

**1 Title: Integrated transcriptomic analysis of human induced pluripotent stem cell-
2 derived osteogenic differentiation reveals a regulatory role of KLF16**

3
4 Ying Ru¹, Meng Ma², Xianxiao Zhou^{1,3,4}, Divya Kriti^{1,5}, Ninette Cohen^{1,6}, Sunita
5 D'Souza^{7,8}, Christoph Schaniel^{7,9,10,11}, Susan M. Motch Perrine¹², Sharon Kuo^{13,14},
6 Oksana Pichurin¹⁵, Dalila Pinto¹, Genevieve Housman^{16,17}, Greg Holmes¹, Eric
7 Schadt^{1,4}, Harm van Bakel^{1,4,18}, Bin Zhang^{1,3,4}, Ethylin Wang Jabs^{1,7,15,19}, Meng Wu^{1,15,19}

8
9 ¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount
10 Sinai, New York, NY, 10029, USA.

11 ²Mount Sinai Genomics, Sema4, Stamford, CT, 06902, USA.

12 ³Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at
13 Mount Sinai, New York, NY, 10029, USA.

14 ⁴Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY,
15 10029, USA.

16 ⁵Present address: Department of Biochemistry and Molecular Biology, Faculty of
17 Medicine, The University of British Columbia, Vancouver, BC V6T 2G3, Canada.

18 ⁶Present address: Division of Cytogenetics and Molecular Pathology, Zucker School of
19 Medicine at Hofstra/Northwell, Northwell Health Laboratories, Lake Success, NY,
20 11030, USA.

21 ⁷Department of Cell, Developmental and Regenerative Biology, Icahn School of
22 Medicine at Mount Sinai, New York, NY, 10029, USA.

23 ⁸Present address: St Jude Children's Research Hospital, Memphis, TN, 38105, USA.

24 ⁹Department of Medicine, Division of Hematology and Medical Oncology, Tisch Cancer

25 Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

26 ¹⁰Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York,

27 NY, 10029, USA.

28 ¹¹Mount Sinai Institute for Systems Biomedicine, Icahn School of Medicine at Mount

29 Sinai, New York, NY, 10029, USA.

30 ¹²Department of Anthropology, Pennsylvania State University, University Park, PA,

31 16802, USA.

32 ¹³Department of Biomedical Sciences, University of Minnesota, Duluth, MN, 55812,

33 USA.

34 ¹⁴Technological Primates Research Group, Max Planck Institute for Evolutionary

35 Anthropology, Leipzig, 04103, Germany.

36 ¹⁵Department of Clinical Genomics, Mayo Clinic, Rochester, MN, 55905, USA.

37 ¹⁶Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago,

38 IL, 60637, USA.

39 ¹⁷Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary

40 Anthropology, Leipzig, 04103, Germany.

41 ¹⁸Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY,

42 10029, USA.

43 ¹⁹Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN,

44 55905, USA.

45

46 **Keywords:** bone, induced pluripotent stem cell, mesenchymal stem cell, osteoblast,

47 RNA sequencing, differential gene expression, transcription factor, single-cell RNA-seq,

48 network analysis, systems biology

49

50 Abstract

51 Osteogenic differentiation is essential for bone development, metabolism, and repair;
52 however, the underlying regulatory relationships among genes remain poorly
53 understood. To elucidate the transcriptomic changes and identify novel regulatory
54 genes involved in osteogenic differentiation, we differentiated mesenchymal stem cells
55 (MSCs) derived from 20 human iPSC lines into preosteoblasts (preOBs) and
56 osteoblasts (OBs). We then performed transcriptome profiling of MSCs, preOBs and
57 OBs. The iPSC-derived MSCs and OBs showed similar transcriptome profiles to those
58 of primary human MSCs and OBs, respectively. Differential gene expression analysis
59 revealed global changes in the transcriptomes from MSCs to preOBs, and then to OBs,
60 including the differential expression of 840 genes encoding transcription factors (TFs).
61 TF regulatory network analysis uncovered a network comprising 451 TFs, organized
62 into five interactive modules. Multiscale embedded gene co-expression network
63 analysis (MEGENA) identified gene co-expression modules and key network regulators
64 (KNRs). From these analyses, *KLF16* emerged as an important TF in osteogenic
65 differentiation. We demonstrate that overexpression of *Klf16* in vitro inhibited osteogenic
66 differentiation and mineralization, while *Klf16*^{+/-} mice exhibited increased bone mineral
67 density, trabecular number, and cortical bone area. Our study underscores the
68 complexity of osteogenic differentiation and identifies novel regulatory genes such as
69 *KLF16*, which plays an inhibitory role in osteogenic differentiation both in vitro and in
70 vivo.

71

72 Introduction

73 Mesenchymal stem cells (MSCs) are multipotent cells that can differentiate into a
74 variety of lineages, including osteoblasts (OBs), chondrocytes, and adipocytes
75 (Pittenger et al., 1999). The multilineage potential, self-renewal capacity, and immune-
76 modulation functions of MSCs have made them a promising therapeutic tool for cell
77 therapies and regenerative medicine (Bruder et al., 1994; Han et al., 2019). Osteogenic
78 differentiation from MSCs plays a pivotal role in bone development, homeostasis, and
79 repair. This process replenishes the pool of osteoblasts and ensures that there are
80 enough cells available for bone formation and repair. It is highly regulated by multiple
81 mechanisms and previous research has highlighted the roles of various genes,
82 including those coding for transcription factors (TFs) and signaling pathways such as
83 TGF- β /BMP, Notch, Hedgehog, Wnt, Hippo, and estrogen receptor signaling (Thomas
84 and Jaganathan, 2022). Precise and robust transcriptional regulation required during
85 osteogenic differentiation is achieved by networks of transcriptional regulators. For
86 example, *RUNX2* is a master transcription factor in early osteogenic differentiation
87 (Komori et al., 1997; Schroeder et al., 2005). TFs *SP7*, *ATF4*, *TEAD4*, and *KLF4* work
88 alongside *RUNX2* to regulate osteogenesis (Nakashima et al., 2002; Suo et al., 2020;
89 Yang et al., 2004; Yu et al., 2021). The broader TF interplay in osteogenic differentiation
90 remains largely uncharted, even with insights from ENCODE (Encyclopedia of DNA
91 Elements) data that reveal the co-association of human TFs in a combinatorial and
92 context-specific manner (Gerstein et al., 2012), revealing a need for further investigation.

93 Transcriptome profiling can uncover the key genes essential for bone formation
94 and growth, offering valuable insights into bone development and the underlying

95 mechanisms of bone-related diseases. However, due to the invasiveness required to
96 obtain human primary MSCs and the suboptimal conditions for preserving available
97 samples, large-scale human transcriptomics projects, such as GTEx, do not include
98 data from bone and cartilage. Furthermore, studies focusing on these tissues are often
99 limited to disease contexts. Induced pluripotent stem cell (iPSC)-derived systems, with
100 their pluripotency and differentiation capabilities, offer an avenue for studying skeletal
101 development. iPSC-derived sclerotome models have been utilized to identify gene
102 expression signatures and regulatory mechanisms crucial to key developmental stages
103 of endochondral ossification (Lamande et al., 2023; Nakajima et al., 2018; Tani et al.,
104 2023). Previous studies for osteogenic differentiation from iPSC-derived MSCs included
105 only one to two MSC cell lines (Matsuda et al., 2020; Rauch et al., 2019), limiting their
106 power for comprehensive transcriptomic analyses. In this study, 20 human iPSC lines,
107 each from a healthy individual, were used to investigate osteogenic differentiation of
108 MSCs by RNA sequencing (RNA-seq) at three key stages: MSC, preosteoblast (preOB),
109 and osteoblast (OB). Our findings highlight dynamic gene expression changes during
110 osteogenic differentiation and identify the regulatory role of TF KLF16 as an inhibitor of
111 osteogenic differentiation. This work provides insights into osteogenic differentiation and
112 sheds light on novel therapeutic targets for bone diseases.

113

114 **Results**

115 **Transcriptome profiles of human iPSC-derived MSCs and OBs are similar to**
116 **those of primary MSCs and OBs**

117 Human iPSCs were generated from peripheral blood mononuclear cells (PBMCs) or
118 skin fibroblasts of 20 healthy individuals (9 males and 11 females) and differentiated
119 into MSC, preOB, and OB stages (Figure 1A; Supplementary Figure 1A, 1B, and 1H;
120 Supplementary Table 1). To abrogate the epigenetic memory from their source cell
121 types and minimize in vitro transcriptional and differentiation heterogeneity, the
122 established iPSC lines were cultured for at least 16 passages and underwent quality
123 control (Polo et al., 2010). The iPSCs were then differentiated into MSCs and sorted for
124 the CD105+/CD45- cell population (Giuliani et al., 2011; Kang et al., 2015)
125 (Supplementary Figure 1C). No significant differences were found in the percentage of
126 this cell population between PBMC and fibroblast origins (Supplementary Figure 1D).
127 After two weeks of expansion, the iPSC-derived MSCs were confirmed to exhibit
128 mesenchymal characteristics, evidenced by strong expression of positive MSC surface
129 markers (CD29, CD73, CD90, and CD105) and minimal expression of negative markers
130 (CD31, CD34, and CD45) at both the protein level by flow cytometry (Supplementary
131 Figure 1E and 1F) and mRNA level by RNA-seq (Supplementary Figure 1G). The MSCs
132 were cultured in osteogenic differentiation medium for 7 days to generate preOBs and
133 for 21 days to produce OBs. After 21 days of osteogenic differentiation, the OBs
134 showed alkaline phosphatase (ALP) enzyme activity and mineralization, indicating the
135 state of functional OBs (Figure 1B).

136 MSCs, preOBs, and OBs derived from the 20 individuals were harvested and a
137 total of 60 RNA-seq libraries were generated (Figure 1A; Supplementary Table 1). We
138 initially employed bulk RNA-seq rather than single-cell RNA-seq (scRNA-seq) because
139 the latter recovers only sparse transcriptional data from individual cells, thus limiting

140 comprehensive expression analysis and the development of robust prediction models
141 for regulatory and gene co-expression networks (Li and Wang, 2021). RNA-seq data
142 revealed increased expression of OB lineage markers (*ALPL*, *COL1A1*, *RUNX2*,
143 *SPARC*, *OMD*, and *OGN*) with characteristic temporal expression patterns during
144 osteogenic differentiation (Figure 1C). We further compared our MSC and OB RNA-seq
145 datasets with previously published human primary MSC and OB datasets, human iPSC
146 datasets, and GTEx datasets for different tissues derived from the three germ layers
147 (ectoderm, endoderm, and mesoderm) and germ cells (Al-Rekabi et al., 2016; Ardlie et
148 al., 2015; Ma et al., 2019; Roforth et al., 2015; Rojas-Peña et al., 2014). By principal
149 component analysis (PCA), our MSC and OB datasets clustered with those of published
150 primary MSCs and OBs, respectively, and were discrete from iPSCs and other tissue
151 datasets, further validating the cell identities of our MSCs and OBs (Figure 1D).

152

153 **Differential transcription profiles of human iPSC-derived MSCs, preOBs, and OBs**
154 During in vitro osteogenic differentiation, we detected the expression of 17,795 unique
155 genes across all stages, of which 79% were protein-coding and 21% were noncoding
156 (Figure 2A). Seventy percent or 9,724 of the protein-coding genes were differentially
157 expressed (DE) between the osteogenic stages (MSC to preOB, or preOB to OB) (fold
158 change ≥ 1.2 and adjusted p value ≤ 0.05). Sixty percent (2,297) of the noncoding
159 genes also were DE (Figure 2A), of which 2,249 (98% of DE noncoding) were long
160 noncoding RNA (lncRNA) genes. By intersecting with the human TF repertoire including
161 over 1,600 known or likely human TFs (Lambert et al., 2018), we found that nine
162 percent (840 genes) of the DE coding genes were TFs (Figure 2B). A chi-square test

163 revealed a significant correlation between TFs and differentially expressed genes
164 (DEGs) ($\chi^2 = 27.4$, p value < 0.0001). There were 5,715 up-regulated genes and 4,409
165 down-regulated genes from the MSC to the preOB stage, and 3,342 up-regulated genes
166 and 2,835 down-regulated genes from the preOB to the OB stage, revealing that the
167 majority of the DEGs were up-regulated during osteogenic differentiation (Figure 2C-E;
168 Supplementary Table 2). Among the top statistically significant DEGs were known
169 osteogenesis-associated genes, including a SMAD ubiquitination regulatory factor
170 *SMURF2* (Kushioka et al., 2020), natriuretic peptide precursor B *NPPB* (Aza-Carmona
171 et al., 2014), hypoxia-inducible factor 3 subunit alpha *HIF3A* (Zhu et al., 2014), a TF
172 *ZBTB16* (Onizuka et al., 2016), a Wnt signaling pathway family member *WNT7B* (Yu et
173 al., 2020), osteoinductive factor *OGN* (Kukita et al., 1990), a SFRP family (modulators
174 of Wnt signaling) member *SFRP2* (Yang et al., 2020), and a glycoprotein member of the
175 glycosyl hydrolase 18 family *CHI3L* (Chen et al., 2017). There were also many genes
176 with no defined function in bone, including those that play critical roles in RNA binding
177 activity (*EBNA1BP2*), maintenance of protein homeostasis (*PSMD2*), arachidonic acid
178 metabolism (*PRXL2B*), hemostasis and antimicrobial host defense (*FGB*), embryonic
179 and induced pluripotent stem cells (*PCSK2*), as well as cell proliferation, differentiation,
180 and migration (*FGFBP1*). A greater number of DEGs with larger changes in gene
181 expression were identified between the MSC and preOB stages compared to the
182 changes between the preOB and OB stages, even though the differentiation interval
183 was half as long between the MSC and preOB stages (one week) as it was between the
184 preOB and OB stages (two weeks) (Figure 2C-E).

185

186 Transcription factor regulatory network during human osteogenic differentiation

187 By PCA, we found that the DEG expression profiles segregated all of the samples by
188 their differentiation stage, with MSCs being most distinct from preOBs and OBs, which
189 is consistent with the larger number of DEGs present between the MSC and preOB
190 stages than the later stages (Figure 3A). Of note, when we used only the 840
191 differentially expressed TFs, their expression profiles also segregated the three
192 osteogenic stages (Figure 3B).

193 TFs regulate not only the transcription of other protein-coding genes and
194 noncoding RNA genes but also their own expression. We focused on the TF biological
195 cooperativity with respect to TF-TF regulation, which has not been extensively
196 investigated in osteogenic differentiation due to the limited availability of DNA-binding
197 data from MSCs, preOBs, and OBs. To predict the TF-TF regulatory network during
198 osteogenic differentiation, we constructed a network using our RNA-seq data and the
199 database ReMap, which integrated and analyzed 5,798 human ChIP-seq and ChIP-exo
200 datasets from public sources providing a transcriptional regulatory repertoire to predict
201 target genes (Chèneby et al., 2020). ReMap covers 1,135 transcriptional regulators,
202 including primarily TFs as well as coactivators, corepressors, and chromatin-remodeling
203 factors, with a catalog of 165 million binding peaks. We filtered out TFs that lacked
204 associations in ReMap, as they are less likely to have direct binding activities in
205 regulating the transcription of TF genes. To further assure the reliability and accuracy of
206 the regulatory relationships in our osteogenic differentiation datasets, we then assessed
207 regulations among them by computing correlation coefficients of gene expression,
208 resulting in 451 TFs in the network. After the application of a partitioning algorithm, the

network was organized into five interconnected modules (Bastian et al., 2009) (Figure 3C; Supplementary Table 3). The network showed that TFs potentially regulate other TFs both internal and external to their respective modules, revealing a highly complex network of transcriptional regulators. We identified the top 100 TFs determined by their betweenness centrality, which coincided with their power to regulate others in the network (Bastian et al., 2009) (Figure 3C; Supplementary Table 3). Among them, TF genes known to regulate osteogenic differentiation were present in different modules, such as *ATF4* (Yang et al., 2004) in Module 1, *FOSL1* (Krum et al., 2010) in Module 2, *ZEB1* (Fu et al., 2020) in Module 3, and *TEAD4* (Suo et al., 2020) in Module 5 (Supplementary Table 3). Of interest, TF *KLF16* in Module 5, which previously was not demonstrated to be involved in bone formation and development, was identified as a key player in the regulatory network. It ranks as the fifth top regulator, directly interacting with the second top regulator *TEAD4*, and shows a high degree of association with 315 TFs (Supplementary Table 3).

Gene ontology (GO) and Reactome pathway (RP) analyses revealed regulatory functions and pathways specific to each module (FDR < 0.05; Figure 3D and 3E; Supplementary Table 3). As examples for the GO biological process (BP) terms, Module 3 was enriched for skeletal muscle cell differentiation and muscle structure development, associated with 15 TF genes (*ZBTB18*, *MEF2A*, *EGR1*, *TCF7L2*, *EGR2*, *EPAS1*, *SRF*, *ZBTB42*, *ETV1*, *FOS*, *RBPJ*, *NR4A1*, *SIX4*, *ID3*, *ATF3*), as well as the term fat cell differentiation and six associated TF genes (*ATF2*, *NR4A1*, *TCF7L2*, *EGR2*, *PPARG*, *GLIS1*). Module 5, containing *KLF16*, was enriched for terms ncRNA

231 transcription and positive regulation of the Notch signaling pathway (Figure 3D), which
232 is essential in bone development and metabolism (Liu et al., 2016).

233 In addition, each of the modules was enriched for GO cellular component (CC)
234 terms of various complexes. Module 2 was enriched for the term ESC/E(Z) complex, a
235 multimeric protein complex that can interact with several noncoding RNAs, a vital gene
236 silencing complex regulating transcription, and an effector of response to ovarian
237 steroids (Dubey et al., 2017) (Supplementary Figure 2A). The modules were enriched
238 for different GO molecular function (MF) terms of receptor binding of various hormones,
239 including androgen, estrogen, glucocorticoid, and thyroid hormone receptor binding
240 (Supplementary Figure 2B), providing more evidence of the crosstalk between bone and
241 gonads (Oury, 2012). By RP analysis, Module 2 was associated with activation of HOX
242 genes during differentiation, aligning with their well-established roles in osteogenic
243 differentiation of MSCs and skeletogenesis (Seifert et al., 2015). Module 5 was
244 associated with signaling by TGF-beta family members, and the
245 SMAD2/SMAD3:SMAD4 heterotrimer regulates transcription (Figure 3E). These results
246 revealed the complex and robust nature of TF regulatory networks during osteogenic
247 differentiation.

