Genetic and chemical hydrophobic modification of transcription factor condensates in human disease

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the

Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Shaon Basu

2022

The work was carried out between 07/2018 and 07/2022 under the supervision of Dr. Denes Hnisz at the Max Planck Institute for Molecular Genetics in Berlin, Germany.

Reviewer 1:

Dr. Denes Hnisz

Reviewer 2:

Prof. Dr. Helge Ewers

Date of defense: December 15th, 2022

Acknowledgements

I would like to thank Dr. Denes Hnisz, my primary supervisor, as well as Dr. Xavier Salvatella, Prof. Dr. Helge Ewers, and Prof. Dr. Alexander Meissner for giving me the opportunity to contribute to exciting science, and for the advice and assistance along the way.

I would like to thank my parents, my sister, and Jana Schulz for the unconditional support during the years of my PhD.

お母さん、お父さん、とシューちゃんの応援のおかげで卒業する事が出来ました。いつも側にいてくれてありがとう。これからも一生懸命頑張ります。

Finally, I would like to thank all the members of the Max Planck Institute for Molecular Genetics: I learned a lot and was inspired by each and every one of you.

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

I also declare that I have not submitted the dissertation in this or any other form to any other institution as a dissertation.

Table of contents

Summary

The work in this thesis examines how hydrophobic modifications within intrinsically disordered regions (IDRs) of transcription factors (TFs), caused by genetic mutations and drug-induced changes, disrupt TF condensation and gene expression programs in various human disease states. I establish as a general paradigm that disease-associated poly-alanine repeat expansions promote the formation of homotypic TF condensates with gel-like material properties, at the expense of heterotypic interactions with other components of the transcriptional apparatus. This effect was elicited by hydrophobic repeat expansions found in the following developmental disorders: synpolydactyly (SPD), cleidocranial dysplasia (CCD), and hand-foot-genital-syndrome (HFGS). Utilizing insights from this paradigm, I hypothesized that the condensation capacity of TFs may be controlled with chemical modifiers of TF-IDRs, in efforts to treat diseases reliant on aberrant transcriptional programs, such as cancer[1]. To this end, this thesis provides evidence that the condensation of an oncogenic transcription factor, the androgen receptor (AR), can be targeted with small molecules that selectively partition into condensates formed by the disordered activation domain of the AR. Increasing the hydrophobicity of the small molecule resulted in higher potency in the arrest of proliferation and AR-driven gene expression programs in a human model of prostate cancer (PCa). Together, these results suggest that hydrophobic modification of transcription factors and small molecules that partition into condensates can be leveraged for therapeutic intent.

Zusammenfassung

In dieser Dissertation wurde untersucht, wie hydrophobe Modifikationen innerhalb von intrinsisch ungeordneten Regionen (IDRs) von Transkriptionsfaktoren (TFs), verursacht durch genetische Mutationen und arzneimittelinduzierte Veränderungen, die TF-Kondensation und Genexpressionsprogramme in verschiedenen menschlichen Krankheitsstadien stören. Dadurch konnte etabliert werden, dass krankheitsassoziierte Poly-Alanin-Expansionen die Bildung von homotypischen TF-Kondensaten mit gelartigen Materialeigenschaften fördern, auf Kosten heterotypischer Wechselwirkungen mit anderen Komponenten des Transkriptionsapparates. Dieser Effekt wurde in folgenden Entwicklungsstörungen charakterisiert: Synpolydaktylie (SPD), cleidocraniale Dysplasie (CCD) und Hand-Fuß-Genital-Syndrom (HFGS). Basierend auf diesen Erkenntnissen folgte die Hypothese, dass die Kondensationskapazität von TFs mit chemischen Modifikationen von TF-IDRs kontrolliert werden kann, um Krankheiten wie z.B. Krebs zu behandeln, die auf abweichende Transkriptionsprogramme angewiesen sind[1]. In diesem Zusammenhang konnte ich in meiner Dissertation belegen, dass die Kondensation eines onkogenen Transkriptionsfaktors, des Androgenrezeptors (AR), mit kleinen Molekülen moduliert werden kann, die sich selektiv in Kondensate integrieren, die von der ungeordneten Aktivierungsdomäne des AR gebildet werden. Eine Erhöhung der Hydrophobie dieses Moleküls führte zu einer stärkeren Hemmung von Proliferations- und AR-gesteuerten Genexpressionsprogrammen, welche in einem menschlichen Prostatakrebsmodell (PCa) nachgewiesen wurde. Zusammengenommen legen diese Ergebnisse nahe, dass die hydrophobe Modifikation von Transkriptionsfaktoren und kleine Moleküle, die sich in Kondensate integrieren, für therapeutische Zwecke genutzt werden kann.

1 Introduction

1.1 Transcriptional condesate model

Components of the transcriptional apparatus can form biomolecular condensates implicated in gene control[2–8], yet the physicochemical properties that govern this process are poorly validated in the context of transcription[9–12]. Biomolecular condensates are sub cellular assemblies that lack membranes, and can enrich for biomolecules with similar cellular functionalities. Many studies provide evidence that a wide range of cellular processes, such as DNA repair, ribosome biogenesis, synaptic transmission, and others involve condensates[10]. These biological processes have been successfully dissected with biochemical and cellular assays with components of well-known membrane-less organelles, such as stress granules[13], P bodies[14], nucleoli[15,16], and neuronal granules[17]. However, due to the nanoscopic[18] and transient nature of condensates formed by transcriptional components[4], there is a dearth of data that characterizes the physicochemical properties that govern the formation and dissolution of transcriptional condensates in the context of genetic activity[19].

Weak multivalent interactions between biomolecules are thought to drive the formation of condensates through a phase separation process, reminiscent to oil droplets separating from water[20–22], and provide a layer of organization that segregates biomolecules within a cell into compartments with specific functionalities[17,23–26]. Intrinsically disordered regions (IDRs) have been shown to facilitate the phase separation of biomolecules[27,28], and transcription factors (TFs) exhibit significant enrichment of IDRs over the rest of the human proteome (up to 94% of all TFs contain an IDR)[29]. IDRs, defined by low complexity amino acid sequences that lack defined structure, harbor functional group combinations that give rise to regions with different chemical properties such as a mixture of different ionic strengths, electron-rich π systems, and hydrophobic regions[8–10]. Understanding how these regions with different chemical features interact to give rise to phase separation, may provide insights to how transcription factors organize themselves to form condensates implicated in their function.

1.2 Relevance to disease genetics

Intriguingly, over 30 inherited disorders are caused by insertion of repetitive DNA sequence elements, most of which (> 20) give rise to disordered repeat expansions in protein coding genes[30]. The majority (15/20) of disease-associated repeat expansions occur in disordered regions of nuclear proteins, many of which are sequence specific to TF-IDRs (**Figure 1A**). Of the catalogued repeat expansions encoded in TFs, poly-alanine repeat expansions require the fewest codons (minimum of 7) to elicit severe developmental malformations in humans[30–32]. As poly-alanine repeat tracts encode for hydrophobic chemical regions, understanding how such repeats affect the physicochemical properties of transcription factor condensates would elucidate a facet of the chemical grammar of biomolecular condensation, and provide insight towards and otherwise enigmatic pathomechanism.

1.3 Relevance to cancer therapeutics

In addition to harboring disease-associated mutations, TF-IDRs are among the most frequently mutated and dysregulated genes in cancer[33], and are potential drug targets for the treatment of cancer[34]. For example, the most frequently mutated and overexpressed genes in cancer (TP53 and MYC respectively) both encode for transcription factors[35]. Despite the substantial effort directed at developing anti-cancer drugs against them, TFs are considered largely 'undruggable.' As TF activation domains tend to be intrinsically disordered, they are intractable for structure-based drug development programs that rely on "lock-and-key" mechanisms[34,36]. Intriguingly, even in cancers that are treatable with small molecules, TF IDRs pose challenges that ultimately lead to drug resistance.

Prostate cancer is the second most prevalent cancer and is the second leading cause of cancer death in the developed world[37,38]. The androgen receptor (AR) is a transcription factor that plays critical roles in controlling gene transcription in response to steroid hormones, and is a major oncogenic driver in prostate cancer[37,38]. The AR consists of an intrinsically disordered N-terminal activation domain (AD), a DNA-binding domain (DBD), and a C-terminal ligand binding domain (LBD). Anti-androgens targeting the LBD are the most common first-line therapy in prostate cancer, but 10-20% of patients progress into a castration-resistant stage which is invariably lethal[39,40]. The mean survival of patients with

castration-resistant prostate cancer (CRPC) is less than three years. A predominant cause of CRPC is the emergence of hyperactive splice-variants of the androgen receptor that conserve the intrinsically disordered AD, but lack the structured LBD and are thus insensitive to anti-androgens [41–43]. As such, understanding the properties that drive condensation of intrinsically disordered regions within oncogenic TFs could facilitate the development of drugs for the most lethal cancers.

1.4 Hydrophobicity paradigm

Using assays developed in this thesis I provide evidence for the paradigm that hydrophobic modifications in TF-IDRs enhance homotypic condensation of transcription factors, at the expense of heterotypic interactions with other biomolecules necessary for transcription. This was established by examination of how hydrophobic disease-associated poly-alanine repeat expansions in TF-IDRs, found in a spectrum of human developmental disorders, affect TF condensation. Utilizing insights gained from this paradigm, I hypothesized that an experimental anti-cancer drug that dampens transactivation driven by the AR AD[44], may partition into AR condensates and modify its condensation capacity. In agreement with my hypothesis, the drug displayed the capacity to partition and chemically modify condensates formed by the AR. Strikingly, increasing the hydrophobicity of the drug elicited in more potent arrest of cancer proliferation and AR-dependent oncogenic gene expression programs.

2 Results

Disclaimer: Parts of the presented results have been published by my colleagues and me in Cell with the title "Unblending of transcriptional condensates in human repeat expansion disease"[8] and in bioRxiv with the title "Androgen receptor condensates as drug targets"[45].

2.1 TF-IDRs facilitate phase separation and harbor disease-associated repeat expansions

Repeat expansions encoding for homo-polymeric alanine tracts are enriched in transcription factors (TFs), and are associated with severe developmental diseases (**Figure 2-1 A**). To study whether disease-associated repeat expansions alter the phase separation capacity of the TFs, I focused on the HOXD13, as its disease genetics have been well dissected in-house at the Max Planck Institute for Molecular Genetics. The seminal work of Dr. Stefan Mundlos and Dr. Daniel Ibrahim[32,46–48] revealed that HOXD13 controls skeletal morphogenesis in the mammalian limb bud during embryogenesis[47], and, when mutated, causes the limb malformation known as synpolydactyly (SPD). This congenital disease is characterized by the fusion of digits and partial or complete digit duplication[32]. Confocal imaging revealed that HOXD13 forms discrete nuclear clusters in mammalian limb bud cells and several human cell lines that express the transcription factor (**Figure 2-1 C**). Examination of limb bud cells by direct stochastic optical reconstruction microscopy (dSTORM) revealed that clusters were 100 nm in diameter on average (**Figure 2-1 B**, bottom right), similar in size to previously described, phase separated, nuclear co-activator and RNAPII condensates that occupy large enhancer elements[3,4]. Collectively, these results suggest that HOXD13 forms nuclear condensates implicated in transcriptional control.

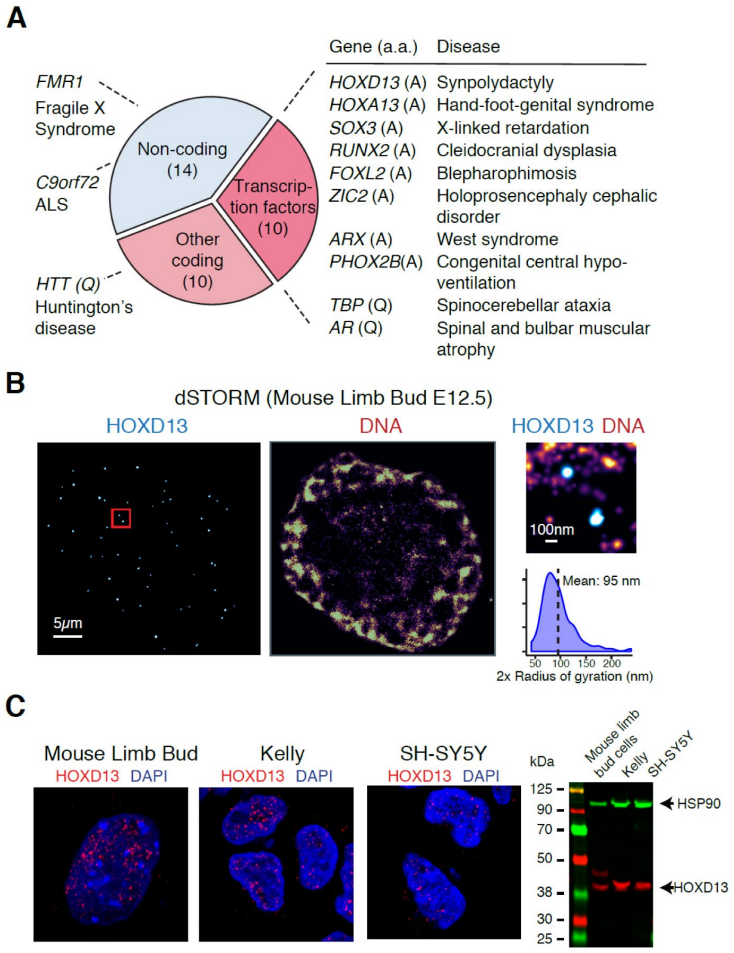

Figure 2-1: The HOXD13 TF harbors disease associated alanine repeat expansions and forms nuclear puncta in mammalian cells. ***A)*** *Catalog of disease-associated repeat expansions in humans. a.a. indicates the amino acid product of the translated trinucleotide repeat codon.* ***B)*** *Stochastic optical reconstruction microscopy (STORM) on HOXD13 in limb bud cells isolated from E12.5 mouse embryo. Right top: zoom inset of HOXD13 puncta. Right bottom: quantification of average diameter (nanometers) of detected HOXD13 puncta in E12.5 limb bud nuclei.* ***C)*** *Left: HOXD13 IF in human cells known to express Hoxd13. Right: Western blot validation of HOXD13 antibody.*

Intrinsically disordered regions (IDRs) in proteins have been identified to facilitate biomolecular condensation by phase separation[25,26], and virtually all TFs contain an IDR[29]. In agreement with this, the N-terminal activation domain of HOXD13 contains an amino-acid sequence predictive of an IDR (**Figure 2-2 A**). To investigate if the HOXD13 IDR facilitates phase separation in cells, I utilized an optogenetic platform (known as the optoDroplet assay)

to investigate whether the HOXD13 IDR can drive phase separation[49]. The assay involves cellular expression of a fusion protein consisting of a target IDR fused to mCherry, and the photosensitive cryptochrome domain of the *Arabidopsis thaliana* CRY2 gene. Excitation of CRY2 with blue light stimulates its self-oligomerization, which leads to localized clustering of the IDR in the fusion cassette. IDRs that drive phase separation facilitate the formation of phase-separated droplets (**Figure 2-2 B**), which do not form in the absence of the IDR[2,49].

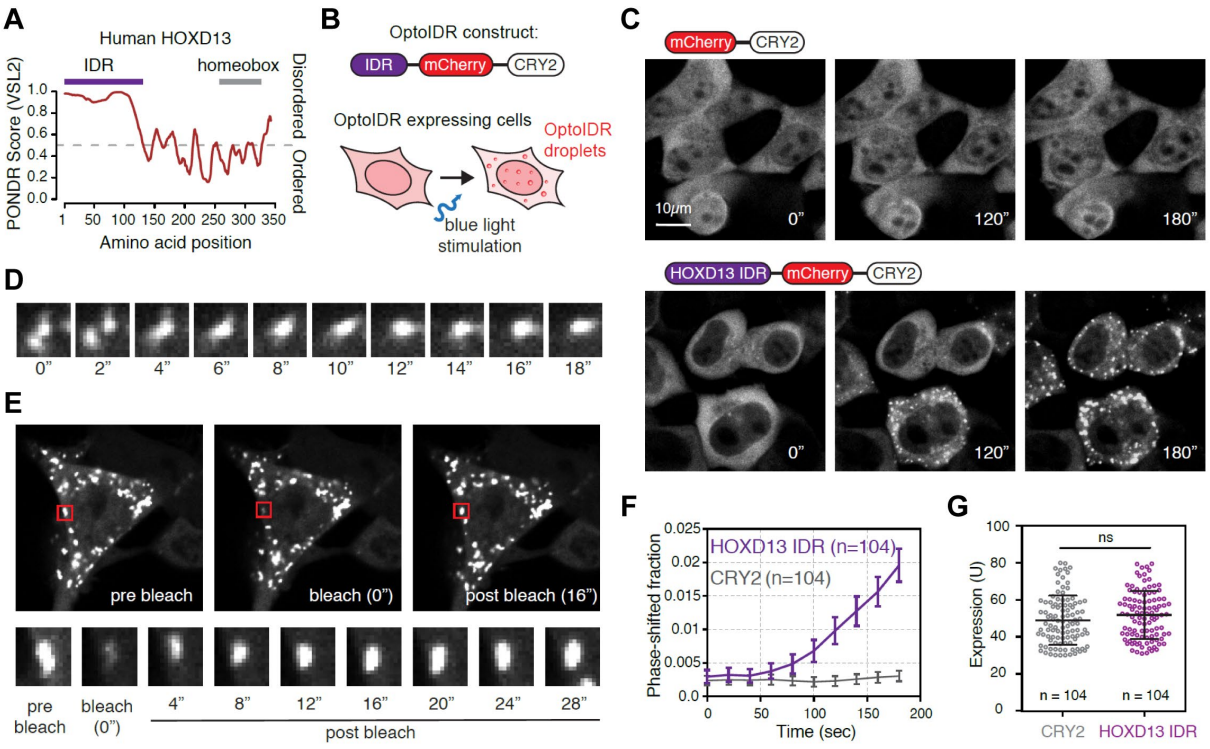

Figure 2-2: HOXD13 TF-IDR facilitates phase separation in the optoDroplet system. *A) PONDR VSL2 prediction of disorder in human HOXD13 amino acid sequence. Purple bar indicates N-terminal intrinsically disordered region (IDR).* ***B)*** *Schematic of optoDroplet system with IDR cassette.* ***C)*** *Time lapse images of optoHOXD13-IDR and control droplet formation in HEK293T cells* ***D)*** *Time lapse images of optoHOXD13-IDR droplet fusion event.* ***E)*** *Time lapse images of florescence recovery after photobleaching (FRAP) of optoHOXD13- IDR droplets.* ***F)*** *Quantification of light induced optoHOXD13-IDR and control droplet formation (phase shifted fraction in methods) as a function of time.* ***G)*** *Expression control showing mCherry signal in cells transfected with optoHOXD13-IDR and control vectors for optoDroplet experiments (n.s. = not significant, Student's t-test).*

The HOXD13-IDR-mCherry-CRY2 fusion (optoHOXD13-IDR) facilitated the formation of droplets upon blue-light stimulation in transiently transfected cells, in contrast to the IDR-lacking control vector (**Figure 2-2 C**). The droplets displayed the capacity to undergo fusion (**Figure 2-2 D**), and fluorescence recovery after photobleaching (FRAP) revealed rapid, liquid-like recovery rate of optoHOXD13-IDR droplets (**Figure 2-2 E**), which are characteristics of liquid-liquid phase-separated (LLPS) condensates[50]. Image-based quantification of cells that expressed similar levels of control and optoHOXD13-IDR vector revealed that the HOXD13 IDR significantly enhanced the condensate formation as a function of time, post-light stimulation (**Figure 2-2 F and G**). Collectively, these results suggest that the HOXD13 IDR has the capacity to drive LLPS in cells.

2.2 Hydrophobic repeat expansions enhance phase separation of the HOXD13 TF-IDR

Hydrophobic and polymorphic alanine repeat expansions in the HOXD13 IDR cause synpolydactyly (**Figure 2-3 A**), a congenital limb malformation in mammals, with disease severity increasing with tract length[32]. Since the alanine repeat expansions occur within the IDR of HOXD13, I hypothesized that the repeat expansions may alter Hoxd13's phase separation capacity. To test this hypothesis, HOXD13 IDRs, including the wild type and several synpolydactyly-associated expansions (+7A, +8A, +9A, +14A), were examined in the optoDroplet system. These experiments were conducted with an additional nuclear localization sequence fusion (NLS, plasmid generation in methods), to ensure optoDroplet constructs were expressed in the cell nucleus. Short expansions greatly enhanced the rate of light-induced droplet formation in cell nuclei that expressed similar levels of the tested construct (**Figure 2-3 B-D**). Longer expansions (+8A, +9A, and +14A) promoted the formation of spontaneous spherical condensates, which did not require light-triggered oligomerization of CRY2 to form (**Figure 2-3 B**, lower panels). These results suggest that synpolydactyly associated hydrophobic alanine repeat expansions enhance the phase-separation capacity of the HOXD13-IDR.

Phase-separated condensates may exhibit gel-like material properties, characterized by reduced internal kinetics[49]. Consistent with these reports, light-induced droplets formed by repeat-expanded optoHOXD13-IDRs (+7A, +8A) exhibited considerably slower FRAP rates than droplets formed by the wild-type IDR (**Figure 2-3 E**). Spontaneous condensates formed by longer repeat expansions (+8A, +9A, and +14A) also exhibited even slower FRAP rates (**Figure 2-3 F**). These results suggest that the alanine repeat expansion alters the material properties of optoHOXD13-IDR droplets, and promote gel-like LLPS, with arrested dynamics that directly correlate to the length of the repeat tract.

