Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution

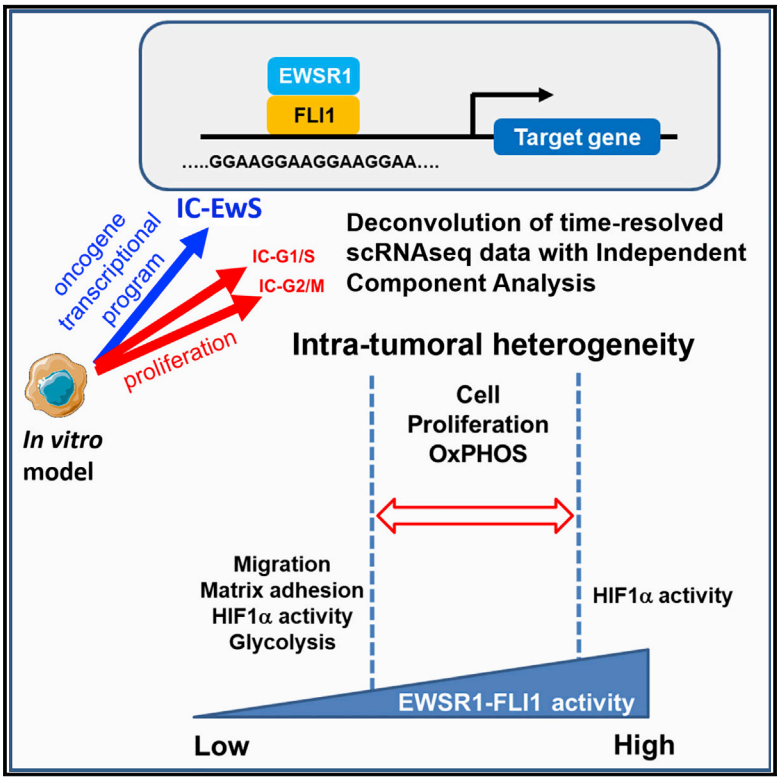
Marie-Ming Aynaud, Olivier Mirabeau, Nadege Gruel, Sandrine Grossetête, Valentina Boeva, Simon Durand, Didier Surdez, Olivier Saulnier, Sakina Zaïdi, Svetlana Gribkova, et al.

Article

Cell Reports

Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution

Graphical Abstract



Highlights

- Time-resolved single-cell RNA-seq of EWSR1-FLI1 induction in an Ewing sarcoma cell line
- Highly specific transcriptomic signature of EWSR1-FLI1 chimeric oncogene
- Sources of transcriptomic heterogeneity in Ewing tumors
- Correlations between EWSR1-FLI1 activity and metabolic and other phenotypic states

Authors

Marie-Ming Aynaud, Olivier Mirabeau, Nadege Gruel, ..., Emmanuel Barillot, Olivier Delattre, Andrei Zinovyev

Correspondence

olivier.delattre@curie.fr (O.D.), andrei.zinovyev@curie.fr (A.Z.)

In Brief

Aynaud et al. reveal a highly specific enhancer-driven transcriptional signature of EWSR1-FLI1 through independent component analysis of time-resolved single-cell RNA-seq analysis. Cell-to-cell variation of the intensity of this signature constitutes a major source of heterogeneity in Ewing tumors and is associated with proliferative, migratory, or metabolic states of the cells.

Aynaud et al., 2020, Cell Reports *30*, 1767–1779
February 11, 2020
https://doi.org/10.1016/j.celrep.2020.01.049

CellPress

Cell Reports
Article

Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution

Marie-Ming Aynaud,[1,14,20] Olivier Mirabeau,[1,20] Nadege Gruel,[1,2] Sandrine Grossetête,[1] Valentina Boeva,[3,4,5,6,7] Simon Durand,[1] Didier Surdez,[1] Olivier Saulnier,[1,15,16] Sakina Zaïdi,[1] Svetlana Gribkova,[5,6,7,8] Aziz Fouché,[9] Ulykbek Kairov,[10] Virginie Raynal,[1] Franck Tirode,[11] Thomas G.P. Grünewald,[1,2,17,18,19] Mylene Bohec,[12] Sylvain Baulande,[12] Isabelle Janoueix-Lerosey,[1] Jean-Philippe Vert,[6,13] Emmanuel Barillot,[5,6,7] Olivier Delattre,[1,*] and Andrei Zinovyev[5,6,7,21,*]

[1]INSERM U830, Équipe Labellisée LNCC, SIREDO Oncology Centre, Institut Curie, 75005 Paris, France
[2]Institut Curie, PSL Research University, Department of Translational Research, 75005 Paris, France
[3]Institut Cochin, INSERM U1016, CNRS UMR 8104, Université Paris Descartes UMR-S1016, 75014 Paris, France
[4]Department of Computer Science, Institute for Machine Learning, Swiss Institute of Bioinformatics (SIB), ETH Zurich, 8092 Zurich, Switzerland
[5]INSERM U900, 75005 Paris, France
[6]Mines ParisTech, PSL Research University, CBIO-Centre for Computational Biology, 75006 Paris, France
[7]Institut Curie, PSL Research University, 75005 Paris, France
[8]Université Denis Diderot, 75013, Paris, France
[9]École Normale Supérieure Paris-Saclay, 94230 Cachan, France
[10]Laboratory of Bioinformatics and Systems Biology, Center for Life Sciences, National Laboratory Astana, Nazarbayev University, Nur-Sultan, Kazakhstan
[11]University of Lyon, Université Claude Bernard Lyon 1, CNRS 5286, INSERM U1052, Cancer Research Centre of Lyon, 69008 Lyon, France
[12]NGS Platform, Institut Curie, 75005 Paris, France
[13]Google Research, Brain Team, 75009 Paris, France
[14]Present address: Lunenfeld-Tanenbaum Research Institute, Mount Sinai, M5G1X5, Toronto, ON, Canada
[15]Present address: Developmental & Stem Cell Biology Program, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada
[16]Present address: The Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada
[17]Present address: Max-Eder Research Group for Pediatric Sarcoma Biology, Institute of Pathology, Faculty of Medicine, LMU Munich, Munich, Germany
[18]Present address: German Cancer Consortium (DKTK), partner site Munich, Munich, Germany
[19]Present address: German Cancer Research Center (DKFZ), Heidelberg, Germany
[20]These authors contributed equally
[21]Lead Contact
*Correspondence: olivier.delattre@curie.fr (O.D.), andrei.zinovyev@curie.fr (A.Z.)
https://doi.org/10.1016/j.celrep.2020.01.049

SUMMARY

EWSR1-FLI1, the chimeric oncogene specific for Ewing sarcoma (EwS), induces a cascade of signaling events leading to cell transformation. However, it remains elusive how genetically homogeneous EwS cells can drive the heterogeneity of transcriptional programs. Here, we combine independent component analysis of single-cell RNA sequencing data from diverse cell types and model systems with time-resolved mapping of EWSR1-FLI1 binding sites and of open chromatin regions to characterize dynamic cellular processes associated with EWSR1-FLI1 activity. We thus define an exquisitely specific and direct enhancer-driven EWSR1-FLI1 program. In EwS tumors, cell proliferation and strong oxidative phosphorylation metabolism are associated with a well-defined range of EWSR1-FLI1 activity. In contrast, a subpopulation of cells from below and above the intermediary EWSR1-FLI1 activity is characterized by increased hypoxia. Overall, our study reveals sources of intratumoral heterogeneity within EwS tumors.

INTRODUCTION

Ewing sarcoma (EwS) is a highly aggressive pediatric bone cancer that is defined by a pathognomonic recurrent somatic mutation: a fusion between the *EWSR1* gene and an *ETS* family member, most frequently the *FLI1* gene (Delattre et al., 1992; Grünewald et al., 2018). This leads to the expression of EWSR1-FLI1, an aberrant and potent chimeric transcription factor. EWSR1-FLI1 can act both as a transcriptional activator and as a repressor, depending on the sequences of DNA binding sites and on the presence of additional co-factors (Bilke et al., 2013; Riggi et al., 2014). EWSR1-FLI1 binds to DNA either at ETS-like consensus sites with a GGAA core motif or at GGAA microsatellites (GGAA-mSats), which are diverted by EWSR1-FLI1 as *de novo* enhancers (Gangwal et al., 2008; Guillon et al., 2009;

Riggi et al., 2014). Through binding to these sites, EWSR1-FLI1 has been reported to act directly or indirectly on many key cellular processes including the cell cycle, apoptosis, angiogenesis, metabolism, and cell migration (Grünewald et al., 2018).

EwS is genetically stable and ranks among the tumors with the lowest mutation rates (Brohl et al., 2014; Crompton et al., 2014; Lawrence et al., 2013; Tirode et al., 2014). Indeed, apart from the *EWSR1-FLI1* fusion, EwS harbors only a few other recurrent mutations, at low frequencies: *TP53* (5%–10%), *CDKN2A* (10%), and *STAG2* (15%) (Brohl et al., 2014; Crompton et al., 2014; Grünewald et al., 2018; Huang et al., 2005; Tirode et al., 2014). Despite this remarkable paucity of somatic mutations, EwS is a very aggressive tumor with a strong propensity to progress, metastasize, and resist treatments, suggesting potent adaptation properties of the cancer cells. Recent data suggest that epigenetic (DNA methylation or histone decorations) and transcriptional heterogeneity can play an important role in explaining the mechanisms of such adaptations. Indeed, from a large cohort of EwS tumors, a study on a genome-scale DNA methylation sequencing described consistent DNA hypomethylation at enhancers regulated by EWSR1-FLI1 and strong epigenetic heterogeneity within tumors (Sheffield et al., 2017). Moreover, variable expression of EWSR1-FLI1 was recently suggested as a source of heterogeneity in cell lines and tumors; cells with a high level of EWSR1-FLI1 expression (EWSR1-FLI1^{high}) are highly proliferative, whereas EWSR1-FLI1^{low} cells demonstrate a strong propensity to migrate, invade, and metastasize (Franzetti et al., 2017). EwS, therefore, constitutes an appropriate model to investigate how a single somatic driver mutation may impact critical cell-fate decisions, ultimately leading to tumorigenesis.

The recent development of single-cell ''omics'' technologies facilitates the investigation of intratumoral heterogeneity at the single-cell level, enabling the exploration of cell-to-cell variations in gene expression (Baslan and Hicks, 2017). These approaches can help characterize the origins of genetic and non-genetic heterogeneity, which can modulate cell response to exogenous and endogenous factors such as the activation of cancer driver genes (Almendro et al., 2013). Such approaches can also decipher essential bi- or multi-modalities in the distribution of the expression of the genes regulating the cell fates (Shalek et al., 2013) or the interplay between the progression through the cell cycle and the action of signaling and/or differentiation pathways (Buettner et al., 2015) that cannot be addressed through bulk-cell analysis.

Here, we first used single-cell analysis to explore the dynamics of EWSR1-FLI1-related expression changes at the single-cell level using a time course upon the EWSR1-FLI1 induction. EwS cell transcriptomic profiles were also compared with a set of single-cell profiles from other reference systems chosen by various aspects of similarity to the EwS cell system: time series experiments, cells corresponding to EwS cell-of-origin, or cells of various tumor types. This analysis was combined with the exploration of changes in histone decorations. Overall, this approach allowed us to distinguish generic transcriptional programs that are shared by most scrutinized systems from system-specific, and particularly EwS-specific, transcriptional programs. These components were then used to investigate single-cells from EwS tumors. This two-step approach illuminates the heterogeneity of EwS tumors, distinguishing different cell populations based on EWSR1-FLI1 activity, proliferation, and metabolic characteristics.

RESULTS

Experimental Design for Collecting EwS Single-Cell RNA Sequencing Time-Resolved Profiles

In order to explore the dynamics of individual cell transcriptomes under *EWSR1-FLI1* expression, we used the previously developed A673/TR/shEF inducible cellular model derived from the A673 EwS cell line, where the expression of *EWSR1-FLI1* can be modulated through doxycycline-controlled short hairpin RNA (shRNA) (Carrillo et al., 2007). Following a down-modulation of *EWSR1-FLI1* (*EWSR1-FLI1*^{low}) by 7 days of doxycycline (DOX) treatment, we performed a time course experiment after the removal of DOX from the medium, leading to *EWSR1-FLI1* re-expression. Using the C1 Single-Cell Auto Prep System (Fluidigm), we made single-cell transcriptome measurements at seven time points (days 7: *EWSR1-FLI1*^{low}, 9, 10, 11, 14, 17, and 22: *EWSR1-FLI1*^{high}) (Figure 1A, left panel). We also tested *in vivo* the impact of *EWSR1-FLI1* on gene expression. From A673/TR/shEF xenografts in a severe combined immunodeficiency (SCID) mouse, single-cell RNA sequencing (scRNA-seq) was conducted without DOX (*EWSR1-FLI1*^{high}) and after 7 days of DOX treatment (*EWSR1-FLI1*^{low}). The modulation of EWSR1-FLI1 protein expression was confirmed by western blot (Figure 1A, right panel). We also conducted scRNA-seq experiments on three EwS patient-derived xenografts (PDXs), established in the laboratory via implantation of tumor samples in SCID mice (Table 1). Finally, we profiled two primary cultures of mesenchymal stem cells (MSCs), the likely EwS cell-of-origin (Tirode et al., 2007) (these datasets are summarized in Table S1).

