Title:

Subcellular sampling and sequencing by iSCseq connects transcriptomics with morphological, physical, and physiological findings

Authors:
Hiroyuki Okada[1,2,3]*, Yuta Terui[4], Yasunori Omata[3,5], Masahide Seki[6], Shoichiro Tani[1,3], Asuka Terashima[5], Sanshiro Kanazawa[7], Masahiro Hosonuma[8,9], Junya Miyahara[3], Kenta Makabe[3], Shoko Onodera[10], Fumiko Yano[11], Hiroshi Kajiya[12], Francesca Gori[2], Taku Saito[3], Koji Okabe[12], Yutaka Suzuki[6], Roland Baron[2], Ung-il Chung[1,13], Sakae Tanaka[3]*, Hironori Hojo[1,13]*

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 Oral Medicine, Infection, and Immunity, Harvard School of Dental Medicine, Boston, MA 02115, USA
[3] Department of Orthopaedic Surgery, the University of Tokyo
[4] Division of Single Cell Solution, Product Strategy Department, Marketing Center, Life Business HQ, Yokogawa Electric Corporation
[5] Bone and Cartilage Regenerative Medicine, the University of Tokyo Hospital
[6] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, the University of Tokyo
[7] Department of Oral and Maxillofacial Surgery, Graduate School of Medicine, the University of Tokyo
[8] Department of Clinical Immuno Oncology, Clinical Research Institute for Clinical Pharmacology and Therapeutics, Showa University
[9] Showa University Pharmacological Research Center
[10] Department of Biochemistry, Tokyo Dental College
[11] Department of Biochemistry, Showa University School of Dentistry
[12] Department of Physiological Science and Molecular Biology, Fukuoka Dental College
[13] Department of Bioengineering, Graduate School of Engineering, the University of Tokyo
*Corresponding author. Email: hokada-tky@g.ecc.u-tokyo.ac.jp; tanakas-ort@h.u-tokyo.ac.jp; hojo@g.ecc.u-tokyo.ac.jp

Abstract:
Multinucleated giant cells, such as osteoclasts (OCs), often contain >10 nuclei. This raises the question whether nuclei residing in the same cytoplasm are similarly regulated or heterogeneous, expressing a different set of genes. We developed intra-single cell sequencing (iSCseq), a new approach combining imaging with confocal microscopy, picking up of any cellular and subcellular components inside both OCs and ordinary small cells using the Single Cellome™ System SS2000 (Yokogawa, Japan), and next-generation sequencing (intra single-cell RNA-seq; iSCseq). Our results revealed diversity between individual nuclei and subcellular components, within the same cell. Subcellular distribution of cellular organs and calcium based on live-cell imaging linked the reality to gene expression with conventional single cell RNA-seq.

One-Sentence Summary:
Intra-single cell sequencing (iSCseq) based on live cell imaging and cellular or subcellular sampling overcomes the anonymity of gene expressions and connects transcriptome with cell physical and physiological findings.

Introduction

Multinucleated giant cells, such as osteoclasts (OCs), often contain >10 nuclei. OCs are haematopoietic 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 specific cellular organization in the presence of multinucleation raises the question of whether nuclei residing in the same cytoplasm are similarly regulated or heterogeneous, expressing a different set of genes and/or different levels of physiological activity. Initial differentiation from the monocyte lineage has been well-examined. Nuclear factor of activated T-cells (NFATc1) is a master regulator of OC differentiation [3].
Single cell (Sc) technology has decoded cellular diversity in terms of gene expressions [4]. Sc technology has also been used to clarify the trajectory of cellular differentiation. Conventional scRNA-seq has already been applied to in vitro RANKL-induced culture [5] and has described the inter-species differences in the OC differentiation pathway between humans and mice [6]. Although gene expression and a few plasmalemmal markers are easily recognized, the original location information of cells disappears. Spatial transcriptomics (ST) provided gene expression with spatial information [7,8]. However, dynamic cellular movement cannot be detected with ST. In addition, conventional Sc techniques do not theoretically allow the separation of OCs into Scs because of their large size and fragility; however, OCs were sequenced using conventional scRNA-seq [5,6]. Furthermore, essential signaling cascades inside cells, such as mitochondrial metabolism and fine-tuning of calcium concentration inside OC [9-11], have not been detected but are predicted. Live-seq challenged the recording of transcriptomes from living cells; however, it did not detect subcellular differences in gene expressions [12].
This study aimed to establish a new technology for examining subcellular transcriptomic heterogeneity in living OCs and small cells.

Results

The basics of iSCseq (intra-single cell sequencing)

Intra single cell sequencing (iSCseq) is a new approach combining confocal live cell imaging, intracellular sampling with a 3 mm or 10 mm probe, and next generation sequencing from ultra-pico input of RNA using commercially available kits (Fig. 1a). iSCseq enabled us to capture the transcriptomes of cellular or subcellular components, including an individual nucleus.
Mature OCs are large cells with a diameter of over 100 mm *in vitro*. Single Cellome™ System SS2000 (Yokogawa, Japan) enabled us to sample not only a single cell, but also subcellular components inside living OCs under visual observation. Using this approach, we sampled single nuclei from the peripheral zone of OC (Fig. 1b, Movie 1) or one nucleus from the aggregation of 4 nuclei for example (Fig. 1c, Movie 2). Fine tuning of the probe tip position and negative pressure also enabled us to select subcellular components from ordinary small single-nucleated cells (Fig. 1d). iSCseq sampled a maximum of 11 batches from one OC and 15 batches from one foreign body giant cell (FBGC, Fig. S1i). FBGCs induced by IL-4 are derived from the same monocyte lineage as OC; however, FBGCs do not resorb bone [13].
Next, we connected fine samples of cellular or subcellular components confirmed by multi-colored confocal images with transcriptomic library construction using the SMART-seq single

cell PLUS kit (Takara Bio, Japan). Sampled cellular or subcellular components were harvested in a microtube containing RNA lysis buffer with RNase inhibitors, centrifuged, and frozen promptly to minimize RNA deterioration. The number of genes detected with a 3 mm probe was equivalent to genes with conventional fluid-based chromium (10x genomics, U.S.), and genes with a 10 mm probe were more than twice those with the chromium (Fig. 1e).

iSCseq connects gene expression with morphology

Gene expression from 280 out of 285 samples (98.2 %) was divided into six clusters on UMAP [14] (Fig. 2a) with python-based deep learning scvi-tools [15-17], considering the percentages of mitochondrial reads (%mt), of ribosomal reads, and cell cycle phase defined by Seurat team [18] (Fig. S2a-c). The RNA spliced/unspliced ratio was evaluated using scVelo [19] (Fig. 2b), and gene ontology (GO) analysis using decoupleR [20] (Fig. 2c), and an automatic cell type annotation using CellAssign [21] (Fig. 2d) illustrated the differentiation pathway to OCs. We called terminal cluster 3 iSCseq transcriptomic OCs (itOCs) from the OC-related gene expression and GO terms (Fig. S2a,b). Clustering results were projected onto the captured images, which provided a morphological annotation (Fig. 2e).

Next, we overviewed the number of nuclei in the original OC using the binned Nuclear Rank. The itOC cluster consisted of less subcellular components from large OCs with 20 and more nuclei, which were called as morphological OCs, than the average proportions (Fig. 2f,g). The geographical positions were dichotomized into peripheral or central nuclei. itOCs contain nuclei from both the peripheral and central zones (Fig. 2h,i).

These results demonstrated the heterogeneity of the gene expression among the nuclei present inside the same mature OC. Large OCs contained fewer itOC ingredients of the terminal phase in differentiation.

Integrative analysis of iSCseq (i2SCseq) with conventional scRNA-seq

To validate our iSCseq approach, we performed an integrative analysis with two previously published scRNA-seq studies of *in vitro* OC differentiation, which were generated using a traditional microfluid-based technique [5,6]. We called this combinatorial approach integrative analysis of iSCseq (i2SCseq). A total of 13,130 cellular or subcellular components, which contained 116 iSCseq results with less than 25 %mt (Fig. S2d) and with more than 400 detected genes, were illustrated on UMAP (Fig. 3a) using the latest scvi-tools [17,22]. To reduce the batch effect of each sample, the number of days of exposure to RANKL and the examiner's identity were considered (Fig. S1g,h). We defined cluster 14 as i2tOCs (integrative iSCseq transcriptomic OCs). The differentiation pathway to i2tOCs was directed downwards from the center left (Fig. 3b-d). The i2tOCs were confirmed by gene expression analysis and GO terms (Fig. S3a,b). iSCseq spots shadowed on integrative UMAP depicted the stage of differentiation of each component (Fig. 3e). Nuclear rank analysis showed that i2tOCs did not contain a higher proportion of fragments from the mOCs (Fig. 3f,g). Interestingly, i2tOCs included ordinary small single-nucleated cells (Fig. 3h). Nuclei positional analysis showed that the i2tOCs contained only central nuclei (Fig. 3i,j). In a representative cell, the peripheral nucleus was assigned to an undifferentiated cluster, whereas the central nucleus of the same OC was assigned to i2tOCs (Fig. 3k).

i2SCseq increased the quantitative resolution of subcellular components in massive datasets. We were able to increase our understanding of the stages of cell differentiation at a higher resolution for the first time using i2SCseq compared with iSCseq alone.

