Title:

Inclusive living subcellular sequencing rendering physical physiological and human pathological features in osteoimmune diversity

Authors:
Hiroyuki Okada[1,2,3]*, Yuta Terui[4], Yasunori Omata[3,5], Asuka Terashima[5], Masahide Seki[6], Shoichiro Tani[1,3], 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 (HO); tanakas-ort@h.u-tokyo.ac.jp (S.Tanaka); hojo@g.ecc.u-tokyo.ac.jp (HH)

Abstract

Recent single-cell transcriptomes have clarified cellular heterogeneity within cell populations. Spatial transcriptomics using fixed tissues connects transcriptomics with positional information on histological specimens. However, these technologies cannot capture the dynamic localization of mRNA inside living cells.

To decode intracellular gene expression in living cells, we developed intra-single cell sequencing (iSCseq), a new approach that combines confocal imaging, picking up any cellular and subcellular components inside both osteoclasts (OCs) and ordinary small cells using the Single Cellome™ System SS2000 (Yokogawa), and next-generation sequencing (intra single-cell RNA-seq; iSCseq). This study aimed to establish a new technology, iSCseq, to examine subcellular transcriptomic heterogeneity in living OCs and small cells.

iSCseq demonstrated subcellular heterogeneity of gene expression in a single cell. iSCseq revealed not only multiple differentiation stages embedded in the same OC but also physical cytoskeletal connections between nuclei and physiological activity of mitochondria and calcium, confirmed by transcriptomic evidence. Inclusive iSCseq with in vivo bone milieu atlas identified new subsets of OCs and clarified the switching role of immunoreceptor tyrosine-based activation motifs in the context of osteoimmune diversity. Inclusive iSCseq with murinized human bone tumors specified another tumor-specific OC subset. The iSCseq approach has the potential to enhance cellular physiology and identify therapeutic targets.

One-Sentence Summary

Inclusive subcellular sequencing based on live-cell imaging and sampling of cellular components has clarified the cellular physical, physiological, and human pathological mechanisms confirmed by confocal images and transcriptomics.

Introduction

Recent single-cell transcriptomic studies have clarified the cellular diversity within the cell population[1]. Spatial transcriptomics using fixed tissues connects transcriptomics with positional information on histological specimen[2,3]. However, these technologies cannot decode the dynamic localization of mRNA inside living cells. Live-seq challenged the recording of transcriptomes from living cells; however, it did not detect subcellular differences in gene expressions[4]. Moreover, it was difficult to elucidate the physiological and pathological implications of sequencing at subcellular resolution.

During the bone remodeling process, hematopoietic and mesenchymal cells are coordinated to maintain bone mineral density (BMD)[5]. Fails in cooperative relationships lead to skeletal-related events such as geriatric, inflammatory, and pathological fractures, which are global challenges in the aged society[6,7]. Although bone resorption inhibitors increase BMD and reduce the fracture risk, their use is often prohibited in patients with renal failure. There remains global controversy regarding the withdrawal of anti-bone resorptive reagents, sometimes causing rapid bone loss[8]. Further investigation of the mechanism of bone resorption is necessary to address the unmet clinical needs.

Osteoclasts (OCs), which have characteristic large portion and often contain >10 nuclei, are haematopoietic stem cell-derived bone-resorbing cells induced by macrophage colony stimulating factor (M-CSF; *CSF1*) and receptor activator of nuclear factor kappa B ligand (RANKL; *TNFSF11*)[9]. Nuclear factor of activated T-cells (NFATc1) is a master regulator of OC differentiation[10]. OC is highly polarized, with a specific conformation of its cytoskeleton in an actin ring that seals off an extracellular bone-resorbing compartment[11]. 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. Essential signaling cascades inside cells, such as mitochondrial metabolism and fine-tuning of calcium concentration inside OC[12-14], have not been detected, but are predicted. In addition, the difference between *in vitro* cultured OC and *in vivo* OC remains unclear because of the difficulty in sampling from the bone. Moreover, the origin of OCs in the hematopoietic lineage has not been quantitatively examined.

To clarify the hidden mechanism inside OCs, scRNA-seq was performed for *in vitro*-cultured OCs[15,16]. However, conventional single-cell techniques theoretically do not allow for the separation of OCs into single-cell suspensions because of their large size and fragility. Therefore, a new platform must be developed to reveal the heterogeneity of OCs.

This study aimed to establish a new technology called intra single cell sequencing (iSCseq) to examine subcellular transcriptomic heterogeneity in living OCs and small cells. Matching the data points obtained by iSCseq with conventional scRNA-seq datasets clarified the physiological and pathological mechanisms of bone resorption in humans.

Results

The basics of iSCseq (intra-single cell sequencing)

Intra single cell sequencing (iSCseq) is a new approach that combines confocal live cell imaging, intracellular sampling with a 3-µm or 10-µm probe, and next-generation sequencing from ultra-pico input of RNA using commercially available kits. iSCseq enabled us to capture the transcriptomes of cellular and subcellular components, including individual nuclei.

Mature OCs are large cells with diameters of > 100 µm *in vitro*. Single Cellome™ System SS2000 (Yokogawa, Japan) enabled the sampling of 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 the OC (Fig. 1a, Movie 1) or one nucleus from the aggregation of four nuclei (Fig. 1b, Movie 2). Fine-tuning of the probe tip position and negative pressure also enabled multiple samplings from ordinary small single-nucleated cells (Fig. 1c). iSCseq sampled a maximum of 11 batches from one OC and 15 batches from one foreign body giant cell (FBGC, Fig. S1d). FBGCs induced by IL-4 are derived from the same monocyte lineage as OC; however, FBGCs do not resorb bone[17].
Next, we connected fine samples of cellular or subcellular components confirmed by multicolored confocal images with transcriptomic library construction using SMART-seq single cell PLUS kit (Takara Bio, Japan). The 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-µm probe was equivalent to that of genes with conventional fluid-based chromium (10x genomics, U.S.), and genes with a 10−µm probe were more than twice those detected with chromium (Fig. 1d).

iSCseq connects gene expression with morphology

Gene expression from 280 of 285 samples (98.2 %) was divided into six clusters on UMAP[18] (Fig. 1e) using Python-based deep learning scvi-tools[19-21], considering the percentages of mitochondrial reads (%mt), ribosomal reads, and cell cycle phase defined by the Seurat team[22] (Fig. S2a-c). The RNA spliced/unspliced ratio was evaluated using scVelo[23] (Fig. 1f), and gene ontology (GO) analysis using decoupleR[24] (Fig. 1g) and automatic cell type annotation using CellAssign[25] (Fig. 1f) illustrated that cellular components in clusters 0 and 1 were involved in the OC differentiation pathway. Terminal cluster 0 was called iSCseq transcriptomic OCs (itOCs) based on the cell-type annotation (Fig. 1h, S1f) and OC-related gene expression and GO terms (Fig. S1g,h). The clustering results were projected onto the captured images, which provided morphological annotations (Fig. 1i).
Next, we reviewed the number of nuclei in the original OC using binned Nuclear Rank. The itOC cluster consisted of fewer subcellular components from large OCs with 20 or more nuclei, which were morphologically called OCs, than the average proportions (Fig. 1j,k). The geographical positions were dichotomized into peripheral or central nuclei. The itOCs contained nuclei from both peripheral and central zones (Fig. 1l,m).
These results demonstrated the heterogeneity of gene expression among nuclei present in the same mature OC. Some single-nucleated small cells, which were believed not OCs in classical morphological definition, had already differentiated into itOCs from the view of transcriptome.

Inclusive iSCseq with *in vitro* scRNA-seq

To validate our iSCseq approach, we performed an integrative analysis with two previously published scRNA-seq studies on *in vitro* OC differentiation, which were generated using a traditional microfluid-based technique[15,16]. A total of 13,130 cellular or subcellular components, including 116 iSCseq results, with less than 25 %mt (Fig. S2a) and with more than 400 genes were illustrated on UMAP (Fig. 2a). To reduce the batch effect of each sample, the number of days of RANKL exposure was also considered (Fig. S2d). Cluster 6 was defined as itOC in integrative analysis. The differentiation pathway to itOCs was directed downwards from the

center left (Fig. 2b-d, S2g-i). The iSCseq spots rendered on integrative UMAP depicted the differentiation stage of each cellular component (Fig. 2e). The itOCs were selected more strictly by integrative analysis than by analysis using only the iSCseq dataset (Fig. 2f). Nuclear rank analysis showed that integrative itOCs did not contain a higher proportion of fragments from large OCs, which were believed to be mature OCs (Fig. 2g,h). Interestingly, integrative itOCs included ordinary, small, and single-nucleated cells (Fig. 2i). Nuclei positional analysis showed that itOCs contained fewer marginal nuclei than average (Fig. 2j,k). In a representative cell, the peripheral nucleus was assigned to an undifferentiated cluster, whereas the central nucleus of the same OC was assigned to integrative itOC (Fig. 2i).
Integrative itOCs could depict the traits of subcellular components at a higher resolution compared with iSCseq alone (Fig. S2f). Even with more strict definition, some small single-nucleated cells were categorized into terminal OCs.

Interventional response evaluated by iSCseq

iSCseq precisely controls the tip of the probe at the xy and z positions. In addition, we were able to manipulate the duration and negative pressure sample by sample. Although the sampling procedure was tuned and equalized, multiple nuclei were occasionally collected thorough a probe (Fig. 3a-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 binned number of nuclei in each OC sample showed that 88% of integrative itOC contained one or more nuclei (Fig. 3d) and that itOC consisted of abundant fractions containing more than five nuclei (ffns) (Fig. 3e). GO terms related to the cytoskeleton were upregulated in integrative itOCs; all were upregulated in ffns (Fig. 3f). The ffns showed a higher expression of acids, such as *Ctsk* and *Mmp9*, however, lower expression of *Atp6v0d2* (Fig. 3g). GO terms related to OC and bone resorption were upregulated in ffns (Fig. 3h).
Physical reactions during interventional sampling linked to GO terms related to cytoskeleton. itOCs contained abundant ffns, which showed higher production of major acids and OC traits. Although cells with large number of nuclei were not always categorized to itOCs when evaluating nuclei number by cell (Fig. 2h), ffns, whose nuclei number were evaluated by component, exhibited OC traits. Thus, iSCseq could provide different evaluation criteria based on live cell imaging.

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 [22,26,27], which means that mitochondrial gene expression was assumed to be roughly equivalent. Single-nucleus RNA-seq regards mitochondrial gene expression as a contamination, which should be discarded. These technologies do not provide the chance to analyze mitochondrial genes.
iSCseq can add transcriptome evidence to morphological and physiological findings based on confocal microscopy. We examined the relationship between mitochondrial gene expression in each fraction and dichotomized mitochondrial density based on fluorescent images obtained by iSCseq experiments.
Fig. 3i and j were representative small and multinucleated cells whose mitochondria were stained purple, with the information of the integrative *in vitro* iSCseq cluster. We can provide information on the mitochondrial fluorescence density in transcriptomic UMAP (Fig. 3k). The integrative itOC cluster contained mitochondria-rich components (Fig. 3l). Mitochondrial

dichotomization based on the images was confirmed using GO terms (Fig. 3m, left panel). In integrative itOC and fractions containing ffns, mitochondrial terms were upregulated (Fig. 3m, center, right). The cellular calcium concentration was evaluated using the same scheme as that used for mitochondrial evaluation (Fig. S3).
Determining the distribution of cellular organs or molecules within a cell is challenging. Live cell imaging using antibodies are strongly affected by experimental factors including antibody affinity. iSCseq approach proved the concept to confirm the fluorescence intensity of images by gene expression. Although the presented data was qualitative, iSCseq suggested the unequal distribution of intracellular organs even in small cells.

Inclusive iSCseq with *in vivo* bone milieu atlas

iSCseq based on live-cell imaging has linked reality to gene expression using anonymous conventional single-cell RNA-seq. Taking advantage of iSCseq, we tried inclusive iSCseq with *in vivo* and *in vitro* scRNA-seq datasets to test whether *in vivo* OCs in the bone microenvironment are different from *in vitro* cultured OC induced by M-CSF and RANKL. First, we built an *in vivo* bone milieu atlas containing >54k cells from 28 public datasets of nine papers[28-36] (Fig. S4a). From the cell type annotation using CellAssign, this atlas recapitulated the bone microenvironment consisting of whole hematopoietic lineage cells, including hematopoietic stem cells and mesenchymal cells (Fig. S4b,n). This atlas can provide insight into differentiation pathways based on RNA velocity[23,37] (Fig. S4c,m) and regulatory cell populations (Fig. S4j,k). Next, we combined *in vitro* scRNA-seq datasets, including iSCseq, and the *in vivo* bone milieu atlas (Fig. S5,6). Integrative iSCseq provided the distribution of 19 cell types from *in vivo* bone environment (Fig. 4a). We shadowed the iSCseq spot on the inclusive UMAP (Fig. 4b). The *in vitro* OC differentiation pathway had a bifurcation (Fig. 4c,e, S7b), and one branch intersected with an *in vivo* monocyte-abundant lineage (Fig. 4d,e, S7c). Two subsets of iSCseq spots from iSCseq terminal cluster 0 were extracted based on positional information around bifurcated pathways (Fig. 4b): in vitro OC points in the purple dashed line and common OC points in the orange dotted dashed line (Fig. 4b,f). No *in vivo* OCs, based on the bone milieu atlas classification, were included in the in vitro OC zone (Fig. 4g). Projection of two types of OCs on iSCseq UMAP showed different distributions even in the same terminal cluster (Fig. 4h). *In vitro* OCs were positioned farthest from the differentiation pathway.
We described the characteristics between two types of OCs. Differentially expressed gene (DEG) analysis showed higher expression of *Ctsk* and *Mmp9* in in vitro OCs; in contrast, expression of immunoreceptor tyrosine-based activation (ITAM) motif adaptor associated receptors was separated (Fig. 4i). Collagen-activated pathways were activated in vitro OCs, whereas ITAM-related kinase activity and T cell costimulation were upregulated in common OCs (Fig. 4j). The morphological, physical, and physiological features were also evaluated between two types (Fig. 4k-t). Common OCs contained more single-nucleated cells (Fig. 4k,l) and single-nucleus fractions (Fig. 4o,p). Mitochondrial and Calcium intensity was higher *in vitro* OCs (Fig. 4q-t). Inclusive iSCseq revealed new subsets of OC, whose differences could not be detected using only *in vitro* assays, based on the distribution of information on large datasets and stream of RNA velocity. The iSCseq points clarified the objects to be compared and provided a larger number of DEGs compared with only chromium analysis (Fig. 1d). The new subsets were characterized by multiple pieces of evidence.

