Cell type-specific network analysis in Diversity Outbred mice identifies genes potentially responsible for human bone mineral density GWAS associations

Luke J Dillard[1], Gina M Calabrese[1], Larry D Mesner[1,2], Charles R Farber[1,2,3]

[1]Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA 22908
[2]Department of Public Health Sciences, School of Medicine, University of Virginia, Charlottesville, VA 22908
[3]Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA 22908

Correspondence to:

Charles R. Farber
E-mail: crf2s@virginia.edu
Center for Public Health Genomics
University of Virginia
P.O. Box 800717
Charlottesville, VA 22908, USA

Tel. 434-243-8584

Abstract

Genome-wide association studies (GWASs) have identified many sources of genetic variation associated with bone mineral density (BMD), a clinical predictor of fracture risk and osteoporosis. Aside from the identification of causal genes, other difficult challenges to informing GWAS include characterizing the roles of predicted causal genes in disease and providing additional functional context, such as the cell type predictions or biological pathways in which causal genes operate. Leveraging single-cell transcriptomics (scRNA-seq) can assist in informing BMD GWAS by linking disease-associated variants to genes and providing a cell type context for which these causal genes drive disease. Here, we use large-scale scRNA-seq data from bone marrow–derived stromal cells cultured under osteogenic conditions (BMSC-OBs) from Diversity Outbred (DO) mice to generate cell type-specific networks and contextualize BMD GWAS-implicated genes. Using trajectories inferred from the scRNA-seq data, we identify networks enriched with genes that exhibit the most dynamic changes in expression across trajectories. We discover 21 network driver genes, which are likely to be causal for human BMD GWAS associations that colocalize with expression/splicing quantitative trait loci (eQTL/sQTL). These driver genes, including *Fgfrl1* and *Tpx2,* along with their associated networks, are predicted to be novel regulators of BMD via their roles in the differentiation of mesenchymal lineage cells. In this work, we showcase the use of single-cell transcriptomics from mouse bone-relevant cells to inform human BMD GWAS and prioritize genetic targets with potential causal roles in the development of osteoporosis.

Keywords: Osteoblasts, Osteocytes, Osteoporosis, Mesenchymal Stem Cells (MSC), Stromal cells, Systems Biology

Introduction

Osteoporosis is a complex disease characterized by low bone mineral density (BMD), bone fragility, and an increased risk of fracture[1]. BMD, a highly heritable trait, is one of the most important clinical predictors of osteoporotic fracture[2,3]. Increasing our understanding of the genetic basis of BMD is critical for the development of approaches for the treatment and prevention of osteoporosis. In recent years, genome-wide association studies (GWAS) have made great progress in unraveling BMD genetics by identifying over 1,100 independent associations[4]. Now the challenge lies in pinpointing causal genes, which is necessary for the translation of genetic findings into novel therapies.

A number of approaches exist to identify genes responsible for GWAS associations[5–8]. Most rely on population-based "-omics" data[9], which are scarce for human bone, to connect associations to causal genes. Recently, our group has used co-expression networks generated from mouse bone transcriptomic datasets to assist in the identification of genes likely responsible for BMD associations. The idea is simple, genes that play a central role in the regulation of a complex trait are often functionally-related and functionally-related genes are often co-expressed[10]. By identifying networks enriched for genes implicated in GWAS we predicted putatively causal genes for hundreds of BMD associations based on their membership in enriched modules. One significant advantage of this approach is its ability to utilize the network connections between candidate genes and genes already established to influence bone function. This enables us to predict how these candidate genes may affect BMD. For example, we generated co-expression networks from bone tissue and primary osteoblasts in mouse genetic reference populations and identified multiple co-expression modules enriched with genes located in BMD associations[11,12]. We then cross-referenced genes in these modules with those regulated by co-localizing expression

69 quantitative trait loci (eQTLs) from the Gene-Tissue Expression project (GTEx)[13,14] to identify
70 "high priority" genes. We showed that one of these genes, *MARK3*, a serine-threonine kinase[15],
71 influenced BMD in an osteoblast-autonomous manner through a role in osteoblast
72 differentiation[11,15]. Recently, we expanded our approach to include directed networks generated
73 via a Bayesian approach using cortical bone RNA-seq data from 192 Diversity Outbred (DO) mice.
74 By combining key driver analysis and GTEx eQTL colocalization data, we identified 19 novel
75 genes, such as *SERTAD4* and *GLT8D2*, which are likely causal for human BMD GWAS
76 associations[16].

77 To date, our analyses have been reliant on networks generated from heterogeneous bulk
78 transcriptomics (RNA-seq) datasets from mouse bone and primary bone cells. However,
79 leveraging single-cell transcriptomics (scRNA-seq) data would offer the added benefit of resolving
80 the transcriptomic profiles of discrete cell types. Additionally, using scRNA-seq data has the
81 potential to provide context by predicting the specific cell types in which causal genes and their
82 associated networks operate. In recent work, we demonstrated the utility of bone marrow-derived
83 stromal cells cultured under osteogenic condition (BMSC-OB) for the generation of population-
84 scale scRNA-seq data on bone relevant cell types[17]. The BMSC-OB model effectively enriches for
85 mesenchymal lineage cells (e.g., mesenchymal progenitors, osteoblasts, osteocyte-like cells) that
86 are highly relevant to the regulation of BMD.

87 In this work, our goal was to prioritize and contextualize genes implicated by BMD GWAS
88 using an expanded large-scale (N=80) BMSC-OB scRNA-seq dataset on bone cell types. We
89 accomplished this by first generating co-expression and Bayesian networks[16] for each BMSC-OB
90 mesenchymal cell type. We subsequently prioritized networks based on their enrichment for genes
91 exhibiting the most dynamic changes in expression across trajectories inferred from the scRNA-

92 seq data, thus highlighting networks likely associated with the differentiation of BMSC-OBs. We
93 then used these networks linked to osteogenic differentiation to prioritize genes with eQTL and/or
94 splicing quantitative trait loci (sQTL) which colocalize with BMD GWAS associations[7,18]. In
95 doing so, this analysis provides additional support for a role of these genes in the regulation of
96 BMD and highlights their potential roles in differentiation of cell types essential to skeletal health.
97
98 **Results**
99
100 **BMSC-OBs derived from DO mice yield diverse cell types that are enriched for**
101 **mesenchymal lineage cells:**
102
103 We cultured BMSCs from a total of 80 Diversity Outbred (DO) mice, a genetically diverse
104 outbred mouse population[19,20] (N=75 from the current study and N=5 from[17]; N = 49 male and
105 N = 31 females). We cultured BMSCs under osteogenic conditions and subsequently performed
106 scRNA-seq, as described in[17]. After stringent processing and quality control (Materials and
107 Methods), the dataset consisted of the expression of 21,831 genes across 139,392 cells. We
108 manually annotated 15 clusters ranging in size from 270 to 27,291 cells and identified cell types
109 of the mesenchymal lineage as well as various other cell types (**Fig. 1A**, **Supplementary Table**
110 **S1**).
111 Based on our prior BMSC-OB scRNA-seq study[17], we expected to identify a large
112 proportion of mesenchymal cells and a smaller fraction of non-mesenchymal cell types. Consistent
113 with this hypothesis, clusters associated with mesenchymal lineages accounted for 74.1% of all
114 cells (**Fig. 1A**). These included mesenchymal progenitor cells (MPCs), late mesenchymal
115 progenitors (LMPs), osteoblast progenitors (OBPs), two mature osteoblast populations (OB1 and
116 OB2), osteocyte-like cells (Ocy), and marrow adipogenic lineage progenitors (MALPs). The non-
117 mesenchymal cell types observed included macrophages, monocytes, granulocytes, T-cells, B-

118 cells, endothelial cells, and osteoclast-like cells (**Fig. 1A**). With regards to the mesenchymal cell
119 types, the only differences in cell clusters relative to our previous report[17] were the presence of
120 MPCs and two mature osteoblast clusters. Upon comparing the two distinct osteoblast clusters,
121 OB1 and OB2 (**Fig. 1A**), both clusters had ubiquitous expression of genes associated with mature
122 osteoblasts (e.g., *Col1a1*, *Bglap, Sparc*, and *Ibsp*) (**Supplementary Table S1**) while many of the
123 "canonical" osteoblast markers were more highly expressed in OB1 compared to OB2
124 (**Supplementary Table S2**). Interestingly, MPCs did not have transcriptomic profiles similar to
125 other mesenchymal progenitor cells previously identified by our group or others[17,21]. All other
126 mesenchymal cell types demonstrated specific expression of canonical marker genes (**Fig. 1A, B**).
127 We next assessed the variability in cell type frequencies across DO mice by quantifying
128 the proportions of each annotated mesenchymal cell type. All other clusters, which mainly
129 consisted of immune cells of hematopoietic origin, were aggregated into one group (Hem) for each
130 mouse. We observed high variability in the raw proportional abundances of cell types derived from
131 each mouse (**Fig. 1C, Supplementary Table S3**). For example, the proportions of osteoblasts
132 (OB1 and OB2) varied significantly among individual DO mice (**Fig. 1D**). All mice used in the
133 current experiment had been extensively phenotyped for a wide range of bone traits (microCT,
134 histomorphometry, biomechanical bone properties, etc.) as part of a previous genetic analysis of
135 bone strength[16]. We correlated cell type frequencies with bone traits, however, none of the cell
136 type proportions were strongly correlated with any bone trait (**Supplementary Table S4, S5**).