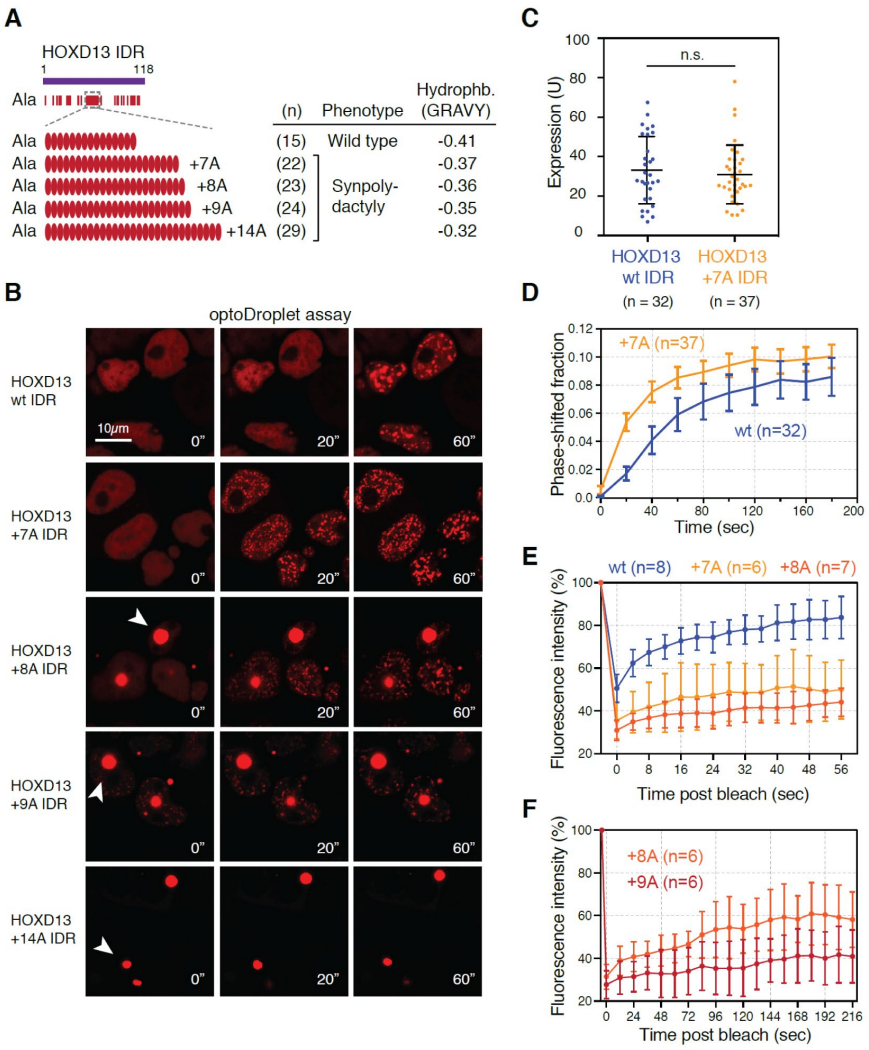

Figure 2-3 Hydrophobic repeat expansions enhance phase separation of the HOXD13 TF-IDR. ***A)*** *Alanine composition within HOXD13 IDR aligned to synpolydactyly associated hydrophobic alanine repeat expansions found in humans.* ***B)*** *Time lapse images of repeat expansion (+7A, +8A, +9A, and +14A) harboring HOXD13 IDR constructs in optoDroplet system (lower panels) vs. wild type optoHOXD13-IDR control (top panels).* ***C)*** *Expression control showing mCherry signal in cells transfected with optoHOXD13-IDR(WT) and optoHOXD13-IDR(+7A) vectors for optoDroplet experiments. (n.s. = not significant, Student's t-test).* ***D)*** *Quantification of light induced optoHOXD13-IDR(WT) and optoHOXD13-IDR(+7A) droplet formation, measured by nuclear phase shifted fraction (methods) as a function of time.* ***E)*** *FRAP signal quantification of light induced optoHOXD13-IDR(WT), optoHOXD13- IDR(+7A), and optoHOXD13-IDR(+8) droplets as a function of time.* ***F)*** *FRAP signal quantification of spontaneous optoHOXD13-IDR(+8A) and optoHOXD13- IDR(+9A) droplets as a function of time.*

To further probe the effect of alanine repeat expansions on HOXD13 phase separation, I purified various HOXD13 IDR-mCherry fusion proteins and investigated their phase separation capacity in vitro (methods). In agreement with the optoDroplet assay results, the +7A and +10A IDR mutants formed more numerous and concentrated droplets in comparison to the wt IDR, at every assayed concentration (**Figure 2-4 A and B**). Additionally, the concentrations at which droplets formed, also known as the saturation concentration (C_{sat}), were progressively lower for IDRs with longer alanine tract lengths (**Figure 2-4 A and B**). Collectively, these results indicate that hydrophobic poly-alanine repeats enhance phase separation of the HOXD13 IDR *in vitro* and in cells, in agreement with the reported correlation between repeat length and disease severity[32,48,51].

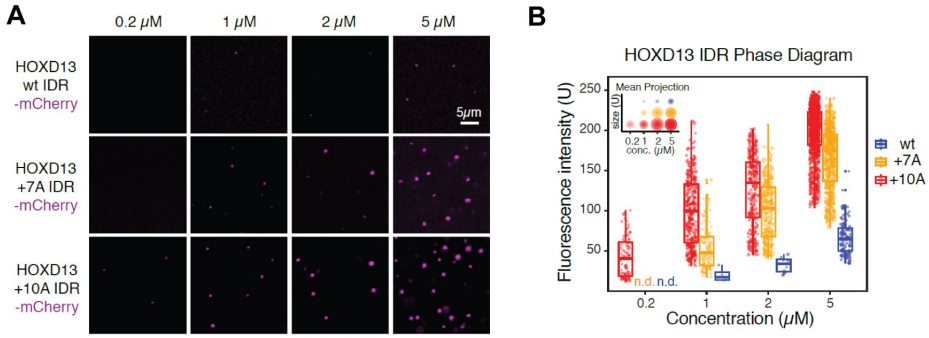

Figure 2-4: Hydrophobic repeat expansions enhance phase separation of the HOXD13 TF-IDR in vitro. ***A)*** *Representative images of purified wildtype and synpolydactyly associated HOXD13 IDR mCherry fusion proteins (WT, +7A, +10A) in droplet formation buffer (methods).* ***B)*** *Phase diagram of HOXD13 IDR-mCherry fusion proteins, at each assayed concentration shown in panel A. Every dot represents a droplet. Inset: projected average size of the droplets as mean ± SD (middle circle, mean; inner and outer, SD). (n.d. = no droplets).*

2.3 Hydrophobic repeat expansions alter heterotypic condensation of HOXD13 TF-IDR in vitro

TF-IDRs were recently shown to engage in heterotypic interactions with the IDR of MED1 subunit of the Mediator co-activator in vitro and in cells[3], a eukaryotic multiprotein complex that operates as a transcriptional co-activator[52]. In light of this finding, I tested whether alanine repeat expansions would alter the ability of HOXD13-IDR to condense with MED1-IDR. To test this, I purified recombinant MED1-IDR-GFP fusion protein and pre-assembled droplets for 30 minutes at room temperature. Preassembled MED1-IDR droplets were then mixed with HOXD13-IDR-mCherry fusion proteins with different alanine tract lengths. Wild-type HOXD13-IDR exclusively enriched into MED-IDR condensates at every tested concentration (**Figure 2-5 A**, top two rows). Intriguingly, though mutant HOXD13-IDRs (+7A and +10A) displayed the ability to enrich into MED-IDR condensates, they also formed homotypic condensates lacking any MED1-IDR (**Figure 2-5 A and B**, white arrows, GFP-signal distribution). The size and concentration of HOXD13-IDR homotypic condensates lacking MED-IDR increased with the length of the mutant alanine tract (white circles, **Figure 2-5 B**). These results suggest that the alanine repeat expansions perturb the molecular composition of HOXD13 IDR condensates *in vitro*.

The mixing experiments with hydrophobic repeat-expanded HOXD13 IDRs suggest that the alanine tracts increase homotypic interactions at the expense of heterotypic interactions with MED1-IDR. Therefore, I hypothesized that disrupting weak hydrophobic interactions could 'rescue' the composition of repeat-expanded HOXD13-IDR condensates. To test this hypothesis, I treated HOXD13-IDR and MED1-IDR condensate mixtures with ATP, a hydrotrope known to solubilize hydrophobic molecules and disrupt biomolecular condensates at millimolar concentrations[53].

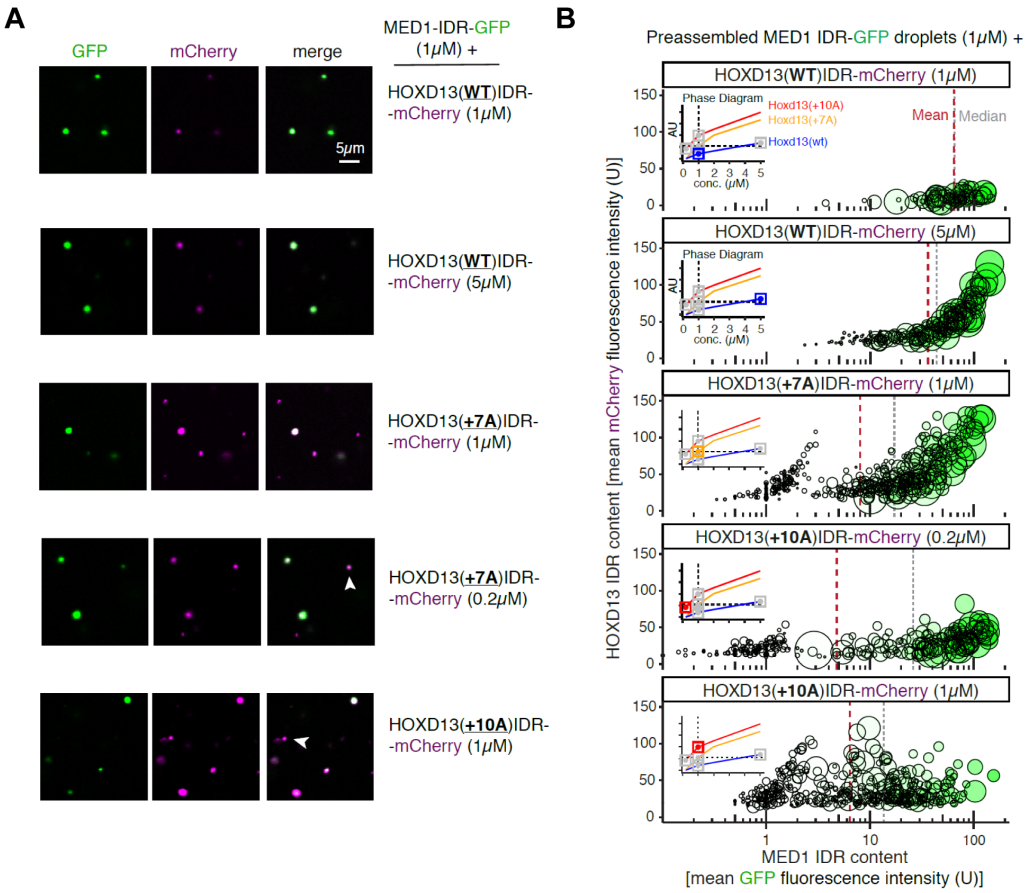

Figure 2-5: Hydrophobic repeat expansions alter heterotypic condensation of HOXD13 in vitro. A) *Representative images of purified wildtype or synpolydactyly associated HOXD13 IDR mCherry fusion proteins (WT, +7A, +10A) and preassembled MED1-IDR-GFP in droplet formation buffer (methods).* ***B)*** *Dual fluorescence projection of GFP and mCherry fluorescence intensities in HOXD13 containing droplets from the mixtures displayed in panel A. Each circle represents one droplet, with size scaling in proportion to the area of the droplet, and green shading scaling to MED1 within the droplet. x-axis is* log_{10} *transformed.*

ATP 'rescued' mutant HOXD13-IDR condensates in a dose-dependent manner (**Figure 2-6 A-C**), and promoted mixing with MED1-IDR. The effect appeared specific to ATP, as other molecules reported to perturb condensates formed by stress granule proteins (lipoic acid, lipoamide, and mitoxantrone[54]) failed to 'rescue' mutant condensates (**Figure 2-6 D**). Taken together, these results suggest that hydrophobic repeat expansion leads to a decrease in the ability of HOXD13-IDR condensates to interact with MED1-IDR in vitro (**Figure 2-6 E**).

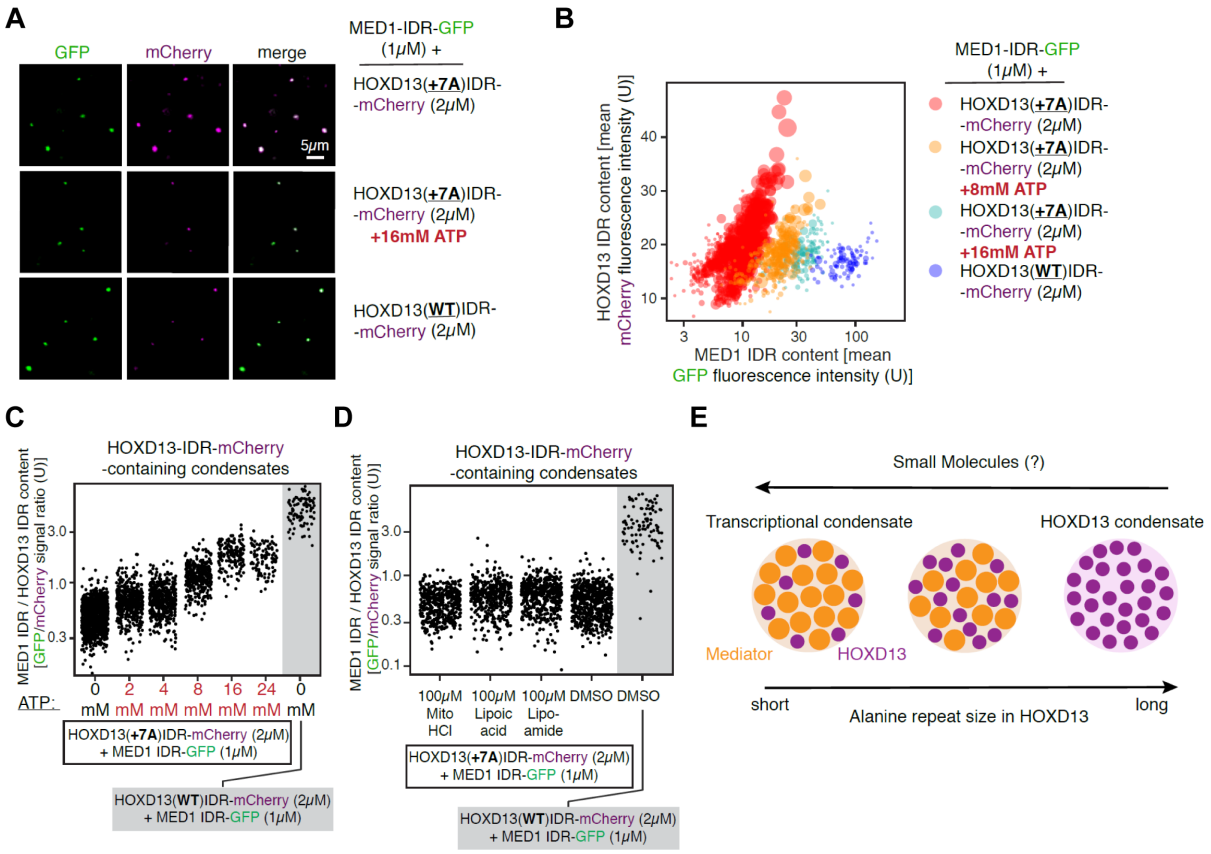

Figure 2-6: Hydrotrope alters heterotypic condensation of HOXD13 in vitro. *A) Representative images of purified wildtype or synpolydactyly associated HOXD13 IDR mCherry fusion proteins (WT, +7A) and MED1-IDR-GFP in the presence or absence of 16 mM ATP.* ***B)*** *Dual fluorescence projection of GFP and mCherry fluorescence intensities in HOXD13 containing droplets in the presence or absence of ATP (representative images displayed in panel A). Each circle represents one droplet, with size scaling in proportion to area of the droplet. x-axis is log_{10} transformed. Colors indicate composition of mixture, indicated in right.* ***C)*** *Quantification of the ratio of MED1 signal (GFP intensity) and HOXD13 signal (mCherry intensity) within HOXD13 droplets from mixtures with indicated titrations of ATP.* ***D)*** *Quantification of the ratio of MED1 signal (GFP intensity) and HOXD13 signal (mCherry intensity) within HOXD13 droplets from mixtures in presence of 100 μM of indicated small molecule or DMSO control at equivalent w/v.* ***E)*** *Model schematic of how hydrophobic repeat expansion may perturb the composition of HOXD13 TF containing condensates.*

2.4 Hydrophobic repeat expansions alter heterotypic interactions and material properties of Hoxd13 condensates in a mouse model of synpolydactyly

To examine the effect of alanine repeat expansions on HOXD13 condensate composition *in vivo*, limb bud cells were harvested from synpolydactyly (spdh) and wild-type (wt) mouse embryos. Spdh mice are homozygous for +7A repeat-expanded *Hoxd13* alleles, and exhibit synpolydactyly[55]. Isolated wt and spdh limb bud cells were then interrogated with dSTORM imaging, H3K27Ac ChIP-Seq, and 1,6-hexanediol treatment followed by HOXD13 immunofluorescence staining (**Figure 2-7 A**).

First, I assessed if transcriptional co-activators are associated with HOXD13-containing condensates in limb bud cells using antibody staining and dSTORM imaging (**Figure 2-7 B and C**). BRD4 is a eukaryotic transcriptional and epigenetic regulator that interacts and condenses with the Mediator complex[2,56], and chemical inhibition of BRD4 dissolves Mediator condensates *in vivo*[4]. dSTORM revealed that BRD4 more frequently overlapped with HOXD13 clusters in wild-type limb bud cells than in spdh limb cells (p-value<0.001, two-tailed t-test, **Figure 2-7 B**). Heterochromatin-associated protein, HP1α[57], was also imaged with HOXD13 as negative control. As expected, the two proteins formed clusters with negligible overlap (**Figure 2-7 D**), suggesting that HOXD13 clusters have altered heterotypic interactions *in vivo.*

MED1 and BRD4 are used to define 'super-enhancers,' or large enhancer elements associated with cell-type specifying genes associated with a high density of co-activators and gene activity[58]. BRD4 co-activator is known to associate with H3K27Ac-enriched enhancer regions via its bromodomain[59], and H3K27Ac is a known histone mark often used to 'foot print' super enhancers via ChIP-sequencing[60]. H3K27Ac ChIP-seq on wt and spdh limb bud cells revealed an overall lower level of enhancer-associated H3K27 acetylation at Hoxd13 binding sites and at marker genes involved in ossification of the inter-digital mesenchyme during embryogenesis (i.e., *Msx2*), in spdh limb bud cells (**Figure 2-7 D**). These results suggest that disease-associated repeat expansions reduce H3K27 acetylation at HOXD13 binding sites.

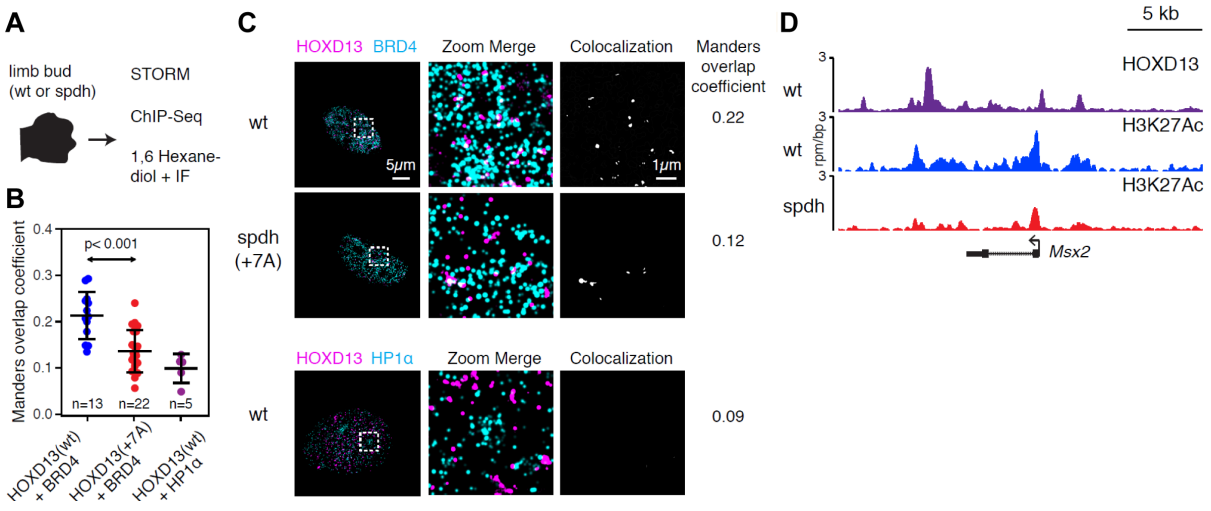

Figure 2-7: Hydrophobic repeat expansions alter heterotypic interactions of Hoxd13 in mouse model of synpolydactyly. A) *Left: Experimental schematic using E12.5 limb bud cells acquired from wt and spdh (Hoxd13 +7A) mice.* ***B)*** *Quantification of STORM composites in (representative images shown in panel C) as Manders overlap coefficients (MOC) of STORM co-localizations. Each dot indicates MOC calculated from localization data collected from a single nucleus. P-value from Student's t-test.* ***C)*** *Representative Co-IF STORM images of nuclei from wt or spdh limb bud cells, with zoom in highlighted with outlined insets, and corresponding MOC calculated from the representative image.* ***D)*** *Hoxd13 and H3K27Ac ChIP-Seq binding profiles at the osteogenic differentiation marker Msx2 locus in wt and spdh mouse limb bud cells.*

To investigate the physicochemical properties of HOXD13 condensates *in vivo*, limb bud cells harvested from wt and sphd mice were briefly exposed to 1,6-hexanediol (1,6-HD). 1,6-HD is a short-chain aliphatic alcohol that dissolves various biomolecular condensates[2,6,61], and is a potent inhibitor of super-enhancer transactivation[2]. After treatment, HOXD13 was visualized with immunofluorescence and confocal microscopy (**Figure 2-8 A**). HOXD13 staining in wt and spdh nuclei revealed a similar number of discrete nuclear puncta (**Figure 2-8 A-C**). Intriguingly, 1,6-HD treatment reduced the number of discrete HOXD13 puncta in wt limb bud cells ($p < 0.0001$, **Figure 2-8 B**), but had a negligible effect on HOXD13 puncta in spdh limb bud cells (n.s., **Figure 2-8 B**). These results suggest that a short alanine repeat expansion (+7A) alters the physicochemical properties of HOXD13-containing puncta in limb bud cells.

As BRD4 and Mediator play key roles in recruiting RNA Polymerase II to TF-bound genes[62], a loss in heterotypic interactions with these molecules would be expected to reduce the transcriptional output of repeat-expanded HOXD13. In agreement with this notion, luciferase reporter assays driven by the *Raldh2* promoter revealed that repeat-expanded HOXD13 displayed significantly reduced transactivation capacity (**Figure 2-8 D**). Taken together, these results suggest that repeat expansions reduce interactions of HOXD13 with transcriptional co-activators and chromatin, change the biophysical properties of HOXD13-containing condensates in disease-relevant limb bud cells, and reduce the transactivation capacity of HOXD13.

