

Reviewed Preprint

v1 • December 5, 2025

Not revised

Reviewed Preprint

v2 • July 27, 2026

Revised by authors

Multimodal single-cell analyses reveal distinct fusion-regulated transcriptional programs in Ewing sarcoma

✉ For correspondence:

sfurlan@uw.edu

† These authors contributed equally

†† These authors contributed equally

Competing interests:

No competing interests declared

Funding:

See [page 18](#)

Reviewing editor:

Lynne-Marie M Postovit, Queens University, Canada

© 2025, Waltner et al. This article is

distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted

use and redistribution provided that

the original author and source are

credited.

Olivia G Waltner¹, April A Apfelbaum^{2,†}, Emma D Wrenn^{2,†}, Shruti S Bhise¹, Sami B Kanaan¹, Rula Green Gladden¹, Mark A Mendoza¹, Roger Volden³, Zev Kronenberg³, Anand Patel⁴, Michael Dyer⁴, Jay F Sarthy^{2,5}, Elizabeth R Lawlor^{2,5,††}, Scott N Furlan^{1,5,††} ✉

¹Translational Science and Therapeutics Division, Fred Hutch Cancer Center, Seattle, United States • ²Ben Towne Center for Childhood Cancer and Blood Disorders Research, Seattle Children's Research Institute, Seattle, United States •

³Pacific Biosciences, Menlo Park, United States • ⁴St. Jude Children's Research Hospital, Memphis, United States •

⁵Department of Pediatrics, University of Washington, Seattle, United States

eLife Assessment

This **valuable** study presents an analysis of the gene regulatory networks that contribute to tumour heterogeneity and tumor plasticity in Ewing sarcoma, with key implications for other fusion-driven sarcomas. The authors employed **compelling** orthogonal approaches, including single-cell sequencing and xenografts, to reveal the existence and plasticity of specific gene regulatory networks (e.g., TGF-beta signaling) within Ewing sarcoma, as well as significant differences that exist between cell lines and patient tumors.

<https://doi.org/10.7554/eLife.109124.2.sa3>

Abstract

Ewing sarcoma (EwS) is a fusion-driven malignancy, peaking in adolescence. Although EwS tumors are driven uniquely by EWS::FLI1 and related fusions, patient outcomes vary greatly. If and how tumor plasticity of EWS::FLI1-regulated transcriptional signatures contribute to disease progression is not known. To address this, we utilized a single-cell co-assay of RNA and chromatin accessibility (ATAC) sequencing to identify gene regulatory networks in EwS. By comprehensively characterizing regulatory elements across cell lines, we identified multiple unique modules of gene regulation. Differential usage and prevalence of these modules was evident across cell lines, associated with distinct epigenetic and transcriptomic signatures, and in specific cases, modifiable by exogenous TGF- β . When we examined primary EwS patient tumors, we observed these same regulatory modules were variably enriched both across and within tumors, highlighting the existence of intratumoral heterogeneity in gene regulatory networks. Our findings demonstrate that multiple, co-existing transcriptional programs shape the phenotypic diversity of EwS and suggest that the balance between these networks may have important implications for clinical outcomes and targeted therapy development.

Summary

Multimodal transcriptional analysis reveal how Ewing sarcoma tumors use distinct gene programs, including one linked to TGF- β , to drive cancer behavior and progression.

Introduction

Ewing sarcoma (EwS) is an aggressive, fusion-driven bone and soft tissue tumor encountered most in adolescents and young adults. Over a third of patients, including nearly all with metastatic or recurrent disease, do not survive^{1,2}. Clinical presentation of EwS is heterogeneous, with 80% of localized cases presenting in the bone, and 20% arising in extraosseous locations. Histologically, EwS tumors are highly undifferentiated, exhibiting phenotypic features of both mesenchymal and neural crest progenitors. Although the cell of origin is controversial, EwS tumors are presumed to be derived from these lineages^{3–6}. Chromosomal translocations that create fusions between the N-terminal domain of *EWSR1* and the DNA-binding domain of ETS transcription factor (TF) family members are the primary oncogenic lesion in EwS¹. The resulting EWS::ETS chimeric oncoprotein reprograms gene regulatory networks through diverse mechanisms, including modifying chromatin state through its function as a pioneer factor at GGAA microsatellite repeats, creating *de novo* enhancers^{7,8}. Apart from their fusions, EwS tumors are relatively bereft of mutations^{1,9}. Neither additional genetic alterations, nor differences in fusion breakpoint and binding partner heterogeneity¹⁰ adequately explain the diverse clinical outcomes that are observed in patients who receive identical standard-of-care therapies. Advances in transcriptomic and epigenomic analyses have in recent years demonstrated the profound contributions of tumor cell plasticity and cell state heterogeneity to tumor metastasis, relapse, and therapy resistance^{11,12}. EwS tumors demonstrate marked inter- and intra-tumoral heterogeneity in their epigenetic states^{13,14} and, as such, these differences may be key determinants of disease progression and clinical outcome.

Recent studies by our group and others have used single cell genomics to study cellular heterogeneity within EwS tumors^{14–23}. These studies reveal that EwS tumors coopt distinct developmental programs to promote proliferation or migratory behavior^{15,16,18,20,21,24,25}. In addition, accruing evidence shows that cooperativity between TFs, EWS::FLI1, and diverse chromatin states may underlie these various oncogenic programs^{20,26–34}. We hypothesized that joint single-cell transcriptional and chromatin profiling technology would reveal new insight into the cell-intrinsic epigenetic and transcriptional heterogeneity of EwS cell states. Using this approach, we have identified a set of diverse gene regulatory networks that are differentially influenced by the EWS::FLI1 fusion and TGF- β in the tumor microenvironment. These distinctive networks vary both across and within tumors from EwS patients providing evidence that a diversity of gene regulatory programs may contribute to the heterogeneity of this disease.

Results

Multiomic analyses reveal distinct chromatin accessibility states across EwS cell lines

Cell lines are one of few available models used to interrogate EwS biology given the rarity of the disease and the lack of appropriately faithful animal models³⁵. One drawback of cell line use is potential loss of intra-tumoral heterogeneity compared to the original tumor due to serial passaging. We reasoned that profiling an array of EwS cell lines would capture the diversity of gene regulatory networks that may underlie the clinical and transcriptional heterogeneity of this disease. Because of the well-described direct influence of the EwS fusion on chromatin remodeling³⁴, we utilized an assay capable of pairing single-cell RNA transcript counts with a chromatin assay for transposase-accessible chromatin (ATAC) (herein referred to as ‘multiomic’ sequencing) on 9 EwS and 5 non-EwS cell lines (Fig.1A [↗](#), Fig.S1A [↗](#)). We recovered 30,105 profiles after filtering for poor quality cells (Fig.1A [↗](#), Fig.S1A [↗](#), Fig.S1B [↗](#)). Multiomic single-cell profiles self-assembled after dimensionality reduction according to cell line identity (Fig.1A [↗](#)).

First, to identify the extent of transcriptional heterogeneity of validated direct EWS::FLI1 target genes¹⁹ (EwS TGs) in individual cells, we calculated expression of EwS TGs at the single-cell level, subtracted by the aggregated expression of control gene sets³⁶. As expected, RNA expression of these EwS TGs was higher in EwS compared to non-EwS cell lines (Fig. S1C [↗](#)). As these published EwS TGs were experimentally derived from the A673 cell line¹⁹, it was unsurprising that A673

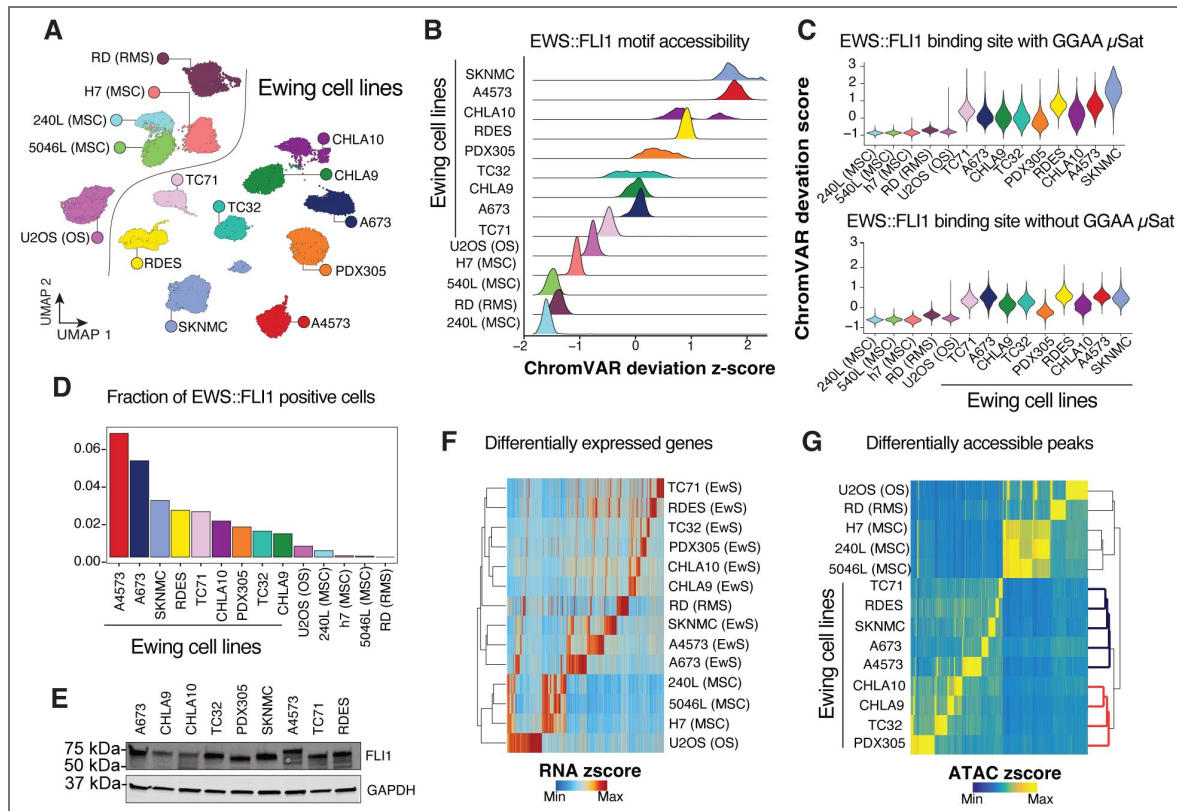


Fig. 1. Multiomic analyses reveal distinct signatures between non-EwS and EwS cells.

(A) UMAP embedding of single cell RNA sequencing and ATAC sequencing of EwS and non EwS cells. **(B)** ChromVAR deviation scores of EWS::FLI1 motif enrichment across EwS and non-EwS cells. **(C)** ChromVAR deviation scores of EWS::FLI1 binding site enrichment, separated by regions with or without GGAA microsatellites (μ Sat; GGAA repeat ≥ 4), grouped by cell line. **(D)** Fraction of cells in each cell line with detection of EWS::FLI1 fusion transcript. **(E)** Western blot of FLI1 across EwS cell lines. GAPDH = loading control. **(F)** Differentially expressed genes ($FDR \leq 0.01$ & $\text{Log}_2FC \geq 1.25$) across EwS/non-EwS landscape. **(G)** Differentially accessible peaks ($FDR \leq 0.01$ & $\text{Log}_2FC \geq 1.25$) of EwS and non EwS cells with row dendrogram showing phenotypic clusters. Row dendrogram shows EwS cell clusters highlighted in bold and red.

exhibited the highest average expression of these genes. To interrogate EwS TGs using chromatin-accessibility data, we generated gene activity scores³⁷ which model accessibility of regulatory elements in the vicinity of a given gene. These scores confirmed that chromatin accessibility of EwS TGs was higher in EwS cell lines compared to non-EwS cells (Fig. S1D [↗](#)). To compare EWS::FLI1 motif enrichment across cells, we calculated single-cell ChromVAR deviation scores³⁷ of GGAA-microsatellites (four or more consecutive GGAA repeats; abbrev. ‘GGAA- μ Sat’) in accessible regions. As anticipated, this analysis revealed low enrichment in non-EwS cell lines. Unexpectedly, stark differences in EWS::FLI1 motif enrichment were observed across EwS cell lines (Fig. 1B [↗](#)). In some cell lines accessibility of GGAA- μ Sats was similar across single cells (e.g. TC71, RDES) while others showed a broader range of accessibility (e.g. SKNMC, TC32, PDX305). CHLA10 cells showed a clear bimodal distribution of EWS::FLI1 motif accessibility (Fig. 1B [↗](#)). This hints that within these cell lines, heterogeneity exists among individual cells with respect to accessibility of GGAA- μ Sats. Next, we collated a list of publicly available EWS::FLI1 fusion binding sites³⁸ with or without a GGAA μ Sat present. We then assessed chromatin accessibility at these published EWS::FLI1 binding sites across all cell lines and discovered that accessibility at these loci varied in a cell-line-specific manner (Fig. 1C [↗](#) & Fig. S1D [↗](#)).

