Title:

intra-single cell sequencing (iSCseq) spotlights transcriptomic and epigenetic heterogeneity inside multinucleated osteoclast

Authors:

Hiroyuki Okada[1,2]*, Yuta Terui[3], Yasunori Omata[2,4], Masahide Seki[5], Shoichiro Tani[1,2], Junya Miyahara[2], Kenta Makabe[2], Asuka Terashima[4], Sanshiro Kanazawa[6], Masahiro Hosonuma[7,8], Fumiko Yano[9], Hiroshi Kajiya[10], Taku Saito[2], Yutaka Suzuki[5], Koji Okabe[10], Roland Baron[2,11], Sakae Tanaka[2], Ung-il Chung[1,12], Hironori Hojo[1,12]

Affiliations:

[1] Center for Disease Biology and Integrative Medicine, Graduate school of Medicine, the University of Tokyo; Bunkyo-ku, Tokyo, 113-8655, Japan.

[2] Department of Orthopaedic Surgery, the University of Tokyo

[3] Biotechnology Section 1, Product Strategy Department, Marketing Center, Life Business HQ, Yokogawa Electric Corporation

[4] Bone and Cartilage Regenerative Medicine, the University of Tokyo Hospital

[5] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, the University of Tokyo

[6] Department of Oral and Maxillofacial Surgery, Graduate School of Medicine, the University of Tokyo

[7] Department of Clinical Immuno Oncology, Clinical Research Institute for Clinical Pharmacology and Therapeutics, Showa University

[8] Showa University Pharmacological Research Center

[9] Department of Biochemistry, Showa University School of Dentistry

[10] Department of Physiological Science and Molecular Biology, Fukuoka Dental College

[11] Department of Oral Medicine, Infection, and Immunity, Harvard School of Dental Medicine, Boston, MA 02115, USA

[12] Department of Bioengineering, Graduate School of Engineering, the University of Tokyo

*Corresponding author. Email: hokada-tky@g.ecc.u-tokyo.ac.jp

Abstract:

Multinucleated giant cells like osteoclasts (OCs) often contain >10 nuclei. This raises the question of whether nuclei residing in the same cytoplasm are all regulated similarly or whether they are heterogeneous, expressing a different set of genes regulated by nucleus-specific epigenetic mechanisms. We combined imaging of living cells with confocal microscope, picking of intracellular components inside a single cell using the Single Cellome™ System SS2000 (Yokogawa, Japan), allowing next generation sequencing at high resolution for mRNA (iscRNA-seq) and for DNA methylation (iscWGBS; whole genome bisulfite sequencing) of any cellular components including a single OC nucleus. Our results reveal that individual nuclei within the same cell are heterogeneous in terms of gene expression and epigenetic regulation, possibly affecting regional cell function.

Main Text:

Multinucleated giant cells like osteoclasts (OCs) often contain >10 nuclei. OCs are hematopoietic stem cell-derived multinucleated bone resorbing cells induced by receptor activator of nuclear factor kappa B ligand (RANKL) (*1*). The main cellular function of OC is bone resorption, and the cell is highly polarized with a specific conformation of its cytoskeleton in an actin ring that seals-off an extracellular bone resorbing compartment closely associated with its function (*2*). This very specific cellular organization in the presence of multinucleation raises the question of whether nuclei residing in the same cytoplasm are all regulated similarly or whether they are heterogeneous, expressing a different set of genes and/or different levels of activity, all regulated by nucleus-specific epigenetic mechanisms.

Single cell (Sc) technology had revealed cellular diversity in terms of gene expressions (*3*). Sc technology also clarified the trajectory of cellular differentiation. Conventional scRNA-seq has already been applied for in vitro Osteoclast culture (*4*). Our group had described the difference in the OC differentiation pathway between human and mouse (*5*). However, conventional Sc techniques does not allow the separation of OCs into Scs due to their large size and fragility. Live-seq challenged recording transcriptome for living cells, however, could not reach subcellular components (*6*).

To overcome these problems and address the question of nuclei heterogeneity within OCs, we developed a combinatorial strategy of intracellular sampling with confocal microscopy and next generation sequencing from ultra-pico input of DNA or RNA called intra-sc sequencing (iSCseq) only with commercially available kits (**Fig. 1A**). iSCseq enabled us to capture the transcriptome and/or epigenome of subcellular components including an individual nucleus.

Mature OCs were large cells with a diameter of over one hundred μm *in vitro*.
Single Cellome™ System SS2000 (Yokogawa, Japan) enabled us to sample not only a single cell but also intracellular components inside OCs under visual observation. We have been able to sample single nuclei stained by Hoechst 33342. Using this approach, we have been able to sample single nuclei in the peripheral zone (**Fig. 1B, video S1**) or from the center of the cell (**Fig. 1C, video S2**). Fine tuning of the probe's tip position and the negative pressure also enabled us to pick whole large OCs (**Fig. 1D**).

Next, we combined multi-colored imaging of living cells with confocal microscope, fine sampling of subcellular components, and next generation sequencing for transcriptome and

epigenome. With SS2000, any intracellular components inside mature OCs including a single nucleus or Ca containing navel-like structure surrounding nuclei could be sampled properly and repeatedly (**Fig. 1A**). Fig.2A shows the positional annotation of subcellular components inside two osteoclasts on the culture dish. We also sampled single cells near OCs. We then constructed cDNA libraries from extracted samples allowing next generation sequencing at high resolution for mRNA (iscRNA-seq) and for DNA methylation (iscWGBS; whole genome bisulfite sequencing) of the single OC nuclei.