We performed unsupervised analysis of the EwS single-cell transcriptomic data by principal-component analysis (PCA). The first principal component (PC1; 14.7% of explained variance) clearly separated *in vitro* (A673/TR/shEF time series and EwS MSCs) and *in vivo* (A673/TR/shEF xenograft and EwS PDXs) datasets (Figure 1B). The second principal component (PC2; 4.2% of explained variance) projection revealed the effect of *EWSR1-FLI1* activation on transcriptomic dynamics. For the A673/TR/shEF time series, *EWSR1-FLI1*^{low} cells (d7) were grouped close to MSCs and were clearly separated from *EWSR1-FLI1*^{high} cells (d22). As early as the second (d9) and the third (d10) days of *EWSR1-FLI1* re-expression, the distribution of single-cell transcriptomes was already significantly different from *EWSR1-FLI1*^{low} cells (d7). Four and seven days after re-induction (d11 and d14) represented intermediate (between *EWSR1-FLI1*^{low} and *EWSR1-FLI1*^{high}) states of single-cell transcriptome distributions. Finally, at d17, most of the single cells converged to the *EWSR1-FLI1*^{high} state of d22. Similarly, *EWSR1-FLI1*^{low} and *EWSR1-FLI1*^{high} states of A673/TR/shEF xenografts could be clearly distinguished. All *EWSR1-FLI1*^{high} cells, including the three PDXs, converged in the transcriptomic space (Figure 1B). The first component was not significantly enriched with any Gene Ontology (GO) gene set, while the second principal component was associated with functional

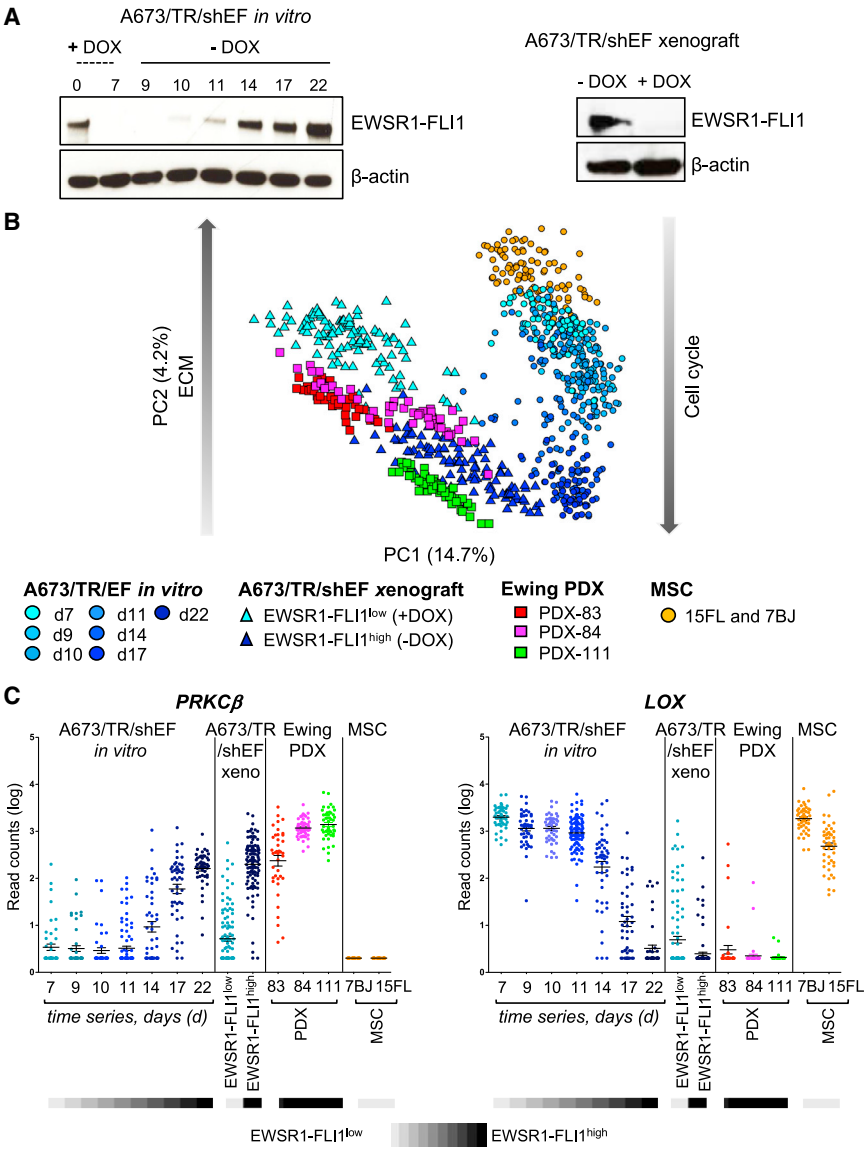


Figure 1. Ewing Sarcoma Single-Cell Transcriptomic Datasets

(A) A673/TR/shEF *in vitro* (left panel): after 7 days of DOX treatment, cells are extensively rinsed; grown in a DOX-free medium for 2 days (d9), 3 days (d10), 4 days (d11), 7 days (d14), 10 days (d17), or 15 days (d22); and isolated for scRNA-seq. The strongly decreased expression of EWSR1-FLI1 at d7 and the re-expression of EWSR1-FLI1 is controlled by western blot. A673/TR/shEF xenograft (right panel): A673/TR/shEF were injected subcutaneously in SCID mice. After 18 days, mice were separated into two groups and treated or not with DOX added to the drinking water for 7 days. EWSR1-FLI1 expression was controlled by western blot.

(B) PCA of single-cell EwS datasets. Cells are indicated by colored circles: from d7 (light blue circles) to d22 (dark blue dot), 348 cells; A673/TR/shEF xenograft cells with EWSR1-FLI1 (dark blue triangle) or without EWSR1-FLI1 (light blue triangle), 215 cells; 3 EwS PDXs (red, magenta, and green square), 142 cells; and MSCs (orange circles), 96 cells.

(C) Expression of two genes known to be modulated by EWSR1-FLI1 plotted for A673/TR/shEF *in vitro*, A673/TR/shEF xenograft, EwS PDXs, and MSCs. Left panel: *PRKCβ* expression, upregulated by EWSR1-FLI1. Right panel: *LOX* expression, downregulated by EWSR1-FLI1. The gray scale on the bottom illustrates the EWSR1-FLI1 level of expression.

See also Figures S1 and S2 and Table S1.

enrichment analysis to EWSR1-FLI1-modulated genes (taken from Kinsey et al., 2006; positive correlation, $p < 10^{-65}$), cell cycle (GO: 0007049, positive correlation, $p < 10^{-26}$), and extracellular matrix (ECM) organization (GO: 0030198, negative correlation, $p < 10^{-25}$).

We also checked the single-cell expression dynamics of eight genes known to be directly modulated by *EWSR1-FLI1* (upregulated genes: *PRKCB*, *LIPI*, *CCND1*, and *NR0B1*; downregulated genes: *IGFBP3*, *IL8*, *LOX*, and *VIM*) (Figures 1C and S1). These results confirmed consistent re-induction dynamics of *EWSR1-FLI1*. Single-cell expression of these genes highlights early and late responsive cells to *EWSR1-FLI1* re-expression at any given time point (Figures 1C and S1).

Collectively, these analyses show that these EwS single-cell transcriptome datasets recapitulated the main results found previously in bulk expression measurements in similar biological systems. However, just as in the bulk data, they do not distinguish between EWSR1-FLI1 direct and indirect transcriptional effects.

Joint Deconvolution of Multiple scRNA-seq Datasets into Independent Components

In order to create a negative control to the EwS-specific datasets and to evaluate the specificity of the sources of cellular heterogeneity, we jointly normalized and merged the EwS-specific single-cell datasets with several other single-cell datasets generated in-house or obtained from public resources (Patel et al., 2014; Trapnell et al., 2014). Altogether, we analyzed 1,964 single-cell transcriptomic profiles from eight different datasets (Table S1). A t-Distributed Stochastic Neighbor Embedding (t-SNE) plot of all cells is shown in Figure S2. The different EwS-specific datasets are grouped together, separated from the other datasets. In both *in vitro* and xenograft cases, cells in which the *EWSR1-FLI1* oncogene has been induced converge to the same area at the center of the plot. Cells of mesenchymal origin (MSCs and myoblasts) are localized close to each other in the plot.

We applied independent component analysis (ICA) to decompose the heterogeneity of scRNA-seq profiles into a relatively small number (30, as it was close to the estimated optimal number;

Table 1. PDX Samples and Sequencing Technologies Used for Single-Cell RNA-Seq Analyses, Related to Table S1

Model/ Patient	PDX ID	PDX Passage	Age at Graft (Weeks)	Days to Observe Tumor Growth	Fusion Type	Localization and Time of Sampling of the Primary Tumor		Secondary Genetic Alterations	Sequencing Technology
IC-pPDX-3	PDX-84	8	13	40	EWSR1 ex7/ FLI1 ex6	humerus	localized, at diagnosis	*CDKN2A*^{DEL}	C1 Fluidigm
	PDX-184	14	N/A	5	EWSR1 ex7/ FLI1 ex6	humerus	localized, at diagnosis	*CDKN2A*^{DEL}	Chromium 10x Genomics
IC-pPDX-5	PDX-111	4	12	82	EWSR1 ex10/ FLI1 ex8	tibia	metastatic, at diagnosis	NA	C1 Fluidigm
IC-pPDX-8	PDX-83	0	16	210	EWSR1 ex7/ FLI1 ex6	sacrum	localized, at diagnosis	*STAG2*^{R614*}	C1 Fluidigm
	PDX-352	5	7	63	EWSR1 ex7/ FLI1 ex6	sacrum	localized, at diagnosis	*STAG2*^{R614*}	Chromium 10x Genomics
IC-pPDX-52	PDX-861	1	17	148	EWSR1 ex7/ FLI1 ex6	sacrum/ ilium	metastatic at diagnosis	NA	Chromium 10x Genomics
IC-pPDX-80	PDX-856	3	19	20	EWSR1 ex7/ FLI1 ex6	chest	localized, at relapse	NA	Chromium 10x Genomics
IC-pPDX-87	PDX-1058	12	10	11	EWSR1 ex7/ FLI1 ex6	sacrum/ ilium	metastatic, at relapse; same patient as IC-pPDX-52	*CDKN2A*^{DEL}; *TP53*^{R175C}	Chromium 10x Genomics

see Method Details) of independently acting factors or independent components (ICs). The rationale for choosing this approach is that ICA can, in principle, deconvolute co-linear signals (Sompairac et al., 2019; Zinovyev et al., 2013). For each IC, the analysis associated a weight for each gene (collectively denoted as metagene) and a score for each sample (denoted as metasample).

We then computed Pearson's correlation coefficients among 30 metasamples and the point biserial correlation coefficients between meta-samples and binary vectors describing different cell subsets (e.g., the Ewing cell subset was described as a vector with value 1 for each cell of EwS origin and 0 otherwise). This correlation matrix was then used to produce a two-dimensional PCA plot that positioned the different ICs and sample groups relative to each other. This analysis enabled us to distinguish generic and cell-type-specific independent sources of heterogeneity (Figure 2A; Table S2), including EwS-specific ICs (IC10 and, to a lesser degree, IC30) as well as components specific to other cell types (Figure 2A). In addition, ICA deconvolution leads to the identification of components not specific to a single cell type, whose correlations with the cell subset binary vectors were small (see bottom-right part of Figure 2A).

Generic and *EWSR1-FLI1*-Specific Components

We then looked for biological processes that could be associated with each of the ICs. For this, we defined two sets of top-contributing genes for each component: one with positive weights and one with negative weights (denoted as ICx+ and ICx−, respectively, where x denotes the component number and ± indicates the positive and negative tails of the weight distribution, respectively), using a threshold of five standard deviations roughly corresponding to a statistical significance of 1%. On these, we then performed gene set enrichment analyses (GSEAs) using the ToppGene suite (Chen et al., 2009) (Table S3; see also http://bioinfo-out.curie.fr/projects/sitcon/mosaic/toppgene_analysis/).This analysis highlighted associations with various generic biological processes—some having remarkably strong enrichment scores—and led us to focus on four gene sets: IC1+, IC2+, IC10+, and IC30+. Thus, IC1+ was associated with chromosome segregation (GO: 0007059; $p = 10^{-80}$) and mitotic nuclear division (GO: 0007067; $p = 10^{-80}$); IC2+ with DNA replication (GO: 0006260; $p = 10^{-69}$); and IC30+ with extracellular matrix organization (GO: 30198; $p = 10^{-14}$). We matched the IC1+ and IC2+ scores to two recently established transcriptomic signatures for the specific phases of the cell cycle (Giotti et al., 2017) and found a strong match between the IC1+ and G2/M scores and between the IC2+ and G1/S scores (Figure S3). Hence, for the sake of clarity, in what follows we will refer to the IC1+, IC2+, and IC30+ gene sets as IC-G2/M, IC-G1/S, and IC-ECM, respectively.

The IC10+ list (220 genes; Table S4) was highly enriched in "genes up-regulated in mesenchymal stem cells engineered to express EWSR1-FLI1 fusion protein" (Riggi et al., 2008) ($p = 10^{-102}$) and several other EwS-related transcriptomic signatures from the MSigDB C2 collection (targets of EWSR1-ETS fusions up [Miyagawa et al., 2008] $p = 10^{-50}$; targets of EWSR1-FLI1 fusion [Hu-Lieskovan et al., 2005] $p = 10^{-48}$; and Ewing family tumor [Staege et al., 2004] $p = 10^{-25}$) and, to a lesser extent, with the EwS gene set (C3536893 entry in DisGeNET database [Piñero et al., 2015] $p = 10^{-9}$). However, unlike previously reported EwS-related gene signatures, IC10+ was not enriched in cell-cycle-related reference gene sets. In the IC10+ gene set, we observed a borderline significance for ossification and axonogenesis, two processes that may be of significance for EwS, which is a bone malignancy for which previous studies have suggested a neural crest origin (Grünewald et al., 2018).