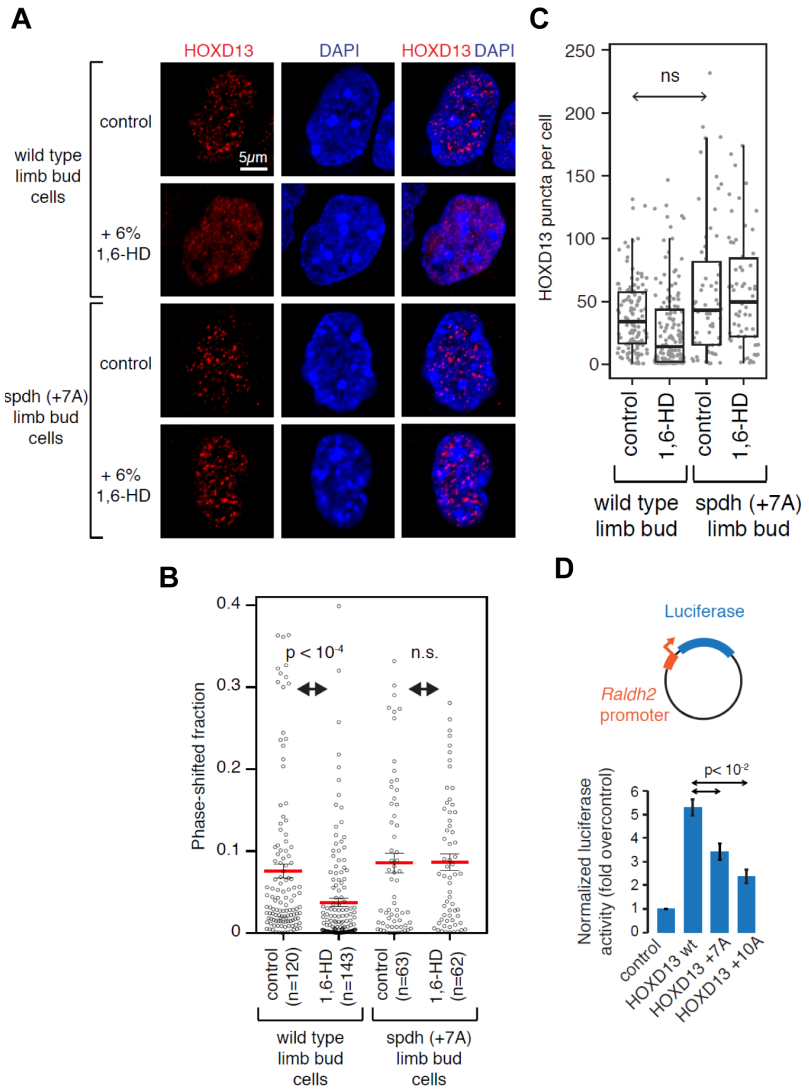

Figure 2-8: Hydrophobic repeat expansions alter material properties of Hoxd13 condensates in mouse model of synpolydactyly. *A) Representative images of Hoxd13 IF in WT and spdh E12.5 mouse limb bud cells with or without treatment with 6% 1,6-hexanediol for 1 minute. P-values from Student's t-test.* ***B)*** *Quantification of the Hoxd13 phase-shifted fraction from 1,6 hexanediol experiment shown in panel A (methods).* ***C)*** *Quantification of the number of Hoxd13 puncta detected in each nucleus from 1,6 hexanediol experiment shown in panel A. (n.s. = not significant, Student's t-test.* ***D)*** *Luciferase reporter assays of HOXD13 WT, +7A, and +10A mutants co-expressed with a Aldh1a2-luciferase construct in COS-7 cells. P-values from Student's t-test. Asita Steige (Max Planck Institute for Molecular Genetics) generated luciferase data shown in panel D.*

2.5 Hydrophobic repeat expansions enhance phase separation of other disease-associated TF-IDRs

Disase-associated alanine repeat expansions occur in IDRs of various transcription factors (TFs) (**Figure 2-1 A**). To test the generalizability of the model, I examined whether the phase separation capacity of other TFs, namely HOXA13 and RUNX2, is enhanced by disease-associated poly-alanine repeat expansions, using the optoDroplet and the *in vitro* droplet system.

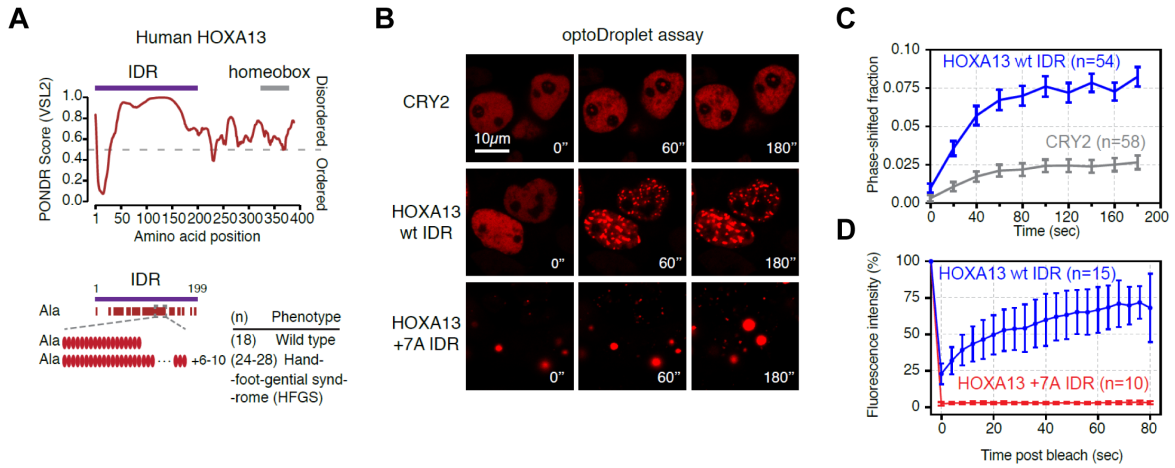

Figure 2-9: Hydrophobic repeat expansions enhance phase separation of Hoxa13-IDR. A) *Top: PONDR VSL2 prediction of disorder in human HOXA13 amino acid sequence. Purple bar indicates N-terminal intrinsically disordered region (IDR). Bottom: Alanine composition within HOXA13 IDR aligned to hand-foot-genital syndrome (HFGS)-associated hydrophobic alanine repeat expansions found in humans.* ***B)*** *Time lapse images of repeat expansion harboring HOXA13 IDR (+7A) constructs in optoDroplet system (lower panels) vs. wild type optoHOXA13-IDR construct (middle panels) and empty optoDroplet control construct (top panels) upon light induction.* ***C)*** *Quantification of light induced optoDroplet control vector, optoHOXA13-IDR(WT) and optoHOXA13-IDR(+7A) droplet formation measured by nuclear phase shifted fraction (methods) as a function of time.* ***D)*** *FRAP signal quantification of light induced optoHOXA13-IDR(WT) and optoHOXA13- IDR(+7A) droplets as a function of time.*

HOXA13 is a transcription factor that contains a homeobox domain, and is implicated in limb-and urogenital tract development in vertibrates[46]. Poly-alanine expansions found in HOXA13 are monogenic mutations linked to hand-foot genital syndrome (HFGS), a congenital disorder involving malformation of the limbs and urogenital tract[63] (**Figure 2-9 A**). Like HOXD13, alanine repeat expansions in HOXA13 occur in its N-terminal IDR[46] (**Figure 2-9 A**). As expected, the wildtype HOXA13-IDR facilitated phase separation in the optoDroplet system (**Figure 2-9 B and C**), and these droplets exhibited liquid-like FRAP rates (**Figure 2-9 D**). In contrast, the +7A HFGS associated repeat expansion enhanced condensation of optoHOXA13-IDR droplets and displayed completely arrested FRAP rates (**Figure 2-9 D**).

RUNX2 is a transcription factor from the RUNT family that controls bone morphogenesis[64]. Poly-alanine expansions found in RUNX2 are monogenic mutations linked to cleidocranial dysplasia (CCD), a disease characterized by severe skeletal defects[64,65] (**Figure 2-10 A**). Like HOXA13 and HOXD13, alanine repeat expansions in RUNX2 occur in its N-terminal IDR[64] (**Figure 2-10 A**). As expected, the wildtype RUNX2-IDR facilitated phase separation in the optoDroplet system (**Figure 2-10 B and C**), and these droplets exhibited liquid-like FRAP rate (**Figure 2-10 D**). In contrast, the +10A CCD-associated repeat expansion enhanced condensation of optoRUNX2-IDR droplets and displayed completely arrested FRAP rates (**Figure 2-10 D**). All constructs used for this phase separation experiment involving the optoDroplet platform were verified for similar nuclear expression levels (**Figure 2-10 E**).

The expression of repeat expanded HOXA13-IDR and repeat expanded RUNX2-IDR significantly reduced transcriptional activity when fused to GAL4 DNA-binding domain in a luciferase reporter system when compared to wild-type IDR ($p < 0.001$, Figure 2 11 A). Furthermore, all disease-associated repeat expansions enhanced droplet formation of mCherry-tagged, purified HOXA13-IDR (Figure 2 11 B and C) and RUNX2-IDR (Figure 2 11 D and E) in vitro. Collectively, these results reveal that pathological alanine repeat expansions found in a diverse set of developmental diseases (SPD, HFGS, CCD), all enhance phase separation capacity and alter transcriptional activity of the associated TF.

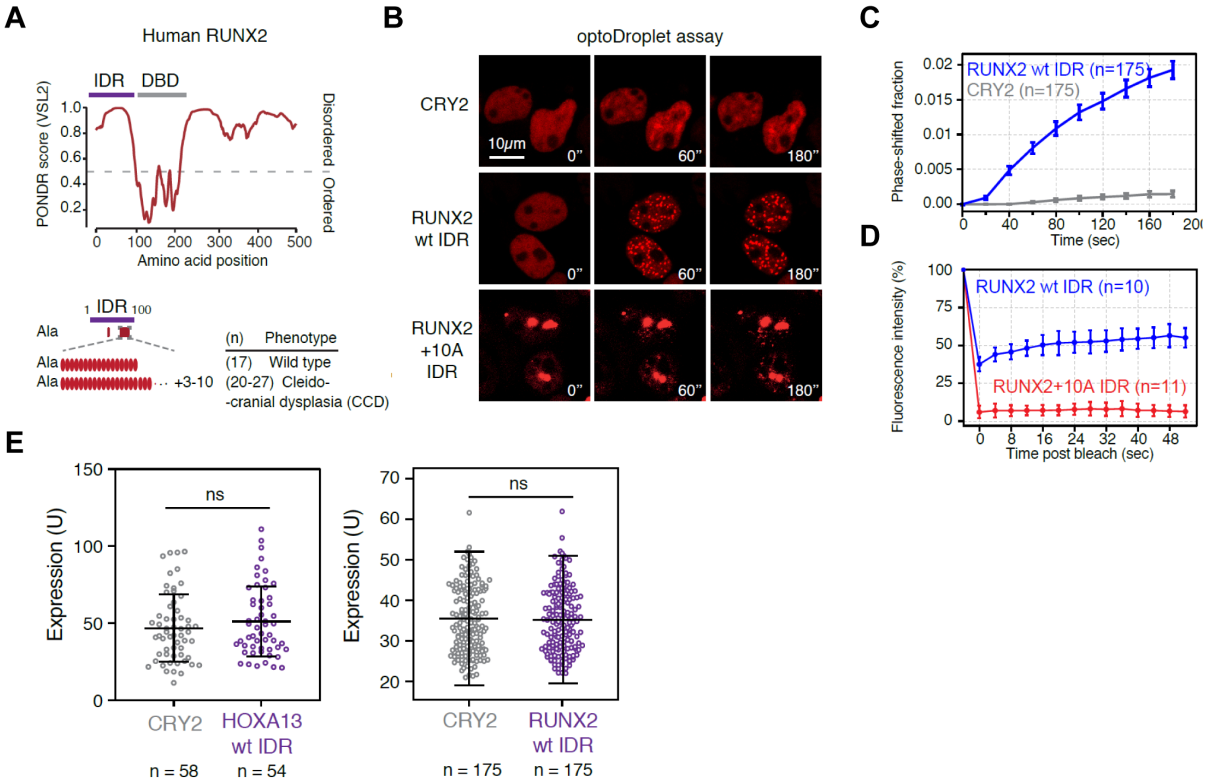

Figure 2-10: Hydrophobic repeat expansions enhance phase separation of TF-IDRs. A) *Top: PONDR VSL2 prediction of disorder in human RUNX2 amino acid sequence. Purple bar indicates N-terminal intrinsically disordered region (IDR). Bottom: Alanine composition within RUNX2 IDR aligned to cleido-cranial dysplasia (CCD)-associated hydrophobic alanine repeat expansions found in humans.* ***B)*** *Time-lapse images of repeat expansion harboring RUNX2 IDR (+10A) constructs in optoDroplet system (lower panels) vs. wild type optoRUNX2-IDR construct (middle panels) and empty optoDroplet control construct (top panels) upon light induction.* ***C)*** *Quantification of light-induced optoDroplet control vector, optoRUNX2-IDR(WT) and optoRUNX2-IDR(+7A) droplet formation, measured by nuclear phase shifted fraction (methods) as a function of time.* ***D)*** *FRAP signal quantification of light-induced optoRUNX2-IDR(WT) and optoRUNX2- IDR(+10A) droplets as a function of time.* ***E)*** *Expression controls showing mCherry signal in cells transfected with vectors for optoDroplet experiments and quantifications shown in panel C (n.s. = not significant, Student's t-test). Dora Knezevic and Henri Niskanen (Max Planck Institute for Molecular Genetics) assisted with data generation for figure panels B-D).*

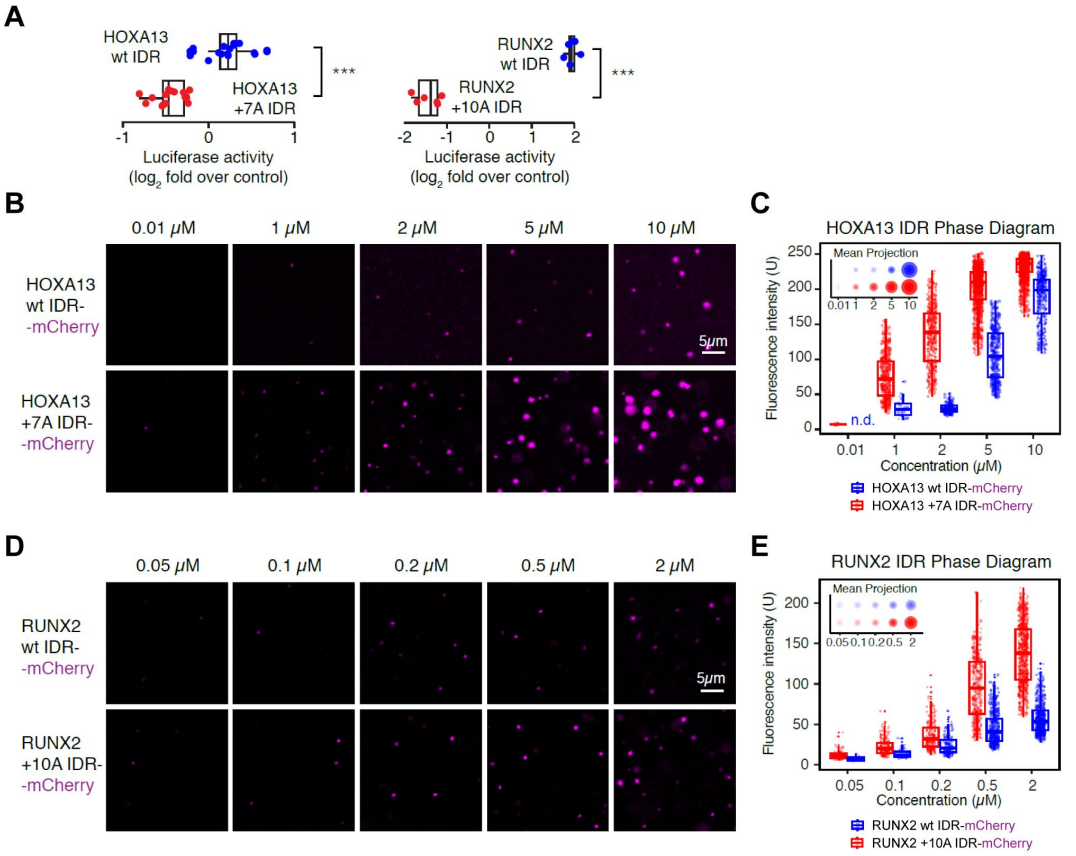

Figure 2-11: Hydrophobic repeat expansions enhance phase separation of TF-IDRs in vitro. ***A)*** *Luciferase reporter activity of the indicated TF IDRs (wt and repeat expanded HOXA13 and RUNX2) fused to GAL4-DBD. *** indicates p-value < 10^{-3} from Student's t-test.* ***B)*** *Representative images of purified wildtype and synpolydactyly associated HOXA13 IDR mCherry fusion proteins (WT, +7A) in droplet formation buffer (methods).* ***C)*** *Phase diagram of HOXA13 IDR-mCherry fusion proteins, at assayed concentrations shown in panel B. Every dot represents a droplet. Inset: projected average size of the droplets as mean ± SD (middle circle, mean; inner and outer, SD). (n.d. = no droplets).* ***D)*** *Representative images of purified wildtype and synpolydactyly associated RUNX2 IDR mCherry fusion proteins (WT, +10A) in droplet formation buffer (methods).* ***E)*** *Phase diagram of RUNX2 IDR-mCherry fusion proteins, at assayed concentrations shown in panel D. Every dot represents a droplet. Inset: projected average size of the droplets as mean ± SD (middle circle, mean; inner and outer, SD). Dora Knezevic and Henri Niskanen (Max Planck Institute for Molecular Genetics) assisted with data generation for figure panel A.)*

2.6 The androgen receptor is an oncogenic TF that forms nuclear condensates

The AR has been reported to form mesoscale nuclear 'speckles' in AR-expressing cells exposed to androgenic hormones, yet the functionality of these speckles has remained an enigma, largely to their extremely small size[66–68]. To overcome this challenge, I developed a stimulated-emission depletion (STED) pipeline to study the attributes of AR 'speckles' in the endogenous human prostate cancer context (**Figure 2-12 A and B**, methods).

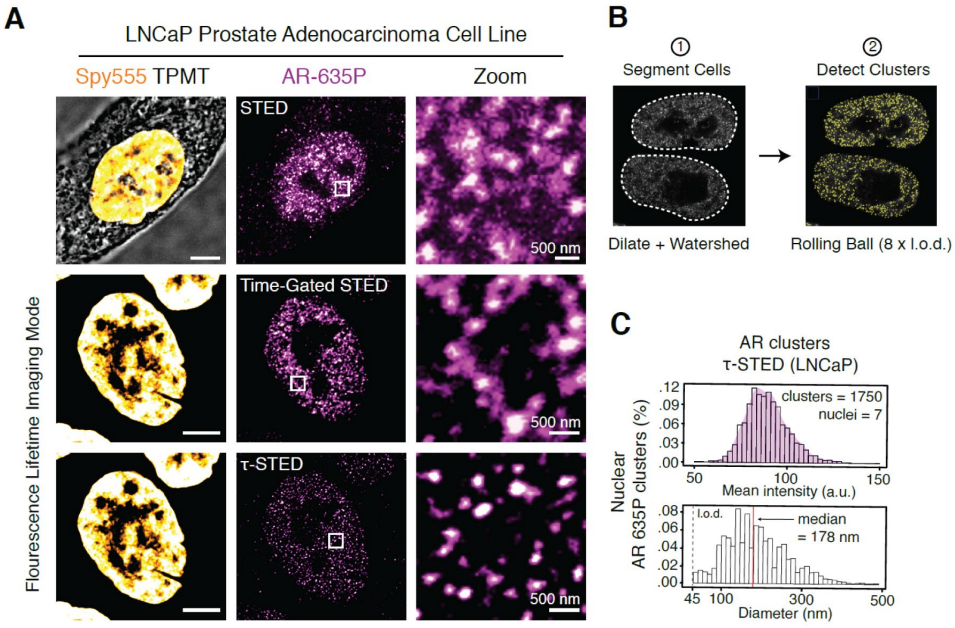

Figure 2-12: The androgen receptor is an oncogenic TF that forms nuclear puncta. A) *Stimulated emission depletion (STED) (top row) and FLIM STED images showing AR clusters in LNCaP nuclei before and after τ-STED deconvolution (middle and bottom row). Left column shows LNCaP nuclear counterstain using Spy555-DNA stain. Scale bar: 5 µm. Right panels show zoom-ins corresponding to intra-nuclear regions indicated by white boxes on panels in the central column. Scale bar: 500 nm.* ***B)*** *Quantification pipeline used to analyze STED image composites, showing segmentation of cells and detection of clusters using rolling ball background subtraction adjusted to 8 x the resolving capacity of the image (48 nanometers / pixel for τ-STED imaging of LNCaP cells). Legend is continued on the next page.*

C) *Top: Quantification of τ-STED intensity signal. Bottom: diameter of endogenous AR clusters in LNCaP cells (1750 AR clusters detected across 7 LNCaP nuclei imaged with same fluorescence time gating). L.o.d indicates the limit of detection. Density$_{max}$ diameter (bin with highest density of AR clusters in the distribution of all detected AR clusters): 123 nm, median diameter: 178 nm.*

To visualize endogenous AR, I performed fixed cell immunofluorescence in LNCaP prostate adenocarcinoma cells, and sequentially imaged stained cells with STED and fluorescence lifetime microscopy with τ-STED deconvolution at super-resolution. LNCaP nuclei displayed hundreds of diffraction-limited (100 – 300 nanometer) AR clusters (**Figure 2-12 A**). The measured AR clusters had a median diameter of 178 nanometers (**Figure 2-12 C**), comparable in size to phase-separated transient clusters formed by RNA Polymerase II and the Mediator co-activator complex.

To gain insights on hormone-dependent nuclear shuttling of the otherwise cytoplasmic AR, I imaged LNCaP cells grown in charcoal-stripped serum (CSS, methods). CSS is depleted in androgenic hormones, and leads to the arrest in proliferation of hormone-sensitive prostate cancer cells[69](**Figure 2-13 A and B**). In similarity to findings using over-expressed GFP-tagged AR[67,68], hormone-deprived LNCaP cells displayed distribution of AR signal throughout the cytoplasm. In contrast, LNCaP cells exposed to androgens exhibited enrichment of clustered AR in the nucleus (**Figure 2-13 C**).

LNCaP cells cultured in CSS medium also expressed lower levels of canonical AR targets known to be associated with super-enhancers found in LNCaP cells[70], when compared to LNCaP cells grown in androgen-containing full-medium (**Figure 2-13 D**). Together, these results suggest that 100 – 300 nanometer, nuclear AR cluster formation is involved in the transactivation of oncogenic AR-target genes, and the proliferation of hormone-sensitive prostate cancer cells.

