RNA-RNA Interactions in RNP Granule Assembly

By

Briana Nicole Van Treeck

B.S. University of New Mexico, Albuquerque NM, 2012

A thesis submitted to the faculty of the graduate school of the University of Colorado in partial fulfillment of the requirement for the degree of Doctor of Philosophy,

Department of Biochemistry, 2018.
This thesis for the Doctor of Philosophy degree entitled:

RNA-RNA Interactions in RNP Granule Assembly

authored by Briana Nicole Van Treeck

has been approved for the Department of Biochemistry

_____________________________
Dr. Roy Parker, Committee Chair

_____________________________
Dr. Chris Link, Second Reader

_____________________________
Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above-mentioned discipline.
ABSTRACT

Van Treeck, Briana Nicole (Ph.D. Biochemistry)
RNA-RNA Interactions in RNP Granule Assembly
Thesis directed by Professor Roy Parker

Ribonucleoprotein (RNP) granules, which are composed of numerous RNAs and proteins, are ubiquitous features of eukaryotic cells. This thesis investigates the assembly mechanisms of P-bodies and stress granules, two cytoplasmic granules that rely on RNAs unengaged in translation to form. Previous studies in the field have focused on the roles of proteins, and specifically their intrinsically disordered regions (IDRs), in targeting components and driving granule assembly. However, domain analysis of four proteins illustrates that having an IDR is neither sufficient nor necessary for targeting to P-bodies. Although IDRs have been shown to be important in some contexts, these results suggest that IDRs are not solely responsible for RNP granule formation. With this in mind, the assembly roles of RNA were considered. RNAs self-assemble robustly in vitro, aided by crowders and salt. Remarkably, when total yeast RNA is allowed to assemble in vitro, the RNAs that are enriched in these assemblies are the same RNAs that are enriched in stress granules in vivo. This, in addition to observations compiled from the literature, suggests that RNA-RNA interactions may play an important role in the assembly of many RNP granules. Preliminary interrogation into RNA sequences important for RNA localization to stress granules has revealed that the addition of certain sequences can significantly alter RNA localization under stress despite total length being held constant. This argues for the presence of stress granule targeting elements within RNA, although further mechanistic investigation is required. Taken together, this thesis works towards understanding RNP assembly and the mechanisms that target their components.
DEDICATION

This thesis is dedicated to my dad, William Van Treeck. He has been a supportive and loving father for the entirety of my life, including my time here at the University of Colorado. After my mother passed away in 2001, my dad was the exemplary human being that I needed and admired. He has sacrificed so much in his life so that I could succeed in mine. I love you!
ACKNOWLEDGEMENTS

It’s rare that a chapter of one’s life gets summed up and packaged so tightly as the PhD experience. It is a time full of great joys, various disappointments and incredible friendships. There are so many people that have helped me be during my time in the biochemistry program at CU.

First and foremost, I have to thank all of the amazing people in the Parker Lab. It is rare to have the opportunity to be surrounded by such talented, intelligent and generous people. Specifically, I would like to thank the three women that keep everything running smoothly: Anne Webb, Denise Muhlrad, and Carolyn Decker. In addition to all of her roles in the lab, Carolyn has been my bay-mate for the entirety of my time in the Parker lab. She was my mentor when I rotated in the lab and is now my greatest role model. Thanks to all of the post-docs (Ross Buchan, Yuriko Harigawa, Sarah Mitchell, Jennifer Garcia, Amber English, Laura Mizoue, Robert Walters, Bhala Rao, Stephanie Moon, Anthony Khong, and James Burke) for being leading examples of high-quality scientists. I would like to especially thank Stephanie Moon, for constantly providing me with opportunities to further my career and Anthony Khong for endlessly bouncing ideas and providing me with snacks from all over the world. Bhala Rao also deserves special acknowledgment for being my pun pal. Thanks to my fellow Parker Lab graduate students: Saumya Jain, Sid Shukla, David Protter, Josh Wheeler, Tyler Matheny, Evan Lester, Devin Tauber and Giulia Corbet. It has been an amazing journey developing my scientific skills with you all. Not only have you all been good co-workers, but you have also been amazing friends. Thank you to Karl and Dylan for
keeping our shelves well stocked, and thank you to Daniel Henthorn, Jay Moore and Gabe Tauber for giving me the opportunity to practice my mentoring skills. I have learned so much by working with you all. Gabe, I have had the privilege of “mentoring” you the longest, although I can hardly say you need my help. You have an extremely bright future ahead; I can’t wait to see what you discover. Finally, thanks to Olke Uhlenbeck for providing new insight into our work during lab meetings.

Of course, I cannot forget the acknowledgment of our captain, Roy Parker. Roy has opened my eyes to what it means to live and breathe science. Through a lot of writing and speaking practice, he has given me confidence to convey my ideas in a clear and concise manner. Confidence through competence. I remember when I was told I could join the lab, Roy said, “You’re coachable, and I think you will be fun to work with”. I can truly say that I have grown so much as a scientist and had a lot of fun along the way. Thank you for everything!

Special thanks to my cohort: Sid Shukla, Abigail Powell, Nicholas Parsonnet, David Protter, Alex Hopkins, Justine Debellius, Arden Doerner, Joe Cardiello, Yu Han, Marie Balboa and Christine Sparks. I am so glad that I got to start my graduate career with such a close-knit group of people.

Thanks to the core facilities at the Jennie Smoly Caruthers Biotechnology Building. Theresa Nahreini and Nicole Kethley do a great job keeping the highly-trafficked tissue culture facility running smoothly. Joe Dragovon is one of the most patient people I have ever met, always willing to provide help to anyone analyzing their
microscopy images. Thanks to Jamie Prior Kershner, and now Amber Scott for doing such an amazing job running the Sequencing Core Facility.

I highly appreciate the thoughts and comments provided by my committee throughout the years. I am so privileged to have Tom Cech, Chris Link, Sabrina Spencer, and John Rinn to provide new insights into my project and my career.

Thanks to the closest friends I have ever had in my life: Abby, Dan, Dave, Joe, Jeanne, Tony, Meg, Zach, Christina, Amy, Nick, Deborah and Jennifer. Thanks especially to Abby for being my adventure buddy and always pushing me to do fun and challenging things outside of lab. Thanks to Deborah for being a close friend, despite the distance that now separates us. And thanks to Jennifer for being such a strong and compassionate person. I’m so proud of the strides you have made, especially in the last year.

Thanks to all of the roommates I have had during my stint in Boulder: Nick, Danny, Marie, Jeff, Tony, Meg, Dave, Joe and Jeanne. It is a huge relief to come home to a house full of people I enjoy. Special thanks to Nick for making the journey to Boulder with me. Starting as my tennis partner in New Mexico, we have become life-long friends.

As cliché as it may seem, I also have to thank my three pets: Castor, Pollex and Citra. Grad school is emotionally trying at times and these furry creatures have
never failed to cheer me up and remind me of my priorities, top of which is mental and physical health.

I cannot forget the acknowledgment of my family. I am so lucky that my family is close-knit. My dad’s mantra was always, “work hard and good things will happen.” He also advised us to find what we loved early in life and pursue it. Having him so close during graduate school has been a true blessing. My brother, Devin, has taught me what it means to be a stand-up human being. He brings such optimism to his life that I try to emulate. My sister, Katie, continues to inspire me, even though she is a complete klutz (like me). When Katie was born, I believed she was a gift given straight to me and I still believe that to this day.

Finally, I want to express how much I have appreciated the love and support of Jake Flood throughout our grad school careers. It has been a great comfort to be able to come home to someone that understands the unique set of challenges that grad school imposes. He is so passionate about his science, so deliberate in all of his actions, and such a well-rounded human being. I simply adore you, Jake.
The poem below is dedicated to the Parker Lab. As a lab we decided on a list of seven attributes, all of which happen to start with ‘C’, that we believe represent us as a whole. These traits were visually represented in our lab crest. I will always strive to maintain these qualities no matter where life takes me.

What else you got?
Briana Van Treeck

The question lingers in the air
But I had reached my final slide
It echoes loud: What else you got?
So I had to search inside

Well, I’ve got curiosity
Had it since I was little
I want to know how and why
And can’t stand an unanswered riddle

And I’ve got creativity
I can think outside that box
I’m always coming up with new ideas
A fraction of which may rock your socks

“What else you got?” Roy said sighing
Don’t worry, I’ve also got the courage of a lion

The courage to stand by my ideas
When I know that they are good
And, the courage to let them go
When I know that I should

And I’m always ready for a challenge
For I myself have a challenging mind and spirit
Give me a mountain, a race and a problem
And I’ll climb it, I’ll finish it, I’ll solve it

Not only that, I’m critical
I won’t believe something simply because I’m told
To fit the puzzle pieces together
I’ll consider all the data my mind can hold

“What else you got” Roy said with interest
So I continued with my fine, fine list
Man oh man have I got community
I’ve got it through and through
The people here are like no other
We make quite the close-knit crew

And sometimes I have clarity
No, I don’t know what it all means, and maybe never will
But I can organize my thoughts
With GARI’s help to fill

And one last time I hear the words
“What else you got?” he says with a huff
“That’s all I got.” I say shrugging
And with a smile he replies, “That’s enough”

So the next time Roy asks, “What else you got?”
I hope you know, you’ve got a lot!
## TABLE OF CONTENTS

### CHAPTER 1: Introduction

1.1. Summary .......................................................................................... 1
1.2. Background ...................................................................................... 2
1.3. Intermolecular RNA-RNA interactions contribute to RNP granule formation .................................................................................. 3
1.4. Biochemical nature of intermolecular RNA interactions .......................................................................................... 8
1.5. A gradient of promiscuous to specific RNA-RNA interactions .................................................................................. 11
1.6. Are RNA-RNA interactions a general feature of diverse RNP granules? .................................................................................. 13
1.7. A combinatorial model for RNP granule assembly .................................................................................. 17
1.8. Regulation of the equilibrium between monomeric and multimeric RNPs .................................................................................. 23
1.9. RNA-RNA interactions and disease .......................................................................................... 26
1.10. RNA self-assembly: Perspectives and Retrospectives .................................................................................. 27

### CHAPTER 2: RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome

2.1. Abstract .......................................................................................... 28
2.2. Introduction ...................................................................................... 29
2.3. Results ............................................................................................ 30
2.4. Discussion ....................................................................................... 33
2.5. Materials and Methods ..................................................................... 53

### CHAPTER 3: Disentangling RNA properties important for recruitment to stress granules

3.1. Abstract .......................................................................................... 66
3.2. Introduction ...................................................................................... 67
3.3. Results ............................................................................................ 68
3.4. Discussion ....................................................................................... 72
3.5. Materials and Methods ..................................................................... 85

### CHAPTER 4: Intrinsically Disordered Regions, From Granule Assembly to Disease

4.1. Abstract .......................................................................................... 100
4.2. Introduction ...................................................................................... 101
4.3. Results ............................................................................................ 102
4.4. Discussion ....................................................................................... 105
4.5. Materials and Methods ..................................................................... 121

### CHAPTER 5: Summary and Future Directions

5.1. Summary .......................................................................................... 132
5.2. Looking Forward ............................................................................... 133
**TABLES**

**CHAPTER 2**

Table 2.1. Sequence of Materials………………………………………………………………………………58

**CHAPTER 3**

Table 3.1. Sequence composition of 300-base inserts…………………………………………………………84  
Table 3.2. Stellaris single molecule FISH pooled probes………………………………………………………90  
Table 3.3. Plasmids used in Chapter 3……………………………………………………………………………96  
Table 3.4. Oligos used for constructing plasmids……………………………………………………………………98