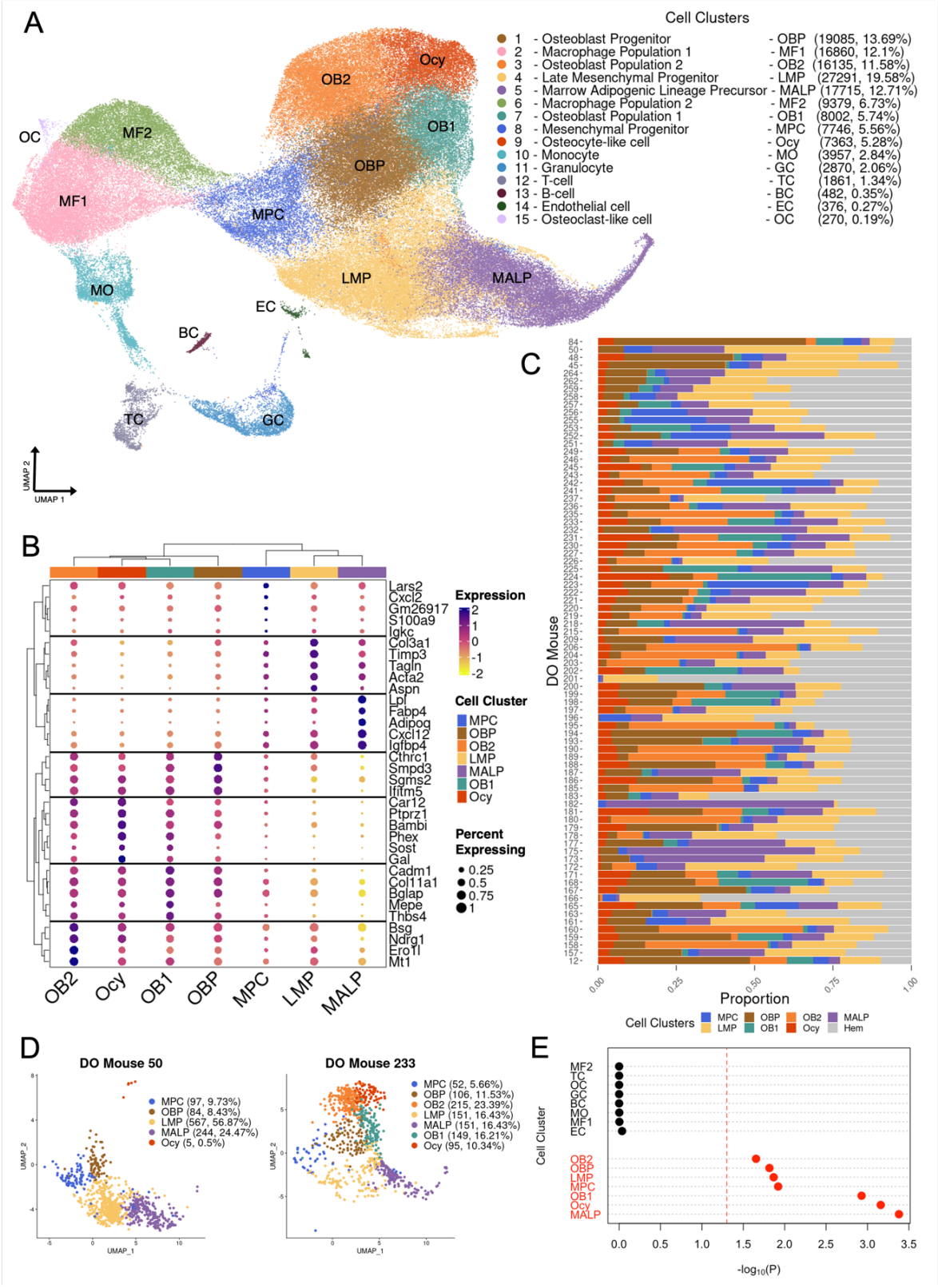

Figure 1. Analysis of single cell RNA-seq (scRNA-seq) data for BMSC-OBs derived from 80 Diversity Outbred (DO)

140 **Mesenchymal lineage cells are enriched in BMD heritability:**
141
142 The primary goal of this work was to prioritize and contextualize genes implicated by
143 BMD GWAS. As a first step towards this goal, we sought to determine which cell types were the
144 most relevant to the genetics of BMD. Using the BMD GWAS and the BMSC-OB scRNA-seq
145 data, we performed a CELLECT[22] analysis to identify cell clusters enriched for BMD
146 heritability. We observed that all mesenchymal cell types were nominally significantly ($P < 0.05$)
147 enriched for BMD heritability (**Fig. 1E**, **Supplementary Tables S6**). The OB1, Ocy, and MALP
148 cell types were significantly enriched for BMD heritability after adjusting for multiple
149 comparisons ($P_{adj} < 0.05$). None of the non-mesenchymal cells identified were significant ($P >$
150 0.05) (**Fig. 1E**). As a result, all downstream efforts focused solely on using data on mesenchymal
151 cell types to inform GWAS.
152

153 **Generating mesenchymal cell type specific Bayesian networks to inform BMD GWAS:**
154
155 We have previously shown that network-based approaches using bulk RNA-seq are
156 powerful tools for the identification of putative causal genes from BMD GWAS data[11,12,16]. Here,
157 our goal was to apply these same approaches using the BMSC-OB scRNA-seq data to prioritize
158 and contextualize genes we previously identified as having a colocalizing expression quantitative
159 trait locus (eQTL; N=512) or splicing QTL (sQTL; N=732) in a tissue from the Genotype-Tissue
160 Expression (GTEx) project[7,14,18]. Genes identified in each study (or both) yielded a list of high
161 priority target genes (N = 1,037). While GTEx does not currently contain data for bone tissue,
162 eQTL can be shared across many tissues and may exert their effects in cell types resident to bone[23].
163 Therefore, utilizing our previous work, we hypothesized that generating cell type-specific
164 networks would yield more biologically relevant representations of processes occurring within

165 particular mesenchymal cell types. Additionally, by integrating GWAS, cell type specific
166 networks, and dynamic gene expression as a function of differentiation we anticipated we would
167 identify points of intervention in which genetic variation impacts genes involved in the
168 differentiation process.

169 Our network analysis begins by partitioning genes into groups based on co-expression by
170 applying iterative weighted gene co-expression network analysis (iterativeWGCNA)[24] to each
171 mesenchymal cell type (Step 1, **Fig. 2**). In total, 535 modules were identified from the BMSC-OB
172 scRNA-seq data, and the number of modules identified for each mesenchymal cell cluster ranged
173 from 26 to 153 with an average of 76 modules per cluster (**Supplementary Table S7, S8**). We
174 sought to infer causal relationships between genes in each cell type-specific co-expression module
175 and subsequently identify networks involved in processes relevant to BMSC-OB differentiation.
176 To this end, we generated Bayesian networks for each co-expression module, thus enabling us to
177 model directed interactions between co-expressed genes based on conditional independence[16]
178 (Step 2, **Fig. 2**).
179

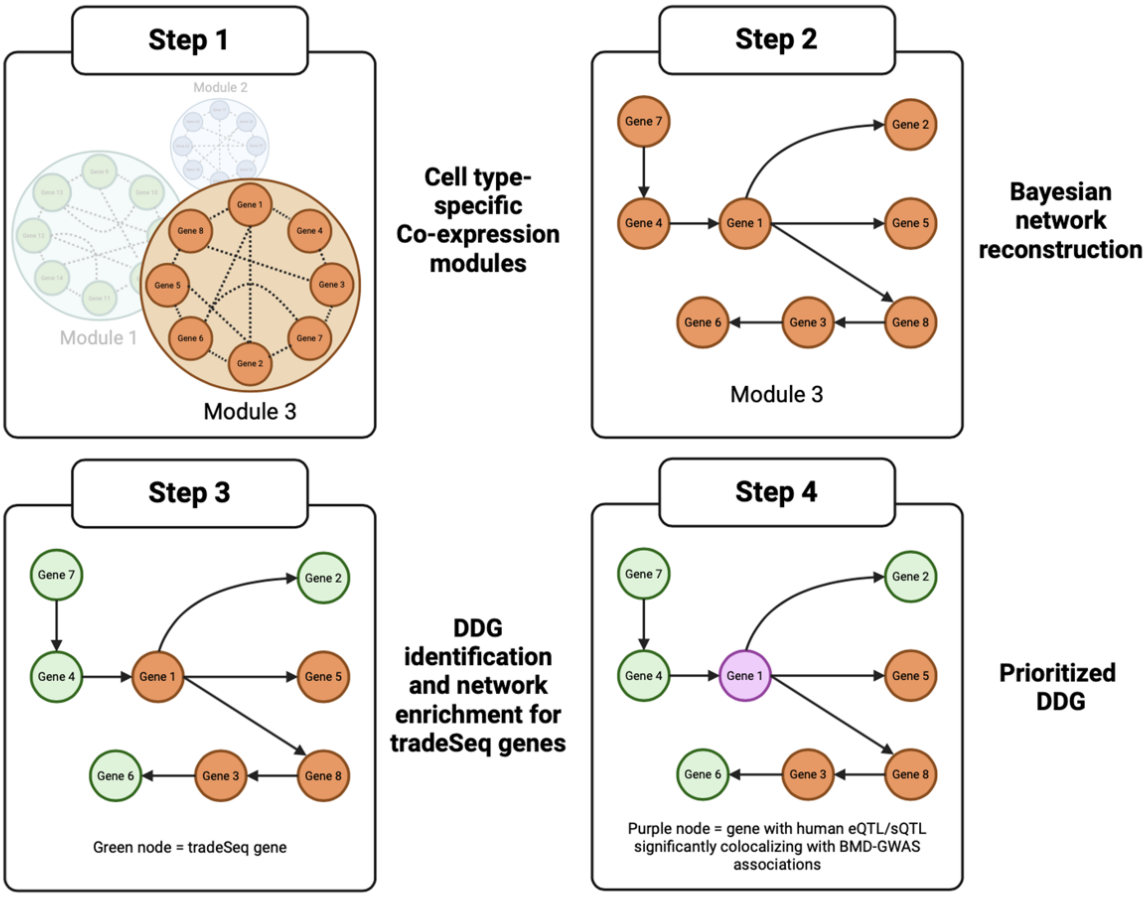

Figure 2. Overview of the network analysis pipeline

Identifying putative drivers of mesenchymal cell differentiation:

We hypothesized that many genes impacting BMD do so by influencing osteogenic differentiation or possibly bone marrow adipogenic differentiation, as suggested by the CELLECT analysis above. Therefore, the next step of our network analysis was to identify cell type-specific Bayesian networks enriched for genes potentially driving mesenchymal differentiation (Step 3, **Fig. 2**). To accomplish this, we first performed a pseudotime trajectory analysis to infer paths of differentiation in the mesenchymal lineage cells. We resolved three pseudotime trajectories (two osteogenic, one adipogenic) originating from the MPC cell cluster and ending in either Ocy, OB2,

192 or MALP cell fates (**Fig. 3A**). It is important to note that given the identification of multiple
193 skeletal stem cells[25–28], we do not view these trajectories as lineages, but rather "differentiation
194 paths" (progenitor to mature and/or terminally differentiated cells) that are likely traversed by
195 different subsets of skeletal stem cells.

