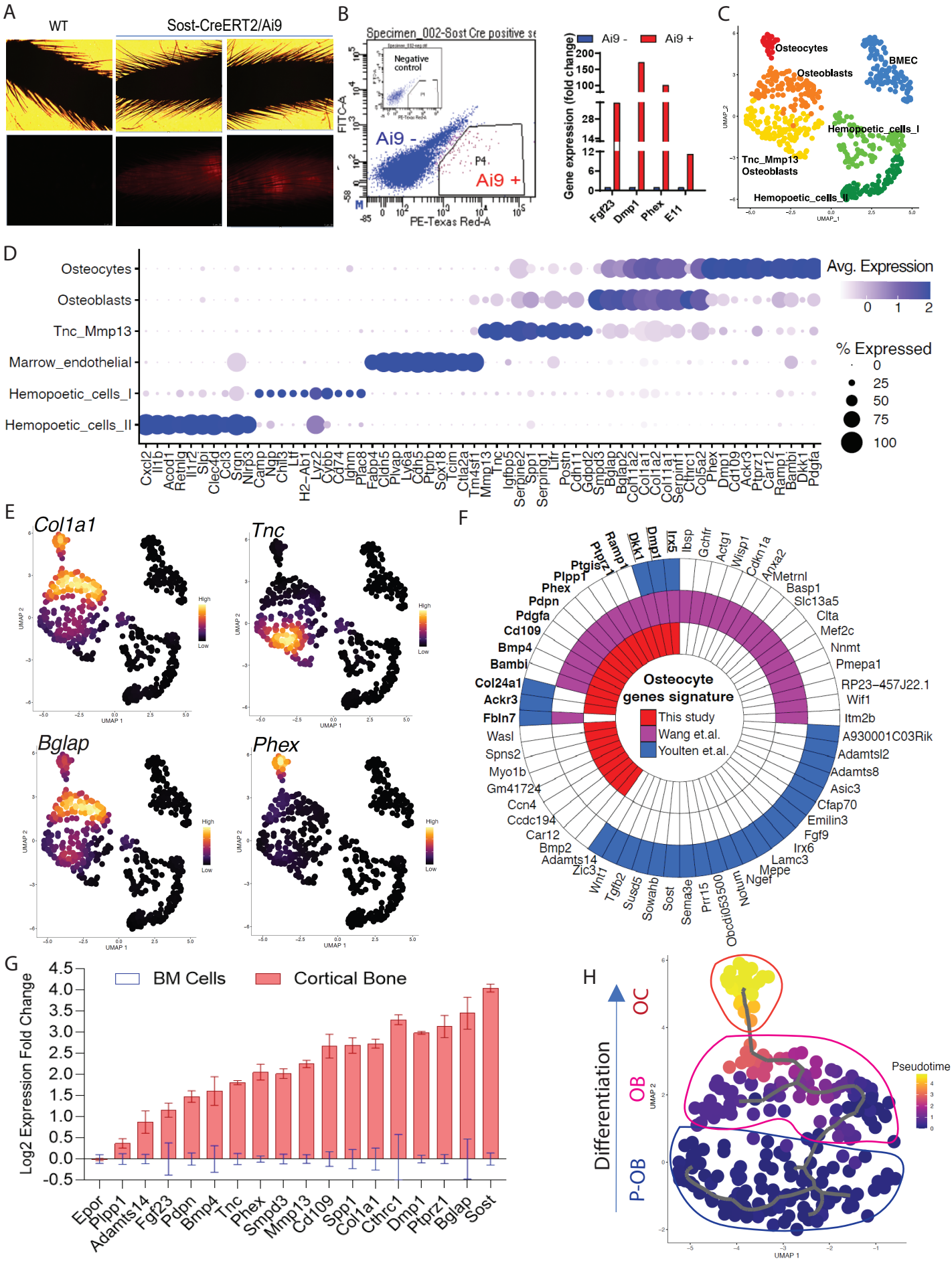
Figure 4. Precocious misregulation of osteoblast/osteocyte genes in CKD. Eight week old wild type C57BL/6 male (shown in blue) and female (represented in red) mice were fed an adenine diet (AD) to induce CKD for 2, or 4 weeks. Mice fed the casein diet for 2 weeks were used as controls. **A.** Key serum biochemical analyses: phosphate, alkaline phosphatase, calcium, blood urea nitrogen, creatinine, and iron are shown. **B.** Cortical porosity (left) and trabecular bone volume (right) were measured using micro computed tomography (μCT). **C.** Representative images of cortical bone μCT. **D-E.** Cortical bone Fgf23 mRNA was measured by real-time qPCR and circulating FGF23 was assessed by ELISA. **F-G.