248

249 **Identification of key regulators in the co-expression network of human
250 osteogenic differentiation.**

251 We performed multiscale embedded gene co-expression network analysis (MEGENA)
252 to identify gene co-expression structures (i.e., modules) as well as key network
253 regulators (KNRs) (Song and Zhang, 2015; Holmes et al., 2020). MEGENA was

254 conducted on the RNA-seq data from the samples at the MSC, preOB, and OB
255 differentiation stages and identified 168 gene co-expression modules ($FDR \leq 0.05$,
256 Figure 4A; Supplementary Table 4). We then identified potential KNRs that were
257 predicted to modulate a large number of DEGs in the network by key regulator analysis
258 (Zhang et al., 2013; Zhang and Zhu, 2013). Many modules enriched for DEGs and their
259 associated KNRs recapitulated known osteogenic factors. For example, Module M204,
260 comprised of 85 genes, was significantly enriched for down-regulated DEGs (MSC to
261 preOB, fold enrichment = 1.80, adjusted p value = 1.58E-2; preOB to OB, fold
262 enrichment = 2.29, adjusted p value = 1.08E-3) (Figure 4A). More than 9% (8 genes) of
263 the DEGs in M204 are related to bone development and metabolism (Supplementary
264 Table 4). Three (*PHLDA2*, *RHOC*, *PFN1*) out of the eight genes were identified as
265 KNRs (Figure 4B). The expression of *PHLDA2* is associated with skeletal growth and
266 childhood bone mass (Lewis et al., 2012). Knockdown of *RHOC* dramatically inhibits
267 osteogenesis (Zheng et al., 2019). *PFN1*, a member of the profilin family of small actin-
268 binding proteins, is known to regulate bone formation. The loss of *PFN1* function causes
269 early onset of Paget's disease of bone, a disease with impaired osteoclast and
270 osteoblast differentiation (Scotto di Carlo et al., 2020). We also examined the overlap of
271 the genes in M204 with the genes from genome-wide association studies with
272 significant signals for bone area (B-area), bone mineral density (BMD), and hip
273 geometry (Hsu et al., 2019; Morris et al., 2019; Styrkarsdottir et al., 2019), and found
274 five genes, *CYFIP1*, *MMD*, *PELO*, *PRSS23*, and *SNX13*.

275 As TFs are known to be fundamental to the osteogenic process, we were
276 particularly interested in finding TFs with novel regulatory roles in osteogenesis. We

277 focused on the modules that were significantly enriched for DEGs and contained
278 differentially expressed TFs that were identified as KNRs (Supplementary Table 4). We
279 identified 80 such KNR TFs, of which 60 have unknown biological functions in
280 osteogenesis (Supplementary Table 4). The top 3 up-regulated (*HIF3A*, *ZBTB16*, and
281 *NR2F1*) (Zhu et al., 2014; Onizuka et al., 2016; Manikandan et al., 2018) and the top 2
282 down-regulated (*TEAD4*, *HMGA1*) (Suo et al., 2020; Wu et al., 2021) KNR TFs have
283 established roles in osteogenesis. The third most down-regulated KNR TF is *KLF16* of
284 Module M204, which previously had little known involvement in osteogenesis
285 (Supplementary Table 4). Thus, based on our cumulative analyses, we hypothesized
286 that *KLF16* plays an important role in osteogenic differentiation.

287 To further evaluate the expression patterns of the top candidate genes identified
288 through our MEGENA, DEG, and TF-TF network analyses at the cell type level, we
289 employed our previously published single-cell RNA-seq (scRNA-seq) data generated
290 from iPSC-induced cells in osteogenic differentiation (Housman et al., 2022). In that
291 study, gene expression was assessed at two stages, MSCs (Day 0) and osteogenic
292 cells (Day 21) – analogous to our MSC (Day 0) and OB (Day 21) stages for bulk RNA-
293 seq, respectively. Although the differentiation culture conditions differed from this study,
294 we found similar differential expression patterns in a pseudobulk analysis of the scRNA-
295 seq time points for the top 5 gene sets of up- and down-regulated KNR TFs
296 (Supplementary Table 4), TF regulators based on TF-TF regulatory networks
297 (Supplementary Table 3), as well as 5 known osteogenic markers (Supplementary
298 Figure 3; Supplementary Table 5). We further analyzed pseudobulk expression data
299 derived from the five different osteogenic cell types at day 21 and found that most

300 genes of interest showed similar levels of expression among these cell types with the
301 most differences occurring in mature osteocytes, including a notable reduction of
302 RUNX2 expression as expected (Thomas and Jaganathan, 2022) (Supplementary
303 Table 6).

304

305 **Regulatory role of *Klf16* in murine osteogenic differentiation.**

306 We further focused on *KLF16*, as it is one of the top TFs in the TF regulatory network
307 and a KNR in the co-expression network. *KLF16* is significantly downregulated from the
308 MSC to the OB stage (Figure 5A), suggesting its inhibitory role in osteogenic
309 differentiation. To validate this, we overexpressed *Klf16* in the murine preOB line
310 MC3T3-E1 using lentiviral vector-mediated gene transfer (Figure 5B). We then
311 performed three independent osteogenic differentiation experiments, each with three
312 technical replicates for each differentiation stage (Day 7, 14, and 21). Overexpression of
313 *Klf16* in MC3T3-E1 cells dramatically suppressed osteogenic differentiation at the early
314 stage (Day 7) with reduced ALP activity. At later stages (Day 14 and Day 21), reduced
315 mineralization was detected by alizarin red and von Kossa staining in the *Klf16*-
316 overexpressing cells compared to the control cells (Figure 5C). The inhibitory effect of
317 *Klf16* overexpression on osteogenic differentiation in vitro was observed consistently for
318 each of the experimental and technical replicates.

319 To explore the role of *Klf16* in bone formation in vivo, we analyzed a *Klf16* knock-
320 out mouse line, generated by deleting a 444 bp segment in exon one using CRISPR
321 technology by the Knockout Mouse Phenotyping Program (KOMP). This mutation
322 deletes the Kozak sequence and the ATG start codon, leaving the final 16 bp of exon

323 one and the splice donor site. We examined adult wild type (WT) and heterozygous
324 (*Klf16*^{+/−}) mice, excluding homozygous mice due to preweaning lethality. No significant
325 difference in the overall length between adult *Klf16*^{+/−} mice and WT controls was found
326 (Figure 5D). However, adult *Klf16*^{+/−} mice had significantly increased lean mass (total
327 tissue mass – fat mass, including bone mineral content) (p value < 0.05), and no
328 significant change in fat mass relative to WT mice (Figure 5D). Furthermore, there was
329 a dramatic increase in whole body bone mineral content (BMC), bone area, and bone
330 mineral density in *Klf16*^{+/−} mice compared to WT control mice revealed by DEXA
331 scanning (p value < 0.01) (Figure 5E). Microcomputed tomography (microCT) scanning
332 of trabecular bone from adult *Klf16*^{+/−} mice showed significantly increased bone volume
333 fraction (BV/TV) (p value < 0.05) and trabecular thickness (Tb.Th) (p value < 0.01), and
334 no significant difference in trabecular number (Tb.N) or trabecular separation (Tb.Sp)
335 (Figure 5F and 5H). In addition, adult *Klf16*^{+/−} mice showed significantly increased
336 cortical bone area (Ct.Ar) (p value < 0.05), cortical periosteal perimeter (Ct.Pe.Pm) (p
337 value < 0.01), and cortical endosteal perimeter (Ct.En.Pm) (p value < 0.01) compared to
338 WT control mice (Figure 5G and 5I). These findings further support the inhibitory role of
339 KLF16 in bone formation.

340 Given that bone phenotypes *in vivo* may be affected by both bone formation and
341 resorption, we analyzed osteoclastogenesis in the *Klf16*^{+/−} mice. The expression of
342 calcitonin receptor (CALCR) as a marker for osteoclasts (Boyce, 2013; Nicholson et al.,
343 1986) was not significantly changed in the epiphysis of the femur, while increased
344 expression was observed in the metaphysis of the *Klf16*^{+/−} mice compared to the
345 controls (Supplementary Figure 4A and 4B). In addition, we examined the expression of

346 receptor activator of NF- κ B ligand (RANKL), which binds to receptor activator of NF- κ B
347 (RANK) on the surface of osteoclast precursors, promoting their maturation into active
348 osteoclasts and regulating osteoclastogenesis (Khosla, 2001; Boyle et al., 2003). In the
349 femurs of the *Klf16*^{+/−} mice, the expression of RANKL did not show a significant change
350 compared to control mice (Supplementary Figure 4C and 4D). These results suggest
351 that osteoclastogenesis is not decreased in the *Klf16*^{+/−} mice; therefore, increased bone
352 mineral content and density in the mutant mice is more likely attributed to enhanced
353 bone formation rather than reduced resorption by osteoclasts.

354

355 Discussion

356 Understanding the cellular and molecular mechanisms underlying both normal and
357 pathological bone formation requires insight into the global transcriptional changes
358 during osteogenic differentiation. To investigate these changes, we utilized human
359 iPSC-derived MSCs as our model system. By profiling RNA expression during in vitro
360 osteogenic differentiation, we observed that the transcriptome profiles of our iPSC-
361 derived MSCs and OBs were consistent with those of their primary counterparts, which
362 suggests the potential applicability of our differentiation system to model in vivo
363 osteogenesis. Our differential gene expression analyses provided insights into
364 transcriptional signatures for each osteogenic stage, characterized by changes in
365 expression levels of numerous coding and noncoding genes. Our systems biology
366 approach revealed complex networks of TFs underlying osteogenic differentiation. TF
367 networks in mammalian cells can be built by combining knowledge of TF binding events
368 within TF genes and functional analysis (Wilkinson et al., 2017). Our research unveiled

369 a multifaceted osteogenic differentiation TF network, segmented into five interactive
370 modules. These modules, consistent with previous findings on functional modularity
371 (Dittrich et al., 2008), showcased specific biological functions. For instance, Module 3
372 underscores TFs associated with fat cell differentiation, suggesting a balance between
373 osteogenesis and adipogenesis within bone structures. This resonates with recent
374 studies emphasizing the role of a network of stem cell TFs in osteogenesis, which act
375 as repressors of adipogenesis (Rauch et al., 2019). Additionally, this module identifies
376 TFs involved in muscle differentiation, indicating overlapping regulatory mechanisms for
377 both bone and muscle development. This concept is further supported by recent
378 discoveries linking mesenchymal progenitors in muscle to bone formation (Julien et al.,
379 2021). This connectivity endorses known interactions between signaling pathways like
380 TGF- β /BMP and Wnt (Hernández-Vega and Camacho-Arroyo, 2021; Thomas and
381 Jaganathan, 2022). Overall, these findings reveal intricate TF regulatory networks and
382 pathway interactions in osteogenic differentiation.

383 In addition, we explored the gene co-expression network, highlighting potential
384 gene modules and key regulators that play an important role in differentiation. We
385 experimentally validated the regulatory role of *KLF16* from Module M204 as a proof-of-
386 concept that our dataset provides a resource for the discovery of novel mechanisms in
387 osteogenic differentiation. Module M204 is enriched for DEGs, of which many are
388 related to bone formation, including KNR genes *PFN1*, *PHLDA2*, and *RHOC*. We then
389 focused on the KNR *KLF16*, which was not previously associated with bone formation,
390 and investigated its involvement in osteogenic differentiation. Overexpression of *Klf16* in
391 vitro reduced ALP activity and matrix mineralization during MC3T3-E1 differentiation.

392 Conversely, heterozygous *Klf16* knockout mice showed altered bone characteristics,
393 including increases in whole body BMD, femoral bone volume to total tissue volume,
394 trabecular bone thickness, and cortical bone area, suggesting KLF16's regulatory
395 function in inhibiting osteogenic differentiation.

396 KLF16, a C2H2-type zinc finger transcription factor, has affinity for GC and GT
397 boxes, and displaces TFs SP1 and SP3 in neural settings (Hwang et al., 2001; Wang et
398 al., 2016). These TFs, when attached to certain promoters, play roles in osteogenic
399 differentiation (Le Mée et al., 2005). Furthermore, KLF16 targets specific metabolic
400 genes, including estrogen-related receptor beta (ESRRB) and PPARA (Daftary et al.,
401 2012; Sun et al., 2021). Our TF regulatory network analysis indicates that *KLF16* is
402 associated with Notch signaling and TGF β /BMP-SMAD signaling pathways, and
403 regulation of ncRNA transcription. This paints a broad canvas of potential regulatory
404 mechanisms for KLF16 in bone dynamics. Furthermore, other members from the KLF
405 family, especially *KLF4* (Module 1), *KLF5* (Module 5), and *KLF15* (Module 3)
406 (Supplementary Table 4), significantly influence skeletal development (Shinoda et al.,
407 2008; Song et al., 2017; Li et al., 2021; Yu et al., 2021; Zakeri et al., 2022). *KLF5*, co-
408 existing with *KLF16* in Module 5, curtails osteogenesis by inhibiting β -catenin through
409 DNMT3B-induced hypermethylation (Li et al., 2021). *KLF4*'s modulation of osteogenic
410 differentiation hinges on the BMP4-dependent pathway, resulting in reduced osteoblast
411 numbers in the deficient mice (Yu et al., 2021). Meanwhile, *KLF15* enhances
412 chondrogenic differentiation via the TF SOX9 promoter (Song et al., 2017). The roles of
413 *KLF6* and *KLF9* (Modules 3 and 2) in bone formation remain less defined (Zakeri et al.,

414 2022). Collectively, these findings underline the KLF family's complex involvement in
415 osteogenesis.

416 Interestingly, *KLF16* was upregulated in the MSCs of elderly patients suffering
417 from osteoporosis compared to age-matched controls in a previous study (Benisch et al.,
418 2012). A following study identified *KLF16* as one of the key TFs whose targets were
419 enriched from the DEGs in the MSCs of the osteoporosis patients versus controls (Liu
420 et al., 2019). The inhibitory role of *KLF16* in osteogenic differentiation identified in our
421 study supports the hypothesis that the overexpression of *KLF16* might contribute to
422 osteoporosis, and that *KLF16* could therefore be a potential therapeutic target.

423 Taken together, the differentiation of iPSC-derived MSCs to OBs provides a
424 valuable model for investigating gene expression and regulatory networks in osteogenic
425 differentiation. Our study sheds light on the intricate, layered, and dynamic regulation of
426 the transcriptomic landscape in osteogenic differentiation and offers a foundational
427 resource for further exploration of normal bone formation and the mechanisms driving
428 pathological conditions. Furthermore, this experimental model might facilitate
429 therapeutic research, with potential applications in the treatment of conditions such as
430 osteoporosis.

431

432 **Methods**

433 **Human subjects**

434 The twenty human subjects included in this study were in good general health. The
435 study protocols and informed consent were approved by the Institutional Review Board
436 (IRB) of Stanford University and the IRB of the Icahn School of Medicine at Mount Sinai.

437 Each subject gave written informed consent for study participation. Complete subject
438 demographic data can be found in Supplementary Table 1.

439

440 **Animal care and use**

441 The *Klf16^{+/−}* mouse strain C57BL/6NJ-Klf16em1(IMPC)J/Mmjjax was generated on the
442 C57BL/6NJ genetic background by the Knockout Mouse Phenotyping Program (KOMP)
443 at The Jackson Laboratory (Stock# 032653) using CRISPR technology. Procedures
444 using mice were in compliance with animal welfare guidelines mandated by the
445 Institutional Animal Care and Use Committee (IACUC) and Institutional Biosafety
446 Committee (IBC) at The Jackson Laboratory. Genotyping was performed by real-time
447 PCR of DNA extracted from tail biopsies. A detailed protocol can be found on The
448 Jackson Laboratory website.

449

450 **Cell lines**

451 MC3T3-E1 Subclone 4 cell line was purchased from ATCC (CRL-2593, sex
452 undetermined). Irradiated mouse embryonic fibroblasts (MEFs) were purchased from
453 GlobalStem (GSC-6001G). Human iPSC lines were generated in this study. The cells
454 were maintained at 37°C with 5% CO₂.

455

456 **Peripheral blood mononuclear cell (PBMC) and dermal fibroblast isolation**

457 PBMCs were isolated as we previously described (Carcamo-Orive et al., 2017). Briefly,
458 blood was drawn in tubes containing sodium citrate anticoagulant solution (BD
459 Vacutainer CPT mononuclear cell preparation tube, 362760). PBMCs were separated

460 by gradient centrifugation. The cells were then frozen in RPMI medium (MilliporeSigma,
461 R7388) supplemented with either 12.5% human serum albumin and 10% DMSO
462 (MilliporeSigma, D2438) or 10% fetal bovine serum (FBS) (MilliporeSigma, F4135) and
463 10% DMSO, and stored in liquid nitrogen for later use. The detailed protocol for dermal
464 fibroblast isolation can be found in our previous publication (Schaniel et al., 2021).
465 Briefly, a skin sample was taken from each of the clinically healthy subjects using a 3
466 mm sterile disposable biopsy punch (Integra Miltex, 98PUN3-11). Each sample was cut
467 into smaller pieces and placed into gelatin-coated tissue culture dishes with DMEM
468 (Thermo Fisher Scientific, 10567022) supplemented with antibiotics, 20% FBS, non-
469 essential amino acids (Thermo Fisher Scientific, 11140050), 2mM L-glutamine (Thermo
470 Fisher Scientific, 25030081), 2 mM sodium pyruvate (Thermo Fisher Scientific,
471 11360070), and 100 µM 2-mercaptoethanol (MP Biomedicals, C194705) to establish
472 fibroblast lines. Fibroblasts were harvested using TrypLE Express (Thermo Fisher
473 Scientific, 12605010) and passaged at a 1 to 4 split ratio. Fibroblasts were
474 cryopreserved in 40% DMEM, 50% FBS, and 10% DMSO.
475

476 **iPSC generation**

477 Erythroblast protocol: The detailed erythroblast reprogramming process can be found in
478 our previous publication (Carcamo-Orive et al., 2017). Briefly, PBMCs were thawed, and
479 the erythroblast population was expanded for 9–12 days until approximately 90% of
480 cells expressed CD36 and CD71. These expanded erythroblasts were reprogrammed
481 by transduction with Sendai viruses (SeV) expressing factors OCT3/4, SOX2, KLF4,

482 and c-MYC using the CytoTune™-iPS 2.0 Sendai Reprogramming kit (Thermo Fisher
483 Scientific, A16518) according to manufacturer's protocol.