Sampled intracellular components including nuclei for transcriptome analysis were harvested into a microtube containing RNA lysis buffer with RNase inhibitors, centrifuged, and frozen promptly. cDNA libraries were constructed for each sample with SMART-seq single cell PLUS kit (Takara Bio, Japan).

iscRNA-seq could detect the expression of more than 400 genes from 82% of the samples. Multidimensional scaling of gene expression showed that there were two groups of genes expressed inside the same OC (**Fig. 2B**). Comparing expressed genes between the two groups with TCC and baySeq (*7*), 43 and 18 genes were extracted as differentially expressed genes (DEGs) in each group (**Fig. 2C**). Gene ontology (GO) analysis of DEGs, revealed that transporter activity and plasma membrane related GO terms were overexpressed in group 1. In contrast, group 2 exhibited histone acetylation and DNA replication related GO terms (**Fig. 2D**). These results demonstrated heterogeneity of gene expression among nuclei present inside the same mature OC. In addition, GO analysis suggested that specific epigenetic mechanisms might control this differential pattern of gene expression.

During the differentiation process, combinations of transcription factors are dynamically changed under epigenetic control, including histone modification (*8*). Nuclear factor of activated T-cells (NFATc1) is a master regulator of OC differentiation (*9*). Status of tri-methylation of lysine 27 and acetylation on histone H3 affect the NFATc1 activity as a transcriptional factor of OC specific genes (*10*). DNA methylation is another epigenetic controller and has been well investigated in developmental and cancer biology (*11*). Bisulfite sequencing detects cytosine methylation at single cell resolution (*12-14*).

For DNA methylome, sampled nuclei or a single nucleus were dropped into microtube containing DNA digestion buffer plus proteinase K. We adopted Pico Methyl-Seq Library Prep Kit (Zymo research, U.S.) for library construction of whole genome bisulfite sequencing

(WGBS). Bismark was used to align and call methylated cytosine on the mm10 reference genome (*15*).

We picked up and performed WGBS for nuclei (No. 51, 52) or a single nucleus (No. 57, 58) corresponding to group 1 (No. 51, 58) or group 2 (No. 52, 57) of iscRNA-seq to determine the role of the relative position within the same OC. iscWGBS showed that methylated patterns in CpG islands were similar among nuclei or a single nucleus (**Fig. 3A**). Principal component analysis showed a different pattern of DNA methylation according to the group (**Fig. 3B**). The proportion of methylation was higher and the coverage length longer in group 1 than in group 2 (**Fig. 3C**). dmrseq (*16*) extracted 42 differentially methylated regions (DMRs) in group 1 and 38 DMRs in group 2 (**Fig. 3D**). In four differentially methylated blocks (DMBs), there was one DMB containing exons in chromatin 15. This DMB had a potential to control secretion of hormone and phosphatidyl-inositol related Calcium physiology in GO terms (**Fig. 3E**). Thus, intra-single cell sequencing (iSCseq) shows that there is heterogeneity of gene expression among the nuclei present in the same OC, and this heterogeneity appears to be determined by DNA methylation.

To validate our iSCseq approach, we then performed integrative analysis with two previously published scRNA-seq studies of *in vitro* OC differentiation which used a traditional microfluid-based technique (*4, 5*). Five datasets were integrated on a UMAP **Fig. 4A** using 72 anchoring genes calculated with Seurat and scTransform (*17, 18*). The positions of our iscRNA-seq with global trajectory and local RNA velocity information supported the finding that mature osteoclasts contain multiple types of subcellular components in terms of differentiation stage (**Fig. 4B**). Feature plots of osteoclast related genes (**Fig. 4C**) and GO terms (**Fig. 4D**) suggested that some, but not all, of the subcellular components exhibited the characteristic of mature OCs as acid producers. Note that we also captured single cells in the vicinity of OCs (No. 18 and 20 in **Fig.2A**) and found strong acid producing cells around mature OCs.

Traditional single cell technology with microfluid devices has clarified cellular heterogeneity in terms of gene expression (*19*). However, there remain limitations as a tradeoff for high throughput. For instance, morphological information is lost through microfluid devices. Besides, multinucleated giant cells like OCs cannot flow in microfluid channels due to their large diameter and fragility. Our integrative approach of scRNA-seq and iscRNA-seq recapitulated the heterogeneity not only of gene expression patterns but also of differentiation stage inside the

same OCs. Geometrical and morphological annotation combined with omics approaches will enhance our understanding of the biology of cells.

iSCseq method enables transcriptomic and epigenetic experiments for living cell population or cellular components. To our knowledge, this is the first report of combinatorial experiments of sampling cellular components repeatedly from the same cell and decoding subcellular heterogeneity of gene expression. In addition, different levels of DNA methylation among nuclei can regulate differentially the expression of genes in individual nuclei. The exact control of tip in z axis with SS2000 was a key to success to collect the desired subcellular components from living cells. Giant cells like OCs constitute useful cell models to investigate hard-to-detect intracellular components in small cells.

In conclusion, using a newly developed single cell methodology, our study demonstrates the heterogeneity of gene expression and its epigenetic regulation among individual nuclei inside the same mature OC. We propose that in multinucleated cells, gene expression levels of each nucleus are finely tuned.

Acknowledgments:

Live cell imaging and intra-cellular sampling were technically supported by Yokogawa Electric Corporation from the planning stage of experiments. Yuta Terui is an employee of Yokogawa Electric Corporation. Our next generation sequencing was succeeded with highly elaborate manufacturer's kit by Takara bio and by Zymo research.

Author contributions:

Conceptualization: HO, SK, MH, HK, KO, RB, S. Tanaka

Methodology: HO, YT, YO, MS, HK, YS, HH

Investigation: HO, YT, YO, S. Tani, JM, KM, AT, FY

Visualization: HO, YT, S. Tani, JM

Funding acquisition: HO, SK, TS, YS, S. Tanaka, UC, HH

Project administration: HO, HH

Supervision: RB, TS, YS, S. Tanaka, HH

Writing – original draft: HO

Writing – review & editing: HO, S. Tanaka, RB, UC, HH

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All videos, data, codes, and materials were included in supplementary materials. Raw fastq files will be public and downloadable (accession numbers will be shared before publication.)