Given the heterogeneity in the magnitude of EWS::FLI1 motif enrichment across EwS cell lines, we hypothesized that EwS cell lines differ in the proportion of cells expressing high levels of the EWS::FLI1 fusion transcript. It has been reported that lower levels of EWS::FLI1 gene expression and activity induce mesenchymal phenotypes in EwS cell lines³⁹. To address this, we designed an approach to enrich EWS::FLI1 fusion transcript from the RNA components of our multiomic capture using biotinylated hybridization probes that tiled the EWSR1 and FLI1 gene loci. We then performed long-read sequencing on the product and counted fusion reads using novel fusion detection software designed specifically for long-reads (see Methods for details). As expected, EWS::FLI1 fusion transcripts were detected in EWS cells (Fig. 1D [↗](#)). Scant non-ES cells showed low levels of the fusion, likely from PCR artifacts of low-complexity products. Overall, however, we observed that EwS cell lines exhibit a high degree of variability in detectable EWS::FLI1 fusion transcript (Fig. 1D [↗](#) & Fig. S1E [↗](#)), possibly reflecting a difference in expression levels among individual cells. Generally, the proportion of cells with detectable transcript correlated with values obtained using single-cell qPCR to assess fusion expression¹⁵. These findings are generally congruent with motif enrichment in that high fusion-expressing cell lines (e.g. A4573, SKNMC, and RDES) exhibit high EWS::FLI1 motif enrichment while cell lines bearing lower levels of fusion transcript and FLI1 product (e.g. PDX305, TC32, CHLA9) exhibit lower EWS::FLI1 chromatin signature (Fig. 1B [↗](#) & Fig. 1D-E [↗](#)). By contrast, other cell lines displayed inverse trends of fusion transcript and product to chromatin enrichment (e.g. A673 and TC71).

To profile the global chromatin and transcriptional heterogeneity using an unbiased approach, we calculated top differentially expressed (DE) genes and differentially accessible (DA) regions across EwS and non EwS cells. As with the EwS TG-focused analysis, this approach showed that top DE/DA features readily distinguished EwS cells from non-EwS cells (Fig. 1F-G [↗](#)). We observed self-aggregation of EwS cells into distinct clusters. Specifically, in terms of top DA regions, CHLA9, CHLA10, TC32, and PDX305 cells were more similar to each other than other EwS cell lines (Fig. 1G [↗](#)). Altogether, these data show that EwS cell lines exhibit heterogeneous gene-regulatory profiles that segregate into groups that display distinct trends in accessibility, potentially reflecting unique underlying gene regulatory networks that are active in discrete cell contexts.

EwS cells utilize gene regulatory programs that drive inter-tumoral heterogeneity

Next, to evaluate how cis-regulatory gene networks differ across EwS cells, we correlated transcript expression to regions of accessibility using the joint measurements of transcript count and accessibility in each cell (so-called “peak-to-gene linkages”)³⁷. Hierarchical clustering of these cis-regulatory networks with a k parameter equal to 3 clearly separated CHLA9, TC32, and PDX305 into one distinct network module and SKNMC, A4573, RDES, and TC71 into another (Fig. 2A [↗](#)), recapitulating what we observed using top DE/DA features (Fig. 1F-G [↗](#)). We selected $k = 3$ to

extend the two-cluster structure observed in Fig. 1F–G, reasoning that A673 represented a clear outlier whose distinct regulatory program would be obscured at lower k . Cell subsets of CHLA10 separated into either cluster 2 or 3 (Fig. 2A) but more cells had cumulative expression of module 2 genes (Fig. 2B). We then assigned the 9 cell lines into three groups based on their dominant module signature (Fig. 2A–B), resulting in assignment of A673 cells as group 1, CHLA9, CHLA10, TC32, and PDX305 cells as group 2, and SKNMC, A4573, TC71, and RDES as group 3 (Fig. 2A). Enrichment analysis using Gene Ontology and Hallmark terms confirmed that the genes associated with these modules have functional associations with distinct cell behaviors (Fig. 2C). While there was some degree of overlap, module 2 showed a relative enrichment in GO terms related to mesenchymal identity, and modules 1 and 3 had enriched neuronal identity markers. Genes involved in oxidative phosphorylation and targets of mTOR were uniquely enriched in module 1, possibly due to the presence of a BRAF^{V2600E} mutation in A673 cells given the reported effects of this mutation on the mTOR pathway^{40–42}.

Next, we investigated if gene modules might harbor distinct TF binding sites. We performed differential motif enrichment using HOMER⁴³ and found a strikingly well-delineated, and module-specific enrichment of specific TF families across the cis-regulatory networks (Fig. 2D). The AP-1 family of motifs (FOS, FOSL2, JUN, AP1), which has been suggested to form ternary complexes with EWS::FLI1⁴⁴, enriched in modules 1 and 2, but not in module 3 (Fig. 2D). Zinc Finger (ZF) TFs (KLF4, KLF5, MAZ, SP1) were most enriched in modules 1 and 3 with limited evidence for activity in regulation of module 2 (Fig. 2D). Module 2 peaks uniquely exhibited motif enrichment for developmental regulators from the Homeobox family of TFs (HOXD13, HOXA9) and downstream transcriptional regulators of the Hippo and Wnt signal transduction pathways (TEAD1, TEAD4). To corroborate these findings, we performed an analysis which correlates, at the single-cell level, TF binding motifs in accessible regions with gene expression (Fig. 2E). This showed that TEAD1, HOXD13, and AP-1 TFs were highly enriched in group 2 cells (Fig. 2E–F). Group 3 lacked motif enrichment for AP-1 TFs but showed robust enrichment in ALX1/4, RUNX3, CEBPB and EN2 TFs (Fig. 2E–F). We assessed protein levels of selected TF regulators by Western blots to evaluate their heterogeneity across the cell lines. The protein abundance of these TFs exhibited substantial variability among the cells tested (Fig. 2G). Together, these multimodal analyses of single cells uncovered at least two dominant gene regulatory circuits. One of these (module 2) displays mesenchymal/EMT-like functionality that is likely to be mediated by high activity of TEAD, AP-1, and HOXD13 TFs. The other (module 3) has enriched neural program functionality most likely mediated by ZF TFs and neural and neural crest development TFs such as RUNX3 and HAND2.

EwS gene regulatory modules are differentially regulated by EWS::FLI1

We next sought to address whether the identified gene regulatory modules are directly or indirectly modulated by EWS::FLI1. We first measured EWS::FLI1, EWS::ERG (the second most common fusion in EwS tumors), and wild-type FLI1 motif enrichment and found all to be enriched in all three modules (Fig. 3A) suggesting that EWS::FLI1 is likely to regulate expression of linked genes across all cells irrespective of the dominant module. However, by annotating these sites using published EWS::FLI1 binding sites (39) and segregating into those with or without a GGAA microsatellite (GGAA- μ Sat), we observed enrichment of GGAA- μ Sats only in module 3 peaks (Fig. 3B). Thus, this analysis suggests that EWS::FLI1 binding is likely to directly influence gene regulatory circuits and its impact on transcriptional state depends on GGAA patterning, as has previously been suggested⁷.

To begin to address the impact of EWS::FLI1 activity on cis-regulatory modules, we sourced data from published lists of EWS::FLI1 “activated” and “repressed” genes and performed gene set enrichment³³ (Fig. 3C). Overall, EWS::FLI1 activated genes were enriched in all modules, whereas repressed genes were enriched only in module 2 (Fig. 3C). When combined with the lack of GGAA- μ Sat enrichment in module 2 (Fig. 3B), these observations are consistent with previous reports that repressed loci usually do not have long stretches of GGAA repeats^{7,45}. Altogether,

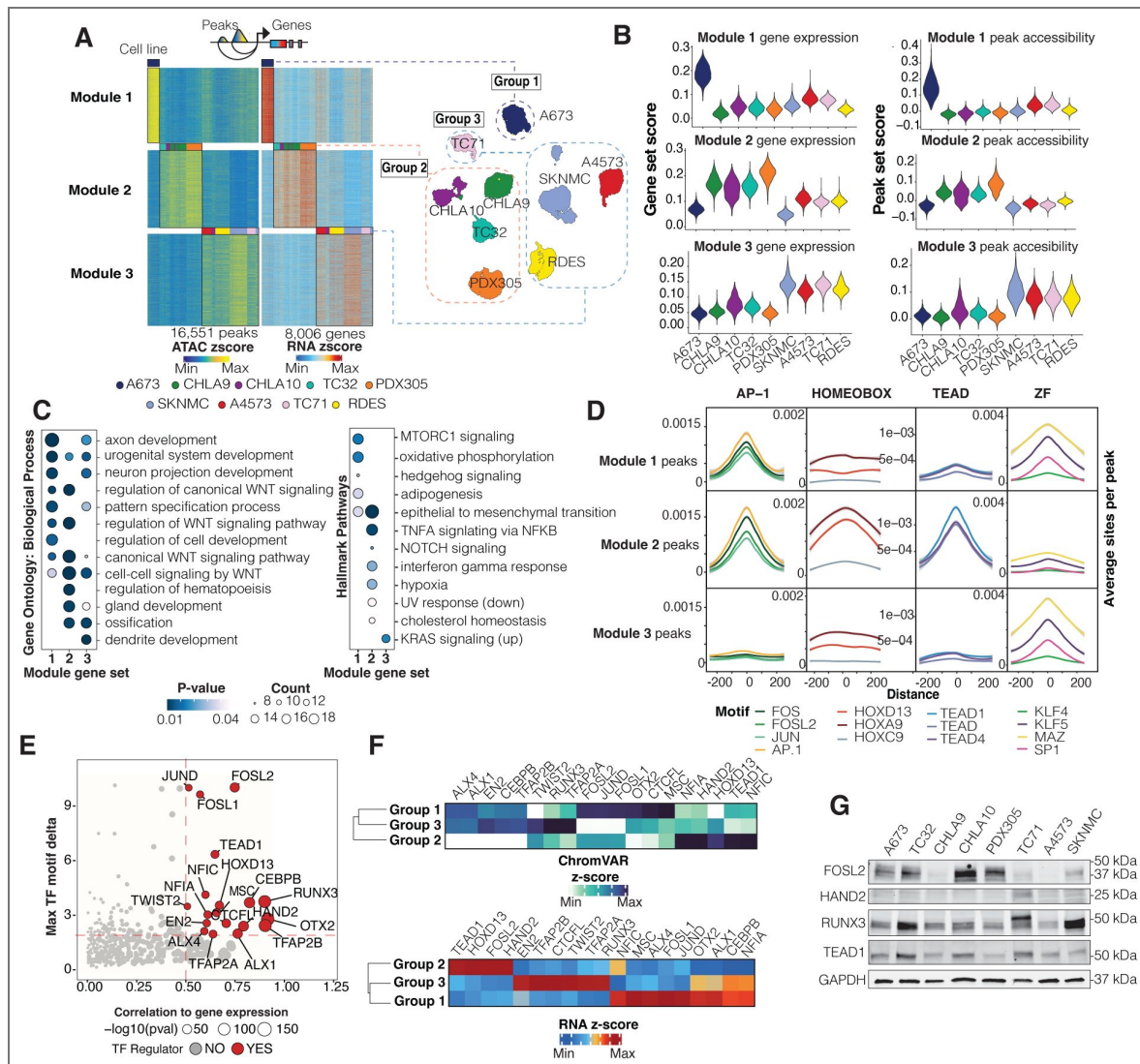


Fig. 2. EwS cells utilize distinct gene regulatory modules

(A) Heatmap of unsupervised kmeans clusters of peak-to-gene links (PGLs) across 9 EwS cell lines. Left- peaks, right-genes, top- colors denote cell lines. Right of heatmap-UMAP of cell lines labeled by subgroup. Group 1 labelled with navy/dashed lines, group 2 in orange and group 3 in light blue. (B) Violin plot of modules 1-3 gene expression and peak accessibility across EwS cell lines. (C) Universal enrichment analysis of module gene (top 300 filtered on peak-to-gene correlation) sets using both MSigDB Hallmark pathways and Gene Ontologies: Biological Process. The top 7 terms filtered by lowest p-value (those greater than 0.05 are not plotted) and count are shown. More significant terms are darker blue. The number of genes in each module that overlap with a gene set term is indicated by the count metric—represented as dot size. (D) HOMER de novo motif enrichment for each module peak set. Results depict the top significantly enriched motifs in each module peak set, and enrichments in other modules. (E) TFs with high expression and motif enrichment in entire single-cell data set of EwS cell lines. Statistically significant TFs correlation to gene expression > 0.5 and a maximum inter EwS subgroup difference in motif deviation zscore (denoted max TF motif delta) > 1.9 highlighted in red. (F) ChromVAR motif deviation scores and gene expression of highly enriched TFs, parsed by EwS subgroup. (G) Western blot of selected TFs across EwS cell lines. GAPDH = loading control.