484 Fibroblast protocol: The detailed fibroblast reprogramming process can be found
485 in our previous publication (Schaniel et al., 2021). Briefly, mycoplasma-free fibroblasts
486 at passage number 3-5 were reprogrammed using the mRNA reprogramming Kit
487 (Stemgent, 00-0071) in combination with the microRNA booster kit (Stemgent, 00-0073)
488 according to the manufacturer's protocol.

489

490 **iPSCs expansion and characterization**

491 To establish iPSCs, the cells were cultured on Matrigel (BD Bioscience, 354230-10, or
492 Corning, 254248) in feeder-free conditions and maintained in mTeSR1 (Stem Cell
493 Technologies, 05850). The iPSCs were passaged every 5-7 days after clones were
494 picked. To ensure the high quality of the iPSC lines used in this study, those lines
495 generated with Sendai virus first underwent virus clearance by continuous passaging.
496 To measure the loss of SeV in the iPSCs generated from erythroblasts with SeV,
497 quantitative RT-PCR was performed according to the SeV reprogramming kit
498 manufacture protocol. For quality control, G-banded karyotyping, ALP staining, iPSC
499 pluripotency marker immunocytochemistry, and embryoid body formation analysis
500 (Yaffe et al., 2016) were performed. Differentiating colonies were routinely eliminated
501 from the cultures. All the lines included in the current study had normal karyotypes,
502 exhibited human embryonic stem cell morphology, and expression of ALP and
503 pluripotency markers NANOG, SOX2, OCT4, SSEA4, and TRA-1-60. Further, embryoid
504 body (EB) formation and differentiation assays demonstrated their competence for

505 differentiation into all three germ layers. We excluded samples where any of these
506 characterizations were abnormal (Daley et al., 2009; Yaffe et al., 2016).

507

508 **Mycoplasma quality control**

509 Mycoplasma testing of the cells was performed intermittently during culture. Cultures
510 were grown in the absence of antibiotics for at least three days before testing. The
511 culture medium or the DNA harvested from the cultures were tested for mycoplasma
512 contamination using a MycoAlert Mycoplasma Detection Kit (Lonza, LT07-418) or e-
513 Myco PLUS PCR Detection Kit (BOCA scientific, 25237).

514

515 **ALP staining**

516 Cells cultured on plates were briefly washed with phosphate-buffered saline (PBS),
517 fixed in 4% PFA for 3 minutes at room temperature, and then stained with Alkaline
518 phosphatase kit II (Stemgent, 00-0055) according to the manufacturer's instructions.

519

520 **iPSC embryoid body analysis**

521 iPSC embryoid body formation and characterization were performed as previously
522 described with modifications (Lin and Chen, 2008). Briefly, iPSCs were cultured in
523 mTeSR1 medium on Matrigel-coated plates, and the medium was changed daily. On
524 the day of EB formation, when the cells grew to 60-80% confluence, cells were washed
525 once with PBS and then incubated in Accutase for 8-10 minutes to dissociate colonies
526 to single cells and resuspended with mTeSR1 containing 2 µM Thiazovivin. To form
527 self-aggregated EBs, single iPSCs were transferred into an ultra-low attachment plate

528 (Corning, CLS3471) via 1:1 or 1:2 passage. EBs were aggregated from iPSCs for three
529 days, then were transferred to 0.1% gelatin-coated 12 well plates and maintained in
530 embryoid body medium (DMEM/F12 supplemented with 10% FBS, 2 mM L-glutamine,
531 0.1 mM non-essential amino acids, and 0.1 mM 2-mercaptoethanol) for another ten
532 days for the spontaneous generation of three germ layers with medium changed every
533 other day. Fluorescent immunostaining was performed to detect the expression of
534 markers of the three germ layers.

535

536 **Fluorescence immunocytochemistry**

537 The cells cultured on plates were briefly washed with PBS, fixed in 4%
538 paraformaldehyde (PFA) in PBS for 15 minutes, washed with PBS three times, and then
539 permeabilized with 0.2% Triton in PBS for 15 mins. The cells were then washed with
540 PBS three times prior to blocking with 1% bovine serum albumin (BSA) (MilliporeSigma,
541 A8412) in PBS for 1 hour at room temperature and incubating with primary antibody
542 1:100 to 1:900 dilutions in 0.1% BSA in PBS for overnight at 4°C. Washing three times
543 was conducted with PBS followed by incubation with a corresponding secondary
544 antibody in 1:500 dilution in 0.1% BSA for 1 hour at room temperature. After washing
545 with PBS, 1:50,000 diluted Hoechst (Thermo Fisher Scientific, H1398) was added to
546 stain nuclei for 5 minutes, followed by washing with PBS three times and imaging. The
547 following primary antibodies were applied to detect the expression of iPSC markers:
548 anti-TRA-1-60 (Invitrogen, 41-1000, 1/100), anti-SSEA4 (Invitrogen, 41-4000, 1/200),
549 anti-NANOG (Abcam, ab109250, 1/100), anti-OCT-4 (Cell Signaling Technology, 2840S,
550 1/400), and anti-SOX2 (Abcam, ab97959, 1/900). To detect the expression of the

551 markers of three germ layers, anti-AFP (Agilent, A0008, 1/100) for endoderm, anti- α -
552 SMA (MilliporeSigma, A5228, 1/200) for mesoderm, and anti-TUBB3 (MilliporeSigma,
553 T2200, 1/200) for ectoderm were used. All the Alexa-conjugated secondary antibodies
554 used were from Invitrogen: anti-mouse Alexa Fluor 488 (A-11029), anti-rabbit Alexa 488
555 (A-21206), and anti-rabbit Alexa 594 (A-11037).

556

557 **G-banding karyotyping**

558 Karyotype analysis was performed on all iPSC lines by the Human Genetics Core
559 facility at SEMA4. iPSCs were cultured on Matrigel-coated T25 flasks before
560 karyotyping and reached approximately 50% confluence on the day of culture harvest
561 for karyotyping. Twenty cells in metaphase were randomly chosen, and karyotypes
562 were analyzed using the CytoVision software program (Version 3.92 Build 7, Applied
563 Imaging).

564

565 **Differentiation of iPSCs to MSCs**

566 In vitro differentiation of iPSCs to MSCs was carried out with commercial cell culture
567 media according to the manufacture protocols with modifications. Briefly, iPSC colonies
568 were harvested with Accutase (Innovative Cell Technologies, AT-104) and seeded as
569 single cells on a Matrigel-coated plate in mTeSR1 and supplemented with 2uM
570 Thiazovivin. The next day, TeSR1 medium was replaced with STEMdiff mesoderm
571 induction medium (MIM) (Stem Cell Technologies, 05221) when cells were at
572 approximately 20-50% confluence. Cells were then fed daily and cultured in STEMdiff
573 MIM for four days. On Day 5, the culture medium was switched to MesenCult-ACF

574 medium (Stem Cell Technologies, 05440/8) for the rest of the MSC induction duration.
575 Cells were passaged as necessary using MesenCult-ACF Dissociation Kit (Stem Cell
576 Technologies, 05426). Cells were subcultured onto Matrigel at the first passage to avoid
577 loss of the differentiated cells before tolerating the switch to MesenCult-ACF attachment
578 substrate (Stem Cell Technologies, 05448). For passage two or above, cells were
579 subcultured on MesenCult-ACF attachment substrate. After three weeks of
580 differentiation, the cells were sorted for the CD105 (BD Biosciences, 561443, 1/20)
581 positive and CD45 (BD Biosciences, 555483, 1/5) negative MSC population (Giuliani et
582 al., 2011; Kang et al., 2015; Sotiropoulou et al., 2006) using BD FACSaria II in the
583 Mount Sinai Flow Cytometry Core Facility, and then expanded in MSC medium
584 consisting of low-glucose DMEM (Gibco, 10567022) containing 10% FBS (Sotiropoulou
585 et al., 2006). After expansion, a few MSC lines were further examined by BD Cantoll for
586 expression of other MSC positive surface markers CD29 (Thermo Fisher Scientific, 17–
587 0299, 1/20), CD73 (BD Biosciences, 561254, 1/20), CD90 (BD Biosciences, 555595, 1/
588 20), absence of other negative surface markers CD31 (BD Biosciences, 561653, 1/20),
589 CD34 (BD Biosciences, 560940, 1/5), and retention of marker CD105 positivity and
590 CD45 negativity as well.

591

592 **Fluorescence activated cell sorting (FACS) and analysis**

593 Single cells were washed with FACS buffer (PBS supplemented with 1% FBS and
594 25mM HEPES (Thermo Fisher Scientific, 1688449)) twice, resuspended to a
595 concentration of 1×10^7 cells/ml in ice-cold FACS buffer, and incubated with the above
596 conjugated antibodies for 30 minutes at 4° in the dark. Stained cells were then washed

597 with FACS buffer three times before FACS or flow cytometry analysis. Flow cytometry
598 data were analyzed with BD FACSDiva software.

599

600 **Differentiation of iPSC-derived MSCs to osteoblasts**

601 iPSC-derived MSCs were plated in a 6-well plate at a density of 3×10^3 cells/cm² in MSC
602 medium. After three days, the culture medium was switched to osteogenic differentiation
603 medium (α MEM (Thermo Fisher Scientific, A1049001) supplemented with 10% FBS, 1%
604 non-essential amino acids, 0.1 μM dexamethasone (MilliporeSigma, D4902), 10 mM β-
605 glycerophosphate (MilliporeSigma, G9422), and 200 μM ascorbic acid (MilliporeSigma,
606 A4544)) and maintained in this medium for 21 days (Barberi et al., 2005; Lee et al.,
607 2015; Pittenger et al., 1999). The culture medium was changed every 2-3 days. Cells
608 were harvested for total RNA isolation and RNA sequencing at Day 0, Day 7, and Day
609 21 of differentiation, as indicated in the main text. ALP staining, and alizarin red staining
610 and von Kossa staining, were employed to examine bone ALP expression and
611 mineralization, respectively.

612

613 **Alizarin red staining and von Kossa staining**

614 Mineralization of osteoblast extracellular matrix was assessed by both alizarin red
615 staining and von Kossa staining. Cells were washed briefly with PBS, fixed with 4% PFA
616 for 15 minutes, and washed with deionized distilled water three times. For alizarin red
617 staining, fixed cells were incubated in alizarin red stain solution (MilliporeSigma, TMS-
618 008-C) with gentle shaking for 30 minutes, followed by washing with water four times for
619 5 minutes each with gentle shaking to remove non-specific alizarin red staining. For von

620 Kossa staining, fixed cells were incubated in 5% silver nitrate solution (American Master
621 Tech Scientific, NC9239431) while exposed to UV light for 20 minutes. Unreacted silver
622 was removed by incubating in 5% sodium thiosulfate (American Master Tech Scientific,
623 NC9239431) for 5 minutes, followed by washing with water twice. Mineralized nodules
624 were identified as red spots by alizarin red staining and dark brown to black spots by
625 von Kossa staining.

626

627 **RNA-seq library preparation and sequencing**

628 For total RNA isolation, cells were washed with PBS and harvested for RNA extraction
629 using miRNeasy mini kit (QIAGEN, 217004) according to the manufacturer's instruction.
630 Illumina library preparation and sequencing were conducted by the Genetic Resources
631 Core Facility, Johns Hopkins Institute of Genetic Medicine (Baltimore, MD). RNA
632 concentration and quality were determined using NanoDrop Spectrophotometer
633 (Thermo Scientific, DE). RNA integrity (RNA Integrity Number, RIN) was verified using
634 Agilent BioAnalyzer 2100 and the RNA Nano Kit prior to library creation. Illumina's
635 TruSeq Stranded Sample Prep kit was used to generate libraries. Specifically, after
636 ribosomal RNA (rRNA) depletion, RNA was converted to cDNA and size selected to
637 150-200 bp in length, then end-repaired and ligated with appropriate adaptors. Ligated
638 fragments were subsequently size-selected and underwent PCR amplification
639 techniques to prepare the libraries with a median size of 150 bp. Libraries were uniquely
640 barcoded and pooled for sequencing. The BioAnalyzer was used for quality control of
641 the libraries to ensure adequate concentration and appropriate fragment size.
642 Sequencing was performed on an Illumina HiSeq 2500 instrument using standard

643 protocols for paired-end 100 bp sequencing. All the samples were processed in one
644 batch.

645

646 **RNA-seq pre-processing and gene differential expression analyses**

647 Illumina HiSeq reads were processed through Illumina's Real-Time Analysis (RTA)
648 software generating base calls and corresponding base call quality scores.
649 CIDRSeqSuite 7.1.0 was used to convert compressed bcl files into compressed fastq
650 files. After adaptor removal with cutadapt (Martin, 2011) and base quality trimming to
651 remove 3' read sequences if more than 20 bases with $Q \geq 20$ were present, paired-
652 end reads were mapped to the human GENCODE V29 reference genome using STAR
653 (Dobin et al., 2013) and gene count summaries were generated using featureCounts
654 (Liao et al., 2014). Raw fragment (i.e., paired-end read) counts were then combined into
655 a numeric matrix, with genes in rows and experiments in columns, and used as input for
656 differential gene expression analysis with the Bioconductor Limma package (Ritchie et
657 al., 2015) after multiple filtering steps to remove low-expressed genes. First, gene
658 counts were converted to FPKM (fragments per kb per million reads) using the RSEM
659 package (Li and Dewey, 2011) with default settings in strand-specific mode, and only
660 genes with expression levels above 0.1 FPKM in at least 20% of samples were retained
661 for further analysis. Additional filtering removed genes less than 200 nucleotides in
662 length. Finally, normalization factors were computed on the filtered data matrix using the
663 weighted trimmed mean of M values (TMM) method, followed by voom (Law et al., 2014)
664 mean-variance transformation in preparation for Limma linear modeling. The limma
665 generalized linear model contained fixed effects for sex (male/female), and cell source

666 (fibroblast/erythroblast) and a random effect term was included for each unique subject.
667 Data were fitted to a design matrix containing all sample groups, and pairwise
668 comparisons were performed between sample groups (i.e., MSC stage, preOB stage,
669 and OB stage). eBayes adjusted p values were corrected for multiple testing using the
670 Benjamin-Hochberg (BH) method (Benajmin and Hochberg, 1995) and used to select
671 genes with significant expression differences (adjusted p value < 0.05).

672

673 **Integration of RNA-Seq data of human primary MSCs and OBs, iPSCs, and**
674 **tissues in GTEx**

675 GTEx gene expression data (version 7) and metadata were downloaded from the GTEx
676 website (Ardlie et al., 2015). Raw counts were extracted for the brain, heart, kidney,
677 liver, lung, muscle, nerve, ovary, testis, and thyroid tissues. After filtering the lowly
678 expressed genes (1 CPM in less than 20% samples), normalization factors were
679 computed using the weighted trimmed mean of M values (TMM) method, followed by
680 the voom mean-variance transformation (Law et al., 2014). The Human iPSC dataset
681 was from our previous publication, deposited in the GEO database with accession
682 number GSE79636 (Carcamo-Orive et al., 2017). Two datasets for each primary human
683 MSC and OB cell type were collected from the GEO accessions GSE94736 (Roforth et
684 al., 2015) and GSE118808 (Kaczynski et al., 2003; Ma et al., 2019), and GSE55282
685 (Rojas-Peña et al., 2014) and GSE75524 (Al-Rekabi et al., 2016), respectively. All the
686 datasets were filtered and normalized in the same way as the GTEx data. Genes from
687 all the datasets were intersected to obtain shared genes. Gene expression levels of the
688 shared genes were extracted from all the datasets. Principal component analysis (PCA)

689 was performed on the extracted and combined gene expression matrix. Multi-
690 dimensional scaling was used to visualize the top two principal components.

691

692 **Gene ontology and Reactome pathway enrichment analyses**

693 Gene ontology (GO) biological process (BP), molecular function (MF), and cellular
694 component (CC), and Reactome pathway enrichment analyses were performed using
695 the PANTHER classification system (www.pantherdb.org) (Mi et al., 2019a). The
696 statistical overrepresentation test in PANTHER was fulfilled by Fisher's exact test
697 together with FDR multiple test correction to identify enriched GO categories and
698 Reactome Pathways among the input genes relative to the indicated reference list as
699 stated in figure legends (The Gene Ontology Consortium, 2019; Mi et al., 2019b).

700 Enrichment tests were filtered using FDR < 0.05.

701

702 **Transcription factor network analysis**

703 The ChIP-seq and ChIP-exo datasets of DNA binding peaks of human transcription
704 regulators (TRs), where the overwhelming majority are TFs, were downloaded from
705 ReMap 2020 (Chèneby et al., 2020). The genomic coordinates of the transcription start
706 sites of human genes were extracted from Ensembl Human Genes GRCh37.p13, and
707 then applied BedTools v2.3.0 to identify the target TR genes of a TR by checking the
708 intersection of the latter TR's binding peaks and the 2kb window before and after the
709 transcription start sites of the target TR genes (Quinlan and Hall, 2010). The Python
710 package `scipy.stats.pearsonr` was used to calculate the correlation coefficient between
711 the expression of TR genes (Hao et al., 2015; Jiang et al., 2014; Virtanen et al., 2020;

712 Zhu et al., 2015). The regulatory relationship between TRs was predicted using two
713 criteria: (1) the binding site peaks of a TR are within the distance of 2 kb upstream or
714 downstream of the known transcription start sites of the target TRs; and (2) the absolute
715 value of the correlation coefficient between the TR genes is greater than 0.6 and
716 adjusted p value is less than 0.05 (Camacho et al., 2005). Gephi 0.9.2 was used to
717 generate and visualize the regulatory network, where each node was defined as a TR
718 gene, and two nodes were connected by an edge when ReMap data demonstrated
719 regulation between the two TRs, and our RNA-seq data also showed expression
720 correlation between them. Furthermore, the network community detection hierarchical
721 algorithm was applied to determine relationships between subsets of the whole network
722 and define modules. Betweenness centrality for each of the nodes was calculated and
723 then used to rank node sizes (Bastian et al., 2009).