Inclusive iSCseq with human bone tumors

In vitro OCs (Fig. 4) with higher expression of Ctsk and Mmp9 did not overlap with OC-annotated cells in physiological bones. This raises the question of whether *in vitro* OC exists in pathological conditions and causes bone destruction. We attempted to identify subsets of OCs in the tumor microenvironment (TME), in which excessive bone resorption often causes pain and pathological fractures.

First, we performed a reanalysis of human scRNA-seq public datasets from giant cell tumor (GCT) of bone in four patients[38] (Fig. S8) and from osteosarcoma in six patients[39] (Fig. S9), which are usually benign and malignant. Copy number variant (CNV) analysis by cell showed that tumor cells were mainly in the mesenchymal lineage, and almost all monocyte-macrophage lineages, including OC, were regarded as normal in both tumors (Fig. S8h,i,j, S9h,i,j).

We then integrated all datasets shown in the previous figures, including iSCseq (Fig. 5a, S10). *In vivo* wild-type cell types of the bone milieu atlas were well separated on the integrated UMAP and could be indicators for cell type identification (Fig. 5b). Blood cells on the left side were almost normal based on CNV analyses (Fig. 5c). We highlighted iSCseq spots on the integrative UMAP coloring of both tumors (Fig. 5d, GCT of Bone; Fig. 5e, osteosarcoma). We focused on the points under OC induction by M-CSF and RANKL and from iSCseq clusters 0 and 1, which were the final two clusters in OC differentiation (Fig. 1e,f). The distribution of the points was clearly separated into four groups (Fig. 5f). Cells from GCT of Bone were distributed around group 2,3,4 (Fig. 5d,f). In contrast, around group 1, there were many malignant osteosarcoma cells (Fig. 5e,f).

Next, we described the four groups of OC lineages in the TME. Group 1 points were mainly from iSCseq terminal cluster 0 and the peripheral zone of iSCseq UMAP (Fig. 5g,h). Although the two types of OC discussed in Fig. 4 did not fully overlap with groups 1 and 2, group 1 contained many in vitro OC; in contrast, group 2 consisted of common OC (Fig. 5i,j). We then performed a gene expression analysis among the four groups (Fig. 4k). Group 1 showed higher expression levels of Ctsk, Mmp9, and Atp6v0d2. In contrast, Fc receptor molecules and Trem2, which are ITAM molecules, were expressed at higher levels in group 2. Groups 3 and 4 showed higher expression of Ctsk and Mmp9 than group 2. From GO analysis, group 1 were the highest acid producers, in contrast, group 3 is related to OC proliferation and regulation of bone resorption (Fig. 5l). Group 3 was positioned mainly in cluster 1, not the final differentiated cluster (Fig. 5g,m). The cluster in which group 3 was mainly positioned (cluster 15 in Fig. S10a) produced acids (Fig. S10f,g) and presented OC-related terms (Fig. S10h,i) at the highest level. The cell population in this cluster was mostly normal cells in the TME. The features based on the images were summarized among the four groups (Fig. S11).

iSCseq identified characteristic bone-resorptive OC subsets in the TME. From these results, we concluded that group 1 to 3 were OCs. Group 1 OCs in malignant osteosarcoma, which mainly overlapped with in vitro OCs, existed in the TME. In addition, group 3 OCs were not assigned to the final iSCseq cluster in OC differentiation, but actively resorbed bone.

Discussion

Traditional single-cell technology using microfluid devices has clarified the cellular heterogeneity in terms of gene expression[40]. However, there is a lack of cellular identity as a trade-off for achieving a high throughput. Moreover, scRNA-seq has 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, considering morphological and physiological annotations (Fig. 6).
Integrative analyses of iSCseq connected conventional scRNA-seq to the cellular or subcellular reality. The origin of the cells or subcellular positions inside a cell can be determined using this combinatorial approach. Large number of scRNA-seq datasets compensate throuputness in iSCseq approach. In contrast, we could discuss a higher number of genes using iSCseq spots compared with traditional scRNA-seq (Fig. 1d). For example, cross-species analysis limits the number of genes in the process of homologue conversion. However, comparison between the focused iSCseq points extracted from high-resolution UMAP allowed the comparison of not only a larger number of genes, but also morphological and physiological features with multi-layered evidence. Cost reduction of sampling and plate-based sequencing is key to random unbiased sampling and sequencing to solve biological and clinical questions. Since scRNA-seq and iSCseq have complementary effects, we recommend using iSCseq in combination to increase the resolution and gain biological meaning of subcellular components.
Multinucleated or large cells are beyond the scope of the trendy single-cell approach. However, iSCseq enabled us to discuss the heterogeneity within cell populations that are difficult to analyze. iSCseq clarified that not all cellular components in in vitro OC culture were involved in the OC differentiation pathway because some subcellular components in morphologically mature OC were assigned to undifferentiated clusters. Conversely, some undifferentiated components were caught or engulfed in OC. Trendy trajectory analysis tends to instill the preconception that differentiation proceeds in a single direction. The iSCseq approach provides insights into a complex differentiation paradigm involving cell fusion.
iSCseq classified some single-nucleated cells as OCs. Dcstamp and Ocstamp, which are necessary for cell fusion[41,42], were upregulated in in vitro OC (group 1 in Fig. 5, S10, 11) and tumor-specific OC (group 3). *In vivo* OCs have been reported to be smaller than *in vitro*-cultured OCs using *in vivo* imaging experiments[43,44]. High expression of Dcstamp and Ocstamp may not be necessary for *in vivo* OC (group 2). The definition of OC should be revised in the omics era. Subsets of OC are a hot topic in the field of skeletal biology. Pathological OC from a different cell source, AtoM, has recently been proposed[45]. Our findings depicted two subsets *in vitro*, and one subset was common in physiological bones. In the tumor microenvironment, in vitro OC existed, and the third subset, still in the process of differentiation, was involved in OC proliferation and regulation of bone resorption. Our inclusive analyses, which connected large datasets with iSCseq, captured new subsets depending on the environment, even from the same cell source.
Moreover, inclusive iSCseq identified candidate switching molecules between subsets of OCs. Epigenetic controls by transcriptional factors NFATc1 and IRF8 are important in OC differentiation[46,47]. At the single cell resolution, IRF8 activity was a candidate regulatory transcriptional factor between in vitro OC and common OC (Fig. S5i). In addition, analyses of differentially expressed genes showed that immunoreceptor tyrosine-based activation motifs (ITAMs) and related costimulatory factors may function are separated between two types (Fig. 4i). ITAMs and costimulatory factors were believed as the third OC differentiation factor from the experiments using knockout or knockdown mice[12,48,49]. iSCseq demonstrated for the first time the physiological roles of ITAMs and related signals.
Based on results of differentially expressed genes and terms (Fig. 4i,j, 5k), plasmalemmal molecules like receptors of costimulatory signals are therapeutic candidate targets for bone resorption. Immunomodulatory drugs targeting Fc receptors[49,50] and TREM2[51,52] coupled with

TYROBP/DAP12 may be good targets to be antagonized for osteoporosis in physiological bones. For excess bone destruction in the tumor microenvironment, Siglec-15[53-55] coupled with TYROBP/DAP12 may be better. Tuning the balance of costimulatory signals depending on the osteimmune environment may be key to increasing bone mineral density.

In conclusion, a newly developed method, iSCseq, demonstrated subcellular heterogeneity of gene expression and its biological significance. iSCseq proved that multiple differentiation stages can be embedded in a cell. The physical responses and physiological distribution of mitochondria and calcium were confirmed using gene expression analysis. Inclusive iSCseq with similar datasets elucidated the physiological mysteries and human skeletal pathogenesis in osteoimmune diversity.

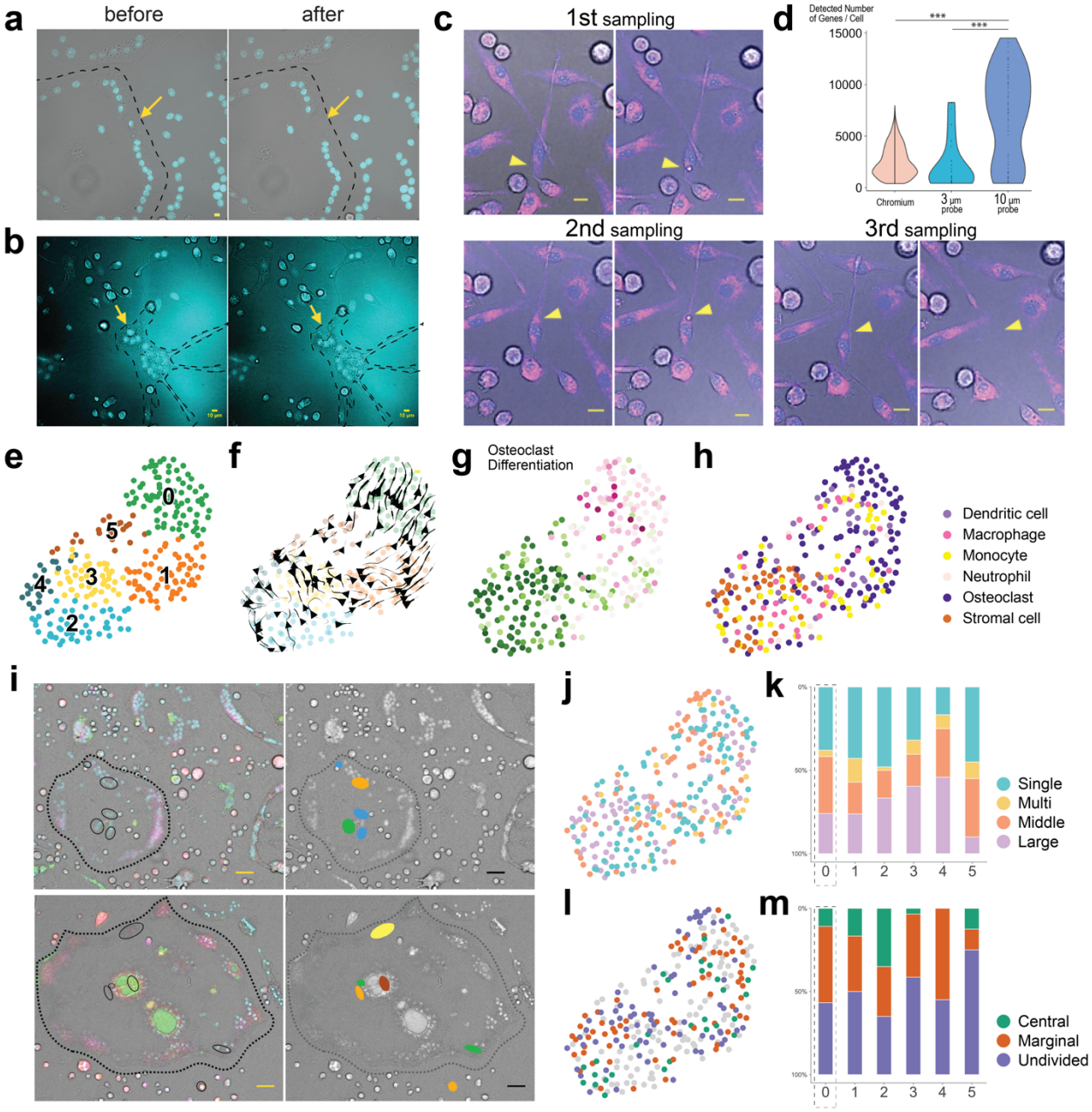


Fig. 1| intra-single cell sequencing (iSCseq) for cellular and subcellular components. **a-c**, Representative cases of sampling each nucleus or subcellular component from a single cell. Blue: nuclei stained with Hoechst 33342. Bar = 10µm. **a,b**, The Area enclosed by the black dotted line represents one osteoclast. **a**, A nucleus from the peripheral zone of large osteoclast (OC). **b**, A nucleus from the aggregation of four nuclei in a OC. **c,** 1st and 2nd sampling for each subcellular component. 3rd sampling for a whole cell with a nucleus. Purple: mitochondria stained with MitoBright LT Deep Red. **d**, Violin plot of genes detected by the sampling and sequencing methods. Left to right, fluid-based chromium (10x genomics), SMART-seq for samples picked by 3 µm probe, and by 10 µm probe. ***p<0.001, One-way ANOVA and Tukey post hoc test. **e**, UMAP consists of 280 cellular and subcellular components. **f**, Streamline presentation based on the spliced/unspliced ratio of RNA using scVelo. **g**, Gene ontology terms including 'Osteoclast

Differentiation' analyzed using decoupleR and AUCell. **h**, Automatic cell-type annotation using CellAssign. **i**, Geometrical information of targeted mature OCs (left). Blue: nuclei stained with Hoechst 33342; Green: Calcium stained with Fluo-4 NW Calcium Assay Kit; Red: acidic environment stained with pHrodo™ Red AM Intracellular pH Indicator; Purple: Tubulin stained with SiR-tubulin kit. The projection of the iSCseq cluster is shown in Fig.**i** right. Bar = 50 µm. **j,k**, Nuclear Rank: bin, 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. **j**, Feature plot of Nuclear Rank. **k**, Cumulative bar plot of the Nuclear Rank in each cluster. **l,m**, Positional information, Central or Peripheral or Undivided in OC. **l**, Feature plot of positional information. **m**, Cumulative bar plot of positional information in each cluster. **k,m**, The clusters surrounded by dotted square are the final clusters in OC differentiation.

