HHS Public Access

Author manuscript

Nat Cancer. Author manuscript; available in PMC 2024 April 30.

Published in final edited form as:

Nat Cancer. 2024 April ; 5(4): 642–658. doi:10.1038/s43018-024-00743-y.

Sarcoma microenvironment cell states and ecosystems are associated with prognosis and predict response to immunotherapy

Ajay Subramanian[1,*], **Neda Nemat-Gorgani**[1,*], **Timothy J. Ellis-Caleo**[2,*], **David G. P. van IJzendoorn**[3], **Timothy J. Sears**[1], **Anish Somani**[1], **Bogdan A. Luca**[4], **Maggie Y. Zhou**[5], **Martina Bradic**[6], **Ileana A. Torres**[5], **Eniola Oladipo**[1], **Christin New**[7], **Deborah E. Kenney**[7], **Raffi S. Avedian**[7], **Robert J. Steffner**[7], **Michael S. Binkley**[1], **David G. Mohler**[7], **William D. Tap**[8], **Sandra P. D'Angelo**[8], **Matt van de Rijn**[3], **Kristen N. Ganjoo**[5], **Nam Q. Bui**[5], **Gregory W. Charville**[3], **Aaron M. Newman**[4], **Everett J. Moding**[1,9]

[1]Department of Radiation Oncology, Stanford University, Stanford, California, USA

[2]Department of Medicine, Stanford University, Stanford, California, USA

[3]Department of Pathology, Stanford University, Stanford, California, USA

[4]Department of Biomedical Data Science, Stanford University, Stanford, California, USA

[5]Division of Oncology, Department of Medicine, Stanford University, Stanford, California, USA

[6]Marie-Josee and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, USA

[7]Department of Orthopedic Surgery, Stanford University, Stanford, California, USA

[8]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA and Weill Cornell Medical Center, New York, New York, USA

CONTACT INFORMATION: Everett J. Moding, M.D., Ph.D., Assistant Professor, Department of Radiation Oncology, Stanford Cancer Institute, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. emoding@stanford.edu.

*These authors contributed equally.

AUTHOR CONTRIBUTIONS STATEMENT

Conceptualization, T.J.S., N.Q.B., G.W.C., A.M.N., and E.J.M.; Methodology, Aj.S., An.S., D.v.I., B.A.L., G.W.C., A.M.N., and E.J.M.; Software, Aj.S., An.S., B.A.L., and E.J.M.; Formal Analysis, Aj.S., T.J.E., T.J.S., A.S., and E.J.M.; Investigation, A.S., N.N., T.J.E., D.v.I, T.J.S., A.S., M.Y.Z., N.Q.B., G.W.C., and E.J.M.; Resources, D.v.I., B.A.L., M.B., I.A.T., E.O., C.N., D.E.K., R.S.A., R.J.S., D.G.M., W.D.T., S.P.D., M.v.R., K.N.G., N.Q.B., and G.W.C.; Data Curation, Aj.S., N.N., T.J.E., and E.J.M.; Writing – Original Draft, E.J.M.; Writing – Review & Editing, Aj.S., N.N., T.J.E., D.v.J., T.J.S., An.S., B.A.L., M.Y.Z., M.B., I.A.T., E.O., C.N., D.E.K., R.S.A., R.J.S., M.S.B., D.G.M., W.D.T., S.P.D., M.v.R., K.N.G., N.Q.B., G.W.C., A.N.M., and E.J.M.; Visualization, Aj.S., T.J.E., An. S., and E.J.M.; Supervision, M.v.R., N.Q.B., A.M.N., and E.J.M.; Funding Acquisition, D.G.M., N.Q.B, and E.J.M.

CODE AVAILABILITY

The original EcoTyper code is available for non-profit academic use at https://github.com/ejmoding/sarcomaecotyper and https://github.com/digitalcytometry/ecotyper. Any additional information is available from the lead contact upon request.

COMPETING INTERESTS STATEMENT

W.D.T. has served as a consultant/advisor for Eli Lilly, Daiichi Sankyo, Deciphera, Foghorn Therapeutics, AmMAx Bio, Novo Holdings, Servier, Medpacto, Ayala Pharmaceuticals, Kowa Research Inst., Epizyme Inc (Nexus Global Group), Bayer, Cogent Biosciences, Amgen, and PER, has received research funding from Novartis, Eli Lilly, Plexxikon, Daiichi Sankyo, Tracon Pharma, Blueprint Medicines, Immune Design, BioAlta, and Deciphera, and has a patent on companion diagnostics for CDK4 inhibitors (4/854,329). S.P.D. has served as a consultant/advisor for EMD Serono, Amgen, Nektar, Immune Design, GlaxoSmithKline, Incyte, Merck, Adaptimmune, Immunocore, Pfizer, Servier, Rain Therapeutics, GI Innovations, and Aadi Bioscience and has received research funding from EMD Serono, Amgen, Merck, Incyte, Nektar, Bristol-Meyers Squibb, and Deciphera. E.J.M. has served as a paid consultant for Guidepoint. The other authors declare no competing interests.

[9]Stanford Cancer Institute, Stanford University, Stanford, California, USA

Abstract

Characterization of the diverse malignant and stromal cell states that make up soft tissue sarcomas (STSs) and their correlation with patient outcomes has proven difficult using fixed clinical specimens. Here, we employed EcoTyper, a machine learning framework, to identify the fundamental cell states and cellular ecosystems that make up sarcomas on a large scale using bulk transcriptomes with clinical annotations. We identified and validated 23 sarcoma-specific, transcriptionally defined cell states, many of which were highly prognostic of patient outcomes across independent datasets. We discovered three conserved cellular communities or ecotypes associated with underlying genomic alterations and distinct clinical outcomes. We show that one ecotype defined by tumor-associated macrophages and epithelial-like malignant cells predicts response to immune checkpoint inhibition but not chemotherapy and validate our findings in an independent cohort. Our results may enable identification of patients with STSs who could benefit from immunotherapy and help develop novel therapeutic strategies.

Introduction

Soft tissue sarcomas (STSs) are rare and diverse neoplasms that arise from connective tissues throughout the body in patients of all ages[1]. Although surgery can cure many patients with localized disease, nearly half of patients with large, high-grade STSs develop metastatic disease[2]. Unfortunately, there are few effective systemic therapies for patients with metastatic STS, and the median survival is less than two years[3–5]. Several tumor characteristics, including tumor size, grade, and histology have been shown to be associated with risk of progression after treatment for localized STS[6,7]. In addition, a gene expression signature related to genomic complexity (CINSARC) has been validated to predict risk of metastasis in STS[8]. However, the role of adjuvant chemotherapy in patients with high-risk STS remains controversial[9], and it will be important to understand additional factors that contribute to progression.

Recent studies have demonstrated that immune checkpoint inhibition (ICI) can achieve durable responses in a subset of patients with metastatic STS, but the majority of sarcomas fail to respond[10–12]. The clinical activity of ICI has been reported to be higher in certain STS histologies such as undifferentiated pleomorphic sarcoma (UPS)[12]. However, other histologies such as leiomyosarcoma (LMS) can also respond to ICI at lower rates[10]. Unfortunately, traditional biomarkers of response to ICI, including tumor mutational burden and PD-L1 expression, do not appear to identify the majority of responders in STS[13].

The tumor microenvironment consists of complex cellular communities that are essential for tumor development and response to therapies like ICI[14]. Unlike most human cancers that arise from epithelial tissues, sarcomas arise from mesenchymal cells, leading to unique interactions with the surrounding stroma. Due to the rarity of STSs, few studies have explored the sarcoma microenvironment at the single-cell level[15]. Furthermore, most clinical samples are formalin fixed and paraffin embedded (FFPE), leading to nucleic acid crosslinking and preventing single cell dissociation. As a result, standard approaches to

study the tumor microenvironment such as flow cytometry and single-cell RNA-sequencing (scRNA-Seq) are not feasible in large, clinically annotated datasets.

Previous studies analyzing a limited number of cell types that may not represent all of the unique cells present within mesenchymal tumors have suggested that the sarcoma microenvironment is prognostic and associated with treatment outcomes[16–18]. For example, a recent study defined five sarcoma immune classes (SICs) associated with patient prognosis using gene expression signatures for nine canonical stromal cell types[18]. One SIC characterized by high immune infiltration and the presence of tertiary lymphoid structures (TLS) was associated with better outcomes in patients with localized sarcomas and patients with metastatic STS treated with anti-PD-1. A follow up study demonstrated that selecting patients with TLS increased the response rate to pembrolizumab in advanced STS[17]. However, the presence of TLS appears to be associated with favorable outcomes regardless of treatment[19].

Here, we implement a machine learning framework called EcoTyper[20] to identify the fundamental cell states and cellular communities that make up STSs and validate our findings with single-cell and spatial transcriptomics analysis. We identify 23 transcriptionally-defined cell states across malignant, immune, and stromal cells in the sarcoma microenvironment and explore their association with patient outcomes. Furthermore, we discover three sarcoma ecotypes defined by co-occurring cell states that are associated with prognosis and predict response to ICI. These findings provide a high-resolution view of the cell states and ecosytems that determine clinical outcomes in human STSs.

Results

Identification and validation of sarcoma cell states

EcoTyper utilizes a machine-learning framework to extract cell type-specific gene expression from bulk transcriptomes, identify transcriptional cell states for each cell type, and define tumor ecosystems (“ecotypes”) made up of co-occurring cell states (Fig. 1). To implement EcoTyper for STSs, we assembled a training cohort of 299 patients with localized STS profiled with bulk RNA-sequencing (RNA-Seq)[16,21] and a validation cohort of 310 patients with localized STS profiled by microarray[8] (Extended Data Fig. 1, Supplementary Tables 1–3).

We initially applied EcoTyper to 153 patients with non-LMS STSs from TCGA to identify sarcoma-specific transcriptional cell states. Using CIBERSORTx[22], we estimated the cell type abundance and cell type-specific gene expression profiles (GEPs) across 9 cell types per sarcoma. Next, we reconstructed the weighted contribution of the fundamental transcriptional states within each cell type using the EcoTyper framework for non-negative matrix factorization (NMF). This framework has been optimized using simulated tumors to choose the number of cell states that maximizes sensitivity and positive predictive value of the cell states discovered while maintaining the stability of clustering results[20]. Using this approach, we identified two to four cell states per cell type, representing 23 distinct cell states (Fig. 2a, Supplementary Table 4). We then performed reference-guided recovery of the

cell states identified in non-LMS STSs from TCGA to estimate the abundance of each cell state in the remaining sarcomas within our training and independent validation cohorts.

To validate the cell states identified from bulk gene expression, we performed scRNA-Seq on four LMS and three UPS from a total of six patients (Supplementary Table 5) and analyzed previously published scRNA-Seq from 12 synovial sarcomas (SS) from 10 patients[15] (Extended Data Fig. 1b). We implemented a hybrid approach leveraging inferred copy number variation and similarity of single-cell transcriptomes to bulk sarcoma gene expression to identify malignant and normal cells within our scRNA-Seq datasets (Extended Data Fig. 2a–d, Supplementary Table 6). Using reference-guided annotation with permutation testing, we were able to significantly recover most of the cell states identified using EcoTyper (Extended Data Fig. 2e, Supplementary Table 7). Most of the cell states that were not significantly recovered were not represented or had low abundance in our scRNA-Seq data. To explore the functions of the identified sarcoma cell states, we defined marker genes for the most abundant cell states present within the scRNA-Seq datasets (Supplementary Table 8). Cell state marker gene expression was reproducible across STS histologies profiled by scRNA-Seq (Fig. 2b). These findings confirm that the EcoTyper framework can reliably identify the fundamental cell states within STSs.

We next annotated each cell state based on the cell state marker genes identified by RNA-Seq and scRNA-Seq when available. To aid in annotation of the malignant cell states, we calculated the overlap of marker genes identified from bulk RNA-Seq with the Molecular Signatures Database (MSigDB) hallmark gene sets (Supplementary Table 9). All cell state annotations and key marker genes are displayed in Supplementary Table 10. The cell states identified by EcoTyper reflected both previously described and undescribed biological functions potentially unique to STSs. For example, we observed three populations of tumor-associated macrophages. State 1 (S01) and S02 monocytes/macrophages expressed markers previously associated with M2-like macrophage polarization with distinct expression signatures between states (*CLEC5A* and *SPP1* in S01 monocytes/macrophages and *ROMO1* and *STAB1* in S02 monocytes/macrophages)[23–26]. In contrast, S03 monocytes/macrophages expressed both classical M2 polarization markers (*MRC1* and *CD163*) and markers associated with M1 polarization and pro-inflammatory responses (*IRF5* and *MYO1F*)[27–29].

Because sarcomas arise from mesenchymal cells, we were surprised to identify four epithelial-like cell states in addition to two malignant fibroblast-like cell states. Epithelial-like sarcoma cells are known to exist within certain sarcoma histologies such as epithelioid sarcoma[30] and SS[31], and sarcomas have been reported to undergo mesenchymal to epithelial transition[32]. Furthermore, the presence of epithelial-like malignant cells was recently confirmed by scRNA-Seq of SS where both mesenchymal and epithelial gene expression modules were observed within translocation-positive sarcoma cells[15]. We explored the presence of fibroblast-like cells and epithelial-like cells within other sarcoma histologies in more depth. We first used CIBERSORTx to estimate the abundance of fibroblast-like cells and epithelial-like cells within the sarcomas from our training cohort. As expected, fibroblast-like cells were detected across all STSs (Extended Data Fig. 3a). Interestingly, although detection of epithelial-like cells varied across samples, epithelial-like cells were present across all sarcoma histologies (Extended Data Fig. 3b). Despite different

mesenchymal cells of origin, the combined abundance of fibroblast-like cells and epithelial-like cells significantly correlated with tumor purity across histologies (Extended Data Fig. 3c). These findings suggest that the fibroblast-like and epithelial-like cell states identified using EcoTyper represent malignant sarcoma cells along the spectrum of mesenchymal to epithelial transition.

To further explore the presence of epithelial-like malignant cells within STSs, we evaluated expression of epithelial markers within LMS and UPS tumor cells by scRNA-Seq. A subset of malignant cells expressed epithelial markers, including epithelial cadherin (CDH1) and epithelial membrane antigen (MUC1, Extended Data Fig. 4a)[33]. Utilizing the gene expression modules defined in SS[15] and previously described gene expression signatures of epithelial to mesenchymal transition (EMT, Supplementary Table 11)[34–36], we observed a spectrum of epithelial differentiation in both LMS and UPS STSs by scRNA-Seq (Fig. 2c–e, Extended Data Fig. 4b). Furthermore, epithelial-like cell state marker genes showed overlap with gene sets associated with oncogenic pathways (KRAS in S03 epithelial cells and both MYC and MTORC1 in S04 epithelial cells, Supplementary Table 9). To explore the expression of epithelial markers at the protein level in STS, we performed immunohistochemistry (IHC) for CDH1, MUC1, and syndecan-1 (SDC1), a transmembrane proteoglycan expressed by normal and malignant epithelial cells (Supplementary Table 12)[37]. At least one of the epithelial markers was detected by IHC in tumor cells for 96% of patients with STS (Fig. 2f,g). These results suggest that epithelial-like malignant cells are present across STS histologies.

Association of sarcoma cell states with patient outcomes

We next explored the association of the EcoTyper cell states with patient outcomes. On multivariable analysis incorporating histology, 12 of the 23 cell states (52%) were associated with progression-free survival in our training cohort (Fig. 3a). Strikingly, the associations between cell state abundance and patient outcomes were highly conserved in the validation cohort (Extended Data Fig. 5). Eight of the 12 cell states that were significantly associated with patient outcomes in the training cohort were also significant in the validation cohort. Furthermore, the direction of the association was preserved for most cell states, and we observed a strong correlation across cohorts despite the different platforms used for bulk gene expression profiling (Fig. 3b).

Consistent with prior reports[38], a higher abundance of *CLEC5A/SPP1*+ M2-like immunosuppressive macrophages (S01) was associated with inferior outcomes in both our training and validation cohorts (Fig. 3c). Higher B cell abundance has previously been reported to be associated with improved outcomes in patients with sarcomas[18], and a higher abundance of *CD69*+ activated memory B cells (S01) was significantly associated with better outcomes in our training cohort and trended towards better outcomes in the validation cohort. We also identified several additional survival associations. For example, *NOSTRIN/PECAM1/CD36+* normal-like endothelial cells (S02) were associated with improved outcomes in both cohorts across sarcoma histologies (Fig. 3d). In addition, epithelial-like malignant cells with upregulated MYC/MTORC signaling (S04) were associated with poor outcomes, and both epithelial-like (S03) and fibroblast-like (S01) malignant cells with

activation of the KRAS pathway were associated with favorable outcomes. These results demonstrate that transcriptionally defined malignant and normal cell states are consistently associated with patient prognosis in STS.

Reconstructing STS cellular communities

We next explored the patterns of cell state co-occurrence using the EcoTyper framework to reconstruct the fundamental cellular communities within STSs. We discovered three distinct "sarcoma ecotypes" made up of four to 10 co-occurring cell states per community (Fig. 4a,b). Although most STSs could be assigned to a dominant sarcoma ecotype, we observed evidence of multiple sarcoma ecotypes within each sample. As a result, we also estimated the abundance of the three sarcoma ecotypes within each sample in a continuous manner. Sarcoma ecotype 1 (SE1) was characterized by several immune cell states lacking activation markers and *SALL1+* pro-angiogenic endothelial cells (S01). In contrast, SE2 consisted of several activated immune cell states, *NOSTRIN/PECAM1/CD36+* normal-like endothelial cells (S02), and malignant cells with activation of the KRAS pathway. Finally, SE3 was characterized by MYC/MTORC1 activated epithelial-like malignant cells (S04) and *CLEC5A/SPP1+* M2-like immunosuppressive macrophages (S01). Remarkably, sarcoma ecotypes were distinct from previously described STS prognostic gene expression signatures (Fig. 4c)[8,18], suggesting the EcoTyper framework enables additional insights into sarcoma biology.

To further characterize the three sarcoma ecotypes, we explored their clinical, cellular, and molecular characteristics (Fig. 5a, Supplementary Table 13). SE1 was enriched in female patients and patients with SS. SE1 sarcomas were immune cold with a low abundance of most immune cell subsets and elevated Hedgehog signaling. In contrast, SE2 sarcomas were more consistent with classically immune hot tumors with a high abundance of B and CD8 T cells and enriched interferon alpha and gamma signaling[39]. SE2 was enriched in male patients and patients with liposarcomas. Finally, SE3 was enriched in older patients and UPS and malignant peripheral nerve sheath tumors (MPNST) compared with other histologies. SE3 tumors were characterized by intermediate immune infiltration and elevated MYC and MTORC1 signaling.

Prior studies have suggested that the tumor microenvironment can change over time and in response to therapy[40,41], which could complicate the implementation of sarcoma ecotypes as biomarkers in STSs. To explore whether sarcoma ecotypes change over time, we compared ecotype abundance across timepoints in patients with more than one sarcoma sample available for analysis. Among these 24 patients, 71% underwent treatment with systemic therapy between sampling, and 45% of paired samples were from primary tumors and a separate location of metastatic disease (Supplementary Table 14). Remarkably, sarcoma ecotype abundance was highly correlated across different timepoints despite a median of 14 months elapsing between samples (Extended Data Fig. 6a).

To explore the possibility that genomic alterations may underly the tumor microenvironment within STS, we compared the SNVs, indels, and copy number alterations (CNAs) across sarcoma ecotypes from TCGA (Fig. 5b). We observed significantly more CNAs and SNVs/indels in SE3 compared to the other sarcoma ecotypes (Fig. 5c and Extended Data Fig.

