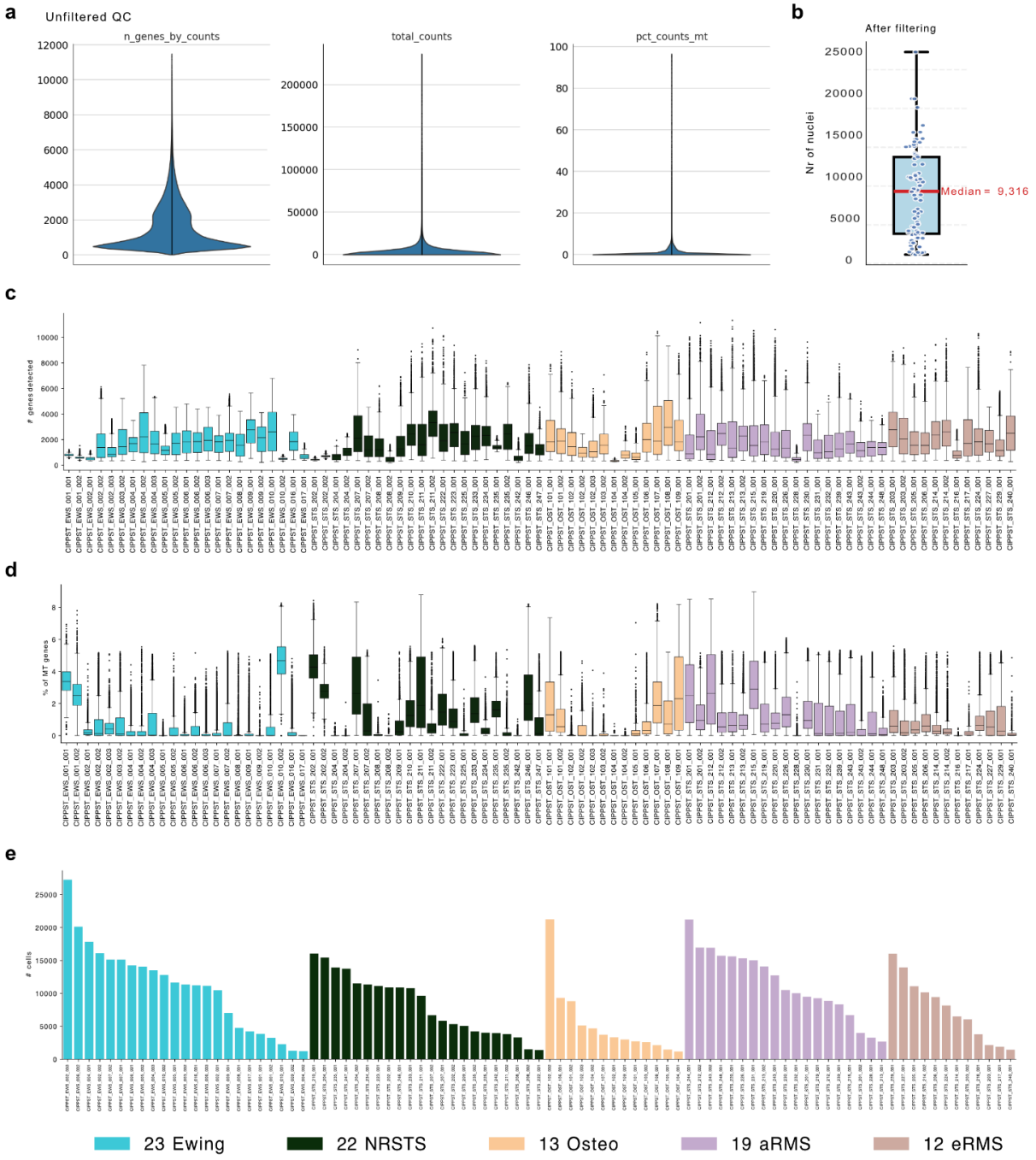
Supplementary Information

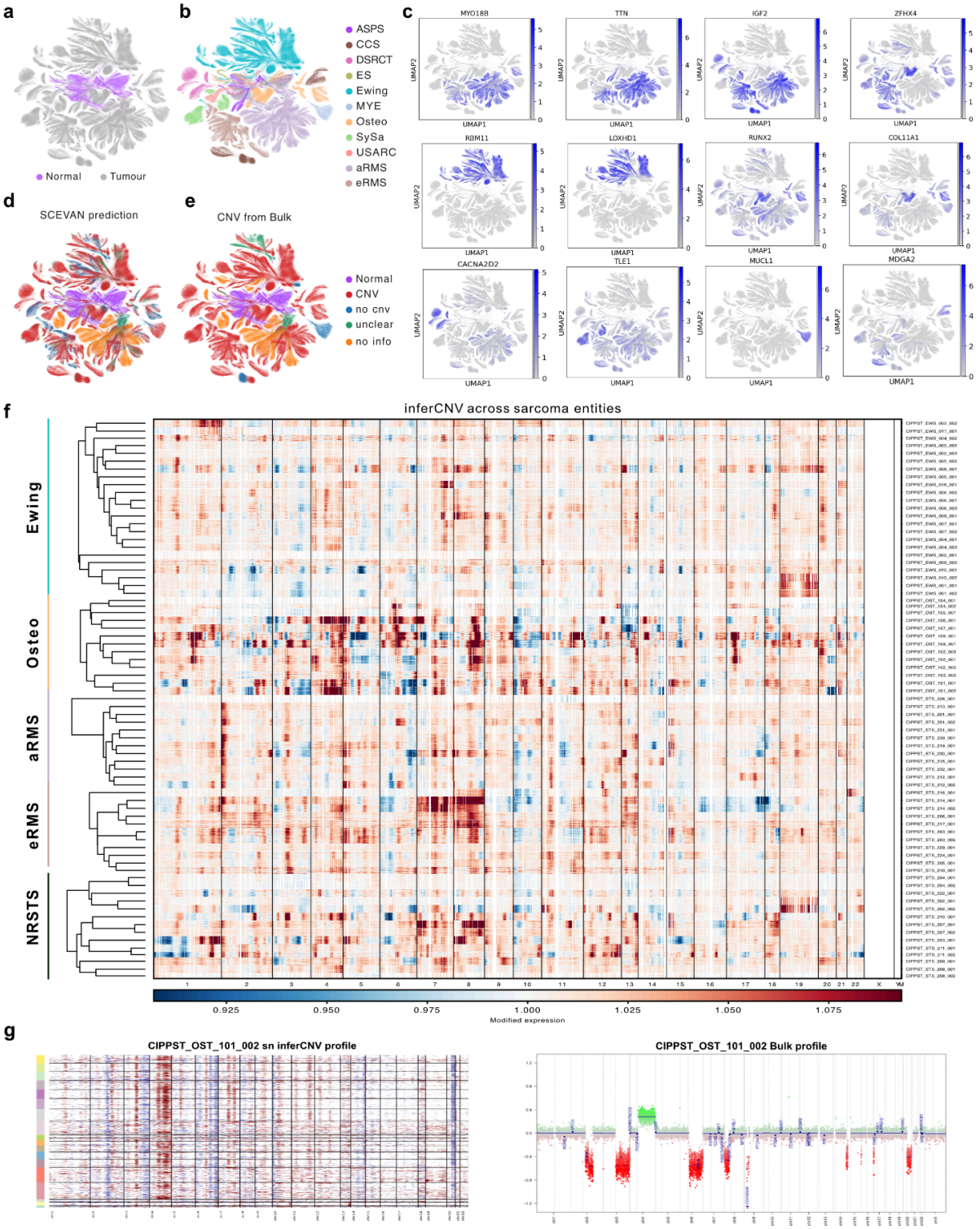
for the manuscript

A single-cell atlas of cancer-educated ecotypes across high-risk pediatric sarcomas

Karla Catacora Castañeda 1,2,3#, Sophia Scheuermann 1,3,4,5,6,7,8#, Tom T. Fischer 1,2,3,4#, Fabienne Engelmann 1,4,5, Anke King 1,2,3, Roland Imle 1,2,3,4,9, Alice Nuernbergk 5, Beate Kristmann 5, Jan-Philipp Mallm 10,11, Benjamin F.B. Mayer 12, Stefanie Volz 1,2,3, Iman Sadeghi 13, Kathrin Schramm 1,3,14, David T.W. Jones 1,3,14, Olaf Witt 1,3,4,15 , Robert J. Autry 1,2,3,4, Michaela Nathrath 16,17, Monika Sparber-Sauer 5,18, Uta Dirksen 19,20,21, Stefan M. Pfister 1,2,3,4, Jörg Fuchs 12, Supat Thongjuea 1,2,3, Thomas G. P. Grünewald 1,3,22,23#, Kristian W. Pajtler 1,2,3,4#*, Kendra K. Maass 1,2,3,4#*, Christian M. Seitz 1,3,4,5,7,15 #**