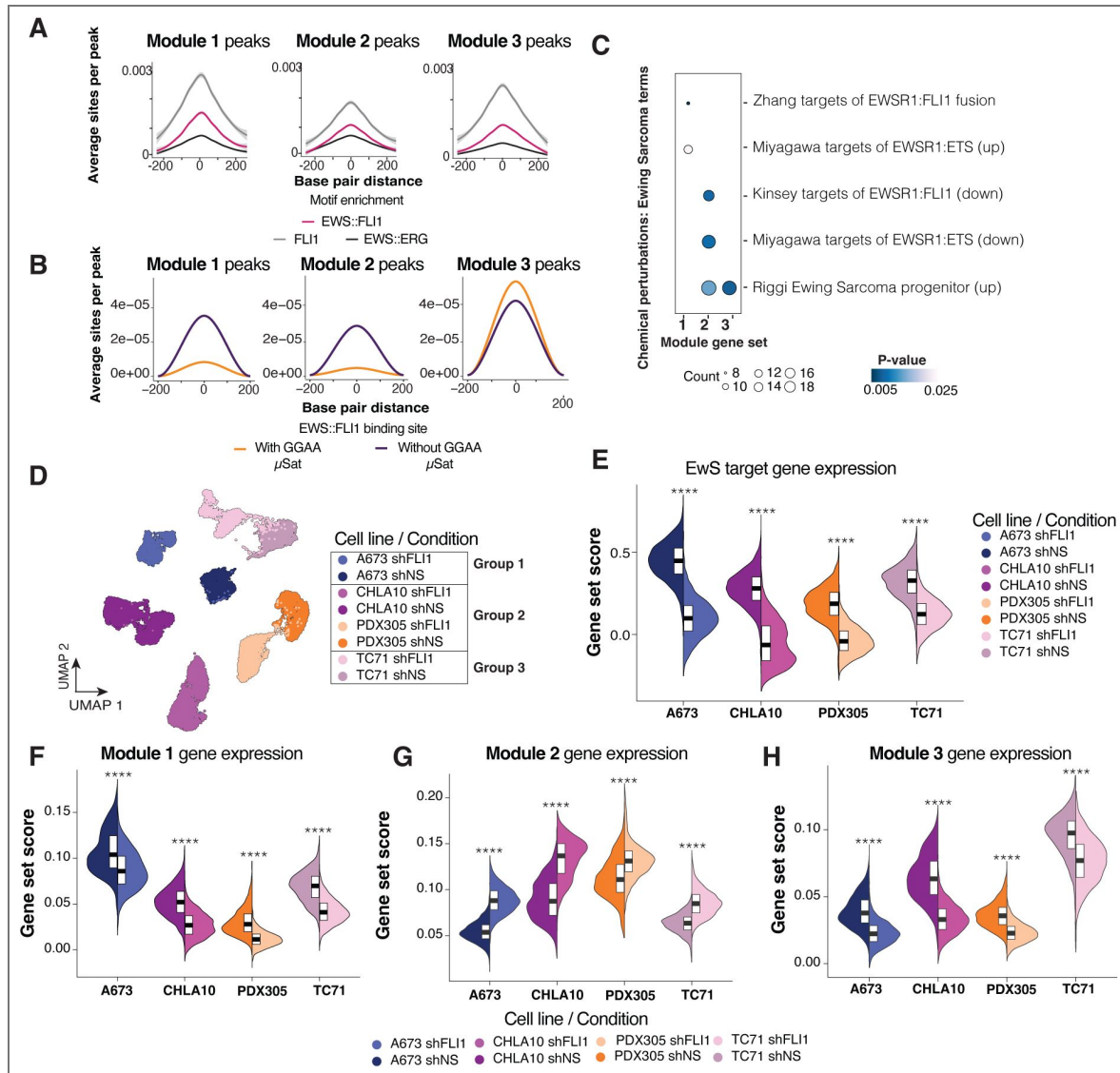


Fig. 3. EwS gene regulatory modules are differentially regulated by EWS::FLI1.

(A) HOMER enrichment of EWS/ETS motifs (EWS::FLI1, pink; FLI::ETS, gray; EWS::ERG, black) in module peak sets. (B) HOMER enrichment of EWS::FLI1 binding sites, separated by regions with (orange) or without (dark purple) GGAA microsatellites. (C) Universal enrichment analysis of MSigDB EWS::FLI1 and Ewing sarcoma terms in module gene sets. (D) UMAP embeddings of shNS and shFLI1 samples. EwS subgroup identity of cell lines represented by dashed boxes. Group 1 (A673) labelled with navy/dashed lines, group 2 (CHLA10, PDX305) in orange and group 3 (TC71) in light blue. (E) Expression of EwS target gene expression across shNS vs. shFLI1 samples. (F-H) Module gene set expression across knockdown samples. **** indicates $p < 1 \times 10^{-4}$ computed by a Wilcoxon Rank Sum test.

these data reveal distinct and polarized fusion-mediated regulatory events across EwS cell lines that differentially depend on GGAA repeats, suggesting that the diversity of fusion function is dependent on the motif landscape at fusion binding sites.

To directly address how different gene modules are affected by EWS::FLI1, we performed single-cell multiomic sequencing on selected cell lines from each EwS subgroup after shRNA-mediated knockdown of the fusion (Fig. 3D). Following quality control (Fig. S2A), we recovered 11,614 cells expressing negative control shRNA (A673 shNS, CHLA10 shNS, PDX305 shNS, TC71 shNS) and 12,466 cells expressing shRNA specific to the FLI1 transcript (A673 shFLI1, CHLA10 shFLI1, PDX305 shFLI1, TC71 shFLI1). All shFLI1 cells showed decreased expression of EwS TGs compared to shNS cells (Fig. 3E), providing evidence that the knockdown effectively induced loss of EWS::FLI1 activity. EWS::FLI1 knockdown had specific, predictable, and reproducible effects on all cell lines. Importantly the same stratification of shNS cell lines into groups with specific module enrichment was observed, indicating that lentiviral transduction and selection did not alter dominant gene programs (Fig. 3F-H, Fig. S2B-D). In addition, each module was affected by EWS::FLI1 knockdown in a manner consistent with predictions based on module gene sets. Specifically, EWS::FLI1 knockdown resulted in decreased expression of module 1 and 3 genes (Fig. 3F,H, Fig. S2B,D) and increased expression of module 2 genes (Fig. 3G, Fig. S2C). These data confirm that EWS::FLI1 activates cis-regulatory networks found in modules 1 and 3 but represses module 2. The results also support our initial hypothesis that EWS::FLI1-regulated gene modules differ among commonly used EwS cell lines. They further show that specific and reproducible patterns of transcriptomic and phenotypic signatures dominate each cell line and these predictably respond to fusion knockdown.

CHLA10 exhibits intra-tumoral heterogeneity of gene regulatory modules

Most EwS cell lines in this study are dominated by a single gene-regulatory program (modules 1, 2, or 3). A notable exception is CHLA10, which contains mutually exclusive sub-programs defined by differences in both EWS::FLI1 motif activity and module utilization (Fig. 1B, Fig. 2A). This distinction is particularly striking because CHLA10 and CHLA9 originated from tumors from the same patient, with CHLA10 being derived post-therapy and following disease relapse. To further investigate this intratumoral heterogeneity, we performed unsupervised Louvain clustering⁴⁶ on EwS cell lines (Fig. S3A) and computed the Pearson correlation coefficient of the most highly variable regions of accessibility. As expected, CHLA9, TC32, and PDX305 cells grouped together based on similarity (Fig. S3B). In contrast, CHLA10 segregated into two clusters (C22 & C23) that each showed greater transcriptional similarity to other cell lines than to each another (Fig. 4A, Fig. S3B), highlighting the extent of its internal heterogeneity. Unsurprisingly, of all cell lines, CHLA10 C23 cells exhibited the greatest correlation to profiles from CHLA9.

Cell-cycle effects contributed to but could not fully explain CHLA10 heterogeneity (Fig. S3C-D). Instead, plotting peak accessibility and transcript abundance of each module revealed that C22 cells enriched for module 3 while C23 cells relied on module 2 (Fig. 4B). Motif enrichment of differentially accessible peaks further underscored the nature of the split: RUNX motifs were enriched in C22 cells whereas AP-1, TEAD, and HOX TF motifs predominated in C23 cells (Fig. 4C). Gene-set enrichment mirrored these findings, C22 expressed neural programs whereas C23 cells were enriched for EMT, migration, and TGF- β signaling gene sets (Fig. 4D-E). Pathway-activity inference analysis between cell clusters⁴⁷ further implicated TGF- β as a dominant upstream driver in C23 (Fig. 4F).

Collectively, the isogenic model CHLA9/CHLA10 encapsulates the spectrum of transcriptional states seen across the other EwS cell lines and pinpoints TGF- β signaling as key mediator of the module 2 transcriptional state.

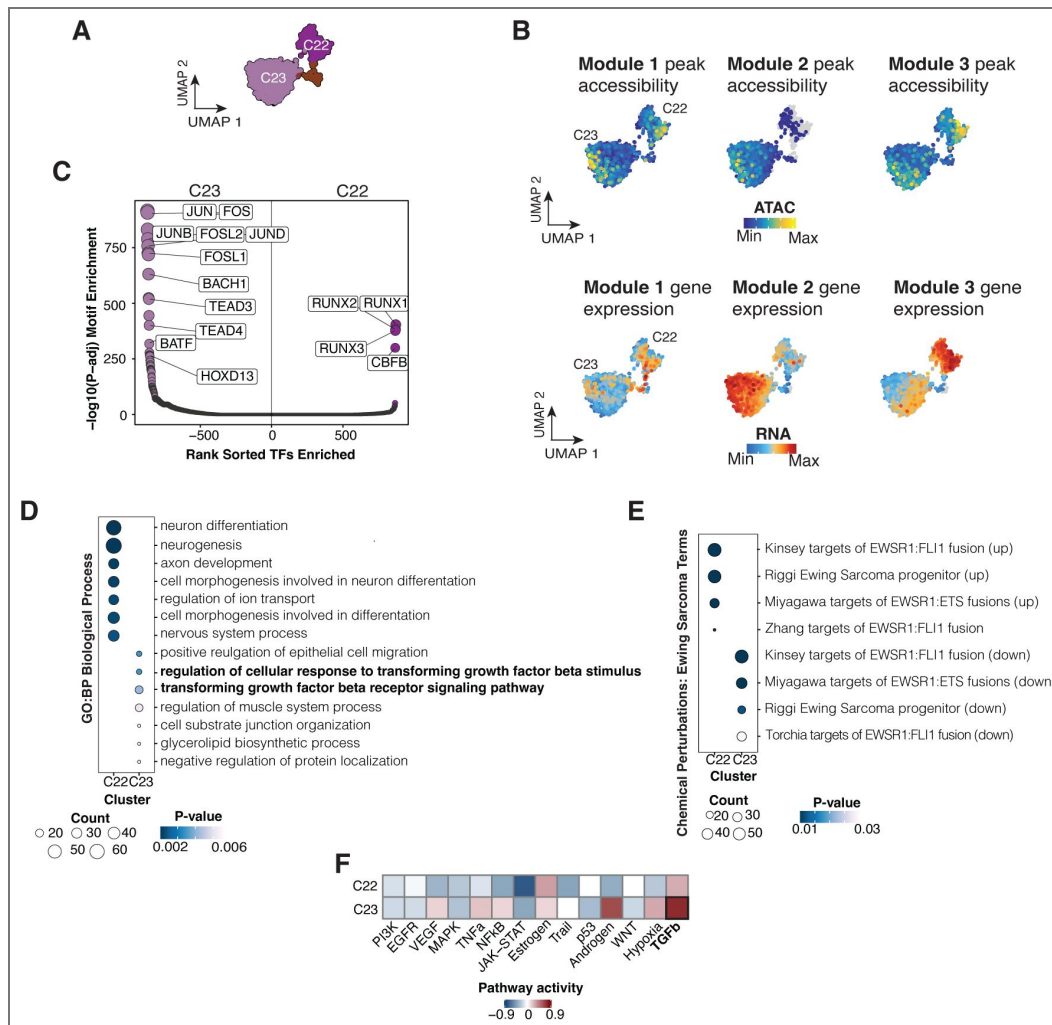


Fig. 4. CHLA10 is a microcosm of EWS::FLI1 mediated heterogeneity

(A) UMAP of CHLA10 colored by cluster, C22 dark purple, C23 light purple. (B) UMAP of module peak accessibility and gene set expression in CHLA10 cells. (C) Differential motif enrichment between C22 and C23, TFs with $-\log_{10}(p\text{-value}) > 250$ highlighted. (D) Universal enrichment of GO: Biological Process terms in differentially expressed genes between C22 and C23, terms associated with TGF- β in bold. (E) Universal enrichment of MSigDB Ewing Sarcoma terms in differentially expressed genes between C22 and C23. (F) Pathway activity scores across C22 and C23, TGF- β highlighted in bold.

Sub-populations of EwS cell lines differentially respond to TGF- β stimulus

Prior reports have shown that TGF- β promotes pro-angiogenic extracellular-matrix (ECM) gene expression in EwS⁴⁸. To interrogate the impact of TGF- β signaling on the newly defined transcriptional states in EwS, we performed multiomic profiling on A673 (module 1-dominant), CHLA10 (module 2 dominant), and TC71 (module 3 dominant) after treatment with recombinant TGF- β 1 for 24 hours and compared them to vehicle-treated controls. After quality control filtering 19,611 cells were retained for analysis (Fig. S4A [↗](#)). Dimensionality reduction revealed substantial shifts in both gene expression and chromatin accessibility in A673 and CHLA10 following TGF- β treatment, whereas TC71 profiles remained relatively stable (Fig. 5A-C [↗](#)). In A673 and CHLA10, TGF- β treatment resulted in a clear polarization of cells towards module 2 usage (Fig. 5A-C [↗](#)), while expression patterns of modules 1 and 3 remained consistent with prior observations (Fig. S4B [↗](#)). The muted response of the TC71 cell line (module-3 dominant) to the influence of TGF- β (Fig. 5A-C [↗](#), Fig. S4B [↗](#)) suggests that pre-existing transcriptional states may condition the sensitivity to TGF- β signaling.

We next examined TF expression and motif accessibility in clusters exhibiting the most pronounced or lowest enrichment of TGF- β -treated cells (indicated by arrows in Fig. 5A-F [↗](#)). Correlation analyses revealed a reciprocal enrichment pattern of many key TFs from Figure 2 [↗](#), where correlation of FOSL2 (but not RUNX3) gene expression and accessibility was generally highest in TGF- β responsive clusters. Western blot validation confirmed TGF- β -induced FOSL2 expression specifically in CHLA10 but not in the relatively insensitive TC71 cells, despite observed SMAD2 phosphorylation (Fig. 5G [↗](#)).

Our recent whole human transcriptome digital spatial profiling (DSP) study of CHLA10 tumor xenografts (Nanostring GeoMx) showed that transcriptionally distinct tumor cell subpopulations are spatially distributed²¹. We therefore interrogated these data to determine if the identified gene regulatory modules in CHLA10 cells were randomly or spatially distributed across tumors in vivo. We first selected genes from modules 1 & 3 that were significantly upregulated (average log₂ fold change > 0.25) by EWS::FLI1 as well as module 2 genes that were significantly repressed by the fusion (average log₂ fold change < -0.25) and applied these fusion-regulated gene sets to the DSP data. Intriguingly, regions of interest (ROIs) that were captured from tumor borders (ROIs 2, 6, 8 & 10) were enriched with tumor cells that displayed heightened module-2 expression, whereas densely packed tumor cores (ROIs 1, 3, 4, 5, 7, 11) favored modules 1 and 3 (Fig. 5H [↗](#), Fig. S4C [↗](#)). Module 3 was also highly favored in subcutaneous compared to tail vein-derived disseminated tumors (Fig. 5H [↗](#)). Cells captured specifically from the border of a metastatic retroperitoneal tumor (ROIs 9 & 10) expressed the highest levels of *FOSL2*, TGF- β -response genes, and ECM-organization programs (Fig. 5I [↗](#)).

