SUPPLEMENTAL INFORMATION

MMP-9 facilitates selective proteolysis of the histone H3 tail at genes necessary for proficient osteoclastogenesis

Kyunghwan Kim,[1,5] Vasu Punj,[2] Jin-Man Kim,[1] Sunyoung Lee,[1] Tobias S. Ulmer,[3] Wange Lu,[4] Judd C. Rice[1] and, Woojin An[1*]

[1]Department of Biochemistry and Molecular Biology, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089, USA

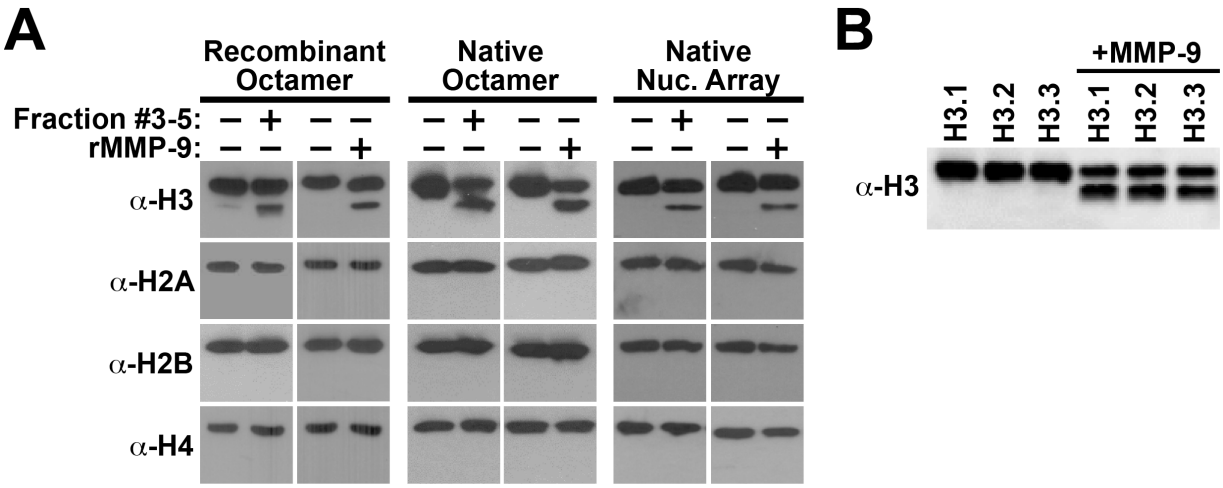
[2]Department of Medicine, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089, USA

[3]Department of Biochemistry and Molecular Biology, Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA 90089, USA

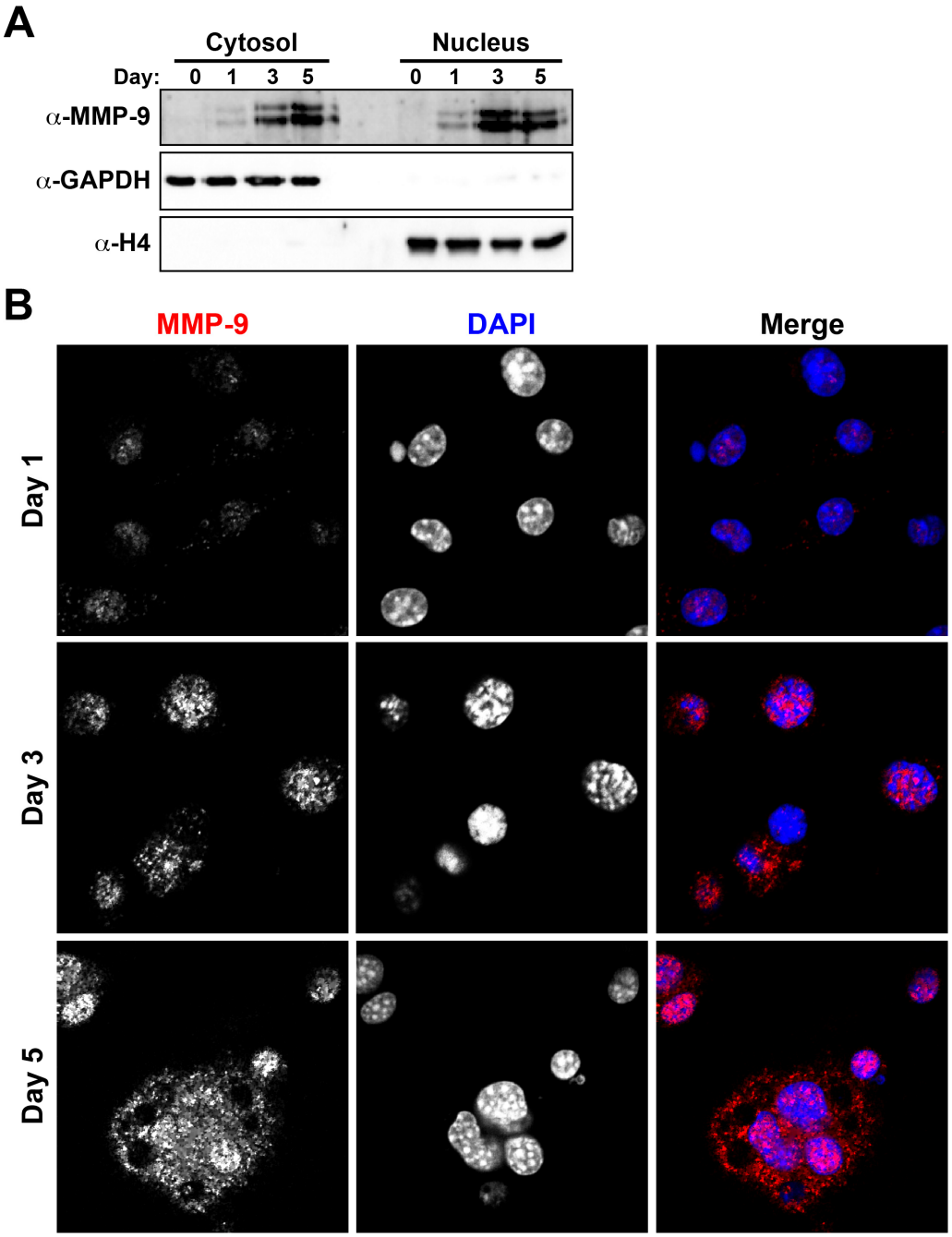
[4]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA 90089, USA