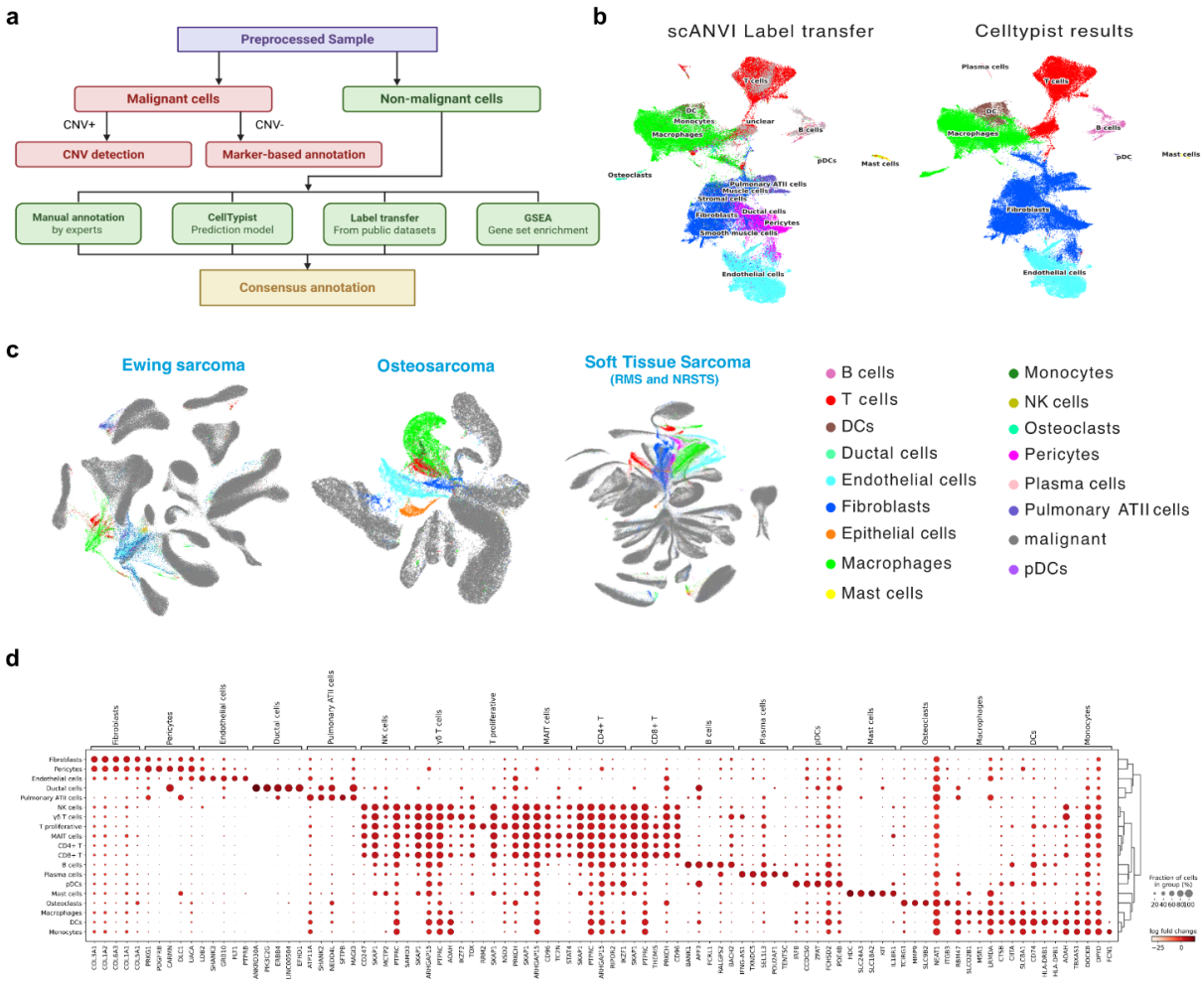


Supplementary Figure 1: Quality control and overview of the single-nucleus RNA sequencing samples. (**a**) Violin plots showing quality control metrics for the single-nucleus RNA sequencing dataset prior to filtering, including number of genes per nuclei with an expression higher than 0 (n_genes_by_counts), total reads per nuclei (total_counts), and percentage of mitochondrial reads (pct_counts_mt). (**b**) Boxplot showing the number of detected nuclei across samples retained after quality control filtering. (**c**) Boxplots showing the number of genes detected per sample after filtering. (**d**) Boxplots showing percentage of mitochondrial reads per sample after filtering. (**e**) Number of nuclei retained for each sample color coded for pediatric sarcoma entity after quality-control and filtering.