6b), suggesting genomic instability may play a role in shaping the microenvironment in these tumors. Several genes were more frequently altered by SNVs/indels or CNAs in SE3 sarcomas compared to other sarcoma ecotypes (Supplementary Table 15). Consistent with the hallmark gene set analysis for S04 epithelial-like cells and bulk SE3 tumors, amplifications of *MYC* and *MTOR* were significantly enriched in SE3 sarcomas. We also observed a significantly increased contribution of the COSMIC mutational signature 2 in SE2 sarcomas (Fig. 5d and Extended Data Fig. 6c), suggesting evidence of APOBEC mutagenesis in these tumors[16,42]. These findings suggest that underlying genomic alterations within malignant cells may drive sarcoma ecosystems, which could explain the consistency of sarcoma ecotypes across different samples from the same patient.

To investigate the co-localization and interaction of cell states with human sarcomas, we profiled four STSs using spatial transcriptomics (Extended Data Fig. 1c). We performed reference-guided recovery to estimate the abundance of each cell state and sarcoma ecotype within each spatially-barcoded spot. In addition to co-occurring within the same tumors, cell states within the same ecotype tended to occur with the same regions of sarcomas (Extended Data Fig. 7a,b). Cell states within the same sarcoma ecotype had a stronger spatial correlation than cell states from different ecotypes (Fig. 5e). Interestingly, sarcoma ecotypes tended to be aggregated within spatially distinct regions of the tumor (Fig. 5f and Extended Data Fig. 7c). Across the 4 sarcomas profiled by spatial transcriptomics, all ecotypes were significantly spatially aggregated (Fig. 5g), suggesting that sarcoma ecotypes represent distinct functional units within human sarcomas.

We next explored intercellular signaling networks within sarcoma ecotypes by identifying putative ligand-receptor pairs across cell states (Supplementary Table 16). We observed complex cell-cell interactions between cell states within the same ecotype (Fig. 5h). TNFA/KRAS activated fibroblast-like cells (S01) overexpressed multiple ligands with corresponding receptors on other cell states within SE2. For example, TNFA/KRAS activated fibroblast-like cells (S01) overexpressed *TNFSF13* and *TNFSF18* whose receptors *TNFRSF17* and *TNFRSF18* were overexpressed on *CD69*+ activated memory B cells (S01) and *GRZMA/PDCD1/TIGIT*+ exhausted cytotoxic CD8 T cells (S01), respectively. Similarly, *TRIM29*+ epithelial-like cells (S01) and *CLEC5A/SPP1*+ M2-like immunosuppressive macrophages (S01) appeared to be at the center of the signaling networks for SE1 and SE3, respectively. These results highlight potential cellular interactions that could shape the STS microenvironment.

Association of ecotypes with outcomes in localized sarcomas

To investigate the association of sarcoma ecotypes with patient prognosis, we compared clinical outcomes in our training cohort and validation cohort based on sarcoma ecotype assignment. In addition, we analyzed sarcoma ecotype abundance as a continuous variable while accounting for sarcoma histology using Cox multivariable analysis. In both training and validation cohorts, we observed that patients assigned to SE2 had superior outcomes than the other sarcoma ecotypes (Fig. 6a and 6c). On multivariable analysis including histology, patients with higher SE2 abundance continued to have significantly better outcomes, and patients with higher SE3 abundance had significantly worse outcomes in

both cohorts (Fig. 6b and 6d). Patients assigned to SE1 appeared to have an intermediate prognosis. In patients from the training cohort with more detailed clinical information available, we observed similar results when including the prognostic factors tumor size, grade, histology, extent of resection, and anatomic site in the multivariable analysis (Supplementary Table 17)[6,7,43]. These results suggest that sarcoma ecotypes are associated with prognosis in localized STSs.

Prediction of ICI response using sarcoma ecotypes

The tumor microenvironment plays a crucial role in the response of solid tumors to systemic therapies such as ICI[44]. As a result, we constructed two cohorts of patients with advanced sarcoma who received systemic therapy with standard of care cytotoxic chemotherapy (n=37) or a combination immune checkpoint blockade with ipilimumab (anti-CTLA4 antibody) and nivolumab (anti-PD-1 antibody) (n=38, Extended Data Fig. 8, Supplementary Tables 18–20). We assessed the association of sarcoma ecotype assignment and sarcoma ecotype abundance with outcomes in both cohorts as described above. Similar to our cohorts of patients with localized sarcoma, patients in SE2 and patients with a higher SE2 abundance had significantly better outcomes after chemotherapy (Fig. 7a,b). In contrast, we observed that patients in SE3, who experienced significantly worse outcomes in patients with localized sarcoma and in our chemotherapy cohort, had the best progression-free survival after ICI (Fig. 7c). In addition, higher SE3 abundance was associated with better progression-free survival after ICI on multivariable analysis including histology (Fig. 7d). These findings suggest that SE3 represents a candidate predictive biomarker for response to ICI in patients with advanced sarcoma.

To further explore the potential for SE3 to identify patients with STSs who benefit from ICI, we analyzed treatment response by sarcoma ecotype. Two of the three responders assigned to sarcoma ecotypes in our ICI cohort were assigned to SE3 (Fig. 7e). SE3 abundance was significantly higher among responders than non-responders (Fig. 7f) and among patients achieving 6 months of non-progression after starting ICI (Extended Data Fig. 9a). Furthermore, SE3 abundance outperformed previously reported predictors of ICI response in sarcomas, TLS and PD-L1 expression, for predicting both response and 6-month non-progression (Fig. 7g and Extended Data Fig. 9b). We observed higher PD-L1 expression and more frequent TLS in STSs assigned to SE2 compared with SE1 (Extended Data Fig. 9c,d). However, elevated PD-L1 expression and the presence of TLS were both detected among patients assigned to SE3. Notably, SE3 abundance did not predict response to chemotherapy (Extended Data Fig. 9e–g). These findings illustrate the promise for SE3 abundance to select patients with STS who would benefit from ICI but not chemotherapy.

Validation of ICI response prediction

To validate the ability of SE3 abundance to predict response to ICI in STSs, we analyzed an independent cohort of 29 patients treated on 3 clinical trials evaluating combinations of anti-PD-1 antibodies with experimental immunotherapies (Extended Data Fig. 8, Supplementary Table 21)[45–47]. Of the 4 responders in the ICI validation cohort assigned to sarcoma ecotypes, 3 patients were assigned to SE3 (Fig. 8a). Furthermore, SE3 abundance was significantly higher in responders than non-responders (Fig. 8b) and strongly predicted ICI

response (AUC 0.89, Extended Data Fig. 10a). To further determine if SE3 abundance can be used to enrich for patients likely to respond to ICI, we defined the optimal cutpoint for SE3 abundance in the training cohort (0.305) and applied the same cutoff to the validation cohort. 28.5% of patients in the training cohort and 42.9% of patients in validation cohort above the threshold responded to ICI while no patients below the threshold responded in either cohort (Fig. 8c). We also explored the ability of SE3 abundance to predict response to ICI in patients with metastatic urothelial bladder carcinoma and melanoma, but we did not observe a significant association in these cancers (Extended Data Fig. 10b, Supplementary Table 22). These results validate the ability of SE3 abundance to predict response to ICI in STSs across independent cohorts.

Previous studies have demonstrated that tumor-associated macrophages can activate the adaptive immune system and directly kill tumor cells in response to ICI in other solid tumors[48,49]. However, ICI is traditionally thought to enable anti-tumor responses through activation of CD8 T cells[50]. We examined our spatial transcriptomics data to explore the possibility that CD8 T cell populations not assigned to SE3 could play a role in ICI response within sarcomas. Among the CD8 T cell states identified by EcoTyper, *GRZMA/PDCD1/TIGIT*+ exhausted cytotoxic CD8 T cells (S01) were in the closest proximity to SE3 spots (Fig. 8d). Interestingly, although S01 CD8 T cells were assigned to SE2 by co-occurrence in bulk tumors, they had similar proximity to SE3 spots by spatial transcriptomics (Fig. 8e). To further evaluate the role of sarcoma cell states in ICI response, we compared cell state and ecotype abundance pre-treatment and on-treatment in 19 patients from the ICI validation cohort with paired samples available (Extended Data Fig. 10c). We observed stable to slightly increased S01 CD8 T cell abundance (Fig. 8f) with a significant decrease in the S01 M2-like immunosuppressive macrophages associated with SE3 after starting ICI (Fig. 8g). These results suggest that ICI reduces the abundance of S01 M2-like immunosuppressive macrophages, which may enable S01 CD8 T cells to generate anti-tumor immune responses.

Discussion

In this study, we implemented a machine learning framework to decode the fundamental cell states and cellular communities that exist within STSs. In contrast to prior studies that have investigated broadly defined cell types and states identified in other cancers[16,18], our computational approach enabled the identification of cell states specific to STSs. We observed a strong association of both individual cell states and sarcoma ecotypes with patient outcomes across independent datasets, confirming the importance of the sarcoma microenvironment in tumor progression and response to treatment. Finally, we demonstrated that SE3 abundance is a candidate predictive biomarker for identifying patients with advanced STSs who benefit from immune checkpoint but not chemotherapy.

EcoTyper overcomes the limitations of most clinical specimens and avoids the need for physical dissociation, which is known to affect gene expression and the representation of certain cell populations[51]. We were able to validate most cell states identified by EcoTyper using scRNA-Seq with the remaining cell states poorly represented in our scRNA-Seq datasets. Furthermore, we confirmed the co-localization of the cell states within sarcoma ecotypes using spatial transcriptomics, demonstrating the potential for cellular interactions

that may affect sarcoma biology. Sarcoma ecotypes were distinct from SICs[18] and CINSARC groups[8], which have both been previously associated with patient prognosis. In contrast to SICs that were defined based on tumor immune infiltration alone and CINSARC groups based on genes associated with genomic complexity within malignant cells, sarcoma ecotypes were defined based on the co-localization of transcriptionally defined states across malignant and stromal cells.

Although sarcoma ecotypes aggregated within distinct regions of the tumor, ecotype abundance was highly correlated across sarcoma samples. Therefore, profiling a limited number of tissue cores or slides appears to be sufficient to characterize the general ecotype composition of a patient's sarcoma. Interestingly, sarcoma ecotypes were associated with underlying genomic alterations with SE3 sarcomas having a higher abundance of SNVs/indels and CNAs and SE2 sarcomas having a higher contribution of the APOBEC mutagenesis-related COSMIC signature 2. Previous studies have also associated APOBEC-mutagenesis with increased immune infiltration into tumors and improved survival in other cancers[52,53]. The lower mutational burden and lower APOBEC-mutagenesis in SE1 may explain the low immune infiltration into these tumors. Due to the strong association of sarcoma ecotypes with patient prognosis and their ability to predict response to ICI, our findings could enable personalized treatment in patients with STSs. However, it will be important to validate these findings in additional retrospective and prospective cohorts. Although RNA sequencing is increasingly used for STSs to detect gene fusions[54], we are also exploring surrogate markers of sarcoma ecotypes.

Because the majority of patients do not respond to ICI[10,12], biomarkers will be crucial to unlock the potential of ICI in sarcomas. PD-L1 expression has been associated with response to ICI in sarcomas[13], and the phase II PEMBROSARC trial demonstrated improved 6-month non-progression rates and objective response rates in STSs selected based on the presence of TLS[17]. In our study, SE3 abundance outperformed both biomarkers for predicting ICI response and 6-month non-progression. Furthermore, we were able to validate the ability of SE3 abundance to predict ICI response in an independent cohort treated with different immunotherapies in combination with ICI. In contrast to SICs and the presence of TLS, which are prognostic of favorable outcomes in both patients with early-stage disease and after treatment with ICI[18], we demonstrated that SE3 predicted favorable outcomes after treatment with ICI but not chemotherapy or in localized STSs. As a result, measuring SE3 abundance may represent a powerful approach to identify patients with advanced sarcoma who should be treated with ICI. Although a higher SE2 abundance was associated with significantly better PFS after chemotherapy, no sarcoma ecotypes predicted chemotherapy objective response. Therefore, additional factors outside of the STS microenvironment may also contribute to chemotherapy response.

The proportion of patients assigned to SE3 was highest in UPS, which may explain the high clinical activity of ICI in this histology[12]. Notably, we also observed a high SE3 abundance in ICI responders with MPNST and LMS. Although sarcomas have a lower tumor mutational burden than most epithelial tumors[16], SE3 sarcomas had more SNVs/indels and CNAs than SE2 and SE1 sarcomas. Interestingly, SE3 sarcomas did not display the immune hot phenotype observed in many epithelial tumors that respond favorably to

ICI[39]. Instead, SE3 sarcomas consisted of an intermediate level of immune infiltration characterized by M2-like immunosuppressive macrophages (S01), MYC/MTORC1 activated epithelial-like cells (S04), mature dendritic cells (S01), and pro-inflammatory neutrophils (S02). PD-1/PD-L1 blockade has been shown to increase macrophage phagocytosis[49], and tumor-associated macrophages may also inhibit anti-tumor immune responses by CD8 T cells[55]. Although a CD8 T cell population was not assigned to SE3 in bulk tumors, we observed strong evidence of co-localization between SE3 and an exhausted cytotoxic CD8 T cell state (S01) by spatial transcriptomics. Notably, SE3 abundance did not predict response to ICI in bladder cancer or melanoma, suggesting a distinct mechanism of ICI response in sarcomas. Together, our findings support a model where SE3 sarcomas are primed for anti-tumor immune responses with a higher mutational load and nearby antigen presenting dendritic cells and tumor-responsive CD8 T cells. However, the presence of M2-like immunosuppressive macrophages prevents tumor rejection in the absence of ICI. Future mechanistic experiments in preclinical models will be helpful to explore these cellular interactions and design therapies that increase the number of patients with STSs who respond to ICI.

Epithelioid sarcomas and biphasic SS are known to express epithelial markers[30,31]. Although the presence of epithelial-like sarcoma cells has not been explored in depth in other STSs, prior studies have observed expression of epithelial markers by IHC[56,57]. Here, we demonstrated that malignant sarcoma cells within LMS and UPS express epithelial markers by scRNA-Seq. Furthermore, by using previously described gene expression signatures of EMT, we demonstrated a spectrum of epithelial differentiation across STS histologies. Finally, we confirmed expression of epithelial markers by IHC across a broad panel of STS histologies. Mesenchymal to epithelial transition (MET) plays a critical role in carcinoma disease progression and metastasis[58]. In addition, EMT may promote resistance to ICI[59], suggesting that the presence of epithelial-like sarcoma cells may contribute to CD8 T cell recognition. As a result, further investigation into the role of MET in sarcoma biology and response to therapies like ICI is warranted.

Our sample size was limited by the rarity of STSs and relative lack of publicly available, clinically annotated gene expression data when compared with other cancer histologies. As a result, we included diverse sarcoma histologies in our cohorts. We focused on non-LMS STSs profiled by TCGA for our initial cell state and ecotype discovery to minimize batch effects and improve the accuracy of our deconvolution results. In the future, additional approaches to account for technical variation between samples and across platforms may enable identification of cell states and ecotypes in larger cohorts of STSs. Because sarcoma histology is known to affect prognosis and response to therapy[12], we accounted for histology in our multivariable models. Here we focused on the most common STS histologies[43], and future studies will be necessary to determine if sarcoma ecotypes are associated with patient outcomes in other sarcoma subtypes. Although we confirmed our findings across multiple retrospective cohorts, future prospective and interventional studies will be necessary before sarcoma ecotypes can be utilized as a biomarker to guide clinical management. We aimed to identify cell states and cellular communities conserved across STS histologies, but there are likely additional malignant and non-malignant cell populations that may be less common or specific to one or a few sarcoma histologies. For example, there may be malignant cell

states specific to the cell of origin for different STSs. Additional studies utilizing single cell approaches will be helpful to identify these cell populations, which may contribute to the diversity and biology of STS.

In summary, we have defined the landscape of fundamental transcriptional cell states and cellular ecosystems within STSs. Our findings have confirmed a crucial role for the tumor microenvironment in sarcoma progression and response to therapy. Furthermore, we have identified a candidate predictive biomarker of response to ICI in STSs that could enable personalization of systemic therapy in patients with advanced disease if validated in future prospective studies.

METHODS

Human subjects

All research was conducted in accordance with the Declaration of Helsinki. Study protocols were approved by the Institutional Review Boards at Stanford University and Memorial Sloan Kettering. Written informed consent was obtained for all samples collected and all data included in this study.

Training cohort

Pre-processed bulk RNA-seq profiles of STSs from The Cancer Genome Atlas (TCGA) were downloaded[60] and scaled to transcripts per million (TPM). BAM files from the Steele et al. Cancer Cell 2019 UPS dataset[21] were downloaded from the European Genome-Phenome Archive (dataset ID EGAD00001004439) and converted to FASTQ files. FASTQ files were quasi-aligned to Gencode version 27 transcript models, and expression levels were summarized using Salmon[61]. After restricting to protein coding genes, expression levels per gene were summarized as TPM. Although the EcoTyper framework can be utilized to robustly recover cell states and ecotypes across gene expression platforms and datasets[20], de novo discovery of cell states and ecotypes performs most robustly using samples with minimal batch effects. Consistent with the initial publication from TCGA, we observed that LMS clustered separately from other STSs by UMAP and tSNE analysis[16]. When we included LMS in our initial cell state and ecotype discovery, we observed expression of myogenic markers in many immune cell states and were unable to validate the identified cell states, suggesting inaccurate deconvolution. To enable identification of cell states and ecotypes from a more homogenous cohort of patients profiled with the same sequencing platform, we performed cell state and ecotype discovery on UPS, liposarcomas, SS, and MPNST profiled by TCGA. We then separately recovered the identified cell states and ecotypes from the LMS profiled by TCGA and the UPS from the Steele et al. Cancer Cell 2019 dataset to complete our training cohort for correlation with patient outcomes. Author-provided tumor purity and clinical information were used for analysis. SIC labels for the TCGA sarcomas were provided by the authors[18].

Validation cohort

Microarray data from the Chibon et al. Nature Medicine dataset were downloaded from the Gene Expression Omnibus (GEO, accession GSE21050). Microarray data were normalized

using the MAS5 algorithm from the 'affy' R package[62], mapped to NCBI Entrez gene identifiers using Brainarray version 21.0, converted to HUGO gene symbols, and restricted to protein coding genes for input to cell state and ecotype recovery. Progression-free survival was not available for this cohort, so distant metastasis-free survival was analyzed as a similar endpoint because metastases represent the majority of progression events in localized sarcomas[2].

Stanford chemotherapy and ICI cohorts

Patients with advanced STS treated with standard of care cytotoxic chemotherapy and/or combination immune checkpoint blockade with ipilimumab and nivolumab were identified retrospectively from an institutional database. Only patients with histologies profiled by TCGA (UPS, liposarcomas, SS, MPNST, and LMS) were included in the analysis. Patients with a diagnosis of spindle cell sarcoma, pleomorphic sarcoma with giant cells, malignant fibrous histiocytoma, fibrosarcoma, and myxofibrosarcoma were also included and analyzed as UPS due to the broad definition of this diagnosis and to be consistent with the external cohorts[63]. A subset of patients treated with ICI received cryotherapy to one lesion in between cycle 1 and cycle 2 of ICI as part of a phase II clinical trial (NCT04118166)[64]. For these patients, response was assessed on lesions not treated with cryotherapy.