** Real-time qPCR was used to measure the mRNA expression of Tnc/Mmp13 osteoblast (pre-osteoblasts), osteoblast and osteocyte gene sets. Data are shown as fold change (2-ΔΔCt) relative to the housekeeping gene *β-Actin* and normalized to the experimental control ('AD0'). The 'AD 0' mice were fed with a control diet (Casein) and sacrificed at 2 weeks; 'AD 2' mice were fed with a CKD diet (adenine) and sacrificed at 2 weeks; and the 'AD 4' were mice fed with the adenine diet and sacrificed at 4 weeks. Data are represented as mean +/- standard deviation. *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.001$ compared to Control.

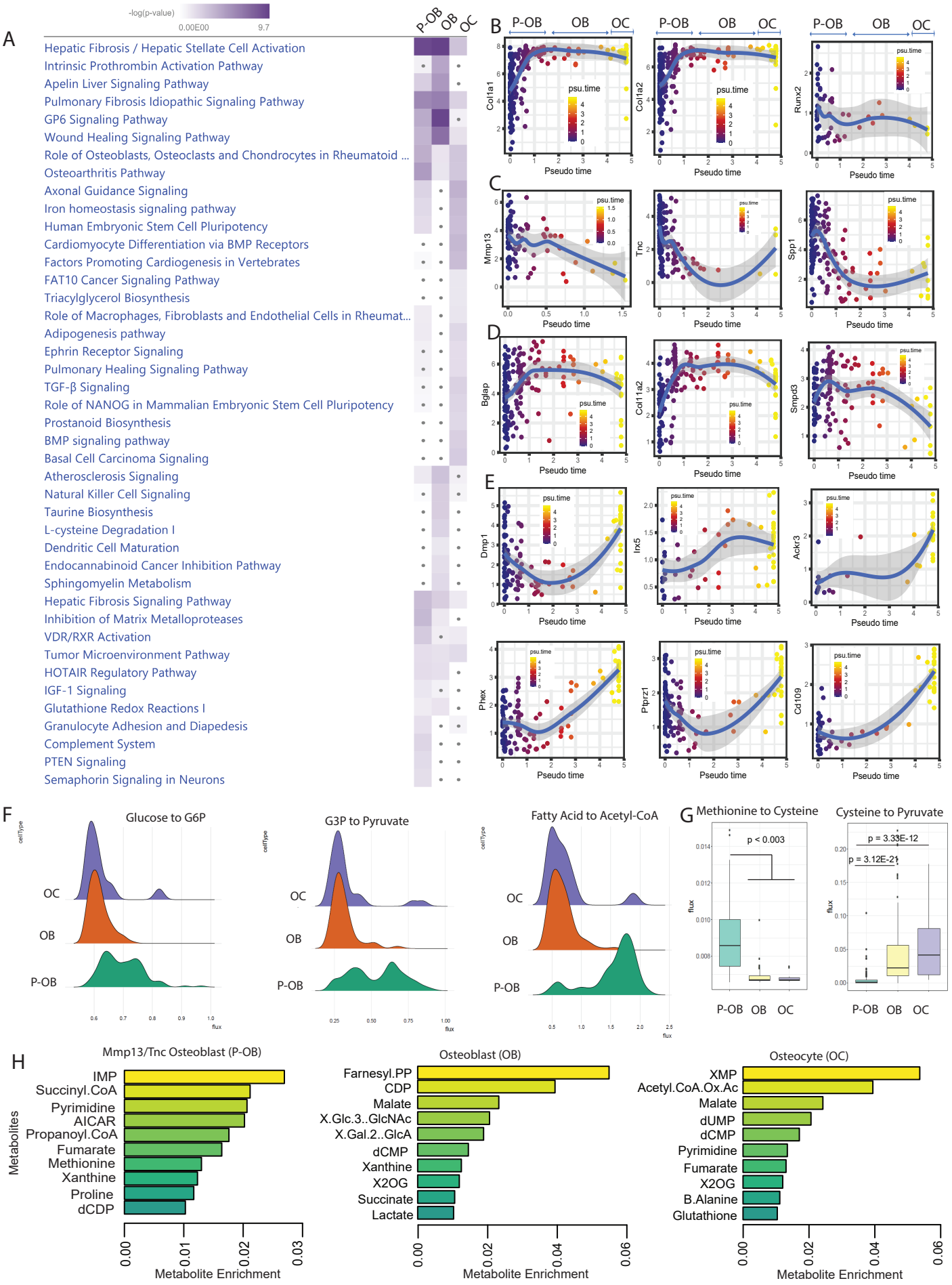
Supplementary Figure 1. A. Predictive upstream regulators in Tnc/Mmp13 osteoblast (pre-osteoblasts), osteoblast and osteocyte. **B.** scFEA UMAP. **C.** Heatmap indicates the distribution of predicted cell-wise flux of glycolytic, TCA, serine metabolism, fatty acids metabolism and glutamate metabolism relative