Mapping the Single-cell Differentiation Landscape of Osteosarcoma

Danh D. Truong[1]*, Corey Weistuch[2]*, Kevin A. Murgas[3]*, Joseph O. Deasy[2], Antonios G. Mikos[4], Allen Tannenbaum[5,6+], and Joseph Ludwig[1+]

*Authors contributed equally

1 Department of Sarcoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

2 Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY

3 Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY

4 Department of Bioengineering, Rice University, Houston, TX

5 Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY

6 Department of Computer Science, Stony Brook University, Stony Brook, NY

+ Co-corresponding authors:

Allen Tannenbaum, Stony Brook University, Stony Brook, NY 11894. Phone: (631) 632-8654; E-mail: allen.tannenbaum@stonybrook.edu

Joseph Ludwig, The University of Texas MD Anderson Cancer Center, Houston, TX 77030. Phone: (713) 792-4265; E-mail: jaludwig@mdanderson.org

The authors declare no conflicts of interest.

Abstract

The genetic and intratumoral heterogeneity observed in human osteosarcomas (OS) poses challenges for drug development and the study of cell fate, plasticity, and differentiation, processes linked to tumor grade, cell metastasis, and survival. To pinpoint errors in OS differentiation, we transcriptionally profiled 31,527 cells from a tissue-engineered model that directs MSCs toward adipogenic and osteoblastic fates. Incorporating pre-existing chondrocyte data, we applied trajectory analysis and non-negative matrix factorization (NMF) to generate the first human mesenchymal differentiation atlas. This 'roadmap' served as a reference to delineate the cellular composition of morphologically complex OS tumors and quantify each cell's lineage commitment. Projecting these signatures onto a bulk RNA-seq OS dataset unveiled a correlation between a stem-like transcriptomic phenotype and poorer survival outcomes. Our study takes the critical first step in accurately quantifying OS differentiation and lineage, a prerequisite to better understanding global differentiation bottlenecks that might someday be targeted therapeutically.

Statement of Significance:

OS treatment kills proliferating cells without addressing the root cause: dysregulated differentiation. By deconvolving OS tumors by cell type and differentiation archetype, we identified core gene sets linked to cell fate and patient survival. The ability to quantify, and eventually modulate, such archetypes facilitate a novel OS-specific drug-screening strategy.

Introduction

Based upon their significant genetic, phenotypic, and lineage-specific diversity, the World Health Organization splits sarcomas into more than fifty unique sarcoma subtypes[1]. Many sarcomas harbor specific chromosomal translocations, oncogenes, or lost tumor suppressors that are used both as diagnostic markers and as potential therapeutic targets[2]. Although genetically heterogeneous, sarcomas are often classified based on their apparent differentiation status and cell types within the adult mesenchymal lineage that they most resemble. High-grade osteosarcoma (OS), a heterogeneous class of poorly differentiated bone sarcomas, is further subclassified based on predominant tissue features resembling osteoblastic, chondroblastic, fibroblastic, and telangiectatic[3,4].

With a few rare exceptions, such as p53-mutated tumors in people with Li Fraumeni syndrome, most OS have a complex karyotype that is caused by chromothripsis and lacks a recurring gene that can be targeted [5]. Although a variety of genetic mutations and genomic structural events (translocations, copy alterations) have been linked to OS, these are often unique to each patient[6]. This presents an obvious challenge in developing therapeutic options broadly applicable to OS patients. Nevertheless, one commonality among high-grade OS is their genetic drift from normal osteoblasts and their varied degrees of dedifferentiation observed along mesenchymal lineages[6].

In contrast to mutation-based drivers that might be predicted to affect cell fate, differentiation, and plasticity more subtly, we hypothesized that accumulated DNA damage acts to collectively disrupt the gene regulatory networks (GRNs) and epigenetic patterns that are critically important to constrain normal osteoblasts in their fully differentiated phenotype[7,8]. The pivotal role a constant epigenetic state (or memory) plays in dictating cell fate, and the adverse effects noted when this goes awry, has been described as an 'emerging cancer hallmark' in Hanahan & Weinberg 2022[9].

Although single-cell RNA-sequencing (scRNA-seq) has emerged as a promising approach to capture gene expression profiles of individual cells, identifying differentiation states and quantifying stemness still pose unique challenges. Recent tools such as StemID[10], SCENT[11], SLICE[12], and CytoTRACE[13] have been developed to quantify stemness, but they cannot identify lineage-specific differentiation states. Unlike hematopoiesis that begins during embryogenesis and continues throughout adulthood and is easily studied with routine blood draws, a similar mesenchymal 'Waddington landscape' doesn't exist and can't readily be studied since the majority of connective tissue development occurs during early fetal development. In rare scenarios (e.g., during bone fracture repair), MSCs transform into chondro/osteogenic precursors that form a callus and subsequent bone, but this hasn't yet been studied using single-cell technologies in vivo. One method to enrich for cells in transit from one lineage to another, and the approach taken herein, relies upon an ex vivo tissue engineered model adapted from established methods developed by the regenerative medicine field in their attempt to create functional connective tissues[14]. In addition to the engineering challenges that we address here, projecting cancer onto a normal differentiation landscape is fundamentally challenging because tumor expression profiles only partially overlap with normal tissue expression profiles.

The present work demonstrates a joint experimental and computational strategy for mapping osteosarcoma onto its underlying mesenchymal differentiation landscape. We developed a high-resolution scRNA-seq reference map composed of osteogenic, adipogenic, and chondrogenic lineages from human primary mesenchymal stem cells, described herein as the Mesenchymal Tissue Landscape (MTL). We hypothesized that the MTL would catalog the various differentiation states possible in OS. To quantify differentiation states in the MTL, we applied a Normalized Nonnegative Matrix Factorization (N-NMF) based archetype analysis to identify recurring gene expression profiles that accurately captured lineage-specific temporal dynamics[15]. These profiles were then used to estimate the relative compositions of each cell lineage and differentiation status within OS tumor data, including patient-derived xenograft (PDX) and tumor biopsy samples[16,17]. We show that the estimated predominant cell types of each tumor broadly agreed with pathologist labels of predominant tissue features (osteogenic, chondrogenic) and present a more intricate analysis that allows one to quantify features of multiple cell lineages within each tumor. That data allowed us to examine the gene expression patterns with respect to patient outcomes and other clinical features, finding that a signature of advanced differentiation was associated with improved survival outcomes and, conversely, that a stem-like signature bodes a poor prognosis.

The capacity of our approach to determine lineage and differentiation states in distinct datasets suggests that it can be more generally applied to different cancers along the mesenchymal differentiation landscape or possibly even carcinomas where differentiation landscapes remain in their infancy.

Results

Intratumoral Heterogeneity in Osteosarcoma PDXs

As proof of concept before using human tissues, we sequenced three OS PDXs to explore the heterogeneity of cell types and differentiation states using single-cell transcriptomics. The OS PDXs were isolated upon reaching 150mm^3 and underwent rapid dissociation. After filtering for high-quality cells by scRNA-seq metrics, we obtained 19,538 cells for analysis. Initial visualization of the uniform manifold approximation and projection (UMAP) showed that cells clustered mainly by PDX sample (Fig. 1A). Since OSs

exhibit marked intratumoral heterogeneity[6], we selected classical markers of osteoblasts, chondroblasts, and fibroblasts that describe the predominant cell states possible within each PDX (Fig. 1C). Significant heterogeneity in osteoblastic markers among the three PDXs was observed. The OS1 PDX displayed the strongest expression of the osteoblastic markers *RUNX2*, *CDH11*, and *SPP1* but also contained a subpopulation of fibroblastic cells and another that lacked strong osteoblastic or fibroblastic markers. Conversely, the OS31 PDX was predominantly chondroblastic and had the strongest expression of the chondroblastic markers *COL2A1*, *SOX9*, and *ACAN*. SA98 was not predominantly any one differentiation state but had expression of osteoblastic, chondroblastic, and fibroblastic markers. All PDXs also had a subpopulation of cycling cells, as demonstrated by the expression of *MKI67* and *TOP2A*. We also explored markers of EMT and stemness but did not detect a significant trend in the PDXs, which might explain the expected rarity of cancer stem cells.

Given the striking degree of intratumoral heterogeneity, we used a graph-based clustering method to identify 15 distinct clusters (Fig. 1B). While many clusters are composed of one PDX, some comprise more than one, suggesting the presence of conserved states between different OS PDXs. Using the clusters, we performed differentially expressed gene (DEG) analysis and ranked the top-expressed markers for each cluster (Fig. 1D). Clusters 4 and 13 are comprised of mostly SA98 and OS1 PDXs with fibroblastic genes (*COL3A1*, *FBN1*, *ACTA2*). In addition, we also observed evidence of fibroblastic expression for SA98 in Clusters 3 and 7 (*POSTN, FGFR1, FGFBP2*), some evidence of a chondroblastic phenotype (*COL11A1*), and a lack of osteoblastic expression. This suggests that SA98 may be a chondroblastic-fibroblastic hybrid. Clearly, clusters 8, 10, and 12 were indicative of cells undergoing cycling. Enrichment analysis showed four possible major gene expression programs that are active in these three PDXs. Many clusters (0, 2, 3, 4, 10, 14), which were predominantly composed of OS31 and SA98, were associated with ECM-related gene sets (Fig. 1E). Interestingly, the predominantly osteoblastic OS1 lacked many of the ECM gene sets. Conversely, Cluster 1 from OS1 had heightened activity of osteoblast differentiation and bone development genes. Interestingly, cluster 13 was enriched with contractile-related processes, including genes such as *TPM2* and *ACTA2*.

