Original Article

Dynamic Molecular Profiles of Bone Marrow-Derived Osteoblasts at the Single-Cell Level

(osteoblasts / molecular profiles / marrow / single-cell RNA sequence)

M. CAO[1], J. XIE[2], Y. HU[3], X. GAO[4], Z. XIE[5]

[1]Department of Joint Surgery, the Sixth Affiliated Hospital of Xinjiang Medical University, Xinjiang Province, China
[2]Department of Spinal Surgery, the Sixth Affiliated Hospital of Xinjiang Medical University, Xinjiang Province, China
[3]Department of Orthopaedics, the Second Hospital of Nanjing, Gulou District, Nanjing, China
[4]Department of Orthopaedics, Shenzhen City, Nanshan District, Guangdong Province, China
[5]Department of Orthopaedics Surgery, the First Affiliated Hospital of Xinjiang Medical University, Xinjiang Province, China

Abstract. Osteogenesis is an important process of bone metabolism, and abnormal osteogenesis leads to various skeletal system diseases. Osteoblasts, the main cells involved in bone formation, are central elements in the study of bone metabolic diseases. Single-cell RNA sequencing is an important tool for studying the transcriptome of cells and can help to elucidate various cellular and molecular functions at the single-cell level, providing new avenues for life science research. Here we explore the heterogeneity of osteoblasts and try to reveal the developmental trajectory of osteoblasts, thereby contributing to efforts to describe the mechanism of osteogenesis. In this study, single-cell sequencing data of murine bone marrow cells were used to identify osteoblasts. Finally, osteoblasts were divided into four groups, each differing in characteristic genes and signal pathways. We also identify clues of the changes of some genes in the process of osteoclast formation, providing directions for further study. Collectively, our findings suggest that bone marrow osteoblasts can be divided into several subgroups, which represent different stages of cells, and that the specific genes of each subgroup respond to the molecular mechanisms of cell development. This data will likely be of great help in resolving diseases of the skeletal system.

Introduction

Bone loss occurs more rapidly and tends to be more severe in older people; however, how aging affects bone loss and bone replacement is still not completely clear (Clarke and Khosla, 2010). A better understanding of intrinsic age-related changes in human bone cells may offer new approaches to mitigate or avoid bone loss. Bone marrow mesenchymal stem cells or bone marrow stromal cells (MSCs) have been shown to be precursors of several cell lineages, including osteoblasts, chondrocytes, myoblasts, adipocytes, and fibroblasts (Lin et al., 2019; Wolock et al., 2019). Research on the mechanism of osteogenesis and bone differentiation of MSCs at the single-cell level has attracted increasing attention. The single-cell RNA sequencing (scRNA-seq) data of this project were derived from bone marrow mesenchymal cells of mice of different ages (Zhong, 2020). Based on scRNA-seq data, the molecular information on osteoblast differentiation was mined, with the aim of better understanding the osteogenesis process.

In this study, we classified the types of cells derived from the mouse bone marrow and identified cell clusters with osteoblast characteristics. The key regulatory genes and signal pathways in these cell populations were identified by single-cell transcriptome analysis, and pseudo-time analysis was used to investigate the interactions and reciprocal regulation between key marker genes at different developmental stages. We also identified several marker genes that play important roles in the regulation of osteoblast development. These results will pro-

vide a new approach to clarify the regulatory mechanism of osteoblast differentiation.

Material and Methods

Single-cell RNA-seq dataset

The single-cell sequencing data of this study are derived from the mouse bone marrow mesenchymal cell project (accession number: GSE145477), which includes batches of single-cell libraries for sequencing 1-, 1.5-, 3-, and 16-month-old mice. Approximately 20,000 cells were loaded to obtain a single library of approximately 10,000 cells for each age group by the chromium controller (V3 chemistry version, 10 × genomics Inc., San Francisco, CA).

Single-cell RNA-seq data processing and statistical analysis

The FASTQ files were transformed into Cell Ranger-specific FASTQ files and processed individually with the reference mouse genome (mm10) as the reference genome. Doublets or cells with poor quality (genes > 6000, genes < 200, or > 10 % genes mapping to the mitochondrial genome) were excluded. All de-duplicated UMIs were processed using quality control, preprocessing, dimensionality reduction, and differential expression analysis. The results of unsupervised clustering were visualized using the t-SNE project with R (Seurat package version 3.0) (Stuart et al., 2019).

Identification of cluster cell types

To annotate cell clusters, we first identified the differentially expressed genes in each cluster using the PanglaoDB (Franzen et al., 2019) and CellMarker (Zhang, 2019) databases. The cell clusters were then annotated according to curated known cell markers. Those cell clusters that consistently expressed the same cell marker were merged. In addition, the SingleR algorithm was used to assist in the validation of identified cell types. Spearman's SingleR expressed per cell was calculated between the expression profile of each spectrum of the reference sample, and a correlation analysis of the reference dataset only gene variation (variable genes) was performed.