Supplementary Materials: **Materials and Methods; Figure S1; Video S1, S2**

References and Notes

1. H. Yasuda *et al.*, Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL. *Proc Natl Acad Sci U S A* **95**, 3597-3602 (1998).
2. S. Tanaka *et al.*, c-Cbl is downstream of c-Src in a signalling pathway necessary for bone resorption. *Nature* **383**, 528-531 (1996).
3. M. D. Luecken, F. J. Theis, Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* **15**, e8746 (2019).
4. M. Tsukasaki *et al.*, Stepwise cell fate decision pathways during osteoclastogenesis at single-cell resolution. *Nat Metab* **2**, 1382-1390 (2020).
5. Y. Omata *et al.*, Interspecies Single‐Cell RNA‐Seq Analysis Reveals the Novel Trajectory of Osteoclast Differentiation and Therapeutic Targets. *JBMR Plus*, (2022).
6. W. Chen *et al.*, Live-seq enables temporal transcriptomic recording of single cells. *Nature*, (2022).
7. T. Osabe, K. Shimizu, K. Kadota, Accurate Classification of Differential Expression Patterns in a Bayesian Framework With Robust Normalization for Multi-Group RNA-Seq Count Data. *Bioinform Biol Insights* **13**, 1177932219860817 (2019).
8. T. Yasui *et al.*, Epigenetic regulation of osteoclast differentiation: possible involvement of Jmjd3 in the histone demethylation of Nfatc1. *J Bone Miner Res* **26**, 2665-2671 (2011).
9. H. Takayanagi *et al.*, Induction and activation of the transcription factor NFATc1 (NFAT2) integrate RANKL signaling in terminal differentiation of osteoclasts. *Dev Cell* **3**, 889-901 (2002).
10. N. Izawa *et al.*, Cooperation of PU.1 With IRF8 and NFATc1 Defines Chromatin Landscapes During RANKL-Induced Osteoclastogenesis. *J Bone Miner Res* **34**, 1143-1154 (2019).
11. Y. Yin *et al.*, Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).
12. M. Farlik *et al.*, Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep* **10**, 1386-1397 (2015).
13. P. Zhu *et al.*, Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet* **50**, 12-19 (2018).
14. S. Gravina, X. Dong, B. Yu, J. Vijg, Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol* **17**, 150 (2016).
15. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).
16. K. Korthauer, S. Chakraborty, Y. Benjamini, R. A. Irizarry, Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* **20**, 367-383 (2019).
17. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).
18. S. Choudhary, R. Satija, Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol* **23**, 27 (2022).
19. X. Zhang *et al.*, Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell* **73**, 130-142 e135 (2019).

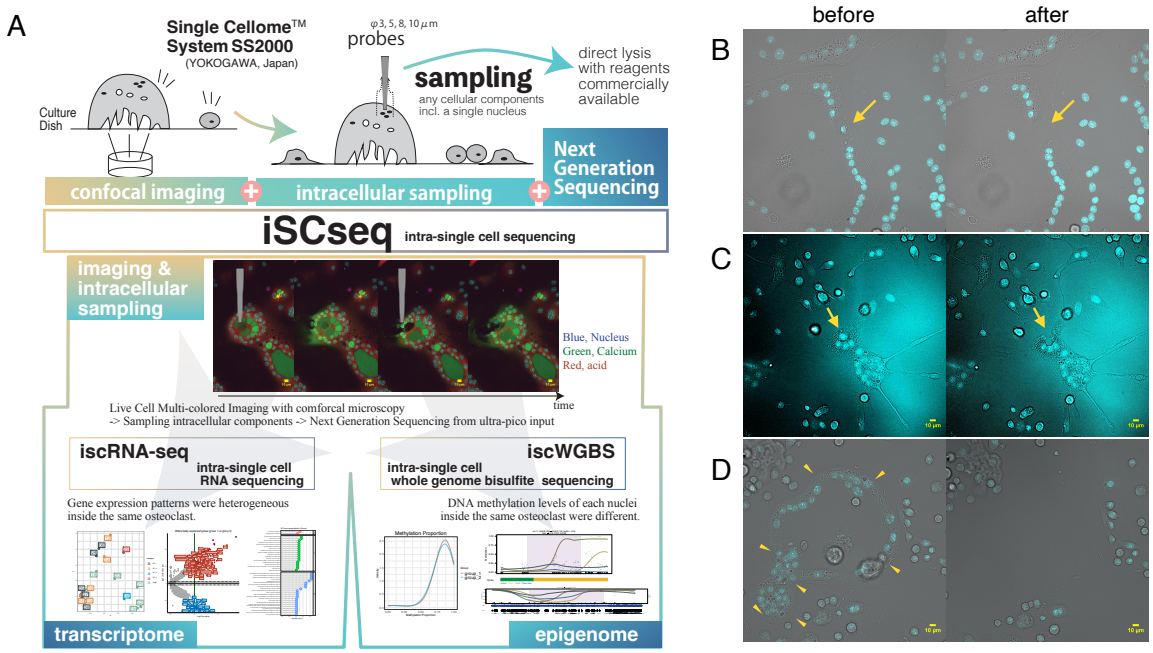

Figure. 1. Outline of intra-single cell sequencing (iSCseq) and snapshot of sampling cellular components from mature osteoclast.

(**A**) Schematic presentation of sampling experiments. Single Cellome™ System SS2000 (YOKOGAWA) took multi-colored images with confocal microscopy and sampled any nucleus or cellular component by a sharp probe with a diameter of three to ten μm, whose z position was finely adjusted. On the center, representative case of sampling intracellular components from living osteoclasts (OCs) with Calcium assay. Blue, Nuclei stained with Hoechst 33342; Green, Calcium stained with Fluo-4 NW Calcium Assay Kit; Red, acid environment stained with pHrodo™ Red AM Intracellular pH Indicator. (**B-D**) Representative cases of sampling each single nucleus inside multinucleated OCs from peripheral zone (**B**) and from serried nuclei (**C**), and sampling whole large osteoclast (**D**). Blue, Nuclei stained with Hoechst 33342.

