Title: Single cell RNA-sequencing of Ewing sarcoma tumors demonstrates transcriptional heterogeneity and clonal evolution.

Authors: Andrew Goodspeed[1,2], Avery Bodlak[3], Sarah Nelson-Taylor[3], Naoki Oike[3], Timothy Porfilio[3], Ryota Shirai[3], Deandra Walker[3], Amy Treece[4], Jennifer Black[5], Nathan Donaldson[6], Carrye Cost[3], Tim Garrington[3], Brian Greffe[3], Sandra Luna-Fineman[3], Jenna Demedis[3], Jessica Lake[3], Etienne Danis[1,2], Michael Verneris[3], Masanori Hayashi[3,7].

Affiliations:

[1]Biostatistics and Bioinformatics Shared Resource, University of Colorado Cancer Center, Aurora, Colorado

[2]Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, Colorado

[3]Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado

[4]Department of Pathology and Laboratory Medicine, Children's of Alabama, Birmingham, AL

[5]Department of Pathology, University of Colorado School of Medicine, Aurora, Colorado

[6]Department of Orthopedics, University of Colorado School of Medicine, Aurora, Colorado

[7]Corresponding Author

Masanori Hayashi, MD

Research Complex 1 North, Room 4402K

12800 E 19th Ave, Mail Stop 8302, Aurora, CO 80045

Telephone: 303-724-7447

Fax: 303-724-3363

Email: masanori.hayashi@cuanschutz.edu

The authors have declared that no conflict of interest exists.

Abstract: Ewing sarcoma is the second most common bone cancer in children, accounting for 2% of pediatric cancer diagnoses. Patients who present with metastatic disease at the time of diagnosis have a dismal prognosis, compared to the >70% 5-year survival of those with localized disease. Here, we utilized single cell RNA-sequencing to

characterize the transcriptional landscape of primary Ewing sarcoma tumors and surrounding tumor microenvironment (TME). Copy-number analysis identified subclonal evolution within patients even prior to treatment. Primary tumor samples demonstrate a heterogenous transcriptional landscape with several conserved gene expression programs, including those composed of genes related to proliferation and EWS targets. We also were able to identify the composition of the TME and molecularly dissect the transcriptional profile of circulating tumor cells in peripheral blood at the time of diagnosis.

[Main Text:]

Introduction

While the intensification of conventional multi-modal therapy has significantly improved the survival rate for localized Ewing sarcoma from less than 10% in the surgery-only era to approximately 70%, many patients still die from metastatic disease[1, 2, 3]. With intensive multi-modal therapy using chemotherapy, radiation, and surgery, many patients, even those with initial metastatic disease achieve a radiographic remission; however, metastatic relapse after a remission frequently occurs. Accordingly, an initial good response to conventional therapy does not guarantee a cure.

While the Ewing sarcoma genomic landscape involves few mutations, the pathognomonic t(11;22) EWS/Fli1 translocation drives 90% of the disease[4, 5, 6, 7]. By studying EWS/FLI1 transcriptional regulation, over 1,000 genes have been identified to be directly or indirectly induced or repressed[8, 9, 10, 11]. While these results have spurred a significant effort focused on EWS/Fli1-induced genes that might be targeted to inhibit tumor progression, clinical translation has been difficult[12, 13]. More recent reports have indicated the presence of transcriptional heterogeneity in cell lines and xenografts[14, 15, 16]. While EWS/Fli1 expression is inarguably essential to the survival of Ewing sarcoma cells, emerging experimental and clinical evidence suggests that Ewing sarcoma tumors have significant intra-tumoral EWS/Fli1 transcriptional heterogeneity[15, 16]. Additionally, conditional knockdown systems of EWS/Fli1 have indicated that low-level EWS/Fli1 promotes a pro-metastasis cell phenotype, compared to high expression[15]. A major unmet need is to better understand tumor heterogeneity and how it potentially contributes to poor prognosis, with the goal of identifying therapeutic targets that are essential to metastasis and recurrence.

The understanding of the Ewing sarcoma Tumor microenvironment (TME) is still evolving. Previous work in this cancer type has suggested the TME is infiltrated by immunosuppressive myeloid cells, functionally impaired antigen presenting cells, especially conventional dendritic cells, and T cells with features suggestive of dysfunction[17, 18]. Ewing sarcoma tumor cells have been shown to lack expression of HLA class I molecules, potentially leading to decreased T cell recognition [17, 19]. Furthermore, T cells in Ewing sarcoma cells have been indicated to have high LGALS3 and LAG3 expression, suggesting dysfunction[17, 18]. Tumor-TME interaction has been

suggested to involve Wnt signaling[20], USP6[21], and VEGF[22, 23] but whether these interactions are present beyond cell line and mouse PDX models is unclear.

Despite these advances in the investigation of Ewing sarcoma tumor biology, clinical support and translation has been poor. For instance, the knowledge of intra-tumoral transcriptional activity in untreated patient tumors is limited, and we know very little about how this may drive treatment resistance or clonal evolution. Here, we used single cell RNA-sequencing (scRNA-seq) to investigate the transcriptional heterogeneity of untreated Ewing sarcoma patient tumors, as well as circulating tumor cells (CTCs). We found that patient tumors contain heterogenous clusters of tumor cells with distinct genetic expression programs, which correlate with patient outcomes. Evaluation of the primary tumors, the TME, and the CTCs revealed several potential therapeutic targets.

Results

scRNA-seq reveals transcriptional heterogeneity in Ewing Sarcoma primary tumors

To characterize the intratumoral landscape of Ewing sarcoma, we performed droplet-based scRNA-seq on biopsies of primary tumor samples from seven patients with newly diagnosed Ewing Sarcoma (Supplementary Fig. 1A). All patients had a confirmed EWS-FLI1 or EWS-ERG translocation, with karyotypes available (Table 1). All samples were from bone primary tumors and were obtained prior to any systemic therapy. One patient presented with metastasis, while six patients presented with localized disease. After doublet removal and filtering of low-quality cells, the scRNA-seq dataset consisted of 6228 cells with 98936678 Unique Molecular Identifiers (UMIs) across 22518 genes (Supplementary Fig. 1B-C, Supplementary Table 1).

Single cells were visualized using uniform manifold approximation and projection (UMAP) followed by unbiased clustering (Supplementary Fig. 2A). We classified these clusters into nine major cell types (Fig. 1A-B, Supplementary Fig. 2B, Supplementary Table 2). In addition to Ewing sarcoma, we found the tumor microenvironment (TME) to consist of myeloid cells, endothelial cells, T cells, B cells, osteoclasts, and two separate cancer associated fibroblasts (CAFs): matrix, and inflammatory CAFs (Fig. 1C-D). Many of these cell types were present in all tumor samples, although inflammatory CAFs were only observed in patient 001 (Fig. 1B, Supplementary Fig. 2C-D).

The Ewing sarcoma cells were identified through several means. These cells expressed high levels of canonical Ewing sarcoma-associated genes such as such as MAPT[24], KDSR[25], NKX2-2[26, 27], NR0B1[33, 28, 29], STEAP1[30, 31], and PRKCB[32] (Fig 1C-D, Supplementary Fig. 2E). The classification of these cells as malignant was further supported by single cell variational aneuploidy analysis (SCEVAN) (Supplementary Fig. 2F) and by the enrichment of several EWS-FLI1 signature gene sets (Supplementary Fig. 2G). Among the Ewing sarcoma cells, a proliferating population was identified by the expression of MKI67 and PCNA (Fig. 1C-D).

A benefit of scRNA-seq is the ability to characterize several forms of tumor heterogeneity. Here, we utilized inferCNV to infer genome-wide copy number alterations (CNAs) across the seven Ewing sarcoma tumors[33] (Fig. 1E, Supplementary Fig. 2H).

Previous Ewing sarcoma studies have identified several common CNAs, including gains in chr 1q chr 2, chr 5, chr 7, chr 8, chr 12, chr 20, and loss of chr 9p, chr 16q[6, 34, 35, 36, 37, 38, 39]. In our dataset, the Ewing sarcoma cells have several of these alterations such as gains of chr1q, chr2, chr5, chr12, and chr20. All primary tumors had conventional karyotyping performed (Table 1) which closely corroborated with the inferred CNAs from the scRNA-seq.

Finally, we explored the expression of previously reported potential therapeutic targets for Ewing sarcoma in our data set (Fig. 1F). Several preclinical and clinical targets have been identified for Ewing sarcoma with varying levels of success. TME-, cell surface-, and cell intrinsic signaling-related targets were plotted across cell type as well as within tumor cells by sample. We found that compared to other cell types in the primary tumor, Ewing cells across all samples express moderate to high levels of PAPPA but low levels of most HLA genes. For immunotherapeutic targets, Ewing sarcoma cells uniquely express higher levels of ADGRG2, IL1RAP, and LINGO1. CD99 and STEAP1 is also highly expressed in Ewing sarcoma cells, although there is significant expression in CAFs as well. While mRNA expression and cell surface expression does not always correlate[40], it is quite notable that the expression of many of these genes is heterogenous across tumor samples, especially for IL1RAP and LINGO1. We found that many of the potential cell signaling-related therapeutic targets are expressed heterogeneously within the tumors, and across patient samples, including CDK4, PARP1, and DDX3X. While RNA expression of these targets is not an ideal readout for activity or dependency, our data confirms what has been widely observed in the clinic; despite the genetic makeup that suggests otherwise, Ewing sarcoma does not seem to have a single dependency that can be targeted with a single agent.

Ewing sarcoma tumors demonstrate transcriptional heterogeneity

Next, we sought to define the transcriptional heterogeneity of Ewing sarcoma tumors. The tumor cells identified in Fig. 1 were isolated, integrated across patients, and then we performed non-negative matrix factorization (NMF) analysis in a sample-specific manner to identify intratumor gene expression programs (GEPs) driving the heterogeneity in tumor cells. This revealed a total of 25 GEPs (Supplementary Table 3), which were grouped into six main meta-programs through hierarchical clustering (Fig. 2A, Supplementary Table 4). Most of the expression programs were represented in multiple tumors, suggesting that they represent diverse functions shared among patients. Next, we performed over-representation analysis to identify characteristics of each identified meta-programs (Fig. 2B). Based on the gene expression and gene set enrichment within each merged GEP, we termed the 6 merged GEPs MAPK, EWS, Proliferation, UPR (unfolded protein response), Ribosome, and OxPhos (oxidative phosphorylation) (Fig. 2B-C). These six merged GEPs were next projected on the UMAP space, where each individual GEPs highlighted a distinct subpopulation of cells within the integrated tumor cells (Fig. 2D-E).

Merged GEP EWS and merged GEP proliferation is predictive of patient outcome

Next, we investigated whether genes enriched in each merged GEP correlates with survival in previously published datasets. The two microarray data sets used in this analysis consisted of 44 Ewing sarcoma tumors which included 32 primary tumors, five recurrences, and seven metastases (GSE17679)[41] and 85 primary Ewing sarcoma tumors which were collected from patients enrolled on COG and EuroEwing trials (GSE63157)[42]. GSE17679 was used as a discovery cohort where we found that patients with low merged GEP EWS scores or a high proliferation scores had significantly poorer survival (Fig. 3A-D). The other four merged GEPs, MAPK, UPR, Ribosomes, and Oxphos scores did not correlate with overall survival (Supplementary Fig 3).

We further evaluated merged GEP EWS and Proliferation in the validation cohort, GSE63157. Similar to GSE17679, patients with low merged GEP EWS scores had significantly poorer survival. However, while high merged GEP Proliferation scores correlated with poorer survival in GSE17679, it did not reach statistical significance in GSE63157 (Fig. 3B, E, F). Previously, there have been reports that high or low EWS-FLi1 expression in cell line models can lead to a switch between a proliferative non-motile phenotype and a motile and metastatic cell phenotype[14, 15]. However, whether the heterogeneity of EWS/Fli1 gene expression regulation contributes to patient outcome has been unknown. The merged GEP EWS that we have identified is a unique gene set that is enriched for EWS/FLi1 target genes, which correlates to survival.

Individual Ewing sarcoma tumors demonstrate subclonal evolution

As mentioned previously, while analyzing for CNA changes, we identified that most patient tumors were found to have distinct large chromosomal gains and losses that largely agreed with previous findings. However, other than a report utilizing molecular inversion probes, there has been little published on CNA subclones within Ewing sarcoma[43]. Capitalizing on scRNA-seq to analyze individual cells, we sought to identify and investigate Ewing sarcoma subclones, which we found in tumors from patients 5, 38, and 61.

In patient 5, two subclones were identified through inferCNV that diverged following an early chr8 gain (Fig. 4A-B). This led to one subclone, termed Subcluster.1.1, which emerged with a chr1 gain, chr5 gain and a chr11p loss while a smaller cluster emerged with a chr1p loss, chr2 gain, chr 9 gain, and chr16q loss, suggesting both chromosomal gains and losses contribute to tumor evolution. These two subclones demonstrate significant transcriptional heterogeneity (Fig. 4C). In particular, subclone 1.2 expressed higher levels of CD44 and FN1 (Fig. 4D-E). CD44 is a mesenchymal stem cell marker induced upon EWS-FLI1 inhibition in A673 cells[44] and can potentially drives Ewing sarcoma migration and invasion[45]. Interestingly, the CD44 gene is located on chr11p13, which was not a region identified to be amplified, suggesting that subclonal evolution led to an indirect upregulation of CD44. FN1 is one of the most abundantly secreted proteins in Ewing sarcoma cell lines TC32 and CHLA10[46], and upregulation within this subclone could be a mechanism of interaction with the tumor microenvironment by Ewing sarcoma cells.

Similarly, three subclones were identified within the tumor from patient 61 (Fig. 4F). Early alterations were defined by loss of chr1 and gains of 2, 8, and 20. Subclones were identified by the additional gain of chr5 and heterogenous loss and gain of chr12 (Fig. 4G). The subclones display heterogenous expression of VIM and one subclone contained higher expression of IGF1 (Fig. 4H-J). Interestingly, the diagnostic tumor chromosome analysis of the tumor from patient 61 also identified a subclone with additional gains of chr5, 12, 13 (Table 1). This subclone appears to be the combination of the two minor subclones in the scRNA-seq where collectively, CNAs in chr5 and 12 were identified. This finding validates the ability of identifying subclones in both scRNA-seq and clinical karyotyping but they may differ in resolution. The CNA analysis in the tumor from patient 38 identified three subclones but they did not lead to distinct transcriptional profiles, suggesting that perhaps some acquired CNAs drive a differential expression profile more than others (Supplementary Fig. 4).

Tumor immune infiltrates demonstrate distinct patient-dependent microenvironment phenotypes

Tumor infiltrating immune cells, which include cells such as NK/T cells and tumor associated macrophages have been demonstrated to contribute significantly to the progression and metastasis of many cancers. Specific to Ewing sarcoma, the TME has been suggested to be pro-inflammatory with high chemokine expression[47]. To further investigate the composition and characteristics of the Ewing sarcoma TME, we focused on the two most abundant immune cells, T, and myeloid cells.

After further clustering the T cells, we were able to distinguish four subtypes within the TME, consisting of tissue resident memory, CD8+ effector, regulatory, and CD4+IL7R+ T cells (Fig. 5A, Supplementary Table 5). CD8+ effector T cells expressed CD3, CD8, KLRK1, KLRD1, and CD7[48, 49]. Notably, these cells did not express exhaustion markers such as CTLA4, TIGIT, LAG3, or PDCD1 (gene for PD1), indicating a lack of exhausted cells as has been previously found[18]. CD4+ IL7R T cells were identified by the expression of CD3, CD4, IL7R, CD40LG, MAF, and KLRB1 (encodes CD161)[50]. The CD8+ tissue resident memory T cell cluster was differentiated by the expression of CD8, CD69, KLF6, KLRB1, and CD161[18, 51]. Finally, expression of genes such as FOXP3, BATF, and CD4 expression was used to identify regulatory T cells. The cells identified here also expressed inhibitory molecules known in regulatory T cells such as CTLA4 and TIGIT, which may be contributing to an immunosuppressive microenvironment (Fig. 5B-D)[52]. Expression patterns of several other genes also suggest an immunosuppressive environment.