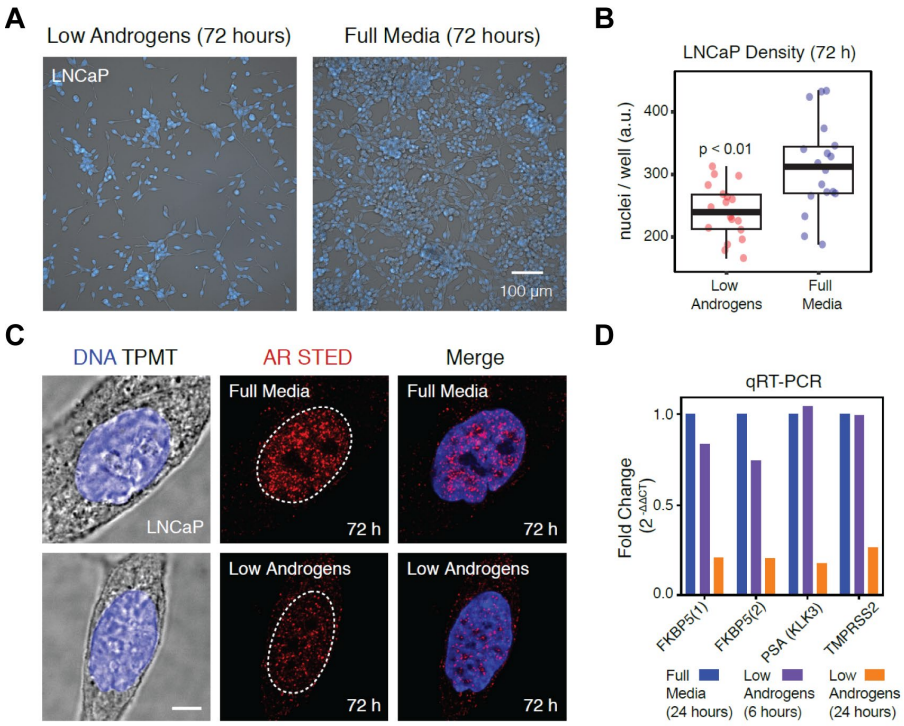

Figure 2-13: The androgen receptor is an oncogenic TF that controls gene expression in prostate cancer cells. ***A)*** *Representative images of Hoechst stained LNCaP cells grown in either charcoal-stripped medium ('Low Androgens') or unstripped medium ('Full Androgens') for 72 hours.* ***B)*** *Quantification of LNCaP cell count (# of nuclei / well) after 72 hours of growth in indicated condition. P-value from Student's t-test.* ***C)*** *AR IF and STED microscopy on LNCaP cells grown in either charcoal-stripped medium ('Low Androgens') or unstripped medium ('Full Androgens') for 72 hours. Dashed white line indicates nuclear boundary.* ***D)*** *qRT-PCR of FKBP5, PSA, and TMPRS22 AR transcript targets using primer pairs for each loci, in LNCaP cells grown in either charcoal-stripped medium ('Low Androgens') or unstripped medium ('Full Androgens') for 6 and 24 hours. Values indicate* $2^{-\Delta\Delta Ct}$ *(Log fold change of gene target versus β-Glucuronidase housekeeping gene in full media sample normalized to values from stripped media sample).*

2.7 Disordered regions in the AR activation domain facilitate phase separation in cells

The full-length AR contains an intrinsically disordered N-terminal activation domain (AD), a structured central DNA-binding domain (DBD), and a structured C-terminal ligand binding domain (LBD) (**Figure 2-14 A and B**). In cells expressing full-length AR-eGFP exposed to ligand (dihydrotestosterone [DHT]), the receptor translocated to the nucleus to form mobile, diffraction-limited clusters, as observed by live-cell STED microscopy (**Figure 2-14 C**).

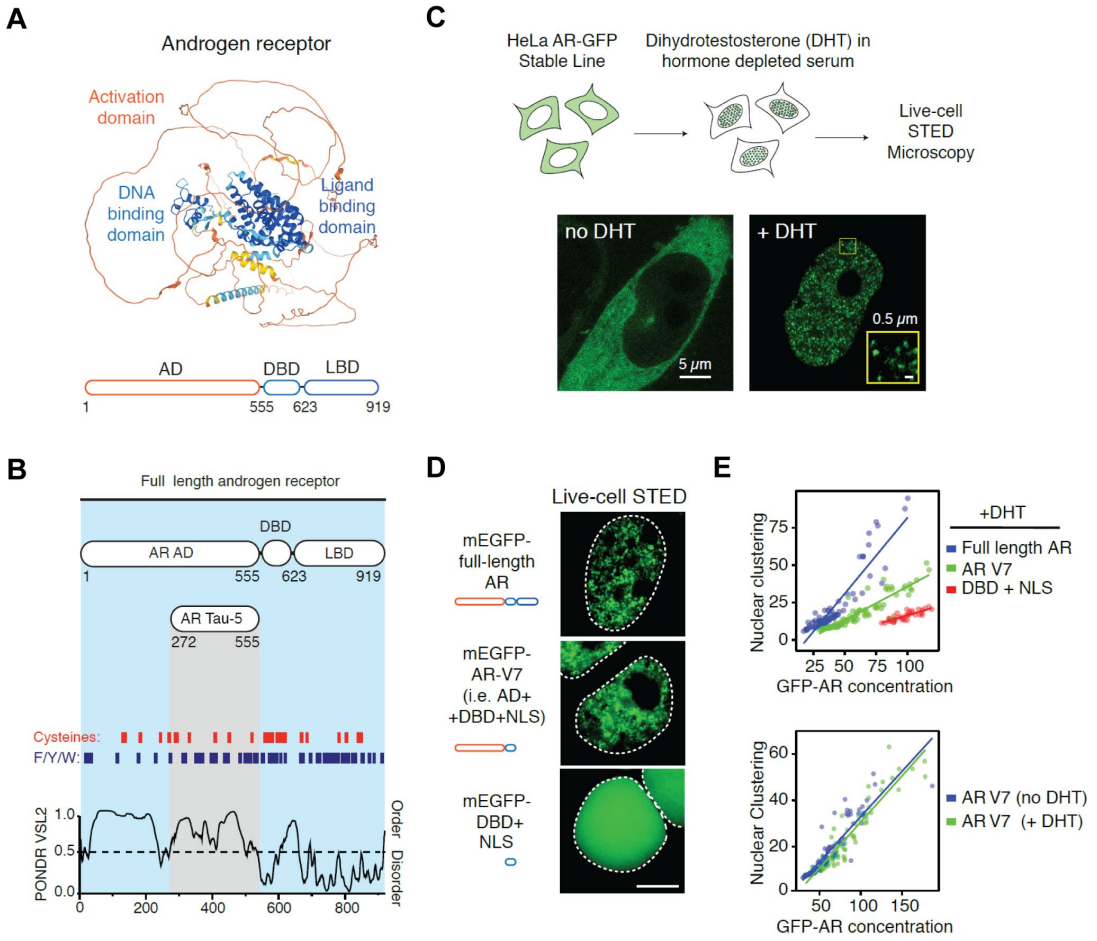

Figure 2-14: Disordered regions in the AR activation domain facilitates phase separation in cells. A) *Top: Structure of AR predicted with AlphaFold. The model is coloured by structure prediction confidence from high confidence (dark-blue) to low confidence (orange-yellow). Bottom: Schematic of full-length AR and its domains. Legend is continued on the next page.*

B) *Top: Scheme showing the individual domains of the AR and AR-Transactivating unit 5 (Tau5) fragment, aligned to AR amino acid sequence (UniprotKB - build P10275). Middle: Red dashes indicate the location of cysteines in the AR and blue dashes indicate the location of phenylalanine, tyrosine, and tryptophan residues within the AR. Bottom: PONDR VSL2 residue scores showing predicted disorder in AR sequence by amino acid.* ***C)*** *Top: Experimental schematic using HeLa AR-eGFP stable cell line. Bottom: Live-cell STED imaging of HeLa cell nucleus stably expressing AR-eGFP, treated with 1 nM dihydrotestosterone (DHT) or vehicle for 4 hours. Yellow insets show zoom-ins of DHT-stimulated AR-eGFP clusters (24-nanometer limit of detection).* ***D)*** *Live-cell STED imaging of HEK293T cells transfected with the indicated AR constructs tagged with mEGFP. Cells were imaged after treatment with 10 nM DHT or vehicle for four hours. Scale bar: 5 µm. Dashed line indicates the nuclear boundary.* ***E)*** *Quantification of live-cell experiment transfected with constructs shown in panel D. y-axis indicates the standard deviation, and x-axis indicates the mean intensity of pixels in the corresponding nucleus. Each dot represents measurements from an individual cell, and lines represent standard regression fits to the indicated data spread (N = 2).*

To dissect the molecular basis of AR phase separation, I tested which domains of the AR facilitated nuclear cluster formation in cells. In transiently transfected HEK293T cells, the full-length AR and AR-V7 splice variant, which both contain the disordered AR-AD and -DBD, formed nuclear clusters. Intriguingly, a construct lacking the AD did not (**Figure 2-14 D**). AR-V7 formed nuclear clusters even in the absence of the hormone (**Figure 2-14 D**), due to splice-out of the hormone-binding LBD for a constitutively active 16 amino-acid nuclear localization signal (NLS)[71]. AR-V7 is an AR splice variant used as an early prognostic biomarker for AR-driven CRPC that is resistant to first-line therapy with anti-androgens[37,42,72]. Cells with higher expression of AR and AR-V7 displayed increased nuclear clustering (**Figure 2-14 E**), suggestive of a biomolecular condensation process reliant on liquid-liquid phase separation. Together, these results suggest that the disordered AR AD facilitates LLPS of the receptor in cells.

2.8 Experimental drug enhances phase separation of the AR activation domain *in vitro*

The search for etiological treatment of CRPC brought to light the existence of EPI-001, a weak anti-androgen enriched in *Geodia lindgreni* marine sponge extracts found in lakes contaminated with industrial waste in British Columbia, Canada[44]. EPI-001 displays the ability to interact with the intrinsically disordered activation domain of the AR AD[73], and the compound contains two aromatic rings and a chlorohydrin group thought to bind covalently to the AR AD[44,73,74] (**Figure 2-15 A**). EPI-001 analogs are currently under clinical investigation in prostate cancer (NCT04421222), but their mechanism of action is unclear, and the residues that may be modified by EPI-001 in the AR AD are not known. As the AR AD facilitates AR phase separation (**Figure 2-14 D and E**) and transcriptional activity[71,75], I hypothesized that EPI-001 would alter these functions.

As expected, in LNCaP cells, EPI-001 displayed anti-proliferative and AR transcriptional inhibitory effects, at sub-millimolar concentrations of compound (**Figure 2-15 B-D**). Intriguingly, sub millimolar concentrations of EPI-001 (250 μM) enhanced the phase separation capacity of a transactivating unit within the AR AD responsible for oncogenic gene activity in CRPC, known as AR transactivating unit 5 (AR-Tau5)[69,71] (**Figure 2-15 B**), tested using an in-vitro model of AR Tau-5 tethered to mCherry. Of note, other nuclear IDPs known to promote LLPS, including NPM1, HP1α, or MED1 IDR[2,57,76] did not display enhanced phase separation in the presence of EPI-001 at the same concentration (**Figure 2-15 G-I**). This effect on enhanced phase separation *in vitro* is reminiscent of the effect of disease-associated poly-alanine expansions found in aforementioned TF-IDRs.

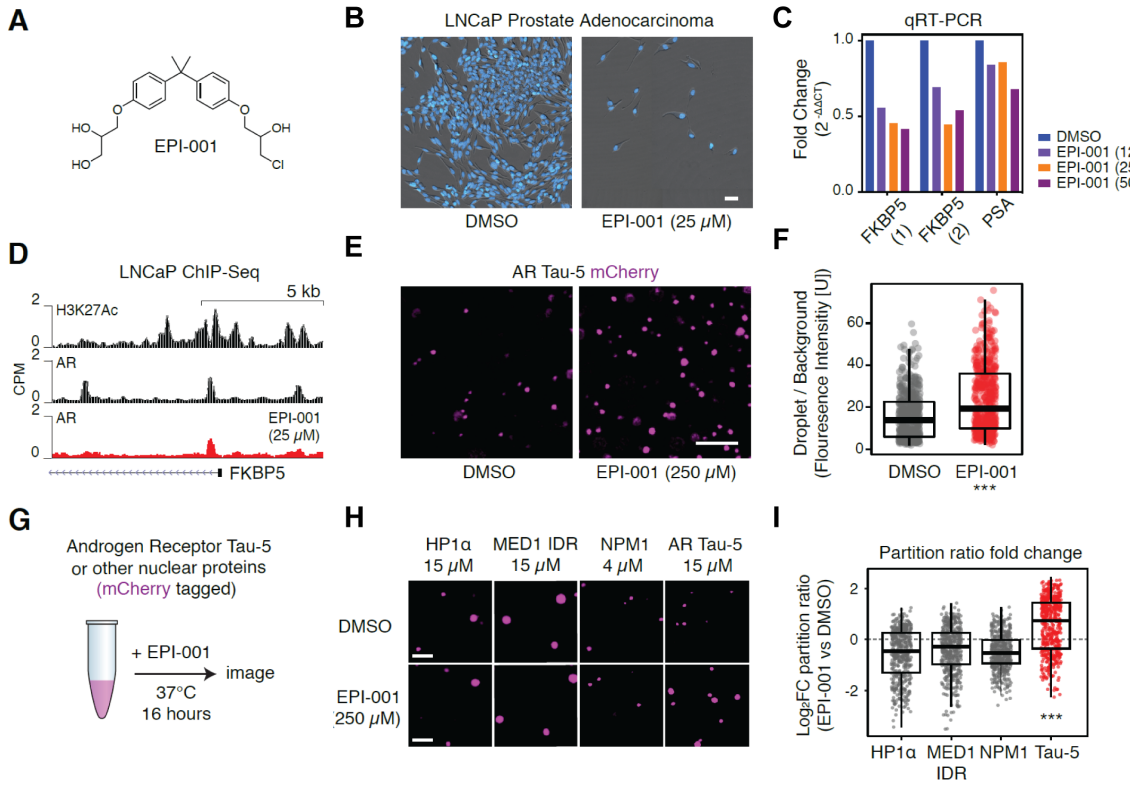

Figure 2-15: Experimental drug enhances phase separation of the AR activation domain in vitro. A) *Chemical structure of EPI-001.* ***B)*** *Representative images of Hoechst-stained LNCaP cells grown in media spiked with 25 μM EPI-001 or DMSO vehicle control for 96 hours. Scale bar: 50 μm.* ***C)*** *qRT-PCR of FKBP5 and KLK3 transcript targets using primer pairs for each loci, in LNCaP cells with 25 μM EPI-001 or DMSO vehicle control for 96 hours. Values indicate* $2^{-\Delta\Delta Ct}$ *(Log fold change of gene target versus β-Glucuronidase housekeeping gene in EPI-001 sample normalized to values from DMSO sample).* ***D)*** *AR and H3K27Ac ChIP-Seq binding profiles at the AR target gene FKBP5 in LNCaP cells treated with either 25 μM EPI-001 or DMSO vehicle for 24 hours.* ***E)*** *Representative images of purified AR transactivating unit-5 (AR Tau-5) mCherry fusion protein in the presence of 250 μM EPI-001 or DMSO in droplet formation buffer (methods). Scale bar: 5 μm.* ***F)*** *Quantification of the partition ratio of AR Tau-5 droplets in conditions shown in panel E (fluorescence intensity within droplet / fluorescence intensity of background, methods). Each dot represents a single droplet, imaged across 2 biological replicates. *** indicates P-value < 0.001 from Student's t-test.* ***G)*** *Scheme of droplet formation assays with AR Tau5-mCherry and other mCherry-tagged nuclear proteins with 250 μM EPI-001 or DMSO after incubation at 37°C for 16 hours. Legend is continued on the next page.*

H) *Representative images of purified mCherry tagged fusion proteins in droplet formation buffer (methods), after treatment described in panel J. Scale bar: 5 μm.* ***I)*** *Quantification of the log_2fold partition ratio of droplets in presence of EPI-001 shown in panel H (fluorescence intensity within droplet / fluorescence intensity of background, methods), normalized to droplets in DMSO. Each dot represents a single droplet, imaged across 3 biological replicates. *** indicates P-value < 0.001 from Student's t-test.*

EPI-001 has been reported to react with side chains of the AR AD through its chlorohydrin moiety, but the exact residues in AR AD that may be covalently modified by EPI-001 have not been mapped[74]. To identify the residues modified by EPI-001, we incubated recombinant AR AD with EPI-001 and DMSO control, and subjected the mixtures to mass spectrometry. EPI-001 adducts were detected in at least 10% of mapped peptides for 5 out of 11 cysteines within the AR AD (**Figure 2-16 A**). The most modified cysteines were cysteine 265, and cysteine 404 and 518 in AR Tau-5 (**Figure 2-16 A**). The contributions of aromatic adducts on cysteines were then investigated by targeted mutagenesis. For this purpose, two AR Tau-5 proteins were generated: one in which the two most heavily modified cysteines were mutated to phenylalanine and tyrosine, and one in which all cysteines detected to have EPI-001 adducts were mutated to either phenylalanine or tyrosine (**Figure 2-16 A**). The mutations substantially enhanced droplet formation by mCherry-tagged AR Tau-5 (**Figure 2-16 B and C**). Together, these results suggest that EPI-001 enhances phase separation of the AR AD, in part by reacting with cysteines and enhancing the aromatic character of the AR AD.

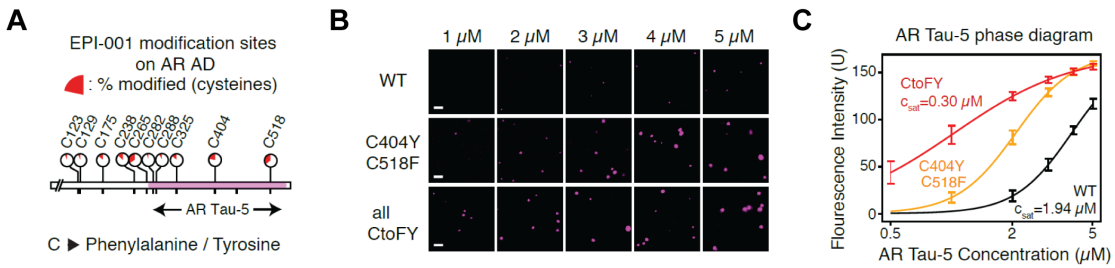

Figure 2-16: Aromatic substitutions enhance phase separation of AR activation domain in vitro. A) *Top: Schematic showing cysteines in AR AD detected to be covalently modified by EPI-001 in vitro after incubation at 37°C for 16 hours. Red indicates % of peptide fragments detected by mass-spectrometry experiment (N = 3) with a mass shift equivalent to one covalent adduct of an EPI-001 molecule. Bottom: Strategy employed to phenocopy EPI-001 adducts with aromatic residues (phenylananine, tyrosine) by mutagenesis of cysteines in AR Tau-5.* ***B)*** *Representative images of purified AR Tau-5 mCherry and EPI-001 phenocopy mutants, in droplet formation buffer (methods). Scale bar: 5 μm.* ***C)*** *Phase diagram of AR Tau-5 and EPI-001 phenocopy mutant proteins calculated as a 4- parameter log-logistic dose response curve (mean fluorescence intensity of droplet vs assayed concentrations shown in panel B).* C_{sats} *calculated from log-logistic dose response curve. The Salvatella lab (Institute for Research in Biomedicine, Barcelona) generated mass spectrometry data shown in figure panel A.*

2.9 Hydrophobic small molecules are more potent inhibitors of AR gene activity

To date, no first or second-generation EPI-001-derived compound has been approved for the treatment of CRPC, in part due to an enigmatic mechanism of action that renders chemical optimization against the AR AD a challenge[77,78]. I hypothesized that as EPI-001 seems to effect AR condensation, we could rationally adjust the design of EPI-001 to make more potent compounds for the treatment of CRPC. To this end, a series of analogues of linearized EPI-001 analogues were created, with substitutions in positions R_1 and R_2 to increase hydrophobicity (compounds 1aa – 1bb, **Figure 2-17 A**), and tested the compounds in LNCaP cells transfected with aa luciferase reporter driven by an AR-dependent promoter and enhancer combination[44,74,79].

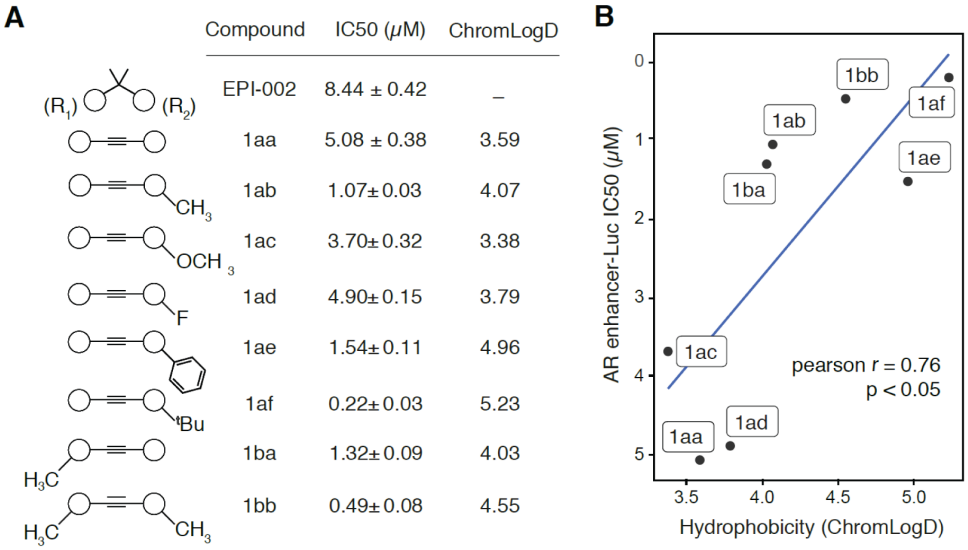

Figure 2-17: Hydrophobic small molecules are more potent inhibitors of AR transactivation capacity. ***A)*** *Left: Molecular structure of engineered molecules derived from the stereoisomer of the lead compound, EPI-001. Middle: IC50 values of engineered molecules calculated from dose response measurements using an AR enhancer driven PSA luciferase assay in LNCaP cells. Right: ChromLogD, hydrophobicity values of engineered molecules experimentally measured by HPLC (octanol phase / water phase).* ***B)*** *Regression fit of AR-luc IC50 in LNCaP cells vs. hydrophobicity (ChromLogD) of engineered molecules. Pearson's correlation coefficient: r = 0.76, p < 0.05. Legend is continued on the next page.*

The Salvatella lab (Institute for Research in Biomedicine, Barcelona) generated hydrophobicity and luciferase activity data of compounds shown in figure panel A.

In agreement to the hypothesis, the more hydrophobic the substitution, the more potent the compound was on dose-dependent inhibition of AR-driven luciferase activity in LNCaP cells (r = 0.76, p < 0.05, **Figure 2-17 B**). For example, introduction of a methyl group at R_1 (1ab) or R_2 (1ba) increased potency from IC_{50} 5 μM down to IC_{50} ~1 μM. However, the introduction of this group in both positions (1bb) exhibited an even higher potency of IC_{50} 0.5 μM in the luciferase reporter system. The most hydrophobic compound of the series, with a tert-butyl substitution at R_2 (1af), or three methyl groups, brought the IC_{50} down to 0.22 μM, boasting highest potency out of the entire series.

To gain an understanding of how the improved compounds affected the phase separation process in cells, compound (1ae), a more hydrophobic derivative of linearized EPI-001 with a phenyl group at R_2, was treated on cells stably expressing AR-eGFP. 1ae was chosen for this purpose, due to its easier solubility and handling in aqueous solution, unlike other compounds in the series. A pilot experiment using live-cell STED microscopy revealed that cells treated with 1ae displayed an overall higher number of nuclear, diffraction-limited AR-eGFP clusters, consistent with the notion that these compounds are able to partition and perhaps enhance phase separation of AR in cells (**Figure 2-18 A**). Much like poly-alanine repeat expansions causing a change in the composition of transcription factor containing condensates, perhaps EPI-001 derived molecules are able to sequester AR into homotypic condensates, at the expense of heterotypic condensates, thereby achieving inhibition of AR-dependent transcription at CRPC oncogenes (**Figure 2-18 B**).