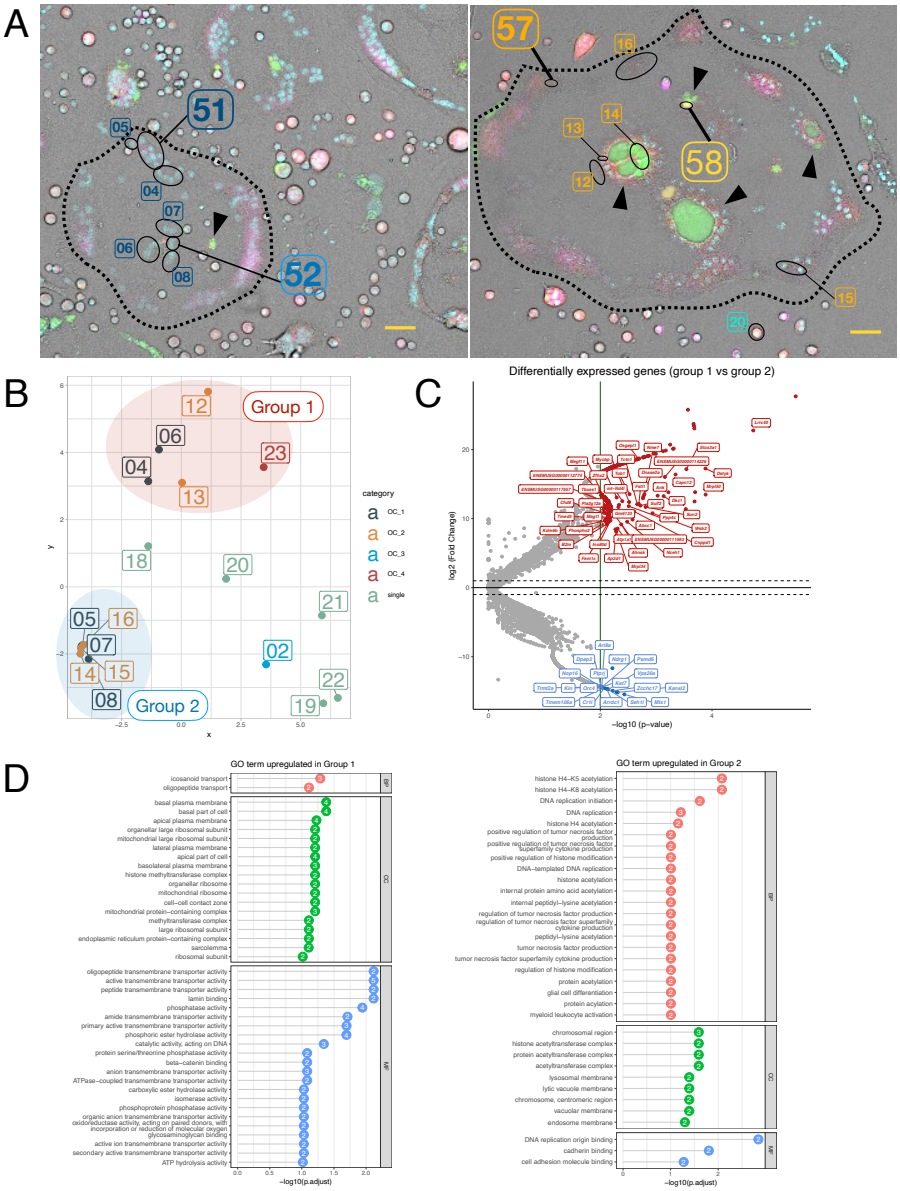

Figure. 2. Gene expression patterns in each nucleus are heterogeneous even in the same osteoclast– evidence from intra-single cell RNA-seq (iscRNA-seq).

(**A**) Geometrical information of targeted mature osteoclasts (OCs). Blue, Nuclei stained with Hoechst 33342; Green, Calcium stained with Fluo-4 NW Calcium Assay Kit; Red, acid environment stained with pHrodo™ Red AM Intracellular pH Indicator; Purple, Tublin stained with SiR-tublin kit. Bar = 50um. (**B**) Multi-dimensional scaling of gene expression patterns from 17 samples including a single nucleus. (**C**) Differentially expressed genes comparing between in group 1 and group 2 cellular components. Vertical axis is log2 fold change of gene expressions and horizontal axis is minus log10 p values. (**D**) Highly represented gene ontology terms in group 1 (left) and in group 2 (right). Circled numbers represent the number of differentially expressed genes related to the GO term.