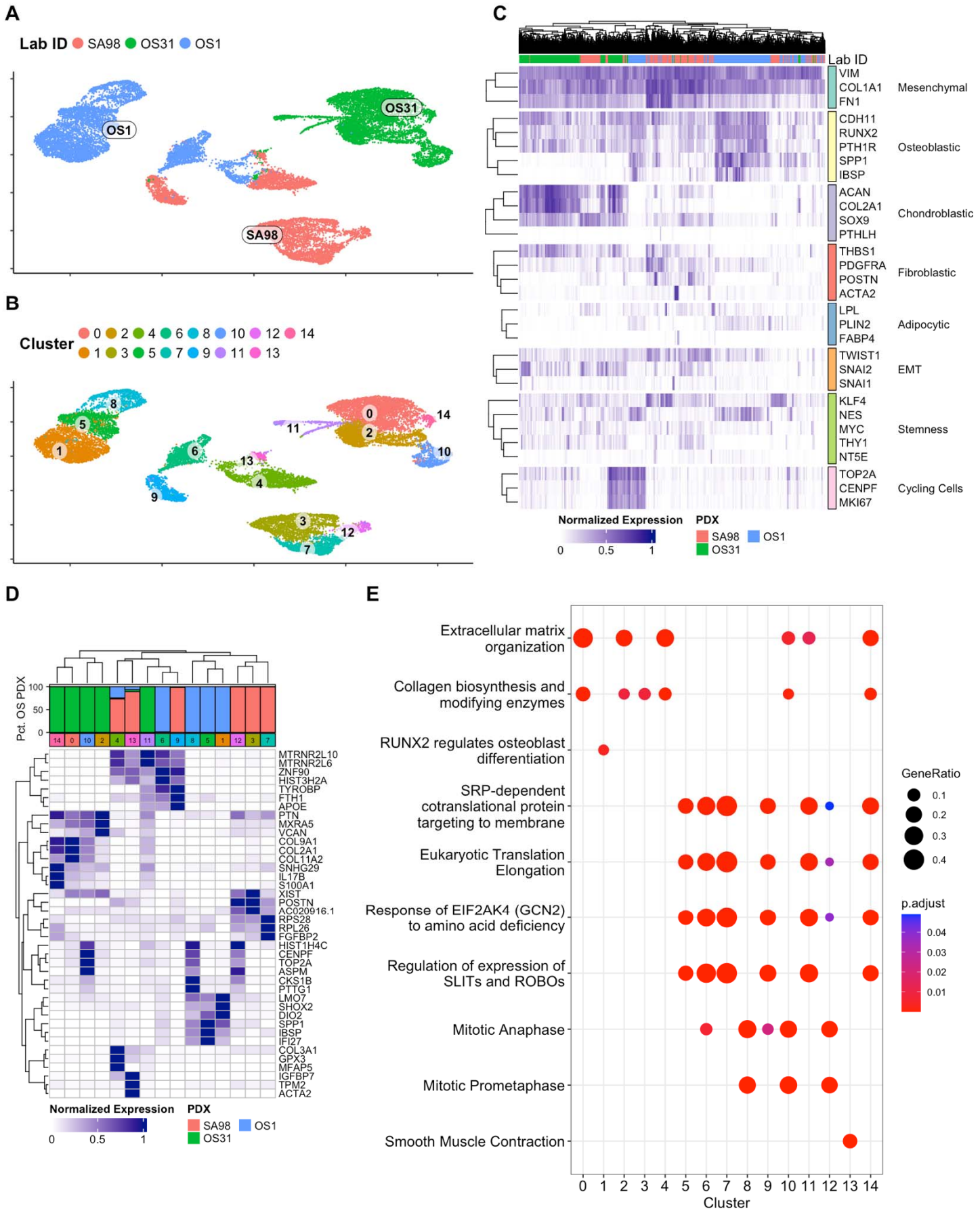
A
Lab ID SA98 OS31 OS1
OS1
OS31
SA98
B
Cluster 0 2 4 6 8 10 12 14 1 3 5 7 9 11 13
C
Lab ID
VIM
COL1A1
FN1
Mesenchymal
CDH11
RUNX2
PTH1R
SPP1
IBSP
Osteoblastic
ACAN
COL2A1
SOX9
PTHLH
Chondroblastic
THBS1
PDGFRA
POSTN
ACTA2
Fibroblastic
LPL
PLIN2
FABP4
Adipocytic
TWIST1
SNAI2
SNAI1
EMT
KLF4
NES
MYC
THY1
NT5E
Stemness
TOP2A
CENPF
MKI67
Cycling Cells
Normalized Expression
0 0.5 1
PDX
SA98 OS1
OS31
D
Pct. OS PDX
100 50 0
14 0 10 2 4 13 11 6 9 8 5 1 12 3 7
MTRNR2L10
MTRNR2L6
ZNF90
HIST3H2A
TYROBP
FTH1
APOE
PTN
MXRA5
VCAN
COL9A1
COL2A1
COL11A2
SNHG29
IL17B
S100A1
XIST
POSTN
AC020916.1
RPS28
RPL26
FGFBP2
HIST1H4C
CENPF
TOP2A
ASPM
CKS1B
PTTG1
LMO7
SHOX2
DIO2
SPP1
IBSP
IFI27
COL3A1
GPX3
MFAP5
IGFBP7
TPM2
ACTA2
Normalized Expression
0 0.5 1
PDX
SA98 OS1
OS31
E
Extracellular matrix organization
Collagen biosynthesis and modifying enzymes
RUNX2 regulates osteoblast differentiation
SRP-dependent cotranslational protein targeting to membrane
Eukaryotic Translation Elongation
Response of EIF2AK4 (GCN2) to amino acid deficiency
Regulation of expression of SLITs and ROBOs
Mitotic Anaphase
Mitotic Prometaphase
Smooth Muscle Contraction
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Cluster
GeneRatio
0.1 0.2 0.3 0.4
p.adjust
0.04 0.03 0.02 0.01

Figure 1: Single-Cell Sequencing Reveals Intratumor Heterogeneity in Patient-Derived Xenograft Models of Osteosarcoma. A: Single-cell gene expression UMAP of three OS PDX samples. **B:** Louvain clustering identified 15 clusters of cells in the three PDXs. **C:** Heatmap of gene expression markers (normalized by maximum across all cells) of all individual cells. Genes were grouped according to specific lineage markers or biological processes. **D:** Average cluster gene expression (normalized by maximum across all clusters) of the set of top 5 differentially expressed genes in each cluster. **E:** Dot plot of pathway analysis scores for each cluster.

Constructing a Map of the Mesenchymal Transition Landscape

Since such few cells transit mesenchymal linages postnatally, it isn't yet possible to detect and profile these rare cell types in sufficient numbers using existing technologies in children or adults. Instead, we utilized an *in vitro* model of human mesenchymal differentiation[14](Fig. 2A). Initially, scRNA-seq was used to profile the mesenchymal differentiation landscape of MSCs under adipogenic or osteogenic conditions to map the changing transcriptome as cells reached their intended terminal fate. We applied a differentiation model involving biochemical and biophysical signals to direct osteogenic and adipogenic tissue differentiation[14]. We then collected cells at defined time points to generate a single-cell time-course profile of each lineage.

Because our differentiation strategy only modeled osteogenic and adipogenic differentiation, we could not directly estimate the chondrogenic lineage. Chondrogenesis is an important mesenchymal-origin cell lineage and constitutes a major component of some chondroblastic OS tumors[6,18]. For completeness, we bolstered our experimentally generated osteogenic and adipogenic MTL with a publicly available chondrogenic lineage single-cell dataset[19]. In brief, the study generated human iPSC-derived chondroblast precursor cells that were chemically stimulated to undergo chondrogenic differentiation. Throughout their MSC-to-chondrocyte lineage transition, gene expression was assessed by scRNA-seq. With similar time-course data, a batch correction procedure was performed to integrate the chondrogenic lineage with the experimentally generated osteogenic and adipogenic lineages. Subsequently, these data were examined to investigate lineage-specific and temporal gene expression dynamics on the mesenchymal tissue landscape (MTL; Fig. 2B-E).