196 To identify genes likely impacting BMSC-OB differentiation, we used tradeSeq to identify
197 genes that exhibit dynamic changes in expression along pseudotime[29]. Prior to performing the
198 tradeSeq analysis, we parsed the pseudotime trajectories into regions that encompass cells
199 associated with each cell type along their respective trajectories (**Fig. 3B**). We defined multiple
200 cell type boundaries (nine in total) using pseudotime values, which represent points along a
201 trajectory. The tradeSeq analysis was performed for each boundary (**Supplementary Table S9**).
202 For example, trajectories bifurcate in the LMP cell cluster (**Fig. 3A**); therefore, cells belonging to
203 the LMP cluster can map to adipogenic and/or osteogenic trajectories depending on their
204 placement along pseudotime. Between a cell type boundary, cells spanning a specific cluster (e.g.,
205 LMP) and mapping to a specific trajectory (e.g., osteogenic trajectory) are used as input to analyze
206 gene expression between the start and end points of the cell type boundary (e.g., LMP_to_OBP).
207 We analyzed gene expression within the established cell type boundaries for all trajectories and
208 identified genes that exhibit the most significant differences in expression between the start and
209 end points of the cell type boundaries. The total number of significant trajectory-specific tradeSeq
210 genes ($P_{adj} < 0.05$) ranged from 87 to 1,697 across the 9 cell type boundaries (**Supplementary**
211 **Table S9, S10-12**). The expression of representative marker genes for all cell types as a function
212 of pseudotime were consistent with boundaries defined for each cell type (**Fig. 3C**).

213 To provide further support that tradeSeq-identified genes are involved in differentiation,
214 we performed a cell type-specific expression quantitative trait locus (eQTL) analysis for each

215 mesenchymal cell type. We identified 563 genes (eGenes) regulated by a significant *cis*-eQTL in
216 specific cell types of the BMSC-OB scRNA-seq data (**Supplementary Table S14**). In total, 73
217 eGenes were also tradeSeq-identified genes in one or more cell type boundaries along their
218 respective trajectories (**Supplementary Table S9**).

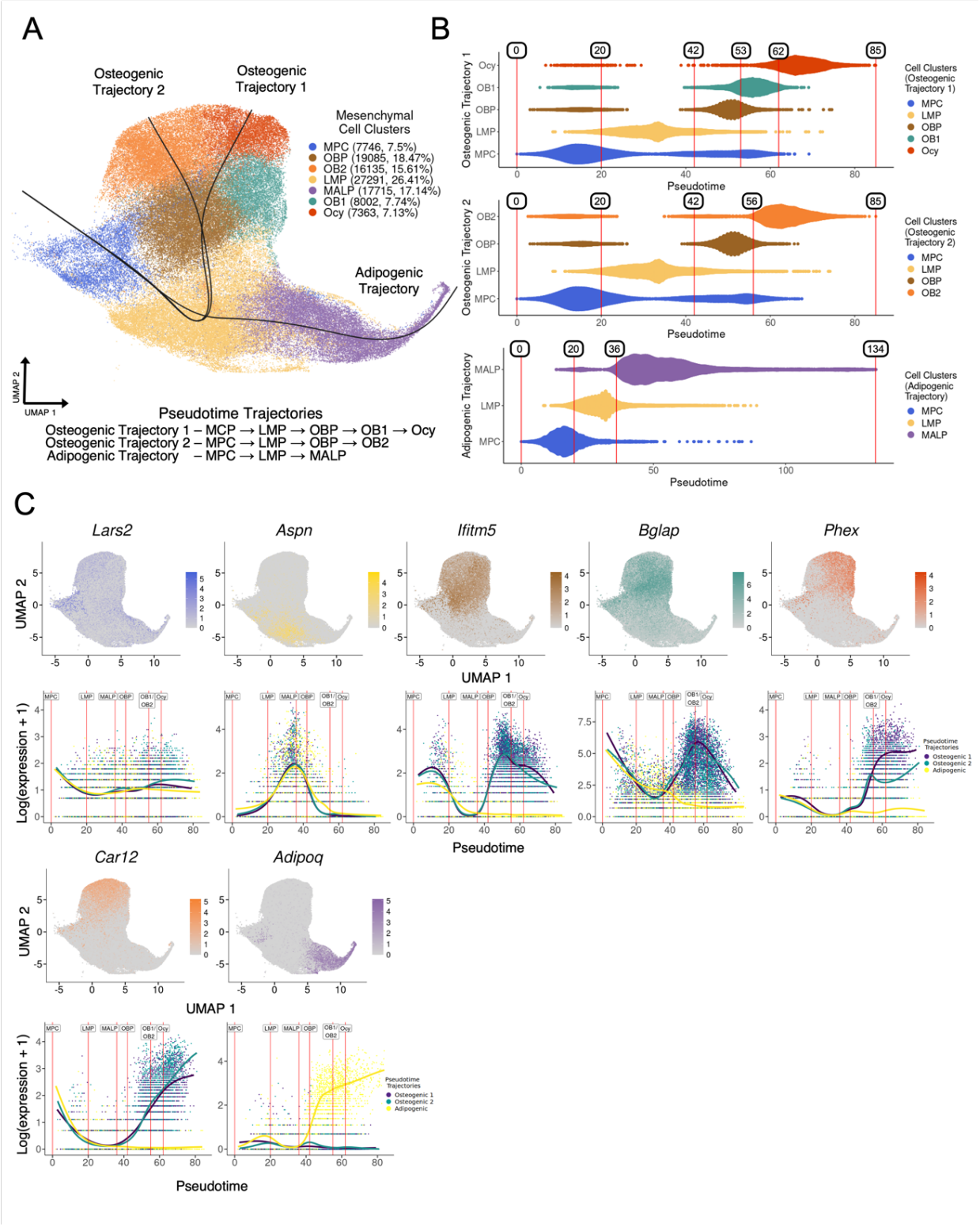

Figure 3. Pseudotime Trajectory Inference analysis and establishment of cell type boundaries for tradeSeq analysis

223 We hypothesized that if tradeSeq genes were responsible for driving mesenchymal
224 differentiation, then the eQTLs that perturb their expression would also impact the proportion of
225 cells at different stages along the cell trajectories. Despite being significantly underpowered for
226 this analysis due to our relatively smaller sample size (N = 80), we identified two cell type-specific
227 eGenes where the genotype responsible for the *cis*-eQTL effect was also associated with cell type
228 proportions. The first of these genes was Pyruvate Kinase, muscle (*Pkm*), which was identified as
229 a significant global tradeSeq gene ($P_{adj} = 8.35 \times 10^{-8}$; **Supplementary Table S13**) associated with
230 the transition from LMPs to OBPs along an osteogenic trajectory (**Fig. 4A**). Moreover, *Pkm* served
231 as an eGene in the LMP cell cluster (LOD = 9.72; **Fig. 4B, Supplementary Table S14**). Mice
232 inheriting at least one PWK allele at this locus (N = 15) demonstrated lower *Pkm* expression (**Fig.
233 4C**) and a notable reduction in mature osteoblasts (OB1) and osteocyte-like cells (Ocy) proportions
234 (P = 0.030 and P = 0.026, respectively), while LMP proportions were unaffected (**Fig. 4D,
235 Supplementary Table 15**).

236 Similarly, S100 calcium binding protein A1 (*S100a1*) was an OBP to OB1 transition
237 tradeSeq gene ($P_{adj} = 0.023$; **Fig. 4A, Supplementary Table S13**) and an eGene in the OBP cell
238 cluster (LOD = 10.12; **Fig. 4B, Supplementary Table S14**). Mice inheriting at least one 129 allele
239 at this locus (N = 30) had higher *S100a1* expression, while the opposite was observed for mice
240 inheriting NZO alleles (N = 14) (**Fig. 4C**). Additionally, mice inheriting at least one 129 allele
241 showed a significant decrease in LMP proportion and increase in OB1 proportion (P = 0.008 and
242 P = 0.016, respectively) (**Fig. 4D, Supplementary Table S15**), while no significant differences
243 were observed in cell type proportions among mice inheriting NZO alleles at this locus
244 (**Supplementary Fig. 2, Supplementary Table S15**). These data support the role of tradeSeq-
245 identified genes in the differentiation of mesenchymal cell types.