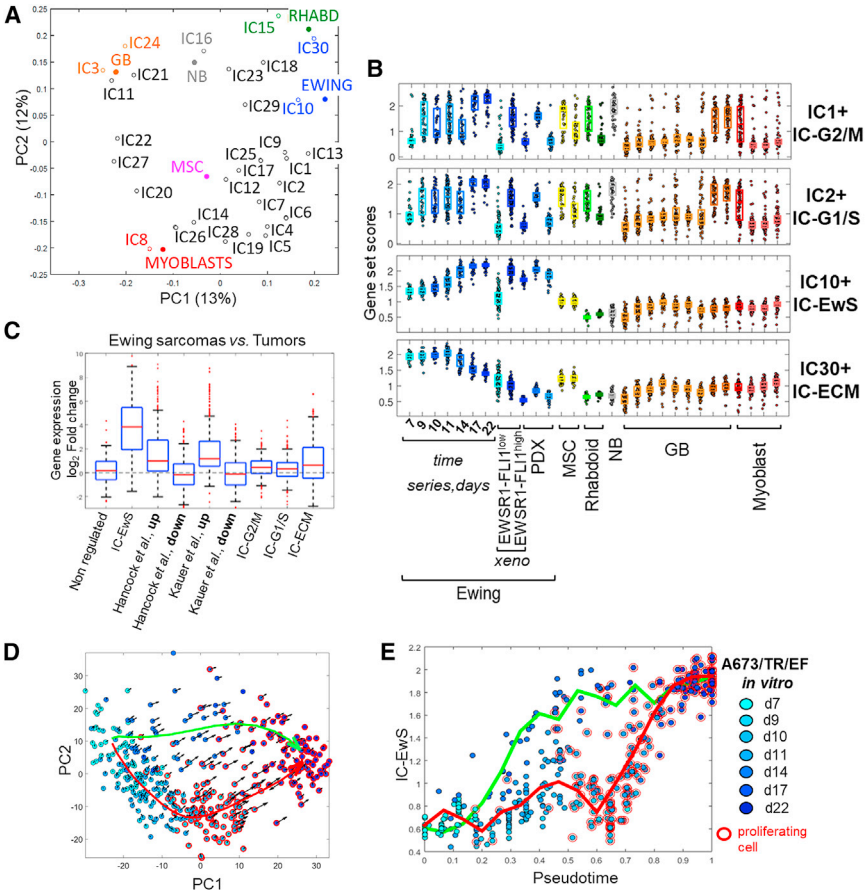


Figure 2. Deconvoluting the Cell-Cycle Phases and Specific Transcriptional Program of EWSR1-FLI1 Activity

(A) PCA plot visualizing the matrix of correlations computed between independent components (ICs) and the binary vectors distinguishing different groups of cells: EwS (blue); neuroblastoma, NB (gray); rhabdoid, RHABD (green); myoblast (red); glioblastoma, GB (orange). Component IC30 appeared closer to the RHABDOID cell subset than to EWING.

(B) Gene set score distribution across all cell datasets for four selected ICs (IC1+: IC-G2/M; IC2+: IC-G1/S; IC10+: IC-EwS; and IC30+: IC-ECM). The scores are computed as the average value of the gene expression of the most variable genes in the set (see Method Details). For the EwS dataset, a blue scale illustrates EWSR1-FLI1 level of expression: from EWSR1-FLI1^{low} (light blue) to EWSR1-FLI1^{high} (dark blue). For the rhabdoid dataset, a green scale demarcates SMARCB1^{-} (light green) from SMARCB1^{+} (dark green). For the GB dataset, GB cell lines are in light orange, and GB tumors are in dark orange. For the myoblast dataset, a red color scale illustrates the myoblast differentiation time course.

(C) Specificity of IC-EwS gene set for EwS. Gene expression analysis is applied on a cohort of 24,364 non-EwS tumors and 156 EwS tumors (all measured by Affymetrix HG-U133Plus 2.0 array). Boxplot of gene expression $\log_2$ fold change of EwS versus other tumors of non-regulated genes (n = 100), IC-EwS genes (n = 220), the up- (n = 503) and downregulated genes (n = 293) described by Hancock and Lessnick (2008), the up- (n = 367) and downregulated genes (n = 252) described by Kauer et al. (2009), and the the IC-G2/M (n = 212), IC-G1/S (n = 291) and IC-ECM (n = 252) genes.

(D) The transcriptional response to EWSR1-FLI1 oncogene activation can be recapitulated by two diverging/converging trajectories of transcriptomic dynamics. The arrows represent RNA velocity vectors. Green and red curves show two branches of cellular trajectories estimated by ElPiGraph.

(E) Dynamics of the activation signature of EWSR1-FLI1 as a function of two branches of pseudotime. Here, green and red values represent the local average score for the cells characterized by a similar value of pseudotime along the two trajectories. We used IC-G1/S and/or IC-G2/M components to define the subset of cells in the proliferating state (see Method Details).

See also Figures S3 and S4 and Tables S2, S3, and S4.

The IC10+ gene set was thus interpreted as highly EwS specific; therefore, from here on, we will refer to it as IC-EwS.

We then assigned a gene set activation score to each cell regarding the different ICs (an average expression of most variable genes in the gene set; see Method Details). The score distributions in this dataset allowed us to make the following conclusions: (1) IC-G2/M and IC-G1/S scores are distributed across all datasets, peaking in the states that can be associated with active proliferation (Figure 2B); (2) within each dataset, IC-G2/M and IC-G1/S scores are highly variable and follow a bi-modal distribution at intermediary points of the time course; (3) IC-EwS and IC-ECM high-score values are almost exclusively associated with EwS cells; and (4) IC-EwS and IC-ECM scores clearly distinguish EWSR1-FLI1^{high} and EWSR1-FLI1^{low} cell states and change monotonically with time, increasing or decreasing, respectively. This is observed in the *in vitro* and xenograft-inducible cellular systems. Figure 2B visualizes IC-G2/M, IC-G1/S, IC-EwS, and IC-ECM scores across the studied datasets.

To further test the specificity of IC-EwS and IC-ECM gene expression in EwS, we performed a gene expression analysis on a cohort of 156 EwSs and 24,364 other tumors (Gene Investigator; Hruz et al., 2008). The IC-EwS gene set strikingly discriminated EwS from all other samples (Figure 2C), more powerfully than did the other two gene signatures, previously defined by bulk transcriptomic data analysis of genes regulated by *EWSR1-FLI1* and enriched in EwS (Hancock and Lessnick, 2008; Kauer et al., 2009). This observation can be explained by the presence of cell-cycle genes in the previously established signatures that limit their specificity toward EwS (Figure S4). This analysis also showed that the IC-ECM gene set is not specific to EwS (Figure 2C).

Altogether, these data highlight that the IC-EwS gene set is highly specific to EwS both in model systems (cell line and PDX) and in tumors.

RNA Velocity and Pseudotime Analysis

RNA velocity analysis was then performed, and the dynamics of single-cell transcriptomes was visualized on a PCA plot

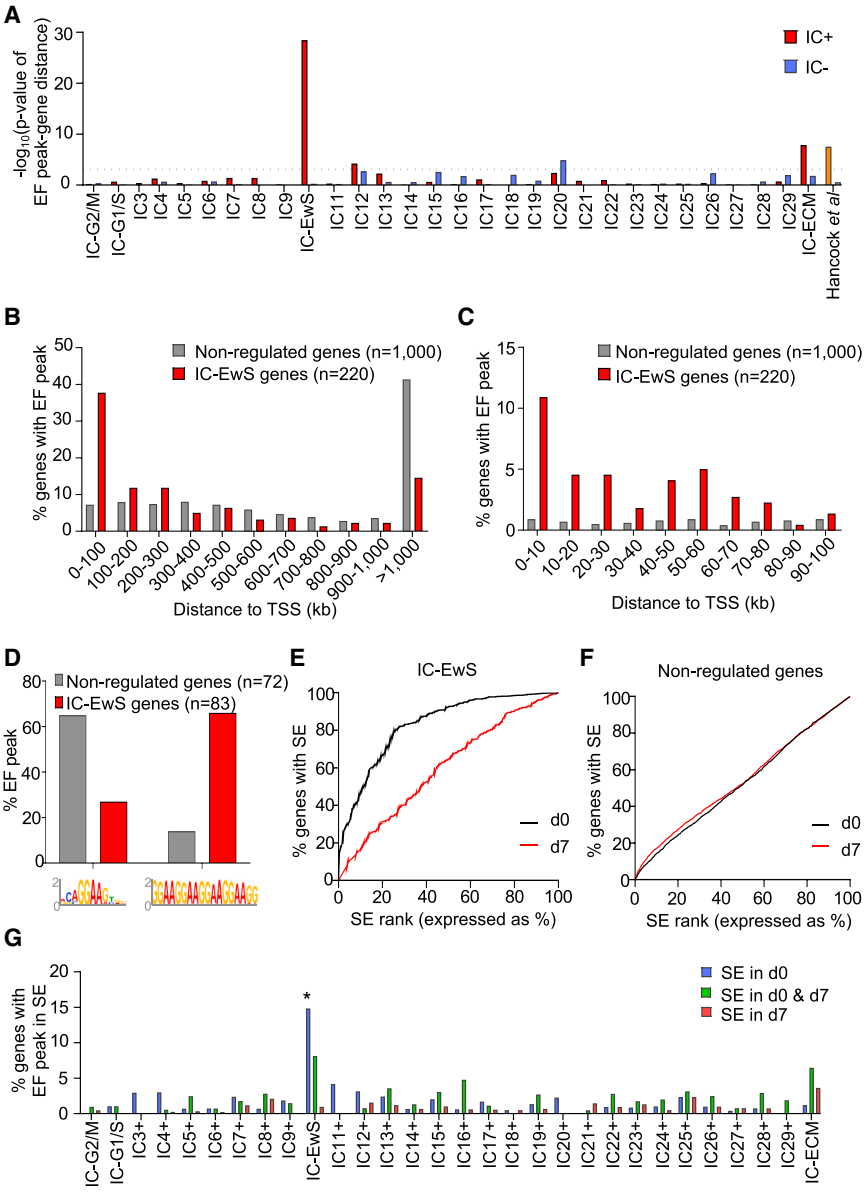


Figure 3. Characterizing the IC-EwS Gene Set

(A) Barplot showing for each IC the log(1/p value) of comparisons of the "gene to EF peak" distance as compared to control genes ("non-regulated genes"; n = 1,000). ICs+ are in red and ICs− in blue. The Hancock and Lessnick (2008) *EWSR1-FLI1* up and down signatures are also shown.

(B) Comparison of the enrichment of genes with EF peaks in IC-EwS (n = 220) as compared to control genes (n = 1,000) from 0 to 1 Mb from the TSS.

(C) Comparison of the enrichment of genes with EF peaks in IC-EwS (n = 220) as compared to control genes (n = 1,000) from 0 to 100 kb from the TSS.

(D) Analysis of the two types of EF peaks (ETS and GGAA-mSats4) in control and IC-EwS genes. Only EF peaks <100 kb of TSS are considered (i.e., 72 peaks for control and 83 for IC-EwS genes).

(E) SE rank curve showing enrichment of top SEs (low rank) in IC-EwS genes at d0 (black line) as compared to d7 (red line).

(F) SE rank curve for the control gene set.

(G) Percentage of genes with EF peaks in SEs (as defined by the ROSE software) for the different ICs. The star indicates the group that is significantly different from the control (Fisher test with Bonferroni correction, $p = 5.10^{-4}$).

See also Tables S4 and S5.

RNA velocities seem to reflect the direction and the amplitude of the rate at which the EWSR1-FLI1 activity changes in individual cells.

Characterization of the Specific EWSR1-FLI1 Activity Signature

To further characterize the IC-EwS signatures, we performed EWSR1-FLI1 chromatin immunoprecipitation sequencing (ChIP-seq) on A673/TR/shEF at d7, d9, d10, d11, d14, and d17 (Figure 1A). EWSR1-FLI1-specific peaks (EF-peaks) were defined as peaks that significantly varied upon EWSR1-FLI1 re-expression ($p < 0.005$).