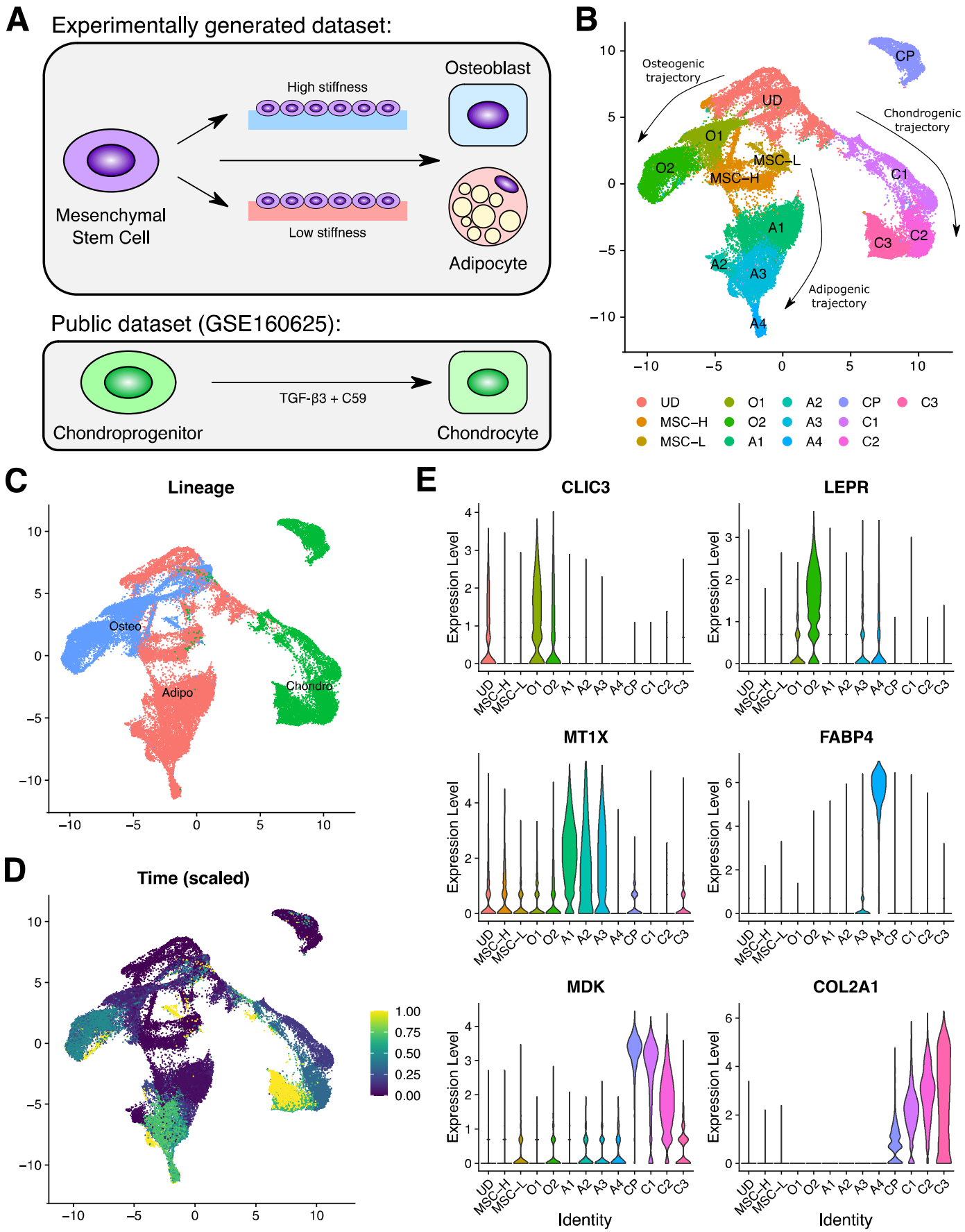

Figure 2: Mesenchymal Differentiation Landscape. A: Schematic of datasets utilized to construct the Mesenchymal Differentiation Landscape. Osteogenic and adipogenic lineages were experimentally generated

over a time course culturing on hydrogels of varying stiffness to influence differentiation trajectory. The chondrogenic lineage was sourced from a publicly available dataset (GSE160625) which measured a time course of chondrogenesis in cultured chondroprogenitor cells treated with a combination of TGF-β3 and C59. These two datasets were integrated to construct a mesenchymal differentiation map containing three lineages. **B:** UMAP of integrated differentiation landscape with designated clusters along three distinct lineages. Clusters were manually annotated as undifferentiated/dividing (UD), chondroprogenitor (CP), mesenchymal stem cell (MSC), osteogenic (O1-O2), adipogenic (A1-A4), chondrogenic (C1-C3). **C:** UMAP colored by lineage. **D:** UMAP colored by experimental time, scaled to the endpoint of each experiment (d21 for osteo, d14 for adipo, d42 for chondro). **E:** Violin plots of marker genes for early and late stages of each lineage showing distinct temporal patterns.

Using the integrated cell embedding produced by Harmony, we generated a Uniform Manifold Approximation and Projection (UMAP) to visualize the topological structure of our single-cell data[20]. Using Louvain graph-based clustering of the Harmony embedding[20], we generated 13 distinct clusters representative of MSCs, osteoblasts, adipocytes, chondrocytes, and transitional states. We utilized the known condition and time-point data to manually annotate the clusters (Fig. 2B-E)[14]. We generated a correlation matrix between average expressions within each identified cluster and condition. Our analysis identified three osteo-lineage clusters (O1-O3), four adipo-lineage clusters (A1-A4), three chondro-lineage clusters (C1-C3), in addition to undifferentiated/dividing cells (UD), a distinct chondroprogenitor cluster (CP), and two populations of MSCs (MSC-H and MSC-L).

Gene Markers Validate the MTL Map and Quantify Differentiation along Multiple Mesenchymal Lineages

To determine cluster-specific markers and differentially expressed genes (DEGs), we performed pairwise differential expression analysis in Seurat for each cluster against the rest of the clusters (Supplementary Table 1). We observed high expression of cell-cycle genes (*CENPF, MKI67, TOP2A*) in UD and CP clusters and to some degree in cluster C1. This was further illustrated by scoring a cell-cycle gene set and identifying the G2M, S, and G1 cell-cycle phases. We observed clear localization of cycling cells in G2 and S phases in the UD and CP clusters. Clusters MSC-H and MSC-L, which consisted mostly of MSCs prior to induction, were enriched for MSC-specific markers (*ENG, NT5E, PRRX1, THY1*). Since MSCs were cultured on gels of different stiffnesses, we were able to identify YAP/TAZ-regulated genes (*ANKRD1, CTGF, CYR61, IGFBP5, TEAD1*) enriched in cluster MSC-H versus cluster MSC-L (Supplemental Fig. S1).

Adipogenic Lineage

Cluster A1, containing early time-points within the adipo-lineage trajectory[21], was enriched with metallothioneins (MTs). MTs are expressed in adipose tissue and have been shown to regulate adipogenic differentiation[22]. Consequentially, clusters A1, A2, and A3 were highly enriched with MTs (*MT1X, MT1E, MT1M, MT2A*), echoing similar literature results within 24h of adipogenic induction[23]. Clusters A1 and A2 were enriched with *WNT5A, PAPPA,* and *FTH1*. *WNT5A* is a part of the non-canonical Wnt pathway activated during adipogenesis[24]. Both *PAPPA* and *FTH1* are up-regulated during activation of peroxisome proliferator-activated receptors (PPARs)[25]. Cluster A3 appeared to be heavily enriched with extracellular matrix (ECM) genes (*COL3A1, COL6A1, COMP, DCN, LUM*). Though not initially enriched during the early phase of adipogenesis, these genes appear to peak within cluster A3. Likewise, various groups have observed that adipogenesis follows a biphasic pattern concerning ECM proteins, where collagens, laminins, biglycan, and lumican, demonstrated a decrease upon adipogenic induction but return to basal levels after a certain time[26]. This was followed by a phase of cell growth and fat storage. As such, cluster A4 is highly enriched for adipocyte gene signatures (*ACACB, ADIPOQ, APOE, FABP4, G0S2, FABP5, LPL, PLIN4, PLIN1*). Though the adipogenic lineage served as a negative control for our OS-focused study, it might prove useful in future studies aimed at unraveling the epigenetic drivers of well-differentiated and dedifferentiated liposarcoma, common sarcoma subtypes that lack effective biologically targeted therapies.

Osteogenic Lineage

Our analysis showed that cluster O1 up-regulated genes related to both YAP/TAZ signaling and early osteogenesis, such as *ANKRD1, AXL, CTGF, DKK1,* and *HHIP.* Activation of YAP/TAZ signaling has been shown to directly target *ANKRD1, AXL, CTGF,* and *DKK1* when cells are grown at high stiffness prior to osteogenesis[27-29]. In addition, the expression of *HHIP* suggests regulation of Hedgehog signaling. Previous

reports stated that YAP activation drove GLI2 nuclear accumulation, which directly activated GLI targets[30,31]. GLI target up-regulation (*CCND1, FOXC2, GLIPR1,* and *IGFBP6*) within cluster O1 of our model acts to validate this finding experimentally. Within clusters O1 and O2, we observed increased expression of ECM genes (*ACAN, COL1A1, COL1A2, COL8A1,* and *ELN*). Furthermore, expressions of *CRYAB, LEPR,* and *TAGLN* indicate positive regulation of osteogenesis[21,32,33]. *COL1A1* and *COL1A2* are characteristic of osteoprogenitor cells, while *ACAN* was expressed in osteochondroprogenitor cells. Additionally, many genes (*DCN, FN1, GSN, IGFBP2, IGFBP6*, and *SERPINE2*) of the osteoblast secretome were detected within cluster O2[34]. Cluster O2 was also highly enriched with additional osteoblast secretome-related genes (*CHI3L1, FBLN1, SAA1, THBS2*). *SAA1* expression has been shown to be induced by osteogenic conditions while also promoting osteogenesis and bone mineralization[35]. In addition, IGFPs, including *IGFBP2, IGFBP4,* and *IGFBP7,* and *IGF2* are some of the osteogenic growth factors up-regulated in cluster O2[36]. Expression of transcription factors *FOS* and *CTNNB1* (β-catenin) indicates activation of mechanosensing pathways in osteoblasts[36]. Osteoblast-specific glycoproteins and ECM proteins (*ECM2, GPNMB*) were highly expressed in cluster O2[37-40]. Likewise, WNT1-Inducible-Signaling Pathway Protein 2 (*WISP2*) was highly expressed in cluster O2 and has been shown to both promote osteogenesis and repress adipogenesis[41,42].