to pre-osteoblasts (Tnc/Mmp13) values. The heatmap uses a column Z-score to show significant differences between pre-osteoblasts (Tnc/Mmp13), osteoblasts, and osteocytes. **D-F.** Ridgeline plots indicate the distribution values of metabolic flux in pre-osteoblasts (P-OB), osteoblasts (OB), and osteocytes (OC). Each ridgeline represents the flux between two metabolites (x-axis) for different cells that are plotted on the y-axis. **G.** Ridgeline plots indicate the distribution values of metabolic flux in pre-osteoblasts (Tnc/Mmp13), osteoblasts (OB), osteocytes (OC), hematopoietic cells (Hem I and Hem II), and bone marrow endothelial cells (BMEC).

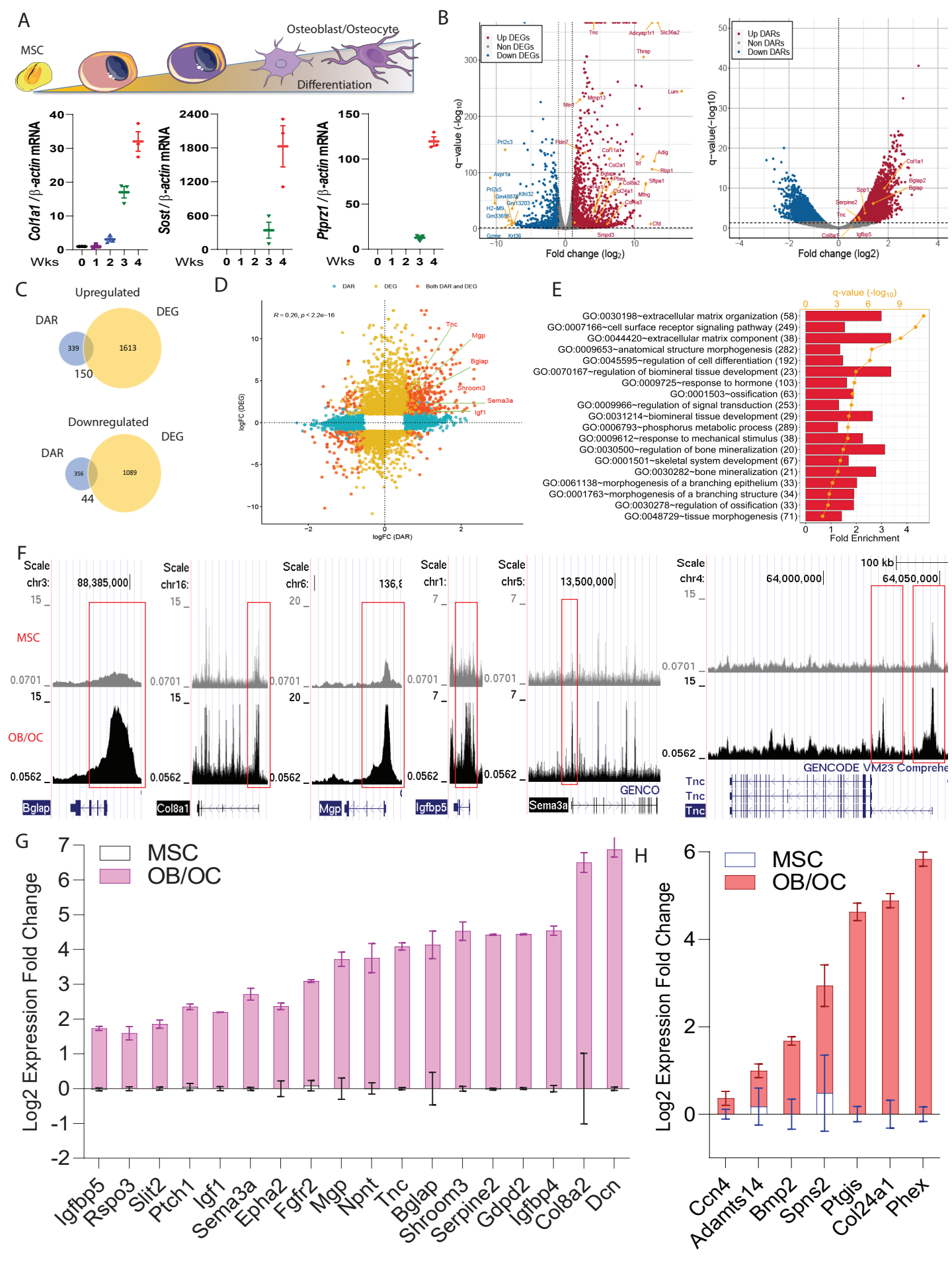
Supplementary Figure 2. A-F. Chromatin accessibilities with corresponding gene expression of Spp1 (p = 0.00933495), Col11a1 (p = 1.9363E-05), and Col11a2 (p = 7.2591E-06) in differentiated (OB/OC) *versus* undifferentiated (MSC) cells.

Supplementary Figure 3. A-C. Assessment of chromatin accessibilities at *Mmp13* (p = 0.01025011), *Serpine2* (p = 3.4093E-09), and *Lifr* (p = 1.2547E-07) genomic regions. These genes were detected as highly enriched in pre-osteoblasts. **D-E.** Assessment of chromatin accessibilities at *Serpinf1* (p = 0.00099491), and *Cthrc1* (p = 0.013) genomic loci. These genes were predicted to be highly enriched in osteoblasts. **D-E.** Assessment of chromatin accessibilities at *Ackr3* (p = 2.503E-06), *Cd109* (p = 0.00179923), *Ptgis* (p = 6.8707E-06), *Spns2* (p = 0.00521604), and *Bmp2* (p = 0.00540629) genomic regions. These genes were detected primarily

in osteocytes). Top tracks (gray) corresponded to undifferentiated cells (MSC) and lower tracks (black) corresponded to differentiated cells (OB/OC).