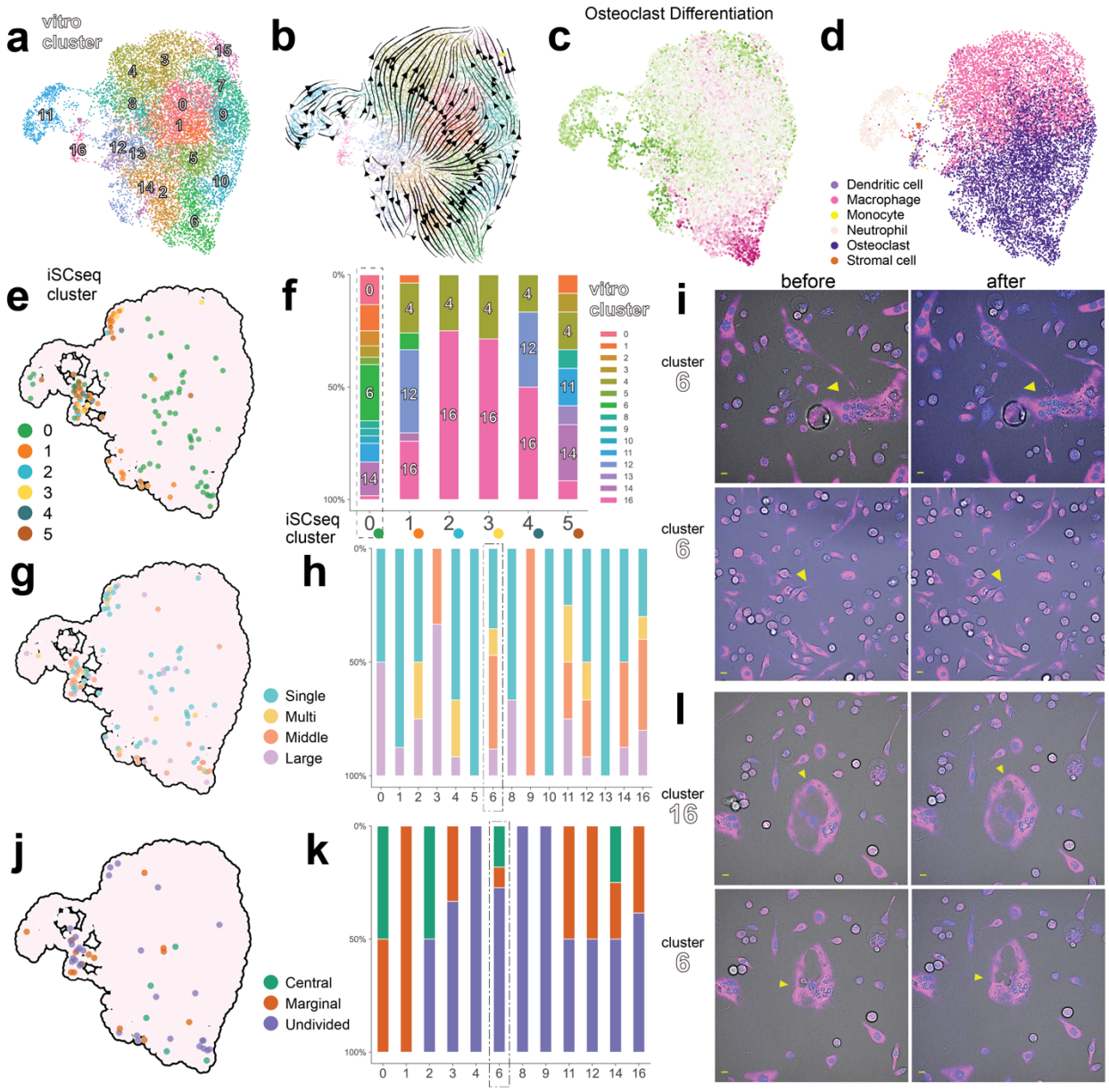


Fig. 2| Inclusive iSCseq with *in vitro* scRNA-seq. **a**, Integrative UMAP of iSCseq and scRNA-seq datasets. **b**, Streamline presentation based on RNA velocity using scVelo. **c**, Gene ontology 'Osteoclast Differentiation' analyzed using decoupleR and AUCell. **d**, Automatic cell-type annotation using CellAssign. **e**, iSCseq spots on integrative UMAP. Colors corresponding to the clusters obtained by iSCseq are shown in **Fig. 1. f,** Cumulative bar plot of integrative *in vitro* clusters for each iSCseq cluster. Terminal cluster 0 in osteoclast (OC) differentiation in iSCseq is surrounded by a dashed square. **g,h**, Nuclear Rank: bin, number of nuclei in the origin of the cell. Large, 20 or more nuclei; Middle, 5 to 19 nuclei; Multi, 2 to 4 nuclei; Single, 1. **g**, Feature plot of the Nuclear Rank. **h**, Cumulative bar plot of the Nuclear Rank in each cluster. **i**, Representative cases of a single-nucleated cell assigned to the final cluster 6 on the differential pathway of OC. **j,k**, Positional information of samples, Central or Peripheral or Undivided in OC. **j**, Feature plot of positional information. **k**, Cumulative bar plot of positional information in each cluster. The

clusters surrounded by dotted dashed squares are the final cluster 6 in OC differentiation of inclusive iSCseq with *in vitro* scRNA-seq (**h,k**). **l**, Representative sampling cases of a nucleus from a multinucleated OC. Blue, Nuclei stained with Hoechst 33342; Purple, MitoBright LT Deep Red. Bar = 10 μm.

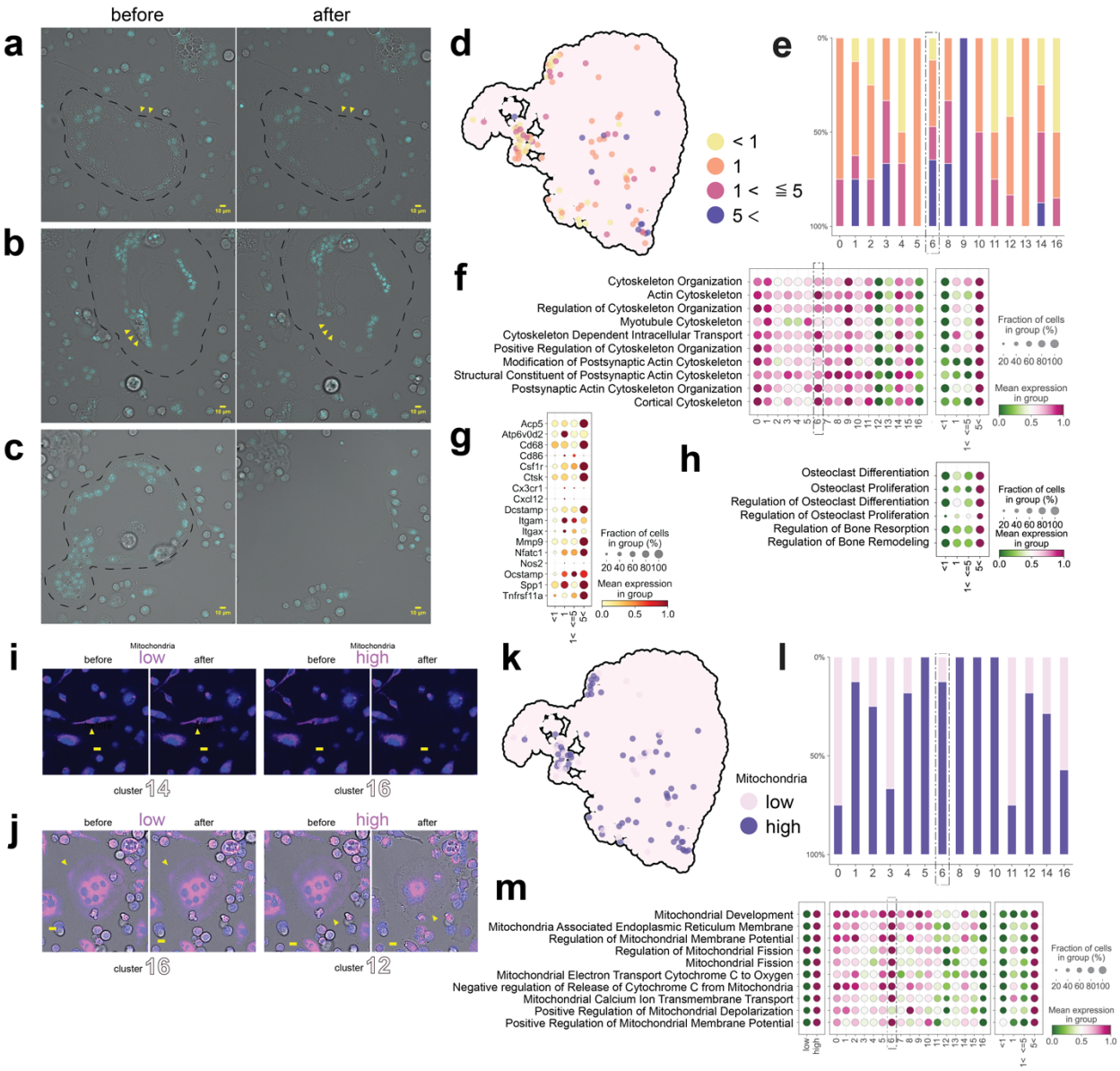


Fig. 3| Morpho-physiological analysis of physical responses and of mitochondria with transcriptome.

a-c, Representative nuclei sampling cases from multinucleated cells. Blue: nuclei stained with Hoechst 33342. The area enclosed by the black dotted line represents one osteoclast (OC). Bar = 10 µm. **a, b**, Yellow arrowheads indicate sampling objects. **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,e**, Binned number of nuclei in each sample. Feature plot (**d**) and cumulative bar plot in each integrative iSCseq cluster shown on **Fig.2a** (**e**). **f**, Dot plot of the top 10 gene ontology (GO) terms including cytoskeleton by decoupleR and AUCell by integrative iSCseq cluster (left) and by binned number of nuclei in each sample (right). **g,h**, Dot plot of OC-related genes (**g**) and of OC-related GO terms (**h**) by the binned number of nuclei in each sample. **i,j**, Representative sampling cases of single nucleated cell (**i**) and multinucleated OC (**j**). Blue, Nuclei stained with Hoechst 33342; Purple, MitoBright LT Deep Red. Bar = 10 µm. **k,l**, High or low information of mitochondrial intensity from

sampling images. Feature plot (**k**) and cumulative bar plot in each integrative iSCseq cluster (**l**). **m**, Dot plot of the top 10 gene ontology terms including mitochondria by decoupleR and AUCell by high or low mitochondrial intensity from images (left), by integrative iSCseq cluster (center), and by binned number of nuclei in each sample (right). Final cluster 6 in OC differentiation is surrounded by dotted dashed squares (**e,f,l,m**).

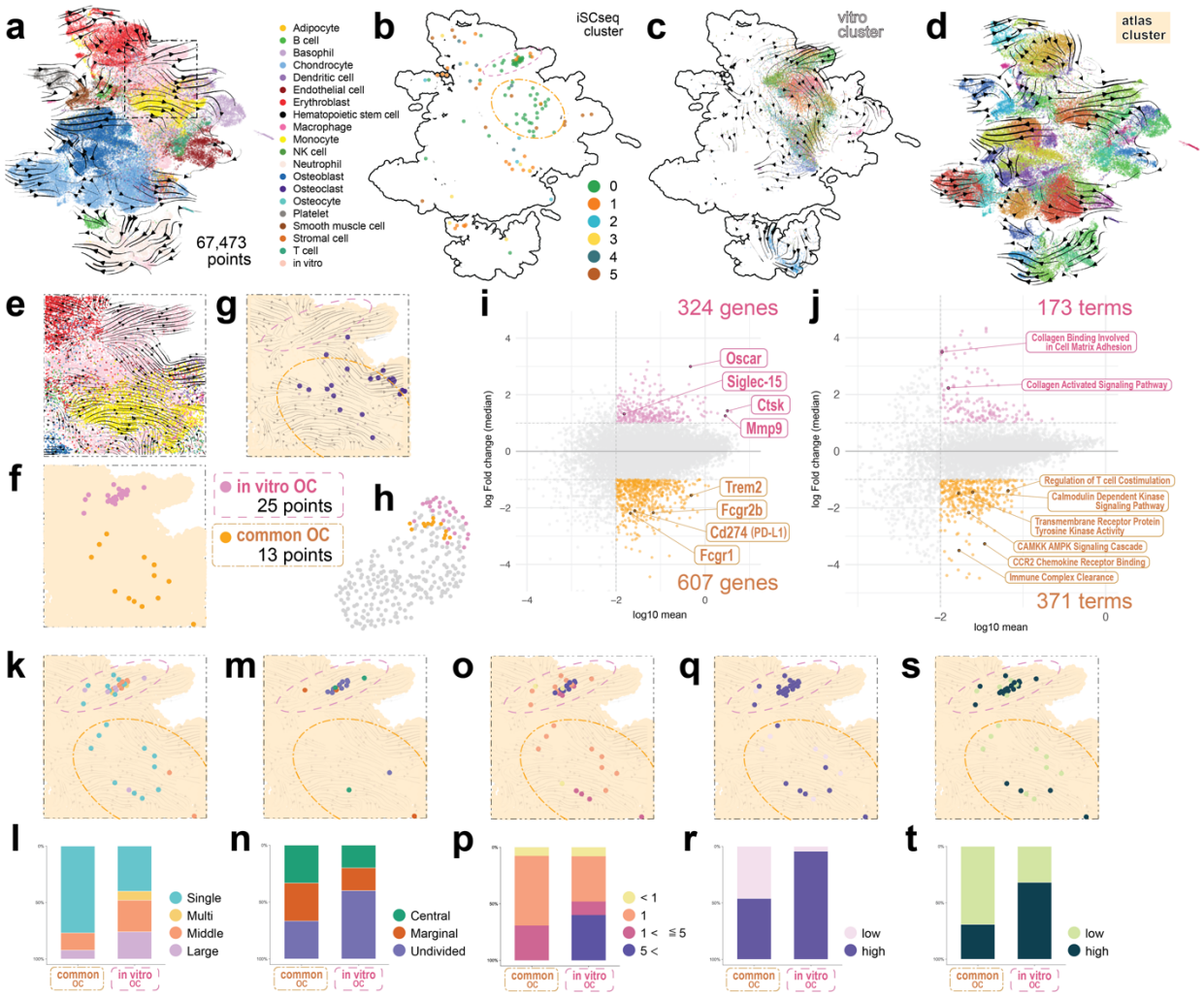


Fig. 4| Inclusive iSCseq with *in vivo* bone milieu atlas. **a**, UMAP of inclusive iSCseq with the bone milieu atlas, which contained iSCseq spots, *in vitro* scRNA-seq cells, and bone milieu atlas cells, colored with cell type annotation of *the in vivo* bone milieu atlas (**Fig. S4**) and RNA velocity stream. *In vitro* spots, including iSCseq, are colored sakura. **b**, iSCseq cluster information presentation on inclusive iSCseq. The two areas where spots in the final cluster of OC differentiation are accumulated are enclosed by a purple dashed circle (in vitro OC points) and an orange dotted dashed circle (common OC points). **c**, *In vitro* iSCseq and scRNA-seq spots in **Fig.2a** with cluster numbers and scVelo streamlines. **d**, Bone milieu atlas cells in **Fig. S4a** with cluster numbers and scVelo streamlines. **e**, Enlarged view of the area surrounded by the double-dotted line square in **Fig. a**. **f**, Two groups of OC spots in the final iSCseq cluster in the enlarged view. **g**, Spots of OC annotated by Cell Assign in the bone milieu atlas in the enlarged view with scVelo streamlines and circles shown in **Fig. b**. **h**, Feature plot of two types of OCs defined in **Fig. f** on iSCseq UMAP (**Fig. 1e**). **i**, M-A plot of differentially expressed genes between the two types in **Fig. f**. **j**, M-A plot of differentially regulated gene ontology terms between the two OC types. **k-t**, Enlarged feature plots (**k,m,o,q,s**) and cumulative bar plots (**l,n,p,r,t**) of the basic characters discussed in **Fig. 2,3,S3** in the double dotted square in **Fig. a** focused on two types of OCs. **k,l**, Nuclear Rank: bin, the number of nuclei in the origin of the

cell. Large, 20 or more nuclei; Middle, 5 to 19 nuclei; Multi, 2 to 4 nuclei; Single, 1. **m,n**, Positional information, Central or Peripheral or Undivided in OC. **o,p**, The binned number of nuclei in each sample. **q-t**, High or low intensity of mitochondria (**q,r**), and calcium (**s,t**) from sampling images.