Chondrogenic Lineage

Within the chondrogenic differentiation experimental time course, cluster C1 consisted of primarily day 7, cluster C2 consisted mostly of day 14, and cluster C3 consisted of days 28 and 42. In the chondrogenic lineage, clusters C1, C2, and C3 showed increased expression of chondrogenic markers (*COL9A1, MATN4, SOX9*)[19,43]. All three chondrogenic clusters also expressed frizzled-related proteins (*EPYC, FRZB, LECT1*)[44]. Clusters C1 and C2, containing earlier time points, exhibited a few early markers of chondrogenic differentiation (*SOX2, SOX6*), while clusters C2 and C3, containing later time points, exhibited increased expression of additional chondrogenic markers (*ACAN, COL2A1, COL3A1*)[45]. The CP cluster was observed to show distinct transcriptional signatures compared to the rest of the mesenchymal landscape, with expression of genes indicating chondrogenic potential (*SOX2, SOX4*) and some neural crest markers (*FOXD3, PAX3, PAX6, OTX2*) and WNT signaling genes (*MAPK10, WNT4*) reflecting the derived progenitor status of this cluster[46,47].

Archetype Analysis Defines Lineage-Specific and Time-Dependent Expression Profiles within the MTL

Next, we sought to quantitate time-dependent signatures of mesenchymal stem cell differentiation along the MTL-derived lineages. We applied archetype analysis in order to identify temporal and lineage-specific gene expression signatures and subsequently estimate the relative contribution of each archetype in individual cells[48]. Twelve distinct differentiation signatures (Supplemental Fig. S2) captured the MTL's gene expression variability (Fig. 3A). The estimated abundance of each archetype varied with time and with experimental manipulation corresponding to specific differentiation lineages. Archetypes 1-4, for example, each peaked at a different experimental time (measured in days) and thus encode differentially expressed genes common to both adipocytes (red) and osteoblasts (blue) as they differentiate. By contrast, Archetypes 5-7 were specific to early adipocytes, late adipocytes, and osteoblasts respectively. Chondrocytes (green), by contrast and perhaps explained by the distinct experimental design involved in their profiling, were associated with archetypes 8-12, each peaking at a distinct experimental time.

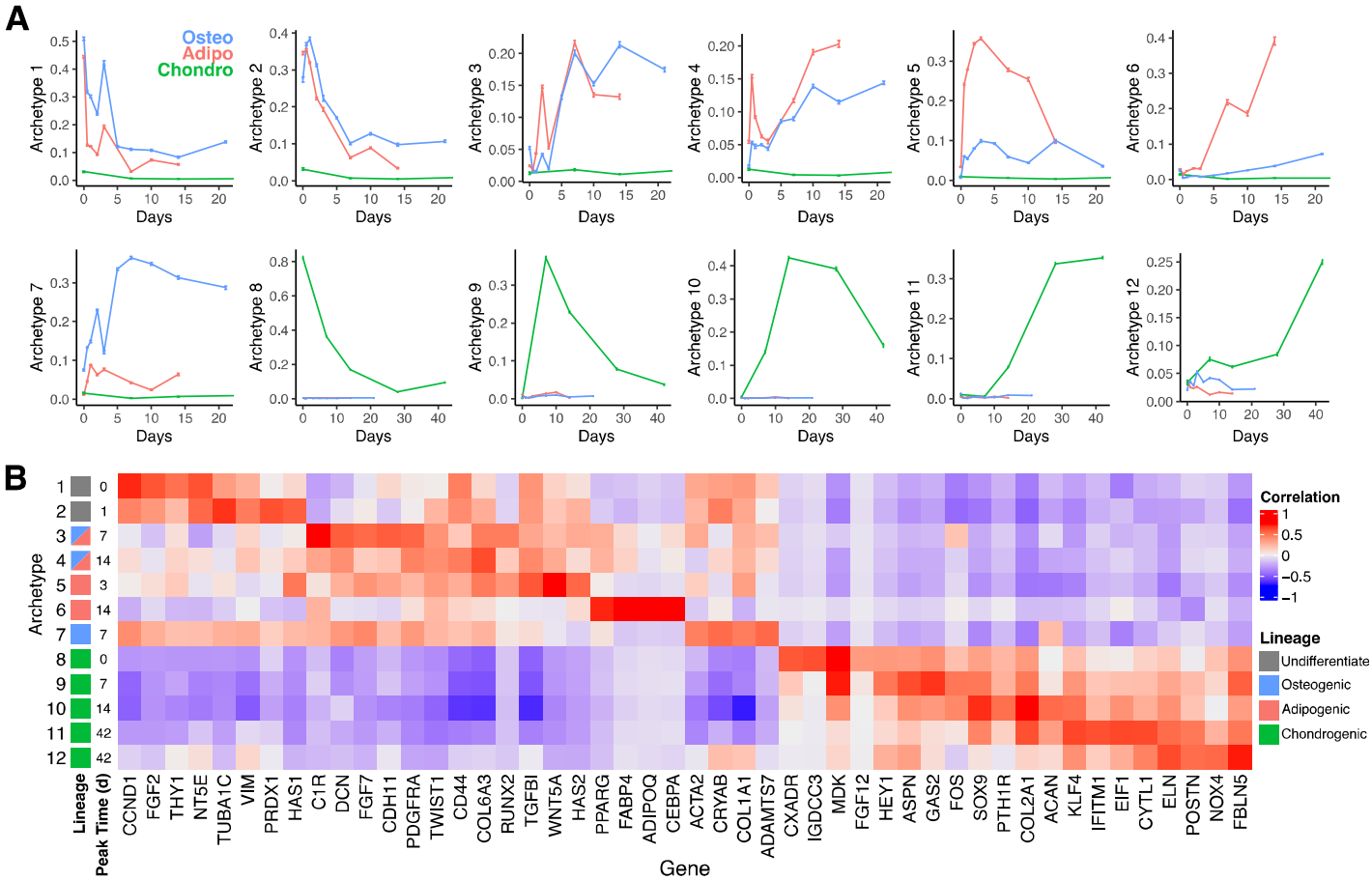

Figure 3: MTL archetypes. A: Average archetype time courses stratified by cell lineage. Archetypes scores for each cell were computed using normalized nonnegative matrix factorization. **B:** Table summarizing the dominant lineage and peak time represented by each archetype. **C:** Heatmap of representative gene-archetype correlations. Four representative genes were selected among the top Pearson correlates of each archetype based on their known biological relevance. The full list of correlates is provided in SI Table 1.

To confirm that these archetypes capture established lineage-specific and stemness-associated features of gene expression, we conducted a genome-wide Pearson correlational analysis between the expression levels of each gene in our dataset and the single-cell scores assigned to each archetype (Fig. 3B). Consistent with the time-course averages, we observed that Archetype 1 exhibited a strong association with well-known stemness genes such as *FGF2* and *THY1*, confirming that this archetype represented an early, undifferentiated stem signature. Furthermore, Archetypes 6, 7, and 12 were associated lineage-specific adipogenic (*FABP4*), osteogenic (*COL1A1*), and chondrogenic (*ACAN*) marker genes, respectively (see Fig. 1C).

An Early Differentiation Stem-like State Predominates in Human Osteosarcoma

To understand osteosarcoma heterogeneity in the context of mesenchymal stem cell differentiation, we projected single-cell expression data from human osteosarcoma samples onto the 12 archetypes defined by the MTL. With the signatures trained on the MTL, we estimated the distribution of differentiation states across the OS PDX and OS tumor cohorts (Fig. 4A, B). Unlike mature cell lineages of the MTL derived from normal cells, individual OS cells exhibited greater phenotypic diversity and a mixture of archetypes from different lineages. Some cells in SA98, for example, scored highly for both archetypes 3 and 9, while some in OS31 scored highly for archetypes 2, 6, and 10. The extent of archetype promiscuity within the same tumors raised the possibility that OS has sufficient epigenetic plasticity to enable the multi-phenotypic morphology observed in most high-grade OS tumors.