**CHAPTER 4**

Table 4.1. Foci formation from overexpression of IDRs correlates with stable amyloid-like assemblies and not with phase separation………………………………………………………………………………115  
Table 4.2. Plasmids used in chapter 4………………………………………………………………………………127  
Table 4.3. Oligos used for the construction of plasmids……………………………………………………………131

**APPENDIX**

Table A.1. Self-assembly conditions of proteins *in vitro*…………………………………………………………153  
Table A.2. Self-assembly conditions of RNA *in vitro*………………………………………………………………156  
Table A.3. Self-assembly conditions of RNA and protein *in vitro*………………………………………………158
FIGURES

CHAPTER 1

Figure 1.1. Protein-protein interactions promote RNP granule formation.........................4
Figure 1.2. RNA contributes to RNP granule formation..............................................7
Figure 1.3. RNA-RNA interactions can be promiscuous or specific..............................14
Figure 1.4. A four-phase model of RNP granule assembly incorporates protein-protein and RNA-RNA interactions and has specific implications........................................18
Figure 1.5. Concentrations and properties of granule components influence the formation and properties of granules themselves.................................................................21
Figure 1.6. A model for the modulation of interactions between RNPs.........................24

CHAPTER 2

Figure 2.1. RNAs remain in stress granules following robust RNase treatment.............34
Figure 2.2. RNA-RNA base pairing influences RNA localization.................................36
Figure 2.3. Various RNAs self-assemble in vitro..........................................................38
Figure 2.4. Total Yeast RNA droplets and tangles have characteristic properties........39
Figure 2.5. RNA binding proteins are recruited to RNA assemblies............................41
Figure 2.6. Homopolymer droplets have distinct biochemical properties....................43
Figure 2.7. Assemblies of yeast total RNA in vitro largely recapitulate the stress granule transcriptome ..........................................................45
Figure 2.8. RNAs in self-assemblies in vitro largely recapitulate the stress granule transcriptome ........................................................................46
Figure 2.9. A four-phase model of stress granule assembly.........................................48
Figure 2.10. Pathogenic dipeptides increase RNA assembly........................................50
Figure 2.11. Pathogenic dipeptides increase RNA assembly in vitro..........................51

CHAPTER 3

Figure 3.1. smFISH as a tool for assaying RNA localization to stress granules.............73
Figure 3.2. smFISH is not compatible with plasmid transfection.................................75
Figure 3.3. smFISH is compatible with genomic integration........................................76
Figure 3.4. Insertion of complementary regions to stress granule enriched RNAs alters localization of some reporter constructs under stress.............................................78
Figure 3.5. Correlation between stress granule recruitment of luc-N-ORF or luc-A-3UTR reporter RNAs and their complementary targets................................................81
Figure 3.6. Knock-down of NORAD has no effect on recruitment of luc-N-ORF...........82

CHAPTER 4

Figure 4.1. IDRs are neither sufficient nor necessary for P-body localization..........106
Figure 4.2. Model of RNP granule assembly and contributions of IDRs...................108
Figure 4.3. Overexpression of various IDRs can drive the assembly of SDS-resistant aggregates in vivo..................................................................................110
Figure 4.4. Galactose induction shows increased SDS-resistant aggregates over time that correlate with microscopically visible foci.......................................................116
Figure 4.5. RNA-binding has no effect on the formation of SDS-resistant assemblies formed by Pub1IDR-GFP-MS2.
CHAPTER 1: INTRODUCTION

1.1. Summary

Eukaryotic cells contain large assemblies of RNA and protein, referred to as ribonucleoprotein (RNP) granules, which include cytoplasmic P-bodies, stress granules, neuronal and germinal granules, as well as nuclear paraspeckles, Cajal bodies and RNA foci formed from repeat expansion RNAs. Recent evidence argues that intermolecular RNA-RNA interactions play a role in forming and determining the composition of certain RNP granules. We hypothesize that intermolecular RNA-RNA interactions are favored in cells yet are limited by RNA-binding proteins, helicases, and ribosomes, thereby allowing normal RNA function. An over-abundance of intermolecular RNA-RNA interactions may be toxic since perturbations that increase RNA-RNA interactions such as long repeat expansion RNAs, arginine-containing dipeptide repeat polypeptides, and sequestration or loss of abundant RNA-binding proteins can contribute to degenerative diseases.
1.2. Background

Eukaryotic cells contain a growing number of non-membrane bound organelles consisting of RNA and protein, which we will generically refer to as RNP granules. RNP granules include stress granules, P-bodies, germ granules, and neuronal granules in the cytoplasm as well as paraspeckles, the nucleolus and Cajal bodies in the nucleus (Buchan, 2014; Kiebler and Bassell, 2006; Voronina et al., 2012; Fox et al., 2002; Gall, 2000). RNP granules are ubiquitous and conserved in eukaryotes. Understanding the assembly mechanisms of these non-membrane bound organelles offers opportunity for new insight into cellular organization and regulation.

Historically, studies aimed at determining the assembly mechanisms of RNP granules have focused on proteins. In numerous cases mutating or deleting key RNP granule proteins reduces RNP granule formation (reviewed in Buchan, 2014; Protter and Parker, 2016). For example, the Edc3 protein strongly enhances P-body formation in *Saccharomyces cerevisiae* (Decker et al., 2007), the G3BP protein is required for stress granule formation in mammalian cells during oxidative stress (Kedersha et al., 2016; Tourrière et al., 2003), and the MEG1 and MEG3 proteins are required for P-granule formation in *C. elegans* (Wang et al., 2014). These proteins all bind RNA and are thought to connect individual RNPs into larger assemblies through protein-protein interactions.

Four different types of protein-protein interactions promote RNP granule formation. Some are stereospecific interactions between well-folded domains (*Figure 1.1A*), such as the dimerization of G3BP in stress granule assembly (Tourrière et al.,
Figure 1.1. **Protein-protein interactions promote RNP granule formation.** Proteins can interact in four ways that contribute to the multivalency of RNP granules. (A) Classical stereospecific interactions between well-folded domains on proteins. (B) Specific proteins bind short linear motifs (SLiMs), conserved sequences within intrinsically disordered regions (IDRs). Often, structure emerges from the IDR upon binding. (C) IDRs can also interact specifically with other IDRs through interaction domains with key amino acid characteristics. For example, LARKs are short repeated sequences in IDRs that contain tyrosine and can interact weakly with other LARKs on neighboring proteins. (D) IDRs can also provide promiscuous interactions, potentially through $\pi-\pi$ or cation-$\pi$ interactions, which enhance assembly once components are at high-local concentrations.
2003) or Edc3 in P-body assembly (Ling et al., 2008). Some granule-promoting interactions occur through intrinsically disordered regions (IDRs) of proteins, which are enriched in RNP granules (Decker et al., 2007; Kato et al., 2012; Reijns et al., 2008). Granule assembly can be promoted by the interactions of conserved short linear motifs (SLiMs) within IDRs with the surface of other well-folded protein domains (Figure 1.1B; reviewed in Jonas and Izaurralde, 2013). Moreover, short repeated sequences in IDRs containing tyrosine, referred to as low-complexity aromatic-rich kinked segments (LARKs), are enriched in RNP granule components and may form local structures that interact with other LARK containing IDRs, thereby providing additional interactions (Figure 1.1C, Hughes et al., 2018; Murray et al., 2017). Finally, IDRs can also interact with other proteins in a promiscuous manner (Figure 1.1D), perhaps through cation-π or π-π interactions (Vernon et al., 2018), and by being tethered to a specific interaction module can enhance the assembly of RNP granules (Protter et al., 2018).

Additional evidence that protein-protein interactions contribute to RNA granule formation is that many purified RNP granule proteins, and/or their IDRs, can undergo self-association in vitro in processes referred to as liquid-liquid phase separation (LLPS) and/or hydrogel formation (Appendix Table A.1.). Although many proteins can undergo LLPS in vitro, in only a few cases have these assemblies been shown to correlate with RNP granule formation within cells. For example, the interactions of Dcp2, Pdc1, and Edc3 have been shown to promote P-body formation in S. pombe, and also promote LLPS in vitro (Fromm et al., 2014). The self-partitioning of RNP
granule components *in vitro* is consistent with these protein-protein interactions contributing to RNP granule formation.

RNAs are also required for the formation of some RNP granules. For example, the transcription, and presence, of *NEAT1* lncRNA is required for paraspeckle formation (Clemson et al., 2009; Mao et al., 2010). Similarly, stress granules and P-bodies require non-translating mRNAs for their formation (Liu et al., 2005; Sheth and Parker, 2003; Teixeira et al., 2005; Pillai et al., 2005) since inhibiting translation initiation increases P-body and stress granule assembly while trapping mRNAs in polysomes reduces P-body and stress granule assembly (Buchan et al., 2008; Teixeira, 2005; Kedersha et al., 1999; Kedersha et al., 2000). Similarly, the formation of RNA foci from repeat expansion RNAs only occurs when the RNA reaches a critical length (Lee et al., 2013; Wojciechowska and Krzyzosiak, 2011). A common model to explain how RNAs promote RNP granule formation is that RNAs provide scaffolds for multivalent RNA-binding proteins, which then, through homotypic or heterotypic protein-protein interactions, connect individual RNP complexes to form higher-order assemblies (Figure 1.2A).

Herein, we review evidence suggesting intermolecular RNA-RNA interactions directly promote the assembly of RNP granules, leading to a combinatorial model in which RNP granule formation is the result of a summation of protein-protein, protein-RNA and RNA-RNA interactions. This has implications for the intracellular conditions that affect RNP granule formation, how these assemblies are regulated, and the steady state nature of RNP complexes.
Figure 1.2. RNA contributes to RNP granule formation. (A) RNA can serve as a scaffold for multivalent RNA-binding proteins. These proteins can then interact with each other as described in Figure 1.1. (B) RNAs can interact non-specifically with each other through Watson-Crick base-pairing, non-canonical base-pairing, and helical stacking. (C) Molecular crowding has a greater impact on the effective concentration of larger molecules, where the available solvent is much more reduced for larger molecules. This is visualized by the two panels; the left illustrates the accessible solvent to a 55 kDa protein (dark red area) and the right a 7.5 kb RNA (dark blue area). Darker coloring denotes accessible locations of the center of the protein or RNA, respectively. (D) The association of larger molecules and complexes is favored by depletion attraction, which is a force only exerted in crowded conditions. The association of two larger molecules decreases the excluded volume and increases the entropy of the smaller macromolecules also in solution. Counterintuitively, the aggregation of larger complexes is entropically favored in crowded environments like the cell.
1.3. Intermolecular RNA-RNA interactions contribute to RNP granule formation

One argument for intermolecular RNA-RNA interactions promoting RNP granule formation is the robust self-assembly of RNA \textit{in vitro} (Appendix Table A.2). Specifically, multiple RNAs, or mixtures of RNAs, are capable of protein-free self-assembly, including all four RNA homopolymers (Aumiller et al., 2016; Van Treeck et al., 2018; Chapter 2), total yeast RNA (Van Treeck et al., 2018; Chapter 2), specific mRNAs (Boundedjah et al., 2012; Langdon et al., 2018), or RNAs corresponding to repeat expansion RNAs (Jain and Vale, 2017). The self-assembly of RNA can occur at concentrations as low as 2 µg/mL (Appendix Table A.2), which is lower than typical concentrations used to demonstrate LLPS of RNA-binding proteins (0.5-10 mg/ml; Appendix Table A.1). Moreover, the concentration of exposed mRNA open reading frames during a stress response, when ribosomes run-off mRNAs and stress granules form, is estimated to be \( \sim \)150-800 µg/mL in yeast and 80-300 µg/mL in human cells, which are concentrations at which RNAs will robustly self-assemble \textit{in vitro} under physiological salt and polyamine concentrations (Van Treeck et al., 2018; Chapter 2).