Ligands of CD161, LLT1 (encoded by CLEC2D) and CD69 (also known as CLEC2C) are expressed in the majority of the T cell subtypes. Most studies in immune cells suggest LLT1-CD161 interactions lead to immunosuppressive environments, although the specific relationship in memory T cells remains to be seen[53]. Recently, CD69 has been shown to inhibit terminal differentiation of CD8+ T cells[54]. MIF expression, along with the receptors CD74 and CXCR4, are also found which can lead to chemotaxis[55, 56, 57]. However, MIF binding to CD74 in T cells has been shown to reduce activation and proliferation[58].

In the myeloid population, 4 subtypes were identified, which consisted of non-inflammatory CD1C+ dendritic Cells, IL1B+ inflammatory monocytes, FCN1+ inflammatory monocytes, and CD14+complement macrophages (Fig. 6A, Supplementary Table 6). The non-inflammatory CD1C+ dendritic cell cluster expressed previously reported genes for this cell type such as CD1C, CLEC10A, FCER1A, HLA-DQA1, HLA-DQB1, and HLA-DPB1[18, 59, 60, 61]. The IL1B+ inflammatory monocyte cluster is notable for the high expression of inflammation related genes such as S100A8, A100A9, S100A12, IL1B, and C5AR1, indicating this cell cluster is a pro-inflammatory monocyte population[18, 62]. The FCN1+ inflammatory monocyte cluster was identified by the higher expression of CD14, CLEC12A, CTSS, FCN1, VCAN, and LYZ, with very few cells expressing FCGR3A (gene for CD16), similar to inflammatory monocyte populations as previously reported[18, 62]. Finally, we were able to identify a CD14+ complement macrophage cluster that expresses classical macrophage related genes such as CD68, CD163, as well as complement-related genes C1QC, C1QA, and C1QB, indicating that this population is similar to previously reported C1QC+ tumor associated macrophages[18, 63]. Interestingly, these cells express APOE, PLTP, CD163, which along with C1QC indicate a strong anti-inflammatory phenotype. which could oppose the other tumor infiltrating inflammatory myeloid cells that were able to identify (Fig. 6B-D).

Similar to the T cell subtypes, the myeloid cells express several markers expected to contribute to an immunosuppressive environment. TIM3 (encoded by HAVCR2), which is moderately expressed in each of the myeloid subpopulations in this study, has been shown to reduce the response to tumors[64]. CD80, a ligand of the T cell stimulatory receptor, CD28, is lowly expressed in the myeloid cells of our study as well as those from another Ewing study[17] which suggests dysfunctional myeloid cells may lead to reduced T cell response. Finally, high MIF expression is similar to that in T cells which can reduce T cell activation and proliferation[58].

Diverse cell intrinsic and intercellular interactions are possible in the Ewing sarcoma TME

Following the determination of specific cell types in TME, we next sought to identify potential cellular interactions using CellChat[65]. Here, we were able to identify multiple significant ligand-receptor interactions involving Ewing sarcoma tumor cells (Fig. 7A, Supplementary Table 7). Of the tumor-intrinsic interactions, the most commonly occurring across samples were WNT5A-FZD1, NPY-NPY1R/NPY5R, and MDK-NCL (Fig. 7B). The Wnt pathway, in particular the non-canonical Wnt ligand WNT5A, has been previously reported as highly expressed in metastatic tumors, and as a potential driver of ROR1 signaling in Ewing sarcoma cells[66]. WNT5A has also been identified as a significant gene that is upregulated upon EWS-FLi1 induction in RD cells[24]. Ewing sarcoma cells have been known to highly express NPY, as well as the receptors NPY1R and NPY5R[67, 68]. MDK (midkine) has been found to be part of the Ewing sarcoma secretome[46], however the mechanism of action, and whether NCL (Nucleolin) is a signaling receptor for midkine in Ewing sarcoma is unknown.

Additionally, there are several tumor-immune interactions possible in multiple tumor samples (Fig. 7A-B). The most common across samples were interactions

between the macrophage migration inhibitory factor (MIF) expressed in Ewing sarcoma tumor cells and CD44, CD74, and CXCR4 expressed in several different infiltrating immune cells, including both T cells and myeloid cells. MIF is highly expressed across many cancer types, including glioblastoma[69], gastric cancer[70], neuroblastoma[71], and osteosarcoma[72], although this has not been reported in Ewing sarcoma. MIF was originally discovered as a cytokine that regulates inflammation and immune activation, and is known to bind to the CD74/CD44 complex[73], CXCR2 and CXCR4[74]. In mouse models of soft tissue sarcoma and melanoma, MIF has been reported to interact with CD74 expressing myeloid cells in the TME causing remodeling to a pro-tumor TME[75, 76].

In terms of immune-intrinsic interactions, only 3 interactions were identified as probable in at least 3 patient tumor samples, all involving cell-intrinsic interactions within tissue resident memory T cells (Fig. 7C). MIF expression, as well as the receptors CD74 and CXCR4, are highly expressed in this cell type and as mentioned above, can promote a pro-tumor TME[75, 76]. We also identified that several T cell subtypes express the receptor CD161 (encoded by KLRB1) and the ligands, LLT1 (encoded by CLEC2D) and CD69 (also known as CLEC2C) (Fig. 5B-C). Here, our data agrees that these interactions could be occurring in a cell-intrinsic manner in memory T cells, supporting an immunosuppressed TME[53, 54].

Ewing sarcoma CTCs form distinct clusters in peripheral blood samples

To further investigate mechanisms of metastatic progression, we next sought to identify and interrogate Ewing sarcoma CTCs using scRNA-seq. Eight patients had peripheral blood collected, including three patients where primary tumors were also analyzed. After collection, blood samples were enriched for CTCs using Cellsieve microfiltration membranes as previously described[77] (Supplementary Fig. 5A). Following similar filtering to the primary tumors, 7139 cells with 9371019 UMIs across 19342 genes were identified in the blood samples. Unbiased clustering revealed ten distinct cell types, which were identified by expression of canonical cell type markers (Fig. 8A-C, Supplementary Table 8).

A cluster of cells were identified as Ewing sarcoma CTCs based on expression of KDSR, NKX2-2, STEAP1, and NPW, all genes known to be expressed in Ewing sarcoma, but not found in other circulating hematopoietic cells. Interestingly, these Ewing sarcoma CTCs were only present in patient 5 (Fig. 8D). Since we had identified subclones within the primary tumor from this patient (Fig. 4A), we next assessed whether the CTCs represented either of those subclones using inferCNV. We were able to identify what appeared to be gains in chr8, chr9, and chr12 suggesting that the CTC potentially represents both subclones found in the primary tumor (Supplementary Fig. 5B). Next, we investigated characteristic genes expressed in the identified CTCs. Recent work from the Ewing Sarcoma Cell Line Atlas (ESCLA)[78] has identified key subsets of genes, such as core EWS-ETS binding sites, super-enhancer associated genes, and differentially expressed genes upon EWS-ETS knockdown. Here, we found that the CTCs highly expressed EWS-ETS bound genes such as FOS, CITED2, TCIM, ZNF704, PAPPA and ITM2A. Furthermore, the CTCs highly expressed CCND1, FCGRT, NRN1, TRPM4, and MYOM2, genes that were found either to be associated

with super-enhancers or differentially expressed upon EWS-ETS knockdown (Fig 8E). Interestingly, one of the most highly expressed genes in the CTC cluster was TSPAN8, a family of the tetraspanin superfamily that has been found to be overexpressed in multiple cancers, and have been indicated to drive metastasis [79, 80, 81, 82, 83] (Supplementary Fig. 5C-D). To confirm if TSPAN8 is a direct EWS-FLI1 target, we further examined the publicly available ESCLA ChIP-seq dataset. Here we identified a FLI1 binding peak proximal to TSPAN8 in all 14 cell lines (Fig. 8F, Supplementary Fig. 6), further suggesting that TSPAN8 is an EWS-FLI1 driven gene that could be contributing to CTC maintenance and could be a potential therapeutic target.

Discussion

In this work, we performed scRNA-seq on untreated Ewing sarcoma patient tumor samples, as well as CTC samples, to generate a dataset detailing the heterogenic copy-number and transcriptional profiles of tumor cells, as well as the composition of the TME. Previous cell line and xenograft studies have identified transcriptional programs driven by EWS-FLI1 or cell proliferation but these phenomena have not been confirmed in Ewing sarcoma patients[14, 15, 16]. With the challenges of these types of analyses in bulk RNA-seq data, our scRNA-seq dataset was interrogated for similar transcriptional programs. We also found EWS- and proliferation-driven signatures in addition to four others. Notably, GEP-EWS and GEP-Proliferation were found to be correlate with survival in one microarray dataset and while GEP-EWS was significant in another, GEP-proliferation only trended towards significance. Together, these findings underscore the clinical impact of transcriptional heterogeneity.

The prognostic association of GEP-EWS is interesting because despite EWS-FLI1 and ESWR1 being well-studied in Ewing sarcoma, no related signature or protein has been previously shown to be prognostic to our knowledge. While higher activity of GEP-EWS correlating with better survival was initially surprising based on the many known pro-tumorigenic roles of EWS-FLI1, the role of this fusion is highly nuanced [84]. EWS-FLI1 enhances metabolic and transcriptomic cell plasticity in Ewing sarcoma cells where the presence of plasticity itself may contribute to pro-tumorigenesis and treatment resistance[15, 16]. Additionally, it was recently shown EWS-FLI1 and SIX1 co-regulate an anti-metastatic gene program, which correlates with better overall survival in Ewing sarcoma[85]. Therefore, the association we observe in microarray datasets could be because of reduced metastatic potential or because lower expression of EWS-FLI1-associated genes, in a tumor where expression is typically high, may be indicative of plasticity potential.

In this study, we were also able to characterize the TME of Ewing sarcoma, which contributes important aspects of tumor biology and potential therapeutic targets[86]. Priming of T cells against PAX3[87], STEAP1[88], and PAPPA[89] have shown preclinical effect against Ewing sarcoma cells. PAX3 and PAPPA, but not STEAP1, were uniquely and highly expressed in tumor cells, demonstrating these could be a potential therapeutic target. While we recognize that mRNA expression does not always correlate to cell surface protein expression, we were surprised to find that many potential immunotherapeutic targets currently investigated, such as CD276 (gene for B7H3)[90],

CD99[91], ROR1[92], and IGF1R[92], had significant heterogeneity within and across patient tumors at a single cell definition; this heterogeneity could play an important role in the success of related treatments.

Much of what is known about the Ewing sarcoma TME is that it is often highly immunosuppressive and this study corroborates many of those findings. We found tumor cells to express only low levels of HLA genes, consistent with previous work showing a lack of HLA expression by Ewing sarcoma cells, potentially leading to decreased T cell recognition[19]. We also found evidence that the myeloid cells of the Ewing sarcoma TME may limit T cell activation through expression of TIM3 and the low expression of CD80[64,17]. The T cell subtypes themselves also seem to be supporting immunosuppression through LLT1-CD161, CD69-CD161, and MIF-CD74/CXCR4 interactions which all have previous evidence of supporting pro-tumor environments[53, 54, 58].

The two CAF populations we identified in our study, termed matrix and inflammatory, are similar to CAF subpopulations found in Ewing sarcoma[17] and other cancer types including triple-negative breast[93] and pancreatic cancer[94]. COL1A1+ CAFs have been shown to remodel the extracellular matrix which can contribute to tumor invasion and progression in breast cancer[95, 96]. Inflammatory CAFs that secrete CXCL12, which is highly expressed in the inflammatory CAFs in our study, recruit regulatory T cells, which is another mechanism to support an immunosuppressive environment[97]. However, the specific roles of these two CAF populations are not well-characterized in Ewing sarcoma and whether they perform similarly as in other cancer types remains to be seen. Taken in aggregate, our scRNA-seq analysis of the TME agrees with previous studies that the Ewing sarcoma TME is significantly immunosuppressed.

Historically, detecting and analyzing circulating Ewing sarcoma cells has been challenging because of low expression of markers used in typical carcinoma CTC detection, such as the epithelial cell adhesion molecule EpCAM[98, 99]. To overcome this, there have been several published approaches. First, using surface expression of CD99, flow cytometric analysis has been used to demonstrate that peripheral blood mononuclear cells from patients with Ewing sarcoma can have a higher number of CD99+CD45-mononuclear cells, compared to control[100, 101]. However, there are known populations of early monocytic progenitor cells within the bone marrow, which has made it difficult to isolate and molecularly characterize Ewing sarcoma CTCs due to high background cell contamination[102]. Ewing sarcoma cells have been found to express cell-surface vimentin, and flow cytometric isolation of Ewing sarcoma CTCs by this marker has been reported to be prognostic[103, 104]. Finally, utilizing the significant size difference of sarcoma cells to leukocytes, size-based CTC detection approaches have been performed using platforms such as ISET and CellSieve[77, 105]. We have previously reported that CellSieve can detect CTCs in many patients with high-grade sarcoma, which also includes Ewing sarcoma patients[77, 106]. In our report, we were able to molecularly characterize a population of Ewing Sarcoma CTCs for the first time. While our initial aim was to directly compare matched tumor and CTCs, the sequencing quality difference due to different sample preparation methods (tumor vs blood) prevented a direct comparison. Unexpectedly, we were able to identify TSPAN8 as one of the most highly expressed genes in the CTCs. TSPAN8 has been investigated as a driver of

metastasis and invasion in other cancer types[80, 82, 107, 108], although there have been no studies performed in Ewing sarcoma. Furthermore, TSPAN8 has also been investigated as a potential cell surface target in other cancer types [109, 110]. Future studies should explore TSPAN8 as an antibody-based or CAR T cell-based therapeutic target, as well as further investigation into the role of EWS-FLI1 regulated TSPAN8 expression in tumor metastasis.

Materials and Methods

Tumor collection and processing

Newly diagnosed Ewing sarcoma patients were pathologically confirmed at Children's Hospital Colorado and consented to institutional review board-approved protocol, (COMIRB 00-206) where tumor and blood samples were collected. Tumor analysis was performed under institutional review board-approved protocol, COMIRB 22-1456. Untreated primary tumor samples were biopsied and immediately preserved using DMSO:FBS(10:90) medium in a Corning CoolCell LX container chilled -1°C/minute until reaching - 80°C, followed by archival storage at -150°C. Frozen tumor tissue was thawed at 37°C and gently dissociated using an optimized collagenase-IV/DNaseI/RPMI digestion protocol. Tumor-digest suspensions were strained through a 70µm filter membrane & washed, then treated using the ErythroClear™ Red Blood Cell Depletion Kit (STEMCELL Technologies) followed by DAPI-exclusion FACS (Beckman Coulter - MoFlo™ XDP). The enriched-tumor single cell suspensions were each constructed into barcoded cDNA libraries using the Chromium Next GEM Single Cell 3' platform (10x Genomics). Target recovery was established at 4,000 cells maximum for GEM capture and single cells were sequenced at a targeted depth of 50,000 read pairs per cell.

