이학박사학위논문

ARID1A의 상분리에 의한 유잉 육종의 암 발생 기전 연구

Studies on the oncogenesis of Ewing's sarcoma mediated by phase separation of ARID1A

2025년 2월

서울대학교 대학원

생명과학부

김 용 률

Studies on the oncogenesis of Ewing's sarcoma mediated by phase separation of ARID1A

by

Yong Ryoul Kim

Advisor

Professor Sung Hee Baek, Ph.D.

A Thesis for the Degree of Doctor of Philosophy

February 2025

School of Biological Sciences
Seoul National University

ARID1A의 상분리에 의한 유잉 육종의 암 발생 기전 연구

Studies on the Oncogenesis of Ewing's Sarcoma Mediated by Phase Separation of ARID1A

지도교수 백 성 희

이 논문을 이학 박사 학위논문으로 제출함

2024년 12월

서울대학교 대학원

생명과학부

김 용 률

김용률의 박사학위논문을 인준함

2024년 12월

위 원 장 김 종 서 (인)

부위원장 백 성 희 (인)

위 원 김 근 일 (인)

위 원 김 익 수 (인)

위 원 장 원 열 (인)

ABSTRACT

Yong Ryoul Kim

School of Biological Sciences

The Graduate School

Seoul National University

Liquid-liquid phase separation (LLPS) facilitates the formation of membraneless organelles within cells, with implications in various biological processes and disease states. AT-rich interactive domain-containing protein 1A (ARID1A) is a chromatin remodeling factor frequently associated with cancer mutations, yet its functional mechanism remains largely unknown. Here, I find that ARID1A harbors a prion-like domain (PrLD), which facilitates the formation of liquid condensates through PrLD-mediated LLPS. The nuclear condensates formed by ARID1A LLPS are significantly elevated in Ewing's sarcoma patient specimen. Disruption of ARID1A LLPS results in diminished proliferative and invasive abilities in Ewing's sarcoma cells. Through genome-wide chromatin structure and transcription profiling, I identify that the ARID1A condensate localizes to EWS/FLI1 target enhancers and induces long-range chromatin architectural changes by forming

functional chromatin remodeling hubs at oncogenic target genes. Collectively, our findings demonstrate that ARID1A promotes oncogenic potential through PrLD-mediated LLPS, offering a potential therapeutic approach for treating Ewing's sarcoma.

Keywords

ARID1A, EWS/FLI1, BAF complex, Phase separation, Oncogenesis, Transcription, Chromatin structure, Nuclear condensate

Student Number: 2018-31266

CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER I
Introduction

I-1. Phase separation

1.1. General information about phase separation

The mechanism of gene expression within cells regulates not only the development of living organisms but also the onset and progression of various diseases, such as cancer. Recent research trends have shifted from focusing on individual gene expression to analyzing chromatin-wide structural changes epigenetically to uncover the fundamental causes of cell differentiation. Proteins containing intrinsically disordered region(IDR) play a critical role in facilitating the separation of cellular components. This capability stems from their ability to rapidly and reversibly transition between dissolved and aggregated states. Phase separation is driven by weak interactions between biomolecules, such as protein-nucleic interactions, electrostatic stacking, and hydrophobic interactions. This process has been recently reported to play a crucial role in various cellular processes, including the formation of membrane-less organelles, gene regulation, signal transduction, and stress responses. These liquid-liquid phase separations create specific cellular environments where essential biological phenomena, such as the formation of nucleoli, stress granules, and P-bodies, can occur (Bracha et al., 2019). Condensates formed through phase separation by transcription factors, co-activators, and mediator complexes promote spatial interactions between enhancers and promoters. These condensates bring distant regulatory elements, such as enhancers, into close proximity to target promoters, thereby enhancing transcriptional activation and gene expression. These structures are referred to as transcriptional condensates or transcriptional hubs because they

form concentrated hotspots of transcription factors, RNA polymerases, and associated co-factors, enabling efficient transcription initiation and elongation. Phase separation stabilizes these condensates and regulates their fluidity, thereby increasing the speed and efficiency of the transcription process (Peng et al., 2015). Recent studies have shown that liquid condensates not only regulate gene expression but also induce chromatin opening through epigenetic regulation and drive chromatin structural changes, forming a three-dimensional chromatin remodeling hub (Fig. I-1).

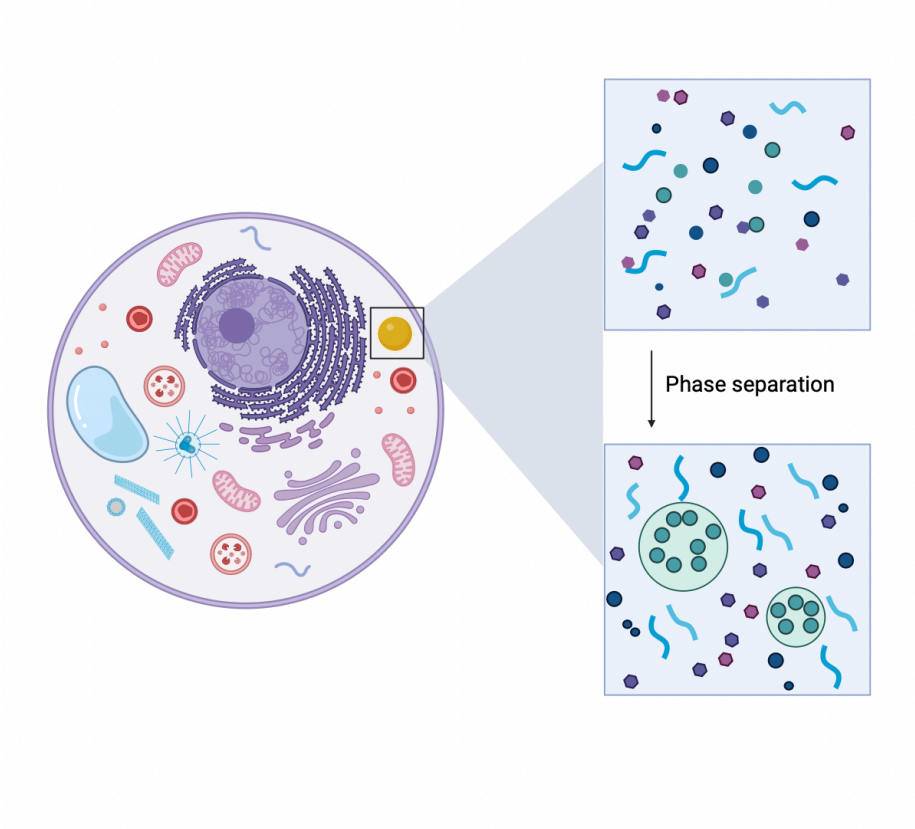

Figure I-1. Diagrammatic illustration of the phase separation in cell

Various cellular components within cell exist as membrane less organelle. These organelles are formed by the process of phase separation in which de-mixing of IDR containing protein occurs.

1.2. Disease related to phase separation

When molecules with phase separation capability adopt pathological forms, aberrant phase separation can lead to the development of various diseases, including cancers and degenerative brain disorders, as recently reported (Burke et al., 2022). Phase separation has drawn significant attention due to its potential to initiate oncogenic transcription programs during cancer progression. However, despite the anticipated correlation between liquid condensates and cancer development, the underlying causes and mechanisms remain largely unknown. Transcriptional hubs, which regulate gene expression patterns within the genome, play a critical role in the expression of homeostasis genes in cells. When these hubs are disrupted, it can lead to disease. Additionally, the formation of onco-condensates, which do not exist in normal cells, can promote the expression of oncogenes in cancer cells and play a crucial role in cancer progression. Several studies have shown that inhibiting the activity of transcriptional activators mediated by condensates can suppress tumorigenesis, highlighting the critical role of liquid phase separation by oncogenic molecules in various cancer types. Phase separation by oncogenic molecules containing intrinsically disordered regions is commonly observed in human tumors and suggests that "nuclear condensates" could serve as potential therapeutic targets for cancer treatment (Fig. I-2).

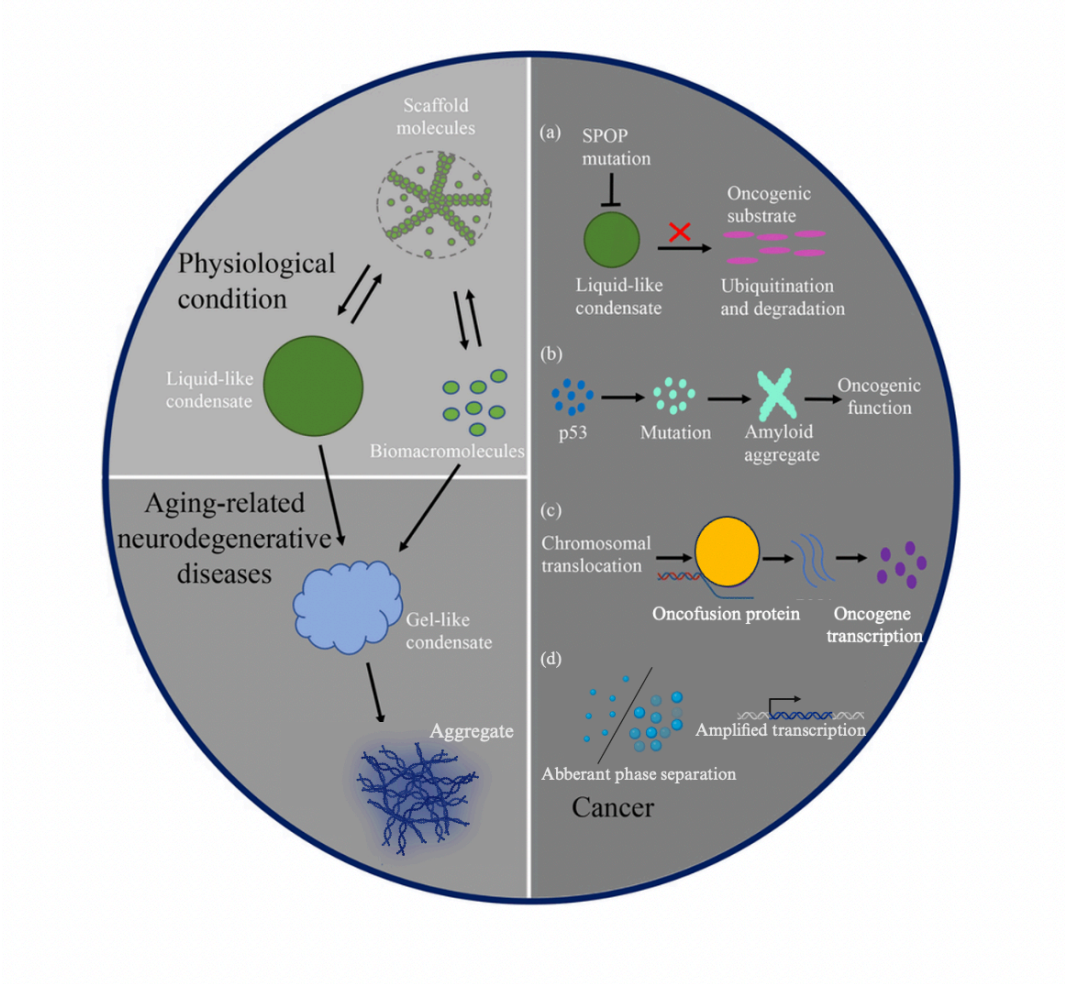

Figure I-2. Human diseases ranging from neurodegenerative disease and cancer are intricately related to phase separation

I-2. Chromatin remodeling complex

2.1. BAF complex

The BAF complex (BRG1/BRM-associated factor complex), also known as the mammalian SWI/SNF complex, is a highly conserved, ATP-dependent chromatin remodeling complex that plays a critical role in regulating gene expression, chromatin structure, and genome stability. By utilizing energy derived from ATP hydrolysis, the BAF complex alters chromatin architecture to facilitate or restrict access of transcription factors and other regulatory proteins to DNA. This complex is integral to various cellular processes, including differentiation, proliferation, DNA repair, and tumor suppression (Mandal et al., 2022).

The BAF complex is composed of several subunits encoded by distinct genes, which assemble into unique configurations depending on cellular context and developmental stage. Key components include ATPase subunits BRG1 (SMARCA4) or BRM (SMARCA2), which provide the ATP-dependent chromatin remodeling activity. Core subunits, SMARCB1 (BAF47), SMARCC1/2 (BAF155/170), and ARID1A/ARID1B or ARID2, play structural and regulatory roles. The diversity of subunits allows for the formation of specialized BAF complexes, including npBAF (neuronal progenitor BAF), nBAF (neuronal BAF), and esBAF (embryonic stem cell BAF), each tailored to the requirements of specific cell types and developmental stages (Wanior et al., 2021). BAF complex is a master regulator of chromatin accessibility and transcription. Its functions include regulation of gene expression by repositioning or ejecting nucleosomes. BAF complex also contributes

to DNA damage repair by facilitating access of repair machinery to damaged DNA regions.

2.2. Disease related to chromatin remodeler complex

Mutations in genes encoding BAF subunits are among the most frequent alterations in human cancers. For instance, ARID1A mutations are common in ovarian, endometrial, and gastric cancers, often leading to loss of tumor suppressor functions. SMARCB1 loss is a hallmark of malignant rhabdoid tumors. PBRM1 mutations are found in renal cell carcinomas, affecting chromatin remodeling and transcriptional regulation (Helming et al., 2014). The disruption of BAF complex activity can lead to widespread transcriptional dysregulation, promoting oncogenesis through aberrant chromatin states, altered DNA damage responses, and activation of oncogenic pathways. Research and Therapeutic Implications Emerging evidence highlights the potential of targeting the BAF complex in cancer therapy. For example, ARID1A-mutant cancers are selectively sensitive to EZH2 inhibitors, providing a targeted treatment strategy. Small molecule inhibitors are underway to develop drugs targeting specific components of the BAF complex or its interactions with chromatin.

I-3. ARID1A

3.1. General information about ARID1A

ARID1A (AT-Rich Interaction Domain 1A) is a critical subunit of the BAF (SWI/SNF) chromatin remodeling complex, a multi-protein assembly that regulates gene expression by altering chromatin structure to allow or restrict access to transcriptional machinery. ARID1A contains an AT-rich interaction domain (ARID), which facilitates DNA binding with a preference for AT-rich sequences, as well as other domains that mediate interactions with chromatin and regulatory proteins. This protein plays a pivotal role in maintaining normal cell functions, including transcriptional regulation, DNA damage repair, cell differentiation, and genome stability. By modulating chromatin accessibility, ARID1A acts as a master regulator of gene expression and epigenetic programming.

3.2. The structure of ARID1A

ARID1A contains 2286 amino acids in both human and mouse, and its amino acid sequences are highly conserved in species. ARID1A consists of various domains (ARID, HIC binding domain and PFAM homology domain). Interestingly, ARID domain is involved in DNA binding and PFAM homology domain is involved in direct interaction with numerous BAF complex subunits. Although ARID1A is a macromolecule, majority of its domain remains unannotated due to lack of ordered structure (Fig. I-3).

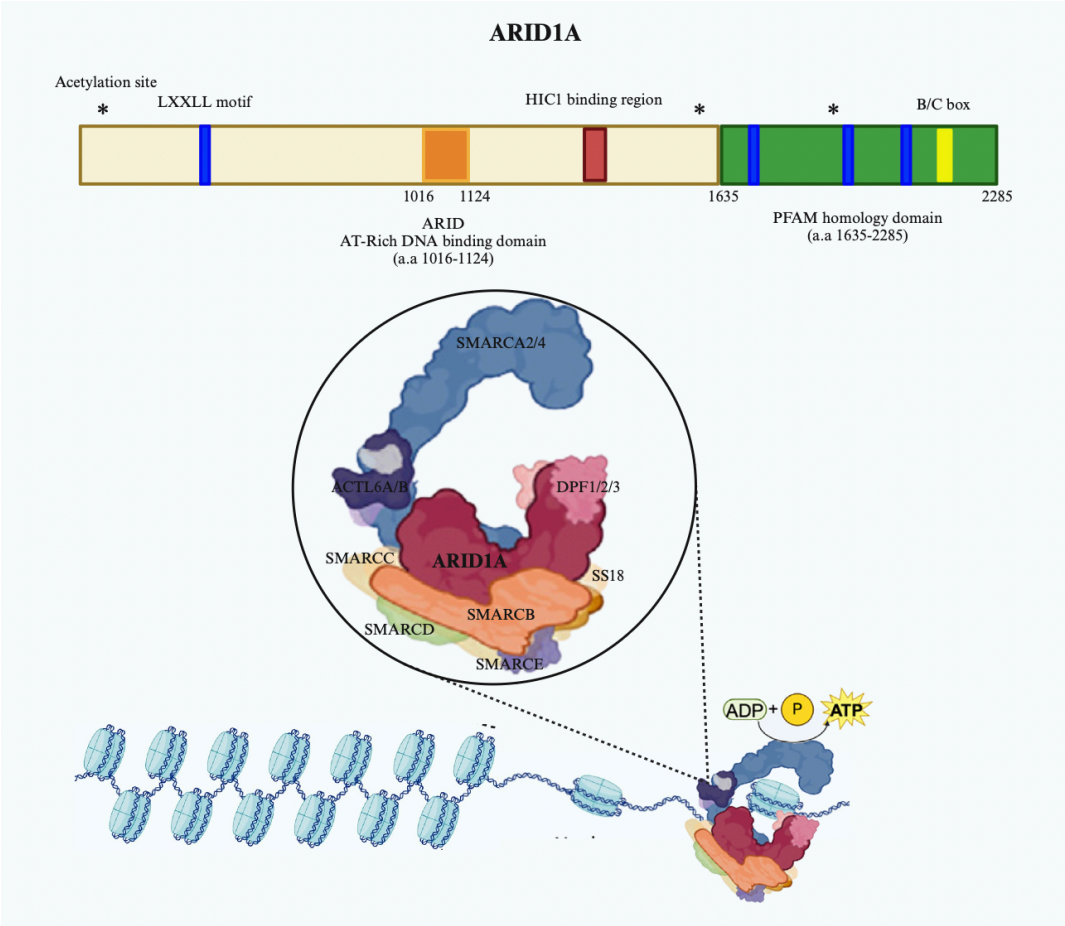

Figure I-3. Illustration of ARID1A structure

ARID1A contains 2285 amino acids in both human and mouse, and its amino acid sequences are highly conserved in species. ARID1A consists of ARID, HIC1 binding domain and PFAM homology domain. ARID1A directly interacts with BAF complex subunits to initiate chromatin remodeling.

3.3. ARID1A as a chromatin remodeler

Accumulating evidence supports that ARID1A functions as a transcriptional activator by opening target gene loci occupied by various transcription factors. In colorectal cancer, ARID1A has been shown to act at the activation protein 1 (AP1)-occupied enhancer and up-regulates associated genes involved in mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinases (ERK) signaling pathway (Sen et al., 2019). Further, ARID1A has been found to occupy luminal transcription factor loci bound by estrogen receptor α (ERα) and Forkhead Box A1 (FOXA1) in breast cancer (Xu et al., 2020). Loss of ARID1A results in luminal to basal transition and resistance to endocrine therapy. ARID1A has been shown to be frequently mutated in cancers, and recurrent mutations in ARID1A have been identified in a wide variety of cancers, including ovarian, breast, and pancreatic cancers (Berns et al., 2018; Nagarajan et al ., 2020; Zhang et al., 2018). These studies suggest the possibility that ARID1A plays critical role in regulating oncogenesis through chromatin remodeling (Fig. I-3).

I-4. EWS/FLI1

4.1. Ewing’s sarcoma

Ewing's Sarcoma is a rare and aggressive cancer that predominantly affects bones or the surrounding soft tissues. It is most commonly diagnosed in children, adolescents, and young adults, typically between the ages of 10 and 20. The disease often arises in long bones such as the femur, tibia, or humerus, as well as the pelvis, ribs, or scapula, and less commonly in soft tissues. Patients typically present with persistent pain, swelling at the tumor site, and, in advanced cases, systemic symptoms like fever, fatigue, and weight loss. Ewing's sarcoma is genetically characterized by a specific chromosomal translocation involving the EWSR1 gene on chromosome 22 and a member of the ETS transcription factor family, most commonly FLI1 on chromosome 11 (Fig. I-4). This translocation generates the EWS-FLI1 fusion protein, an aberrant transcription factor that disrupts normal gene expression and promotes oncogenesis by altering cell cycle regulation, apoptosis, and differentiation (Riggi et al., 2021).

4.2. The structure of EWS/FLI1

EWS/FLI1 is the hallmark fusion oncogene in Ewing's sarcoma, a rare and aggressive cancer that primarily affects bones and soft tissues. This oncogenic fusion results from a chromosomal translocation, most commonly t(11;22)(q24;q12), which fuses the EWSR1 gene on chromosome 22 with the FLI1 gene on chromosome 11. The resulting EWS-FLI1 fusion protein functions as an aberrant transcription factor and is a key driver of tumorigenesis in Ewing's sarcoma by rewiring transcriptional programs and promoting oncogenic pathways. Structurally, the EWS-FLI1 fusion protein combines domains from both parent genes, endowing it with unique oncogenic properties. The EWSR1 component contributes an intrinsically disordered N-terminal transactivation domain, which interacts with transcriptional coactivators and chromatin remodeling complexes. This disordered nature enhances its ability to form phase-separated transcriptional hubs, further amplifying oncogenic transcription (Zuo et al., 2021). The FLI1 component, on the other hand, provides a C-terminal ETS DNA-binding domain, which binds specifically to ETS consensus sequences in DNA. Together, these domains enable the fusion protein to redirect normal transcriptional machinery and activate oncogenic targets that are not typically regulated by wild-type FLI1 (Fig. I-4).

4.3. EWS/FLI1 mediated transcription and oncogenesis

The EWS/FLI1 fusion protein is a potent aberrant transcription factor that drives the oncogenesis of Ewing’s sarcoma by reprogramming transcriptional and epigenetic landscapes. One of its key features is its ability to bind specifically to GGAA microsatellite

repeats, which act as non-canonical enhancers. The repetitive GGAA sequences allow multivalent binding of EWS/FLI1, enabling the formation of transcriptionally active super-enhancers that amplify the expression of oncogenic target genes such as MYC, IGF1R, and CCND1, promoting proliferation, survival, and metabolic reprogramming (Riggi et al., 2021). Additionally, EWS/FLI1 represses tumor suppressor genes like TGFBR2, further tipping the balance towards oncogenesis. The GGAA repeats provide a bait for EWS/FLI1 recruitment, driving aberrant transcription at loci that are otherwise inactive in normal cells (Fig. I-4). Beyond direct transcriptional regulation, EWS/FLI1 interacts with chromatin remodeling complexes, such as the BAF complex, to induce long-range chromatin interactions, creating a permissive chromatin environment for its oncogenic transcriptional network. The fusion protein also promotes phase-separated transcriptional hubs, amplifying its transcriptional activity at critical oncogenic loci. These molecular mechanisms collectively contribute to the hallmark features of Ewing's sarcoma, including rapid proliferation, invasion, and metastasis.

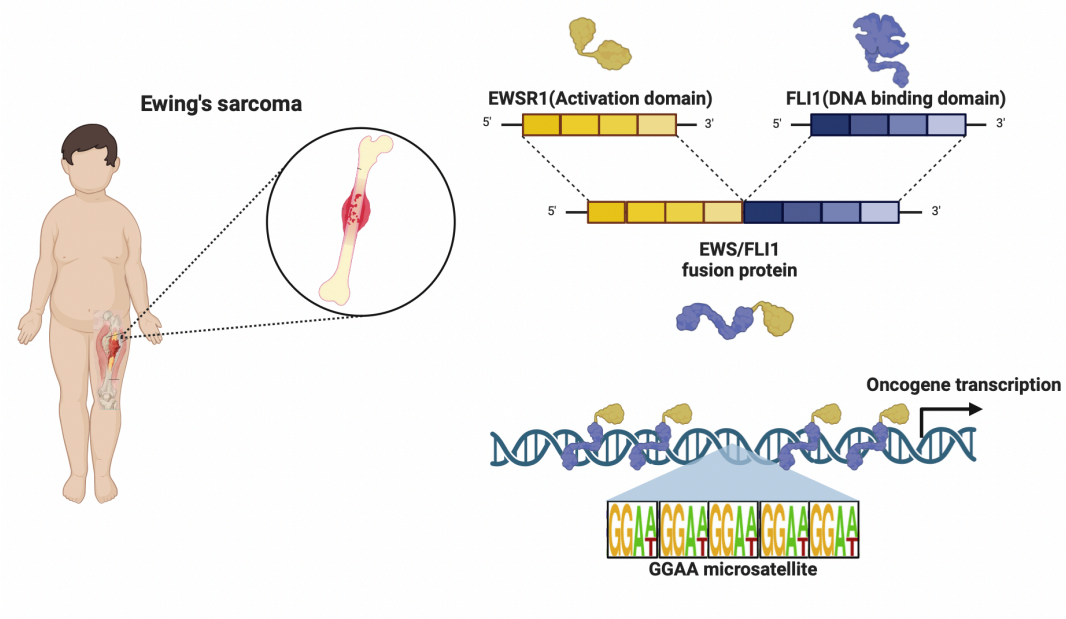

Figure I-4. Illustration of EWS/FLI1 mediated oncogenesis

Ewing's sarcoma is pediatric cancer that occurs in bone. Chromosome rearrangement takes place that results in EWS/FLI1 fusion protein expression. The fusion protein binds to GGAA repeat and leads to oncogene expression.