Interventional response evaluated by iSCseq

iSCseq precisely was controlled the tip of the probes in the xyz position. In addition, we were able to manipulate the duration and the negative pressure sample by sample. Although the sampling procedure was tuned and equalized, multiple nuclei were sometimes collected thorough a probe (Fig. 4a-c). We hypothesized that the sampling reaction could reflect the tightness of the cytoskeleton among the nuclei and be evaluated in terms of gene expression.

Analyses of the binned number of nuclei in each OC sample showed that 89.7% of itOCs and all i2tOCs contained one or more nuclei (Fig. 4d-g). The i2tOCs did not contain fractions containing more than five nuclei (ffn) (Fig. 4g). GO terms related to the cytoskeleton were upregulated in the middle bin of the nuclei; in contrast, they were almost downregulated in ffn (Fig. 4h). The ffn component showed higher expression of acids, such as Cathepsin K and Mmp9 (Fig. S4a,c), and upregulated GO terms related to bone resorption (Fig. S4b,d). A representative OC sample (Fig. 4i) showed the peripheral ring of tubulin (purple) and nuclei (blue). The central nuclei are independent of the tubulin ring.

The reaction in interventional sampling can be explained by the gene expression in each sample. The ffn of OC showed less cytoskeleton-related gene expression compared with one and fewer than 5 nuclei. Physical reactions may partially be explained by gene expression.

Inspective morpho-physiological analysis with gene expression

It is difficult to evaluate mitochondrial genes using conventional scRNA-seq. Seurat used %mt as a normalization factor [18,23,24], which means that the mitochondrial gene expression was assumed to be roughly equivalent. Single nucleus analysis showed the contamination of mitochondrial genes.

iSCseq can color transcriptomes with morphological physiological findings based on confocal microscopy. iSCseq can visualize the existence and subcellular distribution of mitochondrial gene, as confirmed by the gene expression profile.

Fig. 5a and b were the representative small and multinucleated cells whose mitochondria was stained purple. Fig. 5b clearly depicted the difference in mitochondrial ingredients between pairs of before and after images. Furthermore, iSCseq enabled us to evaluate mitochondrial components based on not only the imaging but also the gene expression. Samples in fig. 5a and b are assigned to different cluster in iSCseq alone and in i2SCseq. The itOCs and i2tOCs had more mitochondrial subcellular components in terms of their gene expression profiles (Fig. 5c-f) and GO terms (Fig. 5g left). Mitochondrial dichotomization was confirmed using GO. In fractions containing one or more nuclei, mitochondrial terms were upregulated (Fig. 5g right). The cellular calcium concentration was evaluated using the same scheme as the mitochondrial evaluation (Fig. S5).

It is difficult to determine the distribution of cellular organs or low molecules inside a cell. iSCseq can strengthen the unequal distribution of intracellular organs, even in small cells, based on images and gene expression.

Discussion

Traditional single cell technology using microfluid devices has clarified cellular heterogeneity in terms of gene expression [25]. However, there is a lack of cellular identity as a trade-off for achieving high throughput. Moreover, scRNA-seq focused on the differences between cells. The

subcellular components were beyond the scope. Our iSCseq approach is the first realistic proof-of-concept approach for detecting subcellular heterogeneity within a single cell with morphological annotations.

iSCseq provides morphological and positional annotations via sequencing. The origin of cells or subcellular positions inside a cell can be depicted using this combinatorial approach. Integrative analyses of iSCseq linked conventional scRNA-seq to the cellular or subcellular reality. A higher number of cells or subcellular components provides a higher resolution of the sequenced data. For example, we could not clearly divide FBGCs from mature OCs only with the iSCseq (Fig. S1j). However, integrative analysis with scRNA-seq of all the components from FBGCs was not positioned in the differentiated OC clusters (Fig. S1k). Since scRNA-seq and iSCseq have complementary effects, we recommend using iSCseq in combination and increasing resolution.

In addition, iSCseq has already become a powerful tool for investigating subcellular physiological phenomena. We called the enhanced version of iSCseq integrative, interventional, and inspective subcellular sequencing (i3sCseq) (**Fig. 6**). i3sCseq clarified the different stages of subcellular components at higher resolution. Physical behavior during intervention gave us a hint of cytoskeletal links between the nuclei. The subcellular distribution of mitochondria and calcium was depicted based on two inspections of gene expression and live cell imaging. The transcriptome and cell physiology are difficult to connect. i3sCseq is a promising tool to investigate cellular physiology based on comprehensive gene expression.

Multinucleated large cells were beyond the scope of the trendy single-cell approaches. Our findings showed that cellular components which were regarded as most mature OCs from the viewpoint of gene expression included a small, ordinary, single-nucleated cell. Large multinucleated cells contain various subcellular components at different stages of differentiation. Physical responses during sampling and the physiological distribution of cellular organs or molecules are easy to detect using conventional confocal microscopy and cellular staining. We believe that an OC is a model cell to investigate the unequal distribution and properties of cellular organs inside a cell.

iSCseq proved a concept to combine imaging living cells, sequencing at subcellular resolution, and confirming physical and physiological findings. A greater number of subcellular components would provide greater resolution for investigating subcellular phenomena. Cost reduction of sampling and plate-based sequencing is a key to random unbiased sampling and sequencing.

In conclusion, a newly developed method iSCseq demonstrated the subcellular heterogeneity of gene expression inside a single cell. Multinucleated cells can contain multiple stages of cell differentiation. Physical responses under intervention and physiological distributions of mitochondria and calcium were correlated to the gene expression. i3SCseq clarified cell physical and physiological diversity embedded in a single cell.

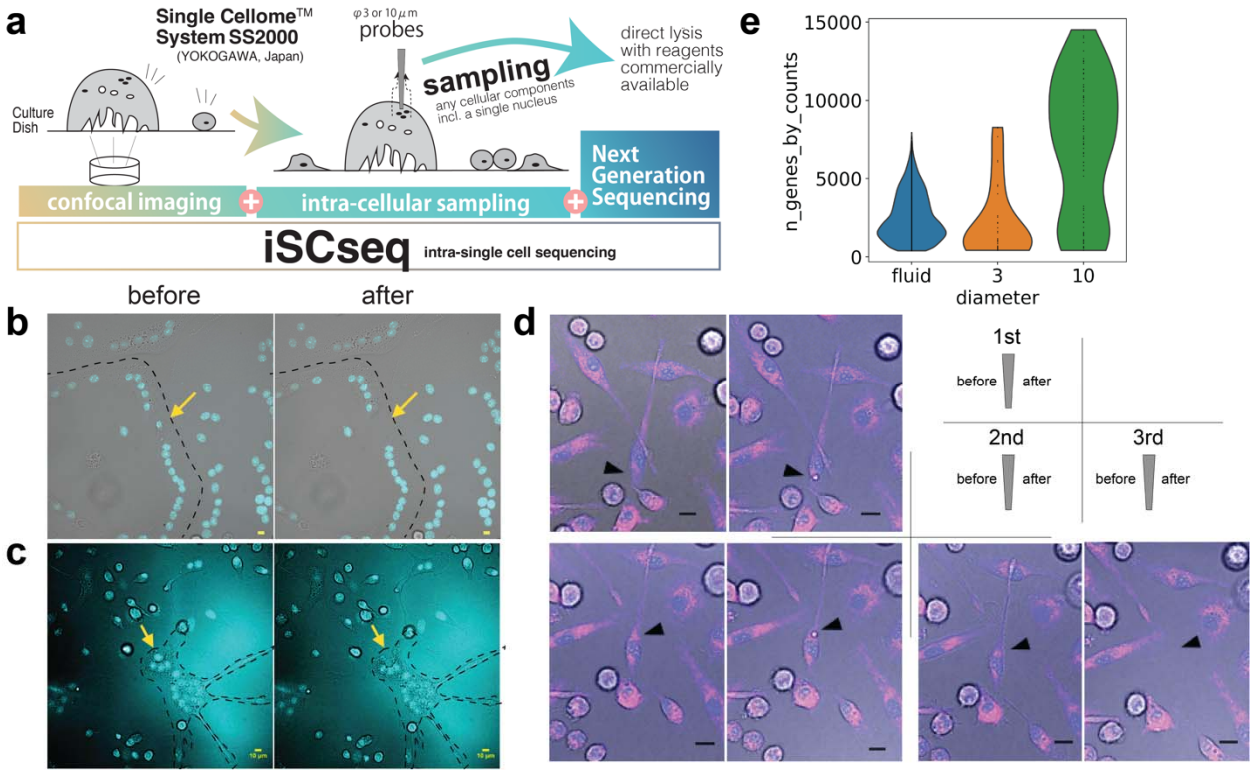


Fig. 1| Outline of intra-single cell sequencing (iSCseq) and snapshot of sampling cellular components from mature osteoclast. a, Schematic presentation of sampling experiments. Single Cellome™ System SS2000 (YOKOGAWA) took multi-colored images with confocal microscopy and sampled any nucleus or cellular component by a sharp probe with a diameter of three to ten μm, whose z position was finely adjusted. **b-d**, Representative cases of sampling each single nucleus or subcellular components from a single cell. Blue, Nuclei stained with Hoechst 33342. Bar = 10um. **b, c**, Enclosed area by black dotted line is one osteoclast. **b**, A nucleus from the peripheral zone of large osteoclast (OC). **c**, A nucleus from the aggregation of four nuclei in a OC. **d**, 1st, and 2nd sampling for each subcellular component. 3rd sampling for a whole cell with a nucleus. Purple, Tubulin stained with SiR-tublin kit. **e**, Violin plot of detected genes by sampling and sequencing method. Left to right, fluid-based chromium (10x genomics), SMART-seq for samples picked by 3 mm probe and by 10 mm probe.