Gene set variation analysis of osteoblasts

Gene set variation analysis (GSVA) is a non-parametric unsupervised analysis method, mainly used to evaluate the result of gene set enrichment of the transcriptome (Hänzelmann, 2013). By converting the expression matrix of genes between different samples into the expression matrix of gene sets, we can evaluate whether different pathways are enriched in different samples. The iconic gene set is downloaded from the publicly available molecular marker database. The signal pathway activity of each cell was scored by the GSVA method following a standardized strategy. The path differ-

ence was calculated using the t-test for two groups, and $P < 0.05$ was considered significant.

Results

Overview of cell subsets identified by single-cell RNA sequencing in bone marrow mesenchymal cells

The number of qualified cells from 1-, 1.5-, 3-, and 16-month-old mice were 7842, 5656, 4337, and 8657, respectively. Unsupervised cluster analysis was performed on 26,510 cells by the graph clustering method in Seurat, and 13 cell clusters were obtained. Using the marker gene in PanglaoDB and CellMarker, we identified nine cell subtypes, including chondrocytes (n = 3653), erythrocytes (n = 1257), B cells (n = 675), neutrophils (n = 4624), osteoblasts (n = 3699, with marker genes *COL1A1*, *BGLAP*, *BGLAP2*, and *CAR3*), fibroblasts (n = 10277), macrophages (n = 1223), endothelial cells (n = 691), and smooth muscle cells (n = 393) (Fig. 1A, 1C, and 1D). By analysing highly expressed genes in each group, we found that *COL1A1* and *BGLAP* were highly expressed in osteoblasts; *Sox9* and *COL2A1* are chondrocyte-specific highly expressed genes, and other cell-specific highly expressed genes are shown in the results (Fig. 1B, 1E and 1F).

Osteoblast heterogeneity based on single-cell RNA sequencing

Transcripts from osteoblasts (cluster 2 and cluster 12) were subsequently analysed, resulting in 3699 single-cell transcriptomes that were analysed further. They were classified into four clusters (clusters 0–3) using t-distributed stochastic neighbourhood embedding (t-SNE) (Fig. 2A and 2B). We selected representative marker genes to identify each different cell type: for cluster 0, *COL1A1*, *CTHRC1*, *COL11A2*, *BGLAP*, and *BGLAP2* were highly expressed; for cluster 1, *SPP1*, *SERPINE2*, *MMP13*, *S100A4*, and *TIMP1* were highly expressed; for cluster 2, *LGF1*, *KCNK2* and *COL8A1* were highly expressed (Fig. 2C). *COL1A1*, *BGLAP* and *BGLAP2* were highly expressed in both cluster 0 and cluster 3. *PLPP1, PDPN*, *PHEX*, *PTPRZ1*, and *DMP1* were highly expressed in cluster 3 (Fig. 2C). These unbiased analyses revealed the diversity of osteoblast cells in the mouse bone marrow mesenchymal cells and reclassified the osteoblast cells into four distinct subsets.