246

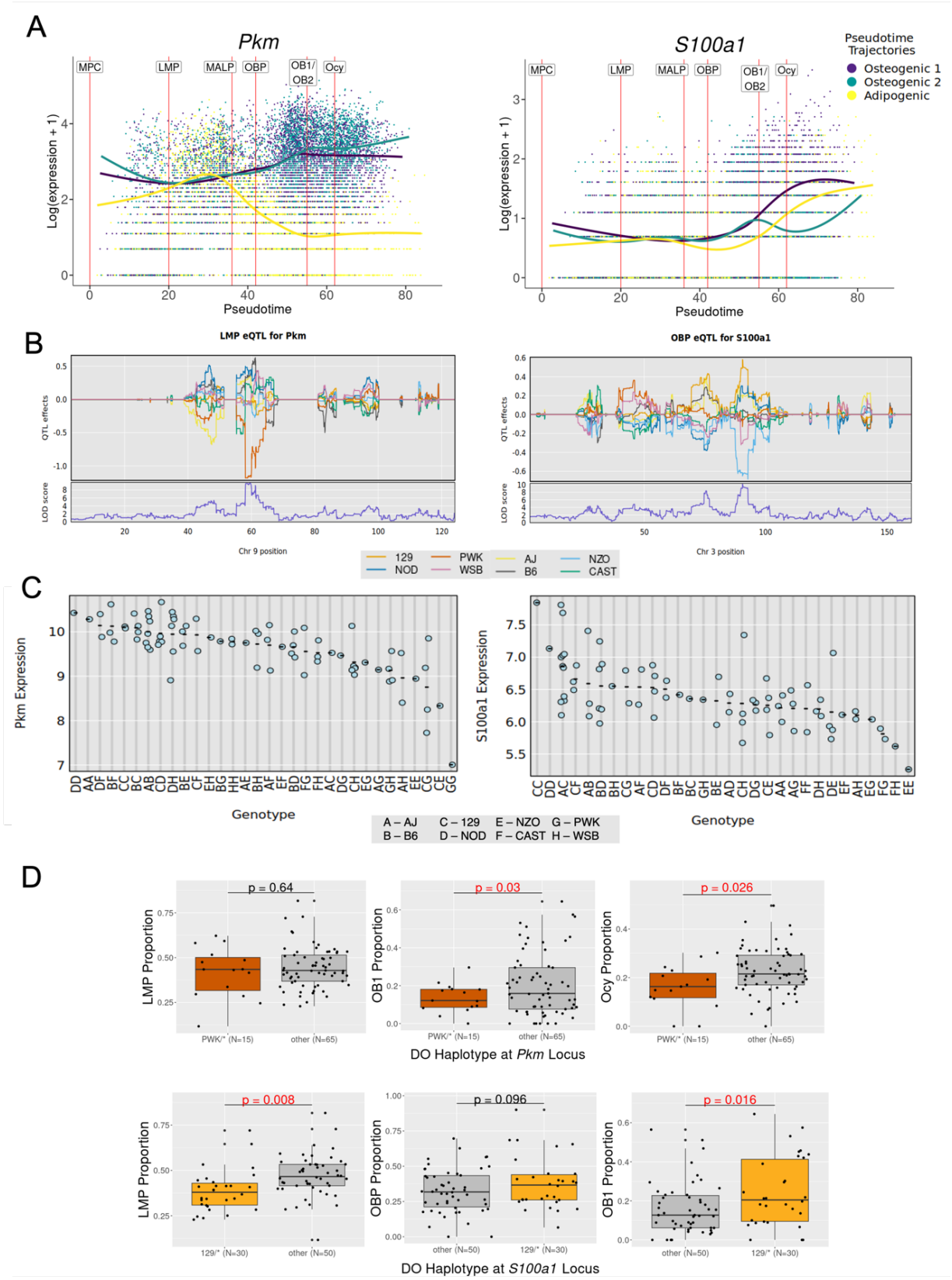

Figure 4. TradeSeq-identified genes associated with BMSC-OB differentiation exhibit eQTL effects.

250 **Identification of differentiation driver genes (DDG):**
251
252 In order to discover BMSC-OB differentiation genes potentially responsible for BMD
253 GWAS associations, the next step of our network analysis leveraged the trajectory-specific
254 tradeSeq genes identified for each cell type boundary (**Supplementary Table S10-12**) to identify
255 differentiation driver genes (DDGs) (Step 3, **Fig. 2**). We identified DDGs by extracting
256 subnetworks (i.e., large 3-step neighborhoods; see Methods) for each gene in each cell type-
257 specific Bayesian network and identifying those subnetworks enriched ($P_{adj} < 0.05$) for trajectory-
258 specific tradeSeq genes for the cell type boundary. The analysis identified 408 significant DDGs
259 (**Supplementary Table S16, S17-19**). We performed a PANTHER[30] Gene Ontology (GO)
260 analysis for the cell type boundaries yielding a sufficient number of DDGs and found that DDGs
261 for the osteogenic cell type boundaries (LMP_to_OBP, OBP_to_OB1, OBP_to_OB2) were
262 enriched for genes associated with the cell cycle (GO:0007049; N = 23, 18, 23; P = 1.12×10^{-6},
263 1.29×10^{-13}, 1.0×10^{-14}, respectively) (**Supplementary Table S20-22**). The DDGs for the
264 adipogenic cell type boundary (LMP_to_MALP, MALP_to_end) were enriched for genes
265 associated with extracellular matrix organization (GO:0030198; N = 10; P = 1.62×10^{-7}) and lipid
266 metabolic processes (GO:0006629; N = 25; P = 1.83×10^{-11}), respectively (**Supplementary Table**
267 **S23-24**). Across all 408 DDGs, 49 (12%) were identified in one or more cell type boundaries as
268 genes with a significant alteration (P < 0.05) of whole-body BMD when knocked-out/down in
269 mice, as reported by the International Mouse Knockout Consortium (IMPC)[31] (**Supplementary**
270 **Table S17-19**).
271 We used our previously generated list of potentially causal BMD GWAS genes (N=1,037)
272 to subsequently prioritize the DDGs (Step 4, **Fig. 2**). Of the 408 DDGs, 21 DDGs in one or more
273 cell type boundaries were genes that have BMD GWAS associations that colocalize with

274 sQTL/eQTL (**Table 1**). The majority of these DDGs were identified in LMPs along both the
275 osteogenic (LMP_to_OBP) and adipogenic (LMP_to_MALP) trajectories (N = 10 and 6,
276 respectively; **Supplementary Table S16, S25**). The remaining DDGs were identified in OBPs
277 along both osteoblast trajectories (OBP_to_OB1, OBP_to_OB2; N = 1 and 3, respectively) and
278 MALPs (MALP_to_end; N = 6). Additionally, 3 of the 21 DDGs (*Tet1*, *Tpx2*, *Timp2*) are IMPC
279 genes that exhibit a significant alteration of BMD (**Supplementary Table S16, S25**).

280 **Table 1: Prioritized Differentiation Driver Genes (DDGs) that have BMD GWAS**
281 **associations that colocalize with splicing/expression QTL (eQTL/sQTL) identified in a**
282 **Genotype-Tissue Expression project (GTEx) tissue.** The tissue with the most significant
283 colocalization (RCP and/or H4PP) is listed for each DGG (26 total, 21 distinct), as determined
284 from Al-Barghouthi *et al.* (2022) and Abood *et al.* (2023) for eQTL and sQTL, respectively[7,18].
285 RCP = Regional Colocalization Probability (GWAS and eQTL colocalization). H4PP = H4
286 Posterior Probability (GWAS and sQTL colocalization).

Trajectory	**Cell type boundary**	**DDG**	**GTEx Tissue with strongest eQTL colocalization** (RCP)	**GTEx Tissue with strongest sQTL colocalization** (H4PP)
1	LMP to OBP	*Tet1*	Adipose (Visceral); 0.3191	-
1	LMP to OBP	*Tpx2*	Testis; 0.2031	-
1	LMP to OBP	*Cdk1*	-	Pituitary; 0.7795
1	LMP to OBP	*Ttyh3*	-	Liver; 0.9350
1	LMP to OBP	*Olfml3*	Artery (Aorta); 0.8048	-
1	LMP to OBP	*Izumo4*	-	Brain (Hypothalamus); 0.9182
1	LMP to OBP	*Sec24d*	Nerve (Tibial); 0.2677	
1	LMP to OBP	*Tmem263*	Adipose (Subcutaneous); 0.5704	Cultured cells (fibroblasts); 0.9716
1	LMP to OBP	*Lmf2*	-	Adrenal Gland; 0.8181
1	LMP to OBP	*Tln2*	Esophagus (Muscularis); 0.9697	-
1	OBP to OB1	*Kremen1*	Heart (Left Ventricle); 0.8686	-
2	OBP to OB2	*Kremen1*	Heart (Left Ventricle); 0.8686	-
2	OBP to OB2	*Ebf1*	-	Testis; 0.8760
2	OBP to OB2	*Lrp4*	Pancreas; 0.7943	-

3	LMP to MALP	*Ttyh3*	-	Liver; 0.9350
3	LMP to MALP	*Fgfrl1*	Cultured cells (fibroblasts); 0.1611	-
3	LMP to MALP	*Ebf1*	-	Testis; 0.8760
3	LMP to MALP	*Ppp1r12b*	-	Nerve (Tibial); 0.8807
3	LMP to MALP	*Rhoj*	Cultured cells (fibroblasts); 0.352	Breast; 0.7844
3	LMP to MALP	*Tln2*	Esophagus (Muscularis); 0.9697	-
3	MALP to end	*Adh1*	-	Esophagus (Gastroesophageal Junction); 0.9999
3	MALP to end	*Fgfrl1*	Cultured cells (fibroblasts); 0.1611	-
3	MALP to end	*Adcy5*	-	Esophagus (Gastroesophageal Junction); 0.8456
3	MALP to end	*Cnn2*	-	Spleen; 0.7743
3	MALP to end	*Mxra8*	-	Pituitary; 0.7545
3	MALP to end	*Timp2*	-	Testis; 0.9429

Network analysis predict Fgfrl1 and Tpx2 as novel regulators of BMD:

Here we highlight two DDGs that putatively impact human BMD via their roles in LMP differentiation along either an adipogenic (*Fgfrl1*) or osteogenic (*Tpx2*) trajectory. Based on our previous work[7], *Fgfrl1* (fibroblast growth factor receptor-like 1) was identified as a DDG with significant human BMD GWAS associations that also colocalized with eQTL identified in the cultured fibroblast GTEx tissue (RCP = 0.1611, **Table 1**). The *Fgfrl1* network was enriched for tradeSeq-identified genes (N = 6 genes, $P_{adj} = 7.5 \times 10^{-3}$) for LMPs along an adipogenic trajectory (**Fig. 5A**). An increase in the expression of all tradeSeq-identified genes for the *Fgfrl1* network was observed (**Fig. 5B, Supplementary Table S12**). Importantly, the expression pattern for the tradeSeq-identified genes were consistent with the cell type boundaries established for LMPs differentiating along the adipogenic trajectory toward the MALP cell state (**Fig. 5B**). Furthermore,