Bulk RNA sequencing

FFPE tumor blocks from patients in the Stanford chemotherapy and ICI cohorts were sectioned and stained with H&E for pathology review and to identify regions with the highest sarcoma content for sampling. Bulk RNA was isolated from two to three 1 mm cores per sample using RNAstorm FFPE RNA Extraction Kits (Cell Data Sciences). Sequencing libraries were prepared using SMARTer Stranded Total RNA-Seq v2-Pico Input Mammalian Kits (Takara Bio USA, Inc.). Libraries were sequenced on a HiSeq 4000 or a NovaSeq 6000 (Illumina) with 150 base pair paired-end reads. FASTQ files were quasi-aligned using Salmon and expression levels were summarized as TPM as described above. Samples with fewer than 1 million mapped reads were excluded from further analysis. For a subset of patients, two separate blocks were sequenced from the same timepoint. For these timepoints, mean cell state abundances across both regions were used for analysis and ecotype recovery as described below. For patients with multiple timepoints available, the sample collected closest to the start of therapy was used for analysis.

ICI validation cohort

The ICI validation cohort consisted of patients with metastatic or locally advanced STSs treated as part of 3 prospective clinical trials combining pembrolizumab and talimogene laherparepvec (NCT03069378)[47], nivolumab and bempegaldesleukin (NCT03282344)[45], and pembrolizumab and epacadostat (NCT03414229)[46]. We restricted our analysis to patients with histologies profiled by TCGA as described above. RNA was extracted from frozen tissues, library prepped using the TruSeq Stranded mRNA LT Kit (Illumina), and sequenced on a HiSeq 4000 (Illumina) as described previously[45]. FASTQ files were quasi-aligned using Salmon as described above, and counts were batch corrected using ComBat-seq from the 'sva' R package to account for variation across sequencing lanes[65]. After restricting to protein coding genes, expression levels per gene were summarized as TPM.

Bladder and melanoma ICI cohorts

Clinical annotations and raw bulk transcriptome counts from patients with metastatic urothelial bladder cancer treated with atezolizumab on the IMvigor210 phase II clinical trial were obtained via the 'Imvigor210CoreBiologies' R package[66]. Clinical annotations and normalized count data for patients with metastatic melanoma treated with anti-PD-1 antibodies[41,67] or anti-CTLA4 antibodies[68,69] were downloaded from the supplementary material of the published manuscripts. Expression levels per gene were summarized as TPM for downstream analysis.

Discovery of sarcoma-specific cell states

We applied EcoTyper to identify sarcoma-specific transcriptionally-defined cell states from the bulk RNA-Seq of non-LMS TCGA samples in the training cohort[20]. We first estimated cell type proportions and imputed cell type-specific GEPs using CIBERSORTx[22]. We utilized two previously validated signature matrices to analyze the major cell populations within sarcomas. LM22 is a signature matrix consisting of 22 human immune cell subsets derived from Affymetrix microarray data[70], and TR4 is a signature matrix consisting of epithelial, fibroblast, endothelial, and immune populations derived from bulk RNA-Seq of flow sorted cell populations[22]. CIBERSORTx was run to impute the fractional abundance of each cell type using B-mode batch correction with LM22 to overcome cross-platform variation, and no batch correction step for TR4. The LM22 subsets were pooled into 9 major lineages: B cells, plasma cells, CD8 T cells, CD4 T cells, natural killer cells, monocytes/macrophages, dendritic cells, mast cells, and polymorphonuclear cells (PMNs)[20]. To unify the signature matrices, the immune cell fractions from LM22 were rescaled to sum to 1 within each sample and multiplied by the total immune cell fraction from TR4, yielding fractions for a total of 12 cell types. We then provided the cell fractions and TPM-normalized expression matrix as input to the high-resolution gene expression purification module of CIBERSORTx. Three cell types with fewer than 50 imputed genes were excluded from additional analysis (plasma cells, CD4 T cells, and natural killer cells).

Non-negative matrix factorization (NMF) combined with specialized heuristics was used to identify and quantitate transcriptionally-defined cell states within the cell type-specific GEPs purified by CIBERSORTx[20]. For each cell type-specific imputed expression matrix, genes were $\log_2$-transformed and scaled to unit variance. Posneg transformation was performed to satisfy the non-negativity requirement of NMF, and NMF was applied to each transformed matrix for 2 to 20 clusters (cell states) using Kullback-Leibler (KL) divergence minimization[71]. To determine the number of cell states for each cell type, we applied a heuristic approach based on classification stability measured using the cophenetic coefficient[71]. For each cell type, the number of cell states was selected by determining the last occurrence in the interval of 2 to 20 for which the cophenetic coefficient dropped below 0.98. The number of cell states immediately adjacent to this crossing point with a cophenetic coefficient closest to 0.98 was selected. Low-quality cell states with <10 marker genes were removed. In addition, an adaptive false positive index was implemented to remove spurious cell states driven by negative features as a result of posneg transformation prior to NMF[20]. In total, 23 distinct cell states with 2 to 4 cell states per cell type were identified and included in downstream analysis.

Discovery of sarcoma ecotypes

We implemented EcoTyper to identify conserved cellular communities, termed ecotypes, defined by co-occurring cell states across sarcoma samples[20]. Each sample was assigned to most abundant cell state per cell type, and a binary matrix with cell states as rows and samples as columns was constructed. A Jaccard index matrix was calculated for all pairwise combinations of cell states, and a hypergeometric test was run to test the hypothesis of no overlap. The Jaccard indices for cell state pairs for which the null hypothesis could not be rejected was set to 0, and unsupervised hierarchical clustering was applied to the Jaccard index matrix. The optimal number of clusters was determined by silhouette width maximization, and clusters with ≤2 cell states were removed from further analysis. Using this approach, we identified 3 clusters, which were termed sarcoma ecotypes (SEs). To estimate the abundance of each SE, the cell state abundance within each of the 3 ecotypes was averaged. The resulting values were normalized to sum to 1 across all SEs in each sample, providing a continuous abundance for each SE. To assign samples to SEs, a two-sided t test with unequal variance was applied to evaluate the difference in estimated abundance between the cell states in each SE relative to the abundance of all cell states in other SEs. The resulting p values were corrected for multiple hypothesis testing. Samples were assigned to the SE with the highest abundance if the q-value was ≤0.25 and the sample was assigned to at least one cell state within the SE. Network plots representing each SE were constructed using the 'igraph' R package with edge weights proportional to the Jaccard index between cell states.

Recovery of sarcoma-specific cell states and ecotypes from independent datasets

We utilized the EcoTyper reference-guided annotation framework to recover and quantitate the cell states and sarcoma ecotypes identified from the non-LMS TCGA samples in other gene expression datasets[20]. Briefly, EcoTyper utilizes the learnt NMF model to determine a coefficient matrix for each cell type with each cell state represented as a weight. The EcoTyper recovery module was applied separately to each dataset (Supplementary Table 1), and each gene was $\log_2$-transformed and scaled to unit variance prior to analysis. We have previously validated the ability of EcoTyper to recover cell states and ecotypes across different bulk (RNA-Seq and microarray) and single cell gene expression platforms[20].

UMAP visualization of sarcoma ecotyper cell states

To generate the UMAP plots displayed in Fig. 2a, we analyzed the CIBERSORTx digitally-purified GEPs for canonical cell types from each sample in the training cohort. Samples were assigned to the most abundant cell state and included in the analysis if the cell state abundance was ≥33% and the most abundant cell state abundance was ≥15% more prevalent than the next most abundant cell state. The digitally-purified GEPs were $\log_2$-transformed and scaled to unit variance within each dataset prior to dimensionality reduction.

Published single-cell RNA-Seq datasets

Previously published SS 10x Genomics and SMART-seq2 scRNA-Seq data from Jerby-Arnon et al. Nature Medicine 2021 were downloaded from GEO (accession GSE131309). Author-provided cell type annotations were used for cell state recovery, and epithelial-like

and fibroblast-like malignant cells were identified using the epithelial and mesenchymal modules as described below.

Single-cell RNA sequencing

Fresh tumors were minced and enzymatically dissociated using the Human Tumor Dissociation Kit (Miltenyi Biotec) using a GentleMACS Octo Dissociator with heaters (Miltenyi Biotec) following the manufacturer's instructions. The dissociated cell suspension was then applied to a 70 um MACS SmartStrainer (130-098-462; Miltenyi Biotec). The quality of the suspension was checked by generating a hematoxylin and eosin (H&E) smear and analyzed using a TC10 automated cell counter (BioRad). Samples were further processed by the Stanford Genomics Shared Resource using the Chromium 3' Gene Expression Solution v3.1 (10x Genomics) per the manufacturer's instructions and sequenced on a NovaSeq 6000 (Illumina).

Analysis of scRNA-Seq data

For the scRNA-Seq samples new to this work, sample demultiplexing, read alignment to the GRCh38 human reference genome, and generation of feature-barcode matrices were performed using CellRanger v6.0.0 (10x Genomics). Feature-barcode matrices were analyzed using Seurat[72]. Cells with ≥15% of reads mapped to mitochondrial genes and ≤200 or ≥9000 expressed genes were excluded from the analysis. Data were normalized and scaled prior to principal component analysis (PCA) using the top 2,000 variable genes by default. The k-nearest neighbors (k-NN) were calculated, and a shared nearest neighbor graph was constructed using the top 25 principal components (PCs) based on the jackstraw procedure. Finally, clusters were identified with a resolution of 1. We observed similar results when choosing different numbers of PCs and resolutions.

Identification of malignant and normal single cells

Malignant and normal single cells were identified separately for each tumor sample. First, the scRNA-Seq profile of each cell was compared to bulk RNA-sequencing of the same sarcoma histology from TCGA and normal cells from the Jerby-Arnon et al. Nature Medicine 2021 10x Genomics scRNA-Seq dataset. For each cell, the average Spearman correlation between its expression profile and the expression profiles for each normal cell was subtracted from the average Spearman correlation between its expression profile and the expression profile from bulk sarcomas. For bulk sarcomas, only genes in the top and bottom deciles of expression were included in the analysis. Cells with a higher differential similarity to bulk sarcomas were considered more likely to be malignant[15]. Second, CNAs were inferred from the scRNA-Seq data using the 'inferCNV' R package[73] with a cutoff = 0.05, HMM = TRUE, tumor_subcluster_partition_method = "qnorm", k_obs_groups = 2, and otherwise default parameters. For SRC171, better overlap with the differential similarity to bulk tumors was observed when observations were split into 3 groups, so k_obs_groups was set to 3 for this sample. Macrophages, endothelial cells, and cancer associated fibroblasts from the Jerby-Arnon et al. Nature Medicine 2021 10x Genomics dataset were used as the normal reference. The presence of CNAs and similarity to bulk sarcoma transcriptomes for each cell was imported into Seurat and visualized using t-distributed stochastic neighbor

embedding (t-SNE). Clusters were assigned as normal or malignant cells based on both metrics.

Normal cell analysis

Normal single cells from sarcomas of the same histology were combined and integrated using the FindIntegrationAnchors and IntegrateData functions in Seurat. The integrated data was scaled, variable features were identified, PCA was performed, and clusters were identified as described above. Marker genes for each cluster were identified using the FindAllMarkers function with min.pct = 0.25, logfc.threshold = 0.25, and only.pos=TRUE. Clusters were assigned to cell types based on canonical and previously described markers (Supplementary Table 6).

Malignant cell analysis

Malignant sarcoma cells from different patients were expected to cluster separately due to underlying genomic alterations, so data integration was not performed prior to the malignant cell analysis. The expression of epithelial markers within malignant sarcomas cells was visualized using the FeaturePlot function in Seurat. Epithelial versus mesenchymal differentiation within malignant cells was queried using the AddModuleScore function to calculate the average expression levels of two gene sets defined previously by Jerby-Arnon et al. from epithelial-like and mesenchymal-like malignant SS cell populations[15]. To generate a score of epithelial versus mesenchymal differentiation, the mesenchymal module score was subtracted from the epithelial module score. Malignant cells with an epithelial versus mesenchymal differentiation score >0 were assigned to epithelial-like cells, and cells with an epithelial versus mesenchymal differentiation score ≤0 were assigned to fibroblast-like cells. We also explored the spectrum of epithelial versus mesenchymal differentiation in malignant sarcoma cells using three signatures of EMT[34–36]. The degree of epithelial versus mesenchymal differentiation for each EMT signature was calculated by subtracting the EMT up score from the EMT down score. Because there is not a hallmark gene set for genes downregulated during EMT, the hallmark EMT score was subtracted from 0 to be consistent with the other analyses. All gene sets are displayed in Supplementary Table 11.

Recovery of EcoTyper cell states from scRNA-Seq

Annotated single-cell transcriptomes were mapped to EcoTyper states as described above. The statistical significance of cell state recovery was evaluated via permutation testing to produce a z-score measuring statistical confidence[20]. Only the cell states identified within each scRNA-Seq dataset were evaluated, and the resulting z-scores were combined into a meta z-score using Stouffer's method.

State-specific marker genes in scRNA-Seq data

To extend upon the marker genes assigned during cell state discovery from bulk RNA-Seq, we identified marker genes using the single cells assigned to each EcoTyper cell state[20]. The single-cell samples were grouped into three datasets by histology: 1) SS profiled by SMART-seq2 from the external Jerby-Arnon et al. Nature Medicine 2021 dataset, 2) LMS profiled by 10x Genomics in this work, and 3) UPS profiled by 10x Genomics in this work.

For each of the cell states with at least 15 assigned cells in each scRNA-Seq dataset, we calculated 6 metrics to prioritize marker genes for selection: 1) the number of datasets n_1 in which the mean gene expression for the cell state was >0, 2) the number of datasets n_2 in which the $\log_2$ fold change of the gene within the cell state was greatest relative to the other cell states within the same cell type, 3) the ratio of n_2/n_1, 4) the number of datasets in which the gene was significantly differentially expressed within the cell state based on the *FindMarkers* function in Seurat (Q<0.05), 5) the aggregate z-score for differential expression of the gene across datasets combined using Stouffer's method, and 6) the mean $\log_2$ fold change of the gene within the cell state across datasets. For each cell state, the metrics were converted to rank space and averaged across measures to generate a composite score. The genes with the highest composite score were combined with manually curated genes in Fig. 2b. All manually curated genes are noted in Supplementary Table 8.

Cell state annotation

Normal cell states were manually annotated based on the bulk RNA-Seq and scRNA-Seq marker genes. Key marker genes and references supporting each assignment are noted in Supplementary Table 10. To aid in the annotation of malignant cell states, we computed the overlap between the bulk RNA-Seq marker genes identified by the EcoTyper pipeline for each cell state and hallmark gene sets using the 'Compute Overlaps' function from the MSigDB web interface (https://www.gsea-msigdb.org/gsea/msigdb/human/annotate.jsp)[34,74]. P values were calculated from the hypergeometric distribution and the q-values displayed in Supplementary Table 9 were calculated using the Benjamini-Hochberg procedure. Key marker genes were chosen from overlapping gene sets based on their expression within single cells assigned to the malignant cell state.

Immunohistochemistry for epithelial markers

A tissue microarray was constructed using one to seven 1 mm FFPE cores per patient. Immunohistochemistry was performed by the Stanford Human Pathology/Histology Service Center for CDH1, MUC1, and SDC1. Tumor cell staining intensity was scored using a 3-point scale (0=negative, 1=weak positive, 2=strong positive).

Biologic features of sarcoma ecotypes

Biologic features associated with each sarcoma ecotype were analyzed in the training cohort. The percentage of patients with each sarcoma histology and proportion of canonical cell types were analyzed based on sarcoma ecotype assignment. Cell type proportions were inferred using CIBERSORTx, averaged, and scaled. Sex and age ≥60 versus <60 years old were analyzed using two-sided Wilcoxon tests comparing the relative abundance of each sarcoma ecotype between groups. Hallmark gene sets were analyzed by scaling each gene to unit variance expression across samples prior to averaging all component genes for each gene set. Enrichment or depletion was analyzed by performing a two-sided Wilcoxon test to compare the average expression for each hallmark gene set in samples assigned to each sarcoma ecotype versus samples assigned to the other sarcoma ecotypes. P values were adjusted for multiple hypothesis testing, and features with a q-value<0.05 were considered significant.

Association of genomic alterations and mutational signatures with sarcoma ecotypes

CNA, SNV, and indel calls from the sarcomas profiled by TCGA were downloaded from the National Cancer Institute Genomic Data Commons Data Portal. The contributions of COSMIC mutational signatures 1, 2, 5, and 13 were analyzed using the 'mutSignatures' R package[75] based on their predominance in STSs[16]. The proportion of patients with CNAs or high impact SNVs/indels in each gene was compared across ecotypes using two-sided Fisher's exact tests. For CNAs, we analyzed COSMIC Tier 1 genes[76] mutated in at least 25% of patients. For SNVs/indels, we analyzed genes mutated in at least 3% of patients. P-values were corrected for multiple hypothesis testing.

Spatial transcriptomics

Fresh sarcoma specimens were harvested and frozen in optimal cutting temperature compound by the Stanford Tissue Procurement Shared Resource. Sarcoma specimens were cryosectioned at −20°C onto gene expression slides by the Stanford Genomics Shared Resource. Permeabilization time was optimized using 10 um sections with the Visium Spatial Tissue Optimization Slide & Reagents kit (10x Genomics). Sequencing libraires were produced using the Visium Spatial Gene Expression Slide & Reagent kit (10x Genomics) according to the manufacturer's instructions and sequenced on a NovaSeq 6000 (Illumina). Sample demultiplexing, tissue and fiducial detection, read alignment to the GRCh38 human reference genome, and generation of feature-barcode matrices was performed using Space Ranger v1.3.0 (10x Genomics).

Recovery of EcoTyper cell states and ecotypes from spatial transcriptomics data

Cell state and ecotype abundance were estimated within each spatially-barcoded spot by imputing the fractional abundance of each sarcoma cell state in the Visium array using EcoTyper[20]. The most abundant cell state per cell type within each spot was set to 1 and the rest to 0. Finally, each cell state was normalized by multiplying it by the parent cell type fraction based on CIBERSORTx. SE abundance was calculated for each spot by averaging the relative fractions of each cell state within the ecotype. SE abundances were scaled so that the 99th percentile value across all 3 SEs was set to 1. Cell state co-localization was analyzed by computing all pairwise Spearman correlation coefficients between cell state abundances within each spot. The distance between spots was calculated by quantitating the Euclidean distance between the spot centers. Spatial aggregation of SEs was investigated using Moran's I[77] based on the relative SE abundance for each spot and the immediately adjacent neighboring spots.

Ligand-receptor analysis

To understand the signaling networks within sarcoma ecotypes, we analyzed the CIBERSORTx digitally-purified cell type-specific GEPs from STSs in the training cohort for enrichment in putative ligand-receptor pairs[78]. Samples were assigned to the most abundant cell state as described above. We assessed whether each ligand and receptor were overexpressed in samples assigned to each cell state versus all other samples using one-sided Mann-Whitney U-tests. P values were corrected for multiple hypothesis testing, and significantly overexpressed ligand and receptor pairs across cell states were identified.

