



**HAL**  
open science

# Exploration of Desmoplastic Small Round Cell Tumors Cellular and Molecular Heterogeneity

Clémence Hénon

► **To cite this version:**

Clémence Hénon. Exploration of Desmoplastic Small Round Cell Tumors Cellular and Molecular Heterogeneity. Cancer. Université Paris-Saclay, 2023. English. NNT : 2023UPASL002 . tel-04861205

**HAL Id: tel-04861205**

**<https://theses.hal.science/tel-04861205v1>**

Submitted on 2 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exploration of Desmoplastic Small Round Cell Tumors Cellular and Molecular Heterogeneity

*Exploration de l'Hétérogénéité Cellulaire et Moléculaire des Tumeurs  
Desmoplastiques à Petites Cellules Rondes*

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 582 :

Spécialité de doctorat : aspects moléculaires et cellulaires de la biologie

Graduate School : Life Sciences and Health

Référent : Faculté de médecine

Thèse préparée dans l'unité de recherche **Prédicteurs moléculaires et nouvelles cibles en oncologie** (Université Paris Saclay, INSERM U981, Gustave Roussy), sous la direction de **Sophie Postel-Vinay, MD, PhD**

Thèse soutenue à Paris-Saclay, le 18 janvier 2023, par

**Clémence HENON**

### Composition du Jury

**Pr Fabrice ANDRÉ**

MD, PhD, Directeur d'unité (U981), Directeur de la recherche, Gustave Roussy, Villejuif, FRANCE

Président du jury

**Pr Jean-Yves BLAY**

MD, PhD, Directeur général du Centre Léon Bérard, Responsable du département de cancérologie médicale et directeur du pôle des sciences cliniques du Centre Léon Bérard, Lyon, FRANCE

Rapporteur & Examineur

**Dr Marc LADANYI**

MD, Chief, Molecular Diagnostics Service, Memorial Sloan Kettering Cancer Center, New-York, UNITED STATES

Rapporteur & Examineur

**Pr Janet SHIPLEY**

MD, PhD, Head of the Division of Molecular Pathology, team leader, Institute of Cancer Research, London, UNITED-KINGDOM

Examinatrice

**Dr Franck TIRODE**

PhD, Directeur de recherche, Centre Léon Bérard, Lyon, FRANCE

Examineur



Would he had been less full of borrowed knowledge!  
Then he would have accepted inspired knowledge from his father.  
When, with inspiration at hand, you seek book-learning,  
Your heart, as if inspired, loads you with reproach.  
Traditional knowledge, when inspiration is available,  
Is like making ablutions in sand when water is near.  
Make yourself ignorant, be submissive, and then  
You will obtain release from your ignorance.

*Molānā Jalāl ad-Dīn Muḥammad Rūmī, Masnavi-ye-Ma'navi, Book IV, Story II*



# Remerciements

Je remercie sincèrement les **membres du Jury, Fabrice André, Jean-Yves Blay, Marc Ladanyi, Janet Shipley, et Franck Tirode**, d'avoir accepté d'évaluer mes travaux et de participer à ma soutenance de thèse. C'est un immense honneur pour moi que de pouvoir défendre mon travail devant vous.

Merci aux membres de mon comité de suivi pour leur mentoring: **Geneviève Almouzni, Olivier Delattre, Antoine Italiano**.

Merci à **Sophie Postel-Vinay**, ma directrice de thèse, qui me suit depuis mon premier semestre d'internat de médecine

**Thomas Eychenne** : tu as été un véritable guide, merci pour ton optimisme et ton soutien sans faille.

**Julien Vibert** : un grand merci, j'ai énormément apprécié travailler avec toi sur ce projet.

**Marlène Garrido, Nicolas Dorvault, Léo Colmet-Daage** : j'ai beaucoup appris à vos côtés, merci pour vos précieux conseils en *wet lab* et bio-informatique.

Aux autres membres actuels et passés de l'équipe : **Asuka Kawai-Kawachi, Madison Lenormand, Clémence Astier, Roman Chabanon, Nicolas Dorvault**, je suis ravie d'avoir fait partie de l'équipe.

A **Virginie Marty, Nicolas Signol** et toute la plateforme PETRA : merci pour tout votre travail rigoureux et enthousiaste.

Aux membres de la plateforme de génomique, **Nathalie Drouin, Maela Francillette, Audrey Naimo**, merci d'avoir pu rendre ce projet possible.

Aux collègues de Curie, **Sarah Watson, Nadège Gruel, Fatima Mechta-Grigoriou, Joshua Waterwall**: merci pour votre aide et nos échanges qui ont été de véritables moteurs dans ce projet.

A tous ceux qui ont participé de près ou de loin à ce projet et ont contribué à faire germer de nouvelles idées: **Antonin Marchais, Jenny Da Costa, Marco Bruschi, Francesco Baschieri, Tudor Manoliu, Charles-Antoine Dutertre, Grégoire Martin, Baptiste Roelens**.

Aux collègues et amis de la pièce 114 et de l'U981 : **Luce Dreno, Loïc Poireaudau, Tala Tayoun, Francesco Facchinetti, Ibrahim Bouakka, Marine Sirerol, Marianne Oulhen, Agathe Aberlenc, Giorgia Guaitoli, Paco Grau-Bejar**, vous avez été de superbes compagnons.

**A tous les collègues du Comité 100 : Axel Le Cesne, Charles Honoré, Matthieu Faron, Benjamin Verret, Rastislav Bahleda**, merci pour votre confiance, je suis très heureuse de prendre cette route avec vous.

Merci à toutes les autres personnes avec lesquelles j'ai pu collaborer pendant ma thèse, et à toutes celles qui ont éclairé mon chemin pendant ces années.

Merci à la **Fondation Philantropia**, et aux **donateurs** du Comité 100: ce travail n'aurait pas vu le jour sans votre aide.

Enfin, merci à **mes amis**, à **ma famille**, pour leur constant soutien dans cette aventure.

*Je dédie humblement ce travail **aux patients et leur famille**.*



## TABLE OF CONTENTS

<b>Synthesis (french) / Synthèse en français</b> .....	<b>3</b>
<b>Abstract</b> .....	<b>8</b>
<b>Publications and Presentations</b> .....	<b>9</b>
<b>List of Figures</b> .....	<b>10</b>
<b>List of Tables</b> .....	<b>11</b>
<b>List of Abbreviations</b> .....	<b>12</b>
<b>Manuscript presentation</b> .....	<b>15</b>
<b>Contributions</b> .....	<b>16</b>
<b>Chapter I. General introduction</b> .....	<b>17</b>
<b>1 Desmoplastic Small Round Cell Tumors (DSRCT)</b> .....	<b>17</b>
<b>1.1 Desmoplastic Small Round Cell Tumors (DSRCT): clinical presentation</b> .....	<b>17</b>
1.1.1 Age and sex ratio .....	17
1.1.2 Incidence .....	18
1.1.3 Clinical presentation and staging .....	18
<b>1.2 DSRCT histological diagnosis and cellular composition</b> .....	<b>20</b>
1.2.1 DSRCT histological pattern and differential diagnosis .....	20
1.2.2 DSRCT differential diagnoses .....	21
1.2.3 DSRCT molecular diagnosis .....	21
<b>1.3 DSRCT current therapeutic management</b> .....	<b>22</b>
<b>1.4 DSRCT oncogenic driver</b> .....	<b>23</b>
1.4.1 EWSR1-WT1: a unique oncogenic driver in DSRCT .....	23
1.4.2 EWSR1-WT1 target genes and DNA-binding sequence.....	26
<b>1.5 DSRCT molecular landscape</b> .....	<b>29</b>
1.5.1 DSRCT relative chromosomal stability .....	30
1.5.2 DSRCT genomic stability .....	30
1.5.3 DSRCT altered transcriptomics pathways .....	33
1.5.4 DSRCT-specific methylation profile .....	37
1.5.5 DSRCT microenvironment .....	37
<b>1.6 DSRCT: molecular vulnerabilities and promising therapeutics</b> .....	<b>38</b>
<b>2 Tumor heterogeneity and cancer cell plasticity</b> .....	<b>42</b>
<b>2.1 Tumor heterogeneity and cancer cell plasticity: general principles</b> .....	<b>42</b>
2.1.1 Definitions .....	42
2.1.2 Tumor cells heterogeneity: evolution models .....	42
2.1.3 Cancer cell plasticity types .....	44
2.1.4 Cancer cell plasticity sources.....	44
<b>2.2 Recent advances in the understanding of cancer cell heterogeneity</b> .....	<b>45</b>
2.2.1 Technical approaches to interrogate tumor cell heterogeneity .....	45
2.2.2 Single-cell omics in cancer research.....	55
<b>2.3 Sarcoma cells' heterogeneity and plasticity</b> .....	<b>56</b>
2.3.1 Sarcoma cells' clonal heterogeneity.....	56
2.3.2 Sarcoma heterogeneity: advances from scRNA-seq approaches.....	57

<b>3</b>	<b><i>Rationale and aims of the presented work</i></b> .....	<b>65</b>
	<b><i>Chapter II. Article</i></b> .....	<b>66</b>
	<b><i>Chapter III. Discussion and perspectives</i></b> .....	<b>194</b>
<b>1</b>	<b><i>Results discussion and open questions</i></b> .....	<b>194</b>
1.1	Dynamic cellular transition states versus clonal heterogeneity .....	194
1.2	Role of EWSR1-WT1 expression level on DSRCT cell heterogeneity .....	195
<b>2</b>	<b><i>Perspectives and ongoing work</i></b> .....	<b>198</b>
	<b>REFERENCES</b> .....	<b>199</b>
	<b>APPENDICES</b> .....	<b>210</b>

## SYNTHESIS (FRENCH) / SYNTHÈSE EN FRANÇAIS

### Introduction

Les Tumeurs Desmoplastiques à Petites Cellules Rondes (DSRCTs) sont un sous-type rare de sarcome de mauvais pronostic, avec une incidence d'environ 0,5 cas par million d'habitants, affectant principalement les jeunes hommes avec un âge médian de 27 ans et un sexe-ratio de trois hommes pour une femme. Les DSRCTs se présentent généralement sous la forme d'une volumineuse tumeur abdominale péritonéale, souvent multinodulaire, et dont les sites métastatiques les plus fréquents sont le foie, les poumons, et les ganglions.

D'un point de vue moléculaire, les DSRCTs sont caractérisés par une faible charge mutationnelle (<1 mutation par Mb). Ils font partie de la famille des sarcomes d'Ewing, dont le pilote oncogénique est un facteur de transcription aberrant résultant d'une translocation impliquant le gène EWSR1 et un partenaire situé en 3' conservant sa capacité de liaison à l'ADN. Le pilote oncogénique pathognomonique des DSRCTs est la protéine chimérique EWSR1-WT1, issue de la fusion entre le gène EWSR1 et le gène WT1 secondairement à la translocation t(11;22) (p13;q12), que l'on retrouve dans 95 % des cas.

Le ciblage pharmacologique de tels facteurs de transcription oncogéniques reste actuellement un défi puisqu'ils ne peuvent être ciblés par les thérapies conventionnelles. Ainsi, peu de progrès thérapeutiques ont été réalisés cette dernière décennie, et le traitement actuel des DSRCTs repose généralement sur une approche multimodale combinant la chirurgie, la chimiothérapie et la radiothérapie, associée à des réponses tumorales insatisfaisantes, à une chimiorésistance secondaire fréquente, et un taux élevé de toxicité.

Malgré la présence d'un pilote oncogénique unique, les DSRCTs sont caractérisées par une différenciation polyphénotypique avec une co-expression de marqueurs myogéniques, épithéliaux, et neuroendocrines, suggérant un certain degré d'hétérogénéité cellulaire ou moléculaire, ou encore une possible pluripotente de la cellule d'origine, qui reste encore indéterminée à ce jour.

L'hétérogénéité et la plasticité tumorale constituent un domaine d'intérêt croissant, comme l'illustre l'ajout récent de la plasticité phénotypique en tant que nouvelle dimension des «*Hallmarks of Cancer*». En effet, la compréhension des mécanismes médiant la plasticité cellulaire tumorale pourrait faire émerger de nouvelles thérapies prometteuses, en ciblant spécifiquement les sous-populations cellulaires les plus agressives, ou encore en ciblant une interaction essentielle entre deux populations cellulaires tumorales interdépendantes.

Dans ce cadre, l'avènement de technologies multiomiques sur cellule unique offre la possibilité sans précédent de capturer et de caractériser l'hétérogénéité des sous-populations de cellules tumorales et de leur microenvironnement.

### **Objectif principal**

L'objectif principal de ce travail est la caractérisation exhaustive de l'hétérogénéité cellulaire et moléculaire des DSRCTs et l'exploration des mécanismes biologiques la sous-tendant, en s'appuyant principalement sur des techniques de séquençage transcriptomique et d'évaluation de l'accessibilité chromatinienne sur cellule unique.

Pour atteindre cet objectif, nous avons formulé les hypothèses suivantes, qui ont dicté les méthodes choisies et décrites dans ce manuscrit :

1. Hypothèse 1 : Les cellules de DSRCT présentent une hétérogénéité phénotypique et des « états cellulaires » plastiques, malgré la présence du seul pilote oncogénique EWSR1-WT1 ;
2. Hypothèse 2 : l'hétérogénéité des cellules DSRCT peut être caractérisée grâce à des méthodes de séquençage sur cellule unique;
3. Hypothèse 3 : l'hétérogénéité des cellules DSRCT est en partie due à des mécanismes intrinsèques, qui peuvent découler de la variabilité du niveau d'expression de EWSR1-WT1, de l'activité transcriptionnelle fluctuante de EWSR1-WT1, de la variabilité des voies de signalisation activées en aval, ou de la variabilité de la reprogrammation épigénétique induite par EWSR1-WT1 ;
4. Hypothèse 4 : la plasticité des cellules de DSRCT est médiée par des facteurs extrinsèques tels que leur microenvironnement, incluant les interactions entre les cellules tumorales et les cellules stromales ou immunitaires.

### **Méthodes**

Dans le but d'explorer l'hétérogénéité intra- et inter-tumorale des DSRCTs, nous avons ainsi caractérisé une série d'échantillons tumoraux issus de patients à l'aide (i) de techniques multi-omiques sur cellule unique incluant le séquençage du transcriptome sur cellule unique (*single-cell RNA-sequencing* ou *scRNA-seq*) et l'analyse de l'accessibilité chromatinienne sur cellule unique (*single-cell Assay for Transposase-Accessible Chromatin with sequencing* ou *ATAC-seq*), (ii) de transcriptomique spatiale, et (iii) d'analyse transcriptomique en « *bulk* » (*RNA-sequencing* ou *RNA-seq*). Nous avons ensuite intégré ces données issues d'échantillons tumoraux de patients à des données de séquençage de lignée cellulaire de DSRCT après immunoprécipitation chromatinienne (*Chromatin ImmunoPrecipitation-sequencing* ou *ChIP-*

*seq*) ciblant EWSR1-WT1, et d'*ATAC-seq* et *RNA-seq* après modulation de l'expression de EWSR1-WT1 par siRNA (small interfering RNA).

## Résultats

Le séquençage de l'ARN sur cellule unique (*scRNA-seq*) de 11 échantillons de patients suivis pour une DSRCT a permis d'identifier des sous-populations récurrentes de cellules tumorales, comprenant des groupes reliés à des lignées développementales distinctes (e.g., épithélial, neural, neuronal, mésenchymateux) et au métabolisme cellulaire (e.g., phosphorylation oxydative, glycolyse anaérobie). En outre, d'autres sous-populations de cellules tumorales ont été identifiées et associées à ce que nous avons appelé des « pseudo-états » cellulaires, reliés au cycle cellulaire, ou caractérisés par une surexpression de gènes codant pour des protéines ribosomales, ou de longs ARNs non codants (*long non-coding RNAs* ou *lncRNAs*).

De façon intéressante, ces sous-populations tumorales étaient caractérisées par des modules de gènes se chevauchant partiellement, en faveur d'un certain degré de plasticité transcriptionnelle. Par ailleurs, nous avons observé peu de diversité entre les différentes sous-populations tumorales en termes de variations du nombre de copies, suggérant que l'hétérogénéité transcriptionnelle observée n'est pas reliée à une évolution clonale ou sous-clonale.

Dans le but d'étudier les mécanismes sous-tendant cette hétérogénéité intra-tumorale, nous avons ensuite cherché à savoir si des mécanismes intrinsèques aux cellules tumorales pouvaient être à l'origine de cette plasticité. Nous avons développé une stratégie exploitant les cDNAs barcodés issus de la technique 10X Genomics, et visant à évaluer spécifiquement le niveau d'expression des transcrits EWSR1-WT1 sur cellule unique. Cette méthode n'a pas permis d'identifier de variation significative de l'expression du transcrit EWSR1-WT1 entre les différentes sous-populations cellulaires tumorales. En s'intéressant ensuite au niveau d'activité d'EWSR1-WT1 plutôt qu'à son taux d'expression, nous nous sommes appuyés sur des données de *ChIP-seq* ciblant EWSR1-WT1 et de *RNA-seq* sur une lignée cellulaire de DSRCT (JN-DSRCT-1) au sein de laquelle l'expression de EWSR1-WT1 a été inhibée par siRNA. Cette modélisation a permis de définir un régulon spécifique de EWSR1-WT1, utilisé comme substitut de son activité transcriptionnelle. L'évaluation de l'activité du régulon EWSR1-WT1 au sein de nos données de *scRNA-seq* a mis en évidence qu'une forte activité de EWSR1-WT1 était associée à des états cellulaires reliés au lignage développemental, tandis que les états cellulaires reliés au métabolisme glycolytique et profibrotiques étaient corrélés à une plus faible activité de EWSR1-WT1. La technique d'*ATAC-seq* associée au *scRNA-seq* (aussi appelée « *single-cell Multiome* ») effectuée sur un échantillon tumoral a identifié que les motifs les plus

différentiellement enrichis entre les sous-populations tumorales étaient essentiellement associés à des facteurs de transcription impliqués dans des processus développementaux comme les familles SOX, FOX et GATA. Par ailleurs, l'activité du régulon EGR1 était inversement corrélée à l'activité du régulon EWSR1-WT1, suggérant un potentiel antagonisme compétitif entre ces facteurs de transcription, dont les séquences consensus sont très proches. Cette observation était par ailleurs concordante avec les données d'*ATAC-seq* sur lignée cellulaire, qui retrouvaient une augmentation de l'accessibilité d'EGR1 lors de l'extinction d'EWSR1-WT1 par siRNA. Nous avons ensuite étudié les facteurs extrinsèques à la cellule tumorale qui pourraient être susceptibles d'influencer l'hétérogénéité intra-tumorale, en s'appuyant sur une technique de transcriptomique spatiale effectuée sur six échantillons de DSRCT issus de patients. Les sous-populations tumorales glycolytiques et profibrotiques étaient spécifiquement localisées au sein de niches hypoxiques à la périphérie de la tumeur, à proximité directe des fibroblastes associés au cancer (*Cancer-Associated Fibroblasts, CAFs*). Le rôle protumoral de ces *CAFs* a été confirmé *in vitro* par le biais de co-cultures de cellules tumorales de DSRCT et de *CAFs*. Enfin, la dernière partie de ce travail a visé à évaluer l'hétérogénéité inter-tumorale des DSRCTs, en réalisant un regroupement ou *clustering* hiérarchique d'une cohorte de 29 échantillons de DSRCT profilés par *RNA-seq*. De façon notable, le groupe de patients présentant le meilleur pronostic était caractérisé par l'expression de gènes reliés à la différenciation épithéliale, l'adhésion cellulaire et la matrice extracellulaire. Nous avons ensuite exploité les signatures de gènes issues des gènes les plus différenciellement exprimés au sein des sous-populations tumorales identifiées en *scRNA-seq*, puis comparer la survie des patients présentant un score élevé ou bas de ces signatures en « *bulk* » *RNA-seq*. Par ce biais, nous avons notamment mis en évidence que les patients caractérisés par une faible expression de la signature spécifique de la sous-population « épithéliale » de DSRCT étaient associés à un pronostic grevé en termes de survie globale, confirmant nos précédentes observations.

## **Conclusion**

Au total, ce travail a permis d'identifier des mécanismes sous-tendant l'hétérogénéité intra- et inter-tumorale des DSRCTs. En particulier, nous avons mis en évidence que la plasticité des cellules de DSRCT résulte de la combinaison de mécanismes cellulaires intrinsèques et extrinsèques, notamment reliés au niveau d'activité transcriptionnelle de EWSR1-WT1 et aux interactions avec le microenvironnement dont les *CAFs*. De plus, l'identification de signatures de gènes associés au pronostic au sein d'une cohorte de 29 patients souligne la pertinence clinique de nos résultats. L'identification de sous-populations cellulaires associées au pronostic

pourraient par exemple justifier le développement de stratégies thérapeutiques visant à contraindre cette plasticité vers un phénotype cellulaire donné et ainsi ouvrir de nouvelles perspectives thérapeutiques pour ces patients.

## ABSTRACT

Desmoplastic Small Round Cell Tumors (DSRCTs) are a rare small round cell sarcoma subtype of poor prognosis driven by the aberrant chimeric transcription EWSR1-WT1. Despite this unique oncogenic driver and the lack of recurrent secondary mutations, DSRCTs harbor a characteristic polyphenotypic differentiation, suggesting the presence of tumor heterogeneity. Using single-cell multiomics, we found that DSRCT tumor cells were characterized by recurrent overlapping phenotypical states, which harbored distinct transcriptional programs, notably related to mesenchymal, neural, and epithelial cell fates, and metabolic switch. Our data further suggest the role of EWSR1-WT1 DNA-binding versatility in DSRCT cells' chromatin plasticity and highlight additional *in vivo* extrinsic signals from microenvironment components like extracellular matrix and cancer-associated fibroblasts which may influence DSRCT cells states and tumor growth. We further explored the clinical applicability of our findings and identified single-cell RNA-sequencing-derived transcriptional signatures that may serve for prognostic risk stratification.

## PUBLICATIONS AND PRESENTATIONS

Vibert J, **Hénon C**, Gruel N, Colmet-Daage L, Kawachi A, Eychenne T, Waterfall J, Adam J, Le Cesne A, Mir O, Honoré C, Delattre O, Watson S & Postel-Vinay S. « Single-cell RNA-sequencing identifies distinct tumor cell subpopulations and immune infiltrate in desmoplastic small round cell tumors (DSRCT) ». *CTOS 2019*.

## LIST OF FIGURES

<b>Fig. 1.</b> DSRCT clinical presentation. ....	18
<b>Fig. 2.</b> DSRCT primary tumor sites and distant metastases. ....	19
<b>Fig. 3.</b> DSRCT histopathological aspect. ....	20
<b>Fig. 4.</b> EWSR1-WT1 fusion transcript, isoforms, and protein domains. ....	23
<b>Fig. 5.</b> Sarcoma histotypes phylogenetic tree (n=60). ....	29
<b>Fig. 6.</b> DSRCT recurrent mutations according to the Whole Exome Sequencing (WES) GENIE DSRCT cohort. ....	33
<b>Fig. 7.</b> Heatmap representing mRNA-based oncogenic signatures scores in five selected sarcoma histotypes. ....	35
<b>Fig. 8.</b> EWSR1-WT1 upregulated and downregulated genes. ....	37
<b>Fig. 9.</b> DSRCT vulnerability pathways and therapeutic opportunities. ....	39
<b>Fig. 10.</b> Tumor heterogeneity models: the clonal evolution, the classical cancer stem cell (CSC), and the plastic CSC theories. ....	43
<b>Fig. 11.</b> Principle of microdroplet generation using oil and microfluidics device. ....	47
<b>Fig. 12.</b> Principle of in-droplet cDNA synthesis and barcode incorporation during the reverse transcription step of the 10X Genomics protocol. ....	48
<b>Fig. 13.</b> ScRNA-seq downstream analyses overview. ....	51
<b>Fig. 14.</b> Clinical applications of single-cell genomics in cancer research. ....	56

**LIST OF TABLES**

**Table 1.** EWSR1-WT1 targets and binding sites. ....28

**Table 2.** DSRCT Targeted Next Generation Sequencing and Whole Exome Sequencing studies. ....32

**Table 3.** Differentially enriched gene sets in DSRCT compared to other sarcoma subtypes. ....36

**Table 4.** Targeted therapies evaluated in DSRCT preclinical studies. ....40

**Table 5.** Specificities of the most used scRNA-seq methods. ....50

**Table 6.** Studies exploiting scRNA-seq technology to explore sarcoma biology. ....63

## LIST OF ABBREVIATIONS

Abbreviation	Correspondence
AC-like	Astrocytes-like
AJCC	American Joint Comity on Cancer
AML	Acute Myeloid Leukemia
AR	Androgen Receptor
ARMS	Alveolar Rhabdomyosarcoma
ASPS	Alveolar Soft Part Sarcoma
ATAC-seq	Assay for Transposase Accessible Chromatin-sequencing
BAF	BRG1/BRM associated factor
CAF	Cancer-Associated Fibroblasts
ccRCC	Clear Cell Renal Cell Carcinoma
CEL-seq	Cell Expression by Linear amplification and Sequencing
ChIP	Chromatin Immunoprecipitation
CI	Confidence Interval
CITE-seq	Cellular Indexing of Transcriptomes and Epitopes by Sequencing
CK	Cytokeratin
CNS	Central Nervous System
CNV	Copy Number Variation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer Stem Cell
CT-scan	Computed Tomography Scan
CTCF	CCCTC-binding factor
CTD	C-terminal Domain
CTL	Cytotoxic T Lymphocytes
CUT&Tag	Cleavage Under Target- Tagmentation
DDR	DNA Damage Response
DOP-PCR	Degenerate Oligonucleotide Primed-PCR
DSB	Double Strand Break
DSRCT	Desmoplastic Small Round Cell Tumor
eGFP	enhanced-Green Fluorescent Protein
EGR1	Early Growth Response Gene 1
EMA	Epithelial Membrane Antigen
EMSA	Electrophoresis Mobility Shift Assay
EMT	Epithelial to Mesenchymal Transition
ES	Epithelioid Sarcoma
ESF	Ewing Sarcoma Family
ESMO	European Society for Medical Oncology
EWSR1	Ewing Sarcoma breakpoint Region 1
FACS	Fluorescence Activated Cell Sorting

FDA	Food and Drug Administration
FISH	Fluorescence In situ Hybridization
G&T-seq	Genome and Transcriptome Sequencing
GEM	Gel-Emulsion Beads
GO	Gene Ontology
GoT	Genotyping of Transcriptome
GSVA	Gene Set Variability Analysis
HLA	Human Leucocyte Antigen
HR	Hazard Ratio
IF	Immunofluorescence
IHC	Immunohistochemistry
KTS	Lysine, Threonine Serine domain
LC-MS	Liquid Chromatography-Mass Spectrometry
LIANTI	Linear Amplification via Transposon Insertion
LOH	Loss of heterozygosity
MALBAC	Multiple Annealing and Looping-Based Amplification Cycles
MARS-seq	Massively parallel RNA Single-cell-sequencing
MATQ-seq	Multiple Annealing and dC-Tailing-based Quantitative single-cell RNA-sequencing
MDA	Multiple Displacement Amplification
MEF	Mouse embryonic Fibroblast
MES-like	Mesenchymal-like
MET	Mesenchymal to Epithelial Transition
MHC-I	Class I Major Histocompatibility Complex
MLPA	Multiplex Ligation-dependent Probe Amplification
MMLV	Moloney Murine Leukemia Virus
MPNST	Malignant Peripheral Nerve Sheath Tumors
MRI	Magnetic Resonance Imaging
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NPC-like	Neural Progenitor-like
NSE	Neural Specific Enolase
NSG	NOD/SCID/Gamma
OPC-like	Oligodendrocyte progenitor-like
OS	Overall Survival
PacBio	Pacific Bioscience
PCA	Principal Component Analysis
PCI	Peritoneal Cancer Index
PCR	Polymerase Chain Reaction
PDAC	Pancreatic Ductal Adenocarcinoma
PDX	Patient Derived Xenograft
PEA/STA	Proximity Extension Assays/Specific (RNA) Target Amplification
PET-CT scan	Positron Emission Tomography- Computed Tomography scan

PLAYR	Proximity Ligation Assay for RNA
Pol II	RNA Polymerase II
RBP	RNA Binding Protein
RIME	Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins
RNA-seq	RNA-sequencing
ROI	Region Of Interest
RRM	RNA Recognition Motif
RT	Reverse Transcription
RT-PCR	Reverse Transcription Polymerase Chain Reaction
scDNA-seq	Single-cell DNA sequencing
sciCAR	Single-cell Chromatin Accessibility and RNA seq
scRNA-seq	Single-cell RNA sequencing
shRNA	Short Hairpin RNA
SIDR	Simultaneous Isolation and parallel sequencing of genomic DNA and total RNA from single cells
siRNA	Small Interfering RNA
SMART-seq	Switch Mechanism at the 5' End of RNA Templates-sequencing
SMOOTH-seq	single-molecule real-time sequencing of long fragments amplified through transposon insertion
SNARE-seq	single-nucleus chromatin accessibility and mRNA expression sequencing
SNP	Single Nucleotide Polymorphism
SRCS	Small Round Cell Sarcoma
SS	Synovial Sarcoma
STS	Soft Tissue Sarcoma
SWI/SNF	SWItch/Sucrose Non-Fermentable
t-SNE	t-distributed stochastic neighbor embedding
TAD	Topologically Associating Domain
TCR	T cell Receptor
TET	Ten Eleven Translocation
TGS	Third Generation Sequencing
TIVA-seq	Transcriptome In Vivo Analysis-sequencing
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier
UPS	Undifferentiated Pleomorphic Sarcoma
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT1	Wilms' Tumor protein 1
ZNF	Zinc Finger domain

## MANUSCRIPT PRESENTATION

---

This thesis is divided into three main parts. The first part introduces the background of DSRCT disease and biology, discusses the role of cellular plasticity in oncology and specifically in sarcoma, and presents the main scientific advances achieved in the understanding of tumor heterogeneity allowed by the massive development of scRNA-seq technology (**Chapter I. General introduction**).

The second part is the main article that emanated from the research work I carried out during my PhD, which is being submitted for peer-review publication (**Chapter II. Article**).

Lastly, the third part provides an overview of the perspectives opened by this research, striving to (1) debate around critical challenges emerging from our results and (2) introduce ongoing and developing work on DSRCT molecular deciphering that we aim to pursue shortly, focusing on (i) the set-up of a drug screening aiming at identifying pharmaceutical compounds amenable to target/destabilize EWSR1-WT1 fusion protein, (ii) the study of the epigenetic and chromatin remodeling role of EWSR1-WT1, and (iii) the generation of novel *in vitro* models that would better reproduce DSRCT tumor cells heterogeneity (**Chapter III. Discussion and perspectives**).

## CONTRIBUTIONS

---

**I warmly thank all the contributors to the article manuscript presented within this document, which is the fruit of fantastic teamwork.**

The conception of the presented idea was from Sophie Postel-Vinay (SPV), my thesis supervisor, and I.

Nadège Gruel (NG) and I performed scRNA-seq wet lab experiments on fresh DSRCT samples. Maria Eugenia Da Costa (MEDC) performed Visium spatial transcriptomics assays from FFPE DSRCT samples.

Thomas Eychenne (TE) and I conceived, optimized, and carried out all cell-line-based experiments, comprising RNA-seq, ChIP-seq, ATAC-seq, the results of which are presented throughout this manuscript, as well as CUT&RUN and HiC, which are introduced in the Perspectives section. TE and I conceived, optimized, and carried out the in-house generated *EWSRI-WTI* targeted scRNA-seq assay.

The optimization of immunofluorescence assays on 2D and 3D grown cells was performed by Marlène Garrido (MG), TE, and I.

Immunohistochemistry assays were performed by Virginie Marty (VM) under the supervision of Carine Ngo (CN) and I.

Leo Colmet-Daage (LCD), Julien Vibert (JV), and I performed bioinformatics analyses. Specifically, RNA-seq, ChIP-seq, and ATAC-seq pipelines were run by LCD while I performed downstream analyses, including multiomics data integration and data interpretation. JV and I completed the analyses for 3'scRNA-seq and 3' snMultiome data, and we both contributed to their interpretation. JV performed additional advanced 3'scRNA-seq analyses comprising regulons and co-regulated gene modules inference. I performed data processing and downstream analyses of spatial transcriptomics assays.

JV and I both designed the 3'scRNA-seq-derived prognostic signatures.

Sarah Watson (SW) and JV contributed to the obtention of an external DSRCT bulk RNA-seq cohort.

All authors contributed to the discussion of the results.

SPV and I supervised the findings of this work and their interpretation. I wrote the article manuscript, reviewed by SPV, TE, and JV.

## CHAPTER I. GENERAL INTRODUCTION

---

Desmoplastic small round cell tumors (DSRCTs) are a rare subtype of small round cell tumors of poor prognosis, with an incidence around 0.5/millions <sup>1</sup>, primarily affecting young people with a median age of 27 years old, with a male predominance (male to female ratio 3:1). DSRCT initial presentation generally consists in an abdominal bulky tumor which may disseminate in the peritoneum in the form of multiple nodules. Molecularly, DSRCTs are part of the Ewing Sarcoma Family (ESF) of tumors, driven by a translocation involving the EWSR1 gene and a 3' partner retaining DNA binding capacity, resulting in a chimeric protein that acts as an aberrant transcription factor. DSRCTs display one unique driver, the chimeric protein EWSR1-WT1 arising from the fusion between EWSR1 and WT1 gene t(11;22) (p13;q12), found in 95% of cases. Although appearing genetically simple, the pharmacological targeting of such an aberrant transcription factor remains challenging, and few therapeutic advances have been made to reverse the poor prognosis of DSRCTs in this last decade compared to other malignancies. Current treatment relies on aggressive approaches combining surgery, chemotherapy, and radiotherapy <sup>2</sup>, resulting in unsatisfying tumor responses, limited prognostic improvement, and high rate toxicity.

Despite this unique oncogenic driver, published data show DSRCT is a polyphenotypic tumor, suggesting cellular/molecular heterogeneity and DSRCT cell pluripotency.

This work will comprehensively address the exploration of DSRCT cellular and molecular heterogeneity to describe the biological mechanisms underlying DSRCT cell oncogenicity.

## 1 DESMOPLASTIC SMALL ROUND CELL TUMORS (DSRCT)

---

### 1.1 DESMOPLASTIC SMALL ROUND CELL TUMORS (DSRCT): CLINICAL PRESENTATION

#### 1.1.1 Age and sex ratio

Desmoplastic small round cell tumor (DSRCT) was first described in 1991 by William L Gerald and Juan Rosai, who posited its clear distinction from other small round cell tumors on clinical presentation and histopathological features criteria <sup>3</sup>. DSRCTs were characterized as a rare subtype of mesenchymal tumors mainly affecting young males with a sex ratio of male to female of 3-5 to 1 and a median age at diagnosis of 22-27 years old <sup>4</sup>.

### 1.1.2 Incidence

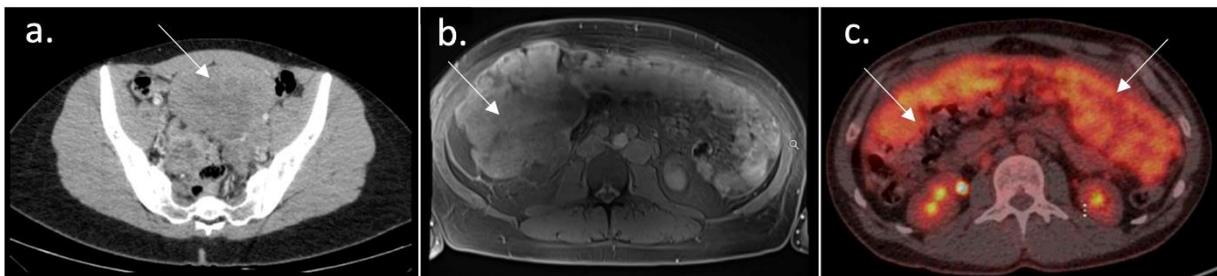
The incidence of DSRCT is estimated at around 0.2 to 0.5 cases per million<sup>5,6</sup>. According to Waqar et al.<sup>7</sup>, DSRCT incidence has risen within the last two decades. Moreover, males seem to have a higher age-adjusted incidence rate (0.22 vs. 0.05,  $p < 0.05$ ).

There is no known genetic risk factor for developing DSRCT. However, several studies described an increased risk among African descendants compared to the Caucasian population<sup>7,8</sup>.

### 1.1.3 Clinical presentation and staging

Patients generally present with an abdominal mass, pain, weight loss, or symptoms secondary to adjacent organ compression, such as obstructive renal failure or bowel obstruction.

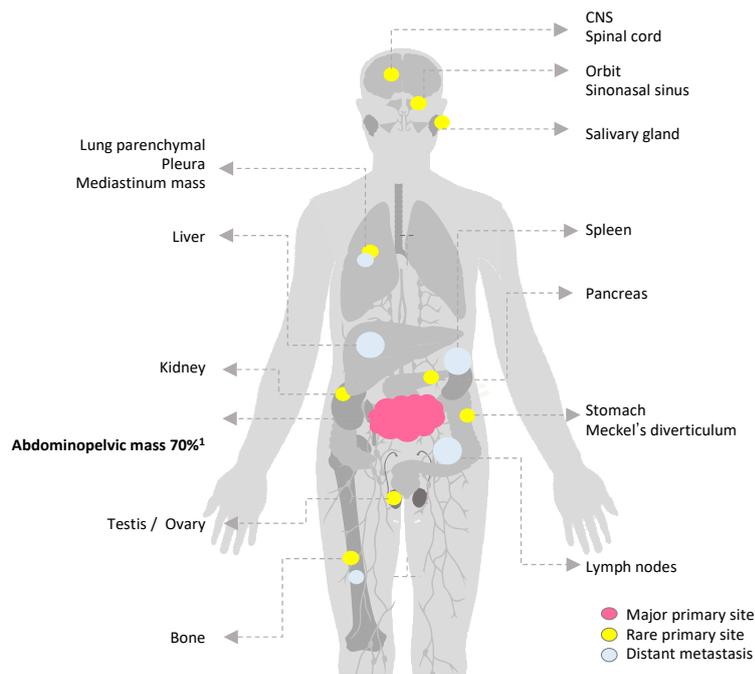
The vast majority (80-90%) of DSRCT tumor primary sites are localized within the peritoneum<sup>9</sup> (**Fig. 1**). Diagnosis is often late and made at abdominal sarcomatosis or metastatic stages in 50% of cases<sup>5,10-12</sup>, the preferential metastatic sites including the liver, the lungs, and mesenteric, retroperitoneal, or mediastinal lymph nodes<sup>1,10,13</sup>.



*Fig. 1. DSRCT clinical presentation.*

*CT (a.), MRI (b.), and 18F-FDG-PET/CT (c.) DSRCT classical abdominopelvic presentation.*

Nonetheless, several cases report extra abdominal localization, raising the primary cell of origin question. Alternative localizations include head and neck (orbital, submental), pleura<sup>14</sup>, and mediastinal, central nervous system<sup>15</sup>, or para testicular<sup>16</sup> primary sites (**Fig. 2**).



*Fig. 2. DSRCT primary tumor sites and distant metastases.*

*Most DSRCTs arise in the abdominal cavity, although rare primary sites comprising the head, neck, and central nervous system have been described. The most frequent DSRCT secondary sites are liver, lung, and lymph nodes.*

Initial clinical examination should involve a physical examination, including inspection and tumor palpation when feasible. Additional complementary exams should assess cardiac, kidney, and liver function, search for complications, and verify systemic treatment and surgery eligibility. The clinical and radiological staging should be performed carefully, as the extent of the disease at diagnosis remains one of the most decisive prognostic factors <sup>17</sup>.

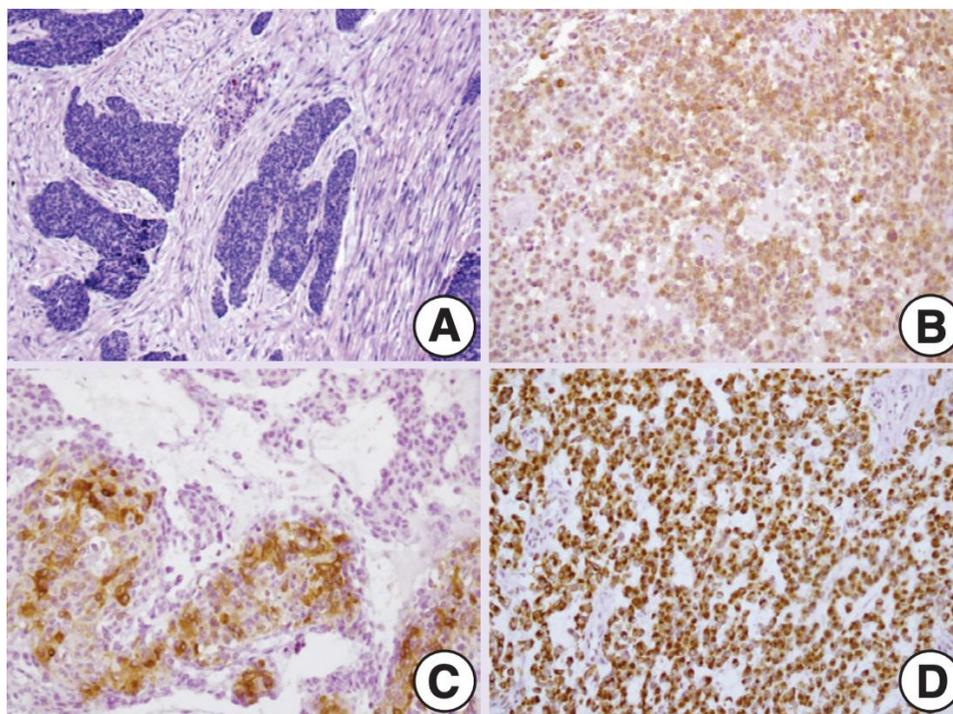
Imaging-based DSRCT staging should include thoraco-abdominal-pelvic-CT, abdominal and pelvic MRI, and 18-fluorodeoxyglucose-PET-CT. Several staging classifications have been developed. The expert consensus AJCC classification (8<sup>th</sup> edition) <sup>18</sup>, set for STS in general, has been validated in DSRCT <sup>19</sup>. It relies on the size of the primary tumor, its depth (superficial or deep), lymph node involvement, and distant metastases. In the latter study, on multivariable analysis, stage IV disease (according to AJCC version  $\leq 7$  for STS) was associated with poorer overall survival (HR=2.12, 95% CI: 1.41-3.18), whereas surgery (HR = 0.68, 95% CI: 0.50-0.91), chemotherapy (HR = 0.52, 95% CI: 0.35-0.78), or radiation therapy (HR = 0.55, 95% CI: 0.33-0.92) were independently associated with improved OS.

Two additional classifications have been designed explicitly for DSRCT disease staging: the first considers liver involvement and ascites, and the second <sup>20,21</sup> is based on the Peritoneal Cancer Index (PCI) and the presence of liver or extra abdominal metastases.

## 1.2 DSRCT HISTOLOGICAL DIAGNOSIS AND CELLULAR COMPOSITION

### 1.2.1 DSRCT histological pattern and differential diagnosis

Histological diagnosis should be made on a sample obtained using a co-axial core needle biopsy using a 12–16-gauge needle targeting the most solid portion of a mass, as recommended by the European Society of Medical Oncology (ESMO)–EURACAN–GENTURIS guidelines<sup>22,23</sup>. Histologically, DSRCT presents with particular features. These tumors are defined by small round blue cells organized in islets and surrounded by a prominent dense desmoplastic stroma composed of reactive fibroblasts, inflammatory myeloid cells, and blood vessels (**Fig. 3**). Several routine IHC stainings are required to orientate a diagnosis towards DSRCT in front of a small round blue cells tumor. These markers include CD45, Cytokeratins (CK, AE1/AE3), Epithelial Membrane Antigen (EMA), PS100, desmin and Ki67. DSRCT is characterized by the coexpression of epithelial (Cytokeratin, EMA), mesenchymal (Desmin, Vimentin), and neural markers such as Neural Specific Enolase (NSE) which are often expressed altogether. DSRCT often displays a weak CD99 staining, which allows its distinction from Ewing sarcoma and is positive for WT1 C-terminal antibody staining due to EWSR1-WT1 fusion protein expression.



*Fig. 3. DSRCT histopathological aspect.*

*H&E staining showing nests of small round blue cells embedded within an abundant desmoplastic stroma (A). Analysis of the tumor by immunohistochemistry (20x) shows diffuse positivity for VEGFR2 (B) and coexpression of (epithelial) EMA (C) and (mesenchymal) desmin (D) markers. From Mora et al. <sup>24</sup>*

### 1.2.2 DSRCT differential diagnoses

Although DSRCT histological presentation is often suggestive, several differential diagnoses should be suspected in front of small round cell tumors, especially when the clinical presentation is not typical. These differential diagnoses include tumors of non-mesenchymal origin such as lymphoma, neuroblastoma, carcinoma (e.g., small-cell lung carcinoma, Merkel cell carcinoma), and mesothelioma. Mesenchymal tumors characterized by small round tumor cells may be grouped into the novel growing entity of Small Round Cell Sarcomas (SRCs) or SRCs mimics <sup>25</sup>. Strictly speaking, DSRCT is not categorized within the SRC group in the 2020 WHO classification. SRCs gather an expanding group of sarcoma, recently added to the 2020 WHO classification of soft tissue and bone tumors <sup>26–28</sup>. They represent a large and heterogeneous entity comprising distinct molecular subtypes now increasingly characterized on a molecular level <sup>29</sup>. SRCs are generally defined by the presence of a chromosomal translocation generating chimeric proteins, most of which act as aberrant transcription factors, such as (1) FET family protein-ETS family protein defining EFTs (e.g., EWSR1-FLI1, EWSR1-ERG, EWSR1-PEA3, FUS-ETV4), (2) EWSR1-non-ETS rearranged sarcoma (e.g., EWSR1-NFATC2), (3) CIC-fused (e.g., CIC-DUX4), or (4) BCOR-fused (e.g., BCOR-CCNB3) sarcoma. Moreover, additional adult sarcoma may also mimic genuine SRCs, including the different subtypes of rhabdomyosarcoma, extraskeletal myxoid chondrosarcoma, myxoid or round cell liposarcoma, and mesenchymal chondrosarcoma.

### 1.2.3 DSRCT molecular diagnosis

DSRCTs are characterized by the pathognomonic fusion gene EWSR1-WT1 arising from the balanced translocation t(11;22) (p13;q12) <sup>30</sup>, considered the sole oncogenic driver in this disease. This genetic abnormality is found in approximately 95% of DSRCT cases. Exceptional cases describe the finding of EWSR1–WT1 in other tumor types whose histological aspect was incompatible with the diagnosis of DSRCT <sup>31</sup>.

Several EWSR1-WT1 variant transcripts have been described so far <sup>32</sup>, the most common isoform arising from a fusion breakpoint between EWSR1 exon 7 and WT1 exon 8.

Fluorescence *In Situ* Hybridization (FISH) is routinely used as a molecular assay to diagnose DSRCT <sup>33</sup>, the most common assay depending on an EWSR1 break-apart probe. In the case of

a translocation involving EWSR1, the two fluorophores positioned in the N-terminal and C-terminal part of the EWSR1 protein will be spatially separated and distinguishable as two differently colored dots.

As an alternative to FISH, multiplex Next Generation Sequencing (NGS)-based RT-PCR is frequently used to identify both partners of a fusion transcript among a list of relevant targets. A classic example of such an assay is the ArcherDX FusionPlex<sup>®</sup> Sarcoma v2 panel, which covers critical fusions in 63 genes often rearranged in sarcoma. Alternatively, NGS-based ligation-dependent multiplex probe amplification (MLPA)<sup>34</sup> requires to know a priori the identity of both fusion partners. In addition, DNA-based hybrid capture may still be used, but is less accurate and cost-effective.

Altogether, DSRCT diagnosis often relies on a set of arguments encompassing typical clinical presentation, coupled with the archetypal morphological aspect and immunohistological staining, finally confirmed molecularly with FISH or RNA-seq identifying the pathognomonic fusion gene or transcript.

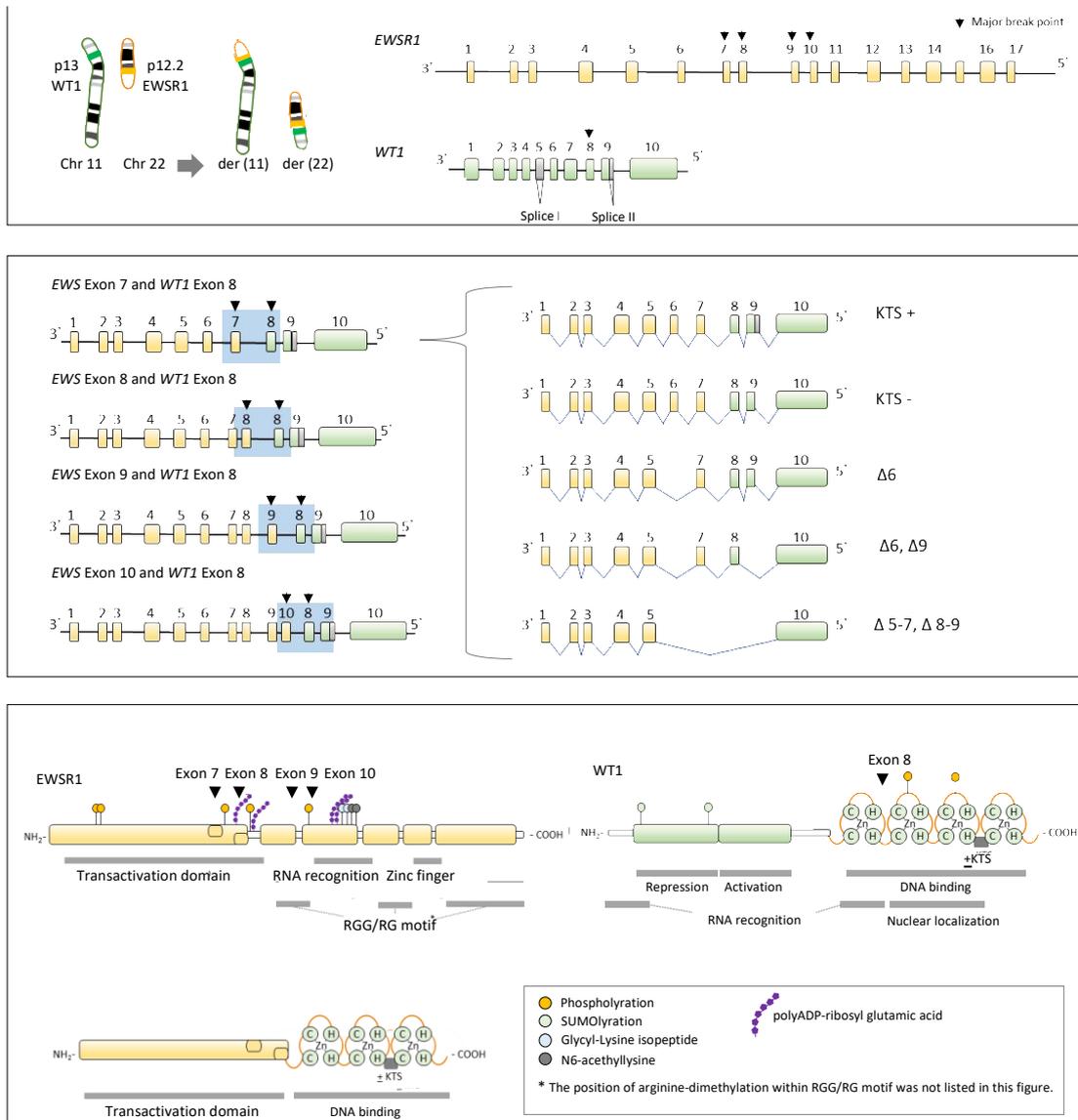
### **1.3 DSRCT CURRENT THERAPEUTIC MANAGEMENT**

DSRCT general therapeutic approach consists of a multimodal strategy<sup>35</sup> combining induction chemotherapy, debulking surgery, followed by maintenance treatment or pan abdominal radiation, the latter being in practice usually reserved for children.

There is no standard regimen, but the association of doxorubicin and high-dose ifosfamide is widely used. Trabectedin<sup>36-39</sup> and pazopanib<sup>40,41</sup>, the only tyrosine kinase inhibitor approved in the management of advanced STS<sup>42</sup>, have also proved their efficacy in advanced DSRCT. No specific targeted therapy towards DSRCT has entered clinical practice to date. Preclinically evaluated compounds and potent future therapeutics are discussed in **Section 1.6**.

## 1.4 DSRCT ONCOGENIC DRIVER

### 1.4.1 EWSR1-WT1: a unique oncogenic driver in DSRCT



**Fig. 4. EWSR1-WT1 fusion transcript, isoforms, and protein domains.**

Upper panel: The  $t(11;22)(p13;q12)$  translocation, full-length EWSR1, and WT1 genes, and the most frequent breakpoints generating EWSR1-WT1 fusion gene in DSRCT. Middle panel: Variability of EWSR1-WT1 fusion transcripts according to gene breakpoints, alternative exonic splicing, and the presence/absence of the KTS domain between the 3<sup>rd</sup> and the 4<sup>th</sup> zinc fingers from the WT1 C-terminal part. Lower panel: Full-length EWSR1 and WT1 protein domains and EWSR1-WT1 chimeric protein.

#### 1.4.1.1 EWSR1-WT1 chimeric gene and t(11.22) translocation

Multiple lines of evidence advocate for the role of EWSR1-WT1 as the unique DSRCT oncogenic driver: (i) EWSR1-WT1 detection in more than 95% of DSRCT cases, (ii) the lack

of recurrent secondary mutations identified in both primary and metastatic sites, (iii) EWSR1-WT1-driven oncogenesis if associated to TP53 deficiency in human embryonic fibroblast <sup>43</sup>, (iv) DSRCT cells dependency on EWSR1-WT1 to grow and survive *in vitro*<sup>44</sup>.

The initial event leading to EWSR1-WT1 fusion gene is the balanced chromosomal translocation t(11;22) (p13;q12). EWSR1-WT1 chimeric protein contains the N-terminal part of EWSR1, which comprises its transcriptional activation domain, whereas the WT1 C-terminal part involves the three last DNA-binding zinc finger domains.

*In vivo* chromosomal translocation are complex events with poorly understood underlying mechanisms. At least two critical steps are required to give rise to a translocation. First, a double-strand break (DSB) should co-occur on both partner genes specific loci. Second, a spatial proximity (even transient) between both DSBs should allow their joining through Non-Homologous End Joining (NHEJ) DNA repair mechanism <sup>45</sup>. The origin of such DSBs can be variable and may be favored by endogenous mechanisms such as a high transcription or replication rate leading to genotoxic stress promoting such DNA cleavage. Thus, one may suspect that DSRCT cell of origin, which is currently not yet identified, may bear a strong EWSR1 and WT1 pathways activation. Interestingly, WT1 is involved in the Epithelial-to-Mesenchymal Transition (EMT) process during development, including EMT occurring in the coelomic epithelium in embryogenesis <sup>46</sup>. Notably, while there is growing evidence that specific germline mutations on protein-coding sequences may participate in Ewing sarcoma tumorigenesis and increase its incidence risk <sup>47-51</sup>, no genetic predisposition factor for developing DSRCT has been established.

The t(11;22) (p13;q12) translocation EWSR1-WT1 chimeric protein hence contains the N-terminal part of EWSR1, which encloses the transcriptional activation domain, whereas WT1 C-terminal part involves the three last DNA-binding zinc finger domains.

#### 1.4.1.2 The 3' fusion partner: EWSR1

EWSR1 belongs to the multifunctional TET/FET proteins family, together with FUS and TAF15. These proteins are DNA/RNA-binding Proteins (RBP) through an RNA Recognition domain (RRM domain), Arginine Glycine Repeats (RRG domains), and a C-terminal zinc finger domain <sup>52</sup>. This RNA-binding capacity confers to EWSR1 an essential function in RNA metabolism and RNA splicing. The N-terminal end of EWSR1 harbors a repeated motif of the Ser-Tyr-Gly-Gln-Gln-Ser hexapeptide <sup>53</sup>. This domain resembles the C-terminal domain (CTD) of the largest subunits of RNA polymerase II (Pol II). It is similar to the activating domain of some transcription factors, conferring EWSR1 a potent transcriptional activation capacity. Significantly, EWSR1 is expressed in most cells and tissues and is mainly located in the

nucleus, where it acts as a transcriptional coregulator through its interactions with several transcription factors, such as Pol II and TFIID<sup>52</sup>.

#### 1.4.1.3 The 5' fusion partner: WT1

*WT1* was initially identified in 1990 as a predisposition gene to Wilms' Tumors<sup>54</sup>, a pediatric kidney cancer. Since then, *WT1* has been confirmed to be mutated in about 15% of Wilms' tumors cases<sup>55</sup> (loss-of-function mutation) and to behave as a tumor suppressor. In the meantime, several molecular functions of WT1 have been uncovered. Notably, WT1 protein has been shown to be involved in embryonic development and tissue homeostasis, cellular switching processes such as EMT or Mesenchymal-to-Epithelial Transition (MET), as well as transcription and epigenetic regulation<sup>56</sup>. WT1 is admitted to be part of the transcription factors family and binds to DNA throughout its four Zinc Fingers (ZNFs). Initial in vitro studies revealed that WT1 can bind the consensus site 5'-GCGGGGGCG-3', which is identical to the EGR1 binding motif<sup>57</sup>. Moreover, Chromatin Immunoprecipitation (ChIP) experiments unraveled WT1 potential regulated targets, most of which are involved in kidney development<sup>56</sup>.

Importantly, the *WT1* gene codes to a transcript containing 10 exons, potentially generating 36 distinct isoforms. In any of the isoforms found in mammals, four C2 H2 Krüppel-like ZNFs are located on the C-terminus part of the protein (**Fig. 4**). One particular key feature within these domains, is the presence of a three amino acid domain (Lysine, Threonine, and Serine; KTS) located within the 9<sup>th</sup> intron. This KTS motif can be inserted between the third and fourth ZNF through alternative splicing. Contrary to other exonic splicing isoforms of *WT1*, whose differential functions remain unclear, +KTS and -KTS isoforms have documented distinct activities<sup>58,59</sup>. In summary, -KTS isoform seems primarily involved in transcriptional regulation as a transcription factor, whereas +KTS isoform mainly participates in post-transcriptional regulation<sup>58,59</sup>.

#### 1.4.1.4 EWSR1-WT1 isoforms

Several studies aimed at characterizing the various EWSR1-WT1 isoforms<sup>30,60,61</sup>. An original work analyzed the EWSR1-WT1 breakpoint sequence in six DSRCT patients. In this study, EWSR1-WT1 breakpoint locus PCR and DNA sequencing revealed a fusion between *EWSR1* intron 7 and *WT1* intron 7. The related fusion transcript linking *EWSR1* exon 7 with *WT1* exon 8 was detected in four of the six patients. This breakpoint bridging *EWSR1* and *WT1* leads to an in-frame fusion of the N-terminal end of EWSR1 and the three last zinc finger domains of WT1. Therefore, the resulting chimeric protein conjugates the potent transcriptional activation

domain of EWSR1 with the three last zinc-finger domains of the WT1 DNA-binding domain. Since its first description in 1992 <sup>62</sup>, several EWSR1-WT1 variant transcripts have been reported, arising from various DNA breakpoints or alternative transcript splicing. Although most EWSR1-WT1 transcripts arise from the fusion between *EWSR1* exon 7 and *WT1* exon 8, different EWSR1-WT1 breakpoints have been described <sup>32</sup>, while the biological or clinical consequences of such variants on DSRCT disease course remain unknown (**Fig. 4**).

Like for wild-type WT1 gene, one particularly documented isoform type concerns the + or - KTS domain. While only the -KTS isoform (and not +KTS) showed transforming potential in NIH3T3 cells *in vitro* <sup>63</sup>, Bandopadhyay et al. highlighted in a subsequent study that both EWSR1-WT1 -KTS and EWSR1-WT1 +KTS expression co-operates with loss or inactivation of p53 to transform MEFs <sup>43</sup>.

#### 1.4.2 EWSR1-WT1 target genes and DNA-binding sequence

Recent works emphasized the dependency of DSRCT cells on EWSR1-WT1 fusion transcript <sup>44</sup> and described that the loss of EWSR1-WT1 activity leads to DSRCT cells' morphologic changes and growth arrest. EWSR1-WT1 has been shown to drive DSRCT oncogenicity through the transcriptional activation of target genes involved in proliferation, cell survival, and invasion, such as *IL-2/15Rβ* <sup>64</sup>, *BAIAP3* <sup>65</sup>, *IGF-2* <sup>43,66</sup>, *PDGFA* <sup>43</sup>, *EGFR* <sup>43</sup>, *LRRC15* <sup>67</sup>, *FGFR4* <sup>66</sup>, *ROCK1* <sup>66</sup>, *PEX5* <sup>66</sup>, *CTCF* <sup>66</sup>, *TSPAN7* <sup>66</sup>, and *TALLA1* <sup>68</sup> contributing to sarcomagenesis and tumor progression.

The first studies exploring EWSR1-WT1 targets and binding sites relied on either electrophoretic mobility shift assays (EMSA), promoter-reporter assays, or original ChIP experiments. These uncovered EWSR1-WT1 target genes and identified distinct binding sites between EWSR1-WT1 isoforms (**Table 1**), though the EWSR1-WT1 preferred DNA-binding motif has not been established yet. For instance, the two EWSR1-WT1 isoforms +KTS or -KTS have specific DNA recognition sequences, EWSR1-WT1 -KTS, showing specific binding to either GC-rich regions <sup>32,69-71</sup> or TCCn repeats <sup>37</sup>. In contrast, EWSR1-WT1 +KTS was shown to target GA-rich regions containing 5'-GGAGG(A/G)-3' motif <sup>67</sup>. Interestingly, according to Liu et al. <sup>32</sup>, most EWSR1-WT1 protein isoforms retain the ability to bind EGR1 GC and TC elements and WT1 consensus sequences, but the activation of EGR1 promoter activity is specific to EWSR1-WT1 -KTS.

Subsequent studies took advantage of Chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-seq), bringing a wider overview of EWSR1-WT1 targets and downstream-

regulated pathways. Hingorani et al. <sup>66</sup> proposed to study the correlation between DSRCT transcriptome and EWSR1-WT1 binding profile using ChIP-seq in a JN-DSRCT-1 cell line <sup>72</sup>. A total of 2,036 peaks were significantly enriched within EWSR1-WT1-binding regions, including 1,284 peaks associated with a protein-coding gene. Novel EWSR1-WT1 targets were thus identified such as *ROCK1*, *PEX5*, *CTCF*, *FGFR4*, *IGF2* and *TSPAN7*. However, this study did not identify *de novo* EWSR1-WT1 binding motifs, although EWSR1-WT1 targeted sequences were close to WT1, EGR1, and EGR2 binding motifs and colocalized with RNA Pol II occupancy. Notably, EWSR1-WT1-specific peaks were mostly enriched within intergenic and intronic regions, suggesting EWSR1-WT1 binding to enhancers rather than promoters. This binding profile is consistent with the one found in other fusion-driven sarcomas, such as Ewing sarcoma <sup>73</sup> and its characteristic EWSR1-FLI1 fusion. A comprehensive description of EWSR1-WT1 downstream activated pathways relying on ChIP and other functional genomics approaches is further detailed in Section **1.5.3**.

Table 1. EWSR1-WT1 targets and binding sites.

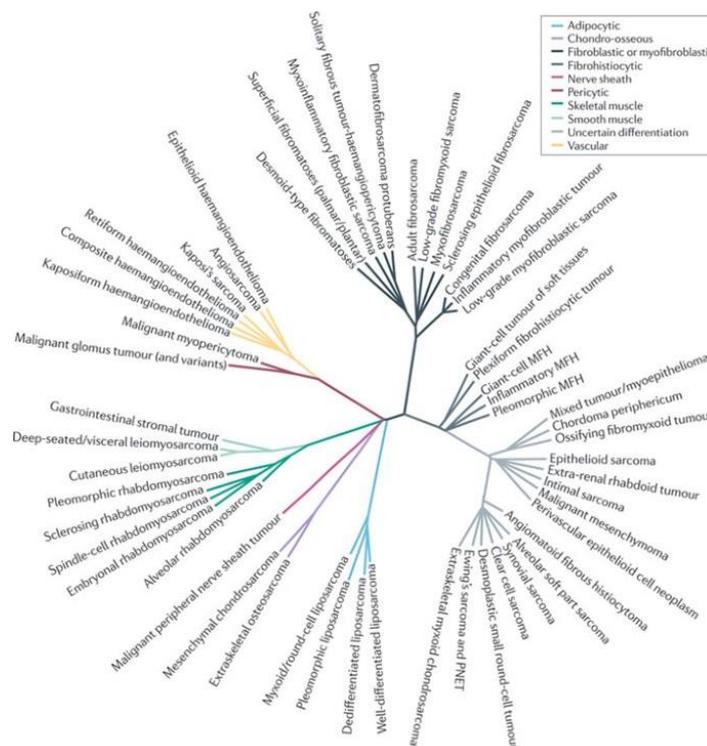
Target	ASCL1	ENT4	LRRC15	BAIAP3	IL-2/15R $\beta$	EGR1	IGF-1	SIK1	EGFR	IGF2	PDGFA	IGF-1R
<b>Reference</b>	69	70	67	65	64	32	71	74	37	37	37,75	76
<b>EWSR1-WT1 isoform</b>	KTS-	KTS +/-	KTS +	KTS-	KTS-	KTS-	KTS+/-	NA	NA	NA	KTS-	NA
<b>Model</b>	Exogenous EWSR1-WT1 expression in MEFs and U2OS cell lines, JN-DSRCT-1 cell line, and FFPE DSRCT tumor sample	Exogenous EWSR1-WT1 expression in U2OS-derived cell lines (UV9, UF5, UED5), JN-DSRCT-1 cell line, and FFPE DSRCT tumor sample	Exogenous EWSR1-WT1 expression in U2OS, FFPE DSRCT tumor sample	Exogenous EWSR1-WT1 expression in U2OS	Exogenous EWSR1-WT1 expression in U2OS and FFPE DSRCT tumor sample	Exogenous EWSR1-WT1 expression in Saos-2, FFPE DSRCT tumor samples	Exogenous EWSR1-WT1 expression in Saos-2 and G401 cell lines	JN-DSRCT-1 cell line	JN-DSRCT-1 cell line	JN-DSRCT-1 cell line	JN-DSRCT-1 cell line, exogenous EWSR1-WT1 expression in U2OS	Exogenous EWSR1-WT1 expression in CHO cell line
<b>Method</b>	Microarray, RT-PCR, promoter-reporter assay, cell proliferation assay, IHC and ChIP	RT-PCR, promoter-reporter assay, IHC, and ChIP	RT-PCR, promoter-reporter assay, IHC, EMSA and ChIP	RT-PCR, promoter-reporter assay, RNA-in situ hybridization, EMSA and ChIP, cell proliferation assay	Promoter-reporter assay, EMSA, IHC	RT-PCR, promoter-reporter assay, EMSA	Promoter-reporter assay, EMSA	ChIP	ChIP, tested upon trabectedin treatment	ChIP, tested upon trabectedin treatment	ChIP tested upon trabectedin treatment, promoter-reporter assay, RNA-in situ hybridization, and IHC	Promoter-reporter assay
<b>Target sequence</b>	10 GC-rich response elements within ASCL1 promoter	<b>GC-rich:</b> (i) 5'-GAGGGGGT C-3', (ii) 5'-GCGGGGGG G-3', (iii) 5'-CTGGGGGC G-3'	<b>GA-rich:</b> 5'-GGAG G(A/G)-3'	<b>E-WRE:</b> 5'-GXXG(T/G)G GGXXG-3'	<b>GC-rich:</b> 5'-(G/C)(C/G)(G/C)TGGGGG-3'	<b>GC-rich:</b> 5'-GCGGGGGC G-3'	<b>GC-rich:</b> 5'-GCGGGGGC G-3'	NA (2kb proximal promoter region of SIK1)	NA	NA	<b>TCC<sub>n</sub> repeats</b>	NA

ChIP: Chromatin ImmunoPrecipitation, EMSA: electrophoretic mobility shift assay, E-WRE: EWSR1-WT1 (KTS-) response element, FFPE: Formalin-fixed, paraffin-embedded, IHC: ImmunoHistoChemistry, RT-PCR: Reverse Transcription-Polymerase Chain Reaction

## 1.5 DSRCT MOLECULAR LANDSCAPE

The previous part presented DSRCT pathognomonic t(11;22) (p13; q12) translocation leading to EWSR1-WT1 fusion and its biological consequences as a unique oncogenic driver, generating a wide transcriptional rewiring.

