Abstract

Assembly mechanisms of nuclear bodies

Edward Michael Courchaine

2020

Eukaryotic cells separate their contents into organelles to perform specific functions in spatially contained compartments. They achieve this by synthesizing lipid bilayers to encapsulate organelles such as mitochondria and endoplasmic reticulum. Other "membraneless organelles" have no lipid bilayer or similar barrier defining their perimeter. Biomolecular condensation has emerged as a mechanism of assembly that could explain much about the formation and function of such organelles. Those found in the nucleus present a particularly interesting case of biological organization: protein, RNA, and DNA must coexist and dynamically carry out their functions. Indeed, these nuclear 'bodies' are often fundamentally tied to transcription and regulation of gene expression. This dissertation aims to identify mechanisms of nuclear body formation, using the Cajal body as its specific focus.

The Cajal body is a conserved structure in metazoans and is found in diverse cell types including embryos and neurons. The Cajal body has been implicated in several biochemical pathways, but snRNP biogenesis has emerged as a fundamentally important function. Developing vertebrate embryos rely on the Cajal body to produce enough snRNPs, and without them they arrest and die. The approach presented here targets known protein components of the Cajal body. I have assessed those proteins for their capacity to

form biomolecular condensates and analyzed the specific contacts between proteins required for Cajal body assembly.

In this study three proteins have been tested for condensation: coilin, the survival of motor neuron protein (SMN), and Nopp140. Most strikingly, the SMN tudor domain makes condensates when bound to dimethylated arginine on protein ligands. This recognition of a post-translational modification by a folded domain represents a new mode for forming the high avidity network required for condensation. This property is shared amongst other tudor domains, suggesting generality. It contrasts the established expectation that primarily intrinsically disordered domains mediate condensate formation.

I have also addressed the specific action of the coilin N-terminal domain in the formation of the Cajal body. The coilin N-terminus is conserved, it is required for proper Cajal body formation, and evidence supports a self-interaction of this domain. Point mutations in this region show striking effects on its mesoscale assembly properties. The same mutants disrupt self-interaction and binding to Nopp140. Depletion of Nopp140 revealed that it, too, is an essential Cajal body protein, and its presence moderates the oligomerization properties of the coilin N-terminus.

Together these results demonstrate that the Cajal body is formed and maintained by multiple mechanisms that include both dynamic condensation and specific protein-protein oligomerization. In general, it is likely that most physiological nuclear bodies also form via the intersection of several mechanisms in the same compartment. My findings suggest that two assembly pathways may antagonize one another, which could provide opportunities for regulation and maintain dynamicity in these structures.

Assembly mechanisms of nuclear bodies

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Edward Michael Courchaine

Dissertation Director: Karla M. Neugebauer

May 2020

Table of Contents

Index of Figures

Index of Tables

Acknowledgements

There is no other way to begin this section than by saying thank you to Dr. Karla Neugebauer, who welcomed me to the lab for a summer rotation in 2014 without even having met me. At the time, I thought I would spend graduate school working on synthetic proteins and bioengineering, and figured a four-week rotation on a molecular biology project might be a fun sidebar. I had never performed a PCR or run a western blot outside of my undergraduate biochemistry lab. I cannot imagine completing my PhD without Karla's generous guidance and support. Being in your lab has truly been a privilege.

To the other members of the Neugebauer Lab, you have all become family to me in the past five years. You have seen me progress as a scientist, experienced what is certainly the best and the worst of me along the way, and I am eternally grateful for your presence in my life. I look forward to working with all of you soon. Tara Alpert, in particular, has been there with me since our first year. Even though our projects never overlapped once in five years, I owe a great deal of my success to your friendship.

The support of the Yale Molecular Biophysics and Biochemistry department cannot be understated. The friendship of my classmates here is one of the most valuable things a person can have, and the mentorship of the faculty has transformed the way I think about science and think about being a scientist. In particular I thank my committee members Drs. Joan Steitz and Julien Berro for their advice and guidance over the years. I also thank Dr. Andrew Miranker for mentoring me as I learned how to be a teacher myself and for being a sounding board when my project was on the rocks.

My collaborators and scientific mentors have been an incredible group of people to work with; you have set a high bar for the teams I hope to work with in the future. Andrew

Barentine, Martin Machyna, and Korinna Straube all deserve an extra measure of appreciation because you have all taught me a bit off science that I didn't know before, and working with you has shown me what it means to have good colleagues.

My family has been nothing but supportive through this process, and I have been fortunate to have them close by in graduate school. I am grateful that you have all remained interested (or at least done an outstanding job of faking it) at all the holiday gatherings where I talked about my work. That may not let up any time soon, but know that you have made it possible for me to get where I am.

To my partner Meaghan, I cannot imagine enjoying science as much as I do without talking about it with you. We met because we both wanted a community to be researchers and human at the same time. I never knew how easy it would be to find both when we work on that together. Cheers to many more projects, late nights reading papers, and adventures both in lab and in life.

A brief note on this moment in history

While dissertations on biochemistry and biophysics are usually agnostic toward current events, I couldn't help but take a few notes on what has been happening in the world while writing.

The President of the United States, Donald Trump, was impeached by the house at the close of 2019. His trial took place while most of my final writing was completed and two days before submitting this dissertation to my committee he was acquitted of both charges, abuse of power and obstruction of congress. Senator Mitt Romney was the only dissenting party vote in favor of removing him from office.

The 2020 presidential election officially began with a botched Iowa caucus on February 3rd. Election day is not until November 3rd.

A novel coronavirus has infected more than 30,000 people in China, with new infections growing by the thousands per day. 2019 was the second warmest year on record.

These points are not intended to be pessimistic. Rather, I want to provide a backdrop that underscores the importance of pursuing knowledge, discovery, and truth in all ways available to us.

Chapter 1. Introduction

1.1. Central question and hypothesis

This work explores the molecular mechanisms of membraneless organelles in the nucleus. I approach this problem with the perspective that biomolecular condensation is likely to be involved, but is not necessarily the exclusive framework to be applied. Thus, I will introduce nuclear bodies broadly alongside their cytoplasmic counterparts with what we have learned about condensates in the past ten years. I will consider the most recent advances separately and discuss how new tools and conceptual advances have enhanced our ability to study condensate systems. Finally, I will introduce specific information about the Cajal body, which will be the primary subject of this work. The background given below justifies this hypothesis: At least one Cajal body protein facilitates assembly by promoting condensation, and at least one protein must define the specificity of that assembly. Subsequent chapters will present experimental evidence in critique of this hypothesis and discuss those findings.

1.2. Membraneless organelles and biomolecular condensation

Outside contributions

Sections 1.2 - 1.4 are adapted from Courchaine et al. (2016) "Droplet Organelles?"[1]. I curated sources, wrote the first draft and coordinated subsequent revisions. Alice Lu contributed to commentary on disease relevance. Karla Neugebauer edited and revised the manuscript.

A proposed mechanism for membraneless organelle formation

Cells spatially organize biochemical reactions, a characteristic that is fundamental to life but often evades analysis. Lipid membranes create discrete chemical environments within canonical organelles and achieve separation of constituents from the bulk cytoplasm. Enclosing membrane-bound compartments requires dedicated machinery to construct, maintain and transport across the lipid bilayer, thus expending energy[2,3]. Many organelles lack lipid bilayers, circumventing these issues and introducing the potential for greater dynamics and alternative mechanisms of regulation. Nuclear structures, such as nucleoli and Cajal bodies, regulate ribonucleoprotein (RNP) assembly in this manner[4,5]. In the cytoplasm, stress granules and P-bodies regulate RNA stability and protein translation in response to cellular stimuli[6]. Accumulating evidence suggests that such non-membrane bound organelles behave as fluid droplets, which undergo a phase transition to a condensed state. The molecular mechanisms that underlie their form and function have been under intense investigation in the past decade, and this introduction focuses on the emergent common principles from that work.

I begin by surveying the literature on systems that exhibit biomolecular condensation and related behavior in order to bridge the *in vivo* studies of cellular structures and the molecular understanding afforded by *in vitro* studies of proteins, RNAs, and their interactions. I will first focus on the work reviewed by me and my colleagues in 2016 including key studies to provide a prelude to those developments. These sections close by returning to the biological consequences of competing models for condensate systems and discusses how they may relate to diseases characterized by the formation of aberrant protein aggregates. My goal is to synthesize an outlook for the relevance of condensation

at both molecular and cellular scales to provide context for the experiments that follow. I emphasize several key issues that had been unresolved at the time of review and attempt to address controversy where it is present. Section 1.5 discusses resolutions made to these problems by subsequent reports.

Nuclear and cytoplasmic bodies are the sites of RNP biogenesis

Many non-membrane bound organelles can be found in the nucleus, and were first observed there over a century ago[7,8]. Nuclear bodies have been investigated extensively regarding the exchange of constituents with surrounding nucleoplasm. For most components, residence times are on the order of seconds with the most stable constituents exchanging within tens of seconds[9,10]. The dynamic nature of nuclear bodies likely underlies their function, which studies of biomolecular condensation strive to explain. The most prominent nuclear body is the nucleolus, which forms on active rDNA loci and is the site of pre-rRNA processing and pre-ribosomal subunit assembly[11-13]. Similarly, the Cajal body forms on active snRNA loci and is the site of snRNA processing, snRNP assembly, and snRNP surveillance[4,14-16]. Nuclear speckles and paraspeckles are more granular in morphology, contain mRNAs and their binding proteins, and form on two long non-coding RNAs (lncRNAs) MALAT1 and NEAT1, respectively[5]. Many of the specific functions of these subdomains are still poorly characterized, but at the descriptive level they are consistent with phase separated systems.

Cytoplasmic bodies are more granular in morphology and have functions often related to translational control and/or mRNA stability. The processing body (P-body) falls into this second category, in which translation is stalled, and transcripts are targeted for degradation by exonucleases or selective reactivation of translation[17,18]. Stress granules are

3

related to P-bodies, in that they contain translationally repressed mRNA, but form in response to heat, osmotic and chemical stress stimuli. Stress granules contain factors that stall translation, as well as a number of RNA binding proteins associated with ALS and other diseases. The normal physiological roles of many stress granule proteins are still unclear but single molecule techniques have revealed that mRNAs are already translationally repressed before entering[19-21]. Evidence suggests that other bodies also undergo a response to environmental challenges, with unknown consequences[22].

Some cytoplasmic RNA granules are developmentally important. In *C. elegans*, the germ cell lineage is specified by the asymmetric inheritance of P granules during mitosis. The exact function of these RNP granules is unknown, but they have some similarity to P-bodies and are functionally implicated in translational control of the germ line[23]. In other metazoans, germ line specification follows a different path, but germ cells still display characteristic RNP structures known as germ plasm, nuage or the chromatoid body[24]. Nuage contains many RNAs as well as several helicases, endonucleases and proteins involved in miRNA-mediated degradation of RNA[25].

RNP bodies and granules display characteristics of phase separated liquids

Even as progress has been made on elucidating the biological functions of these nuclear and cytoplasmic bodies, their physical properties have only recently come to light through high resolution live cell imaging. Like nucleoli, Cajal bodies fuse and split[26]. In 2005, Gall and colleagues speculated that Cajal bodies were actually "semi-fluid spheres suspended in semi-fluid nucleoplasm"[27]. This conclusion was based on observed shape, permeability, and differential protein concentrations between the nucleoplasm and the large Cajal bodies present in *Xenopus* germinal vesicles. In 2009, seminal work on the P granules

of *C. elegans* demonstrated their liquid-like properties and that they localize to the future germ cell cytoplasm by dissolving and condensing rather than by moving as discrete objects through the cytoplasm[28]. Extrachromosomal nucleoli of *Xenopus laevis* were also shown to exhibit properties expected of liquid droplets[29]. Advanced microscopy techniques, including tracking of subcellular structures over time, turned the conjecture that nuclear and other organelles undergo condensation into a well-characterized phenomenon.

These apparent manifestations of condensation *in vivo* were striking and harken back to studies of other biomolecules that undergo phase separation *in vitro*. In 2006, Görlich and colleagues showed that the yeast FG (phenylalanine-glycine) repeat proteins of the nuclear pore complex are capable of condensing into a hydrated gel matrix[30]. These gels showed no fluorescence recovery after photobleaching (FRAP), suggesting that the hydrogel is not fluid and instead may trap constituent molecules in an immobile meshwork. Cross-β amyloid-like interactions are likely to underlie hydrogel formation, and are disrupted by nuclear transport factors. This "melting" of the gel matrix by some proteins but not others effectively constitute a selectivity filter[31,32]. Species other than fungi are less likely to have FG nucleoporins that show strong amyloid character, but all rely on hydrophobic interactions in their formation[32,33]. This *in vitro* molecular behavior is distinct from that observed for constituents of Cajal bodies and nucleoli *in vivo*. Specifically, the GFP fluorescence recovery of coilin and SMN, two core components of Cajal bodies, indicates that they exchange rapidly with bulk nucleoplasm[9]. Such rapid recovery would not be expected of hydrogels. Comparing the findings of *in vitro* studies of the nuclear pore components and of nuclear bodies *in vivo,* it would seem that there are two different ways

that biological systems can separate from the bulk solvent: either as crosslinked hydrogels or as liquid droplets.

1.3. Properties of condensing molecules

In vitro systems show fundamental properties of phase separation

New assays have been devised to probe the underlying physical and molecular nature of phase separating systems. Using purified protein domains derived from signaling proteins of the N-WASP pathway, Rosen and colleagues demonstrated that multivalency is sufficient to drive the phase separation of concatenated SH3 domains that bind to concatenated proline rich motifs[34]. In the same study, RNA with multiple binding motifs was also able to form droplets in combination with the multivalent protein PTB, which exhibit FRAP and form above a critical concentration threshold. Other work has also shown that the properties of phase-separated proteins depend heavily upon phosphorylation state[35-38], as discussed below. One investigation into these effects on FUS, an RNA and DNA binding protein found in both cytoplasm and nucleoplasm, highlighted the importance of post-translational modification as a means of regulating condensation and implicated low-complexity protein domains (LCD) as fundamental to body formation[39].

Proteins with low complexity domains are often found in bodies

The organelles so far discussed almost all contain at least one, if not several, disordered proteins that notably contain characteristic LCDs. LCDs are defined by the overrepresentation of a subset of amino acids in their primary sequence. Many of these proteins are listed in Table 1. It is worth clarifying some of the terminology that has been used to describe LCDs: Several studies have designated LCD sequences based on their sequence similarity to the yeast prion proteins[40,41]. These so-called "prion-like domains"

seem to be related to disease and have a propensity to aggregate[42]; however, this term is too narrow for the present discussion of phase separating systems. "Low complexity domain" is currently the most satisfying term available because it includes the prion-like domains as well as disordered domains with other amino acid compositions that might be important for forming phase separated systems like stress granules and P bodies[43,44].

Two proteins containing LCDs have come to the forefront of much recent research due to their relevance to ALS and frontotemporal dementia: FUS and hnRNP-A1. FUS is associated with several nuclear and cytoplasmic bodies, including paraspeckles and stress granules. Under normal conditions, FUS is found in the nucleus where it takes part in DNA repair and transcriptional regulation[45,46]. For reasons that are not entirely clear, stress stimuli result in the export of FUS to the cytoplasm where it joins stress granules[47]. These studies have identified the importance of both RNA binding properties and the presence of the LCD within FUS. Similar to FUS, nuclear hnRNP-A1 has two RNA recognition motifs (RRMs) and an LCD, is recruited to stress granules, and phase separates *in vitro*[36,48]. Significant inroads have been made into understanding how FUS, hnRNP-A1 and other LCD proteins can generate fluid cellular bodies (Table 1, page 8)[49-51].

Table 1. Protein components of cellular compartments that phase separate in vitro

Protein (aa)	Low complexity motif(s)	Structured domains	Cellular compartment	*In vitro* morphology	Ref.
DDX4 (724)	19aa R/G-rich 16aa R/G/S 9aa P-rich	DEXDc HELICc	Nuage RNP granules	Droplets*	[52]
eIF4GII (914) ***yeast**	38aa N/K/Y 21aa T/P 13aa S/R 17aa E/A 29aa N/S	MIF4G	Stress granules	Droplets*	[53,54]
EWS (656)	19aa Y/A/S 82aa A/Q/T 85aa Q/S/Y 33aa R/G 22aa D-rich 60aa R/G/P 86aa R/G/P	RRM ZnF_RBZ	DNA damage sites	Hydrogel*	[35,55]
Fibrillarin (321)	72aa R/G	Methyl-transferase (fibrillarin)	Nucleolus Cajal body	Droplets	[56]
FUS (526)	156aa S/G/Q **55aa RGG** **76aa RGG**	RRM ZnF	Paraspeckles Stress granules	Droplets Hydrogel	[55,57-59]
hnRNP A1/A2 (372/353)	180aa G/S/R/Q	2 RRMs	Stress granules	Droplets Hydrogel	[48,53,54,60]
LAF-1 (708) ***C. elegans**	130aa G/R 75aa G/R/Q	DEXDc HELICc	P granules	Droplets	[61]
Lsm4 (187) ***yeast**	173aa N/R	Sm	P bodies stress granules	Droplets*	[53]
PTB (557)	13aa S/N/A 18aa A-rich 30aa A-rich	4 RRMs	Nuclear speckles Peri-nucleolar	Droplets	[34,53]
Pub1 (453) ***yeast**	55aa N/M 33aa Q-rich	3 RRMs	Stress granules	Droplets*	[53]
RBM14 (669)	300aa A/R/S/Q/P	2 RRMs	Paraspeckles	Hydrogel	[62]
SRSF2 (221)	113aa R/S	RRM	Nuclear speckles	Hydrogel	[36]
TAF15 (589)	148 S/G/Q/Y 39 RGG 26aa R/G 186aa RGG/YGG	RRM ZnF_RBZ	DNA damage sites	Hydrogel*	[35,55]

Protein (aa)	Low complexity motif(s)	Structured domains	Cellular compartment	*In vitro* morphology	Ref.
TDP-43 (414)	43aa G/F/N 9aa A-rich 17aa Q/N 37aa S/G	2 RRMs	Stress granules	Droplets	[54,57]
Tia1 (386)	23aa S/T/Q	3 RRMs	Stress granules	Droplets*	[53]
Whi3 (729) ***A. gossypii***	Poly-Q	RRM	Cytoplasmic RNP granules	Droplets*	[63]

Protein and its amino acid length (aa) are listed. Unless otherwise noted with an asterisk, all proteins are human. "Yeast" is shorthand for *Saccharomyces cerevisiae*. Low complexity motifs are listed in the order of primary sequence and amino acid enrichment is indicated. Structured domains are according to the SMART database. "Cellular compartment" refers to the structure in which the protein concentrates *in vivo*, in the absence of mutation. Morphologies noted with an asterisk have only been determined using a fragment of the protein, rather than full length. Proteins are listed alphabetically, and selected references are shown.

Electrostatic interactions and biomolecular condensation

What causes an LCD to transition into either a liquid droplet or a solid hydrogel is a matter of debate, and the recent literature considers several contributing factors[53,57,60]. Isolated LCDs from a variety of proteins that scaffold cellular structures are sufficient for phase separating behavior. Strikingly, the low complexity sequences of eIF4GII, hnRNP-A1 and FUS all form liquid droplets without the addition of any other components[53]. The integrity of the full-length LCD is essential for the formation of paraspeckles, and a tyrosine to serine mutation in half of the repeat motifs of RBM14 clearly alters the morphology of the resulting hydrogel[62]. Hydrogel crosslinking may occur during droplet maturation, a process described for numerous intrinsically disordered regions (IDRs) *in vitro*. Typically, crosslinking is induced over time by manipulating protein and/or salt concentration, temperature, and molecular crowding[52,53]. This is in contrast to fluid droplets, which have a notably different structure from solid droplets seen in electron micrographs. Thus, electrostatic interactions may provide an underlying force for liquid-like condensation, though the evidence for this is taken from experiments that were performed above physiological protein concentrations or below physiological salt concentrations.

By demonstrating that the two-phase state of these systems requires a low ionic strength, these studies allow us to speculate on the importance of charge and polarity in the composition of LCDs as well as the relevance of post-translational modifications that can render amino acids either more or less charged. For example, hydrogel recruitment of the SRSF2 LCD, which is serine and arginine-rich, is blocked by phosphorylation[36]. Intriguingly, phosphorylation can also alter the range of structural ensembles and binding

interactions for some IDRs[64], suggesting that some IDRs may become less disordered when post-translationally modified. Experiments employing phosphomimetic or alanine replacements have indeed confirmed that the conformation of the peptide backbone of the FUS IDR is determined by modification state and controls gelation[65]. While existing data indicate that electrostatic interactions between LCDs are important, we have yet to address the fact that condensation often coincides with the presence of another charged macromolecule: RNA.

Roles for RNA *in vitro* and *in vivo*

We have come to understand that nucleic acids play a fundamental role in condensation *in vitro* and *in vivo*. In considering how electrostatic interactions might mediate the formation of liquid droplets, recall that RNA is concentrated in many bodies and granules *in vivo*. Because of its anionic phosphate backbone, RNA is a highly charged molecule that can potentially contribute to electrostatic interactions with positively charged residues in LCDs. Indeed, RNA enhances the fluid properties of droplets formed by the P granule component LAF-1, which contains positively charged arginine-glycine-glycine (RGG) repeats[61]. This increase in fluidity might be related to a general promotion of LLPS by RNA, as is also observed in the formation of fibrillarin droplets[56]. Poly(ADP-ribose), or PAR, is a polynucleotide synthesized upon DNA damage; PAR acts as a signal in the localization of the DNA repair machinery to sites of DNA damage (Table 1, page 8). The chemical structure of PAR strongly resembles RNA and appears to nucleate liquid phase separated regions through electrostatic interactions with RGG repeats found in FUS, EWS and TAF15[55,59]. Like LAF-1, these proteins also contain RNA binding domains and two regions of RGG repeats, but only the N-terminal LCD is sufficient for condensation.

Synergistic effects between these regions may contribute to *in vivo* condensation by binding multiple proteins and RNAs, and associating with other LCDs[57,66,67]. That said, the potential for non-canonical binding of RNA by low complexity domains should not be discounted. Indeed, the Cajal body scaffolding protein coilin was shown to bind RNA in the absence of an annotated RNA binding domain[68]. Recent work has emphasized that RNA-RNA interactions themselves can form condensates and postulates that cells must manage the uncontrolled phase separation of RNAs through base pairing[69,70]. Stress granules very well may represent condensates of mRNA that result from stalled translation and must be cleared[71]. Thus, RNA may play a role as a generic poly-anion or may regulate condensation through specific interactions with proteins or itself.

A second way that RNA can participate in condensation is through protein-RNA interactions mediated by canonical RNA binding domains, such as RNA recognition motifs (RRMs), zinc fingers (ZnFs), and KH-domains. Indeed, condensate scaffolding proteins frequently contain one or more RNA binding domain (Table 1, page 8). RNA sequence can provide multivalency for RNA-protein interactions by presenting repeated motifs for protein binding. For example, the highly structured RNA binding protein, PTB, forms droplets when added to concatemerized RNA target sequences[34]. A study of Whi3, a poly-Q protein that forms cytoplasmic mRNA granules in the fungus *Ashbya gossypii,* showed that droplet formation depends on a functional RRM and binding to the CLN3 mRNA[63]. Full-length Whi3 did not form fibrous structures *in vivo* or *in vitro*, while the LCD alone formed filaments that became less soluble over time. Further, Whi3 induces changes to RNA structure to promote co-condensation[72]. Similar results were obtained in studies of FUS, which was shown to transition from a droplet to filaments over time *in vitro* and to

12

form aggregates in the context of hydrogels[59,62]. The potential role of the FUS RRM in modulating condensation, aggregate and/or filament formation has not yet been directly analyzed.

Because RNA-rich nuclear bodies – Cajal bodies, nucleoli, histone locus bodies and paraspeckles – can form at transcription sites, nascent RNA likely plays a role in their nucleation[5,13,68,73,74]. The combination of an RRM with an LCD should be ideal for scaffolding transcriptionally dependent condensate structures. Indeed, many proteins within the paraspeckle contain N-terminal RRMs and C-terminal LCDs (Table 1, page 8), and both FUS and RBM14 form droplets and aggregates[62]. Recent investigations of hnRNP-A1 indicate that RNA binding by N-terminal RRMs modulates droplet formation promoted by the protein's C-terminal LCD[53,54]. Interestingly, fusion of PTB's RRMs to the IDRs of a variety of RNA binding proteins rendered these IDRs capable of droplet formation in the presence of RNA, and hnRNP-A1's RRM was essential for droplet formation. Thus, nucleic acids may serve in the formation of condensed cellular structures by providing a multivalent platform to seed phase separation and by buffering the protein-protein contacts to promote fluidization.

Molecular interactions create distinct protein states

Why would cells risk the use of systems that promote toxic aggregation? A major focus has been to determine which LCDs tend to form liquid droplets, hydrogels or solid amyloid aggregates. Many neurological diseases include protein aggregation in their pathology. The solid stress granules of yeast are controlled by the disaggregase machinery and must be disassembled before the cell may resume growth[75,76]. In addition, yeast stress granules can be disassembled by autophagy in response to cellular conditions[77]. In higher

13

organisms there are likely to be more subtle control processes that determine when LCDs remain liquid and when they aggregate. At least one report has identified a complex system of recognition for some granules that directs them towards the autophagy pathway[78]. The idea that proteins have multiple states of matter has been raised previously, and it is appealing to consider that each state must be resolved by its own pathway[79].

The dynamic and fluid nature of many condensate systems depends on the attractive potential between the constituent protein and RNA molecules. By buffering ionic strength within a liquid droplet, RNA may be able to maintain the two-phase state near a theoretical "cloud point." Two studies have demonstrated that droplets formed by condensation have properties that correspond to those predicted by Flory-Huggins Theory[52,54]. This takes into account the entropic cost of confining large polymers (proteins and RNA) into a localized region, which is counterbalanced by the enthalpy of electrostatic contacts between components that attract[80,81]. This framework from polymer chemistry has broad applicability in biochemistry. From this perspective, the observed amyloid formation is likely to be a secondary consequence of increased protein concentration rather than condensation occurring as a consequence of limited amyloid-like interactions. Therefore, cells can use this property to rapidly respond to changing conditions, with disassembly and assembly occurring on time scales on the order of seconds. Thus, these results substantiate the relevance of biomolecular condensation as a framework for understanding droplet formation, but leave open the question of how droplets, hydrogels and amyloids may be related to one another on a molecular scale.

The relationship between protein conformation and phase separation

Despite the rapid and substantial progress on this topic, conflicting perspectives on the molecular mechanisms of biomolecular condensation have emerged. These will have consequences for understanding how diseases like ALS arise and what therapies might be developed. Three proteins serve as the best examples of this controversy: FUS, hnRNP-A1, and hnRNP-A2. All contain LCDs and bind nucleic acids (Table 1, page 8). Furthermore, these proteins have been shown to form droplet, hydrogel and amyloid aggregate structures. In several studies, the formation of these structures has been posited to result from slow formation of cross-β interactions between LCDs[53,62]. The intriguing aspect of such a mechanism is that only one mode of interaction causes the formation of both droplets and amyloids, making the relationship between each state a matter of kinetics. While wild-type FUS droplets eventually form fiber structures, disease mutations drastically shorten the time required for FUS aggregation[59]. Experiments have also shown that some LCDs, including that of hnRNP-A1, undergo irreversible droplet condensation after a certain time period or after a number of cycles[53,54]. This lends further support to the idea that cross-β interactions require time to form hydrogels or aggregates. In an investigation of hnRNP-A2, Xiang et al. showed that a cross-β structure is found in liquid droplets, hydrogels and *in vivo*[60]. This use of a chemical foot-printing method would seem to confirm that these assemblies of LCDs lie on a continuum.

Other experiments lend support to an alternative interpretation: the molecular mechanisms for liquid-like and solid-like condensation are separable. The fact that droplet condensation is reversible in certain conditions could indicate that a Flory-Huggins phase separation occurs first, but that the high local concentration of an LCD forms cross-β

interactions[53,54]. This stepwise mechanism would require that the proteins remain disordered in droplets. The disorder of proteins in droplets was observed in an NMR study of FUS showing that its LCD maintains the disordered state after condensation[57]. Here, we are presented with a conundrum, because the same LCD cannot be both disordered (as described here) and simultaneously participate in cross-β interactions (as described above). Properties of hnRNP-A1 complicate matters. The protein contains a hexapeptide repeat that is frequently mutated in ALS and causes amyloid fibers to form, but this repeat is not necessary for condensation[54]. Indeed, this hexapeptide region alone is insufficient for precipitation, but when it is inserted into a globular protein like RNase A it creates robust amyloid fibers[48,82]. At a minimum, these results indicate that phase separation and precipitation are highly context-dependent.