Three observations suggest that RNA self-assembly \textit{in vitro} is relevant to \textit{in vivo} formation of certain RNP granules. First, protein-free yeast RNAs assembled \textit{in vitro} under physiologically relevant salt and crowding conditions largely recapitulate the yeast stress granule transcriptome (Van Treeck et al., 2018, Chapter 2), which is biased towards longer RNAs (Khong et al., 2017). Although it remains possible that there is an unrecognized code in these self-assembling RNAs, the simplest interpretation is that longer RNAs have more sites for intermolecular RNA-RNA interactions, and therefore have enhanced self-association both in cells and \textit{in vitro}.
Second, pathogenic repeat expansion RNAs, are more prone to self-partitioning \textit{in vitro} than length-matched counterparts (Jain and Vale, 2017). This is consistent with repeat expansion RNAs forming highly structured hairpins (Sobczak et al., 2003), the base-pairs of which could easily be rearranged to form between RNAs, creating an RNA network (Jain and Vale, 2017). Moreover, the ability of repeat expansion RNAs to self-assemble \textit{in vitro} correlates with their ability to form RNA foci in cells in terms of their length requirements and sensitivity to increased ammonium acetate or doxorubicin, a nucleic acid intercalator (Jain and Vale, 2017). Third, there is a correlation of mRNA self-assembly \textit{in vitro} with the localization of mRNAs in specific RNP granules in the filamentous fungus \textit{Ashbya gossypii}. Specifically, the \textit{SPA2} and \textit{BNI1} mRNAs, which are enriched in an RNP granule at the growth-tip preferentially self-assemble together \textit{in vitro}, while \textit{CLN3} mRNA, which is found in a nuclear associated RNP granule, preferentially assembles with itself (Langdon et al., 2018).

Additional observations are also consistent with the formation of stress granules, and potentially other RNP granules, being partially driven by promiscuous RNA-RNA interactions. Specifically, when naked RNA is injected into the cytosol, it triggers the formation of stress granules (Mahadevan et al., 2013). Similarly, transfection of the luciferase mRNA (Bounedjah et al., 2012) or short RNAs prone to forming G-quadruplexes into cells promotes stress granule formation (Fay et al., 2017). Notably, even electroporated ssDNA will nucleate stress granules, which could be explained by ssDNA base-pairing with RNAs, or ssDNA serving as a scaffold for granule-promoting RNA-binding proteins that can also bind ssDNA (Bounedjah et al., 2014). In addition, stress granule formation is sensitive to the osmotic strength of the
cell in a manner that correlates with RNA-RNA interactions and not protein-protein interactions. Specifically, hyper-osmotic stress, which increases intracellular salt concentrations and enhances RNA-RNA interactions, promotes stress granule formation (Bounedjah et al., 2012). In contrast, hypo-osmotic stress, which lowers intracellular salt and diminishes shielding of the RNA’s negative backbone, results in rapid stress granule disassembly (Bounedjah et al., 2012). In contrast, LLPS of RNA-binding proteins or their IDRs in vitro is typically increased by lower salt, and decreased by higher salt concentrations (Appendix Table A.1). Together, these observations suggest that intermolecular RNA-RNA interactions can contribute to RNP granule assembly.
1.4. Biochemical nature of intermolecular RNA interactions

The biochemical interactions that drive self-assembly of RNA in vitro, and in cells, are of three types (Figure 1.2B-D). First, both Watson-Crick and non-Watson-Crick interactions between bases promote both intra- and intermolecular RNA-RNA interactions. Second, base stacking, either between single-stranded regions, or co-axial stacking of helices can also promote trans RNA-RNA interactions (Zanchetta et al., 2008). Finally, since RNAs are roughly an order of magnitude larger than their encoded proteins, RNA self-assembly may be enhanced by the crowded cellular environment (Ellis, 2001), which has a greater effect on the self-association of larger molecules or assemblies (Figure 1.2C, D, Marenduzzo et al., 2006).

These observations suggest that whenever there are high concentrations of exposed RNA, intermolecular RNA-RNA interactions can form, therefore contributing to higher-order RNA assemblies. Such events could occur during a stress response when large amounts of open reading frames are freed from ribosomes, or locally at sites of proliferative transcription. Given the degenerative nature of RNA-RNA base pairing, and the size of RNAs, such interactions need not be specific. For example, the average length of an mRNA in a mammalian stress granule is 7.5 kb (Khong et al., 2017), and by a sliding window contains ~7500 hexamers, thereby allowing \((7500)^2\) possible six base-pair interactions with another 7.5 kb RNA. Since one out of 4096 hexamers can perfectly base-pair, any two random 7.5 kb RNAs in stress granules would therefore be predicted to have approximately 14,000 possible six base interactions of perfect complementarity simply by chance. Even if ~99% of these possible sites of interaction are hidden by secondary structures or bound proteins, two
random 7.5 kb mRNAs would be expected to have over 100 potential sites of interaction. This ignores the possible interactions from shorter helices, partial matches, non-canonical base-pairs, base stacking and crowding effects. Since large RNAs can have multiple interactions, any given individual interaction between RNAs can be weak and transient and still, through a summation of multiple weak interactions, form a stable assembly (Banani et al., 2016).
1.5. A gradient of promiscuous to specific RNA-RNA interactions

Cells appear to utilize both promiscuous and specific RNA-RNA interactions in the formation of RNP granules. Stress granules, which form rapidly during a stress response and are heavily biased towards longer mRNAs (Khong et al., 2017), may simply form through random associations between RNAs (Figure 1.3A, Van Treeck et al., 2018; Chapter 2). Similarly, the retention of long mRNAs in the germ plasm of Drosophila is proposed to occur, at least in part, through essentially random base-pairing to multiple piRNAs that reside within the germ plasm simply because longer RNAs have more sites for possible interactions with the piRNAs (Vourekas et al., 2016). In contrast, specific base-pairing in trans between oskar mRNAs, or bicoid mRNAs, is required for their recruitment to RNP granules during Drosophila oocyte development (Figure 1.3B, Jambor et al., 2011; Ferrandon et al., 1997). mRNAs may also have structures that limit trans intermolecular interactions and thereby impart specificity to the assembly of RNP granules. This latter possibility is suggested by the observation that a mutation altering the structure of CLN3 mRNA in Ashbya gossypii increases its interaction with SPA2 and BNI1 mRNAs in vitro and the colocalization of the mRNAs in cells (Langdon et al., 2018).
Figure 1.3. RNA-RNA interactions can be promiscuous or specific. (A) Promiscuous interactions may be prevalent in granules containing a diverse set of RNAs in a high local concentration. Here, RNAs are predicted to assemble with a variety of interactions facilitated by the interaction capabilities of an RNA with any other RNA. An example of this may be the formation of stress granules, in which cells experience a large influx of free RNA following ribosomal run-off. (B) RNA-RNA interactions contributing to assembly can also be specific. Two examples have been described in Drosophila development in which specific RNA-RNA homodimers of either bicoid or oskar RNAs are important for RNP granule assembly. (C) Transcription sites may be a common location of RNP granule assembly, driven in part by the high-local concentration of RNA. As NEAT1 is transcribed, for example, a high local concentration of NEAT1 RNA is achieved. We hypothesize that newly transcribed NEAT1 RNA is capable of forming interactions with neighboring NEAT1 RNAs. As transcripts are released from transcription, a paraspeckle remains. Mature paraspeckles have a clear orientation with the middle of the RNA and certain proteins found in the center, and RNA ends oriented on the outside.
1.6. Are RNA-RNA interactions a general feature of diverse RNP granules?

The prevalence and stability of RNA-RNA interactions raise the possibility that intermolecular RNA-RNA interactions contribute to multivalent assemblies whenever such assemblies contain a high local concentration of RNA molecules. For example, several observations suggest that the formation of paraspeckles, a nuclear RNP granule containing approximately 50 copies of the long isoform of \textit{NEAT1} lncRNA (Chujo and Hirose, 2017), may be instigated by intermolecular RNA-RNA interactions that occur during transcription (Figure 1.3C). First, paraspeckles form during the transcription of \textit{NEAT1}, require \textit{NEAT1} for their assembly, and are sensitive to RNase once formed (Fox et al., 2005; Prasanth et al., 2005). Based on the volume of a paraspeckle, 50 copies of the 23 kb \textit{NEAT1} RNA within paraspeckles (Chujo et al., 2017) equates to a \textit{NEAT1} concentration of approximately 1 mg/mL, which is well above the concentrations for RNA-based self-assembly under physiological conditions \textit{in vitro} (Appendix Table A.2). Interestingly, 1,6 hexanediol, an aliphatic alcohol thought to disrupt some weak protein-protein interactions, causes the loss of the NONO protein from paraspeckles while \textit{NEAT1} foci remain intact, indicating the formation of \textit{NEAT1} RNA foci is independent of NONO (Yamazaki et al., 2018). Finally, numerous RNA-RNA interactions have been identified within the \textit{NEAT1} RNA both \textit{in vivo} (Lu et al., 2016) and \textit{in vitro} (Lin et al., 2018), and although these interactions are assumed to be intramolecular, in high local concentrations, such as the site of transcription or within paraspeckles, these interactions may also be intermolecular and contribute to paraspeckle formation.
We suggest that intermolecular RNA-RNA interactions should be expected to occur in any biological context with high local concentrations of RNA, unless actively limited by the cell (see below). Such situations would include transcription sites with the propensity for intermolecular interactions increased by high rates of transcription, longer transcripts, and specific sequences prone to intermolecular interactions. Other possible contexts where intermolecular RNA-RNA interactions could be important for subcellular assemblies include the nucleolus, which is dependent on rRNA transcription for its formation (Falahati et al., 2016), Barr bodies, where the formation of an XIST RNA-protein complex coating the inactive X chromosome could be promoted by intermolecular interactions between XIST molecules (Lu et al., 2016; da Rocha and Heard, 2017), and viral RNA synthesis factories, which have a high local concentration of RNA (Nikolic et al., 2017).

One intriguing hypothesis is that RNA-RNA interactions could also play a role in the function of eRNAs, which are produced at both enhancers and super-enhancers, are important for enhancer function, and stimulate transcription proportional to their abundance (Hnisz et al., 2017). However, how eRNAs enhance transcription is not clear. A recent model for enhancer and super-enhancer function is that the concentration of transcription factors and eRNAs at these sites allows for the formation of a local phase separation, which leads to downstream transcriptional activation (Hnisz et al., 2017). Given the propensity of RNAs to interact in trans, one prediction is that eRNA-eRNA interactions might play a role in the assembly of this phase transition at enhancers and thereby trigger downstream transcriptional activation.
1.7. A combinatorial model for RNP granule assembly

Given the diversity of interactions promoting RNP granule assembly, we suggest a “four-phase” model wherein RNP granules form when the summation of protein-protein, protein-RNA and RNA-RNA interactions increase over a threshold for assembly (Figure 1.4A). An assembly diagram containing four unique regions illustrates this model. Increasing protein-protein interactions leads to RNP granule assembly as the cellular environment moves up on the Y-axis, to a position in which protein-protein interactions are sufficient to drive assembly. Conversely, increasing RNA-RNA interactions, through rapid influxes of RNA, or the production of RNAs with increased propensity to assemble can lead to formations primarily driven by RNA. This model also provides context for why disruption of protein-protein interactions could disrupt formation of a primarily RNA-driven assembly.