CHAPTER II

Phase separation of ARID1A is crucial for oncogenesis of Ewing's sarcoma

II-1. Summary

Here, I identified that ARID1A largely consists of intrinsically disordered domain. Using bioinformatics tools such as PONDR, PLAAC, and FOLDIndex, I found that two key PrLDs (PrLD1 and PrLD2) are critical drivers of ARID1A's phase separation properties. By tagging ARID1A with GFP and observing its behavior in cells, I confirmed that ARID1A forms liquid-like nuclear condensates. I found that tyrosine residues within PrLDs act as "stickers," crucial for phase separation, as tyrosine-to-serine substitutions (PrLD(Y/S)) abolished LLPS. Truncation experiments confirmed that PrLDs are indispensable for ARID1A's LLPS, while the Pfam domain is essential for recruiting BAF subunits into ARID1A condensates. Focusing on Ewing's sarcoma, I observed that ARID1A is highly expressed and forms nuclear condensates in patient tissues and cell lines, such as A673 and SK-N-MC. Using CRISPR-Cas9, I knocked out ARID1A and saw significant reductions in cell proliferation, migration, and invasion. Reconstitution of these cells with wild-type ARID1A restored their tumorigenic properties, but LLPS-deficient mutants (ΔDD or PrLD(Y/S)) failed to do so. In vivo xenograft experiments supported these results, showing that LLPS-deficient ARID1A mutants lack the ability to drive tumor progression. These findings demonstrate that ARID1A undergoes LLPS to form nuclear condensates, recruiting BAF subunits and driving oncogenic processes in Ewing's sarcoma. By impairing ARID1A's LLPS, I could significantly reduce its tumorigenic potential, highlighting the critical role of PrLD-mediated phase separation in cancer progression.

II-2. Introduction

Ewing's sarcoma, a rare but aggressive cancer, predominantly affects children and young adults, primarily targeting bones and soft tissues. Its hallmark genetic driver is the EWS/FLI1 fusion oncogene, formed by a balanced chromosomal translocation t(11;22)(q24;q12). This fusion protein functions as an aberrant transcription factor, altering normal gene expression patterns to promote oncogenesis (Riggi et al., 2021). Despite advances in understanding its molecular basis, the prognosis for advanced Ewing's sarcoma remains poor, necessitating novel therapeutic strategies (Riggi & Stamenkovic, 2007). The concept of liquid-liquid phase separation has revolutionized our understanding of cellular compartmentalization. Proteins containing intrinsically disordered regions or prion-like domains can undergo LLPS, forming biomolecular condensates that lack membranes but perform essential cellular functions (Hyman et al., 2014; Patel et al., 2015). LLPS-driven condensates are critical for transcriptional regulation, facilitating the assembly of transcriptional hubs and enhancing enhancer-promoter interactions. Aberrant LLPS, however, has been implicated in various diseases, including cancer, where it can amplify oncogenic transcriptional programs.

ARID1A, a subunit of the SWI/SNF (BAF) chromatin remodeling complex, has emerged as a key regulator of chromatin structure and gene expression. Recent studies revealed that EWS/FLI1 undergoes LLPS, driven by its PrLDs, forming nuclear condensates essential for its interaction with other chromatin remodelers and transcriptional machinery (Zuo et al., 2021). These condensates enable chromatin remodeler such as ARID1A to serve as a

structural scaffold, mediating chromatin accessibility and transcriptional regulation (Wanior et al., 2021). This study investigates ARID1A's role in the oncogenesis of Ewing's sarcoma through LLPS. I demonstrate that ARID1A forms nuclear condensates enriched with BAF subunits, facilitating oncogenic transcriptional programs. Importantly, disruption of ARID1A's LLPS significantly impairs the tumorigenic potential of Ewing's sarcoma cells, underscoring its potential as a therapeutic target.

II-3. Results

Prion-like domain drives liquid-liquid phase separation of ARID1A in the nucleus

ARID1A contains a few annotated domains, including the ARID DNA-binding domain, a nuclear localization signal (NLS), and a Pfam homology domain (Wu et al., 2021). Using PONDR and PLAAC, bioinformatic algorithms that are used to identify intrinsic disordered regions (IDRs) and PrLD, respectively, we found that ARID1A is primarily composed of PrLD (Fig. II-1A). ARID1A possesses two PrLDs separated by the ARID domain (hereafter referred to as PrLD1 and PrLD2). Moreover, FOLDIndex, a program that scores protein unfolding, indicated that both PrLD1 and PrLD2 contain smaller regions that are markedly unfolded and disordered compared to surrounding regions (annotated as disordered domain (DD) 1, and 2) (Fig. II-1A) (Prilusky et al., 2005). Lastly, catGRANULE analysis revealed that the phase separation propensity is high for both PrLD1 and PrLD2 (Fig. II-1A). To examine whether ARID1A exhibits phase separation in cells, I expressed GFP-ARID1A into 293T cells and monitored its subcellular distribution. GFP-ARID1A exhibited distinct nuclear condensates (Fig. II-1B). FRAP experiments on the nuclear ARID1A condensates showed rapid molecular rearrangements, confirming that ARID1A condensates are indeed liquid-like (Fig. II-1B). The ARID1A foci often grew in size through frequent fusion events and exhibited highly spherical morphology, suggesting that they have liquid-like physical properties (Fig. II-1C). Furthermore, intracellular ARID1A condensates were dissolved significantly upon the treatment of 1,6-hexanediol- compound that disrupts weak

hydrophobic interactions (Fig. II-1D). GFP-ARID1A generated discrete nuclear foci in a manner dependent on its concentration (Fig. II-1E).

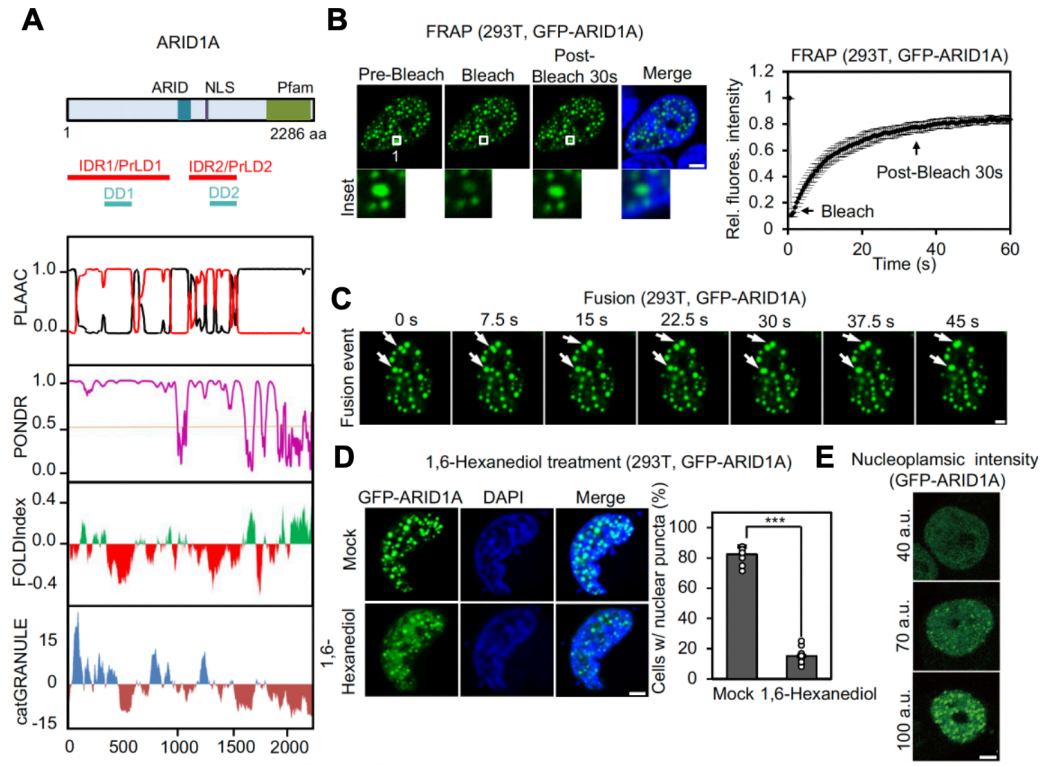

Figure II-1. Prion-like domain drives liquid-liquid phase separation of ARID1A in the nucleus

(A) Domain structure and intrinsic disorder tendency of ARID1A. The top panel shows the domains of ARID1A, along with PLAAC analysis, PONDR analysis, FOLD analysis, and catGRANULE analysis. (B) Representative images of the FRAP experiment conducted in GFP-ARID1A transfected 293T cells. The white box highlights the organelle subjected to targeted bleaching. The bottom presents the quantification of FRAP data for GFP-ARID1A puncta. Bleaching occurred at t = 0 second(s). Initial fluorescence was used as the reference value to calculate relative fluorescence intensity. Data are presented as the means ± SEMs (n = 9). n = individual ARID1A nuclear condensate Scale bar: 5 µm. (C) Live-cell imaging of 293T cells expressing GFP–ARID1A. The arrows indicate representative ARID1A puncta that fused over time. Scale bar: 2 µm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (D) GFP-ARID1A formed nuclear puncta in 293T cells. Cells transfected with GFP-

ARID1A were treated with or without 6 % Hex for 5 min and imaged using confocal microscopy. Nuclei were stained with DAPI. The quantification on the right shows the percentage of cells with nuclear puncta. Data are presented as the mean ± SEM. **$p < 0.01$, ***$p < 0.001$, NS non-significant. Statistics by two-tailed t-test. Twelve transfected cells from each group (mock and hexanediol treatment) were analyzed; n= 12 biologically independent samples. Scale bar: 5 μm. (E) Representative confocal images of 293T cells expressing GFP-ARID1A at different fluorescence intensity. Scale bar: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome.

Aromatic residues within Prion-like domain is responsible for phase separation of ARID1A

I then sought to comprehensively identify the regions of ARID1A that are responsible for driving LLPS. By employing PONDR and PLAAC amino acid sequence-based analysis, I observed that ARID1A possesses two ordered domains including the ARID domain and the Pfam homology domain and two disordered domains including PrLD1, and PrLD2. Proteins that segregate through the phase separation process generally have specific domains required for driving condensation (Rawat et al., 2021; Yu et al., 2021). To dissect the parts of ARID1A which play a key role in inducing LLPS, I generated several truncation variants which included different regions of ARID1A, and observed their cellular localization (Fig. II-2A). GFP-fused full-length ARID1A formed distinct nuclear condensates, whereas control GFP expression showed a widely spread staining pattern throughout the entire cell. Since PrLD1 doesn't have a NLS, it remained in the cytoplasm, forming a big, clearly visible segment in the bright field. Conversely, the NLS-possessing PrLD2 took the form of a distinct condensate within the nucleus. ARID DNA-binding domain and Pfam homology domain, which lack PrLD, failed to undergo LLPS (Fig. II-2A). To confirm that PrLD1 and PrLD2 are the key regions that induce ARID1A LLPS, I generated an ARID1A deletion mutant lacking disordered domains of ARID1A ΔDD). This ΔDD mutant failed to undergo LLPS (Fig. II-2A), proving the indispensable roles of disordered domains of PrLD1 and PrLD2 in ARID1A condensation. Intermolecular interactions driving phase separation of associative polymers have been described using a

stickers-and-spacers framework. Stickers exhibit associative interactions with one another to drive phase separation. For PrLDs, aromatic residues such as tyrosine are shown to act as stickers. On the other hand, spacers are linkers that connect stickers, and play modulatory roles in phase separation. To probe whether tyrosine residues are important for phase separation of ARID1A, I generated an ARID1A variant of which its 52 tyrosine residues in the PrLDs were replaced with serine. Live cell imaging using the ARID1A PrLD(Y/S) mutant revealed that the disruption of tyrosine interactions completely abolished the PrLD-mediated LLPS of ARID1A (Fig. II-2A), confirming that aromatic residues within the PrLD are key drivers of LLPS. Lastly, I studied the concentration-related formation of nuclear ARID1A condensates by tracking the nucleoplasmic concentrations at which ARID1A proteins started to form condensate. As a result, I discovered a discernible threshold at which ARID1A initiates the formation of condensates. Condensate-forming proteins, including ARID1A WT, NLS-PrLD1, and PrLD2, exhibited a gradual increase in the size and number of nuclear condensates as their molecular concentrations surpassed the threshold. (Fig. II-2B). Nevertheless, ARID domain, Pfam homology domain, and LLPS deficient mutants were incapable of forming nuclear condensate, proving the critical role of PrLDs of ARID1A in condensate formation (Fig. II-2B).

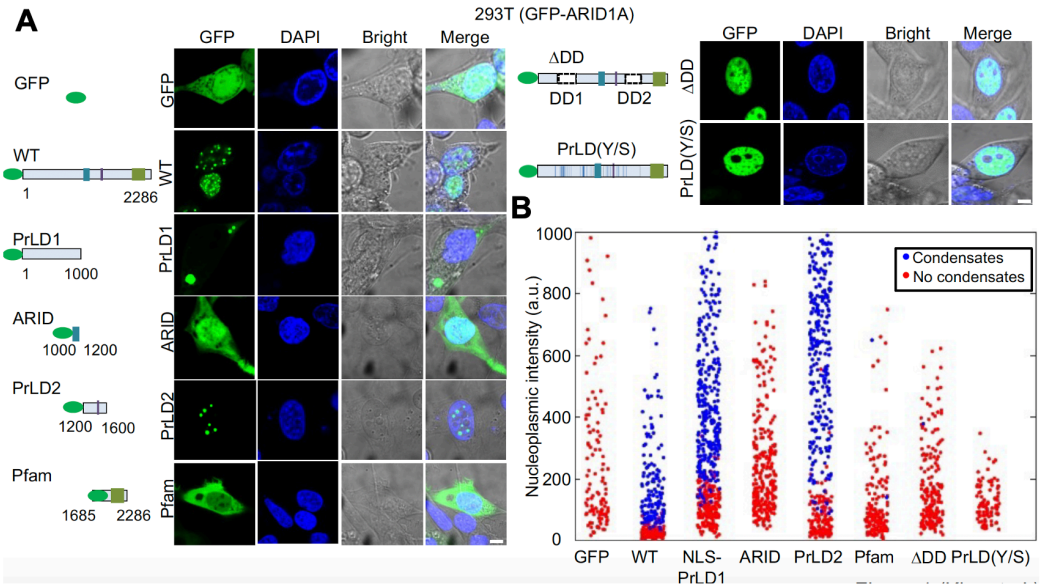

Figure II-2. Aromatic residues within Prion-like domain is responsible for phase separation of ARID1A

(A) Representative confocal images of 293T cells transfected with different forms of recombinant GFP-ARID1A constructs, including GFP, GFP-ARID1A, GFP-PrLD1, GFP-ARID, GFP-PrLD2, GFP-Pfam, GFP-ARID1A PrLD(Y/S), and GFP-ΔDD mutant. Scale bar: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (B) Quantitative phase diagram depicting the intra-nuclear concentration of ARID1A domains and mutants observed. Each dot represents the ARID1A concentration from a unique cell. Red indicates positive phase separation, while blue indicates negative phase separation. (a.u= arbitrary unit).

ARID1A undergoes phase separation *in vitro*

To evaluate LLPS ability of ARID1A in vitro, I purified GFP-labeled ARID1A protein and performed in vitro droplet assay. When examined with fluorescence microscopy, I found that GFP-ARID1A indeed formed highly spherical droplet-like assemblies (Fig. II-3A). Consistent with the liquid state, ARID1A assemblies exhibited rapid shape relaxation after fusing with one another (Fig. II-3A). To further probe protein dynamics within ARID1A assemblies, I performed fluorescence recovery after photobleaching (FRAP) experiments (Bolognesi et al., 2016). Bleached regions showed a near-complete fluorescence recovery (Fig. II-3B), confirming that GFP-ARID1A formed dense liquid droplets with high molecular mobility. As expected for LLPS, I found that ARID1A droplets were microscopically observable only above a threshold concentration of 1-1.25 μM and bigger droplets were observed at higher protein concentrations (Fig. II-3C). The assembled droplets were highly susceptible to 1,6-hexanediol treatment, indicating that hydrophobic interactions play important roles in driving ARID1A phase separation (Fig. II-3D). Taken together, my results indicate that ARID1A undergoes LLPS both in vitro and in cells. In vitro droplet assay using purified recombinant proteins was performed (Fig. II-3E). PrLD1 and PrLD2 regions exhibited strong concentration-dependent phase separation behaviors, similar to the full length ARID1A. In contrast, neither ARID domain nor Pfam homology domain showed phase separation up to 10 μM. In agreement with results from in cellulo experiments, ΔDD and PrLD(Y/S) mutant failed to undergo LLPS (Fig. II-3E).

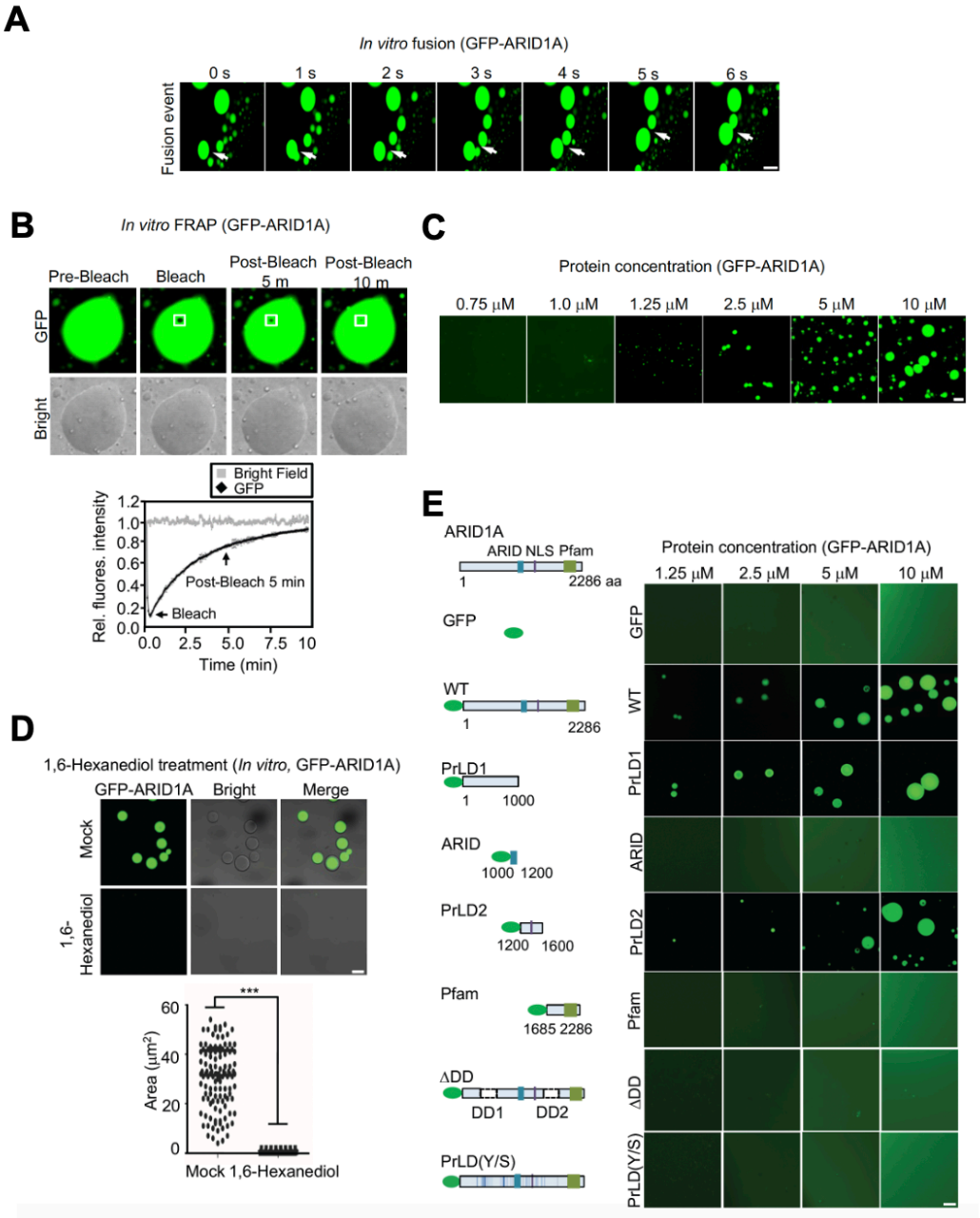
A
In vitro fusion (GFP-ARID1A)
0 s
1 s
2 s
3 s
4 s
5 s
6 s
Fusion event
B
In vitro FRAP (GFP-ARID1A)
Pre-Bleach
Bleach
Post-Bleach 5 m
Post-Bleach 10 m
GFP
Bright
Bright Field
GFP
Rel. fluores. intensity
Post-Bleach 5 min
Bleach
Time (min)
C
Protein concentration (GFP-ARID1A)
0.75 μM
1.0 μM
1.25 μM
2.5 μM
5 μM
10 μM
D
1,6-Hexanediol treatment (In vitro, GFP-ARID1A)
GFP-ARID1A
Bright
Merge
Mock
1,6-Hexanediol
Area (μm²)
***
Mock
1,6-Hexanediol
E
ARID1A
ARID NLS Pfam
1
2286 aa
GFP
WT
1
2286
PrLD1
1
1000
ARID
1000 1200
PrLD2
1200 1600
Pfam
1685 2286
ΔDD
DD1
DD2
PrLD(Y/S)
Protein concentration (GFP-ARID1A)
1.25 μM
2.5 μM
5 μM
10 μM
GFP
WT
PrLD1
ARID
PrLD2
Pfam
ΔDD
PrLD(Y/S)

Figure II-3. ARID1A undergoes phase separation *in vitro*

(A) Live imaging of GFP-ARID1A droplets fusing over time in vitro. Scale bar: 5 μm. (B) Representative confocal images of the FRAP experiment conducted on droplets formed in vitro. The graph illustrates the fluorescence intensity of the bleached droplet. The error bars represent the mean ± standard error of the mean (SEM). Scale bar: 5 μm. (C) Representative confocal images of concentration-dependent droplet formation in vitro. Scale bar: 5 μm. (D) 1,6-hexanediol (Hex; 6 %) disrupted droplet formation in vitro. GFP-ARID1A (10 μM) was analyzed at room temperature (20 °C) with or without 6 % Hex. The upper panel show representative fluorescence images of the droplets, while the lower panel quantifies the size and number of droplets. Each dot represents a droplet. Data are presented as the mean ± SEM of n = 125. Scale bar: 5 μm. (E) Right: Representative confocal images of a concentration series of droplet formation assays evaluating homotypic droplet formation for various proteins, including GFP, GFP-WT ARID1A, GFP-PrLD1, GFP-ARID, GFP-PrLD2, GFP-Pfam, GFP-ΔDD and GFP-ARID1A PrLD(Y/S) mutant. Scale bar: 5 μm. Left: Diagram illustrating the different forms of recombinant GFP-ARID1A proteins tested.

ARID1A requires both PrLD and Pfam homology domain to incorporate BAF subunits into condensate

ARID1A serves as a structural core and scaffold in the structural organization of BAF complex, and c-terminal Pfam homology domain of ARID1A mediates direct molecular contact with other BAF subunits to maintain stable base module of BAF complexes (He at al., 2020). Therefore, I next examined whether the ARID1A condensate possesses the ability to compartmentalize its chromatin remodeler cofactors via phase separation. First, I checked whether loss of ARID1A LLPS can affect its interaction with BAF subunits. I performed co-immunoprecipitation assay and found that both WT and ΔDD mutant maintain its interaction with BAF complex subunits including SMARCA2, SMARCB, SMARCD, SMARCC1, and SMARCE, whereas ΔPfam mutant failed to interact with BAF complex subunits (Fig. II-4A). These data indicate that phase separation ability of ARID1A does not affect its interaction with BAF complex. When overexpressed in 293T cells, unlike ARID1A, other BAF subunits were diffusively distributed (Fig. II-4B). To investigate the capacity of ARID1A to compartmentalize the BAF subunits, I co-expressed various ARID1A truncation constructs with either SMARCB1 (Fig. II-4C) or SMARCD1 (Fig. II-4D). Due to lack of NLS in PrLD1, NLS was artificially added to acquire nuclear condensate of PrLD1. Though NLS-PrLD1 formed a nuclear condensate, SMARCB1 and SMARCD1 remained diffuse, thus the lack of partitioning was revealed. Successful incorporation BAF subunits by ARID1A WT condensate was contrasted by the Y/S mutants of PrLD and ΔDD, as the two couldn't form a nuclear droplet, and SMARCB1 and

SMARCD1 were not segregated appropriately. Lastly, analogous to NLS-PrLD1, ΔPfam formed distinct condensates in the nucleus yet failed to compartmentalize SMARCB1 and SMARCD1 (Fig. II-4C,D). Together, these data indicate that Pfam domain of ARID1A is essential not only for its interaction with BAF complex subunits but also for partitioning of BAF complex subunits into ARID1A condensate.