1.4. Physiological consequences of condensation

Biological roles for liquid phase separation and amyloids

Molecular mechanisms notwithstanding, the range of properties displayed by liquid droplet systems appear to be leveraged by the cell. In the Cajal body, the increased local concentration of its components might be a kinetic driver of snRNP assembly, as predicted by simulation[83]. One notable example of how a condensate might promote biochemistry comes from plants: the carbon fixation enzyme Rubisco forms a condensate with EPYC1, a protein of the pyrenoid of *Chlamydomonas reinhardtii.* The ability to promote reactions would be a powerful use of condensation, but would likely require the droplet to remain highly fluid. The maturing droplets observed by Lin et al. might not be advantageous to reactions, but instead sequester mRNAs to prevent translation during stress, which is the proposed role of the stress granule[19,53].

Even amyloids, which are classically associated with human disease, have important physiological functions[84]. Bacteria and fungi have been shown to exploit amyloid fibrils for their mechanical properties and to use them to promote host invasion processes. Amyloid aggregation is used as a regulatory mechanism for heterokaryon formation in fungi and amyloid-like oligomers are essential for long-term memory formation in *Drosophila*[85]. Amyloid potentiates mammalian melanin production, and a number of endocrine hormones are stored in an amyloid-like state before dissociation upon release[86,87]. More recently, it has been shown that regulated amyloid-like protein aggregates control gene expression during yeast gametogenesis[88]. These apparent functions for amyloid suggest that its quaternary structure is not necessarily aberrant, but rather that other regulatory failures cause disease.

Disease states and failure of condensate regulation

Many of the proteins currently used as models for condensation and fibril formation are linked to ALS and frontotemporal lobar degeneration (FTLD). The most extensively studied are hnRNP proteins, FUS and RBM14, which contribute to ALS pathology[54,59,62]. These proteins all bear known mutations that are either causative in familial disease or correlate with spontaneously arising disease. For FUS, a possible mechanism of neurotoxicity has emerged: ALS linked FUS mutants sequester a variety of proteins involved in RNA metabolism, suggesting that the misregulation of condensation could cause disease through downstream effects on gene expression even if amyloid structures are not toxic on their own[58]. While it remains unclear what contexts allow amyloid promoting peptides to induce fibril formation, disease linked mutations facilitate their formation. Alzheimer's and Huntington's may also have a relationship to condensation

mechanisms[89,90]. These diseases often do not manifest until later in life, suggesting that compensatory mechanisms may fail during aging. One such example is the autophagy process, which is responsible for clearing stress granules that are not disassembled by other means[77]. Defects in the autophagy receptor in ALS patients are associated with amyloid aggregation of stress granule proteins in neurons[91]. Factors dictating whether a stress granule simply reverses phase separation or must be targeted for total enzymatic degradation are unknown.

Understanding condensation provides an opportunity to understand how we might intervene in disease. If we consider the formation of toxic amyloids to be a continuous process, the mechanism must be unified in some way. Therefore, any intervention in amyloid formation might also interfere with the normal functions of condensate organelles. On the other hand, we might consider a model where amyloid formation occurs independently of condensation, but is promoted by the high local concentration of certain proteins. In this scenario we might be able to block amyloid formation without disrupting phase separated bodies. Further, these bodies are regulated by disaggregase activity in addition to the phase separation process. That activity allows cells to guard against toxicity by digesting bodies and fibrils that will not dissolve on their own.

Broader biological consequences

While cellular bodies lacking membranes have been known for decades, whether they behave as liquid droplets *in vivo* has now become a pressing question. Recent advances have brought us a long way towards understanding the intermolecular interactions underlying these phenomena, including the importance of RNA as a structural component and regulator of condensation. Most of the recent work has focused on *in vitro*

characterization of stress granule proteins. If we are to form a generalizable understanding of condensation in cells, we must extend these studies to *in vitro* systems that more faithfully reproduce the molecular complexity of the *in vivo* situation by including protein and RNA interactors. Furthermore, we must examine how non-equilibrium processes mediated by chaperones, helicases, kinases, and proteases alter the dynamics of condensates and amyloid formation.

1.5. New tools and concepts for understanding condensation

Methodological transitions from test tubes to cells

The tools for studying condensates have developed along with our understanding of their formation. Early studies of biomolecular condensation (1990s and ongoing) were quantitative studies of the crystallin proteins of the mammalian eye lens[92,93]. These efforts sought a more complete understanding of protein solubility in relation to the high concentrations seen in the lens and what goes awry in the formation of cataracts. The gel-like barrier in the nuclear pore was studied by similar methods, where aggregation prone proteins were introduced at various concentrations concurrent with the removal of denaturing agents preventing condensation[30,32,33].

As our understanding that membraneless organelles may have related assembly processes, these same methods were applied to understand RNA binding proteins and IDRs by forming droplets and gels in test tubes[53,54,59,62]. The advantages of this approach include precise control over the protein concentration, biochemical purity of reagents, buffer composition, and environmental conditions. However, many studies were criticized for assays that were mismatched to physiological conditions for any particular droplet

formulation. Thus, a theoretical framework for protein condensate formation emerged, but cell-based assays were needed to fully establish the relevance of condensation.

Optogenetic control of biomolecular condensation

Studying endogenous condensates presents the critical challenge of manipulation of a membraneless organelle in a rigorous and reproducible manner. Fortunately, the mesoscale nature of most condensates makes them amenable to study by fluorescence microscopy, and therefore optogenetic techniques have been fruitful. The Cry2 domain from *A. thaliana* forms dimers in response to blue light activation, and is functional in mammalian cells[94]. Self-oligomerization has been successfully used to recapitulate some of the *in vitro* results seen with hnRNP-A1, DDX4, and FUS[95]. This study underscored the requirement for multimerization to form condensates in live cells: the IDRs in this study did not have a high enough affinity for themselves to overcome competition from physiological molecules until multimerization was induced. The essentiality of a specific interaction to augment the role of IDRs was further borne out by *in vitro* studies of RNA-binding proteins, including hnRNP-A1[96]. Confirmation of the optogenetic findings begins to build a model where many protein architectures can provide the multivalent network required to make condensates, and IDRs are one possible component.

The Brangwynne lab has continued to drive optogenetic manipulation of condensates forward since the introduction of Cry2. Material properties of the nucleolus can be manipulated using this system, raising intriguing possibilities about the biological ramifications of the physical properties of this organelle[97]. Others have used Cry2 to study the hours-long effects of aggregation by G3BP1, a stress granule protein, in neurons[98]. Using ferritin clusters, precise control of the interaction number for various proteins can

be used to map phase diagrams under controlled *in vivo* conditions[99]. Adapting Cas9 to interact with the Cry2 binding partner Cib1 has allowed for the probing of condensate interactions with the chromatin itself[100]. These studies define the frontier of *in vivo* condensate cell biology and raise many interesting hypotheses to be addressed.

Condensate purification

Several attempts have been made to isolate membraneless organelles from live cells. Such purifications are not new in and of themselves: isolation of nucleoli has been performed for decades, and a Cajal body isolation was first published nearly twenty years ago[101,102]. The large size of *Xenopus* oocytes makes their nuclear bodies amenable to study under oil, and many foundational observations about membraneless organelles have been made there[27,29,103]. The advent of biomolecular condensation as a likely assembly mechanism for these organelles has recast these purification methods in a new light. Many do not account for the possibility that low affinity interactions likely hold important components in place.

Several recent efforts have shown that cytoplasmic membraneless organelles can be purified and analyzed with contemporary methods. The stress granule, like the nucleolus, appears to have subdomains with different properties and compositions[104]. Purification of the core compartment is achievable through differential centrifugation and affinity purification coupled to proteomic and RNA-seq analysis[104,105]. While these efforts did not fully account for what might be lost to the more dynamic 'shell' compartment of the stress granule, they allowed significant insight into the residents of this organelle. Using a fluorescence sorting method, new information is coming to light for the related cytoplasmic organelle, the P-Body[106]. At this time, cross-linking has not been used

extensively in the isolation of condensates, as there is a high probability it may introduce artifacts. However, overcoming this significant technical challenge may become necessary if transient molecular residents of condensate organelles are to be characterized.

Conceptual advances

The seminal discovery that *C. elegans* P-granules are phase condensates introduced a framework from materials science into biology[28]. Since that time, several other key developments have changed our thinking about how phase condensate systems form and function. While IDRs were justifiably the subject of intense study for many years, the nuances of condensation in real biological systems are becoming clear, including the roles of larger complexes, the definition of specificity in a condensate, and the functional implications of undergoing phase transitions.

The scale and composition of complexes that undergo biomolecular condensation is expanding dramatically. Recent studies have shown that chromatin, including protein and DNA, can undergo a phase transition[107]. Remodeling of the nucleosome by Swi6 triggers condensation in the formation of heterochromatin, implying a regulatory role for this process[108]. Such results represent a leap forward from earlier work where only a single IDR was considered in the formation of *in vitro* condensates. Nucleosomes are octamers of histones that interact tightly with DNA, making chromatin an example of a multi-component system capable of forming a condensate with biological function.

Before these multicomponent systems were studied, the field lacked a conceptual model for how condensates could establish specificity of membraneless organelles in cells. An analysis of Nephrin and SH3 domains has shown that the properties of two components that bind each other can produce a condensate with tunable properties[109]. These studies

were extended to analyze how the relative amounts of certain co-condensing molecules can control which components act as 'scaffolds' and which act as 'clients'[110,111]. Thus, a framework was established for how stoichiometry and available binding sites set up the capability of one molecule to form a condensate and another to recruit into it. This model was later applied to explain how a condensate formed by Nephrin, Nck and N-WASP increase the dwell time of the Arp2/3 complex at membranes, promoting actin assembly. This result demonstrates a kinetic effect of a physiological condensate and brings the arc of SH3 domain and proline rich motif condensation to completion[112].

As a picture of specificity in condensates takes shape, we have learned more about how they may contribute to biological function and direct specific activities. The revelation that RNA itself can form condensates through base pairing provides another platform for specific multivalent interactions[69,71]. The assembly of RISC condensates relies on an IDR in Ago2, but structural study has revealed that there are specific binding sites for tryptophans in that IDR which bind to form the condensate[113]. We are beginning to appreciate how specific amino acids can act as binding sites in the formation of membraneless organelles, especially if those amino acids are the target of post-translational modification. A transition between transcriptional activation condensates and splicing condensates is controlled by phosphorylation of RNA Pol II, demonstrating that relatively small molecular events can control mesoscale assemblies and their biological activities[114].

1.6.   The Cajal body

The Cajal body is an ideal system for studying the relevance of biomolecular condensation on the assembly of a nuclear body. It possesses nearly all of the properties one would expect of a condensed membraneless organelle, and its morphology is

predictable and well documented in many contexts. Removing a single protein, coilin, causes disassembly of the body, providing a convenient tool for manipulation of the organelle[115,116]. Thus, I will review the major functions and components of the Cajal body to provide context to my experimental work presented in following chapters.

Functions of the Cajal body

The Cajal body is a hub of RNP processing in the nucleus. Genomic loci encoding the non-coding RNAs often found in the Cajal body tend to be found in their vicinity[117]. Coilin is associated with these RNA species and is detected at these loci by ChIP, in support of a model where the Cajal body is a site of transcription for these genes[68]. Some of these genes include the U snRNPs that go on to participate in splicing and are themselves enriched in the Cajal body, but also include those for snoRNAs that aid in ribosomal RNA processing[118,119]. If these genes are in fact transcribed in, and the RNAs transferred to the Cajal body, then the general model of nuclear bodies as RNA processing sites applies.

Other functions have also been associated with the Cajal body. A relationship to telomerase assembly and telomere maintenance has been controversial and remains an unresolved question[120-122]. The presence of certain processing factors, like TGS1 and WRAP53 raises questions of whether more RNA modification activity occurs in Cajal bodies than has been described in detail[68,123]. Early screening for export competency occurs in the Cajal body, where the export factor PHAX is added to RNA after associated factors are assembled[124]. The presence of other factors in various contexts is provocative, and as we learn more about this organelle, we may find that it is involved in even more diverse biochemical processes[4].

U snRNP assembly and processing

Splicing of pre-mRNA is a hallmark of eukaryotic gene expression and Cajal bodies play an important role in maintaining the splicing machinery. As noted above, it seems likely that U snRNAs are transcribed directly into the Cajal Body[68,117,125,126]. After transcription, these snRNAs are exported to the cytoplasm to be assembled with their protein components. There, Sm proteins are added and the 5′ end is hypermethylated by Tgs1[127,128]. This process of incorporating protein components requires SMN and the other gemin proteins which together form the SMN complex[129-131]. Once assembly is complete, snRNPs return to the nucleus. There they return to the Cajal body to complete processing and for the U4, U5, and U6 snRNPs to be assembled into the tri-snRNP required for splicing[16]. As snRNPs are expended by splicing reactions, they return to the Cajal body for reassembly[132]. The presence of the body is thought to impart a kinetic advantage to keep up with the splicing load of the cell[83].

The Cajal body is crucial for snRNP assembly in specific cell types and during certain moments of development. The best evidence for this comes from the knockdown of coilin during zebrafish embryogenesis: the embryos arrest, but can be rescued by supplementing them with preformed snRNPs[115]. This effect is not idiosyncratic to zebrafish, as a genetic knockdown in mice also affects fertility of subsequent generations[116,133].

The importance of Cajal body integrity is not limited to development. SMN joins two nuclear bodies after import into the nucleus: the gem and the Cajal body. Depletion of SMN results in defects in the Cajal body and a loss of snRNP enrichment in those compartments[125,134,135]. This pattern of Cajal body defects is seen in the genetic disease

spinal muscular atrophy (SMA)[136]. In SMA, a defect in the gene SMN1 results in a low overall level of SMN and fatality in early childhood for most patients[137]. That patients survive at all is due to the low level of SMN expression from the SMN2 gene, because organisms without any SMN gene fail to complete embryogenesis[138,139]. The necessity for Cajal bodies in various tissues beyond embryos and motor neurons remains to be seen, but the evidence supports the view that cells with high splicing requirements need the high turnover rate of snRNPs provided by this organelle.

Arguments for condensation in the Cajal body

Several lines of evidence support the idea that the Cajal body forms through biomolecular condensation. As discussed in section 1.2, a study from the Gall lab postulated that nuclear bodies could behave as liquid within liquid phase separations as early as 2005[27]. One of the hallmarks of liquid condensates is that they spontaneously fuse to minimize surface area, which has been observed with Cajal bodies[26]. The protein components like coilin and SMN all exchange with the bulk nucleoplasm on time scales of seconds, showing the Cajal body is dynamic[9]. Nucleation of the body can be achieved by anchoring any of several protein components to a specific location in the genome[140]. That finding supports the view that a network of interactions between proteins (and likely RNA) holds the body together, rather than a stepwise assembly process. Finally, many of these proteins also have a self-interaction property, which provides the multivalent platform we expect for condensation[34,141].

Consideration of the relationship of the Cajal body to other nuclear bodies, raises questions about how condensation may be at work. SMN is in the Cajal body, but also joins the nuclear gem[142]. The gem is enriched in U1 and is likely to be a hot spot for assembly

of that specific snRNP[143]. If different RNPs are being processed in these different bodies, variable nucleation sites could explain why the bodies are often separate even though they share components. Indeed, the molecular link between coilin and SMN is an arginine methylation site on coilin that is bound by SMN's tudor domain[144,145]. Without that methylation, the two bodies dissociate and Cajal bodies disassemble. Such molecular events are less well understood for another related body, the histone locus body (HLB). In many cell types HLBs are merged with the Cajal body, while in other cases they remain separate[4,146]. Why this is the case remains unknown.

Specific molecular components of the Cajal body also exhibit features that would suggest condensation as a mechanism. Nopp140 is a nucleolar protein that binds to coilin, is enriched in the Cajal body, and its residence there correlates with severity of SMA[147,148]. Nopp140, coilin, and SMN have stretches of IDR that might promote condensation (Figure 1, page 28). The SMN N-terminus contains a lysine-rich region, while the C-terminus has a proline-rich tract and a tyrosine/glycine repeat motif. Coilin has a disordered stretch linking the N-terminal domain (NTD) and C-terminal domain (CTD). Nopp140 is almost entirely disordered and enriched for serine and acidic amino acids. All three proteins have domains that allow for self-interaction, raising the potential for multivalency[141,149,150]. A large number of RNA binding proteins are present, providing for ample interactions with any transcripts or snRNPs available[68]. The assembly activity of coilin, and possibly other proteins, can be regulated by phosphorylation events, though the precise consequences of modification are still unclear[151]. Together, we have evidence to hypothesize that the Cajal body is assembled through a condensation process; what remains is to directly determine how this might occur.
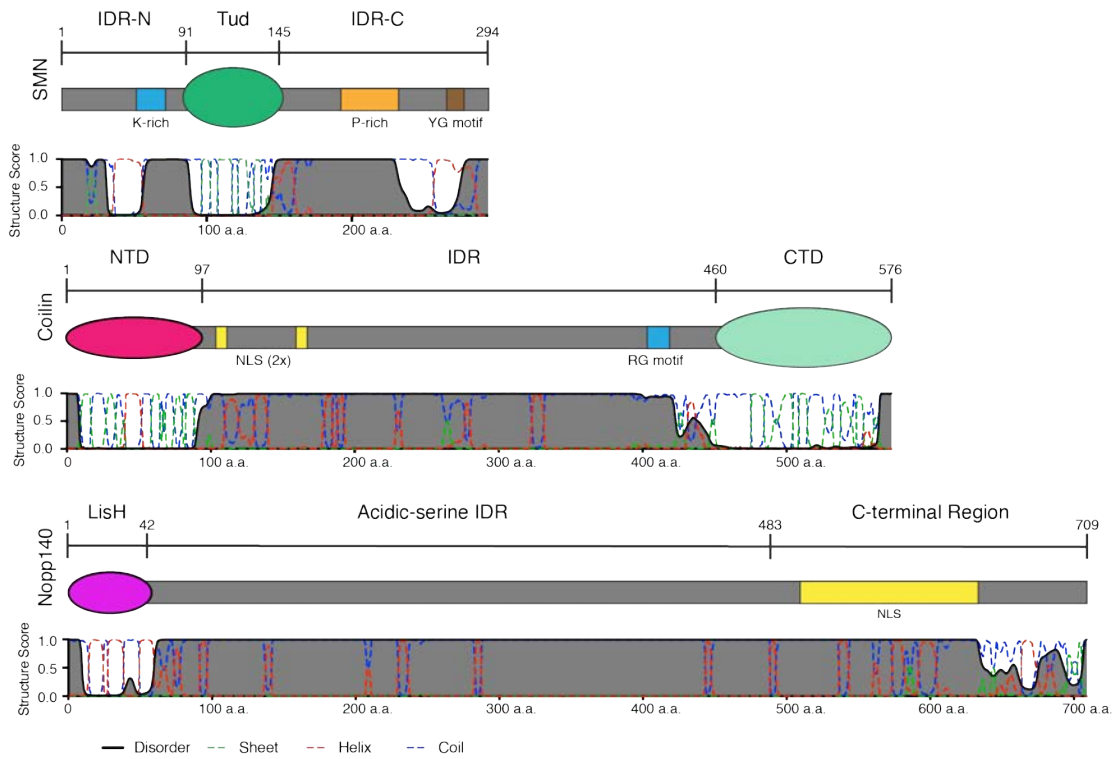
Figure 1. Predicted secondary structure of SMN, Coilin, and Nopp140.

Schematic representations of SMN, coilin, and nopp140 domain architecture and accompanying secondary structure prediction. Structure score is a unitless value for secondary structural properties predicted by the RaptorX algorithm [152]. Abbreviations: Intrinsically disordered region (IDR), Tudor (Tud), N-terminal domain (NTD), C-terminal domain (CTD), Nuclear localization signal (NLS).

Chapter 2. Condensation of Cajal Body Proteins

2.1.   Goals and approach

This chapter will address the condensation properties of three Cajal body proteins, SMN, coilin, and Nopp140. I will begin by discussing our rational for examining SMN, as it is intimately involved in the snRNP biogenesis role of the Cajal body, and then move on to coilin and Nopp140. The majority of this chapter will detail the mechanism of SMN tudor domain condensation. I assert that this mechanism is a new way in which cells regulate their assembly through dynamic and specific interactions. I have chosen an *in vivo* optogenetic approach over the *in vitro* assays discussed in Chapter 1 in order to avoid the difficult of choosing physiologically relevant conditions. I will conclude the chapter with a discussion on the broader implications of these findings for nuclear bodies overall.

Outside contributions

Section 2.2 was adapted from a manuscript titled "Arginine modifications trigger condensation of numerous tudor domain proteins *in vivo*" submitted for publication in January, 2020. Andrew Barentine assisted with image analysis by preparing the mathematical framework and writing programs used to analyze live cell imaging data. Korinna Straube carried out the molecular cloning of constructs used throughout this chapter. Karla Neugebauer and Joerg Bewersdorf oversaw the project and helped revise the manuscript. I wrote the first draft and compiled subsequent revisions. I prepared all cell lines and performed all experiments. The specific contributions of others are noted in the relevant figure legends.

2.2. Condensation properties of SMN

Does SMN have condensation propensity?

As discussed, our understanding of biomolecular condensation has developed from descriptive to mechanistic. Multivalency is a key feature of condensing systems, allowing for the formation of a high avidity network of interactions[28,34,109,153]. IDRs have been the intense focus of many studies that demonstrate their capacity to form this network[53,59-61,154]. Yet how these principles apply to the scaffolding and function of many membraneless organelles, which contain IDRs and a host of other proteins, remains unknown. SMN is a key component in the biogenesis of small nuclear ribonucleoproteins (snRNPs), and a deficiency of SMN results in the fatal childhood disease, spinal muscular atrophy (SMA)[131,155,156]. When SMN is depleted from cells, three membraneless compartments – gems, Cajal bodies, and U-bodies – no longer form [125,134,135,157-159]. How loss of snRNP assembly and/or compartmentalization contributes to SMA is currently unknown.

I asked whether SMN plays a direct role in condensate formation, by isolating each region of the protein and assessing its condensation activity. SMN has N-terminal and C-terminal IDRs flanking a single tudor domain that binds symmetric dimethyl arginine (DMA) on snRNPs and other ligands (Figure 1, page 28)[160,161]. Some IDRs can form droplets spontaneously *in vitro* but require multimerization *in vivo*[59,95]. Thus, we fused SMN regions to the light-activated dimerization domain Cry2. This approach was previously exploited to form "optodroplets" with IDRs in a light-dependent manner when expressed in NIH-3T3 cells (Figure 2, page 32)[95]. Similar to prior data, mCherry-Cry2 did not form clusters, while FUS[IDR] and hnRNP-A1[IDR] did (Figure 3, page28). This effect is presumed to be due to the added multivalency of interactions between the IDRs, as is the

case *in vitro*. Thus, this assay will allow us to probe for multivalent assembly that might be present in any of the domains of SMN or other proteins.

I observed condensation, but not in the regions of SMN we anticipated. Unexpectedly, the SMN IDRs did not cluster, while the tudor domain (SMN^{Tud}) formed obvious clusters throughout the cell upon light activation (Figure 4, page 34). These clusters were heterogeneous and dynamic: when photobleached, they showed 42 -100% recovery with time constants of 30 ± 22sec (mean $\pm$ SD, n=10) (Figure 5A, page 35). Cluster fusion was infrequent but observable (Figure 5B, page 28). When Cry2 was allowed to deactivate after cluster formation, SMN^{Tud} clusters dissipated within a few minutes (Figure 5C, page 35). Taken together, these results suggest that SMN^{Tud} may sometimes nucleate around immobile structures and other times it forms condensates *de novo.* Fusion might be observed more frequently over longer observation times, but most importantly, the clusters are not static or solid, suggesting they are more akin to liquid condensates. We conclude that SMN^{Tud} is sufficient for the formation of dynamic condensates upon light-induced multimerization. This finding was striking, because tudor domains are small (60 amino acids) and structured, rather than disordered and repetitive in sequence like FUS^{IDR} and hnRNP-A1^{IDR}.
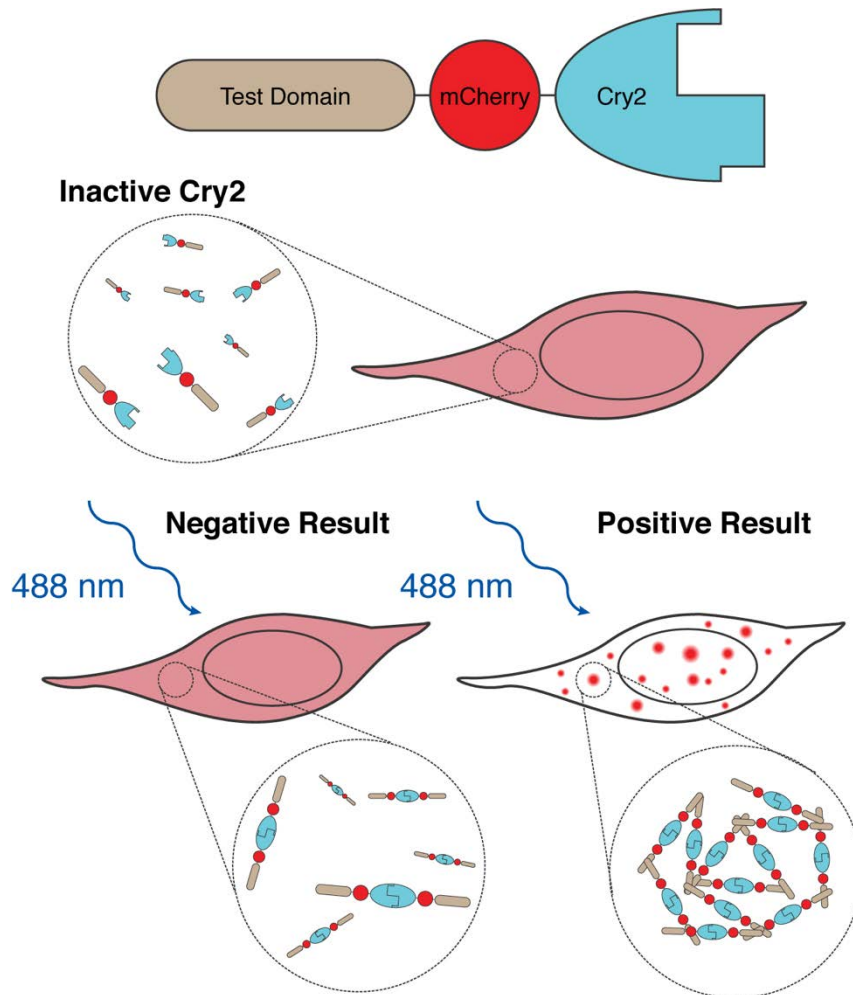
Figure 2. Diagram of the Cry2 condensation assay.

Cry2 dimerizes upon blue light activation, but without sufficient interactions contributed by the test domain, condensation will not occur. If the test domain provides interactions to increase valency, condensation is observed as mCherry fluorescent foci.
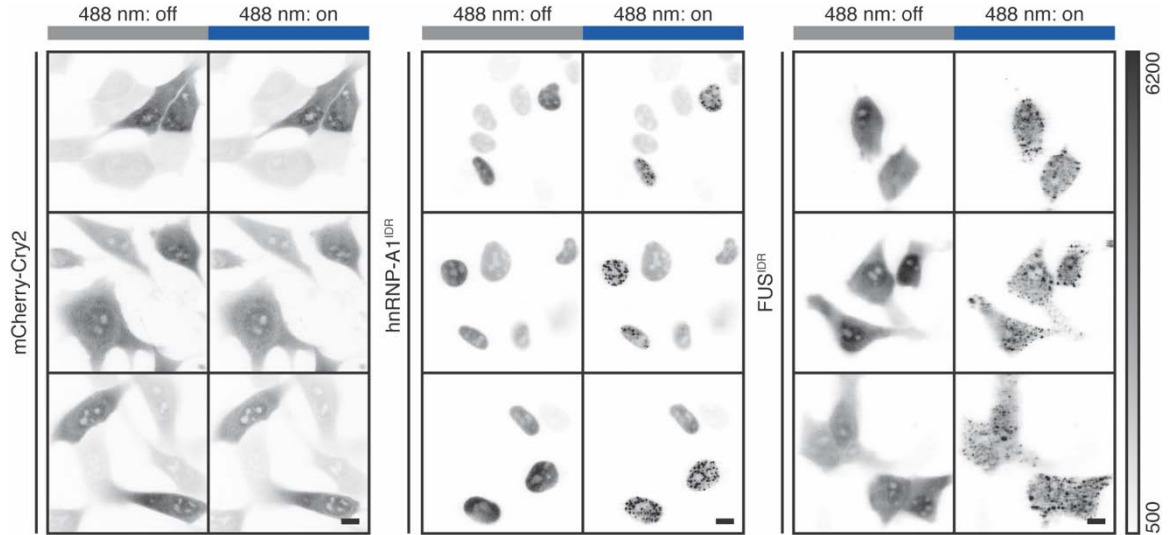
Figure 3. Condensates are formed by IDRs but not mCherry alone.