Developmental pseudotime analysis of marker gene expression in osteoblasts

To reconstruct the developmental trajectory during differentiation, we performed pseudo-temporal ordering of cells (pseudotime) from our scRNA-seq data using Monocle 2 (Trapnell et al., 2014). In total, the pseudo-time path has three branches, and the different cell clusters can be arranged relatively clearly at different branch
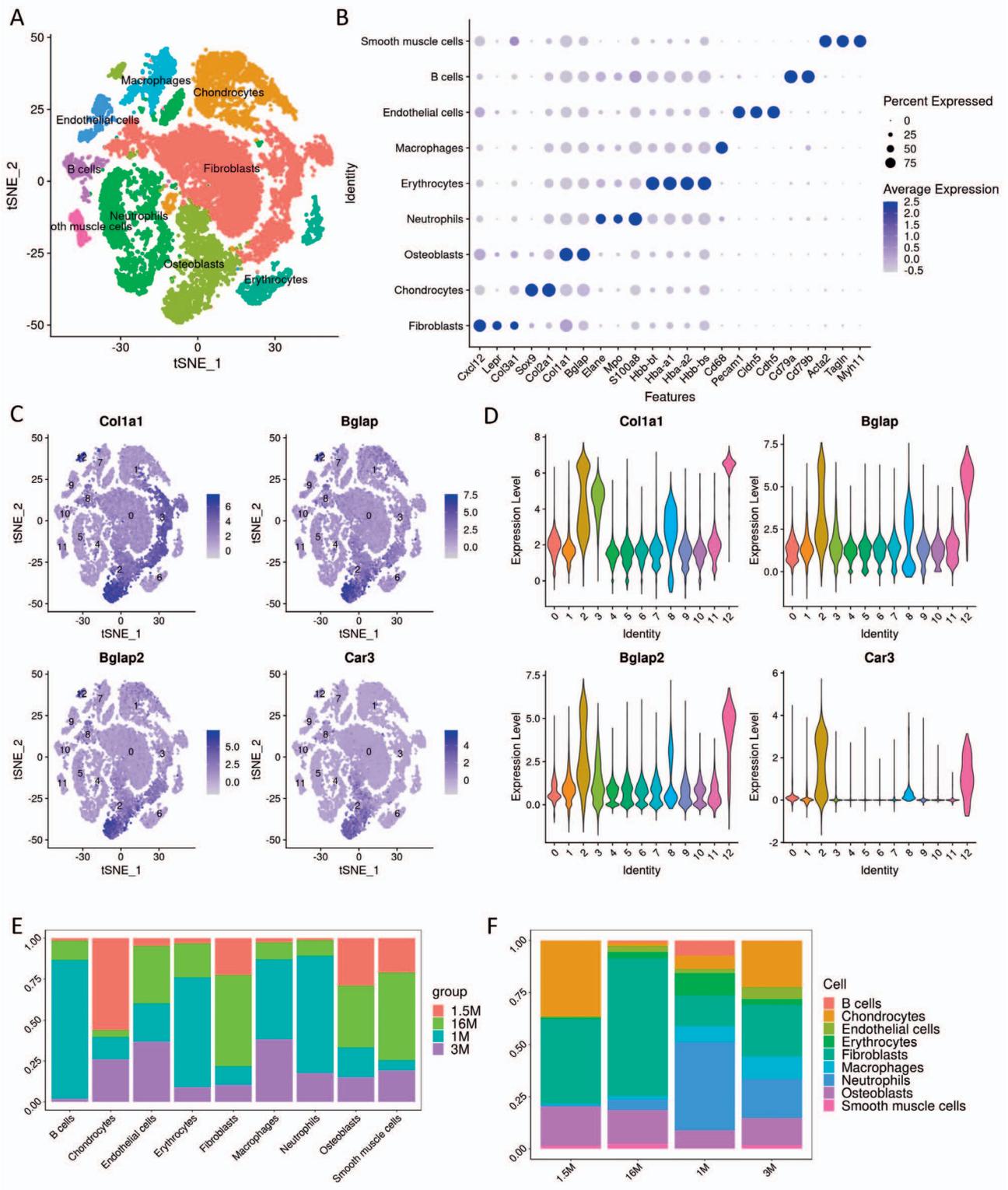
Fig. 1. Clustering and identification of bone marrow mesenchymal cells
(**A**) t-SNE map of bone marrow mesenchymal cell clusters derived from single-cell RNA sequencing. (**B**) The dot plot shows specific genes highly expressed in each cell type. (**C–D**) Characteristic gene expression map of osteoblasts. (**E**) Composition diagram of the sample source of identified bone marrow mesenchymal cell types. (**F**) Diagram of cell composition of different samples.
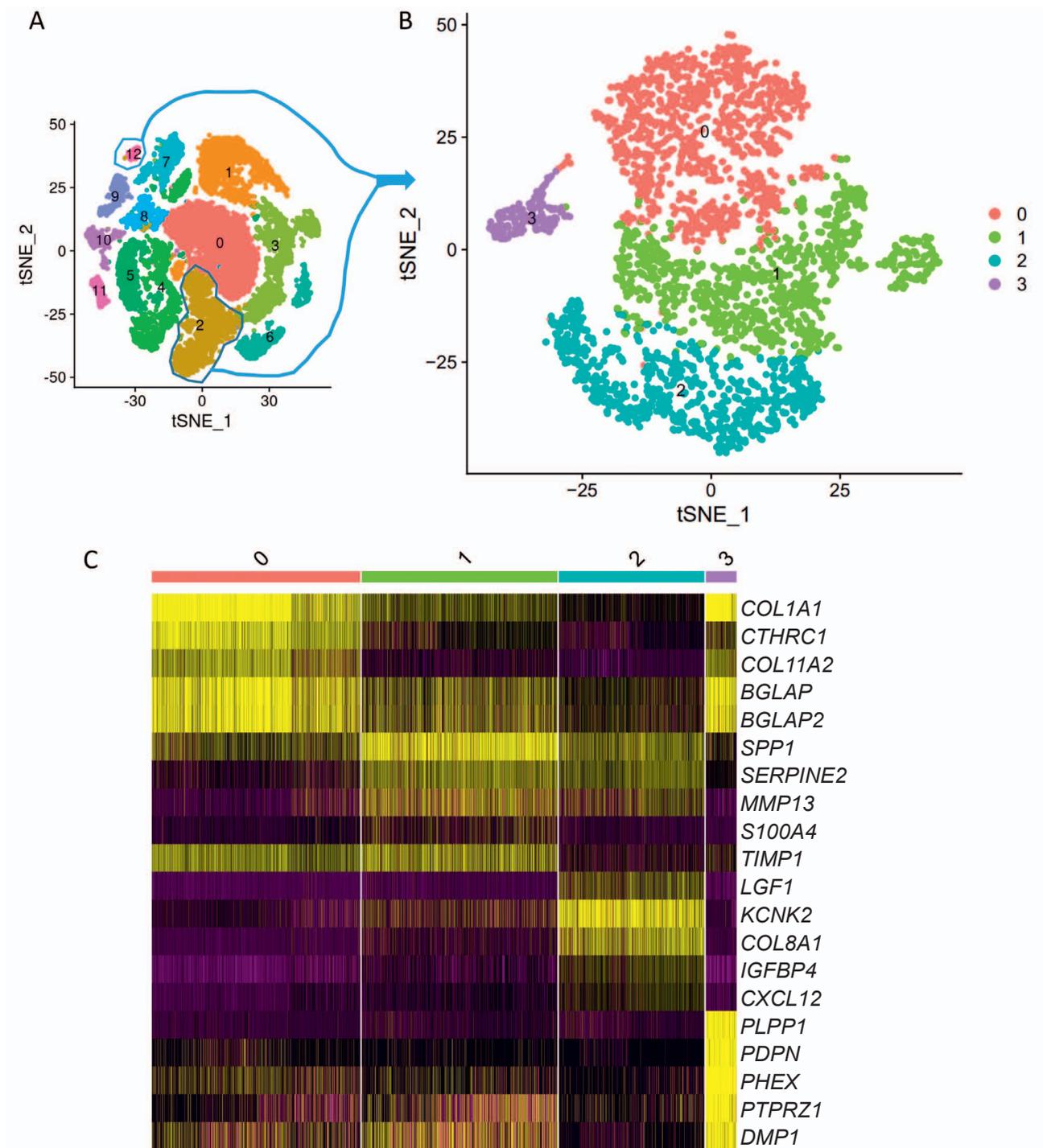
Fig. 2. Transcriptome characteristics of osteoblasts at the single-cell level
(**A-B**) Cluster 2 and cluster 12 were identified as osteoblasts, which could be divided into four groups by cluster analysis.
(**C**) Characteristic expressed genes in the four subsets of osteoblasts.