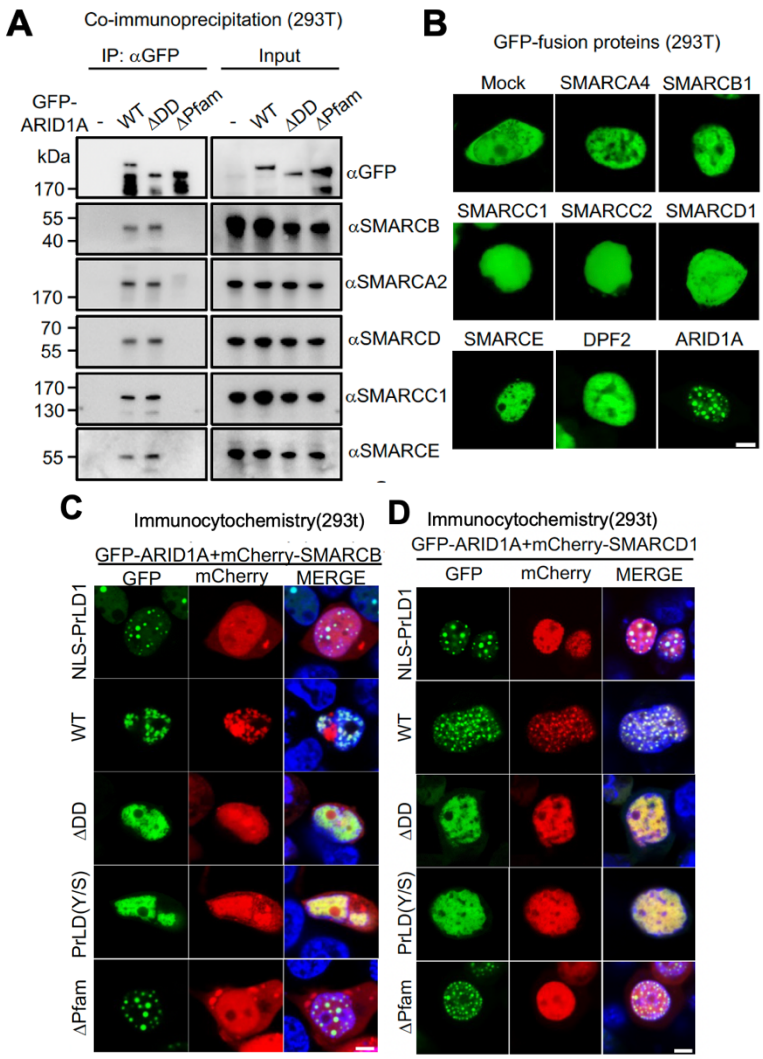

Figure II-4. ARID1A requires both PrLD and Pfam homology domain to incorporate BAF subunits into condensate

(A) Co-immunoprecipitation assay performed to detect the interaction between endogenous BAF complex subunit and ARID1A wild type (WT), ΔDD, or ΔPfam mutant expressed in 293T cells. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (B) Representative confocal images

showing the cellular localization of different GFP-BAF complex subunits. Scale bar: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (C) Representative confocal images demonstrating the colocalization pattern of recombinant ARID1A proteins (green) and SMARCB1 (red). Scale bar: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (D) Representative confocal images illustrating the colocalization pattern of recombinant ARID1A proteins (green) and SMARCD1(red). Scale bar: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome.

Light inducible Corelet system exihibit spontaneous ARID1A phase separation and subsequent BAF complex subunit recruitment

My results suggest the modular domain organization of ARID1A for phase separation: PrLDs drive the formation of condensates while the Pfam domain tunes their compositions. To further test this idea, I took a synthetic approach to build up condensates using the previously developed light-inducible Corelet system (Fig. II-5A). The Corelet components include a 24-mer ferritin core appended with iLID (improved light induced dimer) and IDRs fused with sspB (stringent starvation protein B). Blue light activation leads to dimerization between iLID and sspB, giving rise to IDR oligomers, which can ultimately trigger phase separation (Fig. II-5B) When PrLD1 of ARID1A was used as an IDR fusion to sspB, I observed a strong blue-light dependent formation of nuclear condensates. However, PrLD1 Corelet condensates failed to recruit SMARCB1 and SMARCD1. In a sharp contrast, appending Pfam domain to PrLD1 (PrLD1-Pfam-sspB) altered the composition of Corelet condensates with clear partitioning of these BAF complex subunits (Fig. II-5B).

To validate that ARID1A condensates hold endogenous BAF subunits, I conducted immunocytochemistry using anti-SMARCD1 and anti-SMARCC1 antibodies. Co-immunostaining of GFP-ARID1A condensate and endogenous BAF showed a notable concentration of BAF subunits (Fig. II-5). The molecules were diffusive when a phase separation defective mutant of ARID1A was expressed. In addition, the condensate formed by ARID1A without the Pfam domain was unable to attract BAF components and they stayed diffusive in the nucleus (Fig. II-5C). Thus, my data indicate that ARID1A builds

nuclear condensates via PrLD-induced phase separation and recruits BAF subunits with the aid of Pfam homology domain.

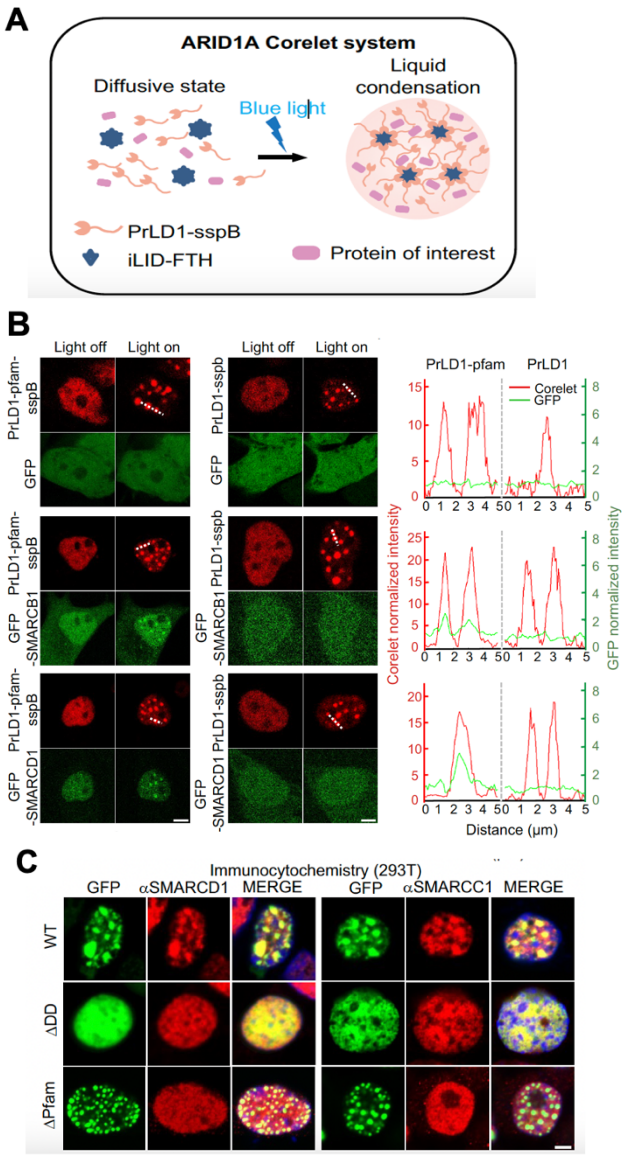
A
ARID1A Corelet system
Diffusive state
Blue light
Liquid condensation
PrLD1-sspB
iLID-FTH
Protein of interest
B
Light off
Light on
PrLD1-pfam-sspB
PrLD1-sspb
GFP
GFP -SMARCB1
GFP -SMARCD1
PrLD1-pfam
PrLD1
Corelet
GFP
Corelet normalized intensity
GFP normalized intensity
Distance (μm)
C
Immunocytochemistry (293T)
GFP
αSMARCD1
MERGE
αSMARCC1
WT
ΔDD
ΔPfam

Figure II-5. Light inducible Corelet system exihibit spontaneous ARID1A phase separation and subsequent BAF complex subunit recruitment

(**A**) Schematics of ARID1A corelet system. (**B**) Representative confocal images of the corelet system using PrLD1-mch-SspB or PrLD1-Pfam-mch-SspB to observe recruitment of BAF complex subunit upon blue light stimulation. Scale bars: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (**C**) Representative confocal images of 293T cells transfected with recombinant ARID1A and immunostained with anti-SMARCD and anti-SMARCC1 antibodies. Scale bars: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome.

ARID1A exhibit aberrant condensate in Ewing′s sarcoma patient tissue

Next, I examined the biological context in which endogenous ARID1A LLPS is observed. Quantitative proteomics to analyze 375 cancer cell lines have generated a cancer cell line encyclopedia (Nusinow et al., 2020). Taking advantage of this resource, I explored the protein levels of ARID1A across multiple types of cancer, and found that ARID1A protein level was significantly high in Ewing′s sarcoma compared to many other cancer types (Fig. II-6A). This unusual high expression of ARID1A in Ewing′s sarcoma was validated by immunoblot analysis in multiple cancer cell lines including Ewing′s sarcoma cell lines (A673 and SK-N-MC) (Fig. II-6B). I sought to explore if there were detectable nuclear condensates in Ewing's sarcoma patient tissues resulting from the concentration-dependent ARID1A LLPS that was observed *in vitro* and inside cells. Using tumor samples from two independent patients with Ewing's sarcoma, I performed immunohistochemistry and imaged the localization of ARID1A. Surprisingly, ARID1A showed increased expression in Ewing's sarcoma patient tissue resulting in visible nuclear foci, while normal bone tissue exhibited much lower punctate pattern throughout the nucleus (Fig. II-6C). Therefore, I decided to test the oncogenic potential of ARID1A in Ewing's sarcoma cell line using cell proliferation, invasion and migration assays. I generated ARID1A knockout (ARID1A-/-) A673 cell line using CRISPR-CAS9 gene editing. Next, I rescued ARID1A knockout cell line with either WT or LLPS-defective mutant DD. Immunocytochemistry data showed that the endogenous ARID1A formed condensate within WT and ARID1A-/-+WT cells (LLPS positive), whereas ARID1A was diffusive in ARID1A-/-+ΔDD cells (LLPS negative) (Fig.

II-6D). A wound healing assay was then performed using LLPS positive and LLPS negative cell lines to assess the motility and ability of each generated cancer cell line to recover the scratch generated on the surface of culture plate. Knockout of ARID1A showed decreased cell migration and motility rate, whereas reconstitution of WT, but not ΔDD mutant, reversed the reduced migration rate (Fig. II-6E).

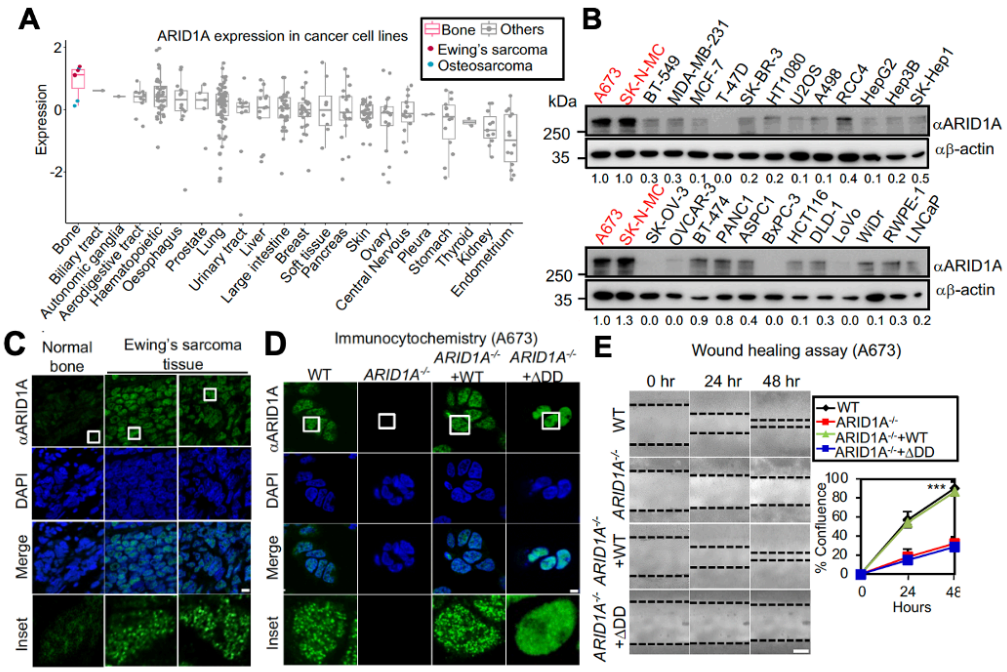

Figure II-6. ARID1A exhibit aberrant condensate in Ewing′s sarcoma patient tissue

(**A**) Expression levels of ARID1A protein in different types of cancer obtained from the cancer cell line encyclopedia. (**B**) Representative immunoblot image measuring ARID1A protein levels in various cancer cell lines. The quantification represents ARID1A/β-actin protein density ratio. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (**C**) Immunohistochemistry results showing ARID1A staining in normal bone tissue and two Ewing's sarcoma patient tissues. Scale bars: 10 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (**D**) Immnuocytochemistry image illustrating endogenous ARID1A localization in WT, ARID1A-/-, ARID1A-/-+WT and ARID1A-/-+ΔDD cells. Scale bars: 5 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (**E**) Left: Wound healing assay conducted on WT, ARID1A-/-, ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines. Right:

Quantification of the wound healing assay. Bars represents the SEM; **$p < 0.01$, ***$p < 0.001$, NS non-significant. n=10 technical replicate of wound closures. Statistical analysis performed using a two tailed Wilcoxon signed rank test on 48hr samples of ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines. Scale bar, 500 μm.

Protein expression levels of BAF subunits are elevated in Ewing's sarcoma

Since BAF complex subunits are enriched inside the ARID1A condensate, I further analyzed protein expression of BAF subunits using CCLE data (Fig. II-7A). As in case of ARID1A, BAF subunits also showed increased protein level specifically in Ewing's sarcoma (Fig. II-7A). Moreover, protein levels of ARID1A and each BAF subunit were positively correlated across various cancer types (Fig. II-7A).

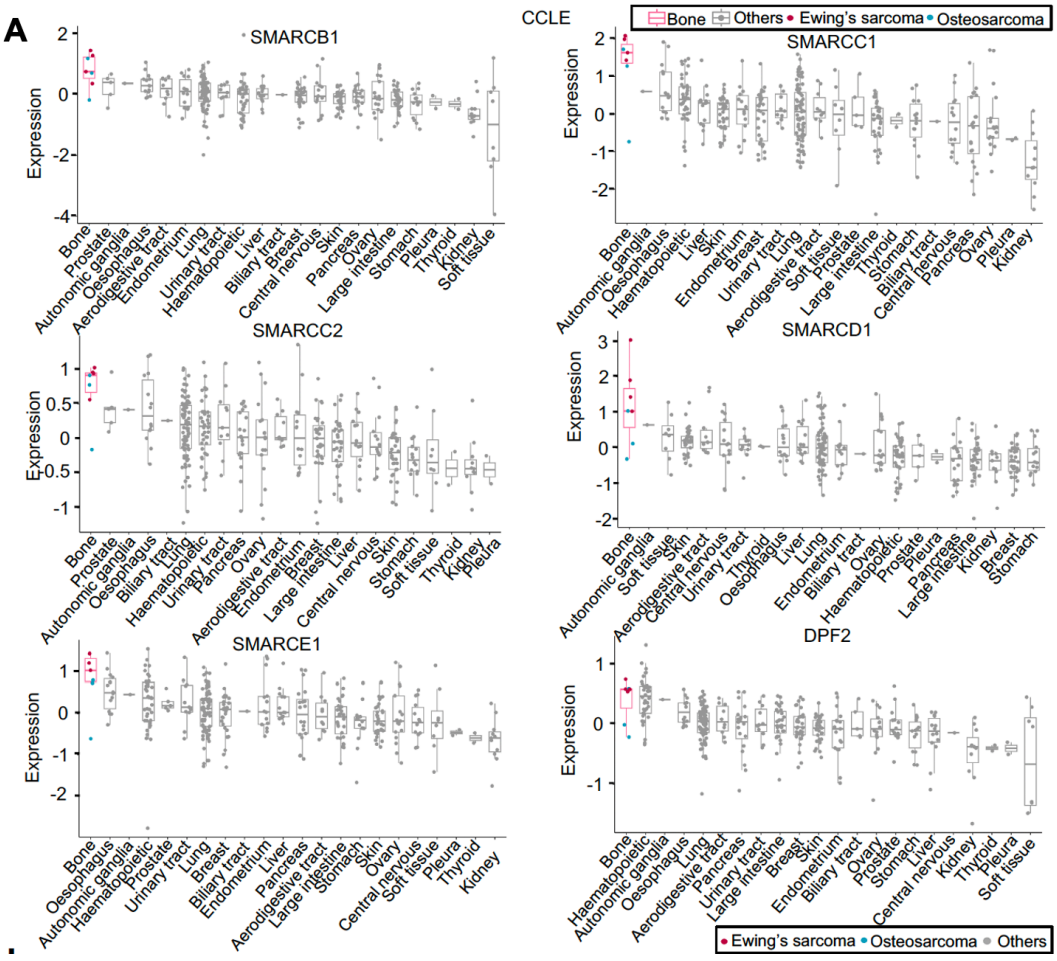

Figure II-7. Protein expression levels of BAF subunits are elevated in Ewing's sarcoma

(**A**) Expression levels of BAF subunits(SMARCB1, SMARCC1, SMARCC2, SMARCD1, SMARCE, DPF2) in different types of cancer obtained from the cancer cell line encyclopedia.

Endogenous ARID1A undergoes phase separation

In order to show that if these nuclear foci indeed represent phase separated condensate, we generated GFP knock-in A673 cell line in which GFP is integrated into genomic loci of ARID1A (Fig. II-8A). I validated that GFP cassette is inserted correctly and that GFP-ARID1A is expressing using various assays (Fig. II-8B-C). Next, I performed high resolution live imaging to identify nuclear condensate formed by endogenous GFP-ARID1A. I found distinctive foci of GFP-ARID1A inside the nucleus of A673 knock-in cell line (Fig. II-8D). I further showed that these foci become diffusive upon 1,6-hexanediol treatment (Fig. II-8E). Next, I performed immunocytochemistry staining BAF subunits while detecting endogenous GFP-ARID1A. We found significant number of BAF subunits co-localized with GFP-ARID1A, indicating enrichment of endogenous BAF subunits inside ARID1A condensate (Fig. II-8F).

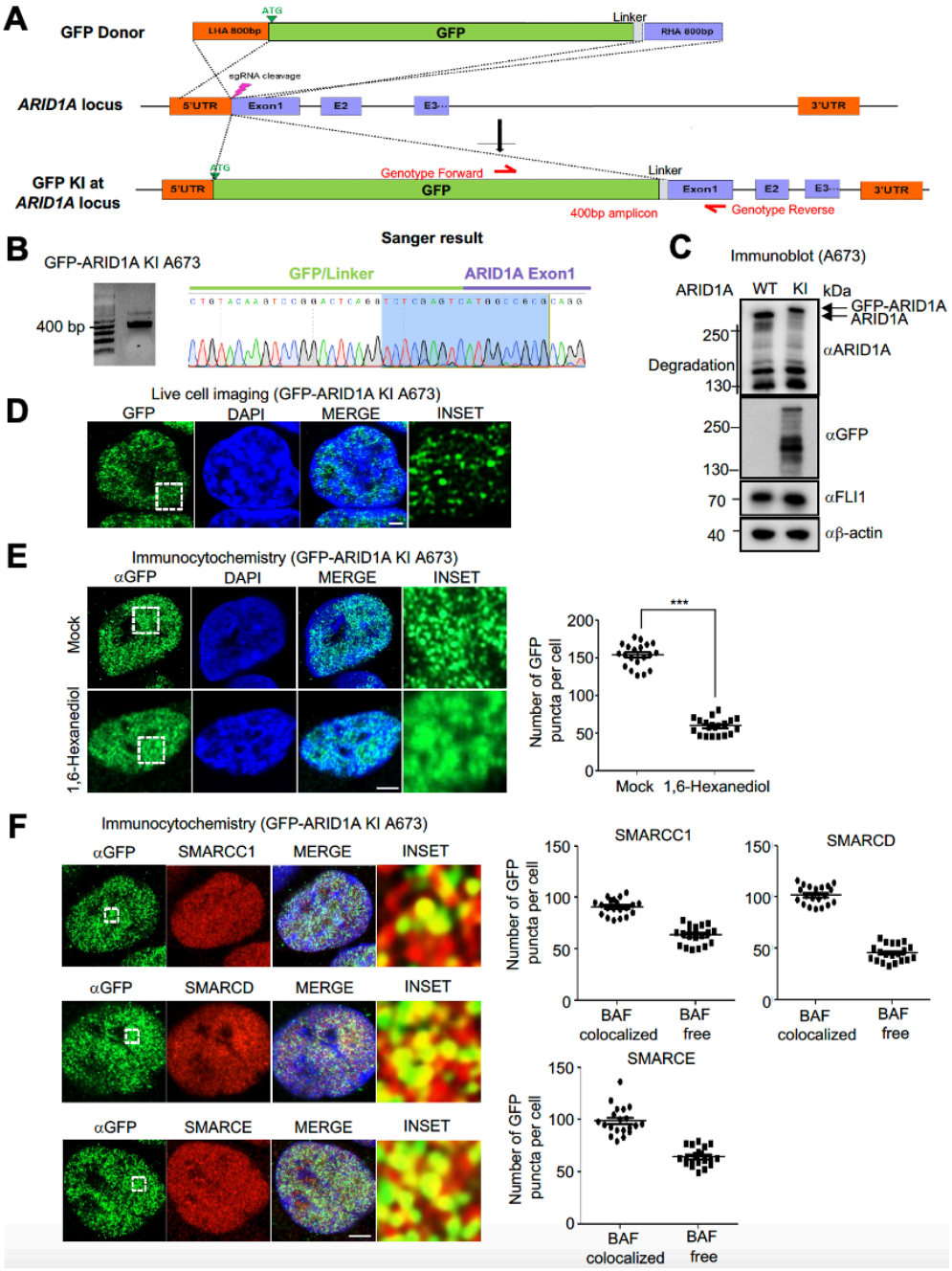
A
GFP Donor
ATG
LHA 800bp
GFP
Linker
RHA 800bp
sgRNA cleavage
ARID1A locus
5'UTR
Exon1
E2
E3···
3'UTR
GFP KI at ARID1A locus
Genotype Forward
Genotype Reverse
400bp amplicon
B
Sanger result
GFP-ARID1A KI A673
400 bp
GFP/Linker
ARID1A Exon1
C
Immunoblot (A673)
ARID1A
WT
KI
kDa
GFP-ARID1A
ARID1A
250
Degradation
130
αARID1A
250
130
αGFP
70
αFLI1
40
αβ-actin
D
Live cell imaging (GFP-ARID1A KI A673)
GFP
DAPI
MERGE
INSET
E
Immunocytochemistry (GFP-ARID1A KI A673)
αGFP
DAPI
MERGE
INSET
Mock
1,6-Hexanediol
***
Number of GFP puncta per cell
200
150
100
50
0
Mock
1,6-Hexanediol
F
Immunocytochemistry (GFP-ARID1A KI A673)
αGFP
SMARCC1
MERGE
INSET
αGFP
SMARCD
MERGE
INSET
αGFP
SMARCE
MERGE
INSET
SMARCC1
SMARCD
SMARCE
Number of GFP puncta per cell
150
100
50
0
BAF colocalized
BAF free

Figure II-8. Endogenous ARID1A undergoes phase separation

(**A**) Scheme for site-specific knock-in (KI) of GFP at the ARID1A locus via CRISPR/cas9 (**B**) Left: DNA electrophoresis gel images of amplicon resulting from PCR at GFP KI site. Right: Sanger sequencing result of PCR amplicon. (**C**) Representative image of A673 WT and KI immunoblot performed using anti-GFP, anti-ARID1A, anti-FLI1, and anti-b-actin antibodies. (**D**) Representative live cell confocal image of GFP-tagged ARID1A knock-in A673 cell line. Scale bar: 5 μm. (**E**) Left: Representative confocal image of GFP tagged ARID1A KI cell treated with or without 6 % Hex for 5 min. Nuclei were stained with DAPI. Right: The quantification on the right shows the number of visible GFP foci per cell. Data are presented as the mean ± SEM. Twenty cells from each group were analyzed; n= 20 biologically independent samples. Scale bar: 5 μm. (**F**) Left: Representative confocal image of GFP tagged ARID1A KI cell immunostained with anti-GFP, anti-ARID1A, anti-SMARCC1, anti-SMARCD and anti-SMARCE antibodies. The quantification on the right shows the number of visible GFP foci per cell that is colocalized with BAF subunit or remain exclusive. Data are presented as the mean ± SEM. Twenty cells from each group were analyzed; n= 20 biologically independent samples. Scale bar: 5 μm.