Micrographs of live cells expressing mCherry-Cry2, hnRNP-A1[IDR], and FUS[IDR]. mCherry does not cluster while the IDRs do. Results produced using the constructs and protocols from Shin, Y., et al. [95]. Grayscale bar is in analog-digital units. Scale bar = 10 μm.
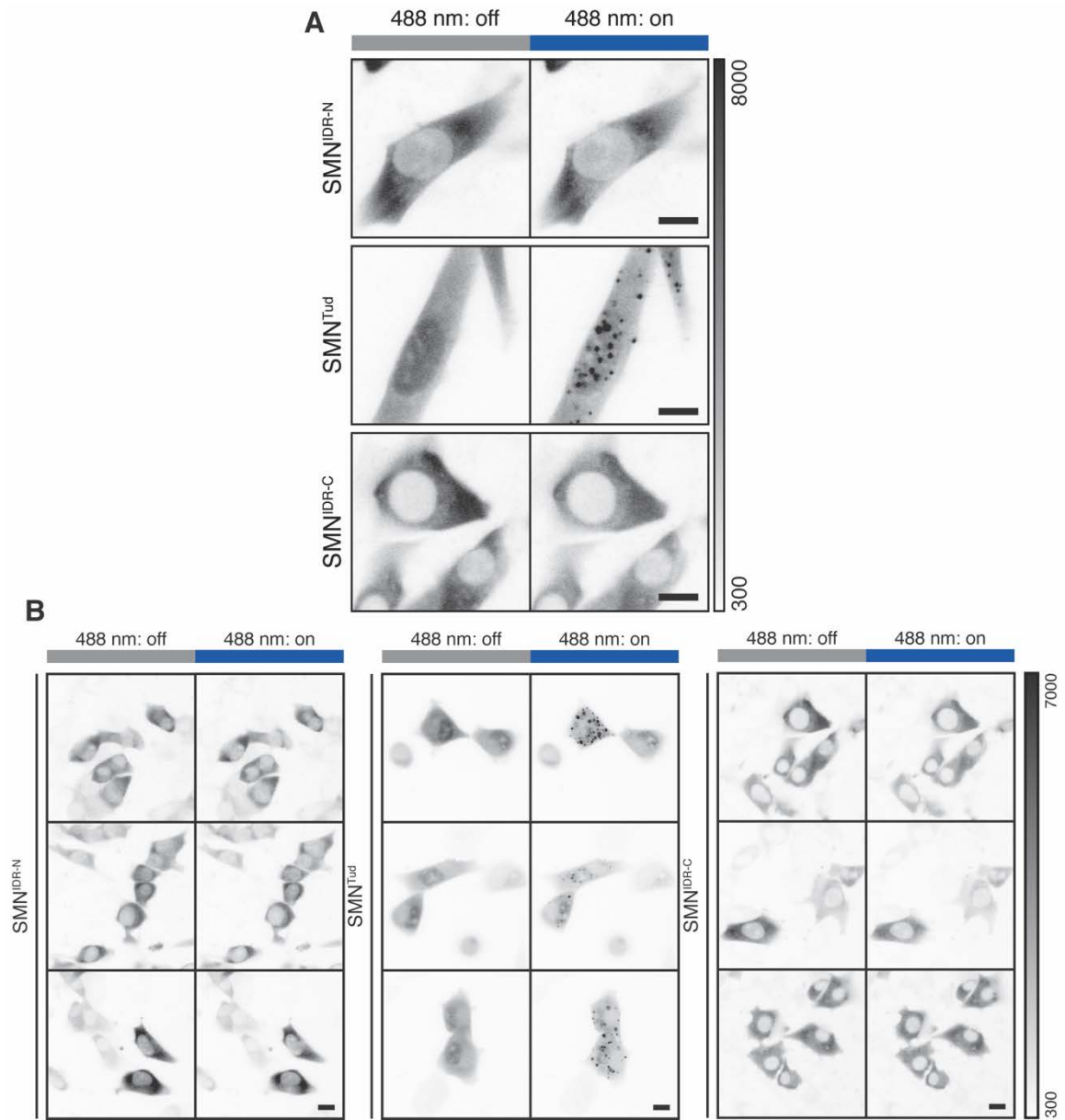
Figure 4. SMN^{Tud} forms clusters after Cry2 activation.

Micrographs of live cells undergoing blue light activation of Cry2. A) Zoom on single cells. B) Fields of view of SMN fragments in the Cry2 assay. Grayscale bar given in analog-digital units. Scale bar = 10 μm.
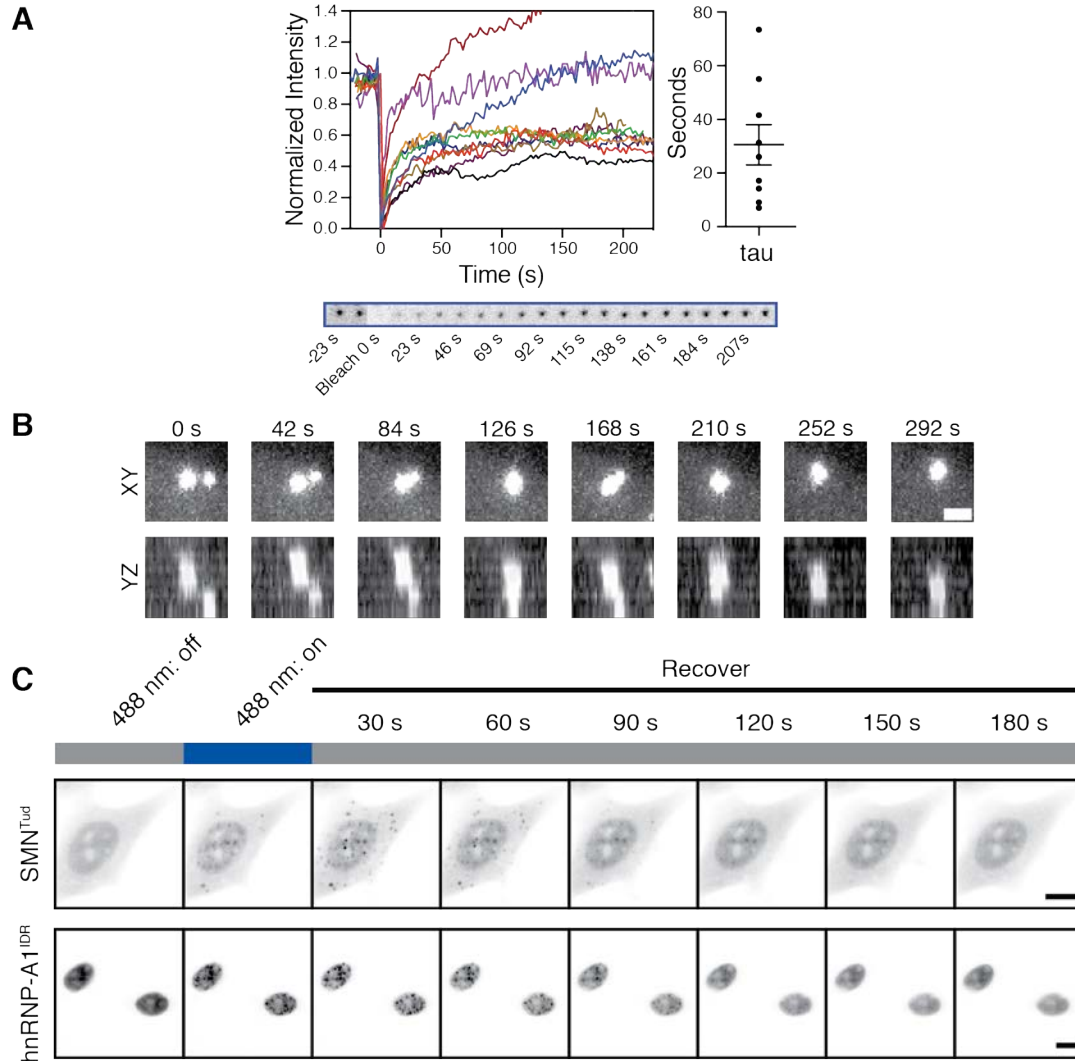
Figure 5. SMN^{Tud} clusters are dynamic condensates.

A) Fluorescence recovery after photobleaching for ten SMN^{Tud} condensed droplets with corresponding tau values from single exponential fits. Example images from blue trace are provided below. B) Example of a fusion event of two condensates. Scale bar = 2 μm. C) Examples of condensate dissipation after Cry2 inactivation in live cells. Blue light pulse was 30 seconds. Scale bar = 10 μm.

35

Does SMN^{Tud} condensation depend on binding its ligand, DMA?

I hypothesized that condensate formation by SMN^{Tud} depends on its recognition of its ligand DMA. Structural studies have provided a precise mechanism for tudor binding to DMA (Figure 6, page 38)[162]. To test this, I developed an image analysis method in collaboration with Andrew Barentine to quantify condensate formation compared to expression level (See Methods, page 65). First, arginine dimethylation was blocked by inhibitors of separate methyltransferases that produce either asymmetric or symmetric DMA (aDMA, sDMA) (Figure 7, page 39). There are three types of arginine protein methyltransferase (PRMT) in mammals[163]. The precise roles for each of the nine PRMTs are still unclear, but PRMT1 produces most aDMA and PRMT5 produces most sDMA[164-166]. Two small molecule inhibitors were used to block the synthesis of aDMA and sDMA. The Type I methyltransferases were inhibited by treatment with MS-023, and PRMT5 was inhibited with EPZ015666, both of which bind the active sites of their targets[167,168]. DMA inhibition using both drugs simultaneously reduced clustering by 35% for those cells that still form clusters above the significance threshold (Figure 8, page 40). Many no longer condense at all.

Next, SMN^{Tud} condensation was tested with a series of mutants: changes in three amino acids in the aromatic cage that accepts DMA (W102L, Y109L, and Y130D), a mutation associated with SMA (E134K), and an uninvolved phenylalanine as a control (F118L) (Figure 6, page 38)[155,162]. The binding site mutations and E134K all eliminate DMA binding in other experiments, and indeed they also eliminate condensation (Figure 9, page 41; Figure 10, page 42)[162]. F118L is on the opposite face of the domain from the binding pocket and is not expected to have an effect on condensation. As I predicted, it did

not (Figure 11, page 43). I therefore conclude that the integrity of the SMN^{Tud}•DMA interaction is required for condensation.
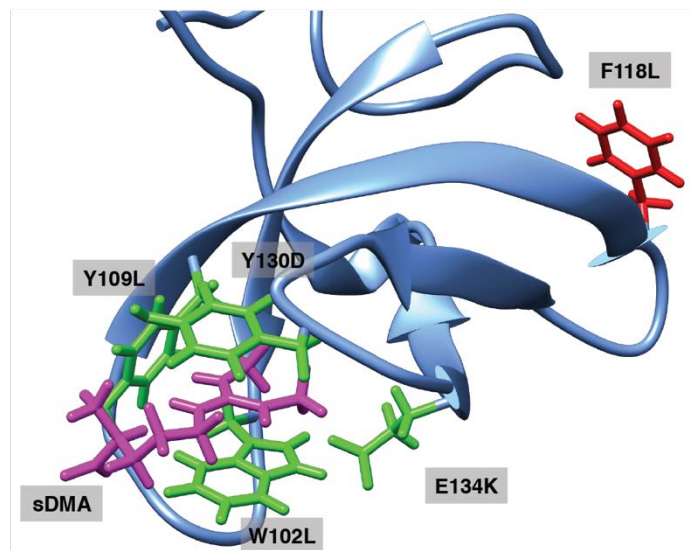
Figure 6. The SMN tudor binding pocket.

Published solution structure of the SMN tudor domain bound to sDMA (magenta). Three aromatic amino acids that make up the binding pocket and one associated with SMA are shown in green. An aromatic residue not involved in DMA binding is shown in red. PDB: 4A4E[162].
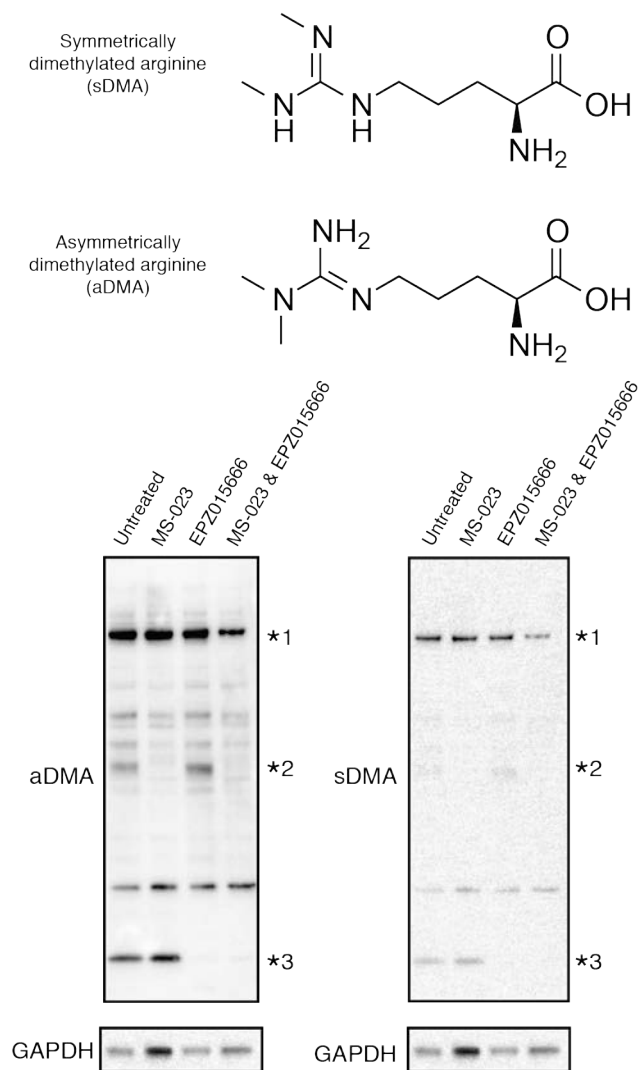
Figure 7. MS-023 & EPZ015666 inhibit DMA synthesis.

Chemical diagrams of sDMA and aDMA are shown above. Whole cell lysates from NIH-3T3 cells from the same biological replicate treated for 48 h with MS-023, EPZ015666, or both, western blotted for aDMA and sDMA shown below using antibodies against each modification (see Table 5). GAPDH blot is from separate lanes and shown next to both DMA blots. Numbered asterisks indicate 1) a band only reduced when treated with both drugs, 2) the major band reduced by MS-023, and 3) a band reduced by EPZ015666.
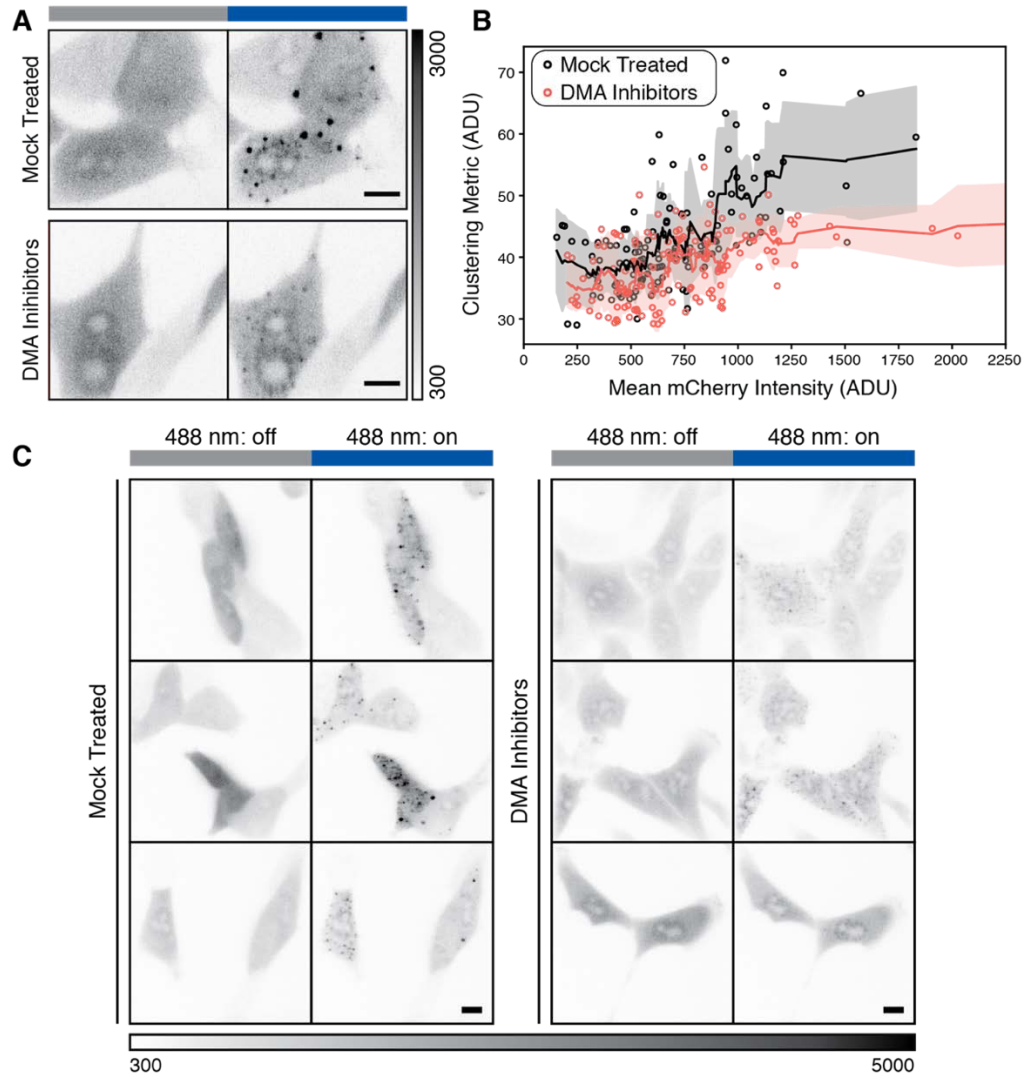
Figure 8. DMA inhibitors reduce SMN^Tud condensation.

A&C) Micrographs of live cells with SMN^Tud either untreated or treated with DMA inhibitors, presented as A) zoomed in on individual cells or C) fields of view. Grayscale bar given in analog-digital units. Scale bar = 10 μm. B) Quantification of live cells untreated or untreated with DMA inhibitors. Mean mCherry intensity and the cluster metric are given in analog-digital units (ADU). Solid line with shading is a rolling mean and standard deviation of 10 points. Each point is one cell. Andrew Barentine contributed the analysis methodology and programming used to generate the clustering metric.
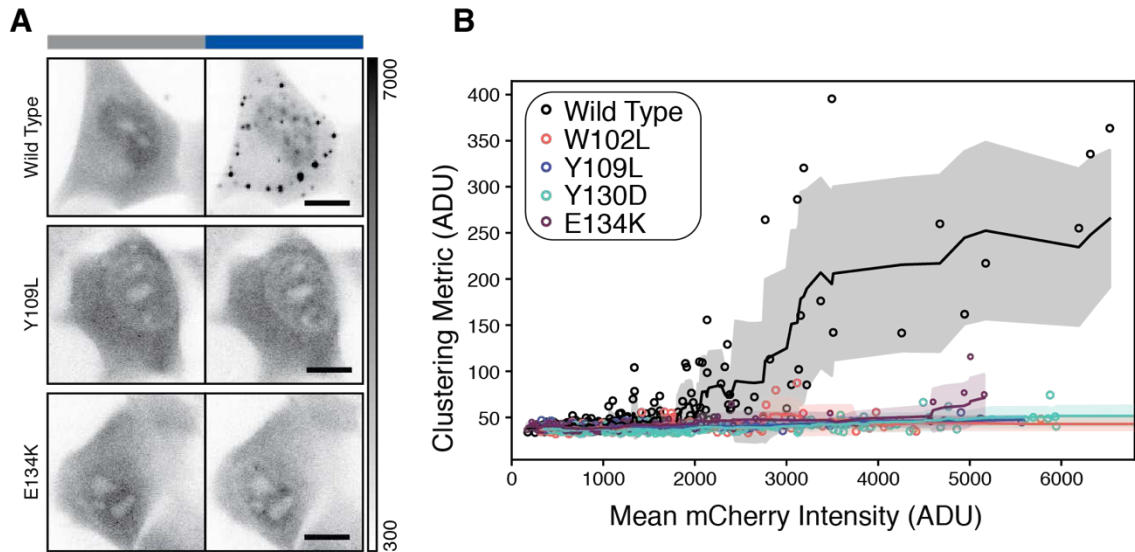
Figure 9. Mutations affecting DMA binding eliminate condensation.

A) Micrographs of live cells expressing SMN^{Tud} wild-type, a mutation to the aromatic DMA binding cage Y109L, and an SMN associated mutation E134K. Grayscale bar given in analog-digital units. Scale bar = 10 μm. B) Quantification of live cells expressing wild-type or mutant SMN^{Tud}. Mean mCherry intensity and the cluster metric are given in analog-digital units (ADU). Solid line with shading is a rolling mean and standard deviation of 10 points. Each point is one cell. Andrew Barentine contributed the analysis methodology and programming used to generate the clustering metric and this plot.

Figure 10. All four mutations affecting DMA eliminate condensation.

Micrographs of live cells expressing SMN^{Tud} or SMN^{Tud} with point mutations. W102L, Y109L, Y130D, and E134K that reduce binding to DMA[162]. Grayscale bar is in analog-digital units and applies to all images. Scale bar = 10 μm.
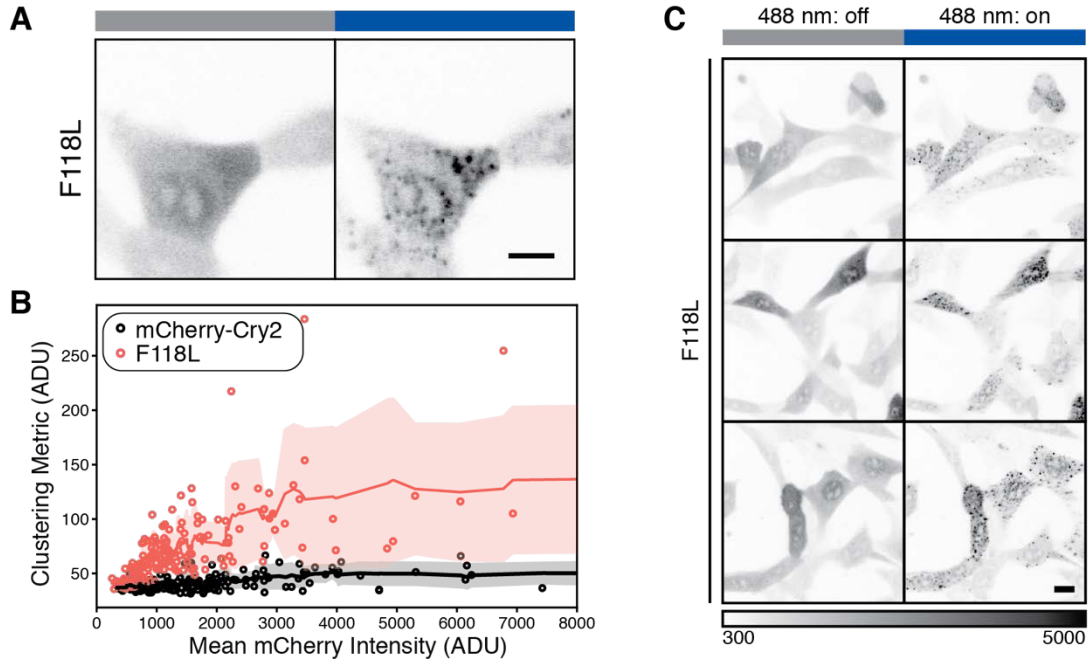
Figure 11. F118L does not prevent condensation.

A&C) Micrographs of live cells expressing SMN^{Tud} F118L where A) is zoomed to a single cell and C) gives fields of view. Grayscale bar is given in analog-digital units (ADU). B) Quantification of live cells expressing SMN^{Tud} F118L compared to control mCherry-Cry2 cells. Units are analog-digital units (ADU). Solid lines and shading are a rolling mean and standard deviation of 10 points. Each point represents one cell. Andrew Barentine contributed the analysis methodology and programming used to generate the clustering metric and this plot.

How do other molecules interact with SMN^{Tud} condensates?

Natural DMA ligands of SMN may co-condense with SMN[Tud] in our assay, providing insight into the relationship between our synthetic condensates and membraneless organelles. Although methylation of snRNP proteins mediates binding to SMN, no snRNP marker was found in SMN[Tud] condensates (Figure 12, page 46)[160]. Indeed, a large number of proteins and interacting moieties that one might expect to interact with SMNTud condensates failed to concentrate there (Table 2, page 47)[169,170]. The absence of many of these proteins prompted me to note that the assembly of condensates could be specific to certain SMN interactors.

The only noteworthy positive marker in SMN[Tud] condensates was coilin. Arginine methylation of coilin is required for proper assembly of the Cajal body and endogenous coilin was detected in SMN[Tud] nuclear condensates (Figure 13A, page 48)[144]. NIH-3T3 cells do not have Cajal bodies under normal conditions, so we rule out incorporation into a pre-existing structure. To determine if coilin association is essential for condensate formation by SMN[Tud], mouse embryonic fibroblasts that lack coilin were tested, and SMN[Tud] condensates still formed (Figure 13B, page 48). These observations confirm that SMN[Tud] can recruit DMA ligands to condensates. Furthermore, ligands besides coilin must support condensate formation by SMN[Tud], implicating a network of modular interactions between SMN[Tud] and DMA modified proteins.

To determine whether condensation is a shared property of tudor domains that bind DMA ligands, we turned to the founding protein of this family: *D. melanogaster* Tudor[171]. The proteins Tudor, Aubergine, Vasa, and the oskar RNA make up germ plasm condensate in flies[24]. DMA in the Aubergine N-terminus is required for germ plasm localization, and

the binding between this region and the eleventh tudor domain of Tudor has been studied by co-crystallization[172]. I designed a Cry2 construct with this tudor domain plus flanking sequence for specificity (*dm*Tudor^{Tud}) and a construct with three repeats of the Aubergine N-terminus fused to GFP (Aub$_3$-GFP) (Figure 14A, page 49). This interaction pair was introduced into NIH-3T3 cells. Western blotting shows that Aub3-GFP is predominantly modified for aDMA (Figure 14B, page 49).

*Dm*Tudor^{Tud} was able to form condensates in the nucleus of NIH-3T3 cells without the addition of any ligand, showing that endogenous murine ligands present in the nucleus suffice (Figure 14C, page 49). I infer that no cytoplasmic protein is recognized by *dm*Tudor^{Tud} due to the lack of condensation there. Co-expression with Aub$_3$-GFP competed with the endogenous protein and prevented condensate formation (Figure 14D, page 49). When lysine was substituted for arginine in Aub$_3$-GFP, condensation was restored, confirming the essentiality of the DMA•tudor interaction (Figure 14D, page 49). The endogenous proteins that allow *dm*Tudor^{Tud} to condense must be nuclear, because no effect was seen in the cytoplasm. The recognition of a mouse protein by *dm*Tudor^{Tud} likely reflects the documented promiscuity of tudor domain recognition of DMA targets[173].