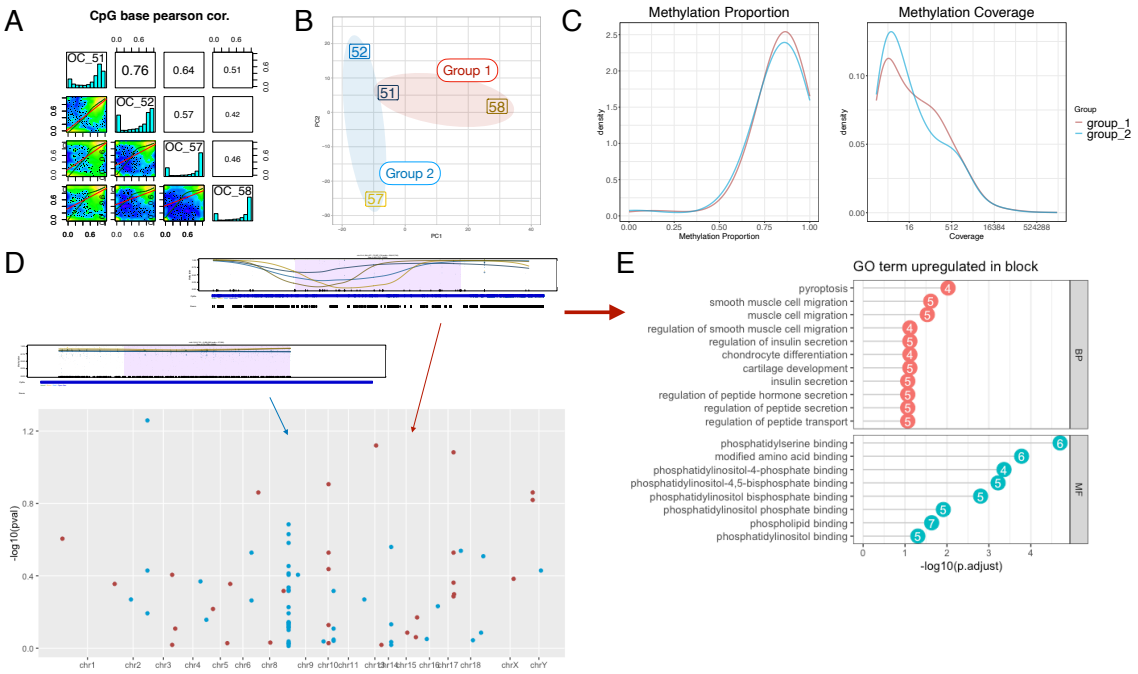

Figure 3. DNA methylation level is heterogeneous inside the same osteoclast–evidence from Intra-single cell whole genome bisulfite-seq (iscWGBS-seq).

(**A**) Correlation plots of DNA methylation patterns on CpG genomic areas reported with Bismark. (**B**) Principal components analysis of DNA methylation in nuclei (No. 51, 52) or in a single nucleus (No. 57, 58). Group 1 or group 2 label is corresponding to Fig. 2B as geometry. (**C**) General methylation proportion (left) and coverage length (right) calculated by dmrseq. (**D**) Manhattan plot of differentially methylated regions on genome (lower) and representative differentially expressed blocks (DMBs) by dmrseq. (**E**) Gene ontology analysis of chr15: 41,244,437 – 70,857,176, the area only containing exons in four DMBs. Circled numbers represent the number of genes included in this region.

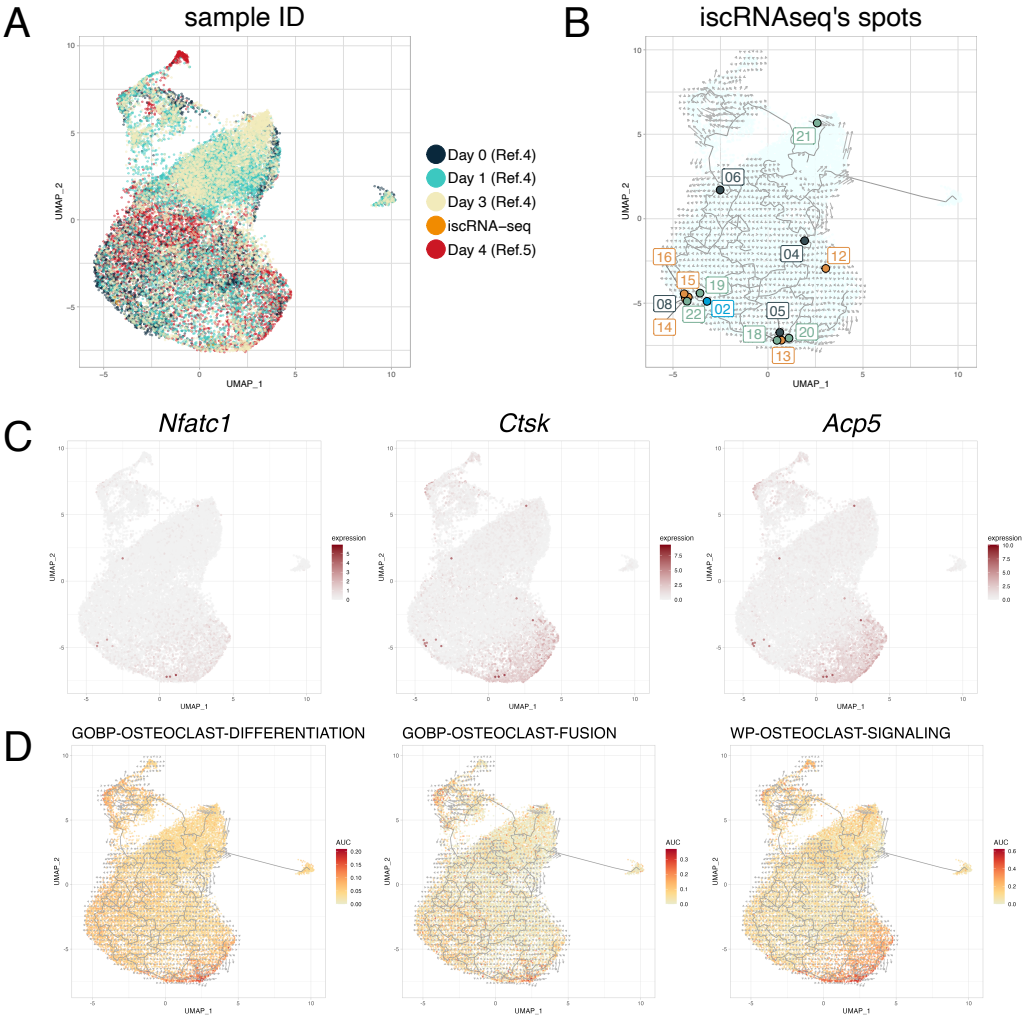

Figure 4. Integrative analysis of intra-single cell RNA-seq (iscRNA-seq) and previous scRNA-seq of *in vitro* osteoclast differentiation.