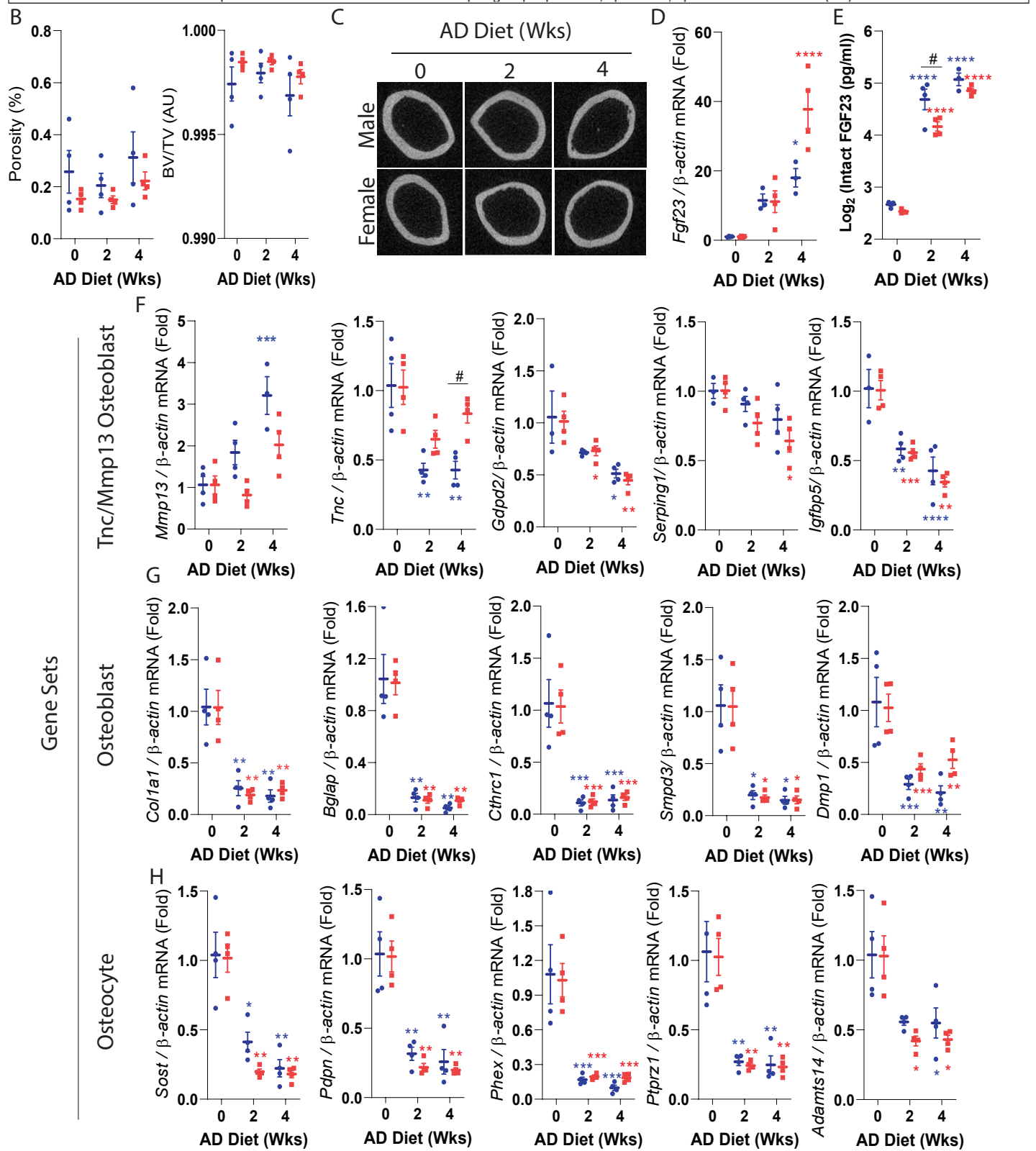

A
WT
Sost-CreERT2/Ai9
B
Specimen_002-Sost Cre positive se
Negative control
Ai9 -
Ai9 +
P4
FITC-A
PE-Texas Red-A
Gene expression (fold change)
Fgf23
Dmp1
Phex
E11
C
Osteocytes
Osteoblasts
BMEC
Hemopoetic_cells_I
Tnc_Mmp13 Osteoblasts
Hemopoetic_cells_II
UMAP_1
UMAP_2
D
Osteocytes
Osteoblasts
Tnc_Mmp13
Marrow_endothelial
Hemopoetic_cells_I
Hemopoetic_cells_II
Avg. Expression
% Expressed
E
Col1a1
Tnc
Bglap
Phex
High
Low
UMAP 1
UMAP 2
F
Osteocyte genes signature
This study
Wang et.al.