Figure 12. snRNPs are not in SMN^{Tud} condensates.
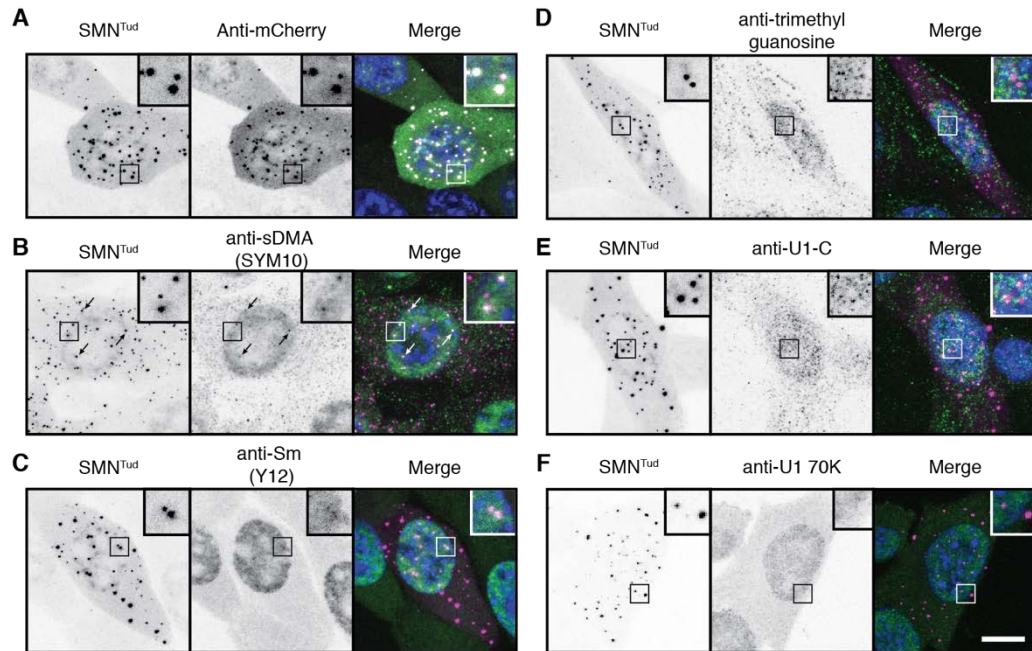
Fixed cells with Cry2-active SMN^{Tud}, which is always visualized by mCherry fluorescent signal. Condensates were stained for A) mCherry, B) sDMA, C) the Sm epitope with Y12, D) the trimethylguanosine 5′ cap E) U1 snRNP-C and F) U1-70k, both U1 snRNP specific proteins. In B) arrowheads indicate nuclear condensates that stain positively with SYM 10. Scale bar = 10 μm.

Table 2. SMN^{Tud} markers tested by immunofluorescence.

Marker	Localized to cytoplasmic condensates	Localized to nuclear condensates
mCherry	Yes	Yes
SYM10	No	Yes
SYM11	No	Yes
ASYM24	No	Yes
Sm(Y12)	No	Yes
Trimethyl guanosine	No	No
SmD3	No	No
U1-C	No	No
U1-70k	No	No
Coilin	No	Yes
SMN	No	No
FUS	No	No
hnRNP U	No	No
hnRNP K	No	No
hnRNP A1	No	No
G3BP1	No	No
R-Loop	No	No
Fibrillarin	No	No
Nopp140	No	Yes*
SFPQ	No	No
Pol II	No	No

Each marker was assayed using a primary antibody corresponding to that epitope. Those marked as "Yes" were observed in that compartment greater than 50% of the time (* indicates Nopp140 was observed but for less than 50% of condensates). "No" indicates that these markers were never observed to concentrate in SMN^{Tud} condensates.
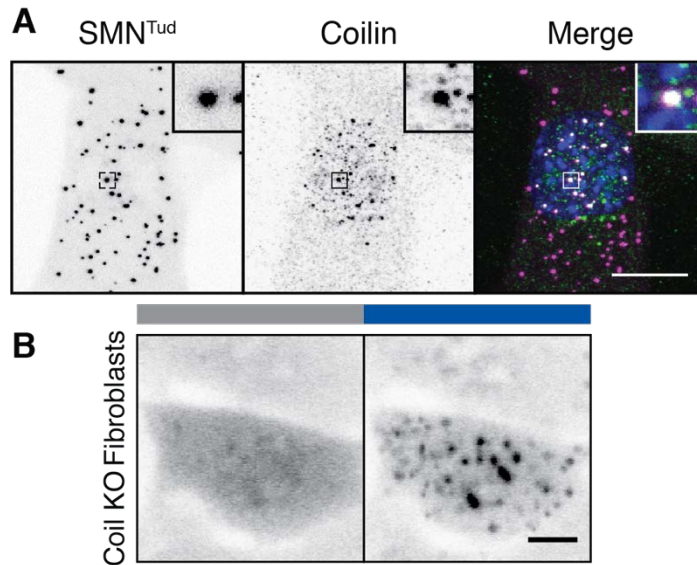
Figure 13. Coilin is a non-essential component of nuclear SMN^{Tud} condensates.

A) Light-activated and formaldehyde fixed cells expressing SMN^{Tud} and stained for endogenous coilin. In merge, blue indicates Hoechst staining. B) Micrographs of Cry2-inactive and Cry2-active SMN^{Tud} in coilin knock-out mouse fibroblasts. Scale bar = 10μm.
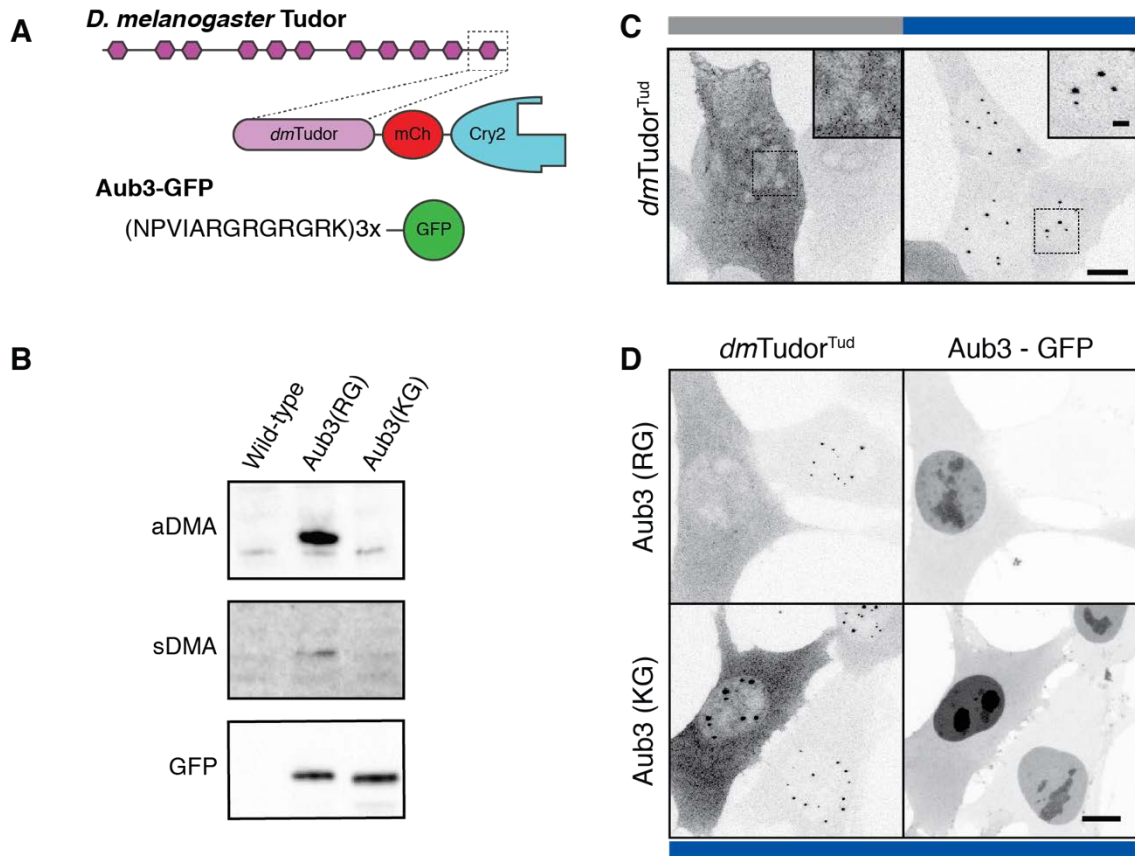
Figure 14. *Dm*Tudor^Tud condensates are disassembled by Aub₃ competition

A) Schematic of *D. melanogaster* Tudor, *dm*Tudor^Tud and Aub3-GFP. B) Western blot of whole cell lysates of NIH-3T3 wild-type, expressing Aub3-GFP, or expressing Aub3-GFP with arginine mutated to lysine. Each panel shows the range of ~25-40 kDa. C) Fixed cells expressing only *dm*Tudor^Tud in Cry2-active and Cry2-inactive states. D) Fixed cells expressing *dm*Tudor^Tud and either Aub3-GFP, or the same peptide with R to K mutations. Cry2 is active in all frames. Scale bar = 10 μm, inset scale bar = 2 μm.

Generality of tudor domain condensation

The human proteome has fifty-five annotated tudor domains from twenty-eight different proteins. Each tudor domain binds a discrete set of DMA ligands due to amino acids surrounding the modified arginine. Therefore, the condensation properties observed for SMN$^{\text{Tud}}$ and *dm*Tudor$^{\text{Tud}}$ could be fortuitous and relatively unique; alternatively, the tendency to condense could be common to tudor domains. To test for commonality, I selected a panel of twelve domains from nine different human proteins involved in diverse biochemical processes (Figure 15, page 51). All have four conserved aromatic amino acids that make up the binding pocket for DMA (Figure 16, page 52). Of these domains, we found that six in addition to SMN$^{\text{Tud}}$ form condensates: Tdrd1$^{\text{Tud}}$#3, Tdrd1$^{\text{Tud}}$#4, Tdrd3$^{\text{Tud}}$, Tdrd6$^{\text{Tud}}$ #5, Tdrd8$^{\text{Tud}}$, and Snd1$^{\text{Tud}}$ (Figure 17, page 53). Expression levels between constructs are variable, so I do not rule out that other tudors might form condensates at higher concentrations.

I chose to interrogate one example of closely related tudor domains: those from SMN and the splicing factor Spf30. I compared the two using the clustering metric in live cells to assess whether expression level alone explains why SMN$^{\text{Tud}}$, but not Spf30$^{\text{Tud}}$, formed condensates. Spf30$^{\text{Tud}}$ did not condense across a range of concentrations in spite of sequence, structural and interactome similarity to SMN$^{\text{Tud}}$ (Figure 16, page 52; Figure 18, page 54)[162]. This suggests that the ability of any tudor domain to mediate condensation depends on the availability, expression level, and methylation status of suitable partner molecules.
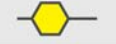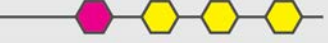
Figure 15. Diagrams of tudor domain containing proteins.

Schematic description of human tudor domain proteins tested for condensation, with tudor domains containing either an intact binding site for DMA (yellow), or lacking one or more of the four aromatic amino acids in the DMA binding pocket (pink). Domain architecture and function correspond to Uniprot annotations.

```
                        10        20        30        40        50        60
                     .         .         .         .         .         .
   Tdrd3  WKPGDECFALYWE--DNKFYRAEVEALHS-SGMTAVVKFIDYGNYEEVLLSNIKPIQTEAWEE-
   Spf30  WKVGDKCMAVWSE--DGQCYEAEIEEIDE-ENGTAAITFAGYGNAEVTPLLNLKP-VEEGRKAK
     SMN  WKVGDKCSAIWSE--DGCIYPATIASIDF-KRETCVVVYTGYGNREEQNLSDLLSPICEVANN-
Tdrd4 #4  FRTEMPCLAEYD---DGLWYRAKIVAIKEFNPLSILVQFVDYGSTAKLTLNRLCQIPSHLMRY-
  Tdrd9   PHPDLVCLAPFADFDKQRYFRAQVLYVS---GNSAEVFFVDYGNKSHVDLHLLMEIPCQFLEL-
Tdrd4 #1  PVQDQACVAKFE---DGIWYRAKVIGLPG--HQEVEVKYVDFGNTAKITIKDVRKIKDEFLNA-
    Sndl  PRRGEFCIAKFV---DGEWYRARVEKVES--PAKIHVFYIDYGNREVLPSTRLGTLSPAFSTR-
Tdrd6 #6  LQRGDMICAVFPE--DNLWYRAVIKEQQP--NDLLSVQFIDYGNVSVVHTNKIGRLDLVNAIL-
Tdrd6 #5  LNPGTLCLAKYT---DGNWYRGIVIEKE-----PKKVFFVDFGNIYVVTSDDLLPIPSDAYDV-
   STK31  LDPNKIYGGLFSE--DQCWYRCKVLKIIS--VEKCLVRYIDYGNTEILNRSDIVEIPLELQFS-
Tdrd1 #3  AEIGQPCCAFFAG--DGSWYRALVKEILP--NGHVKVHFVDYGNIEEVTADELRMISSTFLNL-
Tdrd1 #2  PAIGDICCAQFSE--DDQWYRASVLAYAS--EESVLVGYVDYGNFEILSLMRLCPIIPKLLEL-
Tdrd1 #4  PRIGDACCAKYTS--DDFWYRAVVLGTS---DTDVEVLYADYGNIETLPLCRVQPITSSHLAL-
```

Figure 16. Alignment of tudor domains with intact binding sites.

Amino acid sequences of tudor domains aligned using Clustal Omega[174]. Green highlights indicate aromatic residues in the DMA binding pocket as determined by structural studies.
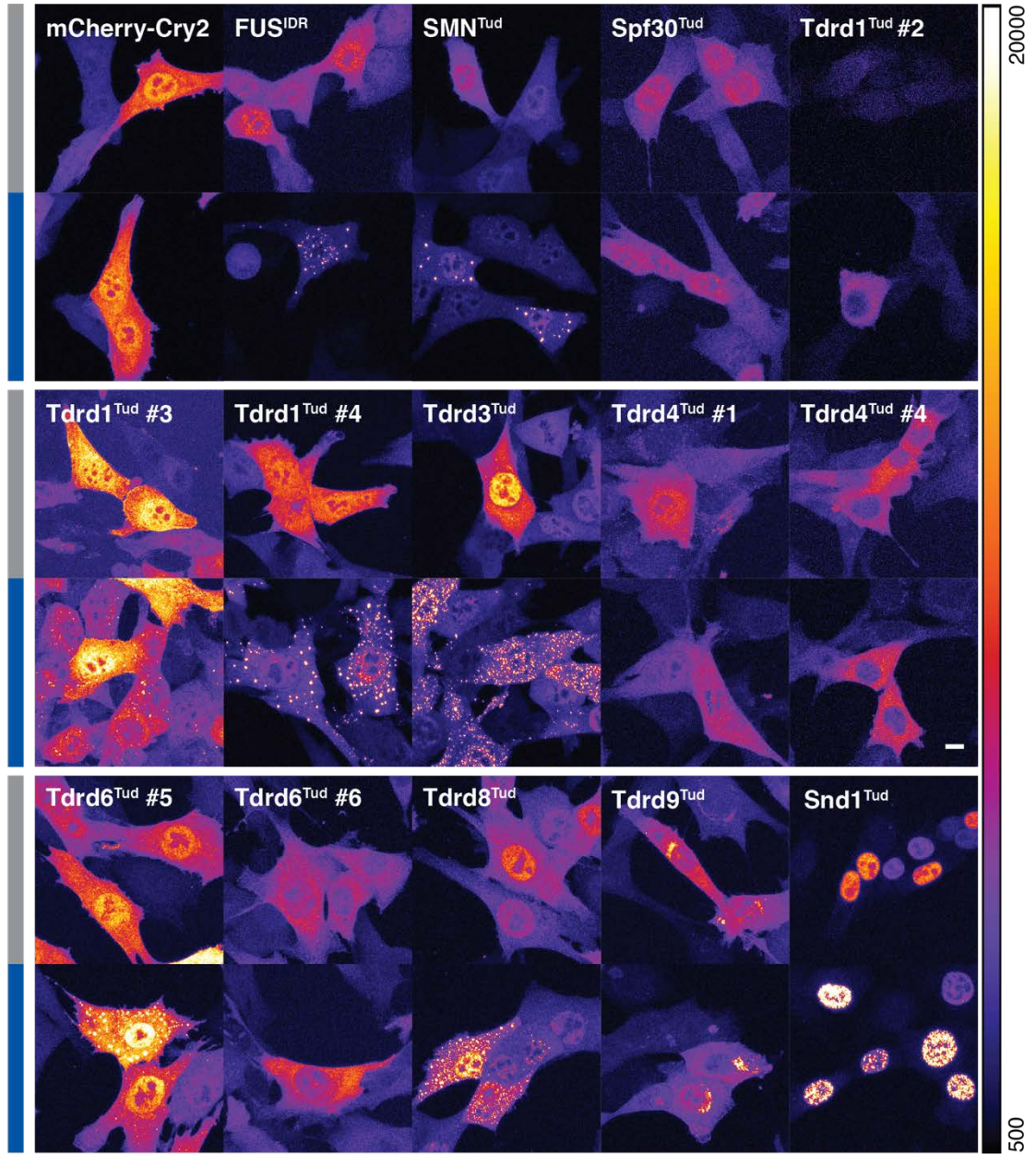
Figure 17. Condensation is shared by other human tudor domains.

Fixed cells expressing Tudor-Cry2 constructs under Cry2-inactive and Cry2-active conditions. Color bar given in analog-digital units. Scale bar = 10 μm.
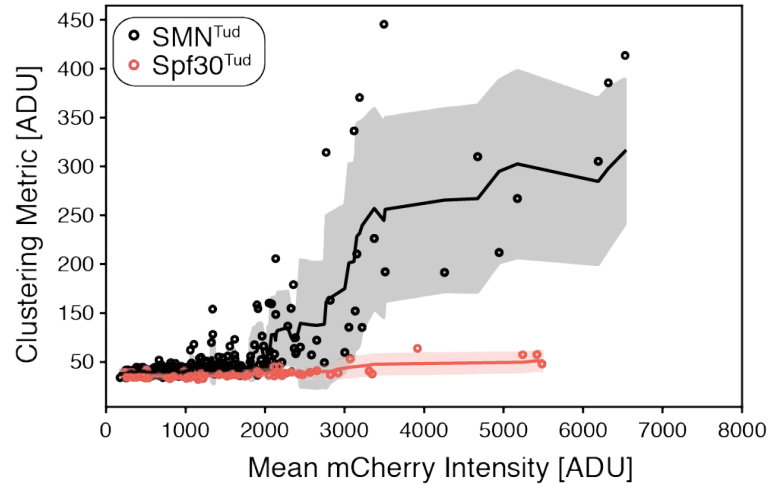
Figure 18. Spf30 does not condense across a range of concentrations.

Quantification of the condensation effect in live cells of SMN^{Tud} compared to Spf30^{Tud}. Units are analog-digital units (ADU). Solid lines and shading are a rolling mean and standard deviation of 10 points. Each point represents one cell. Andrew Barentine contributed the analysis methodology and programming used to generate the clustering metric and this plot.

2.3. Condensation of Coilin and Nopp140

Does coilin have condensation propensity?

As I have discussed, the disordered middle domain of coilin makes it a candidate for condensation. My findings with the SMN tudor domain underscore the importance of testing all regions of a protein. The coilin N-terminal domain is thought to be a self-interaction domain, making it a candidate for increasing multivalency[141]. The IDR region includes the nuclear localization signals (NLS), and a predicted nucleolar localization signal[141,175]. This region has high positive charge density, and this alone might be responsible for the nucleolar association sometimes observed[176-178]. Nevertheless, the coilin IDR remains a candidate. The CTD of coilin is a tudor-like domain, though it is unlikely that it binds DMA without an aromatic binding site[179]. Just prior to the CTD, there is an RG motif that is the target for DMA modification. This motif is the binding site for SMN and has been posited to be a regulatory region of Cajal body morphology[144,180,181]. I again used the Cry2 condensation assay to assess each of these regions.

I found that no region of coilin formed compelling condensates (Figure 19, page 57). The coilin NTD-Cry2 aggregated in cytoplasmic puncta before light activation. These puncta are highly mobile, but aside from rare instances of increased intensity during activation, seem to be independent of the oligomerization state of Cry2. These aggregates were present even if cells were kept in darkness for 24 hrs prior to imaging. Such aggregation is not uncommon for fluorescent protein fusions to either end of the NTD, and in other assays we have dismissed these aggregates as artefactual. The coilin IDR localizes to nuclei and upon Cry2 activation appears to concentrate in nucleoli; this was observed only for extremely high-expressing cells, as the coilin IDR does not express well in stable

cell lines and could be observed only for transiently transfected NIH-3T3 cells. The coilin CTD-Cry2, including the RG motif, is unresponsive to light activation.

I used a number of other constructs of combinations of different regions of coilin in the Cry2 assay to investigate any subtle effects of specific portions of the protein. No construct produced results differing from those shown in Figure 19. All constructs tested for coilin are summarized in Table 3 (page 65).
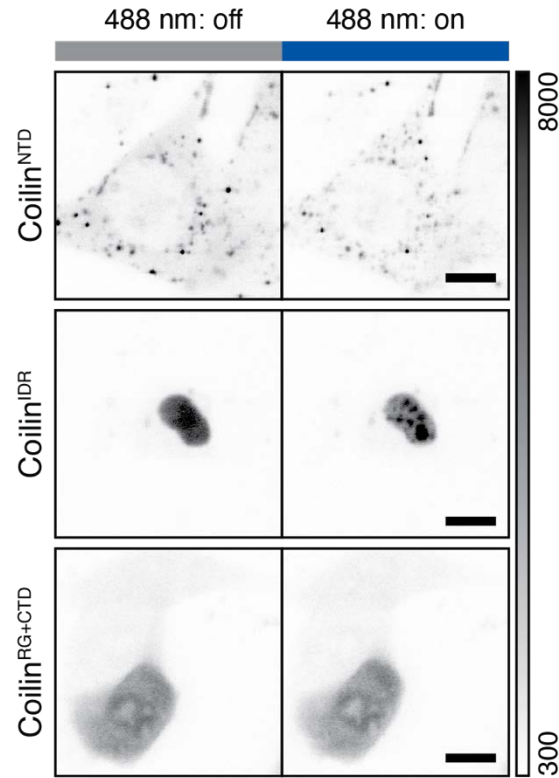
Figure 19. Coilin regions do not form condensates in response to Cry2 activation.

Micrographs of live cells expressing coilin regions fused to mCherry-Cry2 undergoing blue light activation of Cry2. Coilin^{NTD} and Coilin^{RG+CTD} were expressed from stable cell lines. The Coilin^{IDR} construct was transiently transfected due to instability of the construct. Grayscale bar given in analog-digital units. Scale bar = 10 μm.

Does Nopp140 have condensation propensity?

Simply by examining its sequence, Nopp140 appears to be an ideal candidate for forming a liquid condensate. Its LisH domain may have oligomerization properties[150]. Large blocks of charged amino acids connected by serine-rich regions in the Nopp140 IDR are similar to those of other condensate proteins[109,182]. The C-terminal region, predicted to be a NLS, has the highest density of positive charge, but the lysine tracts elsewhere in the protein make it unlikely that the protein is ever outside the nucleus for long. Intriguingly, numerous phosphorylation sites provide an opportunity for regulation[183].

Once again, I used the Cry2 method test for condensation of Nopp140 (Figure 20, page 59). As expected, the IDR appears to form condensates. It is not clear whether this effect is related to Nopp140[IDR] partitioning into the nucleolus in response to Cry2 activation. The LisH domain exhibits a unique response in that its overall localization remains constant, while it subtly and consistently enriches in what appears to be the dense fibrillar center of the nucleolus. Prior to light-induced oligomerization, the construct is mostly excluded from this compartment. The C-terminal region exhibits no substantive response to induction of Cry2.
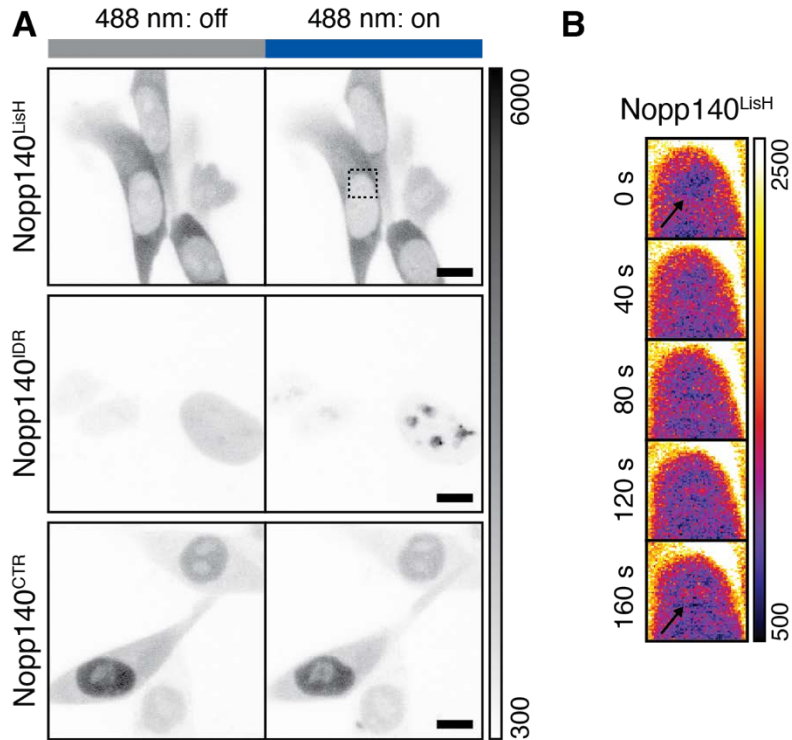
Figure 20. Nopp140 condensation is closely tied to the nucleolus.

A) Micrographs of live cells expressing Nopp140 regions fused to mCherry-Cry2 undergoing blue light activation of Cry2. Grayscale bar given in analog-digital units. Scale bar = 10 μm. B) Inset of Nopp140^{LisH} during activation showing an intensity increase in what is likely to be the dense fibrillar component of the nucleolus. Color bar given in analog-digital units.

2.4. Full length proteins

I carried out the Cry2 assay with full-length SMN, coilin and Nopp140 appended to mCherry-Cry2. The results of these experiments are provided in Table 3 (page 65). While these constructs do serve as a useful control to verify that the mCherry-Cry2 fusion has not altered the folding or localization of these proteins, I view the presence of condensation after Cry2 activation skeptically. If these proteins are incorporated into native complexes, inducing oligomerization via Cry2 activation ought to induce reorganization of those complexes and condensates, making this result mostly trivial. Any number of interacting proteins would contribute to the reorganization and apparent condensation, preventing us from making conclusions about the role of the full-length protein. Only by isolating individual protein regions can one attribute condensation activity to a particular region.

2.5. Discussion

Implications of SMN tudor domain condensation

My findings show that tudor domains are widely involved in biomolecular condensation and further indicate that the DMA post-translational modification triggers assembly when bound to specific ligands. How DMA modifying pathways act to regulate condensates *in vivo* is unexplored. DMA modifications have been found for a vast and growing list of proteins including histones, RNA polymerase II, and G3BP1[78,170,184]. These proteins are associated with chromatin, transcription, and stress granules, respectively. These systems are all reported to be associated with condensates *in vivo,* raising the possibility that tudor domain proteins and DMA play regulatory roles in their assembly and/or disassembly[54,107,114].

60

I propose that the DMA•tudor interaction module represents a means by which cells can form a multivalent interaction network leading to condensation. This module is specific and can be regulated because it depends on the modification of arginine; yet it is promiscuous in the sense that a single tudor domain can bind multiple different proteins containing DMA. Likewise, DMA has been shown to modulate the condensation of IDRs[185]. One demethylase for aDMA has been identified, making the modification reversible and potentially dynamic[186]. Thus, DMA may be a post-translational modification similar to phosphorylation that can dynamically modify condensates in response to cellular signaling[36]. Regulated assembly via modification highlights the potential of multimerized tudor domains to act as drivers of biomolecular condensation in membraneless organelles.

Interpreting the nucleolar relationship of Nopp140 and coilin

IDRs from both Nopp140 and coilin either recruit to, or form new condensates associated with the nucleolus. Further validation using immunofluorescence will be a necessary next step to illuminate the relationship between Nopp140 IDR condensates and preexisting nucleoli. Nopp140 is a canonical nucleolar protein and coilin has a known relationship to this organelle[148,177,187]. It may be impossible to entirely uncouple these proteins from the preexisting organelle to draw conclusions from *in vivo* assays like the Cry2 platform. However, some insights are provided by the negative results with other regions.

Coilin and Nopp140 may turn out to share certain architectural principles. Both N-terminal domains promote self-interaction, albeit by different mechanisms, which I shall return to in Chapter 3[141,150]. The central IDRs could act as linkers with other functional domains, or act to promote condensation, though these are not mutually exclusive. Initially,

the Nopp140 IDR appeared to be the best candidate for condensation due to its low complexity sequence and prevalent blocks of alternating charged amino acids akin to those in FUS, LAF-1 and other proteins (Table 1, page 8)[59,61,188]. Thus, it is satisfying to see that induced multimerization results in the formation of droplets and recruitment to the nucleolus.