Blood collection and processing

At time of diagnosis, peripheral whole blood was collected using EDTA(K2) BCTs and immediately passed through an 8µm-pore microfilter membrane (CellSieve) at a rate of 5mL/minute on a KD Scientific LEGATO 110 syringe pump. Captured cells were washed from the membrane using RPMI1640-10%FBS and immediately preserved with DMSO:FBS(10:90) medium in a Corning CoolCell LX container chilled -1°C/minute until reaching - 80°C, followed by archival storage at -150°C. Frozen cells were thawed at 37°C and treated using the ErythroClear™ Red Blood Cell Depletion Kit (STEMCELL Technologies). Cells were constructed into barcoded single-cell cDNA libraries and sequenced as described above.

scRNA-seq processing

Cell Ranger (v6.0.1)[111] was used to process fastq files to cell and gene count tables using unique molecule identifiers (UMIs) aligning to GRCh38 (compiled by 10X Genomics as 2020-A). Downstream quality control and analysis utilized the Seurat (v4.0.4)[112] pipeline. Cell Ranger-filtered data was further processed by removing genes identified in fewer than 10 cells. Cell barcodes were removed if they had greater than 50,000 UMIs, greater than 7,000 genes, greater than 20% of UMIs from mitochondria,

or fewer than 200 genes. Cell doublets were identified and removed using the scDblFinder R package[113]. This filtering criteria resulted in 23,547 cells across all samples totaling 22,639 genes.

The primary and CTC-enriched blood samples were further processed independently. The gene counts were normalized using Seurat by divided by each cell's total counts and multiplying by 10,000 followed by natural-log transformation. The top 2,000 most variable genes were scaled with total UMI and percentage mitochondria regressed out. Clusters were identified as cell types through multiple methods. The top 100 enriched genes by cluster were used as input for over-representation analysis[114] with gene sets from the MSigDb C8 collection[115], the PanglaoDb[116], and select Ewing sarcoma-related gene sets. Similar enrichment analysis was performed for further indicated comparisons. The SCEVAN R package was used to confirm the identification of tumor cells[117]. Canonical cell type markers were plotted in UMAP space and using dot plots. Myeloid and T cells in the primary samples were subset and analyzed separately.

Copy-number alterations in scRNA-seq

Copy-number alterations were estimated in primary and CTC-enriched blood samples using the inferCNV R package[118]. In each data subset, cells defined as Ewing were input as the tumor while all other cells were identified as normal. Following analysis on the dataset of primary samples, three samples were analyzed further as they appeared to have subclusters. Discrete subclusters were identified using random trees whose branches were manually trimmed to fewer clusters according to the hierarchy. Phylogenetic trees were produced with the uphyloplot2 R package with major chromosomal alterations labeled[119].

Non-negative matrix factorization in scRNA-seq

Intratumor gene expression programs (GEPs) were identified using consensus non-negative matrix factorization (cNMF)[120]. The Ewing cells of each primary tumor were processed separately including only genes expressed in each sample (12000 genes). cNMF was run per sample with the 2000 most variant genes using raw counts first with k ranging from 2-10 and 50 iterations to select a k per sample with high stability and a low error rate. After k selection, each sample was run a final time with 200 iterations.

Sample-level GEPs were defined as the top 50 correlated genes per program. Ewing cells from all samples were scored using AddModuleScore[112] followed by z-score transformation of each program's score and labeling the top program per cell as the program with the highest score. Merged GEPs were identified by correlating sample-level programs using the cell scores. Each merged GEP was defined by the aggregate of genes from the appropriate sample-level programs. Genes spanning multiple merged GEPs were removed. Merged GEPs were defined by over-representation analysis as described above[114]. Cells were scored by merged GEPs as described above. UMAP plots were visualized after integrating the samples with the RPCA method[112].

Evaluation of merged gene expression programs in Ewing sarcoma microarray datasets

Two publicly available Ewing sarcoma microarray datasets were analyzed using the merged GEPs. GSE17679 cohort B contains 44 samples and served as the discovery cohort while GSE63157 served as the validation cohort containing 85 samples[41, 42]. Both datasets were normalized with rma[121]. The most variable probeset was chosen for each gene. The genes for each merged GEP were used to subset the GSE17679 dataset. Z-score transformation across each gene was performed and the samples were scored by the mean expression of the transformed values. Samples were divided into two groups using the median score for each merged GEP. Kaplan-Meier plots and log-rank tests were performed using the survminer and survival R packages[122, 123]. Heatmaps were generated with the ComplexHeatmap R package[124]. The merged GEPs demonstrating a correlation with survival were tested in a similar manner in the GSE63157 dataset.

Cell communication in scRNA-seq

Cell communication between ligands and receptors in the cell types of each primary sample were predicted using the CellChat R package[65]. Select interesting interactions were highlighted based on the significant interactions across multiple samples. Ligand and receptor expression was plotted across cell type by patient sample.

FLI1 ChIP-seq in Ewing sarcoma cell lines

Publicly available FLI1 ChIP-seq data in Ewing sarcoma cell lines (GSE176400) were analyzed for binding near the gene TSPAN8[125]. Illumina adapters and low-quality reads were filtered out using BBDuk (v.38.87) (http://jgi.doe.gov/data-and-tools/bb-tools). Bowtie2 (v.2.3.4.3)[126] was used to align the sequencing reads to the hg38 reference human genome. Samtools (v.1.11)[127] was used to select the mapped reads (samtools view -b - q 30) and sort the bam files. PCR duplicates were removed using Picard MarkDuplicates tool [128]. The normalization ratio for each sample was calculated by dividing the number of uniquely mapped human reads of the sample with the lowest number of reads by the number of uniquely mapped human reads of each sample. These normalization ratios were used to randomly sub-sample reads to obtain the same number of reads for each sample using using samtools view -s. Bedtools genomecov was used to create bedgraph files from the bam files[129]. Normalized Bigwig files were created using deepTools bamCoverage[130] and visualized using the trackplot R package[131].

References:

1. Womer RB, *et al.* Randomized controlled trial of interval-compressed chemotherapy for the treatment of localized Ewing sarcoma: a report from the Children's Oncology Group. *J Clin Oncol* **30**, 4148-4154 (2012).

2. Rosen G, *et al.* Proceedings: Disease-free survival in children with Ewing's sarcoma treated with radiation therapy and adjuvant four-drug sequential chemotherapy. *Cancer* **33**, 384-393 (1974).

3. Nesbit ME, Jr., *et al.* Multimodal therapy for the management of primary, nonmetastatic Ewing's sarcoma of bone: a long-term follow-up of the First Intergroup study. *J Clin Oncol* **8**, 1664-1674 (1990).

4. Brohl AS, *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet* **10**, e1004475 (2014).

5. Crompton BD, *et al.* The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov* **4**, 1326-1341 (2014).

6. Tirode F, *et al.* Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov* **4**, 1342-1353 (2014).

7. Dagher R, *et al.* Molecular confirmation of Ewing sarcoma. *J Pediatr Hematol Oncol* **23**, 221-224 (2001).

8. Riggi N, *et al.* EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* **26**, 668-681 (2014).

9. Hancock JD, Lessnick SL. A transcriptional profiling meta-analysis reveals a core EWS-FLI gene expression signature. *Cell Cycle* **7**, 250-256 (2008).

10. Kauer M, *et al.* A molecular function map of Ewing's sarcoma. *PLoS One* **4**, e5415 (2009).

11. Tomazou EM, *et al.* Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell Rep* **10**, 1082-1095 (2015).

12. Theisen ER, Pishas KI, Saund RS, Lessnick SL. Therapeutic opportunities in Ewing sarcoma: EWS-FLI inhibition via LSD1 targeting. *Oncotarget* **7**, 17616-17630 (2016).

13. Bailey K, *et al.* Emerging novel agents for patients with advanced Ewing sarcoma: a report from the Children's Oncology Group (COG) New Agents for Ewing Sarcoma Task Force. *F1000Res* **8**, (2019).

14. Franzetti GA, *et al.* Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **36**, 3505-3514 (2017).

15. Aynaud MM, *et al.* Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cell Rep* **30**, 1767-1779 e1766 (2020).

16. Khoogar R, *et al.* Single-cell RNA profiling identifies diverse cellular responses to EWSR1/FLI1 downregulation in Ewing sarcoma cells. *Cell Oncol (Dordr)* **45**, 19-40 (2022).

17. Visser LL, Bleijs M, Margaritis T, van de Wetering M, Holstege FCP, Clevers H. Ewing Sarcoma Single-cell Transcriptome Analysis Reveals Functionally Impaired Antigen-presenting Cells. *Cancer Res Commun* **3**, 2158-2169 (2023).

18. Cillo AR, *et al.* Ewing Sarcoma and Osteosarcoma Have Distinct Immune Signatures and Intercellular Communication Networks. *Clin Cancer Res* **28**, 4968-4982 (2022).

19. Shamamian P, Mancini M, Kawakami Y, Restifo NP, Rosenberg SA, Topalian SL. Recognition of neuroectodermal tumors by melanoma-specific cytotoxic T lymphocytes: evidence for antigen sharing by tumors derived from the neural crest. *Cancer Immunol Immunother* **39**, 73-83 (1994).

20. Hawkins AG, *et al.* Wnt/beta-catenin-activated Ewing sarcoma cells promote the angiogenic switch. *JCI Insight* **5**, (2020).

21. Henrich IC, *et al.* Ubiquitin-Specific Protease 6 Functions as a Tumor Suppressor in Ewing Sarcoma through Immune Activation. *Cancer Res* **81**, 2171-2183 (2021).

22. Holzer G, Obermair A, Koschat M, Preyer O, Kotz R, Trieb K. Concentration of vascular endothelial growth factor (VEGF) in the serum of patients with malignant bone tumors. *Med Pediatr Oncol* **36**, 601-604 (2001).

23. Kreuter M*, et al.* Clinical significance of Vascular Endothelial Growth Factor-A expression in Ewing's sarcoma. *Eur J Cancer* **42**, 1904-1911 (2006).

24. Hu-Lieskovan S, Zhang J, Wu L, Shimada H, Schofield DE, Triche TJ. EWS-FLI1 fusion protein up-regulates critical genes in neural crest development and is responsible for the observed phenotype of Ewing's family of tumors. *Cancer Res* **65**, 4633-4644 (2005).

25. Guillon N, Tirode F, Boeva V, Zynovyev A, Barillot E, Delattre O. The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. *PLoS One* **4**, e4932 (2009).

26. Gangwal K*, et al.* Microsatellites as EWS/FLI response elements in Ewing's sarcoma. *Proc Natl Acad Sci U S A* **105**, 10149-10154 (2008).

27. Smith R*, et al.* Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. *Cancer Cell* **9**, 405-416 (2006).

28. Kinsey M, Smith R, Lessnick SL. NR0B1 is required for the oncogenic phenotype mediated by EWS/FLI in Ewing's sarcoma. *Mol Cancer Res* **4**, 851-859 (2006).

29. Kinsey M, Smith R, Iyer AK, McCabe ER, Lessnick SL. EWS/FLI and its downstream target NR0B1 interact directly to modulate transcription and oncogenesis in Ewing's sarcoma. *Cancer Res* **69**, 9047-9055 (2009).

30. Grunewald TG*, et al.* STEAP1 is associated with the invasive and oxidative stress phenotype of Ewing tumors. *Mol Cancer Res* **10**, 52-65 (2012).

31. Grunewald TGP*, et al.* High STEAP1 expression is associated with improved outcome of Ewing's sarcoma patients. *Ann Oncol* **23**, 2185-2190 (2012).

32. Surdez D*, et al.* Targeting the EWSR1-FLI1 oncogene-induced protein kinase PKC-beta abolishes ewing sarcoma growth. *Cancer Res* **72**, 4494-4503 (2012).

33. Slyper M*, et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802 (2020).

34. Mackintosh C*, et al.* 1q gain and CDT2 overexpression underlie an aggressive and highly proliferative form of Ewing sarcoma. *Oncogene* **31**, 1287-1298 (2012).

35. Roberts P*, et al.* Ploidy and karyotype complexity are powerful prognostic indicators in the Ewing's sarcoma family of tumors: a study by the United Kingdom Cancer Cytogenetics and the Children's Cancer and Leukaemia Group. *Genes Chromosomes Cancer* **47**, 207-220 (2008).

36. Hattinger CM*, et al.* Prognostic impact of chromosomal aberrations in Ewing tumours. *Br J Cancer* **86**, 1763-1769 (2002).

37. Armengol G*, et al.* Recurrent gains of 1q, 8 and 12 in the Ewing family of tumours by comparative genomic hybridization. *Br J Cancer* **75**, 1403-1409 (1997).

38. Su XA*, et al.* RAD21 is a driver of chromosome 8 gain in Ewing sarcoma to mitigate replication stress. *Genes Dev* **35**, 556-572 (2021).

39. Maurici D, Perez-Atayde A, Grier HE, Baldini N, Serra M, Fletcher JA. Frequency and implications of chromosome 8 and 12 gains in Ewing sarcoma. *Cancer Genet Cytogenet* **100**, 106-110 (1998).

40. Li J, Zhang Y, Yang C, Rong R. Discrepant mRNA and Protein Expression in Immune Cells. *Curr Genomics* **21**, 560-563 (2020).

41. Savola S*, et al.* High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients. *ISRN Oncol* **2011**, 168712 (2011).

42. Volchenboum SL*, et al.* Gene Expression Profiling of Ewing Sarcoma Tumors Reveals the Prognostic Importance of Tumor-Stromal Interactions: A Report from the Children's Oncology Group. *J Pathol Clin Res* **1**, 83-94 (2015).

43. Jahromi MS*, et al.* Molecular inversion probe analysis detects novel copy number alterations in Ewing sarcoma. *Cancer Genet* **205**, 391-404 (2012).

44. Tirode F, Laud-Duval K, Prieur A, Delorme B, Charbord P, Delattre O. Mesenchymal stem cell features of Ewing tumors. *Cancer Cell* **11**, 421-429 (2007).

45. Fernandez-Tabanera E*, et al.* CD44 Modulates Cell Migration and Invasion in Ewing Sarcoma Cells. *Int J Mol Sci* **24**, (2023).

46. Hawkins AG*, et al.* The Ewing Sarcoma Secretome and Its Response to Activation of Wnt/beta-catenin Signaling. *Mol Cell Proteomics* **17**, 901-912 (2018).

47. Berghuis D*, et al.* Pro-inflammatory chemokine-chemokine receptor interactions within the Ewing sarcoma microenvironment determine CD8(+) T-lymphocyte infiltration and affect tumour progression. *J Pathol* **223**, 347-357 (2011).

48. Tsyklauri O*, et al.* Regulatory T cells suppress the formation of potent KLRK1 and IL-7R expressing effector CD8 T cells by limiting IL-2. *Elife* **12**, (2023).

49. Aandahl EM, Sandberg JK, Beckerman KP, Tasken K, Moretto WJ, Nixon DF. CD7 is a differentiation marker that identifies multiple CD8 T cell effector subsets. *J Immunol* **170**, 2349-2355 (2003).

50. DeMartino J*, et al.* Single-cell transcriptomics reveals immune suppression and cell states predictive of patient outcomes in rhabdomyosarcoma. *Nat Commun* **14**, 3074 (2023).

51. Stelma F*, et al.* Human intrahepatic CD69 + CD8+ T cells have a tissue resident memory T cell phenotype with reduced cytolytic capacity. *Sci Rep* **7**, 6172 (2017).

52. Sobhani N, Tardiel-Cyril DR, Davtyan A, Generali D, Roudi R, Li Y. CTLA-4 in Regulatory T Cells for Cancer Immunotherapy. *Cancers (Basel)* **13**, (2021).

53. Braud VM, Meghraoui-Kheddar A, Elaldi R, Petti L, Germain C, Anjuere F. LLT1-CD161 Interaction in Cancer: Promises and Challenges. *Front Immunol* **13**, 847576 (2022).

54. Koyama-Nasu R*, et al.* CD69 Imposes Tumor-Specific CD8+ T-cell Fate in Tumor-Draining Lymph Nodes. *Cancer Immunol Res* **11**, 1085-1099 (2023).

55. Bernhagen J*, et al.* MIF is a noncognate ligand of CXC chemokine receptors in inflammatory and atherogenic cell recruitment. *Nat Med* **13**, 587-596 (2007).

56. Burger-Kentischer A*, et al.* Expression of macrophage migration inhibitory factor in different stages of human atherosclerosis. *Circulation* **105**, 1561-1566 (2002).