Youlten et.al.
G
BM Cells
Cortical Bone
Log2 Expression Fold Change
Epor
Plpp1
Adamts14
Fgf23
Pdpn
Bmp4
Tnc
Phex
Smpd3
Mmp13
Cd109
Spp1
Col1a1
Cthrc1
Dmp1
Ptprz1
Bglap
Sost
H
Differentiation
OC
OB
P-OB
Pseudotime
UMAP 1
UMAP 2


A
-log(p-value)
0.00E00
9.7
P-OB
OB
OC
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Intrinsic Prothrombin Activation Pathway
Apelin Liver Signaling Pathway
Pulmonary Fibrosis Idiopathic Signaling Pathway
GP6 Signaling Pathway
Wound Healing Signaling Pathway
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid ...
Osteoarthritis Pathway
Axonal Guidance Signaling
Iron homeostasis signaling pathway
Human Embryonic Stem Cell Pluripotency
Cardiomyocyte Differentiation via BMP Receptors
Factors Promoting Cardiogenesis in Vertebrates
FAT10 Cancer Signaling Pathway
Triacylglycerol Biosynthesis
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumat...
Adipogenesis pathway
Ephrin Receptor Signaling
Pulmonary Healing Signaling Pathway
TGF-β Signaling
Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency
Prostanoid Biosynthesis
BMP signaling pathway
Basal Cell Carcinoma Signaling
Atherosclerosis Signaling
Natural Killer Cell Signaling
Taurine Biosynthesis
L-cysteine Degradation I
Dendritic Cell Maturation
Endocannabinoid Cancer Inhibition Pathway
Sphingomyelin Metabolism
Hepatic Fibrosis Signaling Pathway
Inhibition of Matrix Metalloproteases
VDR/RXR Activation
Tumor Microenvironment Pathway
HOTAIR Regulatory Pathway
IGF-1 Signaling
Glutathione Redox Reactions I
Granulocyte Adhesion and Diapedesis
Complement System
PTEN Signaling
Semaphorin Signaling in Neurons
B
C
D
E
psu.time
Pseudo time
Col1a1
Col1a2
Runx2
Mmp13
Tnc
Spp1
Bglap
Col11a2
Smpd3
Dmp1
Irx5
Ackr3
Phex
Ptprz1
Cd109
F
Glucose to G6P
G3P to Pyruvate
Fatty Acid to Acetyl-CoA
cellType
OC
OB
P-OB
flux
G
Methionine to Cysteine
Cysteine to Pyruvate
p < 0.003
p = 3.33E-12
p = 3.12E-21
H
Mmp13/Tnc Osteoblast (P-OB)
IMP
Succinyl.CoA
Pyrimidine
AICAR
Propanoyl.CoA
Fumarate
Methionine
Xanthine
Proline
dCDP
Osteoblast (OB)
Farnesyl.PP
CDP
Malate
X.Glc.3..GlcNAc
X.Gal.2..GlcA
dCMP
Xanthine
X2OG
Succinate
Lactate
Osteocyte (OC)
XMP
Acetyl.CoA.Ox.Ac
Malate
dUMP
dCMP
Pyrimidine
Fumarate
X2OG
B.Alanine
Glutathione
Metabolites
Metabolite Enrichment


A
MSC
Osteoblast/Osteocyte
Differentiation
Col1a1 /β-actin mRNA
Sost /β-actin mRNA
Ptprz1 /β-actin mRNA
Wks
B
Up DEGs