In summary, these findings highlight TGF- β -driven activation of module 2 gene expression in EwS cells via AP-1 transcription factors, particularly *FOSL2*. Furthermore, spatial analysis suggests module 2 programs are enriched at invasive tumor regions, underscoring the role of TGF- β signaling in mediating EwS cell heterogeneity and tumor progression dynamics.

Patient tumors show heterogeneous usage of EwS cis gene regulatory modules

We next sought to determine whether the gene-regulatory networks defined in our cell line panel are similarly deployed across the landscape of early passage patient-derived-xenograft (PDX) and primary patient biopsies. To determine how the three regulatory gene programs manifest in these tumors, we used the fusion-regulated gene sets as defined above. In a series of publicly available PDX tumor profiles¹⁹, module 1 was prominent, while modules 2 and 3 (Fig. 6A [↗](#)) showed more modest and heterogeneous enrichment. On the other hand, in six primary EwS tumor samples⁴⁹

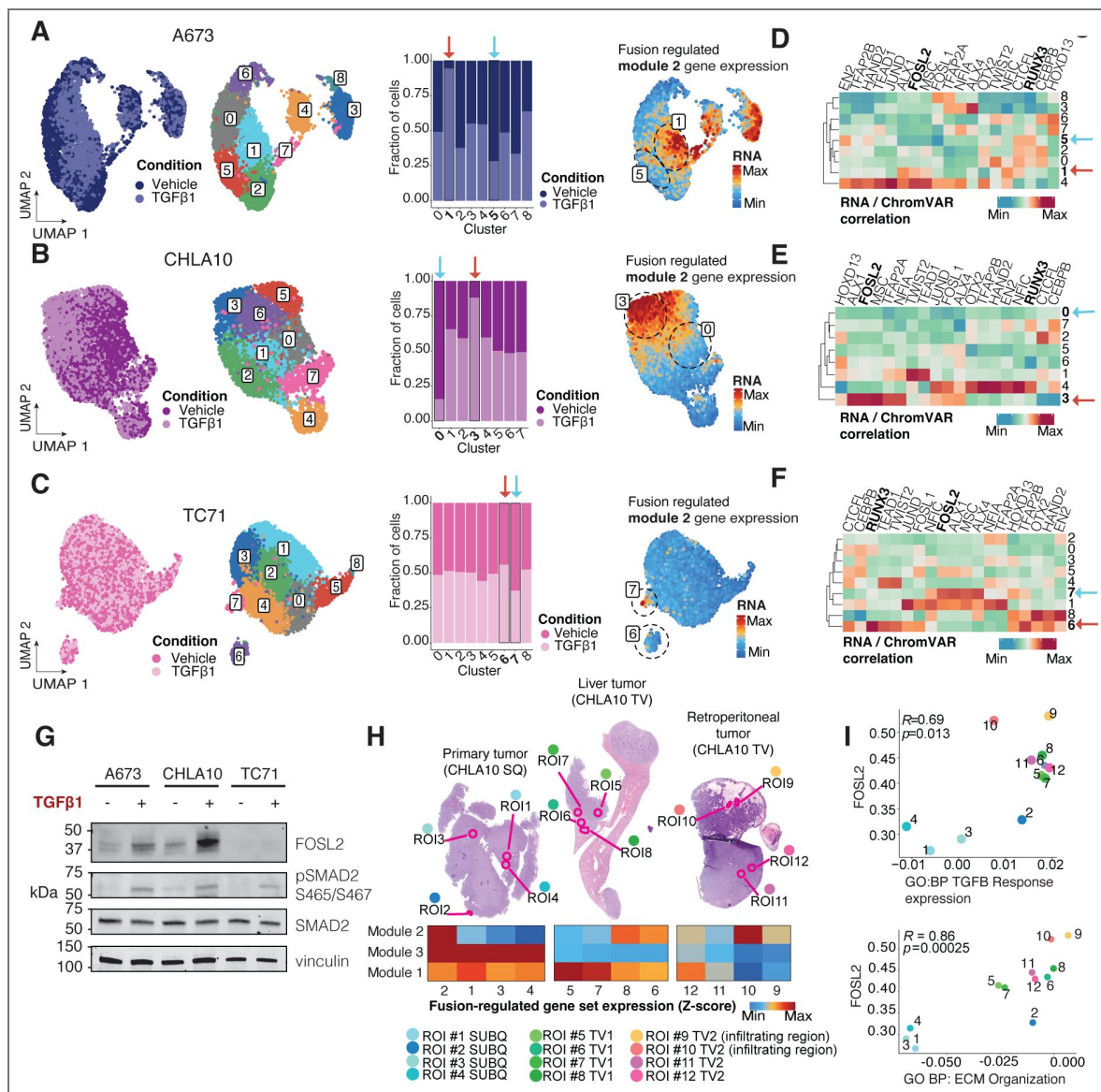


Fig. 5. Sub-populations of EwS cell lines differentially respond to TGF-β stimulus

(A–C) UMAP of A673, CHLA10, and TC71 cells colored by condition (vehicle control or TGFβ1 treatment) – column 1; and cluster – column 2; Fraction of treated versus control cells in each cluster with highest and lowest indicated by arrow – column 3; Expression of fusion regulated module 2 gene set, highest and lowest TGF-β clusters highlighted with dashed lines – column 4. (D–F) Correlation between TF expression and chromVAR TF deviation scores of EwS TFs across clusters, highest (red) and lowest (blue) TGF-β clusters highlighted by arrows. (G) Western blot of active (pS465/467) and total SMAD2 and FOSL2 protein in A673, CHLA10 and TC71 cells after 24 hr TGFβ1 treatment. (H) H&E staining of 3 CHLA10 xenografts in NSG mice generated by subcutaneous or tail vein injection (upper left). Those regions of interest (ROIs) subjected to whole human transcriptome digital spatial profiling are highlighted. Heatmap of module gene expression across ROIs (bottom left). (I) Scatter plot of average expression of Gene Ontology: Biological Process (GO:BP) TGFβ response genes and extracellular matrix (ECM) organization genes compared to FOSL2 expression in ROIs. Pearson R and p-value highlighted (top left).

(Fig. 6B [↗](#), Fig. S4D [↗](#)), profiled by single-nucleus RNA-seq, every case contained discrete subpopulations that preferentially expressed either module 2 or module 3. In contrast to the PDX data, module 1-positive cells were comparatively rare in primary tumors.

We next asked whether these single-cell patterns could be detected in bulk transcriptome data that were generated from three independent patient datasets^{9,50,51}. To achieve this we employed a deconvolution method⁵² that identifies the overlap of signals from bulk measurements with a single-cell transcriptomic reference and captures any proportion of the bulk signal not represented by single-cell data in an ‘intercept’ term (schematized in Fig. 6C [↗](#)). We first validated the method on bulk RNA sequencing data taken from many of the same EwS cell lines we profiled in our experiments and found a high level of concordance with little signal captured in the intercept term (Fig. S4E [↗](#)). We then tested the transcriptomic overlap of EwS subgroups in 85 non-clinically annotated patient tumors that were profiled by bulk RNA sequencing in two independent cohorts (PRJNA253759, PRJNA261990). Group 2 signals were evident in nearly all tumors and group 3 cell signals were also present in more than half of tumors, confirming heterogeneity of gene network usage and cell state in primary tumors (Fig. 6D [↗](#)). Tumors with abundant group 2 signal also displayed significantly higher expression of TGF- β response genes and FOSL2 expression (Fig. 6E [↗](#)), linking our in-vitro observation that TGF- β drives module 2 activation to patient tumors in the clinical setting.

Finally, we examined microarray data from a series of 46 clinically annotated primary tumors (GSE63157) that were uniformly treated and for whom outcome data were available⁵¹. First, we observed significant correlation between expression of FOSL2, TGF- β response and module 2 expression (Fig. 6F [↗](#)). Furthermore, high expression of fusion-repressed module 2, and particularly of its key effector *FOSL2*, identified patients with markedly worse event-free survival, whereas modules 1 and 3 had little prognostic value (Fig. 6G–H [↗](#), Fig. S4F [↗](#)). Module 2, linked to TGF- β responsiveness and FOSL2 activity, emerges as the predominant—and clinically adverse—program in primary EwS, underscoring its potential as both a biomarker of aggressive disease and a therapeutic target.

Discussion

Single-cell genomics has made it possible to chart, with unprecedented resolution, the transcriptional circuitry that underlies intratumoral diversity^{53,54}. In EwS, a cancer that has very few genetic mutations beyond its canonical chromosomal translocation^{19–21,39} and notable clinical heterogeneity, this diversity likely arises from epigenetic plasticity^{55,56}. Our study enhances this understanding by defining two distinct transcriptional and chromatin accessibility states consistently observed across cell lines, patient-derived xenografts (PDXs), and primary tumor samples.

The first transcriptional state, associated with module 3 genes, is enriched for canonical GGAA-microsatellite enhancers, high EWS::FLI1 expression, and neural-crest transcription factors such as TFAP2A and RUNX3. In contrast, the second state associated with module 2 genes lacks GGAA repeats, is suppressed by EWS::FLI1, and involves activation by TEAD and AP-1 family members together with the homeobox factor HOXD13—all elements previously linked to epithelial-to-mesenchymal transition (EMT), invasion, and a low-fusion state^{57–59}.

The mutually antagonistic architecture of these modules is consistent with earlier observations that EWS::FLI1-repressed loci tend to be short-repeat or single ETS sites that recruit additional repressor complexes such as ETV6^{7,60,61}. Further supporting this, recent data³⁴ demonstrate EWS::FLI1’s role in shaping cell identity by controlling promoter-enhancer looping and 3D chromatin networks. Loss of EWS::FLI1 was shown to alter 3D chromatin structure, activating enhancers enriched for AP-1 motifs and facilitating mesenchymal gene expression, consistent with our findings.

Analysis of three independent patient cohorts confirmed module 2 and module 3 signatures in most primary tumors. Clinically, elevated expression of module 2 genes, notably FOSL2, correlated with inferior event-free survival in a prospectively treated cohort of patients with localized EwS.

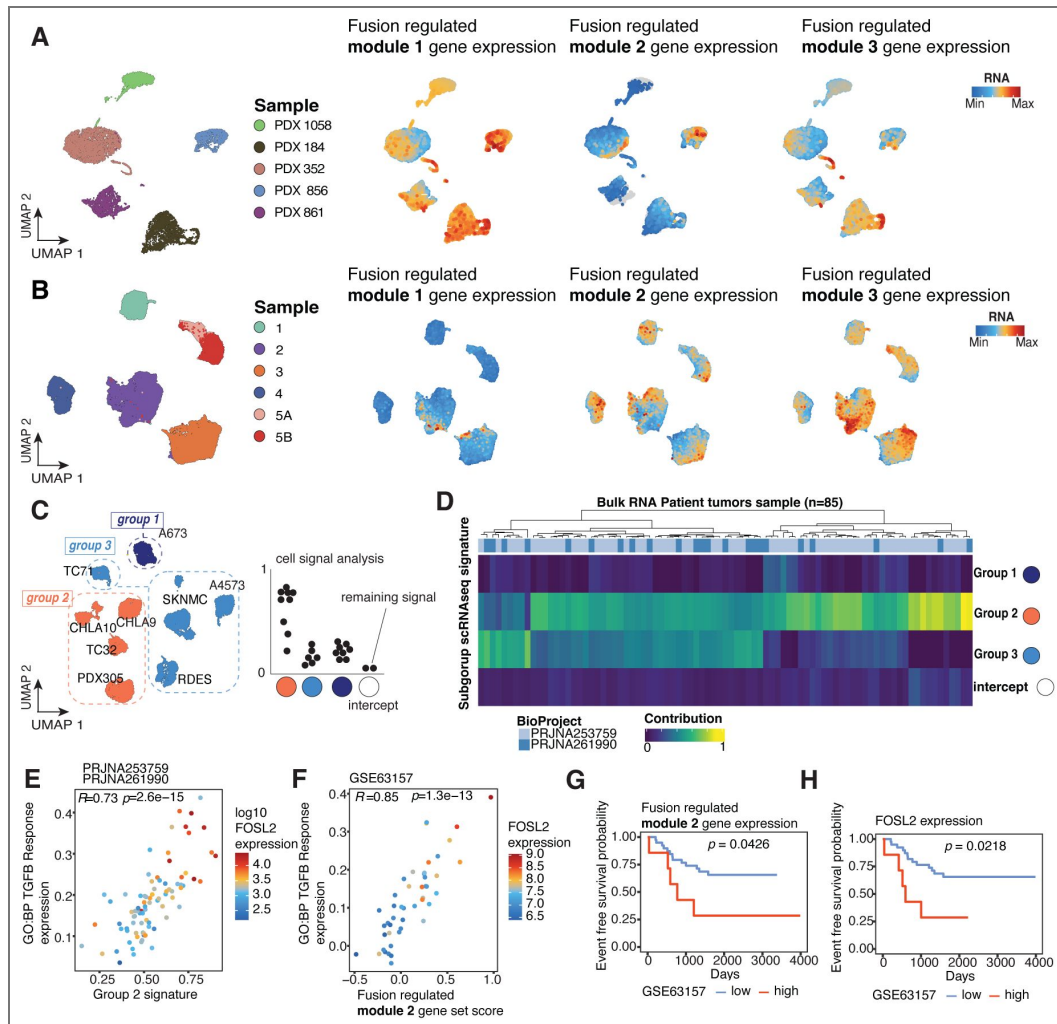


Fig. 6. Human tumor data show usage of EWS::FLI1 gene regulatory modules.