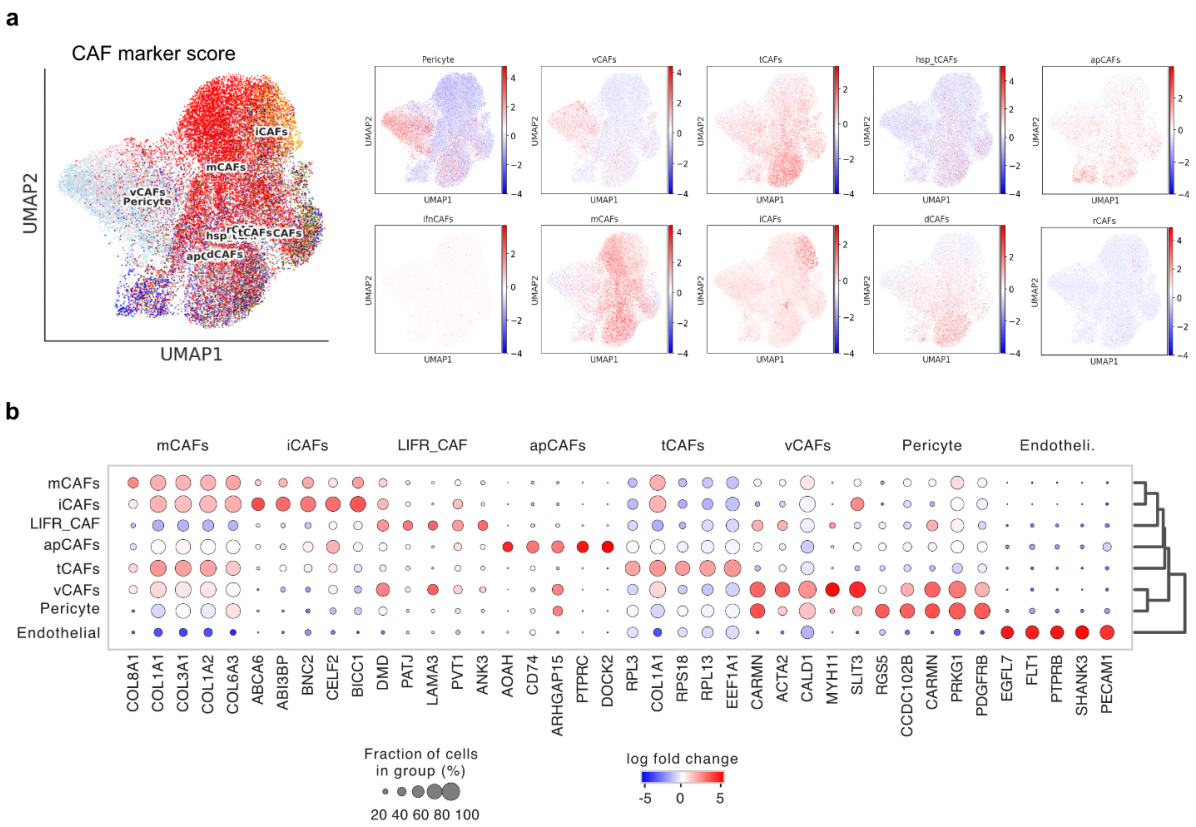


Supplementary Figure 2. Malignant cell type annotation and copy number inference across pediatric sarcoma snRNA cohort. (**a**) UMAP embedding of the integrated single-nuclei dataset colored by tumor and normal cell populations. (**b**) UMAP embedding colored by pediatric sarcoma entity. (**c**) UMAP feature plots showing the expression of canonical sarcoma marker genes used to identify distinct malignant cell populations. (**d**) UMAP highlighting malignant cells predicted by SCEVAN tool based on inferred copy number variation (CNV) profiles. (**e**) UMAP showing malignant cell annotations by matched bulk tissue CNV status characterized by DNA methylation array. (**f**) Chromosomal alteration heatmap estimated using inferCNV tool. T cells were used as a reference to infer patterns

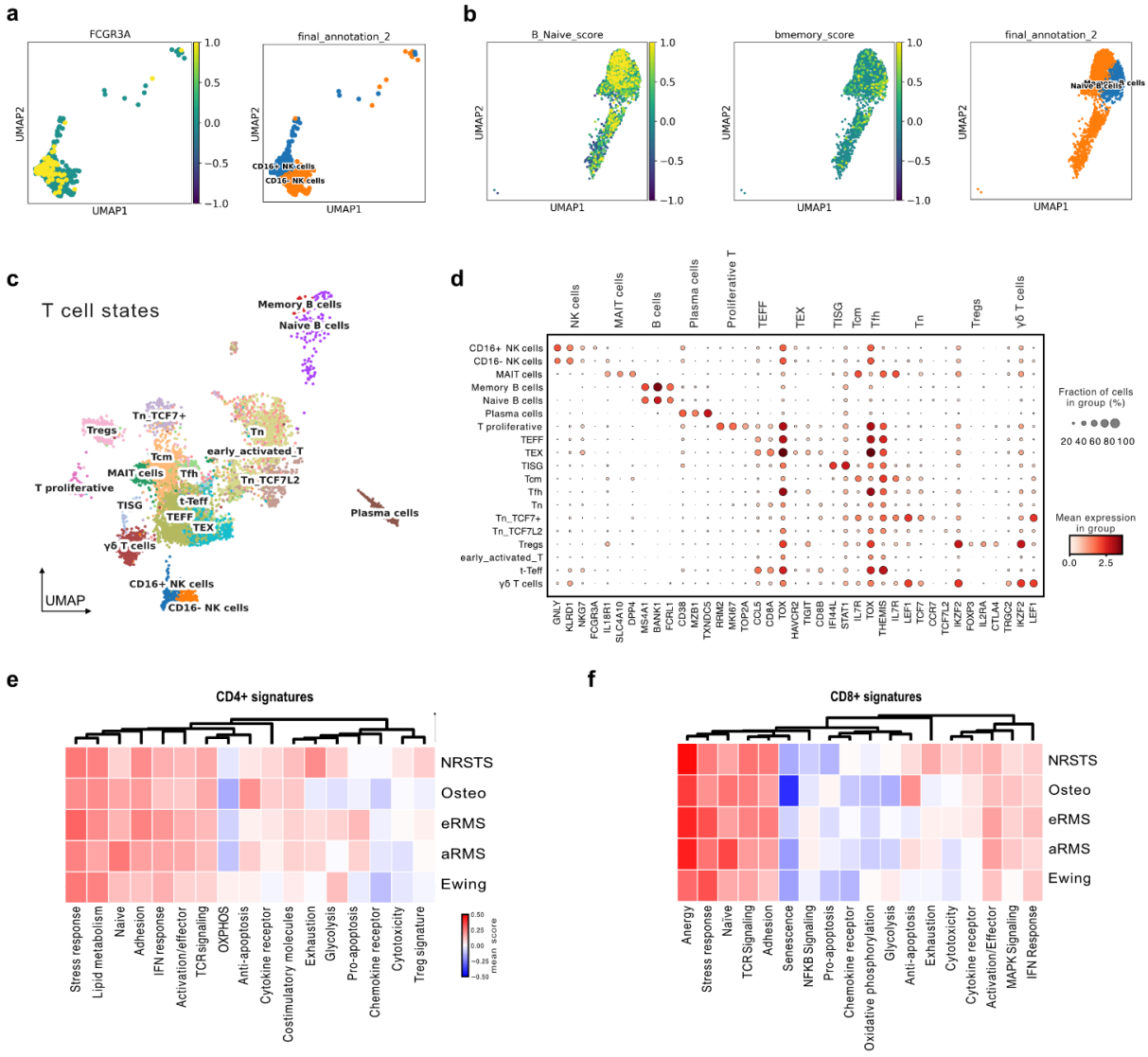
across malignant cells from all four pediatric sarcoma entities (*n*=73 samples). (**g**) Representative example of an osteosarcoma sample (CIPPST_OST_101_002) showing concordant CNV profiles in single-nuclei transcriptomic data (inferCNV) and matched bulk DNA methylation array data.