The apparent effects of recruitment to the nucleolus seen for several regions of these proteins are puzzling. The apparent partitioning into the nucleolar compartment dependent on multimerization by Cry2 further reinforces the essentiality of multivalency for the formation of these compartments[34,111]. That framework does not explain why the Nopp140 LisH domain is mostly excluded from the nucleolus before Cry2 activation, but then enters the central dense fibrillar compartment after induction. This would suggest that complexes must acquire elevated valency to enter the organelle, and is a concept that will require further study to fully understand.

Coilin folded domains

The coilin NTD and CTD did not exhibit any convincing condensation properties. The aggregation of the coilin NTD is consistent with its self-interaction properties (Figure 19, page 57)[141]. I do not interpret this to be a meaningful "condensate" formation, though semantically it is a liquid-to-solid precipitation. Experiments on the NTD in the nucleus could be illuminating and I will return to other properties of this domain in Chapter 3.

The CTD, including the RG motif, is thought to be where coilin interacts with SMN and snRNPs[144,180,189]. Based on what we learned from studying SMN, it would not have been surprising if the CTD formed condensates, but their absence is consistent with the CTD's relationship to snRNPs. The Cajal body must dynamically recruit and recycle

snRNPs, and thus tight binding could be unproductive to that function. SnRNPs are not part of SMN tudor condensates, and thus it seems reasonable that the region of coilin that binds them also would not support condensation (Figure 12, page 46).

Consequences for assembly of the Cajal body and specificity in condensation

Taken together, these findings raise two points about how Cajal bodies may assemble, with implications for other bodies as well. First, condensation of SMN's tudor domain with other proteins, including coilin, represents a mechanism for specificity control by post-translational modification. SMN also associates with U-bodies and gems, each of which are RNP containing bodies[142,158,159]. It is thus conceivable that recognition of DMA by SMN acts to recruit the SMN complex to specific substrates to form each of those bodies. The substrate would then be regulated by the methylation of its arginine, and subsequently bound by SMN. In the case of the Cajal body, coilin is the substrate. The previous studies that have examined methylation typically assume that methylation of snRNP Sm proteins is most important[144,159]. My findings suggest that methylation of other targets could be crucial to the structural integrity of these membraneless organelles.

Second, the promiscuous nature of proteins that co-condense with, or partition into the nucleolus further reinforces the close relationship between the nucleolus and the Cajal body. A key feature of our assay is that NIH-3T3 cells do not often have Cajal bodies. Cells lacking nucleoli also lack ribosomal RNA synthesis and have drastic perturbations in physiology[11,22]. Thus, the Cry2 assay may fall short of providing definitive evidence for or against condensation with proteins associated with the nucleolus. Coilin and Nopp140 have that clear functional and structural relationship to the nucleolus, complicating some interpretations[136,177,178]. One possible outcome of this study is that the nucleolus acts as a

sink for many proteins that form condensates. A specificity-oriented interaction like DMA•tudor may promote Cajal body formation by extracting some of these condensed proteins from the nucleolus into a separate body.

Table 3. Results of Cajal body protein Cry2 results.

Construct	Amino acids	Localization	Description of Cry2 activation result
SMN^{FL}	1-294	Cytoplasmic, some nuclear	Small droplets visible in cytoplasm, and occasionally in nucleus.
SMN^{IDR-N}	1-91	Mostly cytoplasmic	No effect
SMN^{Tud}	92-145	Whole cell	Droplets throughout cell
SMN^{IDR-C}	146-294	Cytoplasmic, some nuclear	No effect
Coilin^{FL}	1-576	Nuclear, forms Cajal bodies	More droplets form in nucleus
Coilin^{NTD}	1-97	Cytoplasmic aggregates	Aggregates remain, may take up some diffuse protein
Coilin^{IDR}*	98-405	Nuclear	Some accumulation in nucleolus
Coilin^{RG+CTD}	406-576	Whole cell	No effect
Coilin^{CTD}	460-576	Whole cell	No effect
Coilin$^{\Delta NTD}$*	98-576	Nuclear	Very little accumulation in nucleolus
Coilin$^{\Delta NLS}$	1-106…200-576	Cytoplasmic aggregates	No effect on aggregates or new droplets
Nopp140^{FL}	1-706	Nuclear, nucleolus	Forms new smaller droplets
Nopp140^{LisH}	1-42	Mostly cytoplasmic	Concentrates in dense fibrillar component of nucleolus
Nopp140^{IDR}	43-585	Nuclear	Forms large condensates, may be concentrated in nucleolus
Nopp140^{CTR}	586-706	Mostly nuclear	No effect

All constructs listed above are appended to mCherry-Cry2. * indicates the construct was not stably expressed and transient transfection was used rather than stable cell line generation. Localization is before Cry2 activation with blue light.

Chapter 3. Control of Coilin NTD oligomerization

3.1. Goals and approach

I now turn to the question of how coilin, the canonical Cajal body protein contributes to assembly. In section 2.2, I demonstrated that the tudor domain of SMN can mediate assembly, but SMN is present in multiple bodies and cannot account for all properties of the Cajal body. The Coilin NTD will be the focus of this chapter, as it is required for normal Cajal body formation, but we lack detailed information on its mechanism[141]. I will address this using a panel of point mutations, fluorescence microscopy and biochemical methods.

Outside contributions

The experiments in this chapter are related to an ongoing project involving many past and present members of the Neugebauer lab. The original hypothesis that motivated the study was put forward by Martin Machyna and he generated and initially characterized the mutants presented here, with the help of Sarah Srivichitranond. I designed the protocol for quantifying morphologies and Jade Enright carried out the manual cell counting. Korinna Straube assisted with sample preparation and molecular cloning. Their specific contributions are noted in the relevant figure legends.

3.2. Assemblages of the coilin NTD

Coilin is required for proper Cajal body formation but we do not fully understand what role it serves in assembly[115,116]. Nevertheless, without coilin and the snRNP assembly role of the Cajal body, viability of vertebrate embryos is compromised[115,116]. The coilin NTD appears to be required for Cajal body assembly, but it is unclear why that is the case[141].

Many reports have shown that the NTD has an array of binding partners, and some coilin binding proteins have yet to be mapped to a specific region of the protein[190]. RNA, Tgs1, and WRAP53 are just a handful of molecules thought to interact with coilin in some way[68,123,191]. Nopp140, the product of the Nolc1 gene, is one such protein believed to interact with the NTD[148]. A recent high-throughput screen identified Nopp140 depletion as a down-regulator of Cajal body integrity. The association of other molecules and essentiality of the coilin NTD provides the motivation to interrogate its role and understand how it might control overall assembly[192].

Morphologies of the coilin NTD and its effects on the Cajal body

The coilin NTD is predicted to be structured, but no atomic model has been solved by any method due to its insolubility (Figure 1, page 28)[193]. Using the RaptorX online server, I generated a predicted three-dimensional atomic structure for the coilin NTD[152]. The NTD takes on an apparent ubiquitin-like fold (Figure 21A, page 69). This predicted structure model serves as a starting point for hypothesizing what specific interactions might be occurring to form a Cajal body.

The coilin NTD is highly conserved[190]. In fact, the human and mouse coilin NTD are identical to one another for nearly 90 amino acids (Figure 21B, page 69). When the NTD is expressed in the cytoplasm of mouse embryonic fibroblasts (MEF) derived from *COIL* knockout mice, it forms fibrillar structures (Figure 22, page 70)[116]. The same construct targeted to the nucleus with the SV40 NLS forms spherical bodies. Point mutations in the NTD at residues chosen for their high conservation in metazoans result in a range of morphologies (Figure 22, page 70). Effects of mutants on cytoplasmic and nuclear morphologies correspond to each other, including the dispersal of both fibrils and

bodies as with R8A and D79A. Remarkably, R36A appears to stabilize and increase the persistence length of fibrils in the cytoplasm, and induce a fibrillar morphology in the nucleus.

We asked whether these morphologies had any bearing on the assembly of Cajal bodies in cells expressing coilin. When expressed in HeLa cells, certain NTD mutants exhibited dominant negative effects on Cajal bodies (Figure 23, page 71). In particular, R8A and D79A disrupt Cajal body formation. Mutating both residues restores the normal number of Cajal bodies in HeLa cells. R36A induced the fibrillar morphology for endogenous full-length coilin in the nucleus. From this I conclude that the NTD is interacting with endogenous coilin and that mutated residues form at least two binding interfaces. The double mutation would then disrupt both interfaces, allowing the restoration of normal Cajal bodies as with R8A/D79A.
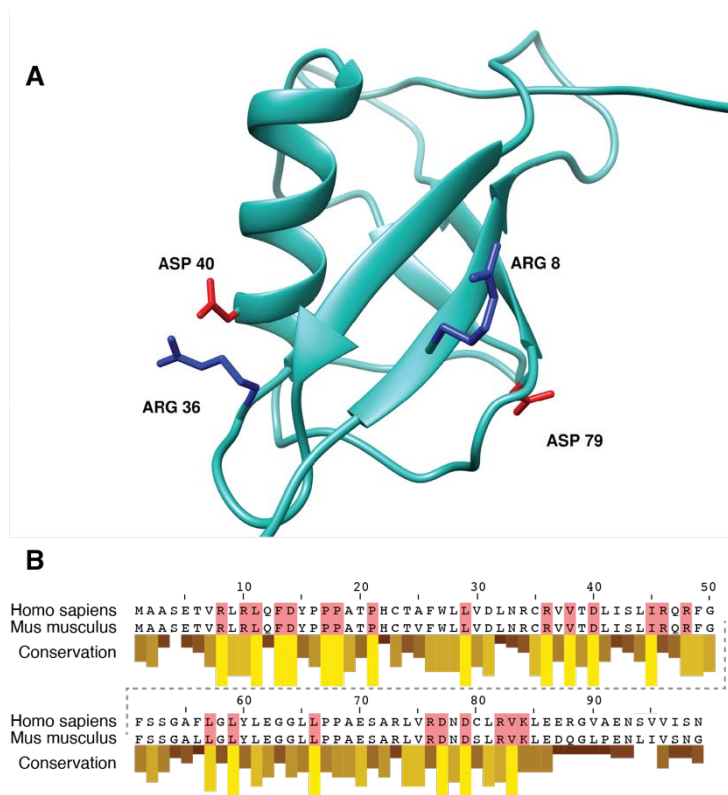
Figure 21. Predicted structure of the coilin NTD.

A) The predicted structure of the NTD resembles a ubiquitin-like fold. Four conserved residues are highlighted, positively charged Arg8 and Arg36 in blue, and negatively charged Asp40 and Asp79 in red. Structure generated using the RaptorX platform[152]. B) Alignment of human and mouse Coilin NTD using Clustal Omega[174]. Conservation score is taken from a multiple alignment of all metazoan coilin sequences. Red residues are conserved residues tested via mutagenesis. Martin Machyna carried out the initial alignment that B) is derived from.
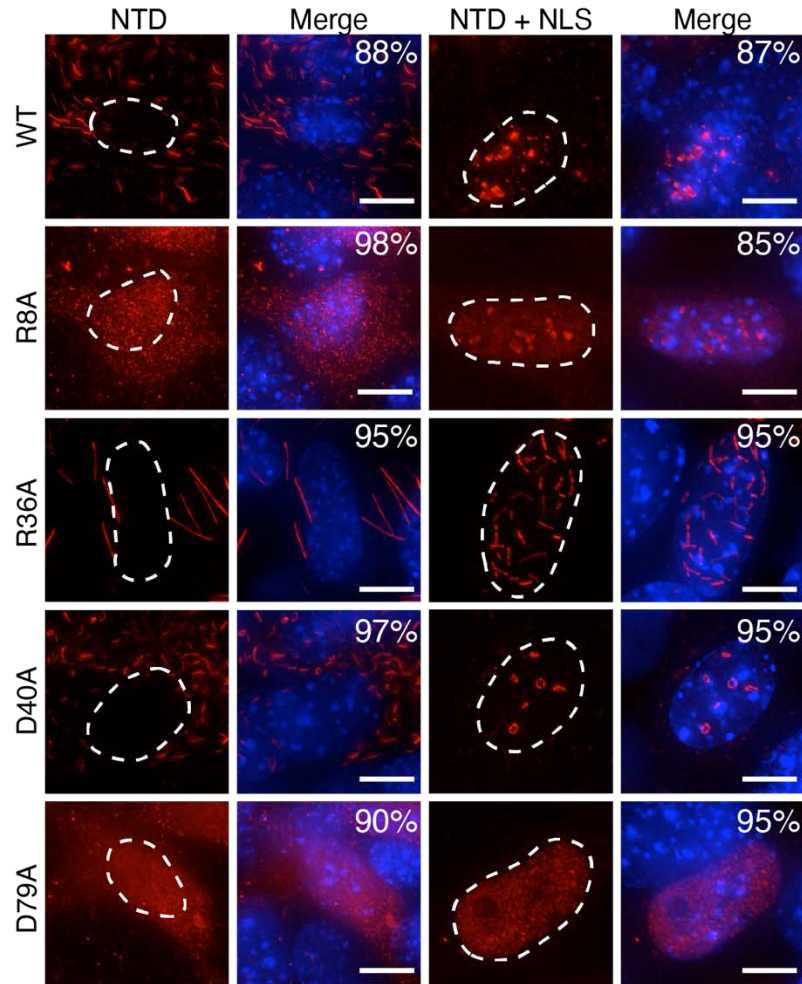
Figure 22. Morphology of the coilin NTD in Coil$^{-/-}$ MEFs.

The coilin NTD expressed in coil$^{-/-}$ mouse embryonic fibroblasts (MEF) both with and without a nuclear localization signal (NLS). Blue channel in merge indicates Hoechst staining. Percentage indicates the fraction of cells with the displayed morphology (n = 300). Scale bar = 10 μm. Martin Machyna and Sarah Srivichitranond made initial mutants and carried out the initial experiments. Jade Enright carried out manual counting of cells.
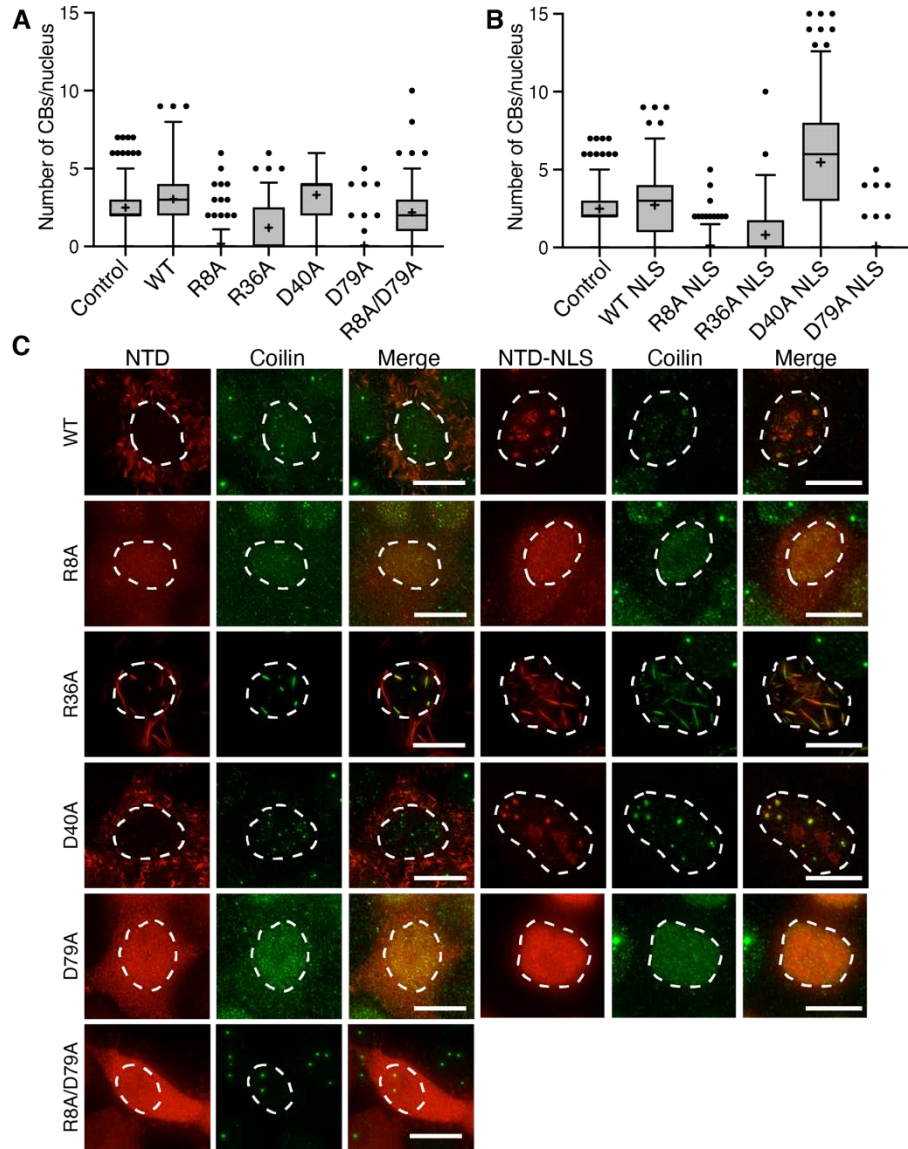
Figure 23. The coilin NTD imposes dominant negative effects on Cajal bodies.

A&B) Cajal body (CB) counts per nucleus in cells expressing coilin NTD in A) cytoplasm and B) nucleus. Median is a solid line, mean is the cross, whiskers indicate 95th percentile, n = 100. C) Representative images of data shown in A and B. Nucleus down as dotted line. Scale bar = 10 μm. Martin Machyna and Sarah Srivichitranond made initial mutants and carried out the initial experiments. Jade Enright carried out manual counting of cells.

3.3.   Interactions of the coilin NTD

After considering these data, I chose to revisit the molecular contacts reported to be associated with the coilin NTD. As discussed, the NTD purportedly forms a self-interaction[141]. From the experiments discussed in section 3.2, I surmise that coilin NTD has at least two binding sites for coilin. The evidence for Nopp140 binding is intriguing, making it a priority to see if that interaction could explain the morphologies evident in the mutant screen[148].

Properties of the coilin self-interaction

There were several possibilities that needed to be addressed regarding the nature of the coilin•coilin interaction. First, I asked whether the fibrillar morphology of coilin NTD wild-type has an underlying amyloid-like cross-β interaction as seen with some other membraneless organelles and condensates[31,48,65]. I chose to use three dyes that stain amyloid aggregates, two that are specific to cross-β (Thioflavin T and Congo Red[194]), and one that stains a wider range of protein aggregates (Proteostat). Coilin NTD fibrils in HeLa cells were negative for all three stains, inconsistent with the hypothesis that amyloid-like interactions underlie the interactions of the coilin NTD (Figure 24, page 74). Viewing these fibrils by electron microscopy shows that they are protein dense, but lack the ultrastructure characteristics of amyloid aggregates (Figure 25, page 75). Thus, I rule out that coilin NTD fibrils form via an amyloid-like mechanism.

I attempted to use *in vitro* methods to study the NTD interaction further. Biophysical methods like fluorescence anisotropy, circular dichroism or single molecule total internal reflection microscopy could provide direct measurements of the NTD self-interaction. Knowing the stoichiometry and kinetics of the self-interaction would provide

a baseline for comparing the effects of other molecules like Nopp140 or RNA. However, a key control experiment showed that these methods would not provide the information that I sought. Immunoprecipitation from insect cell lysates with GFP labeled and unlabeled full-length coilin showed that the coilin interaction forms only between molecules expressed in the same cells. Lysates where labeled and unlabeled coilin are mixed does not show co-precipitation indicating that new coilin•coilin interactions are not forming. From this I conclude that the coilin•coilin interaction is either A) extremely stable and slow to dissociate, or B) requires an active process to form *in vivo*. Regardless, *in vitro* biophysical methods are likely to fail without more information in either case.
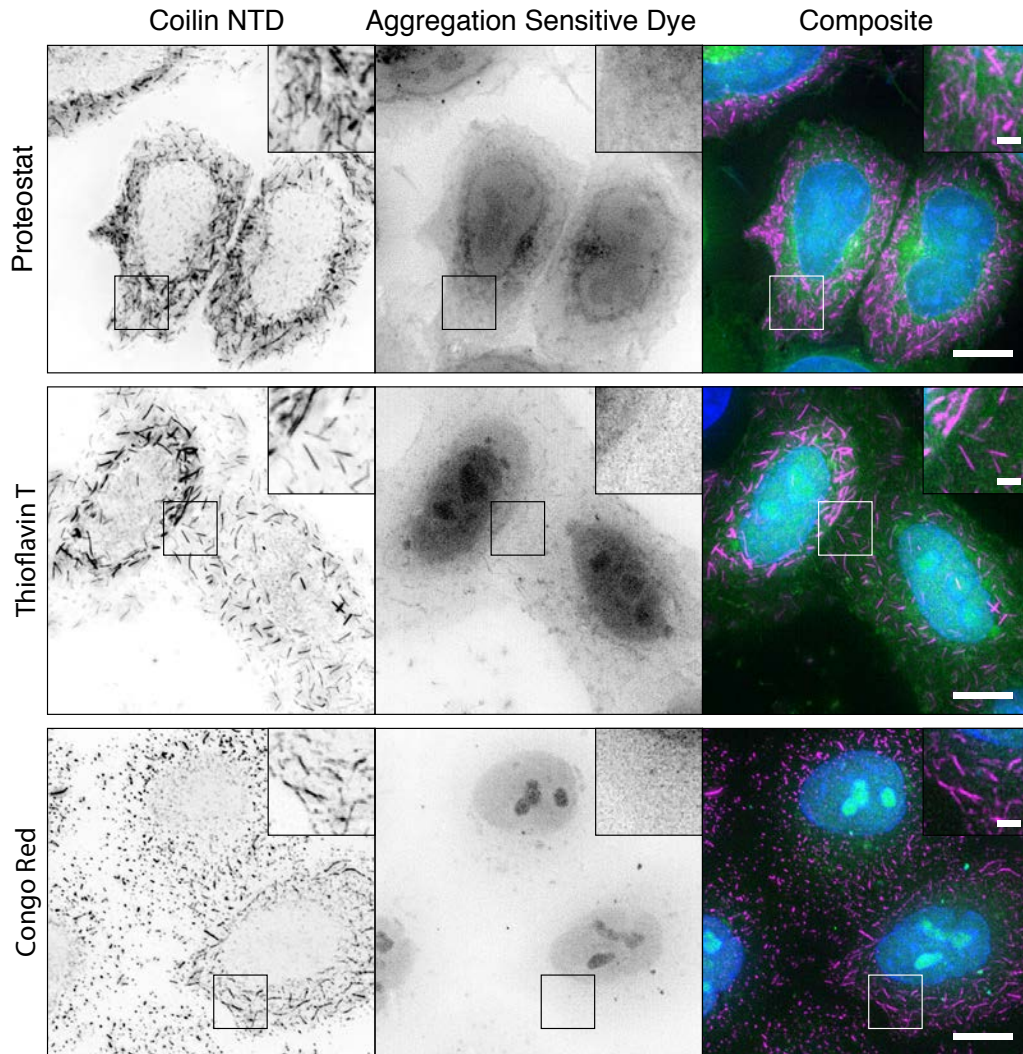
Figure 24. Amyloid dye staining of the coilin NTD.

HeLa cells expressing the coilin NTD, which forms fibrils in the cytoplasm. Proteostat is a generic protein aggregation stain. Thioflavin T and Congo Red are specific stains for amyloid. In the composite image, magenta indicates coilin NTD staining, green indicates the respective amyloid dye. Nuclear Hoechst stain is in blue. Nucleoli stain with these dyes because of the fibrillar amyloid like interactions in their interior. Scale bar = 10 μm, inset scale bar = 2 μm.
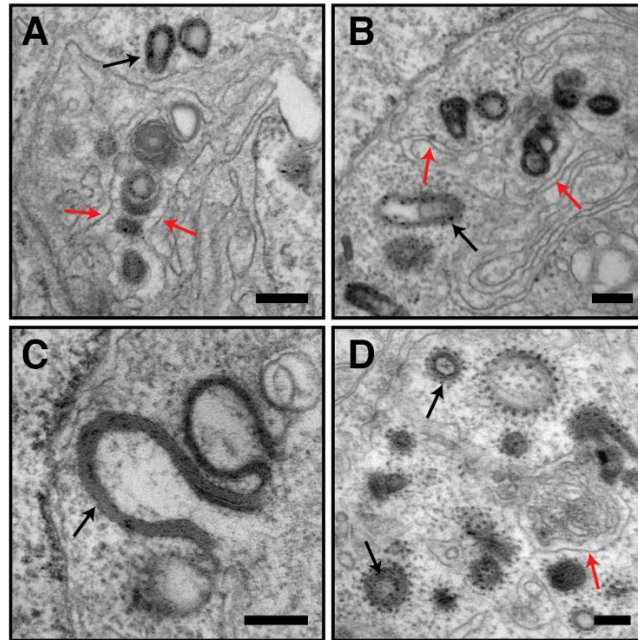
Figure 25. Electron micrographs of coilin NTD fibrils.

Electron micrographs of HeLa cells expressing coilin NTD for 8 hrs (A&B), 16 hrs (C), or 24 hrs (D). Black arrows indicate protein dense objects inferred to be fibrils. Red arrows indicate multilamellar autophagy membranes engulfing fibrils. Scale bar = 200 nm.
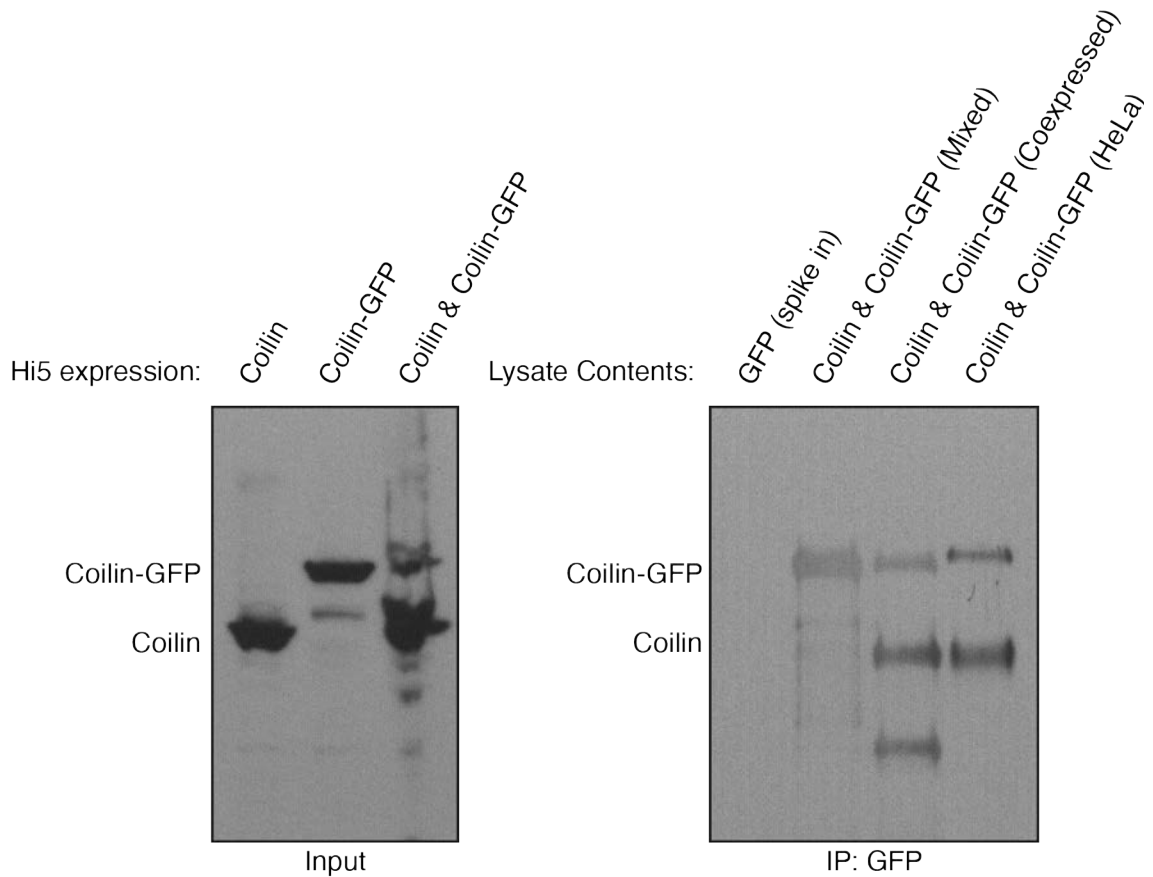
Figure 26. Coilin does not form new interactions with itself in lysate.