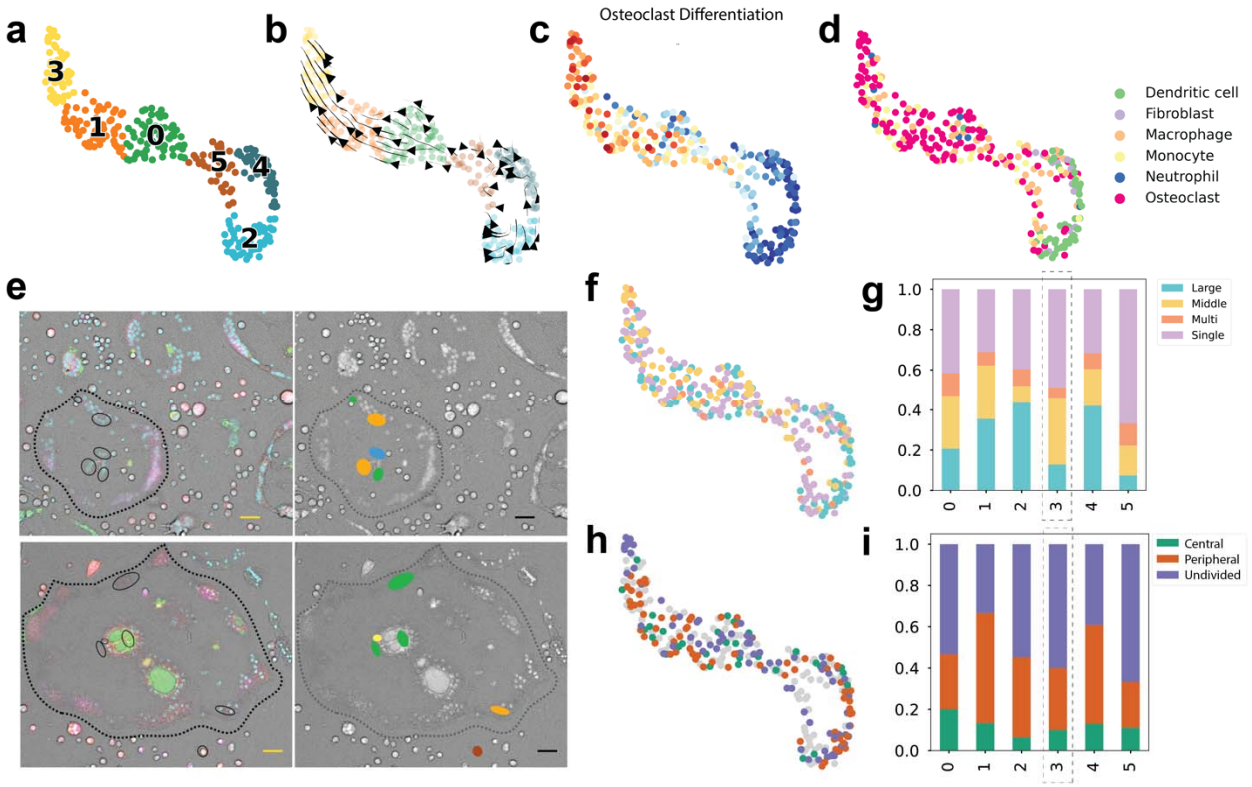


Fig. 2| Result of intra-single cell sequencing (iSCseq) for cellular or subcellular components. **a**, UMAP consisted of 280 cellular or subcellular components. **b**, Streamline presentation based on spliced/unspliced ratio of RNA by the scVelo. **c**, Gene ontology 'Osteoclast Differentiation' analyzed by decoupleR. **d**, An automatic cell type annotation by CellAssign. **e**, Geometrical information of targeted mature osteoclasts (OCs, left). 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, Tubulin stained with SiR-tublin kit. Projection of iSCseq cluster presented on fig. **e** (right). Bar = 50 mm. **f, g**, Nuclear Rank: bin, the number of nuclei in the origin of cell. Large, 20 or more nuclei; Middle, 5 to 19 nuclei; Multi, 2 to 4 nuclei; Single, 1, respectively. **f**, Feature plot of Nuclear Rank. **g**, Bar plot of proportion of Nuclear Rank in each cluster. **h, i**, Positional information, Central or Peripheral or Undivided in OC. **h**, Feature plot of positional information. **g**, Bar plot of proportion of positional information in each cluster. **g,i**, The clusters surrounded by dotted square were the final clusters in OC differentiation.

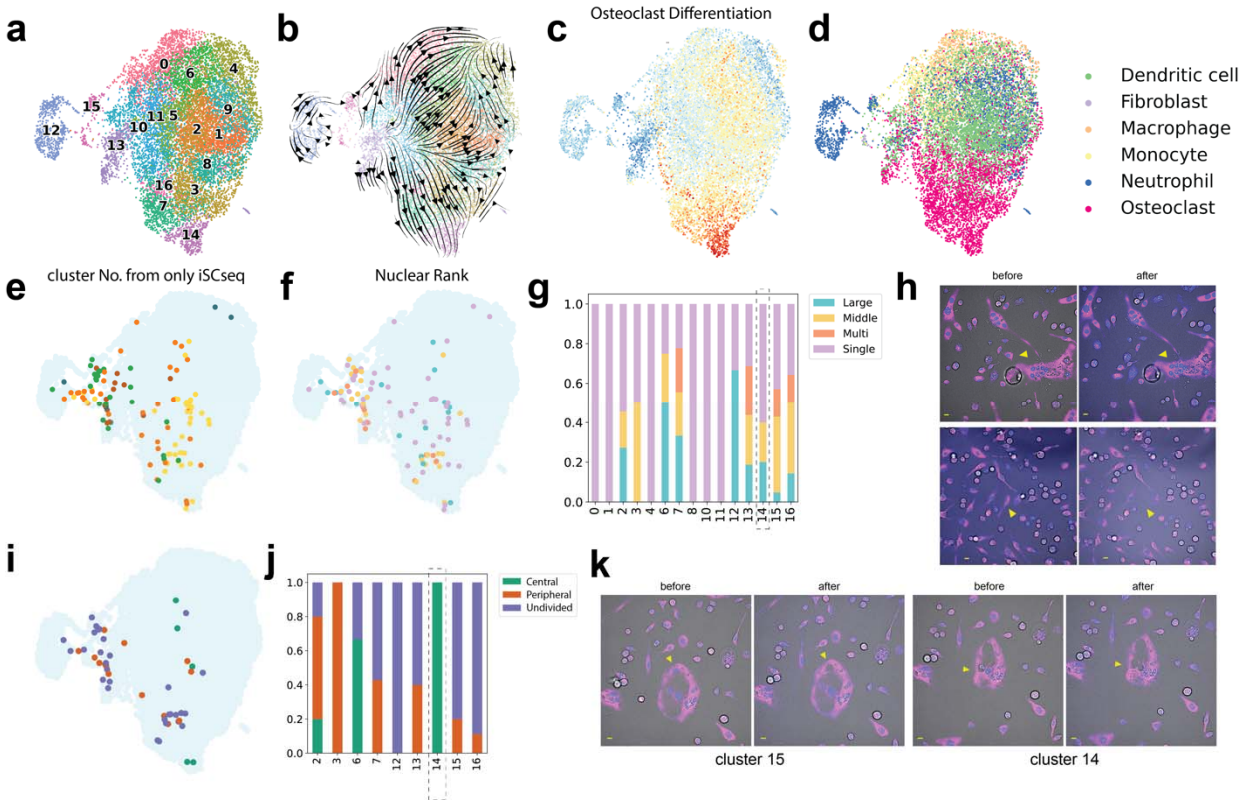


Fig. 3| Integrative analysis of intra-single cell RNA-seq (i2SCseq) and previous scRNA-seq of *in vitro* osteoclast differentiation. a, Integrative UMAP of iSCseq and scRNA-seq datasets from Dr. Tsukasaki's team (after RANKL stimulation day 0, day 1, day 3) and one from Dr. Omata's team (after RANKL day 4). **b**, Streamline presentation based on spliced/unspliced ratio of RNA by scVelo. **c**, Gene ontology 'Osteoclast Differentiation' analyzed by decoupleR. **d**, An automatic cell type annotation by the CellAssign. **e**, Representation of iSCseq spot on i2SCseq UMAP. Colors corresponding to the clusters by iSCseq shown on **Fig. 2a**. **f, g**, Nuclear Rank: bin, the number of nuclei in the origin of cell. Large, 20 or more nuclei; Middle, 5 to 19 nuclei; Multi, 2 to 4 nuclei; Single, 1, respectively. **f**, Feature plot of Nuclear Rank. **g**, Bar plot of proportion of Nuclear Rank in each cluster. **h**, Representative cases of a single-nucleated cell assigned to the cluster on the differential pathway of OC. **i, j**, Positional information, Central or Peripheral or Undivided in OC. **i**, Feature plot of positional information. **j**, Bar plot of proportion of positional information in each cluster. **g,j**, The clusters surrounded by dotted square were the final clusters in OC differentiation. **k**, Representative sampling cases of a nucleus from a multinucleated OC. Blue, Nuclei stained with Hoechst 33342; Purple, Tubulin stained with SiR-tublin kit. Bar = 10 mm.