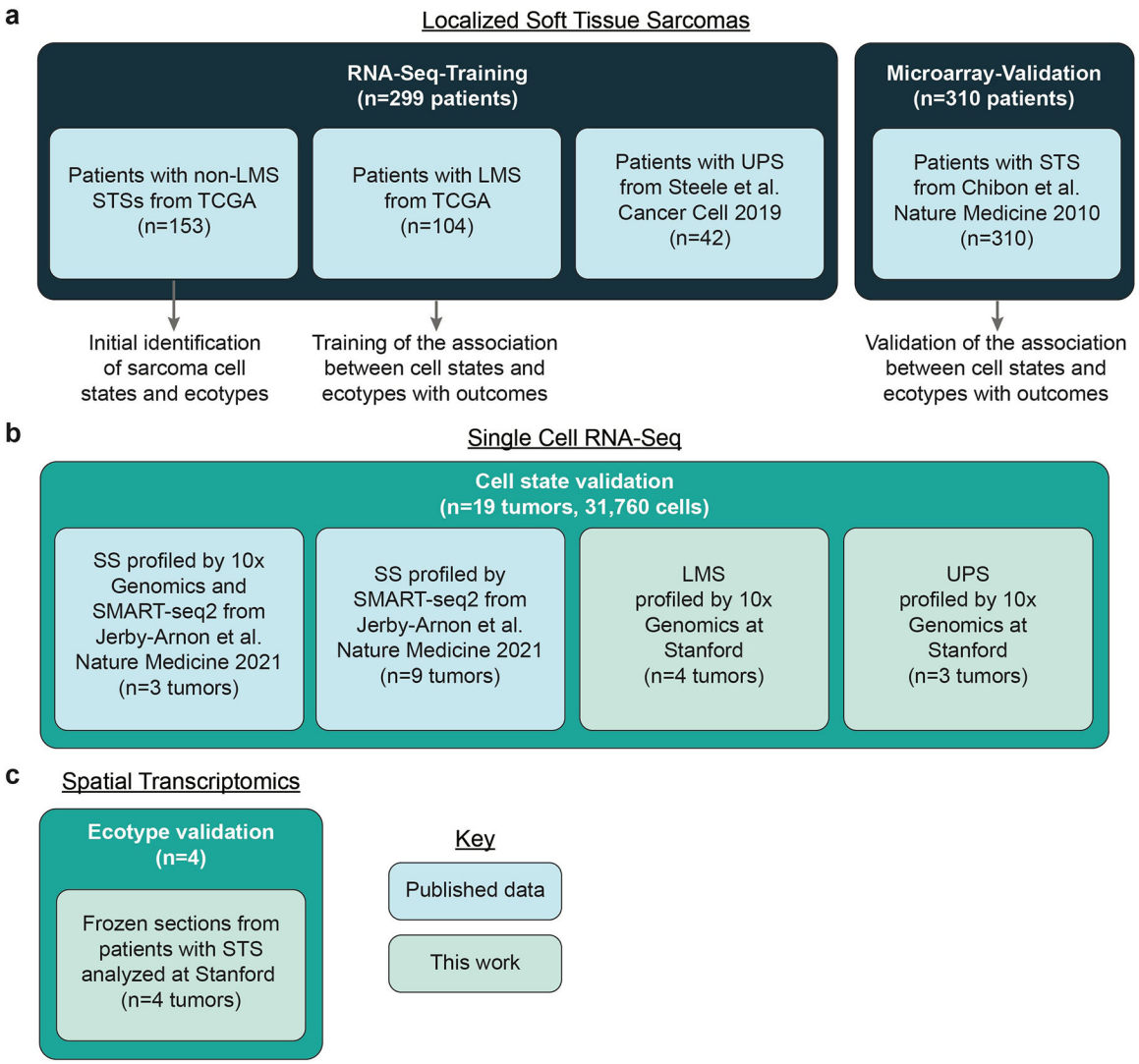
PD-L1 staining and TLS scoring

Immunohistochemistry for PD-L1 was performed on FFPE sections by the Stanford Human Pathology/Histology Service Center using the SP263 clone antibody following antigen retrieval with CC1 buffer for 64 minutes and scored using the combined positive score on a representative histological section of tumor. The presence of TLS was analyzed by morphology on H&E-stained slides of a representative section of tumor and scored as absent or present. Both PD-L1 and TLS scoring was performed by a board-certified anatomic pathologist (G.W.C) blinded to sarcoma ecotypes and clinical outcomes.

Statistics and reproducibility

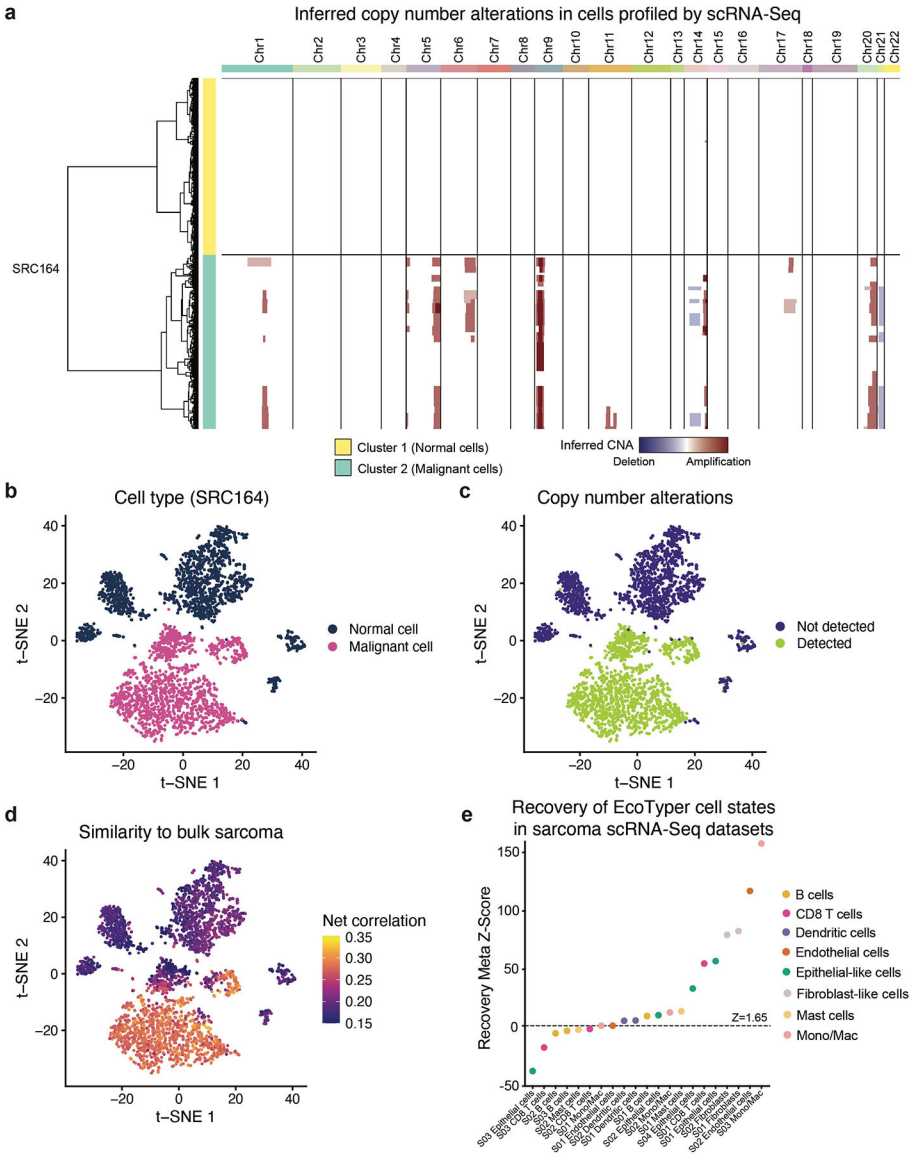
No statistical method was used to predetermine sample size but our sample sizes are similar to those reported in previous publications[8,15,16,18]. Experiments were performed once. The experiments were not randomized. Except for PD-L1 and TLS scoring, data collection and analysis were not performed blind to the conditions of the experiments. Samples failing to meet the quality control metrics described above were excluded from analysis. For the Stanford chemotherapy and ICI cohorts, PFS was defined as the time from the start of treatment to any progression or death. Progression and response were determined radiographically using RECIST 1.1 criteria[79]. Patients without progression or death at the time of last imaging follow-up were censored. PFS and DMFS were calculated using the Kaplan-Meier method and statistical significance was determined using two-sided log-rank tests when comparing two groups or pairwise two-sided log-rank tests with correction for multiple hypothesis testing when comparing multiple groups. Multivariable Cox proportional hazards models were fit with the ‘coxph’ function from the ‘survival’ R package, and the significance of individual variables was assesses using two-sided Wald tests. Sarcoma cell state and ecotype abundance, tumor size, and grade were used as continuous variables for Cox regression, and histology, resection extent, and anatomic site were included as categorical variables. Optimal cutoffs for stratifying patient outcomes based on S01 monocyte/macrophage and S02 endothelial cell abundance were defined in the training cohort using the ‘surv_cutpoint’ function from the ‘survminor’ R package with minprop=0.2 and applied to the validation cohort. The optimal cutpoint for predicting response to ICI using SE3 abundance was defined in the training cohort based on the maximum Youden’s J statistic. Two-sided Fisher’s exact tests were used to compare proportions, and two-sided Mann-Whitney U-tests or Wilcoxon signed-rank tests were used to compare distributions using non-paired and paired samples, respectively. Correlation between variables was assessed using the Spearman correlation coefficient. Data distribution was assumed to be normal, but this was not formally tested, and non-parametric tests were utilized when possible. P-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure. Statistical significance was assumed at $P<0.05$. Statistical analyses were performed with Prism (GraphPad Software) or R through the RStudio environment. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Extended Data



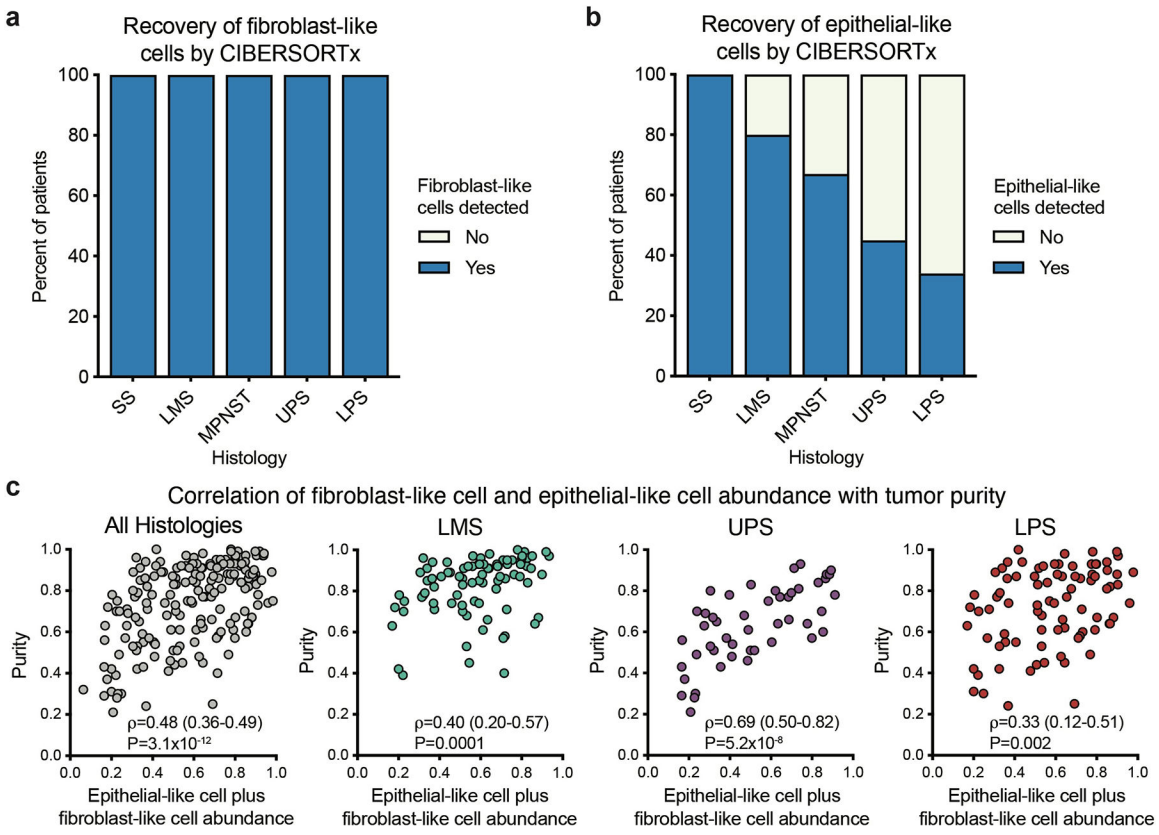
Extended Data Fig. 1. Summary of patient cohorts and datasets utilized for discovery and validation of sarcoma cell states and ecosystems.

a, Two cohorts of patients with localized soft tissue sarcomas (STSs) were analyzed. Soft tissue sarcomas profiled by the Cancer Genome Atlas (TCGA)[16] and a previously published group of patients with undifferentiated pleomorphic sarcoma (UPS)[21] profiled by RNA-sequencing were combined to form the training cohort. Initial identification of sarcoma cell states and ecotypes was performed in non-leiomyosarcoma (LMS) soft tissue sarcomas from TCGA, then the full training cohort was used to assess associations between cell states and ecotypes with patient outcomes. A previously published cohort of patients with localized STS profiled by microarray was used as the validation cohort for associations between cell states and ecotypes with patient outcomes. **b,** Cell states were validated using single-cell RNA-sequencing from 12 previously published synovial sarcomas (SS)[15] and 4 LMS and 3 UPS new to this work. **c,** The spatial distributions of cell states and ecotypes were validated using spatial transcriptomics analysis of 4 STSs.



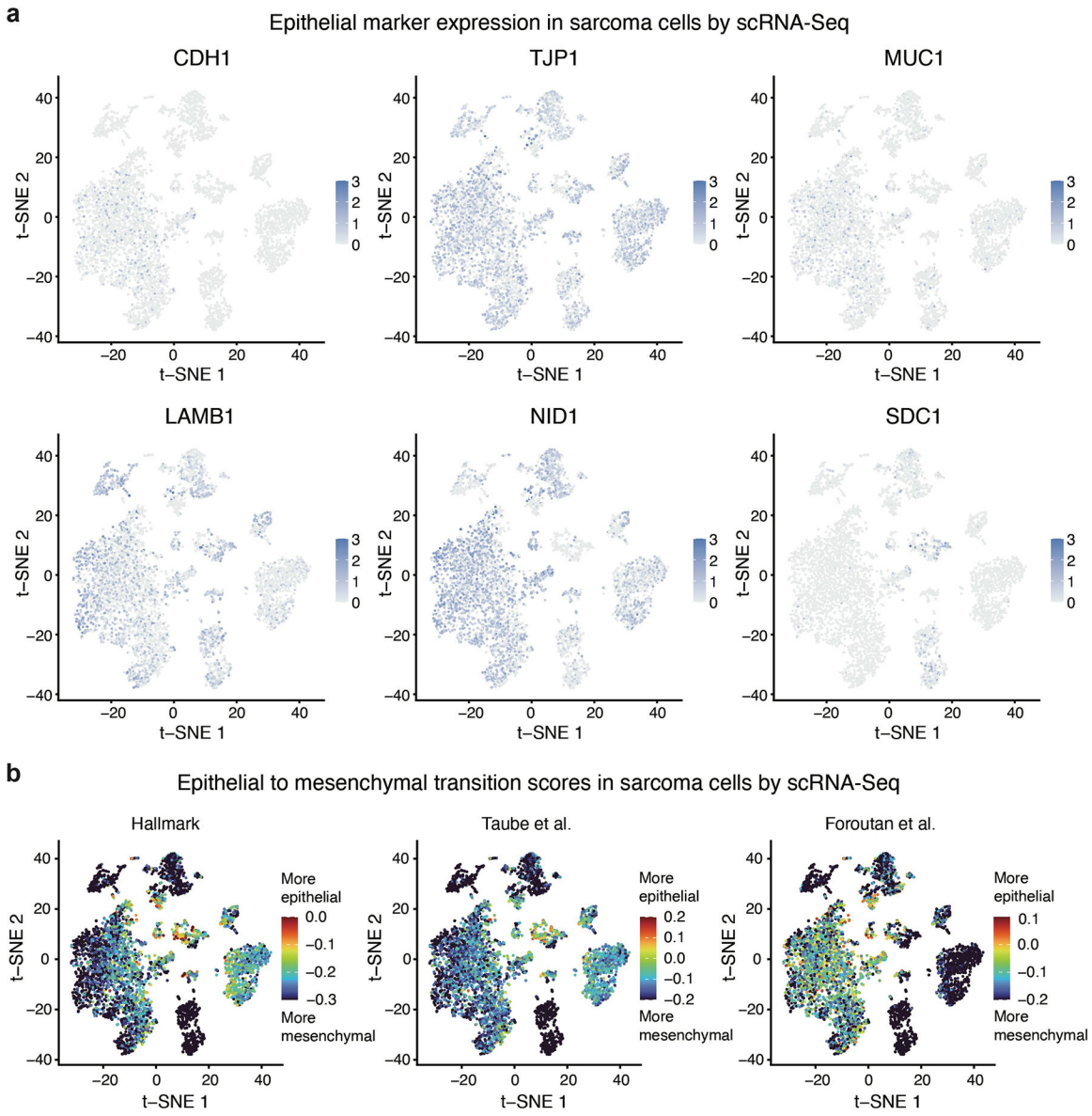
Extended Data Fig. 2. Identification of malignant sarcoma cells and validation of sarcoma ecotyper cell states using scRNA-Seq.

a, Representative plot of inferred copy number alterations (CNAs) in single cells clustered into two groups using inferCNV from SRC164. Amplifications and deletions are shown across each chromosome. **b-d,** t-SNE plots of scRNA-Seq profiles from SRC164 colored by (**b**) assignment to normal or malignant cells, (**c**) detection of CNAs, and (**d**) differential similarity to sarcomas of the same histology profiled by bulk RNA-Seq compared with normal cells profiled by scRNA-Seq. **e,** Significance of EcoTyper cell state recovery across scRNA-Seq datasets measured by permutation testing and aggregated into a meta z-score using Stouffer's method. Z-scores>1.65 (one-sided P value<0.05) are considered significant.



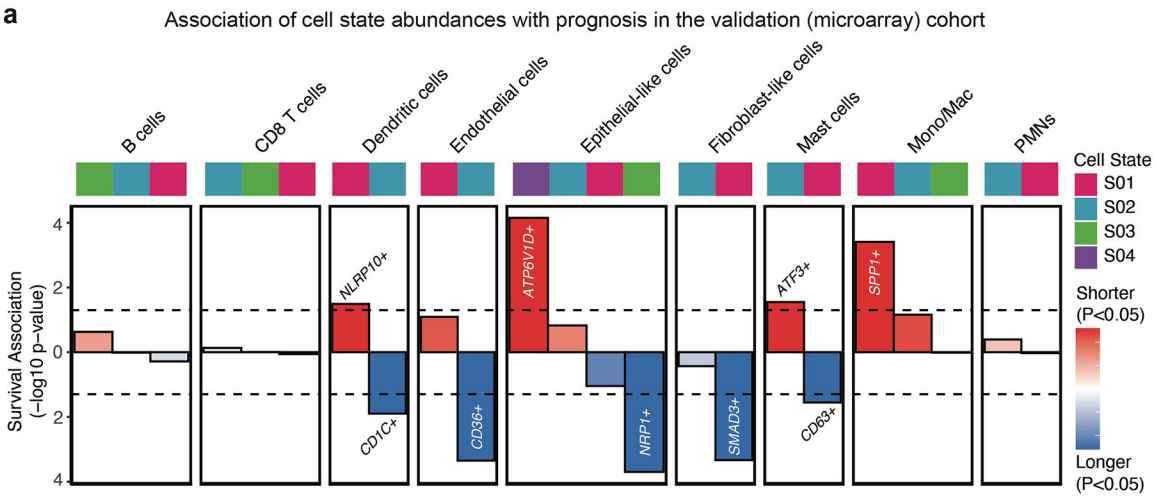
Extended Data Fig. 3. Recovery of malignant cells across soft tissue sarcomas using CIBERSORTx.

a-b, Stacked bar plots displaying the percentage of patients with each soft tissue sarcoma histology from the training cohort with (**a**) fibroblast-like cells and (**b**) epithelial-like cells identified by CIBERSORTx. Synovial sarcoma = SS, leiomyosarcoma = LMS, malignant peripheral nerve sheath tumor = MPNST, undifferentiated pleomorphic sarcoma = UPS, liposarcoma = LPS. **c,** Scatter plots showing the correlation between tumor purity and the combined abundance of epithelial-like cells and fibroblast-like cells by CIBERSORTx. Spearman's correlation coefficient and two-sided P value are displayed on the graph.