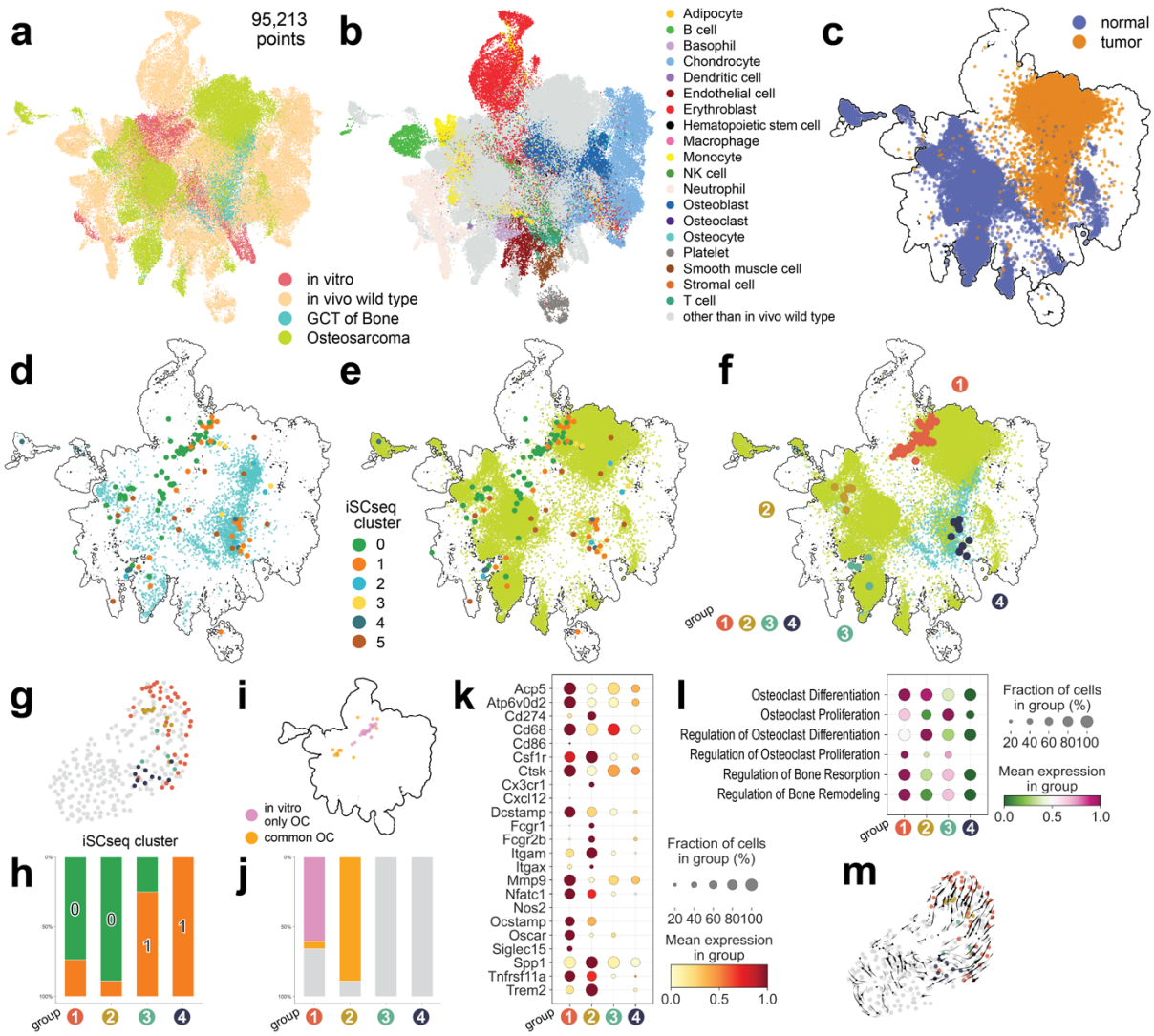


Fig. 5| Inclusive iSCseq analysis with human bone tumors. **a**, UMAP of inclusive iSCseq with the bone milieu atlas, which contained iSCseq spots and *in vitro* scRNA-seq cells (in vitro), bone milieu atlas cells (in vivo wild type), and two types of murinized human tumors. **b**, Cell type annotation in vivo wild type. **c**, Normal or tumor information based on CNV score in each tumor. **d,e**, iSCseq cluster positional information with GCT of Bone cells (**d**) and with Osteosarcoma cells (**e**). **f**, Four groups of osteoclast (OC) induced by M-CSF and RANKL and from iSCseq cluster 0 or 1 with the coloring information of two types of tumor cells. **g**, Feature plot of four groups of OCs defined in **Fig. f** on iSCseq UMAP (**Fig. 1e**). **h**, Cumulative bar plot of iSCseq clusters in the four groups of OCs. **i**, Feature plot of two types of OCs defined in **Fig.4f**. **j**, Cumulative bar plot of two types of OCs defined in **Fig.4f** for the four OC groups. **k**, Dot plot of OC-related genes and selected genes from **Fig. 4i** for four groups of OCs. **l**, Dot plot of gene ontology terms related to OC. **m**, Feature plot of the four groups of OCs and scVelo streamlines (**Fig. 1f**).

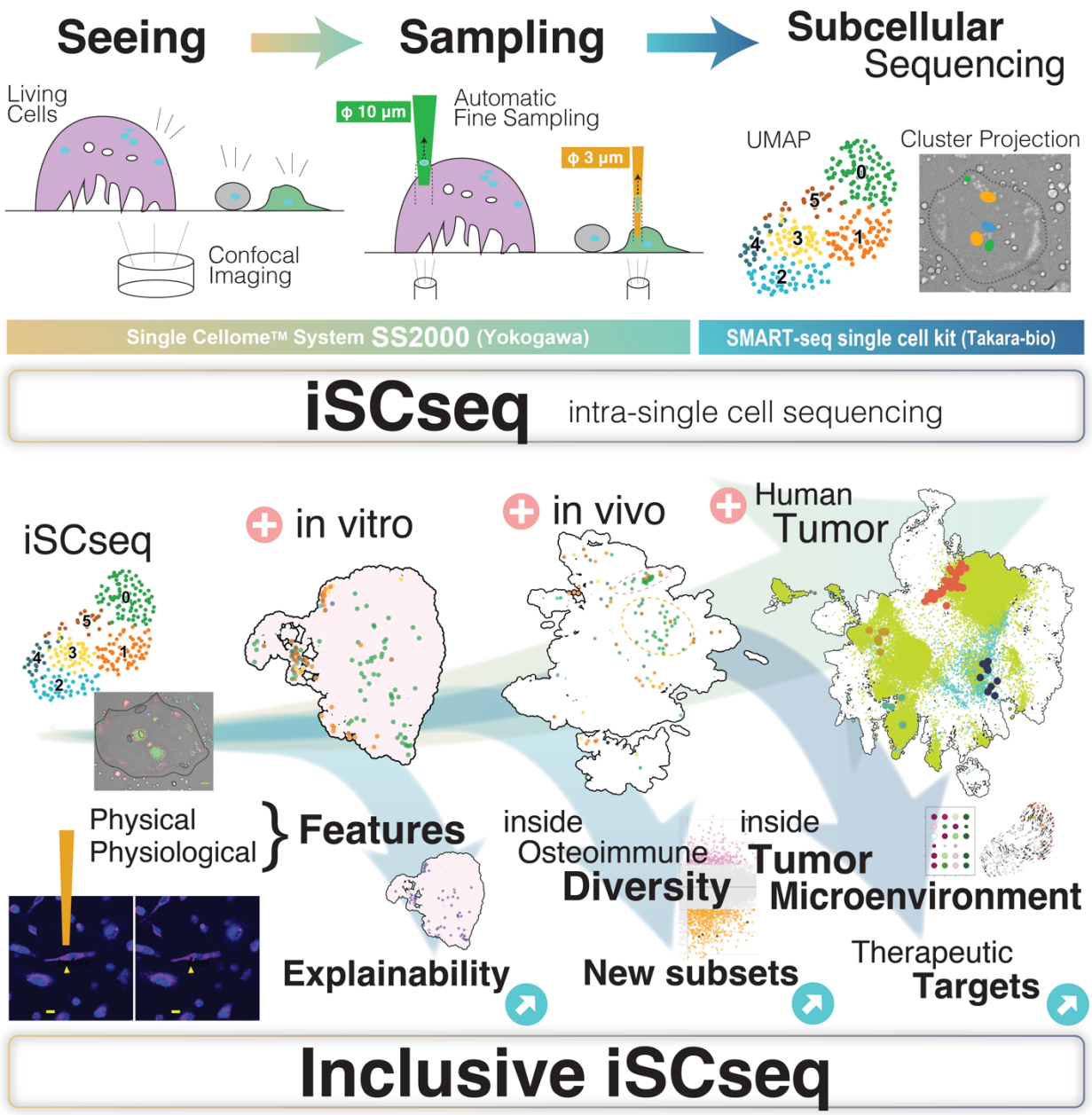
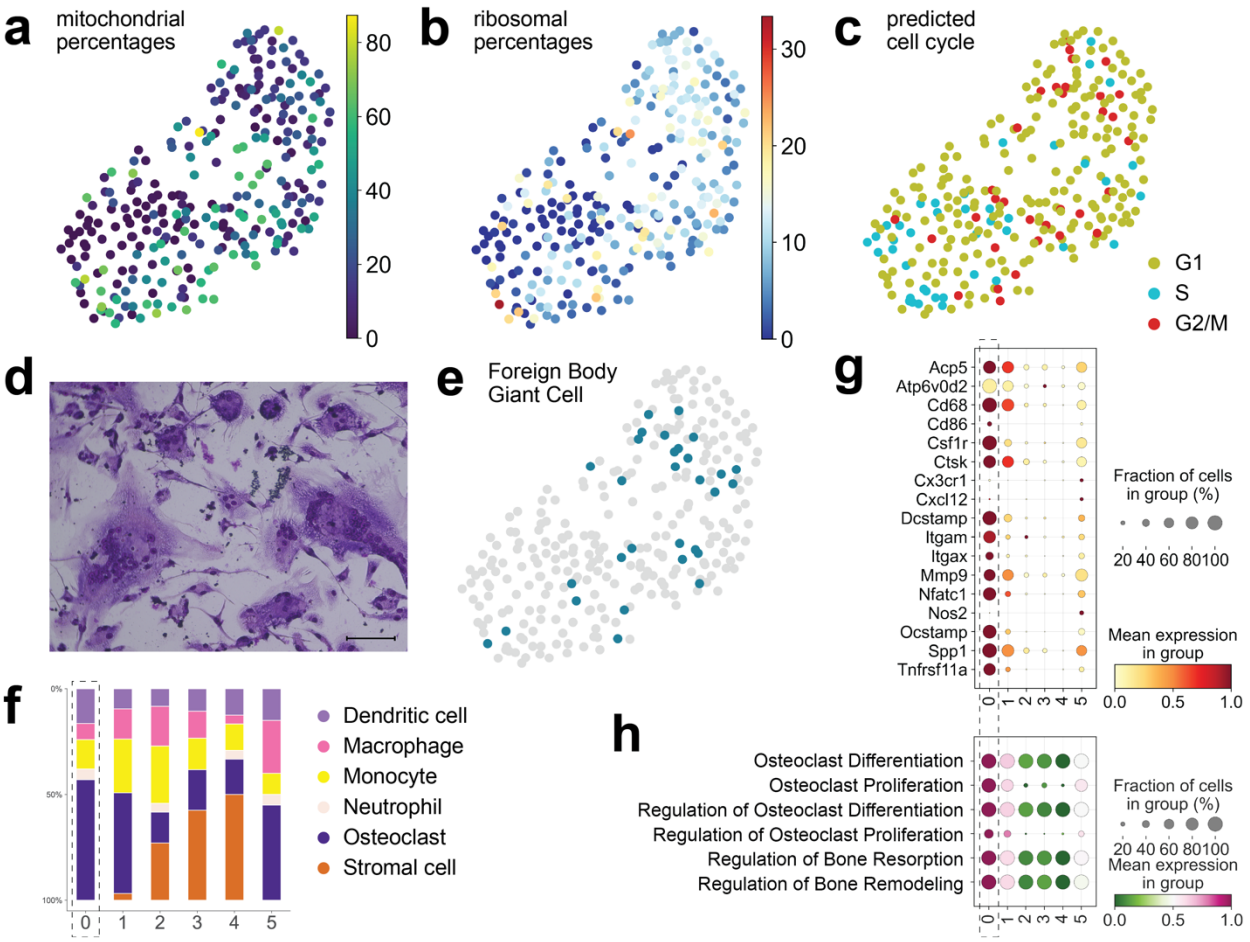
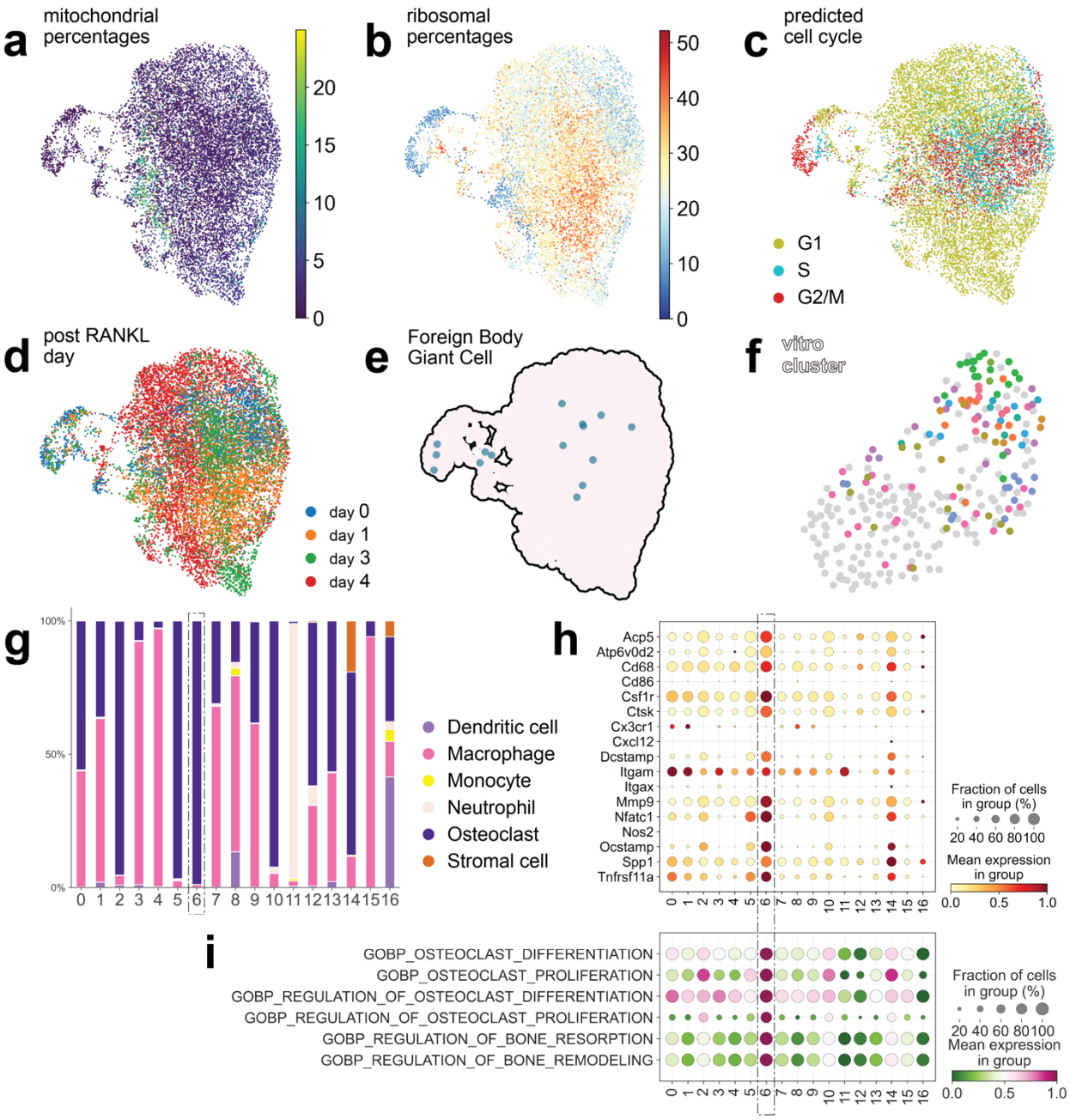