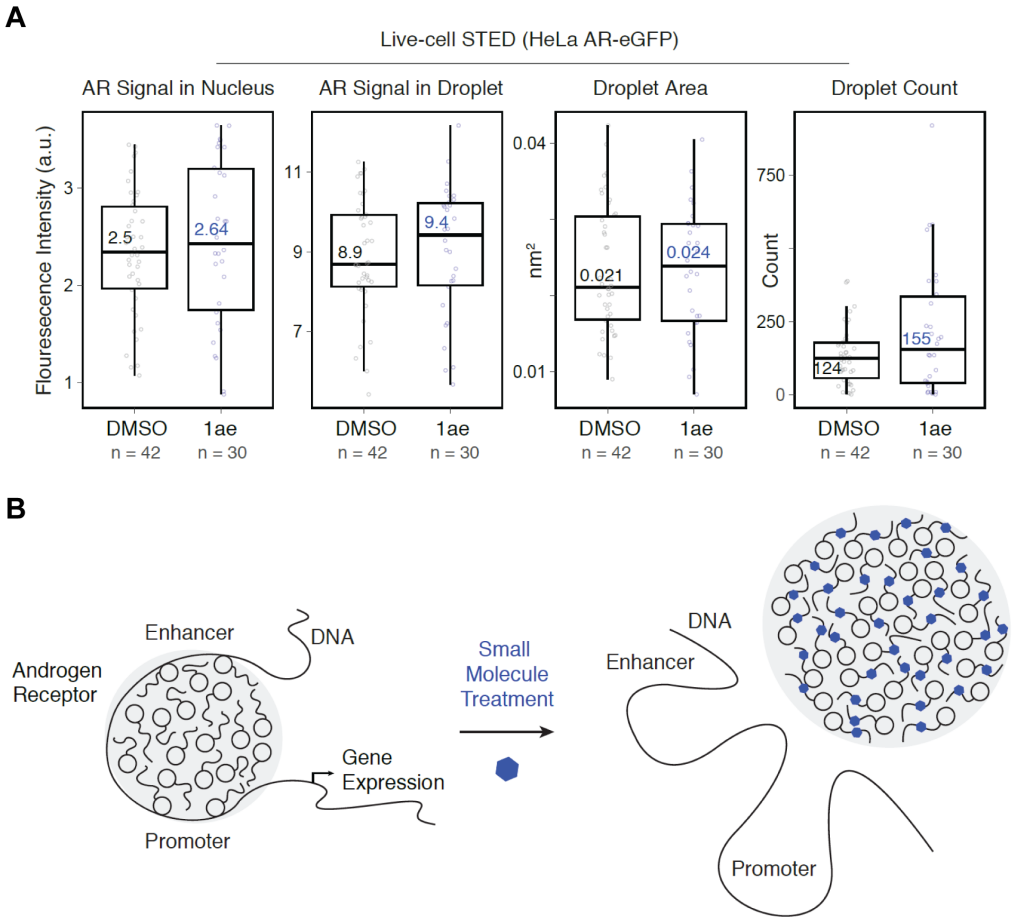

Figure 2-18: Hydrophobic small molecules may perturb phase separation of AR condensates in cells A) *Live-cell STED quantifications of DHT-induced AR-eGFP nuclear condensates in HeLa cells, after 4 hours of co-treatment with 5 μM of the engineered molecule 1ae or DMSO vehicle. (methods).* ***B)*** *Model schematic of how hydrophobic small molecules may perturb the composition and activity of AR-containing transcriptional condensates.*

2.10 Hydrophobic small molecule more potently inhibits proliferation and AR-driven oncogenic programs in a human prostate cancer model

The effect of 1ae and EPI-001 on proliferation of AR-dependent cells was studied using proliferation assays in LNCaP cells after treatment with a dose titration of each compound for 96 hours (**Figure 2-21 A-C**). Proliferation assays revealed that 1ae had an IC50 of approximately 1 µM, in comparison to EPI-001 which displayed and IC50 of 22 µM.

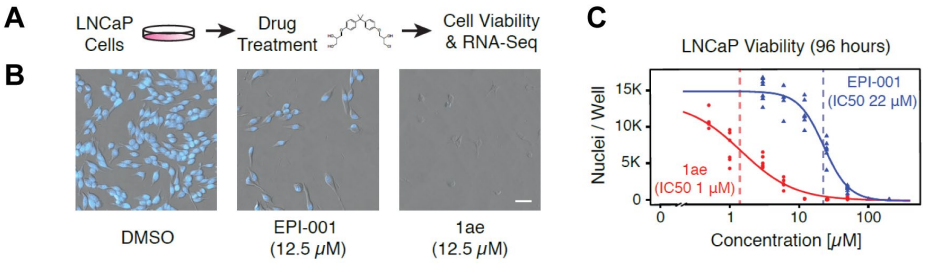

Figure 2-19: Hydrophobic small molecule more potently inhibits proliferation in a human prostate cancer model. ***A)*** *Schematic showing approach used to investigate the effect of compound 1ae on proliferation and transcriptional program in LNCaP prostate cancer cells.* ***B)*** *Representative images of Hoechst stained LNCaP cells grown in media spiked with 12.5 µM EPI-001, 12.5 µM 1ae, or DMSO vehicle control for 96 hours. Scale bar: 50 µm.* ***C)*** *4-parameter log-logistic dose-response curve of viable LNCaP cells as a function of compound concentration after 96 hours of treatment (N = 2).*

To investigate the potency and specificity of 1ae and EPI-001 on AR-dependent transcription in prostate cancer cells, I employed RNA-sequencing after treatment of LNCaP cells with approximated IC10 and IC50 values of the compounds for 6 and 24 hours (**Figure 2-20 A**). 6-hour treatment had a negligible effect on the gene expression profile of prostate cancer cells, in comparison to 24 hours of treatment (**Figure 2-20 A**). Interestingly, 24 hours treatment with 25 µM EPI-001 led to the differential expression of 64 genes, and 24-hour treatment with 5 µM 1ae led to the differential expression of 231 genes, compared to DMSO-treated control cells (cut-offs for differentially expressed gene [DEG] calling: |Log2FoldChange| > 1 and p-value < 1e-10, **Figure 2-20 B**).

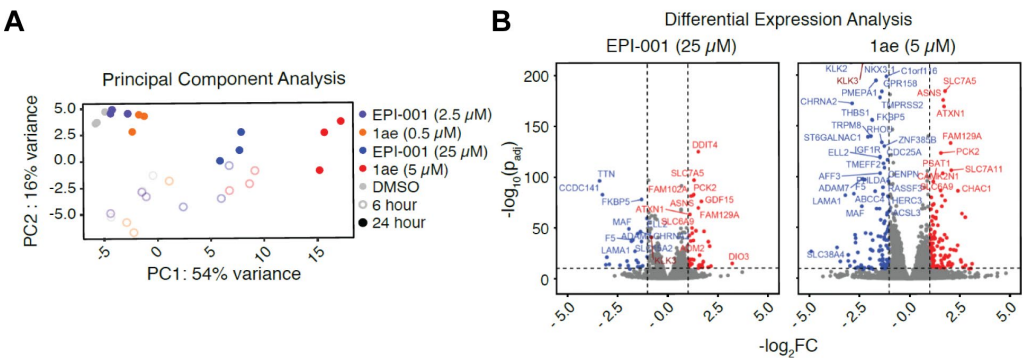

Figure 2-20: Hydrophobic small molecule more potently induces differential gene expression in a human prostate cancer model. A) *Principal component analysis on RNA-Seq transcriptome profile of LNCaP cells treated at indicated conditions for 24 hours (N = 3).* ***B)*** *Differential expression analysis on RNA-Seq transcriptome profile of LNCaP cells treated with either 5 μM 1ae or 25 μM EPI-001 vs. DMSO control for 24 hours. Differentially expressed genes (DEGs) called with $|Log_2FC| > 1$ and p-value < 1e-10 cutoffs.*

Gene set enrichment analysis (GSEA) revealed that perturbed genes were significantly enriched for known AR-targets, for both EPI-001 and 1ae (padj < 0.01) (**Figure 2-21 A**). Both EPI-001 and 1ae dysregulated the same subset (5/50) of pathways tested with GSEA (**Figure 2-21 A**), and downregulated the mean expression of genes in the identified pathways (**Figure 2-21 B**). The significantly dysregulated pathways included the AR response pathway, and other pathways known to be hyperactive in CRPC (MYC, E2F, G2M pathways)[70,80]. Remarkably, 5 μM 1ae treatment led to a more profound reduction in the expression of all down-regulated DEGs elicited by treatment with 25 μM EPI-001 (**Figure 2-21 B**). Of note, other expression pathways found in prostate tissue (i.e., spermatogenesis, **Figure 2-21 B**), were not affected by dosage of EPI-001 or 1ae. These results collectively indicate that 1ae specifically inhibits AR-dependent oncogenic programs in prostate cancer cells, and is more potent in its inhibitory effect than EPI-001 on proliferation and AR-dependent oncogenic transcription.

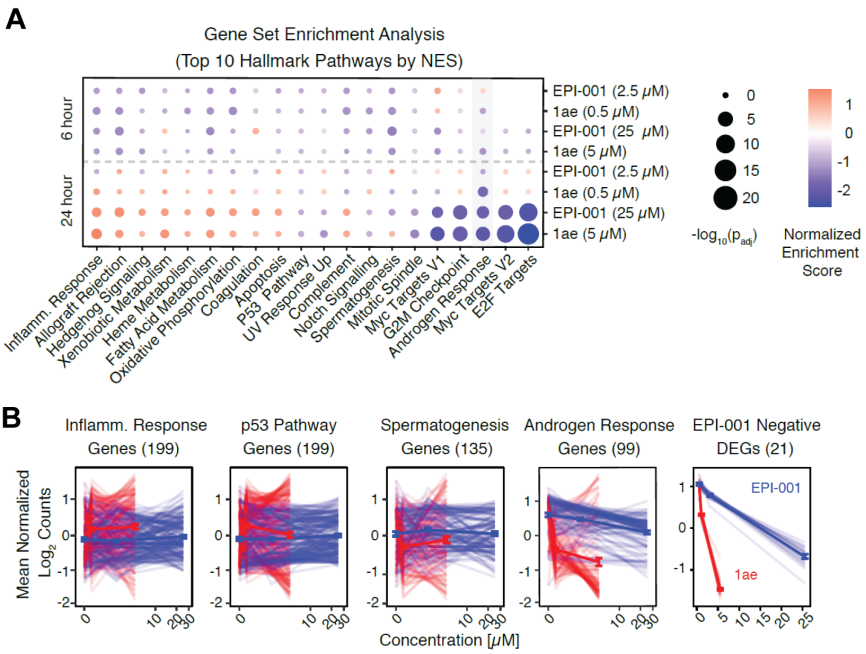

Figure 2-21: Hydrophobic small molecule more potently inhibits AR-driven oncogenic programs in a human prostate cancer model. A) *Gene set enrichment analysis of the curated hallmark androgen response pathway*[81]*, from RNA-Seq transcriptome profiles of LNCaP cells treated with either 5 μM 1ae or 25 μM EPI 001 vs. DMSO control. Insets list top 5 down-regulated genes.* ***B)*** *Line plots of the expression of individual genes in curated hallmark pathways selected from panel A), as a function of drug concentration.*

3 Discussion

The results presented here help establish that hydrophobic sequence features within TF-IDRs, studied using 1) disease-associated sequence mutations and 2) small molecule-mediated sequence interactions, enhance the phase separation capacity of TFs, and their ability to co-condense with components necessary for transcription.

For HOXD13, HOXA13, and RUNX2 TFs, disease-associated poly-alanine repeat expansions (synpolydactyly[32], hand-foot-genital syndrome[82], and cleido-cranial dysplasia respectively[64]) lowered the saturation concentration (c_{sat}) and promoted gel-like material properties of the TF-IDR condensates in biochemical and cell-based assays (**Figure 2-3 F, 2-9 D, and 2-10 D**). *In vitro*, repeat expanded HOXD13-IDR co-condensed with less Mediator than wild type IDR, and in a mouse model of synpolydactyly, repeat expanded HOXD13 TF associated with less histone H3K27 acetylation at canonical Hoxd13 target sites and BRD4 co-activator (**Figure 2-5 and 2-7**).

For the oncogenic TF, the androgen receptor (AR), the experimental anti-cancer drug EPI-001[44] enhanced the phase separation capacity (lowered c_{sat}) of the disordered AR transactivation domain, in a manner that was biochemically phenocopied by hydrophobic aromatic amino acid mutagenesis (**Figure 2-1**). For all assayed TF-IDRs (HOXD13, HOXA13, RUNX2, and AR), chemical modifications were associated with a change in transcriptional output, both in luciferase-transactivation systems and cell-based genomic assays (**Figure 2-8, 2-11 A, 2-17, and 2-21 B**). Together, these results indicate that changing the physicochemical properties of a TF-IDR, has a direct effect on phase separation, and elicits drastic changes in transcriptional programs in various human disease states.

Repeat expansion diseases include several severe, incurable developmental diseases thought to be associated with the presence of protein aggregates[46,83,84] (**Figure 2-1 A**). The transcriptional condensate model described here may explain several observations of the pathology of repeat expansion diseases that cannot be explained by toxic aggregates. For example, for poly-alanine repeat expansions, aggregates or nuclear inclusions were only described in overexpression systems but not in primary tissue samples[47]. Additionally, short +7A repeat expansion of HOXD13 recapitulates the human synpolydactyly phenotype in

mice, yet does not produce protein aggregates in the limb bud [51]. Other devastating diseases believed to be caused by the presence of 'toxic' aggregates composed of hydrophobic and disordered proteins, such as Alzheimer's and prion disease[85–87], may also benefit from the biomolecular condensate framework of thinking. Despite the wide spread subscription to the 'amyloid hypothesis', therapeutic strategies for Alzheimer's disease based on amyloid fibril clearance have been consistently ineffective for the past quarter century[88–90]. As an alternative hypothesis, perhaps pathogenesis in Alzheimer's disease, and other proteinopathies, is caused by a disease-associated shift in the composition of phase-separated, relevant sub-cellular compartments[17,22,25,91,92] and unrelated to the deposition of proteinaceous fibril. Changes in the phase separation capacity and miscibility of repeat expanded proteins with heterotypic condensates are consistent with this notion.

Recent studies suggest that phase separation could provide a mechanism for concentrating co-activators, TFs, RNA polymerase II, and DNA polymers into dynamic transcriptional condensates that associate at cell-type-specific super-enhancers[2–4,7,19]. Understanding the chemical features that drive condensate partitioning of transcriptionally relevant biomolecules could help elucidate how assembly controls genetic activity, and ultimately provide clues as to how to control gene activity with compounds that may exploit the chemical features of transcriptional condensates. My work provides evidence that hydrophobicity is indeed a chemical feature that can be exploited to control the formation and material properties of transcriptional condensates in diverse disease states. This is supported by the finding that longer repeat expansion lengths associate with more severe synpolydactyly phenotypes (**Figure 2-3 A**), elicit increased phase separation of HOXD13-IDR (**Figure 2-3 and 2-4**), cause further decreases in association of HOXD13-IDR with mediator (**Figure 2-5**), and more potently inhibit transactivation capacity of HOXD13-IDR (**Figure 2-8 D**). In a similar vein, chemical substitutions to functional groups on the anti-cancer drug EPI-001, elicit stronger change in inhibition of AR-dependent transcription, the more hydrophobic the substitution is (**Figure 2-17 B**, $p < 0.05$). The most potent compound of the series, 1bb, contains two methyl groups at R_1 and R_2, rendering it the most hydrophobic. In a human model of prostate cancer, a hydrophobic derivative of EPI-001, 1ae, with a phenyl substitution at functional group R_1, was indeed more potent in inhibition of cell proliferation and AR-dependent oncogenic transcriptional programs (**Figure 2-21**). Together, these results provide clues on new ways to control transcription, and prove that

hydrophobicity is an exploitable chemical feature of transcriptional condensates to control gene expression.

To date, the 'druggability' of transcriptional condensates has relied largely on conjecture[11]. The concept of selective partitioning, i.e., the concentration of a chemical into a specific biomolecular condensate within a cell, has only been shown, albeit in an off-target manner, for unspecific chemotherapeutic agents thus far[10,11,93]. For example, the anti-cancer drug cisplatin, partitions into Mediator condensates in a manner dependent on aromatic residues in the MED1 subunit of the mediator complex, despite targeting DNA[10,11,93]. Nevertheless, these results may explain the relative efficacy of platinated drugs, as partitioning into super enhancer-containing condensates (by association with mediator), may be responsible for the efficacy of these agents in diverse cancer states. Understanding the rules that govern selective partitioning into a condensate is attractive for this reason, as it provides a strategy to improve upon potency and specificity of small molecules used to treat disease[10]. I show evidence that a small molecule can specifically partition into condensates formed by the AR to enhance phase separation and exact changes in genomic expression driven by its cognate target. This is captured by experiments in which EPI-001 selectively enhances condensation of disordered regions in AR, as opposed to other disordered proteins commonly found in phase-separated nuclear sub-compartments (**Figure 2-15 H**). Intriguingly, not only did EPI-001 selectively partition into AR condensates, EPI-001 and optimized derivatives deregulated oncogenic expression programs in prostate cancer, specifically driven by AR[80,94], (**Figure 2-21 B**). Understanding how this specificity arises, and the rules that allow EPI-001 to selectively partition into AR condensates, will undoubtedly aid in the rational engineering of newer drugs that target specific components within various biomolecular condensates with different functions, to result in new specific therapies. For the AR, a combination of small structures and the enrichment of aromatic residues within the disordered AR activation domain (**Figure 2-14 B**), hint that π-π stacking[22] may be involved in the initial 'attraction' of EPI-001 and derived molecules, to pockets in the AR activation domain within AR containing condensates.

Finally, I provide evidence that modulation of AR phase separation actually has antitumorigenic effect in an in vivo CRPC model driven by an "undruggable" AR variant. Anti-androgens that are used as first-line therapy against prostate cancer, such as enzalutamide, target the ligand binding domain (LBD) to inhibit activation by androgenic

hormones such as testosterone and dihydrotestosterone[95]. A biomarker for CRPC is the emergence of AR isoforms that splice out the LBD, and only consist of the DNA binding domain (DBD) and the disordered activation domain (AR AD) of the receptor[37]. The emergence of LBD-devoid AR splice variants suggests that inhibition of the AR AD could inhibit prostate cell proliferation in CRPC. The Salvatella took advantage of a previously described small molecule, EPI-001, derivatives of which have been investigated in the clinic[96], and greatly improved its inhibitory potency on AR transactivation by increasing its hydrophobicity (**Figure 2-17**). These results establish the basis for further development of such compounds for biochemical studies on the AR, and potential anti-CRPC drugs. Though the exact reasons why enhanced AR phase separation inhibits AR function remain elusive, these findings are consistent with the notion that enhanced condensation driven by hydrophobic sequence features, may compromise the ability of AR molecules to interact optimally with the rest of the transcriptional apparatus by sequestration into homotypic AR condensates (**Figure 2-18 B**).

In summary, the results in this thesis establish an experimental framework to understand the chemical features of transcription factor condensation, utilizing high-resolution microscopy and genomic-based approaches. I examine the effect of hydrophobic sequence modifications on transcription factor condensation in diverse genetic disease states, ranging from developmental disorders to a very lethal form of human cancer. Finally, I provide a generalizable drug-optimization strategy based on adjusting the chemical features of small molecules that selectively partition into transcription factor condensates.

4 Method Details

Discclaimer: The methods described in this section are taken from parts of Basu et al. 2020[8] and Basu et al. 2022[45], which I wrote and developed.

4.1 Plasmid generation

The following sections describe how the plasmids used in this thesis were generated. Plasmids have been made publicly available via addgene and the full sequences can be accessed via the following link:

URL: https://www.addgene.org/browse/article/28211176/

4.1.1 OptoDroplet vectors

All optoIDR constructs used in **Figures 2-2, 2-3, 2-9, and 2-10** were derived from the pHR-mCherry CRY2WT plasmid (Addgene, 101221) described in (Shin et al., 2017)[49]. To generate pHR mCherry-CRY2-NLS, the SV40 NLS was ordered as a primer pair (Sigma) (see **Table 4-1** for sequences), annealed, and ligated into the pHR-mCherry-CRY2 plasmid using NotI and SbfI restriction sites. To generate optoIDR constructs, sequences encoding the IDRs for HOXD13 wt, HOXD13 -7A, HOXD13- 15A, HOXD13 DEdel, HOXA13 wt, HOXA13 +7A, TBP 38Q, TBP 53Q, RUNX2 wt, RUNX2 +10A were ordered as synthetic DNA from commercial vendors (Genewiz and IDT, see **Table 4-1** for sequences). IDR sequences were flanked by AscI and NdeI restriction sites for cloning, and SpeI and BsiWI sites in the case of RUNX2wt and RUNX2+10A. The IDR-encoding DNA fragments were then ligated into pHR-mCherry-CRY2WT or pHR-mCherry-CRY2-NLS using AscI and NdeI restriction sites. In the case of *RUNX2*, a multiple cloning site was introduced into pHR mCherry-CRY2-NLS via AscI and NdeI to produce pHR-MCS-mCherry-CRY2-NLS and the *RUNX2* constructs were subsequently introduced via SpeI and BsiWI restriction sites. For insertion of the alanine expansions +7A, +8A, +9A, and +14A into the pHR-HOXD13-mCherry-CRY2-NLS construct, single-stranded oligonucleotides encoding respective alanine expansions were purchased (Sigma, see **Table 4-1** for sequences). Expansion oligonucleotides were then inserted into the alanine stretch of pHR-HOXD13-mCherry-

CRY2-NLS using Gibson assembly, using NotI cleavage and the NEBuilder HiFi DNA Assembly master mix, to generate pHR-HOXD13(+7A)-mCherry-CRY2-NLS, pHR-HOXD13(+8A)-mCherry-CRY2-NLS, pHR-HOXD13(+9A)-CRY2-NLS, pHR HOXD13(+14A)-CRY2-NLS. All constructs were sequence verified.

Table 4-1: Sequences of primers, expansion oligonucleotides, and gene fragments used to create OptoDroplet vectors.