To further assess the composition of differentiation states within each OS tumor, we further classified cells by their most prevalent archetype and found that each tumor had a widely variable fraction of cells in different states (Fig. 4C, D). Both OS datasets revealed populations of cells primarily composed of archetypes 2 (Day 0.5 MTL), 3 (Day 7 MTL), 6 (late adipocytes), and 10 (Day 14-28 chondrocytes), with smaller populations containing the other archetypes.

Importantly, the estimated fraction of chondrogenic lineage closely matched the histological classification of chondroblastic origin in PDX OS31 and OS tumors BC20 and BC22. In addition, individual tumors (such as SA98 and BC21) contained multiple populations of cells with multiple lineages and marked variability in their differentiation states, while other tumors (BC5, BC6, and BC17) were composed of mostly early differentiation states without lineage commitment.

An unexpected finding was that the PDX samples were more phenotypically diverse than the human tumors and potentially more differentiated along the osteochondroblastic lineage. If validated prospectively using a larger sample size, this might have implications for preclinical drug testing programs. As expected for high-grade OS clinical samples, our computational approach correctly identified tumor cells in an early progenitor state that exists prior to late-stage osteochondrogenic differentiation. The ability to quantify cellular constituents along this continuous stemness/differentiation spectrum provides a new dimension to characterize OS that might prove more useful than subjective cell type (e.g., chondrocytes or osteoblasts)-based approaches used by pathologists today. Since a minority of tumor cells appear to have undergone differentiation into distinct cell fates, this might explain why pathology-assigned subclassifications of OS into fibroblastic, osteoblastic, or chondroblastic subtypes has little prognostic value and no bearing on a clinician's treatment choice.

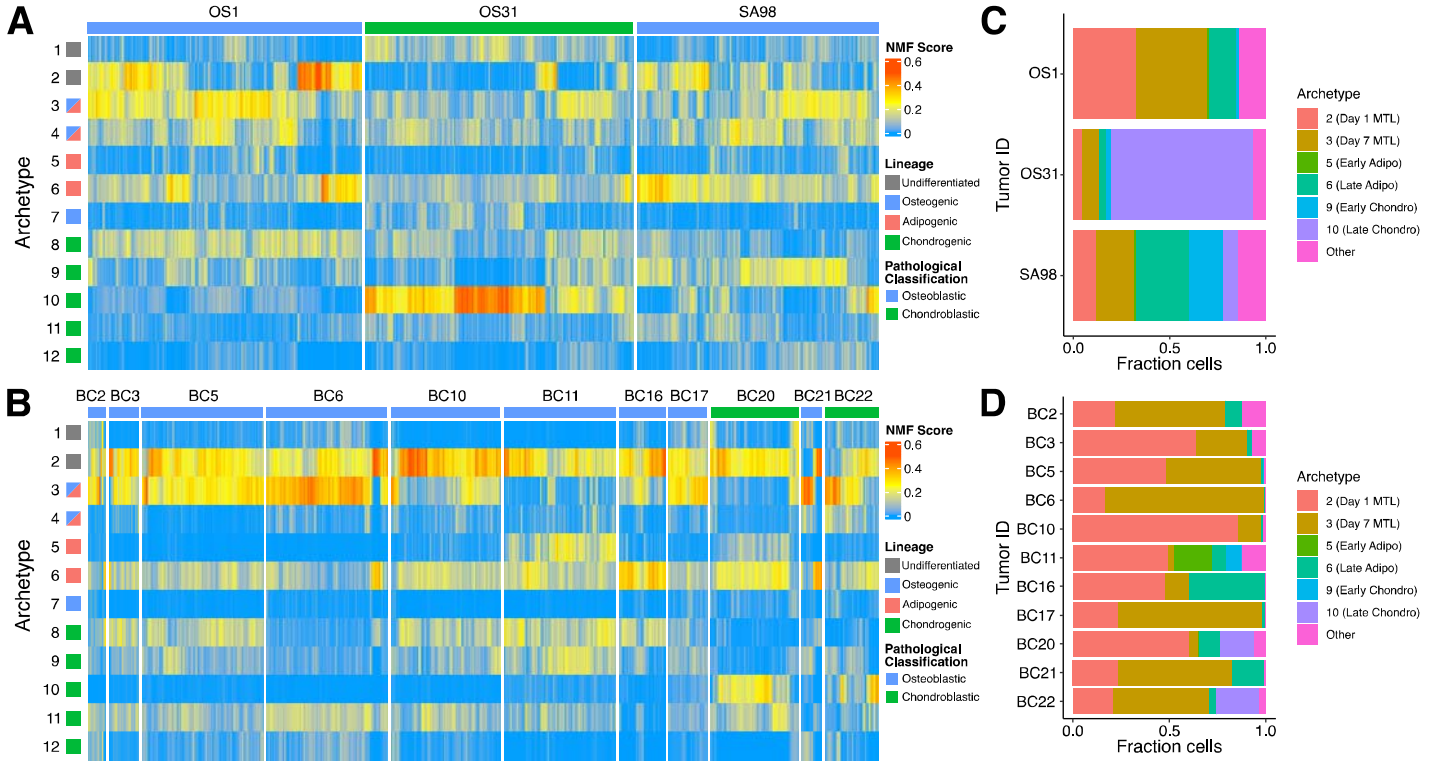

Figure 4: Archetype composition of osteosarcoma tumor samples and PDX models. A: Single-cell archetype score heatmap of 3 PDX models of osteosarcoma, with hierarchical clustering to accentuate cell groups (dendrogram not shown). Row annotation on left indicates lineage of each archetype (same as Fig. 3B). Column annotation indicates pathologist label based on predominant cell type. **B:** Single-cell archetype score heatmap of 11 human osteosarcoma tumor samples. Similar annotation as panel A. **C:** Compositions of each PDX based on the maximum archetype score of each cell. **D:** Compositions of each OS tumor.

A Differentiation Archetype is Associated with Improved Survival in Pediatric Osteosarcoma

We next explored how OS tumor heterogeneity affects clinical outcomes using the NCI's Therapeutically Applicable Research to Generate Effective Treatments (TARGET) OS data subset, made publicly available to accelerate cancer research. Specifically, we projected the N=85 bulk RNA-seq samples from TARGET-OS onto the MTL to identify archetypes associated with differential survival outcomes (Fig. 5A). Despite lacking the intratumoral resolution of single-cell analysis, we found substantial variation in archetype compositions across the cohort. Survival analysis in the cohort revealed that gene expression archetype heterogeneity could explain differential osteosarcoma survival (log-rank test: p=0.033; Fig. 5B). A multivariate analysis determined that archetype 3 (differentiated) was most significantly associated with improved survival (Cox PH: p=0.041; Fig. 5C). In other words, a differentiation signature was associated with improved prognosis, and conversely, tumor stemness, which is frequently associated with progression and osteosarcoma transdifferentiation, was linked to a worse prognosis.

A particularly important finding was that the normal gene expression signatures resolved by the MTL are useful for understanding both intratumoral and intertumoral variability by estimating the composition of OS tumors from bulk gene expression data. A similar approach might be taken to investigate liposarcoma, chondrosarcoma, and – with additional refinements of the MTL to include a myogenic lineage – rhabdomyosarcoma, a highly aggressive sarcoma subtype that occurs predominantly in adolescents and young adults.