Fig. 6| Schematic representation of iSCseq and inclusive iSCseq.
Intra-single cell sequencing (iSCseq) is a new sequencing approach that combines confocal imaging, picking up any cellular and subcellular components inside cells using the Single Cellome^{TM} System SS2000 (Yokogawa), and next-generation sequencing using SMART-seq single cell kit (Takara-bio).
Inclusive iSCseq is a combination of iSCseq and conventional scRNA-seq datasets. Inclusive iSCseq enlarges the information space for gene expression using similar scRNA-seq datasets, clarifies the physical and physiological features, reveals new cellular subsets, and provides therapeutic targets.

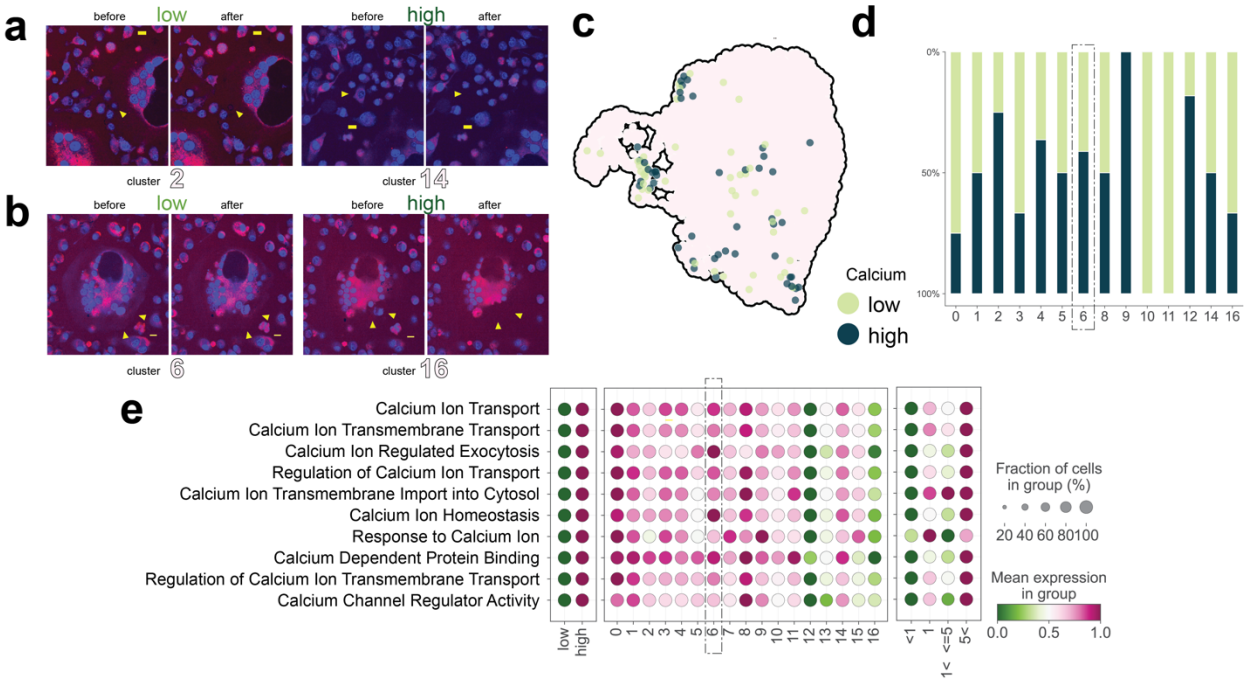


Extended Data Fig. S1| Basic parameters of intra-single cell sequencing (iSCseq).
a-c, Feature Plots of mitochondrial percentage (**a**), of ribosomal percentage (**b**), and of cell cycles (**c**) from iSCseq. **d**, representative foreign body giant cells (FBGCs) stained with May-Grünwald Giemsa. Bar = 50 µm. **e**, Feature plot of FBGCs represented by moss-colored spots. **f**, Cumulative bar plot of cell types using CellAssign. **g**, Dot plot of osteoclast-related genes. **h**, Dot plot of osteoclast-related GO terms calculated by decoupleR by each iSCseq cluster. The final cluster in the OC differentiation pathway is indicated by dashed squares (**g,h**).

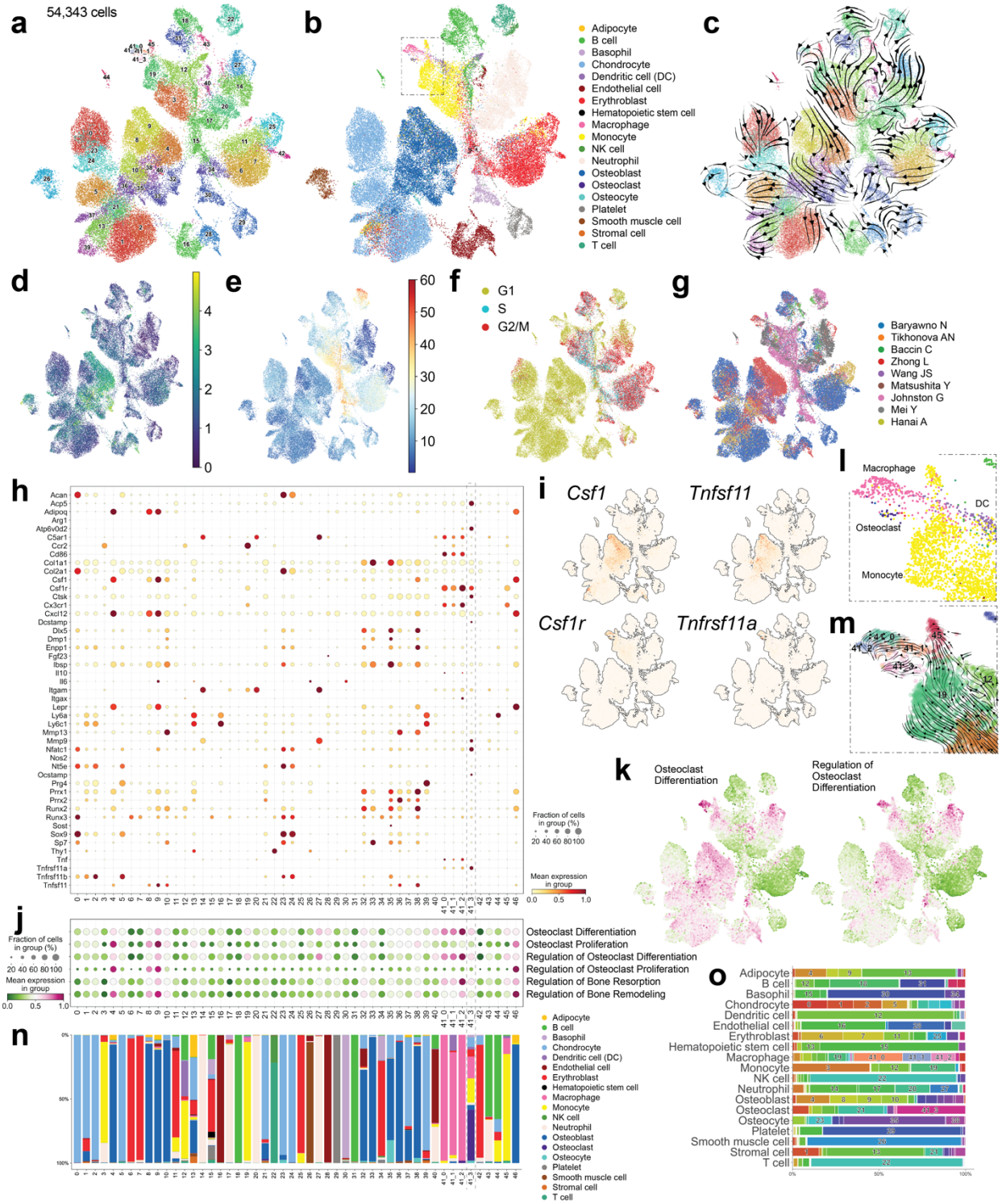


Extended Data Fig. S2| Basic parameters of integrative iSCseq with *in vitro* cultured scRNA-seq datasets.

a-d, Feature Plots of mitochondrial percentage (**a**), of ribosomal percentage (**b**), of cell cycle (**c**), and of post-RANKL stimulation day (**d**) from integrative iSCseq. **e,** Feature plot of FBGCs represented by moss-colored spots. **f**, Integrative clusters spotted on iSCseq UMAP. **g**, Cumulative bar plot of the cell type using CellAssign. **h**, Dot plot of osteoclast-related genes. **i**, Dot plot of osteoclast-related GO terms calculated by decoupleR by each iSCseq cluster. The final cluster in the OC differentiation pathway is highlighted by dotted dashed squares (**g,h,i**).

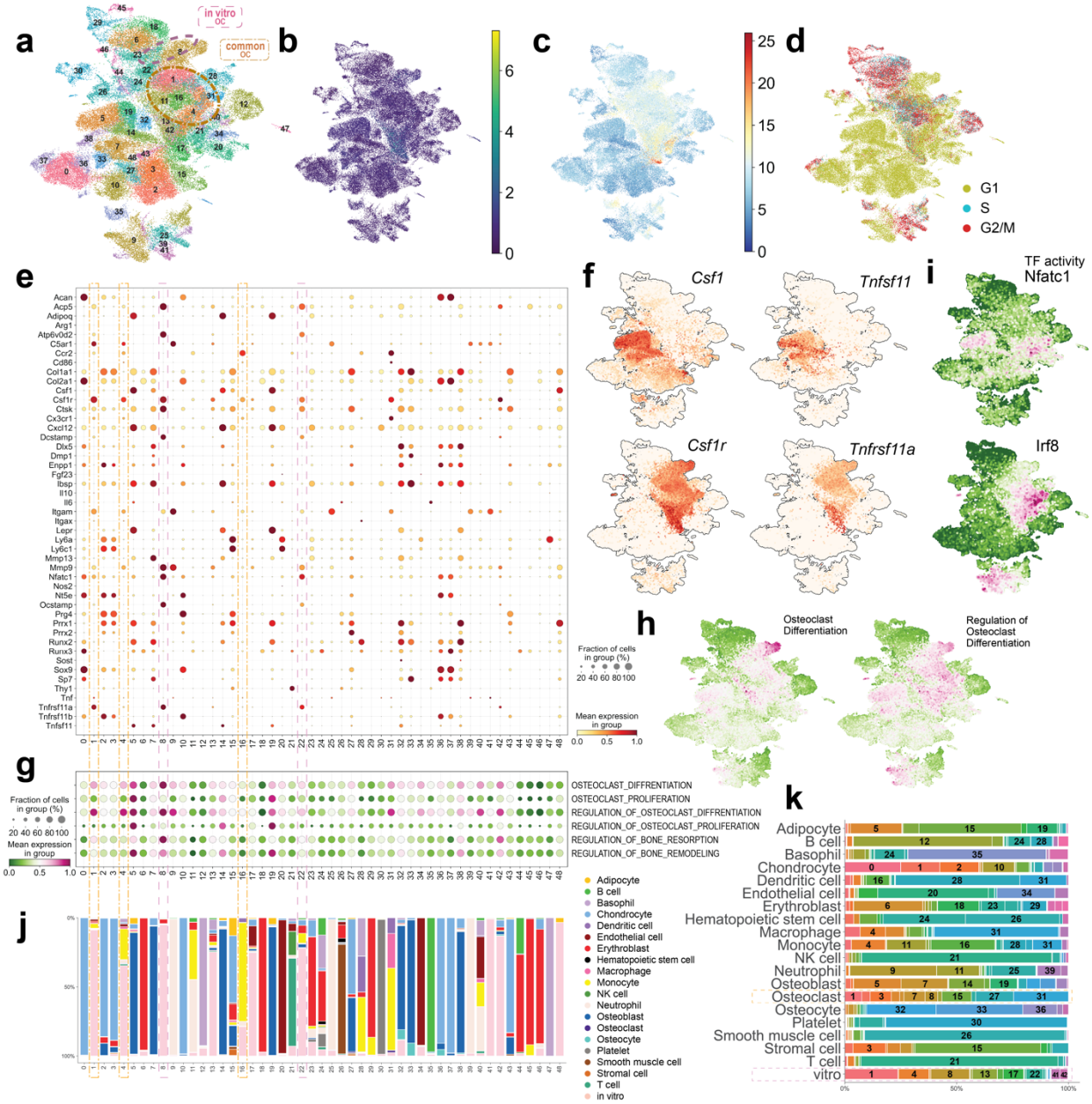


Extended Data Fig. S3| Morphophysiological analysis of calcium with the transcriptome. **a,b,** Representative sampling cases of single nucleated cell (**a**) and multinucleated OC (**b**) with qualitative high or low calcium (Ca) intensity and with integrative iSCseq cluster in **Fig. 2a**. Blue: Nuclei stained with Hoechst 33342; Red: Calbryte^{TM} 590 ABD. Bar = 10 µm. **c,d**, High or low Ca staining in each iSCseq sample. **c**, Feature plot highlighted by iSCseq spot with high or low Ca intensity. **d**, Cumulative bar pot of high or low Ca intensity in each integrative iSCseq cluster. **e**, Dot plot of the top 10 gene ontology (GO) terms including Ca using decoupleR and AUCell by high or low Ca intensity from images (left), by integrative iSCseq cluster (center), and binned number of nuclei in each sample (right). Final cluster 6 in OC differentiation is surrounded by dotted dashed squares **(d,e).**