[5]Department of Biology, College of Natural Sciences, Chungbuk National University, Cheongju, Chungbuk 361-763, Republic of Korea

SUPPLEMENTAL FIGURES

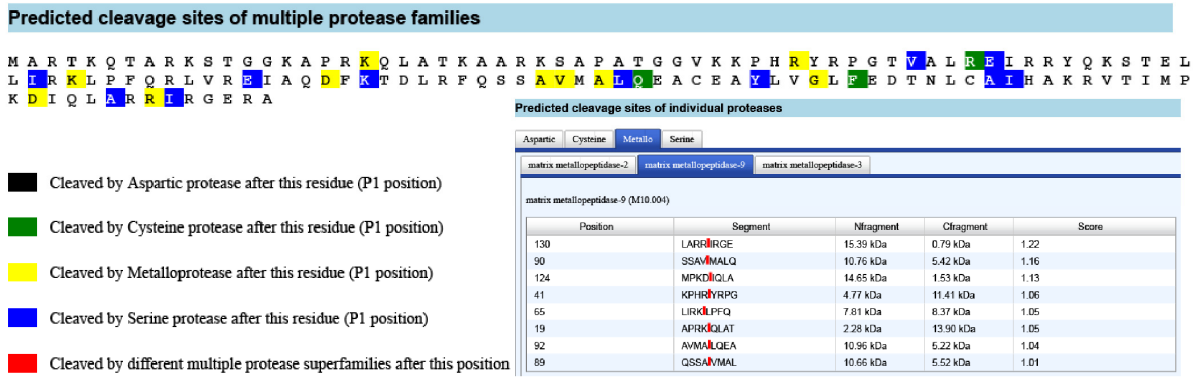


Supplemental Figure 1. **(A)** *In vitro* H3NT cleavage assays performed with the purified glycerol gradient fractions #3-5 (Fig. 1C) or recombinant MMP-9 (rMMP-9) as indicated. Substrates used in the assays were reconstituted recombinant histone octamer (left), reconstituted octamer containing purified native histones from HeLa cells (center) or reconstituted native nucleosome arrays (right). Western blot analysis was performed using antibodies that detect the C-terminal tail of each indicated core histone (left). **(B)** *In vitro* H3NT cleavage assays as described above using recombinant H3.1, H3.2 or H3.3 substrates in the absence (left) or presence (right) of rMMP-9.



Supplemental Figure 2. (A) Western blot analysis of cytosolic (left) and nuclear (right) lysates isolated in parallel from OCP-induced cells at the indicated days for MMP-9, GAPDH (cytosol control) and histone H4 (nucleus control). The volumes and total amount of protein (20 μg) loaded per lane were identical in all samples. **(B)** Confocal microscopy of OCP-induced cells co-stained for MMP-9 (red) and DAPI (blue) at the indicated days (left). OCP cells were grown on glass cover slips, fixed with 4% paraformaldehyde for 15 min and permeabilized with 0.25% Triton X-100 in phosphate-buffered saline (PBS) for 5 min. Cells were blocked with PBS + 2% goat serum (GS) in PBS for 60 min prior to incubation with an MMP-9 antibody (Santa Cruz). Cells were washed with PBS+GS, incubated with a FITC conjugated secondary antibody (Jackson Laboratory) and mounted with VectaShield with DAPI (Vector Laboratories).