Historically, sarcomas have been classified into two categories. The first category comprised sarcomas with near-diploid karyotypes and simple genetic alterations (e.g., translocations), while the second encompassed subtypes with complex karyotypes<sup>77</sup>. DSRCTs are classified into the first category and thus considered simple genomics tumors, harboring few additional mutations and few copy number alterations. In this group are included other fusion sarcomas such as synovial sarcoma (SSX-SS18 translocation), Ewing sarcoma (EWSR1-FLI1 translocation), and other STS driven by one main genomic alteration such as rhabdoid tumors (SMARCA4 or SMARCB1 deficiency) (**Fig. 5**).



*Fig. 5. Sarcoma histotypes phylogenetic tree (n=60).*

*This classification is based on relationships among lineage, driver genetic alteration, prognostic, and other parameters. Fusion sarcomas cluster altogether. From Taylor et al., 2011<sup>77</sup>.*

Since its first description in 1991, the characterization of DSRCT molecular landscape has gained increasing interest throughout the last few years. In addition to the depiction of the

various EWSR1-WT1 isoforms, studies have sought to explore both EWSR1-WT1-induced dysregulated pathways and potentially biologically relevant secondary mutations.

### 1.5.1 DSRCT relative chromosomal stability

Although DSRCTs are characterized by a stable chromosomal landscape, several copy number alterations have been described.

Based on targeted Next Generation Sequencing (NGS) using MSK-IMPACT® panels (n=341, 410, and 468 genes for MSK-IMPACT® versions 1, 2, and 3, respectively), Nacev et al. <sup>78</sup> highlighted several significantly recurrent copy number variations (CNVs) including gains of 1q, 5p, 5q, and 16q within a cohort of 48 DSRCTs. Within the same cohort, recurrent loss of heterozygosity (LOH) <sup>79</sup> was identified in chromosome arms 11p, 11q, and 16q in 18%, 22%, and 34% of 50 patients, respectively, suggesting their potential biological relevance in tumor progression.

Using single nucleotide polymorphism arrays, Sydow et al. <sup>80</sup> further studied 25 samples from 19 DSRCT patients and found CNVs in all samples, the most common of which being gains of 1/1q and 5/5p and loss of 6/6q and 16/16q.

To date, the biological role of such alterations remains largely unknown. Of note, chromosome 6 is enriched in genes associated with nucleosome formation, notably in genes that are members of the histone H1 (HIST1 H) family, some of which are the targets of missense mutations in DSRCT <sup>78</sup>, as well as in HLA class I genes, which may explain decreased HLA expression in DSRCT.

### 1.5.2 DSRCT genomic stability

Despite clear evidence of EWSR1-WT1 oncogenic capabilities, attempts have been made to identify potentially recurrent secondary mutations on DSRCT tumors (**Table 2, Fig. 6**), which may provide hints on EWSR1-WT1 essential partners and uncover antitumoral treatment mechanisms of resistance. Additionally, this could deepen the knowledge of the biological pathways involved downstream EWSR1-WT1 activity and thus help to define novel pharmacological targets.

Seminal work on both Ewing sarcoma and DSRCT molecular characterization relying on targeted exome sequencing highlighted a very low amount of secondary mutations <sup>81</sup>. A first study focusing on pediatric malignancies using NGS revealed gene amplification of *AURKB* and *MCL1* without putative pathogenic significance and on a restricted number of tested genes <sup>82</sup>. Among the 10 DSRCT patients evaluated in the latter study, only two exhibited secondary mutations on *GRB10*, *PTPRD*, *KRAS*, *MET*, and *PIK3CA*. This observation was confirmed later

in a multiplatform profiling study (DNA sequencing and protein biomarkers analysis), in which a limited number of mutations were found for DSRCT samples<sup>83</sup>. However, these studies were based on the analysis of a small panel of genes, which narrows their interpretability. More comprehensive panels have subsequently been used to characterize DSRCT secondary mutations, including FoundationOne Heme and MSK-IMPACT® NGS panels, each involving more than 400 genes (**Table 2**). In the first study, describing 83 DSRCT cases but without matched normal tissue, authors underlined DSRCT low mutational burden in 96% of patients and identified alterations in *FGFR4*, *ARID1A*, *TP53*, *MSH3*, and *MLL3*<sup>84</sup>. In the second study, evaluating 68 DSRCT cases with matched germline tissue, recurrent mutations were rare but were identified in *TERT* (3%), *ARID1A* (6%), *HRAS* (5%), and *TP53* (3%), along with low mutational rate with a median of 0.80 mutations/Mb<sup>79</sup>.

Access to Whole Exome Sequencing technologies (WES) enabled to investigate more exhaustively the mutational landscape of DSRCT. Devecchi et al.<sup>85</sup> analyzed the copy number alterations and somatic mutations of six pretreated DSRCT patients' archival tissues using WES. In total, 137 somatic mutations were detected, but were mostly case-specific. Only two mutations were found recurrent (*MUC19* and *GUSBPI*) but at different gene positions. This work revealed that among the identified mutations, 27% could be classified into two main categories: DNA damage response (DDR), Epithelial-Mesenchymal transition (EMT), and Mesenchymal-Epithelial Transition (MET). Concerning DDR, a significant part of the mutated genes concerned RNA-binding protein. The most recent study exploiting DSRCT WES data reached a similar conclusion, describing DSRCT low mutation rate (around 0.72 mutations per Mega bases) and identifying mutations on previously described genes such as *ARID1A*. Interestingly, the authors also analyzed the phylogenetic evolution of mutations and CNVs upon multiple tumor sites within patients, and found limited and late evolution between distinct tumor sites<sup>86</sup>.

In total, the generation of a comprehensive landscape of DSRCT secondary mutations is hampered by the variability of the sequencing gene panels used in such studies and the inconsistent use of matched normal tissue, which may overestimate somatic mutation rates. Nonetheless, in line with the molecular characterization of most fusion-driven sarcoma<sup>78,87</sup> (e.g., Ewing sarcoma<sup>88,89</sup>, synovial sarcoma<sup>78</sup>, rhabdomyosarcoma<sup>90</sup>), DSRCT exhibits a markedly stable genomic landscape with generally <1 mutation /Mb<sup>79,85,86</sup> and few secondary mutations, which are rarely recurrent across patients (**Fig. 6**). However, the few recurrent mutations identified in DSRCT most often occur in genes involved in DNA damage response,

Epithelial-Mesenchymal Transition/ Mesenchymal-Epithelial Transition, and immune response<sup>85</sup> suggesting a biological relevance of these pathways in DSRCT tumor progression.

*Table 2. DSRCT Targeted Next Generation Sequencing and Whole Exome Sequencing studies.*

Author	Chow et al.	Slotkin et al.	Devecchi et al.	Wu et al.
Year	2020	2021	2018	2022
Reference	84	79	85	86
Number of analyzed samples	83	68	6	22
Panel	FoundationOne Heme assay	MSK-IMPACT®	WES	WES
Normal matched tissue	No	Yes	Yes	Yes
Number of analyzed genes	406	468	exome	exome
Number of identified mutations	137	68	137	460
AQP7				15%
ARID1A	11%	6%		15%
C14orf23				15%
C22orf42				15%
DKZFp434P0216				15%
FADS6				15%
FGFR4	8%	7%		
GUSBP1			33%	
HRAS		4%		
KRTAP9-1				15%
LNP1				15%
LOC283685				15%
MAN2C1				15%
MAST3				15%
MLL3	16%			
MSH3	14%			
MUC16				23%
MUC19			33%	
MUC4				15%
MUC6				15%
OR4C45				15%
ROBO2				15%
TERT		3%		
TMEM128				15%
TNN				15%
TP53	10%	3%		
ZNF773				15%

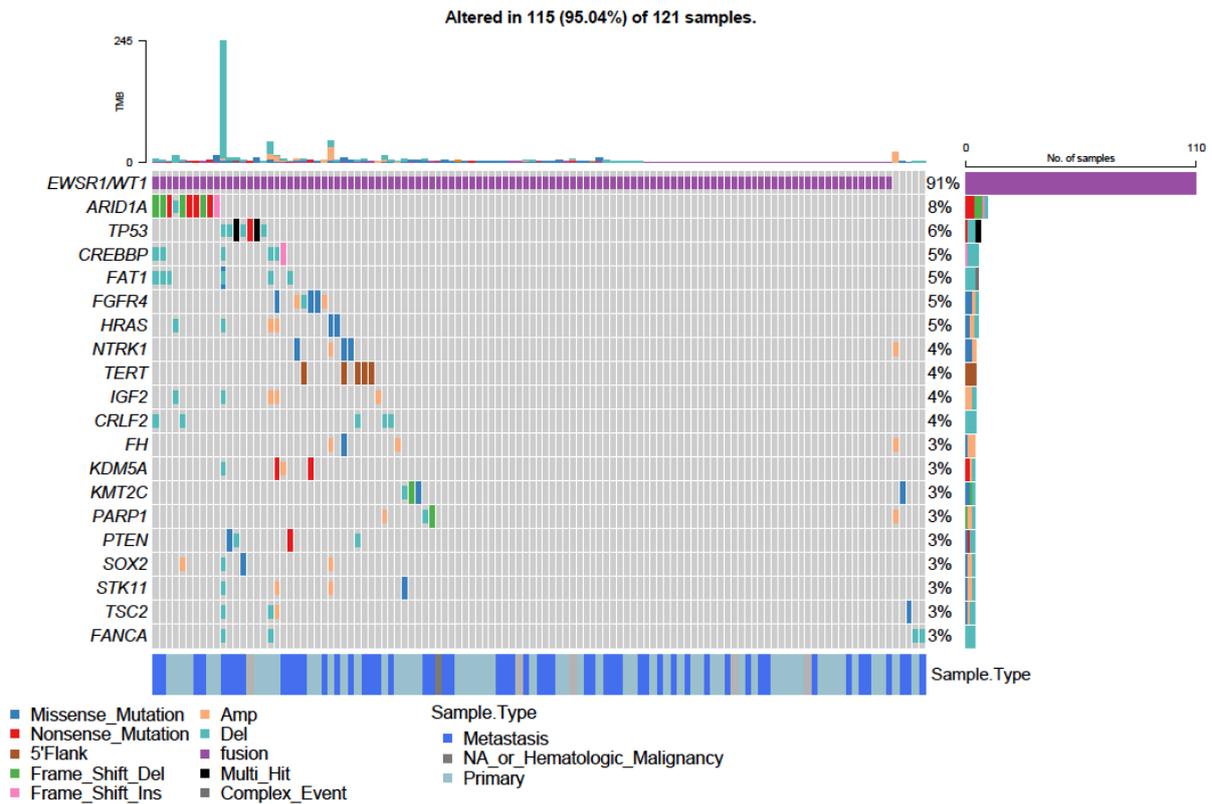


Fig. 6. DSRCT recurrent mutations according to the Whole Exome Sequencing (WES) GENIE DSRCT cohort.

The top 20 recurrent mutations are displayed.

### 1.5.3 DSRCT altered transcriptomics pathways

Since EWSR1-WT1 behaves as a transcriptional activator, identifying downstream genes and pathways specific to DSRCT could open therapeutic opportunities.

The first evidence showing that EWSR1-WT1 fusion protein was leading to gene expression dysregulation was provided by a study using a mouse conditional knock-in model for EWSR1-WT1 expression<sup>69</sup>. The authors constructed two models containing or not the KTS motif. When using Mouse Embryonic Fibroblasts (MEFs) from these mice strains, significant expression changes were detected for thousands of genes upon the induction of either of the two distinct isoforms. Notably, very few overlapping genes were found regulated by both the -KTS and +KTS isoforms, suggesting that both isoforms are likely to be recruited at distinct genomic regions and to regulate specific downstream genes, in line with previous *in vitro* and *in vivo* observations (please refer to Section 1.4.2 for additional information on EWSR1-WT1 binding sequence). One particularly interesting point in this study was that upon EWSR1-WT1 -KTS induction, Gene Ontology (GO) analysis revealed a strong enrichment in neural-related genes.

To confirm this observation, the authors relied on a previously published work <sup>91</sup>, in which the transcriptome of 28 DSRCT tumors was analyzed by microarrays and compared to other tumors, including alternate fusion-driven sarcomas. When doing such a comparison, authors observed that neural pathway genes were significantly enriched in the DSRCT samples. Notably, a subset of upregulated genes upon EWSR1-WT1 -KTS induction in the MEFs was also detected as upregulated in DSRCT tumors compared to other sarcomas. These genes included *ASCL1*, *PLXNB1* and *NTRK3*.

Interestingly, similar observations have been established in Ewing sarcoma, in which EWSR1-FLI1 up-regulates critical genes involved in neural crest development <sup>92</sup>. Later on, several groups performed transcriptome analysis of DSRCT using RNA-seq <sup>93-95</sup>. When compared to some other pediatric tumors, including central nervous system tumors (CNS), glioblastoma, glioma, and other sarcoma, DSRCT transcriptome clusterized within a subcluster of rare sarcomas, suggesting a common behavior with alternate fusion-driven sarcomas, possibly due to the proximity of the cell of origin.

Smith et al. notably reported a signature based on Gene Set Variability Analysis (GSVA) and highlighted that DSRCT top oncogenic signature genes included *MEK*, *EGFR*, *ERB2*, and *RAF* <sup>94</sup> (**Fig. 7**). Interestingly, the *EGFR* pathway had also been previously shown to be enriched in DSRCT compared to other sarcomas <sup>93</sup>. Another pathway that has gained interest as it was found to be upregulated in several independent DSRCT transcriptomic studies is the gene coding for the Androgen Receptor (AR). A recent study confirmed this observation using genomics, proteomics, and functional genomics approaches, and showed that AR targeting in patient-derived xenograft models (PDXs) could represent a valuable therapeutic approach <sup>96</sup>.

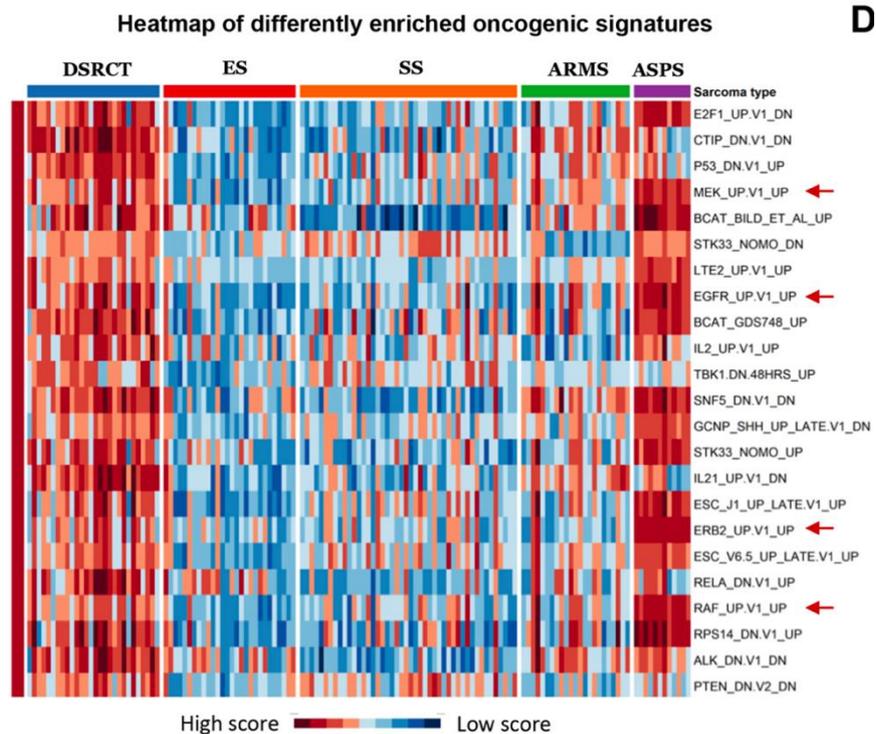


Fig. 7. Heatmap representing mRNA-based oncogenic signatures scores in five selected sarcoma histotypes.

DSRCT: Desmoplastic Small Round Cell Tumor, ES: Ewing Sarcoma, SS: Synovial Sarcoma, ARMS: Alveolar Rhabdomyosarcoma, ASPS: Alveolar Soft Part Sarcoma. From Smith et al. <sup>94</sup>

As stated above, Hingorani et al. <sup>97</sup> combined EWSR1-WT1 ChIP-seq and RNA-seq experiments in JN-DSRCT-1 cell line with microarray transcriptome data from publicly available datasets of fusion-driven sarcomas (rhabdomyosarcoma, alveolar soft part sarcoma, DSRCT, and Ewing sarcoma). Gene set enrichment analysis on DSRCT specifically upregulated genes highlighted an activation of DNA damage response, muscle development, and chromatin remodeling pathways (**Table 3**).

GSEA term	p value	FDR (BH)
Cellular protein complex disassembly	<0.001	0.03
Protein complex disassembly	<0.001	0.03
DNA damage checkpoint	<0.001	0.02
DNA integrity checkpoint	<0.001	0.02
DNA dependent DNA replication	<0.001	0.03
Cellular component disassembly	<0.001	0.03
Establishment and or maintenance of chromatin architecture	<0.001	0.03
DNA repair	<0.001	0.03
Response to endogenous stimulus	<0.001	0.03
Response to DNA damage stimulus	<0.001	0.03
Chromatin remodeling	<0.001	0.03
Positive regulation of transferase activity	<0.001	0.05
Striated muscle development	<0.001	0.06
Skeletal muscle development	<0.001	0.05
B cell activation	<0.001	0.07
Positive regulation of protein modification process	<0.001	0.04
Regulation of multicellular organismal process	<0.001	0.05
Hydrolase activity hydrolyzing O-glycosyl compounds	<0.001	0.05
Amine transport	<0.001	0.05
Hallmark: myogenesis	<0.001	0.02
Muscle development	<0.001	0.02

Table 3. Differentially enriched gene sets in DSRCT compared to other sarcoma subtypes. From Hingorani et al.<sup>66</sup>.

Interestingly, EWSR1-WT1 ChIP-seq enriched peaks were found on known EWSR1-WT1 targets such as *TSPAN* but also on new putative genes (e.g., *IGF2*, *FGFR4*, *CTCFL*, *PEX5* and *ROCK1*). Of particular interest, EWSR1-WT1 associated peaks corresponding genes were enriched for specific pathways, such as Wnt, Notch, and components of the extracellular matrix, that may constitute targetable vulnerability pathways in DSRCT.

Other recent studies using *in vitro* models isolated from DSRCT patients aimed at describing EWSR1-WT1 target genes by using functional genomics approaches<sup>44,98</sup>. The use of a short hairpin RNA (shRNA) or small interfering RNA (siRNA) to silence EWSR1-WT1 followed by RNA-seq was used in both studies to explore the effect of EWSR1-WT1 silencing on gene regulation (**Fig. 8**). Both studies found some overlapping genes likely to be modulated (directly or indirectly) by EWSR1-WT1 and highlighted a strong dependency on *ERG* expression, a *FLI1* family member. Interestingly, both studies highlighted some similarities between DSRCT and Ewing Sarcoma, including the role of EWSR1-WT1 in DNA repair, transcription, TGF $\beta$ , and IGF pathways<sup>44</sup>. In addition, similarly to EWSR1-*FLI1* silencing in Ewing Sarcoma, EWSR1-WT1 knock-down affected tumor cell shape and proliferation, likely by affecting some genes involved in cell adhesion and cell migration.

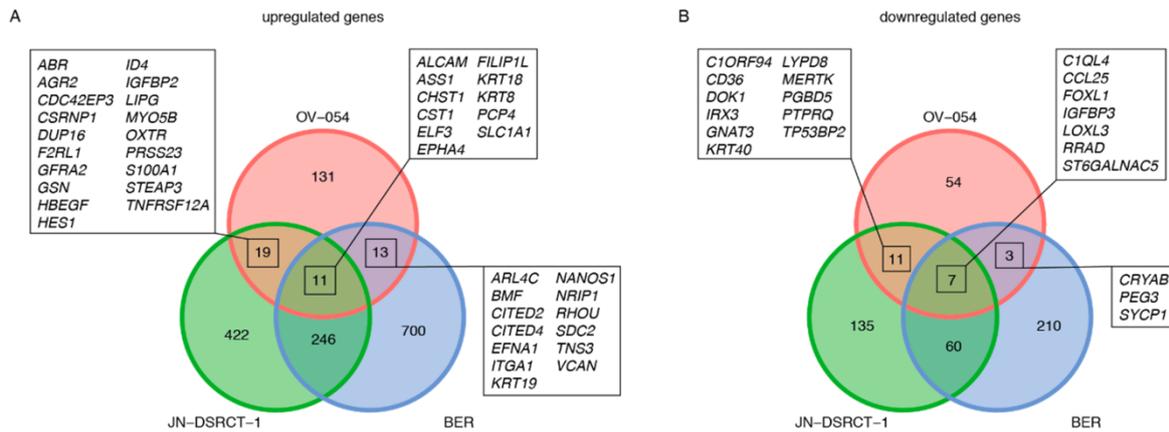


Fig. 8. *EWSR1-WT1* upregulated and downregulated genes.

From Bleijs et al. <sup>98</sup>.

More importantly, it has been recently shown that diverse cancers characterized by chimeric TFs, including Ewing sarcoma and DSRCT, induce the expression of highly specific and novel spliced transcripts, some of which are efficiently translated <sup>99</sup>. While the function of these neoproteins remains unknown, this observation provides new opportunities to identify DSRCT-specific neoantigens that could be used as therapeutic targets.

#### 1.5.4 DSRCT-specific methylation profile

In line with gene expression-based studies, DNA methylation analysis of several small round cell tumors revealed an evident heterogeneity between evaluated sarcoma subtypes <sup>100</sup>. When specifically focusing on fusion sarcoma, DSRCT exhibited a specific signature when compared to Epithelioid Sarcoma (ES), Synovial Sarcoma (SS), Alveolar Rhabdomyosarcoma (ARMS), and Alveolar soft part sarcoma (ASPS).

#### 1.5.5 DSRCT microenvironment

Concerning DSRCT microenvironment, the latter is constituted by a typical dense desmoplastic stroma, which is usually poorly infiltrated by T cells <sup>101</sup>. Interestingly, when comparing DSRCT primary and recurrent tumors, authors observed that the CTL (cytotoxic T-lymphocytes) score, which combines HLA A/B/C and  $\beta$ 2-microglobulin IHC expression, was even lower at relapse compared to primary tumor <sup>101</sup>, suggesting the role of immune escape in tumor recurrence and therapeutic resistance. Transcriptomics studies also provided clues on DSRCT immune landscape. The analysis of neoantigens and Major Histocompatibility Complex class I (MHC I) expression from RNA-seq datasets revealed that DSRCT is characterized by a low MHC I

and neoantigens expression<sup>95</sup>. This observation is consistent with the low mutational burden observed in DSRCT since neoantigens expression often correlates with mutational burden. The RNA-seq-derived inference of immune infiltrate showed that DSRCT, like a large part of sarcoma subtypes, exhibits a low immune infiltrate and is characterized as “cold” tumors<sup>95</sup>.

## 1.6 DSRCT: MOLECULAR VULNERABILITIES AND PROMISING THERAPEUTICS

As previously mentioned, no targeted therapy is currently approved for DSRCT apart from pazopanib. However, recent preclinical studies and clinical trials have shown promising efficacy in DSRCT targeting. Several reviews recently summarized DSRCT potential therapeutic opportunities and current clinical trials<sup>102,103</sup>. The distinct current and future therapeutic strategies to target DSRCT cells are outlined in **Fig. 9**. Briefly, initial strategies focused on the targeting of EWSR1-WT1 downstream pathways such as PI3K/AKT/mTOR, VEGF/PDGF, EGF, IGF, or AR pathway. The recent identification of NTRK3 activation also paved the way for novel NTRK inhibition strategies (**Table 4**). One of the most promising preclinical studies notably involves DNA repair targeting, relying on DSRCT cell inherent replicative stress, as exemplified by the evaluation of SIK1/CHK1 double inhibition.

Notably, with the advent of anti-tumor immune therapy, the place of immune checkpoint inhibitors such as PD-1/PD-L1 inhibitors in the management of sarcoma has been recently questioned. In the AcSé Pembrolizumab phase II study, six patients with advanced DSRCT were treated with pembrolizumab, an anti-PD-1 monoclonal antibody, resulting in one partial response (17%) and a median progression-free survival of 5 months<sup>104</sup>.

Overall, the therapeutic strategies developed to date to target DSRCT resulted in disappointing clinical improvements, which may result from the variety of activated oncogenic pathways downstream EWSR1-WT1 and the scarcity of clearly identified vulnerabilities. Importantly, it is likely that DSRCT cell heterogeneity also plays a crucial role in tumor progression and therapeutic resistance. Indeed, despite a unique oncogenic driver and potentially rare secondary mutations, DSRCT cancer cells may move dynamically across a continuum of cell states that allow them to adapt to microenvironment changes, therapeutic pressure, or tumor site relocation. The comprehensive characterization of DSRCT cell heterogeneity may enable the detection of the most tumorigenic or prometastatic subpopulations that should be specifically targeted.

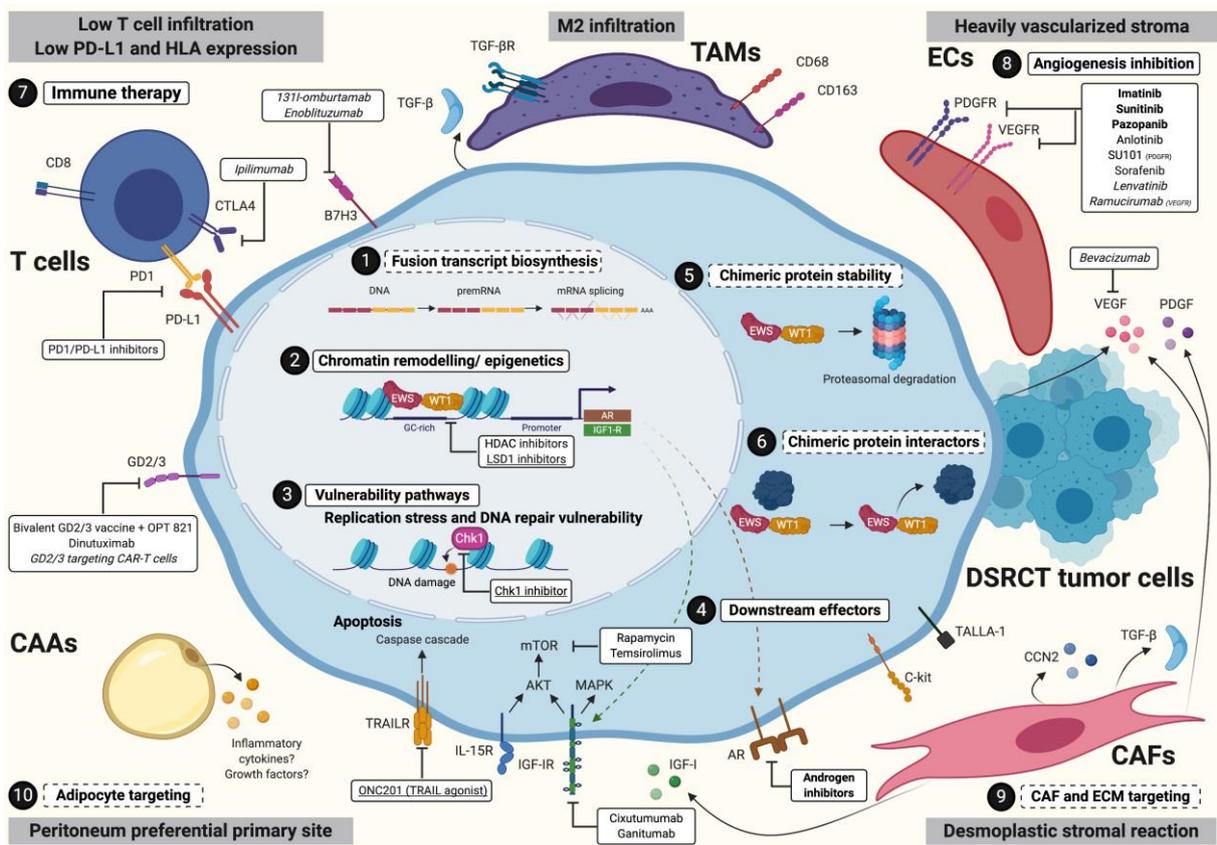


Fig. 9. DSRCT vulnerability pathways and therapeutic opportunities.

Table 4. Targeted therapies evaluated in DSRCT preclinical studies.

Drug class	Compound	In vitro experiments	In vivo experiments	Biological mechanism	Reference
TRAIL agonist	ONC201	IC50: Treatment with increasing concentrations (0-20 $\mu$ M) of ONC201 for 72 h decreases cell proliferation in JN-DSRCT-1 cells with a maximum inhibition achieved at 10 $\mu$ M (IC50=1.66 $\mu$ M)	JN-DSRCT-1 orthotopic xenograft model: the response rate in the low-dose group was 35% with a rebound effect after treatment withdrawal, compared to 60% in the high-dose group	Caspase activation and apoptosis induction	105
		CFA: Treatment of cells with 0.5-1 mM ONC201 for two weeks induced a dose-dependent decrease in the number of DSRCT clones compared with untreated cells			
Chk1 inhibitor	Prexasertib	None in DSRCT	PDX models: complete tumor regression was observed in two DSRCT PDXs	DSRCT replication stress and DNA damage vulnerability, Chk1 induction of DNA damage response	106
KDM1/LSD1 inhibitor	ORY-1001, GSK2879552, SP2509	2D growth assays: JN-DSRCT-1 cells display no sensitivity to treatment with the irreversible inhibitors of KDM1A demethylase function ORY-1001 and GSK2879552 but showed sensitivity to the reversible KDM1A inhibitor SP2509	Not explored	Not explored	107
PARP inhibitor	Olaparib+temozolomide	2D growth and migration assays: olaparib and temozolomide monotherapy showed dose-response growth inhibition of JN-DSRCT-1 cells. Combination therapy exhibited a synergistic effect	JN-DSRCT-1 cell subcutaneous xenograft models: combination therapy showed increased tumor response compared to monotherapies on Days 7, 14, and 21, but PDX reached a similar tumor volume as olaparib monotherapy on Day 28	Olaparib treatment reduced cell viability and migration dose-dependently through caspase activation. Combination treatment led to a cell-cycle arrest and induction of DNA damage and apoptosis, even when combined at low dosages	108
EGFR inhibitor	Afatinib, Cetuximab	2D growth assays: Both monotherapies inhibit cell growth of the three DSRCT cell lines (BER-DSRCT, SK-DSRCT-1, SK-DSRCT-2) dose-dependently with an IC50 around 0.3-0.6 $\mu$ M. Combination therapy does not improve the cytotoxic effect	BER-DSRCT, SK-DSRCT2 cell subcutaneous xenografts, and one PDX: No effect was observed on cell line xenografts treated with afatinib monotherapy. Tumor growth was	Caspase activation and induction of apoptosis	94

			impaired with cetuximab monotherapy and afatinib+cetuximab. Similar results were obtained for PDX.		
<b>NTRK3 inhibitor</b>	<b>Entrectinib and repotrectinib</b>	<b>2D growth assays:</b> Two DSRCT cell lines were tested (JN-DSRCT-1 and SK-DSRCT2). Entrectinib and repotrectinib inhibit cell growth in both cell lines with an IC50 between 0.42 and 0.95 $\mu$ M for entrectinib and 1,64 $\mu$ M for repotrectinib	<b>PDX models:</b> Entrectinib led to tumor growth inhibition with near total regression of one of the three tumors and a tumor growth reduction of 50 and 75%, respectively, when compared to vehicle-treated mice	Loss of ERK signalization upon treatment suggesting a mitogenic signaling dependency breakage in treated DSRCT cells.	109
<b>SIK1+CHK1 inhibitor</b>	<b>YKL-05-099 and prexasertib</b>	<b>2D growth assays:</b> Two DSRCT cell lines were tested (JN-DSRCT-1 and BER). YKL-05-099 inhibited BER growth dose-dependently with an IC50 of 3.5 $\mu$ M. JN-DSRCT-1 was less sensitive to YKL-05-099. The same observation was done for prexasertib (IC50 for BER 1.3 $\mu$ M and 14 $\mu$ M for JN-DSRCT-1). Combination therapy with both YKL-05-099 and prexasertib increases the sensitivity of JN-DSRCT-1 cells with an IC50 of 1.9 $\mu$ M for prexasertib when treated with a fixed dose of YKL-05-099.	<b>PDX models:</b> The pharmacological compounds were not tested in mice in this study, but SIK1 knock-down with inducible shRNA is shown to induce tumor growth inhibition.	Cell cycle regulation, cell cycle arrest.	110
<b>AR inhibitors</b>	<b>Enzalutamide, AR-ASO</b>	<b>2D growth assay:</b> Enzalutamide and AR-Antisense oligonucleotide inhibit JN-DSRCT-1 growth dose-dependently. Enzalutamide IC50 was 0,046 $\mu$ M and AR-ASO IC50 was 0.012 $\mu$ M.	<b>JN-DSRCT-1 cell subcutaneous xenografts or PDX:</b> Enzalutamide inhibited tumor growth significantly during the first two months, but tumor growth accelerated after this timepoint. AR-ASO-treated mice exhibited tumor growth inhibition with a long-lasting response after two months.	Not explored	96

## 2 TUMOR HETEROGENEITY AND CANCER CELL PLASTICITY

---

### 2.1 TUMOR HETEROGENEITY AND CANCER CELL PLASTICITY: GENERAL PRINCIPLES

#### 2.1.1 Definitions

The term ‘tumor heterogeneity’ was historically used to characterize cancer cell genetic diversity and has long been limited to the spectrum of clonal diversity<sup>111</sup>. The role of genetic diversity as a critical player in intratumoral heterogeneity has been primarily established long ago<sup>112</sup>. Indeed, genetic mechanisms have been shown to contribute to cancer cells’ adaptation to the microenvironment, and the latter reversely also participates in genome destabilization<sup>113</sup>. Tumor genetic heterogeneity can arise from genetic or chromosomal instability, from single-nucleotide substitutions to chromosomal rearrangements and whole genome amplification. Importantly, mounting evidence shows that non-genetic mechanisms also play a central role in tumor cells’ heterogeneity and adaptability. We will refer to cancer cell plasticity to designate these non-genetic events driving tumor cells’ spatial and temporal adaptability, whether these mechanisms rely on epigenetic, transcriptional, or translational processes. However, the term ‘cancer cell plasticity’ does not have any consensual definition, and it is generally used to characterize the cells’ dynamical adaptation to change their phenotype without genetic mutations in response to environmental cues or genetic events. These changes can be reversible or not, continuous or working in a switch on/switch off manner<sup>114</sup>.

It is now widely accepted that tumor heterogeneity results from both stochastic genetic events and cancer cell plasticity, giving rise to cell state variations in relationships with microenvironmental signals.

#### 2.1.2 Tumor cells heterogeneity: evolution models

Tumor clonal dynamics have first been widely studied from sequential and primary/metastasis next-generation sequencing of tumor genomes. Initially, two main tumor heterogeneity models had been described: the clonal evolution theory and the cancer stem cell theory (CSC) (**Fig. 10**).

The cancer stem cell model relies on the hypothesis that tumors are hierarchically organized into subpopulations of tumorigenic cancer stem cells arising from a founder cell, constituting an autorenewing niche differentiating into distinct subclones with limited tumorigenic potential<sup>115</sup>. In this model, cancer stem cells are thought to drive tumor initiation, tumor growth,

metastatic progression, and therapy resistance, and can partially recapitulate the heterogeneity of the tumor. The pivotal work that primarily established this CSC theory showed that putative breast cancer stem cells were identified as CD44+/CD24- cells and could, when injected into immunodeficient mice in a low number, initiate tumor formation in ~90% of cases <sup>116</sup>.

The clonal evolution theory, first introduced by Peter Nowell in a landmark publication almost four decades ago <sup>117</sup>, posits that tumors arise from a single cell of origin, which will derive into distinct subclones along the course of tumor progression, resulting from the stochastic acquisition of mutations within the initiating clone, and allowing the selection of the most aggressive or adaptive to change subclones. Several distinct models have further been described, such as the Darwinian clonal evolution, the parallel evolution, and gradualism or macroevolution <sup>112</sup>.

Several observations highlight that tumor heterogeneity is a highly complex phenomenon, suggesting that the two previously presented models might not be mutually exclusive. Actually, the concept of cancer cell plasticity may help to unify these models. Indeed, with the improved understanding of cancer cell plasticity, a third model has emerged, where cancer stem-like cells endure a differentiation/dedifferentiation process and thus have the ability to move between stem and differentiated states in response to intrinsic or extrinsic factors <sup>118,119</sup>.

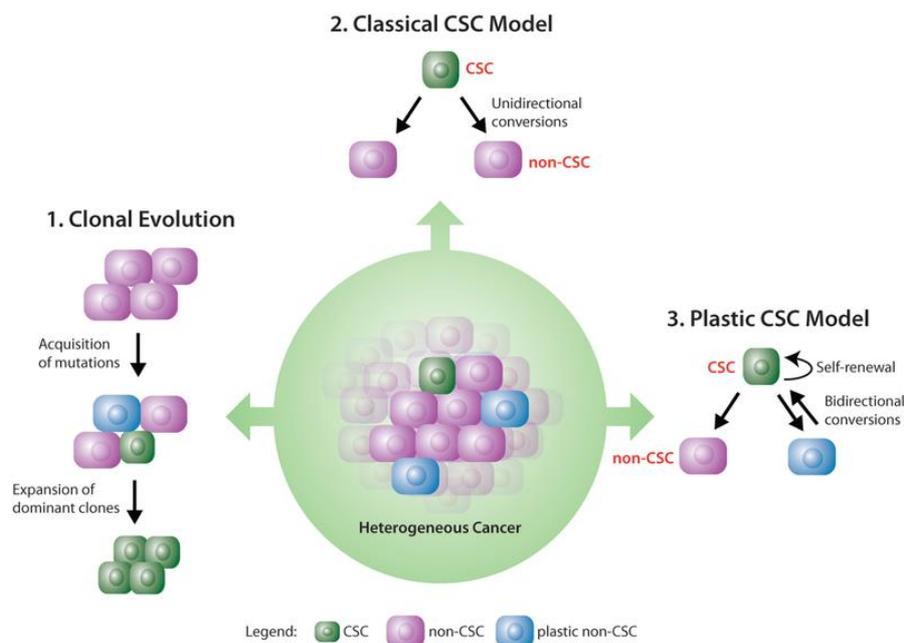


Fig. 10. Tumor heterogeneity models: the clonal evolution, the classical cancer stem cell (CSC), and the plastic CSC theories.

From Marjanovic ND et al., 2013 <sup>120</sup>.

### 2.1.3 Cancer cell plasticity types

Cancer plasticity has recently been added to Hanahan's hallmarks of cancer<sup>121</sup>. Cell plasticity can be defined by continuous transitions among a continuum of phenotypical states. Importantly, this ability is a signature property of stem cells and progenitors during embryonic differentiation. Nonetheless, cell plasticity is not restricted to embryonic cells, and it has been described as an adaptive feature under physiological conditions or pathologic stress in various differentiated adult cells, as well as in cancer cells, whose evolution extensively correlates with developmental paradigms. Hence, cancer cell plasticity is a central mechanism of all stages of tumor development, during premalignant tumorigenesis, tumor initiation, progression, and metastasizing, and under therapeutic pressure.

Several types of cancer cell plasticity have been described, mainly in carcinoma models, among which the best-described example to date relies on the process of Epithelial-to-Mesenchymal Transition<sup>122</sup>. Other examples of cancer cell plasticity have been thoroughly described in exhaustive reviews<sup>122-124</sup>, comprising tumor cells' transdifferentiation (i.e., neuroendocrine differentiation lineage plasticity) or evolutions towards slow-cycling stem cell-like phenotypes<sup>119,122</sup> as a response to escape from therapeutic pressure. Additional types of cancer cell plasticity, outside of developmental or lineage differentiation, may also play a central role in tumor progression and therapeutic resistance. For example, the switch between various metabolic programs may occur in response to proliferative cell states, site-specific changes in nutrient availability, or drug exposure. For example, dynamics between oxidative phosphorylation, anaerobic glycolysis, and fatty acid oxidation constitute hybrid states associated with therapy resistance and metastatic progression<sup>125</sup>.

Interestingly, in line with the recent hypothesis that clonal subpopulations are not necessarily in competition but act in cooperation, cancer cell plasticity potentially reflects tight interdependencies between distinct cell states essential to tumor progression<sup>111</sup>.

### 2.1.4 Cancer cell plasticity sources

Cancer cell plasticity sources can be broadly divided into two origins. First, cell-extrinsic factors involve the microenvironment, including cell-to-cell interactions with immune and stromal cells, tissue composition and architecture, and drug exposure. For example, hypoxia has been described as a source of plasticity by inducing a HIF-1-dependent migration transition from collective-to-amoeboid type in epithelial cancer cells<sup>126</sup>. In addition, several cell-intrinsic factors arise from the acquisition of genetic stochastic mutations or variations in transcriptional

pathways or epigenetic states <sup>118</sup>. Strikingly, additional physical factors may also induce cellular plasticity. For example, integrating a computational disordered chromatin-packing macromolecular crowding (CPMC) model with scRNA-seq, chromatin electron microscopy tomography (ChromEM), and live-cell partial wave spectroscopic (PWS) microscopy, Virk et al. <sup>127</sup> recently showed that disordered chromatin packing itself was amenable to drive and regulate cells' phenotypic plasticity. Notably, the scaling of chromatin packing was able to increase the transcriptional malleability in response to stress, such as antitumor agents <sup>127</sup>. In this study, authors also showed that the extent of tumor cell plasticity was inversely correlated to patients' survival in a subset of solid tumors.

## **2.2 RECENT ADVANCES IN THE UNDERSTANDING OF CANCER CELL HETEROGENEITY**

With the blossom of technologies aiming at characterizing cancer cells at single-cell resolution, the knowledge of cells' states and phenotypic plasticity has recently gained unprecedented resolution.

### **2.2.1 Technical approaches to interrogate tumor cell heterogeneity**

As seen previously, tumor heterogeneity is a complex process but a key feature in cancer biology. Initial experiments investigating tumor heterogeneity first relied on histological analysis to interrogate cells' morphological heterogeneity. Other histological methods such as IHC, immunofluorescence (IF), or *In Situ* Hybridization (FISH) enabled to analyze the heterogeneity of both tumor cells and non-malignant microenvironmental cells (e.g., immune cells or stromal cells). These approaches are instrumental since they preserve spatial tumor organization. However, the latter are low throughput since only a few markers or genes can be interrogated at a time <sup>128</sup>.

Together with the development of NGS technologies, experiments focusing on tumor characterization gained in resolution with the use of Whole Genome Sequencing or RNA-seq, interrogating SNPs, copy number variations, or gene expression in a high throughput fashion. However, using NGS approaches on bulk tumors only authorizes access to mutations or gene expression of a whole population, giving a "mean" overview of heterogeneous cells' exome or transcriptome. On the other hand, such approaches prevent the detection of rare subpopulations or subclones. Accessing the single-cell genome or transcriptome would then allow to characterize intratumor heterogeneity and uncover the biological mechanisms driving such diversity. Former techniques combined technologies such as laser microdissection and previously mentioned NGS techniques to provide a near single-cell resolution.

In the last decade, technological achievements in sequencing, molecular biology, and instrumentations such as microfluidics enabled the development of “true” single-cell methods. Currently, virtually all types of “omics” experiments have been developed to interrogate the transcriptome (single-cell RNA-sequencing), genome (single-cell DNA sequencing), epigenome (single-cell ATAC-sequencing, single-cell ChIP-sequencing), and proteome of single cells <sup>129</sup>. All these methods first rely on the obtention of a suspension of individual cells. Single-cell suspensions are obtained from a tumor biopsy or a surgical specimen after tissue dissociation, usually performed using enzymes such as collagenases, dispase, or hyaluronidase, which digest the extracellular matrix and junction proteins. Acquiring viable cells is critical for the results' quality and interpretation. Thus, protocol optimizations (including the duration and the temperature of tissue dissociation) to avoid cell death during dissociation is critical <sup>130</sup>, as apoptotic or stress-induced signals may significantly affect the cell transcriptome. Single nuclei may also be used in place of single cells <sup>131</sup> and can be advantageous to facilitate the use of frozen samples or tissues that remain hard to dissociate.

#### 2.2.1.1 Non-spatially resolved techniques

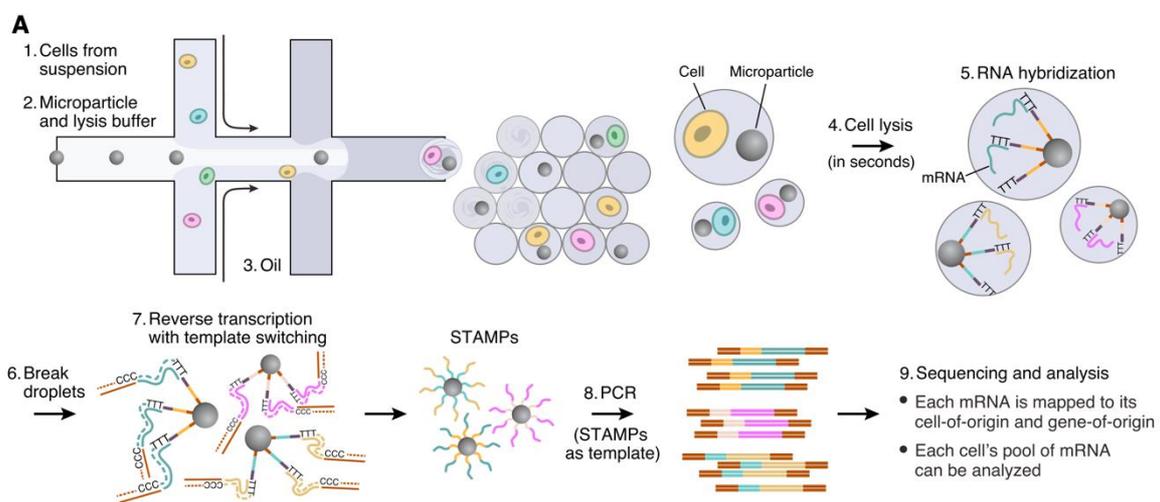
##### 2.2.1.1.1 Single-cell RNA-sequencing

###### 2.2.1.1.1.1 scRNA-seq techniques

Among the different single-cell-based methods, single-cell RNA-sequencing has been the most widely used and developed. The first study reporting the analysis of a single cell's transcriptome was published in 2009 <sup>132</sup>. Since then, several methods have been developed to increase the throughput from a few cells to hundreds of thousands of cells. All of these methods rely on a 4-step protocol comprising (i) the isolation or capture of single cells after dissociation, (ii) a reverse transcription to obtain cDNAs, (iii) cDNAs amplification, and (iv) the construction of the sequencing library.

Single-cell isolation aims at parallelly individualizing numerous RNA-seq experiments. Initially, cells were isolated in microwell plates using micropipettes, dramatically decreasing the throughput <sup>132</sup>. Other methods rely on Fluorescence-Activated Cell Sorting (FACS) to isolate single cells into microwell plates (96 or 384-wells). FACS approaches encompass several methods, such as SMART-seq, MATQ-seq, MARS-seq, and CEL-seq <sup>133–138</sup>. FACS-based strategies also enable selecting specific or rare cell subpopulations by using a specific marker before sorting (e.g., viability markers, cell surface phenotypic marker) but require the manutention of numerous microwell plates to reach a significant number of cells. Another classical method to isolate single cells is the use of microfluidic devices (**Fig. 11**). This

technology relies on the use of equipment built with microchannels carved into a polymer matrix, which allows the manipulation of fluids of microliter quantities and enables the orientation of single cells into microwells (e.g., Fluidigm® technology, Seq-Well protocol<sup>139,140</sup>). Microfluidics devices can also be combined with oils to generate emulsion droplets that capture single cells. These droplets are formed in an aqueous phase, generally containing beads carrying DNA probes used in later steps and RT reagents incorporated with cells. The Drop-seq, 10X Genomics, and inDrop-seq protocols rely on such droplet-based single-cell isolation<sup>141–143</sup> and allow the isolation of thousands of cells into individual droplets and are, to date, the most advanced method in terms of cell throughput.



*Fig. 11. Principle of microdroplet generation using oil and microfluidics device.*

*From Macosko et al., 2015<sup>141</sup>.*

Once the cells are isolated, the next step is a reverse transcription reaction to convert RNA molecules into cDNA strands. Most of the previously described methods use oligodT primers, hence preventing the conversion of non-polyA transcripts into cDNA (e.g., most long non-coding RNA (lncRNA)). This cDNA synthesis step also serves to append a cell-specific molecular barcode on each transcript to identify the cell of origin in the final analysis. These molecular barcodes are unique sequences of  $n$  nucleotides included within the oligodT primer. For droplet-based methods, the single bead incorporated in each droplet together with the cell is covered with this oligodT-cell barcode as exemplified for the 10X Genomics approach in **Fig. 12**.

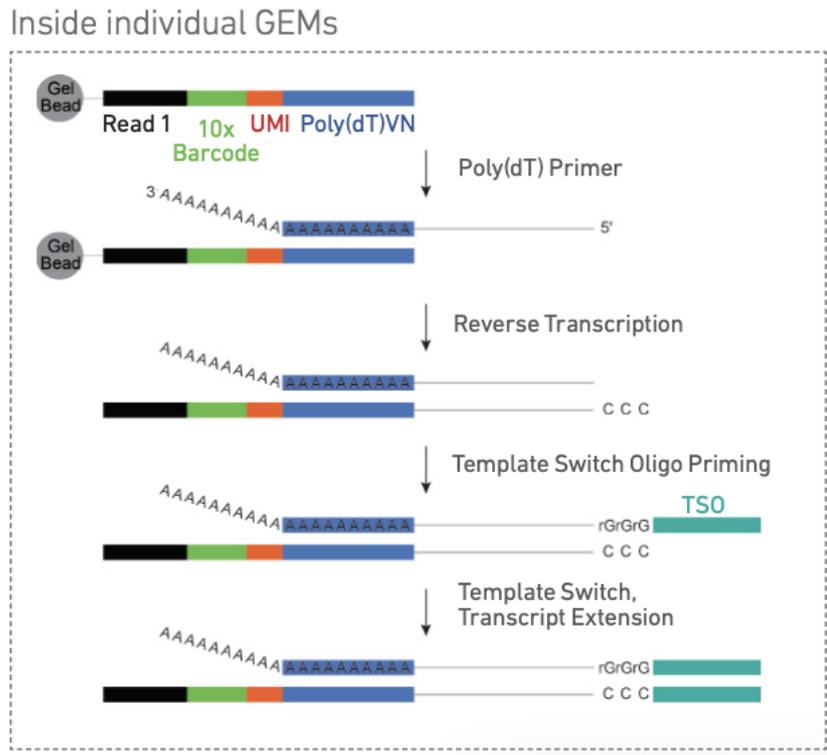


Fig. 12. Principle of in-droplet cDNA synthesis and barcode incorporation during the reverse transcription step of the 10X Genomics protocol.

Because the first strand cDNA contained in a single cell is in too low amount to construct an RNA-seq library, the latter must be amplified. To this aim, two options exist in the abovementioned protocols: an in vitro transcription to linearly amplify the pooled cDNA molecules (e.g., MARS-seq, CEL-seq) or a template switching. The template switching relies on the use of the Moloney Mouse Leukemia Virus (MMLV) reverse transcriptase, whose terminal transferase activity adds a few non-template C nucleotides at the 5' end of the cDNA molecules. These non-templated nucleotides serve as an anchor for an oligonucleotide hybridization that is essential for subsequent PCR amplification. Template switching, which enables full-length recovery of the cDNA, is the most widely used strategy in scRNA-seq protocols (e.g., SMART-seq, Drop-seq, Seq-well, and 10X Genomics).

Finally, the last step consists in generating the sequencing libraries from previously amplified cDNAs. Currently, almost every scRNA-seq method uses the Illumina sequencing technology. This technology is based on short-read sequencing, requiring a mean fragment size of around 500 bp. As cDNA molecules usually display considerable size diversity and are mostly above this 500 bp limit, the library construction requires prior cDNA molecules enzymatic fragmentation. In most techniques, the molecular barcode used to identify the cell of origin is

added either at the 3' or the 5' end of the fragmented cDNA molecules. Consequently, only the 3' or the 5'-end of the transcript will be covered, preventing subsequent analyses such as splicing variant studies. To date, SMART-seq, Fluidigm® and MATQ-seq are the only methods enabling full-length transcript analysis on single-cell experiments. Importantly, to prevent potential PCR amplification biases, cDNA molecules are generally identified with a unique random sequence called UMI (Unique Molecular Identifier). Hence, within downstream analyses, amplified cDNA molecules will *in fine* be quantified using UMI counts instead of transcript counts, avoiding biases stemming from any potential artifactual overamplification. Each scRNA-seq technology contains its own advantages and drawbacks, and the choice of one technique over another should be based on the nature of the tissue of origin (blood/solid tumor, fresh/frozen tissue), the targeted number of cells (hundreds/thousands), the targeted cell types (frequent/rare), the underlying scientific question (e.g., requiring full-length transcripts or admitting 3'/5'-end transcripts), and the cost (**Table 5**). Several studies focused on the comparison of quality control metrics. For instance, Ziegenhain et al.<sup>144</sup> compared six scRNA-seq methods (i.e., CEL-seq2, Drop-seq, MARS-seq, SCRB-seq, Smart-seq, and Smart-seq2) and found out that Drop-seq was more cost-effective for transcriptome quantification of larger number of cells, while Smart-seq2 has the highest sensitivity. In another study, several droplet-based ultra-high throughput single-cell RNA-seq systems have been compared, including 10X Genomics Chromium<sup>145</sup>, which achieved the highest sensitivity compared to alternate methods (e.g., in-Drop, Drop-seq).

Table 5. Specificities of the most used scRNA-seq methods.

Adapted from Jovic et al., 2021<sup>146</sup>.

Platforms	Isolation strategies	Tissue	Cell numbers	Targets	UMI	Amplification methods	Region
Smart-seq	FACS	Dissociated cell	Hundreds of cells	/	×	PCR	Full-length
Smart-seq2	FACS	Dissociated cell	Hundreds of cells	/	×	PCR	Full-length
Fluidigm C1	Micro-fluidic	Dissociated cell	Hundreds of cells	No poly(A) minus RNA detection	×	PCR	Full-length
Drop-seq	Microdroplets	Dissociated cell	Large number of cell	No poly(A) minus RNA detection	√	PCR	3' end
10x Genomics	Microdroplets	Dissociated cell	Large number of cells	No poly(A) minus RNA detection	√	PCR	3' end
MATQ-seq	FACS	Dissociated cell	Hundreds of cells	No poly(A) minus RNA detection	√	PCR	Full-length
Seq-Well	Micro-fluidic	Dissociated cell	Large number of cells	No poly(A) minus RNA detection	√	PCR	3' end
CEL-seq	FACS	Dissociated cell	Hundreds of cells	No poly(A) minus RNA detection	√	IVT	3' end
MARS-seq	FACS	Dissociated cell	Hundreds of cells	No poly(A) minus RNA detection	√	IVT	3' end
inDrop-seq	Microdroplets	Dissociated cell	Large number of cell	No poly(A) minus RNA detection	√	IVT	3' end
DNBelab C4	Microdroplets	Dissociated cell	Large number of cells	No poly(A) minus RNA detection	√	PCR	3' end

#### 2.2.1.1.1.2 scRNA-seq-derived bioinformatics tools

Along with the development of new wet lab technologies enabling to capture single cells' gene expression, the number of bioinformatics tools available for downstream scRNA-seq data analysis flourished as well.

Briefly, scRNA-seq data analysis requires a prior pre-processing step, consisting in performing several quality controls and filtering. After assessing reads quality and duplicates, low-quality cells are filtered, mainly based on the median number of detected genes per cell and the ratio of mitochondrial genes.

Because scRNA-seq data represents a very high dimensional dataset, dimensionality reduction is required and generally first relies on a principal component analysis (PCA) followed either by a t-distributed stochastic neighbor embedding (t-SNE) or a Uniform Manifold Approximation and Projection (UMAP) algorithm.

Downstream analyses are generally grouped into cell-based and gene-based analyses<sup>147,148</sup> (**Fig. 13**). The most basic scRNA-seq pipelines typically focus on cell-clustering, which groups together cells with similar transcriptome, followed by cell type identification according to a differential gene expression analysis performed across previously identified clusters. Further

characterization of cluster-specific transcriptional states usually relies on gene set enrichment analyses based on the top overexpressed genes per each cluster. More advanced tools can focus on the detection of co-expressed genes (i.e., gene modules) or co-regulatory networks (also known as regulons) based on transcription factors' networks. Other tools seek to reconstruct cell hierarchy using trajectory inference methods, the most known of which relies on the concept of "pseudotime", which measures cells' pseudotemporal biological progression. Currently, a myriad of scRNA-seq-based bioinformatics tools have been developed and are continually evolving, and their potential is almost unlimited.

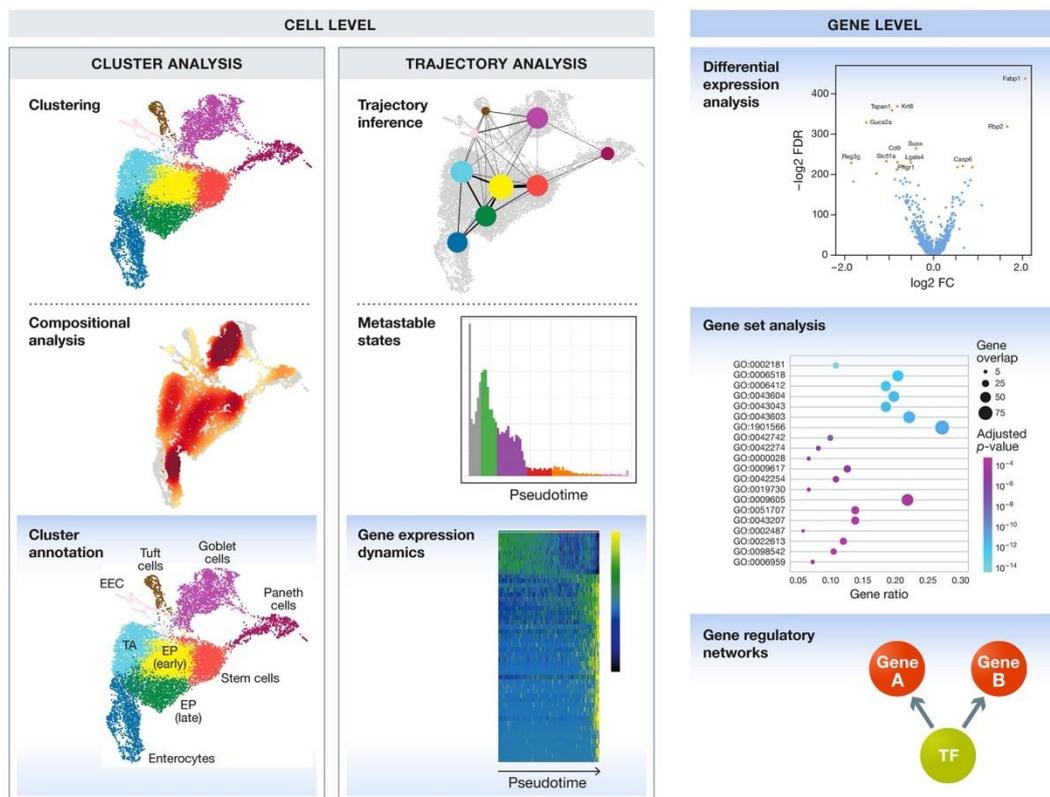


Fig. 13. ScRNA-seq downstream analyses overview.

From Luecken et al., 2019<sup>147</sup>.

#### 2.2.1.1.2 Genome and epigenome-focusing single-cell methods

Several technologies focusing on genome and epigenome characterization at the single-cell level have also emerged. Nonetheless, single-cell whole genome sequencing has to face some technical hurdles. First, the amount of DNA contained within a single cell is way too low to construct a sequencing library and thus requires amplifying the whole genome without introducing biases. To date, available methods for single-cell genome amplification are DOP-PCR (Degenerated Oligonucleotides Primers-PCR)<sup>149</sup>, Multiple Displacement Amplification (MDA)<sup>150</sup>, multiple annealing and looping-based amplification cycles (MALBAC)<sup>151</sup> and

Linear Amplification via Transposon Insertion (LIANTI)<sup>152</sup>. These four methods are all based on short-read sequencing, enabling the detection of SNPs, copy number variation, and small indels, but are not efficient in detecting structural variations such as deletions, insertions, and translocations. An additional method called SMOOTH-seq (single-molecule real-time sequencing of long fragments amplified through transposon insertion) has been recently developed and overcomes this limitation. This method takes advantage of a long-read sequencing technology commercialized by Pacific Bioscience (PacBio) to detect SNV<sup>153</sup>. Despite their high accuracy, these single-cell genome sequencing methods have a substantial limitation: the number of cells that can be analyzed. The sequencing depth required to cover the whole genome of a single cell with optimal accuracy is high and thus limits the number of cells that can be analyzed. So far, the experiments relying on these methods usually examine only hundreds of cells, contrary to scRNA-seq, which can investigate several thousands of cells per experiment. Some studies used single-cell Whole Exome Sequencing (WES), which decreases the coverage needed by focusing only on coding regions of the genome but still suffers from a low cell throughput<sup>154,155</sup>.

The epigenome, including chromatin accessibility, modifications (histone marks or DNA methylation), or topology, can also be interrogated by single-cell techniques. All these single-cell methods are adaptations of already well-established bulk methods, such as single-cell ChIPseq<sup>156,157</sup>, single-cell CUT&Tag<sup>158</sup>, single-cell HiC<sup>159</sup>, or single-cell DNA methylation<sup>160</sup>. Similarly to genome/exome-focused single-cell assays, these methods are also limited in the cell throughput. The most advanced single-cell epigenomic targeting method to date is single-cell ATAC-seq (Assay for Transposase Accessible Chromatin combined with sequencing)<sup>161,162</sup>. Despite being so far limited by some technical issues, all these epigenomics single-cell approaches will eventually become powerful tools to characterize cancer cell heterogeneity, since a dramatic part of cells' phenotype is thought to be driven by non-coding regions of the genome. One can speculate that within the upcoming years, efforts will be made to improve these methods, which, together with their integration with transcriptome or genome single-cell data, will allow to achieve a comprehensive understanding of cancer cells' heterogeneity.

#### 2.2.1.1.3 Single-cell proteomics

Despite revolutionizing molecular cell biology, scRNA-seq still provides surrogate data of actual protein expression. Proteins being the effectors of cell biology, having access to the single-cell proteome is also crucial to fully understand cancer cell biology and thus tumor heterogeneity.

One efficient method to characterize protein expression on single cells is mass cytometry. This technology couples flow cytometry with mass spectrometry. Using non-radioactive isotopes to label antibodies makes it possible to use an increasing number of antibodies simultaneously within the same experiment, since there are no longer conflicts due to the fluorescence spectrum. Thus, it becomes possible to analyze up to 40 proteins at a time for a single cell. Like other methods mentioned above for genome and epigenome profiling, some studies proposed exploiting proteomic approaches already used for bulk analysis to increase cell throughput. These methods rely on proteome analysis from single cells (isolated by FACS or other previously explained methods) using liquid chromatography coupled with mass spectrometry (LC-MS). One method called SCoPE-M and its successor SCoPE2 has succeeded in using an isobaric labeling strategy to analyze approximately one thousand proteins per cell in one thousand cells<sup>163,164</sup>. However, detected peptides corresponded to the most abundant proteins, preventing the detection of lowly expressed proteins. Others also used isobaric labeling but with another cell preparation approach called nanoPOTS to successfully characterize single-cell proteome<sup>165-167</sup>. Single-cell proteome analysis remains in its infancy and is still limited by its throughput, sensitivity, and cost. It will undoubtedly be improved in the upcoming years with accompanying technological advances.

#### 2.2.1.1.4 Multiomics approaches

All previously cited single-cell technologies represent a significant step forward in molecular cell and cancer biology by improving the knowledge on specific cell subpopulations, cell evolution, and cell-to-cell communications. However, the analysis and integration of multiomics data stemming from different assays is a major source of bias. Recently, technologies aiming at coupling several parallel analyses on a single cell have emerged. These methods are referred to as single-cell multiomics technologies. To date, several strategies aiming to simultaneously investigate genome and transcriptome, transcriptome and epigenome, and transcriptome and proteome have been published.

Concerning genome and transcriptome simultaneous analysis, several protocols have been established, such as scTrio-seq, genome and transcriptome sequencing (G&T-seq), simultaneous isolation of genomic and total RNA (SIDR), and TARGETseq<sup>168-172</sup>. These methods differ in how DNA and RNA are selectively recovered and barcoded from single cells. Nevertheless, they are limited in throughput with the same constraints as mono-omic scDNA-seq assays.

Technologies dealing with simultaneous transcriptome and epigenome analysis have also been described. The first multiomic method linking transcriptome and epigenome focused on

transcriptome and methylome co-analysis using scM&Tseq<sup>173</sup>. The epigenome characterization throughout chromatin accessibility analysis coupled with scRNA-seq has also been developed. These approaches gather the combinatorial indexing of single-cell chromatin accessibility and mRNA (sci-CAR), single-nucleus chromatin accessibility and mRNA expression sequencing (SNARE-seq), methylation and transcription sequencing (scNMT-seq), and some commercially available methods (10X Genomics single cell multiome)<sup>174–176</sup>. These technologies enable analyzing thousands of cells at a time and simultaneously collecting gene expression and chromatin accessibility on a single cell.

Lastly, technologies that parallelly investigate single cells' transcriptome and proteome are also under development. Published methods include proximity extension assay/specific RNA target amplification (PEA/STA), proximity ligation assay for RNA (PLAYR), or cellular indexing of transcriptomes and epitopes (CITE-seq)<sup>177–179</sup>. These techniques all use antibodies to target and barcode proteins besides mRNA barcoding. The use of antibodies prevents a whole proteome characterization. Thus, these methods can only analyze targeted proteins' expression besides gene expression, up to 40 proteins for PLAYR.

All these emerging methods bring new opportunities to explore cellular heterogeneity by providing multiple layers of mechanistic dissection underlying single cells' biology. Nevertheless, except for single-cell chromatin accessibility coupled with transcriptome, these technologies remain limited in cell throughput.

#### 2.2.1.2 Spatially resolved technologies

All previously presented single-cell methods are performed from a single-cell suspension, thus without preserving information on cells' localization and surroundings within the tissue of origin. The topology of tissues remains essential to understand tumors' biology. Thanks to technical progress achieved within the last few years, methods to investigate transcriptome and proteome in a spatially resolved way emerged. These methods' main challenge is enabling the indexing or barcoding of transcripts *in situ*. After sequencing with high throughput methods, the transcriptome can be assigned to its original location in the tissue. To date, two different approaches to spatially index RNA molecules have been developed<sup>180</sup>. The first one is the Solid-phase transcriptome capture method. It relies on the use of arrayed barcoded oligo-dT primers that are immobilized on a glass surface. The tissue section is then applied onto the glass surface to perform RNA barcoding by RT. Technologies using this method include 10X Genomics commercial protocol called Visium. The slide used in the Visium assay comprises 5,000 spots of 55µm diameter, conferring a resolution which is not reaching single cells. A

protocol called Slide-seq and its later improvement, Slide-seq V2, uses smaller beads with a diameter of 10 $\mu$ m to achieve a near single cell resolution<sup>181,182</sup>. However, the captured areas still include multiple or at least a few cells, preventing a direct detection of distinct cell types. Noteworthy, several bioinformatic tools have been designed to deconvolute and infer cell types within these areas<sup>183,184</sup>. Other groups developed a method to reach subcellular resolution. For example, by using a high amount of individually barcoded beads and a patterned slide with an increased number of hexagonal arrays, Vickovic et al. could even decipher the subcellular compartmentalization of gene expression<sup>185</sup>. Another approach named Stereo-seq, performing *in situ* RNA capture on DNA nanoballs into patterned flow cells, allowed to reach a resolution lower than a micrometer and to profile whole embryos, small organisms, or several small samples at the same time<sup>186</sup>. Another method based on the same principle, Seq-Scope, recently reached subcellular resolution<sup>187</sup>.

All the previously mentioned technologies are based on Solid-phase transcriptome capture approaches. The other family of spatial transcriptomics methodologies relies on deterministic spatial barcoding methods. The latter techniques are based on the selection of regions of interest (ROI) to be barcoded. Initially, laser microdissection followed by cell capture was used to single cell transcriptome<sup>188</sup>. More sophisticated protocols such as TIVA-seq, NICHE-seq, ZIP-seq, or the commercially available GeoMX from Nanostring are now available<sup>189–191</sup>. Despite using different chemical approaches, all these methods use a reversible marker to define an ROI *in situ* and barcode RNA molecules within this ROI before releasing those barcoded RNA molecules for further sequencing.