ARID1A forms spherical condensate inside the nucleus of Ewing's sarcoma cell

Further examination using 3D confocal imaging revealed that these nuclear ARID1A foci detected in Ewing's sarcoma patient and GFP ARID1A knock-in A673 cell resembles spherical morphology (Fig. II-9A,B).

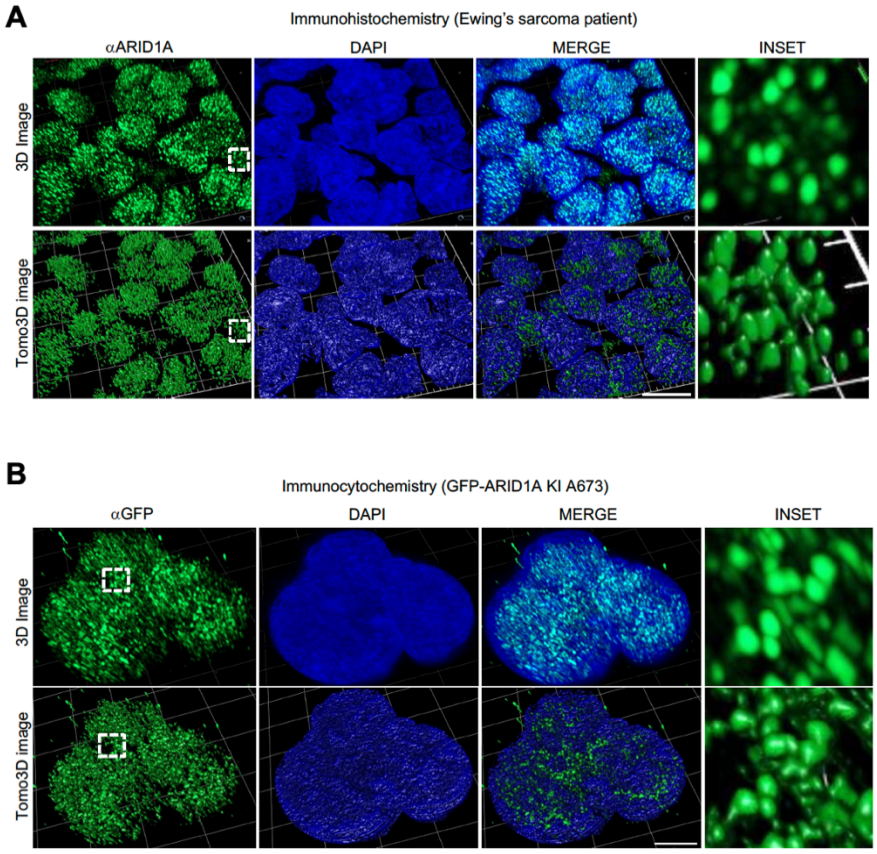

Figure II-9. ARID1A forms spherical condensate inside the nucleus of Ewing's sarcoma cell

(**A**) 3D reconstruction of ARID1A condensates in Ewing's sarcoma patient immunostained with ARID1A antibody. Insets, zoomed-in 3D shapes of a ARID1A condensate on the xy, xz, and yz planes. Scale bar: 10 µm. (**B**) 3D reconstruction of ARID1A condensates in GFP-ARID1A KI A673 cell line immunostained with anti-GFP antibody. Insets, zoomed-in 3D shapes of a GFP-ARID1A condensate on the xy, xz, and yz planes. Scale bar: 5 µm.

Loss of ARID1A LLPS significantly reduces proliferative and invasive property of Ewing′s sarcoma

To further validate that ARID1A LLPS promotes cancer progression and loss of ARID1A condensate functions antagonistically, spheroid formation assay and spheroid invasion assay were performed to measure the ability of cells to form tumor-like solid structure and to evaluate the invasion property of the spheroids, respectively. At post 4 days of spheroid formation, the volume of spheroid formed by LLPS positive cells were larger than that of spheroid formed by negative cells (Fig. II-10A). The spheroid invasion assay showed the similar results (Fig. II-10B). Lastly, in order to assess whether these phenomena are recapitulated in vivo, I performed in vivo xenograft by subcutaneously injecting the cells into nude mice. The tumors generated by ARID1A LLPS positive cells showed notably larger in volume than LLPS negative cells (Fig. II-8C). Immunohistochemistry also confirmed that ARID1A condensate was evident in ARID1A LLPS positive tumors, but was either diffused or undetectable in ARID1A LLPS-negative tumors (Fig. II-8D).

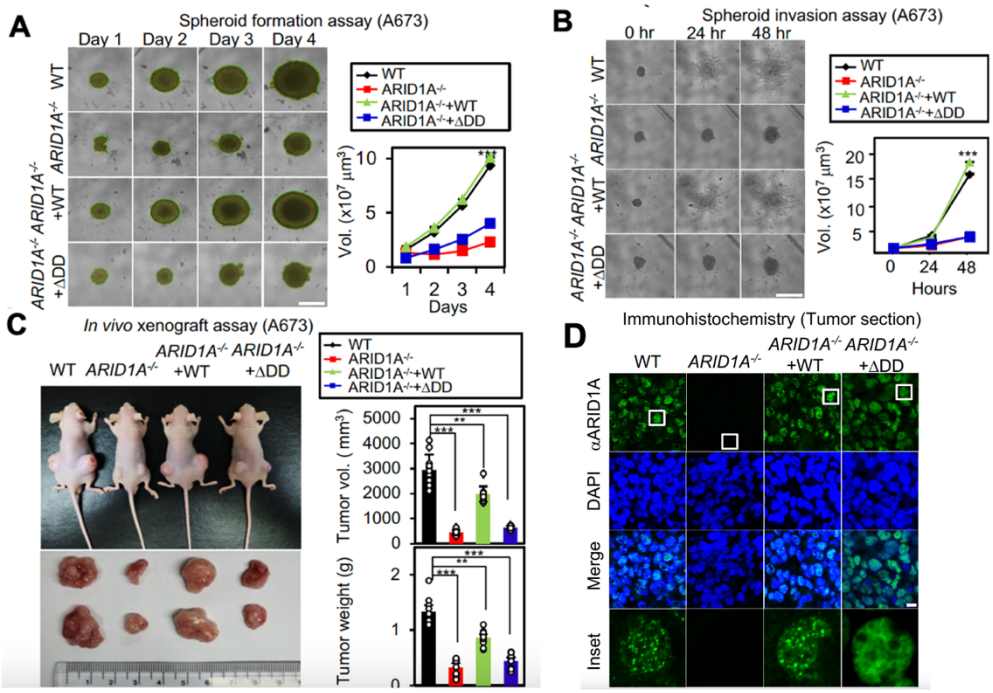

Figure II-10. Loss of ARID1A LLPS significantly reduces proliferative and invasive property of Ewing′s sarcoma

(**A**) Left: Spheroid formation assay performed for four cell lines over 4 days. Right: Quantification of the spheroid formation assay. Bars represents the mean ± SEM; **p < 0.01, ***p < 0.001, NS non-significant. n=10 technical replicates of spheroids. Statistical analysis performed using a two-tailed Wilcoxon signed rank test on Day 4 samples of ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines. Scale bar, 500 µm. (**B**) Left: Spheroid invasion assay conducted on four cell lines over 2 days. Right: Quantification of the spheroid invasion assay. Bars represents the mean ± SEM; **p < 0.01, ***p < 0.001, NS non-significant. n=10 technical replicates of spheroids. Statistical analysis performed using a two-tailed Wilcoxon signed rank test on Day 4 samples of ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines. Scale bar, 500 µm. (**C**) Left: In vivo xenograft assay performed using four cell lines. Nude mice and extracted tumors are shown. Top right: Quantification of the volume of the extracted tumors. Bottom right: Quantification of the weight of the extracted tumors. Bars represents the mean±SEM; **p < 0.01, ***p < 0.001, NS non-significant. n=10 tumor extracts. Statistical analysis performed using a two-tailed

Wilcoxon signed rank test. ARID1A-/-, ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines were individually compared to WT. (**D**) Representative immunohistochemistry images of extracted tumors formed by the four cell lines. Immunostaining was performed using an anti-ARID1A antibody. Scale bars: 10 μm. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. Source data are provided as a Source Data file.

Aromatic residue mediated phase separation of ARID1A is crucial for oncogenic potential of Ewing's sarcoma

Additionally, I conducted spheroid-related assays using cell lines rescued with the PrLD(Y/S) mutant or a mutant with a more restricted number of substitutions, PrLD(Y33/S). As in the case of ΔDD rescued mutant, this substitution-based phase separation failed to rescue proliferative and invasive phenotypes (Fig. II-11A,B). Finally, I conducted spheroid-related experiments utilizing the SK-N-MC cell line, which is another Ewing's sarcoma cell line. Similar to our observations with A673 cell line, I discovered that a decrease in the phase separation of ARID1A resulting from the knockdown led to a reduction in the oncogenic capabilities of the SK-N-MC cells as well (Fig. II-11C,D).

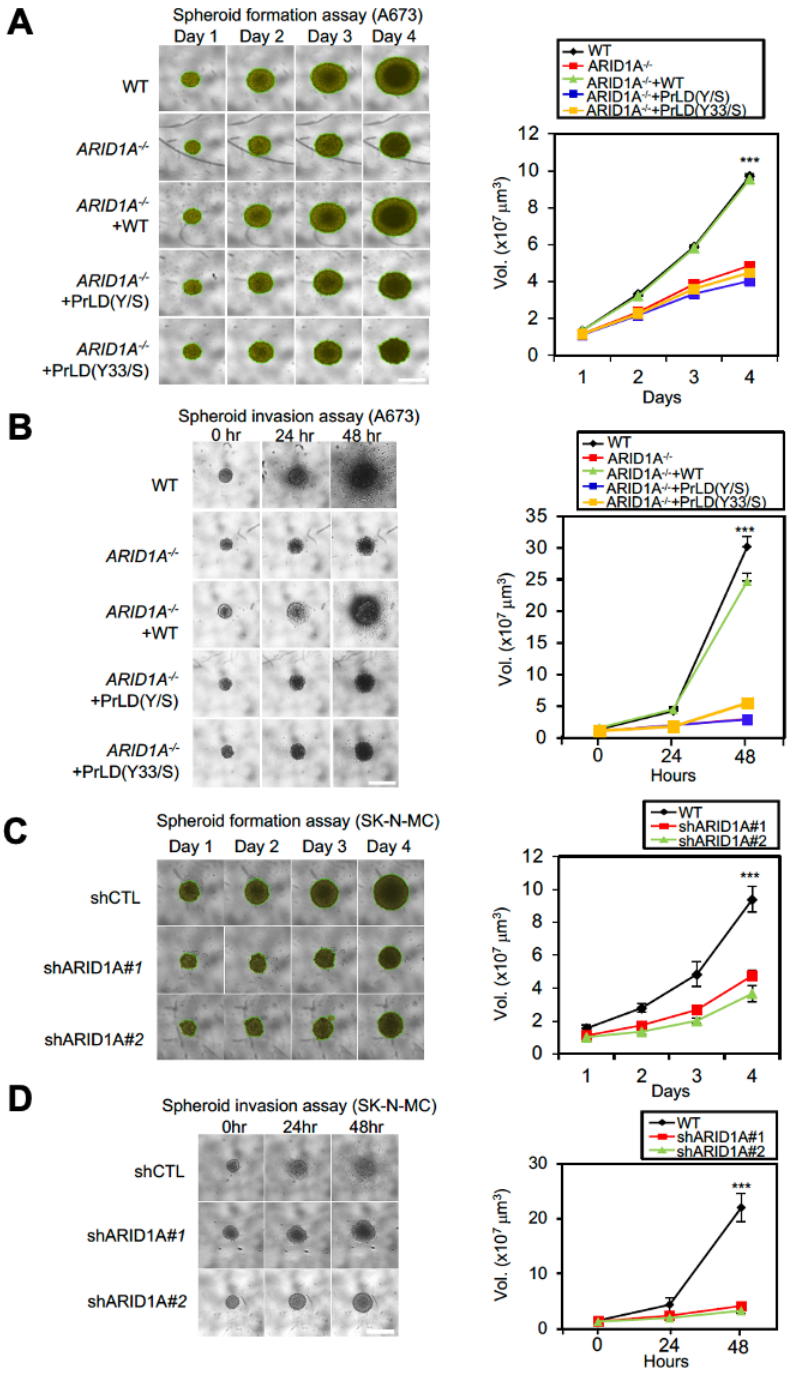
A
Spheroid formation assay (A673)
Day 1
Day 2
Day 3
Day 4
WT
ARID1A-/-
ARID1A-/- +WT
ARID1A-/- +PrLD(Y/S)
ARID1A-/- +PrLD(Y33/S)
WT
ARID1A-/-
ARID1A-/-+WT
ARID1A-/-+PrLD(Y/S)
ARID1A-/-+PrLD(Y33/S)
***
Vol. (x10^7 µm^3)
Days
B
Spheroid invasion assay (A673)
0 hr
24 hr
48 hr
WT
ARID1A-/-
ARID1A-/- +WT
ARID1A-/- +PrLD(Y/S)
ARID1A-/- +PrLD(Y33/S)
WT
ARID1A-/-
ARID1A-/-+WT
ARID1A-/-+PrLD(Y/S)
ARID1A-/-+PrLD(Y33/S)
***
Vol. (x10^7 µm^3)
Hours
C
Spheroid formation assay (SK-N-MC)
Day 1
Day 2
Day 3
Day 4
shCTL
shARID1A#1
shARID1A#2
WT
shARID1A#1
shARID1A#2
***
Vol. (x10^7 µm^3)
Days
D
Spheroid invasion assay (SK-N-MC)
0hr
24hr
48hr
shCTL
shARID1A#1
shARID1A#2
WT
shARID1A#1
shARID1A#2
***
Vol. (x10^7 µm^3)
Hours

Figure II-11. Aromatic residue mediated phase separation of ARID1A is crucial for oncogenic potential of Ewing's sarcoma

(**A)** Spheroid formation assay performed for five indicated A673 cell lines over 4 days. Right: Quantification of the spheroid formation assay. Bars represents the mean ± SEM; ***P<0.001. Statistical analysis performed using a two-tailed t-test. Scale bar, 500 μm. (**B)** Spheroid invasion assay conducted on five indicated A673 cell lines over 2 days. Right: Quantification of the spheroid invasion assay. Bars represents the mean ± SEM; ***P<0.001. Statistical analysis performed using a two-tailed t-test. Scale bar, 500 μm. (**C)** Spheroid formation assay performed for three indicated SK-N-MC cell lines over 4 days. Right: Quantification of the spheroid formation assay. Bars represents the mean ± SEM; ***P<0.001. Statistical analysis performed using a two-tailed t-test. Scale bar, 500 μm. (**D)** Spheroid invasion assay conducted on three indicated SK-N-MC cell lines over 2 days. Right: Quantification of the spheroid invasion assay. Bars represents the mean ± SEM; ***P<0.001. Statistical analysis performed using a two-tailed t-test. Scale bar, 500 μm.

II-4. Discussion

In this chapter, I explored the role of ARID1A's phase separation (LLPS) properties in the oncogenesis of Ewing's sarcoma and demonstrated its critical function in tumor progression. By systematically characterizing ARID1A's prion-like domains (PrLDs) and their ability to mediate phase separation, I revealed that ARID1A forms liquid-like nuclear condensates both in vitro and in cellulo. These nuclear condensates serve as functional hubs for the recruitment of BAF chromatin remodeling complex subunits, which are critical for transcriptional regulation and cancer cell proliferation.

The identification of PrLD1 and PrLD2 within ARID1A as key domains driving LLPS is consistent with previous findings on phase separation-prone proteins, which contain intrinsically disordered regions (IDRs) rich in aromatic residues (Hyman et al., 2014; Banani et al., 2017). My experiments further confirmed that tyrosine residues within the PrLDs act as "stickers" to drive LLPS, as substitution of tyrosine with serine (PrLD(Y/S)) completely abolished ARID1A condensate formation. This observation aligns with the "stickers-and-spacers" framework, where hydrophobic and π-π interactions mediated by aromatic residues are essential for phase separation (Patel et al., 2015). Furthermore, the disruption of LLPS using either ΔDD mutants or PrLD(Y/S) substitutions not only impaired condensate formation but also severely affected the tumorigenic potential of ARID1A in Ewing's sarcoma cells.

One of the most significant findings of this study is the dual role of ARID1A's PrLDs and Pfam domain (Fig. II-12). While PrLDs are indispensable for driving phase separation,

the Pfam domain is critical for incorporating BAF subunits into ARID1A condensates. This modular organization suggests a sophisticated mechanism by which ARID1A forms nuclear hubs to facilitate chromatin remodeling and transcriptional regulation. My findings demonstrated that LLPS-deficient ARID1A mutants (ΔDD or PrLD(Y/S)) fail to recruit key BAF subunits such as SMARCB1 and SMARCD1 into nuclear condensates. The inability to form functional condensates highlights the importance of ARID1A's phase separation in maintaining chromatin accessibility and supporting oncogenic transcriptional programs.

The elevated expression of ARID1A and its nuclear condensates in Ewing's sarcoma tissues, compared to normal bone tissues, underscores its potential role as a tumor-promoting factor. Using CRISPR-Cas9 knockout and rescue experiments, I demonstrated that LLPS-deficient ARID1A mutants fail to restore tumorigenic properties, including cell proliferation, migration, invasion, and in vivo tumor growth. These findings strongly support a model in which ARID1A phase separation promotes the assembly of oncogenic transcriptional hubs, thereby driving Ewing's sarcoma progression. The in vivo xenograft experiments further validated the importance of ARID1A's LLPS in tumor progression. Tumors derived from LLPS-deficient ARID1A cells were significantly smaller and lacked the distinct nuclear condensates observed in wild-type ARID1A tumors. These results emphasize that ARID1A-mediated phase separation is not merely a passive phenomenon but a critical mechanism for oncogenic transformation. Furthermore, my observation of ARID1A's co-localization with endogenous BAF subunits suggests that ARID1A LLPS may enhance chromatin accessibility and facilitate transcription of EWS/FLI1 target genes,

consistent with previous reports on phase-separated transcriptional condensates (Zuo et al., 2021; Wanior et al., 2021). Targeting phase separation through small molecules or disrupting aromatic residues within PrLDs could be a novel strategy to impair ARID1A's function in Ewing's sarcoma. My findings pave the way for future studies to develop strategies that selectively inhibit ARID1A LLPS without affecting its other chromatin remodeling functions. In conclusion, this study establishes a clear link between ARID1A's phase separation properties and its oncogenic function in Ewing's sarcoma. By demonstrating that ARID1A forms liquid-like nuclear condensates, recruits BAF subunits, and drives tumor progression, I provide critical insights into the mechanisms of ARID1A-mediated oncogenesis. These findings underscore the importance of LLPS in cancer biology and highlight ARID1A as a promising therapeutic target for Ewing's sarcoma.

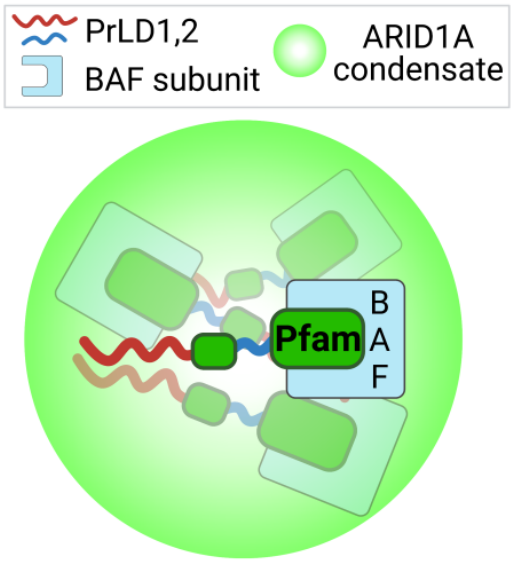

Figure II-12. Schematic model of prion-like domain mediated phase separation of ARID1A and BAF complex recruitment

ARID1A undergoes liquid-like phase separation through prion-like domain. Upon condensate formation, ARID1A recruits BAF complex by PFAM homology domain mediated protein binding. This chromatin remodeling condensate is enriched in Ewing's sarcoma patient tissue and cancer cell.

II-5. Materials and Methods

Reagent

The following commercially available antibodies were used: anti-ARID1A (ab182560), anti-SMARCC1 (ab172638), anti-FLI1 (ab133485), anti-SMARCE (ab70540), anti-H3K27ac (ab4729) (Abcam); anti-Flag (F3165), and anti-β-actin (A1978), anti-ARID1A (AMAb91192) (Sigma-Aldrich); anti-GFP (sc-9996), anti-SMARCB (sc166-165), anti-SMARCD (sc-135843) (Santa Cruz Biotechnology); anti-SMARCA2 (26613-1-AP) (Proteintech). Following commercially available fluorescent-labeled secondary antibodies were used: Alexa Fluor 488 donkey anti-rabbit IgG (A21206) and Alexa Fluor 594 donkey anti-mouse IgG (A21203) (Invitrogen). We used antibodies recommended by the manufacturer for the species and application.

Cell culture and transfection

The A673, SK-N-MC cell lines were purchased from the Korean Cell Line Bank. HEK293T cell was obtained from ATCC. TC-106 was kindly provided by Thomas G.P. Grünewald. Cells were tested for mycoplasma contamination and were routinely treated with BM-cyclin. A673 cells were cultured with RPMI1640 medium containing 10 % fetal bovine serum (FBS) and antibiotics in a humidified incubator at 5 % CO_2 and 37 °C. HEK293T cell line was grown in Dulbecco's modified Eagle's medium containing 10% FBS and antibiotics. Transfection was performed using PEI (Sigma-Aldrich). The cell lines were STR authenticated by STR profiling.

Generation of genetically modified cell line

Ewing's sarcoma cell line A673 was genetically modified using the CRISPR-CAS9 system. A guide RNA (sequence: GGGGCCTGGAGCCCTACGCG) targeting the first exon of ARID1A was cloned into a px330 vector. Transfected cells were then cultured in a 99-well plate at one cell per well. Single-cell colonies were grown, and genotyping was performed. Cells with a complete loss of ARID1A protein expression, as confirmed via western blot, were chosen as ARID1A KO A673. For ARID1A rescue A673 KO cells, the pLenti-puro-ARID1A (plasmid#39478, Addgene) plasmid was transduced into ARID1A KO A673 cells using lentiviral particles. The transduced cells were selected using puromycin. Selected cells were then plated in a 99-well plate at one cell per well. Single-cell colonies were genotyped several times until the ARID1A levels of rescue cells were comparable to those of their wild-type counterparts. For the generation of ARID1A PrLD(Y/S) and ARID1A ΔDD-rescued A673 ARID1A KO cells, pLenti-puro-ARID1A PrLD(Y/S) and pLenti-puro-ARID1A ΔDD were each transduced into KO cells via lentiviral particles. The transduced cells were selected using puromycin and added into a 99-well plate at one cell per well. Single-cell colonies were genotyped several times until sufficient ARID1A PrLD(Y/S) or ARID1A ΔDD expression was observed.