Description	Sequence
SV40 NLS forward primer	/5Phos/GG CCG GAA CTC CCA CCT GCA ACA TGC GTG ACG GAG GCG GTC CAA AAA AGA AGA GAA AGG TAT GAC TGA GGC CGC GAC TCT AGA GTC GAC CTG CA
SV40 NLS reverse primer	/5Phos/GG TCG ACT CTA GAG TCG CGG CCT CAG TCA TAC C TT TCT CTT CTT TTT TGG ACC GCC TCC GTC ACG CAT GTT GCA GGT GGG AGT TCC
+7A oligonucleotide for Gibson assembly	GCGGCTGCTGCTGCTGCTGCAGCCGCGGCGgcagctgcagctgcggccgctGCCGCTAGTGGATTTGCCTATCCTGGGACGAG
+8A oligonucleotide for Gibson assembly:	GCGGCTGCTGCTGCTGCTGCAGCCGCGGCGgcagctgcagctgcggccgcagctGCCGCTAGTGGATTTGCCTATCCTGGGACGAG
+9A oligonucleotide for Gibson assembly:	GCGGCTGCTGCTGCTGCTGCAGCCGCGGCGgcagctgcagctgcggccgcagcagctGCCGCTAGTGGATTTGCCTATCCTGGGACGAG
+14A oligonucleotide for Gibson assembly:	GCGGCTGCTGCTGCTGCTGCAGCCGCGGCGgcagcagctgcagcggctgctgcagctgcggccgcagcagctGCCGCTAGTGGATTTGCCTATCCTGGGACGAG
HOXD13 wt gene fragment:	cgagctctataaaagagctcacaacccctcactcggcgcgccagtcctccgacagactgagtcgcccggggggatctggagctctcgagaattctcacgcgtcaagtggagcaaggcaggtggacagtggatccggagctaccATGTCCCGAGCCGGTAGCTGGGACATGGACGGCCTGCGAGCGGACGGAGGCGGAGCTGGAGGAGCTCCCGCATCAAGTAGTAGCTCAAGTGTGGCCGCTGCCGCGGCTAGCGGACAATGTAGGGGGTTTTTGTCAGCGCCTGTCTTTGCGGGCACACATTCCGGGAGGGCCGCTGCTGCGGCTGCTGCTGCTGCAGCCGCGGCGGCCGCTAGTGGATTTGCCTATCCTGGGACGAGTGAGCGCACTGG

TTCATCATCCTCATCTTCATCCAGTGCGGTAGTCGCCGCTCGGC
CAGAAGCACCACCTGCAAAAGAGTGTCCGGCCCCGACGCCAG
CTGCAGCAGCAGCGGCACCGCCTTCAGCTCCTGCGgtaccGGAG
GCGGGATGGTTTCCAAAGGAGAGGAGGATAATATGGCTATAA
TTAAAGAGTTTATGCGGTTCAAGgtgcatatgGAGG
GCTCCGTAAACGGTCATGAGTTC

HOXA13 wt gene fragment:

cgagctctataaaagagctcacaacccctcactcggcgcgccagtcctccgacagactgagtcgcccg
gggggggatctggagctctcgagaattctcacgcgtcaagtggagcaaggcaggtggacagtggatcc
ggagctaccATGGAGTTGAACAAAAACATGGAGGGGGCGGCTGCG
GCTGCAGCCGCGGCAGCGGCTGCAGCGGCTGCAGGAGCCGGT
GGAGGAGGTTTTCCCCATCCGGCGGCGGCTGCCGCGGGGGGGG
AATTTCAGTGTTGCGGCAGCTGCAGCAGCTGCAGCTGCTGCCG
CAGCTAACCAGTGCCGCAACCTGATGGCGCATCCAGCGCCTCT
TGCGCCGGGGGCCGCATCAGCGTACAGTTCTGCTCCTGGGGAA
GCACCCCCGTCCGCGGCTGCTGCCGCGGCAGCGGCAGCAGCA
GCCGCGGCTGCGGCCGCGGCAGCTAGCTCCAGTGGAGGTCCCG
GACCAGCGGGACCTGCTGGTGCGGAAGCCGCGAAGCAGTGTA
GCCCGTGCAGCGCGGCAGCTCAGAGCTCATCCGGTCCCGCCGC
CCTCCCATACGGATACTTTGGCTCAGGGTACTACCCGTGCGCG
AGAATGGGCCCTCACCCCAACGCCATAAAGTCATGTGCCCAAC
CCGCGTCAGCCGCAGCGGCAGCAGCTTTTGCCGATAAGTACAT
GGACACTGCTGGCCCCGCGGCGGAGGAGTTCAGTAGCAGGGC
GAAGGAGggtaccggaggcgggatggtttccaaaggagaggaggataatatggctataattaaa
gagtttatgcggttcaaggtgcatatggagggctccgtaaacggtcatgagttc

HOXA13 +7A gene fragment:

gctcacaacccctcactcggcgcgccagtcctccgacagactgagtcgcccgggggggatctggagc
tctcgagaattctcacgcgtcaagtggagcaaggcaggtggacagtggatccggagctaccATGG
AGTTGAACAAAAACATGGAGGGGGCGGCTGCGGCTGCAGCCG
CGGCAGCGGCTGCAGCGGCTGCAGGAGCCGGTGGAGGAGGTT
TTCCCCATCCGGCGGCGGCTGCCGCGGGGGGGAATTTCAGTGT
TGCGGCAGCTGCAGCAGCTGCAGCTGCTGCCGCAGCTAACCAG
TGCCGCAACCTGATGGCGCATCCAGCGCCTCTTGCGCCGGGGG
CCGCATCAGCGTACAGTTCTGCTCCTGGGGAAGCACCCCCGTC
CGCGGCTGCTGCCGCGGCAGCGGCAGCAGCAGCCGCGGCTGC
Ggcagctgcagctgcggccgct GCCGCGGCAGCTAGCTCCAGTGGAGGTC
CCGGACCAGCGGGACCTGCTGGTGCGGAAGCCGCGAAGCAGT
GTAGCCCGTGCAGCGCGGCAGCTCAGAGCTCATCCGGTCCCGC
CGCCCTCCCATACGGATACTTTGGCTCAGGGTACTACCCGTGC
GCGAGAATGGGCCCTCACCCCAACGCCATAAAGTCATGTGCCC
AACCCGCGTCAGCCGCAGCGGCAGCAGCTTTTGCCGATAAGTA
CATGGACACTGCTGGCCCCGCGGCGGAGGAGTTCAGTAGCAG

GGCGAAGGAGgtaccggaggcgggatggtttccaaaggagaggaggataatatggctata
attaaagagtttatgcggttcaaggtgcatatggagggatctgtgaacggtcacgagt

4.1.2 Bacterial fluorescent fusion protein vectors

For the generation of pET-45b-IDR constructs (used in **Figures 2-4, 2-5, 2-6, and 2-11**), pET-45b(+) was ordered from Novagen (Sigma Cat No. 71327). mCherry was subcloned into pET-45b(+) using the KpnI restriction site from an OptoIDR vector, to generate pET-45b-mCherry. IDRs were then PCR amplified from OptoIDR vectors using Q5 polymerase (NEB M0494S), and cloned into pET-45b mCherry using the NEBuilder HiFi DNA Assembly master mix (NEB E2621L). This pipeline was used to generate pET-45b-HOXD13-IDR, pET-45b-HOXD13(+7A)-IDR, pET-45b HOXD13(+10A)-IDR, pET-45b-RUNX2(WT)-IDR, pET-45b-RUNX2(+10A)-IDR, pET 45b-HOXA13(WT)-IDR, pET-45b-HOXA13(+7A)-IDR, and pET-45b-TBP38Q-IDR. **Table 4-2** contains primer sequences used to amplify HOXD-constructs (IDR-mCherry forward primer and IDR-mCherry reverse primer) and RUNX-constructs (RUNX2 IDR-mCherry forward primer and IDR-mCherry reverse primer).

The AR Tau-5 sequence was amplified from a mammalian expression vector encoding human AR (Addgene #29235) and subcloned with an mCherry insert into pET51b vector (Novagen 71553) using Gibson assembly (NEB E2611) to create the pET51b-AR-Tau-5- mCherry bacterial expression vector used in **Figures 2-15 and 2-16** (see **Table 5-2** for primer sequences).

pET28a-mEGFP-MED1-IDR vector was provided as a gift from the Young lab, and was used to generate pET28a-mCherry-MED1-IDR using Gibson assembly (Sabari 2019)[2]. HP1a and NPM1 ORFs were subcloned from murine stem cell cDNA into pET45b-mCherry (Addgene #145279) using Gibson assembly to create pET45b-mCherry-HP1a and pET45b-mCherry-NPM1 vectors used in **Figure 2-15** (see **Table 4-2** for primer sequences).

For AR Tau-5 mutagenesis, synthetic genes encoding cysteine point mutations were ordered from Genewiz, and then subcloned into pET51b-mCherry using Gibson assembly to create pET51b-AR-Tau-5-CtoFY-mCherry and pET51b-AR-Tau-5-C404Y-C518F-mCherry vectors used in **Figure 2-16** (see **Table 4-2** for sequences of synthetic genes).

Table 4-2: Sequences of primers and gene fragments to generate bacterial mCherry fusion vectors.

Description	Sequence
IDR-mCherry reverse primer	cgcaagcttcttgtacaattcatccatgc
IDR-mCherry forward primer	taccggtgacagtggatccggagctac
RUNX2 IDR-mCherry forward primer	taccggtGGaagagaggccACTAGTGC
AR Tau-5 forward primer:	aataattttgtttaactttaagaaggagatataccATGCTTTTGGGAGTTCCACCCGC
AR Tau-5 reverse primer:	atagccatgttatcctcctcgcctttagacaccatTGGAAAGTAATAGTCAATGGGC
NPM1 forward	primer:atagcacaggggcatggatgaattgtacaagtacacggATATGGAAGACTCGATGGATATGGACA
NPM1 reverse primer:	cgcagcagcggtttctttaccagactcgagtgcggccgcaACAAGAGATTTCCTCCACTGCCAG
HP1a forward primer:	catagcacaggggcatggatgaattgtacaagtacacggATATGGGAAAGAAGACCAAGAGGACAG
HP1a reverse primer:	cgcagcagcggtttctttaccagactcgagtgcggccgcACAGCTCTTCGCGCTTTCTTT
AR Tau-5 C404Y C518F synthetic gene:	CTTTTGGGAGTTCCACCCGCTGTGCGTCCCACTCCTTGTGCCCCATTGGCCGAATGCAAAGGTTCTCTGCTAGACGACAGCGCAGGCAAGAGCACTGAAGATACTGCTGAGTATTCCCCTTTCAAGGGAGGTTACACCAAAGGGCTAGAAGGCGAGAGCCTAGGCTGCTCTGGCAGCGCTGCAGCAGGGAGCTCCGGGACACTTGAACTGCCGTCTACCCTGTCTCTCTACAAGTCCGGAGCACTGGACGAGGCAGCTGCGTACCAGAGTCGCGACTACTACAACTTTCCACTGGCTCTGGCCGGACCGCCGCCCCCTCCGCCGCCTCCCCATCCCCACGCTCGCATCAAGCTGGAGAACCCGCTGGACTACGGCAGCGCCTGG GCGGCTGCGGCGGCGCAGTATCGCTATGGGGACCTGGCGAGCCTGCATGGCGCGGGTGCAGCGGGACCCGGTTCTGGGTCACCCTCAGCCGCC

	GCTTCCTCATCCTGGCACACTCTCTTCACAGCCGAAGAAG GCCAGTTGTATGGACCGTGTGGTGGTGGTGGGGGTGGTGG CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGAGGCGGGAGCTGTAGCCCCCTACGGCTA CACTCGGCCCCCTCAGGGGC TGGCGGGCCAGGAAAGCGA CTTCACCGCACCTGATGTGTGGTACCCTGGCGGCATGGTG AGCAGAGTGCCCTATCCCAGTCCCACTTTTGTCAAAAGCG AAATGGGCCCCTGGATGGATAGCTACTCCGGACCTTACGG GGACATGCGTTTGGAGACTGCCAGGGACCATGTTTTGCCC ATTGACTATTACTTTCCA
AR Tau-5 CtoFY synthetic gene:	CTTTTGGGAGTTCCACCCGCTGTGCGTCCCACTCCTTATGC CCCATTGGCCGAATACAAAGGTTCTCTGCTAGACGACAGC GCAGGCAAGAGCACTGAAGATACTGCTGAGTATTCCCCTT TCAAGGGAGGTTACACCAAAGGGCTAGAAGGCGAGAGCC TAGGCTTTTCTGGCAGCGCTGCAGCAGGGAGCTCCGGGAC ACTTGAACTGCCGTCTACCCTGTCTCTCTACAAGTCCGGA GCACTGGACGAGGCAGCTGCGTACCAGAGTCGCGACTAC TACAACTTTCCACTGGCTCTGGCCGGACCGCCGCCCCCTC CGCCGCCTCCC CATCCCCACGCTCGCATCAAGCTGGAGA ACCCGCTGGACTACGGCAGCGCCTGGGCGGCTGCGGCGG CGCAGTATCGCTATGGGGACCTGGCGAGCCTGCATGGCGC GGGTGCAGCGGGACCCGGTTCTGGGTCACCCTCAGCCGCC GCTTCCTCATCCTGGCACACTCTCTTCACAGCCGAAGAAG GCCAGTTGTATGGACCGTGTGGTGGTGGTGG GGGTGGTG GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG GCGGCGGCGGCGAGGCGGGAGCTGTAGCCCCCTACGGCT ACACTCGGCCCCCTCAGGGGCTGGCGGGCCAGGAAAGCG ACTTCACCGCACCTGATGTGTGGTACCCTGGCGGCATGGT GAGCAGAGTGCCCTATCCCAGTCCCACTTTTGTCAAAAGC GAAATGGGCCCCTGGATGGATAGCTACTCCGGACCTTACG GGGACATGCGTTTGGAGACTGCCAGGGACCATGTTTTGCC CATTGACTATTACTTTCCA

4.1.3 Mammalian mEGFP Expression Vectors

Monomeric EGFP was subcloned into vectors containing human AR (Addgene #29235) and AR-V7 (Addgene #86856) using Gibson assembly to create mEGFP-AR-FL and mEGFP AR-V7 mammalian expression vectors. The sequence downstream of the AR activation domain in mEGFP-AR-V7 was subcloned into an mEGFP expression vector

(Addgene #18696) using Gibson assembly to create the mEGFP-AR-V7-ΔAD expression vector. Primer sequences can be found in **Table 4-3**.

Table 4-3: Primer sequences to create mammalian mEGFP expression vectors.

Description	Sequence
AR-V7 ΔAD forward primer:	agttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagctgatctgtggagatgaagc
AR-V7 ΔAD reverse primer:	caataaacaagttgggccatggcggccaagcctctacaaatgtggtatggc

4.2 Expression and purification of fluorescent proteins

For fluorescent fusion proteins used in **Figures 2-4, 2-5, 2-6, and 2-11**, plasmids were transformed into BL21(DE3) (NEB M0491S) cells, and grown in home-made automatic-induction medium (AIM), as described in[97]. In brief, 10 mL of overnight culture in minimal media (MDG) was used to seed 100 mL of AIM (ZYM-5052) supplemented with ampicillin and trace metals. This culture was incubated at 37°C with vigorous shaking for 2 hours, after which the temperature was brought down to 16°C. Cells were harvested when the cultures became dark magenta or bright green (24-48 hours later), after which pellets were frozen at -80°C for at least 16 hours. For protein purification, pellets were resuspended in 30 mL of Buffer A (50 mM Tris pH 7.5, 300 mM NaCl) supplemented with 8M Urea (Sigma, U5378) and cOmplete protease inhibitors (Sigma, 11697498001). This denatured suspension was sonicated and clarified by centrifugation at 20,000 g for 20 minutes at 4°C. Supernatants containing fusion proteins were loaded onto a His GraviTrap column (GE HealthCare, 11003399) pre-equilibrated in Buffer A supplemented with 8 M Urea. The loaded column was washed with 30 Column Volume (CV) of 6% Buffer B (50 mM Tris pH 7.5, 300 mM NaCl, 500 mM Imidazole) in Buffer A supplemented with 8 M Urea, and another 20 C.V. of 6% Buffer B to remove denaturant. Proteins were eluted in 3 C.V. of 50% Buffer B, immediately diluted 1:2 in storage buffer (50 mM Tris pH 7.5, 125 mM NaCl, 1 mM DTT, 10% Glycerol), and then concentrated using 30 MWCO centrifugal filters (Merck, UFC803008) by spinning at 7500 g for 30 minutes at 4°C. The resulting fraction was then diluted 1:100 in storage buffer, re-concentrated, and then stored at -20°C. All fusion proteins were purified in the same manner.

mCherry fusion proteins used in **Figures 2-15 and 2-16** were expressed and purified as similarly as described in [8] with the following modifications: 1) Proteins were expressed in Rosetta(DE3)pLysS cells (Sigma 70956) in ZYM5052 media spiked with the appropriate antibiotic and chloramphenicol 2) lysis was performed in non-denaturing conditions 3) AR Tau 5 mutagenesis proteins were gel-filtered using size exclusion chromatography (GE-Healthcare GE28-9909) and 4) eluates were stored in TCEP containing buffer (10% glycerol [w/v], 125 mM NaCl, 50 mM Tris pH 8.0, 1 mM TCEP, 1 mM PMSF, 1x cOmplete inhibitor cocktail) at -80°C prior to experimentation.

4.3 Drug treatment of proteins for droplet experiments and mass spectrometry

For fluorescent fusion protein experiments shown in **Figure 2-15**, EPI-001 (Selleckchem Lot #S795502) was dissolved in analytical grade DMSO (Sigma 94563) to a final concentration of 50 mM. Stocks were aliquot frozen and stored at -80°C. For *in vitro* reactions, EPI-001 stocks were thawed on ice and brought to a 1 mM intermediate concentration in 100 µL of reducing buffer (125 mM NaCl, 50 mM Tris pH 8.0, 1 mM TCEP, 1 mM PMSF, 1x cOmplete inhibitor cocktail). Analytical grade DMSO was also diluted in 100 µL of reducing buffer (1:50 [v/v]) to create a vehicle control. EPI-001 and DMSO intermediates were then cut 1:2 v/v with recombinant protein preparations of mCherry tagged NPM1, MED-IDR, HP1α, or AR Tau-5 in 60 µL of reducing buffer to ensure at least 10 molar excess of EPI-001. Reactions were then incubated for 16 hours at 37°C, before subjection to mass spectrometry and confocal microscopy. Each target was reacted with EPI-001 or DMSO control at least three times on separate days. Proteomics sample preparation of human AR Tau 5 probes from biological triplicates was performed with the preOmics in-stage tip kit (iST kit 96x, Martinsried, Germany), according to the manufacturer's protocol. Each sample was dissolved in 38 µl LC-Load buffer provided by the kit, and half of it was injected on a nano LC-MS/MS system. To avoid any carry-over of peptides, one wash was always run in between all individual samples.

4.4 Isolation of limb bud cells

Limb buds from E12.5 wild type and spdh homozygous embryos used in **Figures 2-7 and 2-8** were micro-dissected individually and digested with Trypsin-EDTA 0.05% (Gibco) for 15 minutes at 37°C and gently dissociated by pipetting after 5, 10, and 15 minutes. Cells were mixed in cell culture media (DMEM, 10% FBS, 2 mM L-Glutamine, 50 U/ml Penicillin/Streptomycin), and a single-cell suspension was obtained using a 40 µm cell strainer (Falcon). About 150,000 limb bud cells from each embryo were seeded in 1000µl medium onto fibronectin-coated glass coverslips in 12 well plates. After 30-60 minutes, additional cell culture medium was added, and cells were grown for 24 hours. After 24 hours, the cells were rinsed twice with PBS and fixed for 15 minutes at room temperature with 4% PFA/PBS.

4.5 Cell treatments

For transient transfection and live cell imaging in **Figures 2-2, 2-3, 2-7, 2-9, 2-10, and 2-10,** HEK-293T cells were seeded onto chambered coverslips (Ibidi, 80826-90), and transiently transfected 20-24 h later using lipoD293 (Signagen, SL100668) according to the manufacturer's protocols. Each transfection series was repeated at least twice.

For 1,6-hexanediol (1,6-HD) treatment in **Figure 2-8**, isolated limb bud cells attached to fibronectin-coated glass coverslips were treated with 6% or 0% 1,6-hexanediol (Sigma, 240117) in 1 ml cell culture media for 1 minute at 37°C. After treatment, cells were washed with PBS and fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) for 10 minutes, and stored at 4°C until processing for immunofluorescence and microscopy. Limb bud cells isolated from two pups per genotype were used for 1,6-HD treatment.

4.6 Western blot

For western blot in **Figure 2-1 C**, Cells were washed once with ice-cold PBS, and lysed in RIPA buffer (Thermo, 88900) supplemented with protease inhibitors (Thermo, 87786). Protein concentration was determined using bicinchoninic acid assay (Thermo, 23225), according to the manufacturer’s protocol. Equal protein amounts were brought to the same volumes with lysis buffer, reducing agent, and LDS loading buffer. Lysates were then

heated to 95°C for 10 minutes, and separated on 4-12% Tris-acetate gels (Invitrogen, NP0322BOX). Protein was transferred onto a nitrocellulose membrane in an iBlot 2 apparatus. After transfer, the membrane was blocked with Licor blocking buffer (Licor, 927-500000) for 1 hour at room temperature with shaking. The membranes were incubated with 1:750 anti-Hoxd13 (abcam ab229234) and 1:3000 anti-HSP90 (BD Biosciences, 610419) antibody diluted in 5% non-fat milk in TBST overnight at 4°C with shaking. The next day, membranes were washed three times with TBST for 5 minutes at room temperature, incubated with fluorescent anti-mouse (IRDye 800CW Donkey anti-Mouse, Li Cor P/N 925-32212) and anti-rabbit (IRDye 680LT Donkey anti-Rabbit, Li-Cor P/N 925- 68023) secondary antibodies at 1:10000 dilution according to the manufacturer's protocol, and washed three times in TBS-T for 5 minutes in the dark. Membranes were imaged on a LICOR Odyssey Clx imager.

4.7 Immunofluorescence and confocal microscopy

Isolated limb bud cells, Kelly cells, and SH-SY5Y cells shown in **Figure 2-1 B and C** were attached to coated glass coverslips were fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) in PBS for 10 min and stored at 4°C in PBS or processed immediately. After two washes in PBS, cells were permeabilized in 0.1% Triton X-100 (Thermo Scientific, 85111) in PBS for 10 min at room temperature (RT). Following three washes with PBS, cells were blocked with the blocking solution (10% fetal bovine serum (FBS), 0.1% Triton X-100 in PBS) for 30 min. Cells were incubated with the primary antibody (anti-HOXD13 Invitrogen PA5-66661 1:250 dilution or anti-HOXD13 abcam ab19866 1:150 dilution) in blocking solution at 4°C overnight. After three 10 min washes in PBS, cells were incubated with the secondary antibody (donkey anti-Rabbit IgG Alexa Fluor 568 Invitrogen A10042 1:1000 dilution) in blocking solution for 1h at room temperature in the dark. Cells were washed three times in PBS, and nuclei were counterstained with 0.24mg/mL 4’,6-diamidino-2-phenylindole (DAPI) in PBS for 3 min at RT in the dark. Following five washes in PBS, coverslips were mounted onto slides with Vectashield (Vector, H-1000) and sealed using transparent nail polish. Images were acquired at the confocal laser scanning microscope (Zeiss LSM880, 63x oil objective, NA 1.4, 1 Airy Unit). Raw images (.czi files) were processed in FIJI (Schindelin et al., 2012). Expression of HOXD13 protein in nuclei of limb

bud cells, Kelly cells, and SH-SY5Y cell lines was examined, as displayed in **Figure 2-1 B**. The abcam ab19866 antibody was used in **Figure 2-1 C.**

HEK293T cells in DMEM 10% FBS were seeded at a density of 40,000 cells / well on glass bottom chambered coverslips (Ibidi 80827). Sixteen hours later, wells were refreshed with 280 µL seeding media and transfected with 50 nanograms of mEGFP expression plasmids shown in **Figure 2-14** using LipoD293 transfection reagent (SignaGen SL100668) according to the manufacturer's protocol. Forty-eight hours later, wells were refreshed with media spiked with 10 nM DHT or equivalent DMSO control (v/v). Four hours after treatment, coverslips were imaged on a LSM880 microscope equipped with a 63x oil DIC objective in a CO_2 incubation chamber and heated stage set to 37°C. 8 - 10 image fields were acquired across two biological replicates for each condition (transfection and treatment).