300 in the surrounding *Fgfrl1* network, two genes (*Plpp3* and *Cfap100*) have significant human BMD
301 GWAS associations that also colocalized with sQTL in GTEx tissues, as reported in our previous
302 study[18]. In the *Fgfrl1* network, many other genes can be associated with adipocyte function (e.g.,
303 *Lpl*, *Plpp3*, *Igfbp4*)[32–34] and the maintenance of cilia (e.g., *Cfap100*, *St5* (*Denn2b*), *Mark1*)[35–37].
304 The other network we identified, the *Tpx2* network, was identified for LMPs along an
305 osteogenic trajectory (**Fig. 5C**). *Tpx2* (TPX2 microtubule nucleation factor) is a DDG with
306 significant human BMD GWAS associations that also colocalized with eQTL identified in the
307 Testis GTEx tissue (RCP = 0.2031, **Table 1**). The network was enriched for tradeSeq-identified
308 genes (N = 9 genes, $P_{adj} = 5.7 \times 10^{-7}$) for LMPs differentiating along the osteogenic trajectory (**Fig.
309 5C**). Furthermore, the expression of the tradeSeq-identified genes for the *Tpx2* network were
310 consistent with the cell type boundaries established for LMPs differentiating along the osteogenic
311 trajectory toward the OBP (osteoblast progenitor) cell state (**Fig. 5D; Supplementary Table S10**).
312 The expression of these genes increase as LMPs differentiate into OBPs and subsequently decrease
313 upon reaching an OBP cell state. Additionally, *Tpx2* exhibited a significant alteration of BMD in
314 both male and female mutant mice (Genotype P-value = 1.03×10^{-3}) from IMPC (**Fig. 5E**). In
315 regards to the constituents of the surrounding *Tpx2* network, additional genes have been tested by
316 the IMPC and result in a significant impact on BMD, such as *Ube2c*, *Top2a*, and *Papss1*. Many
317 other genes in the *Tpx2* network can be associated with cellular division and proliferation,
318 including four of the genes of the kinesin family (*Kif*) motor protein genes[38]: *Kif4, Kif11, Kif15,*
319 *Kif23*.
320

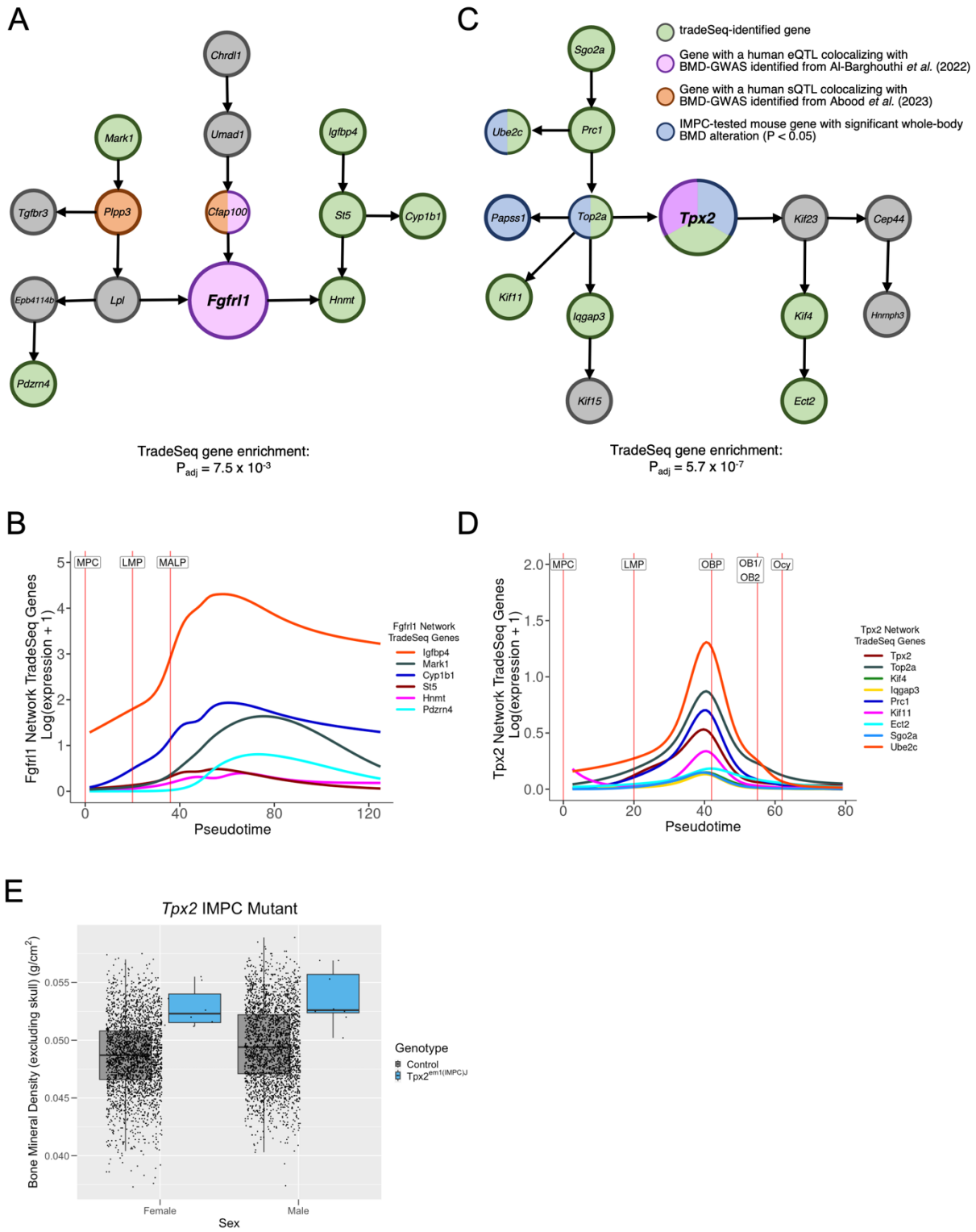

Figure 5. *Fgfrl1* and *Tpx2* are prioritized DDGs and putative drivers of mesenchymal differentiation.

Discussion

BMD GWAS has been successful at identifying thousands of SNPs associated with disease; however, the identification of the causal genes and defining their functional role in disease remains challenging. The integration of "-omics" data, particularly transcriptomics, can assist in overcoming this challenge. Leveraging transcriptomics data has proven invaluable to informing GWAS, as demonstrated in studies that use these data to perform eQTL mapping, transcriptome-wide association studies (TWASs), and co-expression/gene-regulatory network reconstruction. GWAS associations can colocalize with predicted sources of genetic variation that perturb causal gene function or expression, thus providing a potential mechanism through which associations impact disease. While bulk RNA-seq data has been the foundation of such analyses, scRNA-seq data can provide valuable biological context by predicting the cell type in which causal genes are affected. To inform BMD GWAS, the generation of population-scale transcriptomics data at single-cell resolution in bone-relevant cell types can assist in the discovery of novel gene targets. Here, we perform scRNA-seq on 80 DO mice to generate single-cell transcriptomics data of mesenchymal cell types relevant to bone. Using these data, our goal was to prioritize putative causal genes and provide biological context in which these genes potentially influence disease, at cell type-specific resolution. Through our temporal gene expression and network analyses, we identified 21 networks governed by predicted differentiation driver genes (DDGs) that have corresponding human BMD GWAS associations colocalizing with eQTL/sQTL in a GTEx tissue.

We demonstrate that the BMSC-OB model serves as an effective method to enrich for mesenchymal lineage cells, particularly bone-relevant cells. We characterized cells from 80 mice and identified both osteogenic and adipogenic cells derived from the mesenchymal lineage, such as two populations of osteoblasts (OB1 and OB2), osteocyte-like cells (Ocy), and MALPs. Our

348 trajectory inference analysis identified three distinct trajectories in which mesenchymal progenitor
349 cells give rise to both osteogenic and adipogenic cell types, thus portraying biologically relevant
350 and known paths of differentiation of mesenchymal progenitor cells. Temporal gene expression
351 was analyzed along each trajectory, in a cell type-specific fashion, to identify genes that were
352 changing the most as a function of pseudotime (tradeSeq-identified genes). Subsequent *cis*-eQTL
353 analysis indicated that the expression of some tradeSeq-identified genes were associated with the
354 relative proportion of cell types. While instances such as these were rare, they illustrate that the
355 potential consequence of genetic variation impacting the expression of tradeSeq-identified genes
356 may impact differentiation and the abundances of certain cell types along a trajectory.

357 To inform BMD GWAS, we utilized the scRNA-seq data in a network analysis to
358 contextualize causal genes (and their associated networks) by predicting the cell types through
359 which these genes are likely acting. Towards this goal, we generated cell type-specific Bayesian
360 networks from our BMSC-OB scRNA-seq data. Our approach was similar to our previous network
361 analyses where bulk RNA-seq data was leveraged to identify genes with strong evidence of playing
362 central roles in networks[11,12,16]. In contrast, here we utilized scRNA-seq data to identify DDGs and
363 prioritize networks based on the likelihood that they are involved in the differentiation of
364 mesenchymal lineage cells (based on network connections enriched for tradeSeq-identified genes
365 determined from inferred trajectories). Leveraging our previous work[7,18], we prioritized DDGs if
366 they were genes with BMD GWAS associations colocalizing with human eQTL/sQTL in a GTEx
367 tissue. Together, a gene being both a DDG and having BMD GWAS associations that colocalize
368 with eQTL/sQTL is strong support of causality.