724

725 **Gene co-expression network analysis and identification of key network regulators**
726 Gene co-expression network was constructed using the R package MEGENA v1.3.7
727 (Song and Zhang, 2015). The same filtered, normalized, and covariate-adjusted gene
728 expression data matrix used in the differential gene expression analysis was used as
729 input for MEGENA. Specifically, the Pearson correlation was used to calculate
730 significant correlations between gene pairs among the 60 samples. With a cutoff of 0.05
731 FDR, significant correlations were identified by 100 permutations of the gene expression
732 matrix. Next, planar-filtered network construction and multi-scale clustering analysis
733 were performed. Finally, significant modules were identified at a 5% FDR with 100 times
734 network permutations. Modules of smaller than 50 genes or larger than 5,000 genes

735 were excluded from the downstream analyses. Enrichment analysis was performed
736 between the modules and DEG signatures between MSC and preOB stages and
737 between preOB and OB stages through Fisher's exact test. BH procedure was applied
738 to the p values to correct for multiple-testing problem (Benjamini and Hochberg, 1995).
739 Modules were visualized using the Cytoscape (Otasek et al., 2019). Sunburst plots of
740 modules were visualized using the R package *sunburstR* v2.1.5. We used all the
741 MEGENA nodes and edges as input to identify KNRs that were predicted to modulate a
742 large number of downstream DEGs in the network (Zhang et al., 2013; Zhang and Zhu,
743 2013). For each gene in the MEGENA network, we tested whether the network
744 neighborhood of the gene was enriched with the DEG signature. Specifically, we tested
745 if the nodes within a path length of 6 steps of the candidate KNR gene were enriched for
746 the DEGs using Fisher's Exact test. The p values were then corrected by the Bonferroni
747 procedure to adjust for multiple comparisons.

748

749 **Processing of comparative single-cell pseudobulk data**

750 scRNA-seq data (GEO accession: GSE181744) from our previous publication
751 (Housman et al., 2022), which includes mixtures of iPSC-derived, MSC osteogenic
752 differentiations from different species, was used as a comparative dataset. To maximize
753 the comparative utility of these data, we re-processed the raw scRNA-seq data to obtain
754 pseudobulk expression values for all genes annotated in the human genome and for cell
755 types of interest from the six humans and one human technical replicate included in this
756 dataset. Briefly, reads were processed using standard 10X Genomics Cell Ranger 3.1.0
757 pipelines (Zheng et al., 2017) that extracted 10X cell barcodes and UMIs and aligned

758 the remaining reads to the human genome (hg38). Human cells and specific cell types
759 of interest (MSCs, osteogenic cells, preosteoblasts, osteoblasts, embedding osteoblasts,
760 mineralizing osteoblasts, and maturing osteocytes) were isolated from the newly
761 processed data using species assignments and cell classifications (Housman et al.,
762 2022). Lastly, single-cell gene count data were consolidated to produce pseudobulk
763 expression values for each unique grouping of individual, replicate, and cell
764 classification. Specifically, these pseudobulk data were defined as the sum of raw
765 single-cell UMI counts within each individual-replicate for a given cell classification.
766 Computational scripts for these processing steps can be found on GitHub at
767 <https://github.com/ghousman/human-skeletal-scRNA>. Pseudobulk data were filtered
768 and normalized as described above.

769

770 **Generation of *Klf16*-overexpressed MC3T3-E1 cell line**

771 Lentiviral transduction of MC3T3-E1 Subclone 4 cells (ATCC, CRL-2593) was
772 performed as we previously described with modifications (Holmes et al., 2020). Cells
773 were infected by incubating in lentivirus-containing cell culture medium at a multiplicity
774 of infection (MOI) of 100 in the presence of 6 µg/ml polybrene for 24 hours. The
775 selection was performed with 2 µg/ml puromycin for 12 days until EGFP expression was
776 observed in 100% of cells, no further cell death was observed, and no live cells
777 remained in the non-transduction negative control dish. Selected cells were expanded
778 and cryopreserved. EGFP expression was monitored routinely, and additional selection
779 was performed when necessary.

780

781 **MC3T3-E1 osteoblast differentiation assay**

782 Osteoblast differentiation of MC3T3-E1 transfected with lentiviral vectors containing
783 either *Klf16* cDNA or a 300bp nonfunctional stuffer sequence was performed as follows:
784 cells were plated in 6-well plates at a density of 3×10^3 cells/cm² in MC3T3-E1
785 maintenance medium (aMEM supplemented with 10% FBS). After three days, the
786 culture medium was switched to osteogenic differentiation medium (Holmes et al., 2020),
787 and cells were maintained in this medium, which was replenished every two days for 21
788 days. Triplicate experiments, each having triplicate wells, were carried out for both
789 control and *Klf16*-overexpressing cells.

790

791 **RT-qPCR**

792 The RT-qPCR method was used to assess gene expression in stable lentiviral MC3T3-
793 E1 cell lines. When cells were plated in 6 well plates for osteogenic differentiation as
794 described above, one parallel well was used for RNA isolation after being cultured in
795 MC3T3-E1 maintenance medium for three days before switching to osteogenic
796 differentiation medium. Total RNA was extracted with the RNeasy Kit (Qiagen, 74106)
797 according to the manufacturer's protocol. cDNA was synthesized with AffinityScript
798 One-Step RT-PCR Kit (Agilent, 600188). Each cDNA sample was amplified in triplicate
799 using the SYBR Green and Platinum Taq polymerase (Thermo Fisher Scientific, S7567)
800 on a 7900HT Real-Time PCR instrument (Thermo Fisher Scientific, 10966). The
801 housekeeping gene *Actb* was used as the reference. mRNA relative expression was
802 calculated by the $\Delta\Delta Ct$ method. qPCR primers used are listed in Supplementary Table
803 7.

804

805 **Whole body bone mineral density**

806 Bone mineral density (BMD) was scanned with a Dual Energy X-ray Absorptiometry
807 (DEXA) Analyzer (Lunar, Piximus II, GE Medical System). Mouse weight and length
808 were measured (nose to the beginning of tail) before scanning. Each unconscious
809 mouse was placed in the DEXA analyzer. A scout-scan was performed. The analysis
810 was conducted on a whole body scan, excluding the head. The mouse was removed
811 once the image was captured and placed on a heated mat set at 37°C in a cage and
812 closely monitored until consciousness was regained.

813

814 **Microcomputed tomography**

815 Femurs were isolated by removing attached soft tissues, fixed in 4% PFA at 4°C, and
816 imaged with microCT scan (Skyscan 1172a, Skyscan, Belgium) at 50 kV and 200 µA
817 using a 0.5 mm aluminum filter and an image voxel size of 4.5 µm isotropic. Images
818 were captured every 0.7°, with 8x averaging, through 180° rotation of each bone and
819 reconstructed using Skyscan NRecon software. Imaging analysis of the metaphyseal
820 regions of each femur was performed by first determining a reference point of the most
821 proximal slice where the growth plate had begun to disappear. Offsets of 101 slices and
822 501 slices from the point of reference in the growth plate were used for trabecular and
823 cortical analyses, respectively, with Skyscan CTAn software. Volume visualization was
824 performed in Avizo 2020.2 (Thermo Fisher Scientific). Data were reported in the
825 standard format used by the American Society for Bone and Mineral Research (ASBMR)
826 (Parfitt et al., 1987). Each volume was then segmented to isolate osteological material

827 using a regularized deep network (RDN)-based image segmentation algorithm (Yazdani
828 et al., 2020). The segmented images were thresholded and masked in Medtool 4.4 (DPI
829 e.U, Medtool) to isolate trabecular bone from cortical bone (Gross et al., 2014). The
830 external mask (cortical bone) was subtracted from the volume, leaving only the internal
831 mask (trabecular bone) and air. Bone volume fraction (BV/TV), trabecular thickness
832 (Tb.Th), trabecular number (Tb.N), and trabecular separation (Tb.Sp) were calculated
833 from the trabecular bone volumes in 3D using 7 mm sampling spheres on a background
834 grid with 3.5 mm spacing following DeMars et al. (2021) in Medtool. Cortical bone
835 variables cortical area (Ct.Ar), cortical periosteal perimeter (Ct.Pe.Pm), and cortical
836 endosteal perimeter (Ct.En.Pm) were then calculated in the BoneJ extension of ImageJ
837 (Doube et al., 2010) for each volume.

838

839 Immunohistochemistry

840 Femurs from two male *Klf16^{+/−}* mice and two male WT mice at the age of 18 weeks were
841 fixed and prepared for paraffin section. Sections were deparaffinized in xylene and
842 rehydrated through a gradient of ethanol (100%, 95%, and 70%, 5 minutes each) to
843 water. Slides were incubated in 0.2% Triton X-100 for 5 min. After washing in PBS,
844 endogenous fluorescence was blocked using TrueBlack Plus Lipofuscin
845 Autofluorescence Quencher (1:40, Biotium, 23014) for 20 min, and then blocked with an
846 Animal-Free Blocker (Vector Laboratories, SP-5035-100) for 1 hour at RT. The primary
847 antibody anti-CALCR (1:50, Bioss, BS-0124R), or anti-RANKL (1:50, Abcam,
848 ab216484), was applied and incubated overnight at 4°C. After washing in PBS,
849 secondary antibody (Alexa Fluor 488 chicken anti-Rabbit IgG, 1:500, Invitrogen, A-

850 21441) was applied and incubated for 1 hour at room temperature. Slides were then
851 washed in PBS, stained with DAPI, and mounted with Antifade Mounting Media (Vector
852 Labs, H-1700-10). Images were acquired with a Nikon T1-SM microscope.

853

854 **Statistical analyses**

855 Statistical analyses were performed with Microsoft Excel using Student's t-test to create
856 graphs displaying mean values with error bars corresponding to the standard error of
857 the mean (SEM) for *Klf16* in vitro overexpression and animal experiments. p value <
858 0.05 was considered significant. Sample sizes are indicated in figure legends. Image
859 analyses were performed with ImageJ (National Institutes of Health). IHC data were
860 quantified and visualized with GraphPad Prism 10 (GraphPad Software).

861

862 **Data and code availability**

863 The RNA-seq data supporting the findings of this study are deposited in Gene
864 Expression Omnibus (GEO) with accession number GSE200492.

865

866 **Acknowledgements**

867 We thank Paige Cundiff for support in iPSC derivation, Arvind Babu for chromosome
868 analyses, Xuqiang Qiao for assistance in the Dean's Flow Cytometry Core, and Jacqui
869 White and Arie Mobley of The Jackson Laboratory Center for Biometric Analysis for the
870 microCT imaging and DEXA scanning of the *Klf16* mouse line. This work was supported
871 in part through the computational resources and staff expertise provided by Scientific
872 Computing at the Icahn School of Medicine at Mount Sinai. The research reported in

873 this paper was supported by the Office of Research Infrastructure of the National
874 Institutes of Health under award numbers S10OD026880 and S10OD030463. This work
875 was funded by NIH grants P01HD078233 (Ethylin Wang Jabs) and R01DE029832 (Ethylin
876 Wang Jabs, Susan M. Motch Perrine) and F32AR075397 (Genevieve Housman). The
877 content is solely the responsibility of the authors and does not necessarily represent the
878 official views of the National Institutes of Health.

879

880 **Author details**

881 **Ying Ru**

882 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount
883 Sinai, New York, NY, 10029, USA

884 **Contribution:** Conceptualization, Methodology, Validation, Formal analysis,
885 Investigation, Data curation, Writing – original draft, Writing – review & editing,
886 Visualization, Supervision

887 **Competing interests:** No competing interests declared

888

889 **Meng Ma**

890 Mount Sinai Genomics, Sema4, Stamford, CT, 06902, USA

891 **Contribution:** Conceptualization, Methodology, Validation, Formal analysis,
892 Investigation, Data curation, Writing – original draft, Writing – review & editing,
893 Visualization

894 **Competing interests:** No competing interests declared

895

896 **Xianxiao Zhou**

897 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

898 Sinai, New York, NY, 10029, USA

899 Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at

900 Mount Sinai, New York, NY, 10029, USA

901 Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY,

902 10029, USA

903 **Contribution:** Methodology, Validation, Formal analysis, Investigation, Data curation,

904 Writing – original draft, Writing – review & editing, Visualization

905 **Competing interests:** No competing interests declared

906

907 **Divya Kriti**

908 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

909 Sinai, New York, NY, 10029, USA

910 Present address: Department of Biochemistry and Molecular Biology, Faculty of

911 Medicine, The University of British Columbia, Vancouver, BC V6T 2G3, Canada

912 **Contribution:** Formal analysis, Investigation, Data curation, Writing – review & editing

913 **Competing interests:** No competing interests declared

914

915 **Ninette Cohen**

916 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount
917 Sinai, New York, NY, 10029, USA

918 Present address: Division of Cytogenetics and Molecular Pathology, Zucker School of
919 Medicine at Hofstra/Northwell, Northwell Health Laboratories, Lake Success, NY,
920 11030, USA

921 **Contribution:** Investigation, Writing – review & editing

922 **Competing interests:** No competing interests declared

923

924 **Sunita D'Souza**

925 Department of Cell, Developmental and Regenerative Biology, Icahn School of
926 Medicine at Mount Sinai, New York, NY, 10029, USA

927 Present address: St Jude Children's Research Hospital, Memphis, TN, 38105, USA

928 **Contribution:** Methodology - iPSC derivation, Writing – review & editing

929 **Competing interests:** No competing interests declared

930

931 **Christoph Schaniel**

932 Department of Cell, Developmental and Regenerative Biology, Icahn School of
933 Medicine at Mount Sinai, New York, NY, 10029, USA

934 Department of Medicine, Division of Hematology and Medical Oncology, Tisch Cancer
935 Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA

936 Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York,
937 NY, 10029, USA

938 Mount Sinai Institute for Systems Biomedicine, Icahn School of Medicine at Mount

939 Sinai, New York, NY, 10029, USA

940 **Contribution:** Methodology - iPSC derivation, Writing – review & editing

941 **Competing interests:** No competing interests declared

942

943 **Susan M. Motch Perrine**

944 Department of Anthropology, Pennsylvania State University, University Park, PA, 16802,

945 USA

946 **Contribution:** Writing – review & editing, Visualization

947 **Competing interests:** No competing interests declared

948

949 **Sharon Kuo**

950 Department of Biomedical Sciences, University of Minnesota, Duluth, MN, 55812, USA

951 Technological Primates Research Group, Max Planck Institute for Evolutionary

952 Anthropology, Leipzig 04103, Germany.

953 **Contribution:** Writing – review & editing, Visualization

954 **Competing interests:** No competing interests declared

955

956 **Oksana Pichurin**

957 Department of Clinical Genomics, Mayo Clinic, Rochester, MN, 55905, USA

958 **Contribution:** Investigation

959 **Competing interests:** No competing interests declared

960

961 **Dalila Pinto**

962 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

963 Sinai, New York, NY, 10029, USA

964 **Contribution:** Formal analysis, Writing – review & editing

965 **Competing interests:** No competing interests declared

966

967 **Genevieve Housman**

968 Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago,

969 IL, 60637, USA

970 Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary

971 Anthropology, Leipzig, 04103, Germany

972 **Contribution:** Writing – review & editing

973 **Competing interests:** No competing interests declared

974

975 **Greg Holmes**

976 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

977 Sinai, New York, NY, 10029, USA

978 **Contribution:** Conceptualization, Writing – review & editing

979 **Competing interests:** No competing interests declared

980

981 **Eric Schadt**

982 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

983 Sinai, New York, NY, 10029, USA

984 Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY,
985 10029, USA

986 **Contribution:** Conceptualization, Methodology, Writing – review & editing

987 **Competing interests:** No competing interests declared

988

989 **Harm van Bakel**

990 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount
991 Sinai, New York, NY, 10029, USA

992 Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY,
993 10029, USA

994 Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY,
995 10029, USA

996 **Contribution:** Conceptualization, Methodology, Validation, Formal analysis,
997 Investigation, Data curation, Writing – original draft, Writing – review & editing,
998 Visualization, Supervision

999 **Competing interests:** No competing interests declared

1000

1001 **Bin Zhang**

1002 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount
1003 Sinai, New York, NY, 10029, USA

1004 Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at
1005 Mount Sinai, New York, NY, 10029, USA

1006 Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY,

1007 10029, USA

1008 **Contribution:** Conceptualization, Methodology, Validation, Writing – review & editing,

1009 Supervision

1010 **Competing interests:** No competing interests declared

1011

1012 **Ethylin Wang Jabs**

1013 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

1014 Sinai, New York, NY, 10029, USA

1015 Department of Cell, Developmental and Regenerative Biology, Icahn School of

1016 Medicine at Mount Sinai, New York, NY, 10029, USA

1017 Department of Clinical Genomics, Mayo Clinic, Rochester, MN, 55905, USA

1018 Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, 55905,

1019 USA

1020 **Contribution:** Conceptualization, Validation, Writing – review & editing, Supervision,

1021 Project administration, Funding acquisition

1022 **Competing interests:** No competing interests declared

1023

1024 **Meng Wu**

1025 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount

1026 Sinai, New York, NY, 10029, USA

1027 Department of Clinical Genomics, Mayo Clinic, Rochester, MN, 55905, USA

1028 Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, 55905,

1029 USA

1030 **Contribution:** Investigation, Writing – review & editing, Supervision

1031 **Competing interests:** No competing interests declared

1032

1033 **Figure legends**

1034 **Figure 1. Generation of healthy human iPSCs and osteogenic differentiation**

1035 **transcriptomic data.** **(A)** Flowchart of iPSC establishment, iPSC-derived MSC

1036 generation, MSC to OB differentiation (preOBs, preosteoblasts; OBs, osteoblasts), and

1037 RNA-seq data generation and analyses. **(B)** In vitro osteogenic differentiation of MSCs

1038 (Day 0), preOBs (Day 7), and OBs (Day 21) stained with alkaline phosphatase (ALP),

1039 alizarin red, and von Kossa. **(C)** Expression level of known osteogenic genes at the

1040 three osteogenic stages. Data are shown as mean \pm SEM. * MSC vs preOB, # preOB

1041 vs OB, or + MSC vs OB, adjusted p value < 0.0001. **(D)** Principal component analysis

1042 (PCA) of RNA-seq data from our iPSC-derived MSCs and differentiated OBs as well as

1043 previously published human primary MSCs and OBs, iPSCs, and other tissues in GTEx.