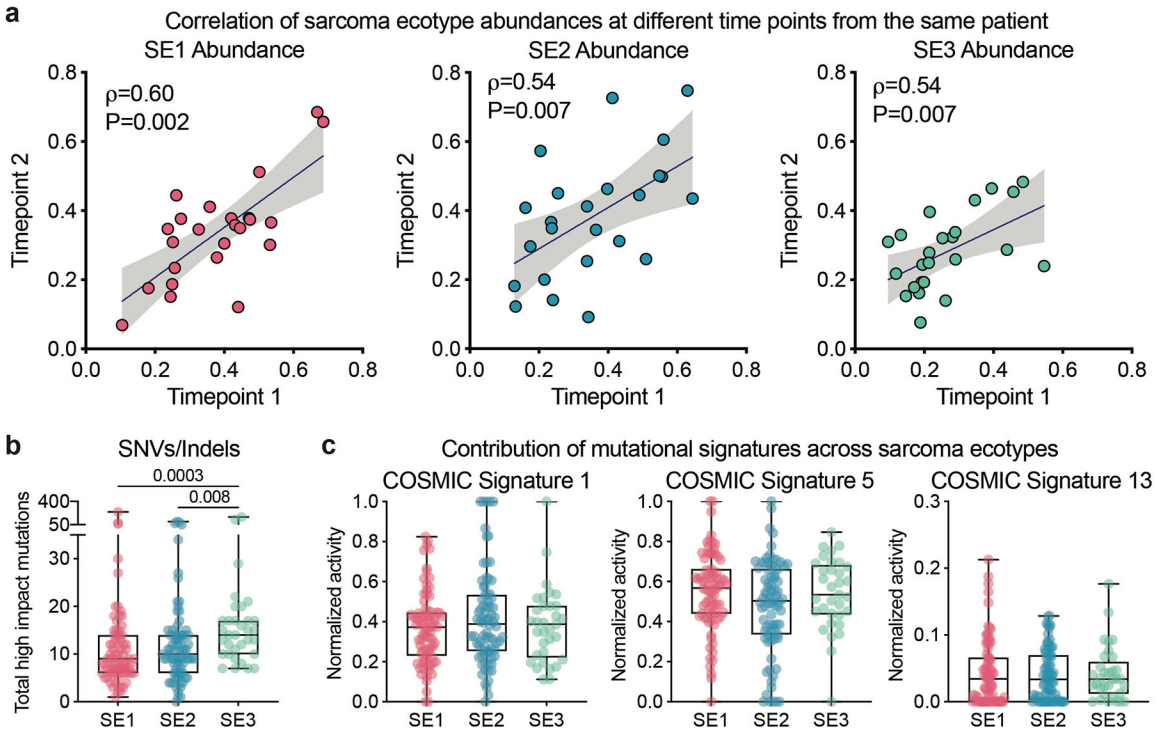
Extended Data Fig. 4. Characterization of epithelial-like malignant cells within soft tissue sarcomas.

a, t-SNE plots displaying expression of epithelial marker genes in malignant sarcoma cells from LMS and UPS tumors profiled by scRNA-Seq. **b,** t-SNE plots of epithelial vs. mesenchymal differentiation in malignant LMS and UPS cells using three previously described signatures of epithelial to mesenchymal transition[34–36]. Sample and histology for each cell are displayed in Fig. 2c. The gene sets are displayed in Supplementary Table 11.



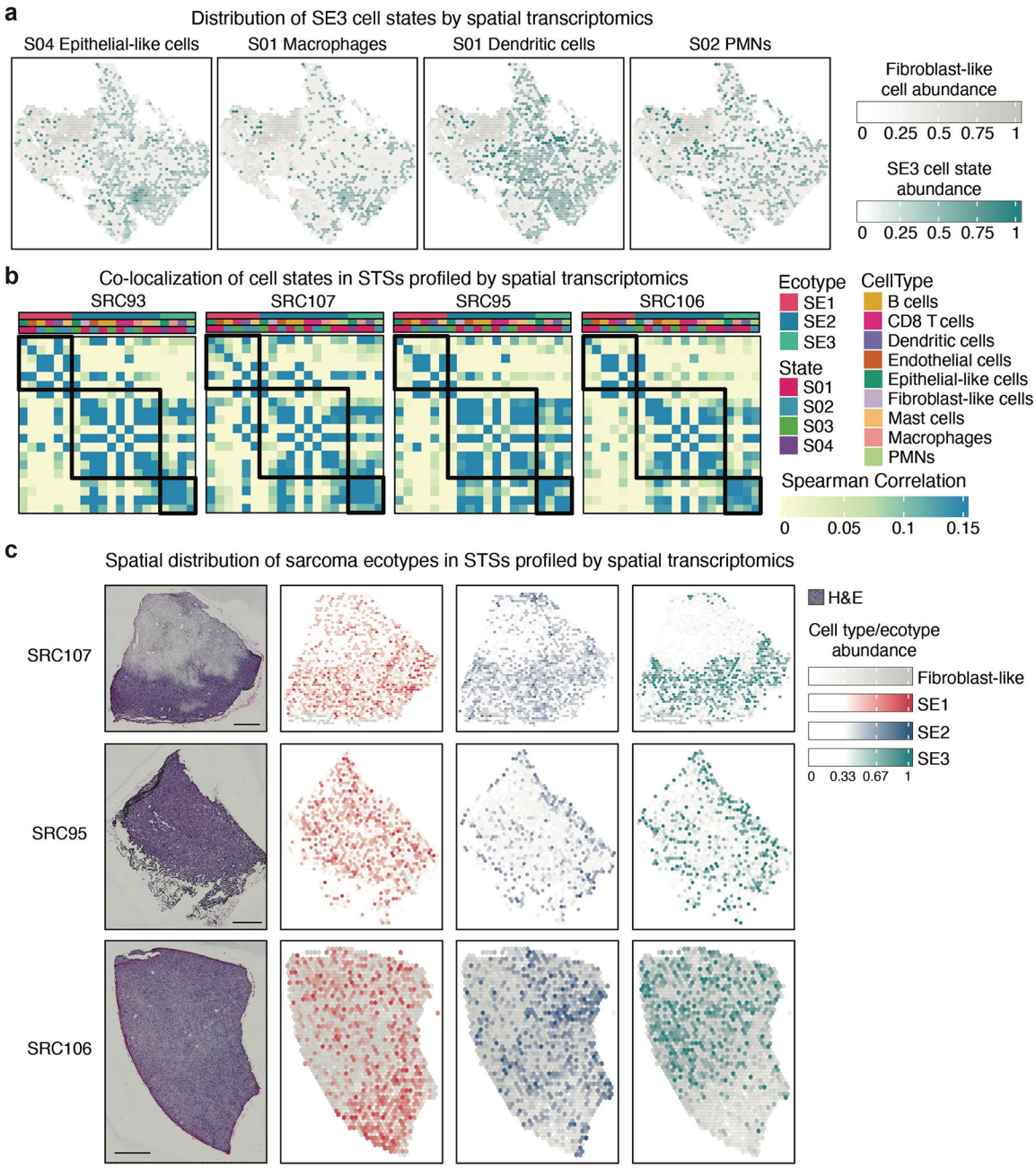
Extended Data Fig. 5. Association of cell state abundances with patient outcomes in the validation cohort.

a, Association of sarcoma-specific cell states identified by EcoTyper with distant metastasis-free survival in the validation cohort. Significance was assess using multivariable Cox proportional hazards models including cell state abundance as a continuous variable along with sarcoma histology. P values were calculated using two-sided Wald tests without correction for multiple comparisons. Marker genes are displayed for significantly associated cell states. Patient survival, histologies, and cell state abundances used in the analysis are shown in Supplementary Table 3. Only patients with survival data available were included in the analysis.



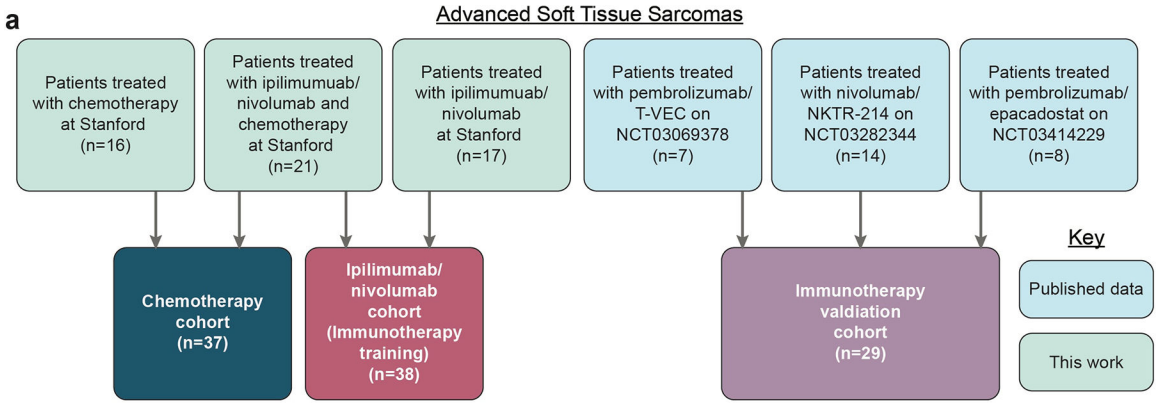
Extended Data Fig. 6. Correlation of sarcoma ecotypes across time and association with genomic alterations.

a, Scatter plots showing the correlation between ecotype abundance across different timepoints from the same patient. Spearman's correlation coefficients and two-sided P values are displayed. The lines of best fit by linear regression and 95% confidence intervals for the lines of best fit are shown on the graph. The data used for this analysis are shown in Supplementary Table 20. **b,c,** Box plots displaying (**b**) the total number of high impact SNVs/Indels and (**c**) normalized contribution of COSMIC mutational signatures 1, 5, and 13 by sarcoma ecotype (n=79 SE1, 78 SE2, and 33 SE3 sarcomas for both panels). P values were calculated using Kruskal-Wallis tests followed by Dunn's tests for pairwise comparisons. Boxes show median and quartiles, and whiskers extend to the minimum and maximum value.



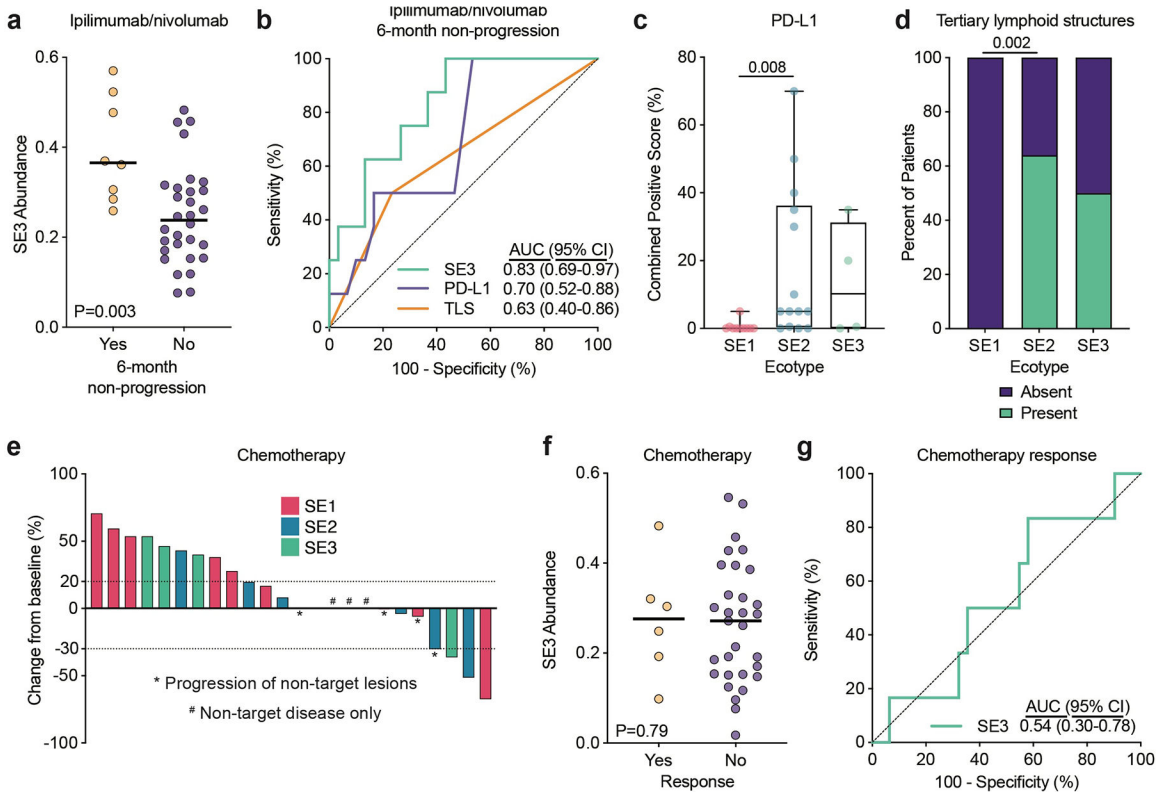
Extended Data Fig. 7. Validation of sarcoma ecotypes using spatial transcriptomics.

a, Distribution of cell states from sarcoma ecotype 3 (SE3) in SCR93, a sarcoma profiled by spatial transcriptomics. Abundance of the cell states that make up SE3 within each spatial transcriptomics spot are shown, and fibroblast-like cell abundance is plotted to show the tumor outline. **b,** Heatmaps displaying Spearman correlation of cell state abundances within spatial transcriptomics spots. **c,** Distribution of sarcoma ecotypes in three sarcomas profiled by spatial transcriptomics. H&E staining along with the abundance of SEs within each spatial transcriptomics spot are shown. Fibroblast-like cell abundance is plotted to show the tumor outline. Scale bars show 1000 μm. A total of four sarcomas were profiled, and SRC93 is displayed in Fig. 5f.



Extended Data Fig. 8. Summary of advanced soft tissue sarcoma cohorts.

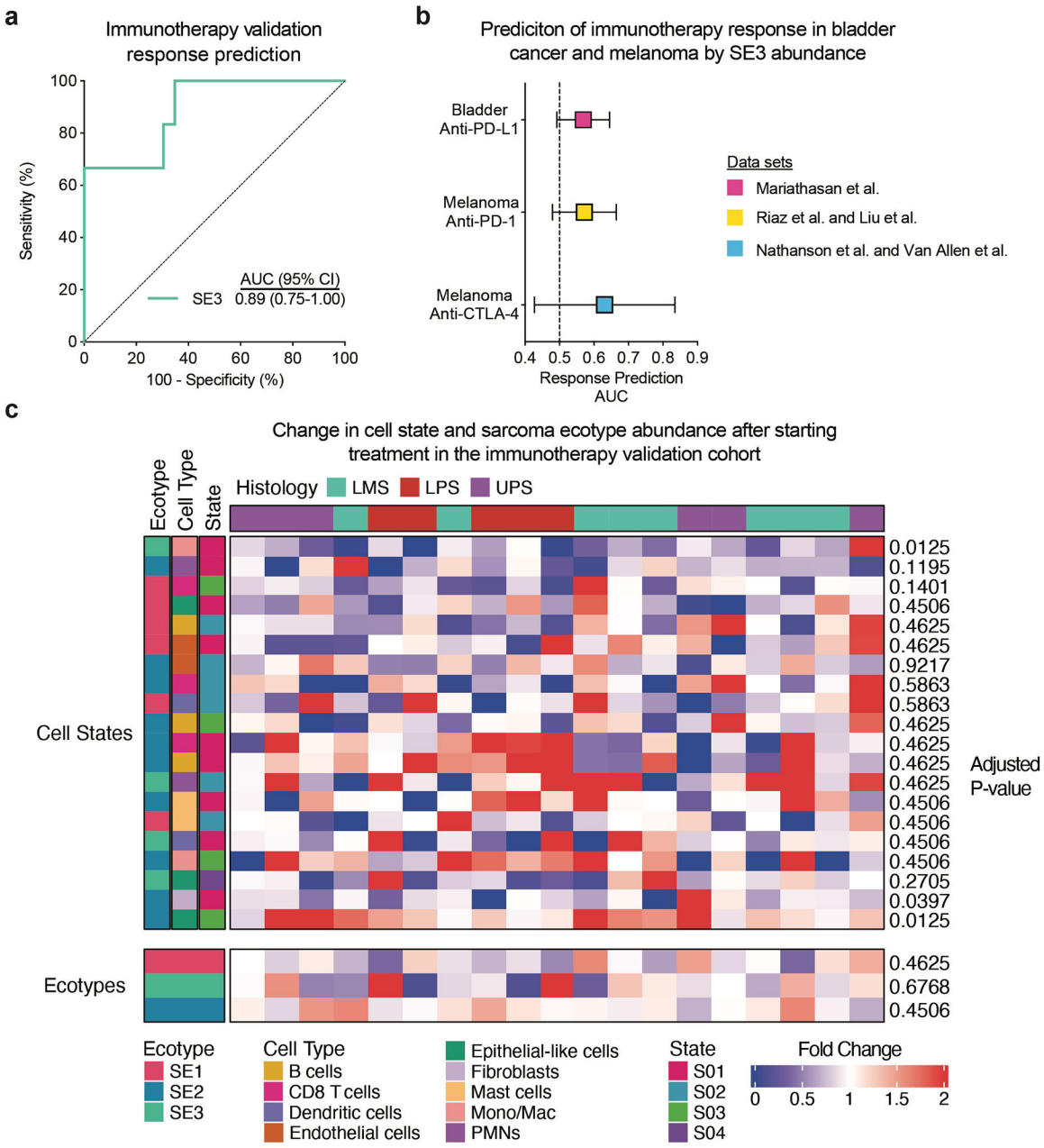
Two cohorts of patients with advanced soft tissue sarcoma treated at Stanford were analyzed based on the type of systemic therapy received. Patients treated with both chemotherapy and ipilimumab/nivolumab were included in both cohorts. The ICI validation cohort consisted of patients with advanced soft tissue sarcomas treated with anti-PD-1 antibodies (pembrolizumab or nivolumab) in combination with experimental immunotherapies (talimogene laherparepvec=T-VEC, bempegaldesleukin=NKTR-214, or epacadostat) as part of 3 clinical trials.