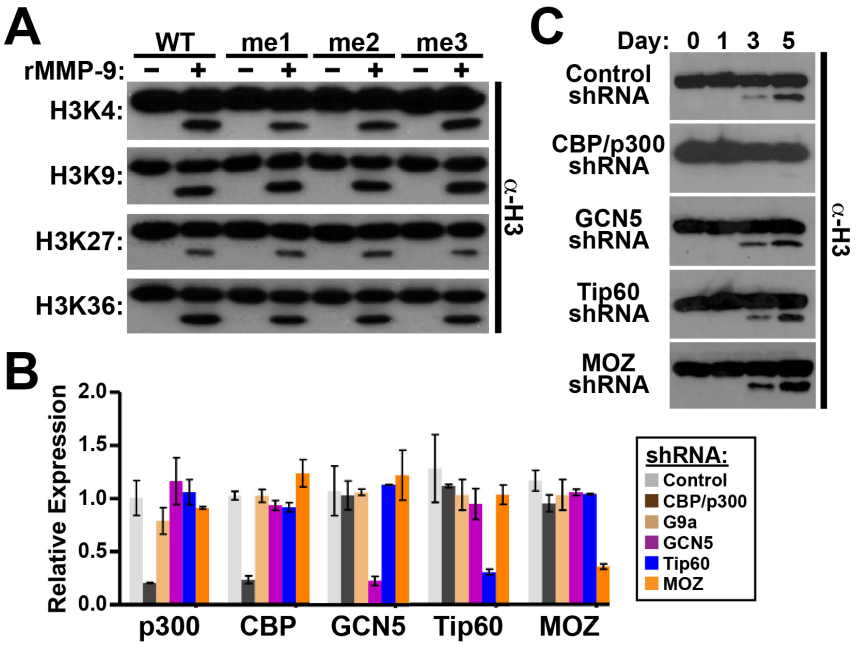
A



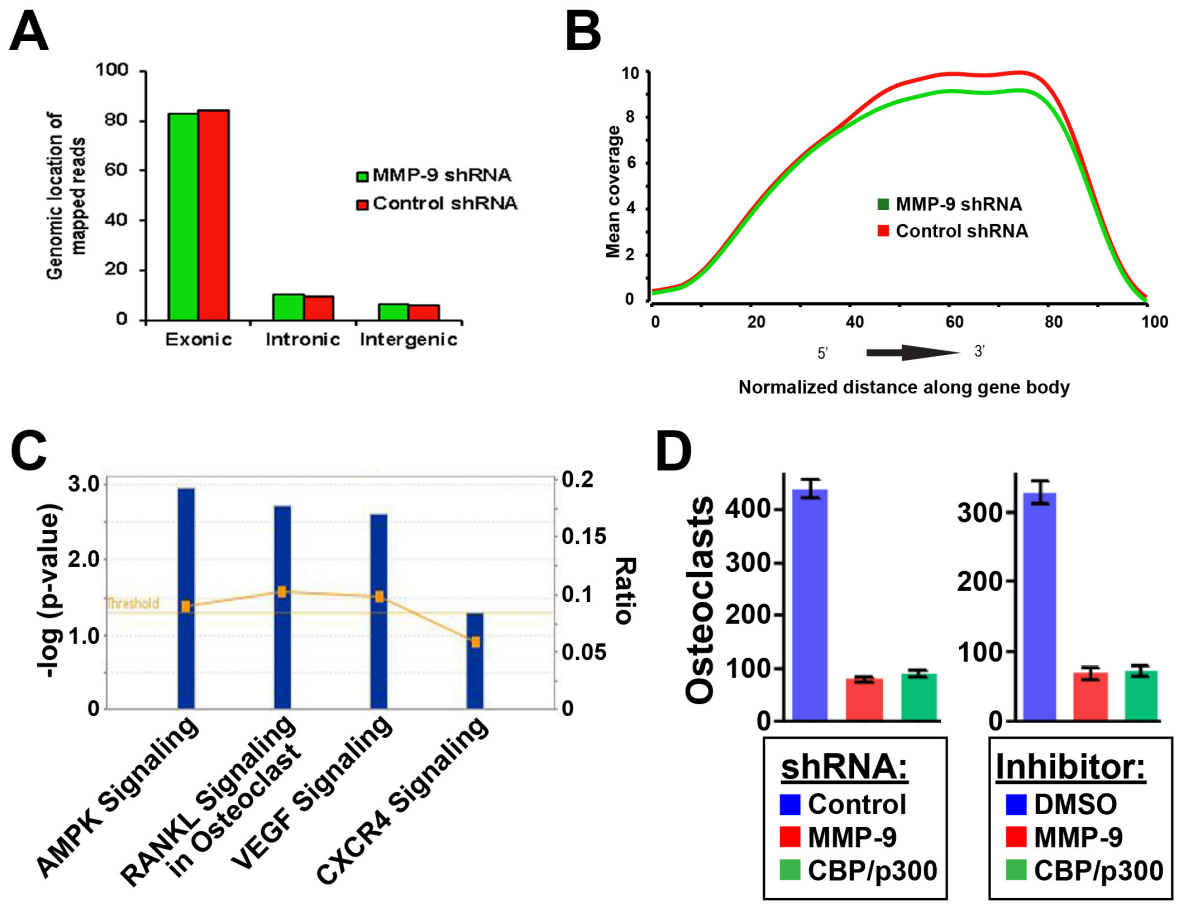
B


```
<html><head></head><body><pre style="word-wrap: break-word; white-space: pre-wrap;">#MMP9:
offset:-5.0,  thresh: 0.5
#Merops:M10.004
#Log-odds probabilities, background corrected (with offsets)
#(no constant tags in averaging)
#The score is calculated using p3-p2-p1-p1'-p2' positions
#(OFFSET is already included in the matrix!!!)
Subsites               S3       S2       S1       S1'      S2'      S3'
Weights         1        1        1        1        1        1
Gly  -3.852667 -3.901634  1.044758 -3.989192 -2.836089 0.0
Ala   1.109962 -0.124959  1.817762 -3.680012 -0.984845 0.0
Pro   3.881524 -3.313585  0.033384 -5.000000 -5.000000 0.0
Cys  -5.000000 -5.000000 -5.000000 -5.000000 -5.000000 0.0
Thr  -2.292369 -1.876678 -2.861343 -3.882145  0.864335 0.0
Ser  -2.901634 -0.327004  0.226730 -1.896952 -1.404764 0.0
Asp  -5.000000 -5.000000 -0.869464 -5.000000 -5.000000 0.0
Asn  -3.579663 -1.629947  1.275359 -2.778729 -2.589181 0.0
Glu  -5.000000 -1.480478 -1.543509 -5.000000 -1.267219 0.0
Gln  -5.000000 -1.452530 -0.826693 -1.649910 -0.167734 0.0
Lys  -5.000000  0.962497 -0.271461 -3.628073 -0.382758 0.0
His  -3.050024 -2.048770  0.636636 -1.029794  0.832291 0.0
Arg  -4.412731  2.000303 -1.932857 -2.614757  1.524154 0.0
Val   1.469224 -0.589290 -3.141908 -0.217407 -0.128551 0.0
Ile  -1.860976 -1.911967 -5.000000  0.358790  0.218149 0.0
Leu  -0.558070  0.281324 -0.815199  2.347796 -1.473090 0.0
Met  -2.204167 -0.359435 -1.135206  1.765721  1.247326 0.0
Phe  -3.541534 -0.012708  0.513814 -0.265848 -0.550190 0.0
Tyr  -5.000000  0.238469 -1.300626  1.247046  1.074855 0.0
Trp  -0.832668  0.750452  1.429943 -1.734323  0.732738 0.0
Rules
</pre></body></html>
```