369 We identified 21 DDGs and associated networks, some of which have little to no known
370 prior connection to bone. We contextualize these causal genes and their networks by not only

371 providing cell type predictions in which they likely operate, but also providing information
372 regarding the biological processes they likely affect. For example, the *Tpx2* network was identified
373 in LMPs differentiating along an osteogenic trajectory. *Tpx2* is a microtubule nucleation factor
374 that interacts with spindle microtubules during cellular division[39]. In our previous study, *Tpx2* was
375 identified as a gene that has BMD GWAS associations that colocalize with eQTL in the Testis
376 GTEx tissue[7]. While GTEx does not maintain bone tissue, eQTL are shared across many tissues[23];
377 therefore, non-bone eQTL may exert their effects in cell types associated with bone, such as LMPs,
378 and evidence of a human eQTL effect indicates that genetic variation can modulate the expression
379 of *Tpx2*. Additionally, when knocked out by IMPC, *Tpx2* exhibited a significant increase in whole
380 body BMD in mice; therefore, providing strong support for *Tpx2* influencing the regulation of
381 BMD in humans. In the surrounding gene neighborhood of the *Tpx2* network, other genes are
382 associated with cellular division as well, such as Topoisomerase 2A (*Top2a*) and the kinesin family
383 (*Kif*) genes[38,40]. Taken together, these results indicate a potential role of *Tpx2* as a mediator of
384 BMD via its role in microtubule maintenance during the expansion of osteogenic cell populations.
385 Additionally, the *Fgfrl1* network was identified in LMPs differentiating along an
386 adipogenic trajectory. Fibroblast growth factor receptor-like 1 (*Fgfrl1*) is presumed to function as
387 a decoy receptor that interacts with FGF ligands[41]. Our previous study also identified *Fgfrl1*, which
388 has BMD GWAS associations that colocalize with eQTL in the cultured fibroblasts GTEx tissue[16].
389 In the surrounding neighborhood of the *Fgfrl1* network, *Lpl*, *Plpp3*, *Igfbp4* have well-established
390 roles in adipocyte function and metabolism[32–34]; however, other genes can be associated with cilia
391 function, such as *Cfap100*, *St5* (*Denn2b*), *Mark1*[35–37]. Interestingly, the maintenance of cilia is
392 essential to the function of both LMPs and pre-adipocytes (MALPs) while mature adipocytes lack
393 cilia[42]. Therefore, the modulation of ciliogenesis and/or cilia function may coincide with *Fgfrl1*

394 signaling. Additionally, given the prioritization of MALPs in the CELLECT analysis and the well-
395 established inverse relationship between marrow adiposity and BMD[43,44], skewed balance of LMP
396 differentiation toward marrow adipogenic cell fates may affect BMD. In summary, the *Fgfrl1*
397 network harbors genes involved in adipogenic function, including cilia, which may contribute to
398 LMP differentiation along an adipogenic trajectory. Together, these results indicate a potential role
399 of *Fgfrl1* as a mediator of BMD via its role in adipogenic differentiation and maintenance of cilia.
400 Analyses performed here are not without limitations to consider. A pitfall of scRNA-seq is
401 the sparsity of the resulting data, which yields an increased frequency of zero values for the
402 expression of some genes in a proportion of cells, also known as "drop-outs"[45]. While statistical
403 approaches can be employed to impute missing data, the accuracy of such methods and whether
404 or not the resulting improvement in transcriptomic signal recovery is enough to warrant such
405 intervention is contentious[46,47]. However, this issue may be partially offset given the larger scale
406 of the scRNA-seq performed in this study and the average expression approach performed for
407 network and eQTL analysis. An additional limitation is that the BMSC-OB model does not capture
408 osteoclasts, another cell type associated with bone tissue. Importantly, results from our CELLECT
409 analysis indicate that BMD heritability was not enriched for genes whose expression was more
410 specific to osteoclast-like cells; however, these cell likely represent immature osteoclasts, as
411 mature multinucleated cells would be too large to be captured for sequencing. Lastly, while our
412 study employed 80 DO mice, the issue of statistical power is still a limitation; however, we
413 demonstrate that the BMSC-OB model is amenable to high throughput and the inclusion of
414 hundreds of mice, thus statistical power will be improved in future studies.
415 In summary, we showcase the use of large-scale scRNA-seq data to inform GWAS by
416 performing a network analysis to contextualize BMD GWAS associations. Through the use of

417 multiple single-cell analyses, we have expanded upon our understanding of the genetics of BMD.
418 Our work exemplifies the power of single-cell transcriptomics coupled with systems genetics to
419 discover the genetic determinants of disease.
420

421 **Methods**
422
423 **Sample preparation and scRNA-seq**
424
425 We prepared our samples in the same fashion as performed previously in Al-Barghouthi
426 and colleagues[17]. In brief, bone marrow was extracted from the femurs of initially 80 DO mice.
427 BMSCs were grown to confluence after 3 days of incubation in 48-well plates and then underwent
428 *in vitro* osteoblast differentiation for 10 days with osteogenic differentiation media (alpha MEM,
429 10% FBS, 1% pen/strep, 1% Glutamax, 50 µg/µL ascorbic acid [Sigma, St. Louis, MO, USA],
430 10 nM B-glycerophosphate [Sigma], 10 nM dexamethasome [Sigma]). After differentiation, single
431 cells were liberated from mineralizing cultures via incubations with 60 mM
432 ethylenediaminetetraacetic acid pH 7.4 (EDTA [Thermo Fisher Scientific], made in DPBS),
433 8 mg/mL collagenase (Gibco) in HBSS/4 mM CaCl2 (Fisher), and 0.25% trypsin–EDTA (Gibco).
434 After single-cell isolation, cells from mice were pooled into groups containing cells from five mice
435 total and concentrated to 800 cells/µL in PBS supplemented with 0.1% BSA (bovine serum
436 albumin). Pooled single cells were prepared for sequencing using the 10× Chromium Controller
437 (10× Genomics, Pleasanton, CA, USA) with the Single Cell 3' v2 reagent kit, according to the
438 manufacturer's protocol. Libraries were sequenced on the NextSeq500 (Illumina, San Diego, CA,
439 USA).
440

441 **scRNA-seq analysis pipeline**
442

443 The data was subsequently processed using the 10× Genomics Cell Ranger toolkit (version
444 5.0.0) using the GRCm38 reference genome[48]. Using Seurat[49] (version 4.1.0), a combined Seurat
445 object containing all cells was generated with the inclusion of features detected in at least three
446 cells and cells with at least 200 features detected. We used Souporcell[50] (version 2.0.0) to
447 deconvolve the genotypes of all mice and to remove doublet cells. Cells were assigned to their
448 associated DO mouse by making a pairwise comparison between allele calls made by the shared
449 variants captured between Souporcell and GigaMUGA genotype arrays generated for all mice in
450 the cohort, as previous performed in Dillard and colleagues[17]. We filtered out cells with more than
451 6200 reads and less than 400 reads, as well as those cells with more than 10% mitochondrial reads.
452 Further, cells were removed if they expressed greater than 20% *Rpl* and 15% *Rps* reads, which
453 equates to cells approximately exceeding the 98 percentile. After filtering, 139,392 cells remained
454 and the resulting object underwent standard normalization, scaling, and the top 3000 features were
455 modeled from a variance stabilizing transformation (VST) using the Seurat. Cell-cycle markers
456 based on Tirosh and colleagues[51] were regressed out using the "CellCycleScoring" and scaling
457 functions. For subsequent dimensionality reduction, 15 principal components (PCs) were
458 summarized. Then, a kNN ($k = 20$) graph was created and the Louvain algorithm was used to
459 cluster cells at a resolution of 0.5. Annotation of cell type clusters was performed manually based
460 on differential gene expression analysis using the Seurat "FindAllMarkers" function
461 (**Supplementary Table S1**).

462 For subsequent WGCNA and eQTL mapping, transcriptomic profiles for each cell type
463 cluster were generated for each sample using a mean expression approach, as performed similarly
464 by others[52,53]. For each sample contributing at least 5 cells to a given cluster, unnormalized unique
465 molecular identifier (UMI) counts of gene expression for all cells in the cluster for the sample were

466 averaged and then rounded to the nearest hundredth decimal place. A total of 80, 80, 77, 67, 50,
467 76, 80 mice contributed enough cells to the MPC, LMP, OBP, OB1, OB2, Ocy, and MALP cell
468 type clusters, respectively. Genes with non-zero expression values in fewer than 15 samples were
469 removed. A total of 11971, 15162, 14857, 13674, 13825, 14136, and 14534 genes remained for
470 the MPC, LMP, OBP, OB1, OB2, Ocy, and MALP clusters, respectively. Samples were
471 normalized by computing CPMs (counts per million) without log transformation for each gene
472 using edgeR[54] (version 4.0.7), then transformed via VST using DESeq2[55] (version 1.42.0), and
473 quantile normalized using preprocessCore (version 1.60.2).
474

475 **Trajectory and tradeSeq Analysis**
476
477 Trajectory inference analysis was performed using Slingshot[56] (version 1.8.0) on the
478 mesenchymal lineage cell clusters (seven total) of the BMSC-OB scRNA-seq data. The starting
479 cluster was set as the MPC cluster and trajectories were inferred using 15 PCs. TradeSeq[29] (version
480 1.4.0) was used to analyze gene expression along the trajectories by fitting a negative binomial
481 generalized additive model (NB-GAM) to each gene using the "fitGAM" function with nknots =
482 10, which was determined by using the "evaluateK" function. Prior to performing the tradeSeq
483 analysis, the scRNA-seq data was downsampled to reduce the size of the dataset to approximately
484 10,000 cells (sampled at random across all seven clusters).
485 All cell type boundaries were established to encompass on average 78% of cells of a cell
486 cluster (**Supplementary Table S9**). To identify genes significantly changing between boundaries
487 in a trajectory-specific fashion, we first performed tradeSeq to compare gene expression within
488 each trajectory (two osteogenic, one adipogenic) to highlight genes with a significant difference
489 in expression between boundaries using the "startVsEndTest" function (**Supplementary Table**

490 **S9, S10-12**). Next, we performed a global test with tradeSeq to compare gene expression between
491 trajectories in order to highlight genes exhibiting a significant difference in expression using the
492 "startVsEndTest" function (**Supplementary Table S9, S13**). All tests were performed with the
493 $\log_2$ fold change threshold (l2fc) = 0.5. For all global and trajectory-specific tests, the P-values
494 associated with each gene were adjusted to control the false discovery rate using the "p.adjust"
495 function from the stats (version 4.2.1) R package and genes were filtered to include those with a
496 $P_{adj} < 0.05$.
497

498 **CELLECT Analysis**
499
500 CELLECT[22] (CELL-type Expression-specific integration for Complex Traits) (version
501 1.1.0) was used to identify likely etiologic cell types underlying complex traits of both the BMSC-
502 OBs scRNA-seq data (**Fig. 1E, Supplementary Table S6**). CELLECT quantifies the association
503 between the GWAS signal and cell type expression specificity using the S-LDSC genetic
504 prioritization model[57]. Summary statistics from the UK Biobank eBMD and Fracture GWAS (Data
505 Release 2018) and cell type annotations from each scRNA-seq data set were used as input. Cell
506 type expression specificities were estimated using CELLEX[22] (CELL-type EXpression-
507 specificity) (version 1.2.1) (**Supplementary Table S26**).