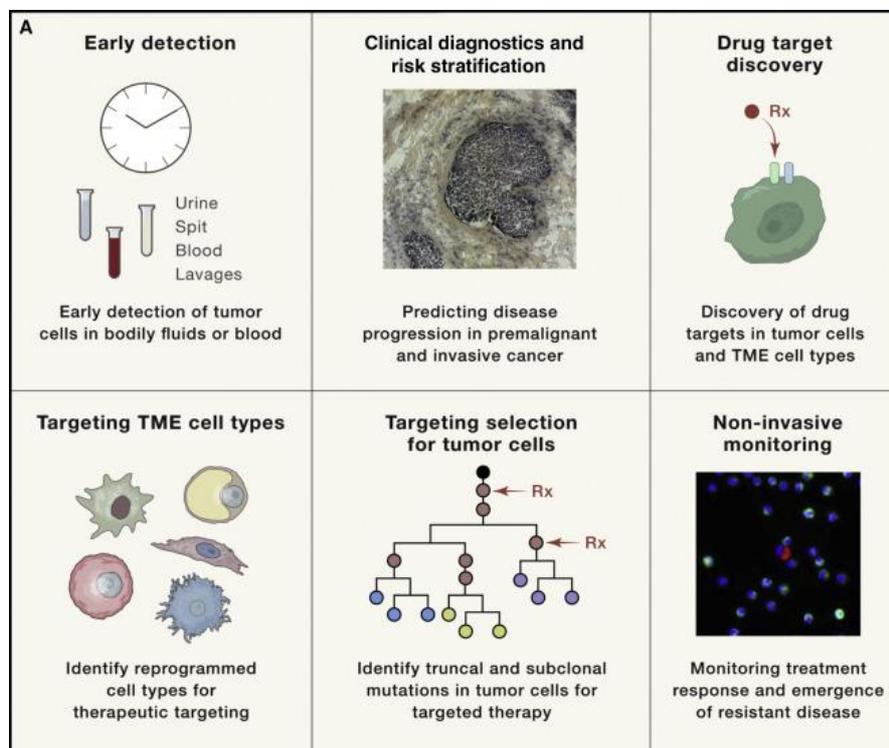
Finally, within the last years, a growing number of methodologies based on imaging have been developed to detect several mRNA molecules within tissues at the subcellular resolution. These methods are of particular interest when the aim is to detect shallow levels of RNA. Several protocols have been published, including FISH-derived high-resolution methods, as reviewed in Moffitt et al.<sup>180</sup>.

### 2.2.2 Single-cell omics in cancer research

The aforementioned single-cell technologies have been used in several fields of biology, dramatically increased the knowledge about cell-type heterogeneity within organs, and enabled the discovery of new or rare cell types. In the oncology research field, the number of publications exploiting single-cell methods is growing exponentially. These are precious techniques to deepen the understanding of cancer cell heterogeneity and plasticity, but also the

characterization of tumor microenvironment. The potential applications of single-cell technologies are numerous<sup>192</sup> and include the dissection of the molecular mechanisms driving cancer progression or treatment resistance (**Fig. 14**). Single-cell technologies may also serve to discover new vulnerability pathways and druggable targets.

Single-cell technologies are also powerful tools in clinical cancer research, for instance, for the non-invasive early detection of circulating tumor cells (CTCs) and to refine clinical diagnoses or ameliorate risk stratifications by characterizing interpatient heterogeneity<sup>193</sup>. These methods are also increasingly integrated into clinical trials. Indeed, single-cell technologies may better help stratify patients according to single-cell biomarkers, monitor treatment response, or improve the understanding of the biological mechanisms underlying the antitumor effect of investigational compounds<sup>194</sup>.



*Fig. 14. Clinical applications of single-cell genomics in cancer research.*

*From Lim et al., 2020<sup>193</sup>.*

## 2.3 SARCOMA CELLS' HETEROGENEITY AND PLASTICITY

### 2.3.1 Sarcoma cells' clonal heterogeneity

While cancer cells' genetic evolution has been widely studied in alternate cancer types like carcinomas, the characterization of sarcoma cells' heterogeneity remains fragmentary. Notably,

the description of sarcoma cells clonal dynamics along the course of tumor progression has been reported in undifferentiated pleomorphic sarcoma (UPS) <sup>195</sup>. In this study, authors relied on a UPS mouse model, multicolor reporters, and CRISPR-Cas9 barcoding to show that distinct clonal evolutions drive different stages of tumor progression. Indeed, primary tumor growth was associated with reduced clonal heterogeneity, whereas increasing tumor clonality conducted local recurrence. Interestingly, metastasizing was driven by a clonal selection of subpopulations in which some essential genes were downregulated, such as *Rasd1*, *Reck*, and *Aldh1a2*.

### 2.3.2 Sarcoma heterogeneity: advances from scRNA-seq approaches

Cell plasticity is particularly interesting for sarcoma, considering their usually low mutation burden and, thus, supposed low genetic heterogeneity. Moreover, the sarcoma microenvironment is a growing subject of interest, whose immunosuppressive features may cause the general limitation of immunotherapy efficacy in sarcoma.

To date, scRNA-seq has been exploited in sarcoma to tackle three main topics: (i) the characterization of sarcoma cells' heterogeneity and protumorigenic states, (ii) the exploration of sarcoma immune and stroma microenvironment, (iii) the unraveling of sarcoma cellular origin. The major published works addressing sarcomas' biology based on scRNA-seq experiments are summarized in **Table 6**.

#### 2.3.2.1 Ewing sarcoma

Ewing sarcoma served as a prototype to characterize cancer cell plasticity in genomically simple fusion-driven sarcoma subtypes. Ewing sarcomas are malignant bone and soft-tissue tumors mainly occurring in young male adults <sup>196</sup>. Like DSRCT, they are characterized by the co-expression of mesenchymal and neuronal features, raising the question of their cell of origin. Ewing sarcoma is also driven by a chimeric aberrant transcription factor, arising from a chromosomal translocation between a FET family member and an ETS family transcription factor, EWSR1-FLI1 fusion accounting for more than 90% of cases. EWSR1-FLI1 both triggers an aberrant opening/activation of heterochromatin regions by the creation of *de novo* enhancers at GGAA microsatellites <sup>197</sup> by interacting with chromatin remodeling complexes such as BAF <sup>198</sup> and represses several gene regions through incompletely understood mechanisms <sup>199</sup>.

Interestingly, EWSR1-FLI1 activity level has been shown to play a central role in Ewing sarcoma cells' heterogeneity. A lower EWSR1-FLI1 activity has been associated with an increase in Ewing cells' mesenchymal features <sup>200</sup>, bestowing enhanced invasive and

metastasizing properties <sup>201,200</sup>. On the contrary, EWSR1-FLI1 high Ewing cells were associated with marked proliferative ability.

The advent of single-cell technology subsequently shed light on the mechanisms underlying Ewing sarcoma cell plasticity. The first study that exploited scRNA-seq to explore Ewing cells' heterogeneity was published in 2020 by Aynaud et al. <sup>202</sup> and focused on exploring EWSR1-FLI1 expression level variability within Ewing cell lines using a time-resolved mapping of EWSR1-FLI1 binding sites after EWSR1-FLI1 shRNA-mediated silencing followed by progressive reexpression. By performing an independent component analysis (ICA) to deconvolute the heterogeneity of Ewing sarcoma cells transcriptional programs, authors characterized several signatures, including cell-cycle related and EWSR1-FLI1 specific signature, which was highly enriched in EWSR1-FLI1 direct targets. Further, using PDX models to integrate previously identified independent components, they identified a coupling between oxidative phosphorylation, proliferation, mRNA splicing, and EWSR1-FLI1 activity, whereas hypoxia and glycolysis appeared as relatively independent mechanisms according to Spearman correlations. Altogether, these data helped to illuminate several mechanisms underlying Ewing cells' heterogeneity, where EWSR1-FLI1 activity appears to play a central role. Additional works - which did not necessarily rely on scRNA-seq experiments - highlighted additional mechanisms underlying Ewing cell plasticity. For instance, focusing on Ewing sarcoma cell plasticity, Keskin et al. sought to identify Ewing sarcoma protumorigenic subpopulations within cells from 3D cell culture <sup>205</sup>. The authors relied on tumor- or PDX-derived primary 3D cell culture *in vitro* models to better mimic *in vivo* tumor conformation and retain tumor cell heterogeneity and plasticity. In these experiments, authors designed a functional microRNA-based live-cell reporter assay from 3D-grown Ewing tumor cells. Based on the previously described inverse relationship between miR-145 expression and Ewing cells pluripotency <sup>206</sup>, they showed that low miR-145 Ewing cells expressed higher levels of SOX2 and did display higher tumorigenicity features. Transcriptome comparison between low versus high miR-145 Ewing cells revealed a gene expression signature of tumor cell aggressiveness, among which *EPHB2*, a transmembrane tyrosine kinase receptor, whose expression within a cohort a 129 patients-derived microarray samples was inversely correlated to patients survival. Additional studies that took advantage of the scRNA-seq technology focused on interrogating the heterogeneity and developmental origin of Ewing sarcoma cells <sup>203</sup>, and another study aimed to identify the diverse cellular responses to EWSR1-FLI1 downregulation in Ewing sarcoma cells <sup>204</sup>.

#### 2.3.2.2 Synovial sarcoma

Jerby Arnon et al. recently took advantage of 10X scRNA-seq, integrated with GeoMX spatial profiling and multiplex IHC to interrogate the interplay between synovial sarcoma (SS) cells and their microenvironment <sup>207</sup>. Starting from the observation that SS is characterized by an exceptionally low T cell infiltration, authors hypothesized that specific SS tumor cell states may prompt immune-evasion mechanisms, despite the known expression of various cancer-testis antigens (e.g., *NY-ESO-1*, *PRAME*, *MAGEA4*, and *MAGEA1*) <sup>208</sup>. Analyzing 12 SS tumors with a total of 16872 cells, authors identified three coregulated gene modules that were consistent across multiple tumors. These three gene modules corresponded to mesenchymal and epithelial cell states and a novel malignant cell state driven by a core oncogenic program controlled by SS18-SSX expression. The latter program, primarily expressed in cycling cells, included the activation of genes from respiratory carbon metabolism and the repression of genes involved in TNF signaling, apoptosis, p53 signaling, and hypoxia. Noteworthy, the expression of this program was associated with the onset of local or distant recurrences. Focusing on the analysis of SS infiltrating immune cells, authors showed that SS CD8<sup>+</sup> T cells overexpressed effector and cytotoxic programs, while macrophages expressed high levels of TNF compared to melanoma CD8<sup>+</sup> T cells. Further exploring the interactions between SS malignant cells and immune evasion, authors showed an inverse spatial correlation between the expression of the core oncogenic program and CD45<sup>+</sup> immune cells infiltration, supporting that this malignant cell state may promote cell exclusion. Strikingly, this core oncogenic program was repressed by microenvironment signaling emanating from surrounding macrophages and T cells including TNF and IFN-gamma, and could be targeted by the combination of HDAC and CDK4/6 inhibitors, hence increasing SS cells immunogenicity and T-cell mediated killing.

#### 2.3.2.3 MPNST

Malignant peripheral nerve sheath tumors (MPNSTs) represent an aggressive subtype of STS that may develop from benign neurofibromas (NFs) and whose incidence is particularly high in patients with type 1 neurofibromatosis (NF1). MPNST and NF cell of origin remains uncertain and is thought to derive from Schwann cells (SCs) or neural-crest-derived pluripotent cells. However, the mechanisms underlying the dynamic evolution of MPNST-initiating cell states during nerve sheath tumorigenesis remain insufficiently explored.

In a recently published article, Wu et al. relied on single-cell multiomics to interrogate the putative MPNST cell of origin and identify the regulators of MPNST oncogenesis by analyzing both the malignant and non-malignant subpopulations present in NF and MPNST samples<sup>209</sup>. Authors performed scRNA-seq profiling of human MPNST, NF, and matching animal models. Using an MPNST mouse model induced by YAP/TAZ signaling activation and *Lats1/2* deficiency, several NF and MPNST subpopulations were characterized. The latter included SC-derived cells, cycling cells, stromal cells (i.e., fibroblasts, endothelial cells, vascular smooth muscle cells), and immune cells (i.e., macrophages, T cells). Notably, the phenotypic heterogeneity of SC-derived cells, fibroblasts, and macrophages significantly increased in malignant tumors compared to benign lesions. For instance, malignant tumors were enriched in activated fibroblasts and protumorigenic anti-inflammatory macrophages. Focusing on SC-derived cells, malignant tumors displayed two specific SC-derived populations: a nestin+ mesenchymal SC precursor-like population (MES-like cells) and a previously undescribed nestin-mesenchymal neural-crest-like population (MES-NC-like cells).

Contrary to previous reports, authors showed that nestin was expressed in various SC-related subpopulations, suggesting that SC lineage cells undergo reprogramming into different tumor cell states and acquire a novel MES-NC-like phenotype during cell transformation. SC phenotypic plasticity was further characterized in murine MPNST models, where 18 metagene activity programs were identified using nonnegative matrix factorization (NMF), which correlated to identified subpopulations. Surprisingly, using a trajectory inference tool, authors showed a reversal of the SC development trajectory, where the nestin- MES-NC-like population was identified as the initiating cell state. Importantly, these findings were further corroborated by the scRNA-seq analysis of human NF1-mutated plexiform NFs and NF1-associated MPNSTs where MES-NC-like and MES-like cells were uniquely found in MPNSTs, and SC lineage-derived cell populations were characterized by a continuum of cell states along the NF-to-MPNST trajectory. Significantly, interrogating the prognostic value of previously defined tumor cell states, the MES-NC-like signature was associated with disease severity and MPNST clinical grades. In addition to scRNA-seq profiling, scATAC-seq was performed on human NF and MPNST tissues and robustly identified three distinct SC-derived epigenomic states comprising non-myelinating SCs and neoplastic SCs in NF, and malignant MES-NC-like cells in MPNST. Finally, interrogating potentially targetable pathways, authors used a Drug Gene Interaction Database and uncovered vulnerabilities associated with the MES-NC-like phenotype, including tyrosine kinase receptors (e.g., IGF1R, EPHA3), *NRG1*, transcriptional regulators (e.g., ZEB1, MEIS2, TEAD1), and protein phosphatases (e.g., EYA2 and PTPRG).

Overall, this exhaustive study on NF to MPNST malignant transformation unraveled the mechanisms underlying NF to MPNST tumorigenesis and unveiled a malignant-specific aggressive cell state that could be pharmacologically targeted.

#### 2.3.2.4 Osteosarcoma

Sc-RNA-seq profiling also helped to describe osteosarcoma (OS) heterogeneity. A study exploiting scRNA-seq characterized 11 osteosarcoma samples from primary and metastatic sites, providing clues on OS malignant and non-malignant subpopulations<sup>210</sup>. Notably, cells' clustering enabled the identification of 11 cell clusters, including several malignant cell subpopulations (e.g., osteoblastic OS cells, proliferating osteoblastic OS cells, chondroblastic OS cells) and microenvironment cells (e.g., osteoclastic cells (OCs), Tumor Infiltrating Lymphocytes (TILs), Tumor-Associated Macrophages (TAMs), fibroblasts, Mesenchymal Stem Cells (MSCs), endothelial cells). Notably, the subclustering of malignant OS cells highlighted that osteoblastic OS cells from primary tumor sites were distinct from those found in metastases, overexpressing *S100A11*, *S100A10*, *PRDX2*, and *PSMD4*. Interestingly, malignant OS chondroblastic cells subclustering highlighted a subpopulation of trans-differentiating cells, with high *RUNX2*, *SPPI1*, and *COL1A1* levels and relatively low levels of *COL2A1* and *SOX9*, suggesting a continuum between chondroblastic cells and osteoblastic cells. This observation was further corroborated by trajectory inference, which showed a bifurcation trajectory driven by genes related to osteoblastic features as well as genes related to epigenetic modifications such as histone methylation and acetylation, suggesting that epigenetic changes may contribute to this transdifferentiation.

This study also shed light on the OS microenvironment. Of note, the subclustering of OCs allowed to identify three distinct clusters (i.e., progenitors, mature OCs, and immature OCs) and showed that the proportion of mature OCs was lower in the chondroblastic, lung metastatic, and recurrent lesions compared to primary OS lesions, indicating that the OC status may depend on extrinsic tumor site-dependent signals. The study of OS immune cells revealed an immunosuppressive microenvironment. Myeloid cells comprised 70-80% of monocytes/macrophages, among which TAMs were divided into three subclusters, dominated by M2 macrophages. Concerning TILs, eight TIL subclusters were identified, including CD4-CD8-, CD8+, and CD4+ T cells, T regs, proliferating T cells, Natural Killer (NK) T cells, and B cells. Notably, CD8+ T cells were characterized by the expression of T cell exhaustion inhibitory receptors, while T regs had high *TIGIT* expression levels.

Importantly, the immunosuppressive microenvironment of OS was confirmed in another study relying on OS samples from chemotherapy-naïve patients <sup>211</sup>. The latter highlighted OS poor T cell infiltration and identified C3\_TXNIP+ and C5\_IFIT1+ macrophages that may participate in CD8+ T cell exhaustion. This work also showed the association between C1\_osteoblastic OS cells - enriched in osteolysis and angiogenesis pathways - and overall survival in a cohort of 85 OS patients.

#### 2.3.2.5 Other

Additional scRNA-seq studies focused on mouse sarcoma models <sup>212,213</sup> to describe the sarcoma microenvironment landscape. For instance, Tessaro et al. sought to characterize the interplay between sarcoma cells and macrophages through the expression of tumor-expressed macrophage-migration inhibitory factor (MIF), which interacts with the surrounding macrophages through the MIF-CD74 axis, hence promoting sarcoma growth <sup>212</sup>. Specifically, the authors identified distinct myeloid cell clusters, representing mouse undifferentiated sarcoma's main infiltrating immune component. Inhibiting MIF in the tumor cells enabled the reshaping of the myeloid microenvironment by promoting anti-tumorigenic inflammatory factors' expression.

In addition, Caldwell et al. recently performed single-nucleus RNA-seq, T cell receptor (snRNA/TCR)-seq, and pool-matched whole-genome sequencing (WGS) in six frozen undifferentiated pleomorphic sarcomas (UPSs) and intimal sarcomas, including samples collected before and after immune checkpoint inhibitors therapy <sup>214</sup>. Notably, the authors highlighted the resurgence of rare cancer cell subclones at immunotherapy resistance in a patient treated for intimal sarcoma. Moreover, the identification of an appropriate T cell clonal expansion in a UPS patient showing primary immunotherapy resistance raised the question of the role of additional immunosuppression mechanisms.

Table 6. Studies exploiting scRNA-seq technology to explore sarcoma biology.

First author	Year	Sarcoma subtype	Tissue type	scRNA-seq method	Main results	Functional validation	Reference
Aynaud et al.	2020	Ewing	Cell line model, engineered A673/TR/shEF model and PDX	Fluidigm C1 Single cell auto prep System	Definition of an exclusive and specific enhancer-driven EWSR1-FLI1 program. Cell proliferation and oxidative phosphorylation metabolism are strongly associated with EWSR1-FLI1 activity. EWSR1-FLI1 intermediate and low activity is associated with hypoxia.	None	202
Hong et al.	2022	Ewing	Tumors (n=4)	10X Genomics Chromium	Ewing sarcoma arises from neural crest cell lineage. Variable timing of oncogenic conversion leads to distinct differentiation trajectories. Interferon-signaling pathway plays a role in tumorigenesis and tumor progression.	None	203
Khoogar et al.	2022	Ewing	Cell lines, functional genetics (siRNA) and PDXs	10X Genomics Chromium	Three distinct transcriptional states were found when comparing proliferating cells to siRNA EWSR1-FLI1 silenced cells, which contributed to ES heterogeneity. In PDXs, around 1% of cells were in a dormant-like state, and 2-4% were in a stem-like and neural stem-like in the proliferating cells.	Primary and secondary spheroids formation, assessed for invasion/motility and autophagy (expression profiling cell sorting and IHC)	204
Wu et al.	2022	MPNST/NF	Tumor (n=4 MPNSTs, n=10 NFs) Mouse models	Drop-seq	Broad range of Nestin-positive Schwann cell subpopulations linked to sarcomagenesis dynamics. Detection of a Nestin-negative mesenchymal neural crest-like population, which correlates with clinical severity and is characterized by a unique regulatory network.	Genetic perturbation with shRNA against ZEB1 and ALDH1 identified as central in the regulatory network	209
Liu et al.	2021	Osteosarcoma	Tumor (N=6)	10X Genomics Chromium	Nine major cell types were identified. Osteoblastic cells were subdivided into five subpopulations, with key transcriptional programs associated with prognosis. One of the osteoblast populations was shown to promote osteolysis and angiogenesis. Osteosarcoma are T cell depleted.	Multiplex IF; Macrophages culture with medium +/- conditioned by OS cells; ELISA detection of RANKL and VEGFA on medium +/-conditioned by OS cells.	211
Zhou et al.	2020	Osteosarcoma	Tumor (N=11)	10X Genomics Chromium	The authors identified six osteoblastic malignant cell subpopulations.	Cytotoxic activity of peripheral blood-derived CD3+ T cells on	210

					Two were proliferative, four were involved in angiogenesis. Lower osteoclast infiltration was observed in the chondroblastic subtype, as well as in OS lung metastases. Pro-inflammatory macrophages (FABP4+) were specifically found in OS lung metastases.	U2OS/143B cells after TIGIT blockade	
Tessaro et al.	2022	Mouse sarcoma	CDXs	10X Genomics Chromium multiplexed with Hashtag	Analysis of the immune tumor microenvironment. Macrophages exhibit different activation states within tumors. Sarcoma cells use macrophage inhibitory factor (MIF) to interact with CD74 macrophages. As a result, macrophages favor the protumorigenic potential of sarcoma cells.	MIF silencing	212
Caldwell et al.	2022	Intimal sarcoma, UPS	Frozen and FFPE tumor tissue	snRNA/TCR-seq	In a patient with complete response to immune checkpoint inhibitor but a later recurrence, rare populations of cells from a clone with copy number alterations emerged during resistance. Pre-existing populations of resistant clones with copy number alteration may drive ICI resistance.	None	214

### 3 RATIONALE AND AIMS OF THE PRESENTED WORK

---

As previously developed, tumor heterogeneity and plasticity are a growing area of interest, as illustrated by the recent addition of phenotypic plasticity as a new dimension of hallmarks of cancer <sup>121</sup>. Cellular plasticity in cancer is a significant challenge and can be considered an Achilles heel to target. Thus, it is essential to understand the mechanisms underlying this heterogeneity and its role in tumor progression. For instance, targeting the most aggressive tumor cell subpopulations may represent a promising opportunity to increase the effectiveness of anti-tumor treatments. Moreover, strategies aiming at confining tumor cells within a specific phenotypic state may hinder the essential interdependency between tumor cell states that promote self-renewing, proliferation, migration, and therapeutic resistance.

Catching the wave of scRNA-seq technology advent, which provides the unprecedented ability to capture and characterize cells' heterogeneity and phenotypical states, and inspired by published works focusing on the prototypical Ewing sarcoma biology, we focused on the exploration of DSRCT heterogeneity at the cell and molecular level and dissected the molecular mechanisms underlying such diversity.

To achieve this goal, we formulated the following hypotheses concerning DSRCT cell biology and heterogeneity, which dictated the elected methods further described within this manuscript to test these hypotheses.

1. **Hypothesis 1.** DSRCT cells display phenotypical heterogeneity and cell state plasticity despite the presence of the sole oncogenic driver EWSR1-WT1;
2. **Hypothesis 2.** DSRCT cell diversity may be uncovered by scRNA-seq experiments;
3. **Hypothesis 3.** DSRCT cell heterogeneity is partly driven by intrinsic mechanisms that may arise from EWSR1-WT1 expression level variability, EWSR1-WT1 fluctuant transcription factor activity, the versatility of downstream activated transcriptional pathways, and stochastic epigenetic reprogramming;
4. **Hypothesis 4.** DSRCT cell plasticity may be induced by extrinsic factors such as tumor site-specific ecosystem and microenvironment cells' signaling.



## Title

# Single-cell multi-omics profiling reveals heterogeneous transcriptional programs and microenvironment in Desmoplastic Small Round Cell Tumors

## Authors

Clémence Henon<sup>1,2,3</sup>, Julien Vibert<sup>4,5,6</sup>, Thomas Eychenne<sup>1</sup>, Nadège Gruel<sup>4,6</sup>, Leo Colmet-Daage<sup>1</sup>, Carine Ngo<sup>1,7</sup>, Marlène Garrido<sup>1</sup>, Nicolas Dorvault<sup>1</sup>, Maria Eugenia Marques Da Costa<sup>8,9</sup>, Virginie Marty<sup>10</sup>, Nicolas Signolle<sup>10</sup>, Antonin Marchais<sup>8,9</sup>, Asuka Kawai-Kawachi<sup>1</sup>, Madison Lenormand<sup>1</sup>, Clémence Astier<sup>1</sup>, Roman Chabanon<sup>1</sup>, Benjamin Verret<sup>2,11</sup>, Rastislav Bahleda<sup>3</sup>, Axel Le Cesne<sup>2,12</sup>, Fatima Mechta-Grigoriou<sup>13</sup>, Matthieu Faron<sup>14</sup>, Charles Honoré<sup>14</sup>, Olivier Delattre<sup>4</sup>, Joshua J Waterfall<sup>5,6</sup>, Sarah Watson<sup>4,6</sup>, and Sophie Postel-Vinay<sup>\*1,3</sup>

## Affiliations

<sup>1</sup>ATIP-Avenir group, INSERM U981, Gustave Roussy, Paris Saclay University, Villejuif, France

<sup>2</sup>Department of Medical Oncology, Gustave Roussy, Villejuif, France

<sup>3</sup>Drug Development Department, DITEP, Gustave Roussy, Villejuif, France

<sup>4</sup>INSERM U830, Équipe labellisée LNCC, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, SIREDO Oncology Center, Institut Curie Research Center, Paris, France

<sup>5</sup>INSERM U830, Integrative Functional Genomics of Cancer Lab, PSL Research University, Institut Curie Research Center, Paris, France

<sup>6</sup>Department of Translational Research, PSL Research University, Institut Curie Research Center, Paris, France.

<sup>7</sup>Department of Pathology, Gustave Roussy, Villejuif, France

<sup>8</sup>INSERM U1015, Gustave Roussy, Paris Saclay University, Villejuif, France

<sup>9</sup>Department of Pediatric and Adolescent Oncology, Gustave Roussy, Villejuif, France

<sup>10</sup>Experimental and Translational Pathology Platform (PETRA), AMMICa, INSERM US23/UAR3655, Gustave Roussy, Villejuif, France.

<sup>11</sup>Breast cancer translational research group, INSERM U981, Gustave Roussy, Villejuif, France

<sup>12</sup>International Department of Medical Oncology, Gustave Roussy, Villejuif, France

<sup>13</sup>INSERM U830, Equipe labellisée LNCC, Stress et Cancer, PSL Research University, Institut Curie Research Center, Paris, France

<sup>14</sup>Surgery Department, Gustave Roussy, Villejuif, France

**Wordcount (without Legends and References): 17,045 words**

**6 Figures**

**0 Tables**

**17 Supplementary Figures**

**30 Supplementary Tables**

*For confidentiality reasons, Supplementary Tables are not included in the online version of this manuscript.*

**123 References**

## Summary

**144/150 words**

Desmoplastic Small Round Cell Tumors (DSRCTs) are a rare small round cell sarcoma subtype of poor prognosis driven by the aberrant chimeric transcription factor EWSR1-WT1. Despite this unique oncogenic driver and the paucity of recurrent secondary mutations, DSRCTs harbor a characteristic polyphenotypic differentiation, suggesting the presence of tumor heterogeneity. Using single-cell multiomics, we found that DSRCT tumor cells were characterized by recurrent overlapping phenotypical states, which harbored distinct transcriptional programs, notably related to mesenchymal, neural, and epithelial cell fates, and metabolic switch. Our data further suggest the role of EWSR1-WT1 DNA-binding versatility in DSRCT cells' chromatin plasticity and highlight additional *in vivo* extrinsic signals from microenvironment components like extracellular matrix and cancer-associated fibroblasts which may influence DSRCT cells states and tumor growth. We further explored the clinical applicability of our findings and identified single-cell RNA-sequencing-derived transcriptional signatures that may serve for prognostic risk stratification.

**Keywords (up to 10)**

DSRCT; desmoplastic small round cell tumor: sarcoma: EWSR1-WT1; heterogeneity; single cell; microenvironment

## Introduction

Desmoplastic Small Round Cell Tumor (DSRCT) is a rare Soft Tissue Sarcoma (STS) subtype that is related to the recently described small round cell sarcomas (SRCSs) entity <sup>1</sup>. DSRCT most often primarily arises in the abdominal or pelvic cavity (70-90%), where it rapidly spreads in the form of multiple nodules and preferentially metastasizes in the liver (35-50%), lung, bones, lymph nodes, or testis. DSRCT preferentially affects young male adults, with a 5:1 sex ratio and a median age of around 27 years old <sup>2</sup>. Despite aggressive multimodal therapeutic approaches involving extensive resection surgery, polychemotherapy, and abdominal radiation, DSRCT is a devastating malignancy whose prognosis remains very dismal, with a median overall survival of approximately two years, strengthening the urgent need for developing patient-customized targeted approaches <sup>2-4</sup>.

At the molecular level, DSRCTs are characterized by the pathognomonic translocation t(11;22) (p13;q12), which fuses the *EWSR1* N-terminal part to the *WT1* C-terminal domain. The resulting chimeric protein EWSR1-WT1 acts as an aberrant transcription factor (TF), activating various oncogenic pathways, such as cell proliferation, survival, and migration <sup>5-7</sup>. EWSR1-WT1 represents the primary and unique driver of this disease, which presents a low mutational burden <sup>8-13</sup> with few recurrent secondary alterations (e.g., *AR*, *FGFR4*, *ARID1A* or *TP53* <sup>9-11,13</sup>) but several recurrent copy number variations <sup>12</sup>.

Histologically, DSRCT presents as nests of small round blue tumor cells surrounded by a dense and abundant immune-cold desmoplastic stroma. Interestingly, despite a very homogeneous histological aspect of tumor cells, the latter exhibits a polyphenotypic differentiation, with positive immunohistochemical staining for epithelial, neural, and mesenchymal markers. This observation suggests that DSRCT, whose cell of origin remains unknown to date, may originate from a pluripotent precursor with multilineage potential or that EWSR1-WT1 may drive various multilineage transcriptional programs that would be activated stochastically or in response to external stimuli from the microenvironment. Recent findings in Ewing sarcoma, the prototypic fusion-driven SRCS characterized by EWSR1-FLI1 fusion in 85% of cases, revealed that variations in EWSR1-FLI1 activity influenced metabolic, proliferative and migratory states of Ewing tumor cells <sup>14,15</sup>. Such heterogeneity has

not yet been explored in DSRCT, and its understanding may allow to customize patients' treatment and thereby ultimately improve their outcomes.

To gain insight into tumor cells and microenvironment heterogeneity in DSRCT, we took advantage of multiple patient samples' single-cell multiomic, digital spatial, and bulk molecular profiling, combined with preclinical modeling of EWSR1-WT1 activity. We focused on the understanding of (i) tumor cell plasticity, (ii) EWSR1-WT1-induced transcriptome rewiring, and (iii) the interplay between DSRCT tumor cells and tumor microenvironment. We found that DSRCT tumor cells are characterized by recurrent overlapping phenotypical states, which harbor distinct transcriptional programs, notably related to developmental cell fate and metabolic switch. Our data further suggest the role of EWSR1-WT1 DNA-binding domains versatility in DSRCT cells' chromatin plasticity and highlight additional *in vivo* extrinsic signals and cell-to-cell interactions, likely to influence DSRCT cell state transitions and tumor growth. We further explored the clinical applicability of our findings and identified single-cell RNA-sequencing (scRNA-seq) derived transcriptional signatures that may serve for prognostic risk stratification.

## **Material and Methods**

### **Human subjects and Ethical Considerations**

This study was performed per the European General Data Protection Regulation (GPDR) following Regulation (EU) 2016/679 of the European Parliament and the Council of April 27, 2016.

Patients with DSRCT treated at Gustave Roussy and Curie Institute alive at the time of analysis gave their preoperative informed consent to allow tumor residuals usage for scientific purposes. Clinical characteristics of patients involved in this study are summarized in **Supplemental information - Table S1**.

### **Cell line models**

JN-DSRCT-1 cell line <sup>16</sup> is a kind gift from Professor Janet Shipley (Institute of Cancer Research, London).

R is a primary cell line derived from a DSRCT PDX purchased from Champions Oncology Inc.

JN-DSRCT-1 and R cell lines were maintained in 2D adherent culture within DMEM-F12 (Gibco) supplemented with 10% FBS, 1% Penicillin-Streptomycin (Gibco), 1% Sodium Pyruvate (Gibco), 1% Sodium Bicarbonate (Gibco), 1% Non-Essential Amino Acids (Gibco) and 1% HEPES (Gibco). Cell passage was performed at 1/10 twice a week. Used cells were controlled for mycoplasma-free status.

### **Spheroid formation**

JN-DSRCT-1 cells were grown as spheroids in 96w ultra-low attachment cell culture plates in supplemented DMEM/F-12 medium. DSRCT cells were first seeded at 1,000 cells per well before 15 min 500 g centrifugation. The formation of the spheroid was assessed at H24 after cells' seeding.

### **siRNA Knock-Down (KD)**

JN-DSRCT-1 cell line was transfected with a custom siRNA targeting EWSR1-WT1 (3' GAT CTT GAT CTA GGT GAG A 5'), CCND1 (Horizon Discovery ON-TARGETplus Human CCND1 siRNA-smart pool, reference L-003210-00-0005) or non-targeting siRNA (Horizon Discovery ON-TARGETplus Non-targeting siRNA #1, reference D-001810-01-05), according to manufacturer's instructions. After cells' seeding and obtention of 50% confluency, transfection was performed using Lipofectamine™

RNAimax (Invitrogen™ reference 13778150), and the medium was replaced the day after. A 48-hour silencing timepoint was used for each described experiment.

### **DSRCT patients' derived tumors 3' single-cell RNA sequencing (3'scRNA-seq)**

#### ***Tumor dissociation and scRNA-seq with 10X Genomics Chromium assay***

Fresh tumor material from patients treated for DSRCT was profiled using the 3'-end counting 10X Genomics Chromium assay.

Briefly, fresh DSRCT tumor and juxta-tumor material was collected into RPMI 1640 medium with GlutaMAX™ Supplement (Gibco™, reference 61870010) or MACS Tissue Storage Solution (Miltenyi Biotec, reference 130-100-008) with a delay <30 min after surgery. Tissue samples were further cut into small chunks and incubated at 37°C in RPMI medium containing DNase I (Sigma, reference DN25-100MG) at 2.2 µg/mL (final concentration) and Liberase™ Thermolysin Low (TL) (Roche, reference 5401020001 at 4.5 µg/mL (final concentration) during 20 to 30 min. After tissue dissociation, the mixture was washed in 1x PBS + 0.04% BSA before centrifugation. Resuspended cells in 1x PBS + 0.04% BSA were filtered twice with a 70 µm and 30 µm cell strainer to obtain a single cell suspension. After another centrifugation, cells were resuspended in 1x PBS + 0.04% BSA. When necessary, a red blood cell lysis was performed with a 2- to 3-min incubation in 1X Red Blood Cell Lysis buffer (Biolegend, reference 420301) protected from light before washing in 10 mL PBS and performing a last centrifugation.

Cell concentration and cell viability were controlled on an automated cell counter with trypan blue and were loaded into the 10X Genomics cassette for a targeted cell recovery of 5,000 cells per sample following the recommendations from the protocol. Cell encapsulation, reverse transcription, and library generation were performed using 10X Genomics standard protocols <sup>17</sup>.

Paired-end sequencing was performed on an Illumina NovaSeq sequencer for a targeted depth of 400 million reads per sample.

#### ***Data analysis***

After quality control, we performed read mapping and expression quantification followed by regularized negative binomial regression-based normalization using SCTransform <sup>18</sup>. ScRNA-seq data were analyzed using the Seurat R package (R v3.5.1). The clustering of cell subpopulations was done using a k-nearest neighbors (k-NN) graph method based on the Euclidean distance on PCA, followed by Louvain algorithm

optimization. Visualization and exploration of the data were performed with the Uniform Manifold Approximation and Projection (UMAP) non-linear dimensional reduction technique (Seurat).

Samples from synchronous distinct localizations were combined by merging the raw count matrices of individual Seurat objects. The integration of samples stemming from different patients or timepoints was performed using Harmony <sup>19</sup> algorithm (<https://github.com/immunogenomics/harmony>). Of note, an additional 3'scRNA-seq assay was performed secondarily on a matched sample ("PZB" sample) exploited for concomitant single-nucleus gene expression and ATAC-seq assay using 10X Genomics Chromium Single Cell Multiome ATAC + Gene Expression technology (snMultiome) and was not included in the Harmony-integrated "*Int\_sc*" dataset.

Differential gene expression analysis was performed using Mann-Whitney Wilcoxon test. We identified the most differentially expressed genes based on p-value and log2FC and applied Gene Ontology (GO) terms to characterize each subpopulation. We performed gene set enrichment analysis of gene ontology (GseGO) on the most differentially expressed genes (p-value <5%) using R package gprofiler2.

### ***DSRCT neotranscripts expression***

To quantify single-cell expression of DSRCT-specific neotranscripts, we ran CellRanger "count" using a custom index built by appending sequences of the neotranscripts to the reference transcriptome. Counts for neotranscripts were log-normalized, and the average log-normalized expression level was plotted with FeaturePlot.

### ***Hotspot***

The Hotspot tool <sup>20</sup> (<http://www.github.com/yoseflab/Hotspot>) was used to identify informative gene modules across clusters from the "*Int\_sc*" dataset.

Hotspot is an algorithm that computes gene modules in three steps: finding informative genes with high local autocorrelation, evaluating the pairwise correlation between these genes, and clustering the results in a gene-gene affinity matrix. The Hotspot depth-adjusted negative binomial model was run using the count matrix and the 50 first principal components. A k-NN graph was then calculated using 30 neighbors, and the 500 genes with the highest significant autocorrelation (false discovery rate < 0.05) were selected. Pairwise local correlation between these genes was computed, and gene modules were created by agglomerative clustering with a minimum number of genes per module set to 15 and a false discovery rate threshold of 0.05. Hotspot module

scores for each cell were calculated by first centering the UMIs using the depth-adjusted negative binomial model. The centered values were then smoothed using the weighted average of their 30 nearest neighbors. These smoothed values were then modeled with PCA using the first principal component, and the cell loadings were reported as the module scores.

### ***CellPhoneDB***

To infer cell-cell communications between identified tumor and microenvironment cells' clusters, we took advantage of CellPhoneDB<sup>21</sup> tool (<https://github.com/Teichlab/cellphonedb>), developed to investigate cells' crosstalks from a repository of interacting ligands and receptors.

### ***NicheNet***

We applied NicheNet<sup>22</sup> computational method (<https://github.com/saeyslab/nichenetr>) to DSRCT tumor cells, immune cells, and cancer-associated fibroblasts to further interrogate cell-to-cell interactions. NicheNet presents the advantage of computing the activity of ligands-reception interactions by inferring their gene regulatory network at the single-cell level.

### ***InferCNV***

An inference of single cells' copy number variations was performed using InferCNV tool (<https://github.com/broadinstitute/infercnv>). Raw gene expression data were first extracted from each patient-integrated Seurat object. For each sample, non-malignant reference cells were selected based on the expression of immune cells' and cancer-associated fibroblasts' markers (**Supplemental information - Table S3**), and tumor cells were grouped according to annotated Louvain clusters as defined above. The cutoff for the minimum average read count per gene was set to 0.1, as recommended for 10x data. The 'cluster\_by\_groups' setting was used to perform separate clustering for each cluster as defined in the cell annotations file. All other options were set to their default values. Each CNV was annotated as a gain or a loss to a p- or q-arm using GRCh38 reference genome.

### ***CytoTRACE***

We used CytoTRACE algorithm<sup>23</sup> to infer the degree of differentiation of DSRCT tumor cells using CytoTRACE R package version 0.1.0. CytoTRACE (Cellular (Cyto) Trajectory Reconstruction Analysis using gene Counts and Expression) aims at predicting the differentiation state of cells from scRNA-seq data by leveraging the number of detectably expressed genes per cell. Subsampling of 1,000 cells was used

to run CytoTRACE function as described in <https://cytotrace.stanford.edu/>. CytoTRACE results were then visualized on UMAPs colored according to the inferred degree of differentiation.

### **StemID**

The cells' degree of differentiation was predicted using StemID <sup>24</sup>, which measures intracellular entropy based on cells' median transcriptome entropy. StemID tool was used as described in [https://github.com/dgrun/StemID/blob/master/Reference\\_manual\\_RaceID2\\_StemID.pdf](https://github.com/dgrun/StemID/blob/master/Reference_manual_RaceID2_StemID.pdf). Briefly, StemID relies on the concept that the multiplicity of states coexisting within a single cell is reflected by the uniformity of the transcriptome, captured by Shannon's entropy. StemID computes a StemID score, which illustrates the level of cell multipotency.

### **CellRank**

Spliced and unspliced read counts were generated using velocity (v0.17.16) "run10x" function <sup>25</sup>. Spliced and unspliced reads were then filtered and normalized using scVelo (v0.2.3)<sup>26</sup> "filter\_and\_normalize" function [min\_shared\_counts=20, n\_top\_genes=2,000]. A k-NN graph was built using the 30 nearest neighbors on the 30 first principal components. RNA velocity scores were calculated using the scVelo "dynamical" model of transcriptional dynamics. Single-cell velocities were then projected onto UMAP embeddings. Cell-to-cell transition probabilities were calculated using the CellRank velocity kernel <sup>27</sup>. Using scVelo dedicated functions, we could compute predicted initial and terminal states, latent time, lineage probabilities, driver genes, and a directed PAGA (partition-based graph abstraction) model.

### **SCENIC+**

SCENIC+ is a recent development of the SCENIC tool that takes advantage of multiomic data <sup>28</sup>. It predicts genomic enhancers along with candidate upstream transcription factors (TF) and links these enhancers to candidate target genes. Specific TFs for each cell type or cell state are thus predicted based on the concordance of TF binding site accessibility, TF expression, and target gene expression as contained in multiomic data (scRNA-seq and scATAC-seq). We used SCENIC+ (v0.1.dev447+gd4fd733) for implementation.

### **EWSR1-WT1 or full length-WT1 targeted scRNA-seq**

An in-house designed assay was developed to evaluate the expression level of *EWSR1-WT1* transcripts at the single-cell level.

The leftover barcoded cDNA library generated from 10X Genomics protocol was used for selective enrichment in cDNA originating from *EWSR1-WT1* transcript. To this aim, a PCR#1 was done using either (i) an *EWSR1-WT1* breakpoint-specific biotinylated primer or (ii) a full-length *WT1*-specific biotinylated primer, and a universal Read1 specific primer – a sequence added on the transcript during the reverse transcription of the 10X Genomics protocol. The resulting PCR was purified using Streptavidin beads. Then, a PCR#2 was performed using PCR#1 product as a matrix, using a primer specific to *WT1* C-terminal end, located at about 450 bp from the transcript end. The second primer was the same as in PCR#1. The PCR#2 product was purified and used as a matrix for PCR#3 to add the P5 and P7 adapters to construct an Illumina-compatible library.

The downstream bioinformatics analyses of *EWSR1-WT1* targeted scRNA-seq relied on a negative selection method to consider the potential unspecific amplification of full-length *WT1* with the *EWSR1-WT1* targeted scRNA-seq assay.

## **DSRCT patients' derived tumors single-nucleus 3' RNA and ATAC sequencing (snMultiome)**

### ***Tumor dissociation, nuclei isolation, and permeabilization***

Fresh tumor material from one patient treated for DSRCT was profiled using the single-cell Multiome ATAC + Gene Expression 10X Genomics, Inc Chromium assay.

The fresh DSRCT tumor specimens were collected into MACS Tissue Storage Solution kept at 4°C, with a delay of less than 30 min after surgery. The tissue samples were then cut into small chunks before proceeding to nuclei isolation and nuclei permeabilization according to 10X Genomics Demonstrated protocol for Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + GEX Sequencing (CG000375 Rev B). Briefly, the tumor's small chunks were incubated on ice into NP40 lysis buffer (Tris-HCl (pH 7.4) 10 mM, NaCl 10 mM, MgCl<sub>2</sub> 3 mM, Nonidet P40 0.1%, DTT 1 mM, RNase inhibitor 1 U/μl, in Nuclease-free Water) during 5 min. The suspension was then passed through a 70 μm strainer into a 2 mL Eppendorf tube before 500 g centrifugation for 5 min at 4°C. The supernatant was removed, and the nuclei pellet was washed in 1 mL of PBS supplemented with 1% BSA with 1 U/μL RNase inhibitor on ice for 5 min. After an additional 500 g centrifugation at 4°C for 5

min, nuclei were resuspended in 1 mL PBS + 1% BSA + 1 U/mL RNase inhibitor and counted using a Malassez counting chamber on Invitrogen™ EVOS XL imaging system after DAPI staining to confirm complete nuclei isolation from total cells and assess nuclei concentration. We further achieved nuclei permeabilization by performing an additional 5 min 500 g centrifugation at 4°C and incubating the nuclei pellet in 0.1X Lysis Buffer (1X Lysis Buffer: Tris-HCl (pH 7.4) 10 mM, NaCl 10 mM, MgCl<sub>2</sub> 3 mM, Tween-20 0.1%, Nonidet P40 0.1%, Digitonin 0.01%, BSA 1%, DTT 1 mM, RNase inhibitor 1 U/μl in Nuclease-free Water, diluted into Lysis Dilution Buffer: Tris-HCl (pH 7.4) 10 mM, NaCl 10 mM, MgCl<sub>2</sub> 3 mM, BSA 1%, DTT 1 mM, RNase inhibitor 1 U/μL in Nuclease-free Water) on ice during 2 min. After washing the pellet in wash buffer (Tris-HCl (pH 7.4) 10 mM, NaCl 10 mM, MgCl<sub>2</sub> 3 mL, BSA 1%, Tween-20 0.1%, DTT 1 mM, RNase inhibitor 1 U/μL, in Nuclease-free Water), the nuclei were resuspended in the appropriate volume of chilled Diluted Nuclei buffer (Nuclei Buffer (20X), DTT 1 mM, RNase inhibitor 1 U/mL in Nuclease-free Water) and counted using both a Malassez counting chamber and Bio-Rad automated cell counter.

### ***Chromatin transposition, ATAC, and RNA libraries' construction***

Next, the appropriate volume of cell nuclei suspension was extracted to target the effective encapsulation of 5,000 nuclei. The further steps were performed following 10X Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338).

Briefly, the transposition of the native chromatin is performed using 10X Genomics Chromium ATAC Enzyme B, which contains the transposase. Adapter sequences are simultaneously added to the ends of the DNA fragments. Gel Beads-in-emulsion (GEMs) generation, and ATAC and Gene Expression libraries construction have been further performed according to 10X standard protocol.

### ***Sequencing***

The sequencing of the obtained ATAC and RNA libraries was performed on NovaSeq 150 bp with paired-end flow cell for a targeted depth of 400 million reads for the gene expression library and 500 million reads for the ATAC library.

### ***Data analysis***

Demultiplexing, reads alignment, filtering, ATAC peak calling, and generation of feature-barcode matrices were done using Cell Ranger ARC.

We relied on Seurat v4.0.4 and Signac v1.5.0 packages for further downstream analyses. First, genome annotation was performed using hg38 EnsDb.Hsapiens.v86

reference genome. Low-quality nuclei were filtered out before peak calling on either the pseudo-bulk data or separately on each further defined cell cluster using MACS2. The number of counts per peak, features, and UMIs were then calculated before a normalization using the SCTransform method, which allows normalization and variance stabilization of molecular count data from sc and snRNA/ATAC-seq experiments. Nuclei clustering was then performed either based on RNA features, ATAC peaks features, or both using the weighted-nearest neighbors (WNN) method<sup>29</sup>. Single-nuclei data and RNA/ATAC/WNN-derived clusters were visualized on Uniform Manifold Approximation and Projection (UMAP). Further downstream analyses comprised differential gene expression, differential chromatin accessibility, and motif enrichment analyses.

Motif enrichment analyses relied either on the JASPAR 2020 human transcription factors motifs database or EWSR1-WT1 ChIP-defined *de novo* motifs. Two different strategies were used: the first consisted of calculating a motif activity Z-score per cell using the ChromVAR pipeline, followed by identifying the top averaged Z-scores per cluster. The second strategy focused on the top marker peaks inferred from differential accessibility analysis and searched for motifs over-representation within this set of peaks by computing the number of peaks containing the given motifs within each cell cluster and comparing it to the total number of peaks containing the motifs within the pseudo-bulk data using a hypergeometric test.

## **DSRCT patients' derived tumors bulk RNA-sequencing**

### ***Samples collection and RNA-sequencing***

DSRCT frozen specimens collected and archived at Gustave Roussy Biological Resource Center from December 1991 to April 2021 were exploited to constitute the DSRCT test cohort analyzed in this study.

According to applicable law, when alive, patients whose samples were to be used had to provide a non-opposition form to allow the use of their archived biological samples. RNA extraction was performed on tumor samples according to Qiagen AllPrep DNA/RNA kit standard protocol. After RNA Integrity Number (RIN) quality control, RNA samples were used to construct total RNA sequencing libraries. Libraries were sequenced on a NovaSeq sequencer with paired-end 150 bp reads with a targeted depth of 30 million reads per sample.

We also took advantage of an external cohort comprising DSRCT and additional selected soft tissue sarcoma RNA-seq data stemming from Curie UGS.

### ***Data analysis***

The RNA-sequencing data were analyzed using a custom bioinformatics pipeline. Alignment was performed with the STAR suite on hg19 (GENCODE version 19). RNA counts were quantified using the GeneCounts tool after adaptor trimming with Atropos. After filtering out tumor samples outliers and low tumor cellularity samples (<5%) according to the corresponding H&E staining, a hierarchical clustering was performed among GR DSRCT samples (test set) using Pearson's correlation coefficient with Ward D2 linkage algorithm on the total sequenced genes.

Deconvolution of stromal and immune microenvironment as performed using CIBERSORTx<sup>30</sup> (<https://cibersortx.stanford.edu/>) and immune xCell<sup>31</sup> tools.

Differential Gene Expression (DGE) analysis was performed using DESeq2 to compare transcriptome profiles from DSRCT samples to those of a cohort of various STS subtypes (**Supplemental Information – Table S27**).

To deconvolve cluster-specific cell subsets from DSRCT tumors bulk RNA-seq data, CIBERSORTx<sup>30</sup> was used to derive a signature matrix from scRNA-seq data. Using the Seurat subset function, five hundred or 1,000 cells were extracted from each cluster in the Harmony-integrated 3' scRNA-seq dataset. Cluster-labeled cells were used to obtain a single-cell reference matrix (scREF-matrix) that was employed as an input on the CIBERSORTx online server using the “Custom” option. We set Default values for replicates (n=5), sampling (0.5), and Min. Expression (0.0). Additional options for kappa (999), q-value (0.01), and number of barcode genes (300–500) were kept at default values. The imputation of cell subtype fractions defined on the scREF-matrix was performed on TPM values from DSRCT tumors bulk RNA-seq data using the “Impute Cell Fractions” function with the “Custom” option and ran in absolute mode. A total number of 500 permutations was performed to test for statistical significance.

### ***Prognostic signatures generation and signatures' score calculation***

The top 100 upregulated genes within each Harmony integrated samples' clusters (“Int\_sc” clusters) defined by Seurat graph-based clustering using the k-NN method were selected to define clusters' specific signatures. A score for each cluster-specific signature was then calculated using the arithmetic mean of Z-scores of each gene expression expressed in variance stabilizing transformed (VST) raw counts from bulk RNA-seq data of the test cohort (n=29). Secondly, using the optimal cutpoint method,

the DSRCT samples bulk RNA-seq dataset was segregated into High and Low signature scores. We then performed a survival analysis of the patients from whom the tumor samples were collected to compare High versus Low signature scores using the Kaplan-Meier model and tested for significance by log-rank test.

Subsequently, we validated our findings on an external validation cohort comprised of 21 RNA-seq samples from frozen DSRCT tumors (**Supplemental Information – Table S28**).

## **Spatial transcriptomics with 10x Visium**

### ***Sample and library preparation***

The assessment of RNA quality was primarily performed on selected DSRCT FFPE samples. Ten 10  $\mu$ m-sections of FFPE tissue were used for RNA isolation using the RNeasy FFPE kit (Qiagen), according to manufacturer instructions. The samples with DV200 >30% were selected and used for further Visium spatial gene expression sequencing.

The spatial transcriptomic analysis was performed using 10X Genomics Visium technology, according to the Visium Spatial Gene expression for FFPE samples protocol with the human whole transcriptome probe set.

A representative tissue area of 6.5x6.5 mm was previously selected on H&E slides.

The latter slide contains four capture areas, each composed of a ~5,000 circular spots-array containing printed DNA oligos for mRNA capture, constituted by a PCR handle, a unique spatial barcode, a unique molecular identified (UMI), and a poly-dT-VN tail. The resolution of 10x Visium spatial transcriptomics enables to capture the mRNA from 10 to 20 cells in each single 55  $\mu$ m circular spot.

Five- $\mu$ m sections of 3'scRNA-seq matched DSRCT tumors FFPE samples were cut after dehydration and placed on 10x Genomics Visium Gene Expression slides as recommended. Visium slides were incubated in a section dryer oven and kept in a desiccator at RT overnight. Deparaffinization was performed within the next day according to 10X Visium general protocol by placing slides in a section dryer oven at 60°C for 2 H, followed by successive baths in xylene and ethanol gradient concentrations. Next, H&E staining and coverslipping were performed, and Visium slides were further imaged at 10x magnification using Olympus VS120. Images were processed using FiJi ImageJ software.

After coverslip removal, slides were placed into the Visium cassette. Decrosslinking was achieved by incubating Visium slides with TE buffer (Ready-to-use, Genemed, Gentaur) in a thermocycler at 70°C for 60 min. A pre-hybridization step with a 15-min incubation within perm Enzyme B (10X Genomics, PN-3000602/ 3000553) and Tween-20 at RT was followed by the hybridization of the human whole transcriptome probes panel on complementary target RNA transcripts on tissue sections. This panel consists of a pair of specific probes for each targeted gene containing Read2S and polyA sequences. After overnight incubation, several washes with FFPE post-Hyb wash buffer 10X Genomics, PN-2000424) were done, followed by the addition of a ligation enzyme (10X Genomics, PN-2000426/ 2000425) and incubation in a thermocycler at 37°C for 60 min to seal the junction of probes-RNA transcripts pairs. Probes' ligation was followed by several washes with post-ligation wash buffer (10X Genomics, PN-2000420/ 2000419). Finally, RNA digestion was performed to allow the probes' release. The latter step included samples' incubation with RNase enzyme (10X Genomics, PN-3000605/ 3000593) at 37°C for 30 min and permeabilization with Perm Enzyme B for 40 min at 37°C to release the probes, which were then captured on Visium slides surface. Probes' extension was performed by incubating samples for 15 min at 45°C with the Extension enzyme (10X Genomics, PN-2000427) before elution using KOH 0.08M and 1M Tris-HCL to stabilize the reaction. Libraries were right after generated by adding to each sample 50 µl of Amp Mix (10X Genomics, PN-2000047) and 5 µl of dual index Kit/plate Ts Set A well ID (10X Genomics, PN-3000511) and further amplified using the number of cycles determined by qPCR before libraries construction. Finally, samples were cleaned up using SPRIselect reagent (Beckman Coulter) and stored in EB buffer at -20°C until sequencing. Quality assessment and quantification of cDNA libraries were done using BioAnalyzer before sequencing.

Generated Visium libraries were pooled and loaded in a single SP Illumina flow cell. Sequencing was performed on NovaSeq PE 50 at a sequencing depth of 50k read pairs per spot covered with tissue using 10X Genomics recommended run parameters.

### ***Data analysis***

The Space Ranger pipelines were used for demultiplexing Visium-prepared raw base call files generated by Illumina sequencers into FASTQ files using spaceranger mkfastq. Next, spaceranger count was used to perform tissue and fiducial alignment from the microscope slide image and barcode/UMI counting to generate feature-barcode matrices for downstream analyses. The number of spots covered with tissue

was calculated using Loupe Browser. Overall, tissue sections from DSRCT samples encompassed a total of 4,236, 3,452, 3,583, 3,214, 3,537, and 3,686 spots containing included barcodes on the capture area for *GR2*, *GR4*, *GR4\_PC*, *BOM*, *GR7*, and *PZB* samples respectively.

Downstream analyses relied on the Seurat package (v4.1). They included spot clustering based on a k-NN graph method with Louvain algorithm optimization, differential gene expression analysis, and exploration of spatially variable features. We further used an anchor-based label transfer workflow<sup>32</sup> using 3'scRNA-seq matching samples identified clusters as a reference. To account for spatially resolved pathway activations, we calculated the average expression levels of selected HALLMARK gene signatures subtracted by the aggregated expression of control gene sets using Seurat's "AddModuleScore" function.

## **WES**

### ***Samples' preparation***

DSRCT tumor frozen samples corresponding to those used for 3'scRNA-seq, along with matching germline tissue from either peripheral blood mononuclear cells (PBMCs) or non-tumoral tissue, were collected. DNA extraction was performed using Qiagen DNeasy Blood and Tissue Kit or QiAmp DNA FFPE Tissue Kit for non-tumoral FFPE samples when patients' derived PBMCs could not be obtained.

### ***Sequencing***

After BioAnalyzer quality control, genomic DNA was sheared, and exons were captured by Agilent Sure Select Human All Exon V6 kit. Sequencing libraries were prepared and sequenced with Paired-end sequencing (150bp) on an Illumina Novaseq 6000 sequencer with 40 million reads per sample.

### ***Data processing***

WES data were processed using the nf-core/sarek pipeline (<https://github.com/nf-core/sarek>). Following alignment to hg38 with BWA allowing up to 4% of mismatches, bam files were cleaned using Genome Analysis Toolkit (GATK) with duplicate marking and base quality score recalibration.

### ***Downstream analyses***

Copy number variation between normal and matched tumor tissue was computed using CNVkit<sup>33</sup> tool (<https://github.com/etal/cnvkit>).

## **JN-DSRCT-1 cell line Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)**

### ***Assay for Transposase-Accessible Chromatin***

JN-DSRCT-1 cells were grown to target ~80% confluence in 6-well plates after a 48 h EWSR1-WT1 or CCND1 siRNA mediated knock-down or Dharmacon non-targeting siRNA (Dharmacon ON-TARGET plus non-targeting siRNA#1) transfection. Independent biological duplicates were performed for this experiment.

Although working on cell lines, ATAC was performed according to Corces et al.<sup>34</sup> OMNI ATAC protocol optimized from standard Buenrostro et al. protocol<sup>35</sup>. Briefly, cells were harvested to target 100,000 cells per sample. Nuclei extraction was done by cell lysis by incubating cells for 3 min on ice in cold lysis buffer (Resuspension buffer (10 mM Tris-HCl, pH 7.5; 10 mM NaCl; 3 mM MgCl<sub>2</sub>); 0.1% NP-40; 0.1% Tween-20; 1% Digitonin). The suspension was then washed in Wash buffer (resuspension buffer with 10% Tween-20) before centrifugation at 500 g for 10 min at 4°C to isolate pellets. The DNA transposition was performed by incubating nuclei during 45 min at 37°C in a thermomixer at 1,000 rpm in a transposition reaction mix containing 2X TD buffer, 1X PBS, with 0.1% Tween-20, 0.01% Digitonin, 10% Tn5 Transposase (Tagment DNA Enzyme 1). DNA purification was then performed using Qiagen MinElute Reaction Cleanup kit. Library amplification was performed using NEBNextUltraII library (New England Biolabs) while determining the necessary total number of PCR cycles by qPCR. Next, libraries were purified with KAPA pure beads (Roche) before assessing the library quality on Agilent BioAnalyzer with High sensitivity DNA kit. Libraries were finally sequenced at 200M depth per sample on NovaSeq with Paired-end 150 bp reads.

### ***Data analysis***

ATAC-seq data were analyzed according to the nf-core/atacseq pipeline<sup>36</sup>. Raw reads QC was performed using FastQC, followed by adapter trimming with Trim Galore!. After reads' mapping to a reference genome (hg38) using BWA, duplicate reads were discarded using picard. BigWig files were generated using BEDtools for IGV visualization. Genome-wide immunoprecipitation enrichment relative to input was performed using deepTools. Broad and narrow peaks were called using MACS2 and annotated relative to gene features using HOMER. Finally, differential binding analysis was performed using DESeq2. Motif analyses were subsequently performed on significantly enriched peaks using the MEME suite. Enrichment in known transcription

factors was evaluated among the JASPAR2020 database using the “ame” tool. EWSR1-WT1 *de novo* motifs were explored using the “dreme” tool.

## **JN-DSRCT-1 cell line WT1 C-terminal Chromatin Immunoprecipitation with sequencing (ChIP-seq)**

### ***Chromatin Immunoprecipitation, library generation, and sequencing***

JN-DSRCT-1 cells were grown to ~80% confluence in a 15 cm dish. The chromatin was crosslinked for 15 min at room temperature (RT) using methanol-free formaldehyde (1% final). The formaldehyde was quenched by adding Glycine solution at a final concentration of 125 mM for 5 min at room temperature (RT). After two washes with ice-cold PBS supplemented with cOmplete Protease Inhibitor Cocktail, EDTA-free (Roche), cells were scraped and centrifuged for 5 min at 300 g before freezing at 80°C for subsequent use. After thawing on ice, cells were resuspended in Farnham lab (FL) buffer (5 mM PIPES pH 8; 85 mM KCl; 0.5% Igepal CA-630) supplemented with cOmplete Protease Inhibitor Cocktail, EDTA-free (Roche) at a final concentration of 1M cells/mL. Furthermore, nuclei extraction was performed according to NEXSON (Nuclei Extraction by SONication) protocol <sup>37</sup>. Cell suspensions were sonicated in 12x12 mm/1 mL microtubes with AFA fiber using Covaris S220 focused-ultra sonicator at 75W peak power, 2% duty factor, and 200 cycles/burst at 4°C for 2 min. Isolated nuclei were resuspended in 1 mL of shearing buffer (10 mM Tris-HCl pH 8; 0.1% SDS; 1 mM EDTA) supplemented with Complete Protease Inhibitor Cocktail EDTA-free (Roche). Then, the chromatin was sheared using Covaris S220 focused-ultra sonicator at 140W peak power, 5% duty factor, 200 cycles/burst during 20 min at 4°C. When needed, sheared chromatin was kept at -80°C for subsequent use the day after. To control the quality of the chromatin shearing, an aliquot of sheared chromatin was incubated overnight with proteinase K at 65°C to reverse the crosslink. DNA was then purified using Qiagen PCR purification kit and analyzed on a Bioanalyzer (DNA High Sensitivity kit) to check for size distribution.

Immunoprecipitation was performed on sheared chromatin diluted with 10X dilution buffer (0.01% SDS, 1.1% Triton X100, 1.2 mM EDTA, 16.7 mM Tris pH8, 167 mM NaCl) by incubating WT1 C-terminal (Genetex GTX12549) targeting antibody at 0.5 mg/mL final concentration or rabbit isotype at 1 mg/mL final concentration overnight at 4°C. The next day, Dynabeads protein G beads (ThermoFischer Scientific) were blocked with PBS-BSA 0.5% at RT for 30 min. Antibody-incubated chromatin was

added upon blocked Dynabeads to capture immune complexes. Immunoprecipitates were then washed twice using successively low salt buffer (0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris pH8, 150 mM NaCl), high salt buffer (0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris pH8, 500 mM NaCl), LiCl wash buffer (250 mM LiCl, 1% NP-40, 1% Na-Deoxycholate, 10 mM Tris pH8, 1 mM EDTA) and TE 1X buffer. The precipitated chromatin was eluted from beads by heating at 65°C in Elution buffer (25 mM Tris pH 7.5, 5 mM EDTA, 0.5% SDS). The crosslink was reverted from eluted samples by overnight incubation at 65°C with Proteinase K. Next, DNA was purified on Qiagen PCR purification columns. Library generation was finally performed according to NEBNext Ultrall DNA kit for Illumina general protocol. Quality control of amplified libraries profiles was performed on Agilent Bioanalyzer before sequencing samples at 200M reads depth on NovaSeq with Paired-end 150pb.

Independent biological duplicates were performed for this experiment.

### ***ChIP data analyses***

#### *Peak calling, gene annotation, and differential binding analysis*

ChIP data analyses were performed according to nf-core/chipseq pipeline<sup>36</sup>. Briefly, a raw read QC (FastQC) was performed before adapter trimming with Trim Galore!. Reads were further mapped to a reference genome (hg38) using BWA, and duplicate reads were discarded using picard. BigWig files were generated using BEDtools to allow IGV fragments visualization. The distribution of peaks was annotated using HOMER. Gene-wide immunoprecipitation enrichment relative to input was performed using deepTools. Broad and narrow peaks were called using MACS2 and were annotated relative to gene features using HOMER. Finally, differential binding analysis was performed using DESeq2.

#### *ChIP gene set enrichment analysis*

Gene set enrichment (GSE) was performed on WT1 ChIP-seq data using two methods. First, peak-associated genes with a peak fold enrichment >2 were selected to generate a gene list, which was further used as an input for ToppFUN analysis (ToppGene Suite) focusing on Gene Ontology gene sets. Secondly, we performed GSE evaluation on a list of ChIP-seq-derived genomic regions, defined as the differentially enriched broad peaks' genomics ranges deriving from DESeq2 analysis on WT1 C-ter ChIP compared to isotype. For the latter analysis, we used ChIP-Enrich package<sup>38</sup>, which accounts for biases induced by the properties of ChIP-seq data, including increasing Type I error secondary to multiple testing (various numbers of peaks for a single gene) and the

gene-length bias. Briefly, the ChIP-Enrich pipeline is divided into four steps. First, loci of interest are defined as the regions spanning the midpoints between the TSSs of adjacent genes so that each peak genomic range is assigned to the gene with the nearest TSS. Secondly, the proportion of each gene locus covered by ChIP-seq peaks is calculated. Thirdly, a logistic regression is performed for each GO gene set, including a correction for locus length. Finally, for each GO gene set, p-values for enrichment or depletion are adjusted for multiple testing.

#### *EWSR1-WT1 ChIP-seq signature*

An EWSR1-WT1 ChIP-seq signature (n=176 genes) reflecting EWSR1-WT1 transcriptional activity was inferred by selecting peaks' corresponding genes displaying a fold enrichment >5 compared to control isotype ChIP (DESeq2). The calculation of a ChIP-seq signature score at the single-cell level on the 3'scRNA-seq Harmony integrated data ("*Int\_sc*") was performed using the AddModuleScore function (Seurat), which computes the average expression levels of a program of interest, subtracted by the aggregated expression of randomly selected control features.

#### *k-mers overrepresentation analysis*

We searched for k-mers enrichment within the 100 bp-wide sequences around EWSR1-WT1-specific ChIP-seq peaks' summits. EWSR1-WT1-specific ChIP-seq peaks were inferred from significantly ( $p < 0.05$ ) enriched peaks in EWSR1-WT1 ChIP compared to control isotype ChIP using DESeq2. We compared the frequency of all 6-mers oligonucleotides ( $n = 6^6$ ) found within EWSR1-WT1 ChIP-seq peaks' summits to their frequency within the whole genome.

#### *Motif enrichment analysis*

Motif enrichment analyses were subsequently performed on significantly enriched peaks using the MEME suite. Enrichment in known transcription factors was evaluated among JASPAR2020<sup>39</sup> database using the "ame" function. EWSR1-WT1 *de novo* motifs were characterized using "the streme" function and were matched to known TF motifs using Tomtom motif comparison tool.

#### *ATAC-seq and ChIP-seq integration*

We analyzed overlapping peaks between EWSR1-WT1-specific binding sites inferred from isotype versus EWSR1-WT1 differential binding analysis ("EWSR1-WT1 ChIP module") and EWSR1-WT1-induced differentially accessible peaks generated from EWSR1-WT1-silenced versus non-silenced JN-DSRCT-1 ATAC-seq data ("EWSR1-WT1 ATAC module").

## **Immunohistochemistry**

DSRCT cases, including those corresponding to the samples processed for scRNA-seq, were selected by an expert sarcoma pathologist. Serial sections of FFPE tissue were cut into 3  $\mu\text{m}$ -thick sections.