sites of the pseudotime path (Fig. 3A). In general, the different developmental processes of osteoblast cells can be seen from cluster 2 to cluster 0 (Fig. 3B). Cluster 1 is concentrated in an early branch in the pseudotime path, which is clearly distinct from other cell types. In principle, the profile of the distribution of different cell types along a developmental trajectory allows the initial determination of the relationship of these cells across developmental stages. We found that cluster 1 vs cluster 2 and cluster 0 vs cluster 1 could not be clearly distinguished on the pseudotime path, indicating that the cellular developmental stages of osteoblasts are more complex. To investigate the pseudotime path of genes in each cluster, we performed heatmap analysis of all highly
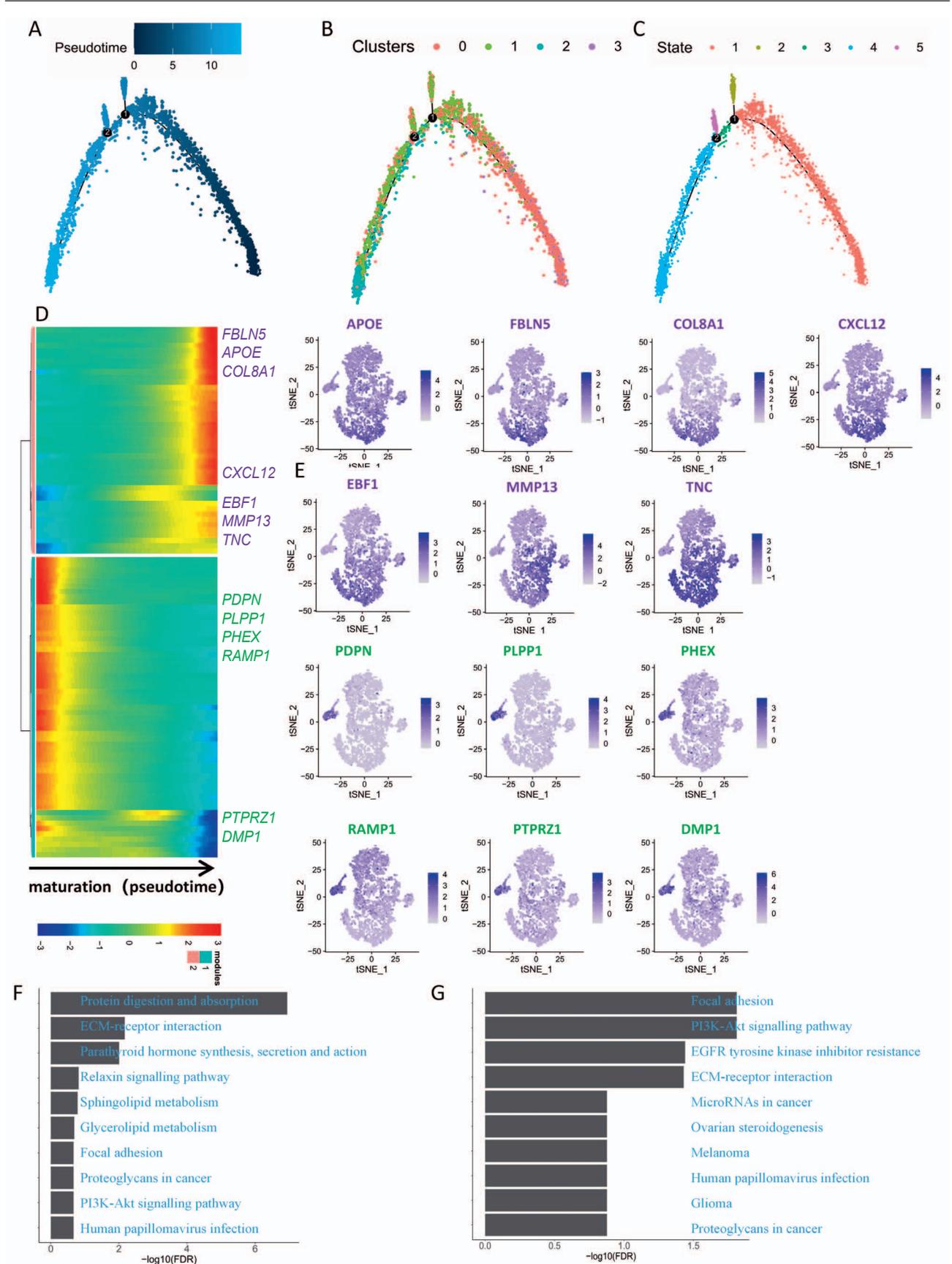
Fig. 3. Pseudotime analysis in osteoblasts
(**A-C**) Monocle pseudotime trajectory of osteoblasts between different samples. (**D-E**) Expression heatmap showing genes with the q values < 0.05; the pseudo-temporal patterns could be divided into two classes; the t-SNE map of gene expression is also displayed. (**F**) KEGG pathway enrichment analysis of the first group from pseudotime analysis in osteoblasts. (**G**) KEGG pathway enrichment analysis of the second group from pseudotime analysis in osteoblasts.