(**A**) Integrative UMAP of iscRNA-seq and scRNA-seq datasets from Dr. Tsukasaki's team (post RANKL stimulation day 0, day 1, day 3) and one from Dr. Omata's team (post RANKL day 4). (**B**) Highlights of iscRNA-seq spots on the integrative UMAP with trajectory by Monocle3 and RNA velocity by Velocyto.R. Colors and numbers corresponds to ones in fig. 2A. Arrows of RNA velocity was calculated referring to Dr. Omata's dataset. (**C**) Feature plots of osteoclast specific genes on the UMAP. (**D**) Feature plots of gene ontology term calculated by AUCell on the UMAP.

Materials and Methods

Mice

Wild-type mice on a C57BL/6 background were purchased from Oriental Kobo Industrial Company (Tokyo, Japan). Mice were euthanized by cervical dislocation to obtain bone marrow cells.

Primary bone marrow cell cultures and in vitro osteoclast formation assay

Bone marrow cells of lower limbs from mice were harvested in MEMalfa with L-Glutamine and Phenol Red (Fujifilm Wako Pure Chemical corporation, Osaka, Japan) containing 10% fetal bovine serum (FBS; JRH Biosciences, Kansas, U.S.) with 25 ng/mL macrophage colony-stimulating factor (M-CSF; R&D Systems, Minnesota, U.S.) and 1% penicillin-streptomycin (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.) for three days to produce BMMs. The BMMs were cultured in on 35mm Glass Bottom Dish (Collagen Coat) φ14mm (Matsunami Glass, Osaka, Japan) in the presence of 50 ng/ml M-CSF (PeproTech Ltd., New Jersey, U.S.), 50 ng/ml soluble RANKL (Fujifilm Wako Pure Chemical corporation). More than three nucleated with larger sizes compared to other types of cells were counted as osteoclasts.

Live cell staining

One night before imaging and sampling from osteoclast, Tubulin was stained with 500nM SiR-tubulin Kit (Cytoskeleton, Denver, Colorado, U.S.). One hour before sampling, Nuclei were stained with 0.2 µg/mL Hoechst 33342, Trihydrochloride, Trihydrate (Invitrogen, Waltham, Massachusetts, U.S.) in PBS incubated at 37°C for 10 minutes. Cells to be sampled were stained with pHrodoTM AM Ester (Thermo Fisher) 2µL and PowerLoadTM concentrate (Thermo Fisher) 20µL in 2mL MEMalfa per a 35mm dish incubated at room temperature for thirty minutes. Then, pH probe was washed away thoroughly with PBS. Finally, cells to be sampled were loaded with calcium indicator Fluo-4 (Fluo-4 NW Calcium Assay Kit, Invitrogen) as per manufacturer's instruction.

Sampling of intracellular components

Cellular components including a single nucleus were sampled with Single Cellome™ System SS2000 (Yokogawa Electric Corporation, Japan). Before sampling, we obtained fine multi-colored images with confocal microscopy to confine the sampling target including a single nucleus. When sampling cellular components, SS2000 finely tuned the position of probes in Z axis and sucked any cellular components into a 10µm glass probe with the set negative pressure. The exact control of tip with SS2000 was a key collect subcellular components from living cells. After sampling cellular components, probes maintain a little positive pressure not to pull in additional components, and finally carry onto microtubes with prechilled dissolution buffer at the right depth.

The buffer for RNA sequencing was 11.5µL lysis buffer with 0.4% volume of RNase inhibitor. Both reagents were included in SMART-Seq Single Cell PLUS kit (Takara Bio, San Jose, CA, U.S.). Just after sampling into the lysis buffer, microtubes containing samples were centrifuged briefly at 2,000 rpm and frozen at -80°C.

The buffer for DNA bisulfite sequencing was 20µL DNA digestion buffer (Zymo Research, Irvine, CA, U.S.) with 10% volume of proteinase K (Zymo). Just after sampling, microtubes were centrifuged at 2,000 rpm and kept at 4°C.

Library construction for intra-single cell RNA-seq with micro input

SMART-Seq Single Cell PLUS kit (Takara Bio) was basused for cDNA library construction basically as manufacturer's instruction. Frozen samples dissolved in lysis buffer with 0.4% volume of RNase inhibitor in microtubes were thawed, 1μL 3' SMART-Seq CDS Primer IIA was added to each microtube, then within the same microtubes incubated at 72°C for three minutes promptly. Following cDNA synthesis was performed as manufacturer's protocol. In cDNA amplification step, PCR cycles was 20, which is the maximum cycles for minimum input of RNA. Purification of amplified cDNA was done with AMPure XP PCR purification kit (Beckman Coulter, Brea, California, U.S.). cDNA quality was checked with Bioanalyzer High Sensitivity DNA kit (Agilent, Santa Clara, CA, U.S.) before library preparation steps. Samples of more than 10 times cDNA concentration compared with negative control by smear analysis within 200-9,000 bp range was included. Only one in 17 samples was excluded for the criteria. The maximum to minimum ratio of cDNA concentration (conc) was over 700 in the included cDNA samples. Samples were categorized to high (> 600 pg/μL), middle, and low (< 60 pg/μL), conc groups. Samples in high and middle conc group were diluted to the save as minimum conc in each group. Low conc samples were not diluted.

SMART-Seq Library Prep Kit including SMART-Seq Single Cell PLUS kit was used for library preparation. Library amplification step, PCR cycles were determined corresponding to the concentration categories; 14 cycles for high conc, 16 for middle conc, 19 for low conc, respectively. After library amplification, samples were pooled, then purified with AMPure once. Pooled amplified library was qualified with Bioanalyzer High Sensitivity DNA kit. Final concentration was 3,781 pg/μL.