Extended Data Fig. 9. Association of sarcoma ecotypes with response to ICI and chemotherapy.

a, Plot of SE3 abundance in patients with and without 6-month non-progression after starting ipilimumab and nivolumab (n=8 patients with and 30 patients without 6-month non-

progression). P value was calculated using a two-sided Mann-Whitney U-test. **b,** Receiver operating characteristic curves for prediction of 6-month non-progression on ipilimumab and nivolumab by SE3 abundance, PD-L1 expression, and the presence of tertiary lymphoid structures (TLS). Area under the curve (AUC) and 95% confidence intervals (95% CI) are displayed on the graph. **c,** Box plot of PD-L1 combined positive score across sarcoma ecotypes (n=10 SE1, 14 SE2, and 4 SE3 sarcomas). P values were calculated using Kruskal-Wallis tests followed by Dunn's tests for pairwise comparisons. Boxes show median and quartiles, and whiskers extend to the minimum and maximum value. **d,** Stacked bar plot of the presence of tertiary lymphoid structures across sarcoma ecotypes. P values were calculated using two-sided Fisher's exact tests. **e,** Waterfall plot showing the best response by RECIST criteria for patients with advanced STSs treated with chemotherapy based on sarcoma ecotype assignment. Horizontal dotted lines represent the criteria for progressive disease (20% increase) and partial response (30% decrease). Patients with only non-target disease are plotted at 0%. **f,** Plot of SE3 abundance in patients with and without a response to chemotherapy (n=6 responders and 31 non-responders). P value was calculated using a two-sided Mann-Whitney U-test. **g,** Receiver operating characteristic curves for prediction of response to chemotherapy by SE3 abundance. AUC and 95% CI are displayed on the graph. Patient survival, treatment response, tumor characteristics, ecotype assignments, and ecotype abundances used in the analysis of the chemotherapy and ipilimumab/nivolumab cohorts are shown in Supplementary Tables 18–20. For panels c, d, and e, patients were analyzed based on ecotype assignment, and patients not assigned to an ecotype were not included. For panels a, b, f, and g, patients were analyzed based on ecotype abundance, and all patients were included in the analysis.



Extended Data Fig. 10. Validation of ICI response prediction by SE3 abundance and changes in sarcoma cell states and ecotypes during treatment with ICI.

a, Receiver operating characteristic curve and area under the curve (AUC) with 95% confidence intervals (95% CI) for prediction of treatment response in the ICI validation cohort. **b,** Areas under the curve for prediction of response to immune checkpoint inhibition based on SE3 abundance in patients with metastatic bladder cancer and melanoma (n=348 bladder anti-PD-L1, 172 melanoma anti-PD-1, and 51 melanoma anti-CTLA-4). Error bars display the 95% confidence interval. **c,** Heatmap of fold change in cell state and sarcoma ecotype abundances on-treatment in the ICI validation cohort. Adjusted P values comparing pre-treatment and on-treatment samples are displayed. P values were calculated using two-

sided Wilcoxon signed-rank tests and corrected for multiple hypothesis testing. Patient treatment response and ecotype abundances used in analysis of the ICI validation cohort and bladder/melanoma cohorts are shown in Supplementary Table 21 and Supplementary Table 22, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the patients and families who participated in this study. We would like to acknowledge N. Pillay and the University College London for providing sequencing data included in the analyses of this manuscript. This work used the Stanford Genomics Shared Resource which is supported by NIH grants S10OD025212 and 1S10OD021763. This work was supported by the Department of Defense (W81XWH-22-1-0161, E.J.M.), the National Cancer Institute (K08CA25542501, E.J.M.), the Stanford Cancer Institute, an NCI-designated Comprehensive Cancer Center (N.Q.B., E.J.M.), the My Blue Dots organization (E.J.M.), and the Tad and Diane Taube Family Foundation (D.G.M., M.v.R., E.J.M.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The schematic in Fig. 1 was created with BioRender.com.

DATA AVAILABILITY

The bulk RNA-Seq data, single-cell RNA-Seq data, and spatial transcriptomics data new to this manuscript have been deposited at GEO under accession numbers GSE213065, GSE212527, and GSE212526, respectively. Previously published RNA-Seq, microarray, and single-cell RNA-Seq data that were re-analyzed here are available through the TCGA Research Network (http://cancergenome.nih.gov/), through GEO under accession numbers GSE21050 and GSE131309, through the European Genome-Phenome Archive under dataset ID EGAD00001004439, through the 'Imvigor210CoreBiologies' R package[66], and in the supplementary material of the published manuscripts[41,67–69]. Source data for all figures has been provided as Source Data files. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

1. Dufresne A, Brahmi M, Karanian M, and Blay J-Y (2018). Using biology to guide the treatment of sarcomas and aggressive connective-tissue tumours. Nat Rev Clin Oncol 15, 443–458. 10.1038/s41571-018-0012-4. [PubMed: 29666441]
2. Wang D, Zhang Q, Eisenberg BL, Kane JM, Li XA, Lucas D, Petersen IA, DeLaney TF, Freeman CR, Finkelstein SE, et al. (2015). Significant Reduction of Late Toxicities in Patients With Extremity Sarcoma Treated With Image-Guided Radiation Therapy to a Reduced Target Volume: Results of Radiation Therapy Oncology Group RTOG-0630 Trial. J. Clin. Oncol 33, 2231–2238. 10.1200/JCO.2014.58.5828. [PubMed: 25667281]
3. Judson I, Verweij J, Gelderblom H, Hartmann JT, Schöffski P, Blay J-Y, Kerst JM, Sufliarsky J, Whelan J, Hohenberger P, et al. (2014). Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. Lancet Oncol. 15, 415–423. 10.1016/S1470-2045(14)70063-4. [PubMed: 24618336]
4. Ryan CW, Merimsky O, Agulnik M, Blay J-Y, Schuetze SM, Van Tine BA, Jones RL, Elias AD, Choy E, Alcindor T, et al. (2016). PICASSO III: A Phase III, Placebo-Controlled Study of Doxorubicin With or Without Palifosfamide in Patients With Metastatic Soft Tissue Sarcoma. J Clin Oncol 34, 3898–3905. 10.1200/JCO.2016.67.6684. [PubMed: 27621408]

5. Tap WD, Papai Z, Van Tine BA, Attia S, Ganjoo KN, Jones RL, Schuetze S, Reed D, Chawla SP, Riedel RF, et al. (2017). Doxorubicin plus evofosfamide versus doxorubicin alone in locally advanced, unresectable or metastatic soft-tissue sarcoma (TH CR-406/SARC021): an international, multicentre, open-label, randomised phase 3 trial. Lancet Oncol 18, 1089–1103. 10.1016/S1470-2045(17)30381-9. [PubMed: 28651927]
6. Callegaro D, Miceli R, Bonvalot S, Ferguson P, Strauss DC, Levy A, Griffin A, Hayes AJ, Stacchiotti S, Pechoux CL, et al. (2016). Development and external validation of two nomograms to predict overall survival and occurrence of distant metastases in adults after surgical resection of localised soft-tissue sarcomas of the extremities: a retrospective analysis. Lancet Oncol 17, 671–680. 10.1016/S1470-2045(16)00010-3. [PubMed: 27068860]
7. Gronchi A, Miceli R, Shurell E, Eilber FC, Eilber FR, Anaya DA, Kattan MW, Honoré C, Lev DC, Colombo C, et al. (2013). Outcome prediction in primary resected retroperitoneal soft tissue sarcoma: histology-specific overall survival and disease-free survival nomograms built on major sarcoma center data sets. J Clin Oncol 31, 1649–1655. 10.1200/JCO.2012.44.3747. [PubMed: 23530096]
8. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, Tirode F, Lucchesi C, de Reynies A, Kauffmann A, Bui B, et al. (2010). Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nature Medicine 16, 781–787. 10.1038/nm.2174.
9. Blay J-Y, and Le Cesne A (2009). Adjuvant chemotherapy in localized soft tissue sarcomas: still not proven. Oncologist 14, 1013–1020. 10.1634/theoncologist.2009-0126. [PubMed: 19808771]
10. D'Angelo SP, Mahoney MR, Van Tine BA, Atkins J, Milhem MM, Jahagirdar BN, Antonescu CR, Horvath E, Tap WD, Schwartz GK, et al. (2018). Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. Lancet Oncol 19, 416–426. 10.1016/S1470-2045(18)30006-8. [PubMed: 29370992]
11. Somaiah N, Conley AP, Parra ER, Lin H, Amini B, Solis Soto L, Salazar R, Barreto C, Chen H, Gite S, et al. (2022). Durvalumab plus tremelimumab in advanced or metastatic soft tissue and bone sarcomas: a single-centre phase 2 trial. Lancet Oncol 23, 1156–1166. 10.1016/S1470-2045(22)00392-8. [PubMed: 35934010]
12. Tawbi HA, Burgess M, Bolejack V, Van Tine BA, Schuetze SM, Hu J, D'Angelo S, Attia S, Riedel RF, Priebat DA, et al. (2017). Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. Lancet Oncol 18, 1493–1501. 10.1016/S1470-2045(17)30624-1. [PubMed: 28988646]
13. Keung EZ, Burgess M, Salazar R, Parra ER, Rodrigues-Canales J, Bolejack V, Van Tine BA, Schuetze SM, Attia S, Riedel RF, et al. (2020). Correlative Analyses of the SARC028 Trial Reveal an Association Between Sarcoma-Associated Immune Infiltrate and Response to Pembrolizumab. Clinical Cancer Research 26, 1258–1266. 10.1158/1078-0432.CCR-19-1824. [PubMed: 31900276]
14. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 24, 541–550. 10.1038/s41591-018-0014-x. [PubMed: 29686425]
15. Jerby-Arnon L, Neftel C, Shore ME, Weisman HR, Mathewson ND, McBride MJ, Haas B, Izar B, Volorio A, Boulay G, et al. (2021). Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. Nat Med 27, 289–300. 10.1038/s41591-020-01212-6. [PubMed: 33495604]
16. Cancer Genome Atlas Research Network (2017). Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell 171, 950–965.e28. 10.1016/j.cell.2017.10.014. [PubMed: 29100075]
17. Italiano A, Bessede A, Pulido M, Bompas E, Piperno-Neumann S, Chevreau C, Penel N, Bertucci F, Toulmonde M, Bellera C, et al. (2022). Pembrolizumab in soft-tissue sarcomas with tertiary lymphoid structures: a phase 2 PEMBROSARC trial cohort. Nat Med. 10.1038/s41591-022-01821-3.

18. Petitprez F, de Reyniès A, Keung EZ, Chen TW-W, Sun C-M, Calderaro J, Jeng Y-M, Hsiao L-P, Lacroix L, Bougoüin A, et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. Nature 577, 556–560. 10.1038/s41586-019-1906-8. [PubMed: 31942077]
19. Sautès-Fridman C, Petitprez F, Calderaro J, and Fridman WH (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. Nat Rev Cancer 19, 307–325. 10.1038/s41568-019-0144-6. [PubMed: 31092904]
20. Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C, Przybyl J, Espín-Pérez A, Diehn M, Alizadeh AA, et al. (2021). Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell, S0092-8674(21)01061-8. 10.1016/j.cell.2021.09.014.
21. Steele CD, Tarabichi M, Oukrif D, Webster AP, Ye H, Fittall M, Lombard P, Martincorena I, Tarpey PS, Collord G, et al. (2019). Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. Cancer Cell 35, 441–456.e8. 10.1016/j.ccell.2019.02.002. [PubMed: 30889380]
22. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol 37, 773–782. 10.1038/s41587-019-0114-2. [PubMed: 31061481]
23. Li Y, Liu H, Zhao Y, Yue D, Chen C, Li C, Zhang Z, and Wang C (2021). Tumor-associated macrophages (TAMs)-derived osteopontin (OPN) upregulates PD-L1 expression and predicts poor prognosis in non-small cell lung cancer (NSCLC). Thorac Cancer 12, 2698–2709. 10.1111/1759-7714.14108. [PubMed: 34423566]
24. Sun G, Cao Y, Qian C, Wan Z, Zhu J, Guo J, and Shi L (2020). Romo1 is involved in the immune response of glioblastoma by regulating the function of macrophages. Aging (Albany NY) 12, 1114–1127. 10.18632/aging.102648. [PubMed: 31945745]
25. Tong L, Li J, Choi J, Pant A, Xia Y, Jackson C, Liu P, Yi L, Boussouf E, Lim M, et al. (2020). CLEC5A expressed on myeloid cells as a M2 biomarker relates to immunosuppression and decreased survival in patients with glioma. Cancer Gene Ther 27, 669–679. 10.1038/s41417-019-0140-8. [PubMed: 31591460]
26. Viitala M, Virtakoivu R, Tadayon S, Rannikko J, Jalkanen S, and Hollmén M (2019). Immunotherapeutic Blockade of Macrophage Clever-1 Reactivates the CD8+ T-cell Response against Immunosuppressive Tumors. Clin Cancer Res 25, 3289–3303. 10.1158/1078-0432.CCR-18-3016. [PubMed: 30755440]
27. Krausgruber T, Blazek K, Smallie T, Alzabin S, Lockstone H, Sahgal N, Hussell T, Feldmann M, and Udalova IA (2011). IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. Nat Immunol 12, 231–238. 10.1038/ni.1990. [PubMed: 21240265]
28. Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, Gordon S, Hamilton JA, Ivashkiv LB, Lawrence T, et al. (2014). Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity 41, 14–20. 10.1016/j.immuni.2014.06.008. [PubMed: 25035950]
29. Piedra-Quintero ZL, Serrano C, Villegas-Sepúlveda N, Maravillas-Montero JL, Romero-Ramírez S, Shibayama M, Medina-Contreras O, Nava P, and Santos-Argumedo L (2018). Myosin 1F Regulates M1-Polarization by Stimulating Intercellular Adhesion in Macrophages. Front Immunol 9, 3118. 10.3389/fimmu.2018.03118. [PubMed: 30687322]
30. Thway K, Jones RL, Noujaim J, and Fisher C (2016). Epithelioid Sarcoma: Diagnostic Features and Genetics. Adv Anat Pathol 23, 41–49. 10.1097/PAP.0000000000000102. [PubMed: 26645461]
31. Paulino AC (2004). Synovial sarcoma prognostic factors and patterns of failure. Am J Clin Oncol 27, 122–127. 10.1097/01.coc.0000047130.91699.dc. [PubMed: 15057149]
32. Sannino G, Marchetto A, Kirchner T, and Grünewald TGP (2017). Epithelial-to-Mesenchymal and Mesenchymal-to-Epithelial Transition in Mesenchymal Tumors: A Paradox in Sarcomas? Cancer Research 77, 4556–4561. 10.1158/0008-5472.CAN-17-0032. [PubMed: 28811330]
33. Kalluri R, and Weinberg RA (2009). The basics of epithelial-mesenchymal transition. J Clin Invest 119, 1420–1428. 10.1172/JCI39104. [PubMed: 19487818]

34. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, and Tamayo P (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417–425. 10.1016/j.cels.2015.12.004. [PubMed: 26771021]
35. Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A 107, 15449–15454. 10.1073/pnas.1004900107. [PubMed: 20713713]
36. Foroutan M, Cursons J, Hediyeh-Zadeh S, Thompson EW, and Davis MJ (2017). A Transcriptional Program for Detecting TGFβ-Induced EMT in Cancer. Mol Cancer Res 15, 619–631. 10.1158/1541-7786.MCR-16-0313. [PubMed: 28119430]
37. Beauvais DM, and Rapraeger AC (2004). Syndecans in tumor cell adhesion and signaling. Reproductive Biology and Endocrinology 2, 3. 10.1186/1477-7827-2-3. [PubMed: 14711376]
38. Fujiwara T, Healey J, Ogura K, Yoshida A, Kondo H, Hata T, Kure M, Tazawa H, Nakata E, Kunisada T, et al. (2021). Role of Tumor-Associated Macrophages in Sarcomas. Cancers (Basel) 13, 1086. 10.3390/cancers13051086. [PubMed: 33802565]
39. Galon J, and Bruni D (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nature Reviews Drug Discovery 18, 197–218. 10.1038/s41573-018-0007-y. [PubMed: 30610226]
40. Aiello NM, Bajor DL, Norgard RJ, Sahmoud A, Bhagwat N, Pham MN, Cornish TC, Iacobuzio-Donahue CA, Vonderheide RH, and Stanger BZ (2016). Metastatic progression is associated with dynamic changes in the local microenvironment. Nat Commun 7, 12819. 10.1038/ncomms12819. [PubMed: 27628423]
41. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, Hodi FS, Martín-Algarra S, Mandal R, Sharfman WH, et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell 171, 934–949.e16. 10.1016/j.cell.2017.09.028. [PubMed: 29033130]
42. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, and Stratton MR (2015). Clock-like mutational processes in human somatic cells. Nat Genet 47, 1402–1407. 10.1038/ng.3441. [PubMed: 26551669]
43. Brennan MF, Antonescu CR, Moraco N, and Singer S (2014). Lessons learned from the study of 10,000 patients with soft tissue sarcoma. Ann Surg 260, 416–422. 10.1097/SLA.0000000000000869. [PubMed: 25115417]
44. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, Osokin N, Kozlov I, Frenkel F, Gancharova O, et al. (2021). Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell 39, 845–865.e7. 10.1016/j.ccell.2021.04.014. [PubMed: 34019806]
45. D'Angelo SP, Richards AL, Conley AP, Woo HJ, Dickson MA, Gounder M, Kelly C, Keohan ML, Movva S, Thornton K, et al. (2022). Pilot study of bempegaldesleukin in combination with nivolumab in patients with metastatic sarcoma. Nat Commun 13, 3477. 10.1038/s41467-022-30874-8. [PubMed: 35710741]
46. Kelly CM, Chi P, Dickson MA, Gounder MM, Keohan ML, Qin L-X, Adamson T, Condy MM, Biniakewitz M, Phelan H, et al. (2019). A phase II study of epacadostat and pembrolizumab in patients with advanced sarcoma. J Clin Oncol 37, 11049–11049. 10.1200/JCO.2019.37.15_suppl.11049.
47. Kelly CM, Antonescu CR, Bowler T, Munhoz R, Chi P, Dickson MA, Gounder MM, Keohan ML, Movva S, Dholakia R, et al. (2020). Objective Response Rate Among Patients With Locally Advanced or Metastatic Sarcoma Treated With Talimogene Laherparepvec in Combination With Pembrolizumab: A Phase 2 Clinical Trial. JAMA Oncol 6, 402–408. 10.1001/jamaoncol.2019.6152. [PubMed: 31971541]
48. Duan Z, and Luo Y (2021). Targeting macrophages in cancer immunotherapy. Sig Transduct Target Ther 6, 1–21. 10.1038/s41392-021-00506-6.
49. Gordon SR, Maute RL, Dulken BW, Hutter G, George BM, McCracken MN, Gupta R, Tsai JM, Sinha R, Corey D, et al. (2017). PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. Nature 545, 495–499. 10.1038/nature22396. [PubMed: 28514441]

50. Wei SC, Duffy CR, and Allison JP (2018). Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. Cancer Discovery 8, 1069–1086. 10.1158/2159-8290.CD-18-0367. [PubMed: 30115704]

51. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron CS, Robin C, and van Oudenaarden A (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods 14, 935–936. 10.1038/nmeth.4437. [PubMed: 28960196]

52. Faden DL, Ding F, Lin Y, Zhai S, Kuo F, Chan TA, Morris LG, and Ferris RL (2019). APOBEC mutagenesis is tightly linked to the immune landscape and immunotherapy biomarkers in head and neck squamous cell carcinoma. Oral Oncol 96, 140–147. 10.1016/j.oraloncology.2019.07.020. [PubMed: 31422205]

53. Glaser AP, Fantini D, Wang Y, Yu Y, Rimar KJ, Podojil JR, Miller SD, and Meeks JJ (2017). APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. Oncotarget 9, 4537–4548. 10.18632/oncotarget.23344. [PubMed: 29435122]