expressed genes, for which the pseudo-temporal patterns could be divided into two groups (Fig. 3D). In the first group, the expression of all marker genes gradually increased with the pseudotime analysis. In contrast, the expression of marker genes in group 2 was reduced at the end of the counter timing analysis axis. In the first group, marker genes were mainly from cluster 2 and included *FBLN5*, *APOE*, *COL8A1*, *CXCL12*, *EBF1*, *MMP13*, and *TNC*, indicating that cluster 2 was at a more mature stage of development (Fig. 3D). KEGG analysis showed that the genes in the first group were mainly enriched in protein digestion and absorption: ECM receptor interaction, parathyroid hormone synthesis, secretion and action, relaxin signalling pathway, sphingolipid metabolism, glycerolipid metabolism, focal adhesion, proteoglycans in cancer, PI3K Akt signalling pathway, and human papillomavirus infection (Fig. 3F). In the second group, marker genes were mainly from cluster 3, including *PDPN*, *PLPP1*, *PHEX*, *RAMP1*, *PTPRZ1*, and *DMP1*, indicating that cluster 3 is at an earlier developmental stage (Fig. 3D). KEGG analysis

showed that the genes in the second group were mainly enriched in focal adhesion, PI3K Akt signalling pathway, EGFR tyrosine kinase inhibitor resistance, ECM receptor interaction, microRNAs in cancer, ovarian steroidogenesis, melanoma, human papillomavirus infection, glioma, and proteoglycans in cancer (Fig. 3G).