57. Zernecke A, Shagdarsuren E, Weber C. Chemokines in atherosclerosis: an update. *Arterioscler Thromb Vasc Biol* **28**, 1897-1908 (2008).

58. Yan X, Orentas RJ, Johnson BD. Tumor-derived macrophage migration inhibitory factor (MIF) inhibits T lymphocyte activation. *Cytokine* **33**, 188-198 (2006).

59. Villani AC*, et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, (2017).

60. Heger L*, et al.* Subsets of CD1c(+) DCs: Dendritic Cell Versus Monocyte Lineage. *Front Immunol* **11**, 559166 (2020).

61. Bourdely P*, et al.* Transcriptional and Functional Analysis of CD1c(+) Human Dendritic Cells Identifies a CD163(+) Subset Priming CD8(+)CD103(+) T Cells. *Immunity* **53**, 335-352 e338 (2020).

62. Tang-Huau TL*, et al.* Human in vivo-generated monocyte-derived dendritic cells and macrophages cross-present antigens through a vacuolar pathway. *Nat Commun* **9**, 2570 (2018).

63. Cheng S*, et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809 e723 (2021).

64. Dixon KO*, et al.* TIM-3 restrains anti-tumour immunity by regulating inflammasome activation. *Nature* **595**, 101-106 (2021).

65. Jin S*, et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).

66. Potratz J*, et al.* Receptor tyrosine kinase gene expression profiles of Ewing sarcomas reveal ROR1 as a potential therapeutic target in metastatic disease. *Mol Oncol* **10**, 677-692 (2016).

67. Kitlinska J*, et al.* Differential effects of neuropeptide Y on the growth and vascularization of neural crest-derived tumors. *Cancer Res* **65**, 1719-1728 (2005).

68. Lu C*, et al.* Hypoxia-activated neuropeptide Y/Y5 receptor/RhoA pathway triggers chromosomal instability and bone metastasis in Ewing sarcoma. *Nat Commun* **13**, 2323 (2022).

69. Presti M*, et al.* Overexpression of macrophage migration inhibitory factor and functionally-related genes, D-DT, CD74, CD44, CXCR2 and CXCR4, in glioblastoma. *Oncol Lett* **16**, 2881-2886 (2018).

70. He LJ*, et al.* Macrophage migration inhibitory factor as a potential prognostic factor in gastric cancer. *World J Gastroenterol* **21**, 9916-9926 (2015).

71. Cavalli E*, et al.* Overexpression of Macrophage Migration Inhibitory Factor and Its Homologue D-Dopachrome Tautomerase as Negative Prognostic Factor in Neuroblastoma. *Brain Sci* **9**, (2019).

72. Wang C*, et al.* Macrophage migration inhibitory factor promotes osteosarcoma growth and lung metastasis through activating the RAS/MAPK pathway. *Cancer Lett* **403**, 271-279 (2017).

73. Shi X*, et al.* CD44 is the signaling component of the macrophage migration inhibitory factor-CD74 receptor complex. *Immunity* **25**, 595-606 (2006).

74. Schwartz V*, et al.* A functional heteromeric MIF receptor formed by CD74 and CXCR4. *FEBS Lett* **583**, 2749-2757 (2009).

75. Tessaro FHG*, et al.* Single-cell RNA-seq of a soft-tissue sarcoma model reveals the critical role of tumor-expressed MIF in shaping macrophage heterogeneity. *Cell Rep* **39**, 110977 (2022).

76. Figueiredo CR*, et al.* Blockade of MIF-CD74 Signalling on Macrophages and Dendritic Cells Restores the Antitumour Immune Response Against Metastatic Melanoma. *Front Immunol* **9**, 1132 (2018).

77. Hayashi M, *et al.* Size-based detection of sarcoma circulating tumor cells and cell clusters. *Oncotarget* **8**, 78965-78977 (2017).

78. Orth MF, *et al.* Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation. *Cell Rep* **41**, 111761 (2022).

79. Park CS, *et al.* Therapeutic targeting of tetraspanin8 in epithelial ovarian cancer invasion and metastasis. *Oncogene* **35**, 4540-4548 (2016).

80. Kim TK, *et al.* Generation of a human antibody that inhibits TSPAN8-mediated invasion of metastatic colorectal cancer cells. *Biochem Biophys Res Commun* **468**, 774-780 (2015).

81. Anami K, *et al.* TSPAN8, identified by Escherichia coli ampicillin secretion trap, is associated with cell growth and invasion in gastric cancer. *Gastric Cancer* **19**, 370-380 (2016).

82. Zhu R, *et al.* TSPAN8 promotes cancer cell stemness via activation of sonic Hedgehog signaling. *Nat Commun* **10**, 2863 (2019).

83. Li J, *et al.* SOX9 is a critical regulator of TSPAN8-mediated metastasis in pancreatic cancer. *Oncogene* **40**, 4884-4893 (2021).

84. Li M, Chen C. Regulation of Metastasis in Ewing Sarcoma. *Cancers (Basel)* **14**, (2022).

85. Hughes CJ, *et al.* SIX1 and EWS/FLI1 co-regulate an anti-metastatic gene network in Ewing Sarcoma. *Nat Commun* **14**, 4357 (2023).

86. Morales E, Olson M, Iglesias F, Dahiya S, Luetkens T, Atanackovic D. Role of immunotherapy in Ewing sarcoma. *J Immunother Cancer* **8**, (2020).

87. Rodeberg DA, Nuss RA, Elsawa SF, Erskine CL, Celis E. Generation of tumoricidal PAX3 peptide antigen specific cytotoxic T lymphocytes. *Int J Cancer* **119**, 126-132 (2006).

88. Schirmer D, *et al.* Transgenic antigen-specific, HLA-A*02:01-allo-restricted cytotoxic T cells recognize tumor-associated target antigen STEAP1 with high specificity. *Oncoimmunology* **5**, e1175795 (2016).

89. Kirschner A, *et al.* Pappalysin-1 T cell receptor transgenic allo-restricted T cells kill Ewing sarcoma in vitro and in vivo. *Oncoimmunology* **6**, e1273301 (2017).

90. Majzner RG, *et al.* CAR T Cells Targeting B7-H3, a Pan-Cancer Antigen, Demonstrate Potent Preclinical Activity Against Pediatric Solid Tumors and Brain Tumors. *Clin Cancer Res* **25**, 2560-2574 (2019).

91. Rocchi A, *et al.* CD99 inhibits neural differentiation of human Ewing sarcoma cells and thereby contributes to oncogenesis. *J Clin Invest* **120**, 668-680 (2010).

92. Huang X, *et al.* IGF1R- and ROR1-Specific CAR T Cells as a Potential Therapy for High Risk Sarcomas. *PLoS One* **10**, e0133152 (2015).

93. Wu SZ, *et al.* Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J* **39**, e104063 (2020).

94. Ohlund D, *et al.* Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214**, 579-596 (2017).

95. Levental KR, *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891-906 (2009).

96. Huo CW, *et al.* High mammographic density is associated with an increase in stromal collagen and immune cells within the mammary epithelium. *Breast Cancer Res* **17**, 79 (2015).

97. Mezzapelle R, *et al.* CXCR4/CXCL12 Activities in the Tumor Microenvironment and Implications for Tumor Immunotherapy. *Cancers (Basel)* **14**, (2022).

98. Cristofanilli M, *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* **351**, 781-791 (2004).

99. Tombolan L, *et al.* Pediatric sarcomas display a variable EpCAM expression in a histology-dependent manner. *Transl Oncol* **13**, 100846 (2020).

100. Dubois SG, Epling CL, Teague J, Matthay KK, Sinclair E. Flow cytometric detection of Ewing sarcoma cells in peripheral blood and bone marrow. *Pediatr Blood Cancer* **54**, 13-18 (2010).

101. Vo KT, *et al.* Impact of Two Measures of Micrometastatic Disease on Clinical Outcomes in Patients with Newly Diagnosed Ewing Sarcoma: A Report from the Children's Oncology Group. *Clin Cancer Res* **22**, 3643-3650 (2016).

102. Dworzak MN, *et al.* Flow cytometric assessment of human MIC2 expression in bone marrow, thymus, and peripheral blood. *Blood* **83**, 415-425 (1994).

103. Satelli A, *et al.* Universal marker and detection tool for human sarcoma circulating tumor cells. *Cancer Res* **74**, 1645-1650 (2014).

104. Dao L, *et al.* Prognostic Value of Cell-Surface Vimentin-Positive CTCs in Pediatric Sarcomas. *Front Oncol* **11**, 760267 (2021).

105. Chinen LT, *et al.* Isolation, detection, and immunomorphological characterization of circulating tumor cells (CTCs) from patients with different types of sarcoma using isolation by size of tumor cells: a window on sarcoma-cell invasion. *Onco Targets Ther* **7**, 1609-1617 (2014).

106. Llosa NJ, *et al.* Reduced-Intensity Haploidentical Bone Marrow Transplantation with Post-Transplant Cyclophosphamide for Solid Tumors in Pediatric and Young Adult Patients. *Biol Blood Marrow Transplant* **23**, 2127-2136 (2017).

107. Wang H, Rana S, Giese N, Buchler MW, Zoller M. Tspan8, CD44v6 and alpha6beta4 are biomarkers of migrating pancreatic cancer-initiating cells. *Int J Cancer* **133**, 416-426 (2013).

108. El Kharbili M, *et al.* Tspan8-beta-catenin positive feedback loop promotes melanoma invasion. *Oncogene* **38**, 3781-3793 (2019).

109. Schafer D, *et al.* Identification of CD318, TSPAN8 and CD66c as target candidates for CAR T cell based immunotherapy of pancreatic adenocarcinoma. *Nat Commun* **12**, 1453 (2021).

110. Heo K, Lee S. TSPAN8 as a Novel Emerging Therapeutic Target in Cancer for Monoclonal Antibody Therapy. *Biomolecules* **10**, (2020).

111. Zheng GX, *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).

112. Hao Y, *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

113. Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2021).

114. Wu T, *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

115. Subramanian A, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

116. Franzen O, Gan LM, Bjorkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, (2019).

117. De Falco A, Caruso F, Su XD, Iavarone A, Ceccarelli M. A variational algorithm to detect the clonal copy number substructure of tumors from scRNA-seq data. *Nat Commun* **14**, 1074 (2023).

118. Tickle T TI, Georgescu C, Brown M, Haas B. inferCNV of the Trinity CTAT Project. (2019).

119. Kurtenbach S, Cruz AM, Rodriguez DA, Durante MA, Harbour JW. Uphyloplot2: visualizing phylogenetic trees from single-cell RNA-seq data. *BMC Genomics* **22**, 419 (2021).

120. Kotliar D, *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**, (2019).

121. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363-2367 (2010).

122. T T. A Package for Survival Analysis in R. R package version 3.5-7.) (2023).

123. T T. Package for Survival Analysis in R.) (2023).

124. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).

125. Tareke KA, Awoke AG. Comparing surface water supply index and streamflow drought index for hydrological drought analysis in Ethiopia. *Heliyon* **8**, e12000 (2022).

126. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

127. Li H, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

128. Institute B. Picard Tools.).

129. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

130. Ramirez F, *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).

131. Pohl A, Beato M. bwtool: a tool for bigWig files. *Bioinformatics* **30**, 1618-1619 (2014).

Funding: This work was supported by National Institutes of Health grant (K12HD068372) (M.H.), the Hyundai Hope on Wheels Young Investigator Award (M.H.), the St. Baldrick's Foundation Scholar Award (M.H.), the Congressionally Directed Medical Research Program (CA200508) (M. H., A. G.), and Tanabe-Bobrow Foundation award (M.H.). The University of Colorado Cancer Center

Biostatistics & Bioinformatics Shared Resource and the Genomics Shared Resources are supported by the University of Colorado Cancer Center (P30CA046934).

.

Author contributions: M.H. developed concepts, designed experiments. A.G., A.B., M.H. wrote the manuscript. A.G., A.B., E.D., S.NT., J.L., M.V., M.H., performed the experiments and analyzed the data. M.H., A.T., J.B., N.D., C.C., T.G., B.G., S.LF., J.S., performed sample collection, N.O., T.P., R.S., D.W., gave technical support and conceptual advice.

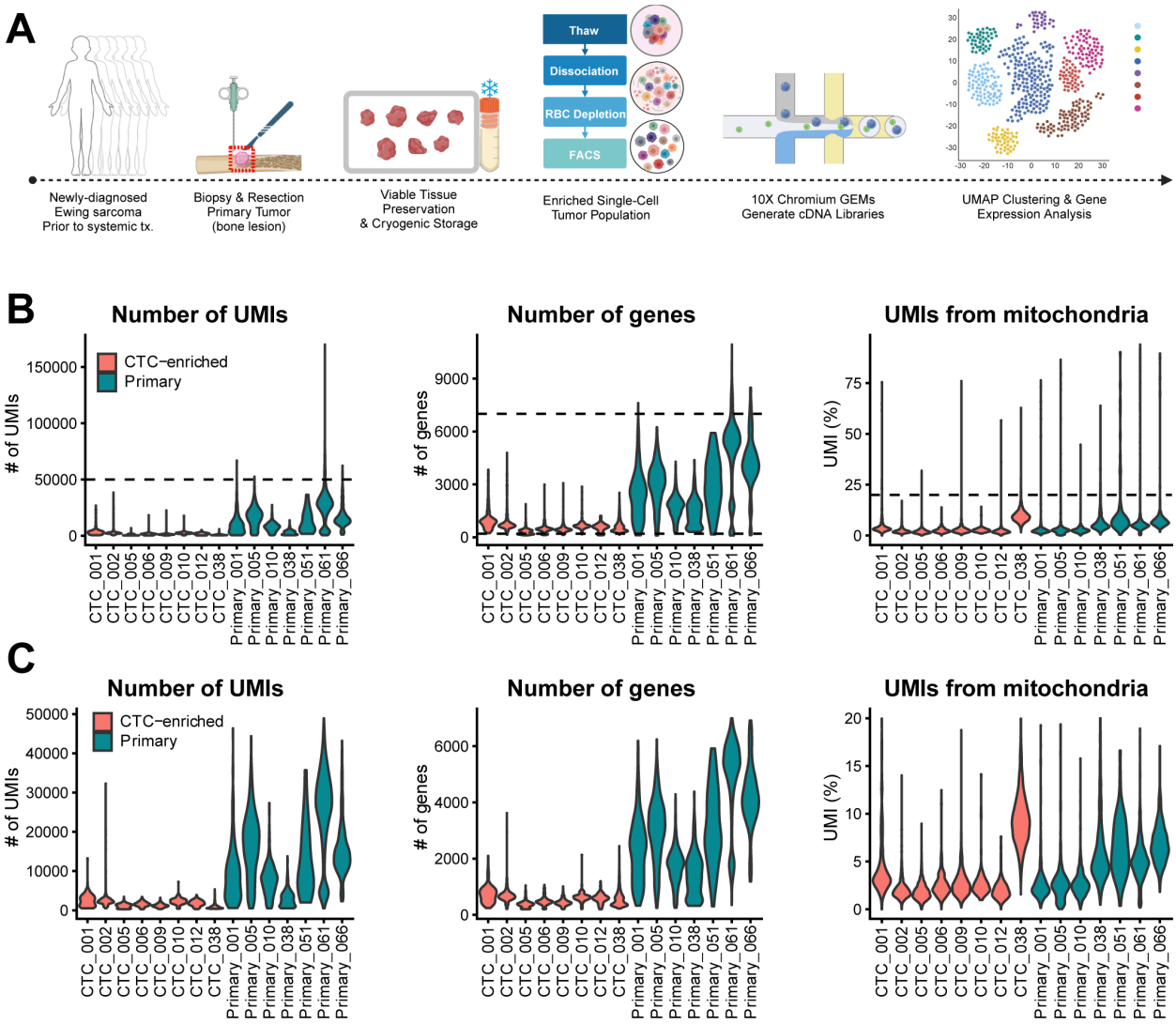
Competing interests: There are no competing interests to report.