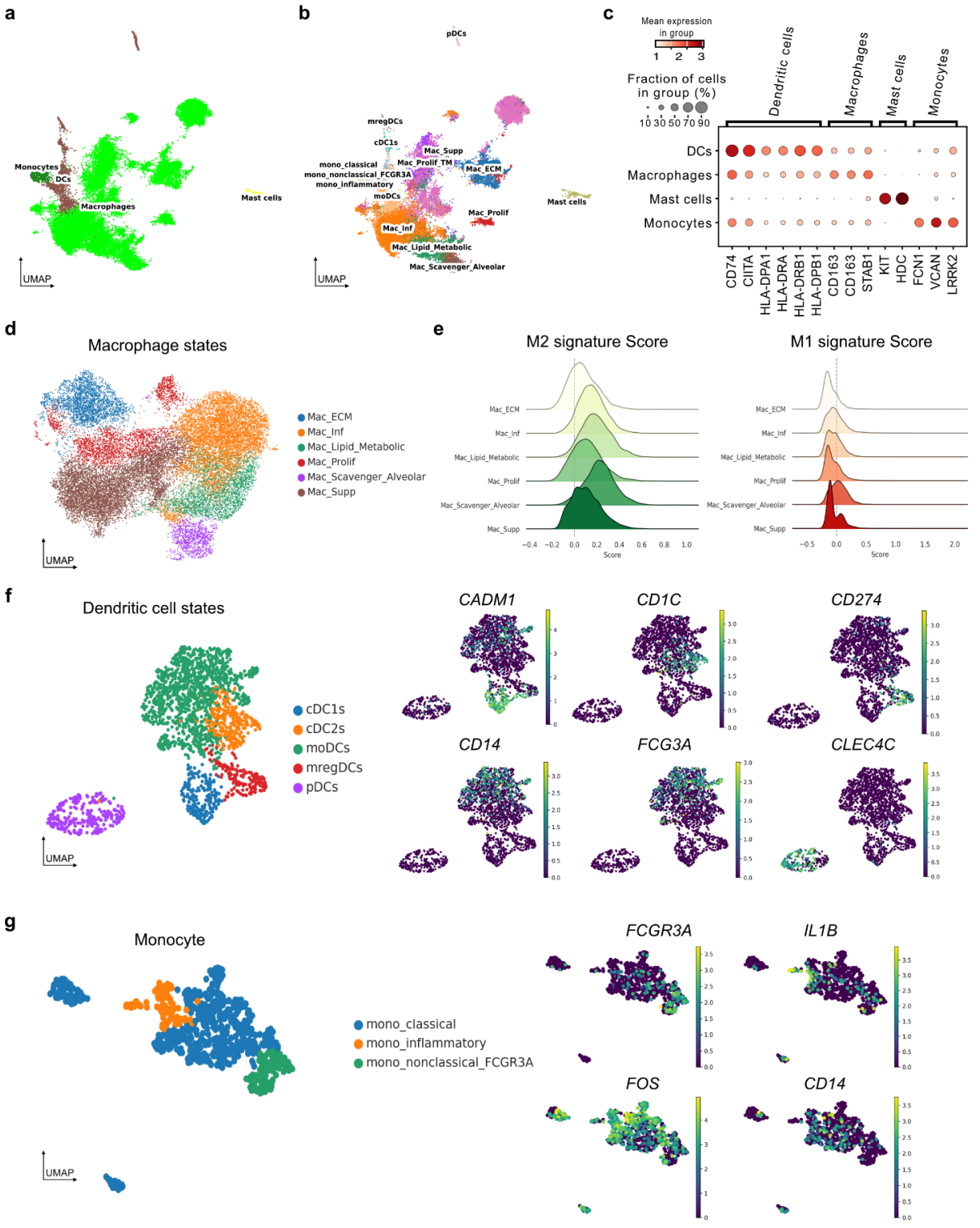
Supplementary Figure 3: Bioinformatic analysis workflow and cell type annotation of the non-malignant compartment. (a) Schematic overview of the bioinformatic workflow used for cell type annotation of the snRNA-seq dataset. **(b)** UMAP embeddings of the non-malignant compartment (*n*=94,082 nuclei) illustrating cell type annotation by reference-based label transfer using scANVI and cell type prediction using CellTypist. **(c)** UMAP embedding showing cell type annotation in bone and soft tissue sarcoma entities. **(d)** Dot plot showing differentially expressed marker genes for each annotated cell type.



Supplementary Figure 4: Stromal compartment state annotation and compositional distribution of CAF states. (**a**) UMAP embedding of all stromal cells, showing the predicted cell state annotation based on CAF marker scores (left), and the score distribution for each individual subpopulation projected across the UMAP (right). (**b**) Dot plot showing differentially expressed marker genes across the identified stromal cell populations. Dot size represents the fraction of cells within a group expressing the respective gene and color encodes the log-fold change in expression.