Establishment of EGFP-tagged ARID1A knock-in A673 cell line

For transfection, A673 cells were rinsed with DPBS (Gibco) and detached with 0.25 % Trypsin-EDTA (LS015-10, WELGEN). After detachment of A673 cells, Trypsin-EDTA

was inactivated by adding RPMI 1640 (HyClon) with 20% FBS. A673 cells were washed with DPBS for two times. Washed A673 cells were counted and resuspended by Resuspension Buffer R in Neon Transfection System μl 100 Kit (MPK10096, Invitrogen) to concentration of 1 x 10 7 cells in 1 ml. 3 μg of Cas9 vector, 1 μg of gRNA vector targeting ARID1A and 2 μg of EGPF-ARID1A knock-in donor plasmid were added to 100 μl of resuspended A673 cells. Plasmid and A673 cell mixture were electroporated by Neon Transfection System with 1650 volts, 10 ms pulse length and 3 pulses condition following the supplier's instructions. After 96 hours from transfection, EGFP + A673 cells were sorted with Flow Cytometer (SH800S, SONY). Sorted cells were seeded into 96 well cell culture plate for single cell isolation.

Immunofluorescence staining and live-cell imaging

Cells were seeded onto a confocal dish and fixed with 1 % formaldehyde in PBS for 15 min. Fixed cells were permeabilized with 0.5 % Triton X-100 in PBS (PBS-T) at room temperature for 10 min. Blocking was performed with 3 % bovine serum in PBS-T for 1 h. For staining, cells were incubated with primary antibodies at room temperature for 2 h, followed by incubation with fluorescently-labeled secondary antibodies and DAPI for 1 h. VECTASHIELD Antifade Mounting Medium was used for mounting, and cells were visualized under a confocal microscope (Zeiss, LSM700). For live-cell imaging, 293T cell was transfected with the GFP-ARID1A construct a day before and imaged. For hexanediol treatment, 293T cells were seeded in a confocal dish, and the nuclei were stained with

Hoechst 33342. 6 % 1,6-hexanediol was directly added to cells under a microscope, and images were continuously acquired.

FRAP

FRAP was performed using a Zeiss LSM700 microscope with a 594 nm laser. Bleaching was performed over at r_{bleach} = 1 μm using 100% laser power. Images were acquired every second. For quantification, multiple sets of FRAP experiments were performed on independent GFP-ARID1A condensates. The same interval of image acquisition was applied for each set of experiments. The fluorescence intensity was acquired by Zen program (Zeiss). The relative fluorescence intensity was calculated using the initial fluorescence intensity as a reference point. The mean and standard deviation were subsequently calculated.

Protein expression and purification

Prokaryotic plasmid containing His-GFP-tagged recombinant protein (PrLD1, ARID, PrLD2, Pfam) or His-mCh-EWS/FLI1 were transformed into M15(pREP) cells. After induction with isopropyl-β-D-thiogalactoside, the bacterial pellet was lysed in buffer (50 mM Tris-HCl [pH 7.5], 200 mM NaCl, 1mM dithiothreitol (DTT), and 1 % Triton X-100). The lysate was then sonicated and centrifuged. The supernatant was incubated overnight with Talon beads. The bead-protein complex was washed with lysis buffer three times and eluted using elution buffer (50 mM Tris-HCl [pH 7.5], 200 mM imidazole, and 1 mM DTT). Eluted protein was assessed via Coomassie staining, and a single protein band was

confirmed. For purification of GFP-ARID1A, GFP-ARID1A PrLD(Y/S) and GFP-ARID1A ΔDD, HEK293T cells were used for transfection. GFP-tagged proteins in cell lysates were enriched using GFP-Trap magnetic bead (Chromotek). The protein-bead conjugate was washed stringently for 5 times using high salt wash buffer and eluted using acidic elution buffer with a composition suggested by manufacturer. The eluted protein was immediately neutralized using neutralization buffer. The size and purity of the eluted proteins were analyzed by SDS-PAGE and Coomassie blue staining.

Droplet assay

In vitro droplet formation assay was performed as previously described (Sabari et al., 2018). Briefly, recombinant proteins were concentrated and desalted using Amicon Ultra centrifugal filters (30K MWCO, Millipore). Eluted proteins were diluted to varying concentrations in phase separation buffer containing 50 mM Tris-HCl [pH 7.5], 10% glycerol, 10 % PEG8000, and 1 mM DTT. The protein solution was loaded onto a confocal dish and imaged using a Zeiss LSM700 microscope. Droplet size was quantified by measuring the circumference of droplets using a Zen image viewer.

Characterization of saturation concentrations of ARID1A variants

To measure the saturation concentrations of various ARID1A variants, we used HEK 293 cells transfected with corresponding GFP-tagged ARID1A variants 24 hours prior to imaging. The nuclear boundaries of cells were detected manually. For individual cells, average fluorescence intensities within nucleus were measured using image processing and

analysis program NIH Image J (http://rsb.info.nih.gov/ij) and the presence of condensates of ARID1A variants was checked.

Corelet colocalization experiments

ARID1A KO A673 cells stably expressing PrLD1 or PrLD1-pfam Corelet constructs were seeded onto a confocal dish and transfected with plasmids of individual BAF complex components 24-48 hours prior to imaging. Media was exchanged 8 hours after transfection. To prevent any unwanted pre-activation of Corelets, the mCherry channel was used to locate cells expressing proper levels of Corelet constructs. Once located, blue light activation was performed simultaneously with data acquisition through sequential imaging of GFP and mCherry channels. The initial images represented conditions prior to blue light activation. Dual color imaging was performed every 3 seconds for 15 minutes. Line profiles were obtained using using image processing and analysis program NIH Image J (http://rsb.info.nih.gov/ij) and normalized with the average intensity of the dilute phase. Moving averages were applied when necessary.

ARID1A protein expression

ARID1A level expression levels were acquired from the raw data of a previously published proteome analysis (Nusinow et al., 2020). Briefly, 375 cell lines were grouped by cancer type, and ARID1A expression levels were determined for each type. The resulting

individual, average and standard deviation of protein expression value of ARID1A is presented as a box plot.

Wound healing assay

Wound healing scratching motility assay was performed in WT, ARID1A$^{-/-}$, ARID1A$^{-/-}$+WT and ARID1A$^{-/-}$+ ΔDD A673 cell lines. Cells were seeded in 12-well culture plates and cultured until they reached confluence. Cells were scratched with a 200 µl micro-pipette tip and incubated at 37ºC for 48 hr. Photomicrographs of the closed gap were captured at 0 hr, 24 hr, and 48 hr using JuLI Stage Real-time live cell imaging system (NanoEntek; http://www.NanoEntek.com). Migration distance of the cells was quantified by distance of gap using wound healing analysis package provided by JuLI Stage software. Values are expressed as means ± s.e.m.

Spheroid formation and spheroid invasion assays

Spheroid formation assay was performed in WT, ARID1A$^{-/-}$, ARID1A$^{-/-}$+WT and ARID1A$^{-/-}$+ ΔDD A673 cell lines. 2000 counted cells from each cell line were pipetted into Ultra-Low Attachment 96 well plate (Corning Costar). Subsequently, the plate was centrifuged at low speed for 10 min and sequestered cells were visualized under microscope. For spheroid formation assay, cells were incubated at 37 ºC for 24 hr before imaging. Photomicrographs of the spheroid growth were captured at each day until day 4 using JuLI Stage Real-time live cell imaging system (NanoEntek; http://www.NanoEntek.com). Spheroid volume was quantified by automated spheroid analysis package provided by JuLI Stage software. For

spheroid invasion assay, next day, matrigel (BD Biosciences, Bedford, MA) was added directly to the media containing the spheroid. Matrigel was solidified inside the incubator. Photomicrographs of the spheroid growth were captured at 0 hr, 2 4hr, and 48 hr using JuLI Stage Real-time live cell imaging system (NanoEntek). Spheroid volume was quantified by automated spheroid analysis package provided by JuLI Stage software.

Immunohistochemistry

To detect ARID1A expression in human tissue samples, paraffin-embedded human normal bone tissue (US Biomax, BO244g) and Ewing's sarcoma tissue (US Biomax, T263, T264a) were deparaffinized, hydrated, and heated in retrieval buffer (10 mM sodium citrate [pH 6.0]) over 10 min for antigen retrieval, and then incubated with ARID1A antibodies (Abcam, ab182560, 1:200). Subsequently, tissues were incubated with fluorescently-labeled secondary antibodies and DAPI for 1 h. VECTASHIELD Antifade Mounting Medium was used for mounting, and tissues were visualized under a confocal microscope (Zeiss, LSM700). Patient age, gender, and diagnosis information are available on the company's website: https://www.biomax.us/tissue-arrays/Bone_Cartilage/T264a; https://www.biomax.us/tissue-arrays/Bone_Cartilage/BO244g.

Xenograft

For tumor formation *in vivo*, 10^7 cells with equal volume of matrigel (BD Biosciences, Bedford, MA) were injected subcutaneously at the left and right flank bilaterally into 6-week-old athymic *nu/nu* female mice (Charles River). Tumors were measured weekly, and

the experiment was terminated at week 5. Total of 10 tumors from 5 mice were excised for each cell line and weighed. Statistical differences in tumor weights were determined by Wilcoxon signed rank test using the Graphpad prism. These experiments were carried out with the approval of the Institutional Animal Care and Use Committee (IACUC) of Seoul National University. Tumor sections were stained and imaged as described above. Image quantification was performed using image processing and analysis program NIH Image J (http://rsb.info.nih.gov/ij).

CHAPTER III

ARID1A forms chromatin remodeling hub to alter chromatin architecture and promote EWS/FLI1 target gene transcription

III-1. Summary

ARID1A is a chromatin remodeling factor that plays a crucial role in regulating gene expression and chromatin accessibility. Although ARID1A is frequently mutated in cancers, the underlying mechanisms by which it promotes oncogenesis remain unclear. Recent studies have highlighted the importance of liquid-liquid phase separation (LLPS) in organizing nuclear architecture and regulating transcriptional programs. LLPS facilitates the formation of biomolecular condensates that compartmentalize regulatory activities, yet its role in chromatin remodeling and oncogenesis is not fully understood. Here, I identify ARID1A as a critical driver of LLPS in Ewing's sarcoma. The prion-like domains (PrLDs) of ARID1A mediate LLPS, forming nuclear condensates that localize to chromatin regulatory elements. ARID1A LLPS significantly enhances chromatin accessibility at EWS/FLI1-bound cis-regulatory elements (cREs), promoting transcriptional activation of oncogenic target genes. I have shown that ARID1A directly interacts with EWS/FLI1 through its PrLDs, enabling the formation of co-condensates that compartmentalize chromatin remodeling machinery. Importantly, disruption of ARID1A LLPS impairs the long-range chromatin interactions necessary for oncogene activation. Finally, I demonstrate that ARID1A LLPS is crucial for the oncogenic potential of Ewing's sarcoma cells. By driving chromatin remodeling and transcriptional activation of cancer-related genes, ARID1A LLPS facilitates tumor progression both in vitro and in vivo. These findings establish ARID1A LLPS as a key mechanism underlying oncogenesis in Ewing's sarcoma and highlight it as a potential therapeutic target.

III-2. Introduction

LLPS is a biophysical process by which proteins and nucleic acids organize into membraneless condensates, and it has emerged as a critical mechanism underlying nuclear organization and gene regulation (Banani et al., 2017). Despite extensive studies on LLPS, its functional role in transcriptional regulation and oncogenesis remains an active area of research.

One of the important contexts in which LLPS has been implicated is Ewing's sarcoma, an aggressive bone and soft tissue cancer driven by the EWS/FLI1 fusion protein. EWS/FLI1 acts as an oncogenic transcription factor that reprograms chromatin and induces the expression of oncogenic gene networks (Riggi et al., 2021). As a chromatin remodeler, ARID1A is well-positioned to influence the activity of EWS/FLI1, yet the precise mechanisms remain unclear. My recent work suggests that ARID1A may function beyond its canonical chromatin remodeling activity by undergoing LLPS to form nuclear condensates that regulate transcription and chromatin architecture.

In this study, I explore the role of ARID1A in the oncogenesis of Ewing's sarcoma through its ability to undergo LLPS. Using genome-wide transcriptional and chromatin accessibility profiling, I show that ARID1A LLPS drives the activation of EWS/FLI1-bound enhancers and oncogenic gene expression. By forming nuclear condensates, ARID1A functions as a coactivator for EWS/FLI1, promoting long-range chromatin interactions and remodeling chromatin to enhance transcription. Since ARID1A is frequently mutated in cancers and

EWS/FLI1 is a hallmark oncogene of Ewing's sarcoma, understanding how ARID1A LLPS contributes to transcriptional regulation offers insights into potential therapeutic strategies for treating Ewing's sarcoma and related cancers.

III-3. Results

ARID1A LLPS leads to change in gene expression and alters chromatin accessibility

I next sought to identify the mechanism by which ARID1A LLPS affects oncogenic potential in Ewing's sarcoma cell lines. Considering the role of ARID1A as a chromatin remodeler, I hypothesized that ARID1A LLPS may activate cancer-related genes by regulating chromatin structure. To identify target genes and *cis*-regulatory element (referred to hereinafter as cREs) affected by ARID1A LLPS, I performed RNA-seq and ATAC-seq on ARID1A LLPS-positive and negative cell lines. To exclude any genetic background variation occurring while cell line generation, I used five different colonies of *ARID1A*$^{-/-}$ +WT and *ARID1A*$^{-/-}$+ΔDD for genome-wide studies (Fig. III-1A,B). As a result, I found ARID1A LLPS-dependent 1,271 differentially expressed genes (DEGs) and 9,686 dysregulated cREs (Fig. III-1C,D,E).

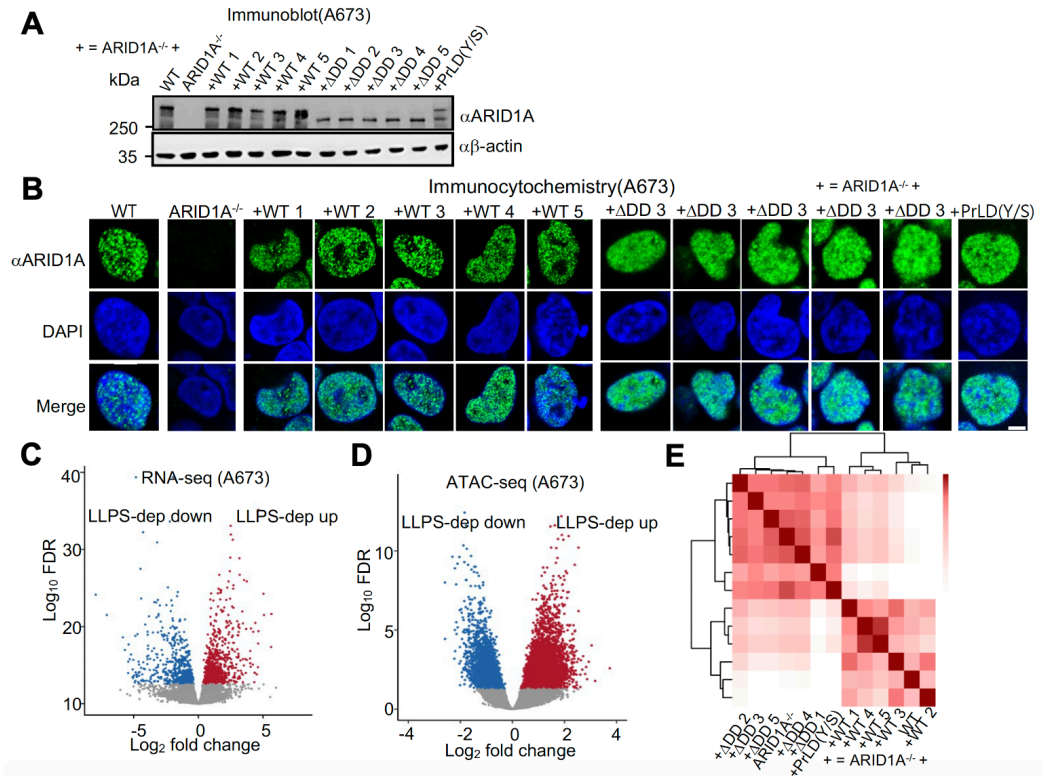

Figure III-1. ARID1A LLPS leads to change in gene expression and alters chromatin accessibility

(**A**) Representative immunoblot image showing ARID1A protein expression in WT, KO, WT rescue replicates, MT rescue replicates used in sequencing. (**B**) Immunocytochemstry data showing anti- ARID1A staining of cell lines (WT, ARID1A-/-, ARID1A-/-+WT, ARID1A-/-+DDD, and ARID1A-/- +PrLD(Y/S)). (**C**) A volcano plot illustrating ARID1A LLPS-dependent DEGs. The LLPS-dependent upregulated genes are colored red, and the downregulated genes are colored blue. (**D**) A volcano plot illustrating ARID1A LLPS-dependently dysregulated cREs. The LLPS-dependent upregulated cREs are colored red, and the downregulated cREs are colored blue. (**E**) A hierarchically clustered heatmap illustrating correlations of the dysregulated cREs among the ATAC-seq samples. The colors indicate Pearson correlation coefficient (PCC) values of the normalized read counts.

ARID1A LLPS upregulates EWS/FLI1 oncogenic transcription program

I performed hierarchical clustering on differentially expressed gene and identified two gene clusters (Fig. III-2A). Notably, a higher number of DEGs were altered with the increase of gene expression dependent on ARID1A LLPS (Fig. III-2A) seemingly associated with the ability of ARID1A LLPS in the recruitment of numerous BAF subunits. Furthermore, Gene ontology analysis revealed that ARID1A LLPS-dependent upregulated genes are significantly enriched by cancer-related terms such as migration, adhesion, and angiogenesis, consistent with the previous experimental results (Fig. III-2B). Moreover, gene set enrichment analysis revealed that genes that are upregulated in Ewing's sarcoma are highly enriched in gene cluster that is positively regulated by ARID1A LLPS (Fig. III-2C).

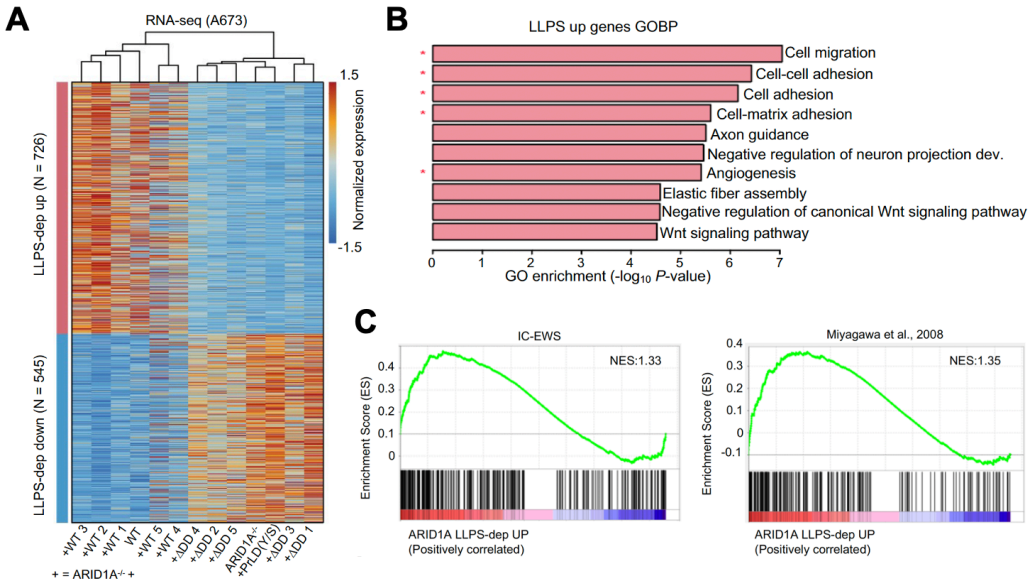

Figure III-2. ARID1A LLPS upregulates EWS/FLI1 oncogenic transcription program

(**A**) A heatmap illustrating expression of DEGs (FDR < 0.05) obtained from the RNA-seq results of WT, *ARID1A*$^{-/-}$, five *ARID1A*$^{-/-}$+WT, five *ARID1A*$^{-/-}$+ΔDD, and *ARID1A*$^{-/-}$ +PrLD(Y/S) A673 cell lines. The colors indicate normalized gene expression. The dendrogram above the heatmap indicates the hierarchical clustering result of the samples. (B) Top 10 enriched gene ontologies in ARID1A LLPS-dependent upregulated DEGs. The cancer-related terms are marked with an asterisk. (C) Gene Set Enrichment Analysis (GSEA) for the genes correlated with the gene expression in ARID1A LLPS-dep UP cluster. NES, normalized enrichment score.

ARID1A shares target gene with EWS/ETV

In order to check whether EWS/FLI and ARID1A LLPS regulates same gene cohort, I compared genes that are upregulated by EWS/FLI1 with genes that are positively regulated by ARID1A LLPS (Fig. III-3A). Gene sets regulated by ARID1A LLPS significantly overlap with genes previously found to be regulated by EWS/ETV (Fig. III-3A). I performed quantitative real time PCR to validate that these genes are indeed positively regulated by ARID1A LLPS(Fig. III-3B)

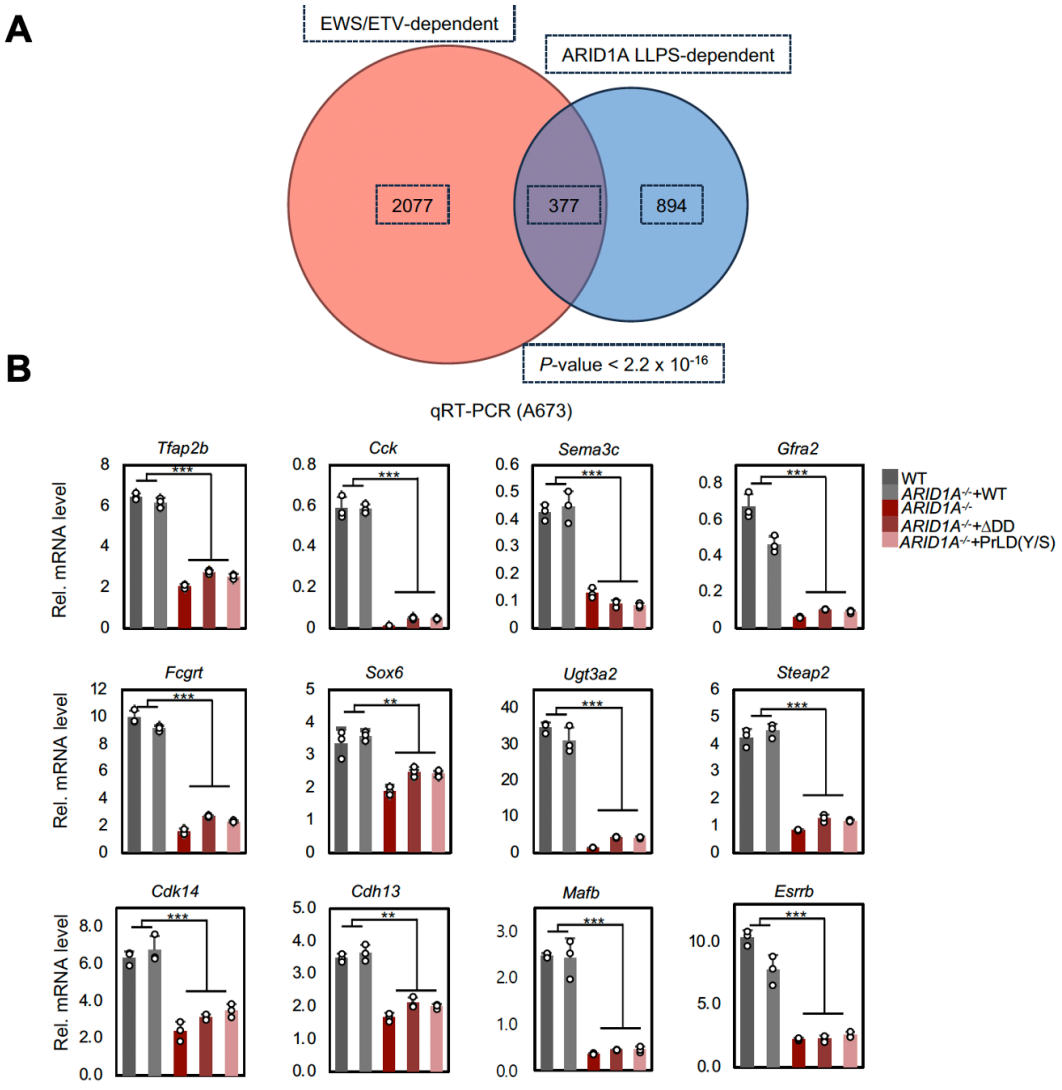

Figure III-3. ARID1A shares target gene with EWS/ETV

(**A**) Venn diagram illustrating overlap between EWSR1-ETS dependent DEGs (Orth et al., Limma-trend FDR < 0.001) and ARID1A LLPS-dependent DEGs. The statistical significances of overlap between the two sets were calculated using two-sided hypergeometric test. (**A**) Relative mRNA expression of EWS/FL1I target genes with quantitative real time-PCR (qRT-PCR).