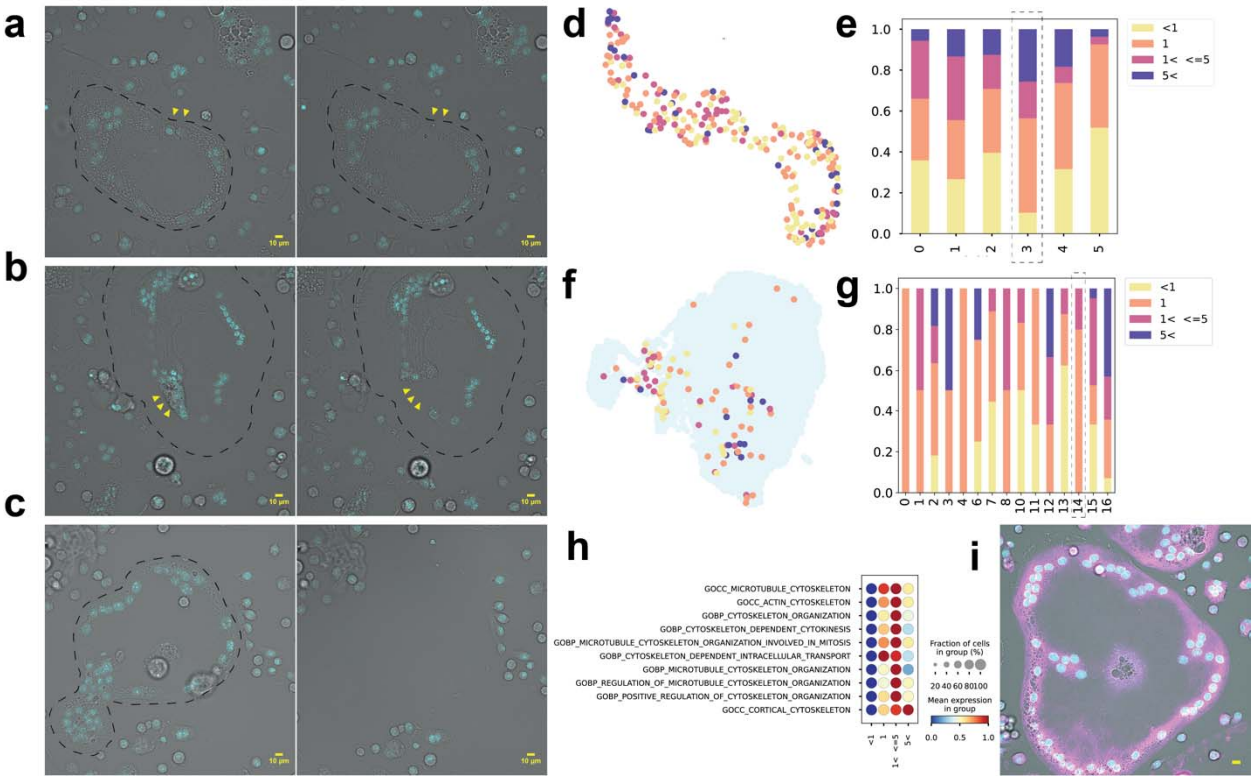


Fig. 4| Interventional response analysis by iSCseq. a-c, Representative nuclei sampling cases from multinucleated cells. Blue, Nuclei stained with Hoechst 33342. Enclosed area by black dotted line is one osteoclast (OC). Bar = 10 mm. **a, b**, yellow arrowheads indicate sampled nuclei. **a**, Two nuclei from the peripheral zone. **b**, Twelve nuclei from the aggregation of nuclei. **c**, Sampling a whole large osteoclast. Feature plots of osteoclast specific genes on the UMAP. **d-g**, The number of nuclei in each sample in the iSCseq (**d, e**) and in the integrative analysis of the iSCseq (i2SCseq) with conventional scRNA-seq (**f, g**). **d, f**, Feature plot of the nucleus number. **e, g**, Bar plot of the nucleus number by iSCseq cluster on **fig. 2a** (**e**) and by i2SCseq cluster on **fig. 3a** (**g**). The clusters surrounded by dotted square were the final clusters in OC differentiation. **h**, Dot plot of the top 10 gene ontology analysis using decoupleR of terms including "CYTOSKELETON" by the binned number of nuclei in cellular or subcellular components of OC. **i**, A representative image of a large OC. Blue, Nuclei stained with Hoechst 33342; Purple, Tubulin stained with SiR-tublin kit. Bar = 10 mm.

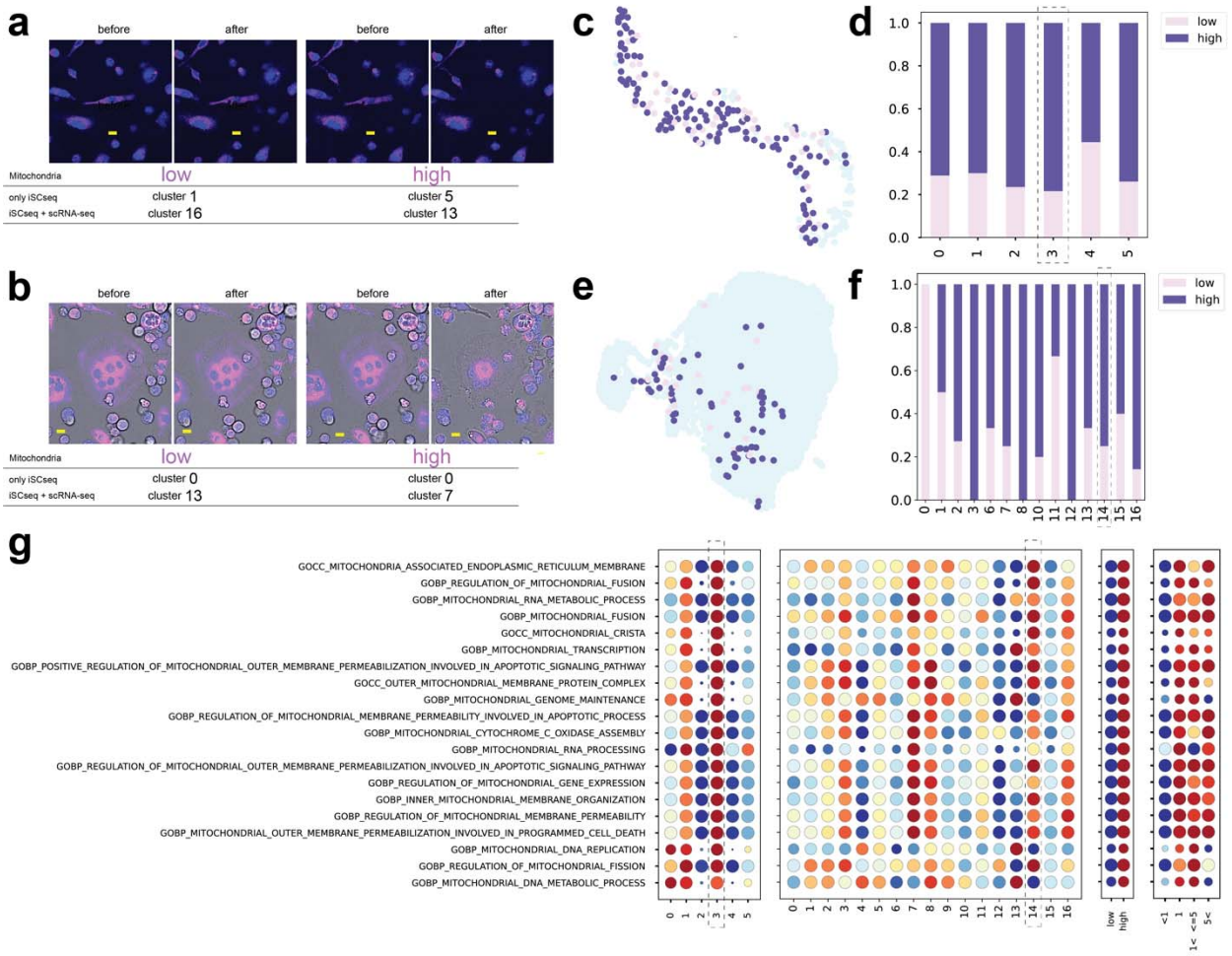


Fig. 5| Inspective analysis for the geographical distribution of mitochondria *in vitro* osteoclast differentiation combined with gene expression. a, b, Representative cases of single nucleated cell (**a**) and multinucleated OC (**b**). Blue, Nuclei stained with Hoechst 33342; Purple, MitoBright LT Deep Red. Bar = 10um. Corresponding iSCseq cluster on **fig. 2a** and integrative analysis of iSCseq (i2SCseq) cluster on **fig. 3a** summarized in each table. **c-f**, High or low information of mitochondrial staining in each sample of iSCseq. **c, d**, Analysis of only iSCseq. **e, f**, Analysis of i2SCseq. **c, e**, Feature plot of dichotomized information of mitochondria staining. **d, f**, Bar plot of high or low mitochondria by iSCseq cluster on **fig. 2a** (**d**) and by i2SCseq cluster on **fig. 3a** (**f**). Clusters surrounded by dotted square are the final clusters in OC differentiation. **g**, Dot plot of the top 20 genes ontology analysis using decoupleR of terms including "MITOCHONDRIA." Left to right, by the iSCseq cluster, by the i2SCseq cluster, by the high or low information based on stained images, and by binned number of nuclei in cellular or subcellular components in OCs.