(A) UMAP embedding of five EWS PDX samples (left) UMAP of fusion regulated module gene sets in PDX samples (right) (B) UMAP embedding of single nuclei sequencing of 6 patient tumor samples (left). Patient tumor scRNA expression of fusion regulated modules (right). (C) Schematic depicting deconvolution method (D) Heatmap EWS subgroup signal in bulk EWS patient tumor RNA-seq. The intercept row represents remaining signal (bottom). Patient tumor samples annotated by GEO BioProject. (E) Scatter plot of bulk RNA patient samples by group 2 signature and GO: BP TGF- β response genes and colored log10 FOSL2 expression. Pearson R and p-value highlighted (top left). (F) Scatter plot of microarray patient samples by module 2 expression and GO: BP TGF- β response genes and colored FOSL2 expression. Pearson R and pvalue highlighted (top left). (G) Survival curves of patient microarray data51 (n = 46), high expression of module 2 gene set in red (top 15% expressing samples, n = 7), low expression in blue (bottom 85%, n = 39) (H) Survival curves of high expression of FOSL2 samples (top 15%, n = 7, bottom 85%, n=39).

This association nominates module 2 expression as a potential biomarker of aggressive disease and links TGF- β signaling to EMT-like transcriptional states and poor prognosis.

Another key insight from our study is the link between cells expressing high levels of module 2 genes and heightened responsiveness to TGF- β . These findings argue that external cues can re-wire specific subsets of EwS cells towards a more extreme mesenchymal/EMT-like state in the absence of genetic change. This is in line with emerging data that cell plasticity in EwS is influenced by a TGF- β -induced transition of EwS cells to cancer-associated fibroblast-like states²¹, which appear to be “unlocked” by TGFB2 response⁶². Thus, therapeutic strategies that disrupt these microenvironmental signals—or directly target downstream effectors such as FOSL2—may limit metastatic spread by constraining module 2 activation. Moreover, by promoting a fibroblast-like phenotype and dampening inflammatory signaling, TGF- β may also contribute to immune evasion and the establishment of an immunosuppressive tumor microenvironment^{22,23,63}.

Our study is constrained by the limited number of primary single-cell datasets currently available; larger, clinically annotated cohorts will be required to validate the prognostic utility of module balance and to map temporal dynamics during therapy and relapse. Moreover, while CHLA9/CHLA10 serves as a useful isogenic model of plasticity, early-passage or organoid systems may preserve an even richer landscape of states. Finally, dissecting the full spectrum of extrinsic factors—beyond TGF- β —that tip the balance between modules will be key to developing microenvironment-targeted interventions.

In summary, our study provides novel insights into the TF regulatory networks that interact with the EWS::FLI1 fusion protein to sustain distinct cellular states and facilitate dynamic transitions between these states. Specifically, we identified core gene network modules and key TFs, notably FOSL2 and other module 2 TFs, which antagonize fusion-mediated repression, promoting a mesenchymal and EMT-like state associated with more aggressive disease phenotypes. Furthermore, our findings highlight cell-extrinsic TGF- β as a critical factor capable of activating FOSL2 and driving the aggressive module 2 program *in vivo*. By uncovering these regulatory mechanisms and their functional implications, we offer a foundation for developing therapeutic strategies aimed at limiting metastasis, overcoming treatment resistance, and ultimately improving clinical outcomes for patients with Ewing sarcoma.

Materials and Methods

Sex as a biological variable

Our study examined male and female animals, and similar findings are reported for both sexes.

Cell Lines and PDX

The EwS cell lines A673, SKNMC, CHLA10, CHLA9, TC32, A4573, RDES, and TC71 were obtained from ATCC and COG (<https://www.childrensoncologygroup.org/>) cell lines repositories. The osteosarcoma cell line U2OS and rhabdomyosarcoma cell line RD were also obtained from ATCC. 5064L and 240L MSCs were obtained from Dr. Darwin Prockop. H7 MSCs were generously gifted from Dr. Alejandro Sweet-Cordero. PDX305 was generated from a recurrent metastatic tumor, as previously described²¹. A673, TC32, A4573, TC71, and SKNMC cell lines were maintained in RPMI 1640 media (Gibco) supplemented with 10% FBS (Atlas Biologicals) and 2 mmol/L glutamine (Life Technologies). RDES was maintained in RPMI 1640 media supplemented with 15% FBS. Isogenic cell lines, CHLA10 and CHLA9, were maintained in IMDM media (Fisher) supplemented with 20% FBS, 2 mmol/L-glutamine, and 1X Insulin-Transferrin-Selenium (Gibco). PDX305 was maintained in RPMI 1640 media supplemented with 10% FBS, 2 mmol/L-glutamine, and 1% Antimycotic-Antibiotic (Gibco). U2OS was maintained in McCoy's 5A media supplemented with 10% FBS and 2mmol/L-glutamine and RD was maintained in DMEM with 10% FBS. All hMSC lines were maintained in MEM-alpha media supplemented with 20% FBS, 2 mmol/L-glutamine, and 1% Antimycotic-Antibiotic. Cells were cultured at 37°C with 5% CO₂. Cells were all confirmed to be mycoplasma free and identities subject to STR-confirmation every 6 months.

Lentivirus Production and Genetic knockdown

pLKO.1 shNS (SHC002) and shFLI1 (TRCN0000005322) were used for FLI1 knockdown multiome experiments (Sigma). Knockdown plasmids were co-transfected with pCD/NL-BH*DDD (addgene, #17531) and pMD2.G (addgene, #12259) plasmids into 293FT packaging cells using PEI. 24 hours post-transfection, sodium butyrate was added to plates for 6 hours. Viral supernatant was collected 48 hours after transfection, filtered through 40 μ M filters, and added the cells for 24-hours. Ewing sarcoma cells were selected by supplementing the media with 1.5 μ g/mL puromycin (Gibco) for 48 hours before collection at 96 hours.

TGF- β treatment prior to single cell sequencing

TC71, A673, and CHLA10 cells were plated at 500,000 viable cells per 10cm plate. 48 hours later recombinant human TGF β 1 (R&D 7754-BH-005, final dose 10 ng/mL in sterile 4 mM HCl and 0.1% BSA) or vehicle control was added. 24 hours later cells were trypsinized and $>1 \times 10^6$ viable cells were subjected to single cell library preparation and sequencing (details below).

Western blots

Cells were washed with ice cold PBS, scraped and centrifuged, then lysed with RIPA buffer (Fisher Scientific) supplemented with protease and phosphatase inhibitors (Sigma) and sonicated (ten 30 second cycles). For TGF- β treated samples, recombinant TGF β 1 (10 ng/mL) or vehicle control (4 mM HCl, 0.1% BSA) were included in the media for either 30 minutes or 24hrs before lysis. Western blots were performed using the Bio-Rad Mini-PROTEAN Tetra System using 40 μ g of lysate and 4-15% Mini-PROTEAN TGX polyacrylamide gels. Following transfer, nitrocellulose membranes were blocked in Odyssey Blocking Buffer (LI-COR) for 1 hour. Membranes were washed and incubated rotating overnight at 4°C with primary antibodies (FOSL2 CST #19967, HAND2 R&D AF3876, RUNX3 CST #13089, TEAD1 CST #12292, pSMAD2 S465/467 CST #3108, SMAD2 CST #3103, GAPDH Invitrogen AM4300, Vinculin CST #13901). Membranes were then washed four times in TBST for 5 minutes each and incubated with secondary antibodies (LI-COR IRDye 700CW or 800CW; 1:10,000) for 1 hour. After 4 additional 5 minute TBST washes membranes were scanned on a LI-COR Odyssey scanner.

Digital Spatial Profiling

Q3 normalized gene counts from GeoMx Digital Spatial Profiling (NanoString) of EwS cell line xenograft tumors formed by CHLA10 injection were obtained from previous work²¹. This processed data was used to query fusion regulated module gene sets with respect to tumor microenvironment and injection site. Gene sets values for each ROI were calculated by converting the DSP Q3 counts matrix into a Seurat object and using the AddModuleScore function.

Single-cell analysis

Cell and library prep

One million EwS cells were resuspended in PBS containing 2% BSA and nuclei were prepared according to manufacturer's specifications. Cell lines in triplicate were pooled 1:1:1 at 1000 cells/uL and libraries were generated using the Chromium Next GEM Single-Cell Multiome ATAC+Gene Expression Kit following the manufacturer's protocol (CG000338 Rev C). The resulting libraries were sequenced using a NovaSeq 6000 targeting a depth of 25K reads per cell (ATAC libraries) and 20K reads per cell (RNA libraries).

Single-cell multiome data analysis

Sequencing reads were demultiplexed and processed using the cellranger ARC pipeline. Cells with poor quality ATAC fragments and multiplets were filtered using ArchR's standard workflow. Cells passing the following were kept for later analysis: percent mitochondrial RNA reads $< 15\%$; $3 < \log_{10}(\text{RNA}/\text{UMI counts}) \leq 4.5$. Souporecell⁵⁴ was used to identify cell lines and remove additional multiplets. Top RNA markers from scRNA data were compared to previously hashed CITEseq

profiles²¹ to match genotype to cell line (GSE236289: GSM7527501, GSM7527502, GSM7527503). ArchR³⁷, Signac⁶⁵, and Seurat⁴⁶ were used for downstream analysis. Dimensionality reduction was performed using addIterativeLSI() for both tile and gene expression matrices. An embedding was created by combineDims() and then addUMAP() with the mulitomic reduction. Group coverages and peaks calls were made using cell line identities. The motif matrix was built using addMotifDeviations(), addBdgPeaks(), and addDeviationsMatrix().

Average gene expression and gene activity scores were computed using addModuleScore(). EWS::FLI1 deviations were computed using the JASPAR2020 motif database. Transcription factor analyses used the cisBP database⁶⁶. Highly utilized TFs were inferred using correlateMatrices(). All heatmaps were made with ComplexHeatmap. Marker peaks and genes were calculated using getMarkerFeatures() and getMarkersHeatmap(). Cis-regulatory elements were identified using addPeak2GeneLinks(). Universal gene set enrichment analysis was performed using clusterProfiler using the function enricher(), with MSigDB⁶⁷. Hallmark and gene ontology references. Motif enrichment analysis on kmean peak data was performed using HOMER. The most enriched motifs were identified using findMotifsGenome.pl and motif enrichment was computed and visualized using annotatePeaks.pl. EWS:ETS binding sites were sourced from previously published data⁶⁸, and analyzed using HOMER and chromVAR with the same workflow as above.

Cells treated with TGF- β agonist and captured for multiomic sequencing were demultiplexed and processed using the same workflow as above. QC metrics used were identical to those above. A similar method of identifying each cell line using genetic demultiplexing was employed. Signac⁶⁴ was used to perform dimensionality reduction using weighted nearest neighbors. Motif matrices were constructed using RunChromVAR(), using cisBP⁶⁶ as a database.

EWS::FLI1 long read sequencing

Starting material was 10x Genomics-generated full-length barcoded cDNA, leftover from what was used to create the Illumina-compatible libraries in the multiome downstream steps. Because 10x Genomics single-cell cDNA preparations typically contain a template switch oligo (TSO) priming artifact (i.e., TSO sequence at both ends of the cDNA molecule) instead of the correct structure, TSO artifact depletion was performed by using biotin-labeled primers for the streptavidin/biotin magnetic selection of molecules with the correct structure, as previously described⁶⁹. To capture the gene transcripts encoding the EWS::FLI1 fusions, an enrichment was performed using xGenTM Hybridization and Wash v2 kit magnetic beads and reagents (IDT #10010351), as described in the manufacturer's protocol, in addition to using 500 ng of cDNA starting material and using biotinylated capture probes designed with high specificity, tiling along exons 5, 6, and 7 of the *EWSR1* transcript, and exon 9 of the *FLI1* transcript. Following hybridization and magnetic capture of EWS::FLI1 molecules, an on-bead PCR was performed to amplify captured molecules, using standard 10x Genomic cDNA amplification protocol (described in 10x Genomics document #CG000315_Rev_B, Step 2.2) with the exception of a number of PCR cycles of 16 and an extension time at 72°C of 3 min. Following this, 500 ng of captured cDNA was DNA-repaired, A-tailed, SMRTbell adaptor-ligated, nuclease-treated, and SMRTbell-cleanup purified using the PacBio SMRTbell Prep kit 3.0 (PacBio #102-141-700) as per manufacturer's protocol. Libraries were sequenced on a PacBio Sequel IIe with 2 hours pre-extension time and 30 hours movie time.

Long-read computational processing pipeline consisted of retrieving error-corrected circular consensus sequence (CCS) BAM files generated by the on-board PacBio Sequel IIe (Pacific Biosciences of California, Inc.). CCS data were then processed by the programs in the IsoSeq3 suite of computational tools: 1) *Lima* to remove sequencing primers, 2) *Tag* to extract UMI and CB sequences, 3) *Refine* to remove polyA sequences, 4) *Correct* to error-correct CBs, and 5) *Dedup* to remove PCR duplicates, 4) *Pbmm2* to align reads. All the IsoSeq3 pipeline steps were used with default settings. Mapped BAMs were then used for pbfusion to detect fusions. Barcodes from long read molecules were extracted from mapped BAMs and merged with data of pbfusion.breakpoints.groups.bed files to count fusion transcript in single cells.