### ***Fluorescent stainings***

Fluorescent multiplex stainings were performed on Bond-RX automated stainer (Leica Biosystems). Then, dewaxing was performed with BOND Dewax Solution. Next, antigen retrieval was performed using BOND Epitope Retrieval Solution 2 (pH=9) for 20 min at 100°C. Protein blocking was performed with PKI blocking (Akoya) for 5 min. Antigen detection was performed using the Opal™ system (Akoya) with Opal anti-mouse and anti-rabbit HRP polymer.

Between each sequence, antigen stripping was performed using BOND Epitope Retrieval Solution 1 (pH=6) for 20 min at 100°C.

A counter-staining was done with DAPI (Akoya). After staining, tissue sections were submitted to serial gradients of xylene and mounted manually with a coverslip using Prolong Diamond.

The following antibodies and respective conditions were used for THY1/CHI3L1/Desmin triplex: anti-THY1/CD90 (1:1,000, rabbit IgG D3V81 clone, Cell Signaling #13801), detected with OPAL480 fluorophore at 1:75; anti-CHI3L1 (1:1600, polyclonal rabbit, Abcam ab77528), detected with OPAL570 fluorophore at 1:100; anti-Desmin (1:200, monoclonal mouse D33 clone, DAKO M0760), detected with OPAL690 at 1:150.

The conditions for FAP/MCAM/ACTA2/ triplex assay were as following: anti-FAP (1:200, polyclonal sheep, R&S AF3715), detected by rabbit anti-sheep (P0163, DAKO) and OPAL520 fluorophore at 1:50; anti-MCAM (1:600, mouse OTI5C4 clone, Origene TA803548), detected with OPAL570 1:600; anti-ACTA2 (1:1,000, mouse 1A4 clone, DAKO M0851), detected with OPAL690 at 1:150.

### ***Mono DAB stainings***

Mono DAB stainings were performed using the automated Benchmark Ultra stainer. Dilution and antibodies for mono DAB IHC stainings were as follows: anti-c-Myc (fixed concentration pre-filled syringe, monoclonal anti-rabbit, Y69 clone, Roche 790-4628), anti-WT1 (1:100, polyclonal rabbit, Zytomed AB\_2864626), anti-Desmin (1:40, monoclonal mouse D33 clone, DAKO M0760), anti-AE1/AE3 (1:75, monoclonal

mouse, Diagnostic BioSystem Clinisciences Mob190-05), anti-CD3 (1:100, polyclonal rabbit, DAKO A0452), anti-THY1 (alias CD90, 1:750, rabbit IgG D3V81 clone, Cell Signaling #13801), and CD56 (alias NCAM, fixed concentration pre-filled syringe, monoclonal rabbit, MRQ-42 clone, Roche 760-4596).

Revelations were performed using UV DAB kit, and counter-stainings were done with hematoxylin and bluing reagent.

### ***Dual chromogenic stainings***

Dual chromogenic CD68/CD163 staining was performed on automated Discovery Ultra stainer using the following antibodies references and dilutions: anti-CD163 (1:100, Diagnostic BioSystem Clinisciences Mob460-05) followed by anti-mouse HRP and DAB incubation, and anti-CD68 (1:200, PG-MI clone, DAKO M0876), followed by anti-mouse HRP and DS Discovery purple incubation.

Counter-stainings were performed with hematoxylin and bluing reagent.

### ***Image analysis***

Slides were scanned on Olympus VS120 automated slide scanner followed by visualization with OlyVIA v2.9. Images were analyzed using QuPath 3.0.0 and ImageJ v1.53.

## **CAFs Coculture**

### ***Colony formation assay***

JN-DSRCT-1 cells were cocultured with mouse Cancer-Associated Fibroblasts (CAFs) isolated from a patient-derived xenograft (PDX). Briefly, the PDX was collected from a Nod Scid Gamma (NSG, Charles River) mouse and dissociated with Tumor Dissociation kit (Miltenyi Biotech). CAFs were isolated with Tumor-Associated Fibroblast mouse kit (Miltenyi Biotech). Isolated cells were seeded on 6-well plates into DMEM-F12 complemented medium, supplemented with 1X Insulin-Transferrin-Selenium (ITS). The day after, a 0.4  $\mu\text{m}$  Transwell<sup>®</sup> device was inserted into each well, and JN-DSRCT-1 cells were seeded (500 cells per well) for colony-forming assay. Mock wells without CAFs were used as a control to compare JN-DSRCT1 cell growth in the presence or absence of CAFs.

Both 6-well plates seeded with JN-DSRCT1 cells and Transwell<sup>®</sup> inserts seeded with CAFs were washed with PBS and stained with 0.5% crystal violet dilute in methanol.

Plates and Transwell® inserts were scanned, and the total colony area was evaluated using ImageJ. The staining was then solubilized in 100% Methanol for 20 min, and triplicate aliquots were collected to measure absorbance at 570 nm.

### ***3D spheroids coculture growth test***

GFP clone #5 JN-DSRCT-1 cells were plated on ultra-low adherence (ULA) 96-well plates at a concentration of 1,000 cells per well in DMEM F12 before 15 min centrifugation at 500 g. Plates were incubated at 37°C in 5% CO<sub>2</sub>, and the formation of the spheroid was assessed 72 H after seeding. PDX-derived CAFs were primarily isolated as previously described using Tumor-Associated Fibroblast mouse kit (Miltenyi Biotech) and stained using Cell Proliferation Staining Reagent (Abcam, Deep Red Fluorescence), according to the manufacturer protocol. Stained CAFs (n=500) were then added to wells containing JN-DSRCT-1 spheroids. Pictures were acquired every two days to assess the spheroid growth on the EVOS™ XL Core Imaging System and were measured according to the product of their two largest diameters. On Day 15, living spheroids were pictured on a spinning-disk microscope (Nikon Eclipse Ti2 equipped with Leica/Gatca CSU-W1 confocal scanner unit, Live-SR module, and incubation chamber) using a 20X oil objective. Spheroids were then measured using ImageJ software v1.53, and tumor cells were distinguished from CAFs based on their fluorescence.

### **2D Immunofluorescence**

JN-DSRCT-1 cells were seeded and grown on a glass coverslip inserted into wells from a 6-well plate and transfected by a siRNA targeting EWSR1-WT1 or a non-targeting siControl (Dharmacon) or mock-transfected. After 48 H, cells were fixed with 4% formaldehyde in PBS for 20 min at RT. After fixation, cells were washed twice with PBS and permeabilized with 0.2% Triton X-100 in PBS for 10 min. After three successive washes in PBS, cells were blocked for 1 H into IFF buffer (2% BSA, 2% FBS in PBS) at RT. Primary antibodies (anti-WT1 C-terminal, 1/500, rabbit polyclonal, GeneTex GTX15249 and anti-Desmin, 1/100, mouse monoclonal D33 clone, Dako M0760) were diluted in IFF buffer and added on the cell layer for overnight incubation at 4°C. This incubation was followed by three washes in PBS. Secondary antibodies (Goat anti-rabbit AlexaFluor 555, 1/5,000, and Goat anti-mouse AlexaFluor 488, 1/5,000) were diluted in IFF buffer, and cells were incubated 1 H at RT for secondary antibody binding. Next, cells were washed twice, once with 1 µg/mL DAPI in PBS and

secondly in PBS. Glass coverslips were mounted on microscope slides on mounting medium (Sigma) and sealed with nail polish. Slides were imaged with epi-fluorescence microscope Eclipse Ti with a 100x magnification (Nikon Instrument Europe B.V., Surrey, England).

### **3D Immunofluorescence**

JN-DSRCT-1 cells were seeded and grown in 3D as previously described. The immunofluorescence assay was performed to evaluate the staining pattern of elected markers within the spheroids. After 15 days of growth in ULA plates, spheroids were fixed for 30 min in PBS-PFA 4%, permeabilized with Triton X-100 0.2% for 10 min, and blocked with PBS-BSA 3%, Triton X-100 0.2% and Tween 20 0.5%. Spheroids were then incubated within 1/100 diluted primary antibodies (anti-WT1 C-terminal (rabbit polyclonal Genetex GTX12549), or anti-Desmin (mouse monoclonal D33 clone, Dako M0760)) overnight at RT. After three PBS washes, spheroids were incubated overnight within secondary antibody solutions (Goat anti-rabbit AlexaFluor 555, 1/5,000 and Goat anti-mouse AlexaFluor 488, 1/5,000). The day after, a DAPI staining was finally performed by incubating spheroids in a 100 µg/mL DAPI solution. Before image acquisition, spheroids were submerged within a sucrose and urea solution to allow tissue clearing with minimal tissue shrinkage. Spheroids were then pictured on a spinning-disk confocal microscope (Nikon Eclipse Ti2 equipped with Leica/Gataca CSU-W1 confocal scanner unit, Live-SR module, and incubation chamber) using a 20X oil objective and appropriate wavelength.

## **Results**

### **scRNA-seq deciphers DSRCT cell-type composition**

To comprehensively characterize DSRCT intratumor heterogeneity, we performed 3' scRNA-seq using the droplet-based <sup>40</sup> 10X Genomics Chromium 3'scRNA-seq assay on 11 fresh human DSRCT samples and one juxta-tumor peritoneal sample (**Fig. 1A, B; Supplemental Information – Table S1**), as well as simultaneous single-nucleus RNA-seq (snRNA-seq) and single-nucleus ATAC-seq (snATAC-seq) (single-nucleus Multiome) (snMultiome) on one DSRCT sample. To further assess intertumor heterogeneity, this was complemented by bulk whole exome sequencing (WES) on

nine single-cell assays matched samples and bulk RNA-sequencing of 29 archived frozen DSRCT patient samples.

As expected, merging the first 11 individual 3'scRNA-seq datasets showed that most tumor cells were grouped by patient identity, while non-malignant cells were gathered by cell type (**Supplemental Information – Fig. S1A, B**). To allow the efficient integration of these individual 3'scRNA-seq datasets<sup>41–46</sup>, we used Harmony<sup>19</sup> algorithm and generated a unique embedded 3'scRNA-seq dataset (“*Int\_sc*” dataset) (**Fig. 1C**).

After quality control and filtering, we could identify 51,671 high-quality cells within the “*Int\_sc*” dataset, including 3,063 cells arising from the juxta-tumor peritoneal sample. On average, we could detect 1,908 genes per cell (**Supplemental Information – Table S3**).

Malignant cells were predominantly identified (n=44,781) (**Fig. 1D, Supplemental Information – Table S2**) and confidently labeled using two complementary DSRCT characteristic genetic features that have been previously described: (1) the top 10 DSRCT marker genes inferred from differential gene expression (DGE) analysis between DSRCT and a subset of alternative sarcomas<sup>47</sup> (**Supplemental Information – Fig. S1C**); (2) the median expression of DSRCT-specific neotranscripts, as described by Vibert et al.<sup>48</sup> (**Supplemental Information – Fig. S1D**).

The assignment of non-tumor cell types composing the integrated “*Int\_sc*” dataset was performed according to the assessment of known microenvironment cells' canonical markers (**Supplemental Information – Table S2**). Non-tumor cell types included, by decreasing order of abundance, cancer-associated fibroblasts (CAFs) (n=2,360), myeloid cells (n=1,966), endothelial cells (n=1,244), T cells (n=1,126) and mesothelial cells (n=784) (**Fig. 1D, Supplemental Information – Table S2**). The relative composition of microenvironment cells was consistent across DSRCT tumor samples. However, several samples showed increased CAFs (*GR2 site#2*, *GR2 site#3*) or T cell infiltration (*GR4\_PC*) (**Fig. 1D**). These proportions were in line with the known histological and immunohistochemical profile of DSRCT, which typically harbors a paucicellular desmoplastic stroma mainly composed of CD68/CD163-expressing macrophages, CD90 (THY1)-positive CAFs, and extremely rare lymphocytes (**Fig. 1E**).

CIBERSORTx deconvolution on a bulk RNA-seq cohort of 29 frozen DSRCT samples showed variable proportions of tumor cells, CAFs, and myeloid cells. In contrast, T

cells, endothelial cells, and mesothelial cells were almost not detectable (**Fig. 1F**, **Supplemental Information - Table S27**).

Differential gene expression (DGE) analysis performed between the identified cell populations highlighted DSRCT malignant cells and non-malignant subpopulations *de novo* signatures (**Supplemental Information – Table S2**). Importantly, while DSRCT tumor cells' top marker genes were concordant with previously described DSRCT-specific genes<sup>47</sup> (e.g., *APOA1*, *CCL25*, *GJB2*, *GAL*), no ubiquitous marker could be universally found in all DSRCT tumor cells (**Fig. 1G**), suggesting some level of tumor cells' heterogeneity.

### DSRCT tumor cell heterogeneity features suggest both lineage and metabolic plasticity

To further explore DSRCT cell heterogeneity, we performed a gene expression-based k-nearest neighbors clustering on the embedded “*Int\_sc*” dataset (**Supplemental Information – Fig. S1E, F**), as well as on single or merged datasets from synchronous distinct tumor sites' samples when available (for *GR2*, *GR7*, and *BOM* datasets) (**Supplemental Information – Fig. S2A, B**). We manually annotated malignant and non-malignant cells' clusters across the “*Int\_sc*” dataset (**Supplemental Information – Table S3**) and individual and merged datasets according to DGE analysis marker genes (**Supplemental Information – Table S4**).

The clustering of the “*Int\_sc*” dataset's tumor cells identified 13 tumor cell clusters (**Fig. 2A**) comprising six lineage-related clusters: (1) a poorly differentiated cluster (*Multidiff\_ASCL1 - 0*), characterized by multilineage genes expression and *ASCL1* marker gene, (2) two mesenchymal clusters (*Mesenchymal\_fibrosis - 8* and *Mesenchymal\_secretion - 11*) respectively defined by *ACTA2/CHI3L1/KRT17* and *CCK/TNNT3/PTGDS* overexpression, (3) one epithelial cluster (*Epithelial\_mesenchymal - 4*), defined by the concomitant expression of epithelial (*CLDN3*, *CDH3*, *MUC16*) and mesenchymal markers (*MB*, *ACTA2*), (4) one neuronal/neuronal cluster (*Neural\_neuronal - 1*) overexpressing *NTRK3*, *TUBA1A* and *GAL*, and (5) one neural stem cluster (*Neural\_stem - 19*), expressing *NES* and *ELAVL2/4*. These multidirectional differentiation states related to distinct developmental cell fates were further corroborated by single-cell label transfer using Azimuth<sup>29</sup> with a fetal development gene expression atlas<sup>49</sup> on the “*Int\_sc*” dataset

**(Supplemental Information – Fig. S1G).** Overall, lineage-related clusters support the characterization of DSRCT as a polyphenotypic differentiated tumor<sup>50</sup> **(Fig. 2B).**

In addition to lineage-oriented DSRCT tumor cell subpopulations, we identified additional clusters associated with specific metabolic pathways. Metabolism-related clusters encompassed two functional DSRCT tumor cells' states: a subpopulation showing activation in genes involved in anaerobic glycolysis metabolism (*Metabolic glycolysis - 5*, overexpressing *ENO1*, *LDHA*, *NDUFA4L2*), and a cluster characterized by serine metabolism-related genes (e.g., *PSAT1*, *ASNS*, *PHGDG*), and tRNA synthetase-related genes (e.g., *YARS*, *CARS*, *GARS*) (*Metabolic\_serine - 15*) **(Fig. 2A, C, Supplemental Information – Table S3A, B).**

We further defined additional cell states that we named “pseudostates,” either related to the cell cycle or pathways of uncertain biological relevance. Pseudostates-related clusters comprised (1) a cycling cells' cluster (*Cycling cells - 3*), expressing *TYMS*, *MKI67*, and *TOP2A*, as well as additional tumor cells' subgroups characterized by the overexpression of (2) ribosomal protein genes (*Ribosomal\_catabolic - 2* and *Ribosomal\_IFN response - 6*), or (3) long non-coding RNAs (lncRNAs) such as *MALAT1*, *MEG3*, *GRID2* and *NEAT1* (*Lnc\_1 - 7*, *Lnc\_2 - 17*) **(Fig. 2A, C Supplemental Information – Table S3A, B).** Considering the significantly lower number of features found in ribosomal protein genes-enriched and lncRNAs-enriched clusters **(Supplemental Information – Fig. S1F)**, their biological relevance remains elusive. The ribosomal protein genes-enriched cluster may correlate to cells showing hyperactivation of ribosomal biogenesis and thus reflect DSRCT tumor cell growth. Concerning the lncRNAs-enriched cluster, considering the frequent identification of such lncRNAs in other single-cell datasets<sup>51,52</sup>, we hypothesized that these clusters encompassed cells that had undergone nuclear membrane damage, thereby leading to an overrepresentation of unspliced transcripts and lncRNAs<sup>53,54</sup>. Although we cannot exclude that these clusters might have some biological relevance in the context of EWSR1-WT1 TF activity, we did not focus on them in subsequent analyses.

Importantly, even if the lineage-related clusters' canonical markers were initially defined based on the most differentially expressed genes, we observed a high degree of overlap in expression across clusters, with very few genes being strictly specific to one cluster **(Fig. 2C, Supplemental Information – Table S3A, B).** This suggested some degree of plasticity within DSRCT tumor cells, allowing them to dynamically transition between a continuum of differentiation or functional states.

The separated analysis of individual or merged single-cell datasets concurred with these findings. Notably, when analyzing merged datasets combining synchronous tumor samples from distinct tumor sites within the same patient, tumor cells grouped by cell states rather than by tumor site, suggesting intrapatient intertumor homogeneity (**Supplemental Information – Fig. S2A, B**). However, this observation should be interpreted cautiously as all analyzed tumor samples arose from peritoneal sites, and no distant metastasis sampling was performed. Twelve DSRCT tumor cells' phenotypes were consistently found across multiple tumor samples, comprising five differentiation or lineage-related states, two metabolic-related states, and five pseudostates. The recurrent differentiation states-related clusters were in line with findings from the embedded “*Int\_sc*” dataset, displaying either neural (*ASCL1*, *TUBA1A*, *GAL*, *GJB2*), mesenchymal (*DES*, *TNNT3*, *CHI3L1*, *ACTA2*, *MSL1*), epithelial (*EPCAM*, *CDH1*, *MUC1*, *KRT*), neuronal (*ELAVL2*, *ELAVL4*, *FOXA2*), or stem cells (*NES*) marker genes (**Supplemental Information – Fig. S2C, Table S4**). Of note, in addition to the three previously described pseudostates (i.e., related to cycling cells, ribosomal protein genes, and lncRNAs), two additional pseudostates-related recurrent clusters were identified across individual or merged samples. The latter showed either overexpression of immediate early genes (e.g., *FOS*, *JUN*, *NR4A1*, *DUSP1*, *EGR1/2/3/4*) or of mitochondrial genes but were not found within the “*Int\_sc*” dataset. While the mitochondrial genes-enriched cluster is likely related to apoptotic cells, the biological interpretation underlying the *FOS/JUN*-enriched clusters should be considered. Indeed, other authors described that such clusters may result from the activation of immediate early response genes secondary to damaging tumor dissociation settings on DSRCT and other sarcoma samples<sup>55</sup>. However, one should also envisage that such a response may also be triggered *in vivo* by external stressful stimuli or within already stress-vulnerable cells. In addition, some DSRCT tumor cell subpopulations were exclusively found in specific samples. For example, the *BOM* dataset displayed the highest level of intratumor heterogeneity and comprised a unique interferon (IFN) response-enriched cluster (*BOM* cluster 8) (**Supplemental Information – Table S4**).

Functional enrichment analysis of Gene Ontology (GO) pathways (i.e., GO Biological Process (GOBP), GO Cellular Component (GOCC), and GO Molecular Function (GOMF)) based on the top 100 differentially expressed genes from each predefined “*Int\_sc*” cluster further confirmed a characteristic enrichment in lineage-related

pathways such as GOBP\_Neuroepithelial cell differentiation (adjusted p-value=3.84E-02), GOBP\_Cell-cell junction organization (adjusted p-value=1.03E-07), GOBP\_Osteoblast differentiation (adjusted p-value= 2.35E-03), or GOBP\_Regulation of trans-synaptic signaling (adjusted p-value= 1.32E-04) respectively for *Neural\_neuronal* - 1, *Epithelial\_Mesenchymal* - 2, *Mesenchymal\_fibrosis* - 8, and *Neural\_stem* - 19 cluster (**Fig. 2D, Supplemental Information – Table S5**). Metabolism-related pathways were significantly enriched in the *Metabolic\_glycolysis* – 5 cluster (e.g., GOBP\_ADP metabolic process, GOBP\_Nucleoside diphosphate phosphorylation). Notably, the *Metabolic\_serine* – 15 cluster displayed GOBP\_Intrinsic apoptotic signaling pathway enrichment in response to hypoxia (**Fig. 2D, Supplemental Information – Table S5**). LncRNAs-enriched clusters were characterized by enrichment in chromatin remodeling-related pathways, such as GOCC\_Nuclear speck (adjusted p-value=4.39E-04 and adjusted p-value=2.13E-02 for *lncRNA* – 7 and *lncRNA* – 17, respectively), as well as GOCC\_nBAF complex (adjusted p-value= 1.11E-02) and GOCC\_SWI/SNF superfamily-type complex (adjusted p-value=1.52E-02) for *lncRNA* – 7 cluster, or RNA helicase activity for *lncRNA* – 17 cluster (adjusted p-value= 2.13E-02) (**Fig. 2D, Supplemental Information – Table S5**).

### Co-regulated genes are widely overlapping across DSRCT tumor cell clusters

Since DSRCTs are solely driven by an aberrant transcription factor, which could itself interact with various partners and coregulators, we investigated modules of co-expressed genes in the integrated dataset (Hotspot algorithm<sup>20</sup>, see Methods section) (**Fig. 2E, Supplemental Information - Table S6, S7, Fig S3A**). Notably, the latter analysis showed both cell clusters' specific modules, such as the Cycling module (*module #3*), whose expression was exclusively increased in the *Cycling cells* - 3 cluster, or *module #13*, significantly activated in the *Metabolic\_glycolysis* – 5 cluster. On the contrary, most other modules displayed homogeneous scores across clusters. In particular, *Cycling cells* - 3 cluster cells were characterized by an enrichment of both gene *modules #1, #2, #10, and #11*. In contrast, Hotspot *modules #1 and #2* were mainly composed of ribosomal protein genes suggesting pathways involved in active translation and ribosome biogenesis. *Module #11* displayed an overrepresentation of mitochondrial encoded genes (e.g., *MT-CO3, MT-CO1, MT-ATP6*) relative to NADH dehydrogenase (ubiquinone) activity and oxidative phosphorylation pathway. *Module*

#10 was composed of 25 genes, including *ASCL1*<sup>56</sup>, which is known to be a direct target of EWSR1-WT1 and acts as a neuronal reprogramming factor. Of note, *module #7* was overexpressed but not restricted to the *Epithelial\_Mesenchymal – 4* cluster and was notably enriched in GOBP\_Epithelium development (adjusted p-value=1.07E-02) and GOBP\_Cell-cell junction organization (adjusted p-value=1.79E-2). Additional modules included *modules #4, #5, #6, #9, and #14*, respectively enriched in cell migration-related pathways, immune response pathways, TF DNA binding gene sets, lncRNAs, and immediate early genes.

Overall, coregulated gene modules inference suggests the existence of DSRCT cells' "hybrid" transcriptional states, characterized by highly overlapping transcriptional programs.

### DSRCT tumor cell clusters do not follow consistent trajectories

To determine whether DSRCT tumor cell clusters could result from distinct sequential time-related evolutionary steps, we further sought to infer cells' trajectories from gene expression across the "*Int\_sc*" dataset. Importantly, directed single-cell fate mapping combining trajectory and RNA velocity inference did not show consistent trajectories across individual or merged datasets clusters, suggesting the existence of more or less reversible dynamic states between which DSRCT cells may oscillate, rather than genuine evolutionary trajectories (*data not shown*). Moreover, the inference of cell-based entropy<sup>24</sup> on the "*Int\_sc*" dataset suggested homogeneous stemness states across clusters (**Supplemental Information – Fig. S3B**). The inference of single cells' degree of differentiation<sup>23</sup> aligned with the latter observation, highlighting poorly differentiated tumor cells compared to non-malignant cells, with a relatively homogeneous differentiation score except for pseudostates-related clusters (*Ribosomal\_catabolic – 2, Ribosomal\_IFN response – 6, Inc\_1 – 7, and Inc\_1 – 17*) which were predicted to have a higher degree of differentiation (**Supplemental Information – Fig. S3C**).

Overall, our findings highlight that DSRCT cells are characterized by heterogeneous lineage commitment processes and metabolism-related states with highly overlapping coregulated gene modules. These data suggest the existence of hybrid unstable phenotypical states between which DSRCT tumor cells may fluctuate.

## Copy number variation (CNV) is homogeneous across DSRCT tumor cell clusters

To identify the potential role of genetic clonal or subclonal evolution in DSRCT tumor cell heterogeneity, we performed Copy Number Variation (CNV) inference <sup>57</sup> and confirmed these results on matched WES data (**Supplemental Information – Fig. S4**).

Overall, CNV inference and WES-identified copy number alterations were highly concordant (**Fig. 2F, G**). Importantly, the *GR4* sample, which originates from the sole non-pretreated patient, displayed almost no CNV across DSRCT tumor cell clusters, the 6p deletion being attributed to a bias since immune cells have been used as a reference cluster and may overexpress MHC genes. The comparison of CNV analysis from samples collected before (*GR4*) and after chemotherapy treatment plus first relapse (*GR4\_PC*) in the same patient did show any increase in CNV complexity. Of note, few recurrent CNVs were observed across samples. A gain of 5p and 5q was identified in three patients (*GR2*, *GR4\_PC*, and *GR7*), while no gain of 1q was detected, although previously described as a recurrent CNV in DSRCT <sup>12</sup>. The CNV inference from synchronous distinct tumor site samples within the same patient showed concordant CNVs across tumor sites (**Supplemental Information – Fig. S4**). Finally, CNVs appeared to be highly homogenous across DSRCT cell clusters in most samples. This observation further supports the hypothesis that DSRCT cell heterogeneity may derive from plastic and dynamic transcriptional programs rather than clonal or somatic evolution.

## DSRCT tumor cell heterogeneity is not driven by *EWSR1-WT1* transcript expression level variation

Previous studies on Ewing sarcoma, another small round blue cell tumor driven by the *EWSR1-FLI1* chimeric transcription factor, have reported that variable *EWSR1-FLI1* activity <sup>14-15</sup> is a significant source of intratumor heterogeneity. We therefore hypothesized that similar mechanisms might also operate in DSRCT. We first speculated that *EWSR1-WT1* transcript expression level variations could drive distinct transcriptional pathways and thus influence DSRCT tumor cells' commitment toward specific archetypal phenotypes. We relied on the hypothesis that genuine 10X Genomics 3'scRNA-seq derived data may not allow uncovering *EWSR1-WT1* expression level differences across cells due to (i) the potential confounding of wild-

type *WT1* and *EWSR1-WT1* transcripts expression (**Supplemental Information – Fig. S5A**), and (ii) an insufficient sequencing depth on lowly expressed transcripts. To answer this question, we took advantage of an in-house developed assay aiming at specifically amplifying *EWSR1-WT1* transcripts from 10X derived barcoded cDNAs (**Supplemental Information – Fig. S5B**). Importantly, we did not detect substantial *EWSR1-WT1* expression variation across DSRCT tumor cell clusters (**Supplemental Information – Fig. S5C**). Because our above results did not support high variability in *EWSR1-WT1* level of expression, we hypothesized that intratumor heterogeneity may be due to alternative molecular mechanisms driving *EWSR1-WT1* activity or to tumor cell-extrinsic signals.

### **EWSR1-WT1 transcriptional activity is linked to modifications in DSRCT cells' chromatin landscape**

To investigate whether variations in *EWSR1-WT1* transcriptional activity or *EWSR1-WT1* DNA binding sites may drive DSRCT heterogeneity, we first sought to study *EWSR1-WT1* target regions and *EWSR1-WT1* role in chromatin accessibility using ChIP-seq and ATAC-seq in the JN-DSRCT-1 cell line in which we conditionally modulated *EWSR1-WT1* expression by siRNA. Notably, the JN-DSRCT-1 cell line was characterized by the coexpression of markers from various 3'scRNA-seq predefined clusters both at the transcriptional level (**Supplemental Information – Table S8**) and protein level, as assessed by immunofluorescence assays (e.g., DES, BAI1/ADGRB1, EGR1, TUBA1A, TNNT, **Supplemental Information – Fig. S6**), supporting that JN-DSRCT-1 is representative of DSRCT tumors and that it might retain pluripotency features. Moreover, when characterized by RNA-seq, the JN-DSRCT-1 cell line clustered within the DSRCT tumor samples among a series of 512 sarcomas, thereby reinforcing its relevance to be used as a preclinical DSRCT model (*data not shown*). To first identify the pathways regulated by *EWSR1-WT1*, we performed *EWSR1-WT1* ChIP-sequencing (ChIP-seq) using a WT1-Cter antibody in the JN-DSRCT-1 cell line. To ensure that wild-type WT1 was not expressed in this cell line and that all identified reads would specifically result from *EWSR1-WT1* transcriptional activity, we verified that WT1 5'-end reads were not detected (*data not shown*). In line with previous reports<sup>58</sup>, the majority of *EWSR1-WT1* target sequences were located within intergenic regions, followed by introns (**Supplemental Information – Fig. S7A**). ChIP-seq analysis evidenced some previously described *EWSR1-WT1* target genes, such as

*CCND1* or *FGFR4*, and revealed novel EWSR1-WT1 targets, such as *EGR1*, extracellular matrix-related genes (e.g., *COL23A1*, *CHI3L1*), and neural or neuroendocrine-related genes (e.g., *GAL*, *ADGRB1*) (**Fig. 3A, Supplemental Information – Table S9, S10**). Moreover, additional targets were identified and found to be related to DNA damage response (e.g., *UBE2V2*<sup>59</sup>), chromatin remodeling (e.g., *DPF1/2/3* which encode for components of BAF and PBAF chromatin remodelers<sup>60</sup>, *CTCF* which is involved in cancer cells' invasive phenotype<sup>61</sup> and androgen-response<sup>62</sup> on other cancer models), and fatty acids metabolic processes (e.g., *ACADVL*, *ECI2*).

Gene set enrichment analysis was further performed on EWSR1-WT1-specific peaks (ChIP-Enrich tool<sup>38</sup>, see Methods section) to recapitulate the functions of EWSR1-WT1 target genes. Peaks-based GO enrichment analysis revealed an enrichment in critical developmental processes including multilineage tissue development (e.g., GOBP\_embryonic heart tube development (adjusted p-value=5.59E-04, OR=1.59E+03), GOBP\_lymphangiogenesis (adjusted p-value=9.62E-04, OR=1.12E+04), GOBP\_skin epidermis development (adjusted p-value=1.94E-03, OR=9.20E+02), GOBP\_spinal cord motor neuron differentiation (adjusted p-value=3.01E-02, OR=1.78E+03), stem cell differentiation and proliferation (GOBP\_stem cell differentiation (adjusted p-value=1.94E-03, OR=9.84E+02), GOBP\_stem cell proliferation (adjusted p-value=5.59E-03, OR=8.54E+02)) and regulation of fatty acids metabolic processes (GOBP\_fatty acyl-CoA metabolic process (adjusted p-value=6.62E-03, OR=9.44E+02), GOBP\_sphingolipid biosynthetic process (adjusted p-value=2.86E-02, OR=3.31E+02)) (**Fig. 3B, Supplemental Information – Table S11**).

Altogether, EWSR1-WT1 ChIP-seq analysis confirmed a role for EWSR1-WT1 in activating lineage-related transcriptional programs and identified novel targets related to chromatin remodeling, extracellular matrix formation, and fatty acids metabolic processes.

The binding activity of TFs may vary according to the recruitment of variable coregulators and their cognate DNA sequence. EWSR1-WT1 binding motif is still not fully characterized and varies according to the presence of the KTS aminoacids at the breakpoint site. Indeed, previous studies (using ChIP, promotor-reporter assays, or

electrophoretic-shift motility assays) have reported that EWSR1-WT1 (-KTS) recognizes a GC-rich sequence<sup>56,63–66</sup>. By contrast, the EWSR1-WT1 (+KTS) isoform optimal binding site was identified as 5'-GGAGG(A/G)-3'<sup>67</sup>. Motif enrichment analysis within peaks that are significantly enriched in EWSR1-WT1 ChIP showed an overrepresentation of TFs from EGR, ZNF, KLF, FOX, E2F, ETS, and HOX families (**Supplemental Information – Fig. S7B, Table S12**). Notably, the most enriched motif corresponded to EGR1 DNA-binding sequence (**Fig. 3C**).

*De novo* motifs discovery, further compared with known TF motifs, highlighted eight significantly enriched *de novo* motifs within EWSR1-WT1-specifically targeted regions (**Supplemental Information – Fig. S7C, Table S13**). The 5'-GCGKGGGAGGVRGV-3', 5'-CCACGCA-3', and 5'-GGAGGAGRAGGAGGAA-3' motifs ranked as the top 3 most enriched *de novo* motifs, respectively best matching with EGR1, EGR2, and ZNF263 TF motifs. Interestingly, WT1 is part of the EGR TF family, and previous publications showed that WT1 binding sites are closely related to EGR1 consensus binding sites<sup>68</sup>, although the sequence of highest affinity binding sites for both proteins was distinct<sup>69–72</sup>. Notably, 6-mers overrepresentation analysis among EWSR1-WT1 ChIP significantly enriched peaks further confirmed that EWSR1-WT1 targeted sequences were enriched in GGA/CCT repeats (**Supplemental Information – Fig. S7D**), in line with EWSR1-WT1 (+KTS) isoform optimal DNA-binding sequence<sup>73</sup>. Overall, EWSR1-WT1 appears to target several variable DNA sequences, including EGR-targeted sequences and GGA/CCT microsatellites, as exemplified in FGFR4 targeted sequence (**Supplemental Information – Fig. S7E**).

To gain functional insight into the effect of EWSR1-WT1 expression on chromatin accessibility, we systematically characterized open chromatin regions using ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) in JN-DSRCT-1 cell line in which we modulated EWSR1-WT1 expression by small interfering RNA (siRNA). We identified 164,043 transposase-accessible DNA elements significantly modulated by EWSR1-WT1 expression level, including 83,524 and 80,519 sites showing decreased (EWSR1-WT1 *ON* ATAC peaks) and increased (EWSR1-WT1 *OFF* ATAC peaks) accessibility upon EWSR1-WT1 silencing, respectively (**Fig. 3D, Supplemental Information – Table S14**). Our experiment was validated by visualizing some genomic regions showing increased accessibility in the presence of EWSR1-WT1 and encoding known EWSR1-WT1 target genes, such as *CCND1*,

*FGFR4*, and *ASCL1*. Other genomic regions whose accessibility was increased when EWSR1-WT1 is expressed encoded transcripts that were identified as top marker genes of predefined scRNA-seq clusters (e.g., *GAL*, *CCL25*, *BAI1*) or involved in the shaping of tumor microenvironment (e.g., *CCL25*, *KRT23* or *ANGPT1*). We also identified several chromatin remodelers or epigenetic modifying enzymes (e.g., *TET2*, *SMARCA2*, *FOXA1*, *KLF4*, and *KLF6*), potentially suggesting a cooperating role of EWSR1-WT1 in driving DSRCT epigenome reprogramming. Interestingly, *AR* was one of the top genes showing increased accessibility when EWSR1-WT1 is expressed, suggesting a positive feedback loop between EWSR1-WT1 and AR signaling. Interestingly, FOXA1/2/3 TFs, whose target sequences are putative binding sites for EWSR1-WT1 (**Fig. 3C, Supplemental Information – Table S12**), are known to activate the transcription of AR target genes<sup>74,75</sup>, potentially bridging the gap between EWSR1-WT1 activity and AR pathway upregulation. We also identified several genomic regions that became increasingly accessible upon EWSR1-WT1 silencing, including some genes involved in mesenchymal stromal cells' phenotype (e.g., *FBN1*, *ALCAM*, and *PDPN*). No modification in wild-type *WT1* accessibility was noted, suggesting that the putative dominant-negative effect of EWSR1-WT1 on wild-type *WT1* did not occur at the chromatin level.

To investigate whether the effects of EWSR1-WT1 in rewiring transcriptional programs were primarily direct or indirect, we sought to explore which genomic regions showing differential accessibility upon EWSR1-WT1 silencing were also direct targets of EWSR1-WT1 by ChIP-seq. Amongst the 1,318 EWSR1-WT1 ChIP enriched peaks, we identified 944 overlapping genomic regions with EWSR1-WT1 ON ATAC peaks (**Fig. 3E**), showing an association between EWSR1-WT1 binding sites and their chromatin accessibility. Those 944 overlapping peaks mapped to 857 unique genes, of which 262 were significantly downregulated upon EWSR1-WT1 silencing (**Supplemental Information – Table S15, S16**), related to (1) GOBP\_Tissue morphogenesis (adjusted p-value=1.01E-02), (2) multilineage cells development (e.g., GOBP\_Epithelium development (adjusted p-value= 2.61E-03), GOBP\_Neuron development (adjusted p-value=3.61E-03), GOBP\_Embryonic heart tube development (adjusted p-value= 5.66E-03), GOBP\_Endothelium development (adjusted p-value= 4.71E-02)), (3) Wnt signaling (GOBP\_Wnt signaling pathway (adjusted p-value= 1.61E-02), GOBP\_Canonical Wnt signaling pathway (adjusted p-value= 4.91E-02))

and (4) GOBP\_Cell-matrix adhesion (adjusted p-value= 4.83E-02) according to ToppFUN enrichment analysis (**Supplemental Information – Table S16**). Moreover, motifs enrichment exploration within EWSR1-WT1 ON ATAC peaks showed a partial correlation with enriched motifs from EWSR1-WT1 targeted regions from ChIP-seq analysis, including FOX and EGR TFs motifs, further reinforcing the chromatin opening role of EWSR1-WT1 at these sites (**Supplemental Information – Table S17**).

Finally, to evaluate the level of activation of EWSR1-WT1 target genes, we performed ChIP-seq studies for the acetylation of H3K27 (H3K27ac) and H3K9 (H3K9ac) - histone marks primarily associated with active enhancers and promoters, respectively. Most EWSR1-WT1 binding sites were associated with H3K27ac and H3K9ac binding, as well as increased accessible chromatin regions, suggesting that EWSR1-WT1 occupancy is mainly associated with active transcriptional states at enhancers and promoters (**Fig. 3F**).

### **Bulk and single-cell level multiomics integration reveals some EWSR1-WT1-dependent cell-intrinsic plasticity**

To explore whether a potential EWSR1-WT1 binding versatility or chromatin accessibility variations may contribute to DSRCT cells' observed heterogeneity, we further integrated ChIP-seq and ATAC-seq results on JN-DSRCT-1 cell line, together with the previously described 3'scRNA-seq dataset, and a parallel single nucleus RNA-seq and ATAC-seq profiling on one DSRCT patient tumor (snMultiome).

To assess whether differential EWSR1-WT1 binding activity correlated with cellular states' heterogeneity in DSRCT, we further defined an EWSR1-WT1 ChIP-seq signature, constituted of 176 genes specifically directly targeted by EWSR1-WT1 (see Methods section; **Fig. 3G, Supplemental Information – Table S18**). Among these, 103 corresponded to features detected in the "Int\_sc" dataset. As a confirmation of this signature's specificity, its score was higher in tumor cells and almost undetectable in tumor microenvironment (TME) cells. We further found that EWSR1-WT1 ChIP-seq signature score was heterogeneous across DSRCT tumor cells - the highest scores being observed in *Neural/neuronal - 1* and *Mesenchymal\_secretion - 11* clusters, altogether suggesting that variable EWSR1-WT1 TF activity may participate to DSRCT tumor cell heterogeneity. Notably, the lowest score among DSRCT tumor cells was

observed for the *Metabolic\_glycolysis – 5* cluster, suggesting that this cluster might be driven by EWSR1-WT1 activity-independent mechanisms (**Fig. 3G**).

Since ChIP-seq and ATAC-seq data generated from the JN-DSRCT-1 cell line suggested some EWSR1-WT1-dependent modifications of the chromatin landscape, we sought to explore the heterogeneity of chromatin accessibility on DSRCT tumors at the single-cell level and link it to transcriptional programs. To do so, we profiled one fresh DSRCT tumor sample using the snMultiome assay (10X Genomics), which allows the integration of the 3' gene expression and chromatin accessibility profile within the same single nucleus. Weighted-nearest neighbors (WNN) clustering on ATAC and gene expression features allowed us to identify 13 clusters (**Fig. 3 H, Supplemental Information – Fig. S8A, B**). When taken separately, ATAC features-based clustering on the one hand, and gene expression-based clustering on the other hand, imperfectly recapitulated DSRCT tumor cell heterogeneity landscape drawn with bimodal WNN clustering (**Supplemental Information – Fig. S8A, C**), demonstrating that the separate characterizations of DSRCT cells transcriptome and chromatin accessibility are not sufficient themselves to recapitulate DSRCT cell heterogeneity. Differential gene expression and differential peak enrichment analyses did not allow to confidently identify the previously described cell subpopulations or DSRCT cell states, except for *WNN\_cluster 9*, *WNN\_cluster 4*, and *WNN\_cluster 3/11/12*, for which top marker genes/peaks were representative of non-malignant cells, mitochondrial genes enriched cluster, and cycling cells, respectively (**Supplemental Information – Table S19, S20, S21**). We further sought to perform a correlation between snMultiome WNN clustering and previously identified “*Int\_sc*” dataset clusters. Because our previous single-cell profiles had been generated based on whole cell-derived transcripts rather than nuclear transcripts only, we speculated that the exclusive analysis of 3'scRNA-seq intronic reads, which supposedly derive from unprocessed nuclear transcripts, would better correlate with single nuclei gene expression from the snMultiome assay. A label transfer of the snMultiome dataset on the Harmony integrated 3'scRNA-seq dataset reference exclusively incorporating intronic reads (“*Int\_sc\_intron*” dataset) allowed uncovering correlations between cells' chromatin accessibility states and previously defined “*Int\_sc*” clusters. While several WNN clusters were tightly correlated to a single “*Int\_sc*” cluster (e.g., *WNN\_cluster 3/11/12* and *Cycling cells – 3*, *WNN\_cluster 5* and *Metabolic\_glycolysis – 5*, *WNN\_cluster 10* and

*Mesenchymal\_fibrosis* – 8), other pairwise correspondences were less straightforward. For example, cells composing *WNN\_cluster 0* were mainly assigned to *Multidiff\_ASCL1* – 0, *Ribosomal\_catabolic* – 2, and *Epithelial\_mesenchymal* – 4 clusters, while *WNN\_cluster 1* and 2 were both preponderantly allocated to the *Epithelial\_Mesenchymal* – 4 cluster (**Fig. 3I**). Notably, the non-malignant nature of *WNN\_cluster 9* was further confirmed, as *WNN\_cluster 9* cells were assigned to *Myeloid cells* – 9, *CAFs\_C1* – 10, *T cells* - 12, *Mesothelial cells* – 14, and *Vascular endothelial cells* – 15 clusters. Interestingly, a small subset of cells from tumor cell WNN clusters, except *WNN\_cluster 10*, 11, and 12, were assigned to the *Inc\_2* – 17 cluster, again questioning its relevance as a biological entity. Compelling concordant findings were observed using the matched *PZB* 3'scRNA-seq sample (**Supplemental Information – Fig. S8D, E**) as a reference for label transfer (**Supplemental Information – Fig. S8F**).

Interrogating the level of chromatin accessibility across clusters, we found out that the number of detected peaks per cell was highly variable across clusters (**Supplemental Information – Fig. S9A, left panel**), the highest number of peaks being called in *WNN\_cluster 12*, and the lowest in *WNN\_cluster 0*. This observation suggests that global chromatin accessibility is variable among DSRCT cell clusters, some of which, like *WNN\_cluster 0*, might be marked by a predominant closed chromatin state.

We analyzed overlapping peaks between *EWSR1-WT1 ON* ATAC peaks or *EWSR1-WT1* ChIP enriched peaks and snMultiome peaks called by cluster. Strikingly, *WNN\_cluster 9*, which encompasses several non-malignant microenvironment cells, showed the lowest ratio of overlapping peaks with both assays, reinforcing that DSRCT cells are characterized by a unique epigenetic landscape driven by *EWSR1-WT1* chimeric protein (**Supplemental Information – Fig. S9A, middle and right panel**).

Interestingly, the analysis of *EWSR1-WT1 ON* and *EWSR1-WT1 OFF* ATAC modules in the snMultiome dataset showed heterogeneity of modules' scores among cells. As an internal control, the non-malignant cells displayed the lowest and the highest scores for *EWSR1-WT1 ON* and *EWSR1-WT1 OFF* ATAC modules, respectively. In addition, *WNN\_cluster 5*, which was labeled as the anaerobic glycolysis pathway-enriched cluster, did display the lowest score of *EWSR1-WT1 ON* ATAC module (**Supplemental Information – Fig. S9B and S9C**). Parallely, the analysis of *EWSR1-*

WT1 ChIP score within the “*Int\_sc*” 3’scRNA-seq dataset also showed a lower score for this cluster. Altogether, these data suggest that the EWSR1-WT1-associated chromatin landscape is heterogeneous among DSRCT cells and that the induction of a metabolic switch towards anaerobic glycolysis may associated with decreased EWSR1-WT1 activity.

Since we hypothesized that chromatin plasticity may participate in DSRCT cell states’ heterogeneity, we further explored whether the top differentially accessible peaks from each snMultiome WNN cluster were enriched in specific motifs.

First, while comparing snMultiome malignant versus non-malignant cells, malignant cells’ marker peaks displayed a highly significant enrichment in EWSR1-WT1 binding motifs such as EGR1, EGR3, SP2, ZNF263, further confirming that EWSR1-WT1 binding is associated to increased accessibility at these sites (**Fig. 3J, Supplemental Information – Fig. S9A, middle and right panel**). Interestingly, RNA-seq data from +/- EWSR1-WT1-silenced JN-DSRCT-1 cells showed that EGR1, and at a lesser extent, EGR2 and EGR3 were downregulated by EWSR1-WT1 (**Supplemental Information – Table S15**). At the same time, EGR1 promoter is directly targeted by the fusion transcript (**Fig. 3A**). These data suggest that EWSR1-WT1 and EGR1 may have competitive transcriptional activity through their binding to common DNA sequences and that EWSR1-WT1 may dominate this competition by inducing EGR1 downregulation.

Secondly, using ChromVAR <sup>76</sup> motif activity inference tool, we highlighted per-cell motif activities across the 10 most enriched motifs within predefined WNN clusters’ differentially accessible peaks (**Fig. 3K, Supplemental Information – Table S23**). For example, *WNN\_cluster 5*, assigned to *Metabolic\_glycolysis – 5*, was mainly characterized by enrichment in FOS/JUN-related motifs, whereas *WNN\_cluster 6* displayed a clear enrichment in GATA-related motifs. Significantly, three main groups of developmental TF families-related motifs were differentially enriched in distinct WNN clusters. SOX (e.g., SOX8, SOX13, SOX9, SOX2, SOX4), FOX (e.g., FOXA1, FOXD2, FOXB1, FOXE1, FOXC1, FOXA3) and GATA (e.g., GATA6, GATA5, GATA2, GATA3) TFs motifs activities were significantly elevated in *WNN\_cluster 2*, *WNN\_cluster 5*, and *WNN\_cluster 6*, respectively, suggesting that accessibility at SOX, FOX, and GATA binding sites shapes DSRCT cell states. Interestingly, *WNN\_cluster 10*, which was previously related to mesenchymal profibrotic tumor cells (*Mesenchymal\_fibrosis – 8*

cluster), displayed increased accessibility in IRF (Interferon Response Factor) TFs motifs, supporting that extrinsic microenvironmental chemokine signaling may drive the latter phenotype. Finally, the Androgen Receptor (AR) motif ranked within the top 10 enriched motifs in *WNN\_cluster 1*.

The inference of enhancer-driven gene regulatory networks from the snMultiome dataset using SCENIC+ partially correlated with the above-described original motifs activity inference. We identified 337 active TFs, among which 253 were activating TFs (i.e., associated with increased chromatin accessibility) targeting 21,695 regions corresponding to 17,229 genes, and 84 were repressors (i.e., associated with decreased chromatin accessibility) targeting 2,722 regions mapping to 2,395 genes. Among activating TFs-driven regulatory networks were notably identified AR, confirming its role in DSRCT oncogenicity <sup>77</sup>. Notably, the FOX family TFs regulatory network activation was shared among most DSRCT tumor cell clusters, in line with EWSR1-WT1 binding on FOX TFs targeting sequences. Moreover, the activity of several regulatory networks was specifically increased in individual WNN clusters. For example, in line with previously described motif activity inference, SOX2 regulon was primarily activated in *WNN\_cluster 2*, supporting the presence of pluripotency and neural features <sup>78</sup>, although this cluster had been associated with *Epithelial\_mesenchymal -4* cluster using the label transfer method. In addition, glycolysis-activated tumor cells (*WNN\_cluster 5*) displayed upregulated activity of stress-related regulatory network including FOS and ATF1-7, as well as regulons involved in epithelial-to-mesenchymal (EMT) transition and metastasis (e.g., BACH1 <sup>79</sup>), which is also known for its ability to restrict dependency on oxidative phosphorylation and promote aerobic glycolysis <sup>80,81</sup>. In addition, ETS family-related (e.g., ETS1, ETV5, FLI1) and ERG/GATA4 regulons were preferentially enriched in mesenchymal profibrotic *WNN\_cluster 10* and *WNN\_cluster 6*, respectively. Most significantly, we observed a decoupling between gene activation and chromatin accessibility for both WT1 and EGR1 regulatory networks, confirming that EWSR1-WT1 may exert a dominant negative effect through a competitive antagonism at WT1 and EGR1 targeted loci (**Fig. 3L**).

Next, hypothesizing that EWSR1-WT1 DNA binding variability may participate in DSRCT tumor cell heterogeneity, we studied the activity of EWSR1-WT1 *de novo* ChIP-inferred motifs within the snMultiome dataset. Interestingly, the latter showed a

patent differential enrichment in EWSR1-WT1 *de novo* motifs across clusters (**Supplemental Information – Fig. S9D**). Most *de novo* EWSR1-WT1 targeted motifs were mainly enriched in cycling cell clusters. On the other hand, *de novo* motif m3\_AARTAAAYA, which was related to FOXC2 TF, was mainly enriched in *WNN\_cluster 5*, which correlated to *Metabolic\_glycolysis - 5* cluster. Strikingly, FOXC2, previously known as Mesenchyme Forkhead 1 (MFH-1), is known to be a key player of mesenchymal cell fate during embryogenesis<sup>82</sup> and is notably involved in epithelial-mesenchymal transitions (EMTs) in breast cancer<sup>83</sup>, further reinforcing the hypothesis that cells activating anaerobic glycolysis pathway would also enter a mesenchymal differentiation trajectory.

Altogether, we identify that DSRCT cell states are characterized by distinct chromatin accessibility landscapes, partly driven by EWSR1-WT1 variable DNA targeting and remodeling. We highlight complementary mechanisms through which EWSR1-WT1 induces a profound transcriptomic and epigenomic rewiring, comprising (i) a potential EWSR1-WT1 and EGR1 competitive antagonism where EWSR1-WT1 binds on GC-rich DNA sequences and EGR1 binding motifs while inducing EGR1 transcriptional downregulation, and (ii) the binding on GGA microsatellites where it acts as a potent transcriptional activator.

While our data primarily support that the various DSRCT cell states may result from intrinsic mechanisms notably driven by EWSR1-WT1 versatile binding sites, we further aimed at elucidating the potential interplay between DSRCT cells and their microenvironment. Relying on the hypothesis that DSRCT cell heterogeneity could also be triggered by *in vivo* extrinsic stimuli, we further dissected DSRCT microenvironment cell subpopulations and characterized the consequences of such interactions on DSRCT cell states and *in vitro* proliferation capacity.

### **DSRCT microenvironment displays immune-tolerant features**

While focusing on immune cells, we found out that myeloid cells were the major immune subpopulation composing DSRCT tumors' microenvironment, according to both "*Int\_sc*" dataset clustering and DSRCT bulk RNA-seq cell type deconvolution (**Fig. 1D, F**). To better characterize the functions of these DSRCT myeloid cells, we performed a deconvolution of this myeloid population using the recently described

MoMac-VERSE atlas<sup>84</sup>, which unifies dendritic cells, monocytes and macrophage subpopulations across human tissues. DSRCT infiltrating myeloid cells were mainly composed of macrophages and a minority of monocytes and dendritic cells (DCs) (**Fig. 4A, Supplemental Information – Table S24**). The main macrophage subtypes were TREM2, FTL, HES1, and RGS1/DLA-DRB1/HL1-DPB1-positive macrophages. Interestingly, TREM2 macrophages have been shown to accumulate in human tumors, display protumorigenic properties, and correlate with exhausted CD8+ tumor-infiltrating lymphocytes<sup>85–87</sup>. In addition, HES1 and FTL overexpression mainly induces anti-inflammatory pathways and directs macrophages towards an M2 phenotype<sup>88–90</sup>. As for monocytes, the latter were mainly represented by IL1B monocytes.

We further aimed at profiling DSRCT T cell subpopulations by exploring their canonical markers<sup>91</sup> and performing differential gene expression analysis after Seurat subclustering (**Fig. 4B, Supplemental Information – Table S24**). A total of 7 clusters was defined, represented by a majority of CD4+ memory T cells, overexpressing *CD4*, *CD40LG*, *CD44*, and *IL7R* (*CD4+ memory T cells – 0* cluster). The second main T cells subtype was CD8+ cytotoxic effector T cells (*CD8A+*, *GZMB+*, *GZMK+*, *GZMA+*, *GZMH+*, *CD8+ effector T cells – 1* cluster). Additional clusters of interest comprised a Natural Killer (NK) cells population overexpressing *NKG7*, *GZMB* and *GZMH* (*Activated NK cells – 3* cluster), a mast cells cluster overexpressing *TPSAB1*, *TPSB2*, and *CPA3* (*Mast cells – 4* cluster), *MKI67+* proliferating T cells (*Proliferating CD8+ T cells – 6* cluster), and a cluster of T cells overexpressing translational machinery involved transcripts such as ribosomal proteins genes, and *TXNIP* (*Translation high T cells – 2* cluster) (**Fig. 4B, Supplemental Information – Table S24**). T cells' expression profile showed an absence or low expression of immune checkpoints such as *PDCD1* (PD-1), *HAVCR2* (TIM-3), and *LAG3*, suggesting that no exhausted T cells could be identified in DSRCT samples. Similarly, markers of naïve like CD8+ T cells (*TCF7*, *CCR7*, *SELL*, *LEF1*) and regulatory T cells (Tregs) markers (*FOXP3*) were not found in our dataset.

Surprisingly, the inference of immune scores using xCell<sup>31</sup> on our cohort of bulk RNA-seq DSRCT samples (**Supplemental Information – Fig. S10A**) showed some discrepancies with 3'scRNA-seq-derived results, showing higher scores for activated myeloid dendritic cells than for M2 macrophages. In addition, while no T helper cell markers were detected (e.g., *IL5*, *IL13*, *MAD2L1*, *BAG2*) in our 3'scRNA-seq dataset,

xCell deconvolution highlighted higher scores for T cells CD4+ Th1, T cells CD4+ Th2, and NK cells signatures than for CD8+ subpopulations (**Supplemental Information – Fig. S10B**).

Overall, DSRCT immune cells' characterization suggests a predominant infiltration of immunotolerant protumorigenic macrophages, with low infiltration of proficient effector cytotoxic CD8+ T cells.

### **DSRCT Cancer-associated Fibroblasts (CAFs) are composed of distinct subclusters**

The initial Seurat clustering performed on all cells distinguished two CAF clusters (*CAFs\_c1 – 10* and *CAFs\_c2 – 13* clusters, **Supplemental Information – Fig. S1E, Table S3**). The *CAFs\_c1 – 10* cluster mainly stemmed from tumor samples, while *CAFs\_c1 – 13* was preponderantly found in the juxta-tumor peritoneal sample *BOM site#1* (**Fig. 1C, Fig. 4C, left panel**). Gene Ontology enrichment analysis highlighted that CAFs from *CAFs\_c1 – 10* cluster were mainly involved in extracellular matrix formation (GOBP\_Collagen fibril organization, adjusted p-value=0.0019; GOBP\_Connective tissue development, adjusted p-value=0.0047) and angiogenesis (GOBP\_Vasculature development, adjusted p-value=0.0089), while CAFs from *CAFs\_c2 – 13* displayed significant enrichment for pathways involved in cell migration and adhesion (GOBP\_Cell migration, adjusted p-value=0.008; GOBP\_Biological adhesion, adjusted p-value=0.038) and complement activation (GOBP\_Complement activation, adjusted p-value=0.0031) (**Supplemental Information – Fig. S11A**). To refine the distinct CAF subpopulations composing DSRCT tumor microenvironment, we performed CAF subclustering from merged *CAFs\_c1 – 10* and *CAFs\_c2 – 13* clusters and defined eight distinct CAF subtypes (**Fig. 4C, right panel**) that were manually annotated based on the top marker genes from differential gene expression analysis: *CD34/MYC fibroblasts – 0*, *Desmoplastic CAFs – 1*, *Undetermined CAFs – 2*, *Lipofibroblasts – 3*, *Myofibroblastic CAFs – 4*, *Pericytes-like CAFs – 5*, *Inflammatory CAFs – 6*, *Adventitial fibroblasts – 7* (**Fig. 4C, E, H, Supplemental Information – Table S25**).

CIBERSORTx deconvolution on DSRCT tumors bulk RNA-seq data confirmed the predominance of *Desmoplastic CAFs – 1* and *Myofibroblastic CAFs – 4* while the proportion of juxta-tumor CAFs cells from *CD34/MYC fibroblasts – 0* remained low (**Fig. 4D, Supplemental Information - Table S27**).

We next defined CAFs' canonical markers based on previously described CAF markers from other tumor types <sup>92-95</sup> (**Fig. 4 H, Supplemental Information – Fig. S11B**). While ACTA2 expression was shared by most CAF subclusters, its expression was minor within *Adventitial fibroblasts – 7*, and absent in *CD34/MYC fibroblasts – 0* and *Undetermined CAFs – 2*. Of note, the latter cluster could not be annotated confidently, and expressed markers such as *LTBP4*, replication-involved genes (*EIF4A1*, *EIF5*), and mitochondrial genes (*ATP5ME*, *ATP5MD*, *MT-ND1*, *MT-ATP6*, *MT-CO3*, *MT-CO2*). *FAP* expression was the highest in *Desmoplastic CAFs – 1*, and *MCAM* was specific to *Pericytes-like CAFs – 5*. The *Inflammatory CAFs – 6* cluster was characterized by the expression of different chemokines such as *LIF*, *CXCL2*, *CXCL8*, and *BDKRB1*, which is known to accumulate under inflammation and hypoxia in other tumor types <sup>96</sup> (**Fig. 4E, H, Supplemental Information – Table S25**).

Interestingly, CAF subclusters that were mostly found within the juxta-tumor sample (i.e., *CD34/MYC fibroblasts – 0*, *Lipofibroblasts – 3*, *Inflammatory CAFs – 6*, *Adventitial fibroblasts – 7*) all showed complement pathways' gene activation (i.e., *C3*, *C7*, *C1R*), and expressed the constitutive lipid droplet protein *PLIN2*. Of note, *CD34/MYC fibroblasts – 0* also showed positivity for *APOD*, *CXCL14*, *PPARG*, and *MGP*, considered pre-adipocytes or adipocyte stem cell markers <sup>97</sup>. Intriguingly, the evaluation of MYC protein expression by immunohistochemistry on DSRCT FFPE samples juxta-tumor areas showed a clear expression pattern in nuclei of adipocyte-like cells localized at the edge of the tumor (**Supplemental Information – Fig. S11C**). In addition, *Adventitial fibroblasts – 7* showed overexpression of *SCARA5*, *CD55*, *SLPI*, *PI16*, *DPP4*, and *CD34* (**Fig. 4E, F, Supplemental Information – Table S25**), which is similar to adventitial stromal cells' expression profile <sup>98</sup>, known to be found in vascular niches and able to produce extracellular matrix <sup>99</sup>. Interestingly, *SCARA5* is known to play an essential role in the commitment of mesenchymal stem cells to adipogenesis <sup>100</sup>, and subcutaneous *DPP4+* mesenchymal cells have been shown to constitute progenitor cells that can differentiate into adipocytes <sup>99</sup>. In line with these observations, we may speculate that *Adventitial fibroblasts – 7* cells may be precursors to adipocyte-like *CD34/MYC fibroblasts – 0* cells.

Next, focusing on the profibrotic and immunoregulatory properties of CAFs <sup>101</sup>, we further found that the expression of desmoplasia-related genes (e.g., *MMP2*, *COL1A1*, *LOX*, *LOXL1*, *VEGFB*) was shared by all CAF subclusters, markedly by *Desmoplastic CAFs – 1* (**Fig. 4G**). Moreover, significant immunosuppressive response-related genes

were overexpressed by CAF subsets. While most CAF subclusters, excepted *Undetermined CAFs - 2* and *Pericytes-like CAFs - 5*, expressed T cells' exclusion features like *MRC2*<sup>102</sup> and M2 macrophages polarization signaling (e.g., *CSF1*<sup>103,104</sup>, *CXCL12*<sup>105</sup>, *IL6*<sup>106</sup>), *Inflammatory CAFs - 6* also displayed strong activation of genes involved in Myeloid-Derived Suppressor Cells (MDSCs) recruitment<sup>107</sup> including *CCL2*, *CXCL2*, and *CXCL8* (**Fig. 4G**).

The application of breast cancer CAFs' signatures<sup>92,93</sup> showed that most CAF subclusters were enriched in the CAF-S1 signature, while only *Pericytes-like CAFs - 5* were enriched in the CAF-S4 signature (**Supplemental Information - Fig. S11D**), herein stressing out the overall immunosuppressive feature of DSRCT microenvironment<sup>92,93,108</sup>. Of note, the CAF-S3 signature was mainly expressed by *CD34/MYC fibroblasts - 0*, consistent with their preponderance in juxta-tumor tissue.

In order to validate our findings at the protein level and assess whether these distinct CAF subpopulations were specifically found in particular DSRCT tumor areas, we developed a fluorescent triplex IHC staining to characterize MCAM, FAP, and ACTA2 (alias  $\alpha$ SMA) protein expression on eight DSRCT samples. We analyzed the fluorescence intensities across three distinct stromal areas: Stromal area n°1 relating to large trabeculae of desmoplastic stroma, Stromal area n°2 characterizing stroma located at the near periphery of tumor nests, and Stromal area n°3 located at the tumor edge within the pseudocapsule (**Fig. 4 H**). While MCAM+ CAFs were located near vessels supporting pericyte-like features, there was a correlation between FAP/ACTA2 fluorescence intensity and CAFs localization within the tumor (one-way ANOVA  $p=0.0092$ ) (**Fig. 4 H, I**). FAP+ CAFs were preferentially located within the pseudocapsule (paired t-test  $p=0.0093$  and  $p=0.0148$ ). Although non-significant ( $p=0.6177$ ), we observed a tendency of FAP+ CAFs accumulation within the large trabeculae of desmoplastic stroma, compared to AML+/FAP-/MCAM- CAFs, which tended to be mostly found at the proximity of tumor cells' bundles (**Fig. 4 H, I**).

### **DSRCT microenvironment cell proximity is associated with specific tumor cell phenotypical and metabolic states, suggesting extrinsic plasticity**

While our previous findings support that DSRCT tumor cells may intrinsically retain pluripotency features and lineage plasticity potential, we further interrogated whether

microenvironment signals may trigger tumor cells to proceed along specific cell states' trajectories.

Focusing on DSRCT tumor cells with mesenchymal features characterized by the expression of *DES*, *CHI3L1*, or *TNNT3*, we performed a triplex immunofluorescent staining on 3'scRNA-seq matched FFPE samples to identify any spatial restriction towards *THY1+* CAFs proximity (**Fig. 5A**). While some DSRCT samples displayed a homogeneous *DES* staining by virtually all tumor cells, several DSRCT samples (e.g., *GR2*) were characterized by a *DES/CHI3L1* expression restricted to either the periphery of tumor cells' islets at the direct proximity of CAFs (**Fig. 5A, bottom left panel**), or at the invasive tumor front within desmoplastic stroma area (**Fig. 5A, bottom right panel**). This *DES+/CHI3L1+* mesenchymal profibrotic DSRCT tumor cell subpopulation likely corresponds to *GR2 Mesenchymal\_TNNT3 – 5*, *Mesenchymal\_DES – 9*, *Neuronal\_mesenchymal – 13*, and *Mesenchymal\_stem – 14* clusters (**Supplemental Information - Table S4**) in which *CHI3L1* or *DES* were significantly overexpressed (**Fig. 5B**).

Interestingly, while JN-DSRCT-1 cells grown in 2D showed lineage-related markers homogeneous stainings (**Supplemental Information – Fig. S6**), JN-DSRCT-1 cells grown in clones or spheroids showed a sharply increased *DES* signal intensity at the periphery compared to core tumor cells on IF assays (**Supplemental Information – Fig. S12**). This functional experiment suggests that interactions between DSRCT tumor cells and non-cellular components of the microenvironment, such as the extracellular matrix, may be sufficient to control some DSRCT cell state trajectories.

To better characterize the potential cell-to-cell interactions between DSRCT tumor cells and their microenvironment, we took advantage of CellPhone DB tool <sup>21</sup>, which infers cell-cell communication from the combined expression of multi-subunit ligand-receptor complexes. We analyzed DSRCT tumor cell clusters versus CAF interactions in the three DSRCT samples that contained the most significant number of CAFs (**Fig. 5C, Supplemental Information – Fig. S13**). CellPhone DB analyses revealed that CAFs interacted the most with *CHI3L1/TNNT3/DES* overexpressing tumor cells. While most of the interactions between CAFs and mesenchymal tumor cells were represented by adhesion ligand-receptor couples (e.g., *COL20A1-a1b1/a11b1*, *FN-a5b1/aVb1/a11b1/aVb5*), proangiogenic (e.g., *FLT1-VEGFA/B*), protumorigenic (e.g., *EGFR-MIF/GRN/COPA/AREG*), and immunoregulatory (e.g., *TGFB1/TRFBR1/2*) interactions were also identified (**Supplemental Information – Fig. S13**).

To clarify the relationship between tumor cells' phenotype and DSRCT tumors' spatial morphology, we performed spatial transcriptomics on 3'scRNA-seq matching samples (**Fig. 5D, E, F, G, Supplemental Information Fig. S14**). The Visium spatial transcriptomics technique allows the study of the spatial whole transcriptome at ~50  $\mu\text{m}$  resolution. After spot clustering, we performed a differential gene expression analysis across identified spot clusters (**Fig. 5E, Supplemental Information - Table S26**). Focusing on the previously described *GR2* sample, we identified specific expression patterns for both tumor and stromal components' related clusters. Notably, *GR2 Spatial clusters 4* and *5* colocalized at the edge of tumor cells' islets. They respectively expressed myofibroblastic CAF-related markers (e.g., *SPARC*, *ACTA2*, *COL1A1*) and tumor cell mesenchymal/anaerobic glycolysis-related markers (e.g., *ENO1*, *CHI3L1*, *VEGFA*) (**Fig. 5D, F**). Strikingly, this pattern was consistent across most analyzed Visium DSRCT samples (**Supplemental Information - Fig. S14, Table S26**). A label transfer performed on matched Visium spatial transcriptomics assays and 3'scRNA-seq datasets used as a reference further confirmed that these spots located at the edge of tumor cells' islets and in direct contact with the stroma corresponded to 3'scRNA-seq clusters showing either mesenchymal, anaerobic glycolysis or ribosomal genes enrichment, respectively labeled as *GR2 Mesenchymal\_TNNT3 – 5*, *GR2 Metabo\_glycolysis – 7*, and *GR2 Ribosomal – 2* clusters (**Supplemental Information – Table S4, Fig. S14**).

Moreover, the application of Hypoxia, Glycolysis, and Oxidative Phosphorylation HALLMARK signatures, along with the previously described EWSR1-WT1 ChIP-seq signature on the *GR2* Visium sample, unveiled an anticorrelation between hypoxia and anaerobic glycolysis on the one hand, and EWSR1-WT1 TF activity and oxidative phosphorylation on the other hand. These findings strongly suggest that DSRCT tumor cells located at the center of the islets display the highest EWSR1-WT1 activity and proliferative capacity while mainly relying on oxidative phosphorylation. On the contrary, the upregulation of anaerobic glycolysis, identified in tumor cells located at the edge of tumor cells' islets, may constitute a metabolic switch mechanism exploited by both DSRCT mesenchymal tumor cells and CAFs in reaction to hypoxia-induced HIF-1 pathway upregulation.