4.8 *In vitro* droplet formation microscopy

For *in vitro* droplet formation experiments in **Figures 2-4, 2-5, 2-6, 2-11, 2-15, and 2-16** recombinant mCherry or mEGFP fusion proteins were measured for concentration, and then diluted or mixed to desired concentrations in storage buffer. For microscopy, these solutions were further diluted 1:2 in either 20% or 10% PEG-8000 in de-ionized water (w/v). 10 µl of this suspension was pipetted onto chambered coverslips (Ibidi, 80826-90), and immediately imaged using a LSM880 confocal microscope equipped with a 63x1.40 oil DIC objective and AiryScan detector. All images were acquired from within the solution interface, and performed before droplets settled onto the bottom of the coverslip, as described in[2,3]. For droplet assays using preformed GFP-MED1 IDR condensates, GFP-MED1 IDR droplets were allowed to form for 30 minutes at room temperature in the presence of 10% PEG-8000 in protein Lo-bind tubes (Eppendorf, 30108094), before proceeding with co-condensation assays. For compound treatments of co-condensates shown in **Figure 2-6**, small molecules [ATP (Jena Bioscience, NU 1010), Mitoxantrone dihydrochloride (Sigma M6545), (±)-α-Lipoic acid (Sigma 62320), and Lipoamide (Medchemexpress HY-B1142)] were directly diluted into droplet mixtures from appropriate stocks with or without vehicle (DMSO, Sigma D2650) to the desired final concentrations, and immediately imaged using confocal microscopy. Droplet formation experiments in **Figures 2-4, 2-6, and 2-11** were repeated at least two times on separate days. Experiments in **Figure 2-5** were repeated seven times on separate days.

For *in vitro* droplet formation experiments in **Figures 2-15 and 2-16**, assays using mCherry tagged proteins were performed as described in Basu et al. 2020[8]. In brief, protein preparations were mixed 1:2 (v/v) with 20% PEG8000 and kept at room temperature for 3 minutes before spotting on chambered coverslips (Ibidi 80826). Suspensions were imaged from the center of each chamber and within the solution interface using a LSM880 microscope equipped with a 63x1.40 oil DIC objective. Droplet imaging using nuclear IDP and EPI-001 incubations and AR Tau-5 mutagenesis constructs were repeated at least three times on separate days.

4.9 OptoDroplet microscopy

All live cell imaging experiments shown in **Figures 2-2, 2-3, 2-9, and 2-10** were performed on an LSM880 confocal microscope (Zeiss) equipped with an incubation chamber with a heated stage at 37°C. Images were acquired with either a Plan-Apochromat 40x0.95 Korr M27 or a 63x1.40 oil DIC objective.

The optoDroplet assay was adapted from (Shin et al., 2017)[25]. Briefly, approximately 20,000 cells were seeded per well on chambered coverslips one day before transfection. The following day, cells were transfected with optoIDR constructs. Cells in each well were transfected with 200ng of plasmid encoding for indicated optoIDR constructs. Forty-eight hours post-transfection, culture medium was refreshed, and cells were imaged on a Zeiss LSM 880 confocal microscope. Droplet formation was induced with scans with the 488 nm laser every 20 seconds for the duration of imaging. For image acquisition, mCherry fluorescence was stimulated at 561nm laser every 20 seconds. The constructs used for the optoDroplet experiments described in **Figure 2-2** did not include the SV40 NLS sequence, and the fusion protein displayed cytoplasmic localization. All other data was generated using constructs that included the SV40 NLS, and the fusion proteins displayed nuclear localization.

For FRAP experiments of light-induced droplets, formation was induced with continuous 488 nm light for 90 seconds. Droplets were then bleached with 561 nm light, and recovery was imaged every 4 seconds in the presence of simultaneous 488 nm light stimulation. Bleaching was performed on a region enclosing a single droplet using 2

iterations of 100% laser power. All experiments using optoIDR constructs were repeated at least two times.

4.10 Stochastic optical reconstruction microscopy (STORM)

For nuclei visualization in **Figure 2-1 B**, switching buffer was supplemented with 5 nM Sytox Orange (Thermo, S11368). The BRD4 antibody was validated in-house using dBET6-induced degradation of endogenous BRD proteins. Images were acquired with a Vutara 352 super resolution microscope (Bruker) equipped with a Hamamatsu ORCA Flash4.0 sCMOS for super-resolution imaging and a 60x oil immersion TIRF objective with numerical aperture 1.49 (Olympus). Data were acquired with TIRF/HILO-illumination at a laser-power density of 62.5 kW/cm2 using an 532 and 639 nm laser. Images for co-IF STORM analyses were acquired on three biological replicates (three limb bud samples from E12.5 mouse embryos for each genotype, harvested from three independent mouse matings).

For STORM composites shown in **Figure 2-7 C**, isolated wild type and homozygous spdh E12.5 limb bud cells attached to fibronectin-coated glass coverslips were fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich, P6148) for 20min at room temperature. After 2 x washes in PBS, fixed samples were treated with permeabilization solution (PBS supplemented with 0.1% Triton X-100) for 10 minutes at room temperature. Samples were then treated with blocking solution (permeabilization solution supplemented with 10% fetal bovine serum) for 30 min. After blocking, samples were incubated with primary anti-Hoxd13 antibody (Thermo, PA5-66661, 1:250 in blocking solution), BRD4 antibody (Clone A-7, Santa Cruz, sc-518021, 1:250 in blocking solution), or HP1α antibody (clone15.19s2, Millipore/Merck, 1:250 in blocking solution) overnight at 4°C, and then washed 3x in PBS. Stained and washed samples were then incubated with secondary antibody (goat anti-Rabbit IgG Alexa Fluor 647 Invitrogen A32733 / goat anti-Rabbit IgG Cy3 Invitrogen A21235, 1:1000 in blocking buffer for Hoxd13 primary and goat anti-Mouse IgG Alexa Fluor 647 Invitrogen A10520, 1:1000 in blocking buffer for BRD4 / HP1α primary) for 45 minutes, after which samples were washed 3 x in PBS. For imaging, samples were placed in a one-well magnetic chamber, covered in switching buffer consisting of 0.15 M 2-mercaptoethanol/0.2 M Tris, pH 8.0 with 4 % (w/v) glucose, 0.25 mg/ml glucose-oxidase, 20 µg/ml catalase.

4.11 Stimulated emission depletion microscopy (STED)

4.11.1 Sample preparation

For STED experiments in **Figure 2-12 and 2-14**, HEK293T and HeLa eGFP-AR cells in DMEM 10% FBS were seeded at a density of 40,000 cells / well on glass bottom chambered coverslips (Ibidi 80827). Sixteen hours later, wells containing HEK293T cells were refreshed with 280 µL seeding media and transfected with 50 nanograms of mEGFP expression plasmids shown in **Figure 2-14** using LipoD293 transfection reagent (SignaGen SL100668) according to the manufacturer's protocol. Forty-eight hours later, wells were refreshed with media spiked with 10 nM DHT. Samples were imaged after 4 hours of DHT treatment.

LNCaP cells (Clone FGC, ATCC CRL-1740) were seeded in RPMI1640 5% FBS onto PLL coated 18mm #1.5 glass covers slips (Sigma P4707, Roth LH23.1) at a density of 100,000 cells / cover slip in a 24 well plate. Sixteen hours later media was refreshed, and cells were incubated further for another 24 hours. For fixation, wells were washed with PBS, then fixed with 1 mL of 4% PFA in PBS for 20 minutes at room temperature. After a second wash in PBS, cells were permeabilized with 0.5% Triton-X, PBS (v/v) (Sigma 93443) and then stained with 1:50 AR primary antibody (AR 441, scbt 7305) and 1:200 STAR 635P secondary antibody (Abberior, ST635P-1001). AR translocation vas validated by staining LNCaP cells grown in RPMI 1640 5% CSS (Gibco, A3382101) with the same protocol. DNA was counterstained with 1:000 Spy555-DNA (spirochrome, SC201) and sample mounted onto glass cover slides in vectashield (Biozol, VEC-H-1900-10).

4.11.2 Live-Cell STED

HEK293T and HeLa cells shown in **Figure 2-14** were imaged on a Leica Stellaris STED DMI 8 microscope equipped with an okolab incubation chamber set to 37°C and 5% CO_2 constant. EGFP imaging was performed using a 473 nm stimulation wavelength laser at 20% power and a 592 nm depletion laser at 20% power. Images were taken using an HC PL APO CS2 63x/1.40 oil objective with a final resolution of 23 nanometers / pixel.

4.11.3 STED FLIM

Fixed and stained LNCaP cells shown in **Figure 2-12** were imaged on a Leica Stellaris STED DMI 8 microscope. Abberior STAR 635P immuno-fluorescence imaging was performed using a 633 nm stimulation wavelength laser at 5% power and a 776 nm depletion laser at 5% power. Images were taken using an HC PL APO CS2 63x/1.40 oil objective with a final resolution of 48 nanometers / pixel. FLIM cutoffs and τ-STED deconvolution strengths were determined automatically using Leica LAS-X software v 2.5.6 to filter background photons with low fluorescence lifetimes.

4.12 Proliferation microscopy

For dose-response curves in **Figure 2-19**, LNCaP cells (Clone FGC, ATCC CRL-1740) in DMEM with 5% FBS were seeded at a density of 4000 cells / well into 96 well flat bottom plates pre-coated with poly-L-lysine (Sigma P4707). Twenty-four hours later, triplicate wells were refreshed with 100 µL of seeding media spiked with 7x serial dilutions of EPI-001 (from 200 µM) (Selleckchem Lot #S795502), 1ae (from 50 µM), or DMSO control, at 0.5% DMSO (v/v) constant. Ninety-six hours later, wells were washed with 200 µL PBS and then fixed with 100 µL of 4% PFA in PBS for 20 minutes at room temperature. Post fixation, LNCaP nuclei in each well were counterstained using 100 µL of Hoescht 33342 (abcam ab228551) diluted to 1:4000 in PBS for 20 minutes at room temperature. After a final wash in PBS, plates were imaged using a Celldiscoverer 7 microscope equipped with a 20x air objective. Twenty-five tile regions (5 x 5 tiles) / well were imaged for each technical replicate (triplicate wells for each dose and compound). Raw LNCaP nuclei counts, assayed as objects detected by automatic otsu thresholding (ZEN Blue v 3.2) on Hoescht signals from in each well, were used to construct dose-response curves for EPI-001 and 1ae. Data were acquired across 2 biological replicates performed on separate weeks.

4.13 Quantitative real-time polymerase chain reaction (qRT-PCR)

LNCaP cells (Clone FGC, ATCC CRL-1740) in DMEM with 5% FBS or 5% CSS were seeded at a density of 300,000 cells / well in 6 well plates. Sixteen hours later, wells were refreshed with seeding media as indicated in **Figure 2-13** and CSS 6 well plates or EPI-001 at the doses indicated in **Figure 2-15** and DMSO control, at 0.5% DMSO (v/v) constant.

After 6 and 24 hours, media was removed and cells were harvested using 300 µL of TRIzol reagent (Thermo 15596026) for each well. RNA was then extracted using a Zymo DirectZol Micro kit (Zymo R2062) according to the manufacturer's protocols. cDNA was synthesized using 1 µg of the resulting RNA and random hexamer primers from the Thermo Scientific RevertAid First Strand cDNA Synthesis (Thermo K1622) kit. cDNA harvested from LNCaP cells treated at each compound, dose, and time point was then probed for transcript targets on 384 well plates using the SYBR Green master mix, and a QuantStudio 7 real-time qPCR machine. The target primer sequences used to probe transcript levels are shown in **Table 4-4**. For calculation of fold change ($2^{-\Delta\Delta Ct}$), Ct values from target regions were normalized to Ct values from control regions in treatment sample, and then normalized against DMSO control sample. Data for compound experiments were acquired across 3 biological replicates performed on separate days.

Table 4-4: Sequences of target primers used in qRT-PCR.

Description	Sequence
FKBP5 forward primer 1:	GCAACAGTAGAAATCCACCTG
FKBP5 forward primer 2:	CTCCAGAGCTTTGTCAATTCC
FKBP5 forward primer 2:	AGGAGGGAAGAGTCCCAGTG
FKBP5 reverse primer 2:	TGGGAAGCTACTGGTTTTGC
PSA (KLK3) forward primer 1:	TGTGTGCTGGACGCTGGA
PSA (KLK3) reverse primer 1:	CACTGCCCCATGACGTGAT
PSA (KLK3) forward primer 2:	AGGCCTTCCCTGTACACCAA
PSA (KLK3) reverse primer 2:	GTCTTGGCCTGGTCATTTCC
TMRPSS2 forward primer:	GGACAGTGTGCACCTCAAAGAC
TMPRSS2 reverse primer:	TCCCACGAGGAAGGTCCC
β Glucoronidase forward primer:	CTCATTTGGAATTTTGCCGATT
β Glucoronidase reverse primer:	CCGAGTGAAGATCCCCTTTTTA

4.14 ChIP-sequencing data generation

For H3K27Ac ChIP shown in **Figure 2-8**, isolated E12.5 limb bud cells (hand plate) from 2 wild type, 2 homozygous spdh, and 2 heterozygous spdh mice (approximately 3.5 million cells each) across 2 separate het x het mouse crosses were fixed in 10 minutes in 1% formaldehyde on ice and quenched with 2.5M glycine. Each pool of 3.5 million cells (wt, homozygous spdh, and heterozygous spdh) was spiked in with 1 million fixed fly S2 cells (Orlando et al., 2014), after which cell suspensions were lysed in lysis buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% Glycerin, 0.5% NP-40, 0.25 Triton X-100, pH 7.5 supplemented with protease inhibitors and Na-Butyrate). After assessing for optimal shearing conditions (47 cycles of 30 seconds of sonication at high settings using Diagenode Bioruptor) and total chromatin amount, 10-15 µg of sheared chromatin was incubated with 4 µL of H3K27Ac antibody (Diagenode C15410174) in a total of 1.2 mL of buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na DOC, 0.5% N-Laroylsarcosine, pH 8.0 supplemented with protease inhibitors and Na-Butyrate) overnight with gentle rotation at 4°C. After incubation, 30 µL of Protein G beads were added to the chromatin and antibody suspensions, and allowed to incubate overnight with gentle rotation at 4°C. Samples were then washed 7 times with 1 mL of RIPA buffer (to 50 mM HEPES-KOH, 1 mM EDTA, 1% NP-40, 0.7% Na-DOC, 500 mM Li-Cl, pH 7.55 supplemented with protease inhibitors and Na-Butyrate). Beads were then washed with TE buffer (supplemented with protease inhibitors and Na-Butyrate) and then centrifuged down to remove TE buffer. Chromatin was then eluted using 210 µL of elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS, pH 8.0) and heated to 65°C with vigorous shaking. Eluates were then treated with 5 µL of Proteinase K overnight at 65°C. The next day, 4 µL of RNAse A was added to the samples, vortexed, spun down and incubated at 37°C for 30 minutes. Chromatin was then extracted using phenol-chloroform, and precipitated using 70% ice-cold EtOH, with centrifugation at max speed for 10 minutes. Supernatant was removed, and the resulting pellet was dissolved in 30-50 uL of de-ionized water. Total yield was assessed by Qubit, and then sent for sequencing. Respective ChIP samples and input controls for wild type, homozygous spdh, and heterozygous spdh were paired-end sequenced using Illumina high-throughput sequencing, to a depth of around 25 million reads.

H3K27Ac ChIP sequencing reads from DMSO-treated LNCaP cells were acquired from GSE125245, published in Rasool et al. 2019[80]. For each androgen receptor ChIP

sequencing experiment shown in **Figure 2-13**, 100 million LNCaP cells were grown in 6 x 15-centimeter plates, performed on separate weeks across two biological replicates. Cells were treated with either vehicle control (0.5% DMSO v/v) or 25 μM EPI-001 (Selleckchem Lot #S795502) for 24 hours. After treatment, cells were fixed in 1% formaldehyde for 10 minutes at room temperature with gentle agitation. The fixation reaction was quenched with glycine, after which plates were washed once with ice-cold PBS. Cells were then harvested in 2 mL of PBS and protease inhibitors on ice. The resulting suspensions were centrifuged for 5 minutes at 2000 g at 4°C, after which supernatant was decanted and pellets stored at -80°C. After harvesting all pellets across biological replicates and treatments, pellets were processed together as described above, with the following modifications: For lysis of cross-linked pellets, 7 cycles of 30 seconds of sonication at high settings was performed using Diagenode Bioruptor. Subsequently, chromatin immunoprecipitations were performed using 5 μL of AR antibody (PG-21, EMD Millipore, 06-680) and 50 μL of Protein A beads (Thermofischer, 10001D), without Na Butyrate in lysis and wash buffers. Respective ChIP samples and input controls for DMSO and EPI-0001 treated LNCaP cells were paired-end sequenced using Illumina high-throughput sequencing, to a depth of 50 million reads to construct sequencing libraries.

4.15 RNA-sequencing data generation

LNCaP cells (Clone FGC, ATCC CRL-1740) in DMEM with 5% FBS were seeded at a density of 300,000 cells / well in 6 well plates. Sixteen hours later, wells were refreshed with seeding media spiked with either EPI-001 or 1ae at the doses indicated in **Chapters 2-19** and DMSO control, at 0.5% DMSO (v/v) constant. After 6 and 24 hours, media was removed and cells we harvested using 300 μL of TRIzol reagent (Thermo 15596026) for each well. RNA was then extracted using a Zymo DirectZol Micro kit (Zymo R2062) according to the manufacturer's protocols. Total RNA-seq libraries were constructed using the KAPA RNA HyperPrep Kit with RiboErase (Roche KR1351) according to the manufacturer's instructions using 1 μg of RNA from each sample replicate, using 10 amplification cycles. Libraries were sequenced on a NovaSeq with paired-end reads of 200 base pairs, with a read depth of 50 million fragments for each library. Triplicate libraries were prepared for each treatment regime (time, dose, compound) from three corresponding biological replicates.

5 Quantification and Statistical Analysis

Disclaimer: The analyses described in this section, are taken from parts of Basu et al. 2020[8] and Basu et al. 2022[45], which I wrote and developed.

5.1 Pairwise comparisons

Pairwise comparisons shown in **Figures 2-1 to 2-21** were performed with Student's t-test or Mann-Whitney U test, as indicated in figure legends, in base R, after assessing for appropriate sample distribution. Differences were considered significant when adjusted p values were less than 0.0001 (****), 0.001 (***), 0.01 (**), or 0.05 (*).

5.2 Dose-response analysis

Raw LNCaP nuclei counts from proliferation experiments in **Figure 2-19 A**, assayed as objects detected by automatic otsu thresholding on Hoescht signal from each well, were used to construct dose-response curves for EPI-001 and 1ae. Segmentation was performed using ZEN Blue version 3.2 on image data acquired across 2 biological replicates performed on separate weeks. Nuclear counts from each well were exported and processed using the DRC[98] package in R to create dose-response curves. Data was modeled with a three-parameter log-logistic function (lower limit 0), and the resulting fit was used to calculate IC_{50} and IC_{10} values for EPI-001 and 1ae, as shown in **Figure 2-19 B**.

5.3 Mass spectrometry analysis

Raw MS data were processed with MaxQuant software (v2.0.1.0)[99] and searched against the human AR amino acid sequence. The following variable modifications on cysteines were included: EPI-001 ($H_{24}C_{26}O_5$). For each cysteine in AR-Tau5 shown in **Figure 2-16 A**, the number of EPI-modified (N_{mod}) specific peptide matches and the number of all (N_{all}) peptide matches was mapped to AR, from which the modification ratio was calculated as follows:

$$r = N_{mod} / N_{all}.$$

5.4 *In vitro* droplet image analysis (c_{sat})

For analysis of *in vitro* droplet experiments in **Figures 2-4, 2-5, 2-6, and 2-11**, 4 to 8 image fields were used for each condition. Droplets were first thresholded by manually gating on pixel intensity in either mEGFP or mCherry channels using Zen Blue 3.0. For preassembled MED1 IDR-GFP scaffold experiments (**Figures 2-5 and 2-6**) and TF IDR mutant phase diagrams (**Figures 2-4 B, 2-11 C and E**), droplet regions were automatically detected by three sigma thresholding on mCherry in Zen Blue. To optimize detection, droplets were also filtered by size and pixel deviation, after which mean fluorescence intensities, area, and diameter of these regions were measured on both channels. For data visualization in **Figures 2-5 B** and **2-6 B to E**, values for diameter, mean intensity, and area were plotted as described in figure legends using R-Studio and the Directlabels package.

For analysis of *in vitro* droplet assays in **Figures 2-15 and 2-16**, with mCherry fusion proteins, droplets were segmented using automatic thresholding (three sigma thresholding) on the mCherry using Zen Blue version 3.2. A secondary region (4-pixel wide ring with a 1-pixel gap from the detected droplet) was used to assay any background signal in the immediate vicinity of the droplet. Droplet and ring regions were measured for mean fluorescence intensities and exported for processing in R. Partitions ratios were calculated as the mean intensity of the droplet region / mean intensity of the corresponding ring region. In **Figure 2-15 I**, the mean partition ratio of droplets in DMSO was used to normalize the partition ratio of droplets in EPI-001. Normalized partition ratio distributions for the indicated nuclear IDP were then compared with pair-wise student's t-test. Droplet values were assayed from 10 image fields for each condition. In **Figure 2-16 C**, phase diagrams of mean droplet intensity as a function of titrated protein concentration were modeled using a four-parameter log-logistic function using the DRC package in R. C_{sats} were approximated as IC_{10} values calculated from the resulting dose-response fits. Droplets values were segmented from 5 image fields for each condition.

5.5 OptoDroplet image analysis (phase-shift ratio)

Image analysis and data wrangling for optoDroplet assays were performed in Zen Blue 3.0 (Zeiss) and R. For optoDroplet analysis in Zen Blue 3.0, nuclear and cytoplasmic mCherry signals were detected by Otsu thresholding and then size-filtered to define primary

regions of interest (ROI). Within these primary regions, optoDroplets were detected using a second fixed pixel intensity threshold. Nuclear mCherry signals were automatically thresholded and size filtered to define primary regions of interest. OptoDroplets were detected as secondary regions of interest by exceeding the primary region intensity threshold limits. All parameters were determined empirically by adjusting the analysis pipeline on 4-5 raw images for each genotype. The pipeline was then used to analyze at least 50 – 200 cells per genotype (indicated as "n" in each figure panel). The phase-shifted fraction was calculated as the total area of detected optoDroplets within a nucleus divided by the area of the corresponding cell's nucleus (**Figures 2-3 C, 2-9 B, and 2-10 B**). This fraction was then averaged over all detected cells (indicated as "n" in each figure panel) for each genotype, and plotted over time (curves shown in **Figures 2-3 D, 2-8 C, 2-10 C**)

Phase-shift ratio at time point t

$$= \frac{\sum_{i=1}^{n} \left\{ \frac{area\ of\ optoDroplets\ within\ ROI\ at\ time\ t_i}{area\ of\ ROI_i} \right\}}{n}$$

In **Figure 2-2 C and F**, where non-nuclear IDR constructs were used, the phase-shifted fraction was calculated as the total area of detected optoDroplets within a cell divided by the area of the corresponding cell. Cells were filtered for expression level (outlier cells were removed). The ratios were averaged over all detected cells per the indicated optoIDR genotypes, and plotted over time. Finally, to compare cell populations that express the optoIDR constructs at similar expression levels across genotypes, a minimum and maximum expression level filter was used for each experiment series. The expression level of fluorescent fusion proteins within cells was quantified as the mean intensity of the fluorescence signal within the primary regions of interest (nucleus or whole cell) determined as described above.