(Figure 2D). At day 7 and day 22, the transcriptomic dynamics were close to steady state, as expected. These two steady states were connected by two types of transitory dynamics, modeled as two diverging and converging cellular trajectories. The difference between the two trajectories was predominantly explained by the different delays of the induction of the IC-EwS score (Figure 2E). In the main branch (red in Figures 2D and 2E), the induction of IC-EwS approximately coincided with the induction of the IC-G1/S and/or IC-G2/M programs. In the minor branch (green in Figures 2D and 2E), the induction of IC-EwS preceded the proliferation. Genes whose pseudotime behavior was significantly different between these two cellular trajectories belonged almost exclusively to the IC-EwS gene set. Of note, the minor trajectory was less defined, since the delay between the induction of IC-EwS and the proliferation programs was highly variable. The measured

For each gene, we calculated the distance between the transcription start site (TSS) and the nearest EF-peak. We then compared the distribution of these distances for genes of the various ICs to the distribution of distances of a set of 1,000 control genes that are not regulated by EWSR1-FLI1 (Table S4). As shown in Figure 3A, this distance is significantly shorter for IC-EwS genes ($p = 10^{-29}$) as compared to other ICs and to the Hancock and Lessnick (2008) EWSR1-FLI1 signature. Indeed, we observed a highly significant enrichment in the percentage of genes with EF-peaks between 0 and 100 kb from the TSS for IC-EwS (38% of genes) as compared to "non-regulated" genes (< 10%), from which we concluded that many of the IC-EwS genes are likely to be directly regulated by EWSR1-FLI1 (Figures 3B and 3C). A slight enrichment was also observed until 300 kb from the TSS (12%) but not for longer

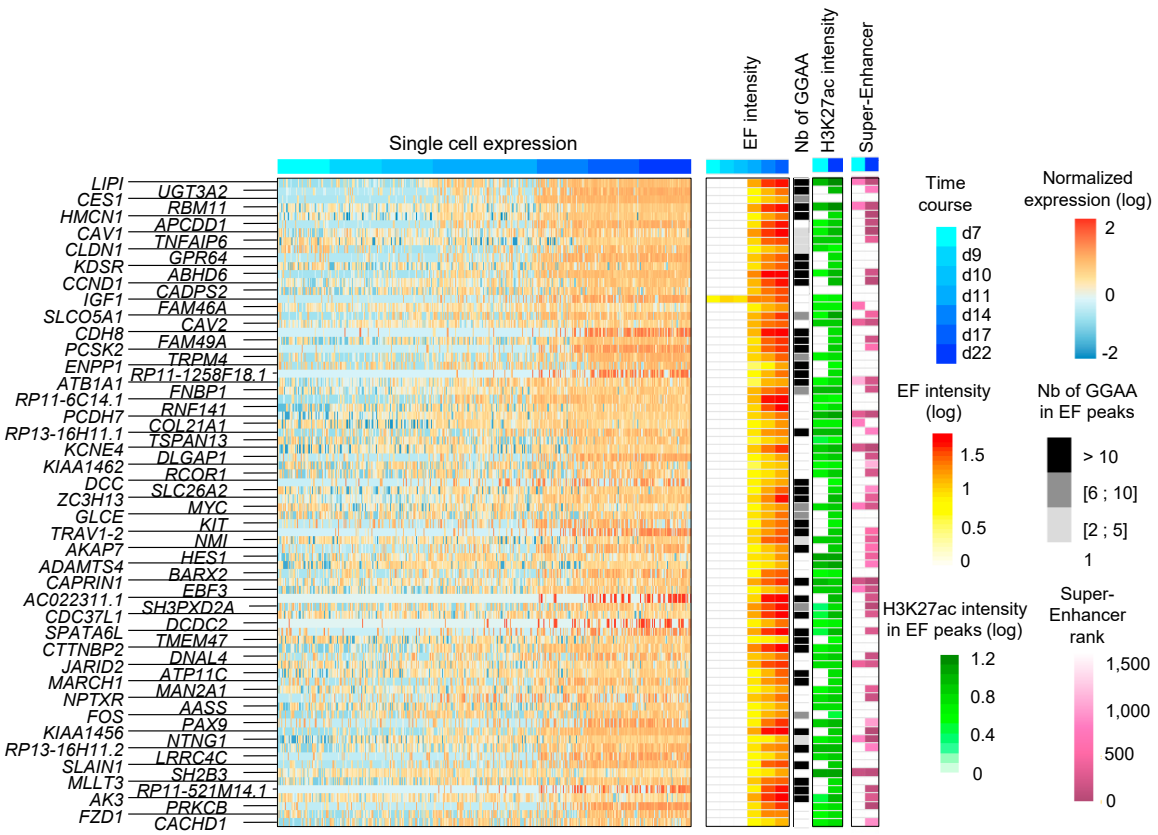


Figure 4. The 78 Genes Identified as Likely Direct EWSR1-FLI1 Targets

From left to right: heatmap of their expression in single-cell normalized expression center scaled and winsorized (5%) in A673/TR/shEF time course from EWSR1-FLI1^{low} (d7, light blue) to EWSR1-FLI1^{high} (d22, dark blue); heatmap of EF-peak intensities in A673/TR/shEF time course from EWSR1-FLI1^{low} (d7, white) to EWSR1-FLI1^{high} (d17, red); number of GGAA in EF-peaks (gray scale); and heatmap of H3K27ac histone mark co-localized with EF peaks (from light green to dark green) and of SE ranking in A673/TR/shEF d0 and d7 (from dark purple to light purple). These genes are ranked by their weights in IC-EwS.
See also Table S4.

distances. TSSs of the Hancock and Lessnick (2008) and the IC-ECM genes are also slightly closer to EF peaks than are the controls but with a much less significant statistical association (Figure 3A). This conclusion remained qualitatively unchanged when 500 or 2,000 control genes were used.

Direct EWSR1-FLI1 target binding sites are shown to be either bona fide ETS motifs or GGAA-mSats (Gangwal et al., 2008; Guillon et al., 2009; Riggi et al., 2014). We used FIMO (Find Individual Motif Occurrences) method (Grant et al., 2011) to analyze the occurrences of these two motifs in EF-peaks located fewer than 100 kb from the TSS of IC-EwS genes (n = 83/220) as compared to control genes (n = 72/1,000). While most EF-peaks of control genes were ETS sites, most EF-peaks of IC-EwS genes contained GGAA-mSats with at least four repeats (GGAA-mSats 4) (Figure 3D).

We also performed ChIP-seq analysis of H3K27ac histone mark to map active chromatin regions at d0 (EWSR1-FLI1^{high}) and d7 (EWSR1-FLI1^{low}). We observed that 91% of EF-peaks are associated with H3K27ac marks at d0, in agreement with previous publications reporting the association of EWSR1-FLI1 binding with active chromatin marks (Boulay et al., 2017; Riggi et al., 2014; Tomazou et al., 2015). When considering only EF-peaks localized in super-enhancer (SE) regions, as defined by the Ranking of Super Enhancer (ROSE) algorithm (Lovén et al., 2013; Whyte et al., 2013), we can define SEs associated with an EF-peak at d0, at d7, and at both time points. We observed that SEs defined at d0 and containing EF-peaks are enriched in the IC-EwS set of genes ($p < 10^{-5}$) (Figure 3G). Moreover, it appeared that the IC-EwS-associated SEs ranked among the strongest SEs (Figure 3E). Such an association is specific to IC-EwS, as it was observed neither for control genes (Figure 3F) nor for any other ICs (data not shown).

Altogether, these analyses allowed us to define the IC-EwS signature as strongly enriched in EWSR1-FLI1 direct target genes. These genes are associated to EF-peaks that (1) significantly vary upon EWSR1-FLI1 expression, (2) are significantly closer to the TSS, (3) are enriched in GGAA-mSat 4, and (4) are significantly enriched in SEs. Given all the analysis steps described above, we defined a set of 78 genes that have most of these characteristics and hence appear to be excellent candidates for being EWSR1-FLI1 direct targets, thus potentially playing key roles in EwS oncogenesis (Figure 4).

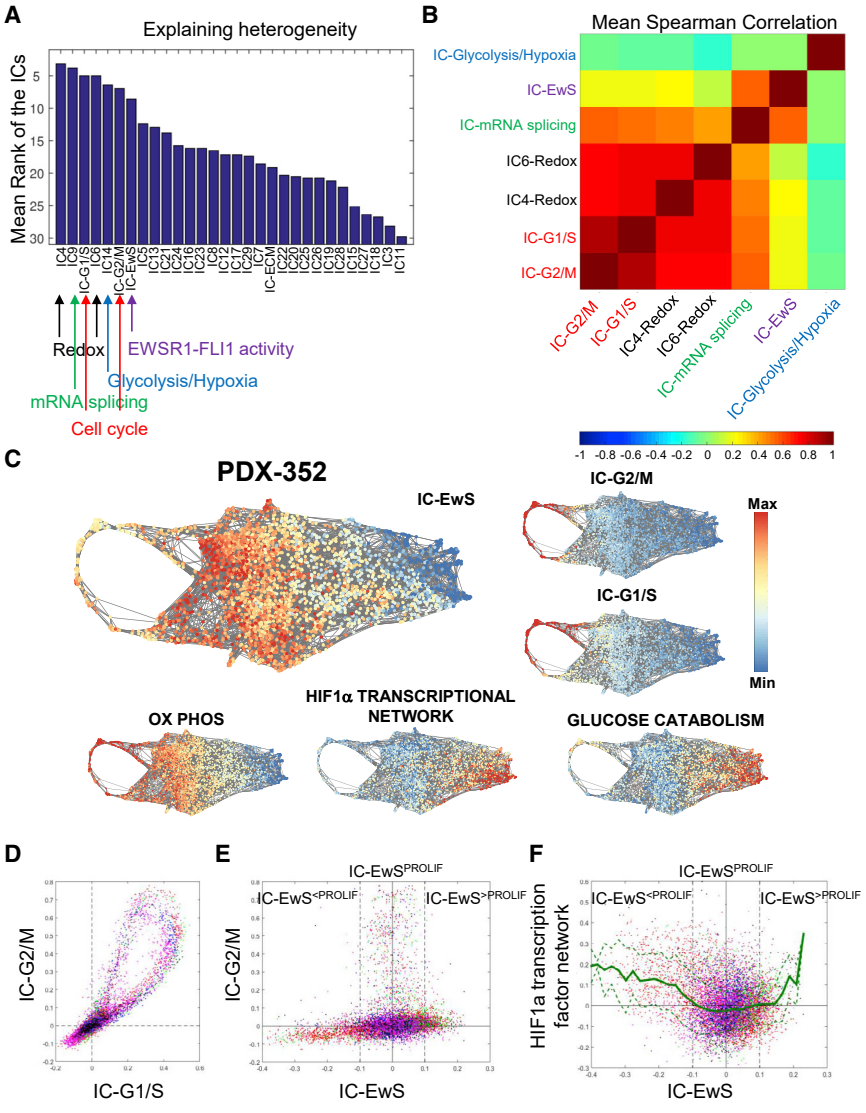


Figure 5. Quantification of PDXs' Heterogeneity Based on Identified Transcriptional Signatures

(A) Ranking of ICs' contributions to intratumoral heterogeneity of expression. In total, 30 ICs (x axis) are ranked based on their ability to explain the gene expression profiles.

(B) Spearman correlations of those ICs that contributed the most to intratumoral expression across the five PDXs.

(C) SPRING representation of the kNN graphs (k = 5) for the PDX-352 dataset. An interactive version of this figure is available at http://bioinfo-out.curie.fr/projects/sitcon/mosaic/SPRING/springViewer.html?datasets/pdx352_nufp.

(D) Visualization of the position of single cells from the five PDXs with respect to the IC-G2/M and IC-G1/S signatures. Scores of non-proliferative cells are centered on zero.

(E) Plot visualizing the connection between IC-EwS and IC-G2/M scores. IC-G2/M cells tend to appear in an optimal range of IC-EwS values ($IC\text{-}EwS^{PROLIF}$). The median of scores of non-proliferative cells is centered on zero for each PDX.

(F) Plot showing the connection between IC-EwS and HIF1α scores. The median of HIF1α scores is centered on zero for each PDX. Solid black line shows the local average of the score, and the dashed lines delimitate the one SD from the mean interval.

See also Figures S5–S7.

Unraveling the Heterogeneity of EwS Tumors at the Single-Cell Level in Tumors

We then investigated whether the aforementioned signatures, defined by *in vitro* systems, may be informative to explore the structure of large single-cell datasets obtained from *in vivo* samples. Five PDXs (PDX-352, PDX-861, PDX-856, PDX-184, and PDX-1058) (Table 1) from *EWSR1-FLI1*-positive EwS were profiled using the 10x genomics sequencing platform. After quality checks and the removal of profiles corresponding to dead cells, a total of 3,595; 1,245; 604; 1,245; and 1,742 scRNA-seq profiles was obtained for PDX-352, PDX-861, PDX-856, PDX-184, and PDX-1058, respectively.

In order to determine which ICs contribute the most to intratumoral heterogeneity, we performed sparse regression analysis of each expression profile on the 30 scores defined by the top-contributing genes of each component. We ranked all components according to their contribution to expression heterogeneity (see Method Details). Seven components were clearly distinguished by this analysis: cell cycle (IC-G1/S and IC-G2/M), mitochondrial biology, oxidative phosphorylation and cellular respiration (IC4+ and IC6+), glycolysis and response to hypoxia (IC14+), mRNA splicing (IC9+), and IC-EwS (Figure 5A). These components showed a consistent pattern of Spearman correlations across the five PDXs (Figures 5B and S6). This pattern suggests a coupling of proliferation and oxidative phosphorylation, coupled activity of EWSR1-FLI1 and mRNA splicing, and a relatively independent role of hypoxia and glycolysis in defining the heterogeneity of EwS tumors.

In order to visualize distances between individual cell transcriptomes, we used the SPRING web-based data visualization interface based on the application of a force-directed graph layout to the graph of similarity between full transcriptomic profiles of individual cells (k Nearest-Neighbor (kNN) graph) (Weinreb et al., 2018). When the IC-EwS score, which can be considered a direct assessment of EWSR1-FLI1 transcriptional activity, was mapped onto the SPRING layout in all PDXs, we indeed observed that this signature largely contributes to the intratumoral heterogeneity (Figures 5C and S5A–S5D). In addition, in all PDXs, IC-G2/M and IC-G1/S define specific groups of cells that form a loop-like structure, most probably reflecting the

Table 2. Genes Upregulated in Cells with High Hypoxia and Low/High IC-EwS Scores as Compared to Cells in the Intermediate IC-EwS^{PROLIF} Range

Gene	Correlation with IC-EwS	IC-EwS$^{<PROLIF}$ versus IC-EwS^{PROLIF}	IC-EwS$^{>PROLIF}$ versus IC-EwS^{PROLIF}
ALDOA	−0.35	++	+++
NDRG1	−0.10	+	++
VEGFA	−0.18	+	++
NRN1	−0.31	+++	++
PGK1	−0.19	+++	++
EGLN3	−0.10	N/S	++
CA9	−0.37	+++	+++
ADM	−0.09	+	N/S
BNIP3	−0.22	+++	++

The labels indicate significantly upregulated genes after regressing out the trend connected to the IC-EwS score. "+++" refers to $p < 10^{-20}$; "++" to $10^{-20} < p < 10^{-10}$; and "+" to $10^{-10} < p < 10^{-5}$. N/S, no significant difference. See also Figure S7.

transcriptional dynamics of the cell-cycle program (Figures 5D and S5A–S5D).