The interactions that allow an RNP granule to assemble should be expected to vary between conditions and granule types. Specifically, one expects that the relative importance of protein-protein to RNA-RNA interactions to vary between different types of RNP granules (Figure 1.4B). For example, stress granules that form from a rapid loss of mRNAs from translation and preferentially recruit long mRNAs may utilize a significant amount of RNA-RNA interactions in their formation. Similarly, RNA foci formed from repeat expansion RNAs with a high propensity to base pair would be expected to be primarily driven by RNA-RNA interactions. In contrast, RNP granules with a low concentration of RNAs, and highly efficient self-assembling proteins, would
Figure 1.4. A four-phase model of RNP granule assembly incorporates protein-protein and RNA-RNA interactions and has specific implications. (A) RNP granules can form through either RNA or protein dominated assembly pathways, or through combinations of these pathways. Increasing key protein-protein interactions can shift monomeric RNPs upwards into an RNP granule regime (black arrow). Increasing RNA-RNA interactions (yellow arrow) can stimulate assembly, but the depletion of assembly RBPs (red arrow) can prevent assembly, even in conditions where RNA-RNA interactions are increased. (B) RNP granules can have different requirements for assembly. The relative contributions of RNA-RNA or protein-protein interactions is expected to vary from one type of RNP granule to the next. Protein-driven granules are expected to be highly influenced by the overexpression or deletion of key protein components. These granules are also expected to be regulated by post-translational modifications and chaperones. In contrast, granules primarily driven by RNA-RNA interactions are predicted to have a high local concentration of RNA, with an enrichment for long RNAs or RNAs with stable, specific interactions. In addition, enzymes acting on RNAs, such as helicases, would be expected to modulate the dynamics of RNA-based assemblies. Most granules will reside somewhere in the middle, with some characteristics properties from both sides of the spectrum.
be predicted to be more dependent on protein interactions. Moreover, the specific protein-protein interactions that contribute to RNP granule assembly can vary. For example, P-body assembly in yeast can be driven by different combinations of proteins depending on the genetic context (Rao and Parker, 2017). Similarly, the deletion of key stress granule proteins in mammalian cells can abrogate stress granule assembly in some stresses and not in others (Kedersha et al., 2016). Thus, one anticipates that the underlying proteome and transcriptome of individual cells allows varying sets of interactions to drive RNP granule assembly in different cell types and biological contexts.

The interactions that allow an RNP granule to assemble should be expected to vary between conditions and granule types. Specifically, one expects that the relative importance of protein-protein to RNA-RNA interactions to vary between different types of RNP granules (Figure 1.4B). For example, stress granules that form from a rapid loss of mRNAs from translation and preferentially recruit long mRNAs may utilize a significant amount of RNA-RNA interactions in their formation. Similarly, RNA foci formed from repeat expansion RNAs with a high propensity to base pair would be expected to be primarily driven by RNA-RNA interactions. In contrast, RNP granules with a low concentration of RNAs, and highly efficient self-assembling proteins, would be predicted to be more dependent on protein interactions. Moreover, the specific protein-protein interactions that contribute to RNP granule assembly can vary. For example, P-body assembly in yeast can be driven by different combinations of proteins depending on the genetic context (Rao and Parker, 2017). Similarly, the deletion of key stress granule proteins in mammalian cells can abrogate stress granule assembly
in some stresses and not in others (Kedersha et al., 2016). Thus, one anticipates that
the underlying proteome and transcriptome of individual cells allows varying sets of
interactions to drive RNP granule assembly in different cell types and biological
contexts.

The self-assembly of RNA and RNA-binding proteins, as well as their
interactions with each other predicts a complex set of conditions whereby the
concentrations, size, and valency of RNA and RNA-binding proteins will influence the
formation and properties of larger assemblies (Appendix Table A.3, Figure 1.5). For
example, the addition of high concentrations of RNA can inhibit the self-assembly of
RNA-binding proteins (Banerjee et al., 2017; Schwartz et al., 2013, Maharana et al.,
2018), perhaps because the RNA competes with the protein-protein interaction
surfaces. However, RNA can also enhance the assembly of RNA-binding proteins (Lin
et al., 2015; Molliex et al., 2015; Patel et al., 2015) at lower concentrations or if in
conditions that allow RNA-RNA interactions to contribute (Figure 1.5B). Similarly, the
ratios and valency of RNA-binding proteins can affect the nature of assemblies driven
by RNA-RNA interactions (Figure 1.5A).
A. Addition of monovalent RNA-binding proteins

Protein recruited

Protein competes for RNA-RNA interactions

Addition of multivalent RNA-binding proteins

Protein recruited

Protein binds RNA and provides additional interactions for assembly

B. Addition of short RNAs

RNA recruited

RNA induces a conformational change that increases protein-protein interactions

Addition of long RNAs

RNA recruited

RNA enhances assembly if conditions allow RNA-RNA interactions and/or scaffolding

RNA competes with protein-protein interactions
Figure 1.5. Concentrations and properties of granule components influence the formation and properties of granules themselves. (A) Effects of monovalent and multivalent RNA-bp on assemblies formed by RNA-RNA interactions. At low concentrations of protein, both monovalent and multivalent RNA-bp are recruited to RNA assemblies, but do not dramatically change the assembly (top left and right) (Van Treeck et al., 2018; Chapter 2). In contrast, at high concentrations a monovalent RNA-bp can inhibit assembly by competing for RNA-RNA interactions (bottom left), while a multivalent RNA-bp can enhance assembly by providing additional cross-linking interactions between RNA molecules (bottom right) (Bounedjah et al., 2014). (B) Effects of RNAs on self-assemblies of RNA-binding proteins. The addition of low concentrations of short or long RNAs should result in the RNA being effectively recruited to the assembly (top left and right). At high concentrations, both types of RNAs can also inhibit assembly by competing with the protein-protein interaction surface (bottom) (Lin et al., 2015; Schwartz et al., 2013; Maharana et al., 2018). However, RNA can also enhance the assembly of RNA-binding proteins (Banerjee et al., 2017; Molliex et al., 2015; Patel et al., 2015) by either providing a scaffold to increase the valency of RNA-protein complex interactions (Figure 1.2A), by triggering a conformational change in the protein that promotes assembly, which can occur with short RNAs (middle left), or through RNA-RNA interactions that promote assembly, which is favored for longer RNAs (middle right).
1.8. Regulation of the equilibrium between monomeric and multimeric RNPs

Given the robust self-assembly properties of RNA, we hypothesize that cellular RNAs are in a continual exchange between monomeric, oligomeric, and multimeric assemblies large enough to be seen in the light microscope and identified as RNP granules. Evidence for intermolecular interactions at scales smaller than RNP granules includes smFISH of the mRNA *DYNC1H1*, which does not show random distribution in the cytoplasm, but concentrates into clusters of 3-7 mRNAs (Pichon et al., 2016). In addition, mapping of RNA-RNA interactions *in vivo* identified over 990 diverse mRNA-mRNA interactions with a substantial fraction being between two different mRNAs (Aw et al., 2016; Lu et al., 2016; Sharma et al., 2016; Gong et al., 2018). Moreover, the imperfect binding of Alu-containing IncRNAs to mRNAs has been suggested to promote Staufen-dependent mRNA decay (Gong and Maquat, 2011). Although more work needs to be done to achieve a more comprehensive and reproducible summary of RNA-RNA interactions in the cell, 170 mRNAs have already emerged as potential interaction hubs with more than 100 documented RNA interaction partners (Gong et al., 2018).

Since the proper functioning of RNAs frequently requires a monomeric RNP, we hypothesize that cells utilize multiple mechanisms to modulate the formation of RNA-based assemblies (Figure 1.6). For example, ribosomes or monovalent RNA-binding proteins can limit RNA-RNA interactions by binding RNA sequences and restricting their availability for RNA-RNA interactions, while multivalent proteins may increase RNP aggregation (Figure 1.5A, Appendix Table A.3). This principle has been
Figure 1.6. A model for the modulation of interactions between RNPs. We hypothesize RNPs may be more prone to associate when their components are in high local concentrations, they are composed of longer RNAs, or contain multivalent RNA-binding proteins or RNAs with increased interaction propensity with other RNAs (such as in repeat expansion diseases). Disassembly may be promoted by increased recruitment of monovalent RNA-binding proteins or short RNAs that limit associations between RNPs. Active remodelers, like chaperones, can also disassemble RNP granules (reviewed in Protter and Parker 2016). In addition, helicases and ribosomes can unwind intermolecular RNA-RNA interactions and block potential RNA-RNA and RNA-protein interactions, respectively.
demonstrated both in vitro and in vivo, where increased concentrations of the monovalent RNA-binding protein YB1 inhibits stress granule formation in cells, and limits the formation of an RNA-TIA1 protein assembly in vitro (Bounedjah et al., 2014). Similarly, knockdown of the abundant RNA-binding protein TDP-43 in human cells, or its ortholog in C. elegans, leads to the accumulation of dsRNAs in either cytoplasmic or nuclear foci (Saldi et al., 2014). This suggests that the binding of proteins to RNA can limit the formation of dsRNA and intermolecular interactions between RNAs that result in RNA foci.

RNA helicases should be expected to disassemble RNA-RNA interactions and thereby limit RNP granule formation. Strikingly, the loss of the RNA helicase CGH-1 in C. elegans generates solid sheets of other P-granule components (Hubstenberger et al., 2013) and the ATPase activity of CGH-1 is necessary for P granule disassembly following extrusion from cells (Smith and Seydoux, 2018). Similarly, ATP hydrolysis mutants in the DEAD-box helicases Dhh1 or Ded1/DDx3, results in constitutive P-bodies or stress granules, respectively (Mugler et al., 2016, Hilliker et al., 2011). The disassembly role of helicases may be difficult to demonstrate in some cases as RNA helicases can also play roles in RNP granule assembly by nucleating protein-protein interactions (e.g. Hilliker et al., 2011).

Cells can also limit RNP granule formation by degrading RNAs; evidence for this effect comes from the increase in P-bodies seen in mammalian or yeast cells when mRNA decapping is blocked (Cougot et al., 2004; Sheth and Parker, 2003).
1.9. RNA-RNA interactions and disease

Several genetic perturbations linked to human disease may function by tipping the equilibrium between monomeric and multimeric RNAs (Figure 1.6). The prevalence of RNA foci in repeat expansion disease is a clear example of a toxic RNA being produced that has the potential for multiple intermolecular interactions and thereby formation of a multimeric RNA assembly in the cell (Jain and Vale, 2017). Moreover, many repeat expansion RNAs, such as ALS-linked G4C2 expansions in the C9orf72 gene, can produce polypeptides composed of dipeptide repeats through a process referred to as repeat associated non-AUG (RAN) translation (Zu et al., 2010). The most toxic of these dipeptides are either (RG)_n or (PR)_n, which can essentially function as polyamines and promote intermolecular RNA-RNA interactions (Van Treeck et al., 2018; Chapter 2), although whether this is the basis of their toxicity remains to be established. Finally, a decrease in the functional pool of abundant hnRNP proteins appears to promote intermolecular RNA-RNA interactions. For example, the sequestration of abundant RNA-binding proteins such as TDP-43 into cytoplasmic foci in numerous degenerative diseases may contribute to toxicity by allowing RNA-RNA interactions, and a stress response to dsRNA (Saldi et al., 2014).
1.10. RNA self-assembly: Perspectives and Retrospectives

One predicts that RNA-RNA interactions will continue to emerge as key players in cellular mechanisms and assemblies. While base-pairing interactions will be easiest to find, important RNA-RNA interactions that are short and transient may remain more elusive. The ubiquity of RNA-RNA interactions argues that such interactions will be prevalent, but also carefully modulated by the cell. As such, mis-regulation of RNA-RNA interactions could result in pathogenic phenotypes. An important area of research will be in defining the range of RNA-RNA interactions in cells, how cells utilize such interactions for normal function, and how they are modulated by cellular machineries.