External Data

The 78 direct EwS TGs were sourced from a previously published dataset¹⁹. Bulk RNAseq patient tumor samples were sourced from dbGaP under accession numbers PRJNA253759 and PRJNA261990. FASTQ files were aligned to reference genome (GRCh38) and using the piperna (<https://github.com/furlan-lab/piperna>) workflow which uses STAR⁷⁰ aligner and generates a count matrix using the Bioconductor summarize overlap function (<https://www.bioconductor.org>) with its default settings. Microarray data was accessed from GSE63157⁵¹. RNA probes were annotated using pd.huex.1.0.st.v2, and normalized using oligo.

scRNAseq Patient Tumors

Patient tumor biopsies were sourced and sequenced at St. Jude's Children's Hospital⁴⁹. 10X reads were demultiplexed and processed using the cellranger pipeline. Ambient RNA was filtered out using cellbender. Cells with poor RNA measurements were filtered using the following thresholds: percent mitochondrial reads < 15% and $3 < \log_{10}(\text{RNA counts}) < 5.25$. Dimensionality reduction and clustering was performed using standard workflow. Surrounding tissues cells were identified by using FindAllMarkers on Seurat clusters. Tumor enrichment was validated by choosing cells that express *CD99*, *NROB1*, and EwS TGs. Identified tumor cells underwent normalization and dimensionality reduction with the same workflow. Average gene set scores were calculated using AddModuleScore().

Deconvolution

Bulk RNAseq patient data were deconvoluted using cellSignalAnalysis (<https://github.com/constantAmateur/cellSignalAnalysis>). EwS subgroups 1-3 were used as transcriptomic references.

Statistics

All statistical analyses were conducted using R or Python, with specific packages indicated in relevant sections of the Methods. Statistical significance was assessed using two-sided tests unless otherwise specified. For comparisons of gene set scores, motif deviations, or expression levels, Wilcoxon rank-sum or Student's *t*-tests were used as appropriate based on data distribution and variance. Correction for multiple hypothesis testing was performed using the Benjamini-Hochberg method where applicable. Gene set enrichment analyses were conducted using the clusterProfiler package with MSigDB Hallmark and Gene Ontology annotations. For scRNA-seq and multiome analyses, differential expression and peak calling were performed using Seurat, ArchR, or Signac, and marker features were identified with adjusted *p*-values < 0.05 considered significant.

Study approval

All work was approved by the Fred Hutchinson Cancer Center Institutional Review Board - Tumor Heterogeneity and Metastasis in Ewing Sarcoma - FHIRB0010627

Data availability

All code and processed data is available on https://github.com/furlan-lab/EwS_multiome. Data has also been deposited on GEO with accession GSE332630 including raw sequence. Single-cell RNA sequencing from primary samples are available at <https://scpca.alexslimonade.org>.

Acknowledgements

The authors thank the Fred Hutchinson Scientific Computing Group and Genomics Core staff. This study was supported by multiple sources of funding, including the National Institutes of Health/National Cancer Institute (NIH/NCI) grant 1L40CA264534-01 (SNF), the 1Million 4 Anna Foundation (SNF, ERL), and NIH/NCI grant R01 CA215981 (ERL). Additional support was provided by NIH/NCI grant F31CA247104 (AAA) and the AACR-QuadW Foundation Sarcoma Research

Fellowship in Memory of Willie Tichenor, Grant Number 22-40-37-WREN (EDW). Funding also came from the Sam Day Foundation (ERL), NIH grants S10-OD-020069 and S10-OD-028685, the Fred Hutch/University of Washington/Seattle Children's Cancer Center Consortium NCI Cancer Center Support Grant P30 CA015704 and the Fred Hutch Translational Data Science Integrated Research Center Award (OGW).

Additional information

Author contributions

Conceptualization: OGW, AA, EDW, JRS, ERL, SNF

Data curation: OGW, AA, EDW

Data generation: AA, EDW, MM, SB, SBK, AP, MD

Formal analysis: OGW, AA, EDW, RV, ZK, SNF

Software: OGW, RV, ZK, SNF

Visualization: OGW, AA, EDW

Supervision: JRS, ERL, SNF

Funding acquisition: ERL, SNF

Project administration: ERL, SNF

Writing—original draft: OGW, AA, EDW, AP, MD

Writing—review & editing: OGW, AA, EDW, RGG, RV, ZK, AP, MD, JRS, ERL, SNF

Funding

Funder	Grant reference number	Author
HHS NIH National Cancer Institute (NCI)	L40CA264534-01	Scott Furlan
HHS NIH National Cancer Institute (NCI)	R01CA215981	Elizabeth R Lawlor MD PhD
HHS NIH National Cancer Institute (NCI)	F31CA247104	April A Apfelbaum

Author ORCID iDs

Anand Patel:  <https://orcid.org/0000-0002-1438-8275>

Scott N Furlan:  <https://orcid.org/0000-0003-3007-9840>

Additional files

[Supplemental Materials](#) 

References

- 1 Riggi N, Suva ML, Stamenkovic I (2021) Ewing's Sarcoma. *N Engl J Med* **384**:154-164 <https://doi.org/10.1056/NEJMra2028910> | PubMed
- 2 Grunewald TGP, et al. (2018) Ewing sarcoma. *Nat Rev Dis Primers* **4**:5 <https://doi.org/10.1038/s41572-018-0003-x> | PubMed
- 3 Tirole F, et al. (2007) Mesenchymal stem cell features of Ewing tumors. *Cancer Cell* **11**:421-429 <https://doi.org/10.1016/j.ccr.2007.02.027> | PubMed
- 4 Riggi N, Suva ML, Stamenkovic I (2009) Ewing's sarcoma origin: from duel to duality. *Expert Rev Anticancer Ther* **9**:1025-1030 <https://doi.org/10.1586/era.09.81> | PubMed
- 5 Lin PP, Wang Y, Lozano G (2011) Mesenchymal Stem Cells and the Origin of Ewing's Sarcoma. *Sarcoma* <https://doi.org/10.1155/2011/276463> | PubMed

- 6 Miller HE, et al. (2020) Reconstruction of Ewing Sarcoma Developmental Context from Mass- Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility. *Cancers* **12** <https://doi.org/10.3390/cancers12040948> | PubMed
- 7 Riggi N, et al. (2014) EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* **26**:668-681 <https://doi.org/10.1016/j.ccell.2014.10.004> | PubMed
- 8 Boulay G, et al. (2017) Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171**:163-178.e119, <https://doi.org/10.1016/j.cell.2017.07.036> | PubMed
- 9 Crompton BD, et al. (2014) The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov* **4**:1326-1341 <https://doi.org/10.1158/2159-8290.CD-13-1037> | PubMed
- 10 Zoubek A, et al. (1996) Does expression of different EWS chimeric transcripts define clinically distinct risk groups of Ewing tumor patients?. *J Clin Oncol* **14**:1245-1251 <https://doi.org/10.1200/JCO.1996.14.4.1245> | PubMed
- 11 Davies A, Zoubeydi A, Beltran H, Selth LA (2023) The Transcriptional and Epigenetic Landscape of Cancer Cell Lineage Plasticity. *Cancer Discov* **13**:1771-1788 <https://doi.org/10.1158/2159-8290.CD-23-0225> | PubMed
- 12 Hanahan D (2022) Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**:31-46 <https://doi.org/10.1158/2159-8290.CD-21-1059> | PubMed
- 13 Tomazou EM, et al. (2015) Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell Rep* **10**:1082-1095 <https://doi.org/10.1016/j.celrep.2015.01.042> | PubMed
- 14 Sheffield NC, et al. (2017) DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. *Nat Med* **23**:386-395 <https://doi.org/10.1038/nm.4273> | PubMed
- 15 Franzetti GA, et al. (2017) Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **36**:3505-3514 <https://doi.org/10.1038/onc.2016.498> | PubMed
- 16 Adane B, et al. (2021) STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer Cell* **39**:827-844.e810, <https://doi.org/10.1016/j.ccell.2021.05.007> | PubMed
- 17 Pedersen EA, et al. (2016) Activation of Wnt/beta-Catenin in Ewing Sarcoma Cells Antagonizes EWS/ETS Function and Promotes Phenotypic Transition to More Metastatic Cell States. *Cancer Res* **76**:5040-5053 <https://doi.org/10.1158/0008-5472.CAN-15-3422> | PubMed
- 18 Surdez D, et al. (2021) STAG2 mutations alter CTCF-anchored loop extrusion, reduce cis- regulatory interactions and EWSR1-FLI1 activity in Ewing sarcoma. *Cancer Cell* **39**:810-826.e819, <https://doi.org/10.1016/j.ccell.2021.04.001> | PubMed
- 19 Aynaud MM, et al. (2020) Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cell Rep* **30**:1767-1779.e1766, <https://doi.org/10.1016/j.celrep.2020.01.049> | PubMed
- 20 Apfelbaum AA, et al. (2022) EWS::FLI1 and HOXD13 Control Tumor Cell Plasticity in Ewing Sarcoma. *Clin Cancer Res* **28**:4466-4478 <https://doi.org/10.1158/1078-0432.CCR-22-0384> | PubMed
- 21 Wrenn ED, et al. (2023) Cancer-associated fibroblast-like tumor cells remodel the Ewing sarcoma tumor microenvironment. *Clin Cancer Res* <https://doi.org/10.1158/1078-0432.CCR-23-1111> | PubMed
- 22 Visser LL, et al. (2023) Ewing Sarcoma Single-cell Transcriptome Analysis Reveals Functionally Impaired Antigen-presenting Cells. *Cancer Res Commun* **3**:2158-2169 <https://doi.org/10.1158/2767-9764.CRC-23-0027> | PubMed
- 23 Goodspeed A, et al. (2025) Single-Cell RNA Sequencing of Ewing Sarcoma Tumors Demonstrates Transcriptional Heterogeneity and Clonal Evolution. *Clin Cancer Res* **31**:2010-2023 <https://doi.org/10.1158/1078-0432.CCR-24-2040> | PubMed

- 24 **Chaturvedi A**, Hoffman LM, Welm AL, Lessnick SL, Beckerle MC (2012) The EWS/FLI Oncogene Drives Changes in Cellular Morphology, Adhesion, and Migration in Ewing Sarcoma. *Genes Cancer* **3**:102-116 <https://doi.org/10.1177/1947601912457024> | PubMed
- 25 **Chaturvedi A**, et al. (2014) Molecular dissection of the mechanism by which EWS/FLI expression compromises actin cytoskeletal integrity and cell adhesion in Ewing sarcoma. *Mol Biol Cell* **25**:2695-2709 <https://doi.org/10.1091/mbc.E14-01-0007> | PubMed
- 26 **Smith R**, et al. (2006) Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. *Cancer Cell* **9**:405-416 <https://doi.org/10.1016/j.ccr.2006.04.004> | PubMed
- 27 **Bilke S**, et al. (2013) Oncogenic ETS fusions deregulate E2F3 target genes in Ewing sarcoma and prostate cancer. *Genome Res* **23**:1797-1809 <https://doi.org/10.1101/gr.151340.112> | PubMed
- 28 **Bledsoe KL**, et al. (2014) RUNX3 facilitates growth of Ewing sarcoma cells. *J Cell Physiol* **229**:2049-2056 <https://doi.org/10.1002/jcp.24663> | PubMed
- 29 **Garcia-Garcia L**, et al. (2021) The Transcription Factor FEZF1, a Direct Target of EWSR1- FLI1 in Ewing Sarcoma Cells, Regulates the Expression of Neural-Specific Genes. *Cancers* **13** <https://doi.org/10.3390/cancers13225668> | PubMed
- 30 **Lin L**, et al. (2019) Super-enhancer-associated MEIS1 promotes transcriptional dysregulation in Ewing sarcoma in co-operation with EWS-FLI1. *Nucleic Acids Res* **47**:1255-1267 <https://doi.org/10.1093/nar/gky1207> | PubMed
- 31 **Howarth Marques**, et al. (2014) Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. *J Clin Invest* **124**:5275-5290 <https://doi.org/10.1172/JCI72124> | PubMed
- 32 **Shi X**, et al. (2020) EWS-FLI1 regulates and cooperates with core regulatory circuitry in Ewing sarcoma. *Nucleic Acids Res* **48**:11434-11451 <https://doi.org/10.1093/nar/gkaa901> | PubMed
- 33 **Kinsey M**, Smith R, Lessnick SL (2006) NR0B1 is required for the oncogenic phenotype mediated by EWS/FLI in Ewing's sarcoma. *Mol Cancer Res* **4**:851-859 <https://doi.org/10.1158/1541-7786.MCR-06-0090> | PubMed
- 34 **Sanalkumar R**, et al. (2023) Highly connected 3D chromatin networks established by an oncogenic fusion protein shape tumor cell identity. *Sci Adv* **9**:eabo3789 <https://doi.org/10.1126/sciadv.abo3789> | PubMed
- 35 **Minas TZ**, et al. (2017) Combined experience of six independent laboratories attempting to create an Ewing sarcoma mouse model. *Oncotarget* **8**:34141-34163 <https://doi.org/10.18632/oncotarget.9388> | PubMed
- 36 **Tirosh I**, et al. (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single- cell RNA-seq. *Science* **352**:189-196 <https://doi.org/10.1126/science.aad0501> | PubMed
- 37 **Granja JM**, et al. (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**:403-411 <https://doi.org/10.1038/s41588-021-00790-6> | PubMed
- 38 **Orth MF**, et al. (2022) Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation. *Cell Rep* **41**:111761 <https://doi.org/10.1016/j.celrep.2022.111761> | PubMed
- 39 **Apfelbaum AA**, Wrenn ED, Lawlor ER (2022) The importance of fusion protein activity in Ewing sarcoma and the cell intrinsic and extrinsic factors that regulate it: A review. *Front Oncol* **12**:1044707 <https://doi.org/10.3389/fonc.2022.1044707> | PubMed
- 40 **Faustino A**, et al. (2012) mTOR pathway overactivation in BRAF mutated papillary thyroid carcinoma. *J Clin Endocrinol Metab* **97**:E1139-1149 <https://doi.org/10.1210/jc.2011-2748> | PubMed
- 41 **Kakkar A**, et al. (2016) Alterations in BRAF gene, and enhanced mTOR and MAPK signaling in dysembryoplastic neuroepithelial tumors (DNTs). *Epilepsy Res* **127**:141-151 <https://doi.org/10.1016/j.eplepsyres.2016.08.028> | PubMed
- 42 **Prabowo AS**, et al. (2014) BRAF V600E mutation is associated with mTOR signaling activation in glioneuronal tumors. *Brain Pathol* **24**:52-66 <https://doi.org/10.1111/bpa.12081> | PubMed