As merging and integrating different tumor single-cell datasets is a well-known challenge in the field of single-cell transcriptomics (Barkas et al., 2019), we decided to perform a tumor-by-tumor analysis. We found that the score distributions for a specific set of genes in our focus of study can be easily aligned (Figures 5D–5F). Using such joint analysis, we focused on some specific relations between biological factors that were generic (such as the cell-cycle stage and the response to hypoxia conditions) and specific to EwS (such as the activity of EWSR1-FLI1 chimera).

The connection between EWSR1-FLI1 activity and proliferation is a well-established fact (Grünewald et al., 2018). Inspection of this relation at single-cell level in EwS indicated that most proliferative cells (in G2/M and/or in G1/S) are clustered within an intermediary range of IC-EwS values. Below and above this range, we observe only a small number of cells in a proliferative state (Figure 5E). As this range of IC-EwS scores also contains non-proliferating cells, we used a Kolmogorov-Smirnov test to confirm that the set of IC-EwS scores from proliferative cells is not a random independent and identically distributed sample from the global distribution of IC-EwS scores ($p < 10^{-22}$). This suggests that an intermediary activity of EWS-FLI1 is associated with cell proliferation (Figure 5E). We thus define IC-EwS^{PROLIF} as a window of IC-EwS scores containing 95% of the total number of proliferative cells (Figure 5E). Other cells, characterized by IC-EwS scores below and above the defined range, and respectively labeled as IC-EwS$^{<PROLIF}$ and IC-EwS$^{>PROLIF}$, were relatively rare (on average, for five PDXs, IC-EwS$^{<PROLIF}$ and IC-EwS$^{>PROLIF}$ groups were 10% and 8% of the total cell number, respectively) (Figure S6).

Using GSEA, it was found that the strongest upregulated signal characterizing the cells outside the IC-EwS^{PROLIF} window was related to hypoxia and glucose catabolism (Figure 5C). Indeed, inspection of the hypoxia score in all PDXs showed that it highlights a subgroup of IC-EwS$^{<PROLIF}$ cells (Figures 5C and S5A–S5D). In all PDXs, this subgroup of IC-EwS$^{<PROLIF}$ cells highly expressed hypoxia-related markers, such as *ALDOA*, *CA9*, *NDRG1*, *VEGFA*, *ADM*, *BNIP3*, and *NRN1* (all members of the HIF1α transcription factor network pathway) (Table 2; Figure S7).

As expected from correlation analyses (Figure 5B), cells residing inside the IC-EwS^{PROLIF} range were characterized by a high oxidative phosphorylation.

A relatively rare population of IC-EwS$^{>PROLIF}$ cells was also characterized by a consistent increase of the HIF1α transcription factor network signature (Figure 5F). However, at the individual gene level, this signal was masked by the increased expression of IC-EwS signature genes. When the IC-EwS score was regressed out from the expression of all genes, only the hypoxia-related signature remained positively enriched in the IC-EwS$^{>PROLIF}$ subpopulation, compared to the rest of the cells. Interestingly, the pro-apoptotic BNIP3 gene was upregulated in both the IC-EwS$^{>PROLIF}$ and IC-EwS$^{<PROLIF}$ cell populations (Table 2; Figure S7). Altogether, these observations suggest an interplay among EWSR1-FLI1 expression level, hypoxia, and apoptosis regulation in cells outside IC-EwS^{PROLIF} range.

DISCUSSION

As with most pediatric cancers, EwS is characterized by a paucity of genetic abnormalities (Gröbner et al., 2018). The accumulation of genetic alterations that frequently result from genetic instability, which are suspected to play major roles in the progression of adult cancers (Bozic et al., 2010), is therefore not expected to be a major player in the progression of pediatric cancers.

Here, based on our recent findings that the variation of expression of *EWSR1-FLI1* constitutes a major source of heterogeneity in EwS (Franzetti et al., 2017), we used a variety of experimental systems to investigate at the single-cell level the gene expression dynamics associated with changes in *EWSR1-FLI1* expression. To our knowledge, this is the first report studying the dynamics of a transcriptome at the single-cell level after the induction of a known cancer driver gene.

We applied unsupervised ICA, which first identified two components that are extremely specific to G1/S and G2/M cell-cycle phases (IC-G2/M and IC-G1/S). These components are not EwS specific and characterize a subset of cells in all experimental systems included in the analysis. In EwS cells, IC-G2/M and IC-G1/S are clearly associated with the expression level of *EWSR1-FLI1.* An IC specific to EWSR1-FLI1 activity was identified that (1) did not highlight any non-EwS cells or tumors in single-cell, tumor bulk, or normal tissue datasets; (2) is strongly enriched by EF1-peaks associated with the presence of GGAA-mSats sequences in the vicinity of the TSS; and (3) is strongly associated with EWSR1-FLI1-dependent SE regions. Based on filtering genes associated with this component, we further identified a set of 78 strong candidate genes for direct regulation by EWSR1-FLI1. Some of these genes— including, but not exclusive to, *CAV1*, *KDSR*, *IGF1*, *MYC*, and *PRKCB*— were previously reported, providing consistency of the IC-EwS signature with previous findings. Previously proposed lists of

EWSR1-FLI1 targets contain numerous cell-cycle genes as a result of the association between *EWSR1-FLI1* induction and cell-cycle gene expression. The IC approach distinguishes the cell-cycle-related ICs that mostly follow an on/off bimodal distribution from the IC-EwS, which shows a monotonous increase upon *EWSR1-FLI1* induction.

Interestingly, as compared to other cell-cycle genes, *CCND1* expression follows the monotonous expression pattern of IC-EwS genes. This reinforces its role as a major player in EWSR1-FLI1-induced activation of the cell cycle. However, a high expression of *CCND1* is not sufficient to induce cell-cycle entry of EwS cells, as many cells with a high expression of *CCND1* do not express G1/S signature. In addition to regulating *CCND1* expression level, previous reports have suggested that EWSR1-FLI1 also regulates the splicing of *CCND1* (Sanchez et al., 2008). Presently, the sensitivity of single-cell transcriptomics is not sufficient to investigate the expression of the different *CCND1* isoforms in individual cells and to establish potential causal relationships between the expression of specific *CCND1* isoforms and cell-cycle entry. Entry into G1/S may also require intrinsic or extrinsic mitogenic signals that remain to be identified in EwS cells.

The identification of the IC-EwS signature constitutes an important improvement as compared to the previously defined EWSR1-FLI1 signatures. When investigated with functional annotation tools, IC-EwS only retrieves weak enrichment annotations as synapses, neurogenesis, or cell adhesion, in agreement with previous observations that EWSR1-FLI1 activates some neural and cell-cell adhesion processes (Franzetti et al., 2017; Hu-Lieskovan et al., 2005). Rather than inducing specific cellular processes, this list contains genes involved in a variety of functions highlighting the pleiotropic effects of EWSR1-FLI1.

The role of length polymorphism of GGAA-mSats sarcoma has been recently documented in the genetic susceptibility to EwS (Grünewald et al., 2015; Machiela et al., 2018). With the identification of a precise set of EWSR1-FLI1 targets regulated by GGAA-mSats described here, the hypothesis of length polymorphisms of GGAA-mSats being also involved in intertumoral heterogeneity can now also be directly tested.

The exploration of five EwS tumors based on these ICs and on the most significant functional reference gene sets they pointed at illuminate some aspects of intratumoral heterogeneity. One distinct group corresponds to actively proliferating cells. The number of cycling cells is variable, from 9% to 30%. These cells demonstrate increased scores for oxidative phosphorylation signatures, in agreement with recent findings that some cyclin/CDK complexes may coordinate mitochondrial respiration and the cell cycle (Wang et al., 2014; Solaki and Ewald, 2018)

We observe a cell-to-cell variability of the IC-EwS signature score, an indirect measure of EWSR1-FLI1 activity. As expected, cells with a low IC-EwS score are not cycling, in agreement with the hypothesis that significant expression of EWSR1-FLI1 is necessary for cell-cycle entry and progression. More surprisingly, cells with the highest IC-EwS scores are not cycling either, suggesting that the proliferation of EwS cells may be induced by an intermediary level of EWSR1-FLI1 expression (called IC-EwS^{PROLIF} in this study).

EWSR1-FLI1 expression is also associated with metabolic heterogeneity. In all tumors, our analyses highlight a subgroup of IC-EwS$^{<PROLIF}$ and IC-EwS$^{>PROLIF}$ cells that are characterized by hypoxia and the upregulation of HIF1α downstream targets. At present, the potential causal relationship between EWSR1-FLI1 activity and hypoxia is unclear. A previous report has shown that hypoxia upregulates EWSR1-FLI1 expression in an HIF1α-dependent manner (Aryee et al., 2010), which may account for the hypoxia expression pattern of cells with the highest IC-EwS expression. To our knowledge, an association between a low EWSR1-FLI1 activity and hypoxia has not been reported previously. However, a recent report described that serum deprivation in association with hypoxia induces migration, matrix degradation, and invasion of EwS cells that are dependent on the SRC kinase activity (Bailey et al., 2016). Though EWSR1-FLI1 expression was not investigated in this report, such characteristics are hallmarks of cells with low EWSR1-FLI1 activity. Investigation of cell-to-cell variations of key factors regulating HIF1α in sarcoma cells, such as YB1 (El-Naggar et al., 2015), and of splicing as well as signal inferences from topological analyses of the interactions of EwS cells with the microenvironment will be necessary to further decipher direct or indirect relationships between EWSR1-FLI1 activity and hypoxia *in vivo*. Interestingly, cell populations with the lowest or highest IC-EwS signatures show an increased expression of the apoptotic and autophagy regulator BNIP3, suggesting a complex interplay among EWSR1-FLI1 expression, hypoxia, apoptosis, autophagy, and mitophagy (Ney, 2015). As these different cell populations may have a strong impact on response to treatment, it will be of interest to follow *in vivo* their evolution in response to therapy. Single-cell technologies, and particularly spatial transcriptomics, will constitute essential approaches to investigate the interplay between cell-intrinsic and cell-extrinsic factors in this heterogeneity (Smith and Hodges, 2019).

Our work documents at the single-cell level that EWSR1-FLI1 heterogeneous activity may have dramatic consequences on the biology of the EwS cells. The mechanisms of such heterogeneity remain to be explored. They may not be univocal. Cell-autonomous, stochastic variation of EWSR1-FLI1 expression may constitute one of these mechanisms. Antagonism of EWSR1-FLI1 with signaling pathways, as recently documented for the Wnt/β-catenin pathway, may also be critical (Pedersen et al., 2016). Finally, the microenvironment may also be an important source of signals regulating EWSR1-FLI1 stability or activity.

The computational biology approach developed in this study can be applied to other contexts when one can expect strong transcriptional heterogeneity of tumoral cell populations manifested in a continuous fashion, rather than being recapitulated in well-defined clusters of cell subpopulations. ICA appeared to be a powerful approach, allowing us to clearly distinguish the transcriptional proliferation program and the transcriptional program of a known cancer driver. Decomposition of single-cell transcriptomes by ICA allows disregarding the strong proliferation signal in downstream bioinformatics analysis without removing it, which can introduce undesirable biases and cancel out the signal of interest. This is an alternative to the use of other methods that have been developed to ''subtract'' the signal related to the cell-cycle signal from the data (Bacher and Kendziorski, 2016; Barron and Li, 2016), a step that would not be suitable in our study, as the cell cycle and the proper oncogene

transcriptional programs appear to be highly associated. Moreover, ICA allowed fine-grained distinction of cell-cycle phases in a completely unsupervised manner that was not achieved before and identified other important sources of expression-related heterogeneity in a cancer-specific manner. This way of applying ICA is in contrast to most of the previous single-cell ICA applications, aimed either at dimensionality reduction or at cleaning out technical artifacts (Sompairac et al., 2019). Our approach allows using the transcriptional signatures extracted from the model systems for quantifying the heterogeneity of tumors at the single-cell level and comparing tumors without matching the single-cell tumoral transcriptomes, which remains a challenging task.

In conclusion, in this study, we characterize the dynamic effect of EWSR1-FLI1 at the single-cell level. We can distinguish, in an unsupervised and unbiased manner, its oncogene-specific transcriptional program (IC-EwS) from a process strongly coupled to it, the induction of proliferation. The IC-EwS allowed us to describe tumoral heterogeneity in EwS PDXs, highlighting three major populations: one corresponding to the optimal window for cell proliferation activation and two others characterized by lower or higher activity of IC-EwS and associated with hypoxia. Further studies, in particular those using cells from primary tumors, will be needed to assess whether the composition of tumors in these different compartments influences the response to treatment and the prognosis of tumors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Patient-derived xenografts
- METHOD DETAILS
  - Tumor dissociation into single-cell suspension
  - Western blot
  - C1 single cell capture and mRNA-seq
  - 10x Genomics single cell capture and mRNA-seq
  - Chromatin-immunoprecipitation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Alignment, counting and sample normalization of reads
  - Exploratory analysis of scRNA-seq data
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.01.049.