Overall, these results provide clues about the mechanisms underlying previously undescribed DSRCT tumor cell metabolic state changes. Because several upregulated

genes in DSRCT mesenchymal tumor cells are also EWSR1-WT1 targets (e.g., *DES*, *CHI3L1*, *ACTA2*, **Supplemental Information – Table S9, S10**), we may posit that an epigenetic reprogramming, driven by EWSR1-WT1 lower TF activity, might be triggered by microenvironmental hypoxic stimuli, hence leading to the activation of a neural-mesenchymal transition program.

### Intertumor heterogeneity and prognostic significance of DSRCT plasticity

While the temporal dynamics governing DSRCT cell state shifts have not been explored *per se*, the study of the heterogeneity of both *GR4* and *GR4\_PC* samples provides information about potential tumor cell evolution under antitumor therapeutic pressure. Indeed, two sequential samples were collected from patient *GR4*, the first obtained in a chemotherapy naïve situation (*GR4*) and the second after adjuvant chemotherapy, first relapse, and first metastatic treatment line (*GR4\_PC*). Importantly, the complexity of DSRCT tumor cell heterogeneity regarding identified clusters was comparable between both samples (**Supplemental Information - Fig. S2**). On the contrary, the CNV inference study did display an increase in copy number alterations, characterized by the acquirement of a gain of 5p/5q, 7p/7q, 10p/10q, 20p/20q, 21p/21q, and 22p/22q, which were corroborated by the CNV analysis from bulk matched frozen tumor/normal samples. Notably, a chr 13p/13q loss was inferred from scRNA-seq data for the *GR4\_PC* sample but was not validated on WES. Conversely, additional gains on chr 1, 15, and 18 were detected on WES analysis but not identified using inferCNV.

“*Int\_sc*” clusters’ deconvolution on our cohort of DSRCT bulk RNA-seq samples highlighted that *Epithelial Mesenchymal – 4* and *Cycling cells – 3* were the most variably represented. Also, the *Neural stem – 19* cluster was undetectable in some samples, while it could reach 10% of the total subpopulations in others (**Fig. 6A**).

To further explore intertumor heterogeneity, we performed a hierarchical clustering using a distance based on Pearson's correlation coefficient with the Ward D2 linkage algorithm, exploiting the totality of sequenced genes on the cohort of 29 frozen DSRCT samples (**Fig. 6B, Supplemental Information - Fig. S10A, Table S27**). We did not identify apparent clustering according to the main recorded clinical data, including sex or tumor primary site. Based on this clustering, we identified three DSRCT patients’

clusters. Of note, the totality of chemotherapy-naïve patients was grouped together within Cluster #3, although this subgroup also included some pre-treated patients. DGE analysis between Cluster #3 and Cluster #1 or #2 notably showed an upregulation of genes involved in nucleosome assembly, chromatin remodeling, and cell cycle positive regulation (**Supplemental Information - Table S28, S29**). On the contrary, genes significantly enriched in Cluster #2 showed enrichment in (i) genes involved in epithelial cell differentiation, cell adhesion, extracellular matrix components, and response to hormone compared to Cluster #3 and (ii) genes involved in lymphocyte activation and antigen binding compared to Cluster #1 (**Fig. 6B, Supplemental Information – Table S28, S29**). Importantly, identified clusters exhibited a significant prognostic value on overall survival (**Fig. 6C**), with median overall survival of 24 months (95% CI: 16-NA), 27 months (95% CI: 16-NA), and 15 months 95% CI (12-NA) for Cluster #1, Cluster #2, and Cluster #3 respectively. Altogether, these findings support that both the activation of a lymphocyte immunostimulatory response through antigen binding and an epithelial-like phenotype are associated with improved DSRCT prognosis. Interestingly, the immunosuppressive features of the worst prognostic group (Cluster #3) were confirmed using immune xCell deconvolution, highlighting a higher CD4+ Th2 score (**Supplemental Information - Fig. S15**).

We next derived a gene signature characterizing each “*Int-sc*” cell cluster, defined by their top 100 marker genes (n=20). To first assess the specificity of the herein defined signatures, for each “*Int\_sc*” cluster, we generated a signature score by computing the arithmetic mean of the bulk RNA-seq-derived Variance Stabilizing Transformation (VST)-normalized raw counts expression from an external cohort of DSRCT and various sarcoma subtypes (**Fig. 6D, Supplemental Information - Fig. S16, Table S30**). Several DSRCT tumor cell cluster signatures (e.g., “*Int\_sc*” *Cycling cells* - 3, “*Int\_sc*” *Ribosomal\_catabolic* - 2) did not show any specificity to the DSRCT transcriptome. Notably, no microenvironment cell populations’ signature score was significantly increased in the DSRCT cohort compared to other sarcoma subtypes. Importantly, the signature scores for *Multidiff\_ASCL1* - 0, *Neural\_neuronal* - 1, *Epithelial\_Mesenchymal* - 4, *Inc\_1* - 7, *Mesenchymal\_fibrosis* - 8, and *Mesenchymal\_secretion* - 11 clusters were significantly higher within the DSRCT cohort.

We further evaluated the prognostic value of the above-described “*Int\_sc*” clusters’ signature scores to assess whether specific clusters could drive higher DSRCT tumor aggressivity, thus deteriorating patients’ overall survival.

To this end, bulk RNA-seq samples were stratified into two groups according to the signature optimal score cutpoint, thus splitting the cohort into two groups defining a “High” or “Low” signature score. A Kaplan Meier overall survival analysis was performed for each signature to evaluate each group's survival probability (**Supplemental Information - Fig. S17**). Notably, a higher score of several 3’ scRNA-seq-derived signatures was significantly associated with worse prognosis (e.g., *Ribosomal\_catabolic* – 2 ( $p=0.036$ ), *Cycling cells* – 3 ( $p=0.0058$ ), *Metabolic\_glycolysis* – 5 ( $p=0.03$ ), *Ribosomal\_IFN response* – 6 ( $p<0.0001$ ), *T cells* – 12 ( $p=0.021$ ), and “*Metabolic\_serine* – 15 ( $p=0.0045$ )), while higher scores of *Epithelial\_mesenchymal* - 4 ( $p=0.038$ ) and *Vascular endothelial cells* - 16 ( $p=0.042$ ) signatures were associated with improved prognosis (**Fig. 6E**). The prognostic significance of these signatures was further tested on an external validation cohort of 21 frozen DSRCT bulk RNA-seq samples. The prognostic value was validated for the *Epithelial\_mesenchymal* – 4 signature ( $p=0.00064$ ), and we observed a signal without reaching statistical significance for *Cycling cells* – 3 ( $p=0.1$ ) and *Ribosomal\_IFN response* – 6 ( $p=0.17$ ) signatures. Together with previous findings from bulk RNA-seq samples hierarchical clustering, these data support that the acquisition of an epithelial phenotype is related to improved prognosis in DSRCT.

## Discussion

This study shows that DSRCT displays some degree of intrinsic heterogeneity partly driven by EWSR1-WT1 activity and extrinsic heterogeneity influenced by interactions with the tumor microenvironment. We could identify three main components of this heterogeneity: (i) lineage plasticity, characterized by highly overlapping differentiation-related phenotypes; (ii) metabolic switches between oxidative phosphorylation, anaerobic glycolysis, and serine metabolism activation; and (iii) pseudostates. Our data support that DSRCT cancer cells permeate through a continuous spectrum of transcriptional states/pseudostates linked by dynamic and reversible transitions rather than follow unidirectional trajectories or clonal/subclonal evolution. We also show that

EWSR1-WT1 transcriptional activity level or binding sites' flexibility may participate in DSRCT cell plasticity by remodeling chromatin accessibility.

Still, several questions need to be addressed. First, it is currently unknown if the neural-to-mesenchymal state of DSRCT cells is revocable or non-reversible. Secondly, DSRCT cell plasticity potency may be both a consequence of their presumed pluripotent cell of origin allowing various differentiation pathways and enabled by EWSR1-WT1 activity. Interestingly, recent data focusing on Clear Cell Sarcoma (CCS), an alternate fusion-driven sarcoma characterized by the EWSR1-ATF1 chimeric TF, suggested that CCS precursor cells retain a degree of plasticity enabling them to differentiate along the melanocytic lineage and that the fusion protein may induce a transcriptional reprogramming toward a more undifferentiated state <sup>109</sup>. Other genomically simple sarcomas, such as synovial sarcoma characterized by SS18-SSX fusion protein, have been shown to undergo a pluripotent differentiation, with the coexpression of both a core oncogenic program driven by SS18-SSX and inducing a differentiation blockade, mesenchymal and epithelial programs <sup>41</sup>. Similarly, in Ewing sarcoma models, EWSR1-FLI1 activity levels direct the neural-to-mesenchymal differentiation <sup>14,15</sup>. Authors also showed that EWSR1-FLI1 induces cell plasticity by deregulating microRNA (miRNA) maturation <sup>110</sup> and decreasing miR-145 expression, which plays a central role in restricting cancer stem cells pluripotency features <sup>111,112</sup>. Noteworthy, our work did not identify evident EWSR1-WT1 transcript expression variability across DSRCT tumor cell clusters. Whether this observation is due to the potential limitations of our in-house developed EWSR1-WT1 single-cell expression assay or reflects genuine DSRCT biology remains to be confirmed. Bleijs et al. <sup>7</sup> performed SORT-seq, a plate-based 3'scRNA-seq technology <sup>113</sup>, on the 2D and 3D-grown DSRCT cell line OV-054. Since OV-054 does not express wild-type WT1, the single-cell level expression of EWSR1-WT1 was inferred by the measure of WT1-mapping reads. The authors did not highlight any EWSR1-WT1 expression level heterogeneity in these experiments. Whether this observation applies to *in vivo* DSRCT cells growing within an evolving microenvironment and submitted to various extrinsic signaling and therapeutic pressure remains questionable.

Beyond the lineage- and metabolism-related components underlying DSRCT cell heterogeneity, we could identify "pseudostates" characterized by enrichment in

lncRNAs, ribosomal protein-coding transcripts, or cell cycle-related genes. The biological reality of the first two transcriptional programs is still unknown. Regarding lncRNAs-enriched clusters, the latter have been described in alternate tumor types and may not be specific to DSRCT. Whether these clusters arise from technical biases, such as the encapsulation of nuclei or membrane-altered cells with cytoplasmic RNAs leakage, or whether a subgroup of DSRCT cells (among other tumor types) may set off a program preferentially activating the transcription of a subset of lncRNAs is still uncertain. Two arguments should be raised against the latter hypothesis. First, no lncRNAs-enriched cluster was identified in the single-nucleus Gene Expression and ATAC-seq experiment or in the Visium spatial transcriptomics assays. Secondly, the label transfer performed on the *PZB* sample using the 3' scRNA-seq intronic reads dataset as a reference and the snMultiome gene expression dataset as a query assigned almost all snMultiome clusters to the 3'scRNA-seq lncRNAs-enriched cluster.

In addition to cell-intrinsic plasticity, DSRCT tumor cell heterogeneity appears also to be triggered by microenvironmental signals such as hypoxia and interactions with CAFs and extracellular matrix components, which are associated with the acquisition of mesenchymal features and the activation of anaerobic glycolysis. This observation supports the need for further investigation of the role of the tumor microenvironment in DSRCT pathogenesis. To our knowledge, this is the first study to characterize DSRCT CAF subpopulations comprehensively. Remarkably, herein defined DSRCT CAF subclusters highly correlate with the recently published pan-cancer single-cell CAFs atlas portrayed by Luo et al.<sup>114</sup>, which identified three predominant CAF components (i.e., CAF<sub>myo</sub> (cancer-associated myofibroblasts), CAF<sub>infla</sub> (inflammatory CAFs), and CAF<sub>adi</sub> (adipogenic CAFs)). Specifically, in pancreatic ductal adenocarcinoma, which is typically characterized by tumor-associated desmoplasia, four CAF subtypes have been described, including myCAFs, iCAFs, antigen-presenting CAFs, and meCAFs harboring a highly activated metabolic state<sup>115–121</sup>. Our results, showing overlapping markers across DSRCT CAF subsets and suggesting a continuum of CAF phenotypes, align with findings from alternate tumor types. Davidson et al.<sup>122</sup> described the evolution of CAF phenotypes from murine melanoma and draining lymph nodes and highlighted the sequential development of “immune”, “desmoplastic”, and “contractile” CAFs along tumor progression. The transition of adipose-derived stromal cells into extra-cellular matrix-producing CAFs has also been described<sup>123</sup>, supporting that an

adipose tissue microenvironment may constitute a favorable niche for cancer development and progression.

Finally, although patients with DSRCT harbor the unique and pathognomonic EWSR1-WT1 fusion, treatment response and survival outcomes may be variable. The underlying biological mechanisms underlying this prognostic heterogeneity remain largely unknown. To date, specific recurrent secondary mutations have not been linked to patient prognosis, and larger series will be required to address this important question, which will also bring insight into the disease's biology. Our work identified critical transcriptional programs linked to patient prognosis. For instance, some DSRCT cell state gene signatures may be surrogates for tumor cell proliferation and cell growth (e.g., *Cycling cells* – 3, *Ribosomal\_IFN response* – 6 signatures) and were linked to adverse prognosis. On the other hand, the *Epithelial\_Mesenchymal* – 4 signature was associated with increased overall survival.

In conclusion, our study sheds light on DSRCT heterogeneity underlying mechanisms, which may have prognostic implications for patients and direct novel therapeutic aiming at corralling cells within specific cell states associated with improved prognosis. Relying on the hypothesis that DSRCT tumor cell subpopulations maintain mutualistic interactions, where each cell subpopulation may benefit from the phenotypical state of each other, we may also expect that the specific targeting of a given crucial subpopulation may destabilize the whole tumor ecosystem. As tumor heterogeneity is one of the major hurdles in cancer treatment, we hope our approach will apply to various alternate tumor subtypes. We envision that such studies will further help to improve the understanding of the epigenetic and transcriptional processes underlying sarcoma biology and pave the way to novel therapeutic opportunities in this deadly disease.

## References

1. Cidre-Aranaz, F. *et al.* Small round cell sarcomas. *Nat. Rev. Dis. Primer* **8**, 66 (2022).
2. Wong, H. H. *et al.* Desmoplastic small round cell tumour: characteristics and prognostic factors of 41 patients and review of the literature. *Clin. Sarcoma Res.* **3**, 14 (2013).
3. Shoushtari, A. N. *et al.* Predictors of overall survival in patients diagnosed with desmoplastic small round cell tumor (DSRCT). *J. Clin. Oncol.* **32**, 10582–10582 (2014).
4. Waqar, S. H. B. & Ali, H. Changing incidence and survival of desmoplastic small round cell tumor in the USA. *Proc. Bayl. Univ. Med. Cent.* **35**, 415–419.
5. Bandopadhyay, P. *et al.* The oncogenic properties of EWS/WT1 of desmoplastic small round cell tumors are unmasked by loss of p53 in murine embryonic fibroblasts. *BMC Cancer* **13**, 585 (2013).
6. Gedminas, J. M. *et al.* Desmoplastic small round cell tumor is dependent on the EWSR1-WT1 transcription factor. *Oncogenesis* **9**, 1–8 (2020).
7. Bleijs, M. *et al.* EWSR1-WT1 Target Genes and Therapeutic Options Identified in a Novel DSRCT In Vitro Model. *Cancers* **13**, 6072 (2021).
8. Bulbul, A. *et al.* Desmoplastic Small Round Blue Cell Tumor: A Review of Treatment and Potential Therapeutic Genomic Alterations. *Sarcoma* **2017**, 1278268 (2017).
9. Chow, W. A. *et al.* Recurrent secondary genomic alterations in desmoplastic small round cell tumors. *BMC Med. Genet.* **21**, 101 (2020).
10. Wu, C.-C. *et al.* Multi-site desmoplastic small round cell tumors are genetically related and immune-cold. *Npj Precis. Oncol.* **6**, 1–14 (2022).
11. Devecchi, A. *et al.* The genomics of desmoplastic small round cell tumor reveals the deregulation of genes related to DNA damage response, epithelial–mesenchymal transition, and immune response. *Cancer Commun.* **38**, (2018).
12. Nacev, B. A. *et al.* Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets. *Nat. Commun.* **13**, 3405 (2022).
13. Slotkin, E. K. *et al.* Comprehensive Molecular Profiling of Desmoplastic Small Round Cell Tumor. *Mol. Cancer Res. MCR* **19**, 1146–1155 (2021).
14. Franzetti, G.-A. *et al.* Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **36**, 3505–3514 (2017).
15. Aynaud, M.-M. *et al.* Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cell Rep.* **30**, 1767-1779.e6 (2020).
16. Nishio, J. *et al.* Establishment and characterization of a novel human desmoplastic small round cell tumor cell line, JN-DSRCT-1. *Lab. Investig. J. Tech. Methods Pathol.* **82**, 1175–1182 (2002).
17. Chromium Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry) -User Guide -Library Prep -Single Cell Gene Expression -Official 10x Genomics Support. <https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v31-chemistry>.
18. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
19. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
20. DeTomaso, D. & Yosef, N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446-456.e9 (2021).
21. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).

22. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
23. Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405–411 (2020).
24. Grün, D. *et al.* De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19**, 266–277 (2016).
25. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
26. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
27. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).
28. González-Blas, C. B. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. 2022.08.19.504505 Preprint at <https://doi.org/10.1101/2022.08.19.504505> (2022).
29. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
30. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
31. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
32. Stuart, T. *et al.* Comprehensive integration of single cell data. *bioRxiv* 460147 (2018) doi:10.1101/460147.
33. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol.* **12**, e1004873 (2016).
34. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
35. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al* **109**, 21.29.1–21.29.9 (2015).
36. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
37. Arrigoni, L. *et al.* Standardizing chromatin research: a simple and universal method for ChIP-seq. *Nucleic Acids Res.* **44**, e67 (2016).
38. Welch, R. P. *et al.* ChIP-Enrich: gene set enrichment testing for ChIP-seq data. *Nucleic Acids Res.* **42**, e105 (2014).
39. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
40. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
41. Jerby-Arnon, L. *et al.* Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med.* **27**, 289–300 (2021).
42. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
43. Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611–1624.e24 (2017).
44. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).

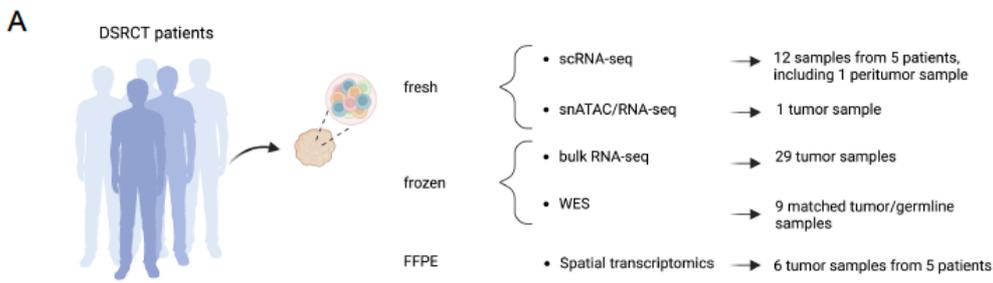
45. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
46. Jerby-Arnon, L. *et al.* A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **175**, 984-997.e24 (2018).
47. Sydow, S. *et al.* Genomic and transcriptomic characterization of desmoplastic small round cell tumors. *Genes. Chromosomes Cancer* (2021) doi:10.1002/gcc.22955.
48. Vibert, J. *et al.* Oncogenic chimeric transcription factors drive tumor-specific transcription, processing, and translation of silent genomic regions. *Mol. Cell* **82**, 2458-2471.e9 (2022).
49. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
50. Thway, K. *et al.* Desmoplastic Small Round Cell Tumor: Pathology, Genetics, and Potential Therapeutic Strategies. *Int. J. Surg. Pathol.* **24**, 672–684 (2016).
51. Rindler, K. *et al.* Single-Cell RNA Sequencing Reveals Tissue Compartment-Specific Plasticity of Mycosis Fungoides Tumor Cells. *Front. Immunol.* **12**, 666935 (2021).
52. Lukowski, S. W. *et al.* A single-cell transcriptome atlas of the adult human retina. *EMBO J.* **38**, e100811 (2019).
53. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).
54. Alvarez, M. *et al.* Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. *Sci. Rep.* **10**, 11019 (2020).
55. Truong, D. D. *et al.* Dissociation Protocols used for Sarcoma Tissues Bias the Transcriptome observed in Single-cell and Single-nucleus RNA sequencing. <http://biorxiv.org/lookup/doi/10.1101/2022.01.21.476982> (2022) doi:10.1101/2022.01.21.476982.
56. Kang, H.-J. *et al.* EWS–WT1 Oncoprotein Activates Neuronal Reprogramming Factor ASCL1 and Promotes Neural Differentiation. *Cancer Res.* **74**, 4526–4535 (2014).
57. Tickle, T. I., Georgescu, C., Brown, M. & Haas, B. inferCNV of the Trinity CTAT Project (2019);<https://github.com/broadinstitute/inferCNV>.
58. Hingorani, P. *et al.* Transcriptome analysis of desmoplastic small round cell tumors identifies actionable therapeutic targets: a report from the Children’s Oncology Group. *Sci. Rep.* **10**, 12318 (2020).
59. Zhao, Y., Long, M. J. C., Wang, Y., Zhang, S. & Aye, Y. Ube2V2 Is a Rosetta Stone Bridging Redox and Ubiquitin Codes, Coordinating DNA Damage Responses. *ACS Cent. Sci.* **4**, 246–259 (2018).
60. Soshnikova, N. V., Sheynov, A. A., Tatarskiy, Eu. V. & Georgieva, S. G. The DPF Domain As a Unique Structural Unit Participating in Transcriptional Activation, Cell Differentiation, and Malignant Transformation. *Acta Naturae* **12**, 57–65 (2020).
61. Janssen, S. M. *et al.* BORIS/CTCFL promotes a switch from a proliferative towards an invasive phenotype in melanoma cells. *Cell Death Discov.* **6**, 1–17 (2020).
62. Salgado-Albarrán, M. *et al.* The epigenetic factor BORIS (CTCFL) controls the androgen receptor regulatory network in ovarian cancer. *Oncogenesis* **8**, 1–12 (2019).
63. Palmer, R. E. *et al.* Induction of BAIAP3 by the EWSR1-WT1 chimeric fusion implicates regulated exocytosis in tumorigenesis. *Cancer Cell* **2**, 497–505 (2002).
64. Wong, J. C. *et al.* Induction of the interleukin-2/15 receptor  $\beta$ -chain by the EWS–WT1 translocation product. *Oncogene* **21**, 2009–2019 (2002).
65. Liu, J. *et al.* Molecular Heterogeneity and Function of EWSR1-WT1 Fusion Transcripts in Desmoplastic Small Round Cell Tumors. *Clin. Cancer Res.* **6**, 3522–3529 (2000).
66. Uboldi, S. *et al.* Mechanism of action of trabectedin in desmoplastic small round cell tumor cells. *BMC Cancer* **17**, (2017).

67. Reynolds, P. A. *et al.* Identification of a DNA-binding site and transcriptional target for the EWSR1-WT1(+KTS) oncoprotein. *Genes Dev.* **17**, 2094–2107 (2003).
68. Rauscher, F. J., Morris, J. F., Tournay, O. E., Cook, D. M. & Curran, T. Binding of the Wilms' tumor locus zinc finger protein to the EGR1 consensus sequence. *Science* **250**, 1259–1262 (1990).
69. Borel, F., Barilla, K. C., Hamilton, T. B., Iskandar, M. & Romaniuk, P. J. Effects of Denys-Drash syndrome point mutations on the DNA binding activity of the Wilms' tumor suppressor protein WT1. *Biochemistry* **35**, 12070–12076 (1996).
70. Drummond, I. A. *et al.* DNA recognition by splicing variants of the Wilms' tumor suppressor, WT1. *Mol. Cell. Biol.* **14**, 3800–3809 (1994).
71. Hamilton, T. B., Barilla, K. C. & Romaniuk, P. J. High affinity binding sites for the Wilms' tumour suppressor protein WT1. *Nucleic Acids Res.* **23**, 277–284 (1995).
72. Nakagama, H., Heinrich, G., Pelletier, J. & Housman, D. E. Sequence and structural requirements for high-affinity DNA binding by the WT1 gene product. *Mol. Cell. Biol.* **15**, 1489–1498 (1995).
73. Reynolds, P. A. *et al.* Identification of a DNA-binding site and transcriptional target for the EWSR1-WT1(+KTS) oncoprotein. *Genes Dev.* **17**, 2094–2107 (2003).
74. Gao, N. *et al.* The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol. Endocrinol. Baltim. Md* **17**, 1484–1507 (2003).
75. Yu, X. *et al.* Foxa1 and Foxa2 Interact with the Androgen Receptor to Regulate Prostate and Epididymal Genes Differentially. *Ann. N. Y. Acad. Sci.* **1061**, 77–93 (2005).
76. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
77. Lamhamedi-Cherradi, S.-E. *et al.* The androgen receptor is a therapeutic target in desmoplastic small round cell sarcoma. *Nat. Commun.* **13**, 3057 (2022).
78. Zhang, S. & Cui, W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* **6**, 305–311 (2014).
79. Igarashi, K., Nishizawa, H., Saiki, Y. & Matsumoto, M. The transcription factor BACH1 at the crossroads of cancer biology: From epithelial–mesenchymal transition to ferroptosis. *J. Biol. Chem.* **297**, 101032 (2021).
80. Wiel, C. *et al.* BACH1 Stabilization by Antioxidants Stimulates Lung Cancer Metastasis. *Cell* **178**, 330-345.e22 (2019).
81. Padilla, J. & Lee, J. A Novel Therapeutic Target, BACH1, Regulates Cancer Metabolism. *Cells* **10**, 634 (2021).
82. Kume, T., Deng, K. & Hogan, B. L. Murine forkhead/winged helix genes Foxc1 (Mf1) and Foxc2 (Mfh1) are required for the early organogenesis of the kidney and urinary tract. *Dev. Camb. Engl.* **127**, 1387–1395 (2000).
83. Mani, S. A. *et al.* Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10069–10074 (2007).
84. Mulder, K. *et al.* Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54**, 1883-1900.e5 (2021).
85. Binnewies, M. *et al.* Targeting TREM2 on tumor-associated macrophages enhances immunotherapy. *Cell Rep.* **37**, 109844 (2021).
86. Nakamura, K. & Smyth, M. J. TREM2 marks tumor-associated macrophages. *Signal Transduct. Target. Ther.* **5**, 1–2 (2020).
87. Molgora, M. *et al.* TREM2 Modulation Remodels the Tumor Myeloid Landscape Enhancing Anti-PD-1 Immunotherapy. *Cell* **182**, 886-900.e17 (2020).

88. Inoue, T. *et al.* Non-canonical cholinergic anti-inflammatory pathway-mediated activation of peritoneal macrophages induces Hes1 and blocks ischemia/reperfusion injury in the kidney. *Kidney Int.* **95**, 563–576 (2019).
89. Shang, Y. *et al.* The transcriptional repressor Hes1 attenuates inflammation via regulating transcriptional elongation. *Nat. Immunol.* **17**, 930–937 (2016).
90. Fan, Y. *et al.* The effect of anti-inflammatory properties of ferritin light chain on lipopolysaccharide-induced inflammatory response in murine macrophages. *Biochim. Biophys. Acta* **1843**, 2775–2783 (2014).
91. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8<sup>+</sup> T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).
92. Pelon, F. *et al.* Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. *Nat. Commun.* **11**, 404 (2020).
93. Costa, A. *et al.* Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell* **33**, 463–479.e10 (2018).
94. Zhang, T., Ren, Y., Yang, P., Wang, J. & Zhou, H. Cancer-associated fibroblasts in pancreatic ductal adenocarcinoma. *Cell Death Dis.* **13**, 1–11 (2022).
95. Geng, X. *et al.* Cancer-Associated Fibroblast (CAF) Heterogeneity and Targeting Therapy of CAFs in Pancreatic Cancer. *Front. Cell Dev. Biol.* **9**, 655152 (2021).
96. Sun, D.-P., Lee, Y.-W., Chen, J.-T., Lin, Y.-W. & Chen, R.-M. The Bradykinin-BDKRB1 Axis Regulates Aquaporin 4 Gene Expression and Consequential Migration and Invasion of Malignant Glioblastoma Cells via a Ca<sup>2+</sup>-MEK1-ERK1/2-NF-κB Mechanism. *Cancers* **12**, 667 (2020).
97. Vijay, J. *et al.* Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. *Nat. Metab.* **2**, 97–109 (2020).
98. Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
99. Merrick, D. *et al.* Identification of a mesenchymal progenitor cell hierarchy in adipose tissue. *Science* **364**, eaav2501 (2019).
100. Lee, H. *et al.* SCARA5 plays a critical role in the commitment of mesenchymal stem cells to adipogenesis. *Sci. Rep.* **7**, 14833 (2017).
101. Monteran, L. & Erez, N. The Dark Side of Fibroblasts: Cancer-Associated Fibroblasts as Mediators of Immunosuppression in the Tumor Microenvironment. *Front. Immunol.* **10**, 1835 (2019).
102. Jenkins, L. *et al.* Cancer-Associated Fibroblasts Suppress CD8<sup>+</sup> T-cell Infiltration and Confer Resistance to Immune-Checkpoint Blockade. *Cancer Res.* **82**, 2904–2917 (2022).
103. Boulakirba, S. *et al.* IL-34 and CSF-1 display an equivalent macrophage differentiation ability but a different polarization potential. *Sci. Rep.* **8**, 256 (2018).
104. Jones, C. V. & Ricardo, S. D. Macrophages and CSF-1. *Organogenesis* **9**, 249–260 (2013).
105. Sánchez-Martín, L. *et al.* The chemokine CXCL12 regulates monocyte-macrophage differentiation and RUNX3 expression. *Blood* **117**, 88–97 (2011).
106. Casella, G. *et al.* IL4 induces IL6-producing M2 macrophages associated to inhibition of neuroinflammation in vitro and in vivo. *J. Neuroinflammation* **13**, 139 (2016).
107. Li, B.-H., Garstka, M. A. & Li, Z.-F. Chemokines and their receptors promoting the recruitment of myeloid-derived suppressor cells into the tumor. *Mol. Immunol.* **117**, 201–215 (2020).
108. Kieffer, Y. *et al.* Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. *Cancer Discov.* **10**, 1330–1351 (2020).
109. Möller, E. *et al.* EWSR1-ATF1 dependent 3D connectivity regulates oncogenic and differentiation programs in Clear Cell Sarcoma. *Nat. Commun.* **13**, 2267 (2022).

110. De Vito, C. *et al.* A TARBP2-dependent miRNA expression profile underlies cancer stem cell properties and provides candidate therapeutic reagents in Ewing sarcoma. *Cancer Cell* **21**, 807–821 (2012).
111. Riggi, N. *et al.* EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev.* **24**, 916–932 (2010).
112. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. & Kosik, K. S. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* **137**, 647–658 (2009).
113. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).
114. Luo, H. *et al.* Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. *Nat. Commun.* **13**, 6619 (2022).
115. Wang, Y. *et al.* Single-cell analysis of pancreatic ductal adenocarcinoma identifies a novel fibroblast subtype associated with poor prognosis but better immunotherapy response. *Cell Discov.* **7**, 1–17 (2021).
116. Helms, E., Onate, M. K. & Sherman, M. H. Fibroblast Heterogeneity in the Pancreatic Tumor Microenvironment. *Cancer Discov.* **10**, 648–656 (2020).
117. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
118. Dominguez, C. X. *et al.* Single-Cell RNA Sequencing Reveals Stromal Evolution into LRRC15+ Myofibroblasts as a Determinant of Patient Response to Cancer Immunotherapy. *Cancer Discov.* **10**, 232–253 (2020).
119. Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
120. Hosein, A. N. *et al.* Cellular heterogeneity during mouse pancreatic ductal adenocarcinoma progression at single-cell resolution. *JCI Insight* **5**, 129212 (2019).
121. Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov.* **9**, 1102–1123 (2019).
122. Davidson, S. *et al.* Single-Cell RNA Sequencing Reveals a Dynamic Stromal Niche That Supports Tumor Growth. *Cell Rep.* **31**, (2020).
123. Zhu, K., Cai, L., Cui, C., de Los Toyos, J. R. & Anastassiou, D. Single-cell analysis reveals the pan-cancer invasiveness-associated transition of adipose-derived stromal cells into COL11A1-expressing cancer-associated fibroblasts. *PLoS Comput. Biol.* **17**, e1009228 (2021).

# Main figures and legends

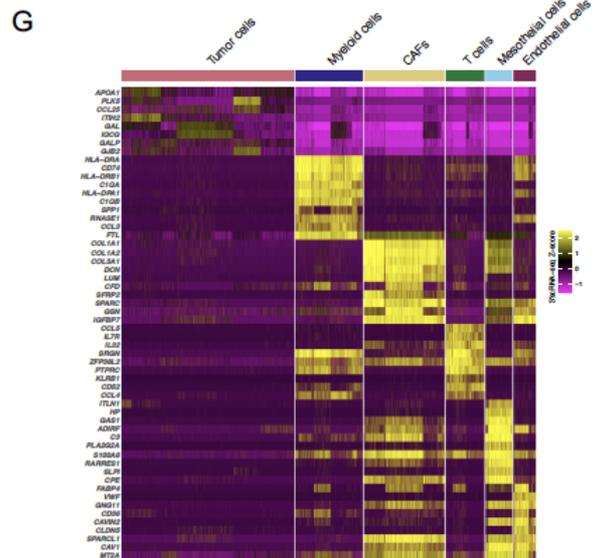
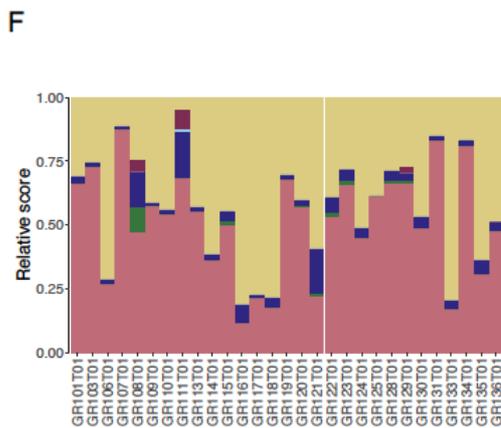
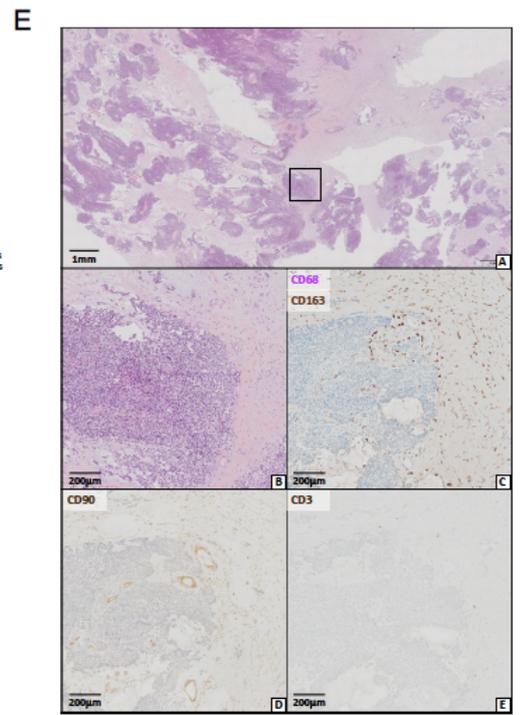
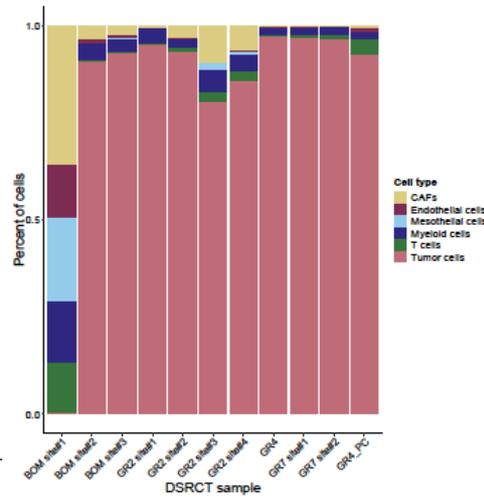
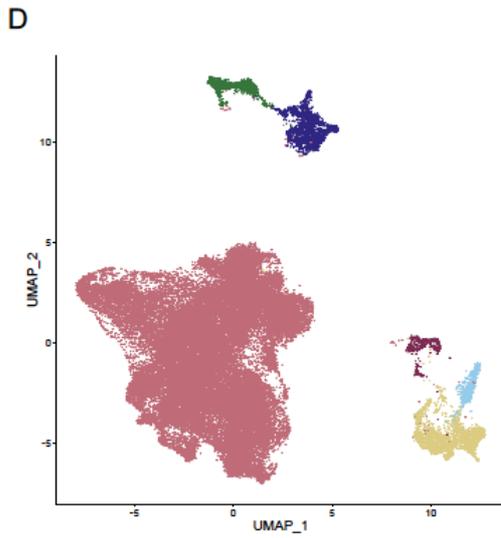
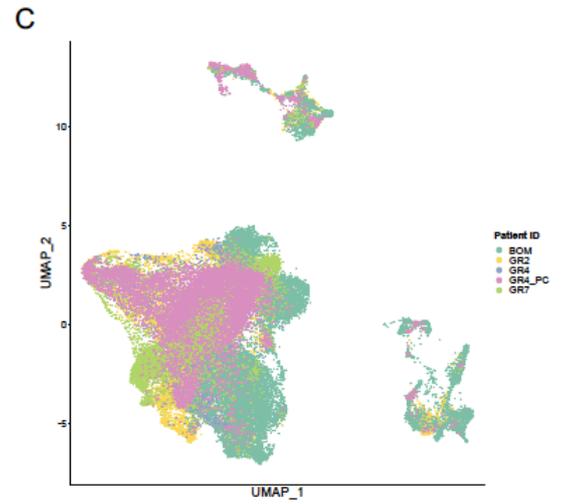


**B**

	GR2 site#1	GR2 site#2	GR2 site#3	GR2 site#4	GR4	GR4_PC	BOM site#1	BOM site#2	BOM site#3	GR7 site#1	GR7 site#2	PZB	PZB
Patient ID	GR2				GR4	GR4_PC	BOM			GR7		PZB	
Age (y)	25				30	31	17			32		31	
Sex	Male				Female	Male	Male			Male		Male	
Prior Systemic Treatment	L2				No	L2	L1			L1		L2	
Sample type	Tumor tissue				Peritumor tissue	Tumor tissue	Tumor tissue			Tumor tissue		Peritumor tissue	
Single-cell technique	3'scRNA-seq				3'scRNA-seq/snATAC-seq	3'scRNA-seq/snATAC-seq	3'scRNA-seq/snATAC-seq			3'scRNA-seq/snATAC-seq		3'scRNA-seq/snATAC-seq	
Bulk RNA-seq	Yes				No	Yes	Yes			Yes		Yes	
WES	Yes				No	Yes	Yes			Yes		Yes	

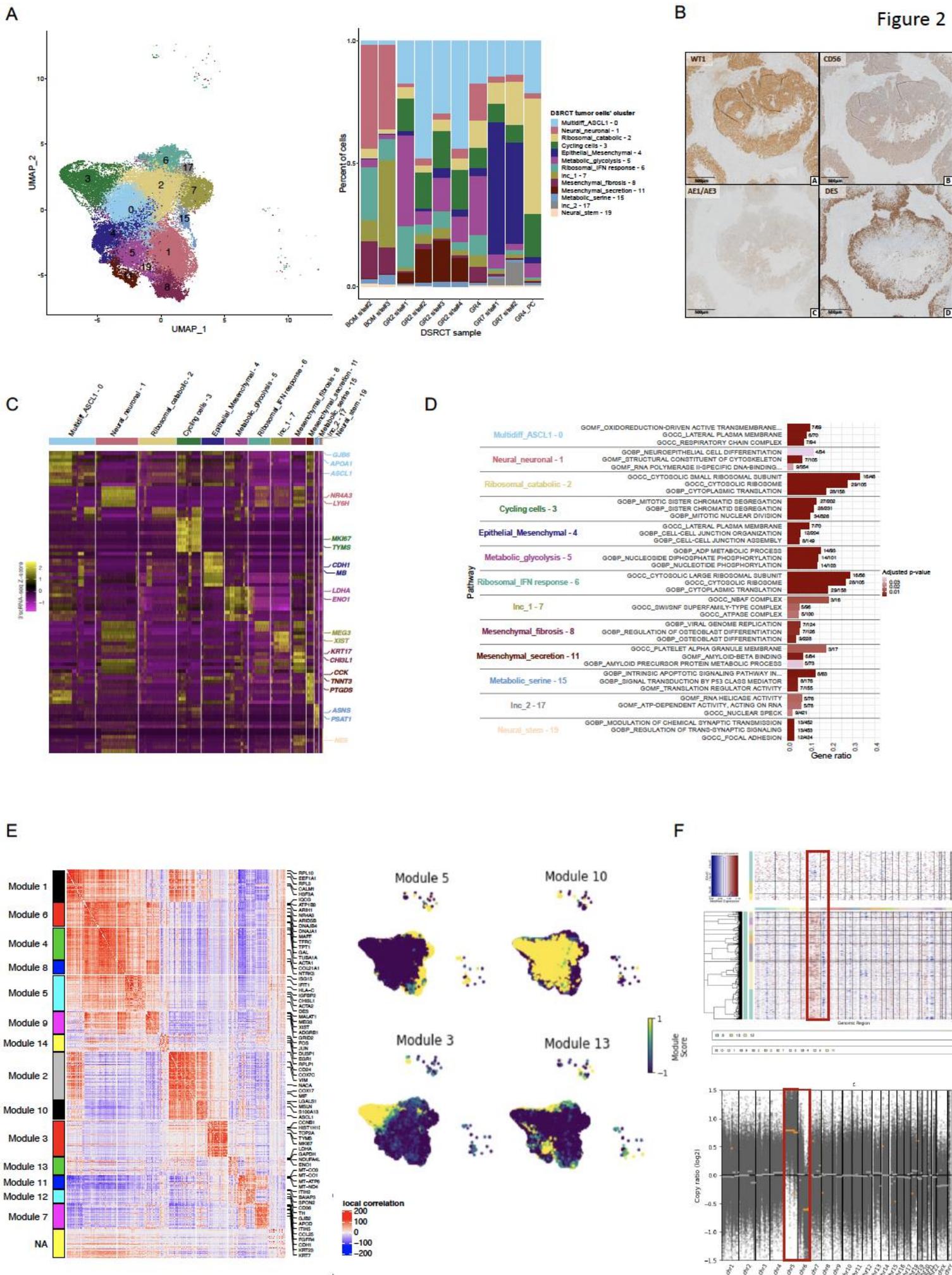
**Legend:**

- Patient ID:** GR2 (yellow), GR4 (purple), GR4\_PC (pink), BOM (green), GR7 (light green), PZB (blue)
- Single-cell technique:** 3'scRNA-seq (light blue), 3'scRNA-seq/snATAC-seq (light green)
- WES:** Yes (white), No (grey)
- Bulk RNA-seq:** Yes (white), No (grey)
- Sex:** Female (dark blue), Male (dark red)
- Sample type:** Tumor tissue (yellow), Peritumor tissue (light yellow)

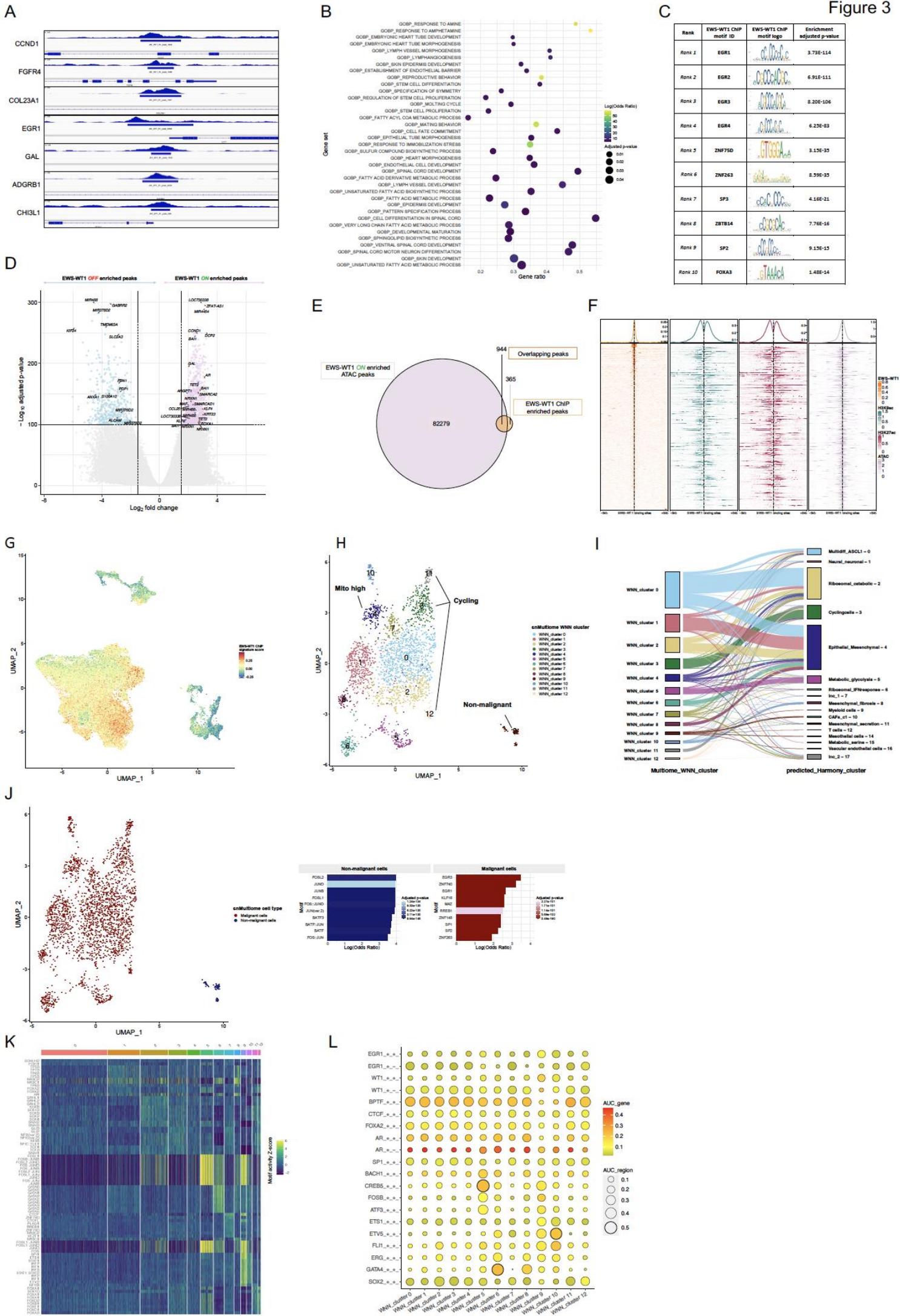


**Fig. 1 | scRNA-seq recapitulates DSRCT cellular composition.** **A**, Experimental design of bulk and single-cell multiomic profiling performed on patient-derived fresh and frozen DSRCT samples. **B**, DSRCT patient clinical characteristics and matching analyzed samples. **C**, Uniform Manifold Approximation and Projection (UMAP) highlighting per patient cells' origin on Harmony integrated 3'scRNA-seq dataset ("*Int\_sc*"). **D**, UMAP highlighting identified DSRCT cell types in the "*Int\_sc*" dataset (left panel). Barplot showing the proportions of identified DSRCT cell types across 3'scRNA-seq samples, including a majority of tumor cells, followed by CAFs and myeloid cells, with a minority of T cells, endothelial cells, and mesothelial cells (right panel). **E**, Hematoxylin and Eosin (H&E), and monoplex immunohistochemistry (IHC) stainings (DAB) on Formalin-Fixed Paraffin-embedded (FFPE) DSRCT samples, displaying DSRCT tumor cells surrounded by a desmoplastic stroma characterized by THY1+ Cancers Associated Fibroblasts (CAFs), CD68+/CD163+ macrophages, and low T cell infiltration (CD3+). **F**, CIBERSORTx deconvolution of the main DSRCT subpopulations on a cohort of 29 bulk RNA-seq DSRCT samples. **G**, Heatmap showing the expression Z-score of the top 10 differentially expressed genes of each DSRCT cell population. High gene expression is shown in yellow, whereas low gene expression is shown in purple. While marker genes were identified for non-tumor cell clusters, no gene was found to be ubiquitously overexpressed across all tumor cell clusters.

Figure 2



**Fig. 2 | DSRCT malignant cell subpopulations' profiles suggest cell plasticity.** **A**, Uniform Manifold Approximation and Projection (UMAP) representation of manually annotated DSRCT malignant cells' clusters from Harmony integrated 3'scRNA-seq dataset ("*Int\_sc*") after clustering using a graph-based K-nearest neighbors' algorithm (left panel) where clusters' annotation was performed based on the top differentially expressed genes and corresponding gene set enrichment analysis. Barplot showing the proportions of DSRCT malignant cell subpopulations across individual samples (right panel). **B**, Monoplex immunohistochemistry (IHC) stainings (DAB) on Formalin-Fixed Paraffin-embedded (FFPE) DSRCT samples highlighting WT1-positive DSRCT malignant cells expressing epithelial (AE1/AE3), mesenchymal (DES) and neural (CD56) markers. **C**, Heatmap highlighting the expression Z-score of top marker genes for each DSRCT malignant cell cluster. High gene expression is shown in yellow, whereas low gene expression is shown in purple. Newly identified canonical markers for each DSRCT malignant cell cluster are shown on the right. **D**, Gene Ontology (GO) pathway enrichment analysis based on differentially expressed genes from identified DSRCT malignant cells' clusters. The top 3 GO pathways are displayed for each cluster on the y-axis. The gene ratio is shown on the x-axis. Significant adjusted p-values are colored in pink (higher adjusted p-value) to red (lower adjusted p-value). **E**, Heatmap representation of Hotspot co-expressed gene modules identified in the "*Int\_sc*" dataset (left panel). Gene modules are specified on the left y-axis. The color highlights the local correlation values between genes from red (high local correlation) to blue (low local correlation). UMAP representation of the "*Int\_sc*" dataset showing Hotspot modules' scores (right panel). Several modules displayed relatively homogeneous scores across clusters (e.g., *Hotspot module #10*). In contrast, other modules were explicitly enriched in particular clusters (e.g., *Hotspot module #3* and "*Int\_sc*" *Cycling cells* – 3 cluster, *Hotspot module #13* and "*Int\_sc*" *Metabolic\_glycolysis* – 5 cluster, *Hotspot module #5* and "*Int\_sc*" *Mesenchymal\_fibrosis* – 8 cluster). **F**, Single-cell level copy number variations (CNV) inference using the InferCNV tool (upper panel) and bulk whole exome sequencing (WES)-derived CNV analysis on matching sample (lower panel). Results for *GR7 site#2* display (i) a gain of 5p/5q and a loss of 6p, confirmed by WES, and (ii) a flat profile across clusters suggesting few/no clonal heterogeneity.

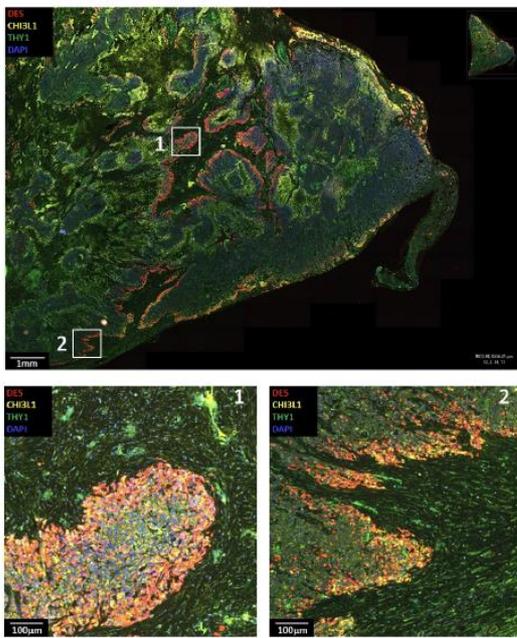


**Fig. 3 | DSRCT heterogeneity is partially driven by variable EWSR1-WT1 activity and epigenetic reprogramming.** **A**, EWSR1-WT1 chromatin immunoprecipitation with sequencing (ChIP-seq) coverage profiles and peaks for known and newly described target genes. **B**, Gene Ontology (GO) enrichment analysis of significantly enriched peaks in EWSR1-WT1 ChIP-seq using ChIPEnrich tool on biological process (GOBP), cellular component (CC), and molecular function (MF) categories. All pathways with significant enrichment (adjusted p-value <0.05) are displayed on the y-axis from lower (top) to higher adjusted p-value (bottom). The gene ratio is shown on the x-axis. **C**, Top 10 enriched motifs from JASPAR 2020 transcription factors (TFs) database in significantly enriched peaks from EWSR1-WT1 ChIP-seq. **D**, Volcano plot showing Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) differentially accessible peaks-corresponding genes in EWSR1-WT1-silenced versus non-silenced JN-DSRCT-1 cells. Peaks that have higher accessibility upon EWSR1-WT1 silencing are shown on the left (EWSR1-WT1 OFF), whereas those that have increased accessibility in non-silenced cells are shown on the right (EWSR1-WT1 ON). Colored dots correspond to the most differentially accessible peaks (-Log<sub>10</sub> (adjusted p-value) >100, Log<sub>2</sub> fold change (Log<sub>2</sub>FC) < -1.5 in blue, Log<sub>2</sub>FC > 1.5 in pink). **E**, Venn Diagram showing overlapping peaks between EWSR1-WT1 ChIP-seq-enriched peaks and peaks with significantly increased accessibility in non-silenced JN-DSRCT-1 cells compared to EWSR1-WT1 silenced cells. **F**, ChIP-seq read coverage heatmaps centered around EWSR1-WT1 ChIP-specific peaks (+/- 5 kb) and ordered by genomic regions for EWSR1-WT1 (first panel), H3K9ac (second panel), and H3K27ac (third panel) ChIP-seq experiments. The fourth panel shows coverage for peaks with increased accessibility in non-silenced JN-DSRCT-1 cells compared to EWSR1-WT1-silenced cells. EWSR1-WT1 targeted regions are characterized by active enhancers (H3K27ac/H3K9ac) and bivalent promoters (H3K9ac) histone marks, and opened chromatin regions. **G**, UMAP plot of “*Int\_sc*” showing EWSR1-WT1 transcriptional activity signature score. The latter signature was derived from the list of EWSR1-WT1 ChIP-seq differentially enriched peak-assigned genes. This signature contains a total of 176 genes, among which 103 genes are expressed in the “*Int\_sc*” dataset. A control signature using a random set of 103 genes is presented in **Supplemental Information – Fig. S7**. **H**, UMAP representation of graph-based K-nearest neighbors’ clustering on single-nucleus Multiome (snMultiome) sample combining single-nucleus ATAC-seq and RNA-seq features using the weighted nearest neighbors (WNN) method. Clusters

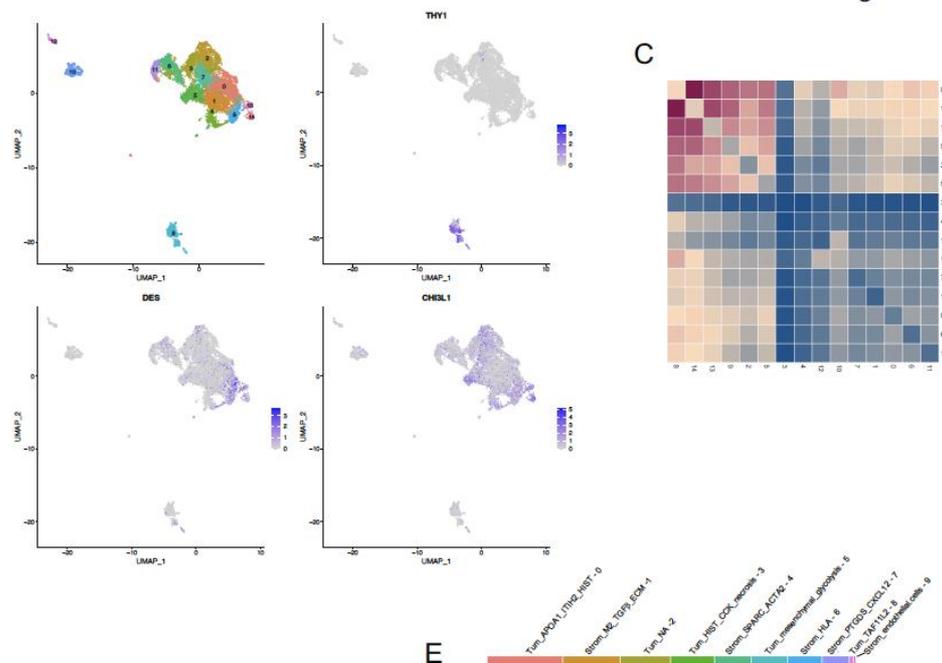
that could be identified according to differentially expressed genes are annotated. **I**, Label transfer of WNN clusters from the snMultiome assay with 3'scRNA-seq "*Int\_sc*" clusters as reference using snMultiome intronic reads. **J**, UMAP representation of the snMultiome assay highlighting malignant and non-malignant cells (left panel) and the top 10 significantly enriched motifs identified from corresponding differentially accessible peaks. **K**, Heatmap showing the ChromVAR-inferred motif activity for the top enriched motifs identified from differentially accessible peaks for each snMultiome WNN cluster. **L**, Bubble plot displaying the area under curve (AUC) of SCENIC+-inferred regulatory networks based on (i) regulons' accessibility (AUC\_region) represented by the size of the dot, and on (ii) regulons' gene expression level (AUC\_gene), represented by the color scale. A selection of regulons is represented based on putative important transcription factors. Noteworthy, we observe a decoupling between WT1 and EGR regulons accessibility and expression, supporting a competitive antagonism with EWSR1-WT1.



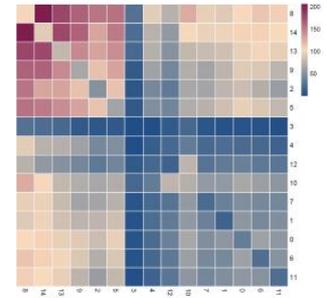
**Fig. 4 | DSRCT microenvironment displays immunosuppressive features and shapes DSRCT cell states.** **A**, DSRCT infiltrating myeloid subpopulations from integrated 3'scRNA-seq dataset (*Int\_sc*), represented by Uniform Manifold Approximation and Projection (UMAP) and pie plot showing the proportion of each immune subpopulation. Myeloid cells were annotated according to the MoMac-VERSE atlas, showing a majority of *HES1+*, *RSG1/HLA-DP+*, *TREM2+*, and *FTL+* macrophages. **B**, UMAP representation of DSRCT infiltrating lymphoid cells from the "*Int\_sc*" dataset and pie plot showing the proportion of each subpopulation. Lymphoid cells were manually annotated after graph-based k-nearest neighbors' clustering, according to the expression of known canonical markers and are composed of CD4+ memory T cells, CD8+ effector T cells, translation-related genes-enriched T cells, and activated NK cells. **C**, UMAP plots displaying the fibroblast population according to the tissue of origin (tumoral or juxta-tumoral) (left panel), and fibroblasts sub-clustering with graph-based K-nearest neighbors' algorithm, highlighting eight subpopulations (right panel). **D**, CIBERSORTx deconvolution of the identified fibroblast subpopulations on a cohort of 29 bulk RNA-seq DSRCT samples. **E**, Heatmap highlighting the single-cell expression Z-score of differentially expressed genes for each defined fibroblast subcluster. High gene expression is shown in yellow, whereas low gene expression is shown in purple. Key marker genes for each fibroblast cluster are shown on the right. **F**, Violin plot showing the expression level profile of cells from each fibroblast subpopulation of selected newly defined canonical markers. **G**, Violin plots displaying the expression level profile of cells from each fibroblast subcluster of profibrotic and immunosuppressive genes. **H**, Immunofluorescence triplex showing ACTA2, MCAM, FAP, and DAPI staining on a DSRCT FFPE sample. The CAFs located within the peritumor pseudocapsule are positive for ACTA2 and FAP, as do the CAFs located within large trabeculae of desmoplastic stroma. On the contrary, ACTA2+/FAP-/MCAM- CAFs surround DSRCT tumor islets, and MCAM+ CAFs encircle intratumor vessels. \*: p-value<0.05 \*\*: p-value<0.01, t-test. **I**, The semi-automated analysis of fluorescence intensity in the distinct stromal areas showed a significant increase FAP/ACTA2 fluorescence ratio in CAFs located in the pseudo-capsule.