- 43 Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**:576-589 <https://doi.org/10.1016/j.molcel.2010.05.004> | PubMed
- 44 Kim S, Denny CT, Wisdom R (2006) Cooperative DNA binding with AP-1 proteins is required for transformation by EWS-Ets fusion proteins. *Mol Cell Biol* **26**:2467-2478 <https://doi.org/10.1128/MCB.26.7.2467-2478.2006> | PubMed
- 45 Johnson KM, Taslim C, Saund RS, Lessnick SL (2017) Identification of two types of GGAA-microsatellites and their roles in EWS/FLI binding and gene regulation in Ewing sarcoma. *PLoS One* **12**:e0186275 <https://doi.org/10.1371/journal.pone.0186275> | PubMed
- 46 Stuart T, et al. (2019) Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-1902.e1821, <https://doi.org/10.1016/j.cell.2019.05.031> | PubMed
- 47 Badia IMP, et al. (2022) decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv* **2**:vbac016 <https://doi.org/10.1093/bioadv/vbac016> | PubMed
- 48 Hawkins AG, et al. (2020) Wnt/beta-catenin-activated Ewing sarcoma cells promote the angiogenic switch. *JCI Insight* **5** <https://doi.org/10.1172/jci.insight.135188> | PubMed
- 49 Hawkins AG, et al. (2024) The Single-cell Pediatric Cancer Atlas: Data portal and open-source tools for single-cell transcriptomics of pediatric tumors. *bioRxiv* <https://doi.org/10.1101/2024.04.19.590243>
- 50 Brohl AS, et al. (2014) The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet* **10**:e1004475 <https://doi.org/10.1371/journal.pgen.1004475> | PubMed
- 51 Volchenboum SL, et al. (2015) 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 <https://doi.org/10.1002/cjp2.9> | PubMed
- 52 Young MD, et al. (2021) Single cell derived mRNA signals across human kidney tumors. *Nat Commun* **12**:3896 <https://doi.org/10.1038/s41467-021-23949-5> | PubMed
- 53 Gavish A, et al. (2023) Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**:598-606 <https://doi.org/10.1038/s41586-023-06130-4> | PubMed
- 54 Marusyk A, Almendro V, Polyak K (2012) Intra-tumour heterogeneity: a looking glass for cancer?. *Nat Rev Cancer* **12**:323-334 <https://doi.org/10.1038/nrc3261> | PubMed
- 55 van Groningen T, et al. (2017) Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* **49**:1261-1266 <https://doi.org/10.1038/ng.3899> | PubMed
- 56 Boeva V, et al. (2017) Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet* **49**:1408-1413 <https://doi.org/10.1038/ng.3921> | PubMed
- 57 Zhang H, et al. (2009) TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition. *J Biol Chem* **284**:13355-13362 <https://doi.org/10.1074/jbc.M900843200> | PubMed
- 58 Katschnig AM, et al. (2017) EWS-FLI1 perturbs MRTFB/YAP-1/TEAD target gene regulation inhibiting cytoskeletal autoregulatory feedback in Ewing sarcoma. *Oncogene* **36**:5995-6005 <https://doi.org/10.1038/onc.2017.202> | PubMed
- 59 Croushore EE, Stipp CS, Gordon DJ (2024) EWS-FLI1 and Activator Protein-1 (AP-1) Reciprocally Regulate Extracellular-Matrix Proteins in Ewing sarcoma Cells. *Int J Mol Sci* **25** <https://doi.org/10.3390/ijms25168595> | PubMed
- 60 Sankar S, et al. (2013) Mechanism and relevance of EWS/FLI-mediated transcriptional repression in Ewing sarcoma. *Oncogene* **32**:5089-5100 <https://doi.org/10.1038/onc.2012.525> | PubMed
- 61 Lu DY, et al. (2023) The ETS transcription factor ETV6 constrains the transcriptional activity of EWS-FLI to promote Ewing sarcoma. *Nat Cell Biol* **25**:285-297 <https://doi.org/10.1038/s41556-022-01059-8> | PubMed

- 62 Wrenn ED, et al. (2025) Autocrine TGFβ2 enforces a transcriptionally hybrid cell state in Ewing sarcoma. *bioRxiv* <https://doi.org/10.1101/2025.06.11.653876> | PubMed
- 63 Kumar RR, et al. (2025) Decoding the immune landscape in Ewing sarcoma pathogenesis: The role of tumor infiltrating immune cells and immune milieu. *J Bone Oncol* **52** <https://doi.org/10.1016/j.jbo.2025.100678> | PubMed
- 64 Heaton H, et al. (2020) Souporecell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods* **17**:615-620 <https://doi.org/10.1038/s41592-020-0820-1> | PubMed
- 65 Stuart T, Srivastava A, Madad S, Lareau CA, Satija R (2021) Single-cell chromatin state analysis with Signac. *Nat Methods* **18**:1333-1341 <https://doi.org/10.1038/s41592-021-01282-5> | PubMed
- 66 Weirauch MT, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**:1431-1443 <https://doi.org/10.1016/j.cell.2014.08.009> | PubMed
- 67 Subramanian A, 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 <https://doi.org/10.1073/pnas.0506580102> | PubMed
- 68 Theisen ER, et al. (2021) Chromatin profiling reveals relocalization of lysine-specific demethylase 1 by an oncogenic fusion protein. *Epigenetics* **16**:405-424 <https://doi.org/10.1080/15592294.2020.1805678> | PubMed
- 69 Al'Khafaji AM, et al. (2024) High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat Biotechnol* **42**:582-586 <https://doi.org/10.1038/s41587-023-01815-7> | PubMed
- 70 Dobin A, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15-21 <https://doi.org/10.1093/bioinformatics/bts635> | PubMed

Peer reviews

Reviewer #1 (Public review):

The investigators elegantly utilized single-cell co-assay of RNA and ATAC seq to unveil the heterogeneous gene regulatory networks in Ewing sarcoma. The authors should be commended on their ability to identify multiple unique modules of gene regulation of Ewing sarcoma utilizing complex computational methods between numerous Ewing sarcoma cell lines. Additionally, they complimented their single cell findings with xenografts as well as primary Ewing sarcoma patient tumors - validating the intratumoral heterogeneous gene regulatory networks of Ewing sarcoma. More importantly, they have revealed that exogenous TGF-β may modify these distinct epigenetic and transcriptional signatures within Ewing sarcoma tumors. Overall, the manuscript highlights an important discovery of the heterogenous gene regulatory programming of Ewing sarcoma and further highlights the role that TGFB plays within the tumor microenvironment of Ewing sarcoma. There are some areas of ambiguity that require clarification to increase the impact of the manuscript.

Comments on the latest revision:

The responses to my review were appropriate and my comments were all addressed. <https://doi.org/10.7554/eLife.109124.2.sa2>

Reviewer #2 (Public review):

Summary:

This work by Waltner, et. al. provides a comprehensive single cell multiomics analysis of plasticity in gene regulatory networks present in Ewing sarcoma using single cell RNA-sequencing (scRNA-seq) and single cell assay for transposase accessible chromatin with sequencing (scATAC-seq). They find that Ewing sarcoma cell lines models have distinct

patterns of chromatin accessibility compared to non-Ewing sarcoma models, and that there is significant variability across Ewing sarcoma cell lines, and sometimes within a single cell line. These differences across models are linked to 3 distinct gene regulatory modules, 2 of which are present across the range of model systems studied here. The first modules present across models is activated when the fusion is expressed and includes genes enriched for the known EWSR1::FLI1 response element, GGAA microsatellites along with other neural crest transcription factors. The other module primarily consists of genes repressed by EWSR1::FLI1, which are activated in EWSR1::FLI1-low states. Interestingly, EWSR1::FLI1-low cells have already been tied to more migratory and metastatic phenotypes and the data here suggest these cells are more responsive to external signals from TGF- β and this may be mediated through FOSL2-mediated gene regulation. This is a technically rigorous study, with a variety of different analytical techniques used to address similar questions and this approach elevates confidence in the answers provided. This is further strengthened by the diverse set of model systems used, including patient-derived cell lines, cell line xenograft models, patient-derived xenografts, mining available single cell data from patient samples, and validation of the gene modules identified in a larger set of patient microarray samples. In whole, this study provides a valuable resource for understanding heterogeneity, plasticity, and gene expression networks in Ewing sarcoma. This may be a useful resource for future studies of metastatic disease and provide a framework for similar questions in other fusion-driven sarcomas.

Comments on revised version.

The authors have addressed comments from my prior review. Thank you!

<https://doi.org/10.7554/eLife.109124.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

We thank both reviewers for their thoughtful and constructive evaluations of our manuscript. We are pleased that the reviewers recognize the value of our multimodal single-cell approach and the diversity of model systems used to define gene regulatory networks underlying heterogeneity in Ewing sarcoma.

Reviewer #1 (Public review):

The investigators elegantly utilized a single-cell co-assay of RNA and ATAC seq to unveil the heterogeneous gene regulatory networks in Ewing sarcoma. The authors should be commended on their ability to identify multiple unique modules of gene regulation of Ewing sarcoma utilizing complex computational methods between numerous Ewing sarcoma cell lines. Additionally, they complemented their single-cell findings with xenografts as well as primary Ewing sarcoma patient tumors - validating the intratumoral heterogeneous gene regulatory networks of Ewing sarcoma. More importantly, they have revealed that exogenous TGF- β may modify these distinct epigenetic and transcriptional signatures within Ewing sarcoma tumors. Overall, the manuscript highlights an important discovery of the heterogenous gene regulatory programming of Ewing sarcoma and further highlights the role that TGFB plays within the tumor microenvironment of Ewing sarcoma. There are some areas of ambiguity that require clarification to increase the impact of the manuscript.

We appreciate Reviewer 1's positive assessment of our work and their recognition of the importance of identifying heterogeneous gene regulatory programs in Ewing sarcoma,

including the role of TGF- β in the tumour microenvironment. We have addressed the areas of ambiguity noted by the reviewer, including clarifying cluster assignments and the selection of $k=3$ for module identification, adding statistical comparisons to relevant figures, correcting figure cross-references, and improving figure labeling for clarity. We have also added higher-resolution images of the spatial profiling data and highlighted relevant correlations between CHLA9 and CHLA10 clusters. We believe these revisions improve the clarity and rigour of the manuscript.

Reviewer #2 (Public review):

Summary:

This work by Waltner et. al. provides a comprehensive single-cell multiomics analysis of plasticity in gene regulatory networks present in Ewing sarcoma using single-cell RNA-sequencing (scRNA-seq) and single-cell assay for transposase accessible chromatin with sequencing (scATAC-seq). They find that Ewing sarcoma cell line models have distinct patterns of chromatin accessibility compared to non-Ewing sarcoma models, and that there is significant variability across Ewing sarcoma cell lines, and sometimes within a single cell line. These differences across models are linked to 3 distinct gene regulatory modules, 2 of which are present across the range of model systems studied here. The first modules present across models are activated when the fusion is expressed and include genes enriched for the known EWSR1::FLI1 response element, GGAA microsatellites, along with other neural crest transcription factors. The other module primarily consists of genes repressed by EWSR1::FLI1, which are activated in EWSR1::FLI1-low states. Interestingly, EWSR1::FLI1-low cells have already been tied to more migratory and metastatic phenotypes, and the data here suggest these cells are more responsive to external signals from TGF- β , and this may be mediated through FOSL2-mediated gene regulation. While there are some minor additional validation studies that can be performed to strengthen a few individual analyses, this is a technically rigorous study, with a variety of different analytical techniques used to address similar questions, and this approach elevates confidence in the answers provided. This is further strengthened by the diverse set of model systems used, including patient-derived cell lines, cell line xenograft models, patient-derived xenografts, mining available single-cell data from patient samples, and validation of the gene modules identified in a larger set of patient microarray samples. In whole, this study provides a valuable resource for understanding heterogeneity, plasticity, and gene expression networks in Ewing sarcoma. This may be useful for future studies of metastatic disease and may also provide a framework for similar questions in other fusion-driven sarcomas.

Strengths:

There are a few core strengths in this study. First is the number and diversity of Ewing sarcoma models studied, spanning commonly used cell lines, patient-derived xenografts, and patient samples. The second is the large array of rigorous and orthogonal approaches used to uncover the identity and function of various gene modules. This includes an array of informatics techniques, as well as specific modulation of cell line models in culture. A third is confirmation that different gene expression programs are present in the same tumor using spatial transcriptomic analysis. Lastly, the authors have made all of their data and code accessible, enabling continued use of this dataset as a resource for others.

Weaknesses:

As highlighted by the authors, this study is somewhat limited by the small number of single-cell data from patient samples that are publicly available. Much of the analysis comes from cell lines. Additionally, they focus only on one type of signal that may modulate cell plasticity, and there are likely to be many others. Lastly, there are a few

weak spots in the data. Some of this likely arises from the underlying complexity of the data, the generally sparse nature of scATAC data, and the biological heterogeneity present in the cell lines studied. The most pronounced weakness was in the analysis of transcription factors that dictate gene expression in the distinct modules, as well as the response to TGF- β . While some specific transcription factors showed module-specific expression consistent with the computational prediction in Figure 2, others did not likely due to additional factors not tested here. Likewise, the same transcription factors did not always show consistent enrichment in the gene modules that responded to TGF- β treatment when analyzed across cell lines. On the whole, these are relatively minor weaknesses and do not diminish the value of this study.