Pathway analysis of the genes enriched in differentially aged cells

To investigate the potential biological functions of genes for osteoblasts in the bone marrow of mice at different ages, we performed pathway enrichment analysis of genes from samples at four ages (Fig. 4). There were significant differences in the signalling pathways of the enriched genes identified in cell types at different ages. We found that the sample derived at one month was especially different from the other three months, and was highly enriched in signalling pathways including sphingolipid metabolism, mammalian circadian rhythm, glycosphingolipid biosynthesis lacto and neolacto series,
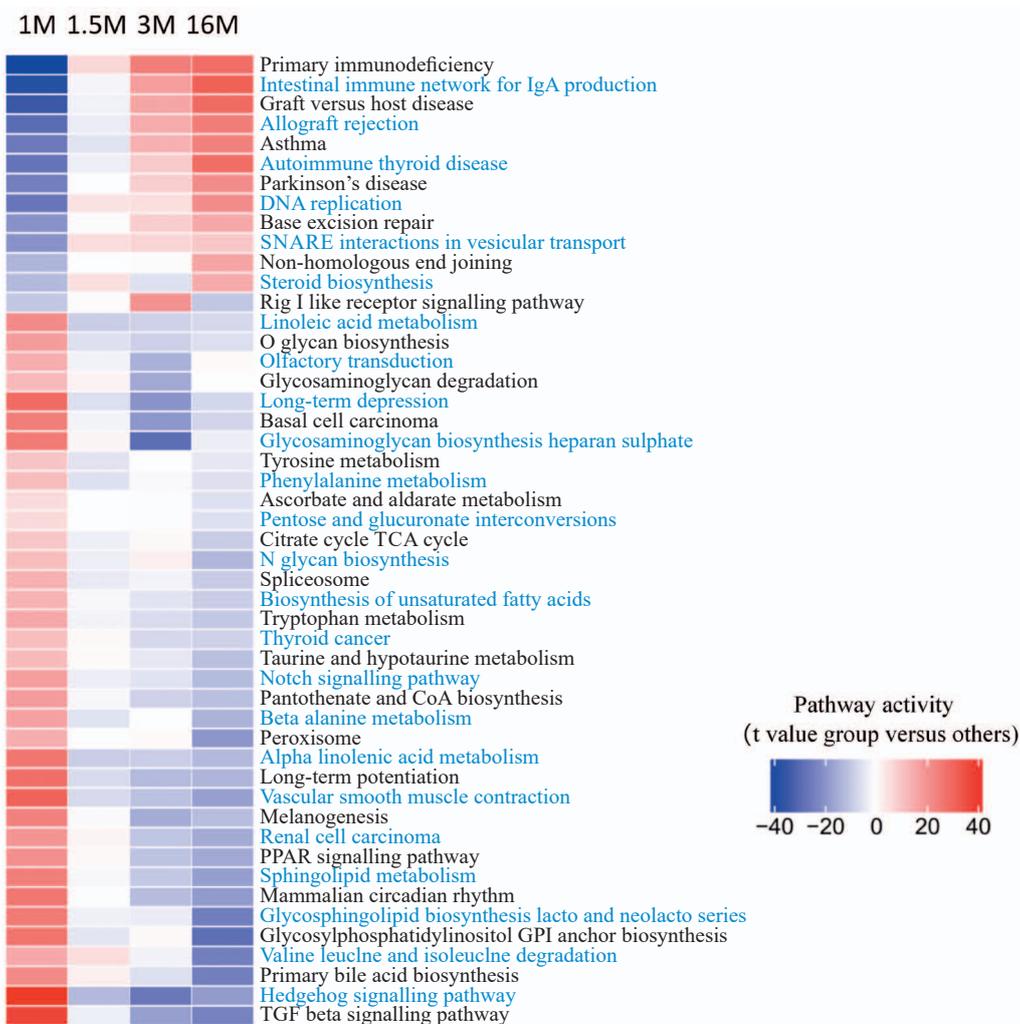

Fig. 4. Heatmap showing the difference in KEGG pathway activities scored by GSVA per cell between different mice. Red – high abundance, blue – low abundance. The colour scale is row (pathway) normalized.

glycosylphosphatidylinositol GPI anchor biosynthesis, valine, leucine and isoleucine degradation, primary bile acid biosynthesis, hedgehog signalling pathway, and TGF beta signalling pathway, whereas these signalling pathways were mostly hypoactive in the 3- month and 16-month samples. Signalling pathways such as primary immunodeficiency, intestinal immune network for IgA production, graft versus host disease, allograft rejection, autoimmune thyroid disease, DNA replication, base exclusion repair, and SNARE interactions in vesicular transport were mostly highly active in the 3- and 16-month samples, but were hypoactive in 1-month samples. These results give us the information for studying the differentiation of osteoblasts and also imply that we must fully consider the influence of age of the animals on the experimental results in animal experiments.

Discussion

In this study, we analysed the heterogeneity of osteoblasts at the single-cell level in mice. Osteoblasts, as a group of cells that play an important role in bone metabolism, undergo a dynamic process from early differentiation to subsequent senescence (Dirckx et al., 2019). We screened osteoblasts from other cells in the bone marrow based on the marker genes *COL1A1*, *BGLAP*, *BGLAP2*, and *CAR3* (Takarada et al., 2017; Yan, 2019; Deng et al., 2020; Wang et al., 2021). *COL1A1*, the gene encoding the pro-alpha1 chain of type I collagen, is highly expressed in cells of the bone, cornea, and dermis (Takarada et al., 2017). Mutations or other abnormalities are associated with osteogenesis imperfecta, classical Ehlers-Danlos syndrome, Caffey disease, and idiopathic osteoporosis (Tanner et al., 2017). BGLAP, as well as BGLAP2, a highly abundant bone protein secreted by osteoblasts that regulates bone remodelling and energy metabolism, functions to bind to calcium and hydroxyapatite, the mineral components of the bone (Liu et al., 2020). CAR3 is highly expressed in osteoblasts and can catalyse the reversible hydration of carbon dioxide to bicarbonate, with the potential to affect mineralization and osteoclast function by regulating pH in the local environment (Staines et al., 2014). In this study, *COL1A1*, *BGLAP*, *BGLAP2*, and *CAR3* were used as marker genes to identify two cell clusters as osteoblasts, or more strictly described, osteoblast-like cells, or cells in the process of transforming into osteoblasts.