Retrospectively, it is self-evident that the robust and degenerate self-assembly of RNA would have been an ideal crucible for the origins of self-replicating RNA. Given an abiotic source of oligonucleotides at a high enough concentration and with abundant counter-ions, essentially any collection of oligonucleotides can self-assemble into a higher-order structure with an increased local concentration of oligonucleotides (Aumiller et al., 2016; Van Treeck et al., 2018; Chapter 2). Moreover, because such assemblies concentrate RNA molecules, they can increase the formation of catalytic moieties by increasing interactions between molecules that promote chemical reactions (Strulson et al., 2012). Such RNA-based assemblies may have preferentially retained longer RNAs while, due to the absence of a membrane, allowed facile entry of oligonucleotide precursors for continued rounds of replication.
CHAPTER 2: RNA SELF-ASSEMBLY CONTRIBUTES TO STRESS GRANULE FORMATION AND DEFINING THE STRESS GRANULE TRANSCRIPTOME

2.1. Abstract

Stress granules are higher order assemblies of non-translating mRNAs and proteins that form when translation initiation is inhibited. Stress granules are thought to form by protein-protein interactions of RNA-binding proteins. Here, RNA homopolymers or purified cellular RNA are shown to form assemblies in vitro analogous to stress granules. Remarkably, under conditions representative of an intracellular stress response, the mRNAs enriched in assemblies from total yeast RNA largely recapitulate the stress granule transcriptome. This suggests that stress granules are formed by a summation of protein-protein and RNA-RNA interactions, with RNA self-assembly likely to contribute to other RNP assemblies wherever there is a high local concentration of RNA. RNA assembly in vitro is also increased by GR and PR dipeptide repeats, which are known to increase stress granule formation in cells. Since GR and PR dipeptides are involved in neurodegenerative diseases, this suggests that perturbations increasing RNA-RNA assembly in cells could lead to disease.
2.2. Introduction

Ribonucleoprotein (RNP) granules are eukaryotic non-membrane bound organelles, composed of RNA and protein. RNP granules are located in the cytoplasm or the nucleus and include P-bodies, germ granules, neuronal granules, paraspeckles, Cajal bodies, and stress granules (Buchan, 2014; Voronina et al., 2012; Kiebler and Bassell, 2006; Anderson and Kedersha, 2006; Fox et al., 2002; Gall, 2000). Stress granules are cytoplasmic assemblies of non-translating mRNAs and RNA binding proteins that form during stress responses where bulk translation initiation is inhibited.

Stress granules are of interest for four reasons. First, stress granules are thought to play a role in the stress response and gene regulation (Buchan and Parker, 2009; Anderson and Kedersha, 2009). Second, related RNP granules exist in neurons and can affect synaptic plasticity (Barbee et al., 2006; McCann et al., 2011). Third, mutations in RNA binding proteins, or stress granule remodeling complexes, that lead to constitutive or increased stress granule formation are causative in amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD), inclusion body myopathy (IBM) and other degenerative disorders (Buchan et al., 2013; Kim et al., 2013; Mackenzie et al., 2017). Fourth, stress granule formation can influence both tumor progression and viral infection (Anderson et al., 2015; Beckham and Parker, 2008; Reineke and Lloyd, 2013; Somasekharan et al., 2015).

Stress granules, and other RNP granules, are thought to form by RNA providing a scaffold for RNA-binding proteins that form protein-protein interactions \textit{in trans} to
drive assembly. For example, mammalian stress granule assembly is promoted by interactions of the TIA1, G3BP1 and ATXN2 proteins (Nonhoff et al., 2007; Kedersha et al., 2016; Gilks et al., 2004). Similarly, assembly of yeast P-bodies is promoted by the dimerization of the RNA binding protein Edc3, as well as an intrinsically disordered region (IDR) on Lsm4p (Decker et al., 2007). Consistent with interactions between RNA-binding proteins promoting RNP granule assembly, several purified RNA binding proteins, in isolation or with RNA, can form self-assemblies in vitro as either coacervates, fibers, or hydrogels, although the specific relationship of these in vitro assemblies to RNP granules remains unclear (Anderson and Kedersha, 2006; Buchan, 2014; Elbaum-Garfinkle et al., 2015; Hennig et al., 2015; Kato et al., 2012; Kiebler and Bassell, 2006; Lin et al., 2015; Nott et al., 2015; Fox et al., 2002; Gall, 2000; Voronina et al., 2012; Zhang et al., 2015). The role of trans RNA-RNA interactions in stress granule formation remains relatively unexplored.

Herein we illustrate that RNA-RNA interactions contribute to the assembly of stress granules and the targeting of RNAs to stress granules. We show that trans RNA-RNA interactions play a role in the recruitment of given RNAs to stress granules. We find that various RNAs are capable of self-interactions, even in the absence of Watson-Crick base pairing. More importantly, we show that mixtures of cellular RNAs partition into self-assemblies in vitro under physiologically relevant conditions and that the RNAs enriched in these assemblies reflect the RNAs enriched in stress granules in vivo. This argues that RNA-RNA interactions both contribute to stress granule assembly and in determining the RNAs enriched in stress granules. Strikingly, we also observed that RNA self-assembly is promoted by the pathogenic dipeptide repeats
GR and PR. This work highlights that RNPs exchange between monomeric and multimeric states within the cell, and the cell utilizes a variety of mechanisms to regulate this equilibrium.
2.3. Results

Four points led us to hypothesize that RNA-RNA interactions in trans might contribute to stress granule formation. First, stress granules form when ribosomes run-off, exposing the previously ribosome occupied coding regions which would be expected to form RNA-RNA interactions both in cis and in trans. Second, long mRNAs partition highly into stress granules, but only show a modest increase in the binding sites of stress granule proteins, suggesting length might contribute to the partitioning of mRNAs into stress granules through trans RNA-RNA interactions (Barbee et al., 2006; Khong et al., 2017; McCann et al., 2011). Third, stress granule cores in lysates are resistant to high salt or aliphatic alcohols, which are known to disrupt many protein-protein, but not RNA-RNA interactions (Buchan et al., 2013; Jain et al., 2016; Kim et al., 2013; Mackenzie et al., 2017). Stress granule cores have been suggested to be independent of RNA based on their resistance to RNase treatment (Anderson et al., 2015; Beckham and Parker, 2008; Jain et al., 2016; Reineke and Lloyd, 2013; Somasekharan et al., 2015), but a re-examination of this observation showed that extensive RNase treatment fails to degrade the RNA within stress granule cores (Figure 2.1). Fourth, trans RNA-RNA interactions contribute to the assembly of multimeric RNPs in Drosophila embryos, and RNA foci containing transcripts with repeat expansions (Nonhoff et al., 2007; Jain and Vale, 2017; Kedersha et al., 2016; Ferrandon et al., 1997; Gilks et al., 2004).

Additional evidence that RNA-RNA interactions could contribute to stress granule assembly came from the partitioning of ncRNAs into stress granules in mammalian cells (Decker et al., 2007; Khong et al., 2017). By data mining the
Figure 2.1. RNAs remain in stress granules following robust RNase treatment. (A) Stress granule enriched fractions (see methods) from yeast transformed with Ded1Δ141-150 were treated with no RNase, RNase cocktail (RNase T1 and RNase A), or RNase III for 2 hrs at 37°C. Stress granules were monitored by Ded1Δ141-150 GFP, and RNAs were stained with the nucleic acid dye SYTO 17. (B) Stress granules were imaged without addition of SYTO 17 to test for bleed-through at the acquisition parameters used in (A). Scale bars are for zoomed out images, not the inserts.
mammalian stress granule transcriptome (Khong et al., 2017), we observed that ncRNAs shorter than 3000 bases, were generally depleted from stress granules, as predicted by length (Figure 2.2A). However, antisense ncRNAs shorter than 3000 bases showed a bimodal distribution with one population enriched in stress granules and one depleted (Figure 2.2B). Of the 14 antisense ncRNAs significantly enriched greater than two-fold in stress granules and with FPKM values above one in total RNA samples, 10 were antisense to long mRNAs that were enriched in stress granules (Figure 2.2D), five of which are illustrated (Figure 2.2F). This is striking since only 14.5 % of mRNAs in the cell are enriched in stress granules (Figure 2.2C). In contrast, of the 23 antisense ncRNAs significantly depleted greater than two-fold from stress granules and with FPKM values above one in total RNA samples, only one had an antisense partner that was enriched in stress granules (Figure 2.2E). Thus, there is a correlation between the localization of antisense ncRNAs and their ability to base pair to longer mRNAs, implying that RNA-RNA interactions may partition specific RNAs into stress granules.