ARID1A LLPS increases chromatin opening at EWS/FLI1 bound cREs

I then performed ATAC-seq experiment using ARID1A LLPS positive and LLPS negative cell lines. Notably, a higher number of cREs were altered with the increase of chromatin accessibility dependent on ARID1A LLPS (Fig. III-4A), seemingly associated with the ability of ARID1A LLPS in the chromatin opening. To dissect underlying mechanism of ARID1A LLPS-dependent cREs dysregulation and its influence on transcription, I sought to identify upstream regulator candidates. To this end, I performed TF motif enrichment analysis for the upregulated cREs. Surprisingly, I revealed that both FLI1 and EWS/FLI1 motif sequences were remarkably enriched in the ARID1A LLPS-upregulated cREs (Fig. III-4B) (Gangwal et al., 2008; Guillon et al., 2009). As EWS/FLI1 is a major driver oncogene in Ewing's sarcoma, I further examined whether ARID1A LLPS-dependent upregulated cREs are co-localized with EWS/FLI1 nucleation sites in Ewing's sarcoma. As a result of EWS/FLI1 ChIP-seq analysis, I found that EWS/FLI1 is directly bound at ARID1A LLPS-dependent upregulated cREs (FIG. III-4C).

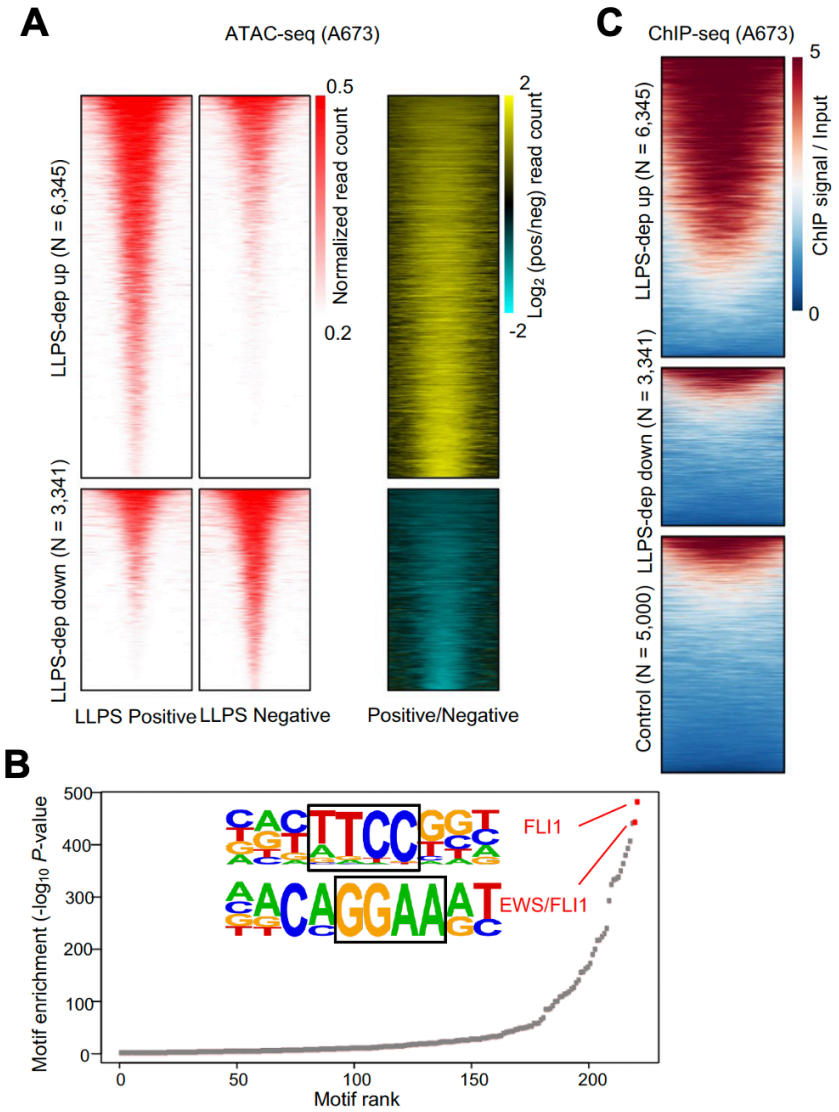

Figure III-4. ARID1A LLPS increases chromatin opening at EWS/FLI1 bound cREs

(A) Tornado plots illustrating ±800 bp regions from each dysregulated cREs (FDR < 0.05) obtained from ATAC-seq of WT, *ARID1A*$^{-/-}$, five *ARID1A*$^{-/-}$+WT, five *ARID1A*$^{-/-}$+ΔDD, and *ARID1A*$^{-/-}$+PrLD(Y/S) A673 cell lines. The colors indicate normalized read counts (left, red) and log2 (LLPS-positive / LLPS-negative) read counts (right, yellow, and cyan). (B) The rank of transcription factor motifs overrepresented in the ARID1A LLPS-dependent upregulated cREs. The top two enriched motifs are highlighted. (C) Tornado plots

illustrating published A673 EWS/FLI1 ChIP-seq signal on the ARID1A LLPS-dependent upregulated cREs, ARID1A LLPS-dependent downregulated cREs, and other randomly selected cREs, respectively. The colors indicate normalized EWS/FLI1 ChIP-seq signal over the input signal.

ARID1A functions as a coactivator of EWS/FLI1

Next, I merged EWS/FLI1 ChIP-seq data with RNA-seq and ATAC-seq data. EWS/FLI1 peaks obtained from A673 Ewing's sarcoma cells showed remarkable overlap with ARID1A LLPS-dependent upregulated cREs (Fig. III-5A) (Adane et al., 2021). These results indicate that the presence of ARID1A LLPS may facilitate the binding of EWS/FLI1 through alteration of chromatin accessibility, which may exert ARID1A LLPS-dependent oncogene activation.

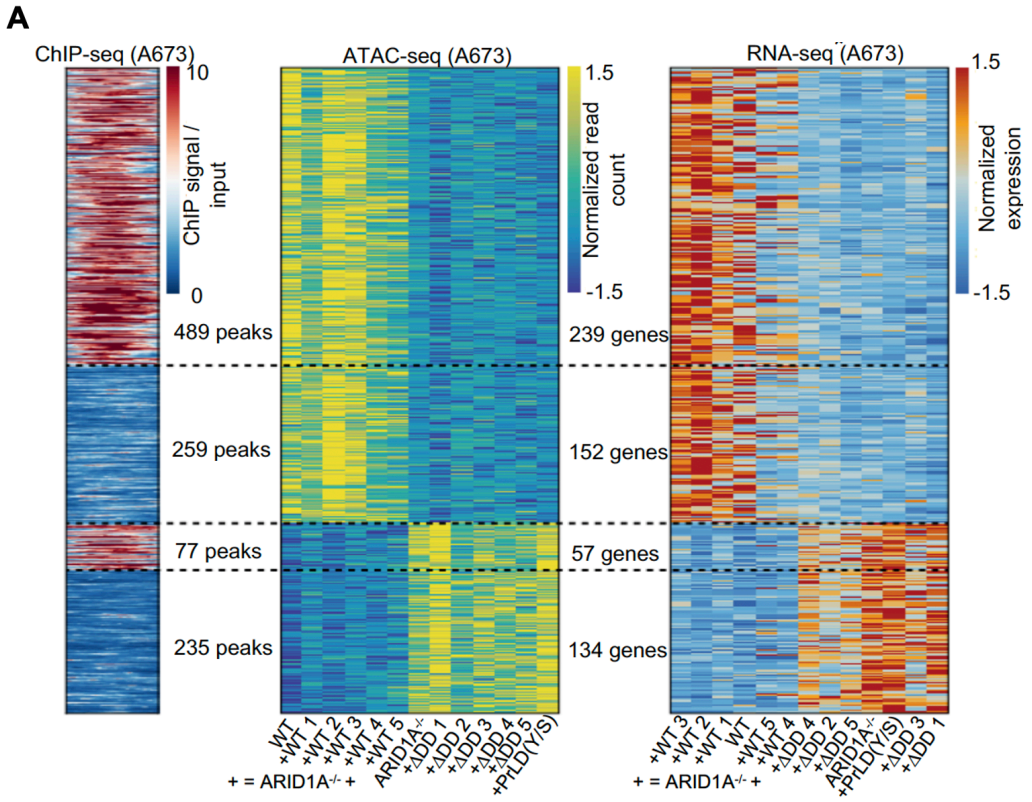

Figure III-5. ARID1A functions as a coactivator of EWS/FLI1

(**A**) Left: A tornado plot illustrating published A673 EWS-FLI1 ChIP-seq signal on the DEG-linked DARs. The colors indicate normalized EWS-FLI1 ChIP-seq signal over input signal (Middle and right). A heatmap illustrating DEG-linked DARs obtained from the ATAC-seq (middle), and DAR-linked DEGs obtained from the RNA-seq (right). The colors indicate normalized read count in the regions and normalized gene expression, respectively.

ARID1A LLPS leads to long-range chromatin interaction change

Reasoning from the significant alteration of chromatin opening, I hypothesized that ARID1A LLPS-dependent cRE activation may directly regulate the oncogene expression in Ewing's sarcoma through modification of chromatin architecture. To test this possibility, I performed *in situ* Hi-C experiments on two ARID1A LLPS-positive and two negative A673 cells (Fig. III-6A) since cREs are known to regulate target genes over large-genomic distance through long-range chromatin interactions (Jung et al., 2019). Analysis of *in situ* Hi-C results revealed that substantial portion of the long-range chromatin contacts between the DEGs and the dysregulated cREs were significantly altered (Fig. III-6B)

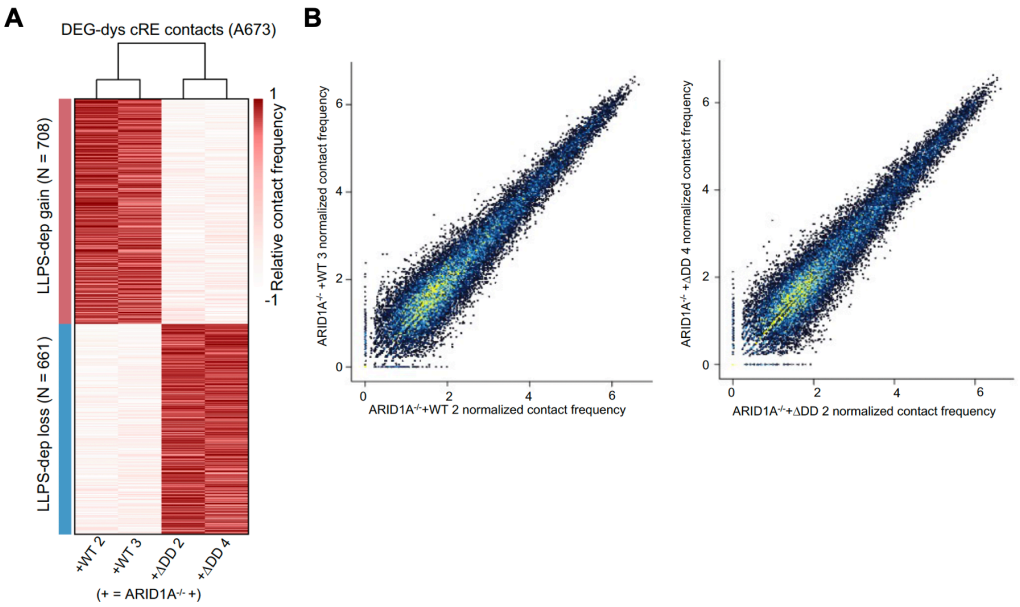

Figure III-6. ARID1A LLPS leads to long-range chromatin interaction change

(**A**) A heatmap showing significantly (*P*-value < 0.05) altered long-range chromatin contacts between the DEGs and the dysregulated cREs in an ARID1A LLPS-dependent manner. (B) Scatter plots showing the reproducibility of normalized Hi-C contacts between the DEGs and the dysregulated cREs in 40 kb resolution.

ARID1A LLPS promote long-range contact of EWS/FLI1 bound cREs

I also found that the genes activated by ARID1A LLPS were markedly linked to the upregulated cREs whereas the repressed genes were linked to the downregulated cREs, (Fig. III-7A) indicating the dysregulated cREs controlled by ARID1A LLPS are closely related to altered gene expression.

Considering that ARID1A LLPS opens EWS/FLI1 binding sites, I further tested whether EWS/FLI1 binding is associated with cREs that control ARID1A LLPS-dependent upregulated genes. Strikingly, over 70% of the upregulated cREs linked to the activated genes are co-occupied by EWS/FLI1 (Fig. III-7B).

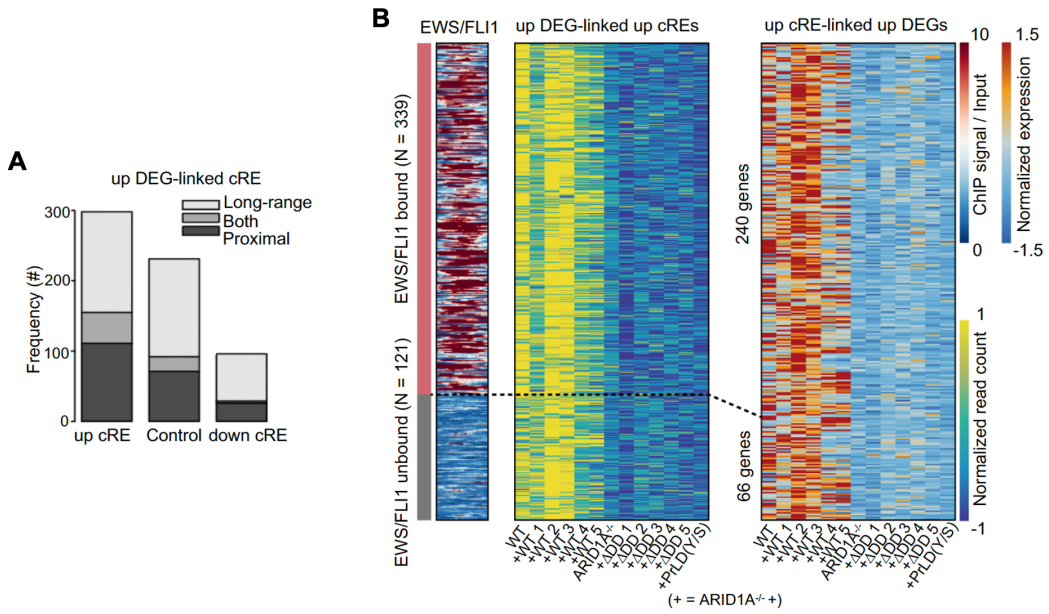

Figure III-7. ARID1A LLPS promote long-range contact of EWS/FLI1 bound cREs

(**A**) A barplot illustrating the number of linkages between upregulated DEGs and upregulated, downregulated, and control cREs, respectively. (**B**) (Left) A tornado plot illustrating published A673 EWS/FLI1 ChIP-seq signal on the upregulated cREs connecting to ARID1A LLPS-dependent upregulated genes. The colors indicate normalized EWS/FLI1 ChIP-seq signal over the input signal. (Middle and right) A heatmap illustrating the upregulated cREs (middle) and the upregulated genes connected to the upregulated cREs (right). The colors indicate normalized read count in the regions and normalized gene expression, respectively. The dashed line indicates linkages between EWS/FLI1-bound upregulated cREs and the upregulated genes.

ARID1A LLPS mediated long-range chromatin contact induces oncogenic gene expression

The promoter of TFAP2B, a gene that was previously found to be induced by EWS/FLI1, showed direct contact with ARID1A LLPS dysregulated cREs (Fig. III-8A) (Riggi et al., 2010). Such enriched EWS/FLI1 binding was not observed when I examined the downregulated cREs linked to the repressed genes. Importantly, the genes putatively activated by the upregulated cREs showed strong enrichment in the cancer-related GO terms compared to the other activated genes (Fig. III-8B). My results indicate that EWS/FLI1-bound cREs, dysregulated in cells with ARID1A LLPS, induce expression of oncogenic genes that increase the oncogenic potential of A673 cells.

Since EWS/FLI1 binding has been previously reported to drive long-range chromatin contacts (Showpnil et al., 2022; Sanalkumar et alk., 2023), I examined whether the increased long-range chromatin contacts are mainly driven by EWS/FLI1 or ARID1A LLPS itself. I utilized *in situ* Hi-C results of EWS/FLI1-depleted A673 cells and investigated chromatin contacts between the activated cREs occupied by EWS/FLI1 and linked upregulated genes. Consistent with the previous studies, I observed that the long-range chromatin contacts were generally weakened compared to the ARID1A LLPS-positive WT rescue cells (Fig. III-8C). Nevertheless, the contacts were significantly maintained when compared to the ARID1A LLPS-negative cells, suggesting that both EWS/FLI1 binding and ARID1A LLPS are necessary for establishing the oncogenic long-range chromatin contacts (Fig. III-9C). Together, my genome-wide analysis suggested that LLPS of ARID1A

activates oncogenes in A673 cells via inducing both the opening of de novo EWS/FLI1-bound cREs and the establishing long-range chromatin contacts.

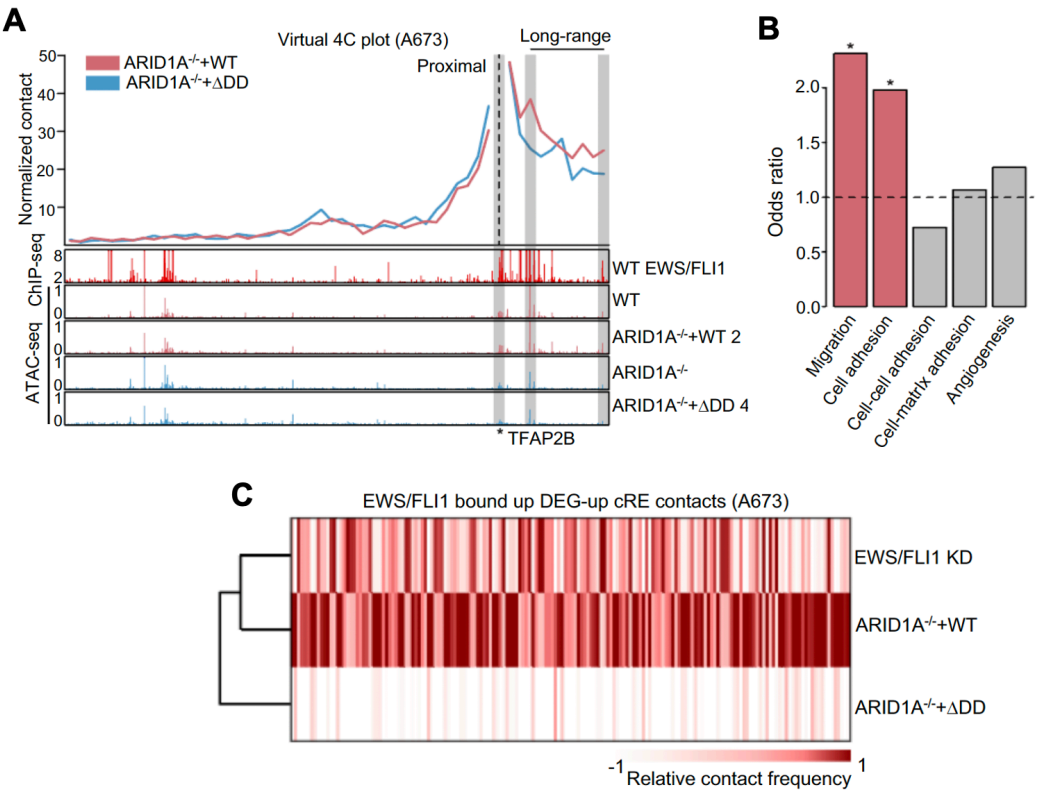

Figure III-8. ARID1A LLPS mediated long-range chromatin contact induces oncogenic gene expression

(**A**) The normalized Hi-C contact frequencies around the TFAP2B gene promoter are illustrated as a virtual 4C plot. The genome tracks of ATAC-seq and published EWS/FLI1 ChIP-seq signal are shown below. The dashed vertical line indicates the viewpoint of the 4C plot and the asterisk indicates the transcription start site of the TFAP2B gene. The shaded regions highlight the linkages between the TFAP2B gene and the EWS/FLI1-bound upregulated cREs via the proximal co-localization or the altered long-range chromatin contacts. (**B**) Odds ratio that an activated gene is included in the cancer-related GO-terms shown in Fig. 4C, comparing the genes linked to the upregulated cREs versus unlinked genes (*P*-values for the enrichment of the linked genes versus the unlinked genes: migration = 0.034, cell adhesion = 0.038, two-sided Fisher's exact test). (**C**) A heatmap comparing

normalized Hi-C contact frequencies of ARID1A LLPS-positive, negative, and published EWS/FLI1 knockdown (KD) A673 cells, respectively. Only the contacts between EWS/FLI1-bound upregulated cREs and their linked upregulated genes are shown.

ARID1A LLPS induces oncogenic transcription program across multiple EWS cancer cell lines

Recent studies suggested that Ewing's sarcoma cell lines show high plasticity (Orth et al., 2022). Therefore, I tested oncogenic potential of ARID1A in different Ewing's sarcoma cell lines. With published EWS/ETV ChIP-seq data across four Ewing's sarcoma cell lines (EW1, RDES, SKES1, and SKNMC), I analyzed the EWS/ETVs binding patterns on ARID1A LLPS-dysregulated cREs obtained from A673 (Fig. III-9A). Notably, I observed significant similarity and conservation of binding pattern of EWS/ETVs across the four EwS cell lines compared to A673 (Fig. III-9A,B) Furthermore, my analysis revealed that EWS/ETVs from other Ewing's sarcoma cell lines exhibited enriched binding at the ARID1A LLPS-dependent upregulated cREs, while EWS/ETV binding was scarce in the downregulated cREs (Fig. III-9B). This result suggests that ARID1A LLPS is crucial in opening the EWS/ETVs binding sites across EwS cell lines. To validate that the phase separation of ARID1A influences transcriptional activity of EWS/ETV in various Ewing's sarcoma cell lines, I first generated ARID1A knockdown (KD) cell lines using another Ewing's sarcoma cell lines, SK-N-MC and TC106 (Fig. III-9C,D).

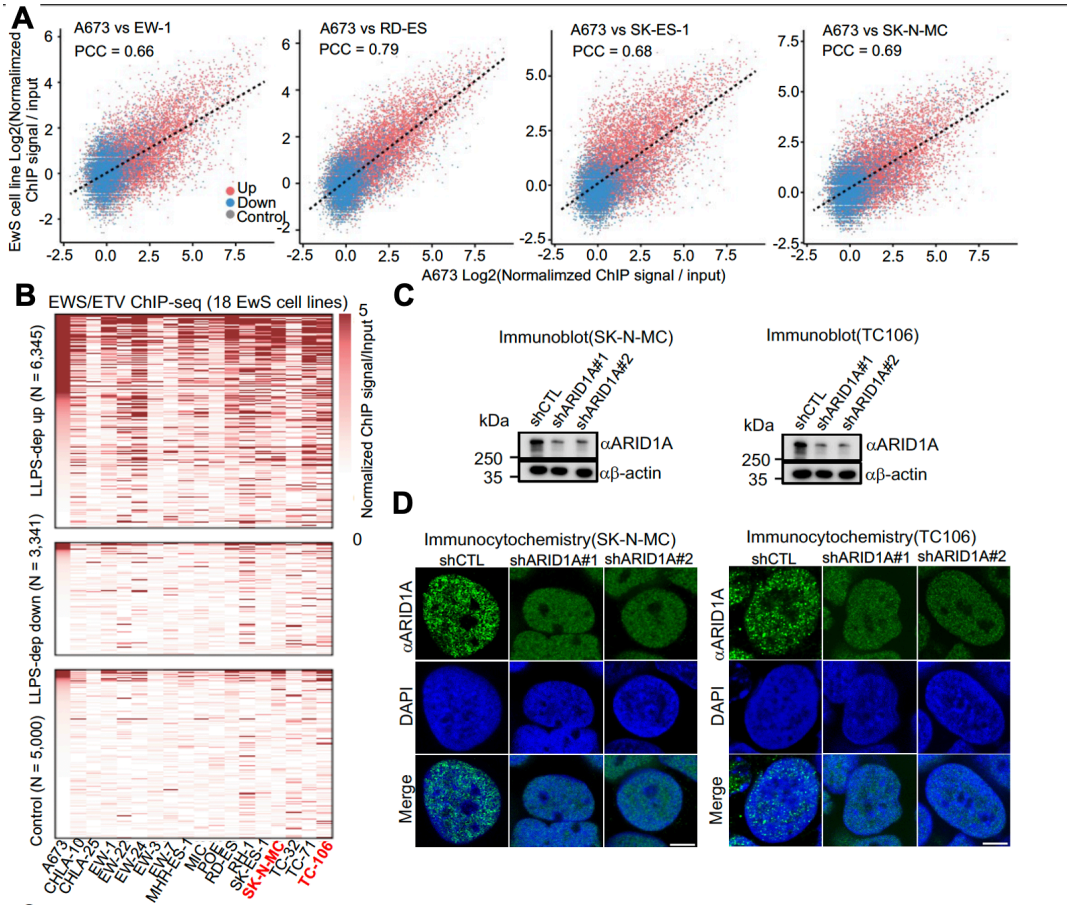

Figure III-9. ARID1A LLPS induces oncogenic transcription program across multiple EWS cancer cell lines

(**A**) Scatter plots highlighting conserved EWS/ETV binding patterns across EwS cell lines (A673, EW-1, RD-ES, SK-ES-1, and SK-N-MC) on the ARID1A LLPS-dependent dysregulated cREs of A673. Published EWS/ETV ChIP-seq signals of each cell line on the upregulated (red), downregulated (blue), and other randomly selected (grey) cREs are shown. The dashed black lines are trend lines of the dots. (**B**) Heat map illustrating EWS/ETV ChIP-seq signal across 18 different Ewing's sarcoma cell lines on the ARID1A LLPS-dependent upregulated cREs, ARID1A LLPS-dependent downregulated cREs, and other randomly selected cREs, respectively. The colors indicate normalized EWS/FLI1

ChIP-seq signal over the input signal. (**C**) Representative immunoblot image showing ARID1A protein expression in shCTL, shARID1A knockdown SK-N-MC and TC106 cell lines. (**D**) Immunocytochemstry data showing anti- ARID1A staining of shCTL, shARID1A knockdown SK-N-MC and TC106 cell lines.