1044

1045 **Figure 2. Gene expression profile during osteogenic differentiation.** **(A)** Gene

1046 expression during osteogenic differentiation with the total number of expressed genes

1047 and percentages of expressed coding and noncoding genes (middle). Number and

1048 percentages of differentially expressed (DE) and non-differentially expressed (non-DE)

1049 coding genes (left; TFs and non-TFs (white pie chart)) and of noncoding genes (right).

1050 **(B)**

1051 DEGs during osteogenic differentiation with total and percentages of coding and

1052 noncoding genes (left), and TFs and non-TFs (right). **(C)** Heatmap showing hierarchical

1053 clustering of 60 RNA-seq datasets from 20 iPSC-derived MSC, preOB, and OB lines

1054 (columns) and significant differentially expressed genes (DEGs) (rows), fold change \geq

1055 1.2 and adjusted p value < 0.05. Up-regulated and down-regulated gene expression is

1056 colored in red and blue, respectively. **(D)** and **(E)** Volcano plots illustrate the distribution
1057 of down- and up-regulated genes (blue and red, respectively) with adjusted p values
1058 and fold changes when comparing gene differential expression from MSC to preOB in
1059 **(D)** and preOB to OB stages in **(E)**. Cutoffs of fold change ≥ 1.2 and adjusted p value <
1060 0.05 were applied to define DEGs. The total number of up-regulated and down-
1061 regulated genes are noted at the top (red and blue, respectively). The genes are labeled
1062 for the top five up-regulated (red, bottom right), downregulated (blue, bottom left), and
1063 statistically significant down-regulated (blue, top left) and up-regulated (red, top right).

1064

1065 **Figure 3. TF regulatory network in osteogenic differentiation.** **(A)** Principal
1066 component analysis (PCA) of all 20 healthy cell lines at three stages of osteogenic
1067 differentiation using all differentially expressed genes (DEGs). **(B)** PCA using only
1068 differentially expressed TF genes. **(C)** TF regulatory network during osteogenic
1069 differentiation. Each node represents a TF, with known bone formation associated
1070 regulators underlined in red. Two nodes are connected by a line where ReMap data
1071 suggest regulation and our RNA-seq data suggest the association between them.
1072 Nodes labeled with the gene name represent the top 100 strongest TFs based on
1073 betweenness centrality. The size of the nodes reflects the regulation strength of the TF,
1074 with the top 5 strongest circled in pink. **(D)** and **(E)** Top significantly enriched GO BP
1075 terms in **(D)** and Reactome pathways in **(E)** of TFs in each network module.

1076

1077 **Figure 4. Gene co-expression network in osteogenic differentiation.** **(A)** Sunburst
1078 plots represent the hierarchy structure of the MEGENA co-expression network

1079 constructed on gene expression during the osteogenic differentiation. The structure is
1080 shown as concentric rings where the center ring represents the parent modules, and
1081 outer rings represent smaller child modules. Subnetwork modules are colored according
1082 to the enrichment of differential gene expression between stages (FDR < 0.05; left,
1083 MSC to preOB; right, preOB to OB; blue, down-regulated; red, up-regulated; cyan, both
1084 up- and down-regulated). The subnetwork branch for Module M204 is outlined and
1085 labeled. **(B)** Co-expression network Module M204. Diamonds indicate KNR genes, and
1086 circles indicate non-KNR genes. Blue indicates DEGs from MSC to preOB stages, red
1087 indicates DEGs from preOB to OB stages, and cyan indicates shared DEGs for both
1088 comparisons. Genes known to be related to bone are in red.

1089

1090 **Figure 5. Inhibitory role of *Klf16* in osteogenic differentiation in vitro and in vivo.**
1091 **(A)** The expression of *KLF16* at three human osteogenic differentiation stages. Data are
1092 shown as mean + SEM. * MSC vs. preOB, # preOB vs. OB, or + MSC vs. OB, adjusted
1093 p value < 0.001. **(B)** Analysis of *Klf16* expression by RT-qPCR in MC3T3-E1 cells
1094 transduced vectors containing either stuffer sequence or *Klf16* cDNA. Data are
1095 presented as the mean \pm SEM (n=3, unpaired t-test p value < 0.01) (Supplementary
1096 Table 7). **(C)** Osteogenic differentiation of MC3T3-E1 cells without or with
1097 overexpression of *Klf16*, stained for ALP at Day 7 and alizarin red and von Kossa at
1098 Day 14 and Day 21. **(D)** Length, fat mass, and lean mass of wild type (WT) and *Klf16*^{+/−}
1099 mice. **(E)** DEXA analysis of whole body (head excluded) bone mineral content (BMC),
1100 bone area (B-area), and bone mineral density (BMD) of WT and *Klf16*^{+/−} mice. **(F)** and
1101 **(G)** Representative microCT images of distal femur trabecular bone in **(F)** (*left*, top view;

1102 right, side view) and cortical bone in **(G)** from WT and *Klf16*^{+/−} mice. Scale bar: 1 mm. **(H)**
1103 and **(I)** Graphs show trabecular bone volume/tissue volume (BV/TV), trabecular
1104 thickness (Tb.Th), trabecular number (Tb.N), and trabecular separation (Tb.Sp) in **(H)**;
1105 cortical bone area (Ct.Ar), cortical periosteal perimeter (Ct.Pe.Pm), and cortical
1106 endosteal perimeter (Ct.En.Pm) in **(I)**. For **(D)**, **(E)**, **(H)**, and **(I)**, data are presented using
1107 BoxPlotR (Spitzer et al., 2014) as the mean ± SEM (WT n = 6, *Klf16*^{+/−} n = 6, 3 males
1108 and 3 females for each group, aged 17 weeks, paired t-test, n.s, not significant, * p
1109 value < 0.05, ** p value < 0.01).

1110

1111 **Supplementary Figure 1. Characterization of healthy human iPSCs and iPSC-
1112 derived MSCs. (A)** Identification of stem cell markers: ALP staining, fluorescent
1113 immunocytochemistry (ICC) of pluripotency markers TRA-1-60, SOX2, OCT4, SSEA4,
1114 and NANOG. Scale bar: 100 µm. **(B)** Embryoid body (EB) formation in suspension from
1115 aggregates of iPSCs. ICC of three germ layer markers: AFP, endoderm; α-SMA,
1116 mesoderm; and TUBB3, ectoderm. Scale bar: 50 µm. **(C)** Fluorescence-activated cell
1117 sorting of CD105+/CD45- MSCs. FITC, fluorescein isothiocyanate; PE, phycoerythrin.
1118 **(D)** The percentage of mesenchymal surface marker CD105+/CD45- cells in total cells
1119 differentiated from iPSCs originated from PBMCs and fibroblasts. Data are shown as
1120 mean + SEM. **(E)** and **(F)** Analyses of MSCs after sorting and expansion with flow
1121 cytometry for MSC negative markers (CD31, CD34, and CD45) in **(E)** and MSC positive
1122 markers (CD29, CD73, CD90, and CD105) in **(F)** labeled with different fluorochromes
1123 (V450, APC, PE, and FITC). Marker expression is presented as histograms (green).
1124 Unstained cells were used as controls (gray). Data are shown as mean + SEM. **(G)**

1125 Gene expression of MSC markers in transcripts per kilobase million (TPM). **(H)** Spindle-
1126 like morphology of iPSC-derived MSCs. Scale bar: 50 μ m.

1127

1128 **Supplementary Figure 2. Gene Ontology (GO) enrichment of TF regulatory**
1129 **network modules. (A)** Top enriched GO cellular component (CC) terms of each
1130 module. **(B)** Top enriched GO molecular function (MF) terms of each module.

1131

1132 **Supplementary Figure 3. Gene expression pattern across osteogenic**
1133 **differentiation stages using pseudobulk single-cell RNA-seq data in Housman et**
1134 **al., 2022.** KNR_TF_up: top five up-regulated KNR transcription factors; KNR_TF_down:
1135 top five down-regulated KNR transcription factors; TFs: top five transcription factors
1136 (TFs) based on TF regulatory network; osteogenic_markers: five known osteogenic
1137 markers.

1138

1139 **Supplementary Figure 4. Expression of osteoclastogenesis markers in the femur**
1140 **bone of WT and *Klf16*^{+/−} mice. (A)** IHC for CALCR in the distal femur bone. The
1141 CALCR-positive areas in epiphysis and metaphysis of two *Klf16*^{+/−} mice and two WT
1142 mice were quantified and shown as the percentage of CALCR-positive area
1143 (fluorescence-positive area/total bone area) in **(B).** **(C)** IHC for RANKL in the distal
1144 femur bone. The percentage of RANKL-positive area were quantified and shown in **(D).**
1145 For **(B)** and **(D)**, three replicate sections from each animal were analyzed. Data are
1146 presented as the mean \pm SEM and visualized with GraphPad Prism 10. n.s: not
1147 significant. * p value < 0.05 by unpaired t-test. Scale bar: 500 μ m.

1148

1149 **Supplementary Source Data for Figure 5, Supplementary Figure 1, and**
1150 **Supplementary Figure 4.**

1151

1152 **Supplementary Table 1**, related to Figure 1. Sample demographic metadata.

1153

1154 **Supplementary Table 2**, related to Figure 2. Differential gene expression between
1155 stages.

1156

1157 **Supplementary Table 3**, related to Figure 3. and Supplementary Figure 2.

1158 Transcription factor regulatory networks and gene ontology enrichments.

1159

1160 **Supplementary Table 4**, related to Figure 4. Significant MEGENA modules.

1161

1162 **Supplementary Table 5**, related to Supplementary Figure 3. Pseudobulk gene
1163 expression at osteogenic differentiation stages using single-cell RNA-seq data in
1164 Housman et al., 2022.

1165

1166 **Supplementary Table 6**, related to Supplementary Figure 3. Gene expression in five
1167 cell types at Day 21 of osteogenic differentiation using single-cell RNA-seq data in
1168 Housman et al., 2022.

1169

1170 **Supplementary Table 7**, related to Figure 5. Primers used for *Klf16* in vitro
1171 overexpression study.

1172

1173 **References**

- 1174 Al-Rekabi, Z., Wheeler, M. M., Leonard, A., Fura, A. M., Juhlin, I., Frazar, C., Smith, J.
1175 D., Park, S. S., Gustafson, J. A., Clarke, C. M., et al. (2016). Activation of the
1176 IGF1 pathway mediates changes in cellular contractility and motility in single-
1177 suture craniosynostosis. *Journal of Cell Science* 129, 483-491.
- 1178 Ardlie, K. G., Deluca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T.,
1179 Trowbridge, C. A., Maller, J. B., Tukiainen, T., Lek, M., et al. (2015). The
1180 Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in
1181 humans. *Science* 348, 648-660.
- 1182 Aza-Carmona, M., Barca-Tierno, V., Hisado-Oliva, A., Belinchón, A., Gorbenko-del
1183 Blanco, D., Rodriguez, J. I., Benito-Sanz, S., Campos-Barros, A. and Heath, K. E.
1184 (2014). NPPB and ACAN, two novel SHOX2 transcription targets implicated in
1185 skeletal development. *PLoS One* 9, e83104.
- 1186 Barberi, T., Willis, L. M., Soccia, N. D. and Studer, L. (2005). Derivation of multipotent
1187 mesenchymal precursors from human embryonic stem cells. *PLoS Medicine* 2,
1188 e161.
- 1189 Bastian, M., Heymann, S. and Jacomy, M. (2009). Gephi: an open source software for
1190 exploring and manipulating networks. *Proceedings of the International AAAI
1191 Conference on Web and Social Media* 3, 361-362.
- 1192 Benisch, P., Schilling, T., Klein-Hitpass, L., Frey, S. P., Seefried, L., Raaijmakers, N.,
1193 Krug, M., Regensburger, M., Zeck, S., Schinke, T., et al. (2012). The
1194 transcriptional profile of mesenchymal stem cell populations in primary
1195 osteoporosis is distinct and shows overexpression of osteogenic inhibitors. *PLoS
1196 One* 7, e45142.
- 1197 Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical
1198 and powerful approach to multiple testing. *Journal of the Royal Statistical Society:
1199 Series B (Methodological)* 57, 289-300.
- 1200 Boyce, B. F. (2013). Advances in the regulation of osteoclasts and osteoclast functions.
1201 *Journal of Dental Research* 92, 860-867.
- 1202 Boyle, W. J., Simonet, W. S. and Lacey, D. L. (2003). Osteoclast differentiation and
1203 activation. *Nature* 423, 337-342.
- 1204 Bruder, S. P., Fink, D. J. and Caplan, A. I. (1994). Mesenchymal stem cells in bone
1205 development, bone repair, and skeletal regeneration therapy. *Journal of Cellular
1206 Biochemistry* 56, 283-294.
- 1207 Camacho, D., de la Fuente, A. and Mendes, P. (2005). The origin of correlations in
1208 metabolomics data. *Metabolomics* 1, 53-63.
- 1209 Carcamo-Orive, I., Hoffman, G. E., Cundiff, P., Chang, R., Quertermous, T., Carcamo-
1210 orive, I., Hoffman, G. E., Cundiff, P., Beckmann, N. D. and Souza, S. L. D. (2017).
1211 Analysis of transcriptional variability in a large human iPSC library reveals

- 1212 genetic and non- genetic determinants of heterogeneity resource analysis of
1213 transcriptional variability in a large human iPSC library reveals genetic and non-
1214 genetic determinant. *Stem Cell* 20, 518-531.
- 1215 Chen, X., Jiao, J., He, X., Zhang, J., Wang, H., Xu, Y. and Jin, T. (2017). CHI3L1
1216 regulation of inflammation and the effects on osteogenesis in a *Staphylococcus*
1217 aureus-induced murine model of osteomyelitis. *The FEBS Journal* 284, 1738-
1218 1747.
- 1219 Chèneby, J., Ménétrier, Z., Mestdagh, M., Rosnet, T., Douida, A., Rhalloussi, W.,
1220 Bergon, A., Lopez, F. and Ballester, B. (2020). ReMap 2020: a database of
1221 regulatory regions from an integrative analysis of Human and *Arabidopsis* DNA-
1222 binding sequencing experiments. *Nucleic Acids Research* 48, 180-188.
- 1223 Daftary, G. S., Lomberk, G. A., Buttar, N. S., Allen, T. W., Grzenda, A., Zhang, J.,
1224 Zheng, Y., Mathison, A. J., Gada, R. P., Calvo, E., et al. (2012). Detailed
1225 structural-functional analysis of the Krüppel-like factor 16 (KLF16) transcription
1226 factor reveals novel mechanisms for silencing Sp/KLF sites involved in
1227 metabolism and endocrinology. *Journal of Biological Chemistry* 287, 7010-7025.
- 1228 Daley, G. Q., Lensch, M. W., Jaenisch, R., Meissner, A., Plath, K. and Yamanaka, S.
1229 (2009). Broader implications of defining standards for the pluripotency of iPSCs.
1230 *Cell Stem Cell* 4, 200-201.
- 1231 DeMars, L. J., Stephens, N. B., Saers, J. P., Gordon, A., Stock, J. T., and Ryan, T. M.
1232 (2021). Using point clouds to investigate the relationship between trabecular
1233 bone phenotype and behavior: An example utilizing the human calcaneus.
1234 *American Journal of Human Biology* 33, e23468.
- 1235 Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T. and Müller, T. (2008).
1236 Identifying functional modules in protein-protein interaction networks: an
1237 integrated exact approach. *Bioinformatics* 24, i223-231.
- 1238 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
1239 Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq
1240 aligner. *Bioinformatics* 29, 15-21.
- 1241 Doube, M., Kłosowski, M. M., Arganda-Carreras, I., Cordelières, F., Dougherty, R. P.,
1242 Jackson, J., Schmid, B., Hutchinson, J. R., Shefelbine, S. J. (2010). BoneJ: free
1243 and extensible bone image analysis in ImageJ. *Bone* 47, 1076-1079.
- 1244 Dubey, N., Hoffman, J. F., Schuebel, K., Yuan, Q., Martinez, P. E., Nieman, L. K.,
1245 Rubinow, D. R., Schmidt, P. J. and Goldman, D. (2017). The ESC/E(Z) complex,
1246 an effector of response to ovarian steroids, manifests an intrinsic difference in
1247 cells from women with premenstrual dysphoric disorder. *Molecular Psychiatry* 22,
1248 1172-1184.
- 1249 Fu, R., Lv, W. C., Xu, Y., Gong, M. Y., Chen, X. J., Jiang, N., Xu, Y., Yao, Q. Q., Di, L.,
1250 Lu, T., et al. (2020). Endothelial ZEB1 promotes angiogenesis-dependent bone
1251 formation and reverses osteoporosis. *Nature Communications* 11, 460.
- 1252 Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X.
1253 J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the
1254 human regulatory network derived from ENCODE data. *Nature* 489, 91-100.
- 1255 Giuliani, M., Oudrhiri, N., Noman, Z. M., Vernochet, A., Chouaib, S., Azzarone, B.,
1256 Durrbach, A. and Bennaceur-Griscelli, A. (2011). Human mesenchymal stem