To determine if RNA-RNA interactions could contribute to stress granule formation, we calculated the approximate concentration of the coding region of mRNAs exposed during polysome collapse. For yeast, we estimated the concentration of exposed coding regions to be between 170-800 μg/mL (see methods), while in the mammalian U-2 OS cell line, the exposed ORFs were estimated at approximately 180 μg/mL (see methods). If RNA-RNA interactions contribute to stress granule formation, then we predicted that RNA at these concentrations would spontaneously assemble under conditions mimicking the intracellular milieu.
Figure 2.2. RNA-RNA base pairing influences RNA localization. (A) All ncRNAs, excluding antisense, and (B) antisense ncRNAs with lengths < 3000 nt and significant sequencing reads were plotted in a histogram showing their log$_2$(fold change) enrichment in stress granule/total RNA. Red bracket highlights the second peak for RNAs enriched in stress granules. (C) Pie chart illustrating the proportion of all mammalian mRNAs enriched (red), depleted (blue) or neither (grey) in stress granules. (D-E) Stress granules localization of binding partners to (D) enriched and (E) depleted antisense ncRNAs. *One enriched antisense ncRNA transcript has a secondary binding partner to a depleted mRNA. (F) Examples of enriched antisense ncRNAs and their binding partners.
To test this prediction, we assessed whether purified protein-free total RNA from *S. cerevisiae* at 150 μg/mL would form assemblies under conditions where we varied the salt and used PEG to mimic molecular crowding. We observed that total yeast RNA readily self-assembled and formed two types of assemblies (**Figure 2.3A**). At higher PEG and lower salt the RNA was observed to form small droplets (**Figure 2.3A,a**). With increasing salt, we observed the formation of more amorphous assemblies that contained a larger percentage of the RNA, which, due to their morphology, we refer to as RNA tangles (**Figure 2.3A, b, c; Figure 2.4A,B**). Based on FRAP of spiked-in fluorescent RNAs, or dilution into lower ionic strength, droplets were more dynamic and less stable than RNA tangles (**Figure 2.4C,D**). Specifically, 97% of the RNA in tangles formed at high salt is immobile (**Figure 2.4C**) and dilution takes nearly an hour to disrupt tangles whereas droplets disperse on the order of seconds (**Figure 2.4D**). Importantly, both droplets and tangles were enriched for the RNA specific dye SYTO RNASelect (**Figure 2.4E**) and depleted for PEG (**Figure 2.4F**). This is consistent with PEG functioning primarily as a crowding agent, which is further supported by Ficoll promoting RNA self-assembly (**Figure 2.4G**). Moreover, in the absence of crowding agents, RNA also self-assembled under physiological concentrations of salt (150 mM), and the polyamines spermine (223 μM) and spermidine (1339 μM), which are known to stabilize RNA-RNA interactions (Aumiller et al., 2016; **Figure 2.3B**). Thus, under *in vitro* conditions analogous to the cytosol during a stress response, purified RNA undergoes self-assembly.
Figure 2.3. Various RNAs self-assemble in vitro. (A) Phase diagram of RNA assembly morphology under varying PEG and NaCl concentrations. Images correspond to labeled positions in phase diagram: (a) Droplets formed at 0 mM NaCl, 10 % PEG, (b) Droplet/tangles formed at 300 mM NaCl, 7.5 % PEG, (c) Tangles formed at 750 mM NaCl, 10 % PEG, (d) No assemblies at 300 mM NaCl, 2.5 % PEG. All conditions contain 1 mM MgCl$_2$ and 150 $\mu$g/mL yeast total RNA. (B) Total yeast RNA in 150 mM NaCl and 1 mM MgCl$_2$ with and without physiologically relevant conditions of spermine (223 $\mu$M) and spermidine (1339 $\mu$M). (C) polyU, polyC, polyA and polyG self-partitions in vitro at 500 $\mu$g/mL respective homopolymer, 10 % PEG and 750 mM NaCl.
Figure 2.4. Total Yeast RNA droplets and tangles have characteristic properties. (A) 150 μg/mL RNA in 10 % PEG, 1 mM MgCl₂ and varying NaCl concentrations (0 mM, 150 mM, 300 mM and 600 mM). (B) Percent of RNA in pellet according to RNA amount by weight (μg/mL) following centrifugation to pellet RNA assemblies. At higher salt concentrations, a larger fraction of the RNA in the mixture is localized to assemblies. (C-D) RNA droplets are more dynamic than RNA tangles. (C) Plot of FRAP recovery (left) of 50 nt fluorescent RNA (PTB RNA) in polyU droplets, total RNA droplets, and total RNA tangles. Representative images of fluorescent recovery following bleaching (right). (D) Total RNA tangles formed at 750 mM NaCl and 10 % PEG and total RNA droplets at 50 mM NaCl and 10 % PEG were diluted to positions on the phase diagram that lacked assemblies. State of assemblies was monitored over time as they dispersed. (E) Total yeast RNA droplets (50 mM NaCl) and tangles (750 mM NaCl) are enriched for RNA by staining with SYTO RNASelect (F) Localization of fluorescently labeled PEG in relation to total RNA droplets and tangles. (G) RNA tangles observed in 750 mM NaCl, 1 mM MgCl₂ and 20 % Ficoll. RNA stained with SYTO RNASelect.
Stress granules recruit a diversity of RNA binding proteins in addition to RNA. Similarly, we observed that RNA self-assemblies recruited RNA binding proteins (Figure 2.5). Specifically, when fused to GFP the IDRs of yeast Lsm4 and eIF4GII, both of which bind RNA (Lin et al., 2015), are recruited to RNA droplets more than the IDR of Fus or GFP alone (Figure 2.5A,B). Similarly, RNA droplets recruit hnRNPA1 (Figure 2.5C). Thus, assemblies formed by RNA interactions in cells would be expected to recruit RNA binding proteins.

RNA-RNA interactions contributing to RNA self-assemblies could be helical stacking (Zanchetta et al., 2008), specific base-pairing (Jain and Vale, 2017), or promiscuous interactions between mRNA involving both traditional Watson-Crick interactions, and additional interactions (Leontis et al., 2002). Evidence that non-Watson-Crick interactions can promote RNA self-assembly is that all four homopolymers self-assemble, with polyU forming rapidly relaxing droplets (as previously observed (Aumiller et al., 2016)), polyC forming slower relaxing droplets, polyA forming asymmetric assemblies with very slow relaxation rates, and polyG forming an aggregate that is presumably based on G-quadruplexes (Figure 2.3C, Figure 2.6A-C). Thus, RNA sequence may impart biophysical characteristics to their assemblies, but more importantly RNA self-assembly is a general property of diverse RNAs and is not restricted to Watson-Crick interactions.

If the self-assembly of RNA in vitro is relevant to stress granule assembly, we predicted that similar mRNAs would assemble into RNA droplets in vitro and stress
Figure 2.5. RNA binding proteins are recruited to RNA assemblies. (A) 500 nM IDRs fused to GFP were added to total RNA assemblies formed with 500 μg/mL total RNA in 3.4 mM Tris pH 7.4, 150 mM NaCl, 1 mM MgCl₂ and 10% PEG. RNA was visualized with spiked in small RNA Cy3-PTB RNA. (B) IDRs with RNA binding capacity are more efficiently recruited to total RNA assemblies than FUS (an IDR lacking RNA binding ability) or GFP alone. (C) MBP-HIS-SNAP-hnRNP A1Δhexa added in indicated concentrations on the left to 3.4 mM Tris pH 7.4, 50 mM NaCl, 10% PEG3350 and 1 mM MgCl₂. Mixtures containing 500 μg/mL yeast total RNA are shown on the left and mixtures without RNA on the right. MBP-HIS-SNAP-hnRNP A1Δhexa is fluorescently tagged with SNAP-Surface 488.
Figure 2.6. Homopolymer droplets have distinct biochemical properties. (A-C) Representative time series showing merging and relaxation events of (A) polyU, (B) polyC, and (C) polyA. Approximate relaxation times are shown on the right. Merging and relaxation times were determined by visual inspection of the amount of elapsed time between when two droplets start merging, losing the distinct edge between them, and when the merged droplet has relaxed back into a sphere.
granules \textit{in vivo}. To test this prediction, we purified RNA droplets formed under physiological salt (150 mM NaCl) and PEG from total yeast RNA in triplicate, sequenced the assembled RNA and compared that RNA population to the mRNAs that partition into stress granules (Khong et al., 2017). This experiment revealed several important observations.

First, triplicates of \textit{in vitro} assemblies and total RNA agreed within themselves but showed clear differences (\textbf{Figure 2.7A}), allowing the identification of both enriched and depleted RNAs in the \textit{in vitro} assemblies (\textbf{Figure 2.8A}). RNAs significantly (p<0.01) and two-fold enriched and depleted under these conditions will be referred to throughout the remainder of the text as \textit{in vitro} enriched and depleted, respectively. On average, the \textit{in vitro} enriched RNAs were significantly longer while the \textit{in vitro} depleted RNAs were shorter (\textbf{Figure 2.8B}). This closely mirrored the length effect seen in the analysis of the yeast stress granule transcriptome (\textbf{Figure 2.8C}; Khong et al., 2017).

A second key observation was a strong overlap between \textit{in vitro} and stress granule enriched mRNAs (\textbf{Figure 2.7B, Figure 2.8D}). Specifically, of the 916 mRNAs enriched in yeast stress granules (Khong et al., 2017), 634 are enriched \textit{in vitro}. Conversely, of the 1111 mRNAs depleted from yeast stress granules (Khong et al., 2017), only 56 are enriched \textit{in vitro}. This correlation extends to RNAs depleted from \textit{in vitro} assemblies and stress granule \textit{in vivo} (\textbf{Figure 2.7C, Figure 2.8E}). Of the 1456 mRNAs depleted from \textit{in vitro} assemblies, only 10 are enriched in stress granules, and of the 1111 mRNAs depleted from stress granules, 573 are also depleted from
Figure 2.7. Assemblies of yeast total RNA in vitro largely recapitulate the stress granule transcriptome. (A) $R^2$ values between each sample library. Triplicates of both 150 mM NaCl and total yeast RNA agree well with each other, while comparing across these classes decreases the $R^2$ values. (B-D) Overlap of 150 mM NaCl (B) enriched, (C) depleted and (D) neither RNAs with RNAs classified as enriched, neither or depleted from yeast stress granules (28). (E) Removal of transcripts with fractional optimal codons outside the range 0.4-0.6 increased the Pearson’s correlation, $R$, from 0.53 to 0.61 when comparing the degree of enrichment in vitro vs. in stress granules.
Figure 2.8. RNAs in self-assemblies in vitro largely recapitulate the stress granule transcriptome. (A) RNAs from in vitro assemblies formed at 150 mM NaCl identified as significantly (p<0.01) and two-fold enriched (1488 RNAs, red dots), and depleted (1456 RNAs, blue dots) when compared to total yeast RNA. (B-C) Box plots showing correlation between transcript length and RNA localization to (B) in vitro assemblies and (C) yeast stress granules (Khong et al., 2017). ***p<0.001 between any three. (D) In vitro enriched mRNAs significantly overlap with mRNAs identified in the yeast stress granule transcriptome and exhibit significant lack of overlap with RNAs depleted from stress granules. (E) In vitro depleted RNAs significantly overlap with RNAs depleted from stress granules and exhibit a significant lack of overlap with RNAs identified in the yeast stress granule transcriptome. (F) The degree of enrichment and depletion between in vitro assemblies and stress granules correlates, Pearson’s Correlation, R= 0.53. RNAs that are more enriched in vitro than in vivo (upper left) tend to have greater proportions of optimal codons (orange).
the RNA self-assembly \textit{in vitro}. This indicates that the biophysical properties of RNA that drive RNA self-assembly \textit{in vitro} correlate with a critical mRNA feature for partitioning mRNAs into stress granules \textit{in vivo}.

Plotting the relative enrichment of RNAs in \textit{in vitro} assemblies vs. in stress granules shows a correlation in the degree of enrichment in both cases with a Pearson’s correlation of 0.53 (Figure 2.8F). In general, RNAs more enriched in stress granules are also more enriched in assemblies \textit{in vitro}, and \textit{vice versa}. The RNAs that show less correlation between \textit{in vivo} and \textit{in vitro} recruitment can be explained by the effects of translation efficiency on mRNA recruitment into stress granules, where efficient translation has been shown to correlate with depletion from stress granules (Khong et al., 2017). Removal of transcripts with fractions of optimal codons below 0.4 or above 0.6 increases the Pearson’s correlation coefficient to 0.61 (Figure 2.7E). This makes sense, as translation would be expected to affect mRNAs partitioning into stress granules \textit{in vivo}, but not affect RNA self-assembly \textit{in vitro}, where there is no translational apparatus. Taken together, these observations suggest that the partitioning of mRNAs into stress granules is modulated in part by the self-assembly capabilities of RNA.

We suggest a working model where stress granules form when the summation of protein-protein, protein-RNA and RNA-RNA interactions, increase over a threshold for assembly. This model can be illustrated in a phase diagram with transitions between “phases” explaining stress granule formation (Figure 2.9). For example,
Figure 2.9. A four-phase model of stress granule assembly. Model illustrating how assemblies may be RNA (bottom right) or protein (top left) dominated, however often a combination of interactions is responsible for assembly within cells (top right). Arrows denote examples from literature that lead to stress granule formation or dissolution. Yellow arrow shows the formation of stress granules through a large influx of non-translating RNAs. Red arrow signifies lack of stress granules when key proteins are deleted. Black arrow denotes formation of stress granules through overexpression of certain RNA binding proteins.
increasing the concentration of exposed RNA, either by inhibiting translation initiation leading to ribosome run-off, or by transfection or injection of RNA into cells (Bounedjah et al., 2014; Fay et al., 2017; Mahadevan et al., 2013), could cause a transition in the cell that leads towards stress granule assembly via increased RNA-RNA interactions (Figure 2.9, yellow arrow). Similarly, deletion of G3BP1 and G3BP2, which are abundant proteins that contribute to stress granule assembly (Kedersha et al., 2016), prevents assembly under most stresses and moves components back into the non-assembled phase in our model (Figure 2.9, red arrow). Conversely, over-expression of TIA1, G3BP1 or FMR1, all of which can contribute to stress granule assembly (Anderson and Kedersha, 2008; Mazroui, 2002; Gilks et al., 2004; Tourrière et al., 2003), may drive the cell into a regime of assembled stress granules by increasing the protein-protein interactions (Figure 2.9, black arrow).