Data and materials availability: Raw and processed scRNA-seq data will be deposited in the Gene Expression Omnibus (GSEXXXX). Here also include any materials that must be obtained through an MTA. Acknowledgments follow the references and notes but are not numbered.

Table 1

ID	Patient Data	Location	Metastases	Fusion	Tumor Karyotype	Recurrence	OS	Samples	
								Tumor	CTC Blood
001	13y M	L fibula	-	EWS-ERG	EWS-ERG only	-		+	+
002	19y F	Pleural mass	+ (spine)	EWS-FLI1		+ (local progression on therapy)	12 mos	-	+
005	15y M	R iliac crest	-	EWS-FLI1	Hyperploidy: 5, 8, 1q	+ (local, 9 mos off therapy)		+	+
006	8y M	Paraspinal	-	EWS-FLI1		-		-	+
009	15y F	Sacrum	+ (R humerus)	EWS-FLI1	Hyperploidy: 2, 4, 12, 16, 20, 21	+ (local, 7 mos off therapy)	29 mos	-	+
010	15y M	R humerus	-	EWS-FLI1	EWS-FLI1 only	-		+	+
012	17y F	Cutaneous (R thigh)	-	EWS-FLI1		-		-	+
038	9y M	L ileum	-	EWS-FLI1	Hyperploidy: 8	+ (metastatic, 3 mos off therapy, lung)	24 mos	+	+
051	19y M	R ulna	+ (liver, spine, sacrum, ilium, femurs)	EWS-FLI1	Hyperploidy: 2, 8, 13, 21	+ (metastatic, 2 mos off therapy, extensive disease in bones and brain)	16 mos	+	-

061	6y M	R scapula	-	EWS-FLI1	Hyperploidy: Clone 1: 2, 8, 20 clone 2: Additional gain of 5, 12, 13	-		+	-
066	14y M	L scapula	-	EWS-FII1	EWS-FLI1 only	-		+	-



Supplementary Figure 1. Quality control of the scRNA-seq dataset
Workflow of primary tumor sample collection and processing for scRNA-seq (A). Distribution of UMIs, genes, and mitochondria in barcodes considered to be cells from each sample prefiltering (B) or post filtering (C). The dashed lines in A indicate the chosen filtering criteria.

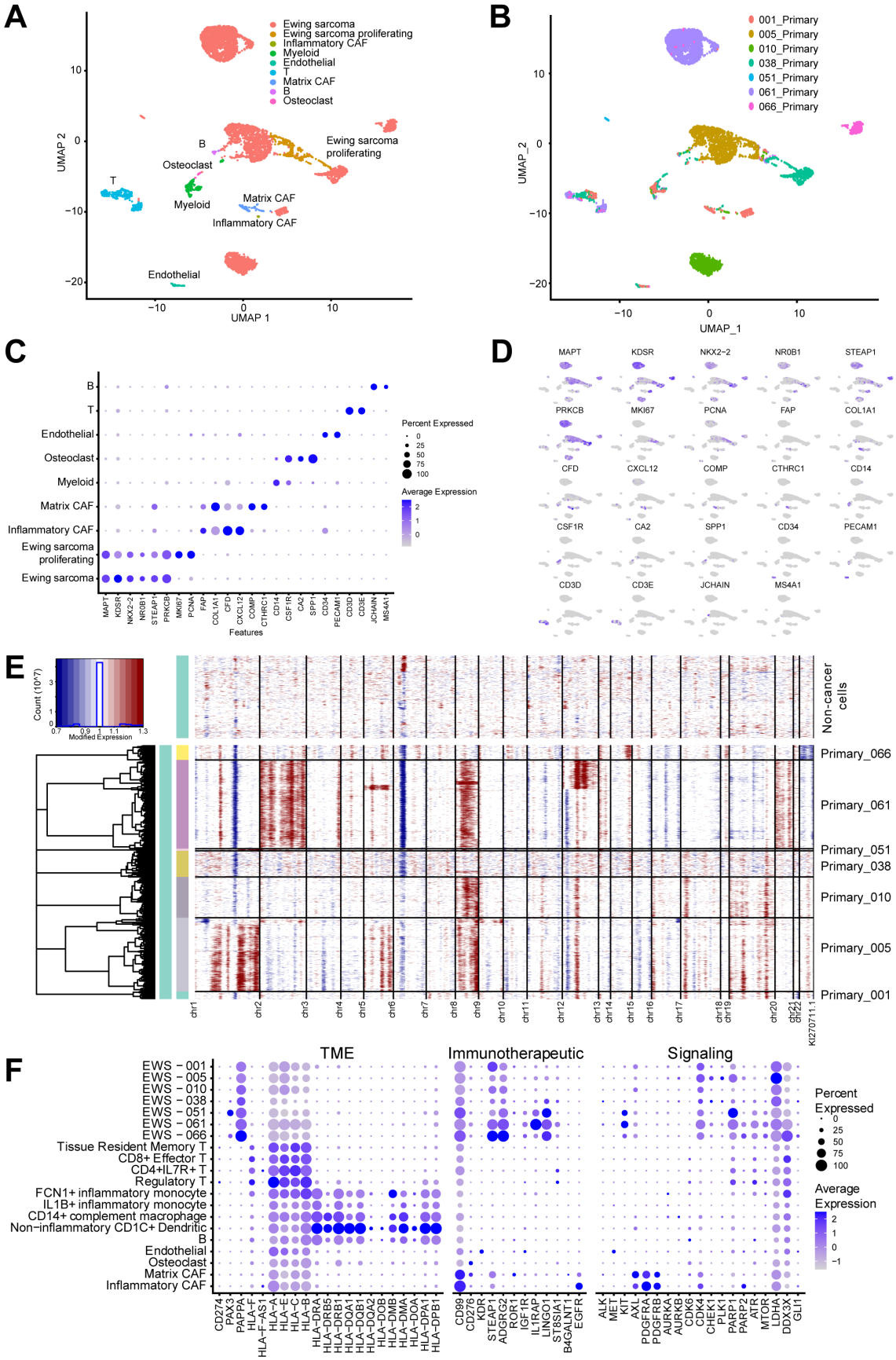
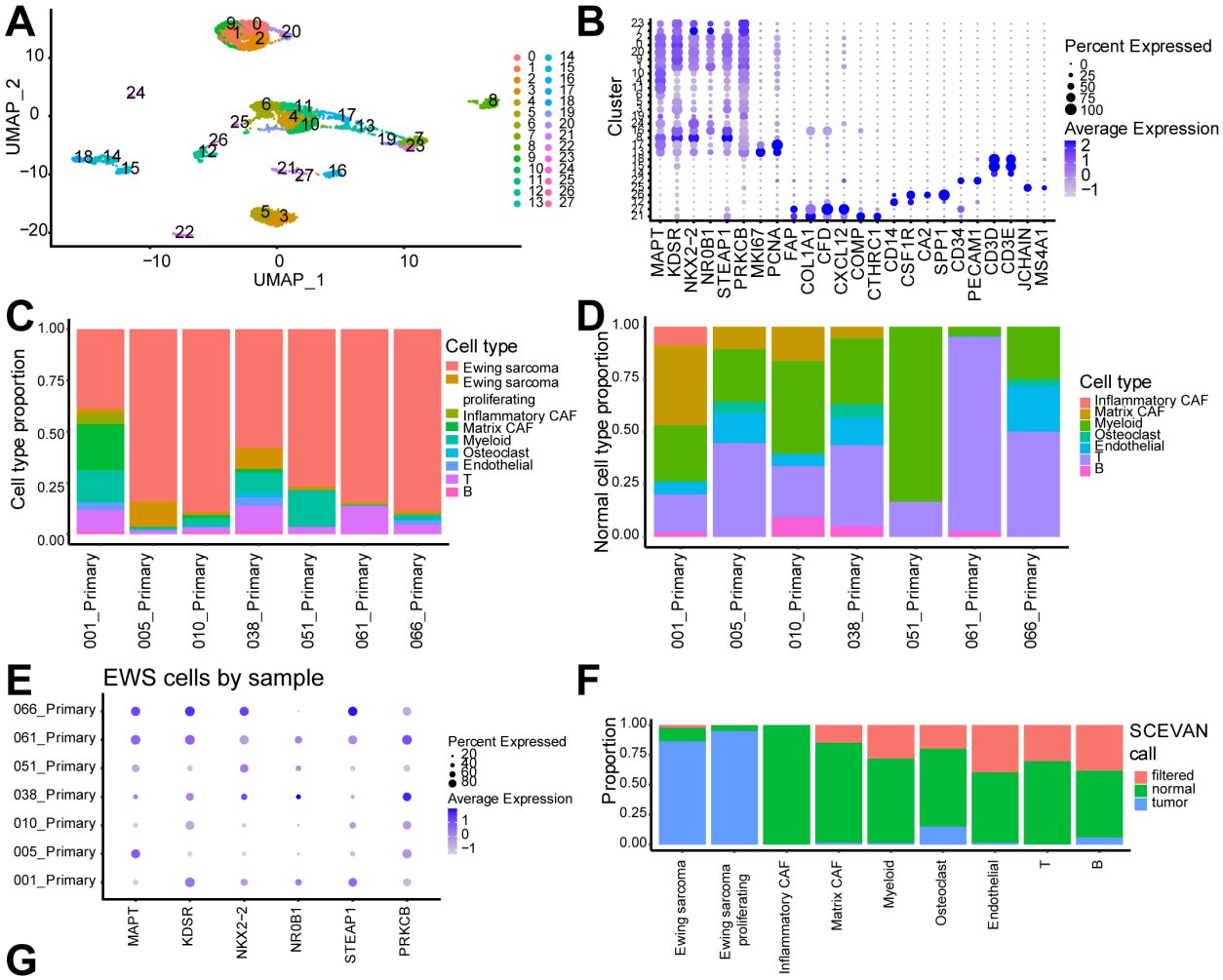

A
Ewing sarcoma
Ewing sarcoma proliferating
Inflammatory CAF
Myeloid
Endothelial
T
Matrix CAF
B
Osteoclast
UMAP 2
UMAP 1
B
001_Primary
005_Primary
010_Primary
038_Primary
051_Primary
061_Primary
066_Primary
UMAP_2
UMAP_1
C
Percent Expressed
Average Expression
Features
D
MAPT
KDSR
NKX2-2
NR0B1
STEAP1
PRKCB
MKI67
PCNA
FAP
COL1A1
CFD
CXCL12
COMP
CTHRC1
CD14
CSF1R
CA2
SPP1
CD34
PECAM1
CD3D
CD3E
JCHAIN
MS4A1
E
Count (10^7)
Modified Expression
Non-cancer cells
Primary_066
Primary_061
Primary_051
Primary_038
Primary_010
Primary_005
Primary_001
F
TME
Immunotherapeutic
Signaling
EWS − 001
EWS − 005
EWS − 010
EWS − 038
EWS − 051
EWS − 061
EWS − 066
Tissue Resident Memory T
CD8+ Effector T
CD4+IL7R+ T
Regulatory T
FCN1+ inflammatory monocyte
IL1B+ inflammatory monocyte
CD14+ complement macrophage
Non−inflammatory CD1C+ Dendritic
B
Endothelial
Osteoclast
Matrix CAF
Inflammatory CAF
Percent Expressed
Average Expression

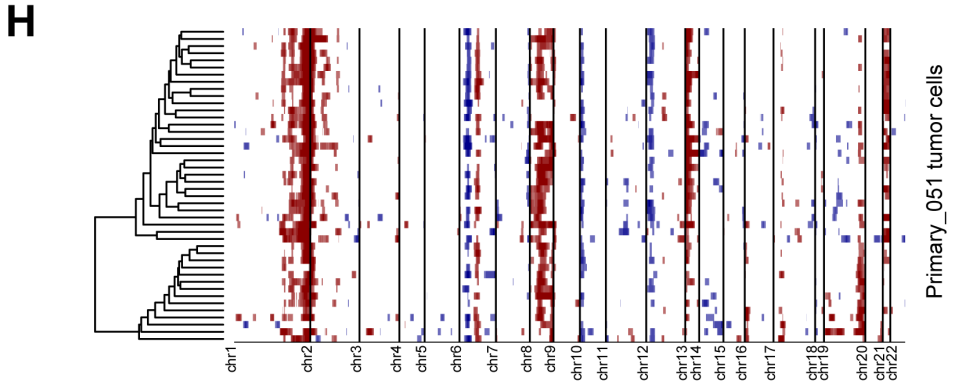
Figure 1. scRNA-seq of Ewing Sarcoma primary tumors reveals several cell types including cancer cells

UMAP projections labeled by the identified cell types (A) or the sample source (B). Expression of canonical marker genes across cell types are shown as a dotplot (C) or UMAP (D). Inferred copy-number alteration estimated in the identified cancer cells separated by sample with all other cell types serving as control cells (E). Expression of select genes within Ewing sarcoma tumors from specific patients and aggregate cell types (F).



G

EWS Gene Set	GeneRatio	BgRatio	Pvalue	FDR
STAEGE_EWING_FAMILY_TUMOR	16/59	32/1757	1.23E-16	4.91E-16
RIGGI_EWING_SARCOMA_PROGENITOR_UP	40/59	414/1757	2.53E-13	5.07E-13
ZHANG_TARGETS_OF_EWSR1_FLI1_FUSION	16/59	87/1757	6.47E-09	8.62E-09
MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_UP	18/59	255/1757	1.07E-3	1.07E-3



Supplementary Figure 2. Distribution of clusters and cell types in primary Ewing Sarcoma tumors using scRNA-seq

Expression of canonical marker genes across clusters are shown as a UMAP (A) or dotplot (B). Distribution of all cell types (C) or non-cancerous cell types (D) by sample. Expression of six canonical Ewing Sarcoma-related genes in the tumor cells identified from each sample (E). SCEVAN tumor classification results distributed by the identified cell type (F). Over-representation analysis of selected Ewing Sarcoma-related gene sets in the tumor cell clusters (G). Inferred copy-number alteration estimated in 051 tumor cells. Same as Figure 1E (H).