508 **WGCNA**
509
510 Cell type-specific mean expression matrices (as obtained above) were used as input to
511 generate signed co-expression network modules (**Supplementary Table S7, S8**).
512 IterativeWGCNA[24] (version 1.1.6) was used from a Singularity container built from a Docker hub
513 image[58]. A soft threshold (power) of 14, which exceeded a R^2 threshold of 0.85 for all cell type
514 clusters, was selected for module construction (**Supplementary Fig. 1**). Modules were generated

515 using iterativeWGCNA with default parameters for the "blockwiseModules" function, a minimum
516 module size of 20 genes, minCoreKME = 0.7, and minKMEtoStay = 0.5.
517

518 **Bayesian network learning**
519
520 Bayesian networks were learned from each of the cell type-specific modules of co-
521 expressed genes with the bnlearn (version 4.8.3). Gene expression matrices containing the genes
522 for each module were used as input to the "mmhc" function which employs the Max-Min Hill
523 Climbing algorithm (MMHC) algorithm[59] to learn the underlying structure of the Bayesian
524 network. From the generated networks, igraph (version 1.6.0) was used to resolve 3-step
525 neighborhoods[60]. Nodes (genes) that were unconnected to a neighborhood or connected to only
526 one neighbor were removed. Neighborhoods were filtered to include those with a size greater than
527 1 standard deviation from the mean across all neighborhoods generated for the network.
528 DDGs (differentiation driver genes) are genes that yield large 3-step neighborhoods that
529 are enriched ($P_{adj} < 0.05$) with tradeSeq-identified genes for a given cell type boundary. We
530 calculated whether each neighborhood contained more tradeSeq-identified genes (for the
531 neighborhoods' associated cell type boundary) than would be expected by chance using the
532 hypergeometric distribution ("phyper" function) from the stats (version 4.2.1) R package. The
533 arguments were as follows: *q*: (number of neighbors in a neighborhood that are also tradeSeq-
534 identified genes for a given cell type boundary) – 1; *m*: total number of tradeSeq-identified genes
535 for a given cell type boundary; *n*: (total number of identified neighborhoods) – *m*; *k*: neighborhood
536 size (total number of neighbors); lower.tail = false. P-values were adjusted to control the false
537 discovery rate using the "p.adjust" function from the stats (version 4.2.1) R package. These

538 pruning steps resulted in a total of 408 DDGs and associated networks for all cell types
539 (**Supplementary Table S16, S17-19**).
540

541 **DO eQTL mapping**
542
543 Prior to performing the eQTL analysis, DNA was extracted from the tails of the 80 DO
544 mice, using the PureLink Genomic DNA mini kit (Invitrogen) and genotyped using the
545 GigaMUGA array by Neogen Genomics (GeneSeek; Lincoln, NE). Processing and quality control
546 of genotype data, including calculation of genotype/allele probabilities, was performed as
547 previously described in Al-Barghouthi and colleagues[16]. Cell type-specific mean expression
548 matrices (as obtained above) for mesenchymal lineage clusters were used as input for the eQTL
549 mapping, which was performed using a linear mixed model (LMM) via the "scan1" function from
550 the qtl2[61] (version 0.30) R package with allowances for the following covariates: sex, age at
551 sacrifice (in days), weight, length, and DO mouse generation. To identify significant eQTL, we
552 calculated a LOD (logarithm of the odds) threshold; for each cell type cluster, we chose 50 genes
553 at random and then permuted them 1000 times using the "scan1perm" function from qtl2. We
554 established the LOD threshold of 9.68 and 9.49 for the autosomal chromosomes and X
555 chromosome, respectively, by taking the average of the median LOD across each cell type. A total
556 of 563 eQTL that exceeded the LOD thresholds and were no more than 1 Mbp from the
557 transcription start site of the associated eGene (**Supplementary Table S14**).
558

559 **Cell type proportion analysis**
560
561 To account for technical sources of variation often retained in scRNA-seq, cell type
562 proportions were transformed using the arcsin (asin) square root transformation from the speckle[62]

563 R package (version 0.0.3). Tests of statistical significance were performed using the propeller t-
564 test and ANOVA functions with default parameters. Sex of the mice and the batch each mouse
565 was associated with for sequencing were modeled as covariates. Transformed values were used as
566 input for computing tests of statistical differences of cell type proportions between mice, as well
567 as correlation to phenotypic traits (**Supplementary Table S3-S5**).
568

Acknowledgements

Research reported in this publication was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under award numbers R01AR68345, R01AR082880, and R01AR077992 to CRF.

Author Contributions

Luke J Dillard: Writing - review & editing; Visualization; Conceptualization; Methodology; Formal analysis; Data curation; Writing - original draft. Gina M Calabrese: Methodology. Larry D Mesner: Methodology. Charles R Farber: Writing - review & editing; Funding acquisition; Supervision; Conceptualization.

Disclosures

The authors declare that they have no conflicts of interest with the contents of this article.

586 **Figure Legends**
587
588 **Figure 1. Analysis of single cell RNA-seq (scRNA-seq) data for BMSC-OBs derived from 80**
589 **Diversity Outbred (DO)**
590 **(A)** Uniform Manifold Approximation and Projection (UMAP) of 139,392 single cells (BMSC-
591 OBs). Cell numbers and corresponding percentages for the fifteen (15) annotated cell clusters are
592 listed in parenthesis to the right of the annotated cluster name. **(B)** Dot plot[63] portraying
593 representative and highly expressed genes for all annotated cell clusters. Dot color indicates the
594 scaled average gene expression while the size of the dot corresponds to the percentage of cells of
595 a given cluster that express a given gene. **(C)** The raw proportional abundances of seven (7)
596 mesenchymal cell clusters and one (1) cluster (Hem) representing the remain cells (i.e.,
597 hematopoietic immune cells) across all 80 DO mice. **(D)** UMAP plots for mesenchymal lineage
598 cell clusters for DO mouse 50 and DO mouse 233. **(E)** CELLECT (CELL-type Expression-specific
599 integration for Complex Traits) cell type prioritization results. The seven (7) mesenchymal cell
600 clusters were significantly enriched ($P < 0.05$) for BMD heritability and the MALP, Ocy, and OB1
601 cell clusters were significantly enriched after correcting P-value for multiple testing. All remaining
602 cell clusters (i.e., hematopoietic immune cells) were not significant ($P > 0.05$)
603
604 **Figure 2. Overview of the network analysis pipeline**
605 *Step 1:* For all seven (7) of the mesenchymal lineage cell clusters (MPC, LMP, OBP, OB1, OB2,
606 Ocy, MALP), cell type-specific co-expression modules were generated using iterative Weighted
607 Gene Co-expression Network Analysis (iterativeWGCNA). *Step 2:* Bayesian networks were
608 learned to generate directed networks and model causal interactions between co-expressed genes.
609 *Step 3:* Differentiation Driver Genes (DDGs) were identified by extracting subnetworks (i.e., large
610 3-step neighborhood) for each gene in each cell type-specific Bayesian network and highlighting
611 those subnetworks that were enriched ($P_{adj} < 0.05$) for trajectory-specific tradeSeq genes for the
612 cell type boundary. *Step 4:* DDGs (and associated networks) were prioritized if the DDG was
613 identified previously as an expression/splicing quantitative trait loci (eQTL/sQTL) that colocalized
614 with BMD GWAS associations. Created with Biorender.com.
615
616 **Figure 3. Pseudotime Trajectory Inference analysis and establishment of cell type**
617 **boundaries for tradeSeq analysis**
618 **(A)** Three (3) trajectories (two adipogenic, one adipogenic) were inferred from the mesenchymal
619 cell clusters of the BMSC-OB scRNA-seq data using Slingshot. All trajectories originate from the
620 MPC and end in either osteogenic (Ocy, OB2) or adipogenic (MALP) cell fates. **(B)** For each of
621 the trajectories, cell type boundaries were generated using pseudotime values along the
622 trajectories, which encompass the majority of cells of a cell type mapping to their respective
623 trajectory. **(C)** Normalized gene expression of select genes associated with each cluster are
624 represented in feature plots (*top*) and each gene plotted as a function of pseudotime (*bottom*) for
625 all pseudotime trajectories (color corresponds to cell type annotation observed throughout).
626 Vertical lines (red) represent the cell type (pseudotime) boundaries established for each cell type
627 (label). The cell type boundary for OB1 and OB2 are represented as one red line/label for
628 visualization purposes.
629
630