Since RNA-RNA interactions contribute to stress granule assembly, we hypothesized that any small molecule or peptide that stabilizes RNA-RNA interactions would promote stress granule assembly. Strikingly, prior work has shown that arginine containing dipeptides GR and PR produced by RAN translation of hexanucleotide (G4C2) repeat expansions of C9orf72 (Ash et al., 2013; Zu et al., 2013) are both toxic to cells and trigger stress granule formation (Boeynaems et al., 2017; Lee et al., 2013). Although these dipeptides perturb many cellular processes (Kwon et al., 2014; Lin et al., 2016; Shi et al., 2017), we predicted they may directly induce stress granules by stabilizing RNA-RNA interactions. Indeed, we observed that (PR)$_{10}$ and (GR)$_{10}$ robustly stimulated RNA self-assembly in vitro, while (GP)$_{10}$ had little effect (Figure 2.10A, B, Figure 2.11). Assemblies with (PR)$_{10}$ or (GR)$_{10}$ were enriched for both RNA
Figure 2.10. Pathogenic dipeptides increase RNA assembly. (A) Dipeptides (GR)\textsubscript{10} and (PR)\textsubscript{10} promote assembly, while (GP)\textsubscript{10} does not. Fluorescent dipeptides (green) and RNA (SYTO 17, red) are both enriched in assemblies. (B) Phase diagram illustrating the assembly of (GR)\textsubscript{10} and RNA. Square signifies lack of assembly. +, sparse and small assembly; ++, moderate assemblies (*constituting either frequent but smaller assemblies or larger assemblies that were sparser); +++, robust assembly. Green indicates protein only assembly.
Figure 2.11. Pathogenic dipeptides increase RNA assembly \textit{in vitro}. (A) \((GR)_{10}\) and \((PR)_{10}\) at 250 mM with and without 150 \(\mu\)g/mL total yeast RNA in 150 mM NaCl and 1 mM MgCl\(_2\). (B) Phase diagram illustrating the assembly of \((PR)_{10}\) and RNA. Square signifies lack of assembly. +, sparse and small assembly; ++, moderate assembly (*constituting either frequent but smaller assemblies or larger assemblies that were sparser); ++++, robust assembly. (C) Representative images of classifications for phase diagram listed above.
and dipeptides (Figure 2.10A), and neither RNA nor protein alone was sufficient for robust assembly under these conditions (Figure 2.10A, Figure 2.11A). Since the most toxic dipeptides are the same dipeptides that stimulate the assembly of RNA in vitro, we suggest that dipeptides may exert some of their toxic effects by promoting RNA-RNA assemblies in the cell, such as through the formation of stress granules.
2.4. Discussion

We present several lines of evidence that stress granules form in part by RNA-RNA interactions, and those interactions can influence the partitioning of mRNA. First, we show that a portion of total yeast RNA effectively self-assembles \textit{in vitro} under conditions mimicking intracellular stress conditions (Figure 2.3). Importantly, this self-assembly of yeast RNA largely reproduces the stress granule transcriptome (Figure 2.8). Moreover, the enrichment of short antisense ncRNAs in mammalian stress granules correlated with their ability to base pair to a longer mRNA enriched in stress granules (Figure 2.2). Finally, dipeptide repeats, that are known to induce stress granules in cells, strongly increase RNA self-assembly \textit{in vitro} (Figure 2.10). In combination with genetic experiments showing proteins can enhance stress granule assembly, we conclude that stress granules assemble by a summation of protein-RNA, protein-protein, and RNA-RNA interactions (Figure 2.9). The precise set of interactions that drive stress granule formation can vary under different conditions as long as the total summation is above the critical threshold for assembly.

There is growing evidence that RNA-RNA interactions can help drive the assembly of stress granules. A role for RNA-RNA interactions could explain why purified stress granule cores are stable against many insults (Jain et al., 2016). In addition, a role for RNA-RNA interactions could suggest why the ATPase activity of Ded1p is required to disassemble stress granules (Hilliker et al., 2011). Our results also highlight the importance of charge shielding in promoting RNA self-assembly, as increasing salt concentrations or adding positively charged molecules such as
polyamines or arginine containing dipeptides greatly increase assembly. This is consistent with observations in the literature suggesting that ionic strength, which has a strong impact on RNA self-assembly in vitro (Figure 2.3, Figure 2.4), could be an important regulator of stress granule formation. Stress granule assembly can be triggered by hypertonic shock and conversely fail to form in conditions of low intracellular osmolality (Boundedjah et al., 2012). Moreover, sorbitol, another osmotic stressor, can partially rescue formation of stress granules in ΔΔG3BP1/2 cells (Kedersha et al., 2016), perhaps because the sorbitol increases the intracellular osmotic strength thereby stabilizing RNA-RNA interactions. In another line of evidence, the addition of G4C2 RNA to U-2 OS cell lysates condenses an assembly containing many stress granule components and cellular RNAs, perhaps by nucleating interactions between various mRNAs in the lysates (Fay et al., 2017). Finally, the prevalence of RNA-RNA interactions helps explain why injection or transfection of concentrated RNA into cells leads to the formation of large, higher-order assemblies (Boundedjah et al., 2014; Fay et al., 2017; Mahadevan et al., 2013). This is analogous to the huge influx of exposed RNAs during a stress response, which may allow for an emergence of interactions that are normally outcompeted by more specific, high-affinity interactions.

The role of RNA-RNA interactions in stress granule assembly has two broader implications. First, one anticipates that RNA-RNA interactions will contribute to other RNP granules. Indeed, given the thermodynamic strength of RNA-RNA interactions, they should be expected to be a stable state in cells and will form in trans whenever there is a sufficiently high local concentration of RNA. For example, we suggest that
very efficient transcription of long RNAs would be expected to drive the formation of RNA-RNA interactions between nascent transcripts. One possible example of this process would be the assembly of paraspeckles, which form at sites of transcription of NEAT1 RNA and contain multiple NEAT1 copies (Clemson et al., 2009).

A second implication is that RNA-RNA interactions can be promiscuous and form between any two RNAs with single-stranded regions. For example, by chance the average mammalian mRNA in stress granules (7.5 kb) (Khong et al., 2017) should have over 300 possible sites of 6 consecutive base pairs with another 7.5 kb mRNA. Even if the vast majority of these sites are lost to intramolecular RNA folding or RNA binding proteins, numerous possible sites for RNA-RNA interaction will remain. Thus, one anticipates that RNA-RNA interactions can arise between many different RNAs. In some cases, evolution will have created definitive interaction sites to give specificity to assemblies, as has been seen with oscar and bicoid mRNAs in Drosophila embryos (Ferrandon et al., 1997).

Although promiscuous RNA-RNA interactions are capable of forming between any two RNAs, it is important to note that there will be a gradient of any given RNAs propensity to assemble. We show that longer RNAs are more enriched in assemblies in vitro and in stress granules (Figure 2.8). In addition, one anticipates that the ability of RNAs to form base-pairing interactions will increase their self-assembly properties. Given these two inputs, short structured RNAs, such as tRNAs (originally sRNAs for “soluble” RNAs), will be excluded from RNA assemblies. Alternatively, longer RNAs,
particularly those with repeat sequences capable of self-base-pairing will be highly efficient at self-assembly. Recent reports have illustrated that longer repeat RNAs more effectively form intracellular and \textit{in vitro} RNA assemblies (Jain and Vale, 2017) and that transfection of G-quadruplex capable G4C2 RNA elicits robust stress granule assembly whereas it's antisense counterpart C4G2 does not (Fay et al., 2017). In this light it is notable that many repeat-containing RNAs, including both pathogenic toxic RNAs and satellite RNAs form specific nuclear foci (de Mezer et al., 2011; Hall et al., 2017; Lee et al., 2013; Mankodi, 2005) This suggests that formation of hyper-stable RNA assemblies in cells is toxic and this provides a possible explanation for why repeat RNAs with strong tendency to base-pair cause disease once they are expanded beyond a certain length.

This work suggests that within cells RNPs exist at an equilibrium between monomeric RNPs and multimeric RNP granules, which is influenced by many parameters. For example, RNA helicases, ribosome association, and monovalent RNA binding proteins are expected to play a role in maintaining RNPs in the monomeric state. Consistent with this view, it is known that depletion of the abundant Tdp-43 ortholog in \textit{C. elegans} leads to the accumulation of dsRNA foci in the nucleus (Saldi et al., 2014). In contrast, longer RNA lengths, high local or transient RNA concentrations, and the propensity of RNA to interact with itself will promote RNA-RNA association. Taken together, we suggest that RNA-RNA assemblies in cells may be a default state and cells prevent this RNA aggregation by active means, ribosomes, and RNA binding proteins.
2.5. Materials and Methods

Materials

Homopolymer RNAs were purchased from Amersham Pharmacia Biotech Inc. (polyA, C, U, G) and Sigma (polyU). Fluorescently labeled PTB (Lin et al., 2015) and short homopolymer oligos were purchased from IDT (Table 2.1). Dipeptides were ordered through New England Peptide (Table 2.1).

Antisense RNA Analysis

Data from the mammalian stress granule transcriptome (Khong et al., 2017) was analyzed for ncRNAs. ncRNAs were split into antisense ncRNAs and all other ncRNAs (Kinsella et al., 2011). For this analysis, only ncRNAs with FPKM values > 1 in the total RNA samples, lengths < 3000 nts, and significant reads were used. ncRNAs were binned by log$_2$(fold change) and graphed. For pie charts, antisense ncRNAs that were significantly enriched or depleted (> two-fold) were found on UCSC genome browser to identify their sense binding partners and the degree of overlap. The localization of the sense binding partners in relation to stress granules was charted.

Preparation of stress granule enriched fraction and RNase treatment

Stress granules were enriched from BY4741 cells transformed with Ded1Δ141-150 as described previously (Jain et al., 2016). Yeast lysates were prepared from cell pellets frozen from 50 mL cultures grown to log phase. Pellets were resuspended in small volumes of lysis buffer (50 mM Tris-HCl, pH 7.4, 100 mM K$_2$Ac, 2 mM Mg(oAc)$_2$, 50 µg/mL Heparin, 0.5 mM DTT, 0.5% NP40 EDTA free protease inhibitor).
Table 2.1. Sequence of Materials. RNA sequence of small fluorescent RNA reporter and amino acid sequence of dipeptide repeats.

<table>
<thead>
<tr>
<th>RNA</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTB RNA</td>
<td>UCUCUAAAAAUCUCUAAAAAUCUCUAAAAAUCUCUAAAAAUCUCUAAAAAUCUCUUAA/A3Cy3Sp/</td>
</tr>
<tr>
<td>Dipeptides</td>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>(GR)10</td>
<td>5Fam-GRGRGRGRGRGRGRGRGRGRGRGRGRGR-OH</td>
</tr>
<tr>
<td>(PR)10</td>
<td>5Fam-PRPRPRPRPRPRPRPRPRPRPRPRPR-OH</td>
</tr>
<tr>
<td>(GP)10</td>
<td>5Fam-GPGPGPGPGPGPGPGPGPGPGPGPG-OH</td>
</tr>
</tbody>
</table>
Approximately 500 µL of acid-washed glass beads (Sigma G8772) were added for bead lysis. Cells were lysed by vortexing at high speed for 2 minutes followed by 2 minutes sitting on ice, repeated three times. A heated 18-gauge needle was used to poke a hole in the bottom of the Eppendorf tube. The Eppendorf tube was placed in a 10 mL conical and the lysate was collected in the 10 mL conical by centrifugation for 2 min at 2000 rpm (805 rcf). The semi-clarified extract was collected and further spun at 15000 rpm (21000 rcf) for 10 min in the cold room. The pellet was resuspended in 50 µL lysis buffer and passed through a final slow spin at 3000 rpm (800 rcf). RNase cocktail (Ambion AM2286) or RNase III (NEB #M0245L) was added to the granule-enriched fraction as per manufacturer’s directions and placed at 37°C for 2 hours before imaging on a microscope.