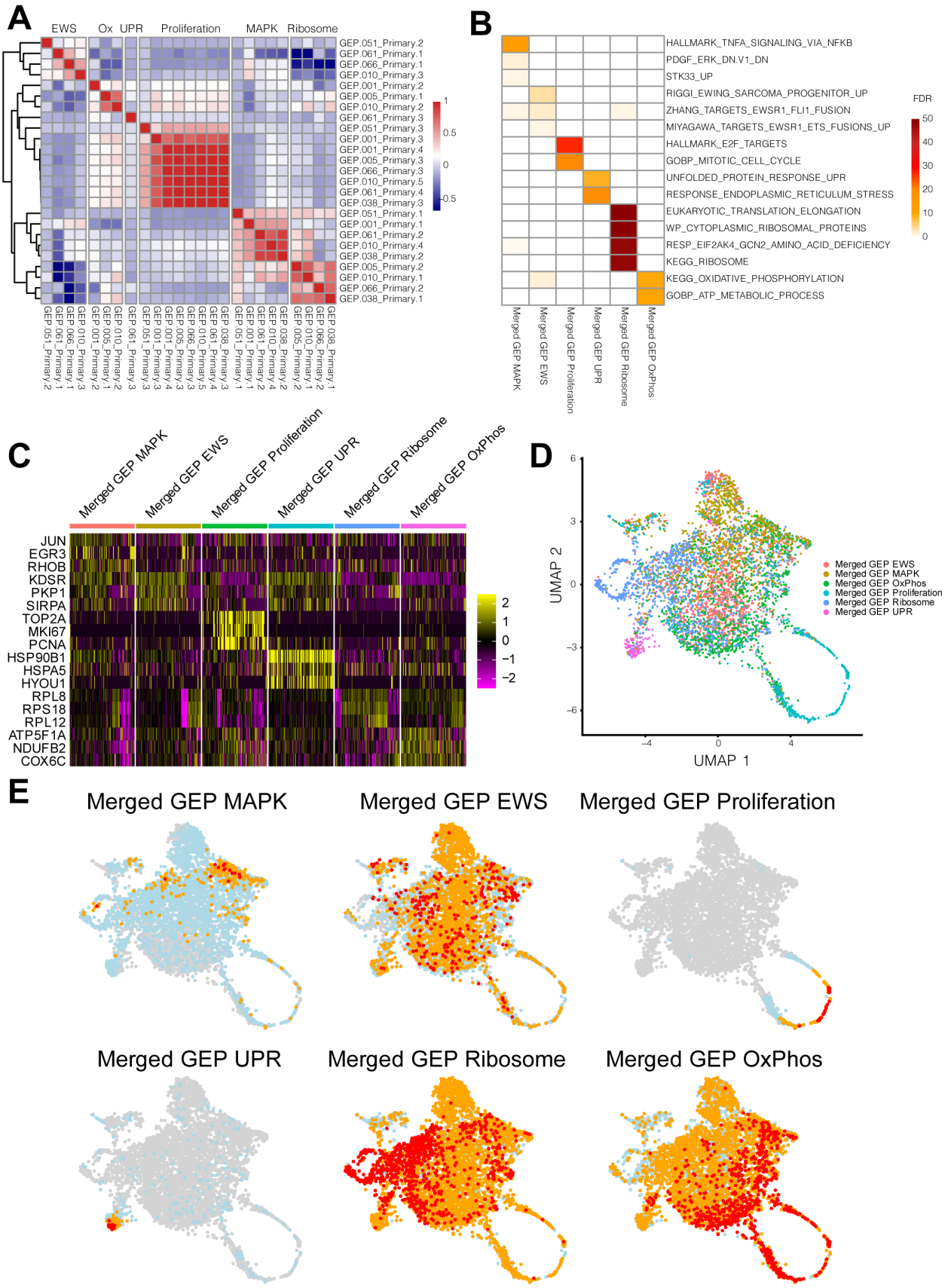

A
EWS
Ox
UPR
Proliferation
MAPK
Ribosome
GEP.051_Primary.2
GEP.061_Primary.1
GEP.066_Primary.1
GEP.010_Primary.3
GEP.001_Primary.2
GEP.005_Primary.1
GEP.010_Primary.2
GEP.061_Primary.3
GEP.051_Primary.3
GEP.001_Primary.3
GEP.001_Primary.4
GEP.005_Primary.3
GEP.066_Primary.3
GEP.010_Primary.5
GEP.061_Primary.4
GEP.038_Primary.3
GEP.051_Primary.1
GEP.001_Primary.1
GEP.061_Primary.2
GEP.010_Primary.4
GEP.038_Primary.2
GEP.005_Primary.2
GEP.010_Primary.1
GEP.066_Primary.2
GEP.038_Primary.1
1
0.5
0
-0.5
B
HALLMARK_TNFA_SIGNALING_VIA_NFKB
PDGF_ERK_DN.V1_DN
STK33_UP
RIGGI_EWING_SARCOMA_PROGENITOR_UP
ZHANG_TARGETS_EWSR1_FLI1_FUSION
MIYAGAWA_TARGETS_EWSR1_ETS_FUSIONS_UP
HALLMARK_E2F_TARGETS
GOBP_MITOTIC_CELL_CYCLE
UNFOLDED_PROTEIN_RESPONSE_UPR
RESPONSE_ENDOPLASMIC_RETICULUM_STRESS
EUKARYOTIC_TRANSLATION_ELONGATION
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
RESP_EIF2AK4_GCN2_AMINO_ACID_DEFICIENCY
KEGG_RIBOSOME
KEGG_OXIDATIVE_PHOSPHORYLATION
GOBP_ATP_METABOLIC_PROCESS
Merged GEP MAPK
Merged GEP EWS
Merged GEP Proliferation
Merged GEP UPR
Merged GEP Ribosome
Merged GEP OxPhos
FDR
50
40
30
20
10
0
C
Merged GEP MAPK
Merged GEP EWS
Merged GEP Proliferation
Merged GEP UPR
Merged GEP Ribosome
Merged GEP OxPhos
JUN
EGR3
RHOB
KDSR
PKP1
SIRPA
TOP2A
MKI67
PCNA
HSP90B1
HSPA5
HYOU1
RPL8
RPS18
RPL12
ATP5F1A
NDUFB2
COX6C
2
1
0
−1
−2
D
UMAP 2
UMAP 1
6
3
0
−3
−6
−4
0
4
Merged GEP EWS
Merged GEP MAPK
Merged GEP OxPhos
Merged GEP Proliferation
Merged GEP Ribosome
Merged GEP UPR
E
Merged GEP MAPK
Merged GEP EWS
Merged GEP Proliferation
Merged GEP UPR
Merged GEP Ribosome
Merged GEP OxPhos

Figure 2. Identification of gene expression programs in Ewing Sarcoma primary tumors.
Correlation matrix of the full dataset scores from each sample-level gene expression program (A). Merged gene expression programs were analyzed using over-representation analysis and the FDR select gene sets are displayed by program (B). A heatmap of select genes with groups downsampled to 100 cells, with yellow indicating higher expression (C). Cells were labeled by the merged gene expression program (D) along with the score (E).

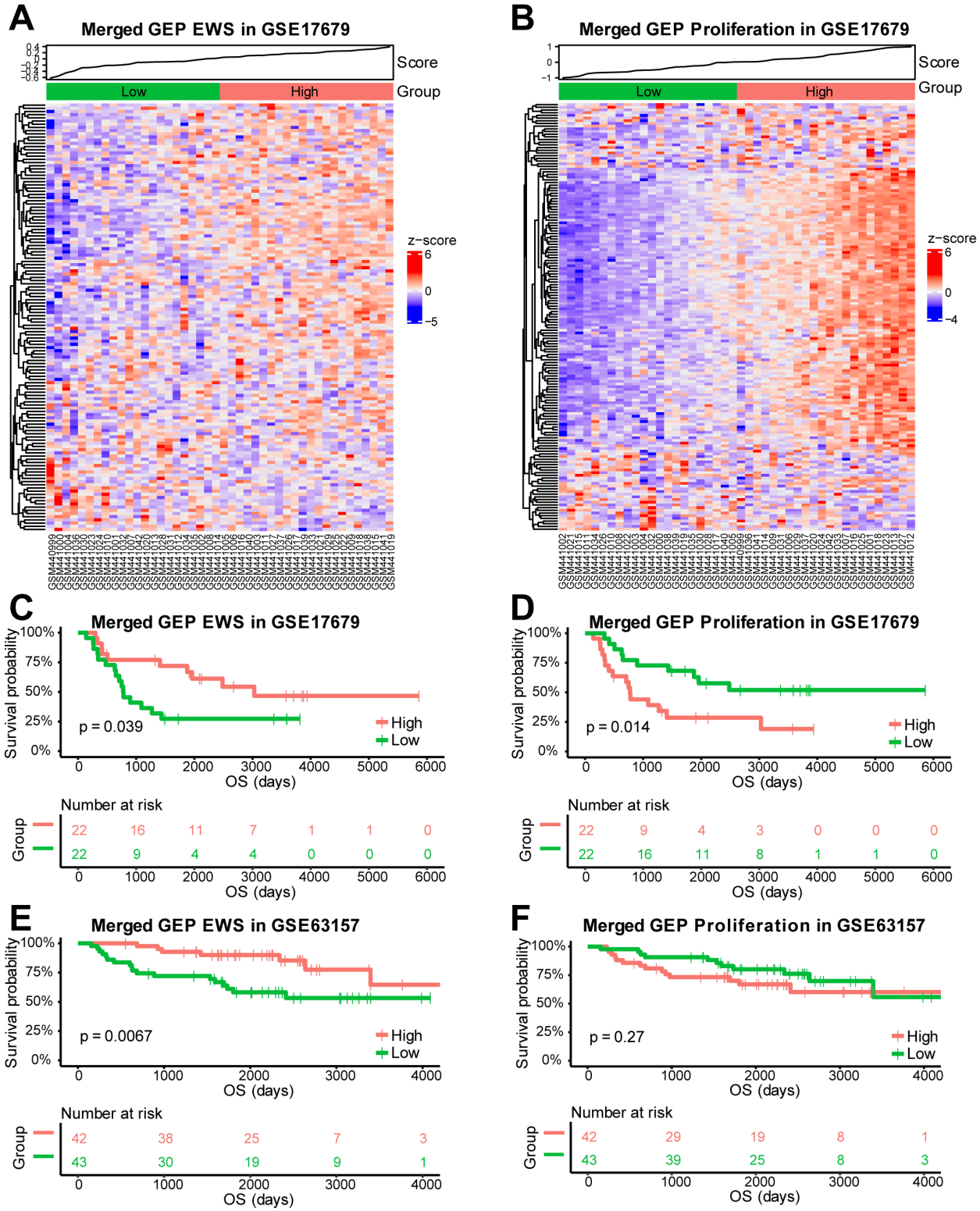
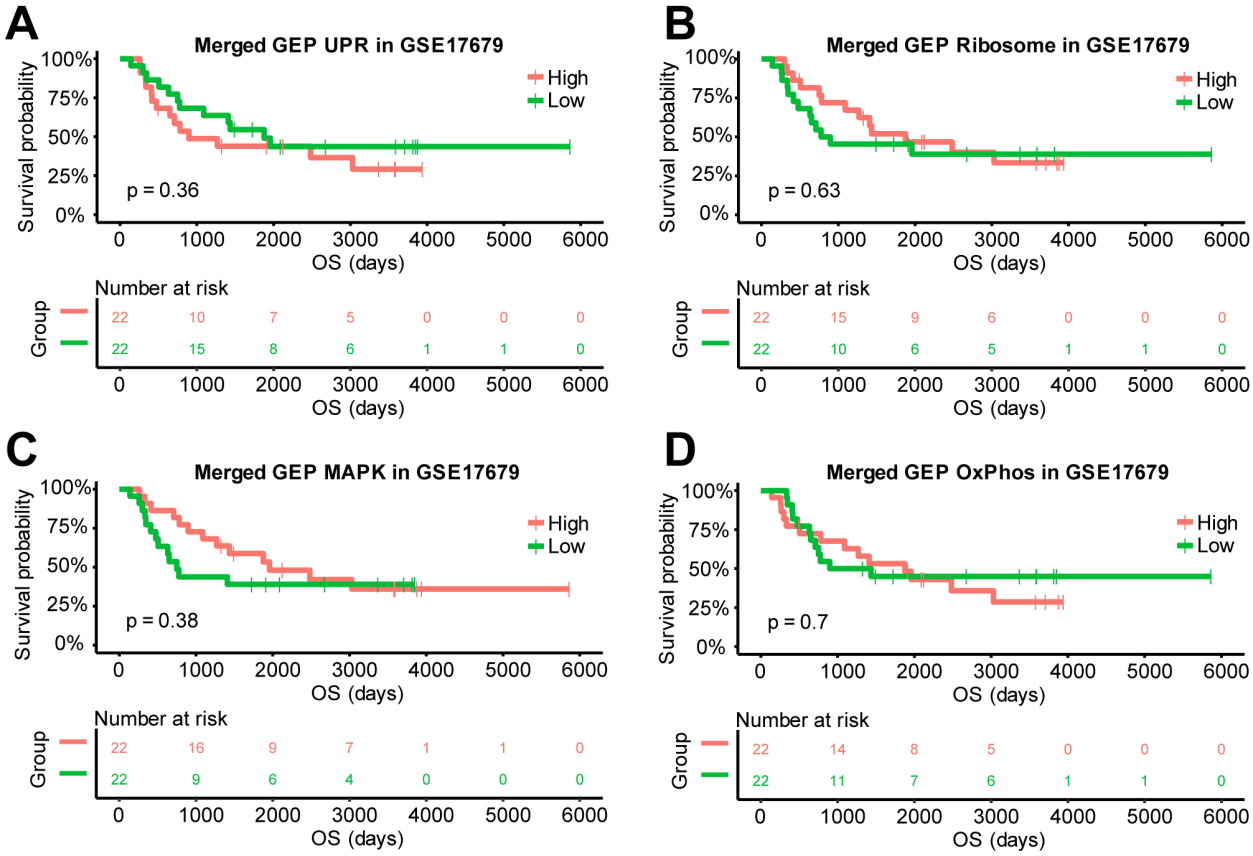


Figure 3. Correlation of the EWS and Proliferation gene expression programs with survival.

Heatmaps displaying the expression of the genes in the EWS (A) and Proliferation (B) gene expression programs in Ewing Sarcoma tumors from GSE17679. Sample scores

are plotted at the top with a median split used to classify samples as Low (green) or High (red) for each program. Kaplan-Meier curves with a log-rank test comparing the overall survival between the Low and High groups identified from EWS (C) or Proliferation (D) gene expression programs in GSE17679 and the validation Ewing Sarcoma cohort, GSE63157 (E,F).



Supplementary Figure 3. Correlation of four gene expression programs with survival in the discovery cohort.

Kaplan-Meier curves with a log-rank test comparing the overall survival between the Low and High groups identified from the UPR (A), Ribosome (B), MAPK (C), or OxPhos (D) gene expression programs in GSE17679.