- 1257 cells derived from induced pluripotent stem cells down-regulate NK-cell cytolytic
1258 machinery. *Blood* 118, 3254-3262.
- 1259 Gross, T., Kivell, T. L., Skinner, M. M., Nguyen, H., and Pahr, D. H. (2014). A CT-
1260 image-based framework for the holistic analysis of cortical and trabecular bone
1261 morphology. *Palaeontologia Electronica* 17, 13.
- 1262 Han, Y., Li, X., Zhang, Y., Han, Y., Chang, F. and Ding, J. (2019). Mesenchymal stem
1263 cells for regenerative medicine. *Cells* 8, 886.
- 1264 Hao, Y., Wu, W., Shi, F., Dalmolin, R. J., Yan, M., Tian, F., Chen, X., Chen, G. and Cao,
1265 W. (2015). Prediction of long noncoding RNA functions with co-expression
1266 network in esophageal squamous cell carcinoma. *BMC Cancer* 15, 168.
- 1267 Hernández-Vega, A. M. and Camacho-Arroyo, I. (2021). Crosstalk between 17 β -
1268 estradiol and TGF- β signaling modulates glioblastoma progression. *Brain
1269 Sciences* 11, 564.
- 1270 Holmes, G., Gonzalez-Reiche, A. S., Lu, N., Zhou, X., Rivera, J., Kriti, D., Sebra, R.,
1271 Williams, A. A., Donovan, M. J., Potter, S. S., et al. (2020). Integrated
1272 transcriptome and network analysis reveals spatiotemporal dynamics of calvarial
1273 suturogenesis. *Cell Reports* 32, 107871-107871.
- 1274 Housman, G., Briscoe, E. and Gilad, Y. (2022). Evolutionary insights into primate
1275 skeletal gene regulation using a comparative cell culture model. *PLOS Genetics*
1276 18, e1010073-e1010073.
- 1277 Hsu, Y. H., Estrada, K., Evangelou, E., Ackert-Bicknell, C., Akesson, K., Beck, T.,
1278 Brown, S. J., Capellini, T., Carbone, L., Cauley, J., et al. (2019). Meta-analysis of
1279 genomewide association studies reveals genetic variants for hip bone geometry.
1280 *Journal of Bone and Mineral Research* 34, 1284-1296.
- 1281 Hwang, C. K., D'Souza, U. M., Eisch, A. J., Yajima, S., Lammers, C. H., Yang, Y., Lee,
1282 S. H., Kim, Y. M., Nestler, E. J. and Mouradian, M. M. (2001). Dopamine receptor
1283 regulating factor, DRRF: a zinc finger transcription factor. *Proceedings of the
1284 National Academy of Sciences of the United States of America* 98, 7558-7563.
- 1285 Jiang, J., Jia, P., Zhao, Z. and Shen, B. (2014). Key regulators in prostate cancer
1286 identified by co-expression module analysis. *BMC Genomics* 15, 1015.
- 1287 Julien, A., Kanagalingam, A., Martínez-Sarrà, E., Megret, J., Luka, M., Ménager, M.,
1288 Relaix, F. and Colnot, C. (2021). Direct contribution of skeletal muscle
1289 mesenchymal progenitors to bone repair. *Nature Communications* 12, 2860.
- 1290 Kaczynski, J., Cook, T. and Urrutia, R. (2003). Sp1- and Krüppel-like transcription
1291 factors. *Genome Biol* 4, 206.
- 1292 Kang, R., Zhou, Y., Tan, S., Zhou, G., Aagaard, L., Xie, L., Bünger, C., Bolund, L. and
1293 Luo, Y. (2015). Mesenchymal stem cells derived from human induced pluripotent
1294 stem cells retain adequate osteogenicity and chondrogenicity but less
1295 adipogenicity. *Stem Cell Research & Therapy* 6, 144.
- 1296 Khosla, S. (2001). Minireview: The OPG/RANKL/RANK System. *Endocrinology* 142,
1297 5050-5055.
- 1298 Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y.,
1299 Bronson, R. T., Gao, Y. H., Inada, M., et al. (1997). Targeted disruption of Cbfa1
1300 results in a complete lack of bone formation owing to maturational arrest of
1301 osteoblasts. *Cell* 89, 755-764.

- 1302 Krum, S. A., Chang, J., Miranda-Carboni, G. and Wang, C. Y. (2010). Novel functions
1303 for NF κ B: inhibition of bone formation. *Nature Reviews Rheumatology* 6, 607-611.
1304 Kukita, A., Bonewald, L., Rosen, D., Seyedin, S., Mundy, G. R. and Roodman, G. D.
1305 (1990). Osteoinductive factor inhibits formation of human osteoclast-like cells.
1306 *Proceedings of the National Academy of Sciences of the United States of*
1307 *America* 87, 3023-3026.
1308 Kushioka, J., Kaito, T., Okada, R., Ishiguro, H., Bal, Z., Kodama, J., Chijimatsu, R., Pye,
1309 M., Narimatsu, M., Wrana, J. L., et al. (2020). A novel negative regulatory
1310 mechanism of Smurf2 in BMP/Smad signaling in bone. *Bone Research* 8, 41.
1311 Lamande, S. R., Ng, E. S., Cameron, T. L., Kung, L. H. W., Sampurno, L., Rowley, L.,
1312 Lilianty, J., Patria, Y. N., Stenta, T., Hanssen, E., et al. (2023). Modeling human
1313 skeletal development using human pluripotent stem cells. *Proceedings of the*
1314 *National Academy of Sciences of the United States of America* 120,
1315 e2211510120.
1316 Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X.,
1317 Taipale, J., Hughes, T. R. and Weirauch, M. T. (2018). The human transcription
1318 factors. *Cell* 175, 598-599.
1319 Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. (2014). voom: precision weights unlock
1320 linear model analysis tools for RNA-seq read counts. *Genome Biology* 15, R29-
1321 R29.
1322 Le Mée, S., Fromigué, O. and Marie, P. J. (2005). Sp1/Sp3 and the myeloid zinc finger
1323 gene MZF1 regulate the human N-cadherin promoter in osteoblasts.
1324 *Experimental Cell Research* 302, 129-142.
1325 Lee, D.-F., Su, J., Kim, H. S., Chang, B., Papatsenko, D., Zhao, R., Yuan, Y., Gingold,
1326 J., Xia, W., Darr, H., et al. (2015). Modeling familial cancer with induced
1327 pluripotent stem cells. *Cell* 161, 240-254.
1328 Lewis, R. M., Cleal, J. K., Ntani, G., Crozier, S. R., Mahon, P. A., Robinson, S. M.,
1329 Harvey, N. C., Cooper, C., Inskip, H. M., Godfrey, K. M., et al. (2012).
1330 Relationship between placental expression of the imprinted PHLDA2 gene,
1331 intrauterine skeletal growth and childhood bone mass. *Bone* 50, 337-342.
1332 Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq
1333 data with or without a reference genome. *BMC bioinformatics* 12, 323-323.
1334 Li, L., Wang, H., Chen, X., Li, X., Wang, G., Jie, Z., Zhao, X., Sun, X., Huang, H., Fan,
1335 S., et al. (2021). Oxidative stress-induced hypermethylation of KLF5 promoter
1336 mediated by DNMT3B impairs osteogenesis by diminishing the interaction with β -
1337 Catenin. *Antioxid Redox Signal* 35, 1-20.
1338 Li, X. and Wang, C. (2021). From bulk, single-cell to spatial RNA sequencing.
1339 *International Journal of Oral Science* 13, 36.
1340 Liao, Y., Smyth, G. K. and Shi, W. (2014). featureCounts: an efficient general purpose
1341 program for assigning sequence reads to genomic features. *Bioinformatics* 30,
1342 923-930.
1343 Lin, Y. and Chen, G. (2008). Embryoid body formation from human pluripotent stem
1344 cells in chemically defined E8 media. In *StemBook*. Cambridge (MA): Harvard
1345 Stem Cell Institute.

- 1346 Liu, P., Ping, Y., Ma, M., Zhang, D., Liu, C., Zaidi, S., Gao, S., Ji, Y., Lou, F., Yu, F., et
1347 al. (2016). Anabolic actions of Notch on mature bone. *Proceedings of the*
1348 *National Academy of Sciences of the United States of America* 113, E2152-2161.
1349 Liu, Y., Li, Y., Liu, X. and Wang, C. S. (2019). Investigation of transcriptome mechanism
1350 associated with osteoporosis explored by microarray analysis. *Experimental and*
1351 *Therapeutic Medicine* 17, 3459-3464.
1352 Ma, J., Wu, J., Han, L., Jiang, X., Yan, L., Hao, J. and Wang, H. (2019). Comparative
1353 analysis of mesenchymal stem cells derived from amniotic membrane, umbilical
1354 cord, and chorionic plate under serum-free condition. *Stem Cell Research &*
1355 *Therapy* 10, 19.
1356 Manikandan, M., Abuelreich, S., Elsafadi, M., Alsalman, H., Almalak, H., Siyal, A.,
1357 Hashmi, J. A., Aldahmash, A., Kassem, M., Alfayez, M., et al. (2018). NR2F1
1358 mediated down-regulation of osteoblast differentiation was rescued by bone
1359 morphogenetic protein-2 (BMP-2) in human MSC. *Differentiation* 104, 36-41.
1360 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
1361 sequencing reads. *EMBnet.journal* 17, 10-12.
1362 Matsuda, M., Yamanaka, Y., Uemura, M., Osawa, M., Saito, M. K., Nagahashi, A.,
1363 Nishio, M., Guo, L., Ikegawa, S., Sakurai, S., et al. (2020). Recapitulating the
1364 human segmentation clock with pluripotent stem cells. *Nature* 580, 124-129.
1365 Mi, H., Muruganujan, A., Ebert, D., Huang, X. and Thomas, P. D. (2019a). PANTHER
1366 version 14: more genomes, a new PANTHER GO-slim and improvements in
1367 enrichment analysis tools. *Nucleic Acids Res* 47, D419-D426.
1368 Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X. and Thomas, P. D.
1369 (2019b). Protocol Update for large-scale genome and gene function analysis with
1370 the PANTHER classification system (v.14.0). *Nature Protocols* 14, 703-721.
1371 Morris, J. A., Kemp, J. P., Youlten, S. E., Laurent, L., Logan, J. G., Chai, R. C.,
1372 Vulpescu, N. A., Forgetta, V., Kleinman, A., Mohanty, S. T., et al. (2019). An
1373 atlas of genetic influences on osteoporosis in humans and mice. *Nature Genetics*
1374 51, 258-266.
1375 Nakajima, T., Shibata, M., Nishio, M., Nagata, S., Alev, C., Sakurai, H., Toguchida, J.
1376 and Ikeya, M. (2018). Modeling human somite development and fibrodysplasia
1377 ossificans progressiva with induced pluripotent stem cells. *Development* 145(16),
1378 dev165431.
1379 Nakashima, K., Zhou, X., Kunkel, G., Zhang, Z., Deng, J. M., Behringer, R. R. and de
1380 Crombrugghe, B. (2002). The novel zinc finger-containing transcription factor
1381 osterix is required for osteoblast differentiation and bone formation. *Cell* 108, 17-
1382 29.
1383 Nicholson, G. C., Moseley, J. M., Sexton, P. M., Mendelsohn, F. A. and Martin, T. J.
1384 (1986). Abundant calcitonin receptors in isolated rat osteoclasts. Biochemical
1385 and autoradiographic characterization. *Journal of Clinical Investigation* 78, 355-
1386 360.
1387 Onizuka, S., Iwata, T., Park, S. J., Nakai, K., Yamato, M., Okano, T. and Izumi, Y.
1388 (2016). ZBTB16 as a downstream target gene of osterix regulates
1389 osteoblastogenesis of human multipotent mesenchymal stromal cells. *Journal of*
1390 *Cellular Biochemistry* 117, 2423-2434.

- 1391 Otasek, D., Morris, J. H., Bouças, J., Pico, A. R. and Demchak, B. (2019). Cytoscape
1392 Automation: empowering workflow-based network analysis. *Genome Biology* 20,
1393 185.
- 1394 Oury, F. (2012). A crosstalk between bone and gonads. *Annals of the New York
1395 Academy of Sciences* 1260, 1-7.
- 1396 Parfitt, A. M., Drezner, M. K., Glorieux, F. H., Kanis, J. A., Malluche, H., Meunier, P. J.,
1397 Ott, S. M. and Recker, R. R. (1987). Bone histomorphometry: standardization of
1398 nomenclature, symbols, and units. Report of the ASBMR Histomorphometry
1399 Nomenclature Committee. *Journal of Bone and Mineral Research* 2, 595-610.
- 1400 Pittenger, M. F., Mackay, A. M., Beck, S. C., Jaiswal, R. K., Douglas, R., Mosca, J. D.,
1401 Moorman, M. A., Simonetti, D. W., Craig, S. and Marshak, D. R. (1999).
1402 Multilineage potential of adult human mesenchymal stem cells. *Science* 284,
1403 143-147.
- 1404 Polo, J. M., Liu, S., Figueroa, M. E., Kulalert, W., Eminli, S., Tan, K. Y., Apostolou, E.,
1405 Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the
1406 molecular and functional properties of mouse induced pluripotent stem cells.
1407 *Nature Biotechnology* 28, 848-855.
- 1408 Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing
1409 genomic features. *Bioinformatics* 26, 841-842.
- 1410 Rauch, A., Haakonsson, A. K., Madsen, J. G. S., Larsen, M., Forss, I., Madsen, M. R.,
1411 Van Hauwaert, E. L., Wiwie, C., Jespersen, N. Z., Tencerova, M., et al. (2019).
1412 Osteogenesis depends on commissioning of a network of stem cell transcription
1413 factors that act as repressors of adipogenesis. *Nature Genetics* 51, 716-727.
- 1414 Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015).
1415 limma powers differential expression analyses for RNA-sequencing and
1416 microarray studies. *Nucleic Acids Research* 43, e47.
- 1417 Roforth, M. M., Farr, J. N., Fujita, K., McCready, L. K., Atkinson, E. J., Therneau, T. M.,
1418 Cunningham, J. M., Drake, M. T., Monroe, D. G. and Khosla, S. (2015). Global
1419 transcriptional profiling using RNA sequencing and DNA methylation patterns in
1420 highly enriched mesenchymal cells from young versus elderly women. *Bone* 76,
1421 49-57.
- 1422 Rojas-Peña, M. L., Olivares-Navarrete, R., Hyzy, S., Arafat, D., Schwartz, Z., Boyan, B.
1423 D., Williams, J. and Gibson, G. (2014). Characterization of distinct classes of
1424 differential gene expression in osteoblast cultures from non-syndromic
1425 craniosynostosis bone. *Journal of Genomics* 2, 121-130.
- 1426 Schaniel, C., Dhanan, P., Hu, B., Xiong, Y., Raghunandan, T., Gonzalez, D. M., Dariolli,
1427 R., D'Souza, S. L., Yadaw, A. S., Hansen, J., et al. (2021). A library of induced
1428 pluripotent stem cells from clinically well-characterized, diverse healthy human
1429 individuals. *Stem Cell Reports* 16, 3036-3049.
- 1430 Schroeder, T. M., Jensen, E. D. and Westendorf, J. J. (2005). Runx2: A master
1431 organizer of gene transcription in developing and maturing osteoblasts. *Birth
1432 Defects Research Part C: Embryo Today: Reviews* 75, 213-225.
- 1433 Scotto di Carlo, F., Pazzaglia, L., Esposito, T. and Gianfrancesco, F. (2020). The loss of
1434 profilin 1 causes early onset paget's disease of bone. *Journal of Bone and
1435 Mineral Research* 35, 1387-1398.

- 1436 Seifert, A., Werheid, D. F., Knapp, S. M. and Tobiasch, E. (2015). Role of Hox genes in
1437 stem cell differentiation. *World J Stem Cells* 7, 583-595.
- 1438 Shinoda, Y., Ogata, N., Higashikawa, A., Manabe, I., Shindo, T., Yamada, T., Kugimiya,
1439 F., Ikeda, T., Kawamura, N., Kawasaki, Y., et al. (2008). Kruppel-like factor 5
1440 causes cartilage degradation through transactivation of matrix metalloproteinase
1441 9. *Journal of Biological Chemistry* 283, 24682-24689.
- 1442 Song, W. M. and Zhang, B. (2015). Multiscale embedded gene co-expression network
1443 analysis. *PLOS Computational Biology* 11, e1004574-e1004574.
- 1444 Song, Z., Lian, X., Wang, Y., Xiang, Y. and Li, G. (2017). KLF15 regulates in vitro
1445 chondrogenic differentiation of human mesenchymal stem cells by targeting
1446 SOX9. *Biochemical and Biophysical Research Communications* 493, 1082-1088.
- 1447 Sotiropoulou, P. A., Perez, S. A., Salagianni, M., Baxevanis, C. N. and Papamichail, M.
1448 (2006). Characterization of the optimal culture conditions for clinical scale
1449 production of human mesenchymal stem cells. *Stem Cells* 24, 462-471.
- 1450 Spitzer, M., Wildenhain, J., Rappsilber, J. and Tyers, M. (2014). BoxPlotR: a web tool
1451 for generation of box plots. *Nature Methods* 11, 121-122.
- 1452 Styrkarsdottir, U., Stefansson, O. A., Gunnarsdottir, K., Thorleifsson, G., Lund, S. H.,
1453 Stefansdottir, L., Juliusson, K., Agustsdottir, A. B., Zink, F., Halldorsson, G. H., et
1454 al. (2019). GWAS of bone size yields twelve loci that also affect height, BMD,
1455 osteoarthritis or fractures. *Nature Communications* 10, 2054.
- 1456 Sun, N., Shen, C., Zhang, L., Wu, X., Yu, Y., Yang, X., Yang, C., Zhong, C., Gao, Z.,
1457 Miao, W., et al. (2021). Hepatic Krüppel-like factor 16 (KLF16) targets PPARα to
1458 improve steatohepatitis and insulin resistance. *Gut* 70, 2183-2195.
- 1459 Suo, J., Feng, X., Li, J., Wang, J., Wang, Z., Zhang, L. and Zou, W. (2020). VGLL4
1460 promotes osteoblast differentiation by antagonizing TEADs-inhibited Runx2
1461 transcription. *Science Advances* 6(43), eaba4147.
- 1462 Tani, S., Okada, H., Onodera, S., Chijimatsu, R., Seki, M., Suzuki, Y., Xin, X., Rowe, D.
1463 W., Saito, T., Tanaka, S., et al. (2023). Stem cell-based modeling and single-cell
1464 multiomics reveal gene-regulatory mechanisms underlying human skeletal
1465 development. *Cell Reports* 42, 112276.
- 1466 The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and
1467 still GOing strong. *Nucleic Acids Research* 47, D330-D338.
- 1468 Thomas, S. and Jaganathan, B. G. (2022). Signaling network regulating osteogenesis in
1469 mesenchymal stem cells. *J Cell Commun Signal* 16, 47-61.
- 1470 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D.,
1471 Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0:
1472 fundamental algorithms for scientific computing in Python. *Nature Methods* 17,
1473 261-272.
- 1474 Wang, J., Galvao, J., Beach, K. M., Luo, W., Urrutia, R. A., Goldberg, J. L. and Otteson,
1475 D. C. (2016). Novel roles and mechanism for Krüppel-like factor 16 (KLF16)
1476 regulation of neurite outgrowth and ephrin receptor A5 (EphA5) expression in
1477 retinal ganglion cells. *Journal of Biological Chemistry* 291, 18084-18095.
- 1478 Wilkinson, A. C., Nakauchi, H. and Göttgens, B. (2017). Mammalian transcription factor
1479 networks: recent advances in interrogating biological complexity. *Cell Systems* 5,
1480 319-331.