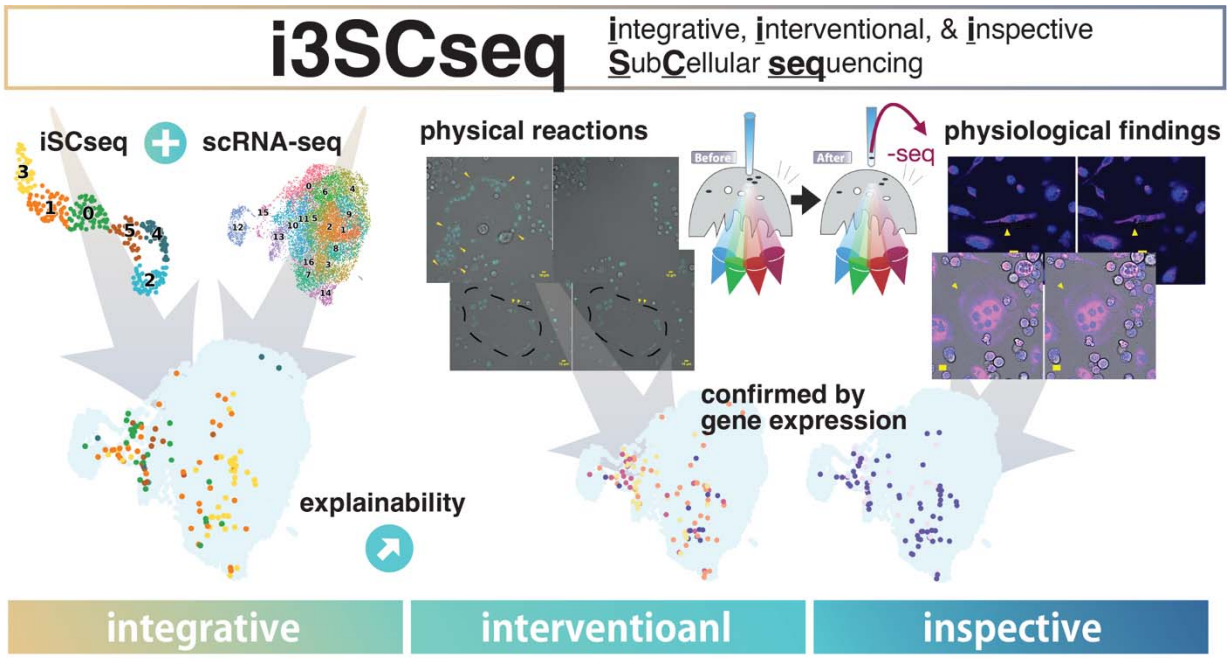
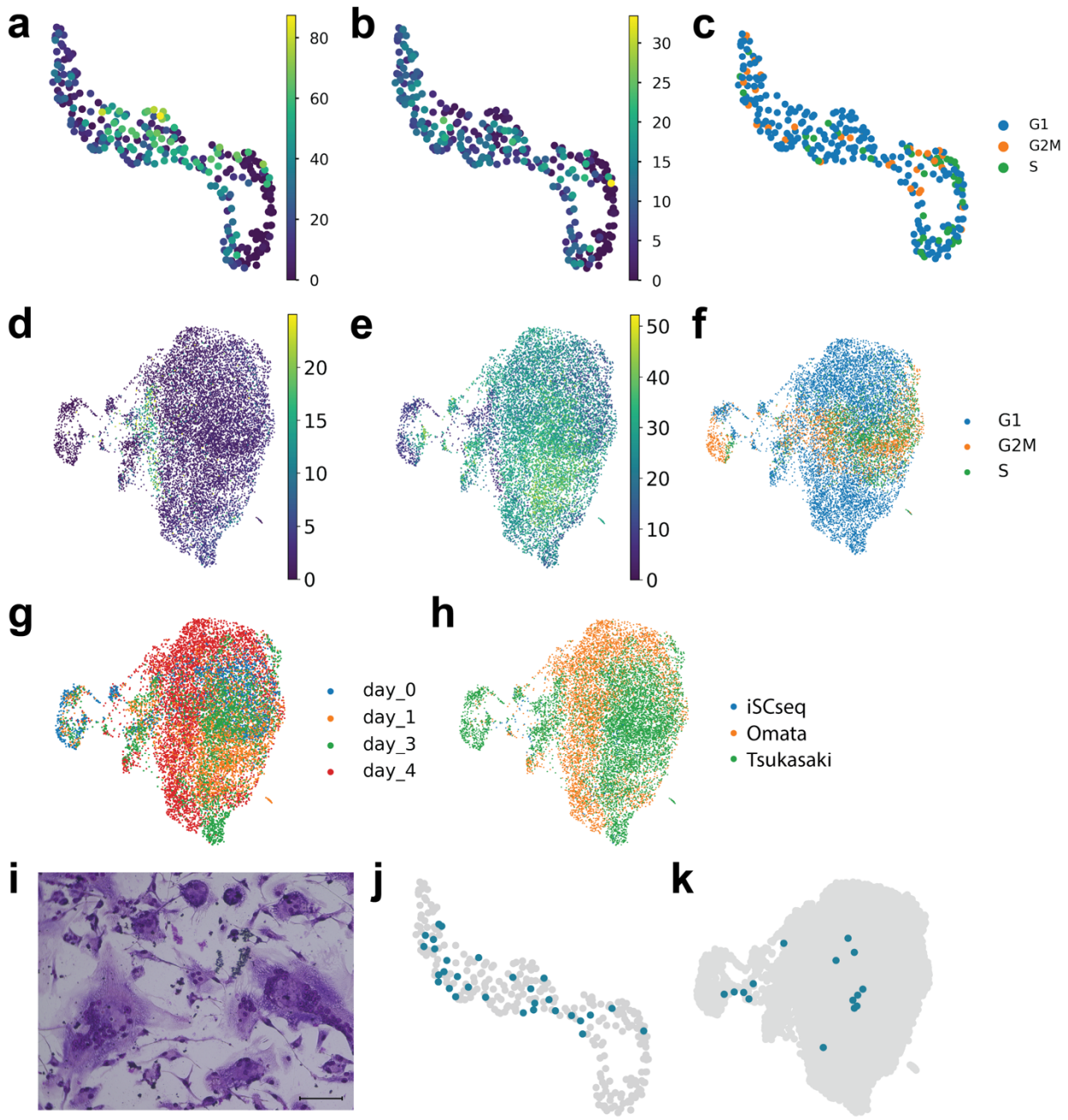
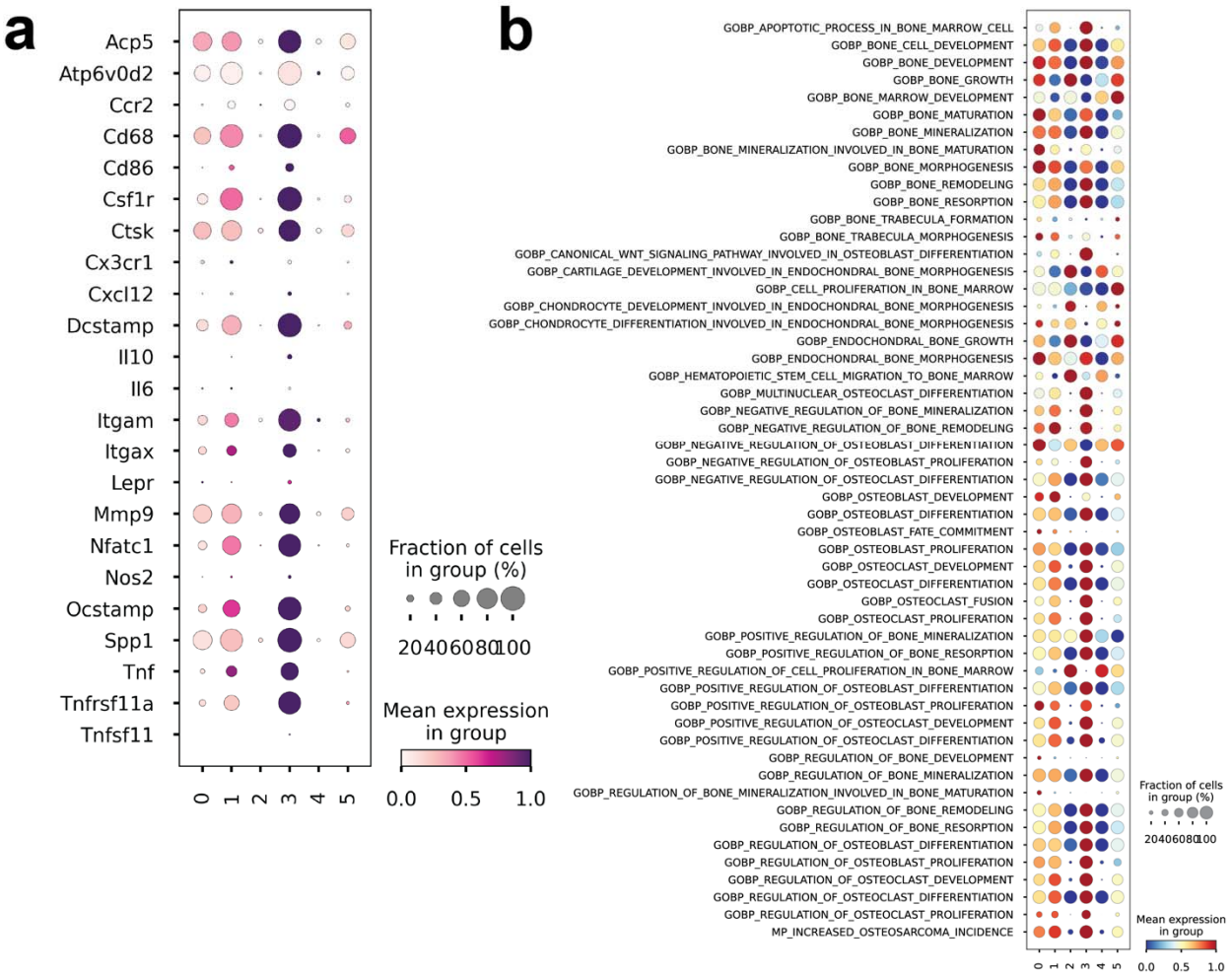