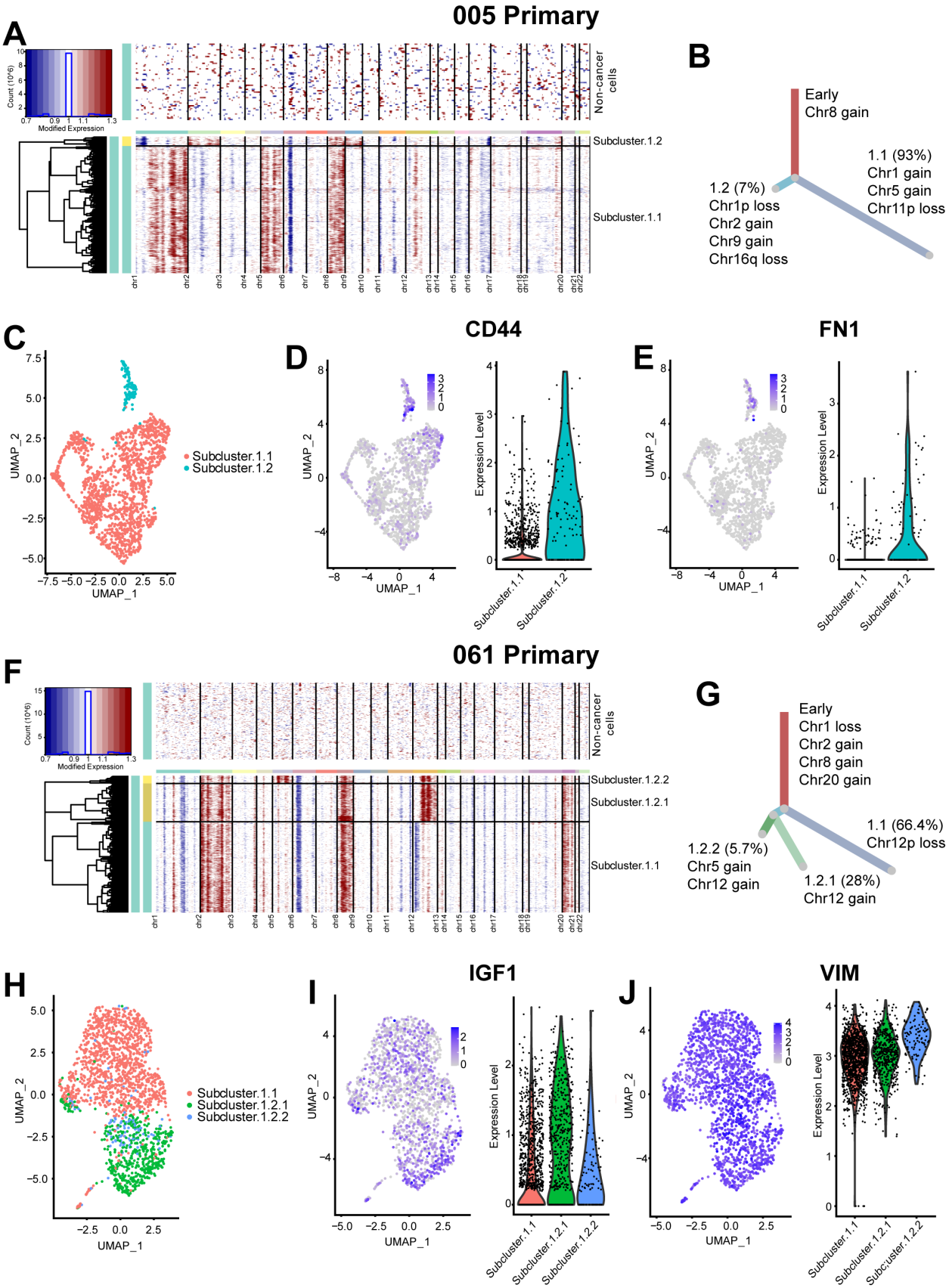
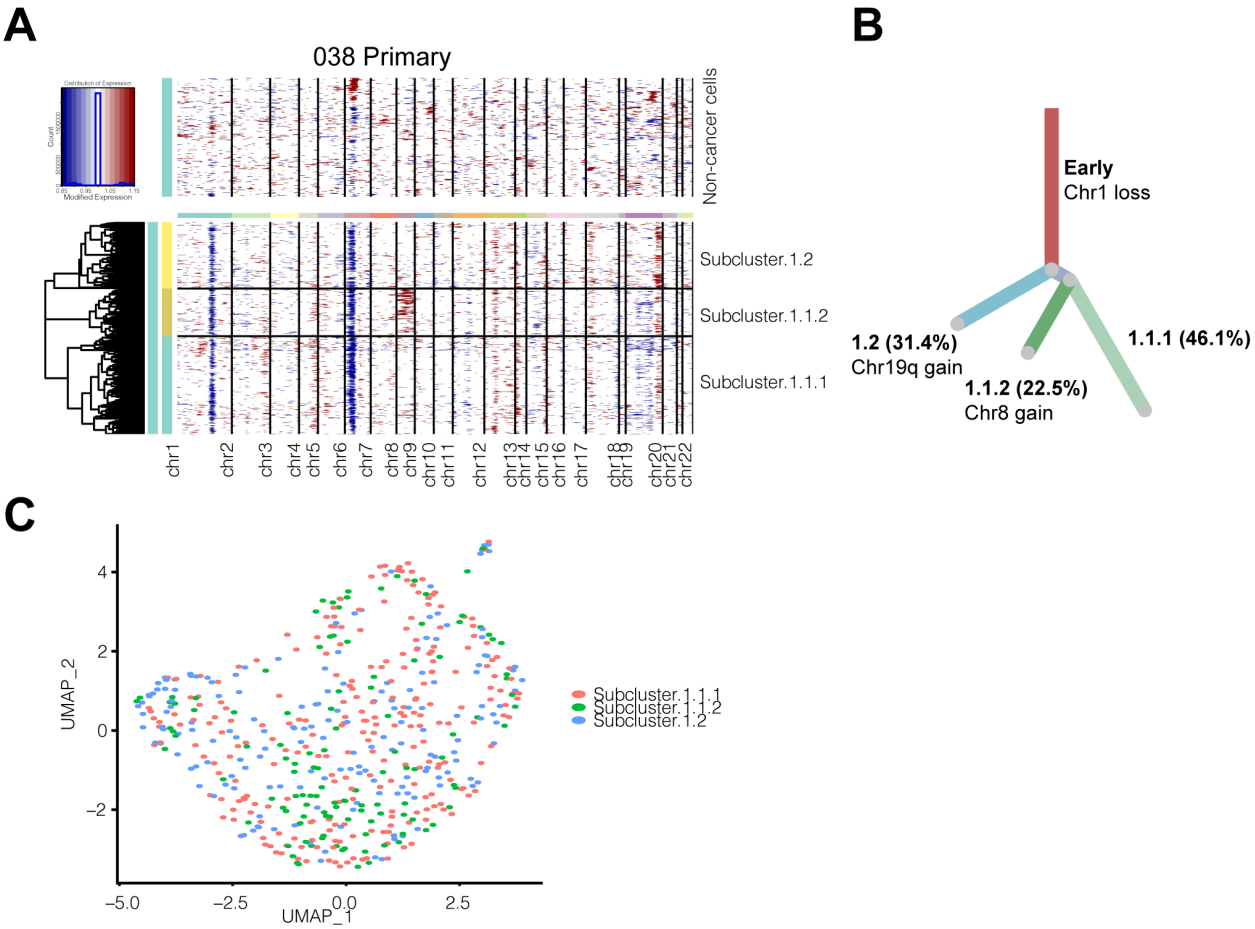

005 Primary
A
Count (10^6)
Modified Expression
Non-cancer cells
Subcluster.1.2
Subcluster.1.1
B
Early
Chr8 gain
1.1 (93%)
Chr1 gain
Chr5 gain
Chr11p loss
1.2 (7%)
Chr1p loss
Chr2 gain
Chr9 gain
Chr16q loss
C
UMAP_2
UMAP_1
Subcluster.1.1
Subcluster.1.2
CD44
D
Expression Level
FN1
E
061 Primary
F
Subcluster.1.2.2
Subcluster.1.2.1
Subcluster.1.1
G
Early
Chr1 loss
Chr2 gain
Chr8 gain
Chr20 gain
1.1 (66.4%)
Chr12p loss
1.2.2 (5.7%)
Chr5 gain
Chr12 gain
1.2.1 (28%)
Chr12 gain
H
IGF1
I
VIM
J

Figure 4. Subclones were identified in Ewing Sarcoma primary tumors by analysis of copy-number alterations.
Two Ewing sarcoma primary tumors displaying subclones: 005 (A-E) and 061 (F-J). Inferred copy number alterations (CNAs) in subclones of each sample (A,F). Phylogenic trees inferring evolution of the subclones based on the CNAs with the percentage of cells in each arm indicated (B,G). Subclones displayed as UMAP (C,H). Select gene expression displayed as UMAP and violin plot (D,E,I,J).



Supplementary Figure 4. Subclones were identified in Ewing Sarcoma primary tumor sample 038 by analysis of copy-number alterations.
Inferred copy number alterations (CNAs) in subclones of sample 038 (A). Phylogenic tree inferring evolution of the subclones based on the CNAs with the percentage of cells in each arm indicated (B). Subclones displayed as UMAP (C).

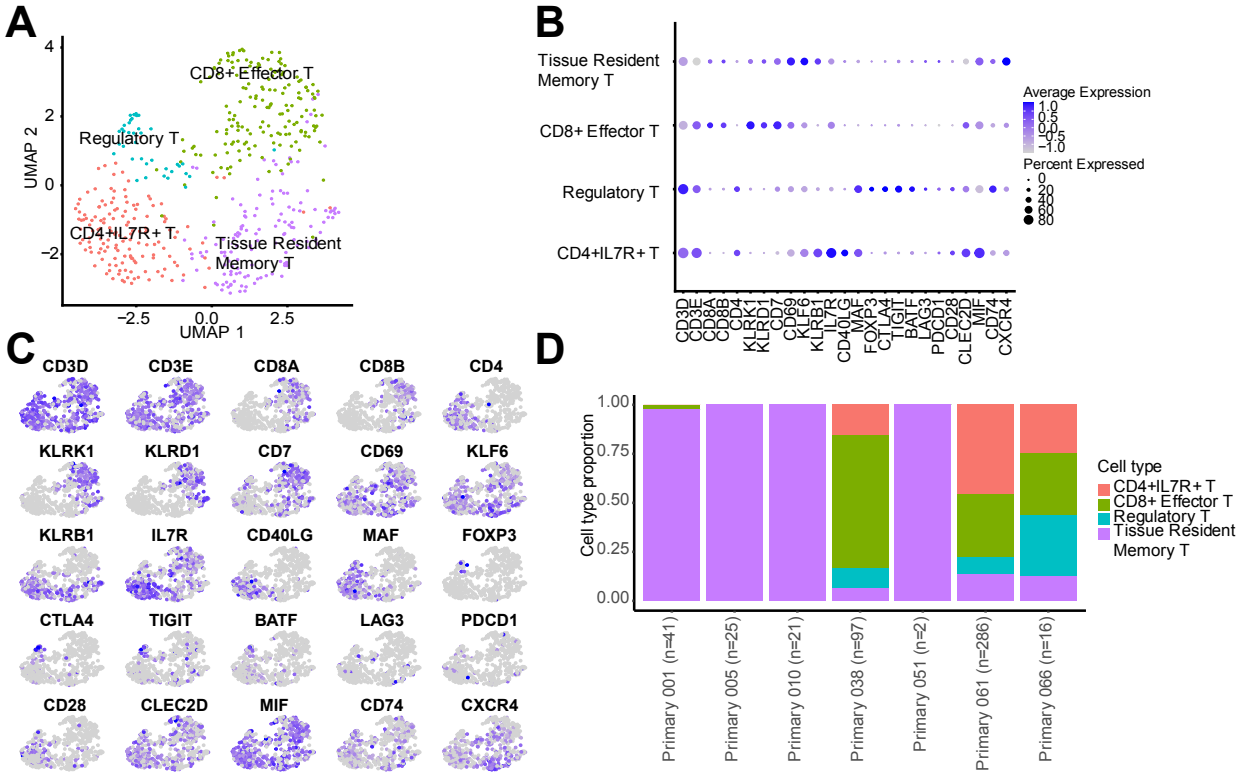


Figure 5. T cell subtypes in Ewing Sarcoma primary tumors.
UMAP showing the four identified T cell subtypes (A). Gene expression defining the subtypes displayed as a dotplot (B) or UMAP (C). Distribution of the subtypes across each sample with the number of total T cells indicated (D).

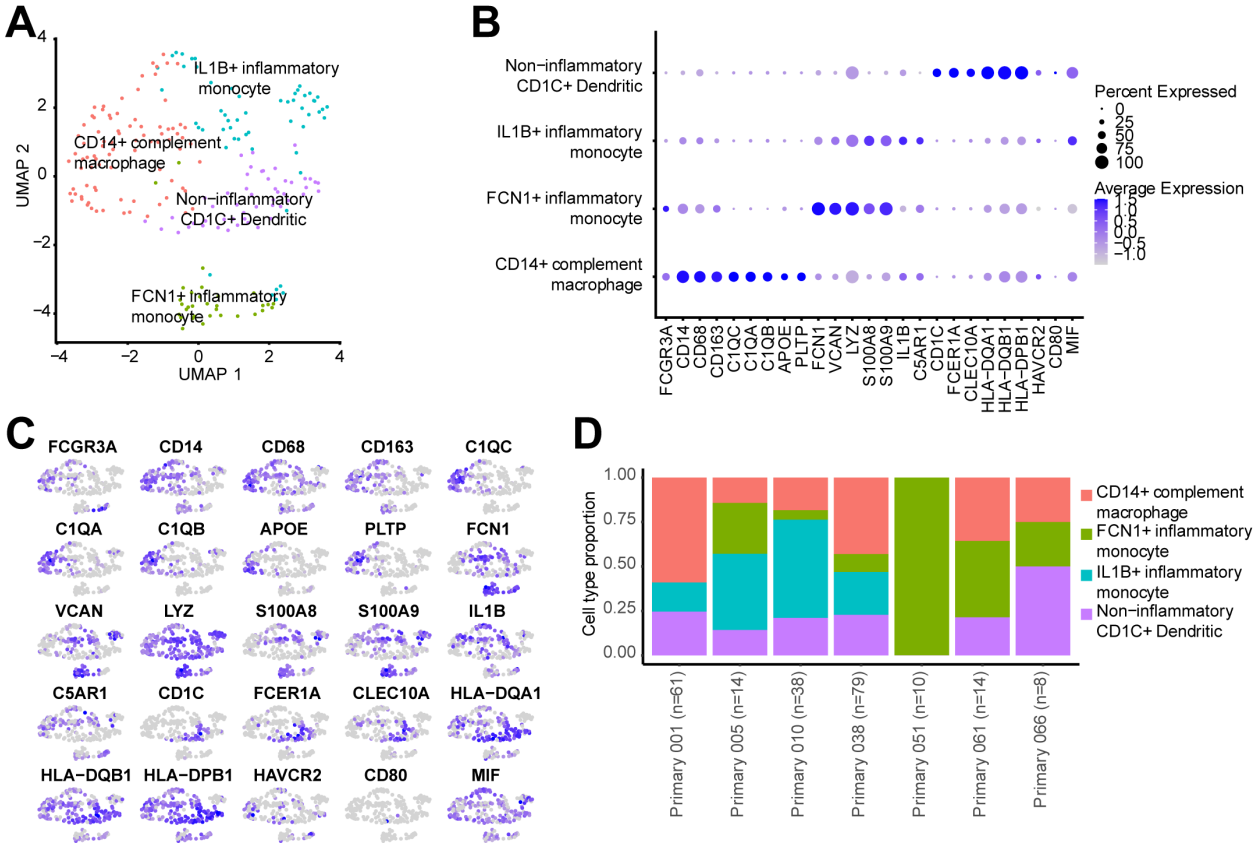


Figure 6. Myeloid subtypes in Ewing Sarcoma primary tumors.
UMAP showing the four identified myeloid subtypes (A). Gene expression defining the subtypes displayed as a dotplot (B) or UMAP (C). Distribution of the subtypes across each sample with the number of total myeloid cells indicated (D).

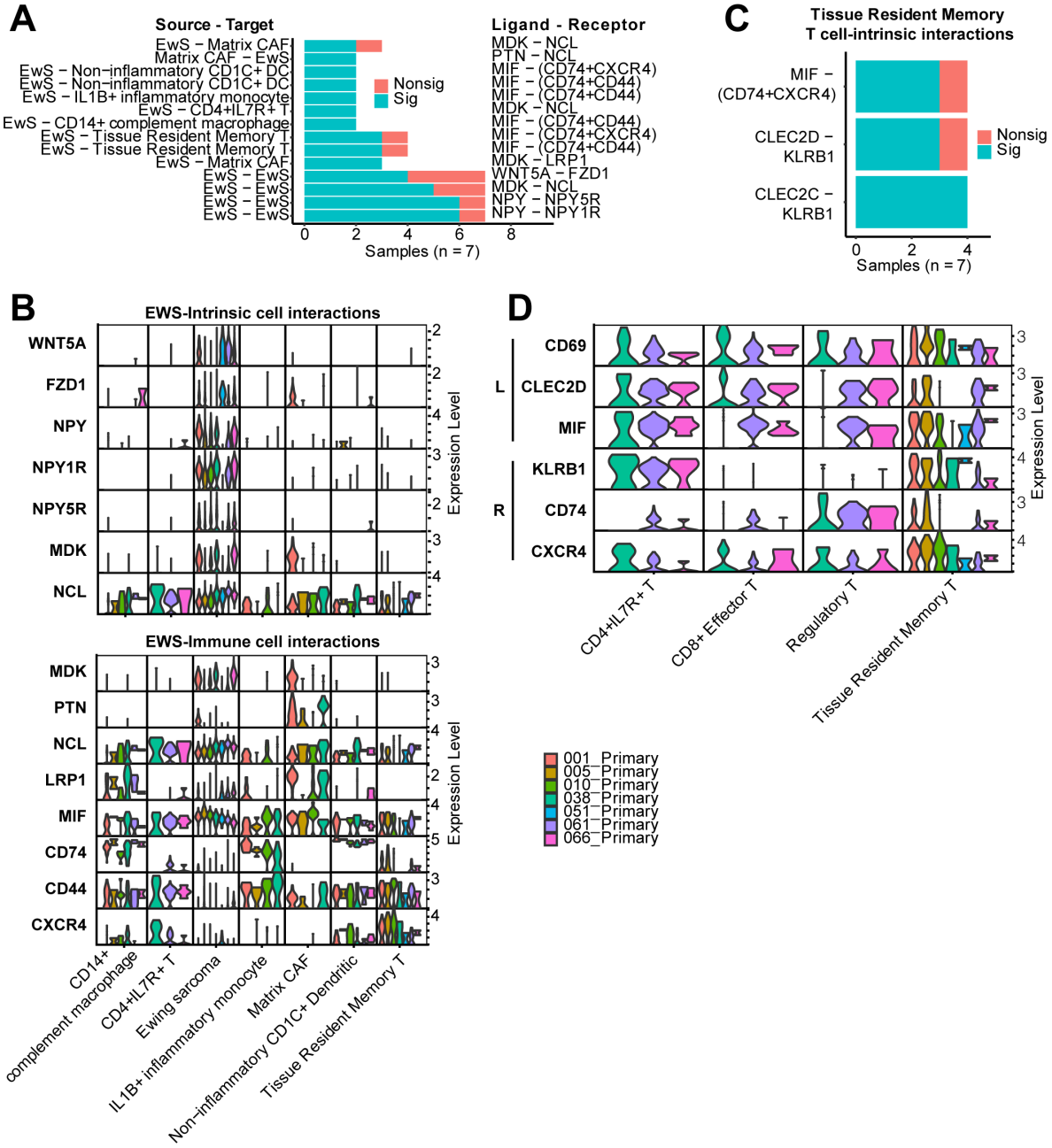


Figure 7. Potential cell-cell communication within Ewing Sarcoma primary tumors.