ACKNOWLEDGMENTS

We thank the personnel of the Unité de Génétique Somatique for their help to generate PDX models. This work is supported by ITMO Cancer SysBio program (MOSAIC project, Nb bio 2014-2: Modeling Single cell vAriability in Cancer progression). U.K. acknowledges the Ministry of Education and Science of the Republic of Kazakhstan for the grant research project "Pan-cancer deconvolution of omics data using Independent Component Analysis" (IRN: AP05135430). This work was also supported by grants from the Institut Curie; the INSERM; the Ligue Nationale Contre le Cancer (Equipe labellisée); the Agence Nationale de la Recherche (ANR-10-EQPX-03); the European PROVABES (ERA-649 NET TRANSCAN JTC-2011); the European Union's Horizon 2020 program (grant 826121, iPC project); and the Agence Nationale de la Recherche as part of the "Investissements d'avenir" program (reference ANR-19-P3IA-0001; PRAIRIE 3IA Institute). This research was supported by FP7 grant "EURO EWING Consortium" 602856 and the following associations: Courir pour Mathieu, Dans les pas du Géant, Les Bagouzamanon, Enfants et Santé, M la Vie avec Lisa, Lulu et les Petites Bouilles de Lune, Les Amis de Claire, L'Etoile de Martin, and the Société Française de Lutte contre les Cancers et les Leucémies de l'Enfant et de l'Adolescent. D.S. is supported by the Institut Curie–SIRIC (Site de Recherche Intégrée en Cancérologie) program. The laboratory of T.G.P.G. is supported by grants from the German Cancer Aid (DKH-70112257) and the Gert und Susanna Mayer Foundation. All experimental work and data generation was conducted at Institut Curie, France.

AUTHOR CONTRIBUTIONS

A.Z. and O.D. designed the study. M.-M.A., O.M., S. Grossetête, N.G., A.Z., and O.D. wrote the manuscript. M.-M.A, N.G., and S.D. performed the single-cell experiments and analyses. M.-M.A, S.Z., and D.S. investigated the chromatin marks. N.G. and S.Z. performed *in vivo* experiments. NG, S.Z. and D.S. performed the xenograft and PDX experiments. O.M., A.Z., S. Grossetête, O.S, S. Gribkova, A.F., J.-P.V., U.K., V.B., and E.B. performed the bioinformatics and the statistical analyses. V.R., M.B., and S.B did all the sequencing experiments. F.T. and I.J.-L. provided important materials or data used in this study. T.G.P.G. established MSC lines.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 19, 2019
Revised: October 7, 2019
Accepted: January 15, 2020
Published: February 11, 2020

REFERENCES

Albergante, L., Mirkes, E.M., Chen, H., Martin, A., Faure, L., Barillot, E., Pinello, L., Gorban, A.N., and Zinovyev, A. (2018). Robust And Scalable Learning Of Complex Dataset Topologies Via Elpigraph. arXiv, arXiv:1804.07580. https://arxiv.org/abs/1804.07580.

Almendro, V., Marusyk, A., and Polyak, K. (2013). Cellular heterogeneity and molecular evolution in cancer. Annu. Rev. Pathol. *8*, 277–302.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Aryee, D.N., Niedan, S., Kauer, M., Schwentner, R., Bennani-Baiti, I.M., Ban, J., Muehlbacher, K., Kreppel, M., Walker, R.L., Meltzer, P., et al. (2010). Hypoxia modulates EWS-FLI1 transcriptional signature and enhances the malignant properties of Ewing's sarcoma cells in vitro. Cancer Res. *70*, 4015–4023.

Bacher, R., and Kendziorski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol *17, 63*.

Bailey, K.M., Airik, M., Krook, M.A., Pedersen, E.A., and Lawlor, E.R. (2016). Micro-Environmental Stress Induces Src-Dependent Activation of Invadopodia and Cell Migration in Ewing Sarcoma. Neoplasia *18*, 480–488.

Barkas, N., Petukhov, V., Nikolaeva, D., Lozinsky, Y., Demharter, S., Khodosevich, K., and Kharchenko, P.V. (2019). Joint analysis of heterogeneous single-cell RNA-seq dataset collections. Nat. Methods *16*, 695–698.

Barron, M., and Li, J. (2016). Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data. Sci. Rep. *6*, 33892.

Baslan, T., and Hicks, J. (2017). Unravelling biology and shifting paradigms in cancer with single-cell sequencing. Nat. Rev. Cancer *17*, 557–569.

Bilke, S., Schwentner, R., Yang, F., Kauer, M., Jug, G., Walker, R.L., Davis, S., Zhu, Y.J., Pineda, M., Meltzer, P.S., and Kovar, H. (2013). Oncogenic ETS fusions deregulate E2F3 target genes in Ewing sarcoma and prostate cancer. Genome Res. *23*, 1797–1809.

Biton, A., Bernard-Pierrot, I., Lou, Y., Krucker, C., Chapeaublanc, E., Rubio-Pérez, C., López-Bigas, N., Kamoun, A., Neuzillet, Y., Gestraud, P., et al. (2014). Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. Cell Rep. *9*, 1235–1245.

Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., et al. (2017). Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. Cell *171*, 163–178.e119.

Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. Proc. Natl. Acad. Sci. USA *107*, 18545–18550.

Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS Genet. *10*, e1004475.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Cantini, L., Kairov, U., de Reyniès, A., Barillot, E., Radvanyi, F., and Zinovyev, A. (2019). Assessing reproducibility of matrix factorization methods in independent transcriptomes. Bioinformatics *35*, 4307–4313.

Carrillo, J., García-Aragoncillo, E., Azorín, D., Agra, N., Sastre, A., González-Mediero, I., García-Miguel, P., Pestaña, A., Gallego, S., Segura, D., and Alonso, J. (2007). Cholecystokinin down-regulation by RNA interference impairs Ewing tumor growth. Clin. Cancer Res. *13*, 2429–2440.

Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. *37*, W305-11.

Chen, H., Albergante, L., Hsu, J.Y., Lareau, C.A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A.N., Bauer, D.E., Aryee, M.J., et al. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat. Commun. *10*, 1903.

Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S., et al. (2014). The genomic landscape of pediatric Ewing sarcoma. Cancer Discov. *4*, 1326–1341.

Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., et al. (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature *359*, 162–165.

Durand, S., Pierre-Eugene, C., Mirabeau, O., Louis-Brennetot, C., Combaret, V., Colmet-Daage, L., Blanchard, O., Bellini, A., Daudigeos-Dubus, E., Raynal, V., et al. (2019). ALK mutation dynamics and clonal evolution in a neuroblastoma model exhibiting two ALK mutations. Oncotarget *10*, 4937–4950.

El-Naggar, A.M., Veinotte, C.J., Cheng, H., Grunewald, T.G., Negri, G.L., Somasekharan, S.P., Corkery, D.P., Tirode, F., Mathers, J., Khan, D., et al. (2015). Translational Activation of HIF1α by YB-1 Promotes Sarcoma Metastasis. Cancer Cell *27*, 682–697.

Franzetti, G.A., Laud-Duval, K., van der Ent, W., Brisac, A., Irondelle, M., Aubert, S., Dirksen, U., Bouvier, C., de Pinieux, G., Snaar-Jagalska, E., et al. (2017). Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. Oncogene *36*, 3505–3514.

Gangwal, K., Sankar, S., Hollenhorst, P.C., Kinsey, M., Haroldsen, S.C., Shah, A.A., Boucher, K.M., Watkins, W.S., Jorde, L.B., Graves, B.J., and Lessnick, S.L. (2008). Microsatellites as EWS/FLI response elements in Ewing's sarcoma. Proc. Natl. Acad. Sci. USA *105*, 10149–10154.

Giotti, B., Joshi, A., and Freeman, T.C. (2017). Meta-analysis reveals conserved cell cycle transcriptional network across multiple human cell types. BMC Genomics *18*, 30.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al.; ICGC PedBrain-Seq Project; ICGC MMML-Seq Project (2018). Author Correction: The landscape of genomic alterations across childhood cancers. Nature *559*, E10.

Grünewald, T.G., Bernard, V., Gilardi-Hebenstreit, P., Raynal, V., Surdez, D., Aynaud, M.M., Mirabeau, O., Cidre-Aranaz, F., Tirode, F., Zaidi, S., et al. (2015). Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. Nat. Genet. *47*, 1073–1078.

Grünewald, T.G.P., Cidre-Aranaz, F., Surdez, D., Tomazou, E.M., de Álava, E., Kovar, H., Sorensen, P.H., Delattre, O., and Dirksen, U. (2018). Ewing sarcoma. Nat. Rev. Dis. Primers *4*, 5.

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

Hancock, J.D., and Lessnick, S.L. (2008). A transcriptional profiling meta-analysis reveals a core EWS-FLI gene expression signature. Cell Cycle *7*, 250–256.

Himberg, J., Hyvärinen, A., and Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. Neuroimage *22*, 1214–1222.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv. Bioinforma. *2008*, 420747.

Hu-Lieskovan, S., Zhang, J., Wu, L., Shimada, H., Schofield, D.E., and Triche, T.J. (2005). EWS-FLI1 fusion protein up-regulates critical genes in neural crest development and is responsible for the observed phenotype of Ewing's family of tumors. Cancer Res. *65*, 4633–4644.

Huang, H.Y., Illei, P.B., Zhao, Z., Mazumdar, M., Huvos, A.G., Healey, J.H., Wexler, L.H., Gorlick, R., Meyers, P., and Ladanyi, M. (2005). Ewing sarcomas with p53 mutation or p16/p14ARF homozygous deletion: a highly lethal subset associated with poor chemoresponse. J. Clin. Oncol. *23*, 548–558.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. *10*, 626–634.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol. *17*, 29.

Kairov, U., Cantini, L., Greco, A., Molkenov, A., Czerwinska, U., Barillot, E., and Zinovyev, A. (2017). Determining the optimal number of independent components for reproducible transcriptomic data analysis. BMC Genomics *18*, 712.

Kauer, M., Ban, J., Kofler, R., Walker, B., Davis, S., Meltzer, P., and Kovar, H. (2009). A molecular function map of Ewing's sarcoma. PLoS ONE *4*, e5415.

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

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

Machiela, M.J., Grünewald, T.G.P., Surdez, D., Reynaud, S., Mirabeau, O., Karlins, E., Rubio, R.A., Zaidi, S., Grossetete-Lalami, S., Ballet, S., et al. (2018). Genome-wide association study identifies multiple new loci associated with Ewing sarcoma susceptibility. Nat. Commun. *9*, 3184.

Medjkane, S., Novikov, E., Versteege, I., and Delattre, O. (2004). The tumor suppressor hSNF5/INI1 modulates cell growth and actin cytoskeleton organization. Cancer Res. *64*, 3406–3413.

Miyagawa, Y., Okita, H., Nakaijima, H., Horiuchi, Y., Sato, B., Taguchi, T., Toyoda, M., Katagiri, Y.U., Fujimoto, J., Hata, J., et al. (2008). Inducible expression of chimeric EWS/ETS proteins confers Ewing's family tumor-like phenotypes to human mesenchymal progenitor cells. Mol. Cell. Biol. *28*, 2125–2137.

Ney, P.A. (2015). Mitochondrial autophagy: Origins, significance, and role of BNIP3 and NIX. Biochim. Biophys. Acta *1853* (*10 Pt B*), 2775–2783.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Pedersen, E.A., Menon, R., Bailey, K.M., Thomas, D.G., Van Noord, R.A., Tran, J., Wang, H., Qu, P.P., Hoering, A., Fearon, E.R., et al. (2016). Activation of Wnt/β-Catenin in Ewing Sarcoma Cells Antagonizes EWS/ETS Function and Promotes Phenotypic Transition to More Metastatic Cell States. Cancer Res. *76*, 5040–5053.

Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., and Furlong, L.I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford) *2015*, bav028.

Riggi, N., Suvà, M.L., Suvà, D., Cironi, L., Provero, P., Tercier, S., Joseph, J.M., Stehle, J.C., Baumer, K., Kindler, V., and Stamenkovic, I. (2008). EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. Cancer Res. *68*, 2176–2185.

Riggi, N., Knoechel, B., Gillespie, S.M., Rheinbay, E., Boulay, G., Suvà, M.L., Rossetti, N.E., Boonseng, W.E., Oksuz, O., Cook, E.B., 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.

Sanchez, G., Delattre, O., Auboeuf, D., and Dutertre, M. (2008). Coupled alteration of transcription and splicing by a single oncogene: boosting the effect on cyclin D1 activity. Cell Cycle *7*, 2299–2305.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Sheffield, N.C., Pierron, G., Klughammer, J., Datlinger, P., Schönegger, A., Schuster, M., Hadler, J., Surdez, D., Guillemot, D., Lapouble, E., et al. (2017). DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. Nat. Med. *23*, 386–395.

Smith, E.A., and Hodges, H.C. (2019). The Spatial and Genomic Hierarchy of Tumor Ecosystems Revealed by Single-Cell Technologies. Trends Cancer *5*, 411–425.

Solaki, M., and Ewald, J.C. (2018). Fueling the Cycle: CDKs in Carbon and Energy Metabolism. Front. Cell Dev. Biol. *6*, 93.

Sompairac, N., Nazarov, P.V., Czerwinska, U., Cantini, L., Biton, A., Molkenov, A., Zhumadilov, Z., Barillot, E., Radvanyi, F., Gorban, A., et al. (2019). Independent Component Analysis for Unraveling the Complexity of Cancer Omics Datasets. Int. J. Mol. Sci. *20*, E4414.