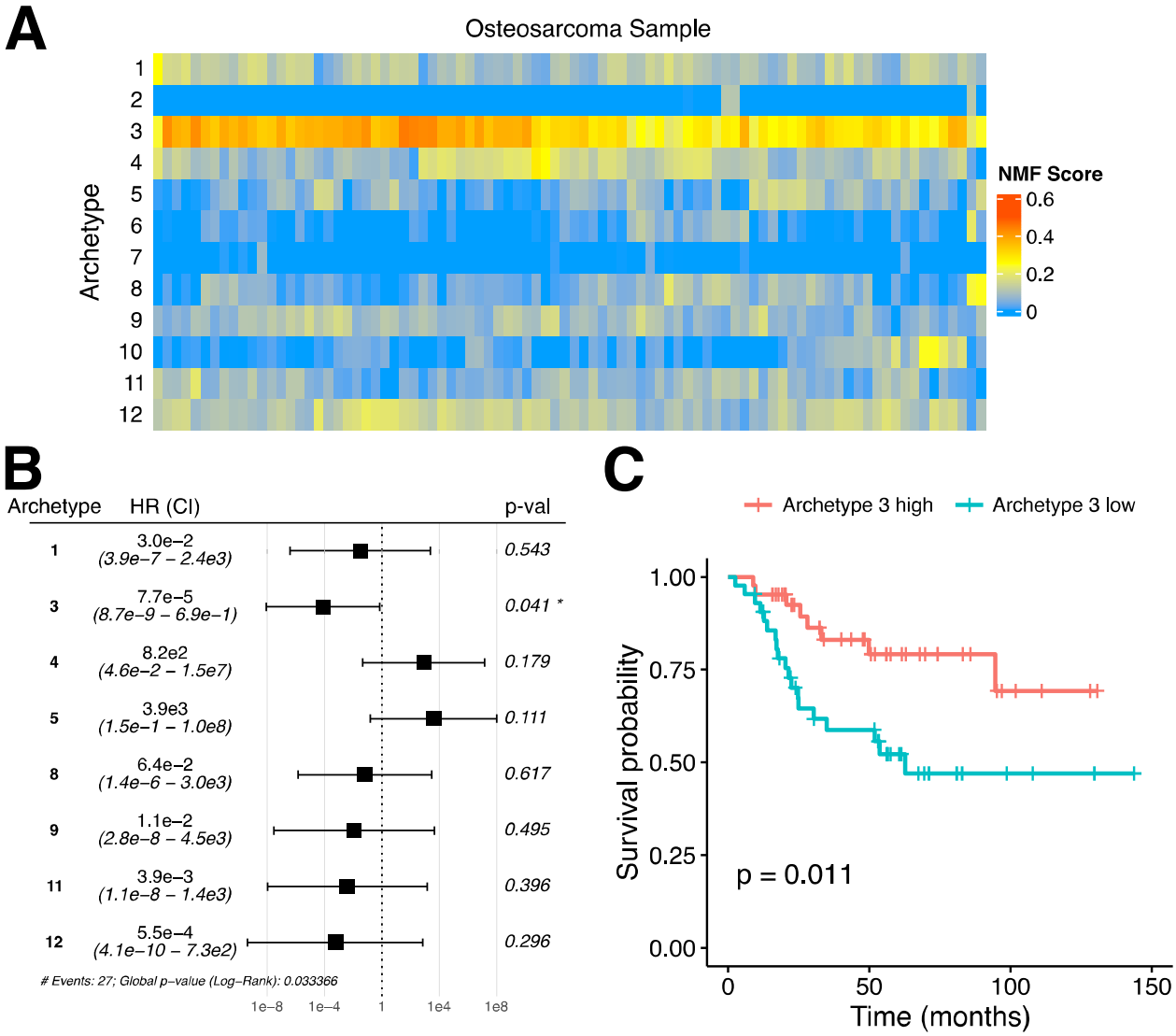

Figure 5: Stemness is associated with poor survival in osteosarcoma. A: Bulk archetype heatmap of TARGET-OS. **B:** Forest plot of the associated Cox regression model with reported hazard ratio (HR) and confidence interval (CI) using the expressed MTL archetypes as regressors. Significance was determined by a multivariable Cox PH test. **C:** Kaplan-Meier plot stratified by the sample-specific Archetype 3 (differentiation) scores grouped as higher or lower than the median. Significance was determined by a univariable Cox PH test.

Discussion

Osteosarcomas are generally classified based on predominant histological features, even though most tumors are comprised of multiple lineages (e.g., osteoblastic, chondroblastic, and fibroblastic). Without the tools to quantify the OS tumor cell composition with a high degree of precision using traditional IHC-based morphometric methods, most high-grade OSs are treated identically using the same cytotoxic chemotherapy regimens developed more than five decades ago. As more powerful proteogenomic methods are applied to tumor samples, one expects to identify prognostic therapeutic biomarkers that ultimately help enable precision-guided medicine that accounts for each patient's unique tumor characteristics. Here, we take an initial step toward that long-term goal by attempting to understand sarcomas in terms of differentiation states from their predicted cell or tissue type of origin. Although OS is thought to derive from an osteogenic cell lineage or mesenchymal progenitor that exists somewhere along the MSC-to-osteochondroblastic continuum, the cancer often presents with osteoblastic, chondroblastic, telangiectatic, and fibroblastic phenotypes. Low-grade OS variants exist that better resemble their mature connective tissue counterparts but are exceedingly rare and were not available, which was a limitation of our analysis.

Given the heterogeneity of cancer cell differentiation, we developed a Mesenchymal Tissue Landscape (MTL) as a high-resolution tissue-engineered reference map that allows for an unparalleled opportunity to profile MSCs by scRNA-seq on their physiological journey toward an osteogenic, adipogenic, and chondrogenic cell fate. Insights generated from the MTL proved valuable in describing mesenchymal differentiation states within OS, revealing distinct lineage-specific gene markers with temporal ordering. Crucially, this MTL provided a framework to understand OS gene expression patterns in terms of the various mesenchymal lineages. Using a similar approach as the one taken here, the adipogenic component of the MTL is under study as a tool to better understand liposarcoma biology and will be reported elsewhere. Further, though the tissue engineering field has demonstrated the ability to push MSCs toward a myoblast phenotype ex vivo by regulating mechanotransduction optimally at 10-12 kPa, this lineage was not included in our initial version of the scRNA-seq MTL, another recognized limitation. As liposarcoma and rhabdomyosarcoma differ significantly from OS in their etiology, genetics, behavior, and treatment, a sarcoma subtype-specific analysis seemed warranted.

The present study interrogated three OS PDX models and the publicly available single-cell gene expression of 11 human OS tumors[16]. In addition to augmenting our cohort, human tumor data ensured that our PDX observations were applicable to clinical tumor samples[49]. To relate mesenchymal differentiation states and tumor heterogeneity in OS, we employed archetype analysis, which represented low-dimensional gene expression patterns in OS relative to the MTL. The archetypes accurately quantified the presence of various differentiation states and lineages in human OS samples. They also showed the potential to quantify the impact of suspected driver alterations and transcription factors on tumor phenotype[15]. Given the link between these archetypes and patient survival, this approach could aid in the development of patient-specific therapeutic strategies for OS. Significantly more OS samples will be required to validate our preliminary findings, potentially enabled through collaboration across the sarcoma research community. Rapid technological changes in the ability to study archival frozen and formalin-fixed paraffin-embedded (FFPE) tissues at the single-cell level, paired with the ability to economically perform sample multiplexing, are on the cusp of enabling these kinds of future large-scale studies.

Notably, though our analysis captured temporal differentiation signatures that may prove useful for lineage tracing during individual tumor evolution, future studies will be required to better understand how tumors dynamically change their fate, plasticity, and lineage in response to biologically targeted therapy as a potential mechanism of resistance. Methods used to infer cell lineage trajectory or plasticity, such as Waddington-OMT and CytoTrace, have shown tremendous promise in understanding temporal gene expression dynamics[13]. As with any nonlinear method, such results may be challenging to interpret and calibrate. However, such methods might be constrained and more easily understood by utilizing archetypes as low-dimensional representations of

lineage-specific gene expression programs. Since the archetypes are trained based on normal mesenchymal differentiation, temporal information derived from archetype analysis can approximate the current differentiation state of sarcoma cells with respect to normal differentiation. Given that many sarcomas may undergo incomplete differentiation or dedifferentiation, this analysis approach allows a way to quantify tumor subpopulations not only of different mesenchymal lineages but also of different temporal stages.

In conclusion, our study has generated a unified and detailed map of the cellular differentiation landscape of three human connective tissue types and demonstrated its potential for characterizing intra- and inter-tumoral OS heterogeneity. The ability to project gene expression data onto this established landscape could facilitate the development of more personalized therapies for OS. Our first-generation MTL lays the groundwork for creating a comprehensive atlas of normal and cancer cell differentiation, which could have far-reaching implications for both basic biology and cancer treatment. The use of archetypes overcomes key technical barriers associated with integrating multiple datasets, enabling the map to naturally evolve as new data is collected. Our study provides a valuable resource for future research into the molecular mechanisms underlying cellular differentiation and disease progression.

Methods

Osteosarcoma Patient-Derived Xenografts

Single-cell RNA-seq data from PDX lines were obtained from a previous study GSE200529[17]. For this study, we selected only “warm” dissociated single-cell RNA-seq data. SA98 (full ID: MDA-SA98-TIS02), OS1, and OS31 are PDX lines maintained by the Pediatric Solid Tumors Comprehensive Data Resource Core[13].

Mesenchymal stem cell culture and differentiation

As *in vitro* differentiation can be inefficient, our differentiation model incorporates biochemical and biophysical cues to better regulate MSC differentiation as described in other works[14,50].

MSC culture and differentiation

Human primary bone marrow-derived MSCs were taken from healthy donors through an IRB-approved partnership with Texas A&M. Cells were differentiated on polyacrylamide-based hydrogels as previously described[14,51]. Briefly, MSCs were plated polyacrylamide gels coated with collagen according to established protocols.

Polyacrylamide-based hydrogel preparation

As previously described[14], we prepare polyacrylamide gels adhered to treated glass slides by glutaraldehyde. The ratio of acrylamide/bis was adjusted for the required stiffnesses. Collagen (10µg/mL) was functionalized to the gels using Sulfo-SANPAH for cell attachment. MSCs were cultured on polyacrylamide-based hydrogels of different stiffnesses (low = 1.5 kPa, high = 25-40 kPa) corresponding to the stiffnesses of the related mesenchymal tissues (adipose ~ 1-2 kPa[50,52,53], osteoid ~ 25-40 kPa[14,54]). Upon adding differentiation induction medium, cells were collected at specified time points until terminal differentiation.

Single-cell suspension, library preparation, and sequencing

Cells undergoing differentiation were collected at specified time-points for sequencing (time points including days 0, 0.5, 1, 2, 3, 5, 7, 10, 14, 21). The cells were washed in PBS before adding trypsin/EDTA (0.25%/380.0 mg/L) followed by 30 min incubation at 37 °C. Afterwards, cells were agitated through pipetting and trypsin activity was briefly terminated by addition of culture media containing serum. Cells were centrifuged and washed with PBS + 0.4% BSA to remove debris and dead cells. Prior to submission for sequencing, cells were re-suspended at 1000 cells/µl.