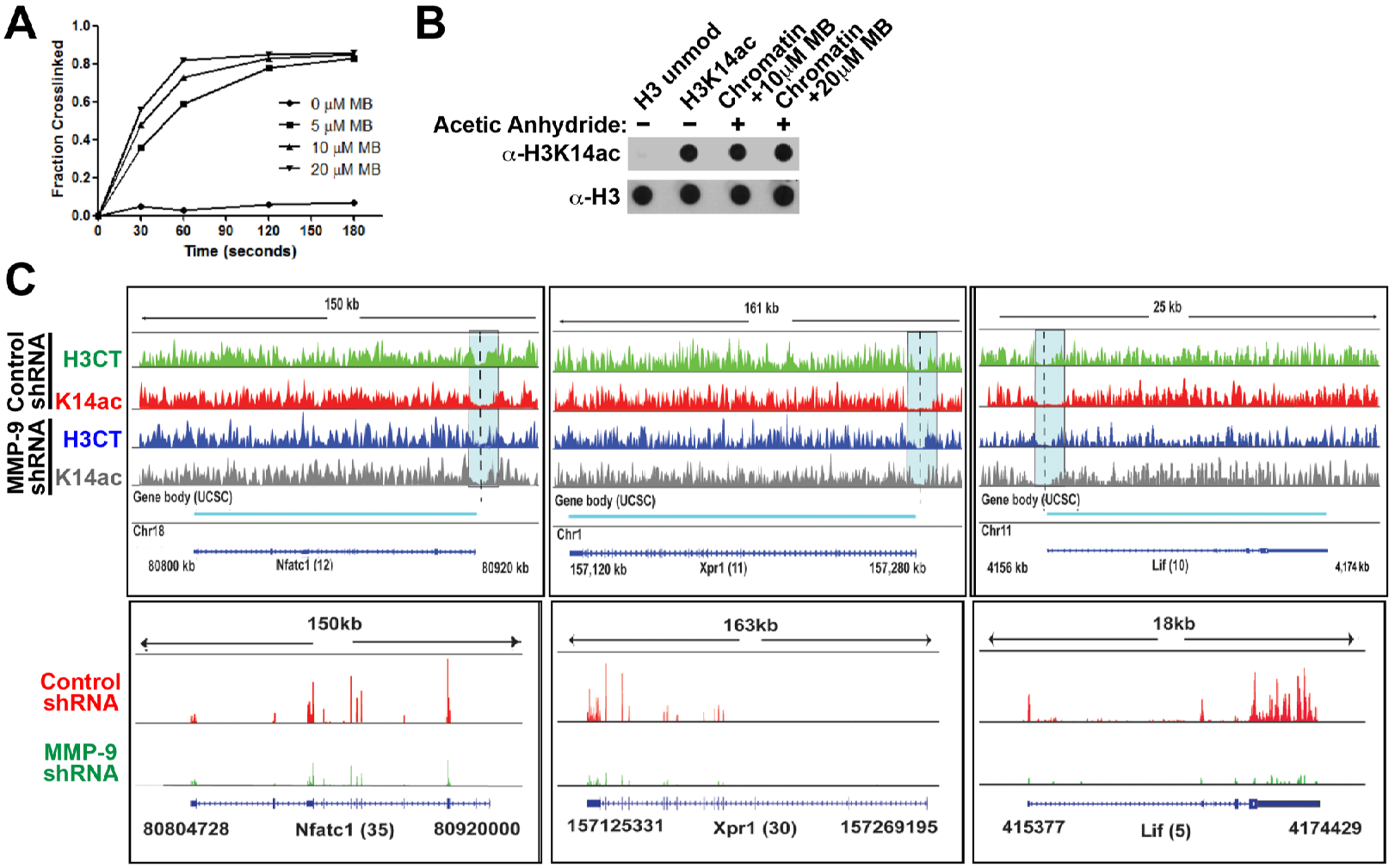

Supplemental Figure 3. (A) The human histone H3 amino acid sequence was analyzed by PROSPER (www.prosper.erc.monash.edu.au/home.html) to identify predicted sites of H3 proteolysis. The only potential cleavage site identified in the H3NT was K18, which was predicted to be cleaved by MMP-9 (inset). **(B)** The log-odds probabilities of amino acids (y-axis) in the P3-P2' position (x-axis) of substrates known to be proteolyzed by MMP-9 was generated from PMAP (www.proteolysis.org).

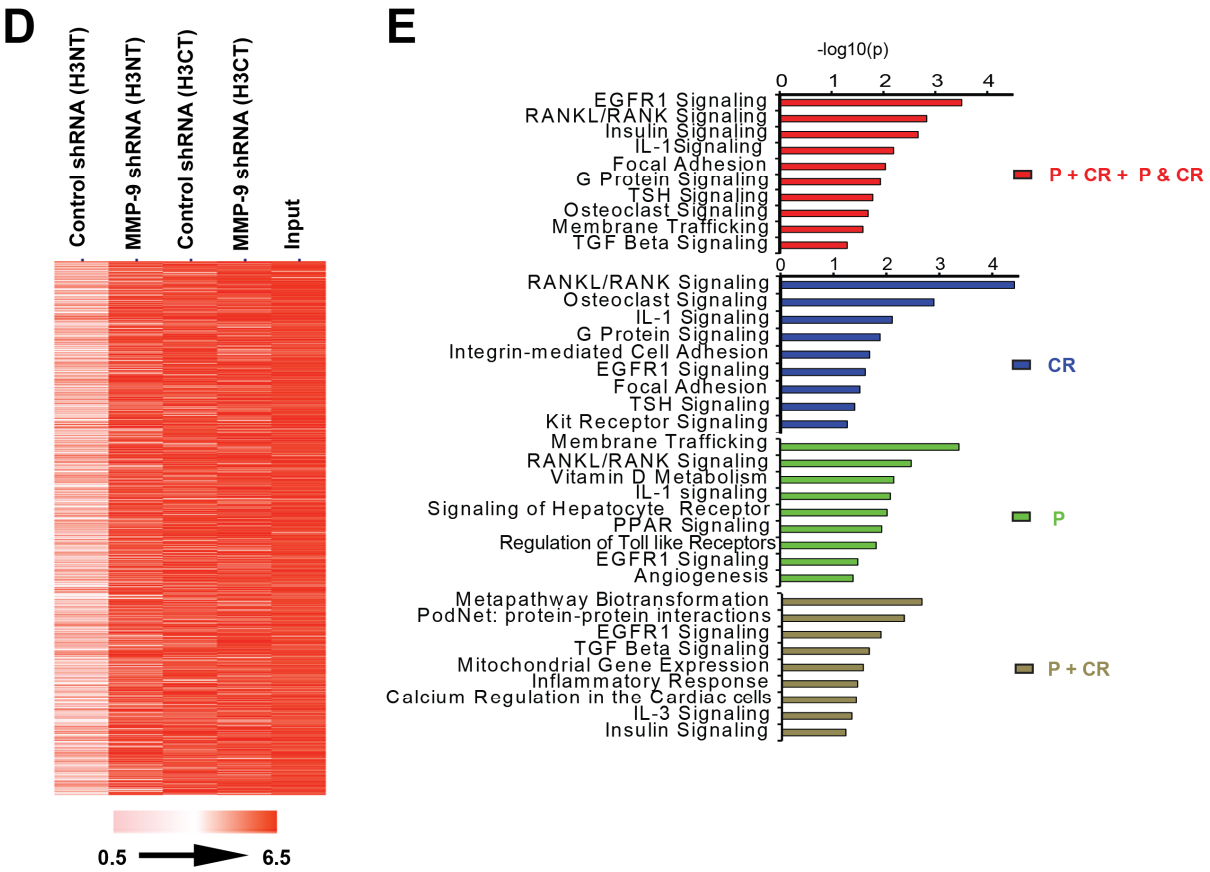


Supplemental Figure 4. (A) Recombinant H3 wild type (WT) and mono-, di- and trimethyl-lysine analogues (top) for H3K4, H3K9, H3K27 and H3K36 (left) were used as substrates in *in vitro* H3NT cleavage assays with rMMP-9. **(B)** OCP cells were transduced with H3K18 acetyltransferase-specific shRNAs as indicated. Knock-down specificity and efficiency of each acetyltransferase shRNA were determined by RT-qPCR to quantitate expression changes relative to control shRNA (y-axis) for each gene (x-axis). **(C)** Chromatin was extracted from OCP cells expressing a control or acetyltransferase-specific shRNA (left) at the indicated days post-induction (top) for Western analysis with the H3CT antibody to detect H3NT proteolysis.