Staege, M.S., Hutter, C., Neumann, I., Foja, S., Hattenhorst, U.E., Hansen, G., Afar, D., and Burdach, S.E. (2004). DNA microarrays reveal relationship of Ewing family tumors to both endothelial and fetal neural crest-derived cells and define novel targets. Cancer Res. *64*, 8213–8221.

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

Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetête-Lalami, S., Rusch, M., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project and the International Cancer Genome Consortium (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. *4*, 1342–1353.

Tomazou, E.M., Sheffield, N.C., Schmidl, C., Schuster, M., Schönegger, A., Datlinger, P., Kubicek, S., Bock, C., and Kovar, H. (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.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Van der Maaten, L.H., and Geoffrey, E. (2008). Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research *9*, 2579–2605.

Wang, Z., Fan, M., Candas, D., Zhang, T.Q., Qin, L., Eldridge, A., Wachsmann-Hogiu, S., Ahmed, K.M., Chromy, B.A., Nantajit, D., et al. (2014). Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. Dev. Cell *29*, 217–232.

Weinreb, C., Wolock, S., and Klein, A.M. (2018). SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics *34*, 1246–1248.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zinovyev, A., Kairov, U., Karpenyuk, T., and Ramanculov, E. (2013). Blind source separation methods for deconvolution of complex signals in cancer biology. Biochem. Biophys. Res. Commun. *430*, 1182–1187.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-FLI1	Abcam	Cat# ab133485; RRID:AB_2722650
Mouse monoclonal anti-beta-actin	Sigma Aldrich	Cat# A-5316; RRID:AB_476743
Mouse anti-IgG horseradish peroxidase coupled	Amersham Bioscience	RPN4201
Rabbit anti-IgG horseradish peroxidase coupled	Amersham Bioscience	RPN4301
Rabbit polyclonal anti-FLI1	Abcam	Cat# ab15289-ChIP grade; RRID:AB_301825
Rabbit Anti-Histone H3K27ac	Abcam	Cat# ab4729-ChIP grade; RRID:AB_2118291
Rabbit IgG	Diagenode	Cat# C15410206 ChIP grade; RRID:AB_2722554
Rabbit anti-CTCF	Diagenode	Cat# C15410210 ChIP grade; RRID:AB_2753160
Biological Samples		
Patient-derived xenografts (PDX)	Institut Curie (see Table 1)	n/a
Chemicals, Peptides, and Recombinant Proteins		
Zeocin	Invitrogen	R25001
Blasticidin	Invitrogen	LSR21001
Hygromycin B	Life Technology	10687010
G418 (geneticin)	Invitrogen	10131035
Doxycycline	Invitrogen	BP26535
MSC-qualified serum GIBCO	GIBCO	12662029
L-glutamine	GIBCO	21051024
bFGF	Sigma	SRP3043
collagenase D	Roche	11088858001
hyaluronidase	Sigma	H1115000
DNase	Sigma	10104159001
Anti-protease cocktail	Roche	04693116001
Chemiluminescence	Pierce	32106
Paraformaldehyde 16%	Electron Microscopy Sciences	15710
Critical Commercial Assays		
SMARTer Ultra Low RNA kit from Illumina sequencing	Clontech	634853
Chromium Single Cell 3′ Library & Gel Bead Kit v2	10x Genomics	PN-120237
iDeal ChIP-seq kit for Transcription Factors	Diagenode	50001
Deposited Data		
scRNaseq and ChIPseq on the inducible system and PDXs	in the paper	Gene Expression Omnibus: GSE130025
Experimental Models: Cell Lines		
A673/TR/shEF	Carrillo et al., 2007	n/a
I2A	Medjkane et al., 2004	n/a
CLB-berlud	Durand et al., 2019	n/a
Experimental Models: Organisms/Strains		
Swiss Nude (for IC-pPDX-3, IC-pPDX-5, IC-pPDX-87)	Charles River	female mice Crl:NU(Ico)-Foxn1nu
CB17 SCID (for IC-pPDX-8)	Charles River	female mice CB17/Icr-Prkdcscid/IcrIcoCrl)

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NSG (for IC-pPDX-80)	Charles River	female mice NOD.Cg-Prkdcscid IL2rgtm1Wjl/SzJ
NSG (for IC-pPDX-52)	Charles River	male mice NOD.Cg-Prkdcscid IL2rgtm1Wjl/SzJ
Oligonucleotides		
CCND1-For (5′GGTGGGAGGTCTTTTTGTTTC3′)	Sigma Aldrich	n/a
CCND1-Rev (5′CACGCAATCCCAGATCAAAAC3′)	Sigma Aldrich	n/a
CDKN1A-For (5′ACTGACTCATCACTACTCCCTC3′)	Sigma Aldrich	n/a
CDKN1A-Rev (5′GTGTGCTATTCCCGCCAG3′)	Sigma Aldrich	n/a
CCND1-For (5′CACAGTGTGGGTATTTCCATCAAGCA 3′)	Sigma Aldrich	n/a
CCND1-Rev (5′GGTGTGTAGGAAAAACAGCTCTCTGGA3′)	Sigma Aldrich	n/a
Sec14L2-For (5′GCCCCCGCTGATGCACTTCC3′)	Sigma Aldrich	n/a
Sec14L2-Rev (5′AAGTGCGCCAGCAGAGCCAG3′)	Sigma Aldrich	n/a
Software and Algorithms		
TopHat (version 2.0.6)	Trapnell et al., 2009	n/a
htseq-count (v. HTSeq-0.5.3p9)	Anders et al., 2015	n/a
Cell Ranger software suite (v. 1.3.1)	10x Genomics	n/a
MATLAB implementations of FastICA	Hyvarinen, 1999	n/a
Icasso, as a part of the BIODICA software available at https://github.com/LabBandSB/BIODICA/	Himberg et al., 2004	n/a
t-SNE analysis	Van der Maaten and Geoffrey, 2008	n/a
SPRING visualization tool	Weinreb et al., 2018	n/a
ToppGene functional analysis tool	Chen et al., 2009	n/a
Bowtie2	Langmead and Salzberg, 2012	n/a
MACS2	Zhang et al., 2008	n/a
scVelo as a part of ScanPy package	Wolf et al., 2018	n/a
ElPiGraph as a part of STREAM package	Chen et al., 2019	n/a

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author Dr Olivier Delattre (olivier.delattre@curie.fr). Questions on the computational methodology should be addressed to the Lead Contact Dr. Andrei Zinovyev (andrei.zinovyev@curie.fr).

This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

All cells are grown at 37°C with 5% of CO_2 with 100 UI/mL Penicillin and 100 μg/mL Streptomycin (GIBCO). A673/TR/shEF (Carrillo et al., 2007) are cultured in DMEM (GIBCO) 10% FBS (Eurobio), with 50 μg/mL Zeocin (Invitrogen), 2 μg/mL Blasticidin (Invitrogen) added *ex-tempo*. I2A cells were grown in RPMI (GIBCO), 10% FBS (Eurobio) with 50 μg/mL hygromycin B (Life Technology), 300 μg/mL G418 (geneticin) and 50 ng/mL DOX (doxycyclin, Invitrogen) added *ex-tempo* when indicated. MSCs from bone marrow EwS patients were isolated by density-gradient centrifugation using Ficoll technique and were cultured in alpha MEM (GIBCO), MSC-qualified serum (GIBCO), 1% L-glutamine (GIBCO) and 1 ng/mL bFGF (Sigma), added *ex-tempo*. CLB-berlud are cultured in RPMI (GIBCO) 10% FBS (Eurobio), 100U/mL.

EWSR1-FLI1 specific small hairpin RNA was induced in A673/TR/shEF cells by adding DOX at 1 ug/mL. After 7 days, DOX was removed and cells were washed three times to allow silencing of the shRNA and induction of *EWSR1-FLI1*. Cells were harvested at seven different times points: 0 day (d7), 2 days (d9), 3 days (d10), 4 days (d11), 7 days (d14), 10 days (d17) and 15 days (d22) after DOX removal. This protocol was used in order to study induction of genes by EWSR1-FLI1 and to avoid a potential effect of DOX on the transcriptome dynamics.

For A673/TR/shEF xenograft, 20 million cells were resuspended in 200 μL of PBS and subcutaneously injected into severe combined immunodeficiency (SCID) mice. When tumor volume reached 1,000 mm3, DOX was added in the drinking water of a subset of mice (+ DOX group) for 7 days.

For I2A cells, DOX was removed and cells were washed three times to induce *SMARCB1* expression.

Patient-derived xenografts

EwS Patient Derived Xenografts (PDX) were established in the laboratory by subcutaneous implantation of tumor samples in immunodeficient female mice, except for IC-pPDX-52 which was male (Table 1). The mice genotypes used in this study were Swiss Nude (Crl:NU(Ico)-Foxn1nu) for IC-pPDX-3, IC-pPDX-5, IC-pPDX-87 models, CB17 SCID (CB17/Icr-Prkdcscid/IcrIcoCrl) for IC-pPDX-8 and NSG (NOD.Cg-Prkdcscid IL2rgtm1Wjl/SzJ) for IC-pPDX-52, IC-pPDX-80 models. The graft location was subscapular (fatpad) at 13 weeks (IC-pPDX-3), 12 weeks (IC-pPDX-5), 10 weeks (IC-pPDX-87), 16 and 7 weeks (IC-pPDX-8), 17 weeks (IC-pPDX-52) and 19 weeks (IC-pPDX-80). Animal care and use for this study were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals. The housing conditions were specific pathogen-free (SPF) for all models.

Experimental procedures were specifically approved by the ethics committee of the Institut Curie CEEA-IC #118 (Authorization APAFIS#11206-2017090816044613-v2 given by National Authority) in compliance with the international guidelines. The establishment of PDX received approval by the Institut Curie institutional review board OBS170323 CPP ref 3272; n° de dossier 2015-A00464-45.

PDX were initially profiled using the Fluidigm™ (PDX-83, PDX-84 and PDX-111) approach and then by the higher throughput 10x Genomics (PDX-184, PDX-352, PDX-856, PDX-861 and PDX-1058) technology when it became available in our institute. Characteristics of PDX are reported in Table 1.

METHOD DETAILS

Tumor dissociation into single-cell suspension

A673/TR/shEF xenografts and EwS PDX were dissected from mice and mechanically dissociated. The finely minced tissue was transferred to a digestion mix consisting of CO_2 independent medium (GIBCO) containing 1 mg/mL collagenase D (Roche), 2 mg/mL hyaluronidase (Sigma) and 25 μg/mL DNase (Sigma), incubated for 45 min at 37 °C and gently resuspended every 10 min. Cell suspension was then filtered using 70 μm and 30 μm cell strainers (Miltenyi Biotec). For A673/TR/shEF xenograft experiments, the tumoral suspension was depleted of infiltrated murine cells using the mouse cell depletion kit from Milteny Biotec. Cells were then adjusted at $1x10^6$ cell/mL in HBSS containing 2 mM EDTA. Viability was measured using trypan blue exclusion.

Western blot

All A673/TR/shEF *in vitro* and xenograft proteins were extracted with RIPA and anti-protease cocktail (Roche). Western blots were hybridized with rabbit monoclonal anti-FLI1 antibody (1:1000, ab133485, abcam) and mouse monoclonal anti-beta-actin (1:10,000, A-5316, Sigma Aldrich). Then, membrane was incubated with anti-mouse/rabbit IgG horseradish peroxidase coupled (1:3,000, Amersham Bioscience). Proteins were detected using chemiluminescence (Pierce).

C1 single cell capture and mRNA-seq

Dissociated cells were captured and processed with the C1 Single-Cell Auto Prep System (Fluidigm™) following the manufacturer's protocol. We started with a cell suspension at a concentration of 0.45×10^6 cells/mL. We identified under the microscope the sites where live single cells were captured. Processing of cells occurred in the C1 instrument to perform steps of cell lysis, cDNA synthesis with reverse transcriptase, and PCR amplification for each cDNA library. Quality of the resulting cDNA was checked using the LabChip GX Touch HT (Perkin Elmer, Waltham, MA).The cDNA synthesis and PCR used reagents from the SMARTer Ultra Low RNA kit from Illumina sequencing (Clontech, Mountain View, CA). After harvest from the C1 device, each cDNA library was tagmented using the Nextera XT DNA Sample Preparation Kit (Illumina). After PCR, cDNA libraries were pooled. All libraries were sequenced with HiSeq2500 (Illumina) using 150 bp paired-end sequencing.

10x Genomics single cell capture and mRNA-seq

Single-cell RNA-seq was performed using the Single Cell 3′GEM Code Single-Cell instrument (10x Genomics, Pleasanton, CA, USA), according to the manufacturer's protocol. Cellular suspension (5,300 cells) was loaded on 10x Chromium instrument to generate 3,000 single-cell GEMs, using the Chromium Single Cell 3′ Library & Gel Bead Kit v2. All libraries were sequenced on Illumina HiSeq 2500 in rapid run mode, using paired-end (PE) 26/98 according to 10x recommendations.

In order to remove profiles corresponding to dead or stressed cells from the analysis of 10x data, mitochondrial percentage score was computed for each cell as the percentage of UMIs captured by the genes from the previously described gene set (Ilicic et al., 2016). In the histograms of this score, a bimodal distribution was observed; therefore, all cells from the higher mode were removed from the analysis. After additional quality checks such as removal of cells with too small total number of UMIs (< 5,000 UMIs per cell,

compared to the median 15,000 number of UMIs per cell) or too high (> 40,000 UMIs), the number of selected cells in each PDX is indicated in the Results section.