For FRAP analysis of OptoIDR constructs (**Figure 2-2 G**), mean pixel intensity of regions of interest were measured using FIJI, and normalized to pre-bleaching intensity. Captured intensity traces were averaged over multiple replicates for each genotype (indicated as n in figure panels), and values were plotted against time.

5.6 Confocal image analysis (image segmentation)

HOXD13 puncta detected in fixed cell immunofluorescence shown in **Figure 2-8 A** were analyzed in Zen Blue 3.0 using a similar workflow described for optoDroplet assay analysis below with the following modifications. First, nuclear regions were automatically detected by Otsu thresholding on the DAPI counterstain signal. HOXD13 puncta within the nuclear regions were then detected using a fixed pixel intensity threshold. Puncta were segmented on morphology, and then filtered based on fluorescence intensity and circularity. After adjusting parameters on 4-5 images, the established pipeline was used to analyze all images from cell-types and treatment conditions (7 images for untreated wild-type limb bud, 7 images for 6% 1-6 hexanediol (1-6 HD) treated wild type limb bud, 18 images for untreated spdh limb bud, 12 images for 6% 1-6 HD treated spdh limb bud). Values for nuclear area, HOXD13 puncta area, HOXD13 puncta mean fluorescence intensity, and HOXD13 puncta count were measured for 120 wild-type cells, 143 1-6 HD treated wild-type cells, 63 spdh cells, and 62 1-6 HD treated spdh cells. The numbers of cells are displayed as "n" in **Figure 2-8 A and B**. For data visualization, phase-shift ratios, puncta count, and mean fluorescent intensities of HOXD13 puncta in each sample were plotted as indicated in figure legends using R.

5.7 STORM image analysis (DBSCAN and manders)

STORM images were localized, visualized, and analyzed using Vutara SRX software, version 6.04.14 (Bruker). For HOXD13 cluster analysis, images were collected with a 50 ms acquisition time and 5,000 images were used to reconstruct the super-resolution composite with minimal background thresholding. The HOXD13 density map was generated by point-splatting localized particles with radial precision. The histogram in **Figure 2-1 B** shows the frequency in size distribution of HOXD13 clusters within the projected cell nucleus. To estimate HOXD13 condensate size, HOXD13 localizations were clustered using density-based spatial clustering of applications with noise (DBSCAN). DBSCAN parameters (specific distance of 100 nanometers) were set using the spatial distribution tool embedded in the Vutara SRX analysis interface, and the radius of gyration for clustered HOXD13 localizations (15-50 particle clusters) was measured. The Nearest Neighbors (NeNa) localization precision[100] was 9 nm for HOXD13 localizations. NeNa reports the average

localization precision of a single molecule localization microscopy (SMLM) experiment. In brief, the routine is dependent on the distance distribution (NN_{adfr}) of nearest neighbors in adjacent frames, and is closely related to the pairwise displacement distribution of Gaussian-distributed localizations of a single, frequently localized molecule[101].

For Co-IF STORM analysis (**Chapter 2-7 A and B**), 4,000 images (2,000 for each probe channel) were collected with a 50 ms acquisition time for each cell (25 cells from wt samples, 22 cells from spdh samples, and 5 cells in negative control sample). HOXD13, BRD4 and HP1α particles were localized with minimal background thresholds on both channels, and then visualized in Vutara SRX software. To assess for overlap between probes, the Manders overlap coefficient (MOC) was calculated using the co-localization tool embedded in the Vutara SRX analysis interface. As the MOC calculates pixel-based overlap between 2 channel images acquired from diffraction-limited microscopy systems[101,102], STORM composites were first converted into a pixel-based image (particle size = 150 nanometers, max cluster size = 500 nanometers) to mimic a diffraction-limited system, and then this image was used to calculate the MOC. Each cell was analyzed for overlap between probes using the same conversion parameters. Only cells with similar localization densities (below 550 localizations / micron^2) were used for MOC comparison, indicated as "n" in **Figure 2-7 B** (n=13 for HOXD13/BRD4 co-localization in wt cells, n=22 for HOXD13/BRD4 co-localization in spdh cells, and n=5 for HOXD13/HP1α co-localization in wt cells). Density maps were generated by point splatting localizations with a size constant of 150 nanometers, to reflect the conversion used for overlap analysis. For visualization of probe overlap, merged density maps were exported and then post-processed in FIJI to isolate the pixel intersection between the two probe channels.

5.8 STED image analysis (rolling ball)

Composites acquired in τ-STED mode were exported as .tiff image fields using Leica LASX version 2.5.6 and analyzed using a custom FIJI pipeline available at https://github.com/BasuShaon/AR/tree/master/STED. In brief, the DNA counterstain or GFP signal was first used to identify and threshold nuclear objects. Clusters within nuclear objects were then detected using the rolling ball algorithm, with the size of the rolling ball set to 8 x the limit of detection (48 nanometers for immunofluorescence, 24 nanometers for live cell), according to standard protocol (Sternberg, 1983)[103].This enabled detection of nuclear AR

clusters for cells imaged with the same τ-deconvolution strength. In **Figure 2-12 B**, Nuclear AR clusters were pooled from 7 LNCaP nuclei, and analyzed for mean intensity and size. In **Figure 2-18 C**, cluster dimensions are projected from a pilot live-cell experiment with compounds on HeLa AR-eGFP cells.

5.9 ChIP-sequencing data pre-processing and visualization

Paired end sequencing reads were first quality checked using FASTQC and then aligned to *Homo sapien* genome build hg19 using bwa-mem with standard settings[104]. Alignment files were then sorted and indexed using samtools[105]. The processed alignments and indices were then used for peak calling with MACS2[106], and the resulting peak files were filtered for blacklisted regions with bedtools[107,108]. Peak and alignment files for ChIP experiments in mouse limb bud (**Figure 2-7 D**) and LNCaP cells (**Figure 2-15 D**), using H3K27Ac and AR antibodies, respectively, were visualized using UCSC genome browser. H3K27Ac ChIP sequencing reads from Rasool et al.[80] (GSE125245) were processed using the above pipeline, and verified for reported LNCaP super-enhancer regions using ROSE[58].

5.10 RNA-sequencing data pre-processing

Paired end sequencing reads were first quality checked using FASTQC and then aligned to *Homo sapien* genome build hg19 using STAR aligner v2.7.5a with standard settings[109]. 1st and 4th columns in ReadsPerGene.out.tab STAR output files (GeneIDs and reverse strand reads) were used to build raw count matrices for each sample library.

5.11 Differential expression analysis

Differential expression analysis between treatment conditions was conducted using the DESeq2 R package[110] a statistical tool that uses shrinkage estimates to compute fold changes. First, raw count matrices from sample libraries were merged into a single object using the 'DESeqDataSetFromHTSeqCount' function with the design set to the treatment condition (time, compound, and dosage). The merged count matrix was then fit to the DESeq statistical model using the 'DESeq' function. The fit and merged matrix was then reduced using a variance stabilizing transformation 'vst' to visualize principal components 1 & 2, as shown in **Figure 2-20 A**. The fold change values in gene expression and corresponding

significance scores were then extracted using the 'results' function by querying any contrast between two conditions. $|Log_2FC| > 1$ and p-value < 1e-10 cutoffs were used to call differentially expressed genes in a contrast, as shown in **Figure 2-20 B**.

5.12 Gene set enrichment analysis

Gene set enrichment analysis was performed using R packages fgsea and DOSE[111,112]. Ranked gene lists were first constructed using log_2FC values for genes in any given DESeq2 contrast by sorting in descending order and filtering out duplicate entries. Ranked lists were then analyzed for the enrichment of 50 hallmark gene sets (collection H) obtained from the molecular signature database msigDB maintained by the Broad Institute using the 'plotEnrichment' and 'plotfgseaRes' functions in fgsea and the 'GSEA' function in DOSE (nperm = 10,000, p-value cutoff < 0.05).

Besides the commonly used gene set enrichment plot for a queried gene pathway I also represent enrichment scores for the top 10 negatively and top 10 positively enriched pathways as a dotplot with gradient scaling to the normalized enrichment score (red = positive NES, blue = negative NES) and size proportional to the statistical significance (p_{adj}) of the calculated enrichment (displayed in **Figure 2-21 A**).

5.13 Mean expression value of genes in hallmark gene sets

Line plots for mean expression values of genes were adapted from Loven et al., 2013[113]. In brief, reads from the merged count matrix were normalized according to the following equation log_2(normalized DESeq counts + 1) to create a normalized count matrix. Normalized counts for each gene in the matrix were then z-score scaled using the 'scale_rows' function from the pheatmap R package. Code integrated with DESeq2 is available at https://github.com/BasuShaon/AR/tree/master/RNAseqLoven. Values of the genes from the below gene sets were then plotted as indicated in **Figure 2-21 B**.

Mgisdb pathways:

http://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H

EPI-001 Negative DEGS:

KLK3, ADAM7, TBX15, FKBP5, PGC, LAMA1, ELL2, CHRNA2, STEAP4, DSC1, UGT2B28, TNS3, BMPR1B, SLC38A4, EAF2, TTN, SLC15A2, CCDC141, HPGD, TMEM100 MAF, F5, TRGC1

6 Bibliography

1. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).

2. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).

3. Boija, A. *et al.* Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842-1855.e16 (2018).

4. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

5. Wei, M.-T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.* **22**, 1187–1196 (2020).

6. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).

7. Shrinivas, K. *et al.* Enhancer features that drive formation of transcriptional condensates. *Mol. Cell* **75**, 549-561.e7 (2019).

8. Basu, S. *et al.* Unblending of transcriptional condensates in human repeat expansion disease. *Cell* **181**, 1062-1079.e30 (2020).

9. Tsang, B., Pritišanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase separation as a missing mechanism for interpretation of disease mutations. *Cell* **183**, 1742–1756 (2020).

10. Kilgore, H. R. & Young, R. A. Learning the chemical grammar of biomolecular condensates. *Nat. Chem. Biol.* **18**, 1298–1306 (2022).

11. Boija, A., Klein, I. A. & Young, R. A. Biomolecular condensates and cancer. *Cancer Cell* **39**, 174–192 (2021).

12. Sabari, B. R. Biomolecular condensates and gene activation in development and disease. *Dev. Cell* **55**, 84–96 (2020).

13. Guillén-Boixet, J. *et al.* RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. *Cell* **181**, 346-361.e17 (2020).

14. Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).

15. Riback, J. A. *et al.* Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214 (2020).

16. Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).

17. Chen, X., Wu, X., Wu, H. & Zhang, M. Phase separation at the synapse. *Nat. Neurosci.* **23**, 301–310 (2020).

18. Mitrea, D. M., Mittasch, M., Gomes, B. F., Klein, I. A. & Murcko, M. A. Modulating biomolecular condensates: a novel approach to drug discovery. *Nat. Rev. Drug Discov.* **21**, 841–862 (2022).

19. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).

20. Pak, C. W. *et al.* Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell* **63**, 72–85 (2016).

21. Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688-699.e16 (2018).

22. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).

23. Case, L. B., Ditlev, J. A. & Rosen, M. K. Regulation of transmembrane signaling by phase separation. *Annu. Rev. Biophys.* **48**, 465–494 (2019).

24. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular condensates in the nucleus. *Trends Biochem. Sci.* **45**, 961–977 (2020).

25. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).

26. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

27. Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerg. Top. Life Sci.* **4**, 307–329 (2020).

28. Darling, A. L., Zaslavsky, B. Y. & Uversky, V. N. Intrinsic disorder-based emergence in cellular biology: Physiological and pathological liquid-liquid phase transitions in cells. *Polymers (Basel)* **11**, 990 (2019).

29. Liu, J. *et al.* Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888 (2006).

30. Darling, A. & Uversky, V. Intrinsic disorder in proteins with pathogenic repeat expansions. *Molecules* **22**, 2027 (2017).

31. Amiel, J., Trochet, D., Clément-Ziza, M., Munnich, A. & Lyonnet, S. Polyalanine expansions in human. *Hum. Mol. Genet.* **13 Spec No 2**, R235-43 (2004).

32. Muragaki, Y., Mundlos, S., Upton, J. & Olsen, B. R. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* **272**, 548–551 (1996).

33. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).

34. Bushweller, J. H. Targeting transcription factors in cancer - from undruggable to reality. *Nat. Rev. Cancer* **19**, 611–624 (2019).

35. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).

36. Biesaga, M., Frigolé-Vivas, M. & Salvatella, X. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Curr. Opin. Chem. Biol.* **62**, 90–100 (2021).

37. Imamura, Y. & Sadar, M. D. Androgen receptor targeted therapies in castration-resistant prostate cancer: Bench to clinic. *Int. J. Urol.* **23**, 654–665 (2016).

38. Estébanez-Perpiñá, E., Bevan, C. L. & McEwan, I. J. Eighty years of targeting androgen receptor activity in prostate cancer: The fight goes on. *Cancers (Basel)* **13**, 509 (2021).

39. Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **25**, 276–308 (2004).

40. Huang, P., Chandra, V. & Rastinejad, F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu. Rev. Physiol.* **72**, 247–272 (2010).

41. Antonarakis, E. S. *et al.* Androgen receptor splice variant, AR-V7, and resistance to enzalutamide and abiraterone in men with metastatic castration-resistant prostate cancer (mCRPC). *J. Clin. Oncol.* **32**, 5001–5001 (2014).

42. Dehm, S. M., Schmidt, L. J., Heemers, H. V., Vessella, R. L. & Tindall, D. J. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res.* **68**, 5469–5477 (2008).

43. Dehm, S. M. & Tindall, D. J. Alternatively spliced androgen receptor variants. *Endocr. Relat. Cancer* **18**, R183-96 (2011).

44. Andersen, R. J. *et al.* Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. *Cancer Cell* **17**, 535–546 (2010).

45. Basu, S. *et al.* Rational optimization of a transcription factor activation domain inhibitor. (2022) doi:10.1101/2022.08.18.504385.

46. Albrecht, A. & Mundlos, S. The other trinucleotide repeat: polyalanine expansion disorders. *Curr. Opin. Genet. Dev.* **15**, 285–293 (2005).

47. Villavicencio-Lorini, P. *et al.* Homeobox genes d11-d13 and a13 control mouse autopod cortical bone and joint formation. *J. Clin. Invest.* **120**, 1994–2004 (2010).

48. Kuss, P. *et al.* Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis. *J. Clin. Invest.* **119**, 146–156 (2009).

49. Shin, Y. *et al.* Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell* **168**, 159-171.e14 (2017).

50. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).

51. Albrecht, A. N. *et al.* A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. *Hum. Mol. Genet.* **13**, 2351–2359 (2004).

52. Kelleher, R. J., 3rd, Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61**, 1209–1215 (1990).

53. Patel, A. *et al.* ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).

54. Wheeler, R. J. *et al.* Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv* (2019) doi:10.1101/721001.

55. Bruneau, S., Johnson, K. R., Yamamoto, M., Kuroiwa, A. & Duboule, D. The mouse Hoxd13(spdh) mutation, a polyalanine expansion similar to human type II synpolydactyly (SPD), disrupts the function but not the expression of other Hoxd genes. *Dev. Biol.* **237**, 345–353 (2001).

56. Jiang, Y. W. *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8538–8543 (1998).

57. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

58. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

59. Winter, G. E. *et al.* BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. *Mol. Cell* **67**, 5-18.e19 (2017).

60. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

61. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, (2018).

62. Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25 (2014).

63. Goodman, F. R. *et al.* Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am. J. Hum. Genet.* **67**, 197–202 (2000).

64. Shibata, A. *et al.* Characterisation of novel RUNX2 mutation with alanine tract expansion from Japanese cleidocranial dysplasia patient. *Mutagenesis* **31**, 61–67 (2016).

65. Matsushita, M. *et al.* A novel in-frame deletion of the RUNX2 gene causes a classic form of cleidocranial dysplasia. *J. Bone Miner. Metab.* **32**, 96–99 (2014).

66. Black, B. & Paschal, B. Intranuclear organization and function of the androgen receptor. *Trends Endocrinol. Metab.* **15**, 411–417 (2004).

67. Tomura, A. *et al.* The subnuclear three-dimensional image analysis of androgen receptor fused to green fluorescence protein. *J. Biol. Chem.* **276**, 28395–28401 (2001).

68. Wu, Y., Kawate, H., Ohnaka, K., Nawata, H. & Takayanagi, R. Nuclear compartmentalization of N-CoR and its interactions with steroid receptors. *Mol. Cell. Biol.* **26**, 6633–6655 (2006).

69. Steinkamp, M. P. *et al.* Treatment-dependent androgen receptor mutations in prostate cancer exploit multiple mechanisms to evade therapy. *Cancer Res.* **69**, 4434–4442 (2009).

70. Takeda, D. Y. *et al.* A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate cancer. *Cell* **174**, 422-432.e13 (2018).

71. Chan, S. C., Li, Y. & Dehm, S. M. Androgen receptor splice variants activate androgen receptor target genes and support aberrant prostate cancer cell growth independent of canonical androgen receptor nuclear localization signal. *J. Biol. Chem.* **287**, 19736–19749 (2012).

72. Hu, R. *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69**, 16–22 (2009).

73. De Mol, E. *et al.* EPI-001, A compound active against castration-resistant prostate cancer, targets transactivation unit 5 of the androgen receptor. *ACS Chem. Biol.* **11**, 2499–2505 (2016).

74. Myung, J.-K. *et al.* An androgen receptor N-terminal domain antagonist for treating prostate cancer. *J. Clin. Invest.* **123**, 2948–2960 (2013).

75. Claessens, F. *et al.* Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nucl. Recept. Signal.* **6**, e008 (2008).

76. Feric, M. *et al.* Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).

77. Zhu, J., Salvatella, X. & Robustelli, P. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. *bioRxiv* (2021) doi:10.1101/2021.12.23.474012.

78. Obst, J. K., Mawji, N. R., Teskey, S. J. L., Wang, J. & Sadar, M. D. Differential gene expression profiles between N-terminal domain and ligand-binding domain inhibitors of androgen receptor reveal ralaniten induction of metallothionein by a mechanism dependent on MTF1. *Cancers (Basel)* **14**, 386 (2022).

79. Banuelos, C. A. *et al.* Ralaniten sensitizes enzalutamide-resistant prostate cancer to ionizing radiation in prostate cancer cells that express androgen receptor splice variants. *Cancers (Basel)* **12**, 1991 (2020).

80. Rasool, R. U. *et al.* CDK7 inhibition suppresses castration-resistant prostate cancer through MED1 inactivation. *Cancer Discov.* **9**, 1538–1555 (2019).

81. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

82. Owens, K. M. *et al.* Analysis of de novo HOXA13 polyalanine expansions supports replication slippage without repair in their generation. *Am. J. Med. Genet. A* **161A**, 1019–1027 (2013).

83. Orr, H. T. & Zoghbi, H. Y. Polyglutamine tract vs. protein context in SCA1 pathogenesis. in *Analysis of Triplet Repeat Disorders* 105–117 (Garland Science, 2020).

84. Zoghbi, H. Y. & Orr, H. T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* **23**, 217–247 (2000).

85. Albrecht, C., Borm, P. J. A. & Unfried, K. Signal transduction pathways relevant for neoplastic effects of fibrous and non-fibrous particles. *Mutat. Res.* **553**, 23–35 (2004).

86. Engelke, A. D. *et al.* Dimerization of the cellular prion protein inhibits propagation of scrapie prions. *J. Biol. Chem.* **293**, 8020–8031 (2018).

87. Selkoe, D. J. Toward a remembrance of things past: deciphering Alzheimer disease. *Harvey Lect.* **99**, 23–45 (2003).

88. Weissmann, C., Enari, M., Klöhn, P.-C., Rossi, D. & Flechsig, E. Transmission of prions. *Proc. Natl. Acad. Sci. U. S. A.* **99 Suppl 4**, 16378–16383 (2002).

89. Morris, G. P., Clark, I. A. & Vissel, B. Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta Neuropathol. Commun.* **2**, 135 (2014).

90. Kametani, F. & Hasegawa, M. Reconsideration of amyloid hypothesis and tau hypothesis in Alzheimer's disease. *Front. Neurosci.* **12**, (2018).

91. Mullard, A. Anti-amyloid failures stack up as Alzheimer antibody flops. *Nat. Rev. Drug Discov.* (2019) doi:10.1038/d41573-019-00064-1.

92. Alberti, S. & Dormann, D. Liquid-liquid phase separation in disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).

93. Nedelsky, N. B. & Taylor, J. P. Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 272–286 (2019).

94. Klein, I. A. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science* **368**, 1386–1392 (2020).

95. Massie, C. E. *et al.* The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *EMBO J.* **30**, 2719–2733 (2011).

96. Tran, C. *et al.* Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787–790 (2009).

97. Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol. Biol.* **1091**, 17–32 (2014).

98. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-response analysis using R. *PLoS One* **10**, e0146021 (2015).

99. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

100. Endesfelder, U., Malkusch, S., Fricke, F. & Heilemann, M. A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. *Histochem. Cell Biol.* **141**, 629–638 (2014).

101. Manders, E. M. M., Verbeek, F. J. & Aten, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).

102. Lagache, T. *et al.* Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nat. Commun.* **9**, 698 (2018).

103. Sternberg. Biomedical Image Processing. *Computer (Long Beach Calif.)* **16**, 22–34 (1983).

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

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

106. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

107. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

108. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: Identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).

109. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

110. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *bioRxiv* (2014) doi:10.1101/002832.

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

112. Yu, G., Wang, L.-G., Yan, G.-R. & He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).

113. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).

7 List of Publications

2020 Shaon Basu*, Sebastian D Mackowiak*, Henri Niskanen, Dora Knezevic, Vahid Asimi, Stefanie Grosswendt, Hylkje Geertsema, Salaheddine Ali, Ivana Jerković, Helge Ewers, Stefan Mundlos, Alexander Meissner, Daniel M Ibrahim, Denes Hnisz. "Unblending of transcriptional condensates in human repeat expansion disease." *Cell* 181, 1062-1079.e30 (2020).

2022 Shaon Basu*, Paula Martínez-Cristóbal*, Mireia Pesarrodona, Marta Frigolé-Vivas, Elzbieta Szulc, Michael Lewis, C. Adriana Bañuelos, Carolina Sánchez-Zarzalejo, Stasė Bielskutė, Jiaqi Zhu, Carla Garcia-Cabau, Cristina Batlle, Borja Mateos, Mateusz Biesaga, Albert Escobedo, Lídia Bardia, Xavier Verdaguer, Alessandro Ruffoni, Nasrin R. Mawji, Juan Wang, Teresa Tam, Isabelle Brun-Heath, Salvador Ventura, David Meierhofer, Jesús García, Paul Robustelli, Travis H. Stracker, Marianne D. Sadar, Antoni Riera, View Denes Hnisz, Xavier Salvatella. "Androgen receptor condensates as drug targets." *bioRxiv* (2022) doi:10.1101/2022.08.18.504385.

* denotes shared first author

8 Appendix – Data and Resource Availability

The Salvatella lab (Institute for Research in Biomedicine, Barcelona) synthesized compounds 1aa – 1bb and provided the HeLa AR-eGFP cell line used for experiments in **Figure 2-14 C**.

Raw data, code, and plasmids have been deposited to the following directories:

Custom code

URL: https://github.com/BasuShaon

Microscopy images:

URL: https://data.mendeley.com/datasets/ztd6wzcv7h/13

Plasmids:

URL: https://www.addgene.org/browse/article/28211176/

RNA sequencing deposition:

URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206853