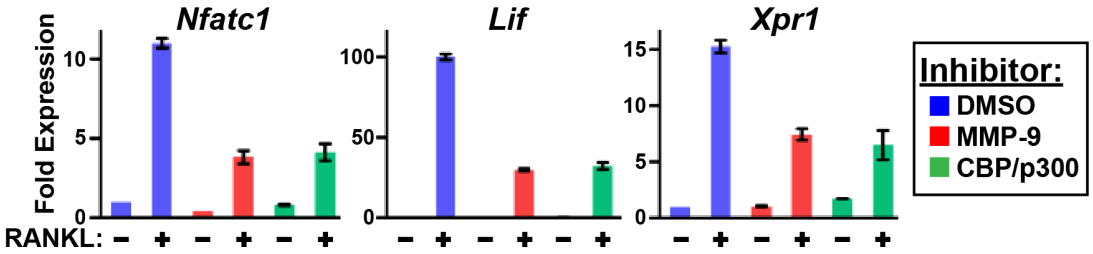


Supplemental Figure 5. (A) Genomic distribution of uniquely mapped RNA-Seq reads from control (red) and MMP-9 (green) shRNA expressing 3-day OCP-induced cells was similar with most reads mapping to coding regions of the mouse genome. **(B)** Gene body coverage plot between control and MMP-9 shRNA expressing 3-day OCP-induced cells demonstrating a similar uniform coverage across gene bodies with no significant 5' or 3' bias. (**C**) Gene ontology analysis of the 758 genes displaying impaired activation in MMP-9 depleted 3-day OCP-induced cells relative to control OCP cells. **(D)** OCP cells transduced with a control (blue), MMP-9-specific (red) or CBP/p300-specific (green) shRNA (left) or cultured with DMSO control (blue), a MMP-9-specific inhibitor (red) or a CBP/p300 inhibitor (green) (right) were treated with RANKL for 5 days (x-axis). Cells were fixed, stained with a TRAP kit (Sigma) and visualized by light microscopy (10x). The number of osteoclasts (y-axis) was determined by counting TRAP-positive cells containing 3 or more nuclei in each sample.


A
Fraction Crosslinked
1.0
0.8
0.6
0.4
0.2
0.0
0
30
60
90
120
150
180
Time (seconds)
0 μM MB
5 μM MB
10 μM MB
20 μM MB
B
H3 unmod
H3K14ac
Chromatin +10μM MB
Chromatin +20μM MB
Acetic Anhydride: – – + +
α-H3K14ac
α-H3
C
Control shRNA
H3CT
K14ac
MMP-9 shRNA
H3CT
K14ac
150 kb
Gene body (UCSC)
Chr18
80800 kb
Nfatc1 (12)
80920 kb
161 kb
Gene body (UCSC)
Chr1
157,120 kb
Xpr1 (11)
157,280 kb
25 kb
Gene body (UCSC)
Chr11
4156 kb
Lif (10)
4,174 kb
Control shRNA
MMP-9 shRNA
150kb
80804728
Nfatc1 (35)
80920000
163kb
157125331
Xpr1 (30)
157269195
18kb
415377
Lif (5)
4174429



Supplemental Figure 6. (A) OCP cells were crosslinked with different concentrations of methylene blue (0, 5, 10 and 20 µM) for the indicated times (x-axis). Chromatin was extracted and DNA-protein crosslinking efficiency was measured by SDS-chloroform-isoamyl alcohol (SDS-CIA) (y-axis). **(B)** Unmodified and K14ac H3 peptides (aa10-35) and methylene blue crosslinked chromatin (10 or 20 µM) treated with acetic anhydride were spotted on nitrocellulose for Western analysis with the indicated antibodies. **(C)** Genomic snapshots of the *Nfatc1* (left), *Xpr1* (center) and *Lif* (right) genes showing the normalized tracks of H3CT control and H3K14ac ChIPac-Seq data (top) and RNA-Seq data (bottom) from control and MMP-9 depleted 3-day OCP-induced cells as indicated. Numbers in parentheses indicate maximum sequencing depth for each sample. Grey boxes indicate the region near TSS (dashed line) displayed in Fig. 6D. **(D)** Heat map showing one dimensional clustering in a single color channel of all genes displaying H3NT cleavage (Fig. 6B and Supplemental Table 2). Euclidean distance and average linkage were used for clustering. Normalized read count values corresponding to each gene are plotted and colored proportional to the level of read counts as indicated in the color bar. **(E)** To investigate potential functional similarities and differences between the identified P, CR and P+CR H3NT-cleaved genes (Supplemental Table 2), the Wikipathway database was queried and a ranked p-value was computed for each pathway from the Fisher exact test based on the bionomical distribution and independence for probability of any gene belonging to any enriched set. The RANKL and several osteoclastogenic pathways were significantly enriched ($p<0.0001$) when all H3NT-cleaved genes were analyzed (red), consistent with RNA-Seq results (Supplemental Figure 5C). These pathways were also significantly enriched independently for P (green) and CR (blue) H3NT-cleaved genes but not for P+CR (brown, $p>0.9$).



Supplemental Figure 7. OCP cells were treated with DMSO control (blue), a selective MMP-9 inhibitor (red) or a selective CBP/p300 inhibitor (green) and cultured with (+) or without (–) RANKL for 3 days (x-axis). RT-qPCR was used to quantitatively measure fold expression changes (y-axis) of the osteoclastogenic genes *Nfatc1*, *Lif* and *Xpr1* relative to non-induced control after normalization to *β-actin* expression.