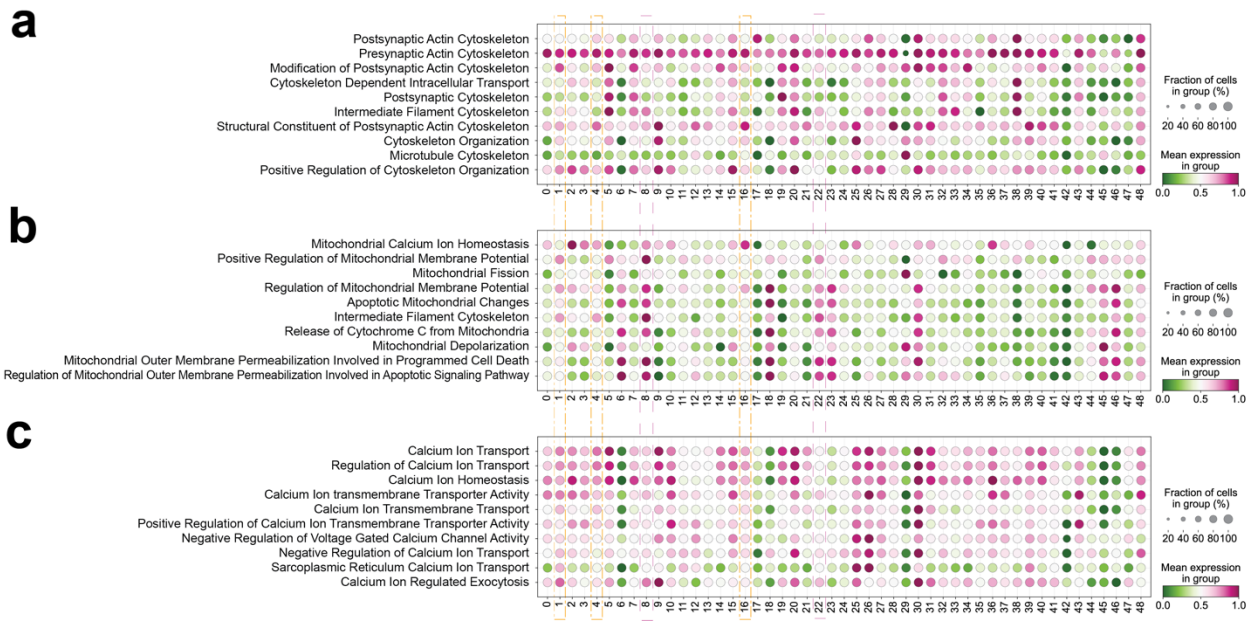
Extended Data Fig. S4| Bone milieu atlas. **a**, UMAP of Bone milieu (bone and bone marrow) atlas from 28 datasets from nine papers, including wild-type or wild-type equivalent control cells. **b**, Automatic cell-type annotation using CellAssign. **c**, scVelo RNA velocity streamlines. **d-g**, Feature Plots of mitochondrial percentage (**d**), of ribosomal percentage (**e**), of cell cycle (**f**), and of paper authors (**g**). **h**, Dot plot of skeletal gene set. The osteoclast-annotated cluster is

enclosed by a dotted square. **i**, Feature plots of gene expression related to osteoclast (OC) differentiation factors. **j**, Dot plot of Gene ontology terms related to OC. **k**, Feature plot of gene ontology terms related to osteoclast differentiation. **l,m**, Enlarged view of the area surrounded by a dotted dashed square in **Fig.b**. CellAssign cell type (**g**) and RNA velocity streamlines (**h**). **n**, Cumulative bar plot of cell types by CellAssign by cluster. **o**, Cumulative bar plot of cell types using CellAssign. Cluster numbers including >10% of cells in each cell type are shown.

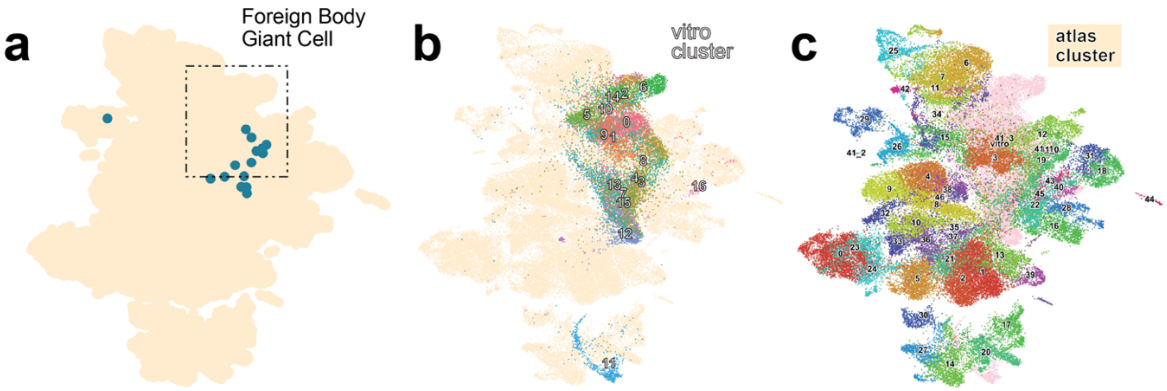


Extended Data Fig. S5| Basic parameters of inclusive iSCseq with bone milieu atlas (Fig. 4). **a**, UMAP with inclusive iSCseq, *in vitro* OC culture, and *in vivo* bone and bone marrow datasets. The two groups of differentiated OCs identified by iSCseq (**Fig. 4m**) are enclosed by a purple dashed circle (in vitro OC points) and an orange dotted dashed circle (common OC points). **B-d**, Feature Plots of mitochondrial percentage (**b**), of ribosomal percentage (**c**), and of cell cycle (**d**). **e**, Dot plot of skeletal gene set. Clusters including the two types of differentiated OCs (mainly defined by circles' position on UMAP and Tnfsf11a upregulation) are enclosed by the corresponding color of squares in **Fig. a**. **f**, Feature plots of gene expression related to osteoclast (OC) differentiation factors. **g,h**, Gene ontology analysis related to osteoclast differentiation was calculated using decoupleR and AUCell. Dot plot (**g**) and feature plots (**h**). **i**, Feature plot of regulatory transcriptional factors calculated using decoupleR and AUCell. **j**, Cumulative bar plot of cell types using CellAssign by cluster. **k**, Cumulative bar plot of inclusive iSCseq clusters.

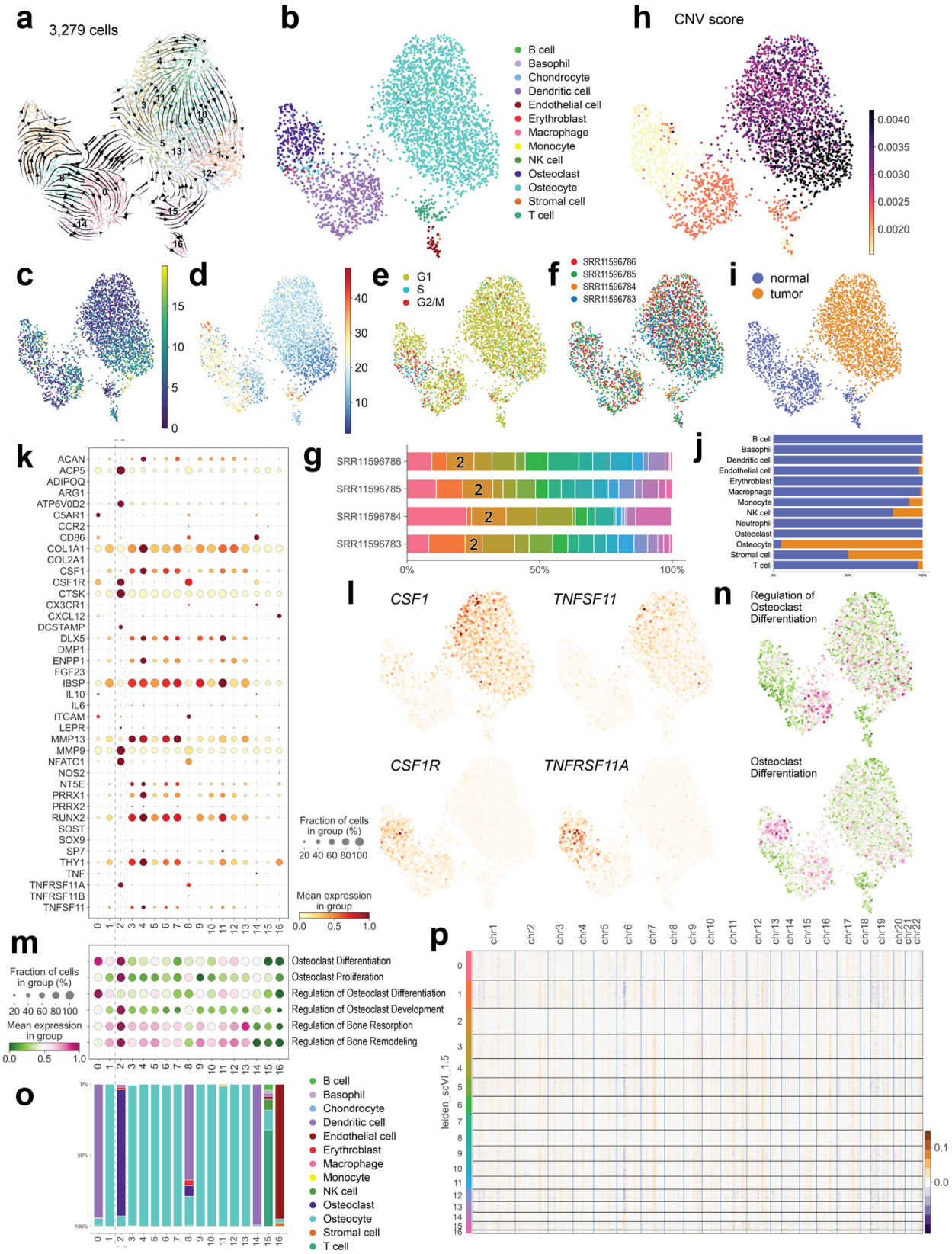
Clusters containing > 10% cells in each cell type were numbered. Clusters in osteoclast and *in vitro* were numbered as >3%.



Extended Data Fig. S6| Gene ontology analysis in inclusive iSCseq with bone milieu atlas (Fig. 4, S5).
Dot plots of the top 10 gene ontology (GO) terms, including cytoskeleton (**a**), mitochondria (**b**), and calcium (**c**), in inclusive iSCseq with *in vitro* cultured OC and *in vivo* bone and bone milieu atlas. Clusters including the two types of differentiated OCs (defined in **Fig. S5e**) are enclosed by the corresponding squares (in vitro OC, purple dash squares; common OC, orange dotted dashed squares).

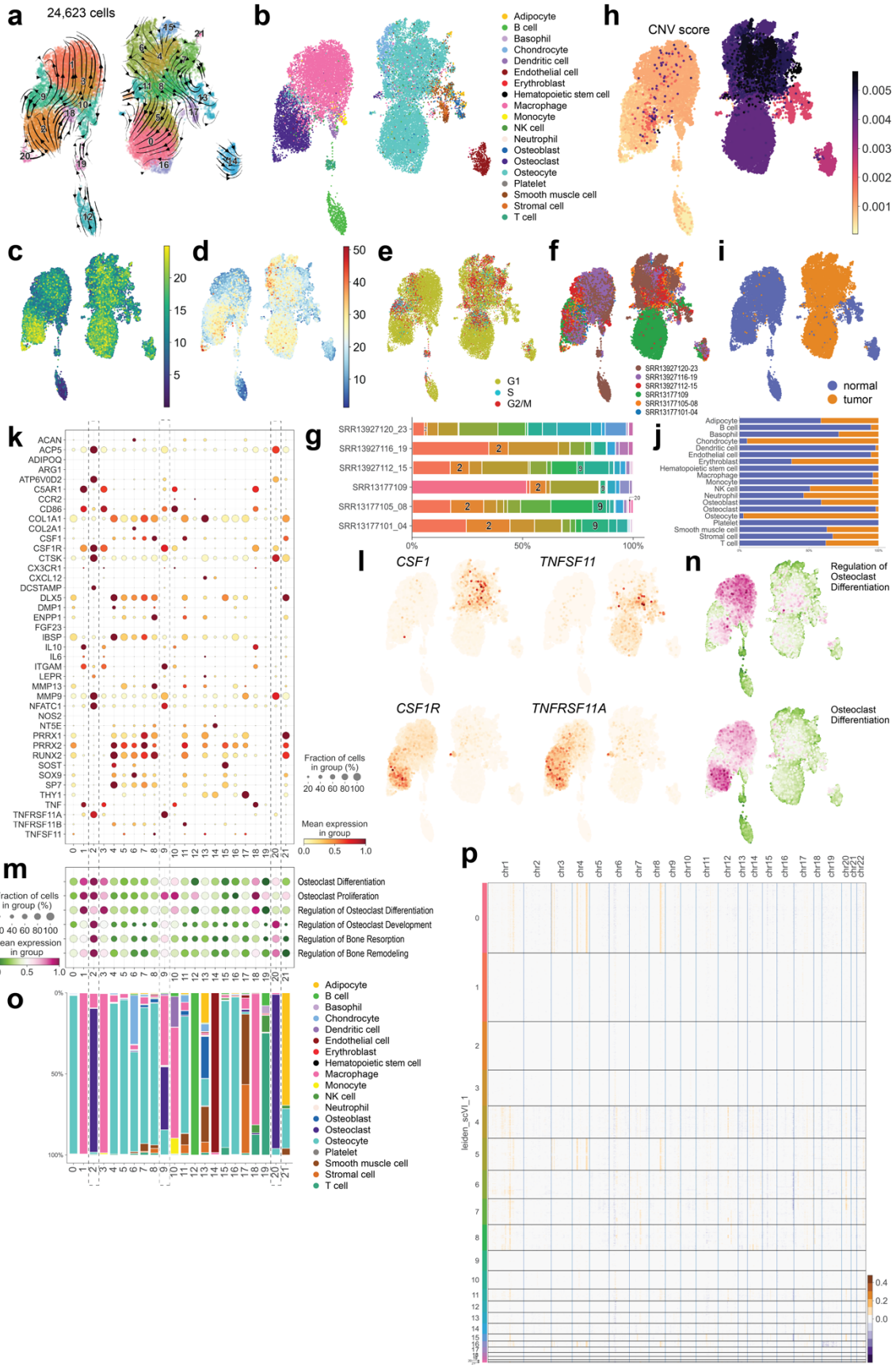


Extended Data Fig. S7| Feature plots of inclusive iSCseq with bone milieu atlas (Fig. 4).
a, Feature plot of FBGCs (**Fig. S1d**) on integrative iSCseq UMAP, which contained iSCseq spots, *in vitro* scRNA-seq cells, and bone milieu atlas cells. The dotted dashed square is the same as in **Fig. 4a**. **b**, *In vitro* iSCseq and scRNA-seq spots shown in **Fig. 2a** with *in vitro* cluster numbers. **c**, Bone milieu atlas cells in **Fig. S4a** with the atlas cluster number.