Immunoprecipitation (IP) of coilin with coilin-GFP. Recombinant zebrafish coilin was expressed in Hi5 insect cells using a baculovirus system (left, inputs). IP with recombinant GFP spiked in to lysate was a negative control. If bait (coilin) and prey (coilin-GPF) are expressed separately and mixed after lysis, no IP is evident. If bait and prey are coexpressed, IP is observed in both Hi5 cells and in HeLa cells stably expressing coilin-GFP.

Point mutation effects on coilin NTD interactions

Having ruled out amyloid-like structure in NTD fibrils and arriving at an impasse with biophysical methods, I chose to use a biochemical approach to assess the coilin•coilin and coilin•Nopp140 interactions. The panel of mutants described in section 3.2 provides a useful tool for interrogating these two molecular contacts. Co-immunoprecipitation (Co-IP) and Förster resonance energy transfer (FRET) provide two measures of interaction, each with its own advantages and drawbacks. Co-IP provides a readout of interaction stability in lysate, but may detect indirect contacts. FRET measures interaction distances between fluorophores under 10 nm, providing more information of how direct contacts may be, but the method is sensitive to artifacts induced by fluorescent protein tagging[195,196].

In Co-IP experiments, the mutations R8A and D79A reduce but do not entirely eliminate the coilin self-interaction (Figure 27, page 78). The double mutation R8A/D79A eliminates coilin binding, supporting my interpretation of the dominant negative effect on Cajal bodies of single mutants and rescue of the double mutant (Figure 23, page 71). Nopp140 interaction is abolished by either R8A or R36A, and is substantially reduced for D79A (Figure 27, page 78). These observations are supported by FRET measurements: R8A reduces binding to coilin and Nopp140, R36A abolishes Nopp140 interaction, and D79A has a reduced effect on the interaction with either molecule (Figure 28, page 79). Notably, R36A preserves coilin interaction and is the mutation that increases the fibrillar nature of the NTD expressed alone.
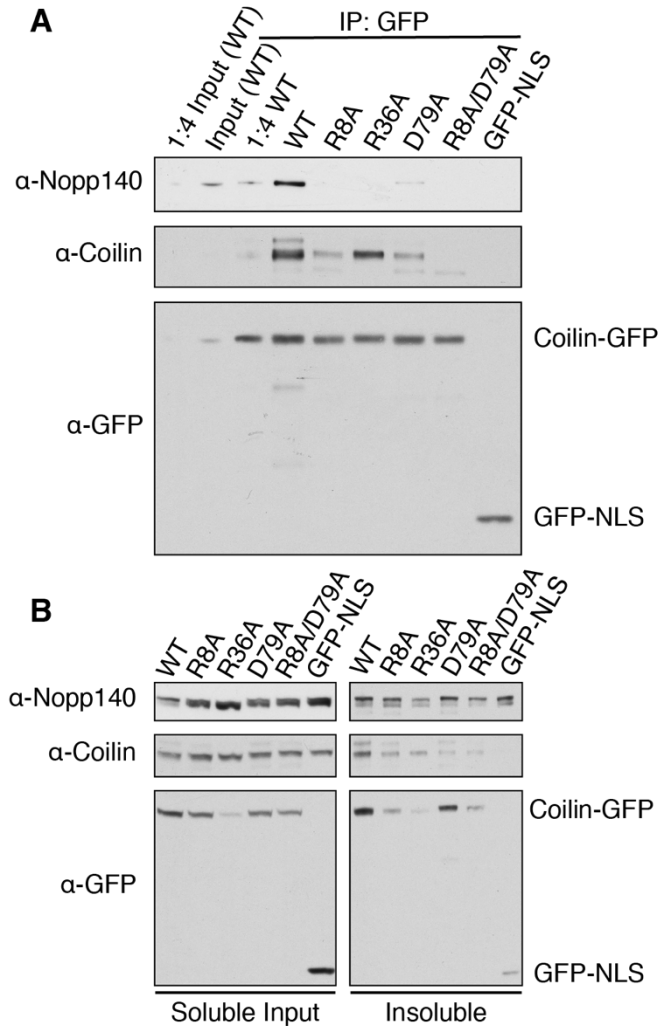
Figure 27. Coilin NTD mutations disrupt co-IP with coilin and Nopp140

Co-IP of coilin and Nopp140 from HeLa cells. The bait molecule is transiently transfected full-length coilin-GFP bearing the designated mutation. The prey molecule is endogenous coilin or Nopp140. A) IP of Nopp140 is disrupted by all three mutations. Coilin interaction is disrupted by R8A or D79A. Input and 1:4 diluted input lanes shown for comparing relative signals. B) Input lysate and insoluble fraction after cell lysis.
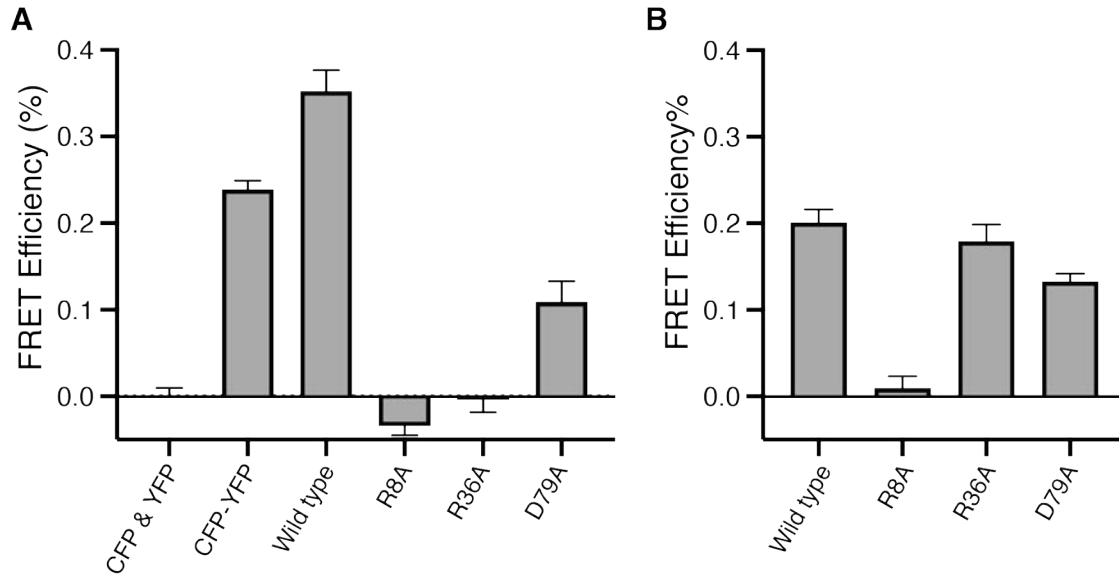
Figure 28. FRET reveals disruptions in NTD binding properties.

A) FRET measurements between full-length coilin and Nopp140. The donor molecule is CFP-Coilin and acceptor is Nopp140-YFP. B) FRET measurements between CFP-coilin and YFP-coilin. FRET efficiency is measured with the acceptor photobleaching method. CFP and YFP simultaneous expression provides a negative control while CFP-YFP fusion protein is a positive control. All measurements taken in transient transfections of HeLa cells. Error bars represent standard error of the mean. Martin Machyna performed the measurements for coilin interaction (B).

3.4.   Nopp140 as a regulator of coilin oligomerization

As the effects of mutations on interactions between coilin and Nopp140 became clear, I reasoned that these could have bearing on the morphologies of the NTD. I therefore asked whether Nopp140 is required for Cajal body formation, and if its absence influences the assembly properties of the coilin NTD. A high-throughput study of nuclear body regulators identified Nopp140 as a possible required factor for Cajal body assembly, but a recent conflicting report on Nopp140 CRISPR knockouts suggests it is not essential[192,197]. I set out to resolve these conflicting reports and incorporate Nopp140 into a model of Cajal body assembly.

Consequences of Nopp140 siRNA knockdown

Nopp140 levels were reduced below the detection limit of western blot by siRNA silencing of its gene, Nolc1, and the effect on Cajal bodies was judged by SMN and coilin immunostaining (Figure 29A, page 82). Nopp140 depletion causes Cajal bodies to disperse, but some gems remain as seen by SMN staining. Coilin knockdown was done in parallel as a positive control. Comparison of the levels of nucleoplasmic coilin between coilin knockdown and Nopp140 knockdown shows that coilin redistributes to the nucleoplasm with reduced levels of Nopp140 (Figure 29B, page 82). SMN puncta are still present in many cells, but rarely associate with residual coilin.

Quantification of the number of coilin and SMN puncta in each knockdown revealed that Nopp140 reduction also reduced the number of coilin puncta and SMN puncta (Figure 30, page 83). Coilin knockdown reduces the number of coilin puncta, as expected, but does provide a useful control for transfection efficiency of the siRNA. The number of SMN puncta in the coilin knockdown does not significantly change, but the variability from

cell-to-cell increases notably. Nopp140 knockdown reduces the number of Cajal bodies by a factor of 1.8, as judged by both SMN and coilin as markers. Further reduction may be limited by transfection efficiency of the silencing oligos. I thus conclude that Nopp140 depletion compromises the integrity of Cajal bodies.

Full-length coilin, even in the absence of Nopp140, may still interact with SMN and other proteins known to bind the coilin CTD. Full-length coilin NTD interactions may be buffered by binding activity elsewhere on the molecule. As shown above, the wild-type coilin NTD makes fibrillar assemblies in the cytoplasm of cells, but forms spherical structures in the nucleus. I hypothesized that the switch between these morphologies is due to the binding of Nopp140, and thus Nopp140 knockdown will allow coilin NTD fibrils to form in the nucleus. Indeed, depleting Nopp140 allowed the coilin NTD to form fibrils that appear to be nucleated in the nucleolus (Figure 31, page 84). In the non-targeting control transfected cells, the coilin NTD remains in spherical bodies that recruit Nopp140 in abundance.
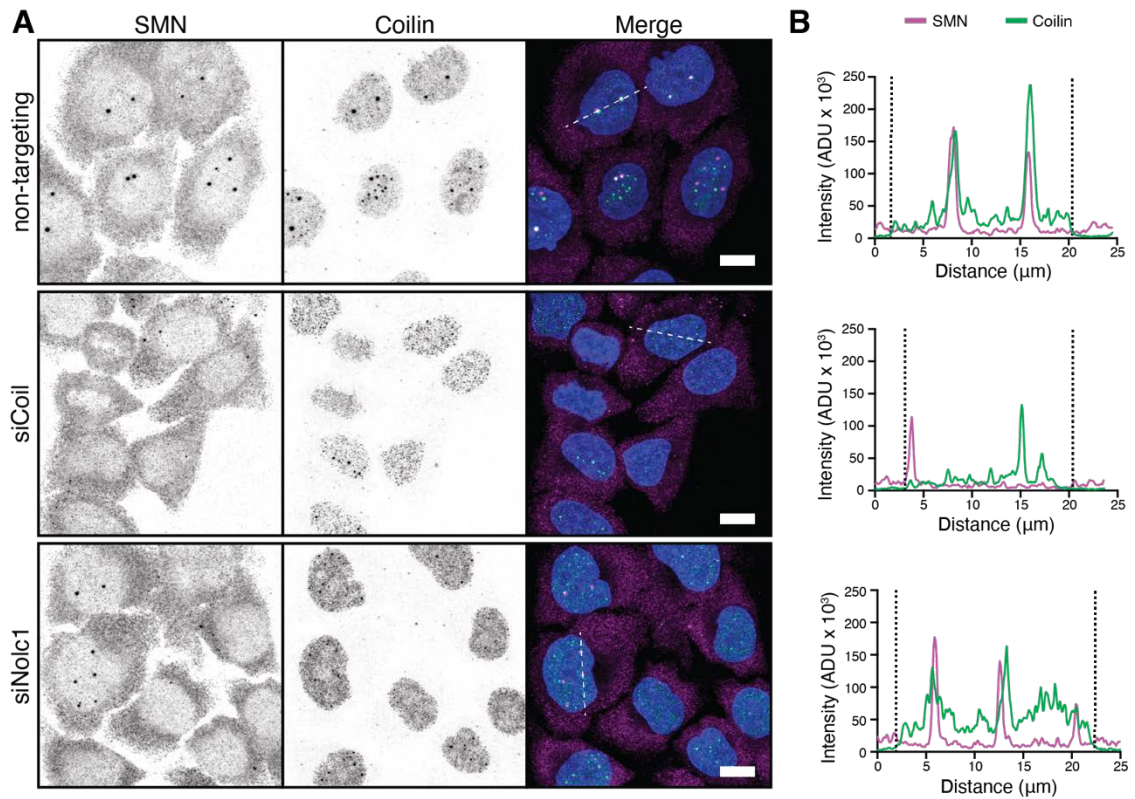
Figure 29. Reduced Nopp140 levels disperse coilin.

A) Micrographs of HeLa cells transfected with non-targeting, and silencing oligos for the *Coil* and *Nolc1* genes. Blue in merge is nuclear Hoechst staining. Scale bar = 10 μm. B) Profile plots through puncta as indicated by dashed lines (A). Dotted borders indicate the limits of the nucleus. Intensities are calculated from summed slices of z-stacks, where units are given in analog-digital units (ADU).
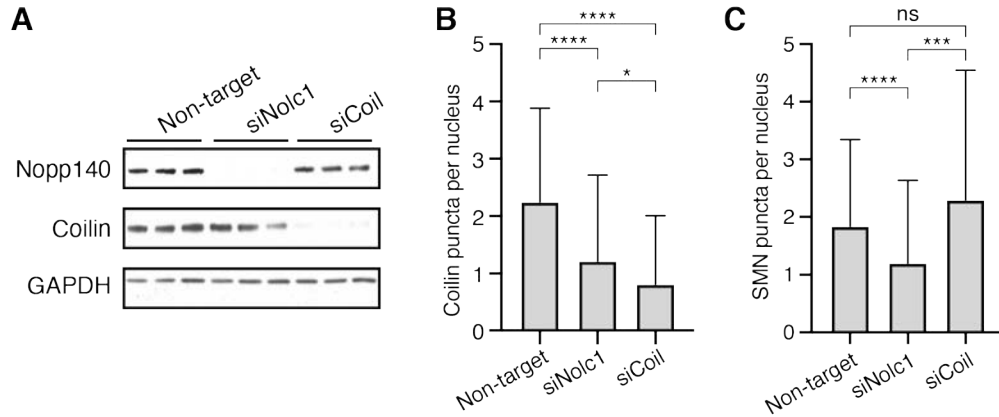
Figure 30. Nopp140 knockdown reduces the number of Cajal bodies.

A) Western blots of three biological replicates of knockdown in HeLa cells. Nopp140 levels were reduced to 0% (below the detection limit). Coilin levels were reduced to 3% compared to the non-targeting oligo. B) Mean and standard deviation for the number of coilin puncta in nuclei for each condition. C) Mean and standard deviation for the number of SMN puncta in nuclei for each condition. For each condition, 30-50 cells were counted for each of three biological replicates and then pooled. (*) indicates $p < 0.05$, (***) indicates $p < 0.001$, and (****) indicates $p < 0.0001$ for Mann Whitney test.
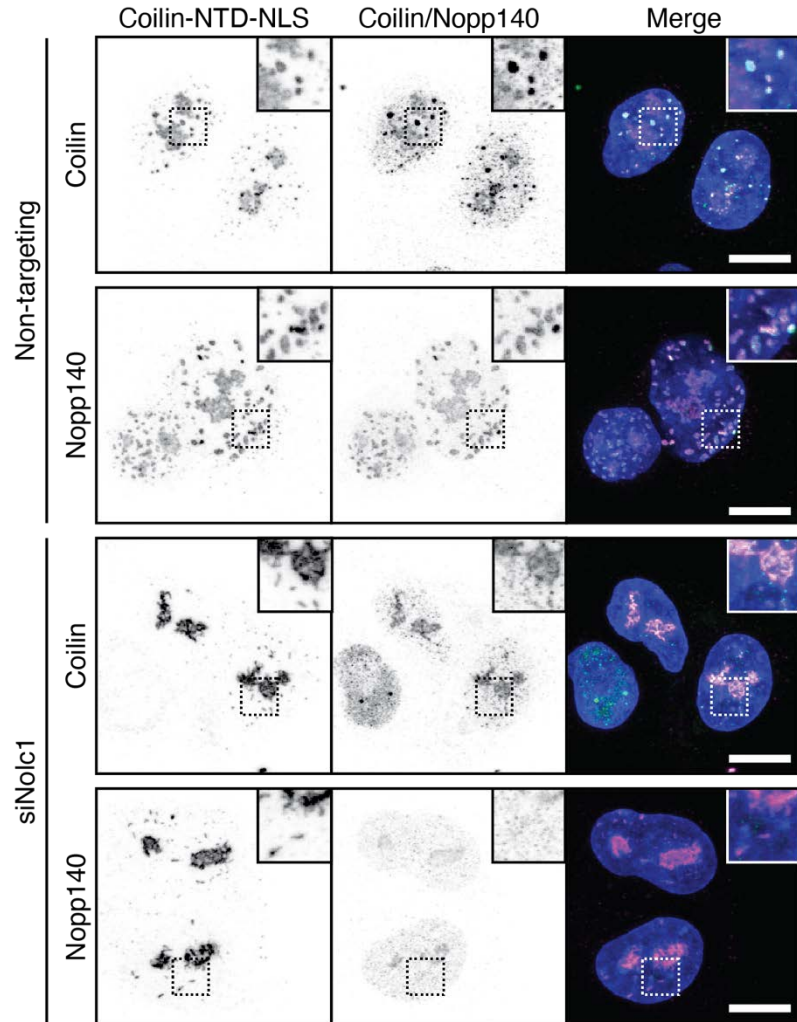
Figure 31. Nopp140 prevents the formation of coilin NTD fibrils.

HeLa cells transfected with silencing oligos that are either non-targeting or Nolc1, the gene encoding Nopp140. Scale bar = 10 μm. Inset is a 2x magnification of the area marked by the dashed line.

3.5. Discussion

Effects of point mutations of the coilin NTD

The coilin NTD is a structured domain that has been documented to fill a self-interaction role required to form the Cajal body[141]. Despite this, very little has been learned about the NTD's mechanistic role. The excitement around condensation has provided new motivation to study aggregation-prone proteins like coilin. I have ruled out an amyloid-like mechanism and the Coilin NTD is not unstructured, forcing me to discard certain models for membraneless organelle formation[31,62,188].

Mutagenesis of the coilin NTD raises questions about how key residues examined in this study might be influencing assembly. R36A appears to stabilize the fibrillar morphology of the coilin NTD in both the nucleus and the cytoplasm; whereas with wild-type, fibrils are only seen in the cytoplasm. Three possibilities arise from this result: 1) the R36A mutation substantially strengthens the coilin interaction with itself or some other molecule responsible for fibril formation, 2) the R36A mutation weakens an interaction that would otherwise prevent fibril formation and this interaction occurs only in the nucleus, or 3) a combination of both scenarios. In interaction assays, coilin interaction is preserved by R36A, but there is no evidence that this mutation makes the coilin interaction stronger. Nopp140 is a known coilin NTD interactor, and R36A disrupts that contact[148]. R36A also induces the fibrilization of endogenous Cajal bodies in HeLa cells, and thus the third scenario seems most likely.

Two other mutations reduce coilin NTD assembly in both the cytoplasm and the nucleus: R8A and D79A. HeLa cells expressing the coilin NTD with only one of these mutations exhibit dominant negative effects in their Cajal bodies, causing them to

dissociate. The effect is eliminated if the double mutant R8A/D79A is transfected. This suggests that these mutations render one interaction interface inactive and unable to participate in productive assembly, yet still the remaining interface occupies enough binding sites to prevent native coilin molecules from forming a Cajal body. My interaction assays support this model. R8A and D79A partially eliminate coilin interaction, and the predicted structure places these residues on separate faces, suggesting a possible explanation of how coilin oligomerizes.

The role of Nopp140 in Cajal body assembly

As I have shown, Nopp140 is required for normal Cajal body assembly and its absence allows coilin NTD fibrils to form in the nucleus of HeLa cells. This result agrees with a high-throughput study of nuclear body regulators, and goes a step further by looking at other morphological consequences. A number of kinases are thought to act on both Nopp140 and coilin to regulate formation of the Cajal body[183,198,199]. A specific binding site for coilin on Nopp140 is not currently known.

The fibrillar nature of the coilin NTD when expressed in the cytoplasm of cells may reflect an underlying molecular architecture of the Cajal body. Full-length coilin is subject to normal regulatory mechanisms. However, when divorced from the rest of the protein, the coilin NTD can be regulated only by molecules that directly target it. If Nopp140 antagonizes the formation of fibrils, it is possible that the Cajal body forms via the merger of two competing mechanisms. Coilin may have a tendency to form oligomers, but these are limited in size by Nopp140 which prefers to form condensates. Together, these clusters could form the core particle of the Cajal body and explain the "coiled" appearance of the Cajal body in electron micrographs[200].

Chapter 4. Outlook and conclusions

Outlook

There are several follow up studies that could further shed light on these experiments. Regarding condensation, tudor domains clearly have an important capacity for the formation of condensates but the interactome of tudor domain proteins remains understudied[155,172]. The low affinity of tudor domains for DMA substrates is essential to their dynamicity but poses a daunting challenge for proteomic study[162]. Determining the relevant protein ligands of tudor domains would aid in the deduction of precise regulatory pathways for condensates depending on tudor•DMA. Characterizing ligands as scaffolds or clients of tudor condensates would further shed light on the functional significance of certain condensation events.

I have characterized the effects of point mutants in the coilin NTD on its interacting partners, but little is known about Nopp140. No binding site for coilin is known and the question of where RNA or chromatin enter into the picture is an open question. Several approaches have shown that coilin is associated with certain genes and RNAs and it may even bind RNA directly, but the data presented here cannot address outstanding questions related to these possibilities[68,117]. A structural approach that validates the predicted model presented here may prove fruitful for working out the precise arrangement of coilin and Nopp140, and for determining if direct nucleic acid binding is involved with either molecule.

Conclusions

This work has focused on addressing the question of how cells separate their contents into functional domains in nuclei without a barrier dividing the "body" from the

bulk nucleoplasm. The Cajal body has been the system of choice for these experiments because it is well studied in terms of its function and constituent molecules[4,190]. Indeed, it has been studied for over 100 years, first observed by silver staining in neurons[8]. Yet specifically how these structures form has remained largely unknown.

My approach has been in two parts. First, I sought to apply the new framework of biomolecular condensate formation to the Cajal body, and to critique its applicability for this particular system. Second, I addressed several outstanding questions regarding the specific molecules residing in the Cajal body. These questions mainly dealt with the conventional molecular contacts between proteins, and aimed to determine which protein components are required for the Cajal body and why they are necessary.

Regarding condensation, I found that tudor domains are sufficient for condensate formation in live cells when they bind a post-translational modification, DMA. This is a new mechanism of condensation that is regulatable and has implications for a range of condensates including but not limited to transcriptional units, Cajal bodies, and stress granules[114,170,201]. In the Cajal body specifically, this condensate mechanism is responsible for uniting coilin and SMN in the processing of snRNPs[145,202].

SMN is not the only Cajal body protein with condensation properties. Nopp140 also makes nuclear condensates and is far more similar in sequence and structure to the intrinsically disordered domains that have become primary candidates for condensation in many cases[109,188]. A major implication of my work is that in the Cajal body, several different mechanisms are at work. They include multiple modes of condensation. SMN's tudor domain would impart a layer of specificity by recognizing only molecules that have methylated arginine in a sequence that it recognizes. Nopp140, being charged and

disordered would impart the non-specific high avidity dynamic interaction seen in other organelles with domains of similar sequence[54,61].

An additional specific interaction is involved in the assembly of the Cajal body that bridges SMN and Nopp140. The coilin NTD is an oligomerization domain that is buffered by its interaction with Nopp140. Near its CTD, coilin is methylated and bound by the SMN tudor domain. It is sensitive to point mutation, suggesting a high degree of specificity not seen in degenerate low-complexity sequences like the IDRs of FUS or hnRNP-A1[39,54]. Depletion of Nopp140 results in disassembly of the Cajal body, adding it to the list of proteins essential for this compartment[115,133,135].

Several common principles emerge that may inform study of other nuclear bodies. If multiple mechanisms are responsible for the formation of any nuclear body, one may ask if they are redundant or dependent upon one another. Separating them from one another may be required to understand the role of any one molecule, or any portion of a molecule. Finally, though most nuclear bodies are mesoscale structures, underlying molecular interactions must be responsible for building those structures. Identifying the molecular basis of each interaction is the first step in understanding how they sum together to form a membraneless organelle.

Chapter 5. Methods

5.1.   Experimental Procedures

Plasmid Construction and molecular cloning

All constructs used in this study are listed in Table 4, page 96. The plasmid system for lentiviral generation included pHR_SFFV (generated by Wendell Lim, Addgene #79121), pMD2.G (generated by Didier Trono, Addgene # 12259), and pCMVR 8.74 (generated by Didier Trono, Addgene # 22036). pHR_SFFV containing the IDRs of FUS and hnRNP-A1 were gifts from Cliff Brangwynne (Princeton University). Constructs were generated with the InFusion HD kit (Takara) by replacing the sequence upstream of mCherry-Cry2 in pHR_SFFV-hnRNPA1-mCherry-Cry2. Aubergine peptide constructs were generated by inserting sequence upstream of GFP in pLV-EGFP (generated by Pantelis Tsoulfas, Addgene #36083, modified by David Phizicky) and expressed from the eIF4a promoter. pEGFP, pECFP, and pEYFP (ClonTech) were generated with the InFusion HD kit with the help of Korinna Straube and Martin Machyna. Point mutations were generated by site-directed mutagenesis.

Cell Culture and cell line generation

The cell lines used in this study include HEK 293FT, NIH-3T3, and *coil* $^{-/-}$-mouse embryonic fibroblasts (MEF), HeLa Kyoto wild-type, Sf9 insect cells, Hi5 insect cells, and HeLa Kyoto coilin-GFP. Mammalian cells were cultured in DMEM supplemented with 10% FBS, penicillin, and streptomycin (Gibco). Cells were incubated at 37°C in a 5% CO_2 atmosphere. For live cell imaging, NIH-3T3 cells were cultured in glass-bottomed 35 mm dishes (MatTek). Just before imaging, the media was replaced with Live Cell Imaging

Solution (Invitrogen) supplemented with 20 mM glucose and 1 µg/mL Hoechst 33342 (Invitrogen).

Pools of cells stably expressing Cry2 constructs were generated by first transfecting either pHR_SFFV or pLV-EGFP with the desired insert, pMD2.G, and pCMV R8.74 into confluent HEK 293FT. Transfection was performed with the Fugene HD reagent (Promega). After 48-72 h, viral supernatant was harvested, filtered, and applied to NIH-3T3 or *coil*$^{-/-}$-MEF. Transient transfection for expression of constructs was performed using the Lipofectamine 2000, or 3000 kit, depending on expression level and transfection efficiency desired.

Sf9 cells were cultured in HyClone SFX (GE), and Hi5 were cultured in HyClone SFX (GE) supplemented with 10% FBS, and 5mM glutamine (Gibco). Sf9 were grown at 28°C as adherent cultures. Hi5 were grown in shaker flasks at 30°C. Baculovirus was generated by transfecting Sf9 cells with Lipofectin and the pFlashBac plasmid along with the desired pOCC8-pOEM vector. After 72 hrs, viral supernatant was collected and either stored for later use or applied to Hi5 cells to induce expression of the desired construct.

Live Cell Imaging of Cry2 Condensates

All live cell imaging was carried out on a Bruker Opterra II Swept Field instrument. The system allows for simultaneous imaging of mCherry fluorescence with the 561 nm laser and field of view activation of Cry2 with the 488 nm laser. The instrument is equipped with a Photometrics Evolve 512 Delta EMCCD camera. A standard protocol was used to make all Cry2 measurements. A PlanApo 60X 1.2 NA water immersion objective (Nikon) and a slit aperture of 70 µm was used to image at four frames per second. Four-frame averaging resulted in an overall frame rate of one frame per second. The 561 nm laser was

set to 80% of full power and the 488 nm was set to 25% of full power. Both were passed through a 20% neutral density filter. Power output was measured at the objective to be 80-90 µW for 561 nm and 20-25 µW for 488 nm with a Thorlabs PM100D power meter and a S170C sensor.