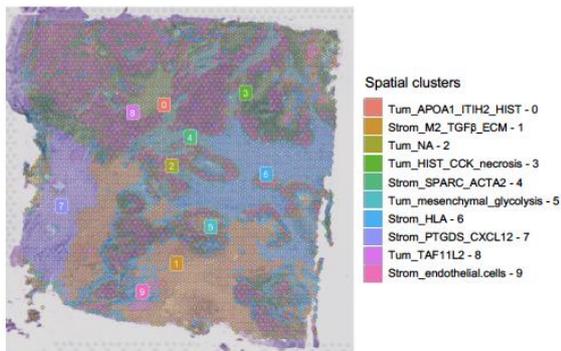
**B**



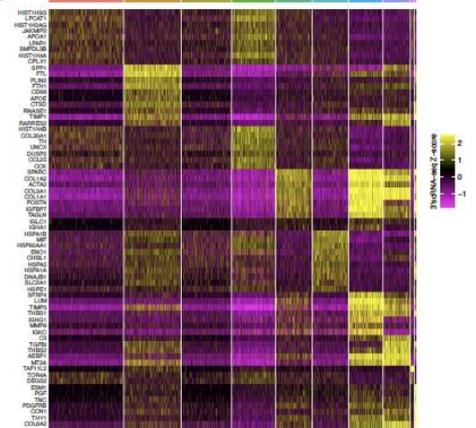
**C**



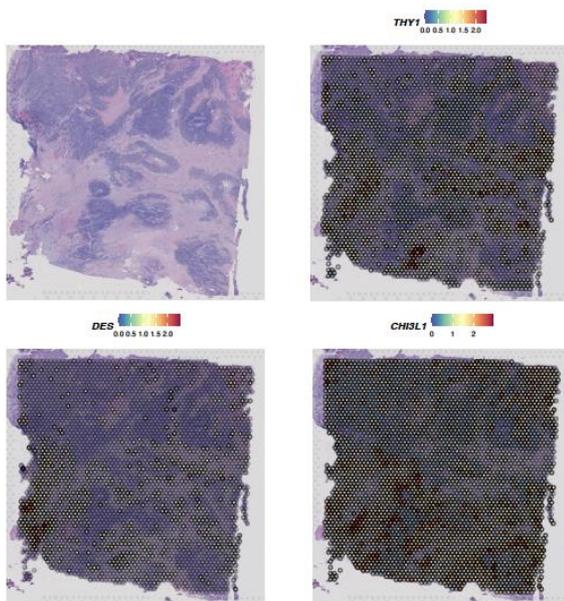
**D**



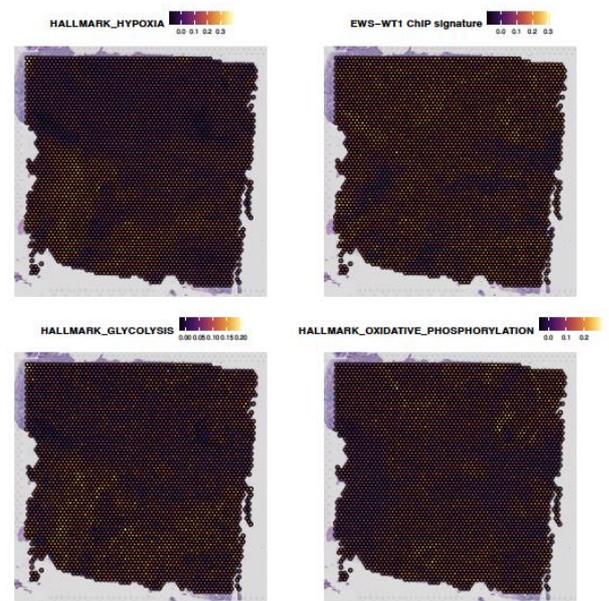
**E**



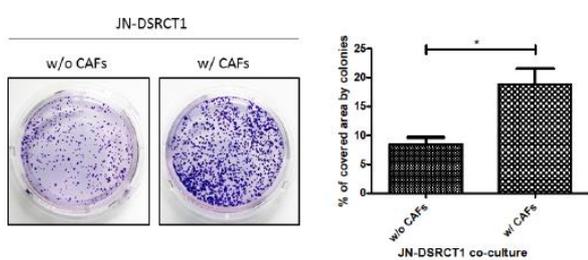
**F**



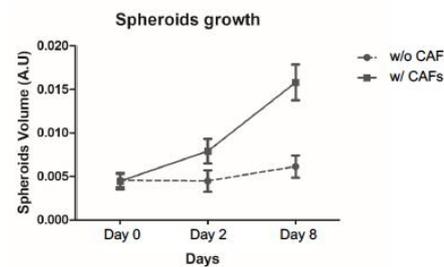
**G**



**H**



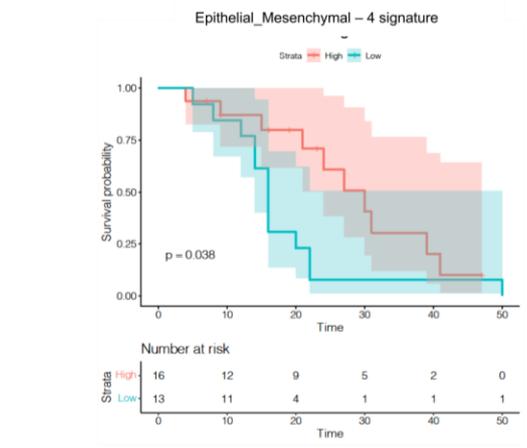
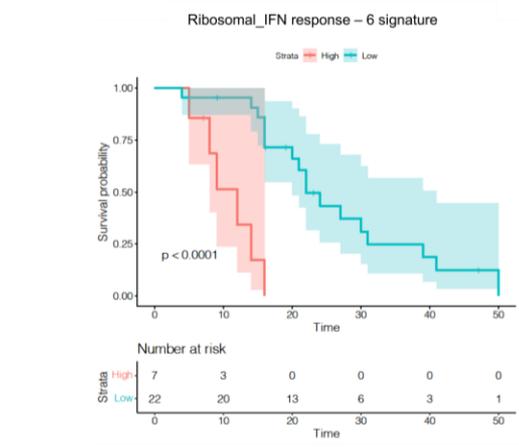
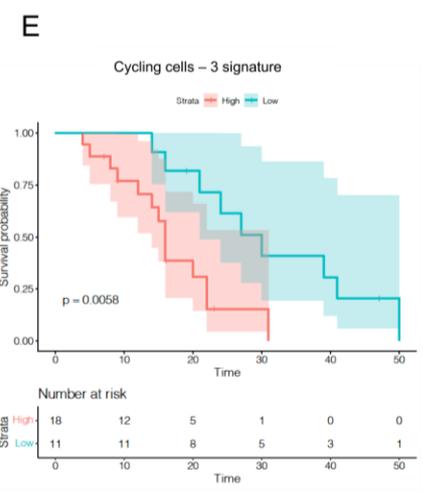
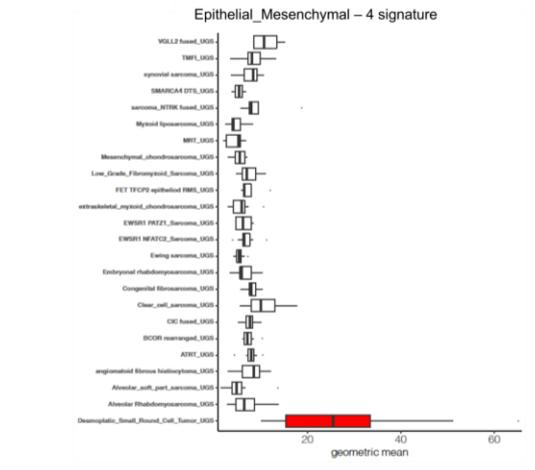
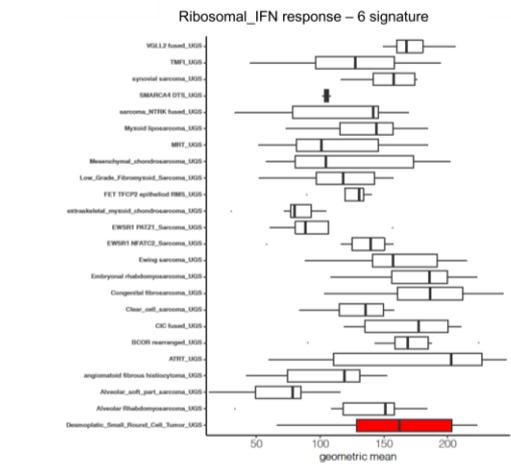
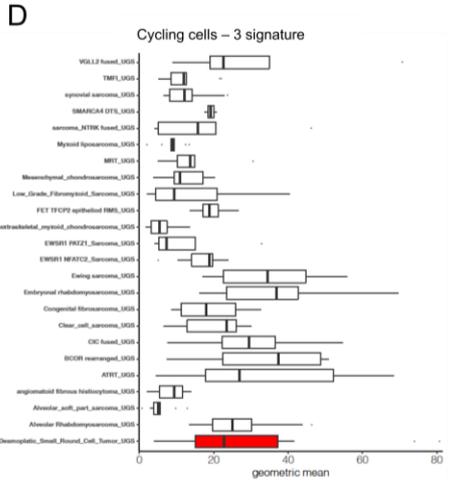
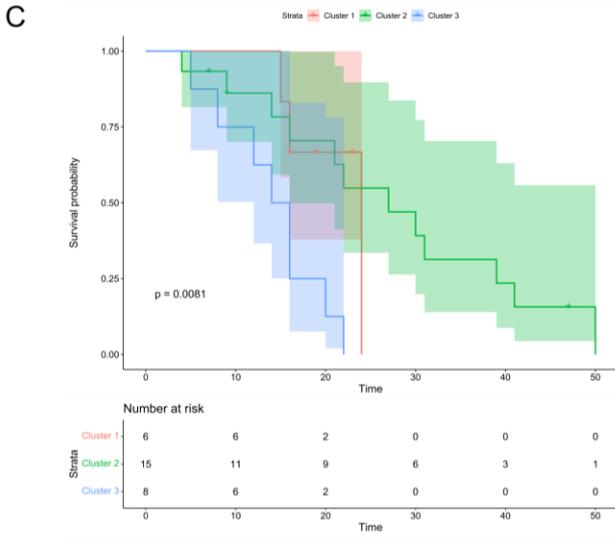
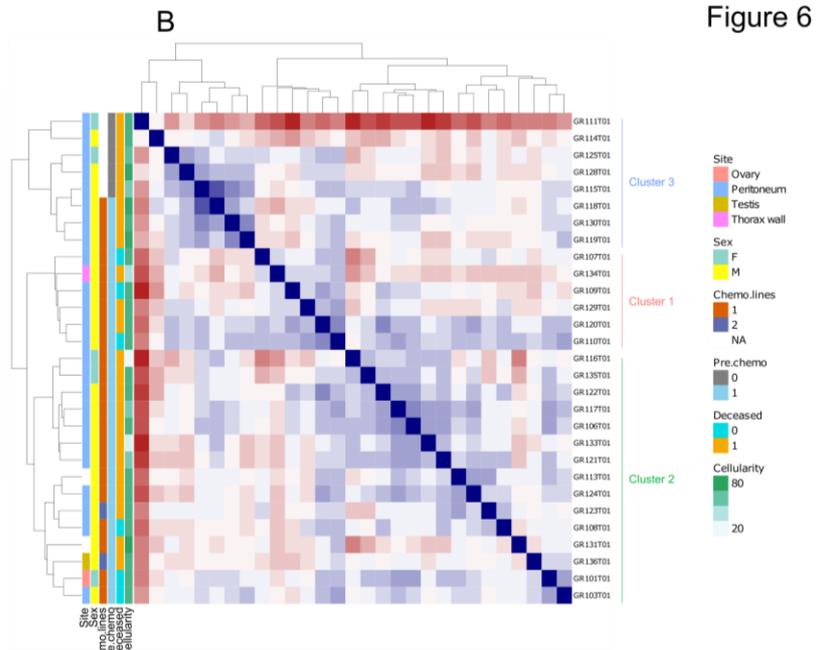
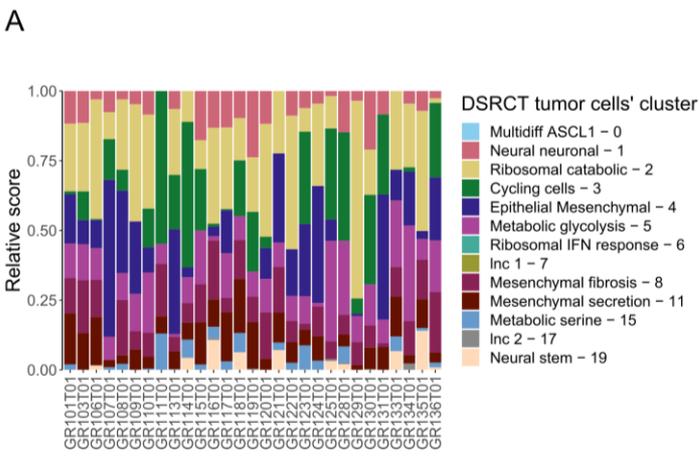
**I**



**Fig. 5 | DSRCT heterogeneity is linked to tumor spatial organization.**

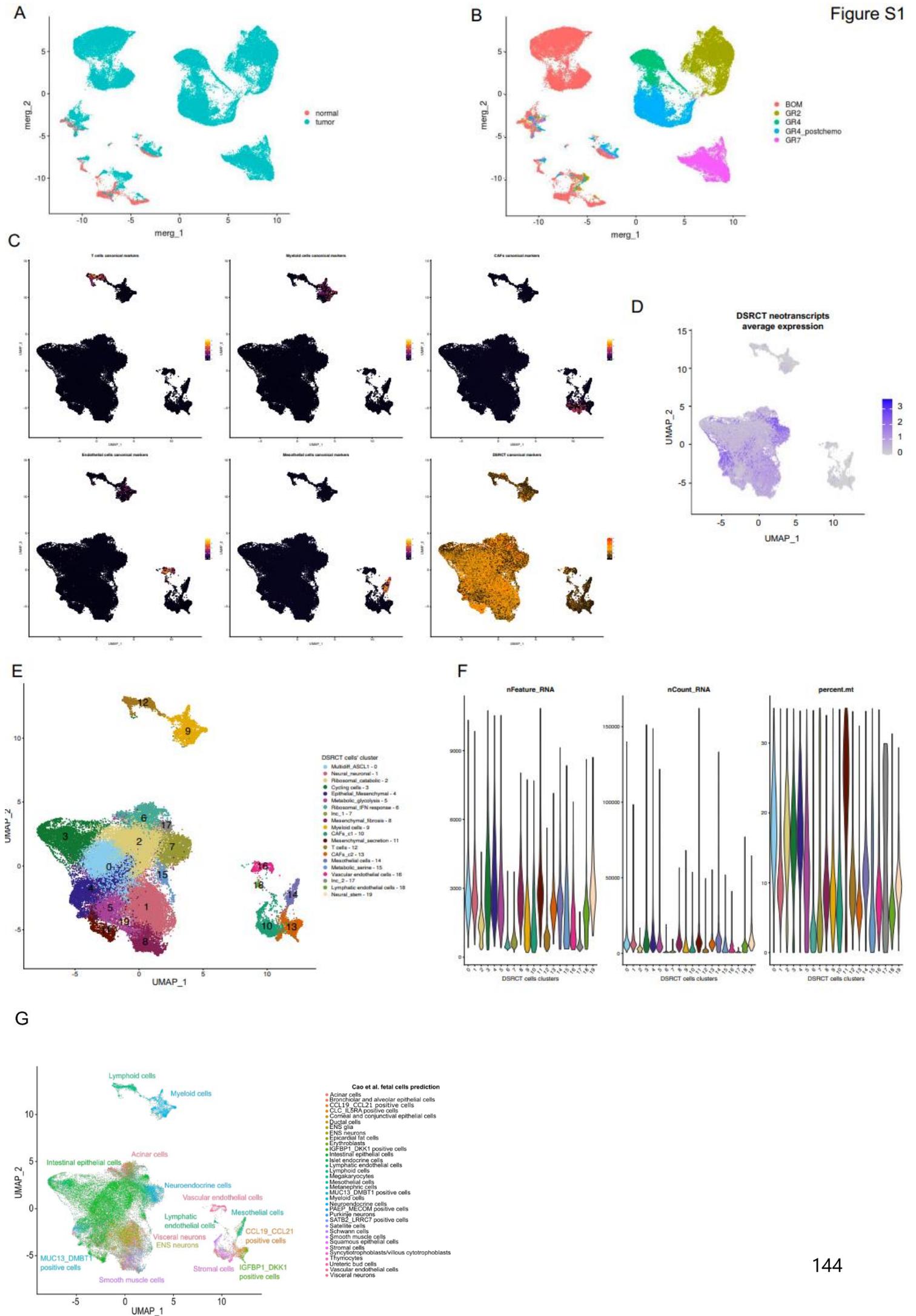
**A**, Immunofluorescent staining characterizing DSRCT tumor cells displaying a mesenchymal phenotype on *GR2* sample, characterized by immunohistochemistry (IHC) staining positivity for DES or CHI3L1, and located either at the periphery of organized DSRCT tumor cells islets or at the invasive tumor front, in direct proximity to cancer-associated fibroblasts (CAFs). **B**, Uniform Manifold Approximation and Projection (UMAP) showing *GR2* 3'scRNA-seq sample clustering (upper left panel), DES (upper right panel), CHI3L1 (bottom left panel), and THY1 (bottom right panel) expression levels. DSRCT tumor cells displaying a mesenchymal phenotype are grouped into *GR2* clusters 5, 9, 13, and 14 and are characterized by *DES* or *CHI3L1* overexpression, whereas CAFs grouped into *GR2* cluster 8 show a distinctive *THY1* expression. **C**, CellPhoneDB heatmap representing the numbers of ligand-receptor interactions between identified cell clusters in the *GR2* 3'scRNA-seq sample. DSRCT tumor cell clusters display the highest number of interactions with CAFs (*GR2* cluster 8) and are characterized by mesenchymal features (e.g., *GR2* clusters 5, 9, 13, and 14). **D**, *GR2* sample Visium spatial transcriptomics assay showing spot clusters defined after graph-based k-nearest neighbors' clustering, overlaid on H&E-stained slide. Tumor areas correspond to *Tum\_APOA1\_ITIH2\_HIST* – 0, *Tum\_NA* – 2, *Tum\_HIST\_CCK\_necrosis* – 3, *Tum\_mesenchymal\_glycolysis* – 5, and *Tum\_TAF11L2* - 8 clusters. While *Tum\_APOA1\_ITIH2\_HIST* – 0, and *Tum\_NA* – 2, are mainly localized within large DSRCT tumor cells' bundles, *Tum\_mesenchymal\_glycolysis* – 5 is mainly localized at the periphery of DSRCT tumor cells' islets, in contact with stromal areas. *Tum\_HIST\_CCK\_necrosis* – 3 is located within necrotic tumor areas. **E**, Heatmap showing single spots' expression of the top differentially expressed genes across identified spot clusters in *GR2* sample Visium spatial transcriptomics assay. High gene expression is shown in yellow, whereas low gene expression is shown in purple. **F**, H&E stained *GR2* slide (upper left panel), associated with the overlaid expression level of DSRCT tumor cell mesenchymal markers (*DES* (bottom left panel), *CHI3L1* (bottom right panel), and Cancer-associated Fibroblasts (CAFs) marker *THY1* (upper right panel)). *DES* and *CHI3L1* are mainly expressed in spots corresponding to DSRCT tumor cells invading stromal areas or at the periphery of tumor cells' islets. **G**, Gene signatures scores plotted on Visium spatial transcriptomics *GR2* sample assay for HALLMARK\_HYPOXIA (upper left panel), HALLMARK\_GLYCOLYSIS (bottom left panel), HALLMARK\_OXPHOS (bottom right

panel) and EWSR1-WT1 chromatin immunoprecipitation with sequencing (ChIP-seq) derived signature (upper right panel). **H**, Colony formation assay (CFA) of JN-DSRCT-1 cell line cocultured with (w/) or without (w/o) CAFs obtained from DSRCT patients-derived xenografts (PDXs). The colonies stained with crystal violet after two weeks of coculture are shown on the left panel. The percentage of well area covered by colonies assessed by ImageJ is displayed on the right panel. The coculture of JN-DSRCT-1 cells with PDX-derived CAFs significantly increases JN-DSRCT-1 colony formation.\*: p-value <0.05, t-test. **I**, JN-DSRCT-1 spheroids growth culture in the presence (w/) or absence (w/o) of DSRCT PDXs-derived CAFs. The median volume of the spheroids (triplicate assay) assessed on Day 0, Day 2, and Day 8 is shown on the y-axis. A.U: Arbitrary Unit.

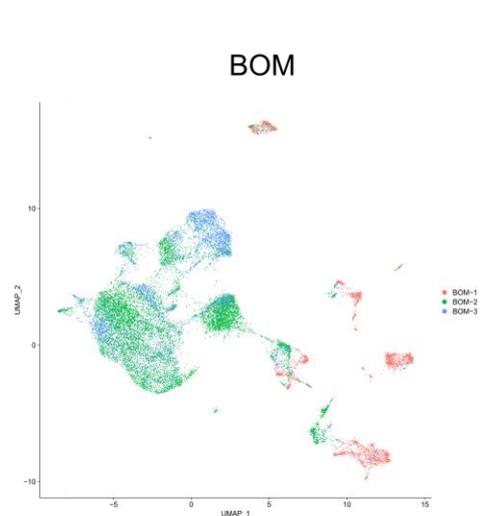
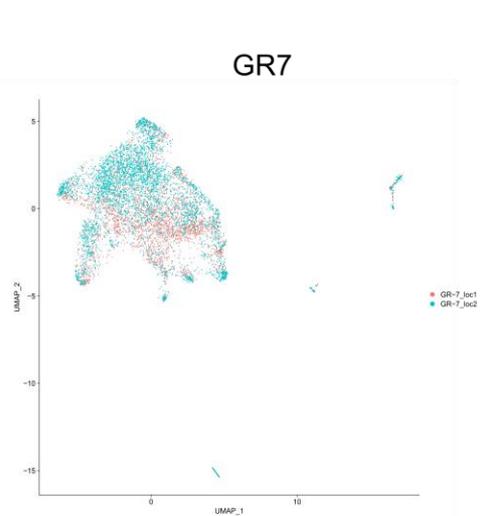
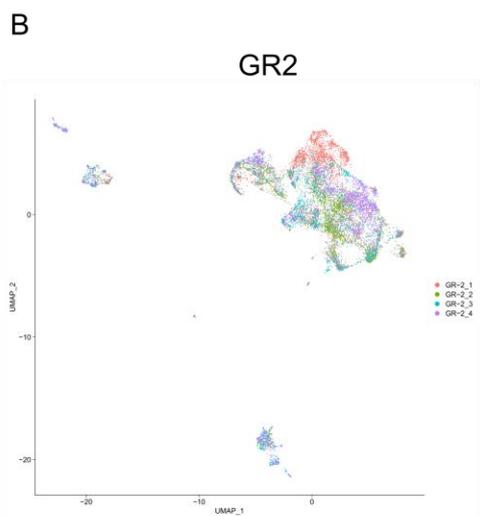
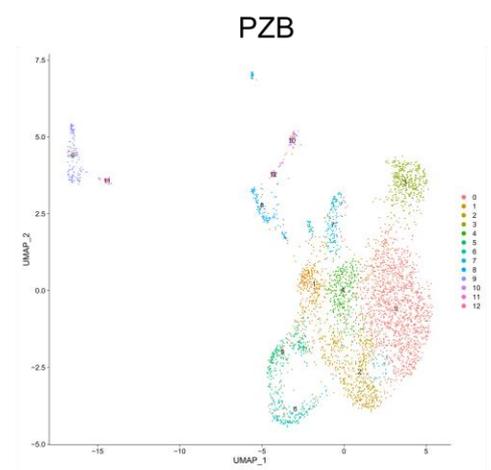
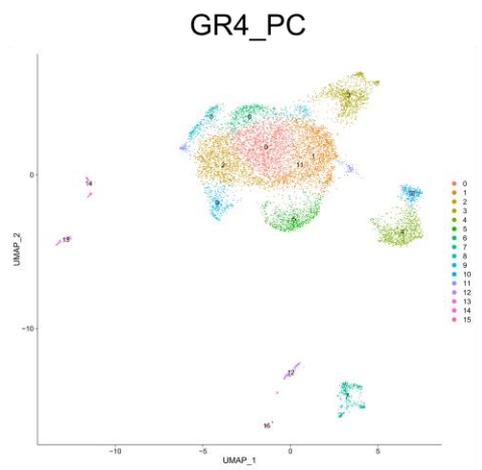
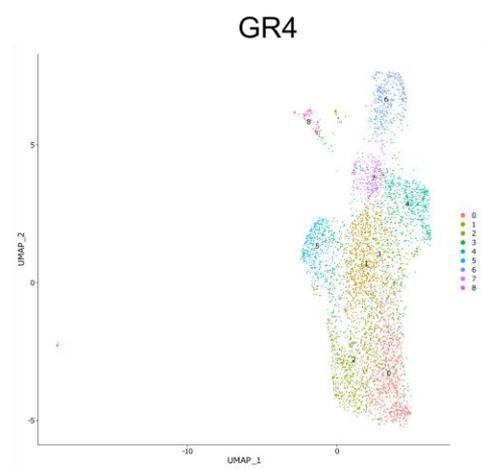
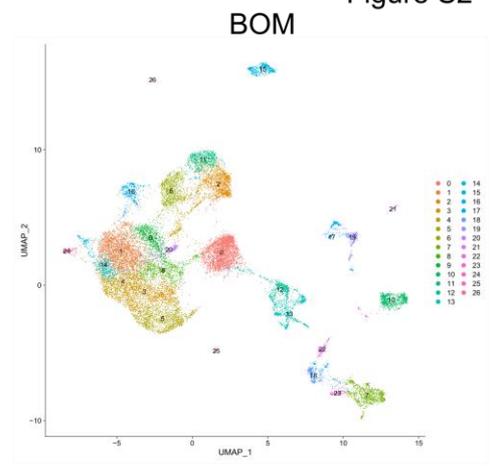
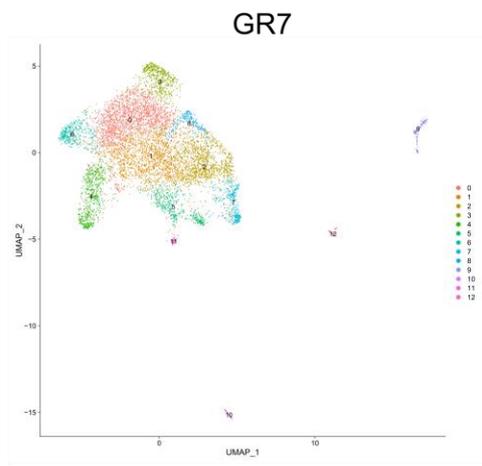
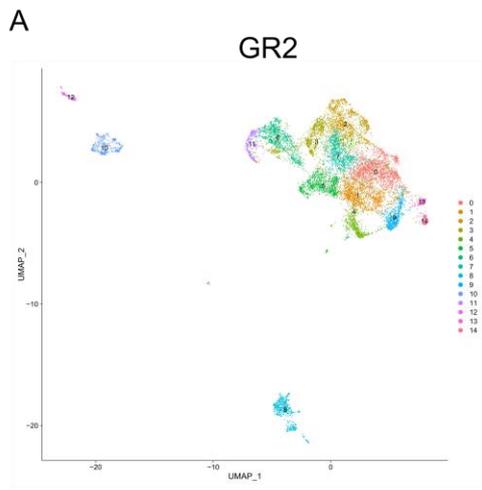


**Fig. 6 | ScRNA-seq-derived gene signatures characterize interpatient heterogeneity and define prognostic groups.** **A**, CIBERSORTx deconvolution of the Harmony-integrated 3'scRNA-seq dataset ("*Int\_sc*")-derived clusters on a cohort of 29 bulk RNA-seq DSRCT samples. **B**, Bulk RNA-seq hierarchical clustering of the DSRCT bulk RNA-seq cohort. Hierarchical clustering of samples was performed using a distance based on Pearson's correlation coefficient using Ward D2 linkage algorithm. DSRCT tumors cluster into three groups. Of note, all non-pretreated patients are grouped within Cluster #3. **C**, Kaplan Meier overall survival plot according to bulk RNA-seq hierarchical clustering-inferred groups. The x-axis shows time in months. "p" relates to the p-value according to a log-rank test. **D**, Selected 3'scRNA-seq-derived signatures on a subset of various sarcoma histological subtypes bulk RNA-seq data. These signatures are defined by the top 100 marker genes from each "*Int\_sc*"-derived cluster. While some signatures seem highly specific to DSRCT (e.g., *Epithelial\_mesenchymal* – 4), others delineate cancer cell common features (e.g., *Cycling cells* – 3, *Ribosomal\_IFN response* - 6). Additional signatures are shown in Supplemental Information – Fig. S16. **E**, Kaplan Meier overall survival plot according to previously defined signature scores in the bulk RNA-seq DSRCT cohort. *Cycling cells* – 3 and *Ribosomal\_IFN response* – 6 signatures, which are hallmarks of cell growth and proliferation, are associated with poorer prognosis. *Epithelial\_mesenchymal* – 4 signature represents a lineage-related state associated with improved prognosis. The x-axis shows time in months. "p" relates to the p-value according to a log-rank test.

## **Supplemental Information – Supplementary figures and legends**



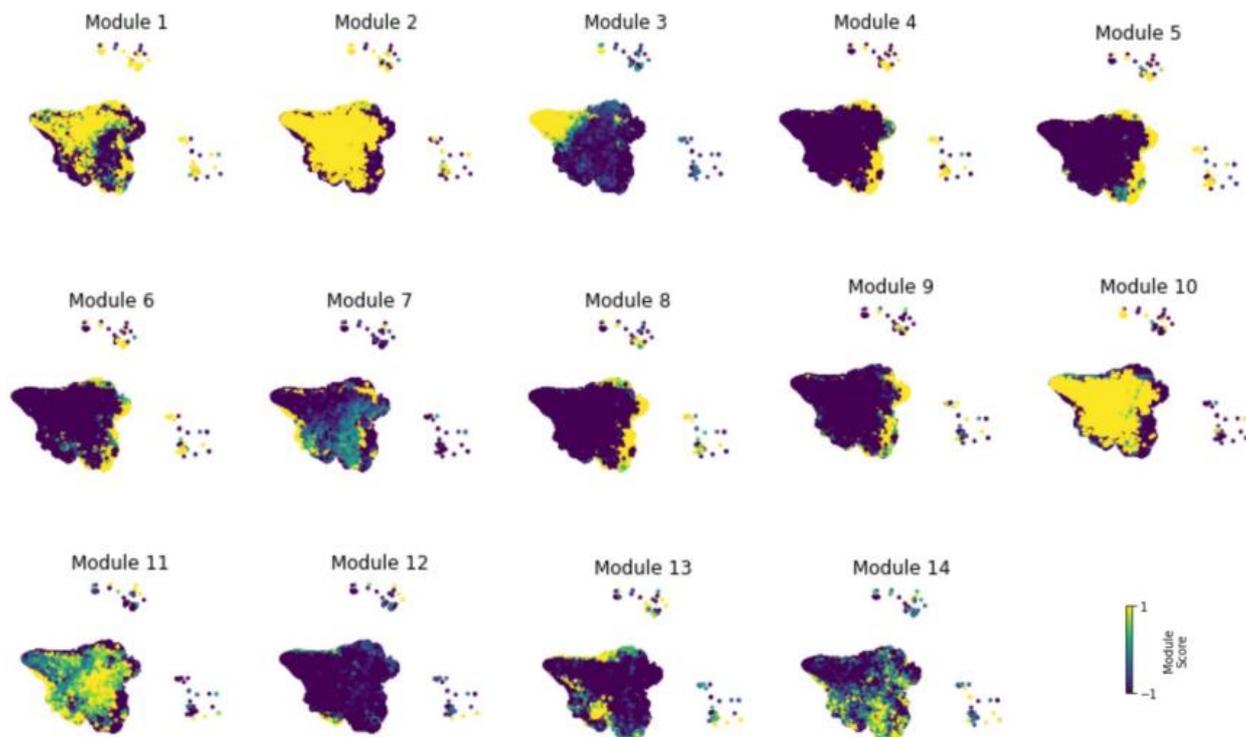
**A**, Uniform Manifold Approximation and Projection (UMAP) highlighting malignant and non-malignant cells on the merged individual 3'scRNA-seq datasets (n=11). **B**, UMAP showing the patient of origin from each individual dataset. **C**, UMAP highlighting the relative expression on the Harmony-integrated dataset ("*Int\_sc*" dataset) of canonical markers for each specified cell population; from top to bottom, left to right: T cells, myeloid cells, cancer-associated fibroblasts, endothelial cells, mesothelial cells, and DSRCT malignant cells. **D**, UMAP highlighting DSRCT-specific neotranscripts average expression on the "*Int\_sc*" dataset. **E**, UMAP recapitulating "*Int\_sc*" cell clusters based on a gene expression-based k-nearest neighbors clustering. **F**, Violin plots representing the distribution of the number of detected genes (nFeature\_RNA), the number of RNA molecules (nCount\_RNA), and the percent of mitochondrial gene counted (percent.mt) per cell in each cell cluster identified on the "*Int\_sc*" dataset. **G**, UMAP showing the "*Int\_sc*" clusters' annotation after label transfer using a fetal development atlas from Cao et al.



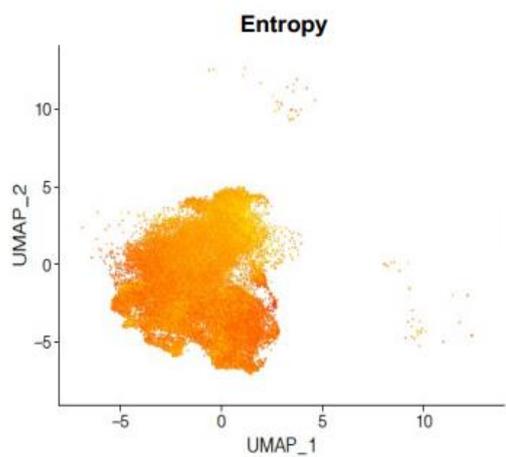


**Supplementary Fig. S2 | scRNA-seq recapitulates DSRCT heterogeneity at the interpatient and inpatient levels. A**, UMAP outlining DSRCT cell clusters based on a gene expression-based k-nearest neighbors clustering for the different individual patient datasets (or merged datasets for patients with synchronous tumor sites). **B**, UMAP representing the site of origin for patients with synchronous tumor site samples (left to right *GR2*, *GR7*, and *BOM* datasets). **C**, Heatmaps representing the top 10 marker genes for all graph-based defined clusters for each individual or site-merged dataset.

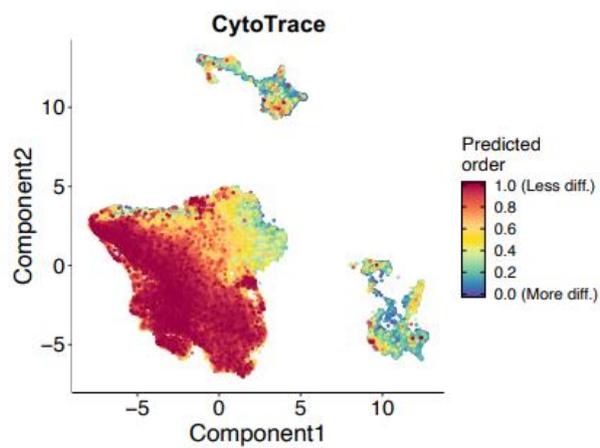
A



B

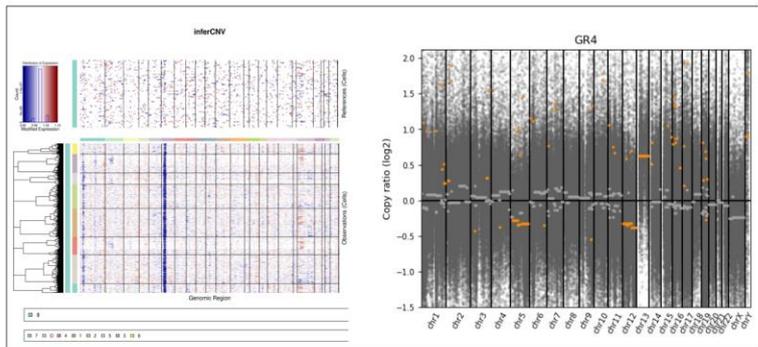


C

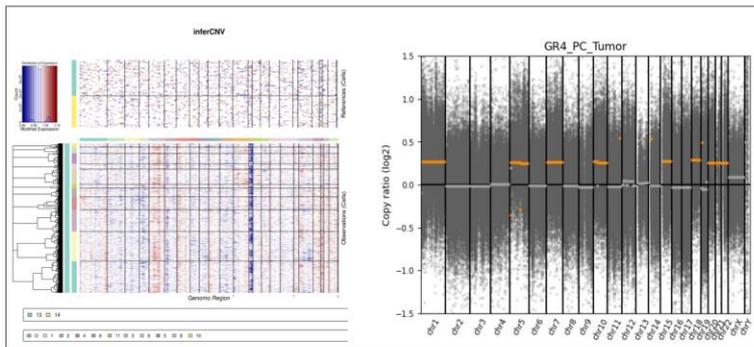


**Supplementary Fig. S3 | DSRCT malignant cells subpopulation profiles suggest cell plasticity.** **A**, UMAP highlighting HotSpot co-expressed module scores on the Harmony integrated dataset (“*Int\_sc*” dataset). **B**, UMAP showing inferred cell-based transcriptome entropy computed using StemID tool on the “*Int\_sc*” dataset, after removal of bias-prone clusters (i.e., lncRNA-enriched, ribosomal protein-coding genes-enriched, cell-cycle-related clusters). **C**, UMAP highlighting the differentiation degree prediction on the “*Int\_sc*” dataset using the CytoTRACE tool.

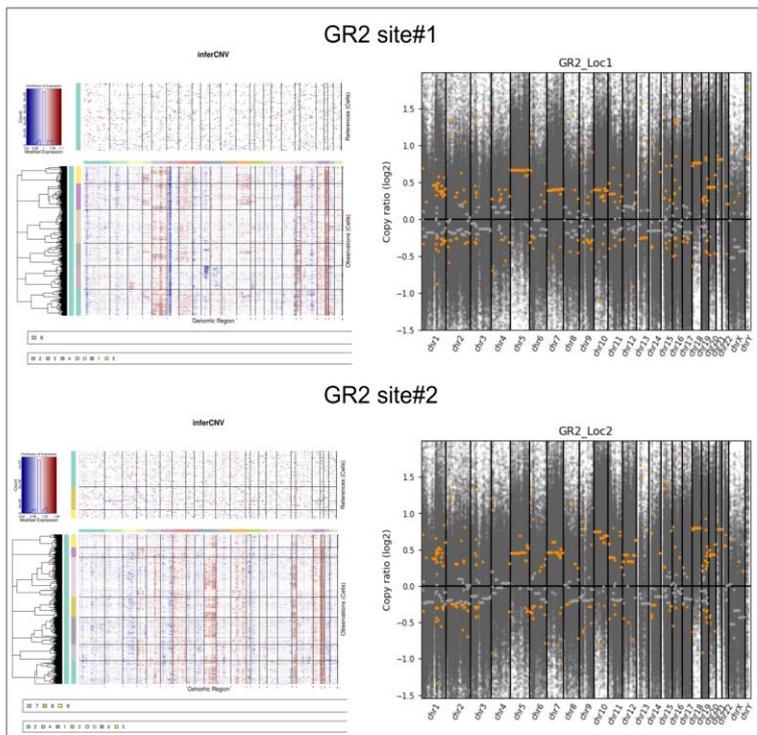
GR4



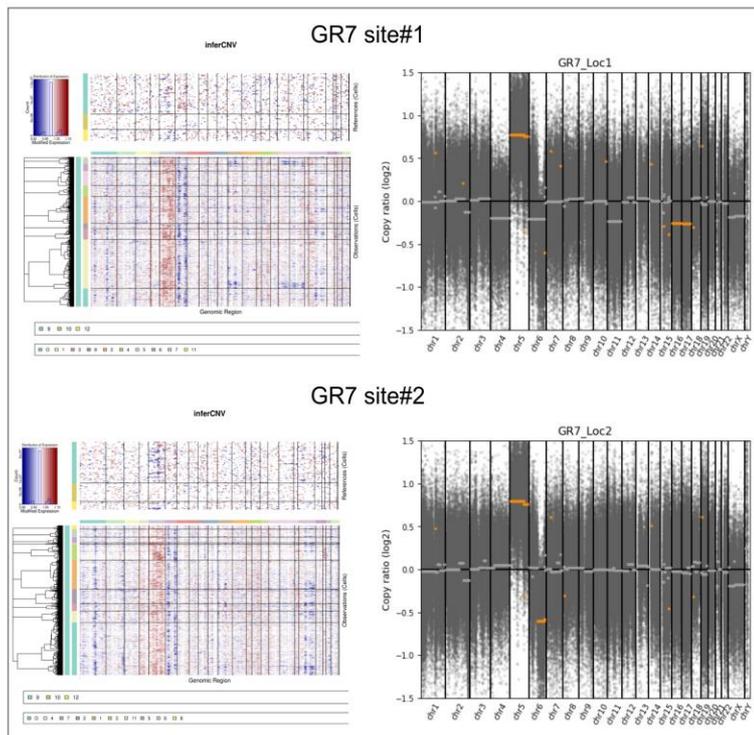
GR4\_PC



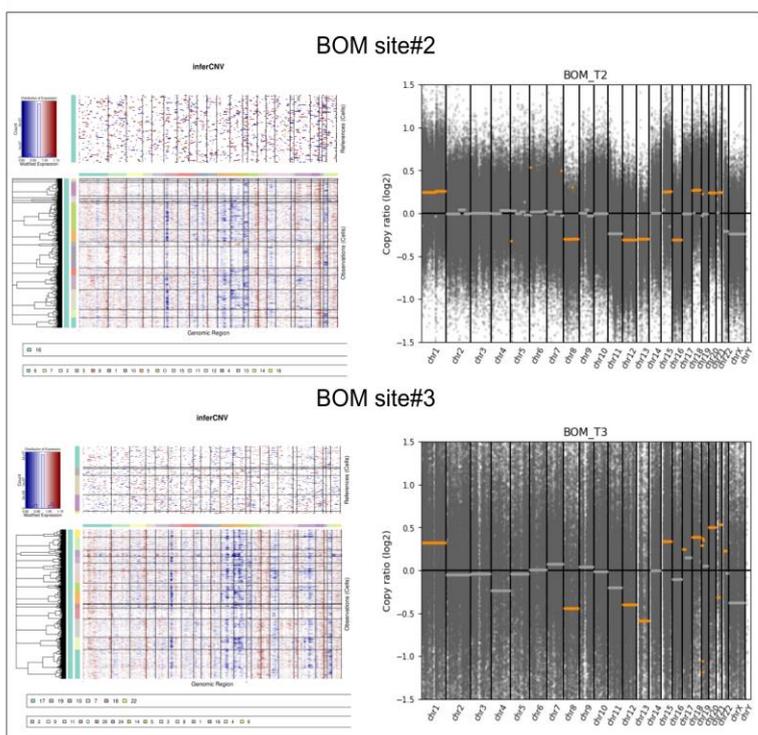
GR2



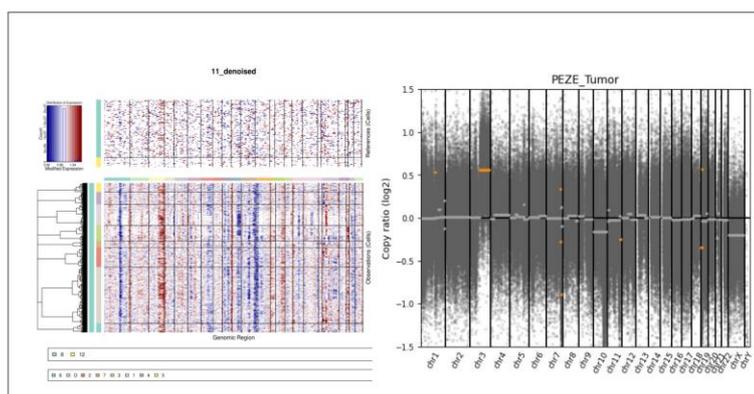
GR7



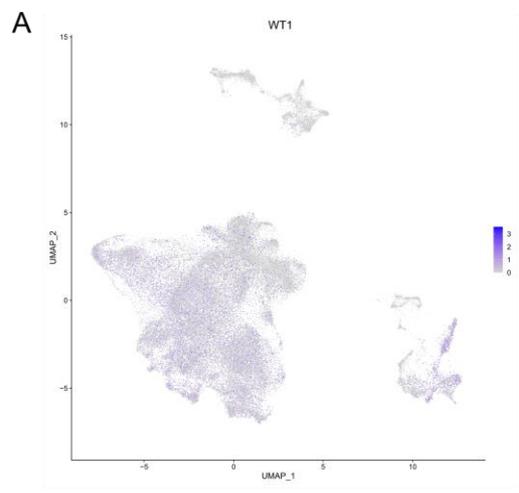
BOM sample



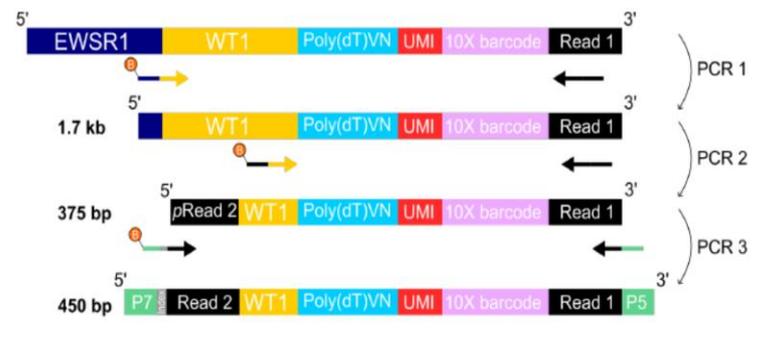
PZB



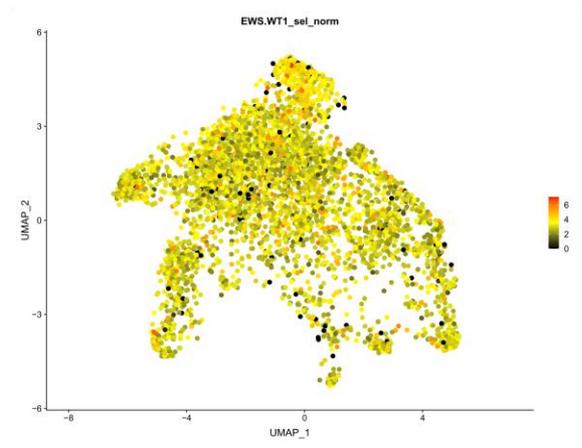
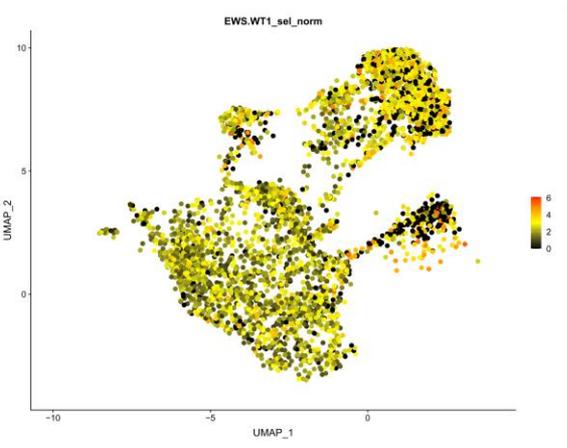
**Supplementary Fig. S4 | Copy number variation (CNV) inference on individual or site-merged 3'scRNA-seq datasets.** Heatmaps representing copy number variation inference from 3' scRNA-seq data on patient samples with InferCNV package. For each patient sample and different tumor sites, when relevant, the heatmap represents the CNV estimation for each genomic position and each cell amongst the dataset. Cells are ordered by cluster. CNV is homogeneous across DSRCT tumor cell clusters and inpatient tumor sites. On the right panels is plotted the  $\log_2(\text{copy ratio})$  for each chromosome location according to Whole Exome Sequencing (WES) on corresponding tumor samples controlled on matched germline samples.



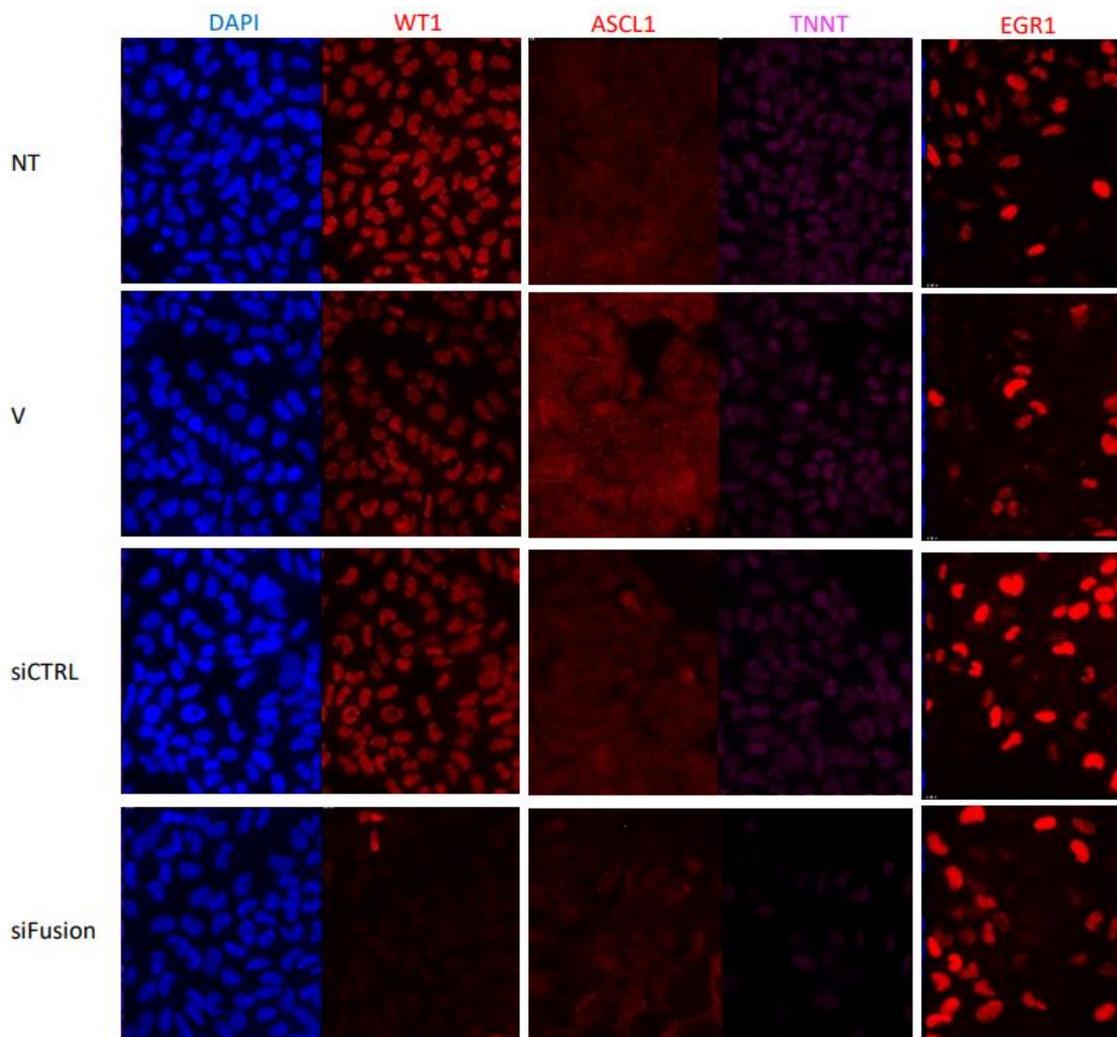
**B**



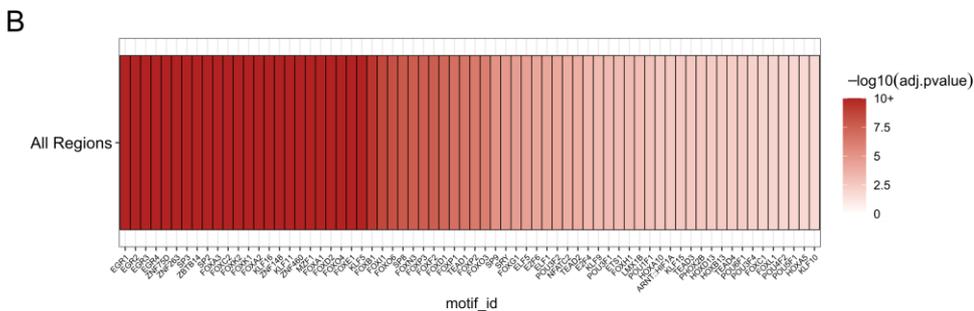
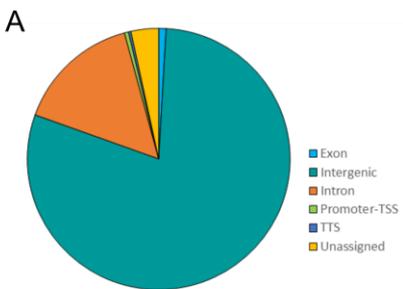
**C**



**Supplementary Fig. S5 | EWSR1-WT1 fusion transcript mRNA is not significantly differentially expressed among cell clusters.** **A**, UMAP showing “WT1” expression within Harmony integrated dataset (“*Int\_sc*” dataset). “WT1” corresponds to both *EWSR1-WT1* and full-length *WT1* transcripts. **B**, Schematic representation of the in-house developed method aiming at interrogating *EWSR1-WT1* fusion transcript expression level at the single-cell resolution. The method relies on a sequential PCR-protocol to enrich the library for *EWSR1-WT1*-specific barcoded cDNAs from 10X Genomics 3’ scRNA-seq. **C**, UMAP highlighting *EWSR1-WT1*-specific expression level on single cells from *BOM site#2* and *GR7 site#2*.

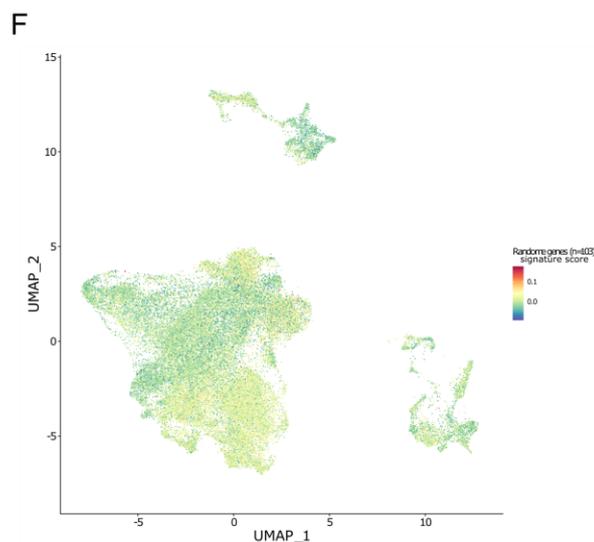
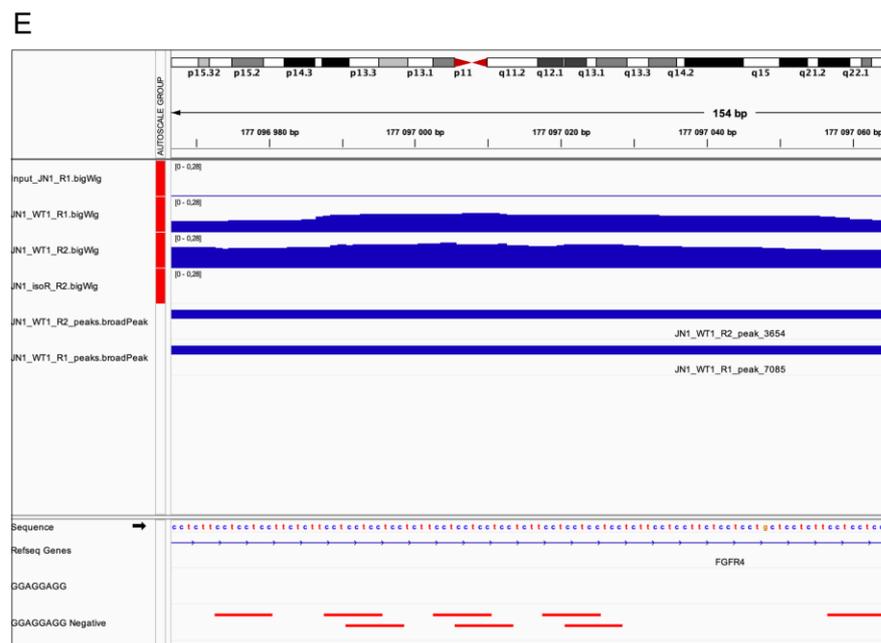
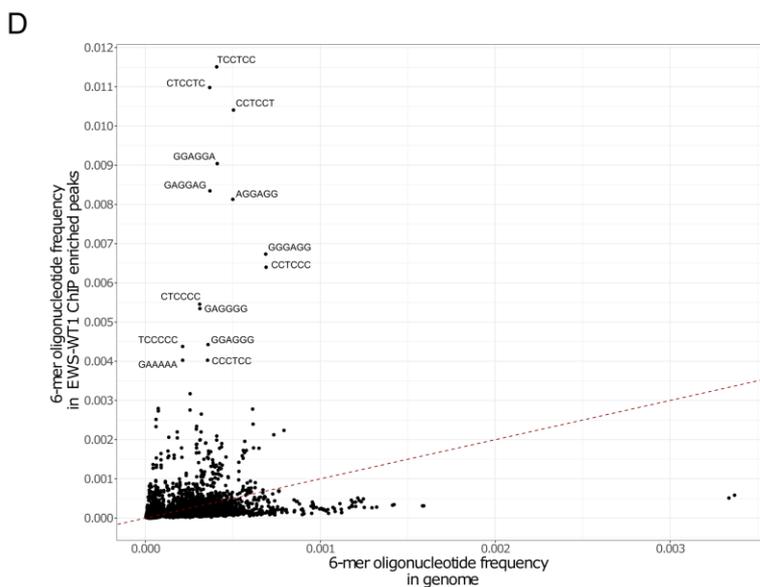


**Supplementary Fig. S6 | JN-DSRCT-1 cells express canonical markers from various 3'scRNA-seq-inferred clusters.** Representative immunofluorescence images for WT1 (C-terminal), ASCL1, TNNT, and EGR1 staining in JN-DSRCT-1 cells non-transfected (NT), mock-transfected (V), non-targeting siCTRL-transfected and siEWSR1-WT1 transfected (siFusion). DAPI staining is used as a nuclear counter-staining.

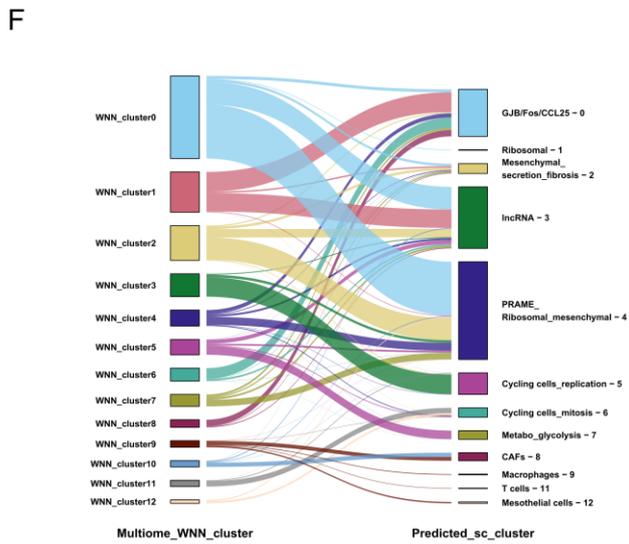
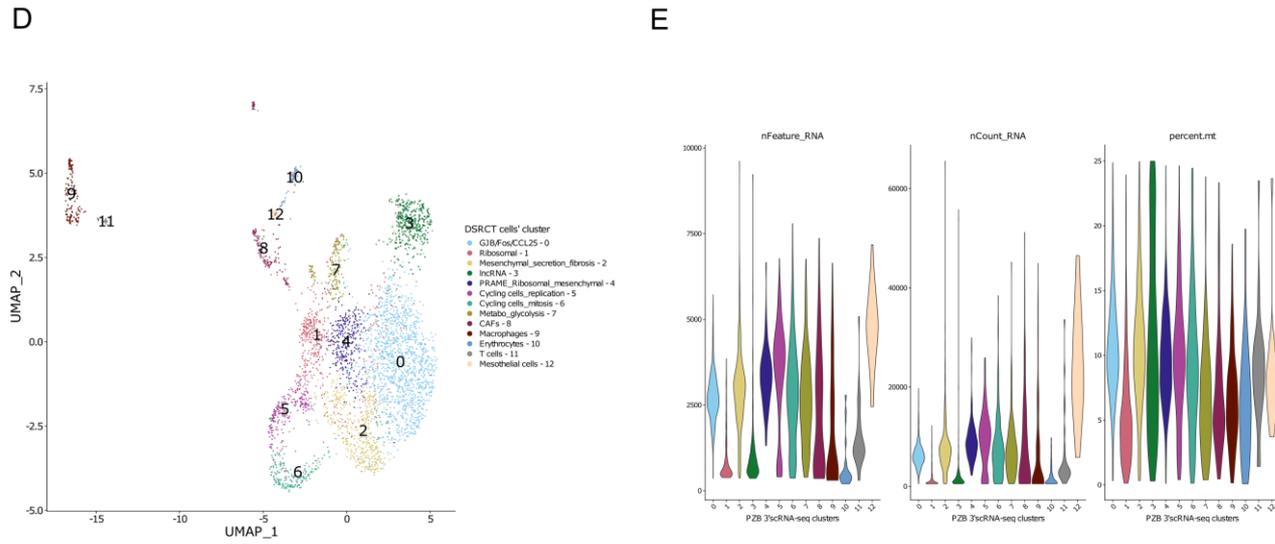
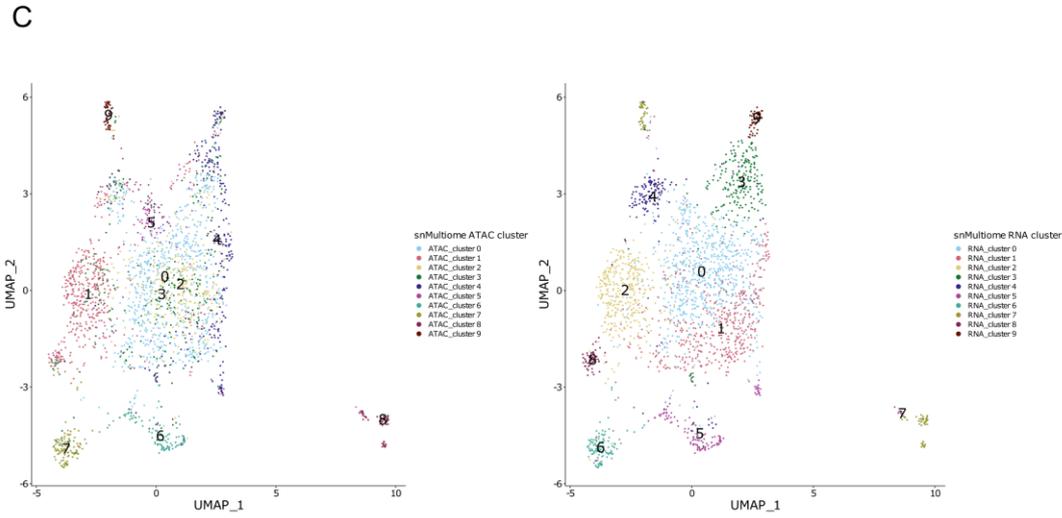
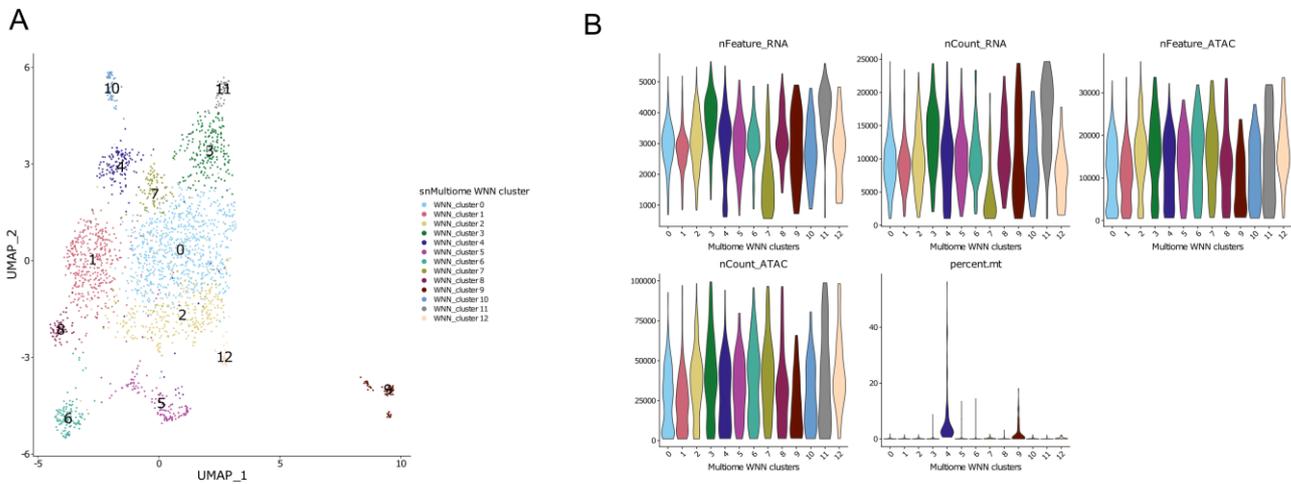


**C**

Rank	EWS-WT1 ChIP de novo motif logo	Enrichment adjusted p-value	EWS-WT1 ChIP de novo motif ID	Best match motif	Best match adjusted p-value
Rank 1		1.84E-151m1	GCGKGGGAGGVRGV	EGR1 [RC]	7.36E-07
Rank 2		9.36E-47m5	CCCACGCA	EGR2	0.00183
Rank 3		5.84E-29m7	GGAGGAGRAGGAGAA	ZNF263	0.0394
Rank 4		6.37E-29m2	GGAGGAGGAGRAAGA	ZNF263	0.0554
Rank 5		1.72E-18m8	CCCTCCTCTCTCC	ZNF263 [RC]	0.477
Rank 6		1.91E-11m3	AARTAAAYA	FOXO2	0.00211
Rank 7		0.000665m4	GAGCCACTTCC	ZNF528	0.897
Rank 8		0.887m6	ACCTGKCT	MYO1D1	0.0373

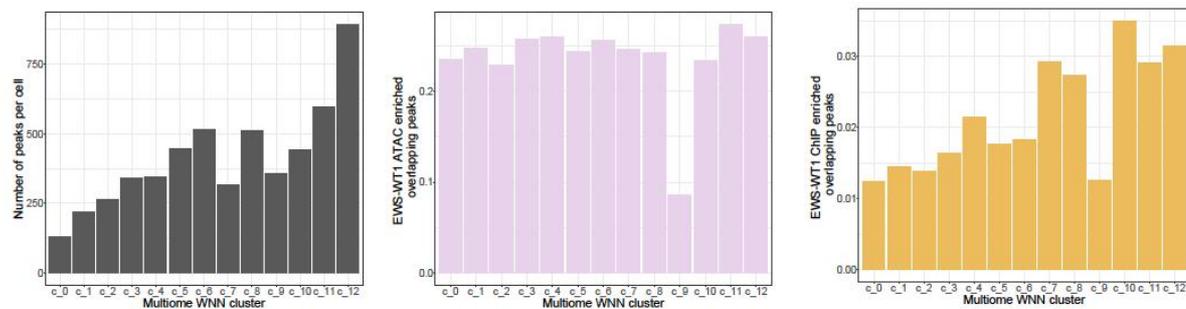


**Supplementary Fig. S7 | EWSR1-WT1 activity is linked to DSRCT cell chromatin landscape modifications.** **A**, Pie chart representing EWSR1-WT1 ChIP-seq peaks' frequency within the genome regions (Exons, Introns, Intergenic, Promoter-TSS (Promoter-Transcription Start Site), TTS (Transcription Termination Site) and unassigned). **B**, Heatmap showing enriched motifs within EWSR1-WT1 ChIP-seq peaks' regions. Motifs were extracted from the JASPAR2020 database. Motifs are ranked from left to right according to  $-\log_{10}(\text{adjusted p-value})$ . **C**, Chart recapitulating significantly enriched *de novo* EWSR1-WT1 binding motifs and their best match with known transcription factor binding sites. Consensus motifs are depicted with the adjusted p-values for the best match comparison. **D**, Scatter plot representing the different possible 6-mers frequencies within the EWSR1-WT1 ChIP-seq peaks' regions (y-axis) according to their frequency in the whole genome (x-axis). Specific enrichment for GGA/CCT repeats is highlighted on the plot. **E**, Integrated Genome Viewer (IGV) track screenshot of EWSR1-WT1 ChIP-seq peaks-read coverage centered on FGFR4 genomic region. Tracks for Input, two EWSR1-WT1 ChIP replicates, and one isotype ChIP are depicted. EWSR1-WT1 binding site maps to GGA/CCT microsatellites. **F**, UMAP highlighting a random genes-based signature on harmony integrated 3'scRNA-seq ("*Int\_sc*"). This signature is used as a negative control for the EWSR1-WT1 ChIP signature.

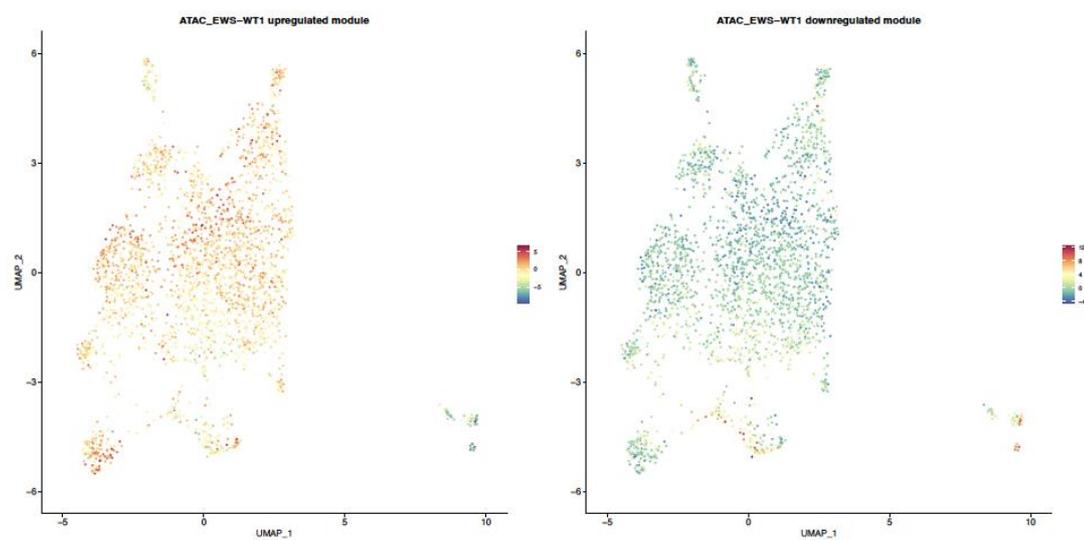


**Supplementary Fig. S8 | Single nucleus Multiome (snMultiome) analysis.** **A**, UMAP representing the Weighted-nearest neighbors (WNN) clustering based on ATAC and gene expression features on the single nucleus Multiome (snMultiome) dataset from *PZB* patient. **B**, Violin plots representing the distribution of the number of genes (nFeature\_RNA), the number of detected RNA molecules (nCount\_RNA), the number of ATAC-seq reads (nFeature\_ATAC), the number of ATAC-seq peaks (nCount\_ATAC) and the percent of mitochondrial genes counted (percent.mt) per cell in each defined WNN cluster on the *PZB* dataset. **C**, UMAP representing clusters defined by ATAC features (left) or Gene Expression (RNA) features (right) only. **D**, UMAP recapitulating DSRCT cell clusters based on a gene expression-based k-nearest neighbors clustering on the *PZB* 3' scRNA-seq dataset. **E**, Violin plots representing the distribution of the number of genes (nFeature\_RNA), the number of detected RNA molecules (nCount\_RNA), and the percent of mitochondrial genes counted (percent.mt) per cell in each cell cluster identified on the *PZB* 3' scRNA-seq dataset. **F**, Label prediction after label transfer of *PZB* snMultiome WNN clusters on *PZB* 3' scRNA-seq clusters' reference. The *PZB* 3'scRNA-seq dataset included only intronic reads to increase nuclei-to-cells comparability.