Next generation sequencing of RNA-seq libraries

Pooled libraries from 17 samples were sequenced at paired-end 150 bp using HiSeq X Ten (Illumina, San Diego, CA, U.S.) by Macrogen Japan (Tokyo, Japan).

Library construction for intra-single cell whole genome bisulfite sequencing (WGBS)

Pico Methyl-Seq Library Prep Kit (Zymo research) was used for making WGBS libraries basically as manufacturer's instruction. Because of low input of DNA such as a single nucleus, in clean-up process, higher concentration of DNA (3:1 ratio of DNA binding buffer) was adopted. In amplification process, number of PCR cycles was 11 referring to the previous paper (*1*). In the amplification process with the index primer, Dual-indexed primers (5μM) 2.0μL in Unique dual index kit (Takara bio) was used instead of 0.5 μL Index Primer in the default Zymo kit. cDNA input was decreased as total volume was 25 μL. cDNA quality was qualified with Bioanalyzer High Sensitivity DNA kit (Agilent, Santa Clara, CA, U.S.). After purification with 1.8x AMPure XP beads (Beckman Coulter), primer dimers around 60 bp was observed. We added size selection of the cDNA library with 1.0x AMPure XP beads. Final concentration of cDNA libraries within 150 to 500 bp ranged from 497 to 1,581 pg/μL.

Next generation sequencing of WGBS libraries

Pooled cDNA libraries from 4 samples were sequenced at paired-end 150 bp using NovaSeq 6000 (Illumina) by Macrogen Japan (Tokyo, Japan).

Bioinformatical analysis of intra-single cell RNA-seq (iscRNA-seq)

Raw fastq files were qualified and trimmed using fastp ver. 0.23.2 (*2*). Preprocessed data was mapped onto the mouse genome mm10 by STAR ver. 2.7.9a (*3*). Gene expression level was summarized into count matrix using RSEM (*4*).

R ver. 4.2.1 (*5*) was used for processing and visualizing count matrix. Trimmed mean of M values normalization was performed by TCC ver. 1.36.0 (*6*) internally with edgeR ver. 3.38.4 (*7*). Multidimensional scaling (MDS) was performed with plotMDS() function in limma ver. 3.52.2 (*8*) and visualized with ggplot2 (*9*).

From MDS results, we extracted two groups containing 5 and 6 samples respectively from 17 samples by similarity of gene expression patterns. Differentially expressed genes (DEGs) were detected by estimateDE() function in TCC (Data S1). DEGs were ranked with baySeq (*10*) and defined as follows; log2 fold change >1 or <-1, p.value < 0.10 and gene expressions detected at least four samples in group 1 and three samples in group 2 categorized in Fig. 2B. Gene ontology analysis for sets of DEGs was performed with enrichGO() function in clusterProfiler ver. 4.4.4 (*11*) with default settings.

Bioinformatical analysis of intra-single cell WGBS (iscWGBS)

Raw fastq files were preprocessed and trimmed using fastp ver. 0.23.2 (*2*). Bismark ver. 0.22.3 was used as DNA methylation caller (*12*). Reference genome for DNA methylation call was constructed from GRCm38/mm10 from Genome Reference Consortium [GCA_000001635.2], then aligned using Bismark internally with bowtie2 ver. 2.4.5 (*13*) setting non_directional option according to Zymo Pico-Methyl technique. DNA methylation was extracted by bismark_methylation_extractor function ignoring the first 10 bp from both reads and making cytosine_report.

From CpG_report by Bismark was assembled into BS object by bsseq (*14*). Differentially methylation analysis was basically performed and inspired by DMRichR ver. 1.7.4 (*15, 16*). Differentially methylated regions (DMRs) between group 1 and group 2 in Fig. 3B were detected by dmrseq ver. 1.16.0 (*15*). List of DMRs is Data S2. Manhattan plot (Fig. 3D) was illustrated using plotGrandLinear() function in ggbio ver. 1.44.1 (*17*). Differentially methylated blocks (DMBs) were defined as regions covering at least 500 CpGs or 50,000 bp. List of DMBs is Data S3. Exons including DMBs were extracted and summarized into GO terms with in clusterProfiler ver. 4.4.4 (*11*).

Integrative analysis of conventional single cell RNA-seq and iscRNA-seq

Seurat ver. 4.0.6 (*18*) and sctransform ver. 0.3.4 (*19*) were used in the integrative analysis of conventional scRNA-seq and iscRNA-seq. Four datasets of scRNA-seq, GSM4419695 (post RANKL stimulation day 0), GSM4419699 (day 1), GSM4419703 (day 3) by Dr. Tsukasaki (*20*), and published undeposited data (day 4) by Dr. Omata (*21*) were included. Before merging datasets, entrez IDs were transformed to gene symbols with EnsDb.Mmusculus.v79 (*22*). Mitochondrial gene percentage and cell cycle score was calculated cell by cell. In datasets integration process, first SCTransform() for each dataset was done with arguments vars.to.regress = c("percent.mt","S.Score","G2M.Score") and vst.flavor="v2", then SelectIntegrationFeatures() with nfeatures = 2000, PrepSCTIntegration(), FindIntegrationAnchors() with dims = 1:8 and k.score =10, and finally IntegrateData() with 72 anchoring genes. After RunPCA(), we determined dims as 1:15 with ElbowPlot(). Next clustering with FindNeighbors(), FindClusters() and tuning resolutions with clustree ver. 0.5.0 (*23*), resolution was set to 0.4. adopted and Uniform Manifold Approximation and Projection

(UMAP) was made for dimension reduction (*24*). Roughly even distribution of percent.mt, cell cycles and sampleID on UMAP were confirmed with DimPlot on UMAP (Fig. S1 and Fig. 4A).