For each field of view, 10 s of inactive Cry2 were recorded, followed by turning on the 488 nm laser to capture 180 s of active Cry2. At the end of the series, a reference image of nuclear Hoechst staining was collected using the 405 nm laser. A stage-top incubator was employed to maintain cells at 37°C throughout imaging.

Fluorescence Recovery after Photobleaching (FRAP)

FRAP measurements were taken using identical imaging conditions as above with the following modifications. The objective was changed to a PlanApo 100X 1.4 NA oil immersion objective (Nikon). Frame averaging was replaced with a maximum projection of 7-11 z-steps at each time point, 0.25 s exposure per frame. After 180 s of activation, a single condensate was photobleached using a pulse of 405 nm laser light. The background was subtracted from a bleached region of interest containing a condensate using an unclustered region of the same cell. The intensity was then normalized to the maximum and minimum intensity values. Each recovery curve was fit using a one-phase exponential to estimate the mobile fraction and characteristic recovery time ($1/e$ recovery, tau).

Western Blotting

Western blots were carried out by preparing whole cell lysate in 0.1% SDS, 0.5% sodium deoxycholate, 150 mM NaCl, 10 mM EDTA, and 10 mM Tris (pH 8.0). Lysates were denatured and reduced by adding NuPAGE LDS sample buffer (Thermo) and heating for 10 min at 85°C. Samples were run out on a 4-12% acrylamide Bis-Tris gel (Invitrogen)

and transferred to 0.2 μm nitrocellulose membranes (BioRad). After blocking, membranes were stained with the antibodies listed in Table 5, page 97.

Fixed cell imaging

Cells were grown on No. 1.5 coverslips (Zeiss) in six-well plates and fixed in 4% paraformaldehyde (Sigma), blocked in 3% Bovine Serum Albumin (Sigma) and permeabilized by 0.1% Triton X-100 (American Bioanalytical). If immunofluorescence was performed, staining was conducted in blocking buffer with the antibodies listed in Table 5, page 97.

For Cry2 activation prior to fixation, we built a custom illumination array designed for six-well plates. Six blue LEDs (470 nm, 3.2 V, Digi-Key 150080BS75000) were arrayed to illuminate each well in a six-well plate for 5 min (Figure 32, page 98). Illumination was maintained during the 10 min fixation step. Inactive Cry2 samples were maintained in the dark for 10 min and fixed in the same manner. Imaging was done on a Leica Sp8 Laser Scanning Confocal.

DMA Inhibitor Treatment

DMA inhibition was achieved by simultaneous treatment with two drugs: MS-023 (Sigma) and EPZ015666 (Sigma)[167,168]. Stock solutions were prepared in DMSO at 50 mM and all subsequent dilutions were done in cell media as described above. MS-023 was added to media at a final concentration of 1.0 μM, and EPZ015666 was added at a final concentration of 5.0 μM. After 48 hrs of drug treatment, imaging was carried out as above, with drug added to the same concentration in the imaging media.

GFP Co-Immunoprecipitation

Co-immunoprecipitation assays were performed using lysates prepared in Pierce lysis buffer (150 mM NaCl, 25 mM HEPES pH 7.5, 1 mM EDTA, 1% NP-40, 5% glycerol, and Roche protease inhibitor). Cells were grown in 10 cm dishes under the specified condition to 90% confluency, harvested and lysed by sonication. Insoluble material was centrifuged out of solution and 4 µg anti-GFP IgG was added to each lysate and incubated overnight at 4°C. The following day, 25 µL protein G magnetic SureBeads (BioRad) were added and incubated for 1 hr. Samples were washed in Pierce lysis buffer three times and eluted into NuPage LDS sample buffer.

Förster resonance energy transfer assay (FRET)

The assay was performed as an acceptor photobleaching FRET experiment[196,203]. 2 µg of each plasmid were transfected into HeLa cells overnight. Cells were fixed in 2% formaldehyde (J.T. Baker) and mounted on slides without Hoechst. Two biological replicates were performed for each condition and the results were pooled. Imaging was performed on a Zeiss 710 laser scanning confocal. Each region of ~30x30 pixels was selected, recorded (I_1), bleached, and then recorded again (I_2). The fluorescence intensity of each measurement was corrected by subtracting the mean background value of each sample calculated from a region of interest outside the cell. FRET efficiency (E_f) was calculated as $E_f = (I_2 - I_1) / (I_2)$.

Amyloid sensitive dye staining

The Coilin NTD was expressed in HeLa cells to form fibrils, which were then fixed as described above. Each coverslip was treated with the corresponding amyloid sensitive dye, 200 nM Thioflavin T (Sigma), 0.005% Congo Red (Sigma), and 3µM Proteostat

according to published protocols or according to the manufacturer's guidelines (Biotium)[194,204]. Fibrils were stained using the myc epitope and a Cy5 conjugated secondary antibody to prevent signal bleed through. Imaging was performed on a GE Life Sciences DeltaVision microscope.

Electron microscopy of HeLa fibrils

The coilin NTD was expressed in HeLa cells as described. Cells were fixed 8, 16, and 24 hours post-transfection using glutaraldehyde/formaldehyde buffer provided by the Yale CCMI. Samples were sectioned and processed by the CCMI Electron Microscopy facility staff.

siRNA Silencing

Coilin and Nopp140 were depleted using Ambion Silencer Select oligos (ThermoFisher) siRNAs targeting the *Coil* (s15662, s15663, s15664) and *Nolc1* (Oligo #: s17632, s17633, s17632) genes. Oligos were prepared according to the manufacturer's instructions. Pools containing equal amounts of each of three oligos were mixed, and 10 pmol total was transfected into 50% confluent HeLa cells in a 12 well dish using Lipofectamine RNAiMAX (Invitrogen). After 24 hours, cells were plated into 6 well dishes. Cells were harvested or fixed after 48 hours of knockdown.

Table 4. Plasmid constructs used in this study

Construct	Uniprot ID	Insert Sequence	Backbone	Promoter
mCh-Cry2	N/A	None	pHR_SFFV-mCherry-Cry2	SFFV
FUS^{IDR}	P35637	1-214	pHR_SFFV-mCherry-Cry2	SFFV
hnRNPA1^{IDR}	P09651	186-320	pHR_SFFV-mCherry-Cry2	SFFV
SMN^{IDR-N}	Q16637	1-91	pHR_SFFV-mCherry-Cry2	SFFV
SMN^{Tud}	Q16637	92-145	pHR_SFFV-mCherry-Cry2	SFFV
SMN^{IDR-C}	Q16637	146-294	pHR_SFFV-mCherry-Cry2	SFFV
dmTudor^{Tud}	P25823	2315-2515	pHR_SFFV-mCherry-Cry2	SFFV
Aub3	O76922	3x(NPVIARGRGRGRK)	pLV-EGFP	eIF4a
Aub3(KG)	O76922	3x(NPVIAKGKGKGKK)	pLV-EGFP	eIF4a
Spf30^{Tud}	O75940	72-132	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd1^{Tud}#2	O9BXT4	541-600	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd1^{Tud}#3	O9BXT4	762-821	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd1^{Tud}#4	O9BXT4	990-1048	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd3^{Tud}	Q9H7E2	555-615	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd4^{Tud}#1	Q9BXT8	726-784	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd4^{Tud}#4	Q9BXT8	1479-1539	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd6^{Tud}#5	O60522	1033-1088	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd6^{Tud}#6	O60522	1352-1411	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd8^{Tud}	Q9BXU1	78-137	pHR_SFFV-mCherry-Cry2	SFFV
Tdrd9^{Tud}	Q8NDG6	944-1004	pHR_SFFV-mCherry-Cry2	SFFV
Snd1^{Tud}	Q7KZF4	729-787	pHR_SFFV-mCherry-Cry2	SFFV
Coilin^{NTD}	Q5SU73	1-97	pHR_SFFV-mCherry-Cry2	SFFV
Coilin^{IDR}*	Q5SU73	98-405	pHR_SFFV-mCherry-Cry2	SFFV
Coilin^{RG+CTD}	Q5SU73	406-576	pHR_SFFV-mCherry-Cry2	SFFV
Coilin^{CTD}	Q5SU73	460-576	pHR_SFFV-mCherry-Cry2	SFFV
Coilin$^{\Delta NTD}$*	Q5SU73	98-576	pHR_SFFV-mCherry-Cry2	SFFV
Coilin$^{\Delta NLS}$	Q5SU73	1-106…200-576	pHR_SFFV-mCherry-Cry2	SFFV
Nopp140^{LisH}	Q14978	1-42	pHR_SFFV-mCherry-Cry2	SFFV
Nopp140^{IDR}	Q14978	43-585	pHR_SFFV-mCherry-Cry2	SFFV
Nopp140^{CTR}	Q14978	586-706	pHR_SFFV-mCherry-Cry2	SFFV
Nopp140^{FL}	Q14978	1-706	pHR_SFFV-mCherry-Cry2	SFFV
SMN^{FL}	Q16637	1-294	pHR_SFFV-mCherry-Cry2	SFFV
Coilin^{FL}	Q5SU73	1-576	pHR_SFFV-mCherry-Cry2	SFFV
Coilin-NTD	Q5SU73	1-97 + myc	pEGFP-N1	CMV
Coilin-NTD-NLS	Q5SU73	1-97 + SV40 NLS +myc	pEGFP-N1	CMV
Zcoilin-GFP	B07SK8	1-538	pOCC8-pOEM-GFP	Gb3
Zcoilin	B07SK8	1-538	pOCC8-pOEM-GFP	Gb3
Coilin-GFP	Q5SU73	1-576	pEGFP-N1	CMV
CFP-coilin	Q5SU73	1-576	pECFP-C1	CMV
YFP-coilin	Q5SU73	1-576	pEYFP-C1	CMV
Nopp140-YFP	Q5SU73	1-576	pEYFP-N1	CMV
CFP	N/A	None	pECFP-C1	CMV
YFP	N/A	None	pEYFP-C1	CMV
CFP-YFP	N/A	YFP from above	pECFP-C1	CMV

(dm) denotes sequence from *Drosophila melanogaster. (Z)* denotes sequence from *Danio rerio. "*Myc*"* sequence: EQKLISEEDL. "SV40 NLS" sequence: PKKKRKV. Korinna Straube prepared or advised cloning.

Table 5. Antibodies Used in this study.

Name	Epitope	Source	Application & Dilution
ASYM24	aDMA	Millipore Sigma (#07-414	WB (1:2000)
Anti-c-myc (9E10)	c-myc	Santa Cruz Biotechnology (#sc-40)	IF (1:200) WB (1:2000)
Anti-c-myc (A-14)	c-myc	Santa Cruz Biotechnology (#sc-789)	IF (1:200)
Coilin (H-300)	Coilin	Santa Cruz Biotechnology (#sc-32860)	IF (1:200)
Anti-Fibrillarin (H-140)	Fibrillarin	Santa Cruz Biotechnology (#sc-25397)	IF (1:100)
FUS/TLS (4H11)	FUS	Santa Cruz Biotechnology (#sc-47711)	IF (1:200)
Anti-G3BP1	G3BP1	Abcam (#ab56574)	IF (1:200)
GAPDH (FL-335)	GAPDH	SantaCruz Biotechnology (#sc-25778)	WB (1:2000)
Anti-green fluorescent protein, rabbit IgG fraction	GFP	Invitrogen (#A11122)	WB (1:2000)
Anti-GFP, Goat, MPI-CBG	GFP	David Drechsel, MPI-CBG	IP (4 μg)
Anti-hnRNP A1	hnRNP-A1	Doug Black Laboratory	IF (1:200)
HnRNP K (D-6)	hnRNP-K	Santa Cruz Biotechnology (#sc-28380)	IF (1:200)
hnRNP U (3Gb)	hnRNP-U	Santa Cruz Biotechnology (#sc-32315)	IF (1:200)
mCherry Antibody	mCherry	Invitrogen (#PA5-34974)	WB (1:2000), IF (1:500)
Alexa Fluor 488-conjugated AffiniPure Donkey Anti-Mouse IgG	Mouse IgG	Jackson ImmunoResearch (#715-545-150)	IF (1:500)
Alexa Fluor 568-conjugated Donkey Anti-Mouse IgG	Mouse IgG	Life Technologies (#A10037)	IF (1:500)
Nopp140 (H-80)	Nopp140	Santa Cruz Biotechnology (#sc-28672)	IF (1:200) WB (1:2000)
Pol II (N-20) X	PolII	Santa Cruz Biotechnology (#sc-899)	IF (1:200)
Anti-DNA-RNA Hybrid (S9.6)	R-Loop	EMD Millipore (#MABE1095)	IF (1:200)
Anti-rabbit IgG Horseradish Peroxidase-Linked Species-Specific Whole Antibody	Rabbit IgG	GE HealthCare (#NA934)	WB (1:10,000)
Alexa Fluor 488-conjugated AffiniPure Donkey Anti-Rabbit IgG	Rabbit IgG	Jackson ImmunoResearch (#711-545-152)	IF (1:500)
Cy5-conjugated AffiniPure Donkey Anti-Rabbit IgG	Rabbit IgG	Jackson ImmunoResearch (711-175-152)	IF (1:500)
SYM10	sDMA	Millipore Sigma (#07-412)	WB (1:1000), IF (1:100)
SYM11	sDMA	Millipore Sigma (#07-413)	WB (1:1000)
Ms mAb to SFPQ	SFPQ	Abcam (#ab11825)	IF (1:200)
Y12	Sm	Gift from Joan Steitz	IF (1:10)
Ms mAb to Gemin 1 (2B1)	SMN	Abcam (#ab5831)	IF (1:200)
Anti-2,2,7-Trimethylguanosine Mouse mAb (K121)	TMG	Calbiochem (#NA02)	IF (1:200)
Rb pAb to SNRPC	U1 snRNP-C	Abcam (#ab82862)	IF (1:200)
CB7	U1-70K	Doug Black Laboratory	IF (1:200)

Abbreviations- WB: Western Blot, IF: Immunofluorescence, IP: Immunoprecipitation

Figure 32. A blue light illumination plate used to activate Cry2 in fixed cell assays.

This illumination plate was used to activate Cry2 in cells plated on coverslips in a six-well dish. Each LED is powered by an adjustable DC power source and is arranged to illuminate the center of each well of the dish.

5.2. Analysis

Image Analysis: Segmentation

Quantification of condensate formation on a per-cell basis requires accurate segmentation of cells within a field of view. A cell profiler pipeline was used to process movies into cell masks. Binary nuclear masks were generated by segmenting the Hoechst image associated with each sequence. Binary cellular masks were generated by performing an Otsu threshold and watershed transform seeded by the nuclear masks on the mCherry signal from inactive Cry2 cells. Masks defined by automated segmentation were manually checked for accuracy; cells with substantial movement from the mask were excluded from analysis.

Image Analysis: Intensity correction

A temporal variance metric was used to quantify clustering without imposing a model on the time-dependence or shape, and in an expression-level independent manner. As photon-counting processes are subject to shot-noise, which is Poisson distributed, the temporal variance of a given pixel increases due to baseline intensity increases, as well as changes in fluorophores moving into or out of the sample region. This effect can be conveniently normalized using the temporal mean, since the variance and expectation value of a Poisson process are equal. Several effects which deviate our raw intensity data from this model were accounted for before calculating the clustering metric.

First, the per-pixel analog-digital offset was characterized and subtracted from all images. We often observe very bright spots appearing on single frames at random positions, which we interpret to be particles or cosmic rays impacting the EMCCD. Such large spikes in intensity can bias variance-based calculations, and because these spikes sometimes occur

within cell masks, we filter them using a temporal sliding-window approach (Figure 33A, page 103). We iterate through time and calculate the median-absolute deviation (MAD) at each xy-pixel within a 10-frame window, where the MAD of a sequence k is simply $\text{median}(|k - \text{median}(k)|)$. Values within this window were identified as spikes if they were both larger than 15 times the MAD and were increases on the previous frame larger than 1500 ADU. Spikes were replaced with the median value of the window, then the window was advanced one frame and the process was repeated until the last frame was included in the window. MAD was employed to determine the variability within a window as it is a robust variability measure, and does not increase substantially due to the very spikes we are trying to filter, unlike the standard deviation or variance.

Finally, we performed a mean-normalization to remove effects of read-out laser intensity fluctuation, bleaching, etc., by multiplying frame j by a factor $\beta = \langle I(t = 0) \rangle_{x,y} / \langle I(t_j) \rangle_{x,y}$ where $\langle k \rangle_x$ denotes an average of k taken over x.

Image Analysis: Clustering Metric

We calculated the temporal variance and average for each pixel during Cry2-activation in live cells. In order to arrive at a metric which is theoretically independent of baseline signal magnitude, and therefore expression levels, we normalized the temporal variance by the mean. This is a convenient normalization for data influenced by shot-noise because the variance of a Poisson process is equal to its mean. The clustering metric is given by

$$\text{Clustering metric} = \left\langle \left| \frac{\text{Var}_t\left[C_{x,y}(t)\right]}{\langle C_{x,y}(t) \rangle_t} \right| \right\rangle_{x,y \in m}$$

where $C_{i,j}(t)$ is the corrected intensity at pixel *x,y* as a function of time, and *m* denotes the set of *x,y* pixels within a mask. An intermediary variance-over-mean image can be generated where individual clusters are visible. Using fluorescence intensity as a proxy for concentration, the clustering metric can be plotted against intensity so that comparisons between constructs may be made (Figure 33B, page 103).

The clustering metric for a homogenous Poisson process should be unity, however our camera data is not only subject to shot noise, but also read noise and noise due to the stochastic nature of electron multiplication *(32)*. A more complete camera noise model may be necessary for comparison of results from different microscopes or camera settings.

Clustering classification

While the clustering metric allows for comparison between the degree of clustering for individual cells, the expression efficiency and/or sampling varies between constructs, complicating statistical analysis between them. To remove the influence of data which did not experience significant clustering, we performed a Mann-Whitney U test between the clustering metric distribution of the negative control (mCherry-Cry2) and each individual clustering metric from other distributions, classifying cells which reject the null hypothesis with $\alpha = 0.2$ as cells which experience clustering. We then calculated the median clustering metric for a construct using only its cluster-positive cells and used that to compute differences in clustering magnitude.

Software

Images were prepared for figures and FRAP calculations were done using ImageJ version 2.0.0-rc-69/1.52p. Image segmentation was done using Cell Profiler version 3.1.9. All other image analysis was done using PYME version 19.12.17 which is available at

python-microscopy.org. The plugin used to compute clustering metrics can be found at github.com/bewersdorflab/quant-condensate. Structure prediction was performed on the RaptorX Property Prediction server using default parameters[152]. Tudor domains were aligned by extracting domain annotations from Uniprot.org and aligning sequences with Clustal Omega using the default parameters.

Image Display

All images within a figure are contrast matched, and an intensity grayscale or color bar is given for reference where applicable. All live cell images are micrographs from single z-slice movies except where otherwise noted. Images and quantification comparisons are only made between samples taken with identical microscope calibrations. Confocal images of fixed samples are maximum intensity projections of z-stacks.
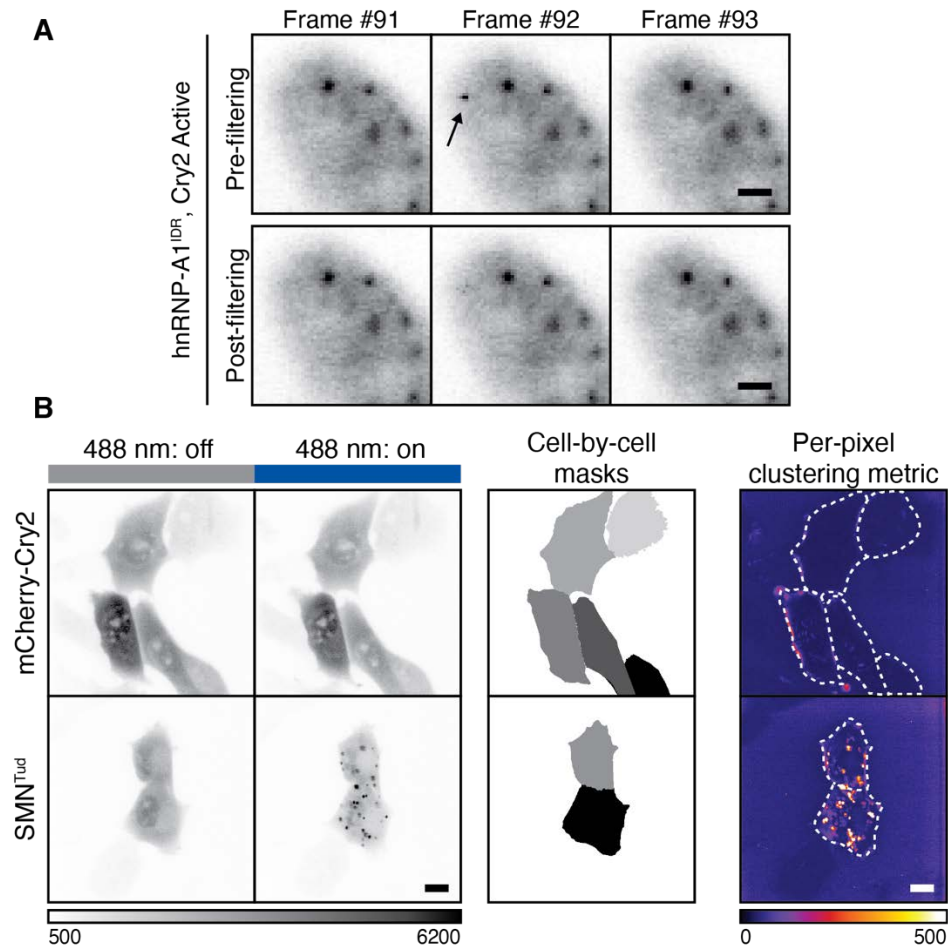
Figure 33. Filtering and quantification of live cell imaging of Cry2 samples.

A) An example of a single frame spike in intensity filtered out using median-absolute deviation. Arrowhead indicates spike. After filtering, the spot is effectively removed. Each frame is a 1.0 s exposure. B) Live cells are recorded for 10 s without Cry2 activation and then 180 s with blue light, activating Cry2. The mCherry signal before activation allows for segmentation of individual cells. The per-pixel temporal variance and mean are computed for the Cry2-active period. These values are used to construct the per-pixel clustering metric (right). Per-cell values are calculated by averaging over each cell mask. Grayscale bar (left) indicates fluorescence intensity, color bar (right) indicates clustering metric, given in analog-digital units (ADU). Scale bar = 10 μm. Andrew Barentine contributed to developing this pipeline.

References

1   Courchaine, E. M., Lu, A. & Neugebauer, K. M. Droplet organelles? *EMBO J* **35**, 1603-1612, doi:10.15252/embj.201593517 (2016).

2   Grossman, E., Medalia, O. & Zwerger, M. Functional architecture of the nuclear pore complex. *Annu Rev Biophys* **41**, 557-584, doi:10.1146/annurev-biophys-050511-102328 (2012).

3   Neupert, W. A perspective on transport of proteins into mitochondria: a myriad of open questions. *J Mol Biol* **427**, 1135-1158, doi:10.1016/j.jmb.2015.02.001 (2015).

4   Machyna, M., Heyn, P. & Neugebauer, K. M. Cajal bodies: where form meets function. *Wiley interdisciplinary reviews. RNA* **4**, 17-34, doi:10.1002/wrna.1139 (2013).

5   Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends in genetics : TIG* **27**, 295-306, doi:10.1016/j.tig.2011.05.006 (2011).

6   Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* **36**, 932-941, doi:10.1016/j.molcel.2009.11.020 (2009).

7   Gall, J. G. Cajal bodies: the first 100 years. *Annu Rev Cell Dev Biol* **16**, 273-300, doi:10.1146/annurev.cellbio.16.1.273 (2000).

8   Cajal, S. R. Un sencillo metodo de coloracion seletiva del reticulo protoplasmatico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. *Trab. Lab. Invest. Biol.(Madrid)* **2**, 129-221 (1903).

9   Dundr, M. *et al.* In vivo kinetics of Cajal body components. *J Cell Biol* **164**, 831-842, doi:10.1083/jcb.200311121 (2004).

10   Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604-609, doi:10.1038/35007077 (2000).

11   Boisvert, F. M., van Koningsbruggen, S., Navascues, J. & Lamond, A. I. The multifunctional nucleolus. *Nat Rev Mol Cell Biol* **8**, 574-585, doi:10.1038/nrm2184 (2007).

12   Pederson, T. The nucleolus. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a000638 (2011).

13   Falahati, H., Pelham-Webb, B., Blythe, S. & Wieschaus, E. Nucleation by rRNA Dictates the Precision of Nucleolus Assembly. *Curr Biol* **26**, 277-285, doi:10.1016/j.cub.2015.11.065 (2016).

14   Frey, M. R., Bailey, A. D., Weiner, A. M. & Matera, A. G. Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. *Curr Biol* **9**, 126-135 (1999).

15   Novotny, I. *et al.* SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. *Cell Rep*, doi:10.1016/j.celrep.2014.12.030 (2015).

16   Stanek, D. & Neugebauer, K. M. Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. *J Cell Biol* **166**, 1015-1025, doi:10.1083/jcb.200405160 (2004).

17   Parker, R. & Sheth, U. P bodies and the control of mRNA translation and degradation. *Mol Cell* **25**, 635-646, doi:10.1016/j.molcel.2007.02.011 (2007).

18   Arribere, J. A., Doudna, J. A. & Gilbert, W. V. Reconsidering movement of eukaryotic mRNAs between polysomes and P bodies. *Mol Cell* **44**, 745-758, doi:10.1016/j.molcel.2011.09.019 (2011).

19   Anderson, P. & Kedersha, N. Stress granules: the Tao of RNA triage. *Trends Biochem Sci* **33**, 141-150, doi:10.1016/j.tibs.2007.12.003 (2008).

20   Li, Y. R., King, O. D., Shorter, J. & Gitler, A. D. Stress granules as crucibles of ALS pathogenesis. *J Cell Biol* **201**, 361-372, doi:10.1083/jcb.201302044 (2013).

21   Moon, S. L. *et al.* Multicolour single-molecule tracking of mRNA interactions with RNP granules. *Nature cell biology* **21**, 162-168, doi:10.1038/s41556-018-0263-4 (2019).

22   Boulon, S., Westman, B. J., Hutten, S., Boisvert, F. M. & Lamond, A. I. The nucleolus under stress. *Mol Cell* **40**, 216-227, doi:10.1016/j.molcel.2010.09.024 (2010).

23   Seydoux, G. & Braun, R. E. Pathway to totipotency: lessons from germ cells. *Cell* **127**, 891-904, doi:10.1016/j.cell.2006.11.016 (2006).

24   Trcek, T. & Lehmann, R. Germ granules in Drosophila. *Traffic* **20**, 650-660, doi:10.1111/tra.12674 (2019).

25   Kotaja, N. & Sassone-Corsi, P. The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat Rev Mol Cell Biol* **8**, 85-90, doi:10.1038/nrm2081 (2007).

26   Platani, M., Goldberg, I., Swedlow, J. R. & Lamond, A. I. In vivo analysis of Cajal body movement, separation, and joining in live human cells. *J Cell Biol* **151**, 1561-1574 (2000).

27   Handwerger, K. E., Cordero, J. A. & Gall, J. G. Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure. *Mol Biol Cell* **16**, 202-211, doi:10.1091/mbc.E04-08-0742 (2005).

28   Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732, doi:10.1126/science.1172046 (2009).

29   Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc Natl Acad Sci U S A* **108**, 4334-4339, doi:10.1073/pnas.1017150108 (2011).

30   Frey, S., Richter, R. P. & Gorlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815-817, doi:10.1126/science.1132516 (2006).

31   Ader, C. *et al.* Amyloid-like interactions within nucleoporin FG hydrogels. *Proc Natl Acad Sci U S A* **107**, 6281-6285, doi:10.1073/pnas.0910163107 (2010).

32   Schmidt, H. B. & Gorlich, D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *Elife* **4**, doi:10.7554/eLife.04251 (2015).

33   Labokha, A. A. *et al.* Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. *EMBO J* **32**, 204-218, doi:10.1038/emboj.2012.302 (2013).

34   Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340, doi:10.1038/nature10879 (2012).

35   Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049-1060, doi:10.1016/j.cell.2013.10.033 (2013).

36   Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139-1145, doi:10.1126/science.1254917 (2014).

37   Wang, J. T. *et al.* Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in C. elegans. *Elife* **3**, e04591, doi:10.7554/eLife.04591 (2014).

38   Su, X. *et al.* Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595-599, doi:10.1126/science.aad9964 (2016).

39   Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767, doi:10.1016/j.cell.2012.04.017 (2012).

40   Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501-2502, doi:10.1093/bioinformatics/btu310 (2014).