- 1481 Wu, X., Wang, X., Shan, L., Zhou, J., Zhang, X., Zhu, E., Yuan, H. and Wang, B. (2021).
1482 High-mobility group AT-Hook 1 mediates the role of nuclear factor I/X in
1483 osteogenic differentiation through activating canonical Wnt signaling. *Stem Cells*
1484 39, 1349-1361.
1485 Yaffe, M. P., Noggle, S. A. and Solomon, S. L. (2016). Raising the standards of stem
1486 cell line quality. *Nature Cell Biology* 18, 236-237.
1487 Yang, H., Li, G., Han, N., Zhang, X., Cao, Y., Cao, Y. and Fan, Z. (2020). Secreted
1488 frizzled-related protein 2 promotes the osteo/odontogenic differentiation and
1489 paracrine potentials of stem cells from apical papilla under inflammation and
1490 hypoxia conditions. *Cell Proliferation* 53, e12694.
1491 Yang, X., Matsuda, K., Bialek, P., Jacquot, S., Masuoka, H. C., Schinke, T., Li, L.,
1492 Brancorsini, S., Sassone-Corsi, P., Townes, T. M., et al. (2004). ATF4 Is a
1493 substrate of RSK2 and an essential regulator of osteoblast biology: implication
1494 for Coffin-Lowry syndrome. *Cell* 117, 387-398.
1495 Yazdani, A., Sun, Y. C., Stephens, N. B., Ryan, T., Monga, V. (2020). Multi-class micro-
1496 CT image segmentation using sparse regularized deep networks. *Conference
1497 Record of the 54th Asilomar Conference on Signals, Systems and Computers,
1498 ACSSC 2020 (IEEE Computer Society)*, 1553–1557.
1499 Yu, F., Wu, F., Li, F., Liao, X., Wang, Y., Li, X., Wang, C., Shi, Y. and Ye, L. (2020).
1500 Wnt7b-induced Sox11 functions enhance self-renewal and osteogenic
1501 commitment of bone marrow mesenchymal stem cells. *Stem Cells* 38, 1020-1033.
1502 Yu, S., Guo, J., Sun, Z., Lin, C., Tao, H., Zhang, Q., Cui, Y., Zuo, H., Lin, Y., Chen, S.,
1503 et al. (2021). BMP2-dependent gene regulatory network analysis reveals Klf4 as
1504 a novel transcription factor of osteoblast differentiation. *Cell Death & Disease* 12,
1505 197.
1506 Zakeri, S., Aminian, H., Sadeghi, S., Esmaeilzadeh-Gharehdaghi, E. and Razmara, E.
1507 (2022). Krüppel-like factors in bone biology. *Cell Signal* 93, 110308.
1508 Zhang, B., Gaiteri, C., Bodea, L. G., Wang, Z., McElwee, J., Podtelezhnikov, A. A.,
1509 Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems
1510 approach identifies genetic nodes and networks in late-onset Alzheimer's disease.
1511 *Cell* 153, 707-720.
1512 Zhang, B. and Zhu, J. (2013). Identification of key causal regulators in gene networks.
1513 In *Proceedings of the World Congress on Engineering*, 5-8.
1514 Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S.
1515 B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al. (2017). Massively parallel
1516 digital transcriptional profiling of single cells. *Nature Communications* 8, 14049.
1517 Zheng, H., Ramnaraign, D., Anderson, B. A., Tycksen, E., Nunley, R. and McAlinden, A.
1518 (2019). MicroRNA-138 inhibits osteogenic differentiation and mineralization of
1519 human dedifferentiated chondrocytes by regulating RhoC and the actin
1520 cytoskeleton. *JBMR Plus* 3, e10071.
1521 Zhu, H., Wang, Q., Yao, Y., Fang, J., Sun, F., Ni, Y., Shen, Y., Wang, H. and Shao, S.
1522 (2015). Microarray analysis of Long non-coding RNA expression profiles in
1523 human gastric cells and tissues with Helicobacter pylori Infection. *BMC Medical
1524 Genomics* 8, 84.

- 1525 Zhu, X., Wang, F., Zhao, Y., Yang, P., Chen, J., Sun, H., Liu, L., Li, W., Pan, L., Guo, Y.,
1526 et al. (2014). A gain-of-function mutation in *Tnni2* impeded bone development
1527 through increasing *Hif3a* expression in DA2B mice. *PLoS Genetics* 10, e1004589.