54. Beaubier N, Bontrager M, Huether R, Igartua C, Lau D, Tell R, Bobe AM, Bush S, Chang AL, Hoskinson DC, et al. (2019). Integrated genomic profiling expands clinical options for patients with cancer. Nat Biotechnol 37, 1351–1360. 10.1038/s41587-019-0259-z. [PubMed: 31570899]

55. Klement JD, Paschall AV, Redd PS, Ibrahim ML, Lu C, Yang D, Celis E, Abrams SI, Ozato K, and Liu K (2018). An osteopontin/CD44 immune checkpoint controls CD8+ T cell activation and tumor immune evasion. J Clin Invest 128, 5549–5560. 10.1172/JCI123360. [PubMed: 30395540]

56. Yang J, Du X, Wang G, Sun Y, Chen K, Zhu X, Lazar AJF, Hunt KK, Pollock RE, and Zhang W (2014). Mesenchymal to epithelial transition in sarcomas. Eur J Cancer 50, 593–601. 10.1016/j.ejca.2013.11.006. [PubMed: 24291235]

57. Tian W, Wang G, Yang J, Pan Y, and Ma Y (2013). Prognostic role of E-cadherin and Vimentin expression in various subtypes of soft tissue leiomyosarcomas. Med Oncol 30, 401. 10.1007/s12032-012-0401-y. [PubMed: 23292832]

58. Yao D, Dai C, and Peng S (2011). Mechanism of the mesenchymal-epithelial transition and its relationship with metastatic tumor formation. Mol Cancer Res 9, 1608–1620. 10.1158/1541-7786.MCR-10-0568. [PubMed: 21840933]

59. Horn LA, Fousek K, and Palena C (2020). Tumor Plasticity and Resistance to Immunotherapy. Trends in Cancer 6, 432–441. 10.1016/j.trecan.2020.02.001. [PubMed: 32348738]

Methods-only references

60. Tatlow PJ, and Piccolo SR (2016). A cloud-based workflow to quantify transcript-expression levels in public cancer compendia. Sci Rep 6, 39259. 10.1038/srep39259. [PubMed: 27982081]

61. Patro R, Duggal G, Love MI, Irizarry RA, and Kingsford C (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419. 10.1038/nmeth.4197. [PubMed: 28263959]

62. Hubbell E, Liu W-M, and Mei R (2002). Robust estimators for expression analysis. Bioinformatics 18, 1585–1592. 10.1093/bioinformatics/18.12.1585. [PubMed: 12490442]

63. Mankin HJ, Hornicek FJ, DeLaney TF, Harmon DC, and Schiller AL (2012). Pleomorphic spindle cell sarcoma (PSCS) formerly known as malignant fibrous histiocytoma (MFH): a complex malignant soft-tissue tumor. Musculoskelet Surg 96, 171–177. 10.1007/s12306-012-0225-0. [PubMed: 23129168]

64. Bui NQ, Nemat-Gorgani N, Subramanian A, Torres IA, Lohman M, Sears TJ, van de Rijn M, Charville GW, Becker H-C, Wang DS, et al. (2023). Monitoring Sarcoma Response to Immune Checkpoint Inhibition and Local Cryotherapy with Circulating Tumor DNA Analysis. Clin Cancer Res 29, 2612–2620. 10.1158/1078-0432.CCR-23-0250. [PubMed: 37130154]

65. Zhang Y, Parmigiani G, and Johnson WE (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom Bioinform 2, lqaa078. 10.1093/nargab/lqaa078.

66. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE, Koeppen H, Astarita JL, Cubas R, et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade

by contributing to exclusion of T cells. Nature 554, 544–548. 10.1038/nature25501. [PubMed: 29443960]

67. Liu D, Schilling B, Liu D, Sucker A, Livingstone E, Jerby-Arnon L, Zimmer L, Gutzmer R, Satzger I, Loquai C, et al. (2019). Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. Nat Med 25, 1916–1927. 10.1038/s41591-019-0654-5. [PubMed: 31792460]
68. Nathanson T, Ahuja A, Rubinsteyn A, Aksoy BA, Hellmann MD, Miao D, Van Allen E, Merghoub T, Wolchok JD, Snyder A, et al. (2017). Somatic Mutations and Neoepitope Homology in Melanomas Treated with CTLA-4 Blockade. Cancer Immunol Res 5, 84–91. 10.1158/2326-6066.CIR-16-0019. [PubMed: 27956380]
69. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM, et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 350, 207–211. 10.1126/science.aad0095. [PubMed: 26359337]
70. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, and Alizadeh AA (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457. 10.1038/nmeth.3337. [PubMed: 25822800]
71. Brunet J-P, Tamayo P, Golub TR, and Mesirov JP (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 101, 4164–4169. 10.1073/pnas.0308531101. [PubMed: 15016911]
72. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29. 10.1016/j.cell.2021.04.048. [PubMed: 34062119]
73. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344, 1396–1401. 10.1126/science.1254257. [PubMed: 24925914]
74. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A 102, 15545–15550. 10.1073/pnas.0506580102. [PubMed: 16199517]
75. Fantini D, Vidimar V, Yu Y, Condello S, and Meeks JJ (2020). MutSignatures: an R package for extraction and analysis of cancer mutational signatures. Sci Rep 10, 18217. 10.1038/s41598-020-75062-0. [PubMed: 33106540]
76. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, and Forbes SA (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nature Reviews Cancer 18, 696–705. 10.1038/s41568-018-0060-1. [PubMed: 30293088]
77. Moran PAP (1950). Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. 10.2307/2332142. [PubMed: 15420245]
78. Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Satagopam VP, Itoh M, Kawaji H, Carninci P, Rost B, et al. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. Nat Commun 6, 7866. 10.1038/ncomms8866. [PubMed: 26198319]
79. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, et al. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer 45, 228–247. 10.1016/j.ejca.2008.10.026. [PubMed: 19097774]

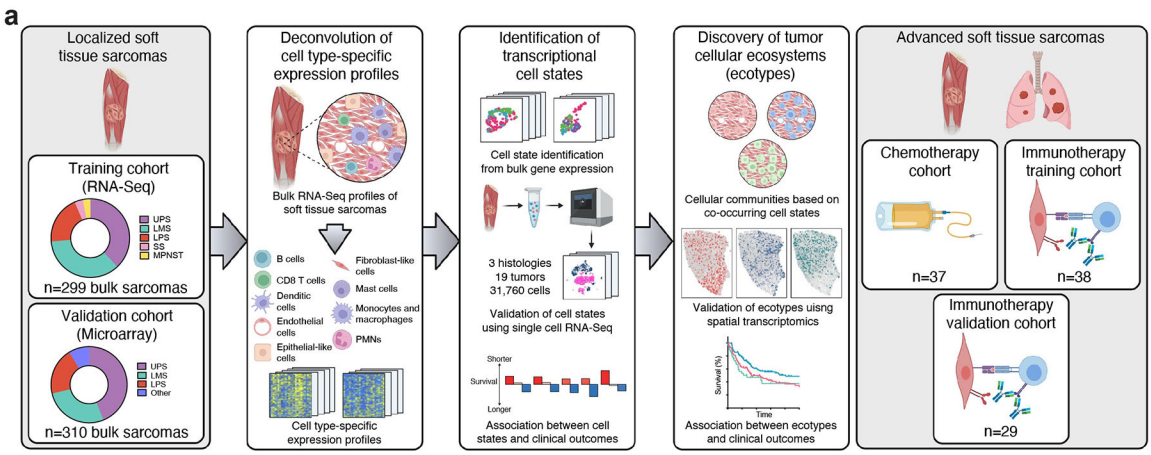


Fig. 1. A machine-learning framework for large-scale identification and validation of sarcoma cell states and ecosystems.

a, Schematic showing the implementation of EcoTyper in patients with soft tissue sarcoma. Cell states and ecotypes were initially discovered in a training cohort of patients with localized soft tissue sarcoma profiled with RNA-Seq followed by validation in a separate cohort of patients with localized soft tissue sarcoma with gene expression analyzed by microarray. Cell type-specific gene expression was purified from bulk transcriptomic data prior to identification of transcriptional states for each cell type. Cell states were validated via single-cell RNA-Seq, and the associations between cell state abundances and patient outcomes were analyzed. Sarcoma ecotypes were discovered by determining the co-occurrence patterns between cell states, and the spatial distribution of the ecotypes and cell states was validated using spatial transcriptomics. Finally, the association between sarcoma ecotypes and patient outcomes was analyzed. Three cohorts of patients with advanced soft tissue sarcoma were analyzed with bulk RNA-Seq, including two independent cohorts of patients treated with immune checkpoint inhibition and a cohort of patients treated with standard of care cytotoxic chemotherapy. See also Extended Data Figs. 1 and 8.

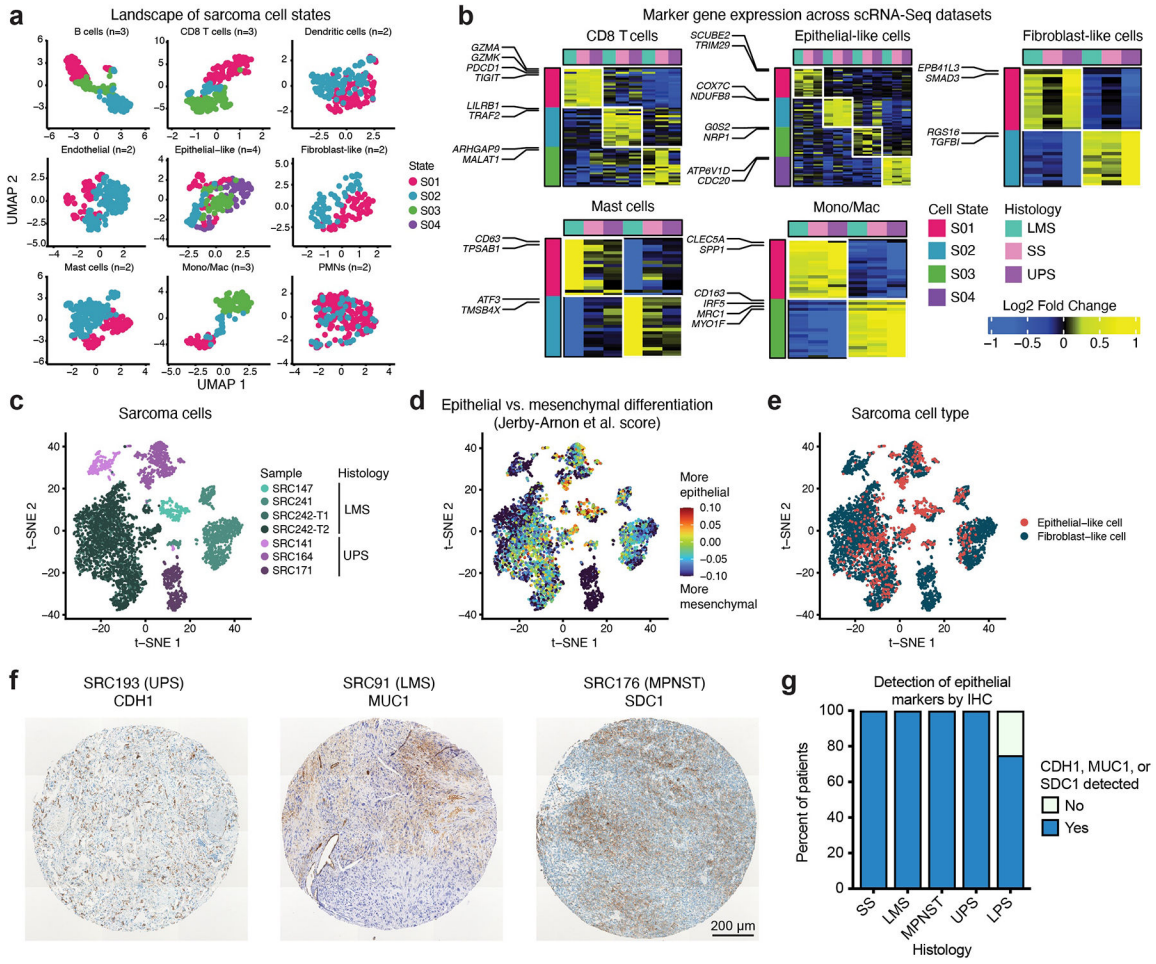


Fig. 2. Discovery and characterization of sarcoma-specific cell states.

a, UMAP plots of cell type-specific GEPs purified by CIBERSORTx from sarcomas profiled by bulk RNA-Seq in the training cohort. Each point represents one sample colored by the most abundant cell state. **b,** Heat maps showing expression of cell state marker genes (rows) across scRNA-Seq datasets spanning three sarcoma histologies (columns). Average $\log_2$ fold change is displayed for each cell state compared with the other cells of the same type. Only cell states with at least 15 assigned cells in each scRNA-Seq dataset are shown. See Supplementary Table 8. **c-e,** t-SNE plots of LMS and UPS malignant cell scRNA-Seq profiles colored by (**c**) tumor sample, (**d**) epithelial differentiation score measured by gene expression modules defined in synovial sarcomas[15], and (**e**) sarcoma cell type assignment based on the epithelial differentiation score. The epithelial differentiation score was calculated by subtracting the mesenchymal module score from the epithelial module score as described in the Methods. **f,** Representative sarcoma cores showing expression of epithelial markers CDH1, MUC1, and SDC1 by immunohistochemistry. A total of 128 cores were stained for each marker. **g,** Stacked bar plots displaying the percentage of patients across soft tissue sarcoma histologies with CDH1, MUC1, or SDC1 detected by immunohistochemistry. Synovial sarcoma = SS, leiomyosarcoma = LMS, malignant

peripheral nerve sheath tumor = MPNST, undifferentiated pleomorphic sarcoma = UPS, liposarcoma = LPS. See Supplementary Table 12.

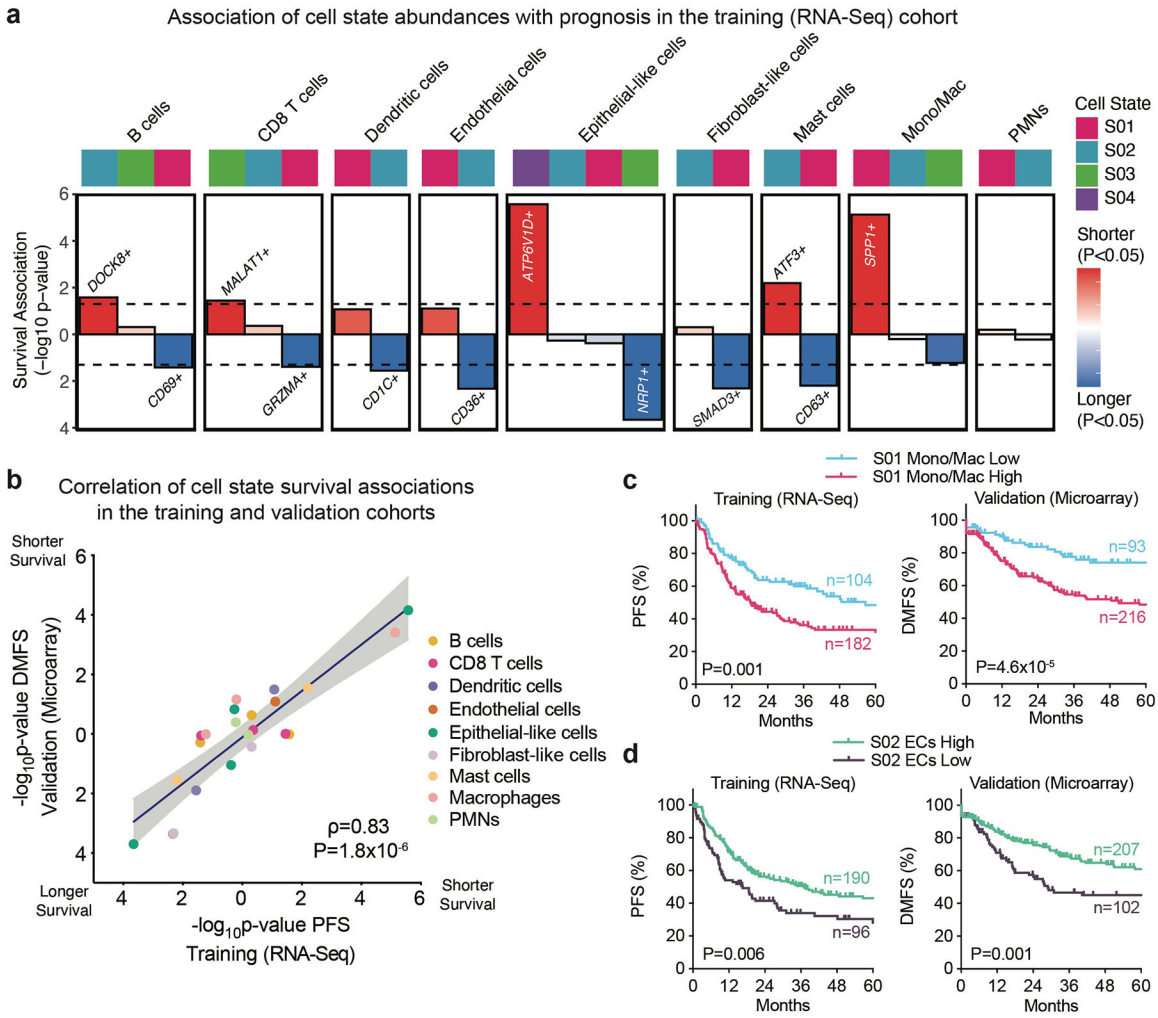


Fig. 3. Association of cell state abundances with patient outcomes across cohorts.

a, Association of sarcoma-specific cell states identified by EcoTyper with PFS in the training cohort. Significance was assessed using multivariable Cox proportional hazards models including cell state abundance as a continuous variable along with sarcoma histology. P values were calculated using two-sided Wald tests without correction for multiple comparisons. Marker genes are displayed for significantly associated cell states. **b,** Scatter plot showing the correlation between cell state survival associations in the training (RNA-Seq) cohort and the validation (Microarray) cohort. Survival associations are displayed as $-\log_{10}$p-values multiplied by 1 if the state is associated with shorter survival and −1 if associated with favorable outcomes. Spearman's correlation coefficient and two-sided P values are displayed. The line of best fit by linear regression and 95% confidence intervals of the line of best fit are shown on the graph. **c,d,** Kaplan-Meier plots showing survival outcomes stratified by the abundance of (**c**) S01 monocytes/macrophages or (**d**) S02 endothelial cells. The optimal threshold was defined in the training cohort and applied to the validation cohort. P values were calculated using two-sided log-rank tests. Patient survival, histologies, and cell state abundances used in the analysis of the training

and validation cohorts are shown in Supplementary Table 2 and Supplementary Table 3, respectively. Only patients with survival data available were included in the analysis.