ARID1A LLPS induces oncogenic gene expression in SK-N-MC and TC106 cell

Additionally, qRT-PCR was performed to demonstrate the transcriptional regulation of EWS/FLI1-bound genes by ARID1A, revealing a significant decrease in the expression of EWS/FLI1 target genes in the SK-N-MC KD cell line (Fig. III-10A). To illustrate the dependency of EWS/ERG on ARID1A, I conducted ARID1A knock-down experiments with subsequent qRT-PCR validation in TC106 cell line. The results indicated that EWS/ERG-bound genes, putative targets of ARID1A LLPS, exhibited decreased gene expression in KD TC106 cell lines (Fig. III-10B). Collectively, these findings demonstrate that ARID1A LLPS is indeed crucial for the oncogenic potential of Ewing's sarcoma, even across different biological backgrounds.

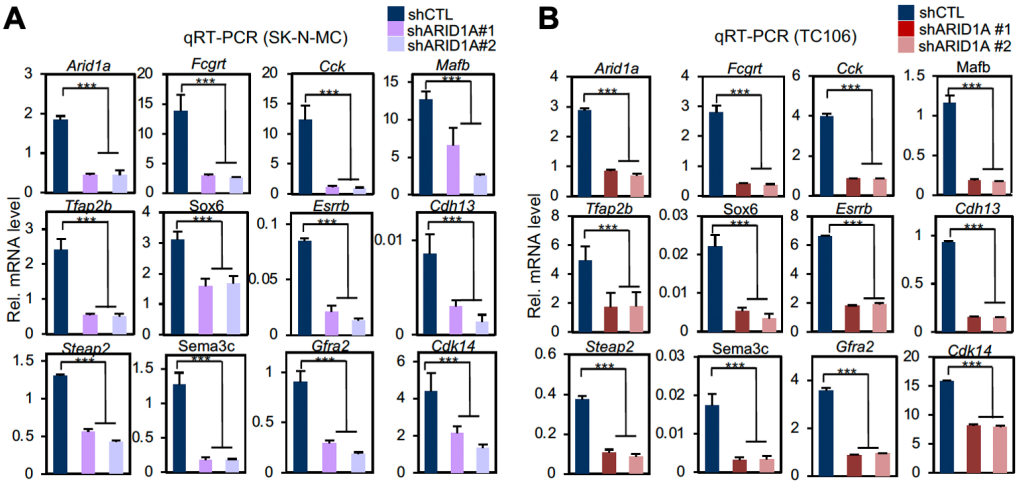

Figure III-10. ARID1A LLPS induces oncogenic gene expression in SK-N-MC and TC106 cell

(**A**) Quantitative RTPCR analysis of ARID1A LLPS dependent genes in SK-N-MC cell lines. Bars, mean ± s.e.m.; ***p < 0.001, Statistics by two-tailed t-test. (**B**) Quantitative RT-PCR analysis of ARID1A LLPS dependent genes in TC106 cell lines. Bars, mean ± s.e.m.; ***p < 0.001, Statistics by two-tailed t-test.

ARID1A interacts with EWS/FLI1 through phase separation

As ARID1A LLPS turned out to induce a significant change in both the chromatin structure and transcriptional profile of EWS/FLI1 target genes, I investigated the direct connection between EWS/FLI1 and ARID1A. Co-immunoprecipitation assay verified a direct interaction between the PrLD1 and PrLD2 regions of ARID1A- responsible for LLPS- and EWS/FLI1 (Fig. III-11A). Additionally, ARID1A WT that is capable of LLPS showed the ability to bind EWS/FLI1, whereas the PrLD Y/S mutant failed to bind (Fig. III-11A), suggesting that the interaction between ARID1A and EWS/FLI1 relies on LLPS. Furthermore, the *in vitro* droplet assay showed that EWS/FLI1 forms co-condensate with ARID1A WT (Fig. III-11B). To further probe the effect of ARID1A LLPS on EWS/FLI1 subcellular localization, I performed immunostaining of endogenous EWS/FLI1 for both ARID1A LLPS-positive and negative cells. In ARID1A LLPS positive cells, there was formation of nuclear condensates of both ARID1A and EWS/FLI1, which were observed to co-localize. However, in ARID1A LLPS negative cells, the number of EWS/FLI1 condensates decreased significantly (Fig. III-11C,D) (Chong et al., 2018; Chong et al., 2022; Zuo et al., 2021).

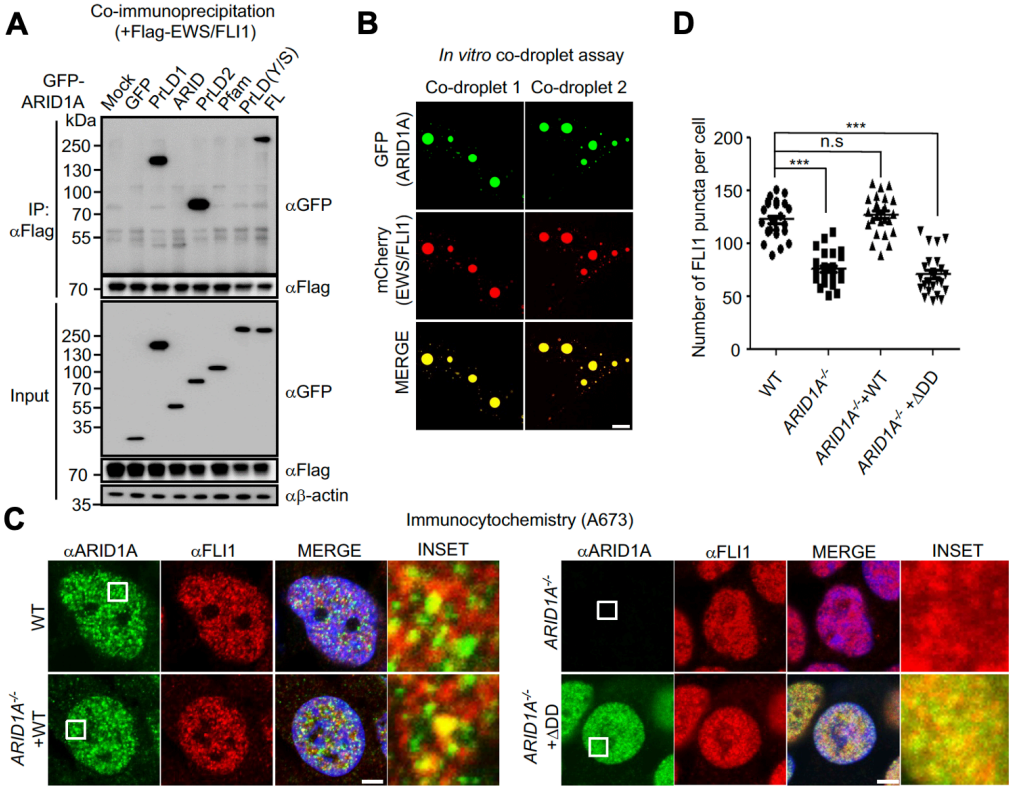

Figure III-11. ARID1A interacts with EWS/FLI1 through phase separation

(**A**) Binding site mapping of FLAG-EWS/FLI1 and GFP-ARID1A recombinant proteins by co-immunoprecipitation assay. Tested proteins include PrLD1, ARID, PrLD2, Pfam, PrLD(Y/S) mutant, and full-length ARID1A. The representative images supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (**B**) Confocal image of an in vitro co-droplet assay demonstrating co-localization of purified GFP-ARID1A and mCherry-EWS/FLI1. Scale bars: 5 μm. The representative images

supported by the relevant statistics have been chosen upon 3 independent preparations with similar outcome. (**C**) Representative confocal images of ARID1A-/-+WT and ARID1A-/-+ΔDD A673 cell lines immunostained with anti-FLI1 and anti-ARID1A antibodies. Scale bars: 5 μm. (**D**) Quantification of number of FLI1 puncta per cell lines in c. n = 32 technical replicates of cells; Bars represents mean ± s.e.m.; **$p < 0.01$, ***$p < 0.001$, NS non-significant. Statistical analysis performed using a two-tailed t-test. ARID1A-/-, ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines were individually compared to WT.

ARID1A LLPS induces binding of EWS/FLI1 to GGAA microsatellite enhancer

I further performed ChIP assay using anti-EWS/FLI1, anti-H3K27ac and anti-SMARCC1 antibodies in the A673 cell lines. I targeted cREs of previously reported EWS/FLI1 induced genes where chromatin accessibility is regulated by ARID1A LLPS (Levetzow et al., 2011; Marchetto et al., 2020). In all cREs we tested, chromatin occupancy of EWS/FLI1 and SMARCC1 significantly decreased upon loss of ARID1A LLPS, along with decreased H3K27ac, active enhancer histone marker (Fig. III-12A). Taken together, my data suggest that nuclear condensates of ARID1A navigate to EWS/FLI1 target genes through co-phase separation with EWS/FLI1 and subsequently compartmentalize BAF complex to form active chromatin remodeling hub and promote Ewing's sarcoma.

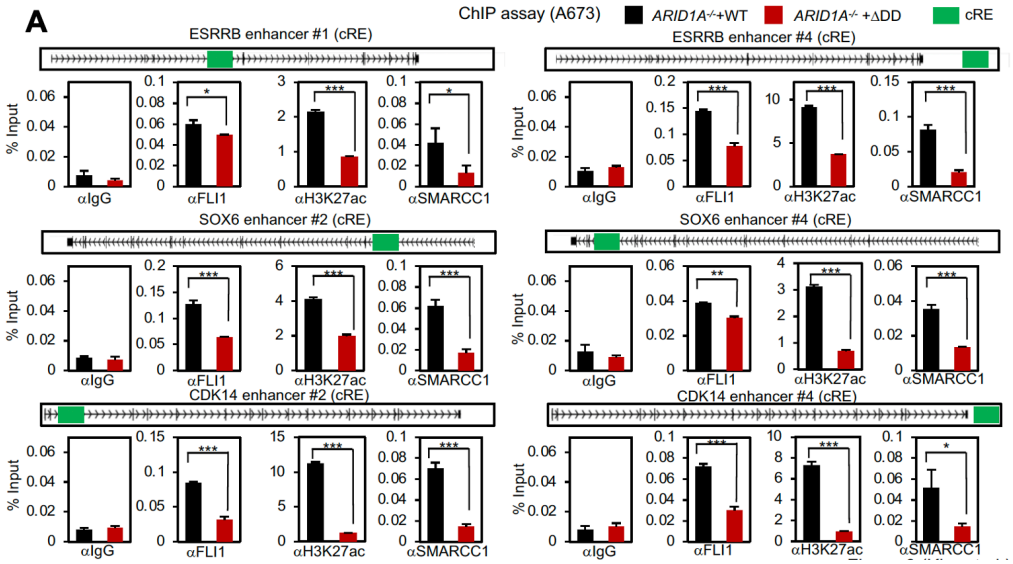

Figure III-12. ARID1A LLPS induces binding of EWS/FLI1 to GGAA microsatellite enhancer

(**A**) ChIP assays performed on EWS/FLI1 bound enhancers in ARID1A-/-+WT and ARID1A-/-+ΔDD A673 cell lines using antibodies against IgG, FLI1, H3K27ac, and SMARCC1. Bars represents mean ± s.e.m.; n = 3 technical replicates; **p < 0.01, ***p < 0.001, NS non-significant. Statistics by two-tailed t-test using ARID1A-/-+WT, and ARID1A-/-+ΔDD A673 cell lines as comparison. Source data are provided as a Source Data file.

III-4. Discussion

Recent studies have highlighted the importance of phase separation in various biological processes, such as RNA processing, autophagy, and cell signaling (Somasekharan et alk., 2922, Fujioka et al., 2020; Su et alk., 2016). Phase-separated condensates form molecular 'hot spots' where molecules with corresponding functionalities get drawn in and set up to execute certain activities. Inside the nucleus, transcriptional condensates with a phase-separated feature demonstrate liquid-like properties and adjust gene expression accordingly. Given that these droplets are an important element of core cellular processes, any irregularities in them could be correlated to maladies, especially cancer.

My study offers clarity on the way ARID1A, with its phase-separation property, advances the oncogenic state of Ewing's sarcoma. ARID1A is composed of PrLDs which have tyrosine residues that act as sticky motifs to enable various interactions and bring about phase separation. Additionally, ARID1A contains a structured Pfam homology domain that directly binds to components of the BAF complex. Surprisingly, phase separation of ARID1A does not interrupt with its ability to connect to BAF molecules, however, these PrLDs are central to compartmentalizing BAF molecules into condensates and creating chromatin remodeling hubs. Samples from Ewing's Sarcoma patients evidenced the presence of visible ARID1A condensates. Disruption of ARID1A LLPS in a cancer cell line weakened the effects of Ewing's sarcoma's oncogenic potential, indicating that the condensates initiated by ARID1A play a role in the cancer phenotype. Consolidation of genome-wide studies, including ChIP-seq, RNA-seq, ATAC-seq, and Hi-C, demonstrated

that the phase separation of ARID1A specifically enhances chromatin accessibility at EWS/FLI1 bound cREs, leading to changes in chromatin architecture and subsequent transcriptional alterations at oncogenic target genes.

Given the capacity of ARID1A to undergo phase separation, this protein may form condensates in cellular settings beyond Ewing's sarcoma. The phase separation of ARID1A in Ewing's sarcoma is indispensable for its engagement with EWS/FLI1, which provides access to EWS/FLI1-linked enhancers and in turn induces chromatin remodeling. ARID1A gains access to enhancers and promoters once it partners with a variety of transcription factors, being AP1, ERα and FOXA. In this fashion, it can be hypothesized that in relevant biological conditions/context, ARID1A could possibly undergo phase separation with distinct transcription factors at chromatin for activation of the gene transcription. It is noteworthy that ARID1A is presented with a high rate of somatic mutations in many diseases, contributing to the accumulation of a significant number of mutations in its PrLDs (Hung et al., 2020; Chan-On et al., 2013). Therefore, it can be speculated that mutations occurring at PrLDs, connected with the disease, might alter the phase separation properties of ARID1A and/or its connection with the transcription factors, thus becoming an important factor in the pathogenesis of human malignancies.

Several transcription factors arising from chromosome rearrangements and translocations, similar to the case of Ewing's sarcoma, have been seen to hijack the BAF complex in cancer. Those transcription factors include EWS/FLI1, TMPRSS/ERG, MN1, FUS/DDIT3, and ENL (Boulay et al., 2018; Sandoval et al., 2018; Riedel et al., 2021; Linden et al., 2019; Nie et al., 2003). Interestingly, EWS/FLI1, MN1, FUS/DDIT3, and

ENL all undergo liquid-liquid phase separation within cancer cells (Guo et al., 2020). Additionally, a prior study indicated that EWS/FLI1 fusion protein interacts with the BAF complex via its PrLD domain. In line with this, our observations reveal co-condensation between ARID1A and EWS/FLI1, facilitated by mutual interactions of their PrLDs. Consequently, it is plausible that oncogenic factors equipped with PrLDs and IDR might co-opt the BAF complex by directly interacting with ARID1A. Although the exact mechanism of recruitment and the nature of the interaction between the BAF complex and oncogenic factors are not fully understood, our results suggest that ARID1A may serve as a connecting link between the oncogenic factor and the BAF chromatin remodeling complex.

Previous attempts have been made to pharmacologically inhibit EWS/FLI1, a major driver of sarcoma, using small molecules. The small molecule trabectedin was found to prevent the localization of EWS/FLI1 within the nucleoplasm and disrupt its function (Harlow et al., 2019). YK-4-279 could block the binding of RNA helicase A with EWS/FLI1, leading to reduced proliferation of Ewing's sarcoma cells (Erkizan et alk., 2009). Nevertheless, transcription factors usually bind extremely tightly to target DNAs and do not feature domains in which small molecules can act on, indicating they are resistant to becoming targeted by potential inhibitor. As a result, more researchers are striving to locate and create safe drugs targeting transcriptional cofactors that are linked indirectly to DNA and move around the genome. High-throughput screening has identified an inhibitor of ARID1A, known as BD98, which can be used to target the ARID1A-specific BAF complex (Marian et al., 2018). Administering BD98 to embryonic stem cells and T cells can effectively

simulate the impact of ARID1A depletion, causing a reduction in the ARID1A-oriented transcriptional program (Guo et al., 2022). Furthermore, this inhibitor has been employed, along with an ATR inhibitor, to cause cell death in colorectal carcinoma cells, illustrating its potential as a cancer therapeutic (Chory et al., 2020). Targeting both ARID1A and its PrLDs could be a promising strategy to restrict EWS/FLI1 from attaching to regulatory enhancers and prevent chromatin contact in the vicinity of oncogenes, making ARID1A PrLDs a potential therapeutic target for Ewing's sarcoma. Combining an ARID1A inhibitor like BD98 with an inhibitor for EWS/FLI1 could mount a more potent form of therapy for patients. Another group has discovered sequence grammar within the IDRs of ARID1A and ARID1B that governs their phase separation within the cell. This condensation of ARID1A/ARID1B creates a distinctive network of protein-protein interactions critical for chromatin navigation and gene activation. Additionally, perturbations in the IDR of ARID1B associated with human diseases have been identified, indicating that the IDR could be a potential therapeutic target region (Patil et al., 2023). These results spotlight the potential of future drug developments, such as targeting ARID1A PrLDs to inhibit abnormal activation of pathogenic genes and halt disease progression.

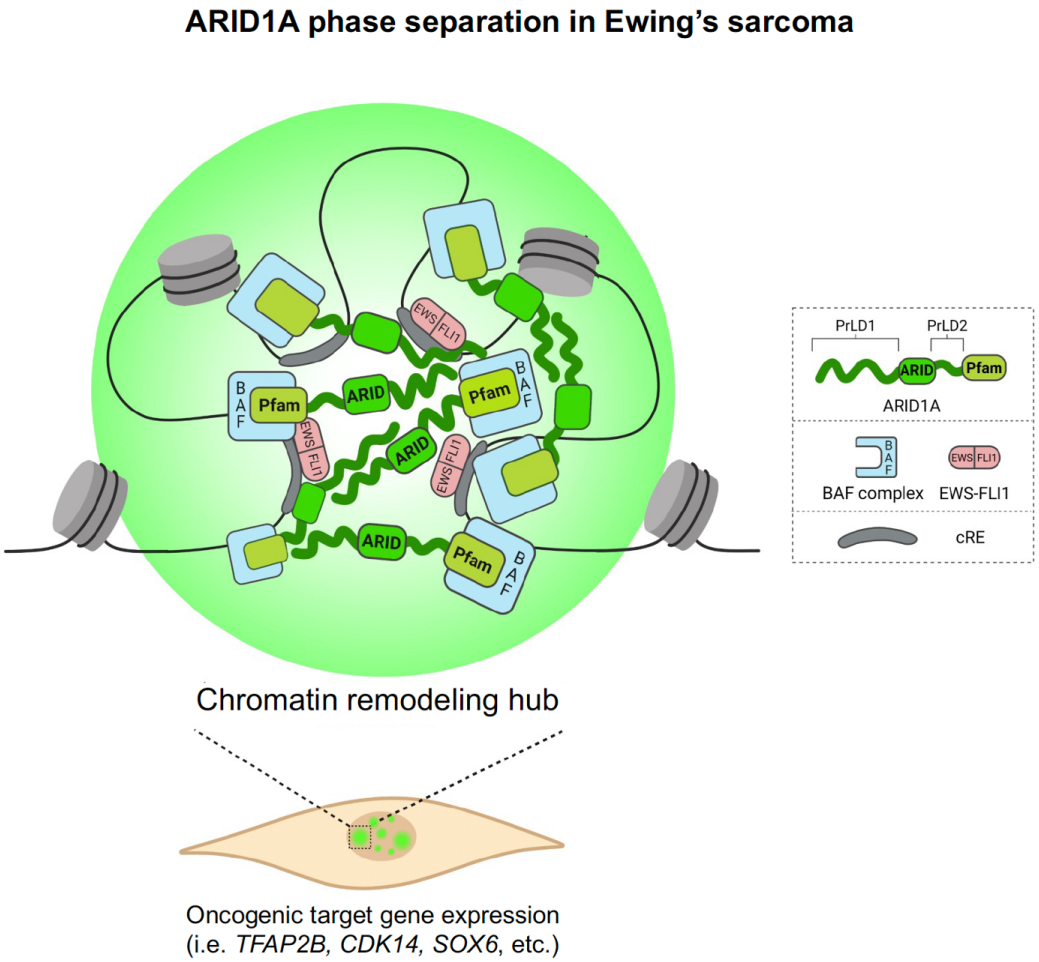

Figure III-13. Schematic model of ARID1A phase separation mediated oncogenesis of Ewing's sarcoma

ARID1A undergoes phase separation through its prion-like domain. Upon condensate formation, ARID1A recruits BAF complex subunits to generate chromatin remodeling hub. Aberrant ARID1A nuclear condensate found in Ewing's sarcoma induces long range chromatin contact that upregulates oncogenic gene transcription, eventually leading to oncogenesis.

III-5. Materials and Methods

RNA-sequencing (RNA-seq)

RNA was extracted from 10^6 harvested cells with a Nucleospin RNA XS kit (Macherey-Nagel, MN740902). RNA-seq libraries were prepared using TruSeq stranded mRNA library prep kit (Illumina, 20020594). RNA-seq libraries were sequenced in a 100bp paired-end mode, with a MGI DNBSEQ-G400 system.

RNA-seq analysis

Reads were aligned to the reference genome (hg38) using STAR software v2.7.8a with default parameters (Dobin et al., 2013). The gene counts were quantified with RSEM (Li et al., 2011). The differentially expressed genes (DEGs) were obtained using DESeq2 with a false discovery rate (FDR) < 0.05 (Love et al., 2014). Among the obtained DEGs, only genes annotated as protein coding gene with confidence levels 1 and 2 were used. Gene ontology (GO) analysis was performed using DAVID with GO biological process (BP) (Sherman et al., 2022).

ATAC-seq analysis

ATAC-seq libraries were prepared for sequencing using Illumina Tagment DNA TDE1 Enzyme and Buffer Kits (#20034197, Illumina, San Diego, CA, USA). The adaptor sequences were trimmed out using Cutadapt (Kechin et al., 2017). The trimmed paired-end sequences were mapped to the human reference genome hg38 using bowtie 2 with

parameters "—very-sensitive –X 1000 –dovetail". The reads with poor mapping quality (MAPQ < 30) and the reads mapped to the mitochondrial genome were discarded. The potential PCR duplicates were marked using MarkDuplicates of Picard, and the reads were shifted using the alignmentSieve function of deeptools with the "—ATACshift" parameter (Ramirez et al., 2016). The accessible regions were defined using MACS2 narrow callpeak, keeping duplicates with a q-value cutoff of 0.01. For downstream data analyses, I merged accessible regions obtained from all samples. ARID1A LLPS-dependent dysregulated cREs were identified by applying DESeq2 to the read counts on the merged accessible regions (FDR < 0.05). Tornado plots of ATAC signal were generated using deeptools bamCoverage and computeMatrix. Enriched motifs of the dysregulated cREs were identified by using HOMER findMotifsGenome knownResults with the parameters '-size given' (Heinz et al.,2010). PCA analysis of ATAC-seq was conducted on the $\log_2$-normalized top 500 highly variable peaks using DESeq2. For visualization into genome track, ATAC reads were depth-normalized among the samples.