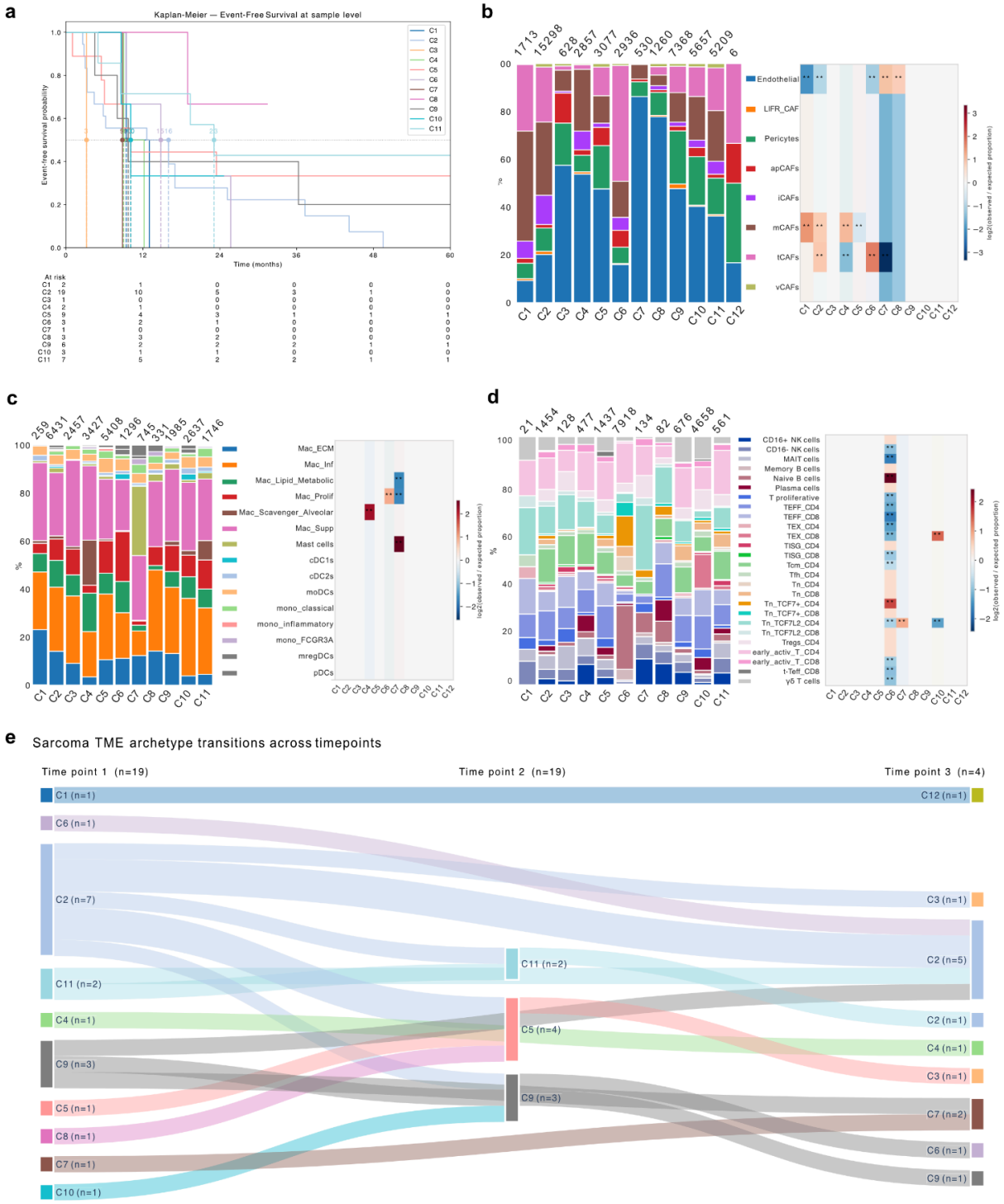


Supplementary Figure 5: Characterization of lymphoid cell types and states across the pediatric sarcoma. (**a**) Feature plot showing expression of *FCGR3A* (*CD16*), highlighting the CD16+ natural killer (NK) cell population. (**b**) Annotation of B cell states based on gene signature scores for memory B cells (*CD80*, *CD86*, *FAS)* and naïve B cells (*CD53*, *IGHD*, *FCER2*). (**c**) UMAP embedding of lymphoid cells colored by cell states. (**d**) Dot plot showing marker genes for definition of lymphoid cell states. (**e**) Heatmap showing CD4+ T cell states signature scores across sarcoma entities. (**f**) Heatmap showing CD8+ T cell states signature scores for across sarcoma entities.