Select interactions involving Ewing sarcoma cells are highlighted with the number of samples where the specific interaction was significant (blue) or nonsignificant (red) (A). Not all interactions could be tested in each sample because a minimum of 10 cells were required. Expression of ligands and receptors for each sample across the cell types are plotted for Ewing sarcoma-intrinsic (top) or Ewing-immune cell interactions (bottom) (B). Significant (blue) or nonsignificant (red) interactions within immune cells found in at least 3 samples (C). Ligand and receptor expression within T cell subtypes by patient sample (D).

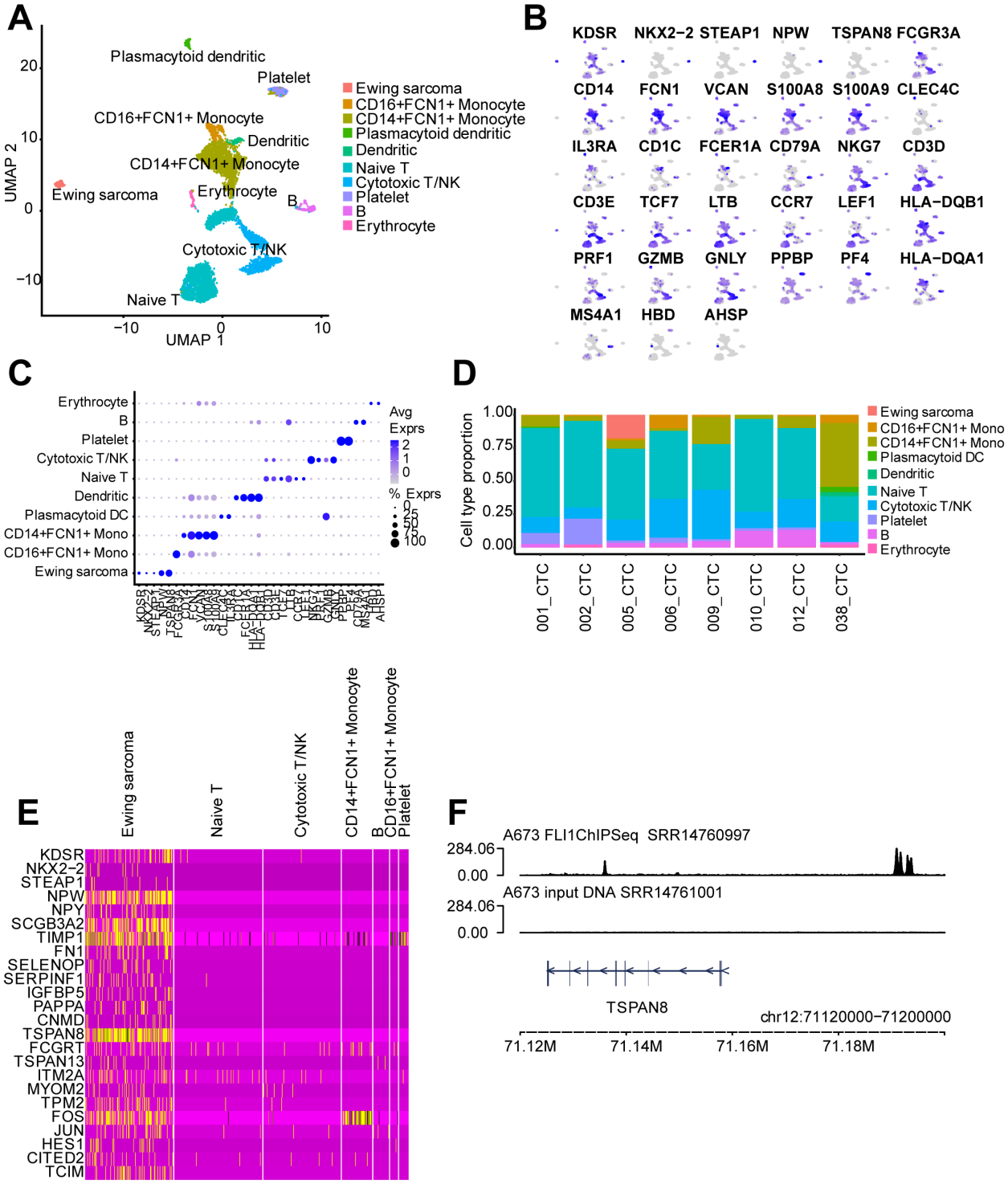
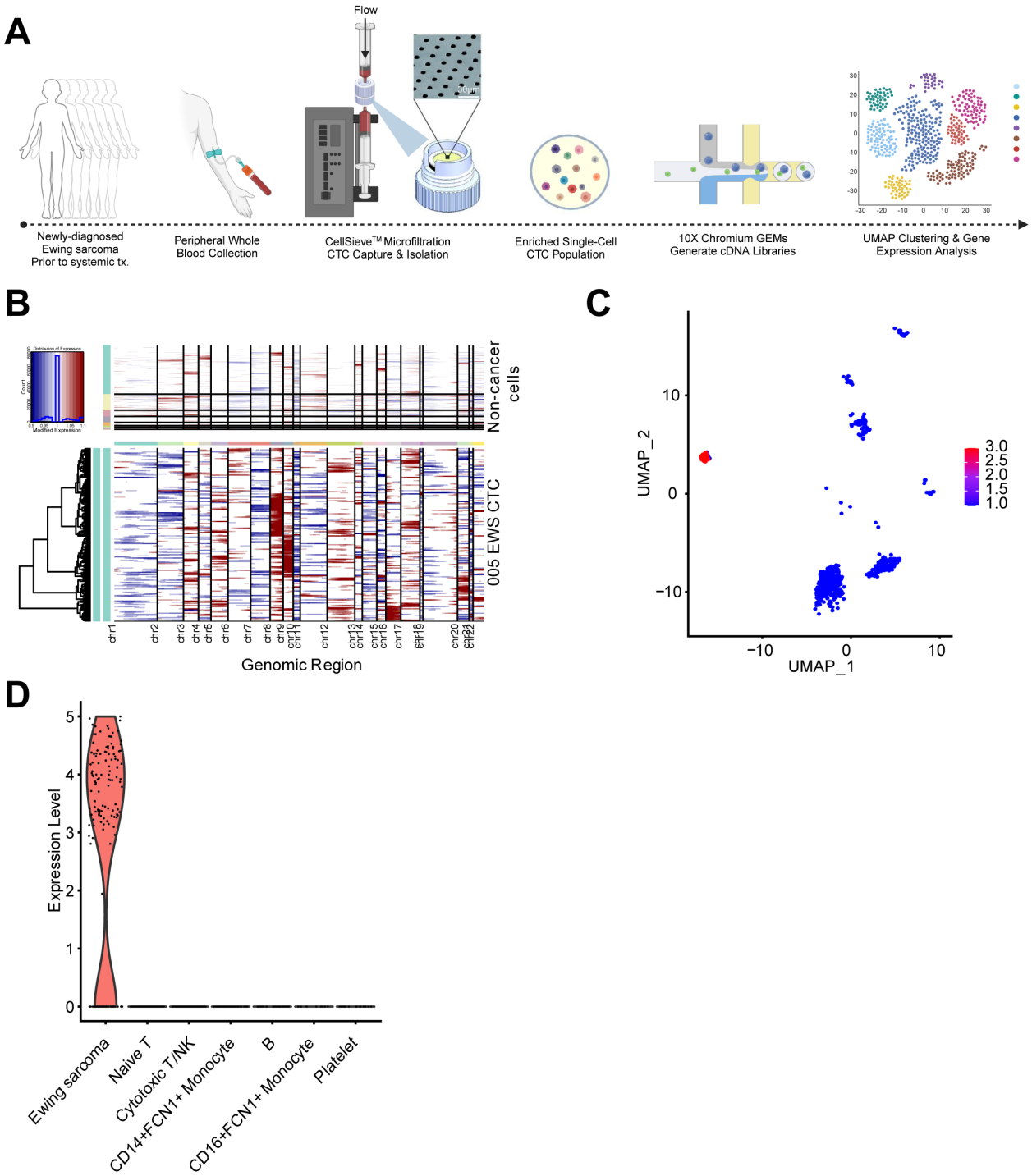
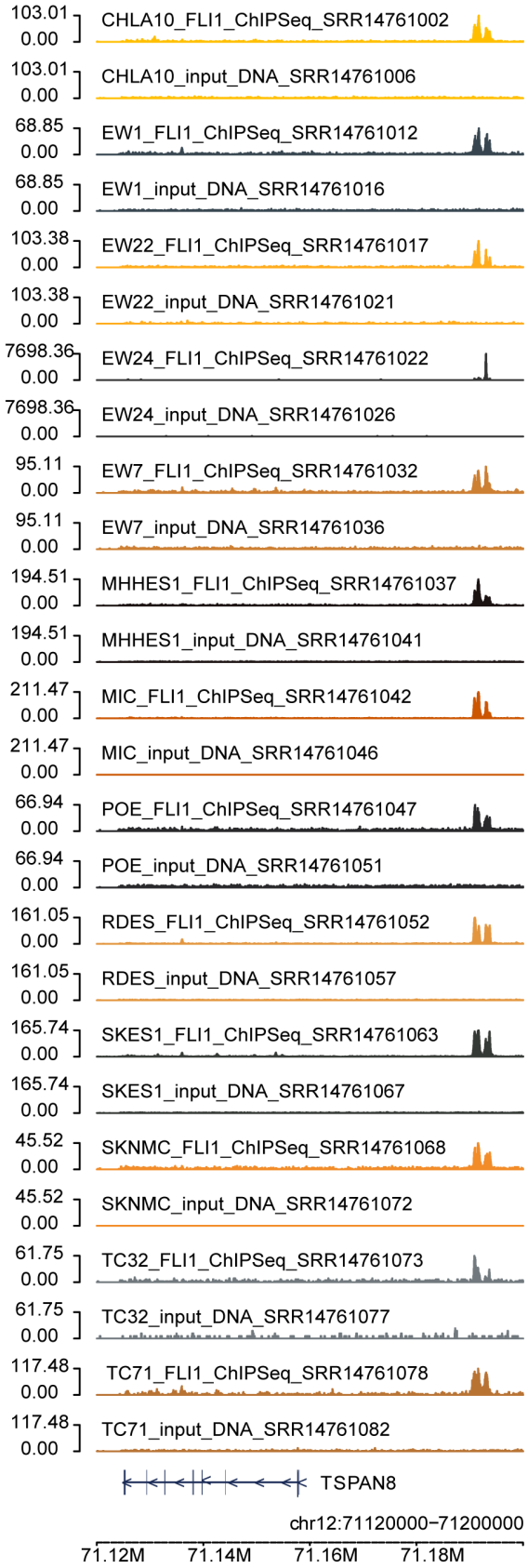


Figure 8. Circulating Ewing Sarcoma cells identified in the blood of Patient 005. Cell types identified in the blood samples displayed as a UMAP (A). Gene expression discriminating the cell types shown as UMAP (B) and dotplot (C) including all samples. Distribution of cell type across sample (D). Gene expression of Ewing-enriched genes displayed as a heatmap including only cells from Sample 005 (E). Binding profile of FLI1 or input DNA near TSPAN8 in the A673 cell line (F).



Supplementary Figure 5. Inferred copy-number alterations and TSPAN8 expression in blood.
Workflow of blood sample collection, enrichment of circulating tumor cells, and processing for scRNA-seq (A). Inferred copy-number alterations of sample 005 Ewing sarcoma cells compared to normal cells (B). TSPAN8 expression in sample 005 as UMAP (C) and violin plot by cell type (D).

A


103.01
0.00
CHLA10_FLI1_ChIPSeq_SRR14761002
103.01
0.00
CHLA10_input_DNA_SRR14761006
68.85
0.00
EW1_FLI1_ChIPSeq_SRR14761012
68.85
0.00
EW1_input_DNA_SRR14761016
103.38
0.00
EW22_FLI1_ChIPSeq_SRR14761017
103.38
0.00
EW22_input_DNA_SRR14761021
7698.36
0.00
EW24_FLI1_ChIPSeq_SRR14761022
7698.36
0.00
EW24_input_DNA_SRR14761026
95.11
0.00
EW7_FLI1_ChIPSeq_SRR14761032
95.11
0.00
EW7_input_DNA_SRR14761036
194.51
0.00
MHHES1_FLI1_ChIPSeq_SRR14761037
194.51
0.00
MHHES1_input_DNA_SRR14761041
211.47
0.00
MIC_FLI1_ChIPSeq_SRR14761042
211.47
0.00
MIC_input_DNA_SRR14761046
66.94
0.00
POE_FLI1_ChIPSeq_SRR14761047
66.94
0.00
POE_input_DNA_SRR14761051
161.05
0.00
RDES_FLI1_ChIPSeq_SRR14761052
161.05
0.00
RDES_input_DNA_SRR14761057
165.74
0.00
SKES1_FLI1_ChIPSeq_SRR14761063
165.74
0.00
SKES1_input_DNA_SRR14761067
45.52
0.00
SKNMC_FLI1_ChIPSeq_SRR14761068
45.52
0.00
SKNMC_input_DNA_SRR14761072
61.75
0.00
TC32_FLI1_ChIPSeq_SRR14761073
61.75
0.00
TC32_input_DNA_SRR14761077
117.48
0.00
TC71_FLI1_ChIPSeq_SRR14761078
117.48
0.00
TC71_input_DNA_SRR14761082
TSPAN8
chr12:71120000–71200000
71.12M
71.14M
71.16M
71.18M

Supplementary Figure 6. FLI1 ChIP-seq binding near TSPAN8.
Binding profile of FLI1 or input DNA near TSPAN8 in several Ewing sarcoma cell lines (A).

Supplementary Tables

Supplementary Table 1. Sequencing information of each scRNA-seq sample.

Supplementary Table 2. Significant gene enrichment by broad cell type in primary tumors.

Supplementary Table 3. Genes of each sample-level gene expression program.

Supplementary Table 4. Genes of each merged gene expression program.

Supplementary Table 5. Significant gene enrichment by T cell subtype in primary tumor.

Supplementary Table 6. Significant gene enrichment by myeloid cell subtype in primary tumors.

Supplementary Table 7. Significnat cell communication interactions involving Ewing Sarcoma cells in primary tumors.

Supplementary Table 8. Significant gene enrichment by cell type in blood samples.