Single-cell capture, lysis, and library preparation were performed using the Chromium Controller system and established protocols (10x Genomics). Libraries were prepared using Single Cell 3' v3 with a capture of 1700 cells as input for an estimated target cell capture of 1000 cells due to cell capture efficiency of 65%. Libraries were sequenced using the Illumina NextSeq 500 to sequence each cell at an estimated 50,000 reads

per cell. The total number of cells captured was approximately 31,527 cells across all time points with a median unique molecular count of 3.6E4 ± 2.2E4 (SD) per cell and median unique feature of 5.3E3 ± 1.4E3 per cell.

Public datasets:

Chondrogenic Differentiation: Data was accessed at NCBI GEO accession code GSE160625[19]. Raw read counts from each condition were loaded using Seurat R package Read10X command and subsequently merged into a single Seurat data file with added metadata columns specifying the experimental condition and time of each cell. To utilize the authors-reported higher efficiency chondrogenic differentiation with C59 treatment, the chondroprogenitor (Cp) and C59-treated conditions (C59_D7, C59_D12, C59_D28, C59_D42) were selected for analysis[19].

Osteosarcoma Tumors (OS11): Data was accessed at NCBI GEO accession code GSE152048[16]. Raw read counts from each condition were loaded using Seurat R package Read10X command and subsequently merged into a single Seurat data file with added metadata column specifying the tumor identity.

Pediatric Osteosarcoma (TARGET-OS): The results published here are in part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (https://www.cancer.gov/ccg/research/genome-sequencing/target) initiative, phs000218. See section “Pediatric Sarcoma Survival Analysis” below for additional details.

Preprocessing of 10x scRNA-seq data

Cell Ranger (10x Genomics) was used to perform demultiplexing, alignment, filtering, barcode counting, UMI counting, and aggregate the outputs from multiple libraries. Following the standard Cell Ranger pipeline, the filtered gene barcoded matrices were read by Seurat v3 for data preprocessing and initial analysis[55,56].

As quality control for all datasets, cells with fewer than 500 detected genes were removed, along with genes detected in fewer than 3 cells. Additional outliers of mapped mitochondrial reads and expressed genes per cells were adaptively filtered using the scuttle R package isOutlier command with default threshold of 3 mean absolute deviations. Finally, an absolute threshold of cells with low unique counts (n<1000) and high mitochondrial genes (n>25%) was applied as a filter.

After the removal of low-quality cells, we normalized the data using Seurat R package SCTransform normalization[57]. In brief, SCTransform performs normalization and variance stabilization using regularized negative binomial regression to remove technical variation, returning corrected read counts. Normalization was performed in a batch-specific manner, where the Seurat object of a given dataset was first split by batch or tumor ID, using SplitObject to create a list of separate Seurat objects. SCTransform command with parameters return.only.var.genes=F, vst.flavor=”v2” was applied to each list object. The resulting normalized data were merged, including the union of genes reported by SCTransform for each batch.

Harmony Batch Correction

To generate a comprehensive visualization of the MTL, we applied Harmony integration to correct for batch-specific variation among the differentiation data[58]. In brief, Harmony utilizes an iterative correction of the principal components (PCs) to remove batch-specific effects and maximize cell clustering. The data was first normalized using sctransform and a principal-component analysis (PCA) was computed using the top variable genes. The PCA embedding was then batch-corrected by the Harmony algorithm to produce an integrated dimensional reduction with batch correction across tumor_id[20]. Harmony integration was implemented using harmony R package RunHarmony command with parameters: group.by.vars = 'Batch.ID', theta = 3, lambda = 0.5, tau = 300.

Visualization and clustering

To visualize the mesenchymal landscape, we produced a 2D map by Uniform Manifold Approximation and Projection (UMAP) on the Harmony reduction, stored as umap_harmony[59]. Then, using the Harmony reduction, we generated a shared nearest neighbor graph and identified clusters of cells using Louvain algorithm[60]. Resulting clusters were manually labeled based on the predominant experimental condition and time points within each cluster, which largely aligned with three mesenchymal differentiation lineages (osteogenic, adipogenic, chondrogenic) and time-course progression. Heatmaps were produced using the R packages ComplexHeatmap and ggplot2.

Marker Gene Analysis

Differentially expressed marker genes among cell clusters were identified using Seurat's FindAllMarkers command with parameters: only.pos=T, logfc.threshold=0.1. This method applied the default Wilcoxon statistical test to perform comparison of all genes in each cluster compared to all other clusters, only considering genes with positive expression and log fold-change (logfc) above the specified threshold and returning a multiple-comparison corrected FDR value. Significant marker genes were subsequently assessed to determine cluster- and lineage-specific markers and confirm the expected biological identify of the cluster cell populations.

Gene Expression Archetype Analysis

We applied a Normalized Non-negative Matrix Factorization (N-NMF) algorithm to define transcriptional programs or "archetypes" which capture the variability of gene expression patterns across the dataset in a low-dimensional subspace[15]. This approach is a semi-supervised machine learning approach, where we first trained N-NMF archetype coefficients on the mesenchymal differentiation dataset and later used these trained archetypes to score the osteosarcoma PDX and tumor datasets. Prior to NMF, we applied quantile normalization to ensure gene expression matched the same distribution across all cells. To minimize the impact of sequencing noise present in lowly expressed genes, we isolated differentially expressed genes across the data by removing genes with variance less than 0.70 (corresponding to the bulk of the genes), selecting a total of 454 highly expressed and variable genes across the data. Then, the optimal NMF rank was determined heuristically by examining the eigenvalue spectrum elbow plot (Supplemental Fig. S1). NMF archetype coefficients were computed using an iterative algorithm to compute a gene-archetype coefficient matrix and cell-archetype score matrix (normalized to sum to 1 for each cell)[61]. Scoring of osteosarcoma cells with trained archetypes was then computed using a similar iterative algorithm but with the gene-archetype coefficient matrix held fixed.

Pediatric Sarcoma Survival Analysis

Pediatric sarcoma gene expression profiles were obtained from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) study OS dataset (TARGET-OS)[62]. Gene expression data along with clinical phenotype and survival data were accessed from UCSC Xena browser, URL: https://xenabrowser.net/datapages/?cohort=GDC%20TARGET-OS[63]. RNA-seq STAR counts were provided as $\log_2$(CPM+1). For archetype analysis, after mapping gene ensembl_ids to HGNC gene symbol, expression data was quantile normalized with respect to the MTL as the target distribution. Archetype analysis was then used to estimate the composition of archetype gene signatures (trained on the MTL as described above) in each OS. With the resulting archetype scores, survival analysis was performed. First, low-expressed archetypes which had means across all samples less than 0.05 were filtered out (4 of 12 archetypes were removed). Finally, Cox multiple regression was used to determine the overall hazard ratio and statistical significance of each archetype, as well as the global log-rank statistic.

Code and Data Availability

All relevant code used for data processing and analysis are available in a public GitHub repository at the following link: https://github.com/Ludwig-Laboratory/sarcoma_differentiation_landscape. The datasets generated analyzed in this study will be available at the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) upon publication.

Ethics Statement

The MSCs used in this study were taken from healthy donors through an IRB-approved partnership with Texas A&M. All experiments were conducted per protocols and conditions approved by the University of Texas MD Anderson Cancer Center (MDACC; Houston, TX) Institutional Animal Care and Use Committee (eACUF Protocols #00000712-RN02).

Acknowledgements

We acknowledge the gracious philanthropic funding from the Joe and Mary Moeller Foundation. This work is also supported in part by an RO1 R01CA180279 grant from the National Cancer Institute (NCI). CW also acknowledges support from the Marie-Josée Kravis Fellowship in Quantitative Biology. Funding provided by Pediatric Solid Tumors Comprehensive Data Resource Core (Grant # RP180819) and support from the

Institutional Tissue Bank generated the osteosarcoma tissue data. The Flow Cytometry and Cellular Imaging Core Facility performed sorting for the osteosarcoma cells. We acknowledge the Advanced Technology Genomics Core (ATGC, Core grant: CA016672) for their support in creating single-cell libraries and sequencing the cells (NIH 1S10OD024977-01). AT acknowledges support from AFOSR grants (FA9550-20-1-0029, FA9550-23-1-0096), Army Research Office grant (W911NF-22-1-0292), and NIH grant (R01-AG048769).

References

1. Jo VY, Doyle LA. Refinements in Sarcoma Classification in the Current 2013 World Health Organization Classification of Tumours of Soft Tissue and Bone. Surg Oncol Clin N Am 2016;25:621-43.

2. Thway K. Pathology of soft tissue sarcomas. Clin Oncol (R Coll Radiol) 2009;21:695-705.

3. Kansara M, Teng MW, Smyth MJ, Thomas DM. Translational biology of osteosarcoma. Nat Rev Cancer 2014;14:722-35.

4. Evola FR, Costarella L, Pavone V, et al. Biomarkers of Osteosarcoma, Chondrosarcoma, and Ewing Sarcoma. Front Pharmacol 2017;8:150.