We performed clustering analysis of the osteoblasts derived from mice at different ages and showed that they could be classified into four categories, with the highest number of cells in cluster 0: *COL1A1*, *CTHRC1*, *COL11A2*, *BGLAP*, and *BGLAP2* were highly expressed in cluster 0, among which *COL1A1*, *BGLAP* and *BGLAP2* were also highly expressed in cluster 3, but their expression was relatively lower in cluster 1 than in cluster 2. In cluster 3, *PLPP1*, *PDPN*, *PHEX*, *PTPRZ1*, and *DMP1* were specifically and highly expressed. The results of the pseudotime analysis divided cells into two cell groups, where the expression of genes in the first cell group was gradually enhanced, while that of the genes in the second cell group was gradually weakening. Combined with the results of pseudotime analysis (Fig. 3), we speculate that the development of osteoblasts is from cluster 2 to cluster 1, to cluster 3, to cluster 0. In the process of osteoblast maturation, the expression of *APOE*, *COL8A1*, *CXCL12*, *MMP13*, and other genes increases gradually, while the expression of *PLPP1*, *PHEX*, *RAMP1*, *PTPRZ1*, and *DMP1* and other genes is relatively weakened. Previous studies have found that circulating apoE levels increase with age in both humans and mice (Huynh et al., 2017). Experiments have demonstrated that circulating apoE plays a powerful inhibitory role in bone repair, and fracture calluses of aged mice show enhanced bone deposition and mechanical strength when circulating apoE levels are reduced (Huang et al., 2019). These results suggest that circulating apoE can inhibit fracture healing by altering the metabolism of osteoblasts, thereby identifying apoE as a new therapeutic target for improving bone repair in the elderly (Huang et al., 2019). DMP1 is highly expressed in bones and plays a key role in mineralization and phosphate metabolism (Martin, 2019). An important role of DMP1 expression in the nucleus of osteoblast precursor cells is to up-regulate osteoblast-specific genes (Shahabipour, 2020). Expression of most of the genes screened in this study is consistent with that reported in the literature, and they can be used as candidate genes for studying osteogenesis.

We analysed samples from different sources by GSVA, which revealed active signalling pathways in 1-month-old mice, including the TGF beta signalling pathway, which is involved in the process of osteogenesis but is not an osteogenesis-specific signalling pathway. Other signal pathways are mostly related to the cell metabolism, including lipid and glucose metabolism, but not directly related to bone formation. The active signal pathways in 16-month-old and 3-month-old mice include primary immunodeficiency, intestinal immune network for IgA production, and graft versus host disease. Although it is impossible to infer the direct relationship between these signalling pathways and osteogenesis, the comparison shows that the cells from 1-month-old mice are in the early stage of development and growth, while the cells from 3-month-old and 16-month-old mice are in a working state. The data are consistent with the characteristics of mature cells.

The innovation of this study lies in analysing the heterogeneity of osteoblasts derived from the bone marrow with the help of single-cell sequencing data, which has greater accuracy than traditional bulk sequencing. This study makes a preliminary analysis of the developmental process of osteoblasts, which can provide clues for further research. A limitation is that only data analysis has been carried out; although many interesting and meaningful genes have been proposed, no cellular and animal experiments were done.

In conclusion, this study performed transcriptome analysis of bone marrow-derived osteoblasts at the sin-

gle-cell level, identified four subtypes of constituent bone cells, and deduced the developmental order of the cells, along with the dynamic changes of related genes according to the pseudotime analysis. The present study lays the groundwork for next mechanistic investigation, in which we plan to identify the genes potentially involved in osteoblast development.

Ethics approval and consent to participate

All data in this study were based on already published studies and data. Therefore, no hospital or animal protection-related ethics approval or patient consent was required.

Authors' contributions

MYC conceived and designed the research. XJ searched and collected literature. YBH analysed the data. GXL and ZRX drafted the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

Clarke, B. L., Khosla, S. (2010) Physiology of bone loss. *Radiol. Clin. North Am.* **48**, 483-495.

Deng, Z. H., Ng, C., Inoue, K., Chen, Z. Y., Xia, Y. H., Hu, X. Y., Greenblatt, M., Pernis, A, Zhao, B. H. (2020) Def6 regulates endogenous type-I interferon responses in osteoblasts and suppresses osteogenesis. *Elife* **9**, e59659.

Dirckx, N., Moorer, M. C., Clemens, T. L., Riddle, R. C. (2019) The role of osteoblasts in energy homeostasis. *Nat. Rev. Endocrinol.* **15**, 651-665.