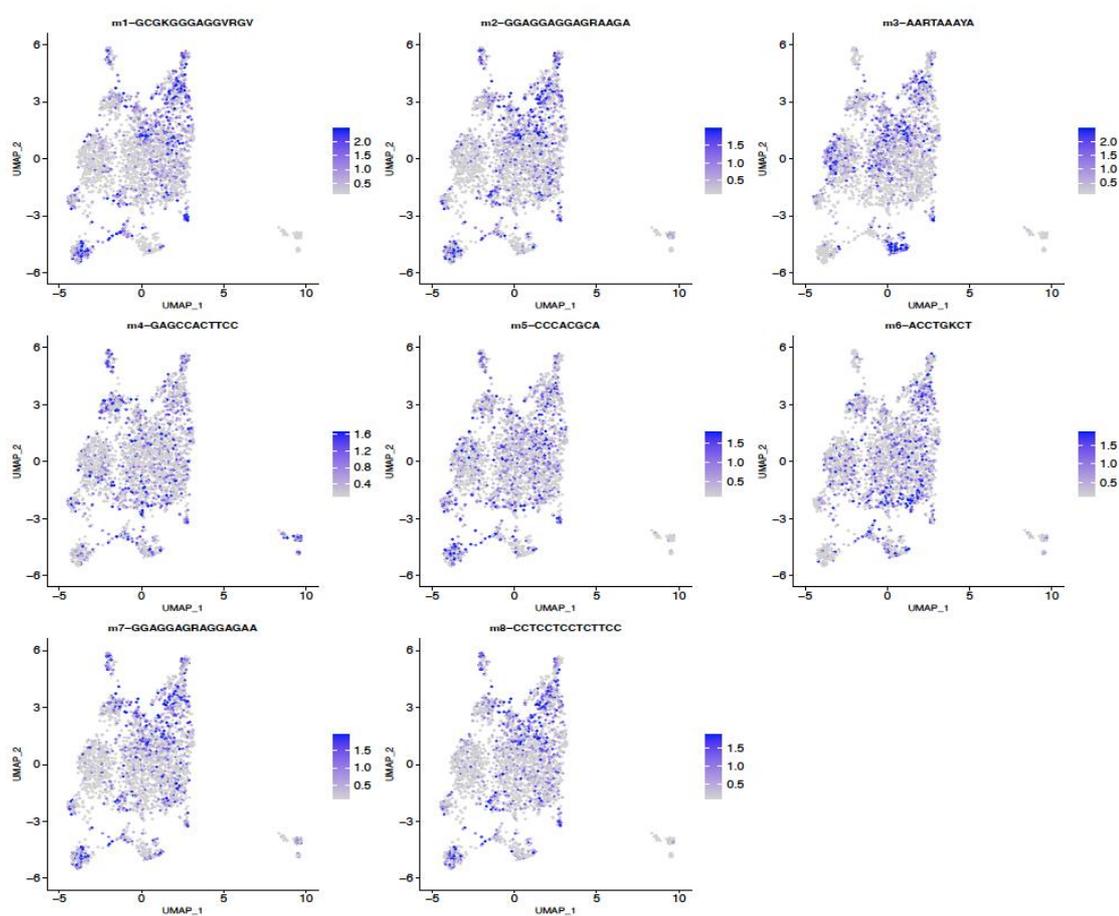
A



B

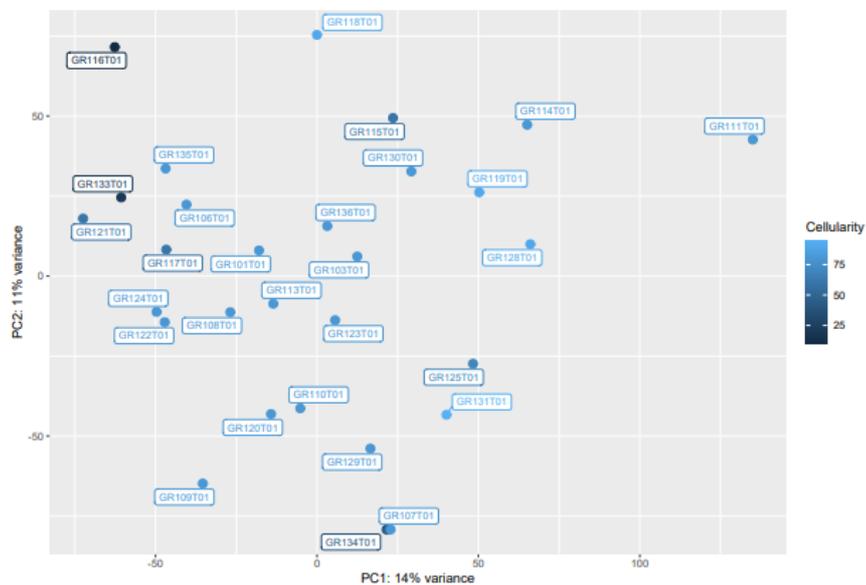


C

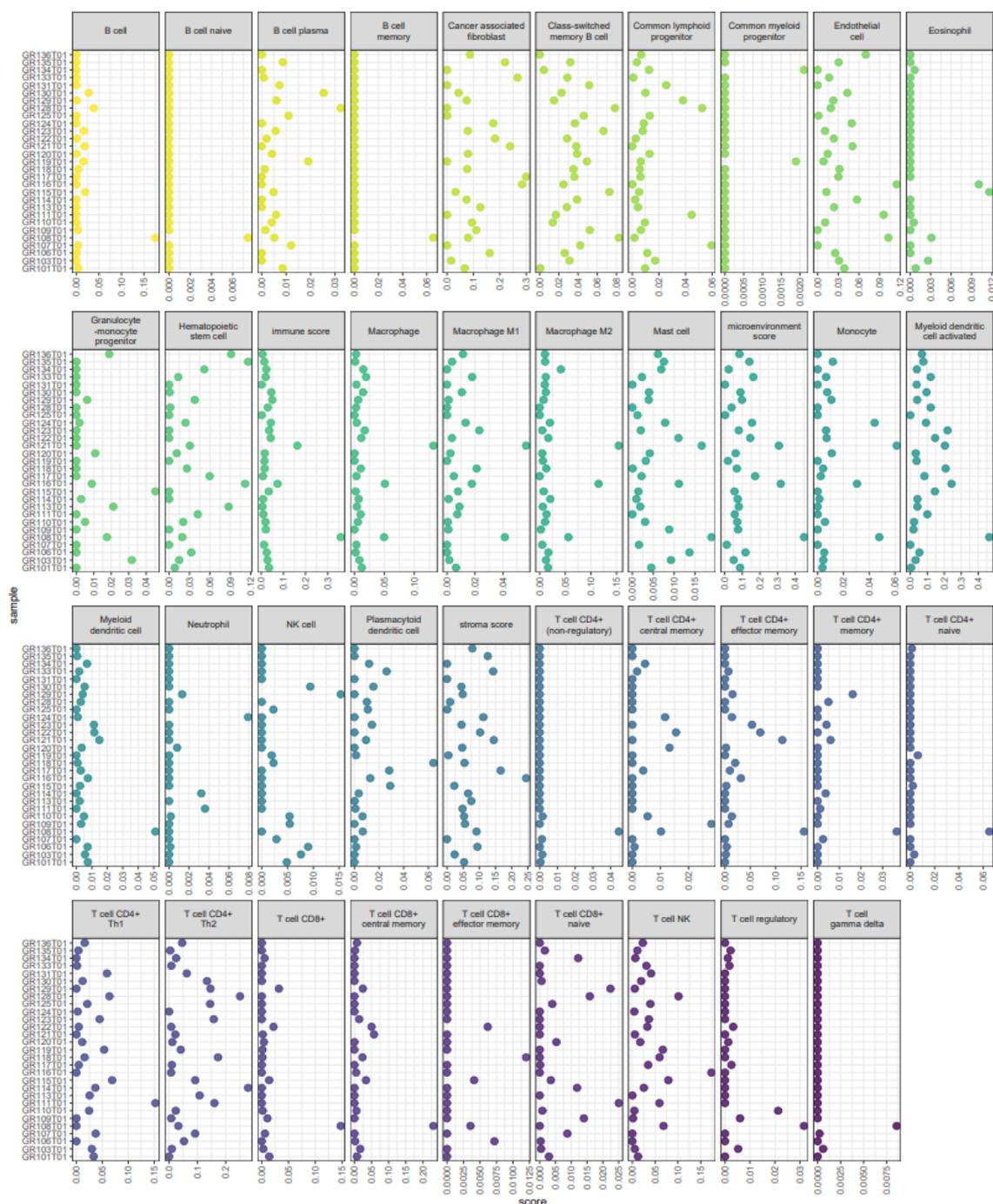


**Supplementary Fig. S9 | Single-nucleus Multiome (snMultiome) reveals distinct chromatin accessibility across clusters. A,** Bar plot representing the number of ATAC peaks called per cell (left), the proportion of total ATAC peaks called per cell overlapping with EWSR1-WT1 *ON* ATAC-enriched peaks (middle), and the proportion of total ATAC peaks called per cell overlapping with EWSR1-WT1 ChIP-enriched peaks within the identified clusters from the snMultiome dataset. **B,** EWSR1-WT1 *ON* (left panel) and *OFF* (right panel) ATAC modules signature plotted on the snMultiome UMAP. *WNN\_cluster 5* (hypoxia-related tumor cells) and *WNN\_cluster 8* (non-malignant cells) display the lowest EWSR1-WT1 *ON* module score and the highest EWSR1-WT1 *OFF* module score. **C,** UMAP representing the activity of EWSR1-WT1 ChIP-inferred *de novo* motifs within the snMultiome dataset. The motif activity was computed using ChromVAR tool. Most *de novo* motifs have higher activity in cycling cells and *WNN\_cluster 6*, except for the m3\_AARTAAAYA motif, which has higher activity in *WNN\_cluster 5*. The latter motif displays significant similarity with FOXC2 binding motif (**Supplementary Fig. S7C**).

A

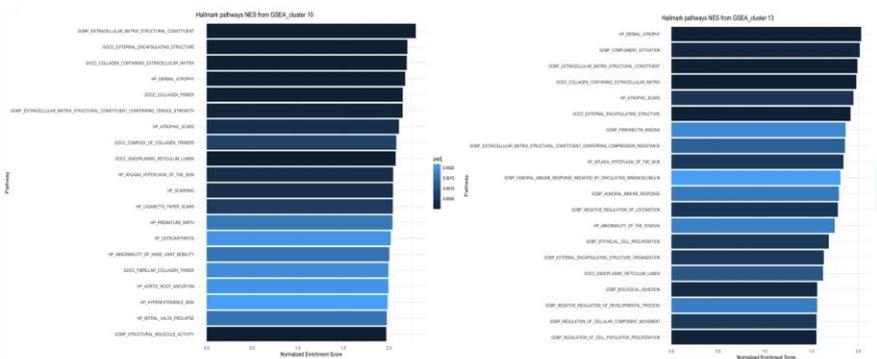


B

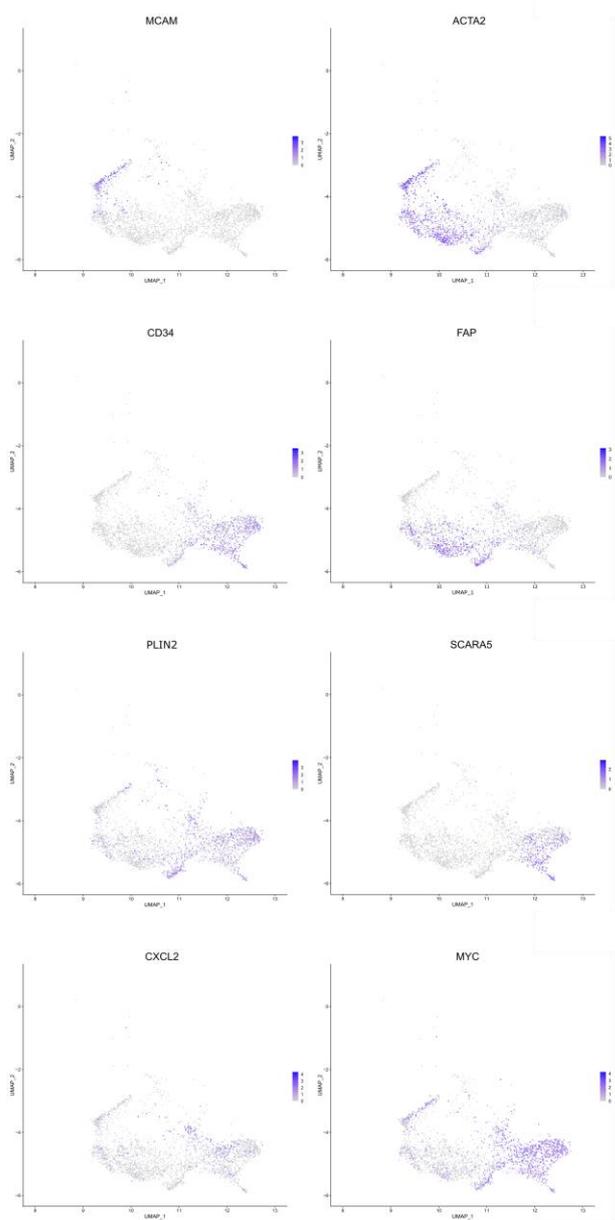


**Supplementary Fig. S10 | Bulk RNA-seq DSRCT cohort and immune landscape deconvolution.** **A**, Principal component analysis (PCA) of the DSRCT bulk RNA-seq dataset from 29 frozen DSRCT samples. The sample cellularity is represented for each sample on the blue color scale. **B**, Immune cell subpopulation deconvolution (Immune xCell ) on the DSRCT bulk RNA-seq cohort.

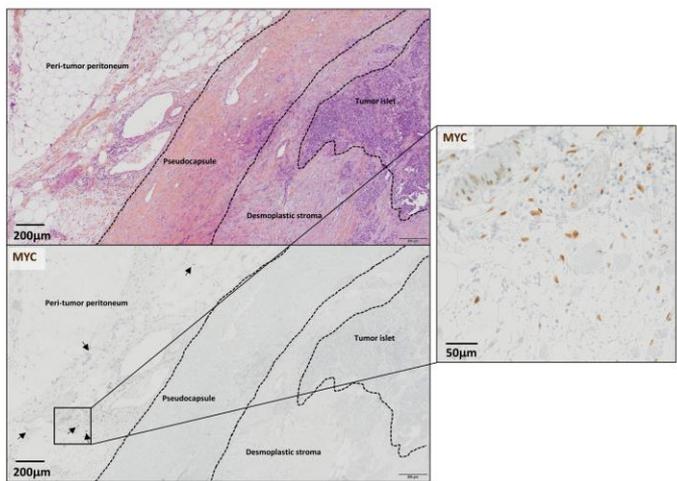
A



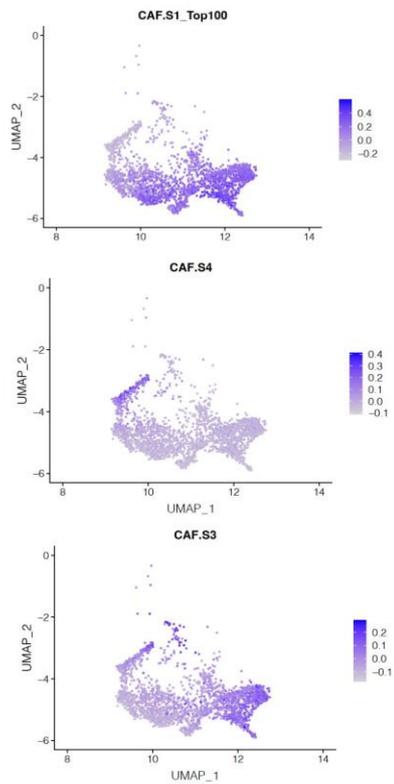
B



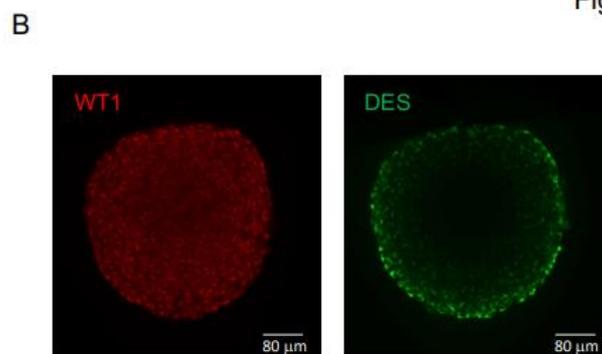
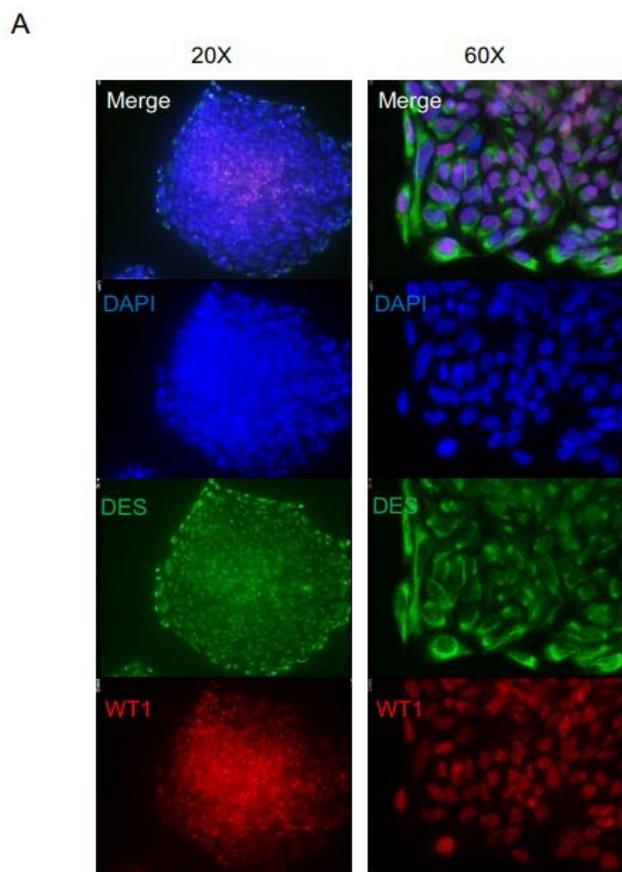
C



D

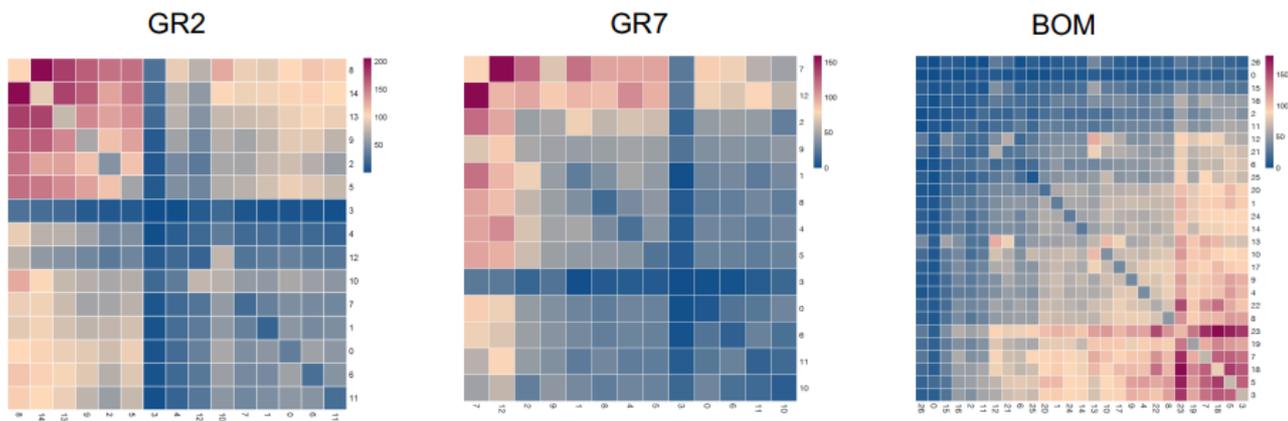


**Supplementary Fig. S11 | DSRCT Cancer-Associated Fibroblasts (CAFs) divide into distinct subclusters.** **A**, Gene Ontology enrichment analysis on both CAF clusters (*CAFs\_c1 – 10* and *CAFs\_c2 – 13* clusters) identified in the Harmony-integrated 3'scRNA-seq ("*Int\_sc*") dataset. Graphs represent the number of genes enriched in each GO term for both CAF clusters. The top 20 GO pathways are highlighted for each cluster. **B**, UMAP specifically representing CAFs identified from the "*Int\_sc*" dataset and highlighting DSRCT CAF-specific markers (*MCAM*, *ACTA2*, *CD34*, *FAP*, *PLN2*, *SCARA5*, *CXCL12*, and *MYC*). **C**, Representative image of the juxta-tumor area from a DSRCT FFPE sample. HES staining (upper panel) was used to define regions of interest (juxta-tumor peritoneum, pseudocapsule, desmoplastic stroma, and tumor islets). MYC immunohistochemistry (IHC) staining is visualized on the lower panel and with a higher magnification on the right panel. MYC-positive cells correspond to adipocytic-like cells and may correlate with the *CAFs\_c2 – 13* cluster showing lipofibroblast features. **D**, UMAP specifically representing CAFs identified from the "*Int\_sc*" dataset and highlighting three different CAF signatures: CAF-S1 (top 100, upper panel), CAF-S4 (middle panel), and CAF-S3 (lower panel) from Mehta-Grigoriou et al. breast cancer CAFs signatures.



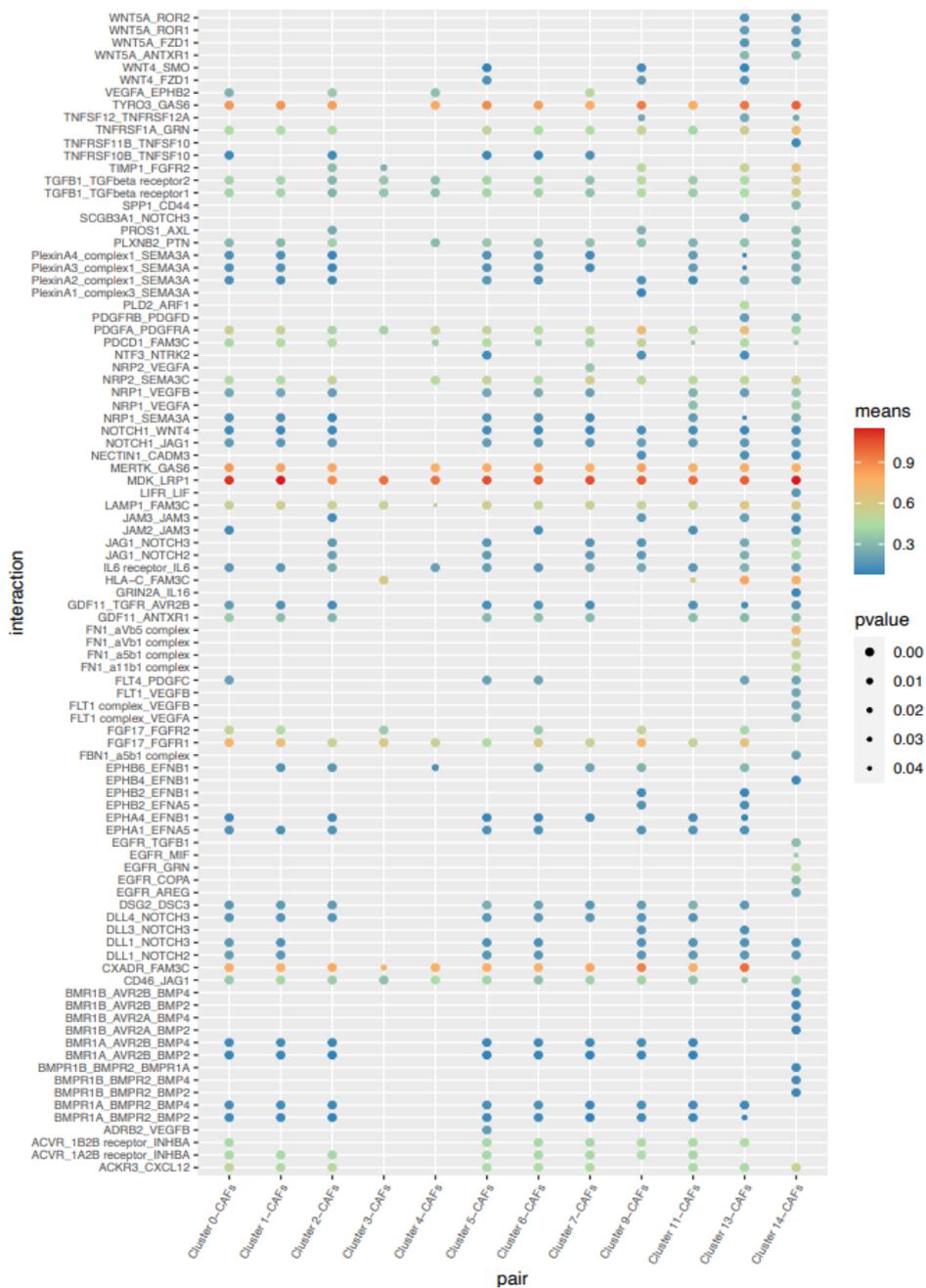
**Supplementary Fig. S12 | Lineage-specific markers' expression in JN-DSRCT-1 cells grown in 2D colonies or 3D culture.** Representative images of Desmin (DES) and WT1 C-terminal (EWSR1-WT1) immunofluorescence stainings on JN-DSRCT-1 cells grown in 2D monolayer and spheroids. **A**, JN-DSRCT-1 grown in colonies and stained for DES and WT at 20X magnification (left panel) and 60X magnification (right panel). DAPI is used as a nuclear counter-staining. DES fluorescence intensity is significantly higher in cells located at the periphery. **B**, Images of a unique stack of a JN-DSRCT-1 spheroid stained for DES and WT1. Cells located at the periphery of the spheroid display significantly increased DES fluorescence intensity.

A



B

GR2



**Supplementary Fig. S13 | DSRCT cells with mesenchymal phenotype interact preferentially with CAFs.** **A**, Heatmaps representing the number of interactions inferred from CellPhoneDB algorithm between each cluster defined in *GR2*, *GR7*, and *BOM* 3'scRNA-seq datasets. **B**, Bubble plot exhibiting the mean value for selected ligands-receptors and its associated p-value for each DSRCT malignant cell cluster-CAF cluster interaction in the *GR2* dataset. The mean value refers to the total mean of the individual partner average expression values for each interacting cluster pair.

GR2 site#4

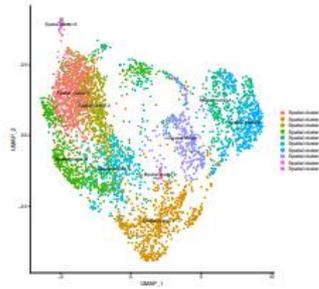
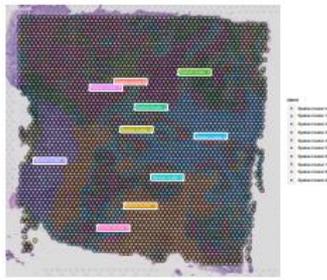
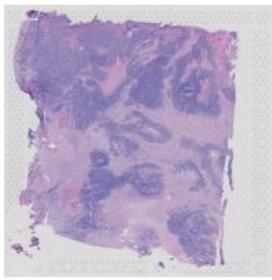
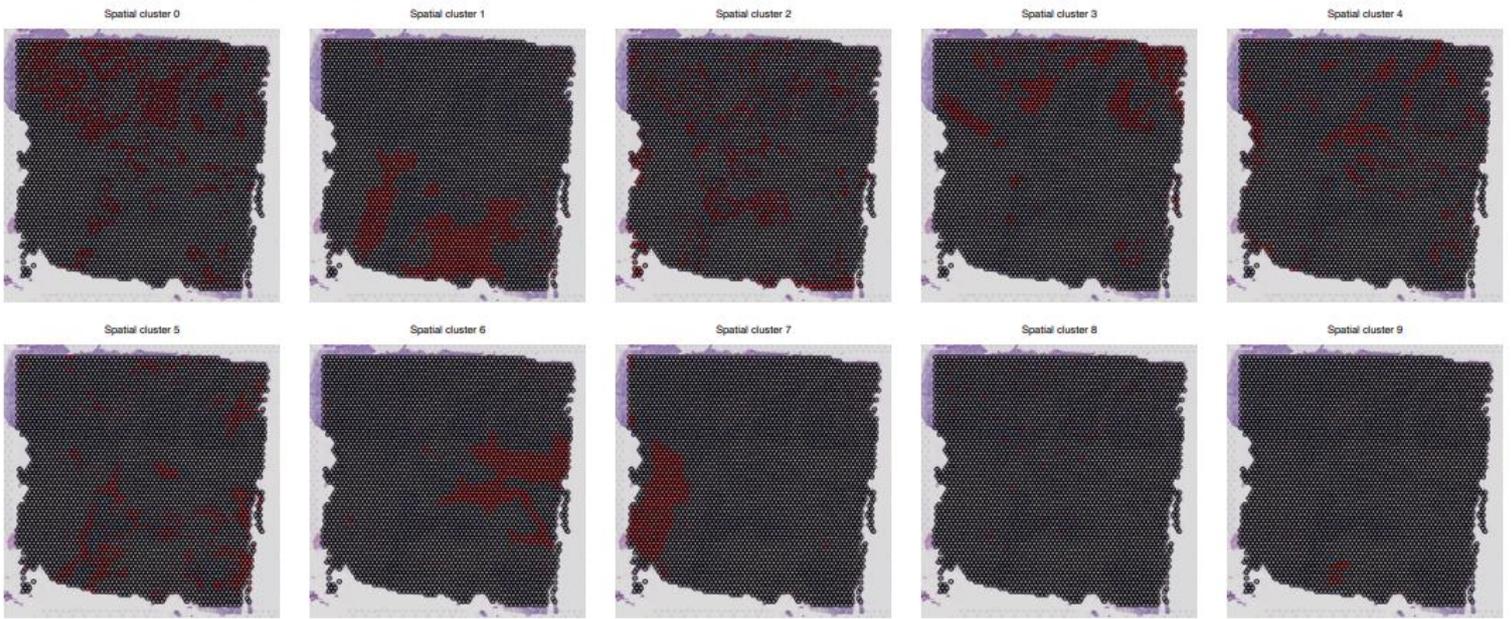
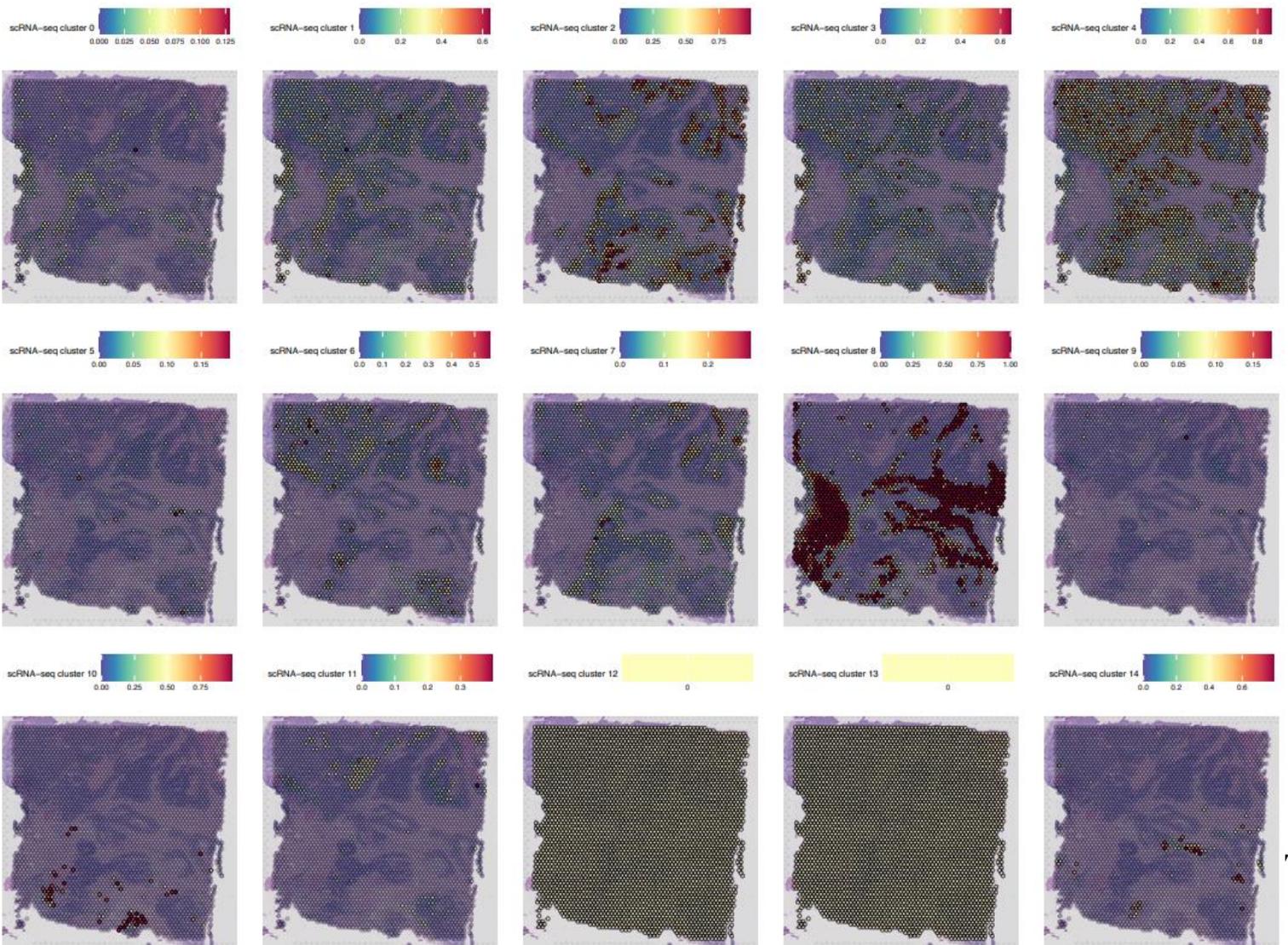


Figure S14  
Part 1/9

GR2 site#4 Visium assay clustering



Label transfer DRSCT GR2 3'scRNA-seq clusters



GR4

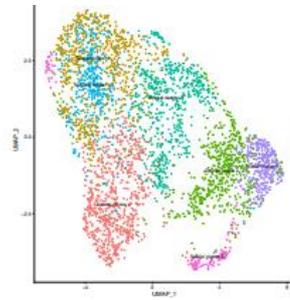
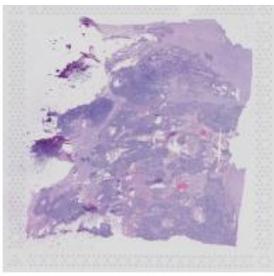


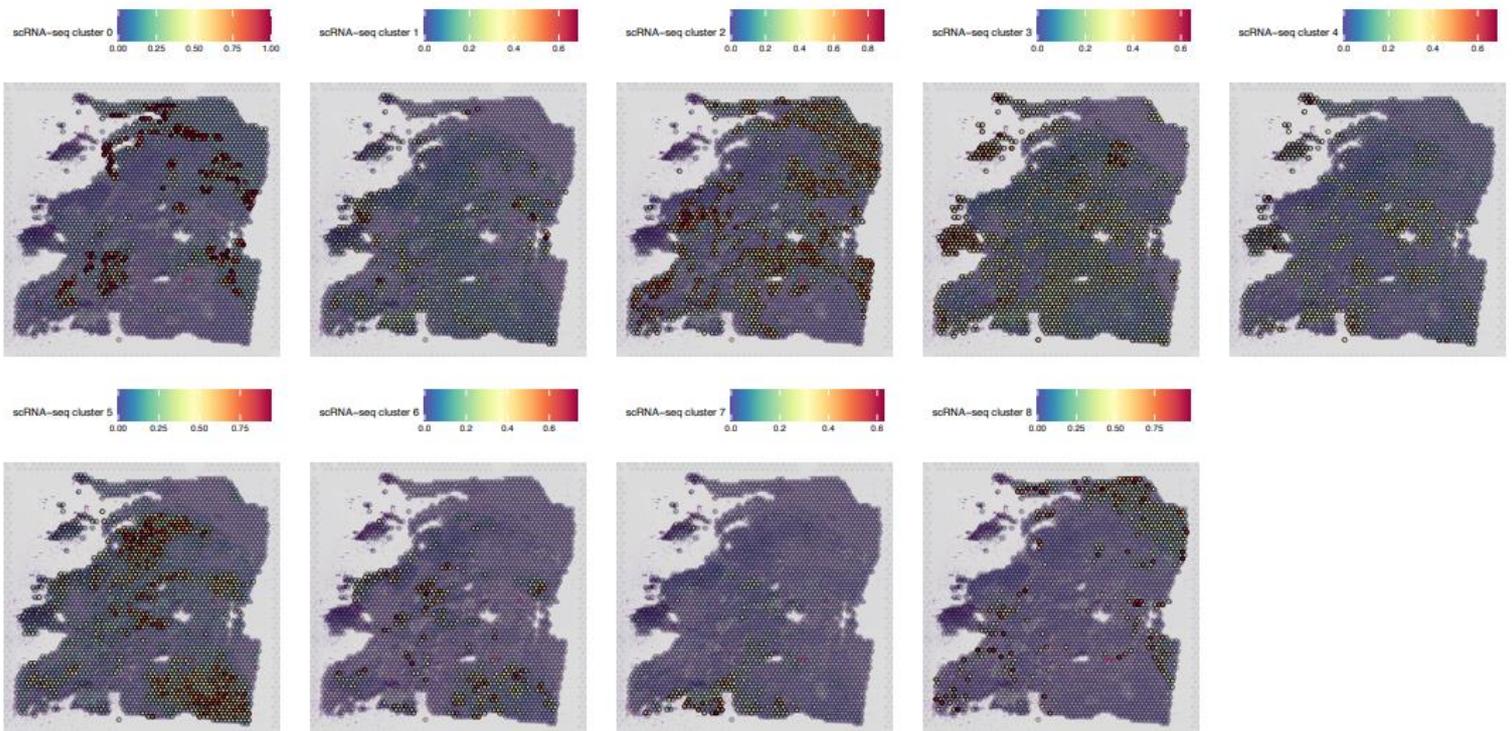
Figure S14

Part 2/9

GR4 Visium assay clustering



Label transfer DRSC T GR4 3'scRNA-seq clusters



GR4\_PC

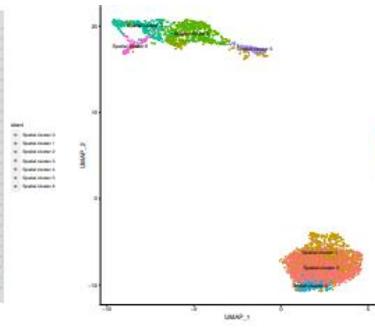
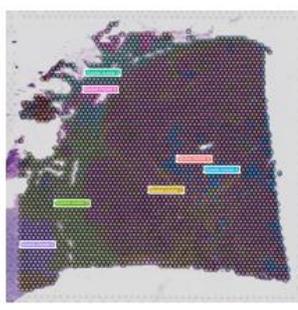
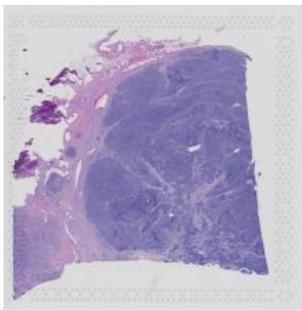
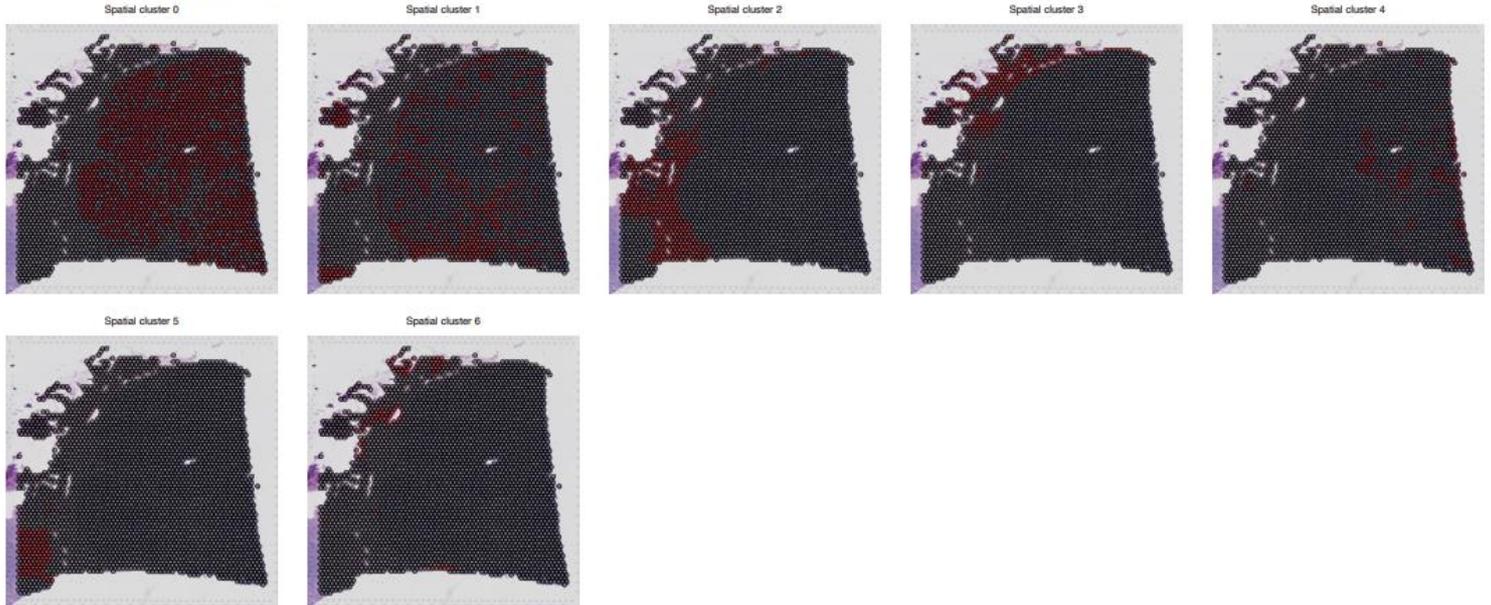


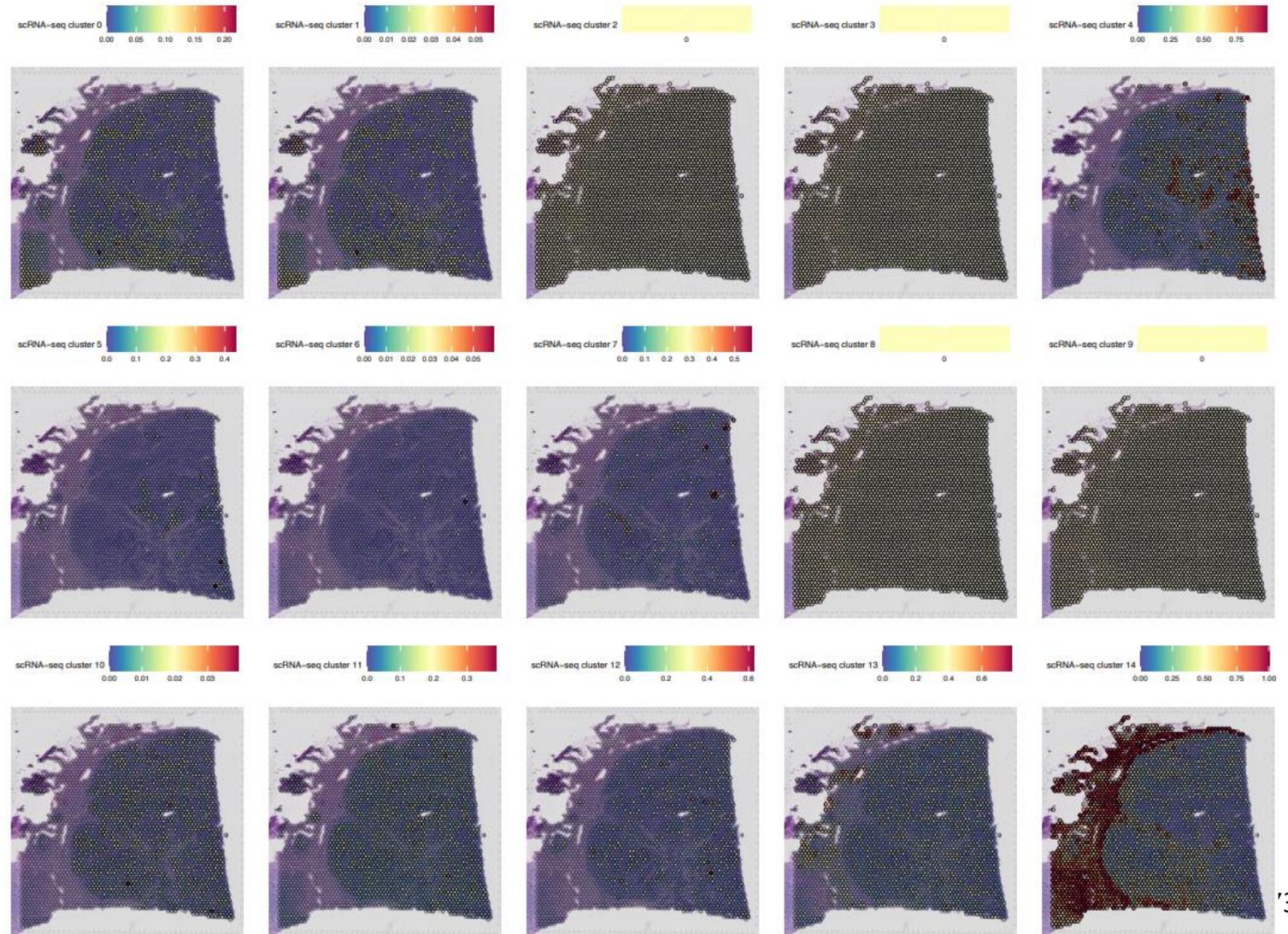
Figure S14

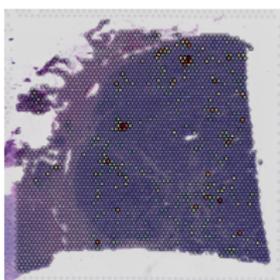
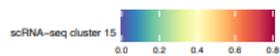
Part 3/9

GR4\_PC Visium assay clustering

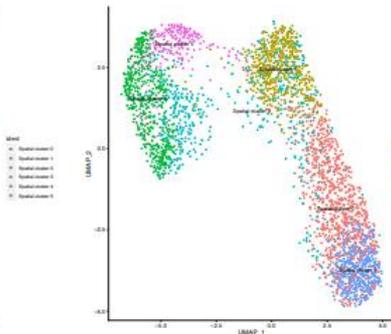
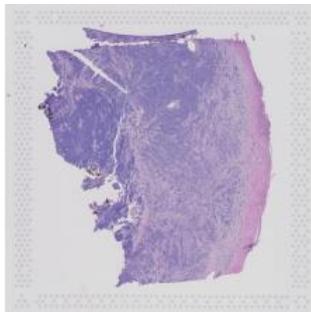


Label transfer DRSC GR4\_PC 3'scRNA-seq clusters

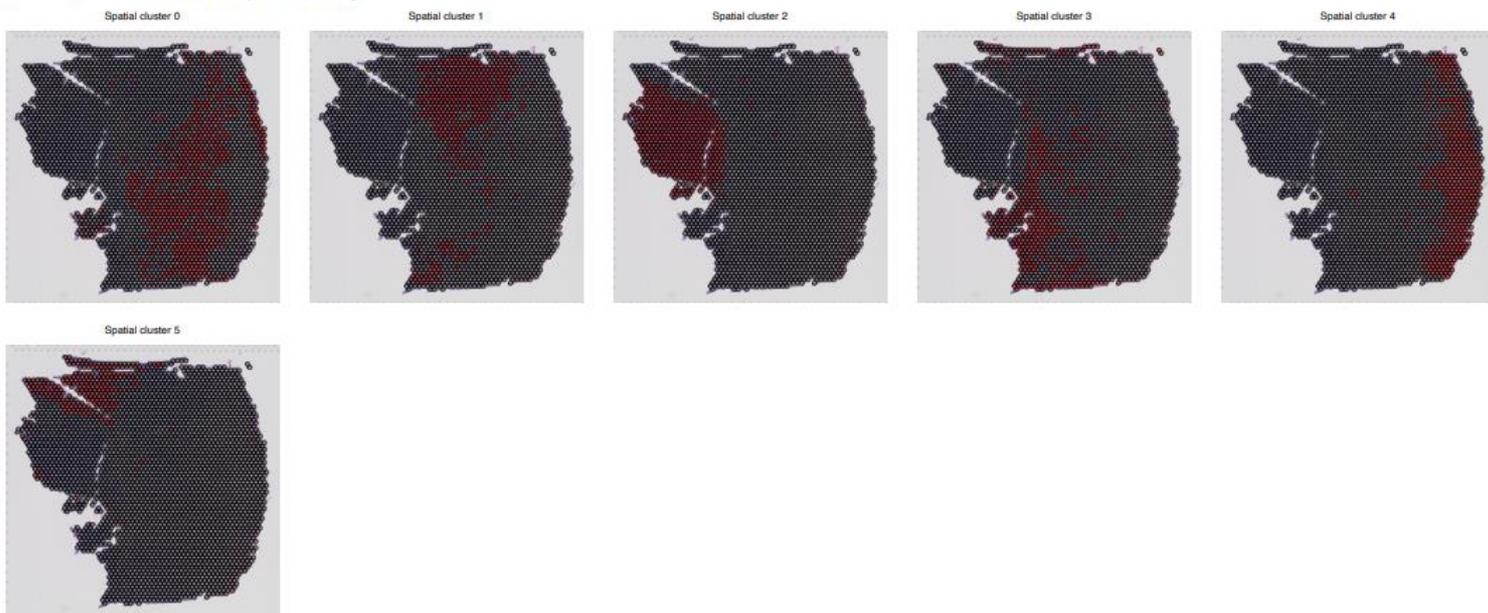




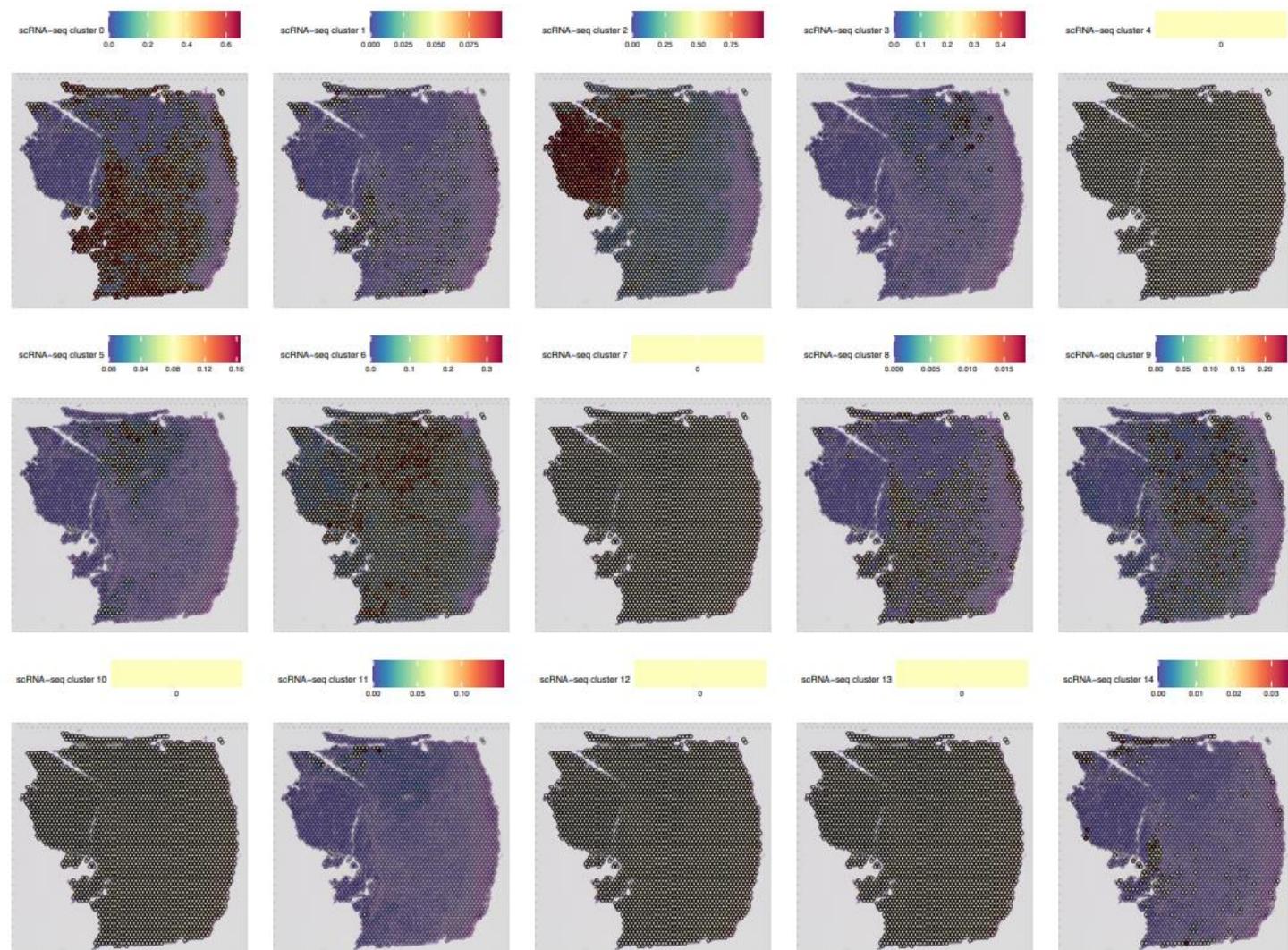
BOM site#2

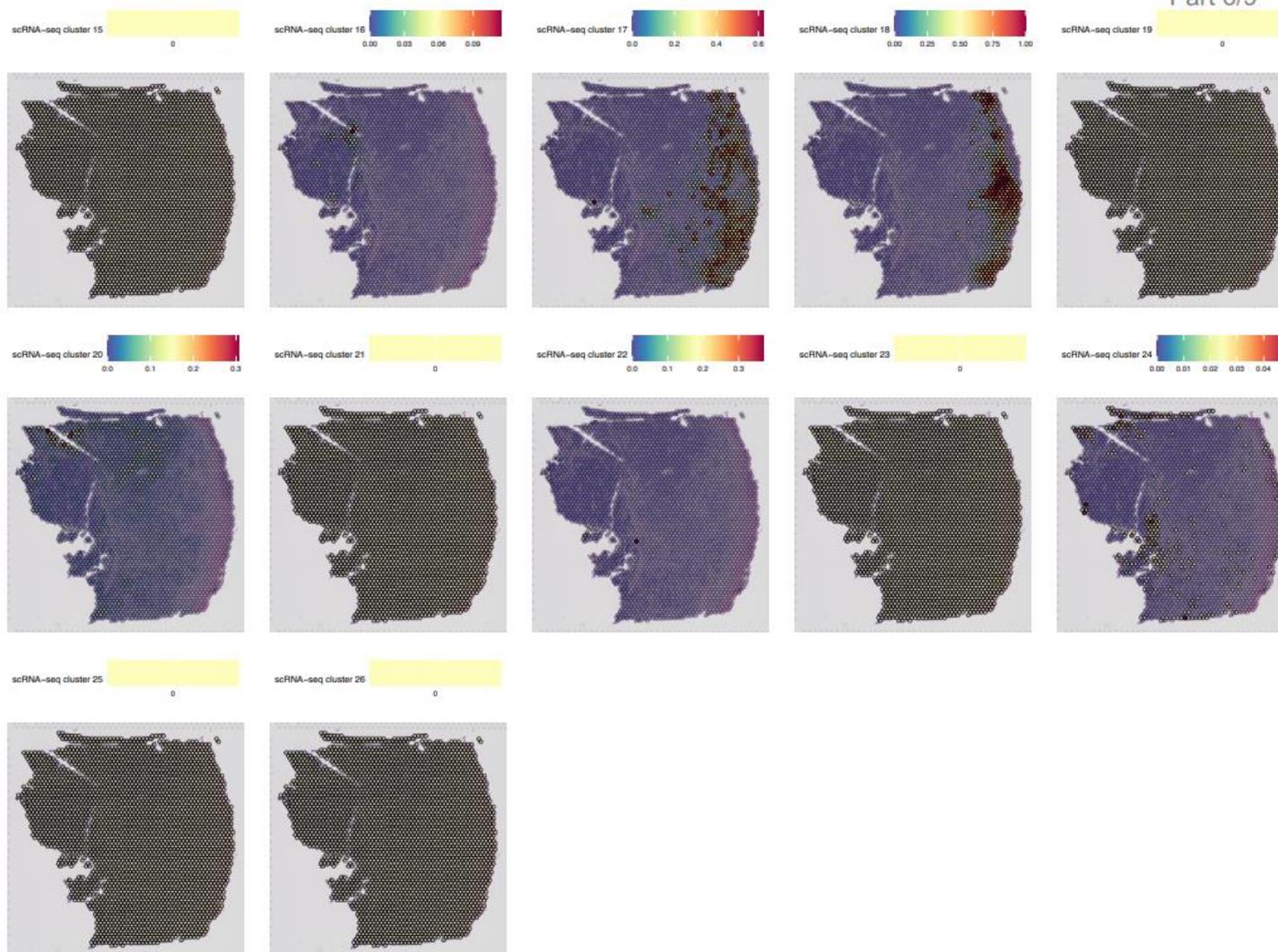


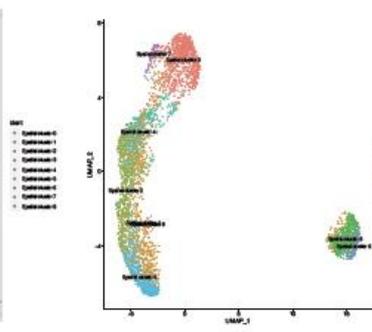
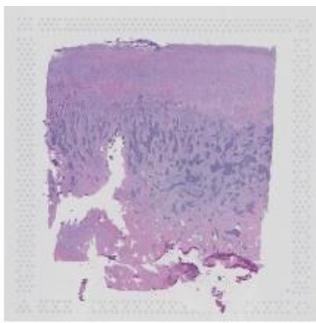
BOM site#2 Visium assay clustering



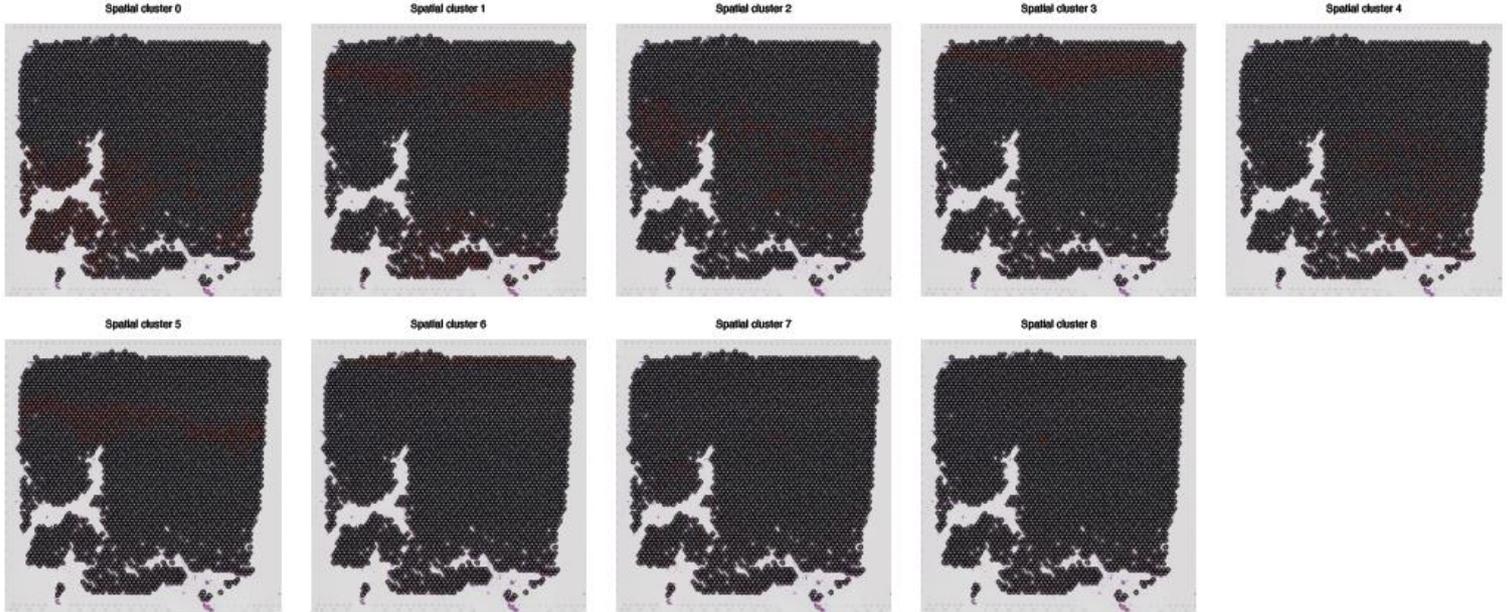
Label transfer DRSCCT BOM 3'scRNA-seq clusters



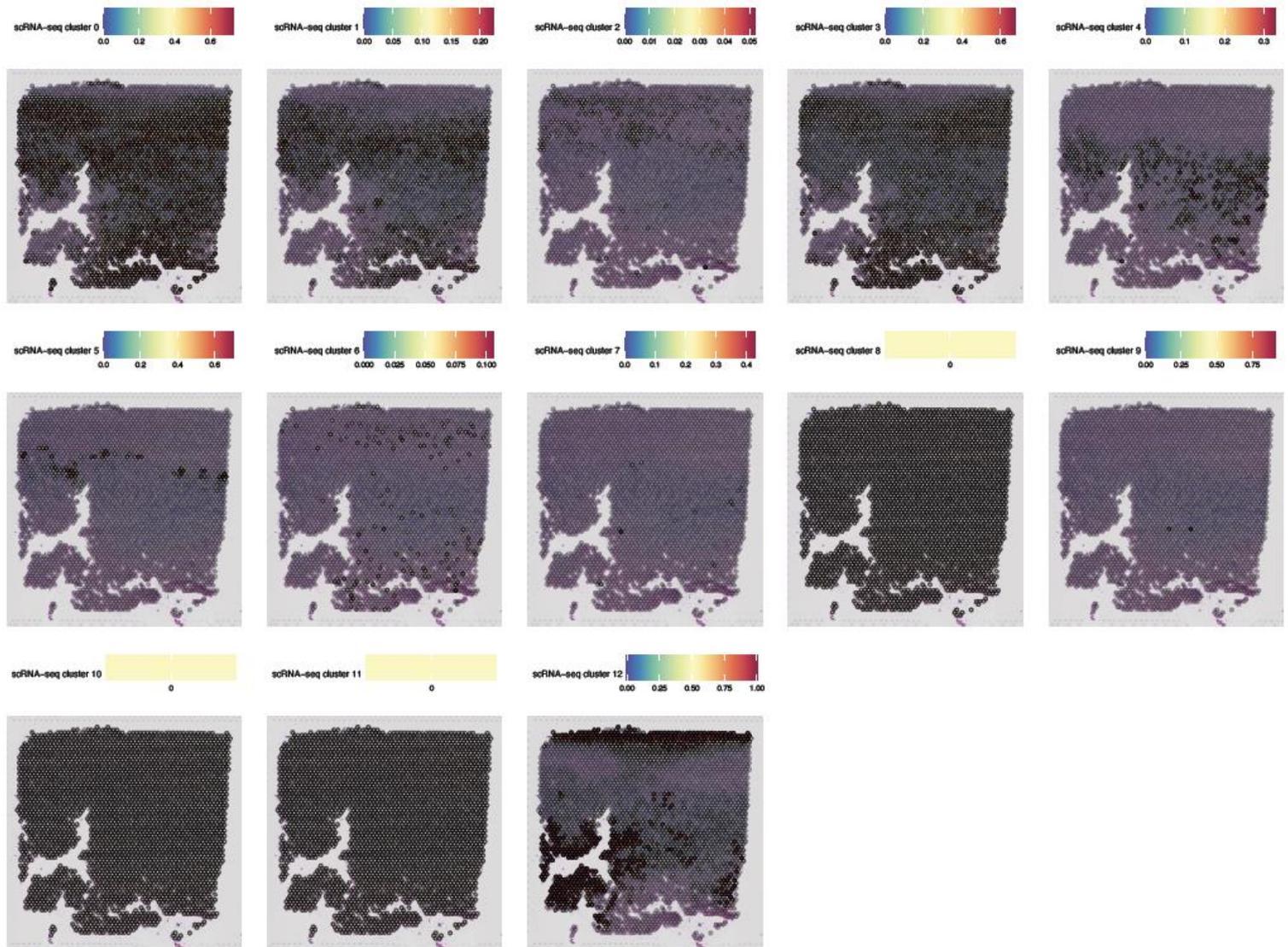




GR7\_site#2 Visium assay clustering



Label transfer DRSCCT GR7 3'scRNA-seq clusters



PZB



Legend  
Spatial cluster 0  
Spatial cluster 1  
Spatial cluster 2  
Spatial cluster 3  
Spatial cluster 4  
Spatial cluster 5  
Spatial cluster 6  
Spatial cluster 7  
Spatial cluster 8  
Spatial cluster 9  
Spatial cluster 10  
Spatial cluster 11

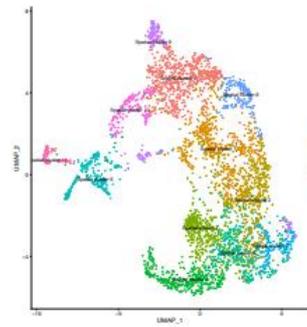


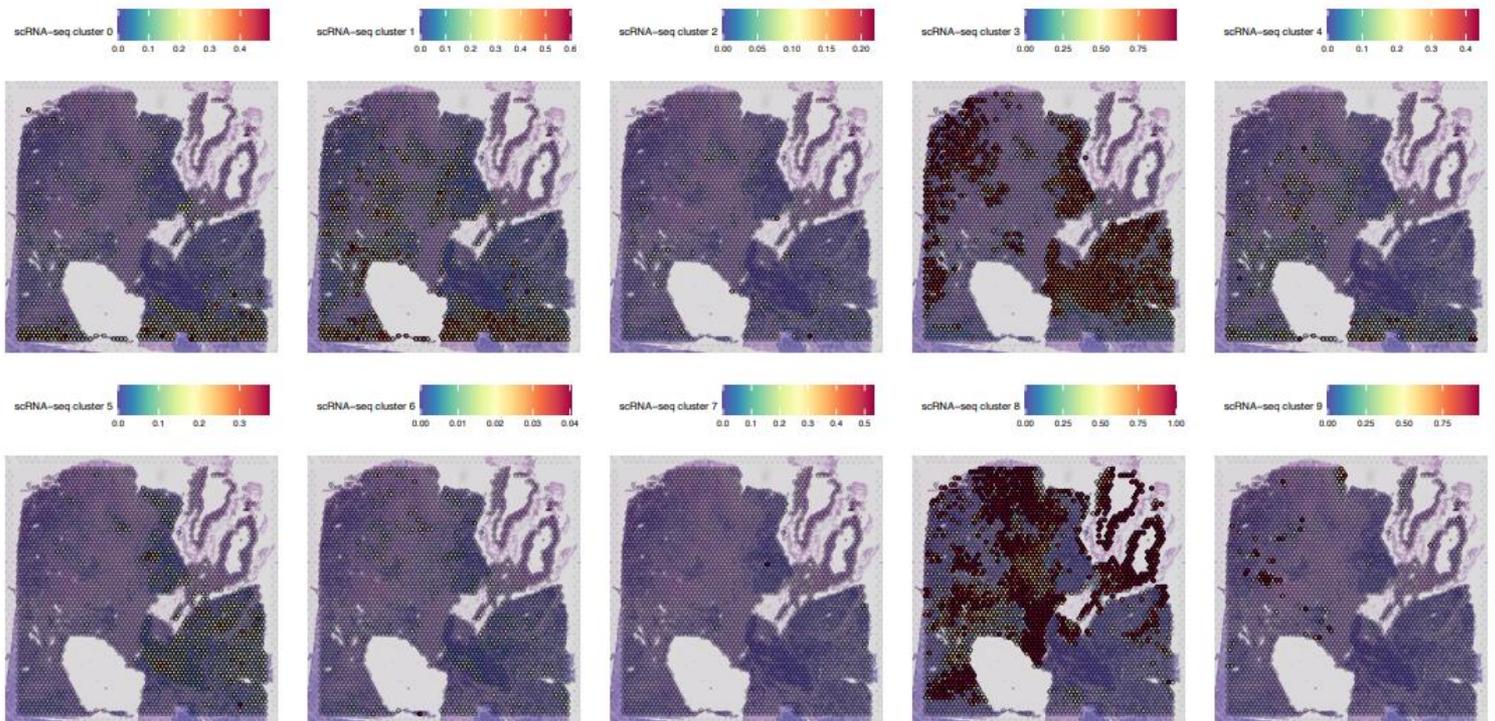
Figure S14

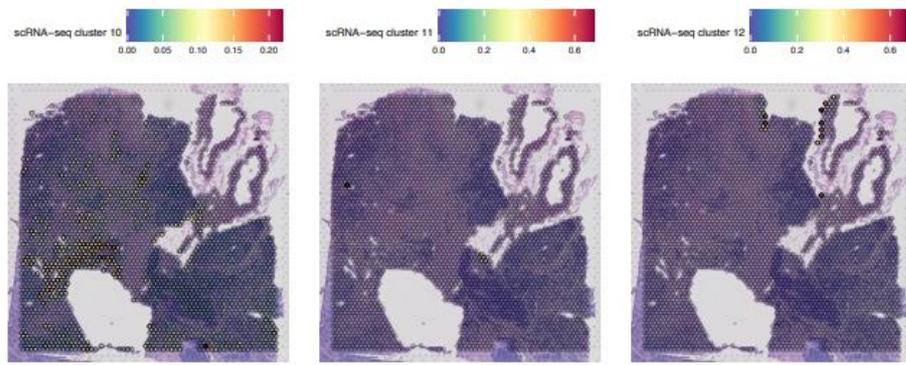
Part 8/9

PZB Visium assay clustering

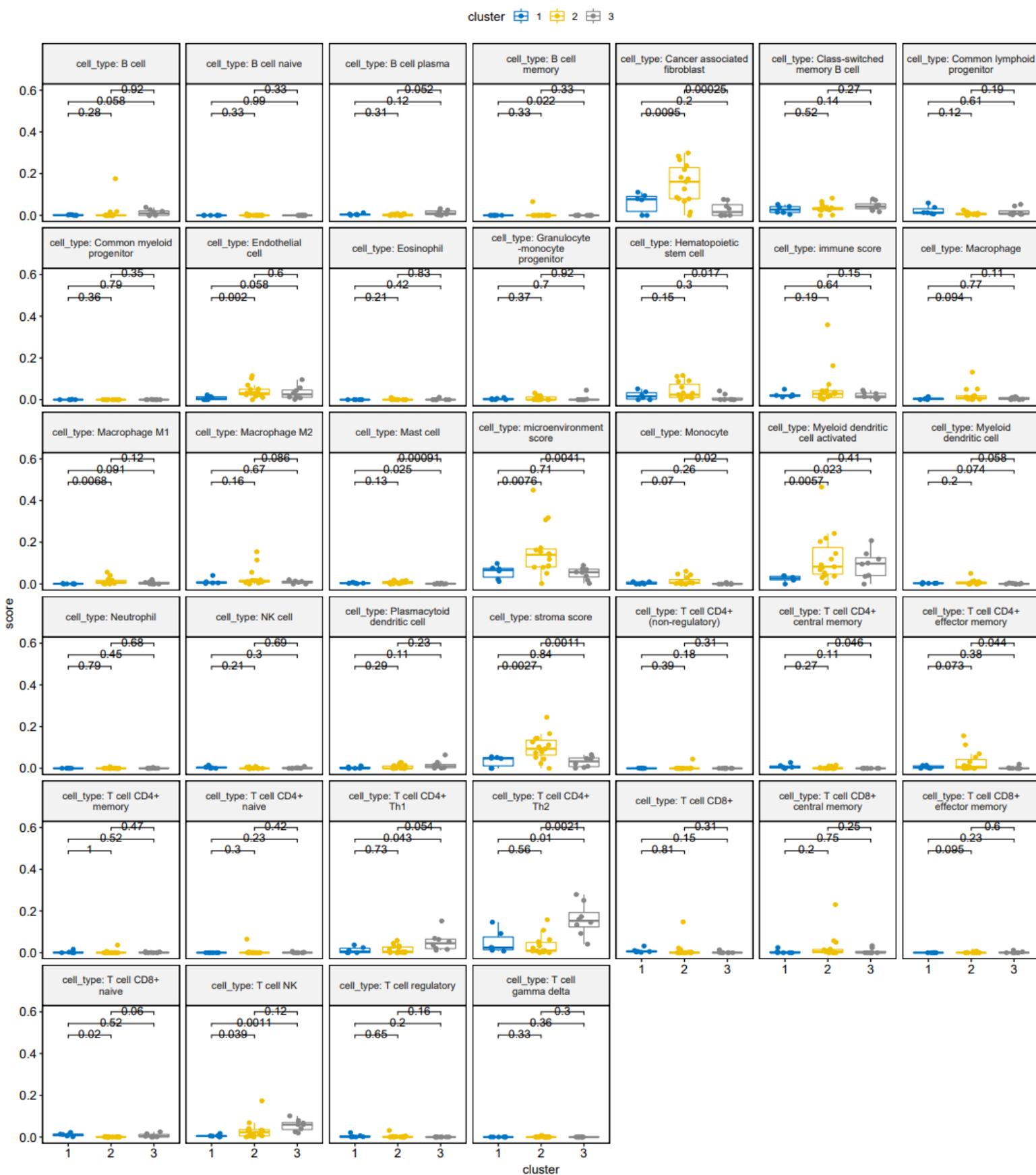


Label transfer PZB 3'scRNA-seq clusters



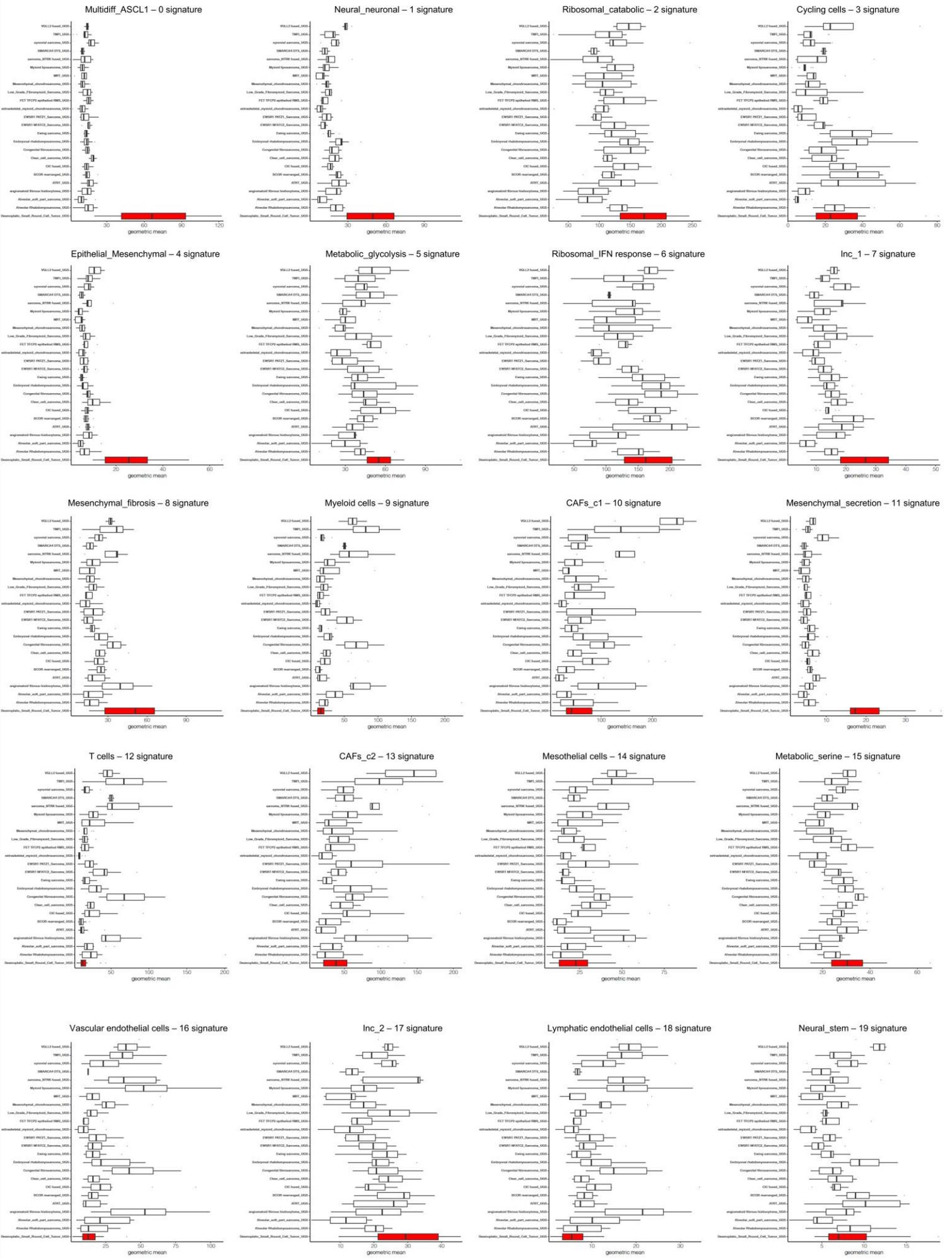


**Supplementary Fig. S14 | Spatial transcriptomics recapitulate DSRCT heterogeneity.** For each patient sample processed with Visium spatial transcriptomics assay (*GR2 site#4*, *GR4*, *GR4\_PC*, *BOM site#2*, *GR7 site#2*, and *PZB*), the upper right panel represents the HES staining of the analyzed sample. The upper central panel depicts the spatial distribution of the defined cluster overlaid on the HES slide beside the UMAP representing these clusters (upper right panel). The middle panel highlights each spatial cluster separately on the slide. The lower panel exhibits per-spot label transfer prediction scores for each previously defined 3'scRNA-seq cluster (**Supplementary Fig. S2**).