Chromatin-immunoprecipitation and sequencing

DNA-protein cross-linking was performed in the presence of 1% of paraformaldehyde on $12x10^6$ cells for each condition during 10 minutes. Cell lysis, chromatin shearing, immunoprecipitation and DNA purification was performed with reagents from iDeal ChIP-seq kit for Transcription Factors (Diagenode, ref: C01010054). Twenty cycles of sonication (30 s high, 30 s off) using TPX tube (Diagenode, ref: 50001) and the Bioruptor (Diagenode) were achieved for chromatin shearing. We took 2 μg of FLI1 rabbit polyclonal antibody (abcam, ab15289-ChIP grade) to perform immunoprecipitation of EWSR1-FLI1 transcription factor and 1 μg of H3K27ac antibody (abcam, ab4729) for histone mark immunoprecipitation. IgG and CTCF ChIP was included as negative and positive control. To check quality of each ChIP reactions, quantitative PCR was realized prior to sequencing on 1/5 of purified DNA. Tested regions correspond to following primers: 1- *CCND1* (5′GGTGGGAGGTCTTTTTGTTTC3′/5′CACGCAATCCCAGATCAA AAC3′); 2- *CDKN1A* (5′ACTGACTCATCACTACTCCCTC3′/5′GTGTGCTATTCCCGCCAG3′); 3- *CCND1* (5′CACAGTGTGGGTATTT CCATCAAGCA 3′/5′GGTGTGTAGGAAAAACAGCTCTCTGGA3′); 4- *Sec14L2* (5′GCCCCCGCTGATGCACTTCC3′/5′AAGTGCG CCAGCAGAGCCAG3′). ChIP and input were sequenced with HiSeq2500 (Illumina) using 100 bp single-end sequencing.

QUANTIFICATION AND STATISTICAL ANALYSIS

Alignment, counting and sample normalization of reads

Reads obtained from sequencing of cells were aligned on the human genome (v. hg19) using TopHat (version 2.0.6) (Trapnell et al., 2009). Reads mapping more than once (parameter -x 1) or having edit distances of more than 3 (-N 3) were discarded.

Counting of reads on annotated genes from the GRCh37 gene build was done using htseq-count (v. HTSeq-0.5.3p9) (Anders et al., 2015) with the following parameters: reads with a quality score less than 10 (-a 10) were discarded and reads partially overlapping with the annotated gene transcript were included in the counts unless they overlapped with another read. In all experiments analyzed the STRANDED = no option was used.

Sample-to-sample normalization was performed by rescaling using DESeq size factors (Love et al., 2014). For all data analyses the number of reads was $\log_{10}(x+1)$ transformed. The EWSR1-FLI1 re-expression experiments on the A673/TR/shEF cells were performed as two series of experiments (days 7, 9, 10, 11 for the first batch, and days 11, 14, 17, 22 for the second batch) which overlapped at day 11. This experimental setting was exploited to interrogate the batch effect after normalization using a linear regression model x ~time.point+batch where x = log2(expression) for each gene. We verified that the number of genes modulated through the batch effect was small compared to the number of genes modulated through the experimental time point, indicating that the batch effect did not affect our results.

In case of 10x Genomics data, the programs ''cellranger mkfastq'' and ''cellranger count'' from the Cell Ranger software suite (v. 1.3.1) provided by 10x Genomics were used for demultiplexing and counting the reads on the reference genome GRCh38. Sample-to-sample normalization was performed using the total number of reads in the log scale. For all data analyses the number of reads was $\log_{10}(x+1)$ transformed. More specifically if X is the count matrix the R code to obtain the normalized matrix X.tpm is the following: median.umi < - median(colSums(X)); X.tpm < - log(t(t(X)/colSums(X))*median.umi+1). For each cell, reads from the k = 5 most similar cells were pooled together to define the new cell measurement, in order to reduce the effect of drop-outs. For pooling, kNN graph was computed on $\log_{10}(x+1)$ transformed data after filtering non-variant genes (variance smaller than 0.01) and reducing the dimension of the data by projecting it into 20-dimensional subspace spanned by the standard PCA components. Pooling the 10x Genomics data served for better visualization of the gene expression patterns on top of the SPRING embeddings but did not affect drastically the main conclusions of the study.

Exploratory analysis of scRNA-seq data

Independent Component Analysis (ICA)

ICA was applied, using FastICA algorithm maximizing non-gaussianity of metagenes (using kurtosis function, symmetric approach and 200 maximum number of iterations) and component stabilization using Icasso algorithm, with an additional procedure for determining the optimal number of independent components (Hyvarinen, 1999; Himberg et al., 2004; Biton et al., 2014; Kairov et al., 2017). In the ICA decomposition X = AS, X is the gene expression (sample versus gene) matrix, A is the (sample versus component) matrix describing the loadings of the independent components, and S is the (component versus gene matrix) describing the weights (projections) of the genes in the components. We used a modified MATLAB implementations of FastICA (Hyvarinen, 1999) and Icasso (Himberg et al., 2004) as a part of the BIODICA software available at https://github.com/LabBandSB/BIODICA/, which contains an algorithm for estimating the optimal number of components to compute (Cantini et al., 2019). Icasso applies FastICA algorithm for finding *m* independent components n = 100 times, and then uses hierarchical clustering to estimate compactness of clusters of the components computed in all runs. The resulting independent components represent medoids of the *m* clusters and are ranked by the reproducibility (cluster compactness) in *n* runs. The orientation of the components was chosen such that the longest tail of the gene projection distributions would correspond to the positive values.

Visualization of correlative relations between the components

In order to visualize the relation between IC metasamples and their relation with specific sample subsets, we first computed a correlation table between them (Table S2). In this table, simple Pearson correlations were computed between metasamples. The cell subsets were represented by binary vectors with 1 indicating the cells within one subset and 0 outside. To associate a metasample to a subset, we've computed the point biserial correlation coefficient, which in this case was equivalent to computing Pearson correlation between a metasample vector and a binary vector. Therefore, each object (either a metasample or a cell subset) was characterized by a vector of correlations with other objects. These vectors were normalized by to the unity L1-norm. Subsequently, a standard PCA analysis was applied to the set of vectors (Figure 2A).

Non-linear multi-dimensional data visualization

t-SNE analysis (Van der Maaten, 2008) was done using R with setting the initial dimension parameter to 100 and the perplexity parameter set to 80.

SPRING visualization was produced through computing the kNN graph ($k = 5$) by applying a standard for SPRING approach (Weinreb et al., 2018) consisting in: 1) filtering genes with the coefficient of variance smaller than 0.05 and the average expression smaller than 0.01, computed for pooled read counts (this filter left from 8 to 9 thousands of genes in our datasets); 2) normalizing the measurements on the library size and 3) reducing the dimension of the dataset to 20 by the standard PCA algorithm.

RNA Velocity and pseudotime computation

RNA Velocity analysis was performed using the scVelo method from the ScanPy Python package (Wolf et al., 2018) with default parameters and specifying groups of cells at day 7 and day 22 for estimating the steady-state parameters of the RNA Velocity model. In order to characterize the diverging/converging pattern of gene expression after induction of EWSR1-FLI1, we fitted the distribution of cells by a closed principal curve using the computeElasticPrincipalCircle function of ElPiGraph R package (Albergante et al., 2018; Chen et al., 2019). Curve fitting was done in the subspace of 30 principal components (i.e., the same subspace where ICA was performed). After the closed principal curve was obtained, two edges at the opposite sides of the data distribution which were characterized by the highest density of cells were cut out, which produced two branches of pseudotime.

Functional enrichment analysis

For interpreting the biological meaning of the sets of top-contributing to each of the ICs genes, we applied the ToppGene functional analysis tool (Chen et al., 2009), limited to reference gene sets no larger than 500 genes (in order to focus on more specific functional categories). The ToppGene analysis was automated through the BIODICA graphical user interface available from https://github.com/LabBandSB/BIODICA/ and recapitulated in the form of an interactive online table http://bioinfo-out.curie.fr/projects/sitcon/mosaic/toppgene_analysis/. The table is organized in two columns reporting the first most enriched functional gene sets for positive and negative part of each IC metagene, in each reference categories (the enriched function is mentioned in the table only if the the Bonferroni-corrected $p < 0.05$ and the number of genes from the function found in the top-contributing list is not smaller than 8). Also the sets of top-contributing genes smaller than 10 were not considered for the enrichment analysis. Each hyperlink in the form of "ICX+/−" leads to a saved detailed enrichment analysis as it was produced by ToppGene. Each hyperlink in the form "X genes" leads to the tested list of top-contributing genes.

The table was further used to select a set of reference signatures for the analysis of the tumor data. Only signatures from GO and Pathway categories enriched with the Bonferroni-corrected $p < 10^{-10}$ were selected for further analysis. On top of this, we added the standard HALLMARK set of transcriptomic signatures from MolSigDB.

For associating the IC-EwS score computed for tumor cells with the reference signatures, we applied the standard pre-ranked GSEA analysis to the t test scores computed between the 10th and 90th percentiles of the IC-EwS score. A classical scoring scheme was used and 1,000 permutations estimating the empirical *p-value*.

Computing and aligning gene set scores

Gene set scores for gene sets were computed in all analyses as average gene expression of the genes composing the signatures, after removal of genes characterized by a small variance (in all analyses, 2000 most variable genes were kept for computing the scores).

In order to compare the gene set scores across the 5 PDXs used in the study, their distributions were standardized. For the conclusions of our manuscript, the most sensitive point was to establish the correct reference point of the IC-EwS score, which would allow its comparison between PDXs. Our assumption was that the proliferation process can be used to establish a reference point for the IC-EwS score, since the proliferation scores and their joint distributions look quite similar across all PDXs (Figure S6). Therefore, as a zero point for all PDXs we chose the median value of the IC-EwS score for the set of proliferative cells, in each PDX separately. Oppositely, for the scores related to the cell cycle, IC-G1/S and IC-G2/M, we chose as a reference point the median value computed for the set of non-proliferative cells. For other scores, such as hypoxia, we compared their relative values with respect to the mean value computed for all cells in a PDX. We underline that the main conclusions drawn in the manuscript did not depend on the exact way the scores were aligned. However, the alignment adjusted for relatively small systematic biases between different PDXs made the data visualization clearer and better illustrated our main statements.

Distinguishing a subset of cells in active proliferation

Proliferating cells were defined based on a IC-G1/S and/or a IC-G2/M score exceeding an empirically determined threshold. In brief, histograms of IC-G1/S and IC-G2/M scores show a bi-modal distribution. In the time-resolved data, these thresholds were set to 1.0 for both scores. For PDX data, the thresholds were set to 0.2 for both scores after aligning the scores as described in the previous section.

Defining non-regulated control gene set

We selected "Non-regulated genes" for which at least 100 reads were detected at d7 and/or d22 and which showed no significant differential expression between d7 and d22 in A673/TR/shEF bulk expression dataset ($0.5 < FC < 2$, $p > 0.01$) ($n = 2,117$). Then, genes were ranked from the lowest to highest fold change. For our analysis, we used the top 100 non-regulated genes for Figure 2C and the top 1,000 non-regulated genes for Figure 3 as a negative control.

ChIP-seq peak detection and annotation

ChIP-seq reads were aligned to the human genome (hg19 version) with Bowtie2 (Langmead and Salzberg, 2012). Peaks were called with MACS2 (Zhang et al., 2008), with option narrow for FLI1 antibody and broad for H3K27ac histone mark. To normalize, we took the input dataset from the same cell line. EWSR1-FLI1 specific peaks were defined as peaks varying upon EWSR1-FLI1 expression ($p < 0.005$). To obtain the *p-value*s for each of the peaks we tested the statistical correlation (lm function in R) between the vectors formed by the EWSR1-FLI1 peak intensities at d7, d9, d10, d11, d14, d17 and the vector c (0, 2, 3, 4, 7, 10). That last vector consists in the number of days of EWSR1-FLI1 re-expression for each of these time points. For each gene, we reported the closest EF-peaks to TSS. We performed a Wilcoxon test to compare the distribution of distances for genes of each IC with the control gene set (Figure 3A). We used FIMO tool (Grant et al., 2011) to scan EF-peaks with ETS motif (JASPAR ID: MA0475.1, $p < 0.1$) and GGAA-mSats (JASPAR ID: MA0149.1, $p < 0.0005$). If several motifs were found, we kept only the best motif. ROSE was used to predict Super-Enhancers from H3K27ac marks (Lovén et al., 2013; Whyte et al., 2013). We applied Fisher's exact test to evaluate the enrichment of EF-peaks in Super-Enhancer (Figure 3D). The Super-Enhancers were associated to the closest expressed gene (Figures 3E and 3F).

Quantifying the contribution of independent component signatures into the expression intratumoral heterogeneity of EwS PDXs

The expression profile of each gene in the PDX dataset was regressed on the scores defined by the top contributing gene sets of 30 independent components identified from the analysis of the inducible system. Elastic net regression was applied with alpha parameter of 0.9, ten-fold cross-validation for optimizing the lambda parameter of lasso, and a maximum of 5 predictors selected. It was counted how frequently a score was selected in the regression. This frequency was used to rank the independent components' contributions to the heterogeneity in each PDX. The mean rank value on all 5 PDXs was used to identify which independent component scores contributed the most to the expression intratumoral heterogeneity.

For the top identified independent component scores, in each PDX, a Spearman correlation matrix was computed.

DATA AND CODE AVAILABILITY

The data from this study was uploaded at GEO under accession number GSE130025.

BIODICA package used to compute stabilized Independent Component Analysis and compute functional enrichment of the top contributing genes is available at https://github.com/LabBandSB/BIODICA/

The code used to process the scRNA-Seq PDX datasets is available at https://github.com/sysbio-curie/EwingSingleCellDataAnalysis