Fig. 6| Schematic presentation of integrative, interventional, and inspective subcellular sequencing (i3SCseq) empowered by intra-single cell sequencing (iSCseq).

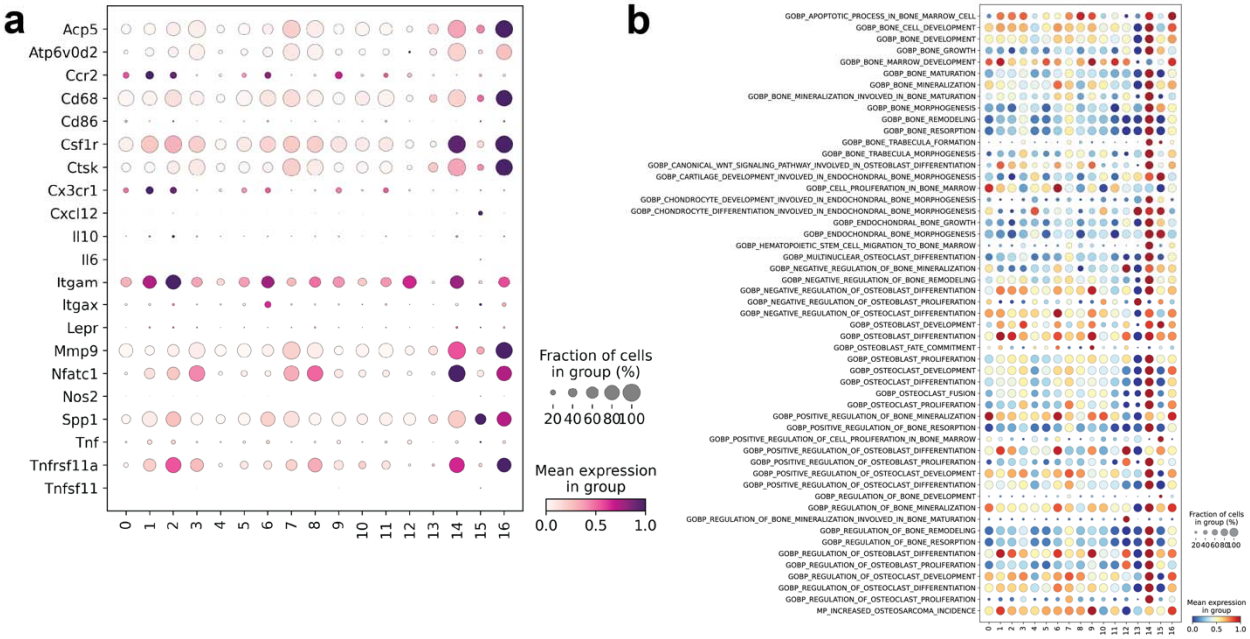


Extended Data Fig. S1| Basic parameters of RNA-seq used in the integration of datasets. DimPlots of mitochondrial percentage (**a, d)**, of mitochondrial percentage (**b, e)**, and of cell cycles (**c, f**) from only iSCseq(**a-c**) or from integrative analysis of iSCseq with conventional scRNA-seq (i2SCseq) (**d-f**). Days of osteoclast (OC) differentiation under RANKL stimulation (**g**) and author's identity (**h**) in each dataset were considered in i2SCseq (**g**). **i**, representative foreign body giant cells (FBGCs) stained with May-Gru□nwald Giemsa. Bar = 100 mm. (**j, k**) Spots of FBGCs, whose origin is the same as OCs but function is different from OC, on UMAP of only iSCseq (**j**) and of i2SCseq (**k**).



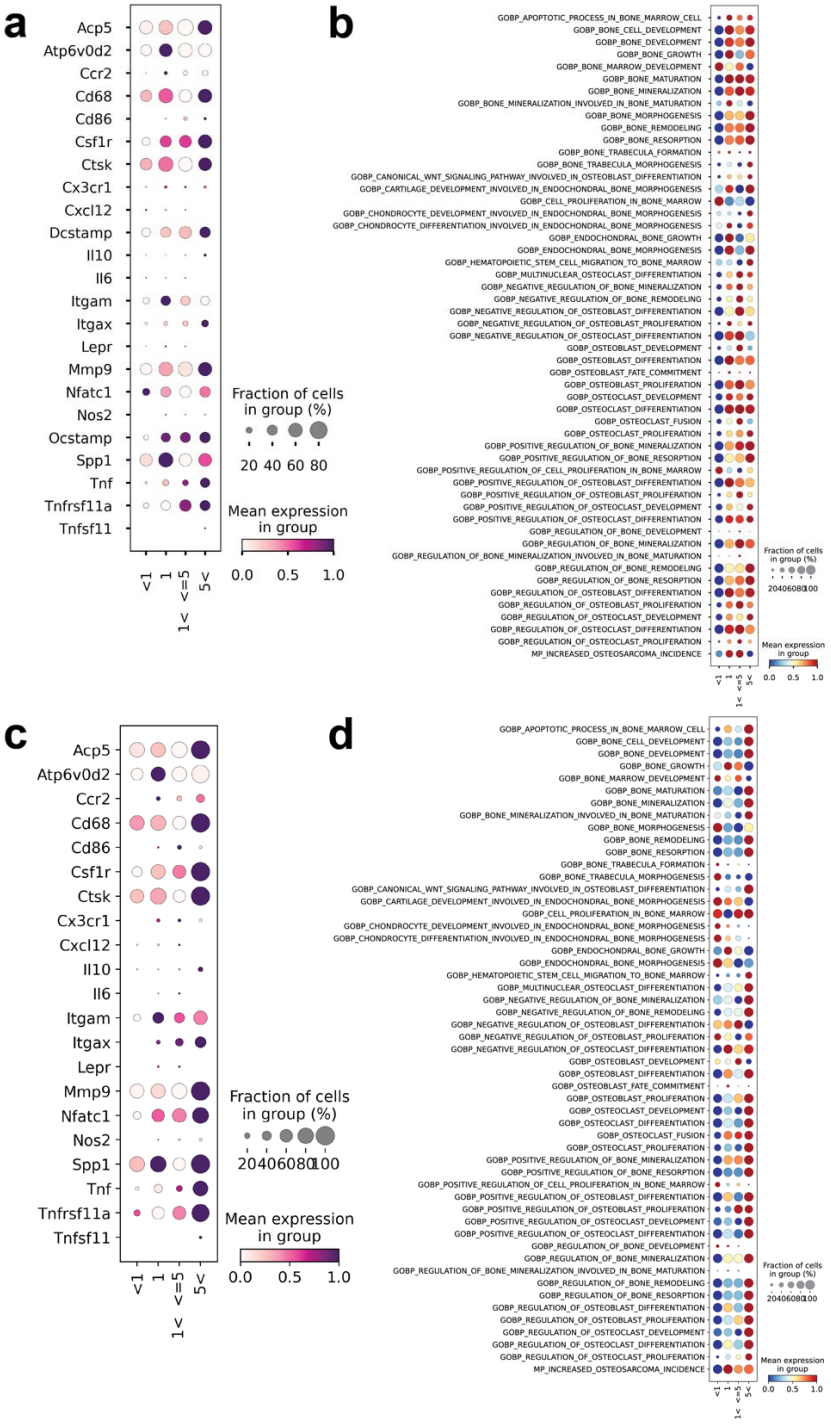
Extended Data Fig. S2| Gene expression and gene ontology analysis from intra-single cell RNA-seq (iSCseq).

a, Dotplot of expressions of genes related to osteoclast. **b**, Gene ontology analysis on the terms containing “BONE” or “OSTEO” using decoupleR by each iSCseq cluster.