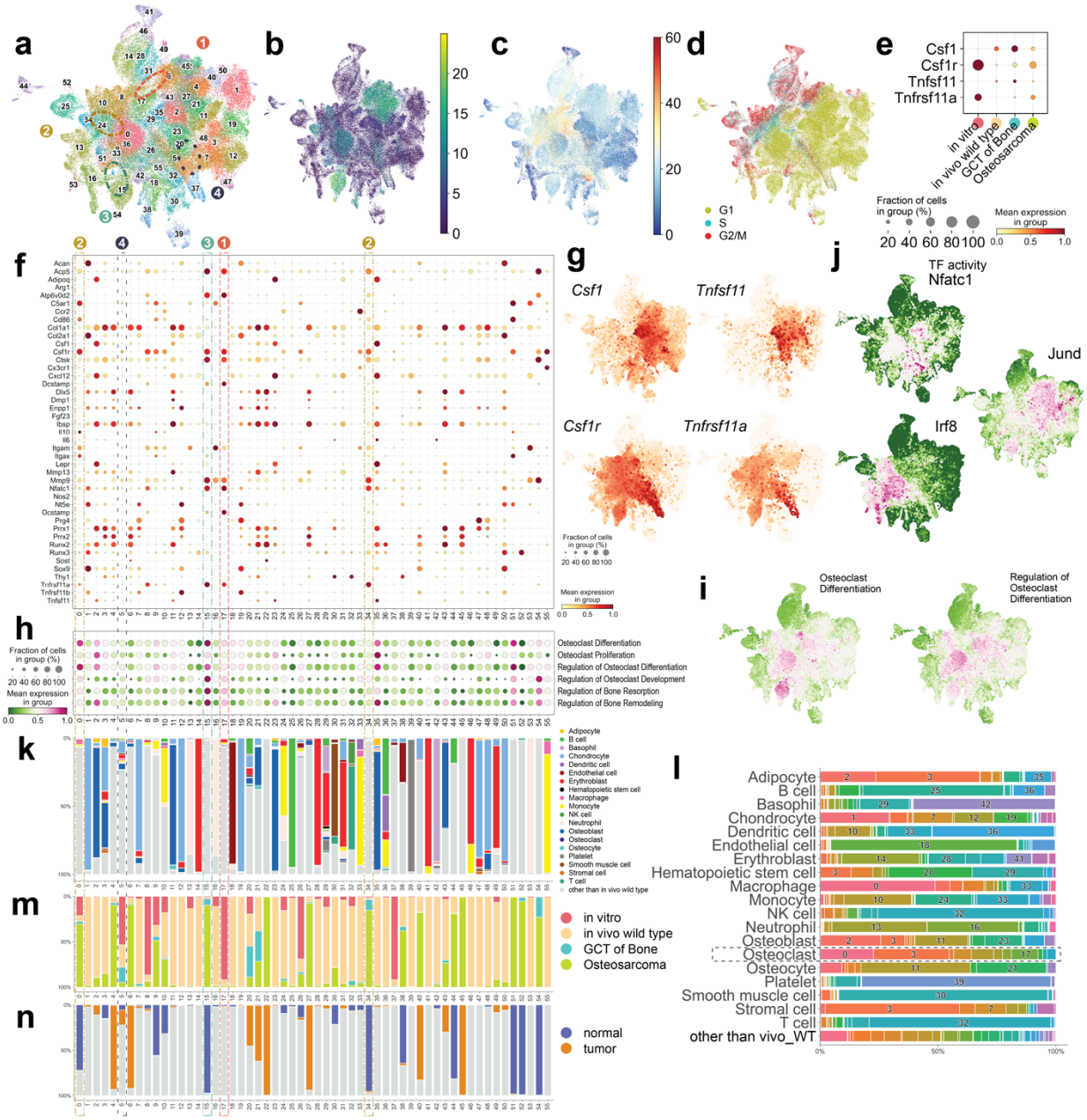
Extended Data Fig. S8| Basic parameters of giant cell tumor of bone datasets from four patients.

Reanalysis of public scRNA-seq datasets from four patients with giant cell tumor (GCT) of the bone. **a**, UMAP with scVelo streamlines. **b**, Automatic cell-type annotation using CellAssign. **c-f**, Feature Plots of mitochondrial percentage (**c)**, of ribosomal percentage (**d)**, of cell cycle (**e**), of patient (**f**). **g**, Cumulative bar plot of cluster components by patient. OC-enriched cluster 2 is highlighted. **h**, Copy number variant (CNV) score calculated using infercnvpy. **i**, Feature plot of normal or tumor information based on CNV score. **j**, Cumulative bar plot of normal or tumor by cell type. **k**, Dot plot of skeletal gene set. **l**, Feature plot of skeletal gene set. **m**, Dot plot of osteoclast (OC)-related terms. **n**, Feature plot of osteoclast (OC)-related terms. OC-annotated cluster 2 is surrounded by a dashed square (**k,m,o**). **p**, Heatmap of copy number variants calculated using infercnvpy.


a
24,623 cells
b
Adipocyte
B cell
Basophil
Chondrocyte
Dendritic cell
Endothelial cell
Erythroblast
Hematopoietic stem cell
Macrophage
Monocyte
NK cell
Neutrophil
Osteoblast
Osteoclast
Osteocyte
Platelet
Smooth muscle cell
Stromal cell
T cell
h
CNV score
0.005
0.004
0.003
0.002
0.001
c
d
e
G1
S
G2/M
f
SRR13927120-23
SRR13927116-19
SRR13927112-15
SRR13177109
SRR13177105-08
SRR13177101-04
i
normal
tumor
k
ACAN
ACP5
ADIPOQ
ARG1
ATP6V0D2
C5AR1
CCR2
CD86
COL1A1
COL2A1
CSF1
CSF1R
CTSK
CX3CR1
CXCL12
DCSTAMP
DLX5
DMP1
ENPP1
FGF23
IBSP
IL10
IL6
ITGAM
LEPR
MMP13
MMP9
NFATC1
NOS2
NT5E
PRRX1
PRRX2
RUNX2
SOST
SOX9
SP7
THY1
TNF
TNFRSF11A
TNFRSF11B
TNFSF11
Fraction of cells in group (%)
20 40 60 80100
Mean expression in group
0.0 0.5 1.0
g
SRR13927120_23
SRR13927116_19
SRR13927112_15
SRR13177109
SRR13177105_08
SRR13177101_04
0%
50%
100%
j
l
CSF1
TNFSF11
CSF1R
TNFRSF11A
n
Regulation of Osteoclast Differentiation
Osteoclast Differentiation
m
Osteoclast Differentiation
Osteoclast Proliferation
Regulation of Osteoclast Differentiation
Regulation of Osteoclast Development
Regulation of Bone Resorption
Regulation of Bone Remodeling
o
p
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
leiden_scVI_1
0.4
0.2
0.0

Extended Data Fig. S9| Basic parameters of osteosarcoma datasets from six patients (Fig. 5d-f).

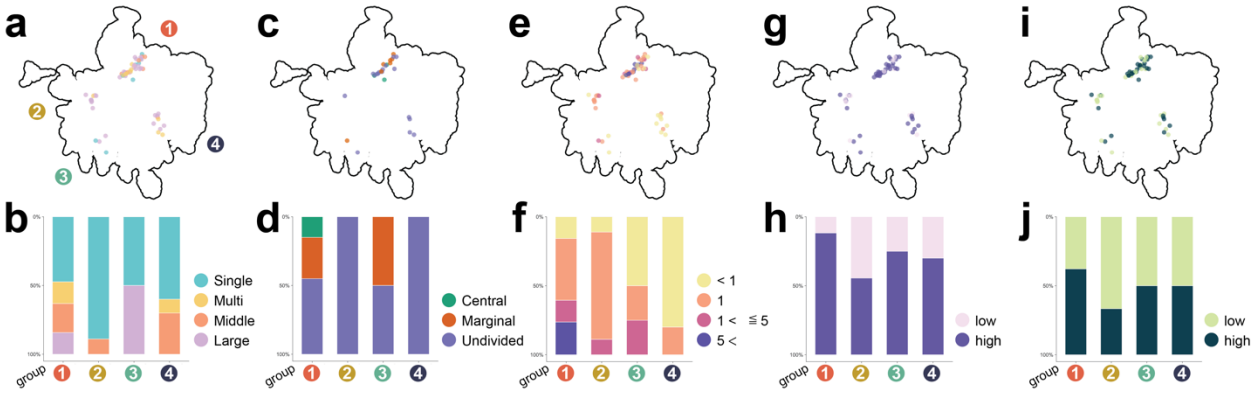
Reanalysis of public scRNA-seq datasets from six patients with osteosarcoma. **a**, UMAP with scVelo streamlines. **b**, Automatic cell-type annotation using CellAssign. **c-f**, Feature Plots of mitochondrial percentage (**c)**, of ribosomal percentage (**d)**, of cell cycle (**e**), of patient (**f**). **g**, Cumulative bar plot of cluster components by patient. OC-enriched cluster 2 is highlighted. **h**, Copy number variant (CNV) score calculated using infercnvpy. **i**, Feature plot of normal or tumor information based on CNV score. **j**, Cumulative bar plot of normal or tumor by cell type. **k**, Dot plot of skeletal gene set. **l**, Feature plot of skeletal gene set. **m**, Dot plot of osteoclast (OC)-related terms. **n**, Feature plot of osteoclast (OC)-related terms. OC-annotated cluster 2 is surrounded by a dashed square (**k,m,o**). **p**, Heatmap of copy number variants calculated using infercnvpy.



Extended Data Fig. S10| Basic parameters of iSCseq with the bone milieu atlas and murinized human tumors (Fig. 5).

a, UMAP of inclusive iSCseq with scRNA-seq of *in vitro* cultured OC, of bone milieu atlas, and of murinized human tumors. The accumulated area of the four groups of differentiated OCs in iSCseq (defined in **Fig. 5f**) are enclosed by circles (group 1 (mainly in vitro OC), orange dashed circle; group 2 (mainly common OC), gold dotted dashed circle; group 3 (tumor-specific OC), blue green double dotted dashed circle; group 4, dark blue dotted circle). **b-d**, Feature Plots of mitochondrial percentage (**b**), of ribosomal percentage (**c**), and of cell cycle (**d**). **e**, Dot plot of OC differentiation factors and their receptors by sequencing categories. **f**, Dot plot of the skeletal gene set. Differentiated OCs shown in **Fig. a (**mainly defined by circles' position on UMAP and Tnfsf11a upregulation**)** are enclosed by the corresponding colored squares. **g**, Feature plots of

gene expression related to osteoclast (OC) differentiation factors. **h,i**, Gene ontology analysis related to osteoclast differentiation calculated using decoupleR and AUCell. Dot plot (**h**) and feature plots (**i**). **j**, Feature plot of regulatory transcriptional factors calculated using decoupleR and AUCell. **k**, Cumulative bar plot of cell types by CellAssign by cluster. **l**, Cumulative bar plot of the inclusive iSCseq cluster by cell type. Clusters containing >10 percent of cells in each cell type were numbered. **m,n**, Cumulative bar plot of sample category (**m**) and normal or malignant information in tumor samples (**n**) by cluster.



Extended Data Fig. S11| Basic characteristics of iSCseq with bone milieu atlas and murinized human tumors (Fig. 5).

Feature plots (**a,c,e,g,i**) and cumulative bar plots (**b,d,f,h,j**) of the basic characters discussed in **Fig. 2,3,S3** for the four groups of OCs defined in **Fig. 5f**. **a,b**, Nuclear Rank: bin, the number of nuclei in the origin of the cell. Large, 20 or more nuclei; Middle, 5 to 19 nuclei; Multi, 2 to 4 nuclei; Single, 1. **c,d**, Positional information, Central or Peripheral or Undivided in OC. **e,f**, The binned number of nuclei in each sample. **g-j**, High or low intensity of mitochondria (**g,h**) and calcium (**i,j**) from sampling images.

Movie S1.

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

Movie S2.

Sampling a single nucleus from four nuclei aggregation inside a medium-sized osteoclast (**Fig. 1b**).

Acknowledgments: Live cell imaging and intra-cellular sampling were technically supported by Yokogawa Electric Corporation from the planning stage of the experiments. Yuta Terui is an employee of Yokogawa Electric Corporation. Our next generation sequencing was successful with a highly elaborate manufacturer's kit from Takara Bio and Clontech. The supercomputer SHIROKANE at the Human Genome Center at the University of Tokyo was partially used in the calculation. The platform for Advanced Genome Science in Japan has 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: The authors have no competing interests to declare.
Data and material availability: All videos, data, and materials are included in the supplementary materials. Raw fastq files of iSCseq were deposited in DDBJ (DNA Data Bank of Japan). The accession numbers are DRA015536 and DRA016857 (will be public after accepted).