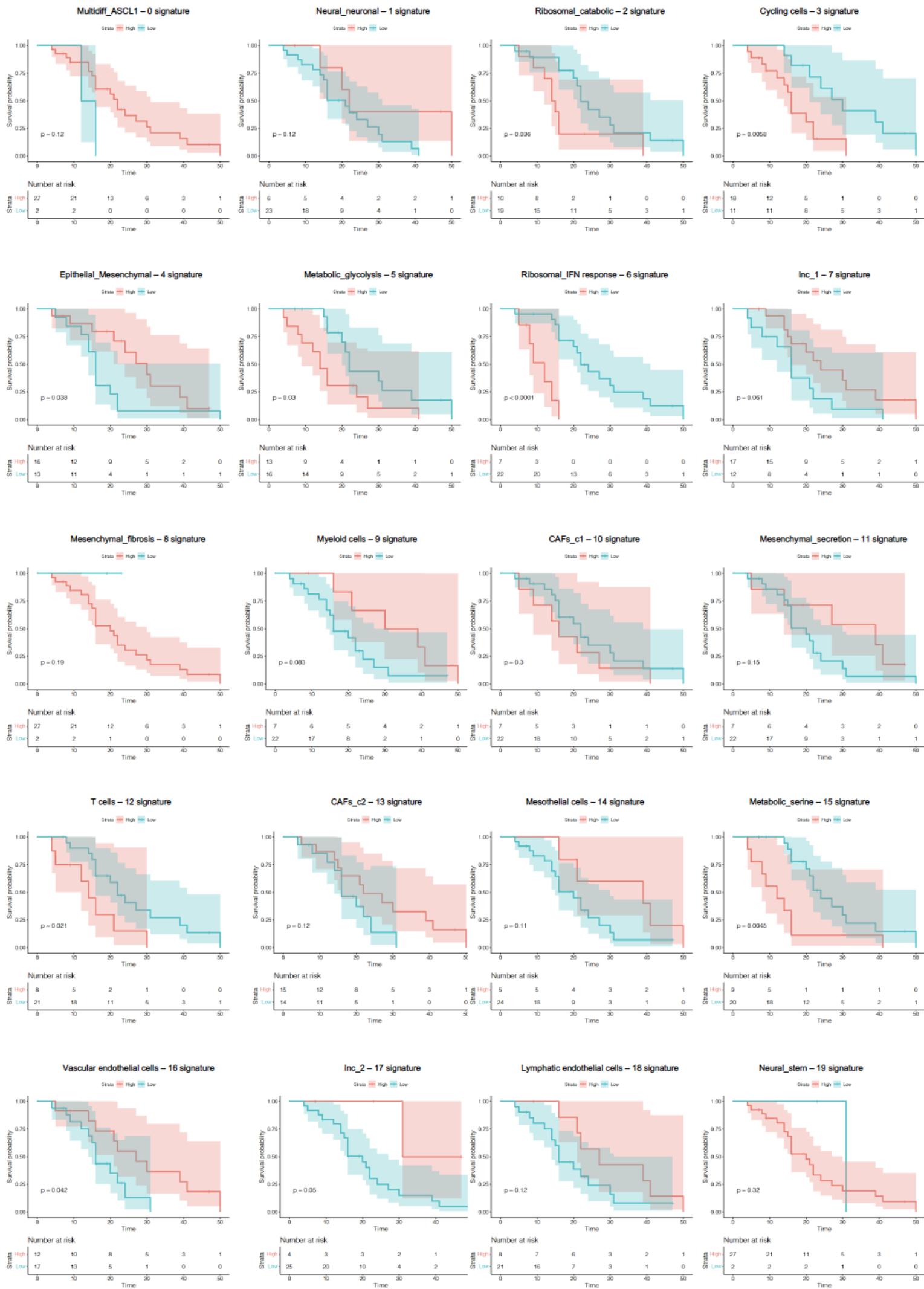


**Supplementary Fig. S15 | Immune xCell scores for each bulk RNA-seq subgroup.**

Clusters #1, #2, and #3 have been defined according to a hierarchical clustering performed on a cohort of 29 DSRCT bulk RNA-seq samples. The boxplots show the deconvoluted immune scores for each subgroup.



**Supplementary Fig S16 | Single-cell RNA-seq inferred signatures' scores across various soft tissue sarcoma subtypes.** Each signature was derived from the top 100 marker genes of the Harmony-integrated 3'scRNA-seq ("*Int\_sc*") dataset. The computation of a score for each signature among different subtypes of soft tissue sarcoma highlights their specificity (or non-specificity) to DSRCT.



**Supplementary Fig. S17 | DSRCT subpopulations' signatures correlate with patient outcomes.** Each signature was derived from the top 100 marker genes of the Harmony integrated 3'scRNA-seq (*Int\_sc*) dataset. After the computation of a score for each DSRCT subpopulation signature on each sample from the DSRCT bulk RNA-seq cohort, the cohort of patients was split in two according to a "High" or "Low" signature score. The Kaplan-Mayer curves represent patient overall survival according to the "High" or "Low" subgroup for each signature.

## Supplemental Information – Supplementary Tables and legends

*For confidentiality reasons, Supplementary Tables are not included in the online version of this manuscript. Still, the tables' legends are detailed below for better understanding.*

**Supplementary Table S1 | Main clinical characteristics of DSRCT patients and samples included in the single-cell experiments cohort.** AI: Adriamycin-Ifosfamide; IE: Ifosfamide-Etoposide; VAC: Vincristine-Actinomycin D-Cyclophosphamide, WES: Whole Exome Sequencing.

**Supplementary Table S2 | DSRCT main cell types' identification according to canonical markers and differential gene expression analysis.** **A**, Main DSRCT cell populations canonical markers. **B**, Top 50 differentially expressed genes across DSRCT main cell populations.

**Supplementary Table S3 | Characterization of cell clusters from the Harmony-integrated 3'single-cell RNA-sequencing dataset (“*Int\_sc*”).** Clusters were defined using the graph-based k-nearest neighbors algorithm. **A**, “*Int\_sc*” cell clusters' top 50 differentially expressed genes. **B**, “*Int\_sc*” clusters quality control metrics.

**Supplementary Table S4 | Top 50 differentially expressed genes of identified clusters for individual or merged 3'single-cell RNA-sequencing datasets.** Merged datasets contain the union of assays performed parallelly on distinct tumor sites from the same patient.

**Supplementary Table S5 | Gene Ontology (GO) terms enrichment analysis for differentially expressed genes of each Harmony-integrated 3'single-cell RNA-sequencing dataset (“*Int\_sc*”) cluster.** Analyzed GO terms comprise GO:MF (Gene Ontology: Molecular Function), GO:CC (Gene Ontology: Cellular Component), and

GO:BP (Gene Ontology: Biological Process). The top 10 GO terms ranked on adjusted p-value for each “*Int\_sc*” cluster are displayed. GO terms enrichment analysis was performed using the G:profiler tool.

**Supplementary Table S6 | Hotspot-inferred co-expressed gene modules from the Harmony integrated 3’single-cell RNA-sequencing dataset (“*Int\_sc*”).**

**Supplementary Table S7 | Gene Ontology (GO) terms enrichment analysis on Hotspot-inferred co-expressed gene modules from Harmony integrated 3’single-cell RNA-sequencing dataset (*Int\_sc*).** GO terms enrichment analysis was performed using ToppFUN portal.

**Supplementary Table S8 | Transcripts per million (tpm) table of JN-DSRCT-1 cell line RNA-sequencing.**

**Supplementary Table S9 | Differential binding analysis between EWSR1-WT1 Chromatin Immunoprecipitation with sequencing (ChIP-seq) versus isotype ChIP-seq on JN-DSRCT-1 cell line.** The differential binding analysis was performed using DESeq2 tool. The experiment was performed in duplicates.

**Supplementary Table S10 | List of significantly enriched peaks in EWSR1-WT1 Chromatin Immunoprecipitation with sequencing (ChIP-seq) on JN-DSRCT-1 cell line.** This list contains all peaks with significant enrichment (adjusted p-value <0.05) in EWSR1-WT1 ChIP-seq compared to isotype ChIP-seq on the JN-DSRCT-1 cell line. The “Gene.Name” column corresponds to peak annotation according to the nearest Transcription Start Site (TSS) using HOMER tool.

**Supplementary Table S11 | Gene Ontology (GO) terms enrichment analysis on significantly enriched peaks in EWSR1-WT1 Chromatin Immunoprecipitation with sequencing (ChIP-seq) on JN-DSRCT-1 cell line.** GO terms enrichment analysis was performed based on the list of EWSR1-WT1 ChIP-seq-significantly enriched peaks detailed in **Supplementary Table S10**, using ChIPEnrich tool. Only GO terms from the Biological Process (GOBP) category are displayed.

**Supplementary Table S12 | Motifs enrichment analysis from significantly enriched peaks in EWSR1-WT1 Chromatin Immunoprecipitation with sequencing (ChIP-seq) on JN-DSRCT-1 cell line.** Motif enrichment analysis was performed using the MEME suite and JASPAR2020 transcription factors (TFs) database on the list of significantly enriched peaks in EWSR1-WT1 ChIP-seq (**Supplementary Table S11**).

**Supplementary Table S13 | *De novo* motifs enrichment discovery from significantly enriched peaks in EWSR1-WT1 Chromatin Immunoprecipitation with sequencing (ChIP-seq) on JN-DSRCT-1 cell line.** **A**, Identified *de novo* motifs within EWSR1-WT1 ChIP-seq-inferred binding sequences. The identification of *de novo* motifs was performed using the “streme” function from the MEME suite on the list of significantly enriched peaks in EWSR1-WT1 ChIP-seq (**Supplementary Table S11**). **B**, Comparison of *de novo* motifs sequences with known transcription factors’ motifs. This motif-based sequence analysis was performed using the JASPAR 2020 database as a reference with the Tomtom motif comparison tool.

**Supplementary Table S14 | Differential peak accessibility analysis from Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) experiment on JN-DSRCT-1 cell line with or without EWSR1-WT1 silencing.** The differential peak accessibility analysis was performed using DESeq2 tool. The experiment was performed in duplicates.

**Supplementary Table S15 | Differential gene expression analysis between EWSR1-WT1-silenced or non-silenced JN-DSRCT-1 cell line.** The experiment was performed in duplicates.

**Supplementary Table S16 | Overlap between peaks with increased accessibility in non-silenced versus EWSR1-WT1-silenced JN-DSRCT-1 cells and EWSR1-WT1 binding sites.** Peaks with increased accessibility in non-silenced versus EWSR1-WT1-silenced JN-DSRCT-1 cells (EWSR1-WT1 ON module) are inferred from differential peak accessibility analysis from Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) experiment (**Supplementary Table S14**). EWSR1-WT1 binding sites are defined by significantly enriched peaks in EWSR1-WT1

Chromatin Immunoprecipitation with sequencing (ChIP-seq) versus control isotype ChIP-seq on JN-DSRCT-1 cell line. Overlapping peaks were identified using the “findOverlapsOfPeaks” function from the ChIPpeakAnno package using a maximum authorized gap of 5 kb and are displayed in **Fig. 3E. A**, Overlapping peaks’ annotated genes between EWSR1-WT ON enriched ATAC peaks and EWSR1-WT1 ChIP-seq enriched peaks. Genes were annotated to identified overlapping peaks based on the nearest Transcription Start Site (TSS) using HOMER tool. In pink are highlighted the genes which are in common between the following lists: (i) genes corresponding to significantly more accessible peaks in non-silenced versus EWSR1-WT1-silenced JN-DSRCT-1 cells in ATAC-seq experiment (adjusted p-value < 1.10E-100 and log2FC >1.5), (ii) annotated genes from EWSR1-WT1 ChIP-seq-inferred significantly enriched peaks, and (iii) significantly overexpressed genes in non-silenced versus EWSR1-WT1-silenced JN-DSRCT-1 cells based on bulk RNA-sequencing experiment. **B**. Gene Ontology (GO) terms enrichment analysis on genes corresponding to overlapping peaks between EWSR1-WT1 ON enriched ATAC peaks and EWSR1-WT1 ChIP-seq enriched peaks. GO terms enrichment analysis was performed using ToppFUN portal.

**Supplementary Table S17 | Motifs enrichment analysis from significantly more accessible peaks in non-silenced (EWSR1-WT1 ON) versus EWSR1-WT1-silenced (EWSR1-WT1 OFF) JN-DSRCT-1 cell in Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) experiment.** Motif enrichment analysis was performed using the MEME suite and JASPAR2020 transcription factors (TFs) database.

**Supplementary Table S18 | EWSR1-WT1 transcriptional activity signature inferred from EWSR1-WT1 Chromatin Immunoprecipitation with sequencing (ChIP-seq) experiment.** EWSR1-WT1 ChIP signature was defined using the top differentially enriched peaks (log2FC<-5) compared to isotype ChIP.

**Supplementary Table S19 | Differentially expressed genes for each cluster of the Single-nucleus Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and RNA-seq (snMultiome) assay.** Clusters were defined using the weighted nearest neighbors’ (WNN) method, which integrates gene expression and chromatin accessibility modalities.

**Supplementary Table S20 | Differentially accessible peaks for each cluster of the Single-nucleus Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and RNA-seq (snMultiome) assay using pseudo-bulk peak calling.** Clusters were defined using the weighted nearest neighbors' method, which integrates gene expression and chromatin accessibility modalities. Pseudo-bulk peak calling refers to macs2 peak calling on the pool of single cells.

**Supplementary Table S21 | Differentially accessible peaks for each cluster of the Single-nucleus Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and RNA-seq (snMultiome) assay using per-cluster peak calling.** Clusters were defined using the weighted nearest neighbors' method, which integrates gene expression and chromatin accessibility modalities. Per-cluster peak calling refers to a separated macs2 peak calling on each cell cluster.

**Supplementary Table S22 | Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)-inferred EWSR1-WT1 ON and EWSR1-WT1 OFF modules.** Genome regions for both EWSR1-WT1 *ON* and EWSR1-WT1 *OFF* ATAC modules are listed. These modules were defined according to differentially accessible peaks between EWSR1-WT1-silenced versus non-silenced JN-DSRCT-1 cells. A per-cell score for both modules was then calculated on the single-nucleus Multiome assay (**Supplemental Information – Fig. S9B**).

**Supplementary Table S23 | Motif enrichment analysis from differentially accessible peaks between weighted-nearest neighbors' defined clusters within the single-nucleus Multiome assay.** Motif enrichment analysis was performed using the "FindMarkers" function from Seurat package and the JASPAR2020 database as a reference. The top 10 enriched motifs for each cluster are displayed.

**Supplementary Table S24 | Top 50 differentially expressed genes for DSRCT immune cell subclusters from the Harmony integrated 3'single-cell RNA-sequencing dataset ("*Int\_sc*").** **A.** Top 50 differentially expressed genes between myeloid cell clusters inferred from MoMac-VERSE atlas label transfer. **B.** Top 50

differentially expressed genes between lymphoid cell clusters defined by graph-based k-nearest neighbors algorithm.

**Supplementary Table S25 | Top 50 differentially expressed genes for DSRCT fibroblast subclusters from the Harmony integrated 3'single-cell RNA-sequencing dataset (“*Int\_sc*”).**

**Supplementary Table S26 | Top 50 differentially expressed genes between clusters from each Visium spatial transcriptomics assay.** Differentially expressed genes between clusters are displayed for Visium *GR2 site#4*, *BOM site#2*, *GR4*, *GR4\_PC*, *GR7 site#2*, and *PZB* datasets.

**Supplementary Table S27 | Main clinical characteristics of the DSRCT bulk RNA-seq test cohort.**

**Supplementary Table S28 | Differential gene expression (DGE) analysis between DSRCT bulk RNA-seq subgroups.** Subgroups are defined according to hierarchical clustering as defined by Fig. 6B. Listed genes are differentially enriched for each specified subgroup comparison (absolute  $\log_2FC > 2$ , adjusted p-value  $< 0.05$ ).

**Supplementary Table S29 | Gene Ontology (GO) terms enrichment analysis on differentially enriched genes in bulk RNA-seq subgroups.** GO terms enrichment analysis was performed using the ToppFUN tool on differentially enriched genes between subgroups (absolute  $\log_2FC > 2$ , adjusted p-value  $< 0.05$ ). Only the top 10 GO terms for Biological Process (GOBP), Molecular Function (GOMF), and Cellular Component (GOCC) are shown.

**Supplementary Table S30 | Independent bulk RNA-sequencing cohort comprising various sarcoma subtypes.** The number of samples per sarcoma histological subtype is shown. The cohort of DSRCTs was used as an external validation cohort to evaluate the applicability of generated prognostic signatures (**Fig. 6F**). AI: Doxorubicin-Ifosfamide; API: Doxorubicin-CDDP-Ifosfamide; HIPEC: Hyperthermic Intra Peritoneal Chemotherapy; IVA: Ifosfamide-Vincristine-Doxorubicin; IVADo: Ifosfamide-Vincristine-Actinomycin-Doxorubicin; L1: First line; L2: Second

line; TEMIRI: Temozolomide-Irinotecan; VAC: Vincristine-Doxorubicin-Cyclophosphamide; VIDE: Vincristine-Ifosfamide-Doxorubicin-Etoposide; VIP: Vincristine-Ifosfamide-CDDP.

## CHAPTER III. DISCUSSION AND PERSPECTIVES

### 1 RESULTS DISCUSSION AND OPEN QUESTIONS

---

#### 1.1 DYNAMIC CELLULAR TRANSITION STATES VERSUS CLONAL HETEROGENEITY

The herein presented data question the mechanisms underlying DSRCT tumor cell heterogeneity.

The term “plasticity” has been used throughout this manuscript, supported by (i) the paucity of CNV variability across tumor cell clusters - suggesting no or few genetic clonal evolution - and (ii) the association between microenvironment signals (e.g., CAFs proximity, hypoxic areas) and DSRCT tumor cell states. However, this terminology may be questionable, as no consensual definition of “plasticity” is universally accepted.

First, the hypothesis that DSRCT cells do not follow clonal or subclonal evolution due to the acquisition of selectively advantageous secondary mutations is worth confirming by additional methods. In this sense, single-cell DNA sequencing would represent a method of choice to validate this hypothesis orthogonally. Nonetheless, available single-cell whole genome or exome assays rely on in-house implementable methods (e.g., SMOOTH-seq<sup>153</sup>, DOP-PCR<sup>149</sup>, multiple displacement amplification (MDA)<sup>150</sup>, multiple annealing and looping-based amplification cycles (MALBAC)<sup>151</sup>, Linear Amplification via Transposon Insertion (LIANTI)<sup>152</sup>) or commercially available library amplification kits<sup>215</sup>. These methods are currently limited in throughput since they depend on manual or FACS single-cell isolation. Hence, the number of analyzed cells is generally low and limited to several hundred. Single-cell sorting is then followed by NGS or third-generation sequencing (TGS) using sequencing technologies such as PacBio or Oxford Nanopore technology to enable structural variations detection<sup>153</sup>, often associated with consequent sequencing costs. On the other hand, targeted single-cell DNA sequencing (e.g., MissionBio Tapestry Platform) generally implements a microfluidic system enabling the analysis of thousands of cells but is limited to the study of a restricted panel of candidate genes, which may be cumbersome in the case of DSRCT study, given the lack of highly recurrent secondary mutations.

Interestingly, recently described methods of Genotyping of Transcriptomes<sup>216</sup> (GoT) may represent an advantageous approach to parallelly study gene expression and exonic mutations by integrating targeted genotyping with high-throughput droplet-based single-cell RNA sequencing. Nonetheless, such methods also rely on amplifying specific loci of interest.

Notably, we are currently evaluating the feasibility of a similar approach, taking advantage of 3'scRNA-seq 10X Genomics Chromium derived barcoded cDNA, followed by adapters ligation and library preparation, and long-read sequencing on Oxford Nanopore sequencer, skipping the step of prior targeted loci amplification. If successful, such an approach may provide further insights into CNV heterogeneity among DSRCT cells' subpopulations.

Secondly, the term plasticity implies the notion of dynamic cell phenotypic states. Although conventional single-cell analysis methods only yield static phenotypic readouts, increasing tools <sup>217</sup> are focusing on trajectory inference, aiming at allocating cells to lineages and ordering them based on "pseudotimes" (i.e., the distance between a given cell and the origin of a specific lineage in the reduced dimensional space) to unravel dynamic processes and cells' transition states, along more or less complex trajectories. Recently implemented methods also aim at identifying genes associated with lineage differentiation <sup>218</sup>. While pseudotime is usually considered a surrogate for actual chronological time, both may not be linearly correlated. In the presented manuscript, we used a directed single-cell fate mapping tool (CellRank <sup>219</sup>), which combines trajectory inference with directional information from RNA velocity (i.e., spliced/unspliced transcripts ratio). Notably, when applied to individual or merged scRNA-seq datasets, no consistent trajectories, initial, or terminal states could be identified across samples. This observation suggests DSRCT cells may follow multiple variable state transitions rather than linear developmental or hierarchical processes. The nature of the signals triggering such processes is still uncertain and may combine stochasticity and microenvironment cues.

Third, the reversible potential of such adaptive plasticity remains to be assessed. Although EMT is generally considered a reversible process <sup>220</sup>, the reversibility of the previously described DSRCT tumor cell states remains unknown. While only functional *in vitro* validation would enable us to answer this question, one should remember that primary cells grown in culture – and, to a more considerable extent, cell lines - may display distinct plasticity landscape topologies compared to *in vivo* cells. A credible approach would be to first concentrate on the neural-mesenchymal trajectory axis and perform a longitudinal evaluation of neural/mesenchymal markers across cells grown in 2D and 3D conditions, more or less upon state-specific induction using appropriate growth factors (i.e., CAFs conditioned medium).

## **1.2 ROLE OF EWSR1-WT1 EXPRESSION LEVEL ON DSRCT CELL HETEROGENEITY**

Part of this study explored whether EWSR1-WT1 transcript expression level would play a significant role in DSRCT cell heterogeneity.

This hypothesis partly relies on analogical reasoning based on Ewing sarcoma biology, in which lower levels of EWSR1-FLI1 were associated with a mesenchymal phenotype with increased invasive properties <sup>200</sup>.

To meet this objective, we took advantage of an in-house developed experiment to amplify EWSR1-WT1 cDNA originating from the barcoded cDNA library of a 10X Genomics 3'scRNA-seq experiment. Indeed, when focusing on WT1 expression at single cell level on the 3'scRNA-seq Harmony integrated data, no apparent WT1 transcript expression heterogeneity could be identified across clusters. Notably, in an original 3'scRNA-seq technology pipeline, wild-type WT1 transcripts cannot be distinguished from EWSR1-WT1 transcripts because only the 3' part of transcripts is kept after the enzymatic fragmentation. Thus, two significant hurdles inherent to 3'scRNA-seq technology may hinder our ability to detect EWSR1-WT1 expression variability. First, although it is widely accepted that wild-type WT1 is repressed in DSRCT tumor cells <sup>66,221</sup>, possibly due to EWSR1-WT1 dominant negative effect, several cases report the persistence of wild-type WT1 expression in DSRCT cells <sup>44</sup>, and it is currently unknown whether wild-type WT1 may still be expressed at low levels. Secondly, although EWSR1-WT1 is the unique oncogenic driver in DSRCT tumors, EWSR1-WT1 overall expression remains low compared to other highly expressed genes, such as *MALAT1*, *CCL25*, or *GJB2*. Thus, one may presume that insufficient sequencing depth on low-abundant transcripts may hamper uncovering such transcripts' expression level variability across cells.

Importantly, the pipeline of our in-house 10X Genomics derivation method relies on the assumption that the primary PCR is highly specific to the EWSR1-WT1 transcript. The design of the forward oligonucleotide used for PCR#1 was generated to ensure such specificity. However, because this primer is aligned on the EWSR1-WT1 breakpoint, one cannot warrant that cDNAs containing matching sequences of both EWSR1 and WT1 partners would not be unspecifically amplified. Notably, the observation that more than 90% of sequenced transcripts aligned to WT1 gene when using EWSR1-WT1 targeting PCR#1 primer confirms that steps of biotin-streptavidin purification - enabled by the biotinylation of PCR#1 and #2 forward primers - efficiently discarded highly abundant transcripts present in the initial barcoded cDNAs pool. To facilitate results' interpretation in view of the risk of full-length WT1 non-specific amplification, we parallelly performed a control assay aiming at specifically amplifying full-length WT1 by using a primer specifically matching a region upstream EWSR1-WT1 breakpoint, which is not retained in EWSR1-WT1 transcript. The downstream analysis of the herein presented experiment attempted to consider this risk of full-length WT1 non-specific amplification by performing a positive selection of sequenced transcripts. Indeed, only UMIs

associated with a number of reads higher in the EWSR1-WT1 amplification assay than in the full-length WT1 amplification assay have been retained in the final analysis. One should note that any overamplification biases that may have been introduced over the three consecutive PCRs performed to generate the final library are overcome since the final transcript abundance is based on UMI counts - instead of transcript counts- and since UMIs are added during the first step of 10X Genomics Chromium protocol. Surprisingly, while performing this positive selection, EWSR1-WT1 expression could still be identified at low levels in some non-tumor cells (e.g., CAFs, immune cells, endothelial cells). This observation may also result from technical issues such as unspecific amplification (e.g., amplification of wild-type WT1 or any other Zinc finger-containing protein transcript showing partial homology with WT1), cell doublets encapsulation, or elusive biological mechanisms. Despite these technical difficulties and interpretational uncertainties, the main conclusion drawn from this experiment is that no substantial EWSR1-WT1 transcriptional expression level variability could be identified across DSRCT tumor cell clusters. Importantly, since the initial development of this technique, other authors described similar approaches based on the targeted amplification of 10X-derived 3'scRNA-seq experiments, with the aim of either measuring the transcription level of fusion transcripts<sup>164</sup> or performing single-cell level genotyping of transcriptomics, thereby underlining the relevance of our strategy. Overall, three main hypotheses might be drawn from our results. First, the prespecified limitations of our in-house single-cell EWSR1-WT1 gene expression targeted experiment may have hindered EWSR1-WT1 variable expression level detection across DSRCT cell clusters. Secondly, DSRCT cell heterogeneity may be driven by flexible EWSR1-WT1 activity secondary to variations in the fusion protein level or stability rather than transcript level modulations. Third, DSRCT cell plasticity may be instead directed by chromatin plasticity as a consequence of EWSR1-WT1 moldable DNA-binding sites, or triggered by microenvironment signals.

While a part of our study focused on exploring the latter hypothesis, several approaches may be envisioned to support these hypotheses. Adopting a similar strategy based on 10X-derived barcoded cDNA, followed by the generation of a library compatible with long-read sequencing (e.g., on Nanopore Promethion sequencing system), would enable to rule out the risk of misassignments of full-length WT1 transcripts to EWSR1-WT1, as the whole EWSR1-WT1 breakpoint region would be sequenced. Moreover, experiments such as inCITE-seq<sup>223</sup> have recently been developed to measure intranuclear protein levels and the nuclei's transcriptome simultaneously in a high-throughput and scalable fashion, hence enabling the joint analysis of transcription factors levels such as EWSR1-WT1 and gene expression *in vivo*. Also,

considering potential EWSR1-WT1 variable binding sites, experiments such as single-cell ChIP<sup>224,225</sup> have emerged and would be of great added value to explore this thesis.

## **2 PERSPECTIVES AND ONGOING WORK**

---

*For confidentiality reasons, the online version of this manuscript does not include this section.*

## REFERENCES

---

1. Wong, H. H. *et al.* Desmoplastic small round cell tumour: characteristics and prognostic factors of 41 patients and review of the literature. *Clin. Sarcoma Res.* **3**, 14 (2013).
2. Bulbul, A. *et al.* Desmoplastic Small Round Blue Cell Tumor: A Review of Treatment and Potential Therapeutic Genomic Alterations. *Sarcoma* <https://www.hindawi.com/journals/sarcoma/2017/1278268/> (2017) doi:10.1155/2017/1278268.
3. Gerald, W. L. *et al.* Intra-abdominal desmoplastic small round-cell tumor. Report of 19 cases of a distinctive type of high-grade polyphenotypic malignancy affecting young individuals. *Am. J. Surg. Pathol.* **15**, 499–513 (1991).
4. Gerald, W. L. *et al.* Clinical, pathologic, and molecular spectrum of tumors associated with t(11;22)(p13;q12): desmoplastic small round-cell tumor and its variants. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **16**, 3028–3036 (1998).
5. Lettieri, C. K., Garcia-Filion, P. & Hingorani, P. Incidence and outcomes of desmoplastic small round cell tumor: results from the surveillance, epidemiology, and end results database. *J. Cancer Epidemiol.* **2014**, 680126 (2014).
6. de Pinieux, G. *et al.* Nationwide incidence of sarcomas and connective tissue tumors of intermediate malignancy over four years using an expert pathology review network. *PloS One* **16**, e0246958 (2021).
7. Waqar, S. H. B. & Ali, H. Changing incidence and survival of desmoplastic small round cell tumor in the USA. *Bayl. Univ. Med. Cent. Proc.* **35**, 415–419 (2022).
8. Worch, J. *et al.* Racial differences in the incidence of mesenchymal tumors associated with EWSR1 translocation. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **20**, 449–453 (2011).
9. Biswas, G. *et al.* Desmoplastic small round cell tumor: extra abdominal and abdominal presentations and the results of treatment. *Indian J. Cancer* **42**, 78–84 (2005).
10. Honoré, C. *et al.* Can we cure patients with abdominal Desmoplastic Small Round Cell Tumor? Results of a retrospective multicentric study on 100 patients. *Surg. Oncol.* **29**, 107–112 (2019).
11. Honoré, C. *et al.* Abdominal desmoplastic small round cell tumor: multimodal treatment combining chemotherapy, surgery, and radiotherapy is the best option. *Ann. Surg. Oncol.* **22**, 1073–1079 (2015).
12. Lal, D. R. *et al.* Results of multimodal treatment for desmoplastic small round cell tumors. *J. Pediatr. Surg.* **40**, 251–255 (2005).
13. Backer, A. *et al.* Desmoplastic small round cell tumour of unknown primary origin with lymph node and lung metastases: histological, cytological, ultrastructural, cytogenetic and molecular findings. *Virchows Arch. Int. J. Pathol.* **432**, 135–141 (1998).
14. Karavitakis, E. M. *et al.* Desmoplastic small round cell tumor of the pleura. *Pediatr. Blood Cancer* **49**, 335–338 (2007).
15. Lee, J. C. *et al.* Clinicopathologic and molecular features of intracranial desmoplastic small round cell tumors. *Brain Pathol. Zurich Switz.* **30**, 213–225 (2020).
16. Saab, R., Khoury, J. D., Krasin, M., Davidoff, A. M. & Navid, F. Desmoplastic small round cell tumor in childhood: the St. Jude Children’s Research Hospital experience. *Pediatr. Blood Cancer* **49**, 274–279 (2007).
17. Zhang, J. *et al.* Analysis of Clinicopathological Features and Prognostic Factors of Desmoplastic Small Round Cell Tumor. *Pathol. Oncol. Res.* **20**, 161–168 (2014).
18. Cates, J. M. M. The AJCC 8th Edition Staging System for Soft Tissue Sarcoma of the Extremities or Trunk: A Cohort Study of the SEER Database. *J. Natl. Compr. Canc. Netw.* **16**,

144–152 (2018).

19. Gani, F., Goel, U., Canner, J. K., Meyer, C. F. & Johnston, F. M. A national analysis of patterns of care and outcomes for adults diagnosed with desmoplastic small round cell tumors in the United States. *J. Surg. Oncol.* **119**, 880–886 (2019).
20. Hayes-Jordan, A., Green, H., Fitzgerald, N., Xiao, L. & Anderson, P. Novel treatment for desmoplastic small round cell tumor: hyperthermic intraperitoneal perfusion. *J. Pediatr. Surg.* **45**, 1000–1006 (2010).
21. Saltsman, J. A. *et al.* A novel image-based system for risk stratification in patients with desmoplastic small round cell tumor. *J. Pediatr. Surg.* **55**, 376–380 (2020).
22. Casali, P. G. *et al.* Soft tissue and visceral sarcomas: ESMO–EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* **29**, iv51–iv67 (2018).
23. Fairweather, M. & Raut, C. P. To Biopsy, or Not to Biopsy: Is There Really a Question? *Ann. Surg. Oncol.* **26**, 4182–4184 (2019).
24. Mora, J. *et al.* Desmoplastic small round cell tumor 20 years after its discovery. *Future Oncol. Lond. Engl.* **11**, 1071–1081 (2015).
25. Cidre-Aranaz, F. *et al.* Small round cell sarcomas. *Nat. Rev. Dis. Primer* **8**, 66 (2022).
26. Sbaraglia, M., Bellan, E. & Dei Tos, A. P. The 2020 WHO Classification of Soft Tissue Tumours: news and perspectives. *Pathologica* **113**, 70–84 (2021).
27. Kallen, M. E. & Hornick, J. L. The 2020 WHO Classification: What’s New in Soft Tissue Tumor Pathology? *Am. J. Surg. Pathol.* **45**, e1–e23 (2021).
28. Publication of the WHO Classification of Tumours, 5th Edition, Volume 3: Soft Tissue and Bone Tumours – IARC. <https://www.iarc.who.int/news-events/publication-of-the-who-classification-of-tumours-5th-edition-volume-3-soft-tissue-and-bone-tumours/>.
29. Watson, S. *et al.* Transcriptomic definition of molecular subgroups of small round cell sarcomas. *J. Pathol.* **245**, 29–40 (2018).
30. Ladanyi, M. & Gerald, W. Fusion of the EWS and WT1 genes in the desmoplastic small round cell tumor. *Cancer Res.* **54**, 2837–2840 (1994).
31. Schoolmeester, J. K. *et al.* EWSR1-WT1 gene fusions in neoplasms other than desmoplastic small round cell tumor: a report of three unusual tumors involving the female genital tract and review of the literature. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **34**, 1912–1920 (2021).
32. Liu, J. *et al.* Molecular Heterogeneity and Function of EWSR1-WT1 Fusion Transcripts in Desmoplastic Small Round Cell Tumors. *Clin. Cancer Res.* **6**, 3522–3529 (2000).
33. Mohamed, M. *et al.* Desmoplastic small round cell tumor: evaluation of reverse transcription-polymerase chain reaction and fluorescence in situ hybridization as ancillary molecular diagnostic techniques. *Virchows Arch. Int. J. Pathol.* **471**, 631–640 (2017).
34. Lanic, M.-D. *et al.* Detection of sarcoma fusions by a next-generation sequencing based–ligation-dependent multiplex RT-PCR assay. *Mod. Pathol.* **35**, 649–663 (2022).
35. Subbiah, V. *et al.* Multimodality Treatment of Desmoplastic small round cell tumor: Chemotherapy and Complete Cytoreductive Surgery Improve Patient Survival. *Clin. Cancer Res.* clincanres.0202.2018 (2018) doi:10.1158/1078-0432.CCR-18-0202.
36. Verret, B. *et al.* Trabectedin in advanced desmoplastic round cell tumors: a retrospective single-center series. *Anticancer. Drugs* **28**, 116–119 (2017).
37. Uboldi, S. *et al.* Mechanism of action of trabectedin in desmoplastic small round cell tumor cells. *BMC Cancer* **17**, (2017).
38. Stacchiotti, S. *et al.* Trabectedin (T) in desmoplastic small round cell tumor (DSRCT): Report of its effect in 3 relapsed patients (pts) and the comparison of different regimens in a patient-derived xenograft (PDX) model. *J. Clin. Oncol.* **39**, e23553–e23553 (2021).
39. Frezza, A. M., Whelan, J. S. & Dileo, P. Trabectedin for desmoplastic small round cell tumours: a possible treatment option? *Clin. Sarcoma Res.* **4**, 3 (2014).

40. Frezza, A. M. *et al.* Pazopanib in advanced desmoplastic small round cell tumours: a multi-institutional experience. *Clin. Sarcoma Res.* **4**, 7 (2014).
41. Menegaz, B. A. *et al.* Clinical Activity of Pazopanib in Patients with Advanced Desmoplastic Small Round Cell Tumor. *The Oncologist* **23**, 360–366 (2018).
42. Lee, A. T. J., Jones, R. L. & Huang, P. H. Pazopanib in advanced soft tissue sarcomas. *Signal Transduct. Target. Ther.* **4**, 1–10 (2019).
43. Bandopadhyay, P. *et al.* The oncogenic properties of EWS/WT1 of desmoplastic small round cell tumors are unmasked by loss of p53 in murine embryonic fibroblasts. *BMC Cancer* **13**, 585 (2013).
44. Gedminas, J. M. *et al.* Desmoplastic small round cell tumor is dependent on the EWSR1-WT1 transcription factor. *Oncogenesis* **9**, 1–8 (2020).
45. Nambiar, M., Kari, V. & Raghavan, S. C. Chromosomal translocations in cancer. *Biochim. Biophys. Acta* **1786**, 139–152 (2008).
46. Carmona, R., Cano, E., Mattiotti, A., Gaztambide, J. & Muñoz-Chápuli, R. Cells Derived from the Coelomic Epithelium Contribute to Multiple Gastrointestinal Tissues in Mouse Embryos. *PLoS ONE* **8**, (2013).
47. Postel-Vinay, S. *et al.* Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma. *Nat. Genet.* **44**, 323–327 (2012).
48. Grünwald, T. G. P. *et al.* Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. *Nat. Genet.* **47**, 1073–1078 (2015).
49. Ballinger, M. L. *et al.* Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol.* **17**, 1261–1271 (2016).
50. Lin, S.-H. *et al.* Low-frequency variation near common germline susceptibility loci are associated with risk of Ewing sarcoma. *PLOS ONE* **15**, e0237792 (2020).
51. Gillani, R. *et al.* Germline predisposition to pediatric Ewing sarcoma is characterized by inherited pathogenic variants in DNA damage repair genes. *Am. J. Hum. Genet.* **109**, 1026–1037 (2022).
52. Tan, A. Y. & Manley, J. L. The TET Family of Proteins: Functions and Roles in Disease. *J. Mol. Cell Biol.* **1**, 82–92 (2009).
53. Morohoshi, F., Arai, K., Takahashi, E. I., Tanigami, A. & Ohki, M. Cloning and mapping of a human RBP56 gene encoding a putative RNA binding protein similar to FUS/TLS and EWS proteins. *Genomics* **38**, 51–57 (1996).
54. Call, K. M. *et al.* Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60**, 509–520 (1990).
55. Charlton, J. & Pritchard-Jones, K. WT1 Mutation in Childhood Cancer. *Methods Mol. Biol. Clifton NJ* **1467**, 1–14 (2016).
56. Hastie, N. D. Wilms' tumour 1 (WT1) in development, homeostasis and disease. *Development* **144**, 2862–2872 (2017).
57. Rauscher, F. J., Morris, J. F., Tournay, O. E., Cook, D. M. & Curran, T. Binding of the Wilms' tumor locus zinc finger protein to the EGR1 consensus sequence. *Science* **250**, 1259–1262 (1990).
58. Barbaux, S. *et al.* Donor splice-site mutations in WT1 are responsible for Frasier syndrome. *Nat. Genet.* **17**, 467–470 (1997).
59. Hammes, A. *et al.* Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell* **106**, 319–329 (2001).
60. Gerald, W. L., Rosai, J. & Ladanyi, M. Characterization of the genomic breakpoint and chimeric transcripts in the EWSR1-WT1 gene fusion of desmoplastic small round cell tumor. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1028–1032 (1995).
61. Gerald, W. L. & Haber, D. A. The EWSR1-WT1 gene fusion in desmoplastic small round cell tumor. *Semin. Cancer Biol.* **15**, 197–205 (2005).

62. Sawyer, J. R., Tryka, A. F. & Lewis, J. M. A novel reciprocal chromosome translocation t(11;22)(p13;q12) in an intraabdominal desmoplastic small round-cell tumor. *Am. J. Surg. Pathol.* **16**, 411–416 (1992).
63. Kim, J., Lee, K. & Pelletier, J. The desmoplastic small round cell tumor t(11;22) translocation produces EWS/WT1 isoforms with differing oncogenic properties. *Oncogene* **16**, 1973–1979 (1998).
64. Wong, J. C. *et al.* Induction of the interleukin-2/15 receptor  $\beta$ -chain by the EWS–WT1 translocation product. *Oncogene* **21**, 2009–2019 (2002).
65. Palmer, R. E. *et al.* Induction of BAIAP3 by the EWSR1-WT1 chimeric fusion implicates regulated exocytosis in tumorigenesis. *Cancer Cell* **2**, 497–505 (2002).
66. Hingorani, P. *et al.* Transcriptome analysis of desmoplastic small round cell tumors identifies actionable therapeutic targets: a report from the Children’s Oncology Group. *Sci. Rep.* **10**, 12318 (2020).
67. Reynolds, P. A. *et al.* Identification of a DNA-binding site and transcriptional target for the EWSR1-WT1(+KTS) oncoprotein. *Genes Dev.* **17**, 2094–2107 (2003).
68. Ito, E. *et al.* A tetraspanin-family protein, T-cell acute lymphoblastic leukemia-associated antigen 1, is induced by the Ewing’s sarcoma-Wilms’ tumor 1 fusion protein of desmoplastic small round-cell tumor. *Am. J. Pathol.* **163**, 2165–2172 (2003).
69. Kang, H.-J. *et al.* EWS–WT1 Oncoprotein Activates Neuronal Reprogramming Factor ASCL1 and Promotes Neural Differentiation. *Cancer Res.* **74**, 4526–4535 (2014).
70. Li, H. *et al.* Adenosine transporter ENT4 is a direct target of EWS/WT1 translocation product and is highly expressed in desmoplastic small round cell tumor. *PLoS ONE* **3**, e2353 (2008).
71. Karnieli, E., Werner, H., Rauscher, F. J., Benjamin, L. E. & LeRoith, D. The IGF-I receptor gene promoter is a molecular target for the Ewing’s sarcoma-Wilms’ tumor 1 fusion protein. *J. Biol. Chem.* **271**, 19304–19309 (1996).
72. Nishio, J. *et al.* Establishment and characterization of a novel human desmoplastic small round cell tumor cell line, JN-DSRCT-1. *Lab. Investig. J. Tech. Methods Pathol.* **82**, 1175–1182 (2002).
73. Riggi, N. *et al.* EWS-FLI1 Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. *Cancer Cell* **26**, 668–681 (2014).
74. Hartono, A. B. & Lee, S. B. Abstract 2386: Identification of SIK1 as a potential therapeutic target for desmoplastic small round cell tumor. *Cancer Res.* **78**, 2386–2386 (2018).
75. Lee, S. B. *et al.* The EWSR1-WT1 translocation product induces PDGFA in desmoplastic small round-cell tumour. *Nat. Genet.* **17**, 309–313 (1997).
76. Werner, H. *et al.* A novel EWSR1-WT1 gene fusion product in desmoplastic small round cell tumor is a potent transactivator of the insulin-like growth factor-I receptor (IGF-IR) gene. *Cancer Lett.* **247**, 84–90 (2007).
77. Taylor, B. S. *et al.* Advances in sarcoma genomics and new therapeutic targets. *Nat. Rev. Cancer* **11**, 541–557 (2011).
78. Nacev, B. A. *et al.* Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets. *Nat. Commun.* **13**, 3405 (2022).
79. Slotkin, E. K. *et al.* Comprehensive Molecular Profiling of Desmoplastic Small Round Cell Tumor. *Mol. Cancer Res. MCR* **19**, 1146–1155 (2021).
80. Sydow, S. *et al.* Genomic and transcriptomic characterization of desmoplastic small round cell tumors. *Genes. Chromosomes Cancer* (2021) doi:10.1002/gcc.22955.
81. Jiang, Y. *et al.* Novel Secondary Somatic Mutations in Ewing’s Sarcoma and Desmoplastic Small Round Cell Tumors. *PLoS ONE* **9**, e93676 (2014).
82. Silva, J. G. *et al.* Clinical next generation sequencing of pediatric-type malignancies in

- adult patients identifies novel somatic aberrations. *Oncoscience* **2**, 187–192 (2015).
83. Movva, S. *et al.* Multi-platform profiling of over 2,000 sarcomas: identification of biomarkers and novel therapeutic targets. *Oncotarget* **6**, 12234–12247 (2015).
  84. Chow, W. A. *et al.* Recurrent secondary genomic alterations in desmoplastic small round cell tumors. *BMC Med. Genet.* **21**, 101 (2020).
  85. Devecchi, A. *et al.* The genomics of desmoplastic small round cell tumor reveals the deregulation of genes related to DNA damage response, epithelial–mesenchymal transition, and immune response. *Cancer Commun.* **38**, (2018).
  86. Wu, C.-C. *et al.* Multi-site desmoplastic small round cell tumors are genetically related and immune-cold. *Npj Precis. Oncol.* **6**, 1–14 (2022).
  87. Cancer Genome Atlas Research Network. Electronic address: elizabeth.demicco@sinaihealthsystem.ca & Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* **171**, 950-965.e28 (2017).
  88. Brohl, A. S. *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet.* **10**, e1004475 (2014).
  89. Crompton, B. D. *et al.* The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* **4**, 1326–1341 (2014).
  90. Shern, J. F. *et al.* Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. *Cancer Discov.* **4**, 216–231 (2014).
  91. Filion, C. *et al.* The EWSR1/NR4A3 fusion protein of extraskeletal myxoid chondrosarcoma activates the PPAR $\gamma$  nuclear receptor gene. *J. Pathol.* **217**, 83–93 (2009).
  92. Hu-Lieskovan, S. *et al.* EWS-FLI1 fusion protein up-regulates critical genes in neural crest development and is responsible for the observed phenotype of Ewing’s family of tumors. *Cancer Res.* **65**, 4633–4644 (2005).
  93. Bulbul, A., Shen, J. P., Xiu, J., Tamayo, P. & Husain, H. Genomic and Proteomic Alterations in Desmoplastic Small Round Blue-Cell Tumors. *JCO Precis. Oncol.* 1–9 (2018) doi:10.1200/PO.17.00170.
  94. Smith, R. S. *et al.* Novel patient-derived models of desmoplastic small round cell tumor confirm a targetable dependency on ERBB signaling. *Dis. Model. Mech.* **15**, dmm047621 (2022).
  95. Byron, S. A. *et al.* Genomic and Transcriptomic Analysis of Relapsed and Refractory Childhood Solid Tumors Reveals a Diverse Molecular Landscape and Mechanisms of Immune Evasion. *Cancer Res.* **81**, 5818–5832 (2021).
  96. Lamhamedi-Cherradi, S.-E. *et al.* The androgen receptor is a therapeutic target in desmoplastic small round cell sarcoma. *Nat. Commun.* **13**, 3057 (2022).
  97. Hingorani, P. *et al.* RNA profiling of desmoplastic small round cell tumors (DSRCTs) using next-generation sequencing. *J. Clin. Oncol.* **34**, 10552–10552 (2016).
  98. Bleijs, M. *et al.* EWSR1-WT1 Target Genes and Therapeutic Options Identified in a Novel DSRCT In Vitro Model. *Cancers* **13**, 6072 (2021).
  99. Vibert, J. *et al.* Oncogenic chimeric transcription factors drive tumor-specific transcription, processing, and translation of silent genomic regions. *Mol. Cell* **82**, 2458-2471.e9 (2022).
  100. Koelsche, C. *et al.* Array-based DNA-methylation profiling in sarcomas with small blue round cell histology provides valuable diagnostic information. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **31**, 1246–1256 (2018).
  101. Wedekind, M. F. *et al.* Immune profiles of desmoplastic small round cell tumor and synovial sarcoma suggest different immunotherapeutic susceptibility upfront compared to relapse specimens. *Pediatr. Blood Cancer* **65**, e27313 (2018).

102. Bexelius, T. S., Wasti, A. & Chisholm, J. C. Mini-Review on Targeted Treatment of Desmoplastic Small Round Cell Tumor. *Front. Oncol.* **10**, (2020).
103. Mello, C. A. *et al.* Desmoplastic Small Round Cell Tumor: A Review of Main Molecular Abnormalities and Emerging Therapy. *Cancers* **13**, 498 (2021).
104. Blay, J.-Y. *et al.* 1271PHigh clinical benefit rates of pembrolizumab in very rare sarcoma histotypes: First results of the AcSé pembrolizumab study. *Ann. Oncol.* **30**, (2019).
105. Hayes-Jordan, A. A. *et al.* Efficacy of ONC201 in Desmoplastic Small Round Cell Tumor. *Neoplasia N. Y. N* **20**, 524–532 (2018).
106. Lowery, C. D. *et al.* Broad Spectrum Activity of the Checkpoint Kinase 1 Inhibitor Prexasertib as a Single Agent or Chemopotentiator Across a Range of Preclinical Pediatric Tumor Models. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **25**, 2278–2289 (2019).
107. Romo-Morales, A., Aladowicz, E., Blagg, J., Gatz, S. A. & Shipley, J. M. Catalytic inhibition of KDM1A in Ewing sarcoma is insufficient as a therapeutic strategy. *Pediatr. Blood Cancer* **66**, e27888 (2019).
108. van Erp, A. E. M. *et al.* Olaparib and temozolomide in desmoplastic small round cell tumors: a promising combination in vitro and in vivo. *J. Cancer Res. Clin. Oncol.* **146**, 1659–1670 (2020).
109. Ogura, K. *et al.* Therapeutic Potential of NTRK3 Inhibition in Desmoplastic Small Round Cell Tumor. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **27**, 1184–1194 (2021).
110. Hartono, A. B. *et al.* Salt-Inducible Kinase 1 is a potential therapeutic target in Desmoplastic Small Round Cell Tumor. *Oncogenesis* **11**, 1–10 (2022).
111. Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta* **1805**, 105 (2010).
112. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
113. Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr.* **4**, 5 (2013).
114. Shen, S. & Clairambault, J. Cell plasticity in cancer cell populations. *F1000Research* **9**, F1000 Faculty Rev-635 (2020).
115. Meacham, C. E. & Morrison, S. J. Tumor heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
116. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3983–3988 (2003).
117. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
118. Fanelli, G. N., Naccarato, A. G. & Scatena, C. Recent Advances in Cancer Plasticity: Cellular Mechanisms, Surveillance Strategies, and Therapeutic Optimization. *Front. Oncol.* **10**, (2020).
119. Cabrera, M. C., Hollingsworth, R. E. & Hurt, E. M. Cancer stem cell plasticity and tumor hierarchy. *World J. Stem Cells* **7**, 27–36 (2015).
120. Marjanovic, N. D., Weinberg, R. A. & Chaffer, C. L. Cell plasticity and heterogeneity in cancer. *Clin. Chem.* **59**, 168–179 (2013).
121. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
122. Yuan, S., Norgard, R. J. & Stanger, B. Z. Cellular Plasticity in Cancer. *Cancer Discov.* **9**, 837–851 (2019).
123. Non-genetic mechanisms of therapeutic resistance in cancer | Nature Reviews Cancer. <https://www.nature.com/articles/s41568-020-00302-4>.
124. Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nat. Rev. Drug Discov.* **19**, 39–56 (2020).

125. Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways | PNAS. <https://www.pnas.org/doi/10.1073/pnas.1816391116>.
126. Lehmann, S. *et al.* Hypoxia Induces a HIF-1-Dependent Transition from Collective-to-Amoeboid Dissemination in Epithelial Cancer Cells. *Curr. Biol. CB* **27**, 392–400 (2017).
127. Virk, R. K. A. *et al.* Disordered chromatin packing regulates phenotypic plasticity. *Sci. Adv.* **6**, eaax6232 (2020).
128. Tellez-Gabriel, M., Ory, B., Lamoureux, F., Heymann, M.-F. & Heymann, D. Tumour Heterogeneity: The Key Advantages of Single-Cell Analysis. *Int. J. Mol. Sci.* **17**, 2142 (2016).
129. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.* **52**, 1428–1442 (2020).
130. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
131. Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Dev. Camb. Engl.* **144**, 3625–3632 (2017).
132. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
133. Goetz, J. J. & Trimarchi, J. M. Transcriptome sequencing of single cells with Smart-Seq. *Nat. Biotechnol.* **30**, 763–765 (2012).
134. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
135. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* 1–7 (2020) doi:10.1038/s41587-020-0497-0.
136. Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* **14**, 267–270 (2017).
137. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
138. Jaitin, D. A. *et al.* Massively parallel single cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
139. Xin, Y. *et al.* Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proc. Natl. Acad. Sci.* **113**, 3293–3298 (2016).
140. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
141. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
142. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
143. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
144. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631–643.e4 (2017).
145. X, Z. *et al.* Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol. Cell* **73**, (2019).
146. Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clin. Transl. Med.* **12**, e694 (2022).
147. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
148. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 1–14 (2018).
149. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–

94 (2011).

150. Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402–412 (2012).
151. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
152. Chen, C. *et al.* Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**, 189–194 (2017).
153. Fan, X. *et al.* SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol.* **22**, 195 (2021).
154. Tang, J. *et al.* Single-cell exome sequencing reveals multiple subclones in metastatic colorectal carcinoma. *Genome Med.* **13**, 148 (2021).
155. Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148**, 886–895 (2012).
156. Gosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
157. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
158. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* 1–11 (2021) doi:10.1038/s41587-021-00869-9.
159. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022).
160. Liu, H. *et al.* DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120–128 (2021).
161. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
162. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
163. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
164. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
165. Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
166. Dou, M. *et al.* High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Anal. Chem.* **91**, 13119–13127 (2019).
167. Tsai, C.-F. *et al.* An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy for Precise Quantitative Single-cell Proteomics. *Mol. Cell. Proteomics MCP* **19**, 828–838 (2020).
168. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* **26**, 304–319 (2016).
169. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes | Nature Methods. <https://www.nature.com/articles/nmeth.3370>.
170. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **33**, 285–289 (2015).
171. Han, K. Y. *et al.* SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. *Genome Res.* **28**, 75–87 (2018).
172. TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and

- Parallel RNA Sequencing - ScienceDirect.  
<https://www.sciencedirect.com/science/article/pii/S266616672030112X>.
173. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).
  174. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
  175. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).
  176. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
  177. Genshaft, A. S. *et al.* Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* **17**, 188 (2016).
  178. Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016).
  179. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
  180. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
  181. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
  182. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
  183. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
  184. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
  185. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
  186. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
  187. Microscopic examination of spatial transcriptome using Seq-Scope - ScienceDirect.  
<https://www.sciencedirect.com/science/article/pii/S0092867421006279>.
  188. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq | Nature Biotechnology. <https://www.nature.com/articles/nbt.3445>.
  189. Lovatt, D. *et al.* Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat. Methods* **11**, 190–196 (2014).
  190. Medaglia, C. *et al.* Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
  191. Hu, K. H. *et al.* ZipSeq: barcoding for real-time mapping of single cell transcriptomes. *Nat. Methods* **17**, 833–843 (2020).
  192. Li, L. *et al.* What are the applications of single-cell RNA sequencing in cancer research: a systematic review. *J. Exp. Clin. Cancer Res.* **40**, 163 (2021).
  193. Advancing Cancer Research and Medicine with Single-Cell Genomics. *Cancer Cell* **37**, 456–470 (2020).
  194. Lei, Y. *et al.* Applications of single-cell sequencing in cancer research: progress and perspectives. *J. Hematol. Oncol. J Hematol Oncol* **14**, 91 (2021).
  195. Tang, Y. J. *et al.* Tracing Tumor Evolution in Sarcoma Reveals Clonal Origin of Advanced Metastasis. *Cell Rep.* **28**, 2837–2850.e5 (2019).
  196. Riggi, N., Suvà, M. L. & Stamenkovic, I. Ewing’s Sarcoma. *N. Engl. J. Med.* **384**, 154–

- 164 (2021).
197. Guillon, N. *et al.* The Oncogenic EWS-FLI1 Protein Binds In Vivo GGAA Microsatellite Sequences with Potential Transcriptional Activation Function. *PLoS ONE* **4**, (2009).
198. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171**, 163-178.e19 (2017).
199. Riggi, N. *et al.* EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* **26**, 668–681 (2014).
200. Franzetti, G.-A. *et al.* Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **36**, 3505–3514 (2017).
201. Chaturvedi, A., Hoffman, L. M., Welm, A. L., Lessnick, S. L. & Beckerle, M. C. The EWS/FLI Oncogene Drives Changes in Cellular Morphology, Adhesion, and Migration in Ewing Sarcoma. *Genes Cancer* **3**, 102 (2012).
202. Aynaud, M.-M. *et al.* Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cell Rep.* **30**, 1767-1779.e6 (2020).
203. Hong, B. *et al.* Single-cell transcriptional profiling reveals heterogeneity and developmental trajectories of Ewing sarcoma. *J. Cancer Res. Clin. Oncol.* (2022) doi:10.1007/s00432-022-04073-3.
204. Khoogar, R. *et al.* Single-cell RNA profiling identifies diverse cellular responses to EWSR1/FLI1 downregulation in Ewing sarcoma cells. *Cell. Oncol.* **45**, 19–40 (2022).
205. Keskin, T. *et al.* A live single-cell reporter assay links intratumor heterogeneity to metastatic proclivity in Ewing sarcoma. *Sci. Adv.* **7**, eabf9394 (2021).
206. Riggi, N. *et al.* EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev.* **24**, 916–932 (2010).
207. Jerby-Arnon, L. *et al.* Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. *Nat. Med.* **27**, 289–300 (2021).
208. Iura, K. *et al.* Cancer-testis antigen expression in synovial sarcoma: NY-ESO-1, PRAME, MAGEA4, and MAGEA1. *Hum. Pathol.* **61**, 130–139 (2017).
209. Wu, L. M. N. *et al.* Single-cell multiomics identifies clinically relevant mesenchymal stem-like cells and key regulators for MPNST malignancy. *Sci. Adv.* **8**, eabo5442 (2022).
210. Zhou, Y. *et al.* Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat. Commun.* **11**, 6322 (2020).
211. Liu, Y. *et al.* Single-Cell Transcriptomics Reveals the Complexity of the Tumor Microenvironment of Treatment-Naive Osteosarcoma. *Front. Oncol.* **11**, (2021).
212. Tessaro, F. H. G. *et al.* Single-cell RNA-seq of a soft-tissue sarcoma model reveals the critical role of tumor-expressed MIF in shaping macrophage heterogeneity. *Cell Rep.* **39**, 110977 (2022).
213. Wisdom, A. J. *et al.* Single cell analysis reveals distinct immune landscapes in transplant and primary sarcomas that determine response or resistance to immunotherapy. *Nat. Commun.* **11**, 6410 (2020).
214. Caldwell, B. *et al.* Multi-modal single-cell profiling of sarcomas from archival tissue reveals mechanisms of resistance to immune checkpoint inhibitors. *J. Clin. Oncol.* **40**, e23518–e23518 (2022).
215. Biezuner, T. *et al.* Comparison of seven single cell whole genome amplification commercial kits using targeted sequencing. *Sci. Rep.* **11**, 17171 (2021).
216. Nam, A. S. *et al.* Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355–360 (2019).

217. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
218. Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
219. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).
220. Savagner, P. The epithelial–mesenchymal transition (EMT) phenomenon. *Ann. Oncol.* **21**, vii89–vii92 (2010).
221. Bleijs, M. *et al.* EWSR1-WT1 Target Genes and Therapeutic Options Identified in a Novel DSRCT In Vitro Model. *Cancers* **13**, 6072 (2021).
222. Chen, C. *et al.* Single-cell multiomics reveals increased plasticity, resistant populations, and stem-cell-like blasts in KMT2A-rearranged leukemia. *Blood* **139**, 2198–2211 (2022).
223. Chung, H. *et al.* Simultaneous single cell measurements of intranuclear proteins and gene expression. 2021.01.18.427139  
<https://www.biorxiv.org/content/10.1101/2021.01.18.427139v1> (2021)  
doi:10.1101/2021.01.18.427139.
224. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
225. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).

## **APPENDICES**

*For confidentiality reasons, the online version of this manuscript does not include this section.*

**Titre :** Exploration de l'hétérogénéité cellulaire et moléculaire des Tumeurs Desmoplastiques à Petites Cellules Rondes.

**Mots clés :** Tumeur desmoplastique à petites cellules rondes, hétérogénéité tumorale, plasticité tumorale, séquençage sur cellule unique

**Résumé :** Les tumeurs desmoplastiques à petites cellules rondes (DSRCTs) sont un sous-type rare de sarcome de mauvais pronostic, dirigé par le facteur de transcription aberrant EWSR1-WT1. Malgré cet oncogène unique et la rareté des mutations secondaires récurrentes, les DSRCTs présentent une différenciation polyphénotypique caractéristique, suggérant la présence d'une hétérogénéité ou plasticité tumorale. En utilisant des techniques multiomiques sur cellule unique, nous avons mis en évidence que les cellules tumorales de DSRCT étaient caractérisées par un spectre continu de phénotypes cellulaires, caractérisés par des programmes transcriptionnels distincts. Les sous-populations de cellules tumorales identifiées étaient notamment liées à des voies de différenciation mésenchymateuse, neurale et épithéliale, ainsi qu'à des modifications des voies de contrôle métabolique.

Nos données suggèrent en outre le rôle de la variabilité des domaines de liaison à l'ADN de EWSR1-WT1 dans la plasticité chromatinienne des cellules de DSRCT. Par ailleurs, nos résultats mettent en évidence le rôle de signaux *in vivo* extrinsèques provenant d'interactions des cellules tumorales avec le microenvironnement, telles qu'avec la matrice extracellulaire et les fibroblastes associés au cancer, influençant ainsi l'expression de ces programmes transcriptionnels et la croissance tumorale. Nous avons également exploré l'applicabilité clinique de nos résultats et identifié des signatures transcriptionnelles dérivées du séquençage ARN sur cellule unique et ayant une valeur pronostique au sein de cohortes de patients atteints de DSRCT.

**Title :** Exploration of Desmoplastic Small Round Cell Tumors cellular and molecular heterogeneity.....

**Keywords :** Desmoplastic small round cell tumor, tumor heterogeneity, tumor cell plasticity, single-cell RNA-sequencing

**Abstract :**

Desmoplastic Small Round Cell Tumors (DSRCTs) are a rare sarcoma subtype of poor prognosis driven by the aberrant chimeric transcription EWSR1-WT1. Despite this unique oncogenic driver and the paucity of recurrent secondary mutations, DSRCTs harbor a characteristic polyphenotypic differentiation, suggesting the presence of tumor heterogeneity. Using single-cell multiomics, we found that DSRCT tumor cells were characterized by recurrent overlapping phenotypical states, which harbored distinct transcriptional programs, notably related to mesenchymal, neural, and epithelial cell fates and metabolic switch programs. Our data further suggest the role of EWSR1-WT1 DNA-binding versatility in DSRCT cells' chromatin plasticity and highlight additional *in vivo* extrinsic signals from microenvironment components like extracellular matrix and cancer-associated fibroblasts, which may influence DSRCT cell states and

tumor growth. We further explored the clinical applicability of our findings and identified single-cell RNA-sequencing-derived transcriptional signatures that may serve for patients' prognostic risk stratification.