SUPPLEMENTAL METHODS

Cell transduction and inhibition

Retroviral particles were generated in Plat-E cells transfected with pMX-H3-Flag (Cell Biolabs) according to the manufacturers' instructions. Lentiviral particles were generated in HEK-293T cells by co-transfecting plasmids encoding VSV-G, NL-BH and pLKO.1-shRNA (Addgene) for MMP-9 (5'-GAGGCATACTTGTACCGCTAT), p300 (5'-CCCTGGATTAAGTTTGATAAA) or CBP (5'-TAACTCTGGCCATAGCTTAAT). OCP cells were transduced and selected with puromycin (2 μg/ml) or cultured with DMSO, 10 nM MMP-9-INI (Santa Cruz) or 1.6 μM C646 (ApexBio) for 3 days prior to differentiation.

Chromatin extraction and Western blot analysis

Chromatin was extracted by first resuspending cells in buffer A (10 mM HEPES, pH 7.4, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 5 mM β-glycerophosphate, 10 mM NaF, protease inhibitors, and 0.2% Triton X-100) and incubating on ice for 8 min. Nuclei were isolated by centrifugation (1,300g for 10' at 4°C), resuspended in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, 5 mM β-glycerophosphate, 10 mM NaF, and protease inhibitors), centrifuged (1,700g for 5 min at 4°C) and the chromatin pellet was washed three times with buffer B prior to sonication in Laemmli buffer. Western blot analysis was performed using antibodies specific for H2A, H2B, H3 and H4 (Abcam); H3K9ac, H3K14ac, H3K18ac and H3K23ac (Active Motif); MMP-9, p300 (Santa Cruz) and CBP (BioLegend); His (Novagen) and FLAG (Sigma).

Recombinant proteins and H3 cleavage assays

His-tagged proteins were generated in Rosetta 2 (DE3) pLysS *E. coli* (Novagen) and purified as previously described (Kim et al. 2008). MMP-9 (aa115-730) was purified from inclusion bodies dissolved with lysis buffer (6 M urea, 0.5 M NaCl, 5 mM imidazole, 20 mM Tris, pH 7.9), refolded in 50 mM HEPES, 0.2 M NaCl, 1M NDSB201 (Sigma) and dialyzed in 50 mM Tris, pH 7.5, 0.1 M NaCl, 5 mM $CaCl_2$, 20 uM $ZnCl_2$ and 30% glycerol. Recombinant and native histone octamers and nucleosome arrays were prepared as previously described (Kim et al. 2012). Cell extracts or recombinant proteins were incubated with histone octamer (1 μg) or nucleosomes (2 μg) in cleavage buffer (20 mM HEPES-KOH, pH 7.8, 1 mM $CaCl_2$, 20 mM KCl) for 0.5-2 hours at 37°C +/- protease inhibitors aprotinin (10 μg/ml), bestatin (130 μm), EDTA (25 mM), leupeptin (100 μM), pepstatin A (1.5 μM), PMSF (1 mM) or L006235 (20 nM). H3 tail peptides (aa10-35) were synthesized by solid-phase Fmoc/tBu chemistry and purity confirmed by ES-MS (EZBiolab). Peptides were incubated with rMMP-9 at seven different concentrations (0, 25, 50, 100, 150, 200, and 250 μM) in cleavage buffer, the reactions stopped by adding 20 mM *o*-phenanthroline (Sigma) and peptide hydrolysis was measured by adding 5 mM fluorescamine followed by detection at λ_{ex} 365 nm and λ_{em} 450 nm using a Plate Chameleon spectrofluorometer (Hidex). Kinetic parameters were determined by Michaelis-Menten analysis.

Chromatography

Nuclear extracts were fractionated on a P11 column (Pharmacia) equilibrated with BC100 buffer (20 mM HEPES-KOH, pH 7.9, 0.5 mM EDTA, 0.05% Nonidet P-40, 10% glycerol, 1 mM DTT, 100 mM KCl) as previously described (Kim et al. 2013b). H3NT active fractions were combined and dialyzed against Buffer R (10 mM HEPES-NaOH, pH7.6, 10 mM KCl, 1.5 mM $MgCl_2$, 10% glycerol, 10 mM β-glycerophosphate, 1 mM DTT), loaded onto a Q-Sepharose column (Phamacia) and eluted with stepwise increased salt concentration in BR buffer. H3NT active fractions were combined and applied to a 5-ml 15-40% glycerol gradient in BR200 buffer containing 0.1% Nonidet P-40, centrifuged in an SW 55Ti rotor (150,000g for 20 hours at 4°C) and fractions (150 μl) were collected from the top of the tube. H3NT active fractions (#3-5) were combined for protein identification by LC-MS/MS analysis.

RNA-Seq

RNA Isolation and Library preparation. RNA was prepared using the Qiagen RNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. For quality control, RNA purity and integrity were verified by denaturing gel electrophoresis, OD 260/280 ratio, and analysis on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Strand specific library perpetration was carried out using a KAPA Stranded mRNA-Seq Kit, with KAPA mRNA Capture Beads (KAPA Biosystem Wilmington, MA). Briefly, 30 ng of total RNA was reverse transcribed to cDNA using a T7 oligo(dT) primer. Second-strand cDNA was synthesized, *in vitro* transcribed, and labeled via incorporation of biotin-16-UTP. Validation of the library preparations was performed on an Agilent Bioanalyzer using the DNA1000 kit. Libraries were quantified using a Roche LightCycler96 with FastStart Essential DNA Green Master mix. Library concentrations were adjusted to 4 nM and pooled for multiplex sequencing. Pooled libraries were denatured and diluted to 15 pM and then clonally clustered onto the sequencing flow cell using the Illumina cBOT Cluster Generation Station and Cluster Kit v3-cBot-HS. The clustered flow cell was sequenced with 1X50 SE reads on the Illumina HiSeq2000 according to manufacturer's protocol. Base conversion was made using OLB version 1.9, de-multiplexed and converted to Fastq using CASAVA version 1.8 (Illumina). This resulted in approximately 50 million reads for control and MMP-9 shRNA expressing OCP cells. The library preparation and sequencing were performed in conjunction with the Sequencing Core Facility of UCLA, Los Angeles.

Bioinformatic analyses of RNA-seq. To check the general quality of raw sequencing reads, we used our in-house RNA-seq workflow (University of Southern California, Epigenome Center) which consists of several open source tools. Briefly, to check the quality of sequencing reads in FastQC, reads were trimmed on both ends based on quality score (>38) of each sequence by using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx toolkit/). Adaptor sequences were also removed. High quality reads were aligned to the mm9 genome using TopHat-2 which uses Bowtie 2 at different steps, allowing one mismatch (in conjunction with gene model from Ensembl release 61) (Kim et al. 2013a). Potential PCR duplicated reads identified with the MarkDuplictaes using *Picard* (https://github.com/broadinstitute/picard) were excluded. Further alignment quality was checked using *SamStat* (Lassmann et al. 2011), *RNA-seQC* (DeLuca et al. 2012), *Seqmonk*, *IGV* and *Qualimap* (Garcia-Alcalde et al. 2012). A total of approximately 21 million unique mapped reads for each control and MMP-9 shRNA expressing OCP cells were obtained.