***In situ* Hi-C**

In situ Hi-C was performed on two ARID1A LLPS-positive (*ARID1A*$^{-/-}$+WT 2 and 3) and two LLPS-negative (*ARID1A*$^{-/-}$+ΔDD 2 and 4) cells. For each sample, 10^6 cells were harvested and crosslinked with 1 % formaldehyde for 9 min at RT in 10 ml PBS and 100 µl FBS. Cells were treated with 250 mM glycine for 5 min at RT and 15 min on ice, to quench the crosslinking. The cells were then lysed with 10 nM Tris-HCl pH 8, 10 mM NaCl, and 0.2 % IGEPAL CA630. The crosslinked chromatin was digested with 100 U MboI, labeled

with biotin-14-dTCP, and ligated with T4 DNA Ligase. The ligated samples were reverse-crosslinked with 2 μg/μl proteinase K, 1 % SDS, and 500 mM NaCl overnight at 65 ºC. The DNA fragments were collected with Ampure XP beads (Beckman Coulter, A63881) and sonicated using Covaris S220 into 300~400bp. The biotin-labeled DNA was pulled down with Dynabeads MyOne streptavidin T1 beads (Invitrogen, 65602) with thorough washings. DNA end repair, un-ligated ends removal, adenosine addition at 3' end (NEB, M0212), ligation of Illumina indexed adapters (NEB, M2200), and PCR amplification was performed to generate Hi-C libraries. The generated libraries were sequenced in 100 bp paired-end mode using MGI DNBSEQ-G400.

***In situ* Hi-C analysis**

Published A673 *in situ* Hi-C data upon EWS/FLI1 depletion was downloaded from GEO database under accession number GSE185125. For both performed and downloaded *in situ* Hi-C data, the sequenced reads were mapped to the human reference genome (hg38) using BWA-mem. Chimeric reads spanning multiple sites of the genome were filtered out. The reads with poor mapping quality (MAPQ < 10) and putative self-ligated reads (genome distance < 15 kb) were discarded. Potential PCR duplicates were marked using MarkDuplicates of Picard (Li et al., 2009). The reads were then assigned into 40 kb genomic bins to generate a 40 kb Hi-C contact map. To consider possible genome-dependent bias, coverage-based contact map normalization was performed with covNorm (Kim et al., 2021). To investigate altered chromatin contacts between the DEGs and the dysregulated cREs mediated by ARID1A LLPS, I collected the normalized Hi-C contacts linked to all pairs of

possible DEGs and dysregulated cREs within 2 Mb from LLPS-positive and negative samples. Quantile normalization of the collected contacts was performed among the samples to normalize depth differences. The contacts were then log-transformed and used as input to LIMMA (Ritchie et al., 2015). I defined significantly increased/decreased contacts by LLPS of ARID1A using the limma-trend algorithm (P-value < 0.05).

Gene-cRE linkage

To investigate the subset of DEGs that are directly regulated by the ARID1A LLPS dysregulated cREs, I defined the potential regulatory linkage between the DEGs and the dysregulated cREs. To account for both proximal and long-range gene-cRE interactions, I considered each gene-cRE pair as linked if they are co-localized (< 40 kb) or if chromatin contact significantly increased between the two elements upon the LLPS of ARID1A in 40 kb resolution.

ChIP assay

Cells were crosslinked with 1 % formaldehyde for 10 min at room temperature. Next, 1.25M glycine was used for quenching and cells were washed two times using PBS. The cells were then scraped and lysed in a buffer containing 50 mM Tris–HCl (pH 8.1), 10 mM EDTA, 1 % SDS, supplemented with a complete protease inhibitor cocktail (Roche). Cells were sonicated with the sonication condition of 70 amplitude, 30min process time, 30s ON and 30s OFF. After sonication, lysates were centrifuged and supernatant was taken. Chromatin extracts containing DNA fragments with an average of 250 bp were then diluted

ten times with dilution buffer containing 1 % Triton X-100, 2 mM EDTA, 150 mM NaCl, and 20 mM Tris–HCl (pH 8.1) with complete protease inhibitor cocktail and subjected to immunoprecipitations overnight at 4 °C. Conjugates were further incubated using BSA blocked 40 µl of protein A/G Sepharose for 2 h at 4 °C. Beads were washed with TSE I buffer (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8.1) and 150 mM NaCl), TSE II buffer (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8.1) and 500 mM NaCl), buffer III (0.25 M LiCl, 1 % NP-40, 1 % deoxycholate, 10 mM Tris–HCl (pH 8.1) and 1 mM EDTA), three times TE buffer (10 mM Tris–HCl (pH 8.0) and 1 mM EDTA) and eluted in elution buffer (1 % SDS and 0.1 M $NaHCO_3$). Reverse-crosslinking was performed by incubating the eluted DNA in 65 °C overnight. RNase and Proteinase K was treated. Lastly DNA was purified using MinElute PCR Purification KIT(QIAGEN). Purified DNA was used for qPCR analysis using primer targeting each enhancer regions of target genes (Supplementary data 2).

ChIP-seq data analysis

Published A673 EWS/FLI1 ChIP-seq and input data were downloaded from GEO database under accession number GSE165783. Reads were mapped to the reference genome (hg38) using BWA-mem and the potential PCR duplicates were marked using MarkDuplicates of Picard[69]. The reads with poor mapping quality (MAPQ < 10) were discarded. EWS/FLI1 peaks were called using MACS2 narrow callpeak, with q-value a cutoff of 0.01. The cREs are considered “EWS/FLI1-bound” if there is an overlap between the cRE and the EWS/FLI1 peak. The tornado plots of EWS/FLI1 on the cREs were generated with

deeptools computeMatrix function with an option “scale-regions” .

CHAPTER IV
Conclusion

Recent advances have highlighted the critical role of liquid-liquid phase separation (LLPS) in cellular compartmentalization and regulation of gene expression. In this study, I establish the importance of AT-rich interactive domain-containing protein 1A (ARID1A) in Ewing's sarcoma progression through its PrLD-mediated LLPS. My findings demonstrate that ARID1A LLPS is pivotal for forming nuclear condensates enriched with chromatin remodeler subunits of the BAF complex. These condensates function as transcriptional hubs, mediating chromatin accessibility and facilitating long-range chromatin interactions at EWS/FLI1 target sites. Genome-wide chromatin and transcriptional profiling reveal that ARID1A LLPS significantly enhances the activation of oncogenic pathways by modifying chromatin architecture. Through the use of CRISPR-Cas9 and mutational analysis, I show that disruption of ARID1A LLPS abrogates its oncogenic functions, including cell proliferation, migration, and invasion, both in vitro and in vivo. This highlights the indispensable role of PrLDs in promoting Ewing's sarcoma tumorigenesis. Furthermore, I identify that ARID1A LLPS supports the co-condensation with EWS/FLI1 and contributes to the recruitment of chromatin remodelers, thereby amplifying oncogenic transcription programs. These insights position ARID1A as a crucial player in chromatin remodeling and oncogenesis, emphasizing its potential as a therapeutic target. Targeting ARID1A PrLDs or its LLPS capacity may offer a novel therapeutic strategy to disrupt transcriptional condensates that drive cancer progression. Future research could focus on elucidating the broader implications of ARID1A phase separation in other cancers and the development of small molecules targeting ARID1A condensate formation. This study not only enhances my understanding of ARID1A's mechanistic role in Ewing's sarcoma but also opens avenues

for innovative therapeutic interventions.

REFERENCES

D. Bracha, M. T. Walls, C. P. Brangwynne, Probing and engineering liquid-phase organelles. *Nature Biotechnology*. 37, 1435–1445 (2019).

L. Peng, E.-M. Li, L.-Y. Xu, Phase separation and transcriptional regulation. *Trends in Biochemical Sciences*. 46, 531–545 (2021).

M. J. S. Burke, L. A. Mullally, Effects of pH alterations on stress- and aging-induced protein phase separation. *Cellular and Molecular Life Sciences*. 79, 380 (2022).

J. Mandal, P. Mandal, T.-L. Wang, I.-M. Shih, Treating ARID1A mutated cancers by harnessing synthetic lethality and DNA damage response. *Journal of Experimental & Clinical Cancer Research*. 41, 256 (2022).

M. Wanior, A. Krämer, S. Knapp, A. C. Joerger, Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy. *Oncogene*. 40, 3637–3654 (2021).

K. C. Helming, X. Wang, C. W. M. Roberts, Vulnerabilities of Mutant SWI/SNF Complexes in Cancer. *Cancer Cell*. 26, 309–317 (2014).

K. Berns, J. J. Caumanns, E. M. Hijmans, A. M. Gennissen, T. M. Severson, B. Evers, G. B. A. Wisman, G. J. Meersma, C. Lieftink, R. L. Beijersbergen, ARID1A mutation sensitizes most ovarian clear cell carcinomas to BET inhibitors. *Oncogene*. 37, 4611–4625 (2018).

S. Nagarajan, S. V. Rao, J. Sutton, D. Cheeseman, S. Dunn, E. K. Papachristou, J.-E. G. Prada, D.-L. Couturier, S. Kumar, K. Kishore, ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response. *Nature Genetics*. 52,

187–197 (2020).

L. Zhang, C. Wang, S. Yu, C. Jia, J. Yan, Z. Lu, J. Chen, Loss of ARID1A Expression Correlates With Tumor Differentiation and Tumor Progression Stage in Pancreatic Ductal Adenocarcinoma. *Technology in Cancer Research & Treatment*. 17, 153303461875447 (2018).

N. Riggi, M. L. Suvà, I. Stamenkovic, Ewing's Sarcoma. *New England Journal of Medicine*. 384, 154–164 (2021).

L. Zuo, G. Zhang, M. Massett, J. Cheng, Z. Guo, L. Wang, Y. Gao, R. Li, X. Huang, P. Li, Z. Qi, Loci-specific phase separation of FET fusion oncoproteins promotes gene transcription. *Nature Communications*. 12, 1491 (2021).

N. Riggi, B. Knoechel, S. M. Gillespie, E. Rheinbay, G. Boulay, M. L. Suvà, N. E. Rossetti, W. E. Boonseng, O. Oksuz, E. B. Cook, A. Formey, A. Patel, M. Gymrek, V. Thapar, V. Deshpande, D. T. Ting, F. J. Hornicek, G. P. Nielsen, I. Stamenkovic, M. J. Aryee, M. N. Rivera, EWS-FLI1 Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. *Cancer Cell*. 39, 114–131 (2021).

Riggi, N., & Stamenkovic, I. The biology of Ewing sarcoma. *Cancer Letters, 254*(1), 1-10. (2007).

Hyman, A. A., Weber, C. A., & Jülicher, F. Liquid-liquid phase separation in biology. *Annual Review of Cell and Developmental Biology, 30*, 39-58. (2014).

Patel, A., Malinovska, L., Saha, S., et al. ATP as a biological hydrotrope. *Science,*

356(6337), 753-756. (2015)

J. N. Wu, C. W. Roberts, ARID1A mutations in cancer: another epigenetic tumor suppressor? Cancer discovery. 3, 35–43 (2013).

J. Prilusky, C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman, J. L. Sussman, FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics. 21, 3435–3438 (2005).

A. Bolognesi, A. Sliwa-Gonzalez, R. Prasad, Y. Barral, Fluorescence recovery after photo-bleaching (FRAP) and fluorescence loss in photo-bleaching (FLIP) experiments to study protein dynamics during budding yeast cell division. Methods Mol. Biol.1369, 25–44 (2016).

P. Rawat, M. Boehning, B. Hummel, F. Aprile-Garcia, A. S. Pandit, N. Eisenhardt, A. Khavaran, E. Niskanen, S. M. Vos, J. J. Palvimo, Stress-induced nuclear condensation of NELF drives transcriptional downregulation. Molecular cell. 81, 1013–1026 (2021).

M. Yu, Z. Peng, M. Qin, Y. Liu, J. Wang, C. Zhang, J. Lin, T. Dong, L. Wang, S. Li, Interferon-γ induces tumor resistance to anti-PD-1 immunotherapy by promoting YAP phase separation. Molecular cell. 81, 1216–1230 (2021).

S. He, Z. Wu, Y. Tian, Z. Yu, J. Yu, X. Wang, J. Li, B. Liu, Y. Xu, Structure of nucleosome-bound human BAF complex. Science. 367, 875–881 (2020).

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. Nature Reviews Molecular Cell Biology, 18(5), 285-298. (2017)

D. P. Nusinow, J. Szpyt, M. Ghandi, C. M. Rose, E. R. McDonald III, M. Kalocsay, J. Jané-Valbuena, E. Gelfand, D. K. Schweppe, M. Jedrychowski, Quantitative proteomics of the cancer cell line encyclopedia. *Cell*. **180**, 387–402 (2020).

Gangwal, K., Sankar, S., Hollenhorst, P., Kinsey, M., Haroldsen, S., Shah, A., Boucher, K., Watkins, W., Jorde, Lynn., Graves, B., Lessnick,. "Microsatellites as EWS/FLI response elements in Ewing's sarcoma." *Proceedings of the National Academy of Sciences*, **105,** 10149–10154 (2008).

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

B. Adane, G. Alexe, B. K. A. Seong, D. Lu, E. E. Hwang, D. Hnisz, C. A. Lareau, L. Ross, S. Lin, F. S. D. Cruz, STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer cell*. **39**, 827–844 (2021).

Orth, M. F., Surdez, D., Faehling, T., Ehlers, A. C., Marchetto, A., Grossetete, S., Volckmann, R., Zwijnenburg, D. A., Gerke, J. S., Zaidi, S., Alonso, J., Sastre, A., Baulande, S., Sill, M., Cidre-Aranaz, F., Ohmura, S., Kirchner, T., Hauck, S. M., Reischl, E., Grunewald, T. G. P. Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation. *Cell Rep*, **41**, 111761 (2022)

I. Jung, A. Schmitt, Y. Diao, A. J. Lee, T. Liu, D. Yang, C. Tan, J. Eom, M. Chan, S. Chee, A compendium of promoter-centered long-range chromatin interactions in the human

genome. *Nature genetics*. **51**, 1442–1449 (2019).

Riggi, N., Suva, M. L., De Vito, C., Provero, P., Stehle, J. C., Baumer, K., Cironi, L., Janiszewska, M., Petricevic, T., Suva, D., Tercier, S., Joseph, J. M., Guillou, L., & Stamenkovic, I. 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).

I. A. Showpnil, J. Selich-Anderson, C. Taslim, M. A. Boone, J. C. Crow, E. R. Theisen, S. L. Lessnick, EWS/FLI mediated reprogramming of 3D chromatin promotes an altered transcriptional state in Ewing sarcoma. *Nucleic Acids Research*. **50**, 9814–9837 (2022).

R. Sanalkumar, R. Dong, L. Lee, Y.-H. Xing, S. Iyer, I. Letovanec, S. La Rosa, G. Finzi, E. Musolino, R. Papait, I. Chebib, G. P. Nielsen, R. Renella, G. M. Cote, E. Choy, M. Aryee, K. Stegmaier, I. Stamenkovic, M. N. Rivera, N. Riggi, Highly connected 3D chromatin networks established by an oncogenic fusion protein shape tumor cell identity. *Sci. Adv.* **9**, eabo3789 (2023).

S. Chong, C. Dugast-Darzacq, Z. Liu, P. Dong, G. M. Dailey, C. Cattoglio, A. Heckert, S. Banala, L. Lavis, X. Darzacq, R. Tjian, Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*. **361**, eaar2555 (2018).

S. Chong, T. G. Graham, C. Dugast-Darzacq, G. M. Dailey, X. Darzacq, R. Tjian, Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic

transcription. *Molecular Cell*. **82**, 2084–2097 (2022).

L. Zuo, G. Zhang, M. Massett, J. Cheng, Z. Guo, L. Wang, Y. Gao, R. Li, X. Huang, P. Li, Loci-specific phase separation of FET fusion oncoproteins promotes gene transcription. *Nature communications*. **12**, 1491 (2021).

Von Levetzow, C., Jiang, X., Gwye, Y., von Levetzow, G., Hung, L., Cooper, A., Hsu, J. H., & Lawlor, E. R. Modeling initiation of Ewing sarcoma in human neural crest cells. *PLoS One*, **6**, e19305 (2011).

Marchetto, A., Ohmura, S., Orth, M. F., Knott, M. M. L., Colombo, M. V., Arrigoni, C., Bardinet, V., Saucier, D., Wehweck, F. S., Li, J., Stein, S., Gerke, J. S., Baldauf, M. C., Musa, J., Dallmayer, M., Romero-Perez, L., Holting, T. L. B., Amatruda, J. F., Cossarizza, A., Grunewald, T. G. P. Oncogenic hijacking of a developmental transcription factor evokes vulnerability toward oxidative stress in Ewing sarcoma. *Nat Commun*, **11**, 2423 (2020).

S. P. Somasekharan, N. Saxena, F. Zhang, E. Beraldi, J. N. Huang, C. Gentle, L. Fazli, M. Thi, P. H. Sorensen, M. Gleave, Regulation of AR mRNA translation in response to acute AR pathway inhibition. *Nucleic acids research*. **50**, 1069–1091 (2022).

Y. Fujioka, J. M. Alam, D. Noshiro, K. Mouri, T. Ando, Y. Okada, A. I. May, R. L. Knorr, K. Suzuki, Y. Ohsumi, Phase separation organizes the site of autophagosome formation. *Nature*. **578**, 301–305 (2020).

X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, R. D. Vale, Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*. **352**, 595–599 (2016).

Y. P. Hung, A. Redig, J. L. Hornick, L. M. Sholl, ARID1A mutations and expression loss in non-small cell lung carcinomas: clinicopathologic and molecular analysis. *Modern Pathology*. **33**, 2256–2268 (2020).

W. Chan-On, M.-L. Nairismägi, C. K. Ong, W. K. Lim, S. Dima, C. Pairojkul, K. H. Lim, J. R. McPherson, I. Cutcutache, H. L. Heng, Exome sequencing identifies distinct mutational patterns in liver fluke–related and non-infection-related bile duct cancers. *Nature genetics*. **45**, 1474–1478 (2013).

Boulay, G., Sandoval, G. J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M. E., Rengarajan, S., Volorio, A., McBride, M. J., Broye, L. C., Zou, L., Stamenkovic, I., Kadoch, C., & Rivera, M. N. Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell*, **171**, 163-178 e119 (2017).

G. J. Sandoval, J. L. Pulice, H. Pakula, M. Schenone, D. Y. Takeda, M. Pop, G. Boulay, K. E. Williamson, M. J. McBride, J. Pan, Binding of TMPRSS2-ERG to BAF chromatin remodeling complexes mediates prostate oncogenesis. *Molecular cell*. **71**, 554–566 (2018).

S. S. Riedel, C. Lu, H. M. Xie, K. Nestler, M. W. Vermunt, A. Lenard, L. Bennett, N. A. Speck, I. Hanamura, J. A. Lessard, Intrinsically disordered Meningioma-1 stabilizes the

BAF complex to cause AML. *Molecular cell*. **81**, 2332–2348 (2021).

M. Lindén, C. Thomsen, P. Grundevik, E. Jonasson, D. Andersson, R. Runnberg, S. Dolatabadi, C. Vannas, M. Luna Santamaría, H. Fagman, A. Ståhlberg, P. Åman, FET family fusion oncoproteins target the SWI / SNF chromatin remodeling complex. *EMBO Reports*. **20**, e45766 (2019).

Z. Nie, Z. Yan, E. H. Chen, S. Sechi, C. Ling, S. Zhou, Y. Xue, D. Yang, D. Murray, E. Kanakubo, M. L. Cleary, W. Wang, Novel SWI/SNF Chromatin-Remodeling Complexes Contain a Mixed-Lineage Leukemia Chromosomal Translocation Partner. *Molecular and Cellular Biology*. **23**, 2942–2952 (2003).C. Guo, Z. Che, J. Yue, P. Xie, S. Hao, W. Xie, Z. Luo, C. Lin, ENL initiates multivalent phase separation of the super elongation complex (SEC) in controlling rapid transcriptional activation. *Sci. Adv.* **6**, eaay4858 (2020).

M. L. Harlow, M. H. Chasse, E. A. Boguslawski, K. M. Sorensen, J. M. Gedminas, S. M. Kitchen-Goosen, S. B. Rothbart, C. Taslim, S. L. Lessnick, A. S. Peck, Trabectedin inhibits EWS-FLI1 and evicts SWI/SNF from chromatin in a schedule-dependent manner. *Clinical cancer research*. **25**, 3417–3429 (2019).

H. V. Erkizan, Y. Kong, M. Merchant, S. Schlottmann, J. S. Barber-Rotenberg, L. Yuan, O. D. Abaan, T. Chou, S. Dakshanamurthy, M. L. Brown, A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. *Nature medicine*. **15**, 750–756 (2009).

C. A. Marian, M. Stoszko, L. Wang, M. W. Leighty, E. de Crignis, C. A. Maschinot, J. Gatchalian, B. C. Carter, B. Chowdhury, D. C. Hargreaves, Small molecule targeting of specific BAF (mSWI/SNF) complexes for HIV latency reversal. *Cell chemical biology*. **25**, 1443–1455 (2018).

A. Guo, H. Huang, Z. Zhu, M. J. Chen, H. Shi, S. Yuan, P. Sharma, J. P. Connelly, S. Liedmann, Y. Dhungana, cBAF complex components and MYC cooperate early in CD8+ T cell fate. *Nature*. **607**, 135–141 (2022).

E. J. Chory, J. G. Kirkland, C.-Y. Chang, V. D. D'Andrea, S. Gourisankar, E. C. Dykhuizen, G. R. Crabtree, Chemical Inhibitors of a Selective SWI/SNF Function Synergize with ATR Inhibition in Cancer Cell Killing. *ACS Chem. Biol.* **15**, 1685–1696 (2020).

A. Patil, A. R. Strom, J. A. Paulo, C. K. Clayton, K.M. Ruff, M. K. Shinn, A. Sankar, K. S. Cervantes, T. Wauer, J. D. Laurent, F. Xu, L. A. Becker, S. P. Gygi, R. V. Pappu, C. P. Brangwynne, C. Kadoch. A disordered region controls CBAF activity via condensation and partner recruitment. *Cell*. **186**, 4936-4955 (2023).

B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, R. A. Young, Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. **361**, eaar3958 (2018).

A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. **12**, 323 (2011).

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

B. T. Sherman, M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi, W. Chang, DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research*. **50**, W216–W221 (2022).

A. Kechin, U. Boyarskikh, A. Kel, M. Filipenko, cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *Journal of Computational Biology*. **24**, 1138–1143 (2017).

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

F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research*. **44**, W160 (2016).

S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. **38**, 576–589 (2010).

H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*. **25**, 1754–1760 (2009).

K. Kim, I. Jung, covNorm: An R package for coverage based normalization of Hi-C and capture Hi-C data. *Computational and Structural Biotechnology Journal*. **19**, 3149–3159 (2021).

M. E. Ritchie, B. Phipson, D. I. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. **43**, e47–e47 (2015).

국문 초록 / ABSTRACT IN KOREAN

액체-액체 상분리 (LLPS) 는 세포 내에서 막이 없는 소기관의 형성을 촉진하며 다양한 생물학적 현상과 질병 상태에 중요한 역할을 한다. 본 연구에서는 ARID1A (AT-rich interactive domain-containing protein 1A) 가 프리온 유사 도메인 (PrLD)을 포함하고 있으며, 이 도메인이 LLPS 를 매개하여 액체 응집체를 형성한다는 사실을 발견하였다. 유잉 육종 환자의 조직 샘플에서 ARID1A 의 LLPS 에 의한 핵 내 응집체가 크게 증가되어 있음을 확인하였다. ARID1A LLPS 를 저해하면 Ewing's 육종 세포의 증식 및 침습 능력이 현저히 감소되었다. 유전체 수준의 크로마틴 구조 분석 및 전사체 분석을 통해 ARID1A 응집체가 EWS/FLI1 타겟 인핸서에 위치하며, 암촉진 타겟 유전자에서 기능적 크로마틴 리모델링 허브를 형성하여 장거리 크로마틴 구조를 변화시키는 것을 확인하였다. 종합하면, 본 연구는 ARID1A 가 PrLD-매개 LLPS 를 통해 암 유발 능력을 촉진한다는 것을 증명하였으며, 이는 유잉 육종 치료를 위한 잠재적 치료 전략을 제시한다.

주요어:
액체-액체 상분리 (LLPS), ARID1A (AT-rich interactive domain-containing protein 1A), 프리온 유사 도메인 (PrLD), EWS/FLI1, 크로마틴 리모델링, 유잉 육종, 암촉진 타겟 유전자

학번: 2018-31266