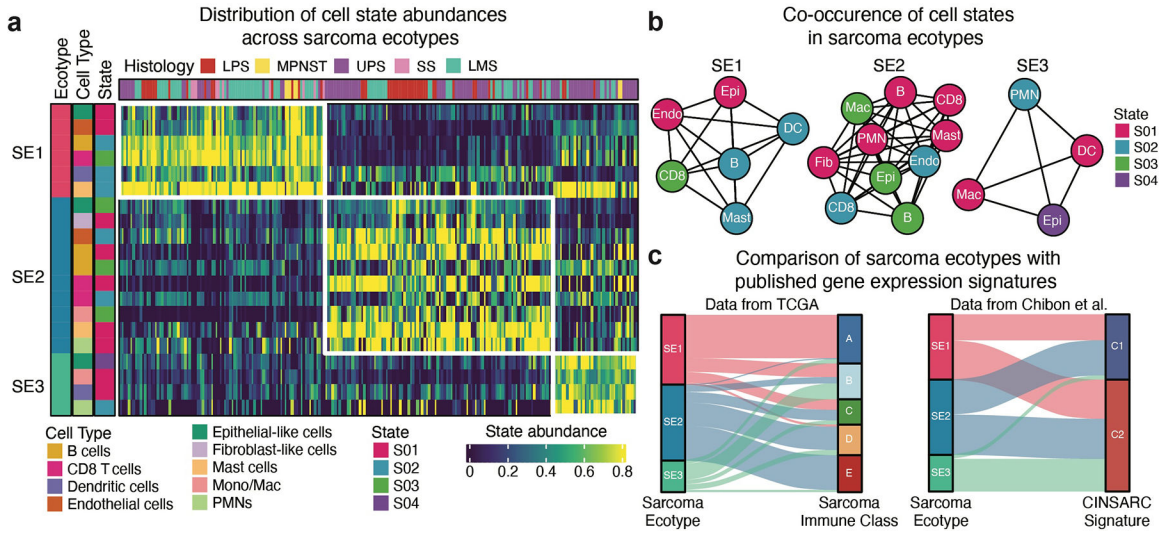


Fig. 4. Discovery of sarcoma multicellular communities.

a, Heatmap of cell state abundances across STSs of different histologies from the discovery cohort separated into three sarcoma ecotypes (SEs). Only cell states and tumor samples assigned to SEs are displayed. **b**, Network diagrams displaying SE composition. The length of each edge represents the Jaccard index across sarcomas from the training cohort. **c,** River plots showing the overlap between sarcoma ecotypes and sarcoma immune classes[18] and CINSARC signatures[8]. The thickness of each bar represents the percent of patients within each group.

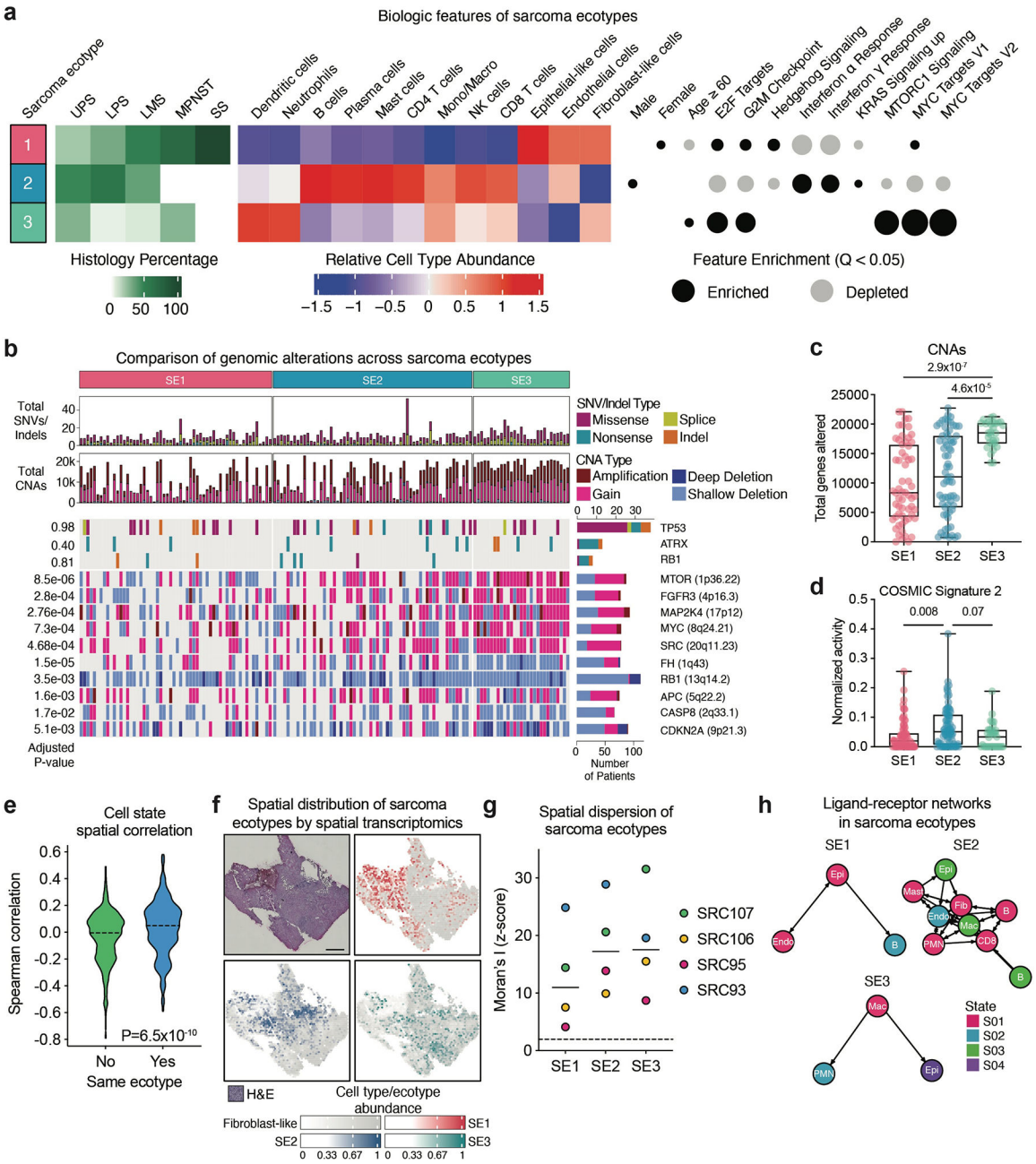


Fig. 5. Characterization of sarcoma ecotypes.

a, Plot of sarcoma ecotype biologic features in the training cohort. Left panel shows the percentage of each STS histology assigned to each sarcoma ecotype. Middle panel shows the relative abundance of 9 canonical cell types. Right panel shows significantly enriched patient features and hallmark gene sets. Additional features are shown in Supplementary Table 13. **b,** Plot of genomic alterations from TCGA by sarcoma ecotype assignment. Top two panels show total number of high impact SNVs/indels and total number of genes with a CNA. Bottom two panels display the genes most altered by SNVs/indels and genes enriched for CNAs in SE3. Adjusted P values were calculated using two-sided Fisher's exact tests corrected for multiple hypothesis testing. For CNAs, representative genes from

each genomic region are displayed. **c,d,** Box plots displaying (**c**) the total number of genes with CNAs (n=60 SE1, 67 SE2, and 29 SE3 samples) or (**d**) the normalized contribution of COSMIC mutational signature 2 by sarcoma ecotype (n=79 SE1, 78 SE2, and 33 SE3 sarcomas). P values were calculated using Kruskal-Wallis tests followed by Dunn's tests for pairwise comparisons. Boxes show median and quartiles, and whiskers extend to the minimum and maximum value. **e,** Violin plot comparing the Spearman spatial correlations between cell states in different ecotypes versus cell states within the same ecotype. P value was calculated using a two-sided Mann-Whitney U test. **f,** Distribution of sarcoma ecotypes in an STS (SRC93) profiled by spatial transcriptomics. H&E staining along with the abundance of SEs within each spatial transcriptomics spot are shown. Fibroblast-like cell abundance is plotted to show the tumor outline. Scale bar shows 1000 μm. A total of four sarcomas were profiled, and the other sarcomas are displayed in Extended Data Fig. 7c. **g,** Spatial aggregation of sarcoma ecotypes by spatial transcriptomics measured using Moran's I (n=4 samples). A z-score>1.96 is considered significantly more aggregated than expected by chance. **h,** Network diagrams displaying putative ligand-receptor interactions between cell states within each sarcoma ecotype. The arrows represent the direction of ligand to receptor signaling.

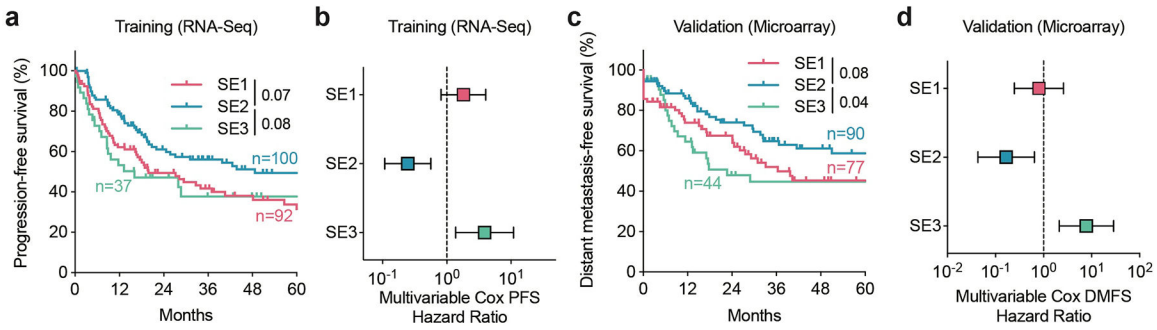


Fig. 6. Association of sarcoma ecotypes with patient outcomes.

a, Kaplan-Meier plot of progression-free survival in the training cohort stratified by sarcoma ecotype assignment. P values calculated using pairwise two-sided log-rank tests with correction for multiple hypothesis testing. **b,** Multivariable Cox proportional hazard ratios for progression-free survival in the training cohort based on sarcoma ecotype abundance including sarcoma histology as a variable (n=238 patients). Error bars display the 95% confidence interval. **c,** Kaplan-Meier plot of distant metastasis-free survival in the validation cohort stratified by sarcoma ecotype assignment. P values were calculated using pairwise two-sided log-rank tests with correction for multiple hypothesis testing. **d,** Multivariable Cox proportional hazard ratios for distant metastasis-free survival in the validation cohort based on sarcoma ecotype abundance including sarcoma histology as a variable (n=309 patients). Error bars display the 95% confidence interval. Patient survival, histologies, ecotype assignments, and ecotype abundances used in the analysis of the training and validation cohorts are shown in Supplementary Table 2 and Supplementary Table 3, respectively. For panels a and c, patients were analyzed based on ecotype assignment, and patients not assigned to an ecotype were not included. For panels b and d, patients were analyzed based on ecotype abundance, and all patients with survival data available were included in the analysis.

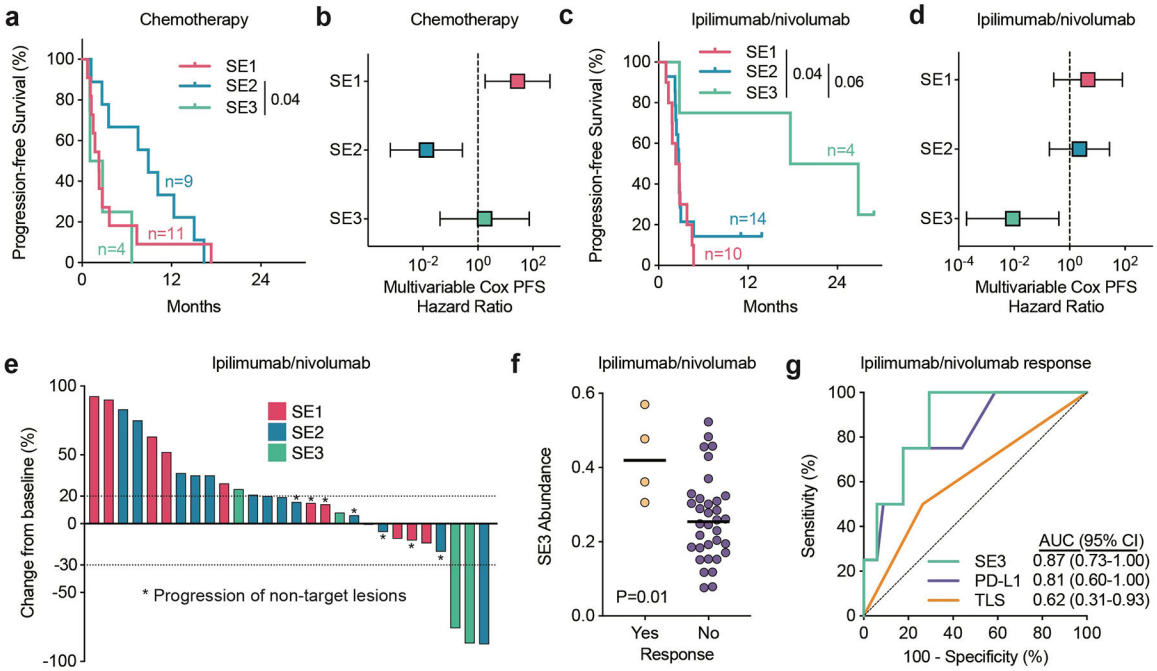


Fig. 7. Predicting STS response to immune checkpoint inhibition with sarcoma ecotypes.

a, Kaplan-Meier plot of progression-free survival in patients with advanced STSs treated with standard of care chemotherapy stratified by sarcoma ecotype assignment. P values calculated using pairwise two-sided log-rank tests with correction for multiple hypothesis testing. **b,** Multivariable Cox proportional hazard ratios for progression free survival in patients with advanced STSs treated with standard of care chemotherapy based on sarcoma ecotype abundance including sarcoma histology as a variable (n=37 patients). Error bars display the 95% confidence interval. **c,** Kaplan-Meier plot of progression-free survival in patients with advanced STSs treated with ipilimumab and nivolumab stratified by sarcoma ecotype assignment. P values were calculated using pairwise two-sided log-rank tests with correction for multiple hypothesis testing. **d,** Multivariable Cox proportional hazard ratios for progression-free survival in patients with advanced STSs treated with ipilimumab and nivolumab based on sarcoma ecotype abundance including sarcoma histology as a variable (n=38 patients). Error bars display the 95% confidence interval. **e,** Waterfall plot showing the best response by RECIST criteria for patients with advanced STSs treated with ipilimumab and nivolumab based on sarcoma ecotype assignment. Horizontal dotted lines represent the criteria for progressive disease (20% increase) and partial response (30% decrease). **f,** Plot of SE3 abundance in patients with and without a response to ipilimumab and nivolumab (n=4 responders and 34 non-responders). P value was calculated using a two-sided Mann-Whitney U-test. **g,** Receiver operating characteristic curves for prediction of response to ipilimumab and nivolumab by SE3 abundance, PD-L1 expression, and the presence of tertiary lymphoid structures (TLS). Area under the curve (AUC) and 95% confidence intervals (95% CI) are displayed on the graph. Patient survival, treatment response, histologies, ecotype assignments, and ecotype abundances used in the analysis of the chemotherapy and ipilimumab/nivolumab cohorts are shown in Supplementary Tables 18–20. For panels a, c, and e, patients were analyzed based on ecotype assignment, and

patients not assigned to an ecotype were not included. For panels b, d, f, and g, patients were analyzed based on ecotype abundance, and all patients were included in the analysis.

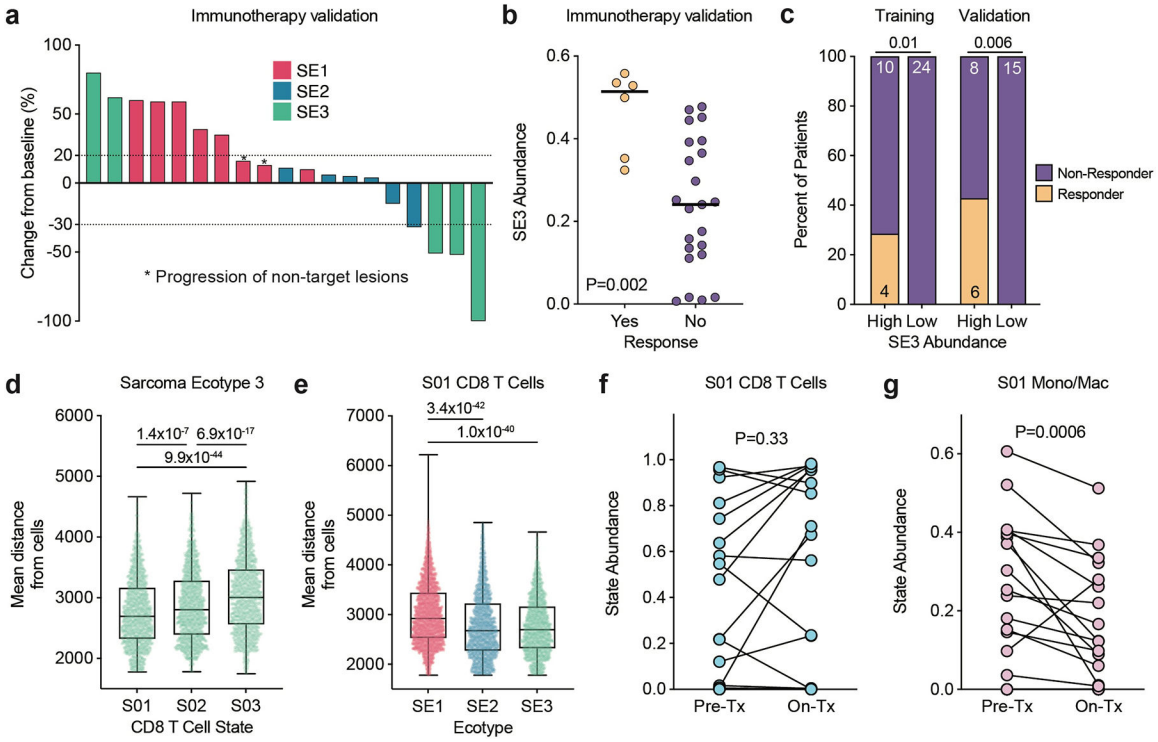


Fig. 8. Validation of SE3 as a predictor of response to ICI in soft tissue sarcomas.

a, Waterfall plot showing the best response by RECIST criteria measured by percentage change in the sum of target lesion diameters for patients with advanced STSs in the ICI validation cohort based on sarcoma ecotype assignment. Horizontal dotted lines represent the criteria for progressive disease (20% increase) and partial response (30% decrease). Only patients assigned to sarcoma ecotypes are displayed. **b,** Plot of SE3 abundance in patients with and without a response to ICI in the validation cohort (n=6 responders and 23 non-responders). **c,** Stacked bar plot demonstrating the percent of ICI responders in the training and validation cohorts stratified by high and low SE3 abundance. The optimal cutoff for SE3 abundance was defined in the training cohort and applied to the validation cohort. P values were calculated using two-sided Fisher's exact tests. **d,e,** Mean distance of (**d**) CD8 T cell states from SE3 spots (n=1963 spots) and (**e**) sarcoma ecotypes from S01 CD8 T cells (n=3352 SE1, 2061 SE2, and 1963 SE3 spots) in sarcomas profiled by spatial transcriptomics. P values were calculated using Kruskal-Wallis tests followed by Dunn's tests for pairwise comparisons. Boxes show median and quartiles, and whiskers extend to the minimum and maximum value. **f,g,** Pre-treatment (Pre-Tx) and on-treatment (On-Tx) abundance of (**f**) S01 CD8 T cells and (**g**) S01 monocytes/macrophages in the ICI validation cohort (n=19 paired samples). P values were calculated using two-sided Wilcoxon signed-rank tests. Patient treatment response, ecotype assignments, and ecotype abundances used in the ICI validation cohort analysis are shown in Supplementary Table 21. For panel a, patients were analyzed based on ecotype assignment, and patients not assigned to an ecotype were not included. For panels b and c, patients were analyzed based on ecotype abundance, and all patients were included in the analysis.