Transcript coverage. We used coverage over the transcript plots to identify any problems associated with library preparation, sequencing process and distribution of reads across the genome or any bias in 5' or 3' expression using *Qualimap* and *RNA-seQC* packages. The majority of the reads (>80%) of all samples were within exonic regions and very low number of reads mapped to intronic or intergenic regions (Supplemental Fig. 5A). To investigate any 5' or 3' bias in the expression, intronic positions were removed, coverage vectors for minus-strand genes were reversed, and remaining values divided into 100 non-overlapping bins were averaged to produce a uniform 100-column matrix of gene coverage for each sample. Genes <100 bp were not included in the analysis. Gene-wise trends in coverage bias and variability were studied by taking the coefficient of variance of the original coverage vector per gene and that of the ratio of the first-quarter mean depth (5′ coverage) to fourth-quarter mean depth (3′ coverage) on the gene body. Global coverage bias per sample was plotted against normalized coverage (Supplemental Fig. 5B) as detailed previously (Graw et al. 2015). To obtain an overall robust gene expression, we used an exon model by counting all reads mapped to only exons and combining all exons to a gene so that we have read count per gene (RPKM) (Mortazavi et al. 2008). The RPKM values were adjusted globally by matching count distributions at the 75th percentile and then adjusting counts to a uniform distribution between samples. Finally differential expression was estimated by selecting transcripts that displayed significant changes ($p < 0.05$) after Benjamini and Hochberg correction using a null model constructed from 1% of transcripts showing the closet average level of observation to estimate

experimental noise (Ring et al. 2015). The gene list was further ranked using fold change criteria. GSEA and leading edge analyses were performed on the ranked list using gene sets from the C2 collection of the GSEA Molecular Signatures Database v3.0 and several gene sets associated with bone remodeling as detailed previously (Kim et al. 2015).

ChIPac-Seq

ChIP-ac. Cells were fixed with 10 μM methylene blue (Sigma) on ice and exposed to white light from a 100 W incandescent bulb at a distance of 3 cm for 3'. Crosslinking was confirmed by SDS-chloroform-isoamyl alcohol (SDS-CIA) of extracted chromatin prior to dialyzation against acetylation buffer (50 mM $NaHCO_3$, pH 8.0, 150 mM NaCl, 0.01% SDS). Chromatin was acetylated with acetic anhydride (20 mM final) in an ice-bath for 1 h while maintaining the pH at 8.0-8.2 by addition of NaOH. Acetylated chromatin was immunoprecipitated with an H3K14ac-specific antibody (Active Motif) or an H3 C-terminal (H3CT) antibody (Abcam) as previously described (Kim et al. 2015).

Library preparation. DNA libraries were constructed from ~30 ng of DNA obtained from each ChIP sample. ChIP DNA and input DNA were first band-isolated on a 2% agarose gel to obtain fragments between 150 and 350 base pairs and DNA was extracted using the QIAquick gel extraction kit (Qiagen) and eluted in 35 μL elution buffer. For input, DNA was diluted 1:5 after gel extraction. The DNA was end-repaired and adapters were ligated to samples for 15 min at room temperature. Adapters in excess were eliminated by gel purification on a 2% agarose gel. A final size selection was performed using a 2% agarose gel to obtain a library with a median length of ~230 bp which is within the recommended size range for cluster generation on Illumina's flow cell. Following PCR, reactions were cleaned using magnetic beads and re-suspended in a small volume of 10 mM Tris 8.5 (Qiagen EB). Libraries were visualized by Agilent Bioanalyzer and quantified using the KAPA Biosystems Library Quantification Kit, according to manufacturer's instructions. Sequencing with 1X50 SE reads was carried out on the Illumina flow cell of HiSeq 2000 as recommended by manufacturer (Illumina). Image analysis and base calling were carried out using RTA 1.13.48.0. Final file formatting, de-multiplexing and fastq generation were carried out using CASAVA v 1.8.2.

ChIPac-Seq quality assessment. All ChIP and input sample qualities were monitored using established strategies (Landt et al. 2012; Bailey et al. 2013). The quality matrices for raw sequence reads were obtained by FASTQc (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). The reads were processed by Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The low sequencing quality nucleotides at the 3′ end were removed, and only the reads longer than 75 bp were retained. Adaptor sequences were also removed using function *cutadapt*. The processed sequences were mapped to the mm9 genome using Burrows-Wheeler Aligner (Li et al. 2008). Further reads aligning to mitochondrial DNA, repetitive elements or unassigned sequences were discarded. This resulted in uniquely aligned reads from approximately 13 to 34 million reads covering the mouse genome. We used deeptools (https://pypi.python.org/pypi/deepTools) to estimate distribution of ChIP signal from background noise (function-*bamFingerprint*) and cross-sample normalization was achieved by scaling read depths to reads using function-*bamCoverage*. Since the DNA polymerases, used in PCR-amplifications during the library preparation, prefer GC-rich regions and may influence the outcome of the sequencing, the GCbias was estimated in each sample but no correction was applied (data not shown). To validate the ChIPac-Seq approach, a function-*bamFingerprint* was used on uniquely mapped reads, which randomly samples genome regions into a specified length (10bp bin) and counts the reads from indexed bam files that overlap with those regions. These counts were then sorted according to their rank (the bin with the highest number of reads will have the highest rank) and the cumulative sum of read counts was plotted (Diaz et al. 2012). In all cases it is possible to differentiate read density of the ChIPac sample from the input sample.