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.), 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 a 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 with 50 ng/ml granulocyte-macrophage colony-stimulating factor (GM-CSF; R&D Systems, Minneapolis, MN, U.S.) and 50 ng/ml IL-4 (R&D Systems). More than three nuclei 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. 1c, Fig. 2i,l, and Fig. 3i,j were stained for nuclei with Hoechst 33342, then exposed to a mitochondrial staining solution containing MitoBright LT Deep Red (Dojindo, Kumamoto, Japan) in 1 mL Fluo-4 AM Assay buffer (HEPES buffer), incubated at 37°C for 45 min, and washed with MEMalfa containing 10% fetal bovine serum.

Cells in Fig. 1i were stained for nuclei by Hoechst 33342, then 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 thoroughly washed with phosphate-buffered saline (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 Fig. S3a,b were stained for nuclei by Hoechst 33342, then exposed to a calcium staining solution containing 1μL MitoBright LT Deep Red Calbryte™ 590AM (AAT Bioquest, California, U.S.) in 1 mL Fluo-4 AM Assay buffer, 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, including M-CSF and RANKL, 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, fine multi-colored images were obtained 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 the key to collecting 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 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 in 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 was the maximum number of cycles required for the minimum input of RNA. The 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 SMART-Seq Single-Cell PLUS Kit, was used for library preparation. The library amplification and PCR cycles were determined according to the concentrations detected by the bioanalyzer. After library amplification, the samples were pooled and purified once using 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 150 bp paired ends using HiSeq X Ten (Illumina, San Diego, CA, U.S.) by Macrogen Japan (Tokyo, Japan) or using 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 using STAR ver. 2.7.10b[56] onto the primary assembly mouse genome GRCm39 and basic gene annotation vM31, or the human genome hg38 and basic annotation v42 downloaded from GENCODE[57]. The gene expression levels of iSCseq were summarized in a count matrix using RSEM[58]. Transcripts Per Kilobase Million (TPM) were calculated for iSCseq RSEM results using 'rsem-generate-data-matrix-modified' in script_rnakato[59] prior to integration with 10x chromium datasets.

Bioinformatics analyses were performed using Python 3.10.11[60]. Scanpy[19] in scverse project[21] was used to process and visualize the count matrix. Scvi-tools were used to filter datasets considering basic parameters (cutoff values shown in Figs of Basic Parameters) and to integrate datasets with a model considering the experimental batch, including the author

information of the dataset, total count, probe diameter, and cell cycle score based on cell cycle genes defined by the Seurat team[22,26,27,61,62], percentages of leads on mitochondrial genes, and ribosomal genes. Scvi-tools are highly recommended for dataset integration in benchmark study[63]. The model was trained using a graphical processing unit (NVIDIA T4) on Google Clouds and Colab. Uniform Manifold Approximation and Projection (UMAP) was performed for dimension reduction[18]. Clustering was performed using the Leiden algorithm[64] at the following resolution: 1 (Fig. 1, S1), 1 (Fig. 2,3, S2,3), 2.5 (Fig. 4a-f, S4), 2 (Fig. 4i-r, S5-7), 1.5 (Fig. 5a-c, S8), 1 (Fig. 5d-f, S9), and 2 (Fig. 5g-l, S10). In bone milieu atlas processing (Fig. 4a), we extracted cluster 41 containing osteoclasts and performed an additional clustering at a resolution of 0.4 into four subclusters, and then subcluster information was shadowed on the prior UMAP.

Automatic cell-type annotation was performed using CellAssign[25], referring to the marker genes on CellMarker 2.0[65] and PanglaoDB[66]. We classified the cells into six cell types, as shown in Fig. 1h and Fig. 2d, for the *in vitro* assay. For the *in vivo* assay, we classified 20 cell types, as shown in Fig. 4b and 5b,e,h. Gene ontology scores and regulatory transcriptional factor activities were calculated using decoupleR[24] and AUcell[67,68] referring to mSigDB[69-71].

Prior to integrating human tumor scRNA-seq datasets, we converted hg38 human gene symbols into corresponding mm39 mouse gene symbols using homologous information from BioMart[72]. A total of 14,826 genes with human and mouse homologues were analyzed in the inclusive analysis in Fig. 5, S10,11. Copy number variant (CNV) score was calculated using infercnvpy, which is a Python version of infercnv[73]. Based on the CNV score, we dichotomized normal or tumor cells.

M-A plots in Fig. 4i,j were based on differentially expressed genes or terms analysis using scvi-tools and drawn using ggplot2[74] on R environment[75].

Public resource of scRNA-seq datasets

In the integrative iSCseq analysis with *in vitro* cultured osteoclast scRNA-seq, we used four datasets GSM4419695 (post RANKL stimulation day 0), GSM4419699 (day 1), GSM4419703 (day 3) by Tsukasaki et al.[15], and published undeposited data (day 4) by Omata et al[16].

For building bone milieu (bone and bone marrow) atlas, we curated 28 scRNA-seq datasets using 10x chromium platform from 9 papers including wild-type or wild-type equivalent control cells; 30,400 cells of GSM3674224-9 by Baryawno et al.[29], 154 cells of GSM2915578 by Tikhonova et al.[34], 1,159 cells of GSM3466897 by Baccin et al.[28], 8,856 cells of GSM4318799-802 by Zhong et al.[36], 1,210 cells of GSM4677821 by Wang et al.[35], 164 cells of GSM4064321 and GSM4064323 by Matsushita et al.[32], 6,051 cells of GSM 4314099-101 by Johnston et al.[31], 3,792 cells of GSM6257588 by Mei et al.[33] and 2,557 cells of GSM6820990 by Hanai et al.[30]. We downloaded fastq files and processed them by STARsolo mapping to the same mm39 reference genome.

We used giant cell tumor of bone scRNA-seq datasets (GSM4492128-31) from 4 patients by Khazaei et al.[38], and osteosarcoma scRNA-seq datasets (GSM4952363-5 and GSM5155198-200) from 6 patients by Liu et al.[39]

Analysis of images

The obtained images were processed and reconstructed using Fiji[76]. Enhanced contrast was used to reduce background noise. High or low signal intensities were determined manually by comparison with other surrounding cells or components inside the same cell.

Statistics

One-way ANOVA and post-hoc Tukey tests were performed using scipy[77] to detect significant differences among the groups. If the p-value was less than 0.01, it was considered significant.

Data availability

Raw fastq files of iSCseq were deposited in DDBJ (DNA Data Bank of Japan). The accession numbers are DRA015536 and DRA016857.

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 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

2 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

3 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

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

5 Tsukasaki, M. & Takayanagi, H. Osteoimmunology: evolving concepts in bone-immune interactions in health and disease. *Nat Rev Immunol* **19**, 626-642 (2019). https://doi.org:10.1038/s41577-019-0178-8

6 Tanaka, S., Tanaka, Y., Ishiguro, N., Yamanaka, H. & Takeuchi, T. RANKL: A therapeutic target for bone destruction in rheumatoid arthritis. *Modern rheumatology* **28**, 9-16 (2018). https://doi.org:10.1080/14397595.2017.1369491

7 Moller, M. *et al.* The Swedish Fracture Register - ten years of experience and 600,000 fractures collected in a National Quality Register. *BMC Musculoskelet Disord* **23**, 141 (2022). https://doi.org:10.1186/s12891-022-05062-w

8 Ferrari, S. & Langdahl, B. Mechanisms underlying the long-term and withdrawal effects of denosumab therapy on bone. *Nat Rev Rheumatol* **19**, 307-317 (2023). https://doi.org:10.1038/s41584-023-00935-3

9 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

10 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

11 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

12 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

13 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

14 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

15 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

16 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

17 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
18 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
19 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
20 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
21 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
22 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
23 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
24 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
25 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
26 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
27 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
28 Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat Cell Biol* **22**, 38-48 (2020). https://doi.org:10.1038/s41556-019-0439-6
29 Baryawno, N. *et al.* A Cellular Taxonomy of the Bone Marrow Stroma in Homeostasis and Leukemia. *Cell* **177**, 1915-1932 e1916 (2019). https://doi.org:10.1016/j.cell.2019.04.040
30 Hanai, A. *et al.* Single-cell RNA sequencing identifies Fgf23-expressing osteocytes in response to 1,25-dihydroxyvitamin D(3) treatment. *Front Physiol* **14**, 1102751 (2023). https://doi.org:10.3389/fphys.2023.1102751
31 Johnston, G. *et al.* Nascent transcript and single-cell RNA-seq analysis defines the mechanism of action of the LSD1 inhibitor INCB059872 in myeloid leukemia. *Gene* **752**, 144758 (2020). https://doi.org:10.1016/j.gene.2020.144758
32 Matsushita, Y. *et al.* A Wnt-mediated transformation of the bone marrow stromal cell identity orchestrates skeletal regeneration. *Nat Commun* **11**, 332 (2020). https://doi.org:10.1038/s41467-019-14029-w
33 Mei, Y. *et al.* Bone marrow-confined IL-6 signaling mediates the progression of myelodysplastic syndromes to acute myeloid leukemia. *J Clin Invest* **132** (2022). https://doi.org:10.1172/JCI152673

34 Tikhonova, A. N. *et al.* The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222-228 (2019). https://doi.org:10.1038/s41586-019-1104-8
35 Wang, J. S. *et al.* Control of osteocyte dendrite formation by Sp7 and its target gene osteocrin. *Nature Communications* **12** (2021). https://doi.org:10.1038/s41467-021-26571-7
36 Zhong, L. *et al.* Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* **9** (2020). https://doi.org:10.7554/eLife.54695
37 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018). https://doi.org:10.1038/s41586-018-0414-6
38 Khazaei, S. *et al.* H3.3 G34W Promotes Growth and Impedes Differentiation of Osteoblast-Like Mesenchymal Progenitors in Giant Cell Tumor of Bone. *Cancer Discov* **10**, 1968-1987 (2020). https://doi.org:10.1158/2159-8290.CD-20-0461
39 Liu, Y. *et al.* Single-Cell Transcriptomics Reveals the Complexity of the Tumor Microenvironment of Treatment-Naive Osteosarcoma. *Front Oncol* **11**, 709210 (2021). https://doi.org:10.3389/fonc.2021.709210
40 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
41 Yagi, M. *et al.* DC-STAMP is essential for cell-cell fusion in osteoclasts and foreign body giant cells. *J Exp Med* **202**, 345-351 (2005). https://doi.org:10.1084/jem.20050645
42 Yang, M. *et al.* Osteoclast stimulatory transmembrane protein (OC-STAMP), a novel protein induced by RANKL that promotes osteoclast differentiation. *J Cell Physiol* **215**, 497-505 (2008). https://doi.org:10.1002/jcp.21331
43 Kikuta, J. *et al.* Dynamic visualization of RANKL and Th17-mediated osteoclast function. *J Clin Invest* **123**, 866-873 (2013). https://doi.org:10.1172/JCI65054
44 Furuya, M. *et al.* Direct cell-cell contact between mature osteoblasts and osteoclasts dynamically controls their functions in vivo. *Nat Commun* **9**, 300 (2018). https://doi.org:10.1038/s41467-017-02541-w
45 Hasegawa, T. *et al.* Identification of a novel arthritis-associated osteoclast precursor macrophage regulated by FoxM1. *Nat Immunol* **20**, 1631-1643 (2019). https://doi.org:10.1038/s41590-019-0526-7
46 Yasui, T. *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). https://doi.org:10.1002/jbmr.464
47 Izawa, N. *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). https://doi.org:10.1002/jbmr.3689
48 Koga, T. *et al.* Costimulatory signals mediated by the ITAM motif cooperate with RANKL for bone homeostasis. *Nature* **428**, 758-763 (2004). https://doi.org:10.1038/nature02444
49 Negishi-Koga, T. *et al.* Immune complexes regulate bone metabolism through FcRgamma signalling. *Nat Commun* **6**, 6637 (2015). https://doi.org:10.1038/ncomms7637
50 Terashima, A. *et al.* Sepsis-Induced Osteoblast Ablation Causes Immunodeficiency. *Immunity* **44**, 1434-1443 (2016). https://doi.org:10.1016/j.immuni.2016.05.012

51 Paloneva, J. *et al.* DAP12/TREM2 deficiency results in impaired osteoclast differentiation and osteoporotic features. *J Exp Med* **198**, 669-675 (2003). https://doi.org:10.1084/jem.20030027
52 Humphrey, M. B. *et al.* TREM2, a DAP12-associated receptor, regulates osteoclast differentiation and function. *J Bone Miner Res* **21**, 237-245 (2006). https://doi.org:10.1359/JBMR.051016
53 Kameda, Y. *et al.* Siglec-15 regulates osteoclast differentiation by modulating RANKL-induced phosphatidylinositol 3-kinase/Akt and Erk pathways in association with signaling Adaptor DAP12. *J Bone Miner Res* **28**, 2463-2475 (2013). https://doi.org:10.1002/jbmr.1989
54 Sato, D. *et al.* Siglec-15-targeting therapy protects against glucocorticoid-induced osteoporosis of growing skeleton in juvenile rats. *Bone* **135**, 115331 (2020). https://doi.org:10.1016/j.bone.2020.115331
55 Tsukazaki, H. *et al.* Anti-Siglec-15 antibody suppresses bone resorption by inhibiting osteoclast multinucleation without attenuating bone formation. *Bone* **152**, 116095 (2021). https://doi.org:10.1016/j.bone.2021.116095
56 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635
57 Frankish, A. *et al.* Gencode 2021. *Nucleic Acids Res* **49**, D916-D923 (2021). https://doi.org:10.1093/nar/gkaa1087
58 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
59 Nakato, R. *Scripts for NGS analysis*, <https://github.com/rnakato/script_rnakato> (2023).
60 Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual*. (CreateSpace, 2009).
61 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
62 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
63 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
64 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
65 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
66 Franzen, O., Gan, L. M. & Bjorkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019** (2019). https://doi.org:10.1093/database/baz046
67 Aibar, S. & Aerts, S. *AUCell: Analysis of 'gene set' activity in single-cell RNA-seq data.*, <https://bioconductor.org/packages/release/bioc/html/AUCell.html> (2016).
68 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

69 Subramanian, A. *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). https://doi.org:10.1073/pnas.0506580102
70 Castanza, A. S. *et al.* Extending support for mouse data in the Molecular Signatures Database (MSigDB). *Nat Methods* (2023). https://doi.org:10.1038/s41592-023-02014-7
71 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011). https://doi.org:10.1093/bioinformatics/btr260
72 Smedley, D. *et al.* BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009). https://doi.org:10.1186/1471-2164-10-22
73 inferCNV of the Trinity CTAT Project. *inferCNV*, <https://github.com/broadinstitute/inferCNV> (2020).
74 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2009).
75 R: A Language and Environment for Statistical Computing, v. 4.3.0, (Vienna, Austria,, 2023).
76 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012). https://doi.org:10.1038/nmeth.2019
77 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272 (2020). https://doi.org:10.1038/s41592-019-0686-2