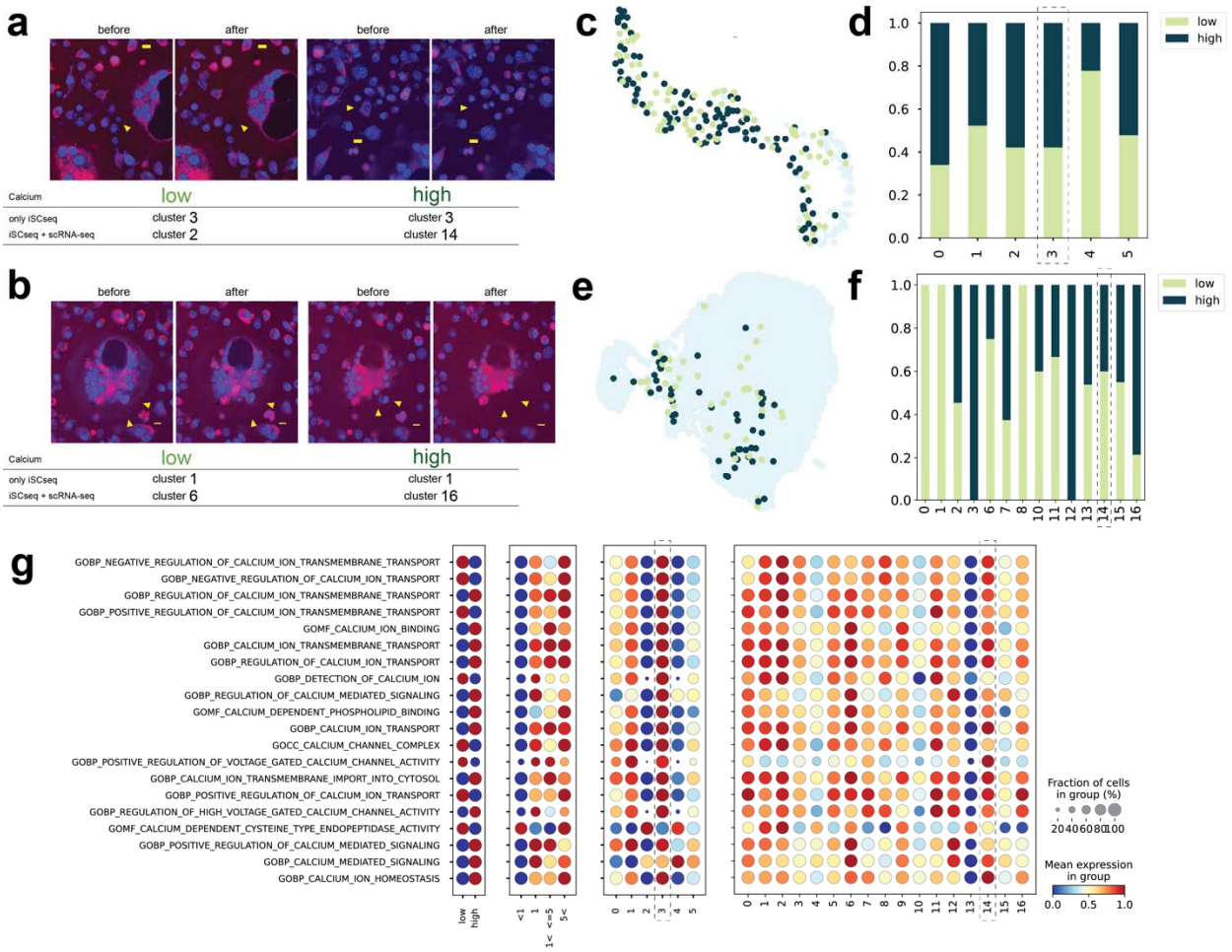
Extended Data Fig. S3| Gene expression and gene ontology analysis from integrative analysis of iSCseq (i2SCseq).

a, Dotplot of the expressions of genes related to osteoclast. **b**, Gene ontology analysis on the terms containing “BONE” or “OSTEO” using decoupleR by each i2SCseq cluster.



Extended Data Fig. S4| Gene expression and gene ontology analysis by binned number of nuclei in each subcellular component.

a, c, Dotplot of expressions of genes related to osteoclast. **b, d**, Gene ontology analysis on the terms containing “BONE” or “OSTEO” using decoupleR. **a, b**, from intra-single cell RNA-seq (iSCseq). **c, d**, from integrative analysis of iSCseq (i2SCseq).



Extended Data Fig. S5| Inspective analysis for the geographical distribution of Calcium *in vitro* osteoclast differentiation combined with gene expression. a, b, Representative cases of single nucleated cell (**a**) and multinucleated OC (**b**). Blue, Nuclei stained with Hoechst 33342; Red, Calbryte^{TM} 590 ABD. Bar = 10 mm. Corresponding iSCseq cluster on **fig. 2a** and integrative analysis of iSCseq (i2SCseq) cluster on **fig. 3a** summarized in each table. **c-f**, High or low information of mitochondrial stain in each sample of iSCseq. **c, d**, Analysis of only iSCseq. **e, f**, Analysis of i2SCseq. **c, e**, Feature plot of dichotomized information of mitochondria stain. **d, f**, Bar plot of high or low mitochondria by iSCseq cluster on **fig. 2a** (**d**) and by i2SCseq cluster on **fig. 3a** (**f**). Clusters surrounded by dotted square are the final clusters in OC differentiation. **g**, Dot plot of the top 20 genes ontology analysis using decoupleR of terms including "CALCIUM." Left to right, by iSCseq cluster, by i2SCseq cluster, by the high or low information based on stained images, and by binned number of nuclei in cellular or subcellular components of OC.

Movie S1.
Sampling a single nucleus in the peripheral zone of a large osteoclast (Fig. 1b).

Movie S2.
Sampling a single nucleus from four nuclei aggregation inside a middle-sized osteoclast (Fig. 1c).

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 Clontech. Supercomputer SHIROKANE at Human Genome Center at the University of Tokyo was partially used in the calculation. Platform for Advanced Genome Science in Japan supported this project.

Funding:

JSPS KAKENHI grant numbers JP 20K22973, 21K16703. 22KK0272, 23K15736 (HO), JP 21K19589 (HH and HO), JP 21K19883 (UC, HH and HO), JP 22K10212 (SK, HH and HO), JP 16H06279 (HO, PAGS: Platform for Advanced Genome Science) and 22H04925 (HO, PAGS ver.2)

AMED grant number JP21zf0127002 (HH and HO) and JP21bm0704071 (HH, MS and HO) "Creation of fully circular bio-adaptive materials driven by bio-polymer big data" Material DX (HH and HO)

KAWAI Foundation for Sound Technology and Music 2021 (HO)

Takeda Science Foundation 2022 (HO)

Mochida Memorial Foundation for Medical and Pharmaceutical Research 2022 (HO)

FY2020 Invitational Fellowships for Research in Japan L20539 (RB)

Author contributions:

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

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

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

Visualization: HO, YT, S. Tani, AT, JM

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

Project administration: HO, HH

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

Writing – original draft: HO

Writing – review & editing: HO, AT, 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.)

Methods

Mice

Wild-type C57BL/6 mice were purchased from the Oriental Kobo Industrial Company (Tokyo, Japan). The 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 the lower limbs of 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, Minneapolis, MN, U.S.) and 1% penicillin-streptomycin (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.) for three days to produce BMMs. To induce Osteoclasts, the BMMs were cultured on 35 mm Glass Bottom Dish (Collagen Coat) φ14 mm (Matsunami Glass, Osaka, Japan) in the presence of 50 ng/ml M-CSF (PeproTech Ltd., New Jersey, U.S.) and 50 ng/ml soluble RANKL (Fujifilm Wako Pure Chemical Corporation). To induce foreign body giant cells, BMMs were cultured in 50 ng/ml GM-CSF (R&D Systems, Minneapolis, MN, U.S.) and 50 ng/ml IL-4 (R&D Systems). More than three nucleated with larger sizes compared to other types of cells were counted as morphological osteoclasts and FBGCs according to the induction protocol.

Live cell staining

One hour before sampling, all nuclei were stained with 0.2 mg/mL Hoechst 33342, Trihydrochloride, Trihydrate (Invitrogen, Waltham, Massachusetts, U.S.) in PBS and incubated at 37°C for 10 min.

Cells in Fig. 2 were stained with pHrodoTM AM Ester (Thermo Fisher) 2μL and PowerLoadTM concentrate (Thermo Fisher) 20 mL in 2 mL MEMalfa per 35 mm dish incubated at room temperature for 30 min. The pH probe was then thoroughly washed away with PBS. Finally, the cells were loaded with the calcium indicator Fluo-4 (Fluo-4 NW Calcium Assay Kit, Invitrogen) and 500 nM SiR-tubulin Kit (Cytoskeleton, Denver, Colorado, U.S.) according to the manufacturer's instructions.

Cells in figures other than Fig. 2 were stained for nuclei by Hoechst 33342, then exposed to a multi-staining solution, which contains Calbryte™ 590AM (AAT Bioquest, California, U.S.) 1μL, MitoBright LT Deep Red (Dojindo, Kumamoto, Japan) 1μL in Fluo-4 AM Assay buffer (HEPES buffer) 1 mL, incubated at 37°C for 45 min, and washed with MEMalfa containing 10% fetal bovine serum. All cells sampled were incubated with the same reagents and under environmental conditions for differentiation.