We thank Reviewer 2 for their thorough and balanced assessment. We agree that the study's strengths lie in the breadth of model systems and orthogonal analytical approaches, and we appreciate the reviewer's acknowledgement that the identified weaknesses are relatively minor.

In response to the reviewer's suggestions, we have made several substantive improvements. First, we have included a new Western blot panel in Figure 1 showing EWS::FLI1 protein levels across cell lines, which provides important context for interpreting the long-read fusion transcript detection data. Second, we have revised text throughout the Results section to improve precision — in particular, clarifying that our chromatin analyses assess accessibility at published EWS::FLI1 binding sites rather than binding per se, and ensuring that our stated hypotheses match the metrics presented in the corresponding figures. Third, we have improved figure color schemes and labeling to aid interpretation and corrected errors in panel labeling.

Regarding the reviewer's observation about transcription factor enrichment patterns across cell lines (particularly RUNX3 in the TGF- β response analysis and FOSL2's modest expression at the protein level in CHLA9), we acknowledge that not all TFs showed perfectly consistent module-specific behaviour across every cell line. As the reviewer notes, this likely reflects the underlying biological complexity and additional regulatory factors not tested here. We have made attempts to temper our language accordingly.

We believe that the revised manuscript, with its additional experimental data, improved figures, and clarified text, addresses the concerns raised by both reviewers and strengthens the overall impact of our findings.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

Specific comments:

(1) Figure 1A: Suggest adding cell labels on the UMAP plot - difficult to tell with different colors - which cluster is which cell line.

We have added labels to Figure 1A (UMAP plot) as suggested.

(2) Figure 1B: Although this may have been slightly addressed later in the manuscript, since CHLA9 and 10 are from the same patient, are the authors surprised to see the differences in EWS::FLI1 motif accessibility, or are these findings further reinforcement of inherent heterogeneity? Should one expect the ChromVAR deviation z-score at least to overlap between CHLA9 and 10, since they're from the same patient, if not, can the authors explain why?

We were indeed surprised to see differences in EWS::FLI motif accessibility inferred from our data particularly from the isogenic lines CHLA9 and 10. While it is impossible to know for sure, we suspect that some treatment related changes increased accessibility in CHLA10. However, when considering global accessibility signatures, a subset of cells from CHLA10 (C23) was most correlated with signatures from CHLA9. We have addressed this by stating

“Unsurprisingly, of all cell lines, CHLA10 C23 cells exhibited the greatest correlation to profiles from CHLA9,” in the 11th paragraph of the results section.

(3) Figure 1B: Can the authors comment on the differences in trends of fusion transcript to chromatin enrichment (e.g., A673 and TC71 inverse trends vs. other cell lines that are direct positive or negative trends).

We are hesitant to draw too many conclusions from transcript counting vs chromatin enrichment given some outliers, but in general, we found that high-fusion transcribing cell lines A4573, SKNMC, RDES were associated with module 3, while the lower transcribing cells CHLA9/10, TC32 and PDX305 were associated with module 2. We did not choose to emphasize this correlation too strongly in the paper as there may be other factors such as additional mutations (such as BRAF^{V600E} in A673) that may have either been present in the original tumour or after serial passaging that may play a role.

(4) Figure S1D: Please make the pt. smaller in order to better visualize the differences and highlight the stark differences of the ChromVAR binding site score.

We have adjusted the point size of the embeddings in Fig S1D to enhance visualization.

(5) Figure S1D: Is it surprising to see a large proportion of PDX305 and minor proportions of CHLA10 and CHLA9 with low ChromVAR binding site score?

Our analyses indeed show that PDX305, CHLA9, and CHLA10 utilise fusion-repressed gene programs, so lower enrichment of EWS: FLI1 microsatellites is consistent with our findings.

(6) Figure 2A: Unclear how $k = 3$ was ultimately selected - was it purely a visualization of how well separated each of the cell lines is? Can the authors further clarify how $k = 3$ was determined to be the most optimal? Is there a UMAP that demonstrated a difference in clustering between groups 1, 2, and 3? Is there a threshold cut-off to assign groups into 1, 2, or 3?

Thank you for pointing out this omission. In Fig 1 we show that unsupervised clustering of DA peaks across cell lines grouped EwS cell lines into 2 clusters. When using peak-2-gene linkages (Fig 2), we chose a $k = 3$ to explore the potential genes/CREs explaining the grouping found in Fig 1 because A673 was a clear outlier. We have added the following sentence to the results section paragraph 6. “We selected $k = 3$ to extend the two-cluster structure observed in Fig. 1F–G, reasoning that A673 represented a clear outlier whose distinct regulatory program would be obscured at lower k .”

(7) Figure 2G: Understanding the substantial heterogeneity of each cell and TF binding motifs - given that CHLA9 is considered group 2, is it unexpected that FOSL2 was not highly expressed compared to the other EwS cell lines within group 2: (TC32, CHLA10, PDX305).

Thank you for astutely pointing out that CHLA9 did defy the trend for FOSL2 protein expression compared with other group 2 lines. We specifically did not comment on this finding in the manuscript as we could not account for its status as an outlier. We address this in the public response above.

(8) Figure S3B: Did the authors perform a similar computational analysis (performed for Figure 4), looking specifically at CHLA9 (Clusters 24 and 25) to determine if there are any overlaps with CHLA10 cluster 23's pathway activity/MSigDB/GO: Biological process terms? If there are potential overlaps, can the authors potentially infer tumor clonal evolution from CHLA9 to CHLA10?

While we didn't do pathway analysis, Figure S3 shows strong pearson correlation of C23 from CHLA10 with both C24 and C25 from CHLA9. We have highlighted this with a red box in the figure to make this more obvious to the reader.

(9) Figure 4: Given the heterogeneity of CHLA-10. Did the authors observe any differences in morphology within CHLA-10 between the predominant modules (modules 2 and 3),

given such stark transcriptional heterogeneity?

We did not observe any morphologic differences within CHLA10 but acknowledge this would be an interesting avenue of further investigation.

(10) Figure 5B: Can the authors also plot out module 3 gene expression to see if the cluster enrichment is unique from module 2 gene expression with TGFB1 and vehicle?

We thank the reviewer for this suggestion and have replaced Fig S4B with violin plots so the enrichments are clearer. Module 3 expression in cluster 3 cells from CHLA10 is among the lowest in the cell line.

(11) Figure 5H: Can the authors provide a higher zoom/resolution of the H&E stain of the ROIs in order to see if there are indeed more stroma/fibrosis in ROI9 and ROI10, and if there are differences in tumor cell morphology within different ROIs that harbor different modules?

We have now included higher resolution and magnification of IHC panels in Figure 5H. These images are included as Supplemental Fig 4C. It is notable that although no dramatic differences in tumour cell morphology are visualized, ROI9 and ROI10 comprise small islands of viable tumour surrounded by necrosis.

(12) Figure 6: Within Volchenboun/Lawlor's dataset, of the 46 clinically annotated primary tumors, 10 of the COG samples contained substantial stromal elements, while all the tumors in the European cohort were >70% viable tumors. Can the authors separate out the stromal-rich (n = 10) samples and analyze the 36 tumor-enriched samples to see if the survival curve is the same as what is shown in Figure 6G/H and S4E?

We thank the reviewer for this suggestion. We observed no differences when stratifying the patients as outlined. Indeed, in the original Volchenboun et al, manuscript it was demonstrated that no prognostic gene signature was identifiable when the stromal-rich tumours were removed from the cohort.

(13) Figure 6: Additionally, can the authors run a similar computational analysis to determine the predominant modules within bulk sequencing of the Volchenboun/Lawlor dataset between each tumor?

The computational approach deconvolution was developed for use with bulk RNA-seq and is based on count data. The linear mixture model at the heart of deconvolution methods requires that the measured signal is proportional to abundance across the full range, and Affymetrix microarray data violate that assumption in a gene-specific, nonlinear way that can't be fully corrected post hoc.

(14) Page 15, line 2: Wrong GSE data cited - currently cited as GSE61357 - should be GSE63157 instead.

This has been corrected, thank you.

(15) Please add statistical comparisons for Figure 3E-H.

We thank the reviewer for highlighting this omission. We used the software package ggpubr to perform Wilcoxon rank-sum tests comparing mean module scores between conditions. We have added statistical labels to the plots. All comparisons were statistically to the level indicated in the figure.

(16) Page 11 Line 21: Figure S3 D-E is not about EMT, migration, and TGFB signaling - I believe the authors are referring to Figures 4D-E

Thank you, we have corrected these errors

Reviewer #2 (Recommendations for the authors):

(1) Data

(a) In Figure 1 and the associated text, there is an analysis of cells expressing EWSR1::FLI1 performed using a locus-specific amplification and long-range sequencing. On page 7, lines 1-5, there is some discussion about how some of these track with overall transcript levels, while others don't. Additionally, a very low fraction of cells is shown to be EWSR1::FLI1 positive. This analysis might also be strengthened by a Western blot to show protein levels and how they vary across the cell lines, which may help explain additional differences in the data. The Abcam antibody ab133485 works well for western blotting of EWSR1::FLI1. While this isn't additional single-cell data, the percentage of cells with transcript detected is not really equivalent to the total expression level. This seems particularly valuable to do, as prior publications (Pishas, et. al., Mol. Cancer. Ther., 2018) show that of the cell lines tested here, TC32 has relatively high EWSR1::FLI1 protein levels, while A673 has relatively low expression. This contrasts with the percent of positive cells here.

We thank the reviewer for this suggestion and have included a new panel, Figure 1E containing the western results for cells harvested in log-phase growth.

(b) Figures 5D-F are not obviously referenced in the section about Figure 5 (page 12 line 4 through pg. 13 line 20). One question to pose to the authors is how to interpret the data here for TC71 in light of the fact that they were relatively insensitive to TGF- β . This shows up obviously in the Western blot in 5G.

Thank you, for drawing attention to this omission. We have corrected the figure reference to include Figure 5D-F. In regards to our interpretation of TC71's lack of response to TGF-B, we direct the authors attention to our statement: "The muted response of the TC71 cell line (module-3 dominant) to the influence of TGF-b (Fig. 5A-C, Fig. S4B) suggests that pre-existing transcriptional states may condition the sensitivity to TGF- β signalling."

(c) For the discussion of Figure 5, the authors say on page 12, line 21, that "RUNX3 was enriched in non-responsive clusters." I'm not entirely convinced that the data support this statement as written. This is true in A673s, but there appears to be no difference between the maximally responsive and maximally non-responsive clusters in CHLA10. TC71 was not particularly responsive, but showed the opposite effect.

We agree this was not clearly written. We have de-emphasized RUNX3 findings here. Our new conclusion in the results section paragraph 15 is: "Correlation analyses revealed a reciprocal enrichment pattern of many key TFs from Figure 2, where correlation of FOSL2 (but not RUNX3) gene expression and accessibility was generally highest in TGF- β responsive clusters."

Can data for A673 cells be included for Figure 5G? Like CHLA10, this cell line had the pattern of FOSL2 enrichment that is concordant with that described in the text.

We thank the reviewer for this suggestion and have included A673 in the western blot.

(2) Figures

(a) In Figure 2A, it is very difficult to distinguish the colors for CHLA9 and TC32 in the left panel. Can the color scheme here be changed to make these easier to distinguish?

We thank the reviewer for pointing this out and we have added labels to the UMAP. We hope this makes the visualization more obvious.

(b) Similarly, the KLF4 and SP1 lines in Figure 2D are a little close and might benefit from having more distinct colors.

We thank the reviewer for pointing this out. We have changed the KLF4 to a green hue.

(c) The lower panel of Figure 5I has 2 samples labeled "11" and no sample labeled "12".

Thank you for catching this error. It has been corrected.

(3) Text

(a) One section early in the response was a little bit confusing and could benefit from some revision to improve clarity. On page 6, lines 9-11, this reads a little bit like they looked at cell-line specific EWSR1::FLI1 binding, but that wasn't the assay that was performed. Perhaps there is a better way to describe this than simply "enrichment of EWS::FLI1 sites."

We thank the reviewer for their efforts to improve clarity of our work.

We changed the following sentence in the 2nd paragraph of the result section:

"We visualized the enrichment of EWS::FLI1 sites across all cell lines and discovered distinct EwS cell-line specific usage (Fig. 1C & Fig. S1D)."

And revised to:

"We then assessed chromatin accessibility at these published EWS::FLI1 binding sites across all cell lines and discovered that accessibility at these loci varied in a cell-line-specific manner (Fig. 1C & Fig. S1D)."

(b) Then at the start of the next paragraph (page 6 line 12), the authors talk about "differences in EWS::FLI1 motif and binding site enrichment" and my first thoughts were whether this was differences in which sites were bound or differences in the strength of enrichment. More precise language would be helpful.

Here in the 3rd paragraph of the result section we changed:

"Given the differences in EWS::FLI1 motif and binding site enrichment"

To:

"Given the heterogeneity in the magnitude of EWS::FLI1 motif enrichment"

(c) Related to the comment about EWSR1::FLI1 positive cells vs. protein levels above, the hypothesis on page 6, line 13 says that there was a hypothesis that different Ewing sarcoma lines have different levels of EWSR1::FLI1 transcript. But the metric shown in Figure 2D is the percentage of cells with detectable transcript, not transcript levels. Be specific about what the data are showing here.

We agree with the need to improve clarity. In the 3rd paragraph, we changed: "we hypothesized that EwS cell lines have different levels of EWS::FLI1 transcript."

To:

"We hypothesized that EwS cell lines differ in the proportion of cells expressing high levels of the EWS::FLI1 fusion transcript."

<https://doi.org/10.7554/eLife.109124.2.sa0>