Non DEGs
Down DEGs
q-value (-log10)
Fold change (log2)
Up DARs
Non DARs
Down DARs
C
Upregulated
DAR
DEG
339
1613
150
Downregulated
356
1089
44
D
DAR
DEG
Both DAR and DEG
R = 0.26, p < 2.2e−16
logFC (DEG)
logFC (DAR)
Tnc
Mgp
Bglap
Shroom3
Sema3a
Igf1
E
q-value (-log10)
GO:0030198~extracellular matrix organization (58)
GO:0007166~cell surface receptor signaling pathway (249)
GO:0044420~extracellular matrix component (38)
GO:0009653~anatomical structure morphogenesis (282)
GO:0045595~regulation of cell differentiation (192)
GO:0070167~regulation of biomineral tissue development (23)
GO:0009725~response to hormone (103)
GO:0001503~ossification (63)
GO:0009966~regulation of signal transduction (253)
GO:0031214~biomineral tissue development (29)
GO:0006793~phosphorus metabolic process (289)
GO:0009612~response to mechanical stimulus (38)
GO:0030500~regulation of bone mineralization (20)
GO:0001501~skeletal system development (67)
GO:0030282~bone mineralization (21)
GO:0061138~morphogenesis of a branching epithelium (33)
GO:0001763~morphogenesis of a branching structure (34)
GO:0030278~regulation of ossification (33)
GO:0048729~tissue morphogenesis (71)
Fold Enrichment
F
Scale
chr3:
88,385,000
chr16:
chr6:
chr1:
chr5:
13,500,000
chr4:
64,000,000
64,050,000
100 kb
MSC
OB/OC
Bglap
Col8a1
Mgp
Igfbp5
Sema3a
Tnc
GENCODE VM23 Comprehe
G
MSC
OB/OC
Log2 Expression Fold Change
Igfbp5
Rspo3
Slit2
Ptch1
Igf1
Sema3a
Epha2
Fgfr2
Mgp
Npnt
Tnc
Bglap
Shroom3
Serpine2
Gdpd2
Igfbp4
Col8a2
Dcn
H
MSC
OB/OC
Log2 Expression Fold Change
Ccn4
Adamts14
Bmp2
Spns2
Ptgis
Col24a1
Phex

A

Table 1. Serum biochemical analysis

	Male			Female		
	Ctl	Adenine 2 weeks	Adenine 4 weeks	Ctl	Adenine 2 weeks	Adenine 4 weeks
Pi (mg/dL)	12.22 ± 1.49	15.01 ± 0.95	17.23 ± 1.39[A]	12.53 ± 1.55	12.85 ± 1.45	14.88 ± 1.74
ALP (U/L)	103.75 ± 21.02	95 ± 10.66	113.75 ± 17.09	131.25 ± 34.16	88.75 ± 8.92	105 ± 11.18
Ca (mg/dL)	11.2 ± 0.29	11.86 ± 0.43	10.25 ± 1.28	10.5 ± 0.32	11.08 ± 0.62	11.26 ± 0.69
BUN (mg/dL)	64.51 ± 15.63	140.00 ± 23.11[B]	176.26 ± 66.17[C]	56.15 ± 6.62	115.88 ± 13.61	175.54 ± 29.14[C]
Creatinine mg/dL)	0.40 ± 0.19	0.69 ± 0.13	0.71 ± 0.20	0.50 ± 0.08	0.65 ± 0.11	0.71 ± 0.06
Iron (μM)	37.13 ± 7.81	15.38 ± 6.05[B]	23.75 ± 4.48[A]	28.38 ± 7.81	8.88 ± 4.44[B]	12.38 ± 2.88[A]

Data presented as mean ± SD. N = 4 mice per group. [A]p < 0.05; [B]p < 0.01; [C]p < 0.001 vs. Control (Ctl)