41   Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146-158, doi:10.1016/j.cell.2009.02.044 (2009).

42   King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res* **1462**, 61-80, doi:10.1016/j.brainres.2012.01.016 (2012).

43   Decker, C. J., Teixeira, D. & Parker, R. Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. *J Cell Biol* **179**, 437-449, doi:10.1083/jcb.200704147 (2007).

44   Sun, Z. *et al.* Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol* **9**, e1000614, doi:10.1371/journal.pbio.1000614 (2011).

45   Wang, W. Y. *et al.* Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. *Nat Neurosci* **16**, 1383-1391, doi:10.1038/nn.3514 (2013).

46   Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**, 126-130, doi:10.1038/nature06992 (2008).

47   Bentmann, E. *et al.* Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J Biol Chem* **287**, 23079-23094, doi:10.1074/jbc.M111.328757 (2012).

48   Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467-473, doi:10.1038/nature11922 (2013).

49   Malinovska, L., Kroschwald, S. & Alberti, S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta* **1834**, 918-931, doi:10.1016/j.bbapap.2013.01.003 (2013).

50   Toretsky, J. A. & Wright, P. E. Assemblages: functional units formed by cellular phase separation. *J Cell Biol* **206**, 579-588, doi:10.1083/jcb.201404124 (2014).

51   Uversky, V. N. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett* **589**, 2498-2506, doi:10.1016/j.febslet.2015.06.004 (2015).

52   Nott, T. J. *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* **57**, 936-947, doi:10.1016/j.molcel.2015.01.013 (2015).

53   Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell* **60**, 208-219, doi:10.1016/j.molcel.2015.08.018 (2015).

54   Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123-133, doi:10.1016/j.cell.2015.09.015 (2015).

55   Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun* **6**, 8088, doi:10.1038/ncomms9088 (2015).

56   Berry, J., Weber, S. C., Vaidya, N., Haataja, M. & Brangwynne, C. P. RNA transcription modulates phase transition-driven nuclear body assembly. *Proc Natl Acad Sci U S A* **112**, E5237-5245, doi:10.1073/pnas.1509317112 (2015).

57   Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol Cell* **60**, 231-241, doi:10.1016/j.molcel.2015.09.006 (2015).

58   Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678-690, doi:10.1016/j.neuron.2015.10.030 (2015).

59   Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066-1077, doi:10.1016/j.cell.2015.07.047 (2015).

60   Xiang, S. *et al.* The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei. *Cell* **163**, 829-839, doi:10.1016/j.cell.2015.10.040 (2015).

61   Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A* **112**, 7189-7194, doi:10.1073/pnas.1504822112 (2015).

62   Hennig, S. *et al.* Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *J Cell Biol* **210**, 529-539, doi:10.1083/jcb.201504117 (2015).

63   Zhang, H. *et al.* RNA Controls PolyQ Protein Phase Transitions. *Mol Cell* **60**, 220-230, doi:10.1016/j.molcel.2015.09.017 (2015).

64   Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensies of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci U S A* **112**, 9614-9619, doi:10.1073/pnas.1512799112 (2015).

65   Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627 e616, doi:10.1016/j.cell.2017.08.048 (2017).

66   Chen, C., Nott, T. J., Jin, J. & Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol* **12**, 629-642, doi:10.1038/nrm3185 (2011).

67   Phan, A. T. *et al.* Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol* **18**, 796-804, doi:10.1038/nsmb.2064 (2011).

68   Machyna, M. *et al.* The coilin interactome identifies hundreds of small noncoding RNAs that traffic through Cajal bodies. *Mol Cell* **56**, 389-399, doi:10.1016/j.molcel.2014.10.004 (2014).

69   Van Treeck, B. & Parker, R. Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. *Cell* **174**, 791-802, doi:10.1016/j.cell.2018.07.023 (2018).

70   Tauber, D. *et al.* Modulation of RNA Condensation by the DEAD-Box Protein eIF4A. *Cell*, doi:10.1016/j.cell.2019.12.031 (2020).

71   Van Treeck, B. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc Natl Acad Sci U S A* **115**, 2734-2739, doi:10.1073/pnas.1800038115 (2018).

72   Langdon, E. M. *et al.* mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922-927, doi:10.1126/science.aar7432 (2018).

73   Shevtsov, S. P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nature cell biology* **13**, 167-173, doi:10.1038/ncb2157 (2011).

74   Bond, C. S. & Fox, A. H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J Cell Biol* **186**, 637-644, doi:10.1083/jcb.200906113 (2009).

75   Cherkasov, V. *et al.* Coordination of translational control and protein homeostasis during severe heat stress. *Curr Biol* **23**, 2452-2462, doi:10.1016/j.cub.2013.09.058 (2013).

76   Kroschwald, S. *et al.* Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *Elife* **4**, e06807, doi:10.7554/eLife.06807 (2015).

77   Buchan, J. R., Kolaitis, R. M., Taylor, J. P. & Parker, R. Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. *Cell* **153**, 1461-1474, doi:10.1016/j.cell.2013.05.037 (2013).

78   Chitiprolu, M. *et al.* A complex of C9ORF72 and p62 uses arginine methylation to eliminate stress granules by autophagy. *Nat Commun* **9**, 2794, doi:10.1038/s41467-018-05273-7 (2018).

79   Weber, S. C. & Brangwynne, C. P. Getting RNA and protein in phase. *Cell* **149**, 1188-1191, doi:10.1016/j.cell.2012.05.022 (2012).

80   Flory, P. I. Thermodynamics of high polymer solutions. *J Chem Phys* **10**, 51-61, doi:Doi 10.1063/1.1723621 (1942).

81   Huggins, M. L. Some properties of solutions of long-chain compounds. *J Phys Chem-Us* **46**, 151-158, doi:DOI 10.1021/j150415a018 (1942).

82   Teng, P. K. & Eisenberg, D. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng Des Sel* **22**, 531-536, doi:10.1093/protein/gzp037 (2009).

83   Klingauf, M., Stanek, D. & Neugebauer, K. M. Enhancement of U4/U6 small nuclear ribonucleoprotein particle association in Cajal bodies predicted by mathematical modeling. *Mol Biol Cell* **17**, 4972-4981, doi:10.1091/mbc.E06-06-0513 (2006).

84   Fowler, D. M., Koulov, A. V., Balch, W. E. & Kelly, J. W. Functional amyloid--from bacteria to humans. *Trends Biochem Sci* **32**, 217-224, doi:10.1016/j.tibs.2007.03.003 (2007).

85   Majumdar, A. *et al.* Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. *Cell* **148**, 515-529, doi:10.1016/j.cell.2012.01.004 (2012).

86   Fowler, D. M. *et al.* Functional amyloid formation within mammalian tissue. *PLoS Biol* **4**, e6, doi:10.1371/journal.pbio.0040006 (2006).

87   Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328-332, doi:10.1126/science.1173155 (2009).

88   Berchowitz, L. E. *et al.* Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. *Cell* **163**, 406-418, doi:10.1016/j.cell.2015.08.060 (2015).

89   Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat Commun* **8**, 275, doi:10.1038/s41467-017-00480-0 (2017).

90   Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247, doi:10.1038/nature22386 (2017).

91   Majcher, V., Goode, A., James, V. & Layfield, R. Autophagy receptor defects and ALS-FTLD. *Mol Cell Neurosci* **66**, 43-52, doi:10.1016/j.mcn.2015.01.002 (2015).

92   Broide, M. L., Berland, C. R., Pande, J., Ogun, O. O. & Benedek, G. B. Binary-liquid phase separation of lens protein solutions. *Proc Natl Acad Sci U S A* **88**, 5660-5664, doi:10.1073/pnas.88.13.5660 (1991).

93   Thomson, J. A., Schurtenberger, P., Thurston, G. M. & Benedek, G. B. Binary liquid phase separation and critical phenomena in a protein/water solution. *Proc Natl Acad Sci U S A* **84**, 7079-7083, doi:10.1073/pnas.84.20.7079 (1987).

94   Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S. & Schaffer, D. V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nat Methods* **10**, 249-252, doi:10.1038/nmeth.2360 (2013).

95   Shin, Y. *et al.* Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171 e114, doi:10.1016/j.cell.2016.11.054 (2017).

96   Protter, D. S. W. *et al.* Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep* **22**, 1401-1412, doi:10.1016/j.celrep.2018.01.036 (2018).

97   Zhu, L. *et al.* Controlling the material properties and rRNA processing function of the nucleolus using light. *Proc Natl Acad Sci U S A* **116**, 17330-17335, doi:10.1073/pnas.1903870116 (2019).

98   Zhang, P. *et al.* Chronic optogenetic induction of stress granules is cytotoxic and reveals the evolution of ALS-FTD pathology. *Elife* **8**, doi:10.7554/eLife.39578 (2019).

99   Bracha, D. *et al.* Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467-1480 e1413, doi:10.1016/j.cell.2018.10.048 (2018).

100   Shin, Y. *et al.* Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **176**, 1518, doi:10.1016/j.cell.2019.02.025 (2019).

101   Lam, Y. W., Lyon, C. E. & Lamond, A. I. Large-scale isolation of Cajal bodies from HeLa cells. *Mol Biol Cell* **13**, 2461-2473, doi:10.1091/mbc.02-03-0034 (2002).

102   Muramatsu, M., Hayashi, Y., Onishi, T., Sakai, M. & Takai, K. Rapid isolation of nucleoli from detergent purified nuclei of various tumor and tissue culture cells. *Exp Cell Res* **88**, 245-251, doi:10.1016/0014-4827(74)90250-x (1974).

103   Deryusheva, S. & Gall, J. G. Dynamics of coilin in Cajal bodies of the Xenopus germinal vesicle. *Proc Natl Acad Sci U S A* **101**, 4810-4814, doi:10.1073/pnas.0401106101 (2004).

104   Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487-498, doi:10.1016/j.cell.2015.12.038 (2016).

105   Khong, A. *et al.* The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol Cell* **68**, 808-820 e805, doi:10.1016/j.molcel.2017.10.015 (2017).

106   Hubstenberger, A. *et al.* P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol Cell* **68**, 144-157 e145, doi:10.1016/j.molcel.2017.09.003 (2017).

107   Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484 e421, doi:10.1016/j.cell.2019.08.037 (2019).

108   Sanulli, S. *et al.* HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* **575**, 390-394, doi:10.1038/s41586-019-1669-2 (2019).

109   Pak, C. W. *et al.* Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell* **63**, 72-85, doi:10.1016/j.molcel.2016.05.042 (2016).

110   Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663, doi:10.1016/j.cell.2016.06.010 (2016).

111   Ditlev, J. A., Case, L. B. & Rosen, M. K. Who's In and Who's Out-Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666-4684, doi:10.1016/j.jmb.2018.08.003 (2018).

112   Case, L. B., Zhang, X., Ditlev, J. A. & Rosen, M. K. Stoichiometry controls activity of phase-separated clusters of actin signaling proteins. *Science* **363**, 1093-1097, doi:10.1126/science.aau6313 (2019).

113   Sheu-Gruttadauria, J. & MacRae, I. J. Phase Transitions in the Assembly and Function of Human miRISC. *Cell* **173**, 946-957 e916, doi:10.1016/j.cell.2018.02.051 (2018).

114   Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548, doi:10.1038/s41586-019-1464-0 (2019).

115   Strzelecka, M. *et al.* Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. *Nat Struct Mol Biol* **17**, 403-409, doi:10.1038/nsmb.1783 (2010).

116   Tucker, K. E. *et al.* Residual Cajal bodies in coilin knockout mice fail to recruit Sm snRNPs and SMN, the spinal muscular atrophy gene product. *J Cell Biol* **154**, 293-307, doi:10.1083/jcb.200104083 (2001).

117   Wang, Q. *et al.* Cajal bodies are linked to genome conformation. *Nat Commun* **7**, 10966, doi:10.1038/ncomms10966 (2016).

118   Kiss, A. M. *et al.* A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res* **30**, 4643-4649, doi:10.1093/nar/gkf592 (2002).

119   Darzacq, X. *et al.* Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* **21**, 2746-2756, doi:10.1093/emboj/21.11.2746 (2002).

120   Chen, Y. *et al.* Human cells lacking coilin and Cajal bodies are proficient in telomerase assembly, trafficking and telomere maintenance. *Nucleic Acids Res* **43**, 385-395, doi:10.1093/nar/gku1277 (2015).

121 Venteicher, A. S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644-648, doi:10.1126/science.1165357 (2009).

122 Cristofari, G. *et al.* Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation. *Mol Cell* **27**, 882-889, doi:10.1016/j.molcel.2007.07.020 (2007).

123 Mahmoudi, S. *et al.* WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies. *PLoS Biol* **8**, e1000521, doi:10.1371/journal.pbio.1000521 (2010).

124 Suzuki, T., Izumi, H. & Ohno, M. Cajal body surveillance of U snRNA export complex assembly. *J Cell Biol* **190**, 603-612, doi:10.1083/jcb.201004109 (2010).

125 Lemm, I. *et al.* Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. *Mol Biol Cell* **17**, 3221-3231, doi:10.1091/mbc.e06-03-0247 (2006).

126 Carmo-Fonseca, M., Pepperkok, R., Carvalho, M. T. & Lamond, A. I. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J Cell Biol* **117**, 1-14 (1992).

127 Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonne, R. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell* **9**, 891-901, doi:10.1016/s1097-2765(02)00484-7 (2002).

128 Yu, Y. T., Shu, M. D. & Steitz, J. A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J* **17**, 5783-5795, doi:10.1093/emboj/17.19.5783 (1998).

129 Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146**, 384-395, doi:10.1016/j.cell.2011.06.043 (2011).

130 Otter, S. *et al.* A comprehensive interaction map of the human survival of motor neuron (SMN) complex. *J Biol Chem* **282**, 5825-5833, doi:10.1074/jbc.M608528200 (2007).

131 Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775-1779, doi:10.1126/science.1074962 (2002).

132 Stanek, D. *et al.* Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. *Mol Biol Cell* **19**, 2534-2543, doi:10.1091/mbc.E07-12-1259 (2008).

133 Walker, M. P., Tian, L. & Matera, A. G. Reduced viability, fertility and fecundity in mice lacking the cajal body marker protein, coilin. *PLoS One* **4**, e6171, doi:10.1371/journal.pone.0006171 (2009).

134 Girard, C., Neel, H., Bertrand, E. & Bordonne, R. Depletion of SMN by RNA interference in HeLa cells induces defects in Cajal body formation. *Nucleic Acids Res* **34**, 2925-2932, doi:10.1093/nar/gkl374 (2006).

135 Strzelecka, M., Oates, A. C. & Neugebauer, K. M. Dynamic control of Cajal body number during zebrafish embryogenesis. *Nucleus* **1**, 96-108, doi:10.4161/nucl.1.1.10680 (2010).

136 Tapia, O. *et al.* Reorganization of Cajal bodies and nucleolar targeting of coilin in motor neurons of type I spinal muscular atrophy. *Histochem Cell Biol* **137**, 657-667, doi:10.1007/s00418-012-0921-8 (2012).

137 Schrank, B. *et al.* Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc Natl Acad Sci U S A* **94**, 9920-9925 (1997).

138 Monani, U. R. *et al.* The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in Smn(-/-) mice and results in a mouse with spinal muscular atrophy. *Hum Mol Genet* **9**, 333-339 (2000).

139 Gavrilina, T. O. *et al.* Neuronal SMN expression corrects spinal muscular atrophy in severe SMA mice while muscle-specific SMN expression has no phenotypic effect. *Hum Mol Genet* **17**, 1063-1075, doi:10.1093/hmg/ddm379 (2008).

140 Kaiser, T. E., Intine, R. V. & Dundr, M. De novo formation of a subnuclear body. *Science* **322**, 1713-1717, doi:10.1126/science.1165216 (2008).

141 Hebert, M. D. & Matera, A. G. Self-association of coilin reveals a common theme in nuclear body localization. *Mol Biol Cell* **11**, 4159-4171 (2000).

142 Young, P. J. *et al.* Nuclear gems and Cajal (coiled) bodies in fetal tissues: nucleolar distribution of the spinal muscular atrophy protein, SMN. *Exp Cell Res* **265**, 252-261, doi:10.1006/excr.2001.5186 (2001).

143 Stejskalova, E. & Stanek, D. The splicing factor U1-70K interacts with the SMN complex and is required for nuclear gem integrity. *J Cell Sci* **127**, 3909-3915, doi:10.1242/jcs.155838 (2014).

144 Hebert, M. D., Shpargel, K. B., Ospina, J. K., Tucker, K. E. & Matera, A. G. Coilin methylation regulates nuclear body formation. *Dev Cell* **3**, 329-337 (2002).

145 Boisvert, F. M. *et al.* Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. *J Cell Biol* **159**, 957-969, doi:10.1083/jcb.200207028 (2002).

146 Nizami, Z., Deryusheva, S. & Gall, J. G. The Cajal body and histone locus body. *Cold Spring Harb Perspect Biol* **2**, a000653, doi:10.1101/cshperspect.a000653 (2010).

147 Renvoise, B. *et al.* The loss of the snoRNP chaperone Nopp140 from Cajal bodies of patient fibroblasts correlates with the severity of spinal muscular atrophy. *Hum Mol Genet* **18**, 1181-1189, doi:10.1093/hmg/ddp009 (2009).

148 Isaac, C., Yang, Y. & Meier, U. T. Nopp140 functions as a molecular link between the nucleolus and the coiled bodies. *J Cell Biol* **142**, 319-329 (1998).

149 Morse, R., Shaw, D. J., Todd, A. G. & Young, P. J. Targeting of SMN to Cajal bodies is mediated by self-association. *Hum Mol Genet* **16**, 2349-2358, doi:10.1093/hmg/ddm192 (2007).

150 Mateja, A., Cierpicki, T., Paduch, M., Derewenda, Z. S. & Otlewski, J. The dimerization mechanism of LIS1 and its implication for proteins containing the LisH motif. *J Mol Biol* **357**, 621-631, doi:10.1016/j.jmb.2006.01.002 (2006).

151 Hearst, S. M. *et al.* Cajal-body formation correlates with differential coilin phosphorylation in primary and transformed cell lines. *J Cell Sci* **122**, 1872-1881, doi:10.1242/jcs.044040 (2009).

152 Kallberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* **7**, 1511-1522, doi:10.1038/nprot.2012.085 (2012).

153 Berry, J., Brangwynne, C. P. & Haataja, M. Physical principles of intracellular organization via active and passive phase transitions. *Rep Prog Phys* **81**, 046601, doi:10.1088/1361-6633/aaa61e (2018).

154 McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev*, doi:10.1101/gad.331520.119 (2019).

155 Buhler, D., Raker, V., Luhrmann, R. & Fischer, U. Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum Mol Genet* **8**, 2351-2357, doi:10.1093/hmg/8.13.2351 (1999).

156 Pellizzoni, L., Kataoka, N., Charroux, B. & Dreyfuss, G. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* **95**, 615-624, doi:10.1016/s0092-8674(00)81632-3 (1998).

157 Raimer, A. C., Gray, K. M. & Matera, A. G. SMN - A chaperone for nuclear RNP social occasions? *RNA Biol* **14**, 701-711, doi:10.1080/15476286.2016.1236168 (2017).

158 Lee, L., Davies, S. E. & Liu, J. L. The spinal muscular atrophy protein SMN affects Drosophila germline nuclear organization through the U body-P body pathway. *Dev Biol* **332**, 142-155, doi:10.1016/j.ydbio.2009.05.553 (2009).

159 Liu, J. L. & Gall, J. G. U bodies are cytoplasmic structures that contain uridine-rich small nuclear ribonucleoproteins and associate with P bodies. *Proc Natl Acad Sci U S A* **104**, 11655-11659, doi:10.1073/pnas.0704977104 (2007).

160 Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A. & Dreyfuss, G. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol Cell* **7**, 1111-1117, doi:10.1016/s1097-2765(01)00244-1 (2001).

161 Wang, J. & Dreyfuss, G. Characterization of functional domains of the SMN protein in vivo. *J Biol Chem* **276**, 45387-45393, doi:10.1074/jbc.M105059200 (2001).

162 Tripsianes, K. *et al.* Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat Struct Mol Biol* **18**, 1414-1420, doi:10.1038/nsmb.2185 (2011).

163 Morales, Y., Caceres, T., May, K. & Hevel, J. M. Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Arch Biochem Biophys* **590**, 138-152, doi:10.1016/j.abb.2015.11.030 (2016).

164 Feng, Y. *et al.* Mammalian protein arginine methyltransferase 7 (PRMT7) specifically targets RXR sites in lysine- and arginine-rich regions. *J Biol Chem* **288**, 37010-37025, doi:10.1074/jbc.M113.525345 (2013).

165 Branscombe, T. L. *et al.* PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. *J Biol Chem* **276**, 32971-32976, doi:10.1074/jbc.M105412200 (2001).

166 Tang, J. *et al.* PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *J Biol Chem* **275**, 7723-7730, doi:10.1074/jbc.275.11.7723 (2000).

167 Eram, M. S. *et al.* A Potent, Selective, and Cell-Active Inhibitor of Human Type I Protein Arginine Methyltransferases. *ACS Chem Biol* **11**, 772-781, doi:10.1021/acschembio.5b00839 (2016).

168 Chan-Penebre, E. *et al.* A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nat Chem Biol* **11**, 432-437, doi:10.1038/nchembio.1810 (2015).

169 Fuller, H. R. *et al.* The SMN interactome includes Myb-binding protein 1a. *J Proteome Res* **9**, 556-563, doi:10.1021/pr900884g (2010).

170   Zhao, D. Y. *et al.* SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* **529**, 48-53, doi:10.1038/nature16469 (2016).

171   Boswell, R. E. & Mahowald, A. P. tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell* **43**, 97-104, doi:10.1016/0092-8674(85)90015-7 (1985).

172   Liu, H. *et al.* Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev* **24**, 1876-1881, doi:10.1101/gad.1956010 (2010).

173   Liu, K. *et al.* Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One* **7**, e30375, doi:10.1371/journal.pone.0030375 (2012).

174   Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).

175   Chan, E. K., Takano, S., Andrade, L. E., Hamel, J. C. & Matera, A. G. Structure, expression and chromosomal localization of human p80-coilin gene. *Nucleic Acids Res* **22**, 4462-4469, doi:10.1093/nar/22.21.4462 (1994).

176   Martin, R. M. *et al.* Principles of protein targeting to the nucleolus. *Nucleus* **6**, 314-325, doi:10.1080/19491034.2015.1079680 (2015).

177   Bohmann, K., Ferreira, J. A. & Lamond, A. I. Mutational analysis of p80 coilin indicates a functional interaction between coiled bodies and the nucleolus. *J Cell Biol* **131**, 817-831, doi:10.1083/jcb.131.4.817 (1995).

178   Tapia, O., Bengoechea, R., Berciano, M. T. & Lafarga, M. Nucleolar targeting of coilin is regulated by its hypomethylation state. *Chromosoma* **119**, 527-540, doi:10.1007/s00412-010-0276-7 (2010).

179   Shanbhag, R., Kurabi, A., Kwan, J. J. & Donaldson, L. W. Solution structure of the carboxy-terminal Tudor domain from human Coilin. *FEBS Lett* **584**, 4351-4356, doi:10.1016/j.febslet.2010.09.034 (2010).

180   Hebert, M. D., Szymczyk, P. W., Shpargel, K. B. & Matera, A. G. Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. *Genes Dev* **15**, 2720-2729, doi:10.1101/gad.908401 (2001).

181   Shpargel, K. B., Ospina, J. K., Tucker, K. E., Matera, A. G. & Hebert, M. D. Control of Cajal body number is mediated by the coilin C-terminus. *J Cell Sci* **116**, 303-312, doi:10.1242/jcs.00211 (2003).

182   Banjade, S. *et al.* Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc Natl Acad Sci U S A* **112**, E6426-6435, doi:10.1073/pnas.1508778112 (2015).

183   Li, D., Meier, U. T., Dobrowolska, G. & Krebs, E. G. Specific interaction between casein kinase 2 and the nucleolar protein Nopp140. *J Biol Chem* **272**, 3773-3779 (1997).

184   Tsai, W. C. *et al.* Arginine Demethylation of G3BP1 Promotes Stress Granule Assembly. *J Biol Chem* **291**, 22671-22685, doi:10.1074/jbc.M116.739573 (2016).

185   Ryan, V. H. *et al.* Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol Cell* **69**, 465-479 e467, doi:10.1016/j.molcel.2017.12.022 (2018).

186   Chang, B., Chen, Y., Zhao, Y. & Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science* **318**, 444-447, doi:10.1126/science.1145801 (2007).

187 Chen, H. K., Pai, C. Y., Huang, J. Y. & Yeh, N. H. Human Nopp140, which interacts with RNA polymerase I: implications for rRNA gene transcription and nucleolar structural organization. *Mol Cell Biol* **19**, 8536-8546, doi:10.1128/mcb.19.12.8536 (1999).

188 Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e616, doi:10.1016/j.cell.2018.06.006 (2018).

189 Xu, H. & Hebert, M. D. A novel EB-1/AIDA-1 isoform, AIDA-1c, interacts with the Cajal body protein coilin. *BMC Cell Biol* **6**, 23, doi:10.1186/1471-2121-6-23 (2005).

190 Machyna, M., Neugebauer, K. M. & Stanek, D. Coilin: The first 25 years. *RNA Biol* **12**, 590-596, doi:10.1080/15476286.2015.1034923 (2015).

191 Bellini, M. & Gall, J. G. Coilin can form a complex with the U7 small nuclear ribonucleoprotein. *Mol Biol Cell* **9**, 2987-3001, doi:10.1091/mbc.9.10.2987 (1998).

192 Berchtold, D., Battich, N. & Pelkmans, L. A Systems-Level Study Reveals Regulators of Membrane-less Organelles in Human Cells. *Mol Cell* **72**, 1035-1049 e1035, doi:10.1016/j.molcel.2018.10.036 (2018).

193 Strzelecka, M. *The essential role of Cajal bodies in embryonic development* Doctor rerum naturalium thesis, Technischen Universität Dresden, (2008).

194 Westermark, G. T., Johnson, K. H. & Westermark, P. Staining methods for identification of amyloid in tissue. *Methods Enzymol* **309**, 3-25, doi:10.1016/s0076-6879(99)09003-5 (1999).

195 Herman, B. Resonance energy transfer microscopy. *Methods Cell Biol* **30**, 219-243, doi:10.1016/s0091-679x(08)60981-4 (1989).

196 Karpova, T. S. *et al.* Fluorescence resonance energy transfer from cyan to yellow fluorescent protein detected by acceptor photobleaching using confocal microscopy and a single laser. *J Microsc* **209**, 56-70, doi:10.1046/j.1365-2818.2003.01100.x (2003).

197 Bizarro, J., Bhardwaj, A., Smith, S. & Meier, U. T. Nopp140-mediated concentration of telomerase in Cajal bodies regulates telomere length. *Mol Biol Cell* **30**, 3136-3150, doi:10.1091/mbc.E19-08-0429 (2019).

198 Sanz-Garcia, M. *et al.* Substrate profiling of human vaccinia-related kinases identifies coilin, a Cajal body nuclear protein, as a phosphorylation target with neurological implications. *J Proteomics* **75**, 548-560, doi:10.1016/j.jprot.2011.08.019 (2011).

199 Cantarero, L. *et al.* VRK1 regulates Cajal body dynamics and protects coilin from proteasomal degradation in cell cycle. *Sci Rep* **5**, 10543, doi:10.1038/srep10543 (2015).

200 Raska, I., Ochs, R. L. & Salamin-Michel, L. Immunocytochemistry of the cell nucleus. *Electron Microsc Rev* **3**, 301-353, doi:10.1016/0892-0354(90)90006-e (1990).

201 Goulet, I., Boisvenue, S., Mokas, S., Mazroui, R. & Cote, J. TDRD3, a novel Tudor domain-containing protein, localizes to cytoplasmic stress granules. *Hum Mol Genet* **17**, 3055-3074, doi:10.1093/hmg/ddn203 (2008).

202 Renvoise, B. *et al.* Distinct domains of the spinal muscular atrophy protein SMN are required for targeting to Cajal bodies in mammalian cells. *J Cell Sci* **119**, 680-692, doi:10.1242/jcs.02782 (2006).

203   Karpova, T. & McNally, J. G. Detecting protein-protein interactions with CFP-YFP FRET by acceptor photobleaching. *Curr Protoc Cytom* **Chapter 12**, Unit12 17, doi:10.1002/0471142956.cy1207s35 (2006).

204   Clement, C. G. & Truong, L. D. An evaluation of Congo red fluorescence for the diagnosis of amyloidosis. *Hum Pathol* **45**, 1766-1772, doi:10.1016/j.humpath.2014.04.016 (2014).