Global trajectory was calculated by Moncle3 (*25, 26*), and local RNA velocity was calculated by velocyto.py ver. 0.17.16 and Velocyto.R 0.6 (*27*). Both information were accumulated onto the same UMAP visualized with ggplot2 (*9*). Feature plots of marker genes were made by Seurat and visualized with ggplot2.

Gene ontology (GO) score for all the GO terms on Mouse Gene Set Resources (*28*) was calculated cell by cell using AUCell ver. 1.18.1, which is internally used in gene regulatory analytic tool SCENIC (*29*), and visualized onto the UMAP. We call this type of integrative GO analysis as differentially expressed gene ontology (DEGO).

Data availability

Raw fastq files of intra-single cell RNA-seq and intra-single cell whole genome bisulfite sequencing were deposited through DDBJ (DNA Data Bank of Japan). Accession number will be provided until the final publication.
Accession number will be provided until the final publication.

Ethical considerations

Animal experiments were reviewed and approved by the University of Tokyo Institutional Animal Care and Use Committee (Permit Number: Med-P17-037).

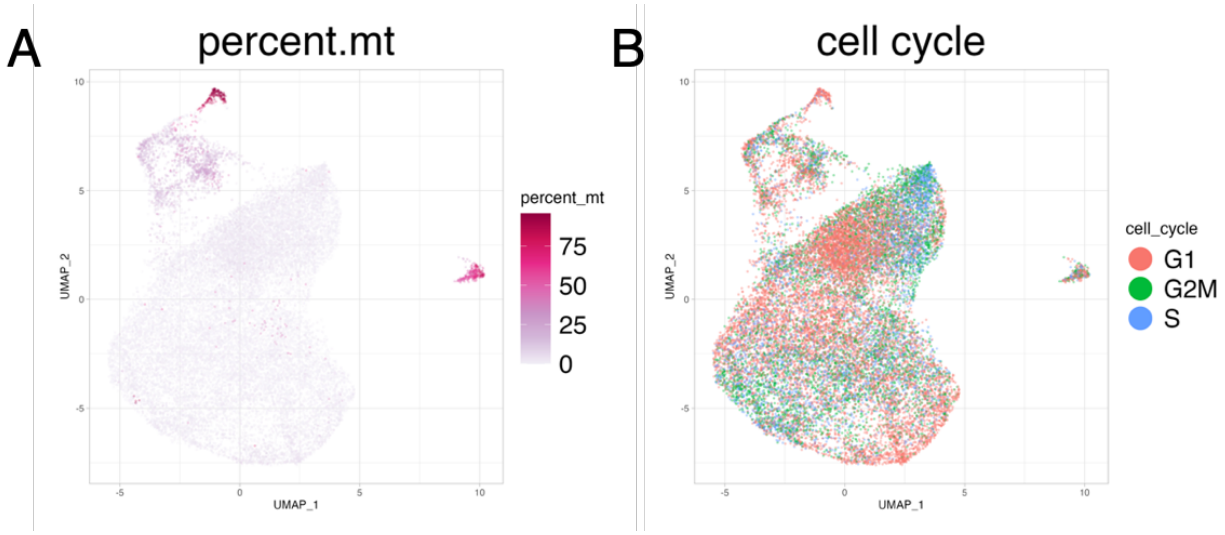

Figure. S1. Integrative UMAP consisting of four datasets from conventional scRNA-seq and intra-single cell RNA-seq.

DimPlot of mitochondrial percentage (**A**) and cell cycles (**B**) on integrative UMAP from iscRNA-seq and scRNA-seq datasets from Tsukasaki's team (post RANKL stimulation day 0, day 1, day 3) and one from Omata's team (post RANKL day 4).

References and Notes

1. S. Gravina, X. Dong, B. Yu, J. Vijg, Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol* **17**, 150 (2016).
2. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
3. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
4. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
5. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2022).
6. J. Sun, T. Nishiyama, K. Shimizu, K. Kadota, TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics* **14**, 219 (2013).
7. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
8. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
9. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York (2016).
10. T. J. Hardcastle, K. A. Kelly, baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**, 422 (2010).
11. T. Wu *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
12. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).
13. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
14. K. D. Hansen, B. Langmead, R. A. Irizarry, BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**, R83 (2012).
15. K. Korthauer, S. Chakraborty, Y. Benjamini, R. A. Irizarry, Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* **20**, 367-383 (2019).
16. B. I. Laufer *et al.*, Low-pass whole genome bisulfite sequencing of neonatal dried blood spots identifies a role for RUNX1 in Down syndrome DNA methylation profiles. *Hum Mol Genet* **29**, 3465-3476 (2021).
17. T. Yin, D. Cook, M. Lawrence, ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol* **13**, R77 (2012).
18. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).
19. S. Choudhary, R. Satija, Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol* **23**, 27 (2022).
20. M. Tsukasaki *et al.*, Stepwise cell fate decision pathways during osteoclastogenesis at single-cell resolution. *Nat Metab* **2**, 1382-1390 (2020).

21. Y. Omata *et al.*, Interspecies Single-Cell RNA-Seq Analysis Reveals the Novel Trajectory of Osteoclast Differentiation and Therapeutic Targets. *JBMR Plus*, (2022).
22. Johannes Rainer, EnsDb.Mmusculus.v79: Ensembl based annotation package. (2017).
23. L. Zappia, A. Oshlack, Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).
24. J. H. Leland McInnes, James Melville,, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*, (2020).
25. C. Trapnell *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).
26. X. Qiu *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).
27. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
28. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
29. S. Aibar *et al.*, SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).
30. T. Osabe, K. Shimizu, K. Kadota, Accurate Classification of Differential Expression Patterns in a Bayesian Framework With Robust Normalization for Multi-Group RNA-Seq Count Data. *Bioinform Biol Insights* **13**, 1177932219860817 (2019).