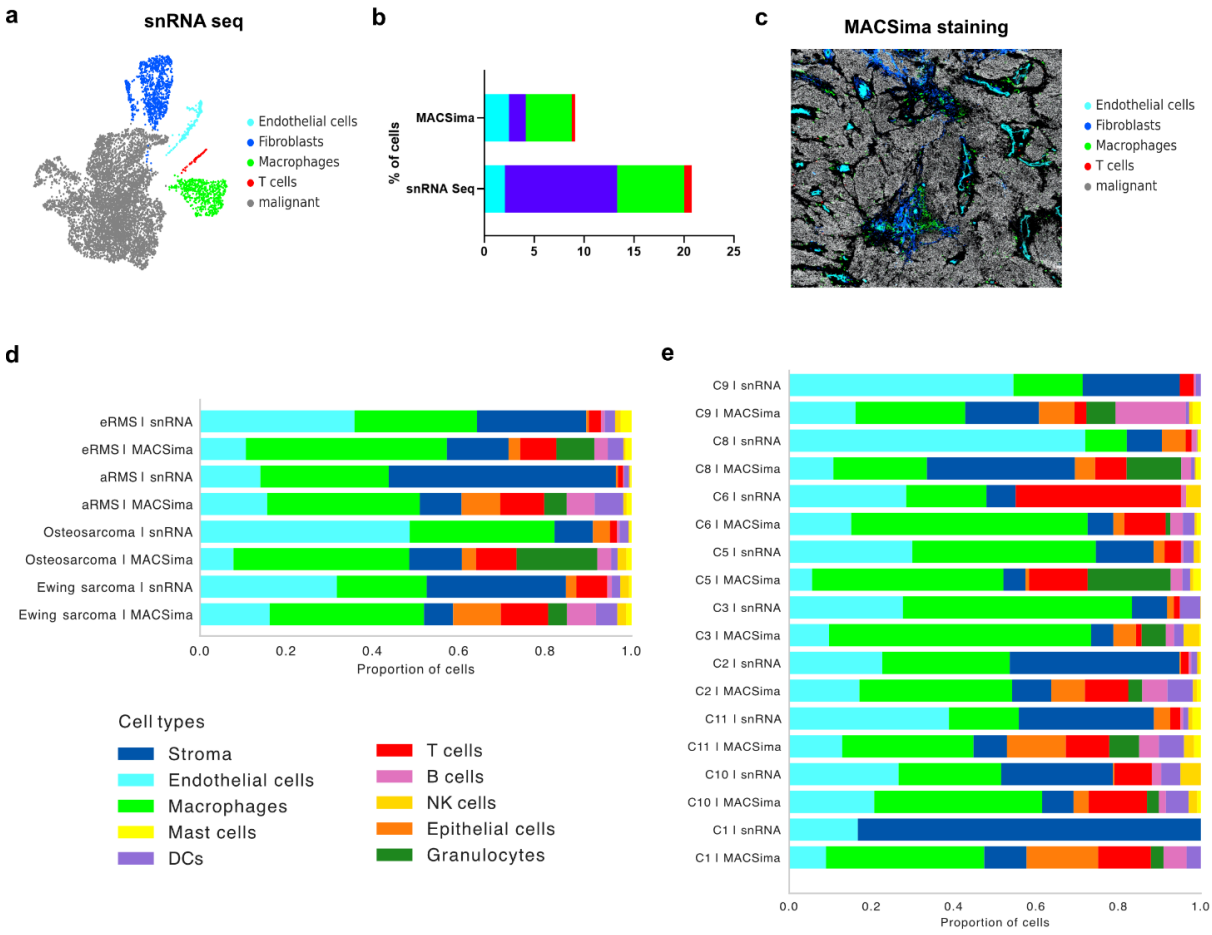
Supplementary Figure 6: Myeloid cell state annotation. (**a**) UMAP embedding of all myeloid cells coloured by broad lineage annotation, identifying monocytes, DCs, and macrophages based on canonical marker genes. (**b**) UMAP embedding of myeloid cells colored by detailed cell state annotation. (**c**) Dot plot showing canonical marker genes used to define major myeloid cell types. Dot size indicates the fraction of cells per group expressing the gene; colour indicates the log-fold change in expression. (**d**) UMAP embedding of macrophages coloured by annotated macrophage state, defined as in (b). (**e**)

showing M2 (left) and M1 (right) polarization signature scores across macrophage states. Signatures do not separate the annotated states, with the M2 score showing a slight positive trend throughout. (**f**) UMAP embedding of dendritic cells coloured by annotated DC state (left), alongside feature plots showing expression of the canonical DC subset markers used for annotation (right). (g) Monocytes subclustered into three functional states with adjacent feature plots showing expression of canonical monocyte subset markers.



Supplementary Figure 7. Cell-state composition, survival, and longitudinal transitions across tumor microenvironment archetypes. (**a**) Kaplan-Meier plot showing post-sampling EFS for an exploratory tumor episode-indexed analysis, evaluating time from tumor sampling to the next documented event. Numbers denote tumor episodes; for patients with serially profiled samples, each episode was included based on their respective archetype classification. (**b-d**) Relative composition of stromal (**b**), lymphoid (**c**), and myeloid (**d**) cell states across all twelve TME archetypes and corresponding enrichment heatmap showing scCODA log2-fold changes. Cell types with a statistically credible effect at the

specified false discovery rate (FDR) are marked with **. (**e**) Sankey plot showing archetypes across different disease episodes over time in patients with longitudinal sample pairs.



Supplementary Figure 8. Comparison of representative complementary MACSima analyses with snRNA-seq data. (**a**) UMAP of snRNA-seq data from a representative aRMS sample (patient ID CIPPST_STS_239_001), colored by cell type and showing the major non-malignant cell populations. (**b**) Descriptive quantification of the percentage of these cell types as detected by MICS versus snRNA-seq. (**c**) Representative MACSima multiplex immunofluorescence image of the corresponding sample illustrating spatial distribution of annotated cell populations within the TME. (**d-e**) Stacked bar plots showing relative proportions of cell types identified by snRNA-seq and MICS, stratified by (**d**) tumor entity (eRMS; aRMS; osteosarcoma; Ewing sarcoma), and (**e**) TME archetype (as defined in Fig. 6).