631 **Figure 4. TradeSeq-identified genes associated with BMSC-OB differentiation exhibit**
632 **eQTL effects.**
633 **(A)** *Pkm* was identified as a significant global tradeSeq-identified gene ($P_{adj} = 8.35 \times 10^{-8}$) for LMP
634 cells along an osteogenic trajectory (LMP_to_OBP) (*left*). *S100a1* was identified as a significant
635 global tradeSeq-identified gene ($P_{adj} = 0.023$) for OBP cells along an osteogenic trajectory 1
636 (OBP_to_OB1) (*right*). **(B)** Plots indicating the cell type-specific expression quantitative trait loci
637 (eQTL) signal for both *Pkm* and *S100a1*. A negative eQTL effect on *Pkm* expression was observed
638 in LMPs for Diversity Outbred (DO) mice with a PWK haplotype background at the *Pkm* locus
639 (*left*). A positive eQTL effect on the expression of *S100a1* was observed in OBPs for DO mice
640 with a 129 haplotype background at the *S100a1* locus, while a negative effect was observed for
641 NZO mice (*right*). **(C)** The expression of *Pkm* and *S100a1* based on DO mouse (expression values
642 transformed via variance stabilizing transformation (VST), as described in Methods). Genotype
643 abbreviations correspond to DO haplotype background (legend) at the respective gene locus. Mice
644 with at least one PWK allele (genotype abbreviation G) tend to have decreased expression of *Pkm*
645 (*left*). Mice with at least one 129 allele (genotype abbreviation C) tend to have increased expression
646 of *S100a1*, while NZO mice (genotype abbreviation E) have decreased expression (*right*). **(D)**
647 PWK mice had a significant reduction in mature osteoblasts (OB1) and osteocyte-like cells (Ocy)
648 proportions relative to other mice ($P = 0.030$ and $P = 0.026$, respectively; t-test), while LMP
649 proportions were unaffected. Asterisks represent any of the other haplotype backgrounds. 129 mice
650 showed a significant decrease in LMP proportion and increase in OB1 proportion ($P = 0.008$ and
651 $P = 0.016$, respectively; t-test), but OBP proportions were unaffected. No significant effects on
652 cell type proportions were observed in NZO mice (**Supplementary Fig. 2**).
653
654 **Figure 5. Fgfrl1 and Tpx2 are prioritized DDGs and putative drivers of mesenchymal**
655 **differentiation.**
656 **(A)** *Fgfrl1* was identified as a Differentiation Driver Gene (DDG) of a network for LMPs
657 differentiating along an adipogenic trajectory. The network is enriched ($P_{adj} = 7.5 \times 10^{-3}$) for
658 trajectory-specific tradeSeq-identified genes for the LMP_to_MALP cell type boundary (*Hnmt,*
659 *St5, Igfbp4, Cyp1b1, Pdzrn4, Mark1*). *Fgfrl1* was previous identified as a gene that has BMD
660 GWAS associations that colocalize with an eQTL in the cultured fibroblast GTEx tissue. **(B)** *Tpx2*
661 was identified as a DDG of a network for LMPs differentiating along an osteogenic trajectory. The
662 network is enriched ($P_{adj} = 5.7 \times 10^{-7}$) for tradeSeq-identified genes for the LMP_to_OBP cell type
663 boundary (*Tpx2, Top2a, Kif4, Iqgap3, Prc1, Kif11, Ect2, Sgo2a, Ube2c*). *Tpx2* is both a tradeSeq
664 gene and previous identified as a gene that has BMD GWAS associations that colocalize with an
665 eQTL in the Testis GTEx tissue. **(C-D)** An increase in the expression of all tradeSeq-identified
666 genes coincides with the cell type boundary in which they were identified as significant along their
667 respective trajectories. **(E)** Box plot displaying whole-body bone mineral density (BMD)
668 measurements (excluding skull) from the International Mouse Knockout Consortium (IMPC) for
669 *Tpx2* mutant mice, which exhibited a significant increase in BMD (Genotype P-value = $1.03 \times 10^{-}$
670 ^{3}) in both male and female mice (N = 8 (M) and 8 (F) mutants; N = 2574 (M) and 2633 (F) controls)
671

References

1. Lin, J. T. & Lane, J. M. Osteoporosis: a review. *Clin. Orthop. Relat. Res.* 126–134 (2004).
2. Peacock, M., Turner, C. H., Econs, M. J. & Foroud, T. Genetics of osteoporosis. *Endocr. Rev.* **23**, 303–326 (2002).
3. Johnell, O. *et al.* Predictive value of BMD for hip and other fractures. *J. Bone Miner. Res.* **20**, 1185–1194 (2005).
4. Morris, J. A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019).
5. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
6. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **13**, e1006646 (2017).
7. Al-Barghouthi, B. M. *et al.* Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for human bone mineral density GWAS associations. *Elife* **11**, (2022).
8. Li, B. & Ritchie, M. D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Front. Genet.* **12**, 713230 (2021).
9. Akiyama, M. Multi-omics study for interpretation of genome-wide association study. *J. Hum. Genet.* **66**, 3–10 (2020).
10. Goh, K.-I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8685–8690 (2007).
11. Calabrese, G. M. *et al.* Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. *Cell Syst* **4**, 46-59.e4 (2017).
12. Sabik, O. L., Calabrese, G. M., Taleghani, E., Ackert-Bicknell, C. L. & Farber, C. R. Identification of a Core Module for Bone Mineral Density through the Integration of a Co-expression Network and GWAS Data. *Cell Rep.* **32**, 108145 (2020).
13. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
14. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
15. Zhang, Q. *et al.* Genomic variants within chromosome 14q32.32 regulate bone mass through MARK3 signaling in osteoblasts. *J. Clin. Invest.* **131**, (2021).
16. Al-Barghouthi, B. M. *et al.* Systems genetics in diversity outbred mice inform BMD GWAS and identify determinants of bone strength. *Nat. Commun.* **12**, 3408 (2021).
17. Dillard, L. J. *et al.* Single-Cell Transcriptomics of Bone Marrow Stromal Cells in Diversity Outbred Mice: A Model for Population-Level scRNA-Seq Studies. *J. Bone Miner. Res.* **38**, 1350–1363 (2023).
18. Abood, A. *et al.* Long-read proteogenomics to connect disease-associated sQTLs to the protein isoform effectors of disease. *bioRxiv* 2023.03.17.531557 (2023) doi:10.1101/2023.03.17.531557.
19. Bogue, M. A., Churchill, G. A. & Chesler, E. J. Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database. *Mamm. Genome* **26**, 511–520 (2015).

20. Churchill, G. A., Gatti, D. M., Munger, S. C. & Svenson, K. L. The Diversity Outbred mouse population. *Mamm. Genome* **23**, 713–718 (2012).
21. Zhong, L. *et al*. Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* **9**, (2020).
22. Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *Elife* **9**, (2020).
23. GTEx Consortium *et al*. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
24. Greenfest-Allen, E., Cartailler, J.-P., Magnuson, M. A. & Stoeckert, C. J. iterativeWGCNA: iterative refinement to improve module detection from WGCNA co-expression networks. *bioRxiv* 234062 (2017) doi:10.1101/234062.
25. Chan, C. K. F. *et al*. Identification of the Human Skeletal Stem Cell. *Cell* **175**, 43-56.e21 (2018).
26. Mizuhashi, K. *et al*. Resting zone of the growth plate houses a unique class of skeletal stem cells. *Nature* **563**, 254–258 (2018).
27. Debnath, S. *et al*. Discovery of a periosteal stem cell mediating intramembranous bone formation. *Nature* **562**, 133–139 (2018).
28. Matsushita, Y. *et al*. A Wnt-mediated transformation of the bone marrow stromal cell identity orchestrates skeletal regeneration. *Nat. Commun.* **11**, 332 (2020).
29. Van den Berge, K. *et al*. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
30. Thomas, P. D. *et al*. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22 (2022).
31. Groza, T. *et al*. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038–D1045 (2023).
32. Enerbäck, S., Ohlsson, B. G., Samuelsson, L. & Bjursell, G. Characterization of the human lipoprotein lipase (LPL) promoter: evidence of two cis-regulatory regions, LP-alpha and LP-beta, of importance for the differentiation-linked induction of the LPL gene during adipogenesis. *Mol. Cell. Biol.* **12**, 4622–4633 (1992).
33. Federico, L. *et al*. Lipid phosphate phosphatase 3 regulates adipocyte sphingolipid synthesis, but not developmental adipogenesis or diet-induced obesity in mice. *PLoS One* **13**, e0198063 (2018).
34. Maridas, D. E., DeMambro, V. E., Le, P. T., Mohan, S. & Rosen, C. J. IGFBP4 Is Required for Adipogenesis and Influences the Distribution of Adipose Depots. *Endocrinology* **158**, 3488–3500 (2017).
35. Sigg, M. A. *et al*. Evolutionary Proteomics Uncovers Ancient Associations of Cilia with Signaling Pathways. *Dev. Cell* **43**, 744-762.e11 (2017).
36. Kumar, R. *et al*. A cell-based GEF assay reveals new substrates for DENN domains and a role for DENND2B in primary ciliogenesis. *Sci Adv* **8**, eabk3088 (2022).
37. Fumoto, K. *et al*. Mark1 regulates distal airspace expansion through type I pneumocyte flattening in lung development. *J. Cell Sci.* **132**, (2019).
38. Miki, H., Setou, M., Kaneshiro, K. & Hirokawa, N. All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7004–7011 (2001).
39. Zhang, R., Roostalu, J., Surrey, T. & Nogales, E. Structural insight into TPX2-stimulated microtubule assembly. *Elife* **6**, (2017).

40. Uusküla-Reimand, L. & Wilson, M. D. Untangling the roles of TOP2A and TOP2B in transcription and cancer. *Sci Adv* **8**, eadd4920 (2022).
41. Trueb, B. Biology of FGFRL1, the fifth fibroblast growth factor receptor. *Cell. Mol. Life Sci.* **68**, 951–964 (2011).
42. Hilgendorf, K. I. Primary Cilia Are Critical Regulators of White Adipose Tissue Expansion. *Front. Physiol.* **12**, 769367 (2021).
43. Fazeli, P. K. *et al.* Marrow fat and bone--new perspectives. *J. Clin. Endocrinol. Metab.* **98**, 935–945 (2013).
44. Veldhuis-Vlug, A. G. & Rosen, C. J. Clinical implications of bone marrow adiposity. *J. Intern. Med.* **283**, 121–139 (2018).
45. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 75 (2017).
46. Cheng, Y., Ma, X., Yuan, L., Sun, Z. & Wang, P. Evaluating imputation methods for single-cell RNA-seq data. *BMC Bioinformatics* **24**, 302 (2023).
47. Yu, X., Abbas-Aghababazadeh, F., Chen, Y. A. & Fridley, B. L. Statistical and Bioinformatics Analysis of Data from Bulk and Single-Cell RNA Sequencing Experiments. *Methods Mol. Biol.* **2194**, 143–175 (2021).
48. Church, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.* **7**, e1000112 (2009).
49. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
50. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
51. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
52. Neavin, D. *et al.* Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol.* **22**, 76 (2021).
53. van der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
56. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
57. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
58. Cartailler, J. P. *Iterativewgcna*. (2022).
59. Tsamardinos, I., Brown, L. E. & Aliferis, C. F. The max-min hill-climbing Bayesian network structure learning algorithm. *Mach. Learn.* **65**, 31–78 (2006).
60. Porter, M. D. & Smith, R. Network neighborhood analysis. in *2010 IEEE International Conference on Intelligence and Security Informatics* (IEEE, 2010). doi:10.1109/isi.2010.5484781.

61. Broman, K. W. *et al*. R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics* **211**, 495–502 (2019).
62. Phipson, B. *et al*. propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).
63. Marsh, S., Salmon, M. & Hoffman, P. *Custom Visualizations & Functions for Streamlined Analyses of Single Cell Sequencing*. (2023). doi:10.5281/zenodo.7534950.