Establishment of reference loci. To calculate the differential tag density of each ChIPac sample, we used a binning approach (Diaz et al. 2012). We normalized tag density to 1x coverage to compensate for the varying sequencing depth and mapping efficiency of each sample. Such normalization is achieved by first determining the actual read coverage (number of mapped reads * fragment length / effective genome size) and then using this number to determine the scaling factor that would yield a 1x coverage, as previously detailed (Kramer et al. 2011). Since each short read represents the end of an immunoprecipitated nucleosome, every read was extended 3' to a total of 200 bp to cover an entire nucleosome prior to analysis. Signal density was calculated in sliding 1 kb windows and normalized aligned reads were considered to be within a window of the midpoint of its estimated fragment. Mid-points in each window were counted, and empirical distributions of window counts were created. Genomic bins containing statistically significant regions were identified by comparison to a Poisson background model assuming that background reads are spread randomly throughout the genome (Kim et al. 2013c).

Identification and analyses of H3NT-cleaved genes. We extracted the tag density in a ±4 kb window surrounding the TSSs using the program ngs.*plot* (https://github.com/shenlab-sinai/ngsplot) (Shen et al. 2014) and *Seqminer* (Ye et al. 2011). To identify H3NT-cleaved regions in specific genomic loci, relative signal densities were defined as read counts per bin fixed divided by the total number of aligned reads in all bins. The loci which displayed differential read count in 1 kb windows between control shRNA H3K14ac over H3CT, as well as MMP-9 shRNA samples (H3K14ac and H3CT) with $\log_2$ ratio > 0.26 or absolute difference ±1.2 were identified and annotated with closet gene (2000 bp in either direction). For visualization, alignment files were transformed into read coverage files (20 bp bin and smooth length 50 bp) to generate bigWig files. Seqmonk and IGV were used at different steps to visualize the data. Unless specified, command line option was used to extract data matrix and figures were generated in R (http://www.r-project.org). K-means clustering of tag densities +/-4 kb of TSSs was determined in R (http://www.r-project.org). Supplementary Table 2 lists the H3NT-cleaved genes identified within each cluster. To study the functional significance of all H3NT-cleaved genes as well as the independent P, CR and P+CR clusters, we quarried the Wikipathway database (Pico et al. 2008). A ranked p-value was computed from the fisher exact test based on the bionomical distribution and independence for probability of any gene belonging to any enriched set. As shown in Supplemental Fig. 6E, all three clusters as well as P and CR clusters (but not P+CR, $p>0.9$) individually displayed a significant enrichment in RANKL and associated osteoclastogenic pathways ($p<0.0001$).

PCR

qPCR was performed using PerfeCta® SYBR Green FastMix (Quanta BIOSCIENCES) and an iCycler IQ5 (Bio-Rad). Primers are: *Nfatc1* (P: 5'-GAAGTGGTAGCCCACGTGAT, 5'-TCTTGGCACCACATAAACCA; CR: 5'-GGGTCAGTGTGACCGAAGAT, 5'-GGAAGTCAGAAGTGGGTGGA; mRNA: 5'-CTCGAAAGACAGCACTGGAGCAT, 5'-CGGCTGCCTTCCGTCTCATAG), *Lif* (P: 5'-CTCTGGCTGTCCTGGAACTC, 5'-CCAGGACCAGGTGAAACACT; CR: 5'-ATCTTGTGGCTTTGCCAACT, 5'-AGTCCTTGCCTGTCTTTCCA; mRNA: 5'-AGAAGGTCCTGAACCCCACT, 5'-AGAAGGTCCTGAACCCCACT-3') and *Xpr1* (P: 5'-AGGACCTTCGGAAGAGCAGT, 5'-CAGCAAGCAGCTCATAACCA; CR: 5'-GGTGGGTTCCACTGAAAGAA, 5'-GGTTCCTCTGACCAAAAGCA; mRNA: 5'-AGGAGCGTGTCCAACATAGG, 5'-CCACGAGATGTTTCCAGGAT). Specificity of amplification was determined by melting curve analysis and all samples were run in triplicate with results averaged.

SUPPLEMENTAL REFERENCES

Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, Madrigal P, Taslim C, Zhang J. 2013. Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS computational biology* **9**: e1003326.

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G. 2012. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**: 1530-1532.

Diaz A, Park K, Lim DA, Song JS. 2012. Normalization, bias correction, and peak calling for ChIP-seq. *Statistical applications in genetics and molecular biology* **11**: Article 9.

Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**: 2678-2679.

Graw S, Meier R, Minn K, Bloomer C, Godwin AK, Fridley B, Vlad A, Beyerlein P, Chien J. 2015. Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples. *Scientific reports* **5**: 12335.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013a. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**: R36.

Kim JM, Kim K, Schmidt T, Punj V, Tucker H, Rice JC, Ulmer TS, An W. 2015. Cooperation between SMYD3 and PC4 drives a distinct transcriptional program in cancer cells. *Nucleic acids research* **43**: 8868-8883.

Kim K, Choi J, Heo K, Kim H, Levens D, Kohno K, Johnson EM, Brock HW, An W. 2008. Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. *J Biol Chem* **283**: 9113-9126.

Kim K, Heo K, Choi J, Jackson S, Kim H, Xiong Y, An W. 2012. Vpr-binding protein antagonizes p53-mediated transcription via direct interaction with H3 tail. *Mol Cell Biol* **32**: 783-796.

Kim K, Lee B, Kim J, Choi J, Kim JM, Xiong Y, Roeder RG, An W. 2013b. Linker Histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation. *Cell reports* **5**: 1690-1703.

Kim K, Punj V, Choi J, Heo K, Kim JM, Laird PW, An W. 2013c. Gene dysregulation by histone variant H2A.Z in bladder cancer. *Epigenetics & chromatin* **6**: 34.

Kramer JM, Kochinke K, Oortveld MA, Marks H, Kramer D, de Jong EK, Asztalos Z, Westwood JT, Stunnenberg HG, Sokolowski MB et al. 2011. Epigenetic regulation of learning and memory by Drosophila EHMT/G9a. *PLoS biology* **9**: e1000569.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**: 1813-1831.

Lassmann T, Hayashizaki Y, Daub CO. 2011. SAMStat: monitoring biases in next generation sequencing data. *Bioinformatics* **27**: 130-131.

Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research* **18**: 1851-1858.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621-628.

Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. 2008. WikiPathways: pathway editing for the people. *PLoS biology* **6**: e184.

Ring A, Mineyev N, Zhu W, Park E, Lomas C, Punj V, Yu M, Barrak D, Forte V, Porras T et al. 2015. EpCAM based capture detects and recovers circulating tumor cells from all subtypes of breast cancer except claudin-low. *Oncotarget*.

Shen L, Shao N, Liu X, Nestler E. 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics* **15**: 284.

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic acids research* **39**: e35.