5. Cortes-Ciriano I, Lee JJ, Xi R, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 2020;52:331-41.

6. Beird HC, Bielack SS, Flanagan AM, et al. Osteosarcoma. Nat Rev Dis Primers 2022;8:77.

7. Lamhamedi-Cherradi SE, Mohiuddin S, Mishra DK, et al. Transcriptional activators YAP/TAZ and AXL orchestrate dedifferentiation, cell fate, and metastasis in human osteosarcoma. Cancer Gene Ther 2021;28:1325-38.

8. Chaffer CL, Brueckmann I, Scheel C, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A 2011;108:7950-5.

9. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov 2022;12:31-46.

10. Grun D, Muraro MJ, Boisset JC, et al. De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. Cell Stem Cell 2016;19:266-77.

11. Teschendorff AE, Enver T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. Nat Commun 2017;8:15599.

12. Guo M, Bao EL, Wagner M, Whitsett JA, Xu Y. SLICE: determining cell differentiation and lineage based on single cell entropy. Nucleic Acids Res 2017;45:e54.

13. Gulati GS, Sikandar SS, Wesche DJ, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. Science 2020;367:405-11.

14. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. Cell 2006;126:677-89.

15. Weistuch C, Murgas KA, Zhu J, et al. Cancer heterogeneity is defined by normal cellular trade-offs. bioRxiv 2023:2023.04. 12.536595.

16. Zhou Y, Yang D, Yang Q, et al. Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. Nature communications 2020;11:6322.

17. Truong DD, Lamhamedi-Cherradi S-E, Porter RW, et al. Dissociation protocols used for sarcoma tissues bias the transcriptome observed in single-cell and single-nucleus RNA sequencing. bioRxiv 2022:2022.01. 21.476982.

18. Wu CC, Beird HC, Andrew Livingston J, et al. Immuno-genomic landscape of osteosarcoma. Nat Commun 2020;11:1008.

19. Wu C-L, Dicks A, Steward N, et al. Single cell transcriptomic analysis of human pluripotent stem cell chondrogenesis. Nature communications 2021;12:362.

20. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289-96.

21. Elsafadi M, Manikandan M, Dawud R, et al. Transgelin is a TGF β-inducible gene that regulates osteoblastic and adipogenic differentiation of human skeletal stem cells through actin cytoskeleston organization. Cell death & disease 2016;7:e2321-e.

22. Kadota Y, Toriuchi Y, Aki Y, et al. Metallothioneins regulate the adipogenic differentiation of 3T3-L1 cells via the insulin signaling pathway. PLoS One 2017;12:e0176070.

23. Ambele MA, Dessels C, Durandt C, Pepper MS. Genome-wide analysis of gene expression during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene expression during adipocyte differentiation. Stem Cell Res 2016;16:725-34.

24. Nishizuka M, Koyanagi A, Osada S, Imagawa M. Wnt4 and Wnt5a promote adipocyte differentiation. FEBS Lett 2008;582:3201-5.

25. Cheon CW, Kim DH, Kim DH, Cho YH, Kim JH. Effects of ciglitazone and troglitazone on the proliferation of human stomach cancer cells. World J Gastroenterol 2009;15:310-20.

26. Mariman EC, Wang P. Adipocyte extracellular matrix composition, dynamics and role in obesity. Cell Mol Life Sci 2010;67:1277-92.

27. Kovar H, Bierbaumer L, Radic-Sarikas B. The YAP/TAZ Pathway in Osteogenesis and Bone Sarcoma Pathogenesis. Cells 2020;9.

28. Dupont S, Morsut L, Aragona M, et al. Role of YAP/TAZ in mechanotransduction. Nature 2011;474:179-83.

29. Tang Y, Feinberg T, Keller ET, Li XY, Weiss SJ. Snail/Slug binding interactions with YAP/TAZ control skeletal stem cell self-renewal and differentiation. Nat Cell Biol 2016;18:917-29.

30. Zheng X, Zeng W, Gai X, et al. Role of the Hedgehog pathway in hepatocellular carcinoma (review). Oncol Rep 2013;30:2020-6.

31. Akladios B, Reinoso VM, Cain JE, et al. Positive regulatory interactions between YAP and Hedgehog signalling in skin homeostasis and BCC development in mouse skin in vivo. PloS one 2017;12.

32. Zhu B, Xue F, Li G, Zhang C. CRYAB promotes osteogenic differentiation of human bone marrow stem cells via stabilizing beta-catenin and promoting the Wnt signalling. Cell Prolif 2020;53:e12709.

33. Zhou BO, Yue R, Murphy MM, Peyer JG, Morrison SJ. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. Cell Stem Cell 2014;15:154-68.

34. Sanchez C, Mazzucchelli G, Lambert C, Comblain F, DePauw E, Henrotin Y. Comparison of secretome from osteoblasts derived from sclerotic versus non-sclerotic subchondral bone in OA: A pilot study. PLoS One 2018;13:e0194591.

35. Ebert R, Benisch P, Krug M, et al. Acute phase serum amyloid A induces proinflammatory cytokines and mineralization via toll-like receptor 4 in mesenchymal stem cells. Stem Cell Res 2015;15:231-9.

36. Marie PJ. Osteoblast biology and mechanosensing. Mechanosensing Biology: Springer; 2011:105-26.

37. Jundt G, Berghauser KH, Termine JD, Schulz A. Osteonectin--a differentiation marker of bone cells. Cell Tissue Res 1987;248:409-15.

38. Abdelmagid SM, Barbe MF, Rico MC, et al. Osteoactivin, an anabolic factor that regulates osteoblast differentiation and function. Experimental cell research 2008;314:2334-51.

39. Merle B, Bouet G, Rousseau JC, Bertholon C, Garnero P. Periostin and transforming growth factor β-induced protein (TGF β I p) are both expressed by osteoblasts and osteoclasts. Cell biology international 2014;38:398-404.

40. Hauschka PV, Lian JB, Cole D, Gundberg CM. Osteocalcin and matrix Gla protein: vitamin K-dependent proteins in bone. Physiological reviews 1989;69:990-1047.

41. Smargiassi A, Bertacchini J, Checchi M, et al. WISP-2 expression induced by Teriparatide treatment affects in vitro osteoblast differentiation and improves in vivo osteogenesis. Mol Cell Endocrinol 2020:110817.

42. Gustafson B, Hammarstedt A, Hedjazifar S, Smith U. Restricted adipogenesis in hypertrophic obesity: the role of WISP2, WNT, and BMP4. Diabetes 2013;62:2997-3004.

43. Hardingham TE, Oldershaw RA, Tew SR. Cartilage, SOX9 and Notch signals in chondrogenesis. J Anat 2006;209:469-80.

44. Bougault C, Priam S, Houard X, et al. Protective role of frizzled-related protein B on matrix metalloproteinase induction in mouse chondrocytes. Arthritis Res Ther 2014;16:R137.

45. Takacs R, Vago J, Poliska S, et al. The temporal transcriptomic signature of cartilage formation. Nucleic Acids Res 2023;51:3590-617.

46. Zhang Y, Pizzute T, Pei M. A review of crosstalk between MAPK and Wnt signals and its impact on cartilage regeneration. Cell Tissue Res 2014;358:633-49.

47. Park S, Seo K, So A, et al. SOX2 has a crucial role in the lineage determination and proliferation of mesenchymal stem cells through Dickkopf-1 and c-MYC. Cell Death & Differentiation 2012;19:534-45.

48. Wang Y-X, Zhang Y-J. Nonnegative matrix factorization: A comprehensive review. IEEE Transactions on knowledge and data engineering 2012;25:1336-53.

49. Shi J, Li Y, Jia R, Fan X. The fidelity of cancer cells in PDX models: Characteristics, mechanism and clinical significance. Int J Cancer 2020;146:2078-88.

50. Engler AJ, Griffin MA, Sen S, Bonnemann CG, Sweeney HL, Discher DE. Myotubes differentiate optimally on substrates with tissue-like stiffness: pathological implications for soft or stiff microenvironments. J Cell Biol 2004;166:877-87.

51. Reticker-Flynn NE, Malta DFB, Winslow MM, et al. A combinatorial extracellular matrix platform identifies cell-extracellular matrix interactions that correlate with metastasis. Nature communications 2012;3:1122.

52. Comley K, Fleck NA. A micromechanical model for the Young's modulus of adipose tissue. International Journal of Solids and Structures 2010;47:2982-90.

53. Iivarinen JT, Korhonen RK, Julkunen P, Jurvelin JS. Experimental and computational analysis of soft tissue stiffness in forearm using a manual indentation device. Medical engineering & physics 2011;33:1245-53.

54. Zaky S, Lee K, Gao J, et al. Poly (glycerol sebacate) elastomer supports bone regeneration by its mechanical properties being closer to osteoid tissue rather than to mature bone. Acta biomaterialia 2017;54:95-106.

55. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.

56. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36:411-20.

57. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biology 2019;20:1-15.

58. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature methods 2019;16:1289-96.

59. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426 2018.

60. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008;2008:P10008.

61. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401:788-91.

62. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 2018;555:371-6.

63. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol 2020;38:675-8.