Sampling of intracellular components

Cellular components, including a single nucleus, were sampled using the Single Cellome™ System SS2000 (Yokogawa Electric Corporation, Japan). Before sampling, we obtained fine multicolored images using confocal microscopy to confine the sampling target, including a single nucleus. When sampling cellular components, SS2000 finely tuned the position of probes in the Z-axis and sucked any cellular components into a 10 μm glass probe with the set negative pressure. The exact control of the tip with SS2000 is a key collection of subcellular components from living cells. After sampling cellular components, probes maintain a small positive pressure

so as 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 11.5 mL of 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.). Immediately after sampling into lysis buffer, microtubes containing samples were centrifuged briefly at 2,000 rpm and frozen at -80°C.

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

SMART-Seq Single Cell PLUS kit (Takara Bio) was used for cDNA library construction according to the manufacturer's instructions. 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. cDNA was synthesized according to the manufacturer's instructions. In the cDNA amplification step, 20 PCR cycles were used, which is the maximum number of cycles for the minimum input of RNA. Amplified cDNA was purified using an AMPure XP PCR purification kit (Beckman Coulter, Brea, California, U.S.). cDNA quality was checked using a Bioanalyzer High Sensitivity DNA kit (Agilent, Santa Clara, CA, U.S.) before library preparation. Samples with more than 10 times the cDNA concentration compared with the negative control by smear analysis within the 200-9,000 bp range were included.

SMART-Seq Library Prep Kit, including the SMART-Seq Single-Cell PLUS kit, was used for library preparation. The library amplification step and PCR cycles were determined according to the concentration detected by the Bioanalyzer. After library amplification, the samples were pooled and purified once with AMPure. The pooled amplified library was qualified using a Bioanalyzer High-Sensitivity DNA kit.

Next generation sequencing of RNA-seq libraries

Pooled libraries were sequenced at paired-end 150 bp using HiSeq X Ten (Illumina, San Diego, CA, U.S.) by Macrogen Japan (Tokyo, Japan) or NovaSeq 6000 by LiSDaC (Life Science Data Research Center at the University of Tokyo) and the Kashiwa sequencing team of the Platform for Advanced Genome Science.

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

Raw fastq files were mapped onto the mouse genome mm10 using STAR ver. 2.7.10b [26]. Gene expression levels of iSCseq were summarized in a count matrix using RSEM [27]. Transcripts Per Kilobase Million (TPM) were calculated using 'rsem-generate-data-matrix-modified' in script_rnakato [28].

Bioinformatics analyses were performed using Python 3.10.11 [29]. Scanpy [15] in scverse project [17] was used to process and visualize the count matrix. Scvi-tools was used to integrate datasets with a model considering the experimental batch, total count, probe diameter, and cell cycle score based on cell cycle genes defined by the Seurat team [18,23,24,30,31], percentages of leads on mitochondrial genes, and ribosomal genes. The model was trained using a graphical processing unit on Google Collaboratory. Uniform Manifold Approximation and Projection (UMAP) was performed for dimension reduction [14]. Clustering was performed using the Leiden algorithm [32] at a resolution of 1. Cell type annotation was performed with CellAssign [21] referring to CellMarker 2.0 [33]. Gene ontology scores were calculated using decoupleR [20] and AUcell [34,35]. The coloring was arranged a palettable [36].

Integrative analysis of conventional single cell RNA-seq and iSCseq

iSCseq dataset was integrated with four datasets of scRNA-seq,GSM4419695 (post RANKL stimulation day 0), GSM4419699 (day 1), GSM4419703 (day 3) by Dr. Tsukasaki [5], and published undeposited data (day 4) by Dr. Omata [6]. Before merging the datasets, entrez IDs were transformed into gene symbols using EnsDb.Mmusculus.v79 [37]. In dataset integration, the author's identity was additionally considered when using scvi-tools, which is highly recommended in benchmark study [22]. Downstream analyses were performed using only the iSCseq data.

Analysis of images

The obtained images were processed and reconstructed using the Fiji software. Enhanced contrast was used to reduce the background noise. High or low signal intensity was determined manually by comparison with other surrounding cells or other components inside the same cell.

Data availability

Raw fastq files of iSCseq were deposited through DDBJ (DNA Data Bank of Japan). Accession numbers 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).

References and Notes

1 Yasuda, H. *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). https://doi.org:10.1073/pnas.95.7.3597

2 Tanaka, S. *et al.* c-Cbl is downstream of c-Src in a signalling pathway necessary for bone resorption. *Nature* **383**, 528-531 (1996). https://doi.org:10.1038/383528a0

3 Takayanagi, H. *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). https://doi.org:10.1016/s1534-5807(02)00369-6

4 Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* **15**, e8746 (2019). https://doi.org:10.15252/msb.20188746

5 Tsukasaki, M. *et al.* Stepwise cell fate decision pathways during osteoclastogenesis at single-cell resolution. *Nat Metab* **2**, 1382-1390 (2020). https://doi.org:10.1038/s42255-020-00318-y

6 Omata, Y. *et al.* Interspecies Single□Cell RNA□Seq Analysis Reveals the Novel Trajectory of Osteoclast Differentiation and Therapeutic Targets. *JBMR Plus* (2022). https://doi.org:10.1002/jbm4.10631

7 Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334-1347 (2021). https://doi.org:10.1038/s41588-021-00911-1

8 Rao, A., Barkley, D., Franca, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211-220 (2021). https://doi.org:10.1038/s41586-021-03634-9

9 Okada, H. *et al.* CTLA4-Ig Directly Inhibits Osteoclastogenesis by Interfering With Intracellular Calcium Oscillations in Bone Marrow Macrophages. *J Bone Miner Res* **34**, 1744-1752 (2019). https://doi.org:10.1002/jbmr.3754

10 Okada, H., Okabe, K. & Tanaka, S. Finely-Tuned Calcium Oscillations in Osteoclast Differentiation and Bone Resorption. *Int J Mol Sci* **22** (2020). https://doi.org:10.3390/ijms22010180

11 Okada, H. & Tanaka, S. Plasmalemmal interface for calcium signaling in osteoclast differentiation. *Curr Opin Cell Biol* **74**, 55-61 (2022). https://doi.org:10.1016/j.ceb.2022.01.001

12 Chen, W. *et al.* Live-seq enables temporal transcriptomic recording of single cells. *Nature* (2022). https://doi.org:10.1038/s41586-022-05046-9

13 ten Harkel, B. *et al.* The Foreign Body Giant Cell Cannot Resorb Bone, But Dissolves Hydroxyapatite Like Osteoclasts. *PLoS One* **10**, e0139564 (2015). https://doi.org:10.1371/journal.pone.0139564

14 Leland McInnes, J. H., James Melville,. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* (2020). https://doi.org:10.48550/arXiv.1802.03426

15 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018). https://doi.org:10.1186/s13059-017-1382-0

16 Gayoso, A. *et al.* A Python library for probabilistic analysis of single-cell omics data. *Nat Biotechnol* **40**, 163-166 (2022). https://doi.org:10.1038/s41587-021-01206-w

17 Virshup, I. *et al.* The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat Biotechnol* (2023). https://doi.org:10.1038/s41587-023-01733-8

18 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *bioRxiv*, 2020.2010.2012.335331 (2020). https://doi.org:10.1101/2020.10.12.335331
19 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020). https://doi.org:10.1038/s41587-020-0591-3
20 Badia, I. M. P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv* **2**, vbac016 (2022). https://doi.org:10.1093/bioadv/vbac016
21 Zhang, A. W. *et al.* Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat Methods* **16**, 1007-1015 (2019). https://doi.org:10.1038/s41592-019-0529-1
22 Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* **19**, 41-50 (2022). https://doi.org:10.1038/s41592-021-01336-8
23 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048
24 Hao, Y. *et al.* Dictionary learning for integrative, multimodal, and scalable single-cell analysis. *bioRxiv* (2022). https://doi.org:10.1101/2022.02.24.481684
25 Zhang, X. *et al.* Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell* **73**, 130-142 e135 (2019). https://doi.org:10.1016/j.molcel.2018.10.020
26 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635
27 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). https://doi.org:10.1186/1471-2105-12-323
28 Nakato, R. *Scripts for NGS analysis*, <https://github.com/rnakato/script_rnakato> (2023).
29 Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual*. (CreateSpace, 2009).
30 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019). https://doi.org:10.1186/s13059-019-1874-1
31 Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol* **23**, 27 (2022). https://doi.org:10.1186/s13059-021-02584-9
32 Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019). https://doi.org:10.1038/s41598-019-41695-z
33 Hu, C. *et al.* CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* **51**, D870-D876 (2023). https://doi.org:10.1093/nar/gkac947
34 AUCell: Analysis of 'gene set' activity in single-cell RNA-seq data. (2016).
35 Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247-2276 (2020). https://doi.org:10.1038/s41596-020-0336-2
36 Davis, M. *palettable. v3.3.3*, <https://jiffyclub.github.io/palettable/> (2023).
37 Johannes Rainer. EnsDb.Mmusculus.v79: Ensembl based annotation package. (2017).