Franzen, O., Gan, L. M., Bjorkegren, J. L. M. (2019) PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, baz046.

Hänzelmann, S., Castelo, R., Guinney, J. (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 1-15.

Huang, R., Zong, X., Nadesan, P., Huebner, J. L., Kraus, V. B., White, J. P., White, P. J., Baht, G. S. (2019) Lowering circulating apolipoprotein E levels improves aged bone fracture healing. *JCI Insight* **4**, e129144.

Huynh, T. P. V., Liao, F., Francis, C. M., Robinson, G. O., Serrano, J. R., Jiang, H., Roh, J., Finn, M. B., Sullivan, P. M., Esparza, T. J., Stewart F. R., Mahan, T. E., Ulrich, J. D., Cole, T., Holtzman, D. M. (2017) Age-dependent effects of apoE reduction using antisense oligonucleotides in a model of β-amyloidosis. *Neuron* **96**, 1013-1023.

Lin, H., Sohn, J., Shen, H., Langhans, M. T., Tuan, R. S. (2019) Bone marrow mesenchymal stem cells: aging and tissue engineering applications to enhance bone healing. *Biomaterials* **203**, 96-110.

Liu, X., Liu, Y., Mathers, J., Cameron, M., Levinger, I., Yeap, B. B., Lewis, J. R., Brock, K. E., Brennan-Speranza, T. C. (2020) Osteocalcin and measures of adiposity: a systematic review and meta-analysis of observational studies. *Arch. Osteoporos.* **15**, 145.

Martin, A. (2019) Bone and heart health in chronic kidney disease: role of dentin matrix protein 1. *Curr. Opin. Nephrol. Hypertens.* **28**, 297-303.

Shahabipour, F., Oskuee, R. K., Shokrgozar, M. A., Naderi-Meshkin, H., Goshayeshi, L., Bonakdar, S. (2020) CRISPR/Cas9 mediated GFP-human dentin matrix protein 1 (DMP1) promoter knock-in at the ROSA26 locus in mesenchymal stem cell for monitoring osteoblast differentiation. *J. Gene Med.* **22**, e3288.

Staines, K. A., Zhu, D., Farquharson, C., MacRae, V. E. (2014) Identification of novel regulators of osteoblast matrix mineralization by time series transcriptional profiling. *J. Bone Miner. Metab.* **32**, 240-251.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M. 3rd, Hao, Y., Stoeckius, M., Smibert, P., Satija, R. (2019) Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e1821.

Takarada, T., Xu, C., Ochi, H., Nakazato, R., Yamada, D., Nakamura, S., Kodama, A., Shimba, S., Mieda, M., Fukasawa, K., Ozaki, K., Iezaki, T., Fujikawa, K., Yoneda, Y., Numano, R., Hida, A., Tei, H., Takeda, S., Hinoi, E. (2017) Bone resorption is regulated by circadian clock in osteoblasts. *J. Bone Miner. Res.* **32**, 872-881.

Tanner, L., Vainio, P., Sandell, M., Laine, J. (2017) Novel COL1A1 mutation c. 3290G > T associated with severe form of osteogenesis imperfecta in a fetus. *Pediatr. Dev. Pathol.* **20**, 455-459.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., Rinn, J. L. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381-386.

Wang, L., Yang, H., Huang, J., Pei, S., Wang, L., Feng, J. Q., Jing, D., Zhao, H., Kronenberg, H. M., Moore, D. C., Yang, W. (2021) Targeted Ptpn11 deletion in mice reveals the essential role of SHP2 in osteoblast differentiation and skeletal homeostasis. *Bone Res.* **9**, 6.

Wolock, S. L., Krishnan, I., Tenen, D. E., Matkins, V., Camacho, V., Patel, S., Agarwal, P., Bhatia, R., Tenen, D. G., Klein, A. M., Welner, R. S. (2019) Mapping distinct bone marrow niche populations and their differentiation paths. *Cell Rep.* **28**, 302-311.e305.

Yan, B., Liu, D., Zhu, J., Pang, X. (2019) The effects of hyperuricemia on the differentiation and proliferation of osteoblasts and vascular smooth muscle cells are implicated in the elevated risk of osteopenia and vascular calcification in gout: an in vivo and in vitro analysis. *J. Cell. Biochem.* **120**, 19660-19672.

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., Ping, Y., Li, F., Shi, A., Bai, J., Zhao, T., Li, X., Xiao, Y. (2019) CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721-D728.

Zhong, L., Yao, L., Tower, R. J., Wei, Y., Miao, Z., Park, J., Shrestha, R., Wang, L., Yu, W., Holdreith, N., Huang, X., Zhang, Y., Tong, W., Gong, Y., Ahn, J., Susztak, K., Dyment, N., Li, M., Long, F., Chen, C., Seale, P., Qin, L. (2020) Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* **9**, e54695.