Fig. 1

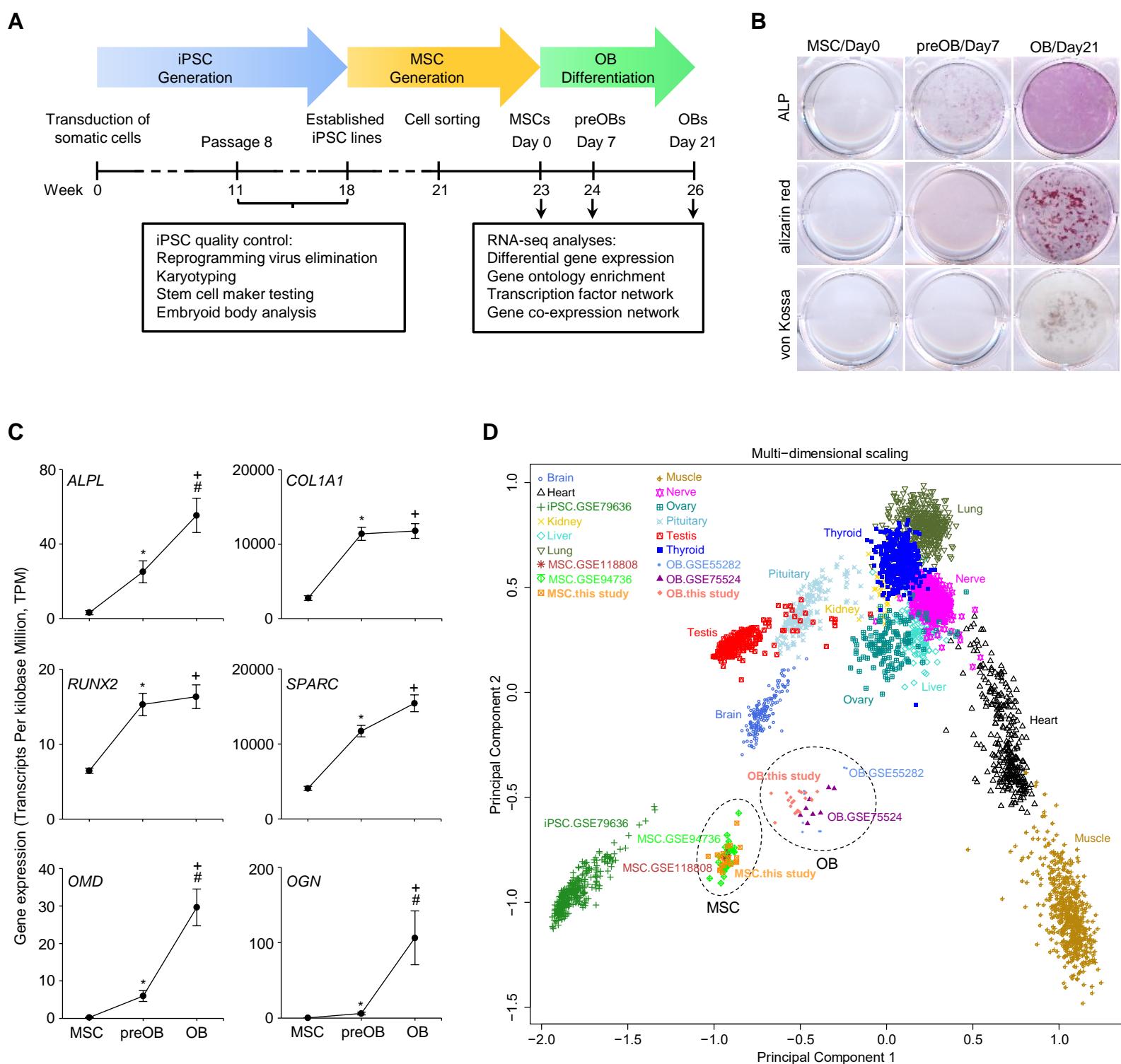


Fig. 2

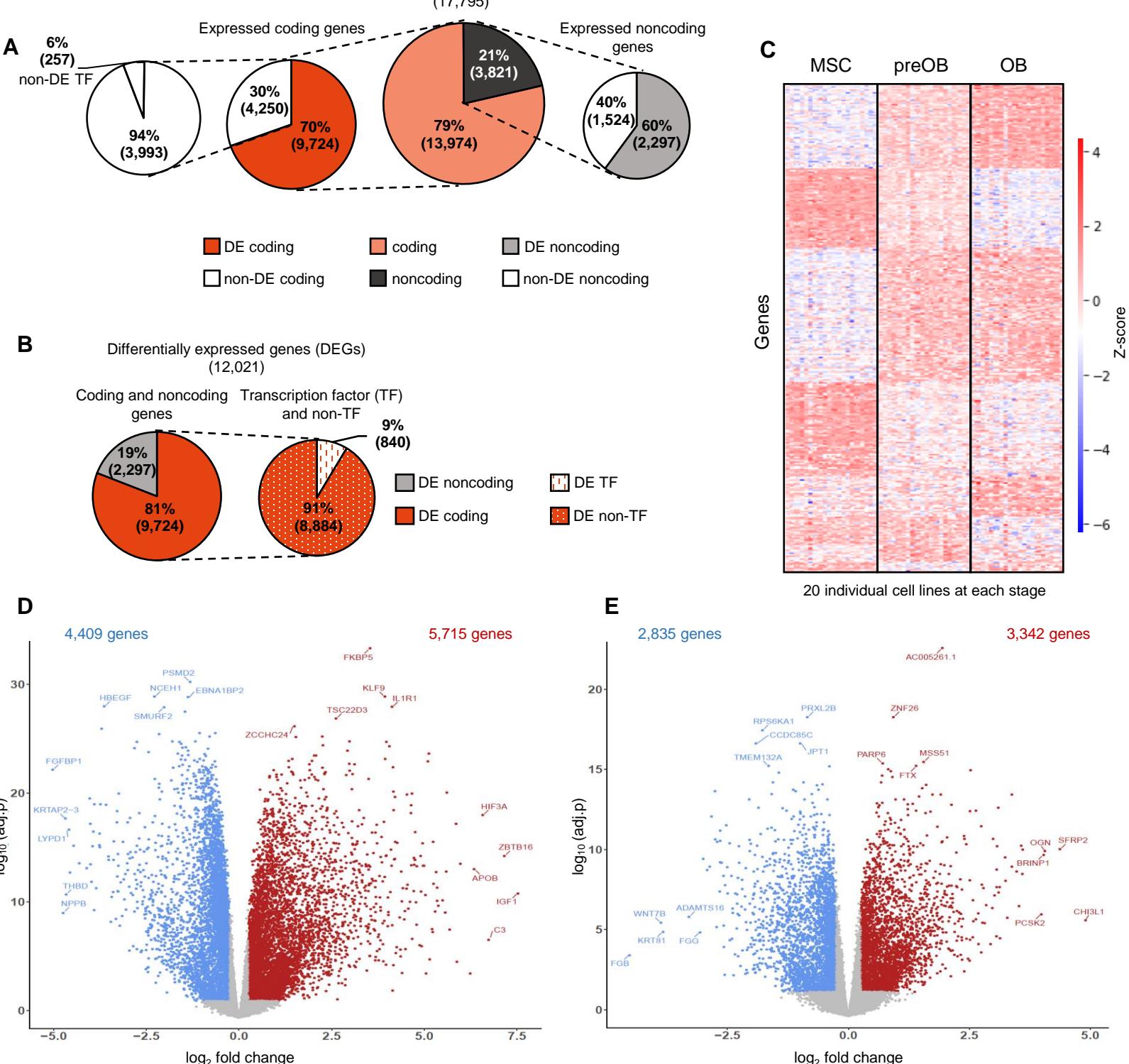


Fig. 3

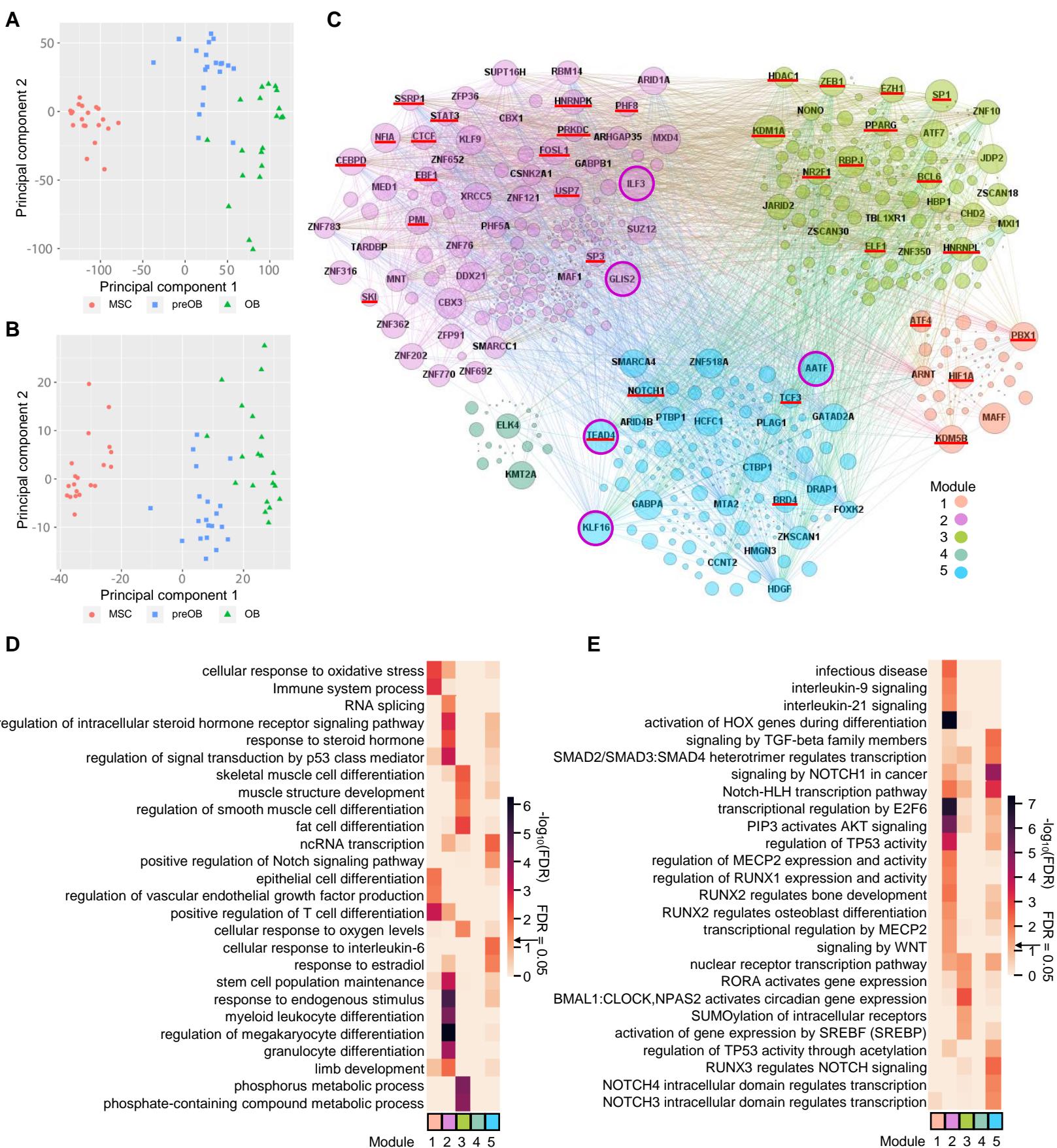


Fig. 4

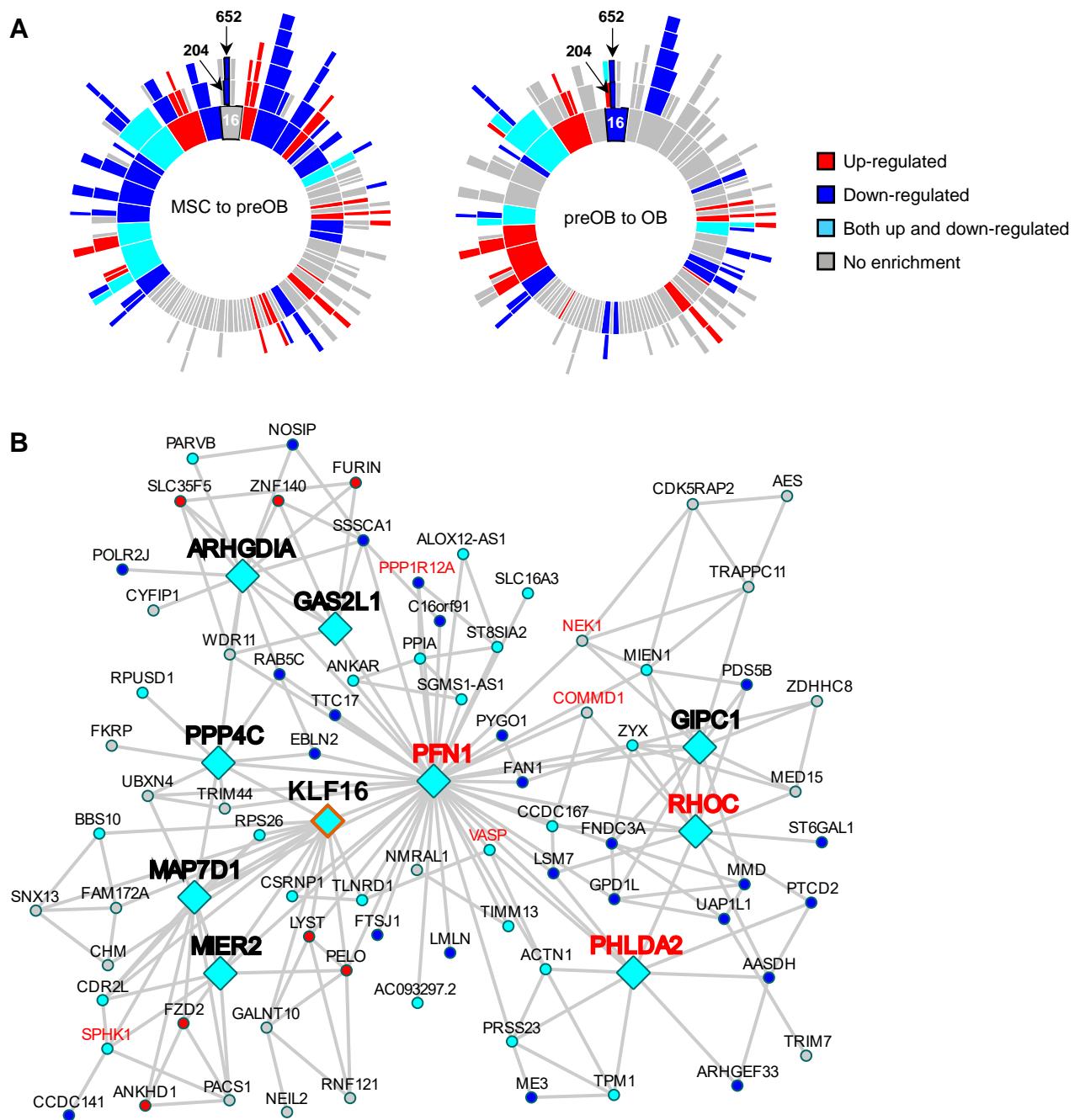
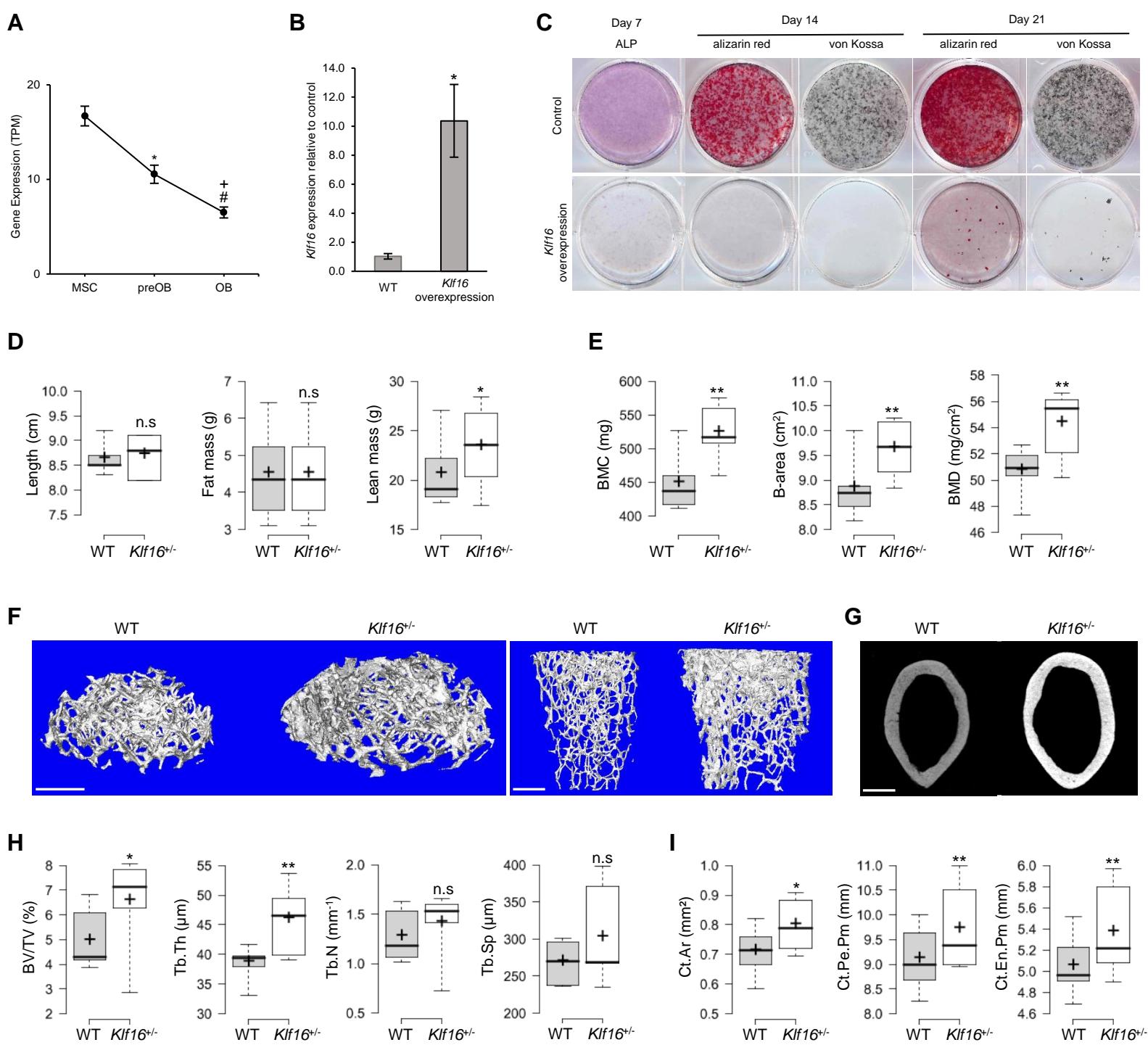
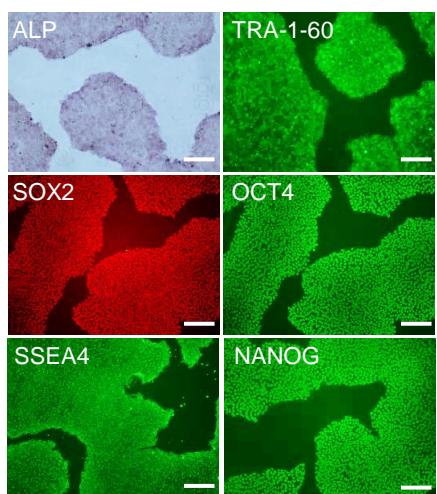


Fig. 5

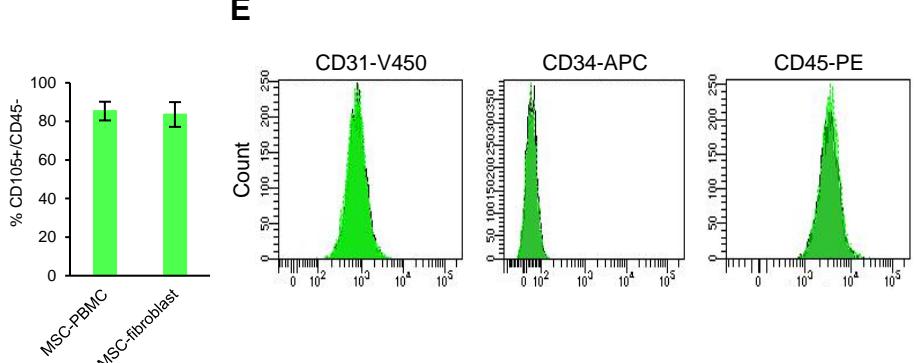


Supplementary Fig. 1

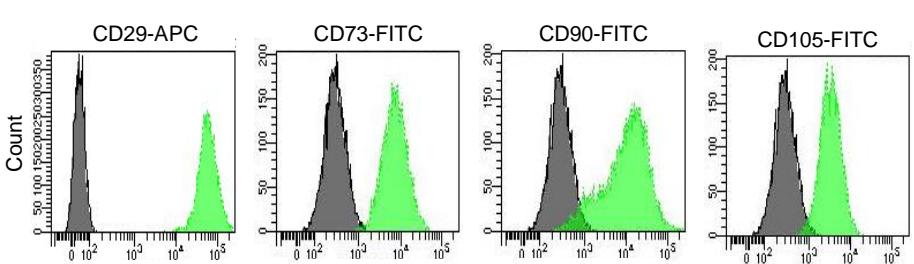
A



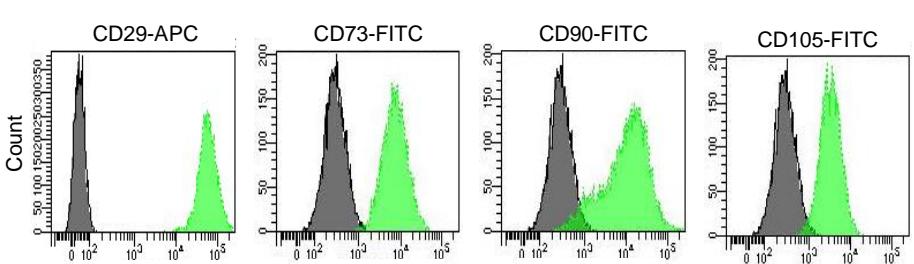
D



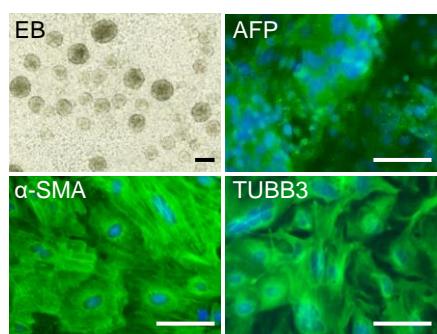
E



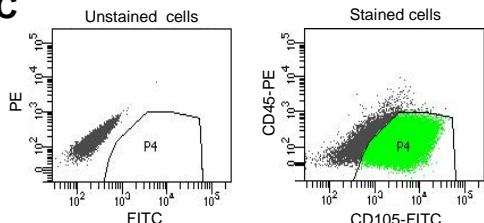
F



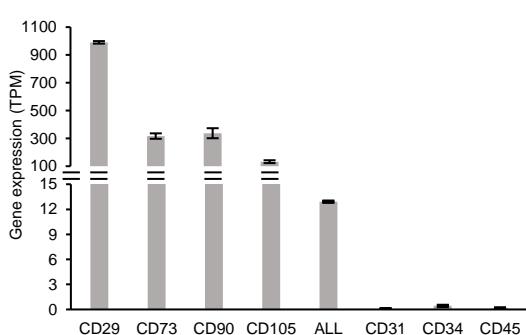
B



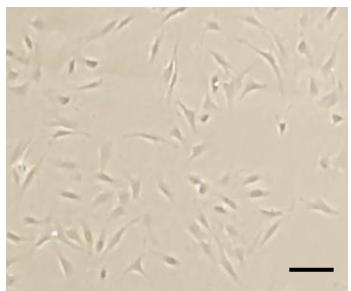
C



G

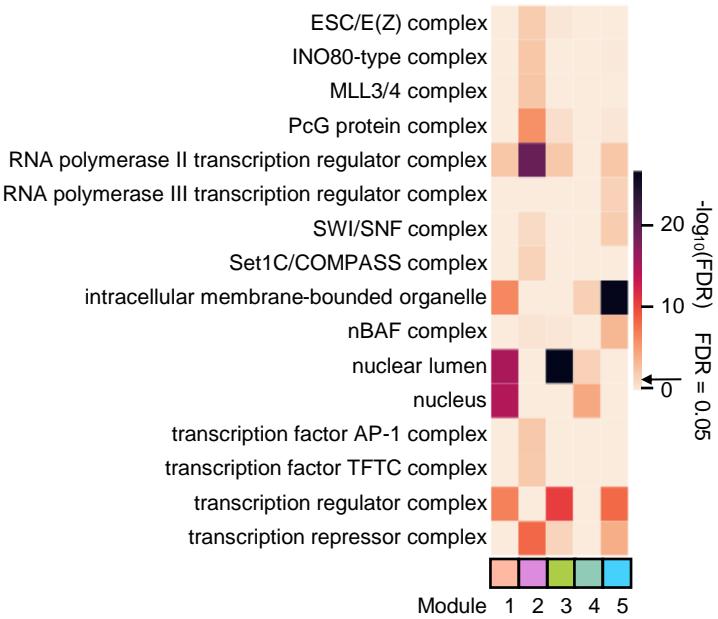


H

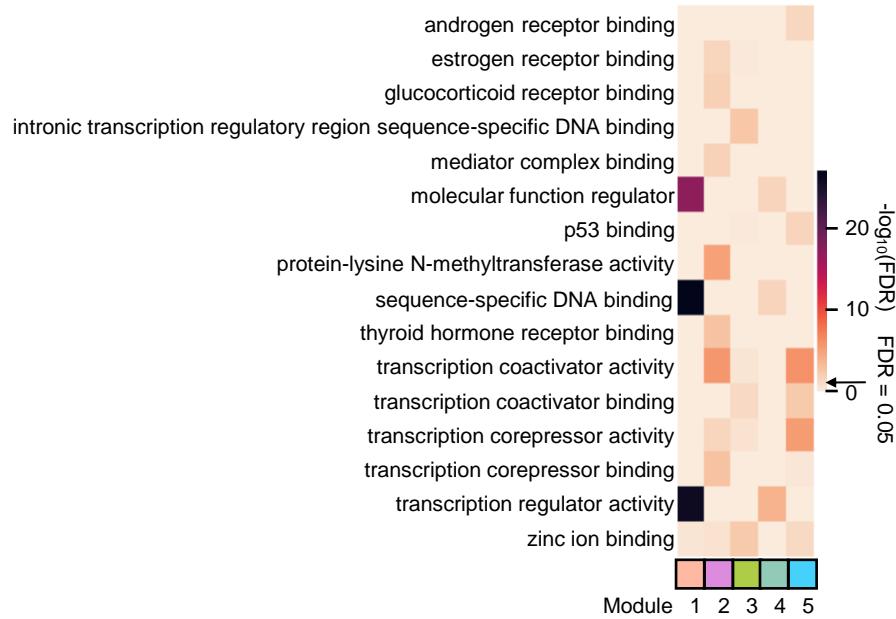


Supplementary Fig. 2

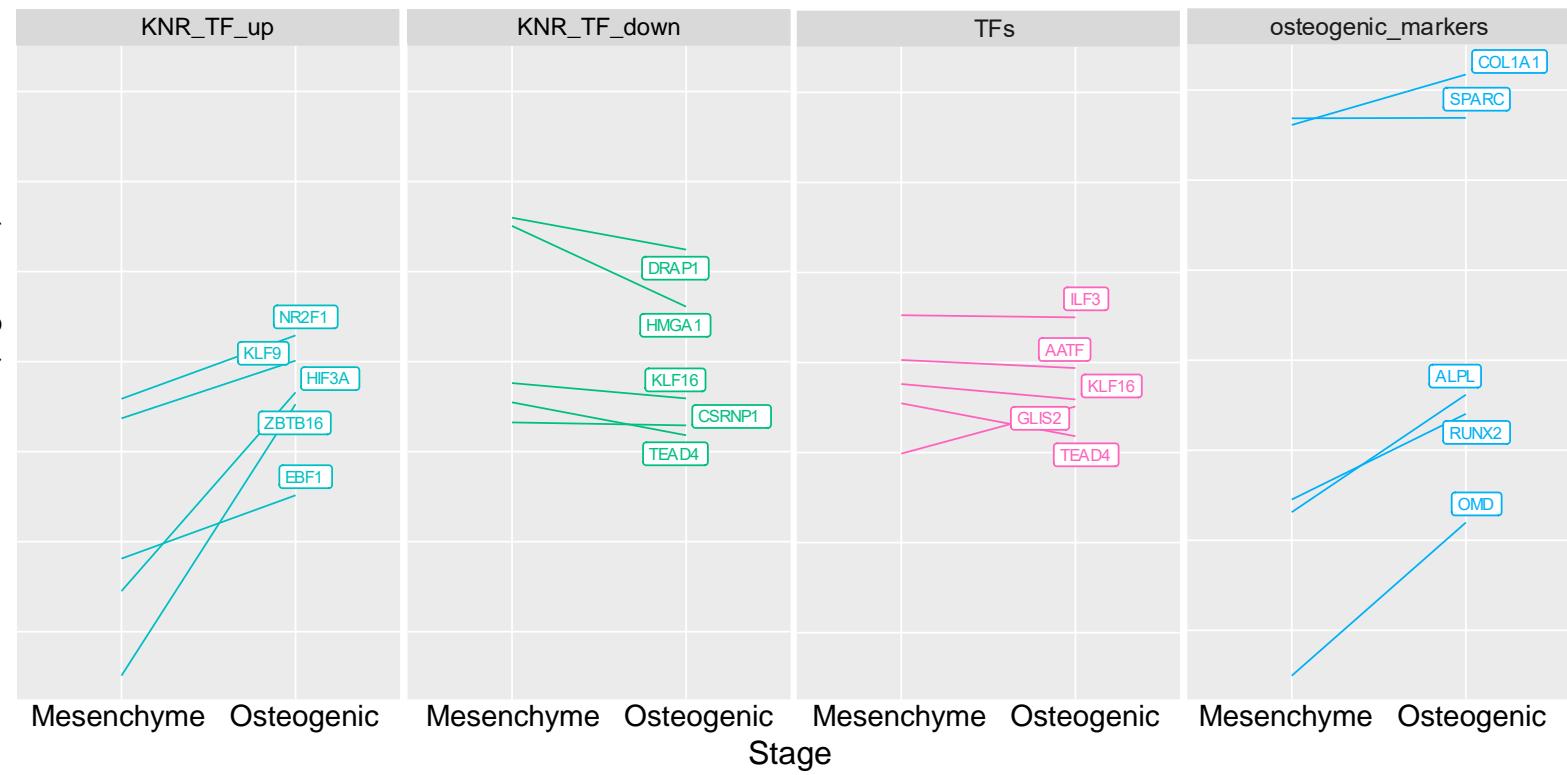
A



B



Supplementary Fig. 3



Supplementary Fig. 4