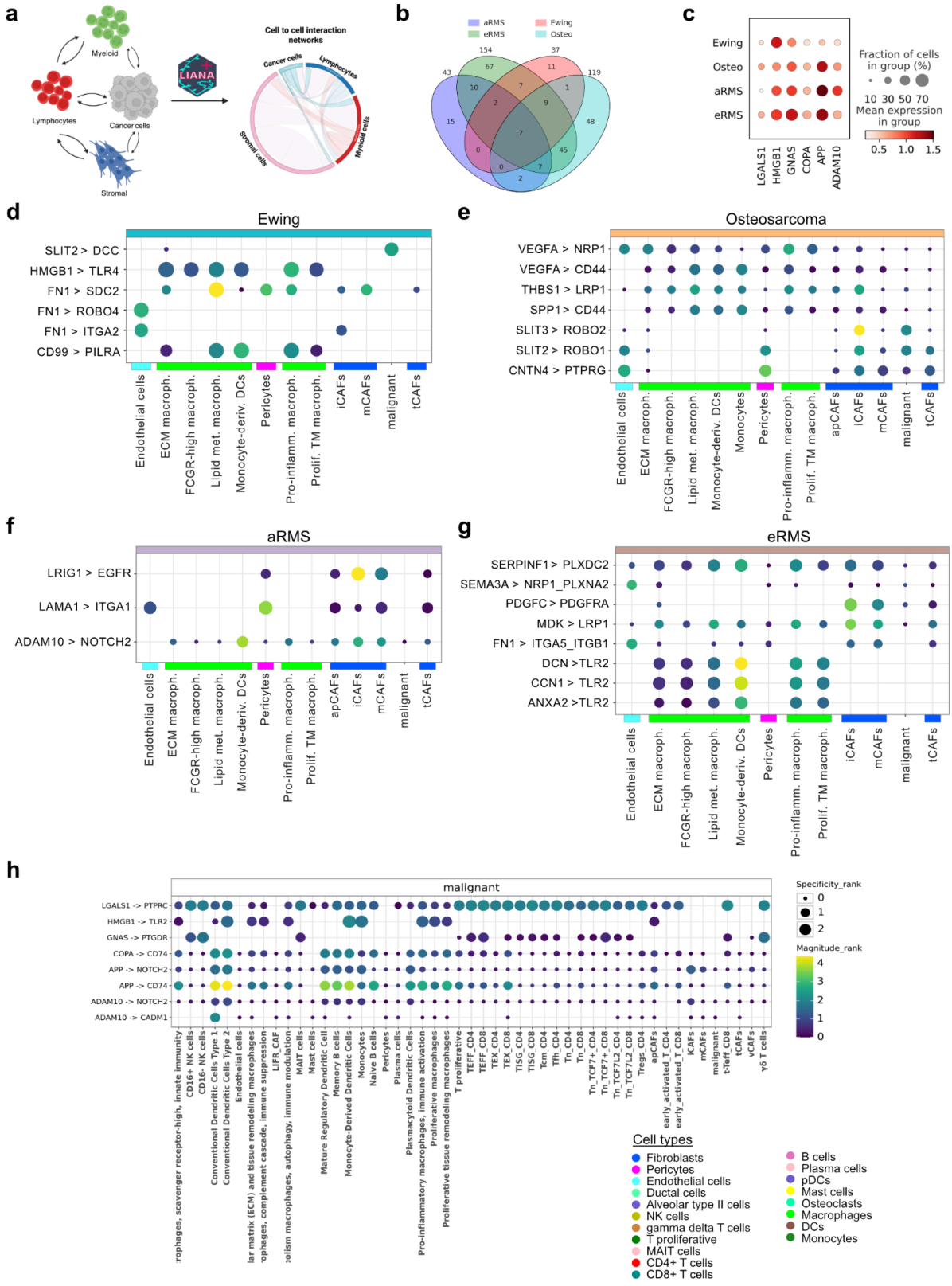


Figure 9. Shared and unique ligand-receptor interactions across sarcomas. (**a**) Schematic of the analysis workflow using LIANA to infer ligand - receptor interactions. (**b**) Overlap of inferred interactions across aRMS, eRMS, Ewing sarcoma, and osteosarcoma. (**c**) Dot plot showing expression of selected ligands across tumor types. (**d–g**) LIANA dot plots of entity-specific ligand-receptor interactions between malignant cells

and the TME in (**d**) Ewing sarcoma, (**e**) osteosarcoma, (**f**) aRMS, and (**g**) eRMS. (**h**) Ligand–receptor interactions from malignant cells to non-malignant cell types shared by all four sarcoma entities (dot size, specificity rank; colour, magnitude rank).

Figure 10. Longitudinal transitions between sarcoma interaction phenotypes. (**a**) Sankey (alluvial) plot showing transitions of samples from distinct interaction phenotypes across sequential timepoints for patients with longitudinal samples. Bar height at each timepoint indicates the number of patients assigned to each interaction group; flow width represents the number of patients transitioning between groups from one timepoint to the next. (**b**) UMAP embeddings of all samples, with lines connecting longitudinal sample pairs from individual patients that cluster with interaction group I0, I1, I2, or I3, respectively, at a later timepoint. (**c**) UMAPs stratified by tumour type (Ewing sarcoma, osteosarcoma, and soft-tissue sarcoma) and coloured by interaction group assignment. Arrows connect longitudinal sample pairs from the same patient, illustrating interaction group transitions over time within each tumour type.