Calculation of exposed RNA during stress response

Calculation of between 170-800 µg/mL exposed RNA due to ribosome run-off was based on the following assumptions for yeast: 1) A yeast volume of ~40 µm³. 2) 15,000-70,000 mRNAs per yeast cell. 3) Average length of yeast mRNA of approximately 1000 nucleotides. 4) ~85 % of yeast mRNA is composed of the ORF. 5) Average molecular weight of nucleotide in RNA is 330 g/mol. 6) 100 % ribosomal run-off.

For mammalian cells, the calculation of 180 µg/mL was based on the following numbers: 1) 1.08 x 10⁹ total nucleotides of mRNA in a U-2 OS cell (from number of mRNA molecules x transcript length (Khong et al., 2017)). 2) Volume of 2000 µm³ for a U-2 OS cell. 3) Average weight of 330 Da for an RNA nucleotide. 3) ORF represents ~60% of the average RNA.
**Total RNA Self-Assembly**

Total RNA was extracted from yeast by standard hot phenol/chloroform extraction followed by EtOH precipitation. RNA was resuspended in RNase free water. All experiments were carried out in 1 mM MgCl$_2$ and in varying PEG (0-10 %) and NaCl (0-750 mM) concentrations as described in the text. RNAs were preheated for 1 min (>60°C) and cooled on ice before addition to rest of mixture.

**Homopolymer Self-Assembly**

polyU, polyA, polyC and polyG stock solutions were resuspended in water. All experiments were carried out in 10 % PEG (MW3350), 1 mM MgCl$_2$ and 750 mM NaCl. RNAs were preheated for 1 min (>60°C) and cooled on ice before adding to the rest of the mixture. Reactions were performed in glass-bottom chambers and allowed to sit for at least 15 minutes before imaging on the microscope.

**Approximate time of merging and relaxation**

To approximate time of merging and relaxation of homopolymer droplets, time lapses were taken to encompass the full series of droplet merging events, from initial contact between two droplets to relaxation back to a sphere following merging. Because of the lack of accuracy in measurement (due to intervals between images) values are represented as approximate times of merging and relaxation.

**Microscopy of RNA Assemblies**
Mixtures were placed in 96-well glass bottom plates with high-performance #1.5 cover glass (Fisher Scientific). Images were acquired on a DeltaVision epi-fluorescence microscope with 100x objective (Applied Biosystems) equipped with a SCMOS camera. Images for FRAP analysis were acquired using a Nikon A1R laser scanning confocal microscope.

*RNA binding proteins and IDR Recruitment to RNA Assemblies*

MBP-GFP-IDR-HIS proteins were added to a final concentration of 500 nM to RNA assemblies formed with 500 μg/mL total yeast RNA, 3.4 mM Tris pH 7.4, 150 mM NaCl, 1 mM MgCl₂ and 10 % PEG. For the graphs showing the fold enrichment over the background, the mean intensity from the darkest region of the image was considered “background.” The graph shows the mean of 10 line traces of equal length across different droplets from 1 image. The shaded areas represent the standard deviation. SNAP-hnRNPA1 was added at varying concentrations to 500 μg/mL yeast total RNA in 3.4 mM Tris pH 7.4, 50 mM NaCl, 1 mM MgCl₂ and 10 % PEG. Proteins were purified as previously described (Lin et al., 2015).

*Total RNA Pelleting for Tape Station and Sequencing*

Total yeast RNA obtained from bead lysis followed by phenol chloroform extraction and EtOH precipitation was assembled in vitro with 150 mM NaCl and 10 % PEG and allowed to sit for 15 minutes at room temperature. Mixtures were spun at 21,000 RCF for 30 minutes. Pellets were resuspended in RNase free water. For sequencing, RNA pellets and total RNA were treated with DNA-free DNA removal kit (Thermo Fisher Scientific) and sent to the University of Colorado BioFrontiers Institute Next-Gen
Sequencing Core Facility for Ribo-Zero treatment, library construction and NextSeq run.

Sequencing Data Analysis

Read quality was assessed using fastqc. Illumina adaptors were trimmed using Trimmomatic 0.32 in paired-end (PE) mode (Bolger et al., 2014). An index genome was built with Bowtie 0.12.7 using the build command and the S288C reference genome version (R64-1-1.fasta file), which was acquired through Saccharomyces Genome Database (SGD) (Langmead et al., 2009). Reads were aligned using Bowtie 0.12.7 using the following options: -S -v 2 -m 3 –best. Reads mapping to each gene were counted using HTSeq with parameters –t gene –I Name –f sam –s reverse, using the R64-1-1 annotation file (Anders et al., 2015). Normalization and differential expression analysis were then performed using DESeq 1.22.1 (Anders and Huber, 2010). Sequences have been uploaded to NCBI GEO and can be viewed with the accession number GSE99170.

Length and optimal codon analysis

Data for the length analysis (Fig. 4C) was obtained from a previous report (Nagalakshmi et al., 2008). The fraction of optimal codons for each gene (Fig. 4G) was calculated using a custom script based on codon stabilization coefficients outlined in a previous report (Presnyak et al., 2015). In brief, the amount of each codon was tallied on a gene-by-gene basis. Optimal codons were assigned a value of one while non-optimal codons were assigned a value of zero. The total number of optimal codons
was then divided by the total number of codons in order to find the fraction of optimal codons. All scatterplots for the sequencing analysis were created using Tableau.

**Statistical Tests**

Two sample proportion tests of binding partner localization were compared to proportions of all mRNAs. To calculate the statistical significance of differences in average length a 2-sample z-test was performed. In order to calculate the statistical significance of the overlap between *in vitro* RNA and the stress granule transcriptome, a 2-population proportion test was performed in R using the option “alternative=greater” to calculate the significance of overlap and the option “alternative=less” to calculate the p-value of the lack of overlap between the indicated datasets. Statistics on Venn Diagrams was done with a Fisher’s Exact Test.

**FRAP**

PTB RNA (100 nM) was spiked into total RNA tangles (500 μg/mL yeast total RNA, 10% PEG, 750 mM NaCl, 1 mM MgCl₂), total RNA droplets (500 μg/mL yeast total RNA, 10% PEG, 50 mM NaCl, 1mM MgCl₂) or polyU droplets (500 μg/mL polyU, 10% PEG, 1 mM MgCl₂). Fluorescence Recovery After Photobleaching (FRAP) experiments were performed on a Nikon A1R laser scanning confocal microscope. Regions containing multiple droplets or large tangle assemblies were bleached, leaving most droplets or tangles within the field of view unbleached. Images were quantified in FIJI (Schindelin et al., 2012). “Correct 3d Drift” was used to correct for x/y drift. Following drift correction, 3 droplets/regions within tangles were manually selected and their mean intensities measured from the photobleached region.
similarly sized droplets/regions of tangles from the non-bleached region were likewise selected and measured. Background photobleaching due to imaging was observed in the non-bleached assemblies. Background photobleaching was calculated as
\[
\text{Background}_i = \frac{f_i}{f_0}, \text{where } f_i \text{ is the intensity at time } i, \text{ and } f_0 \text{ is the average intensity of the first three acquired time points (prior to intentional bleaching). Bleached-droplet intensities were correct for background-bleaching:}
\]
\[
\text{Corrected}_i = \frac{\text{Bleached Droplet Mean Intensity}_i}{\text{Background}_i}
\]
Fractional Recovery at each time point was calculated as :
\[
\text{Fractional Recovery} = \frac{\text{Corrected}_i}{\text{Corrected}_0}
\]
Curves were fit using MATLAB to:
\[
\text{Fractional Recovery} = (\text{mobile fraction})\times((1 - e^{-\tau t})+\text{offset})
\]
where
\[
\tau = \frac{t_{1/2}}{\ln(0.5)}
\]
and “offset” accounts for incomplete bleaching. Because droplets and tangles in different conditions recruited PTB RNA to different degrees, acquisition parameters were changed between conditions.

RNA Assembly Dilution
Total RNA was assembled into droplets (50 mM NaCl) or tangles (750 mM NaCl) and diluted with 3x the volume of RNase free water. Water was added drop-wise to limit disturbance. Assemblies monitored on DeltaVision over time.

Dipeptides Experiments
(GR)$_{10}$, (PR)$_{10}$, and (GP)$_{10}$ fluorescently tagged with 5Fam were purchased from New England Peptide. Dipeptides were added at the concentrations listed in the presence of yeast total RNA and 150 mM NaCl. No crowding agent was added. Mixtures sat for one hour preceding visualization.
CHAPTER 3: ANALYSIS OF RNA-RNA INTERACTIONS IN RNA RECRUITMENT TO STRESS GRANULES

This chapter presents ongoing work.
3.1. Abstract

Stress granules are non-membrane bound organelles that assemble in the cytoplasm upon exposure to stresses that inhibit translation initiation. The recent elucidation of the stress granule transcriptome revealed several antisense RNAs that were enriched in stress granules despite their predicted depletion. This, along with the ability of RNAs to self-assemble \textit{in vitro}, raised the possibility that intermolecular RNA-RNA base-pairing could be a driving force for targeting certain RNAs to stress granules. To test if RNA base-pairing was sufficient to increase RNA recruitment we designed a single-molecule reporter RNA assay to monitor how changes in an RNA affect its recruitment into stress granules. Two important observations emerged: 1) adding sequences to a reporter can increase recruitment into a stress granule and 2) this effect does not appear to be due to the designed base-pairing as the complementary target is not required for targeting and reporter RNAs are generally not close enough in proximity to their complementary targets to support persistent base-pairing. This suggests that relatively short sequences can have an effect on RNA localization perhaps due to protein binding partners or RNA structural features.
3.2. Introduction

Stress granules are higher-order assemblies that form during a variety of stresses including heat shock, nutrient deprivation, hypo-osmotic conditions and oxidative stress, all of which inhibit bulk translation initiation (reviewed in Protter and Parker (2016)). Stress granules, as well as other ribonucleoprotein (RNP) granules, are considered non-membrane bound organelles, allowing exchange of components with the surrounding milieu. Despite this dynamicity, certain RNAs and proteins are enriched in stress granules over others (Jain et al., 2016; Khong et al., 2017), presumably due to either an increased number of interactions (e.g., higher valency; Banani et al., 2016), or the formation of very stable interactions. Understanding the mechanisms underlying the recruitment and retention of stress granule enriched components will lead to tools for elucidating stress granule function and may shed light on how compositional changes could contribute to pathological transitions of persistent stress granules.

The targeting of RNP granule proteins relies on different domains for each individual protein. Many proteins have been interrogated with the emergence of some common themes (Anderson and Kedersha, 2008). First, RNA-binding ability is a major player in RNP granule targeting. This makes sense as RNP granules are highly enriched for RNA. For example, ALS-mutant FUS incorporates into stress granules, whereas abrogating the RNA-binding of FUS disrupts this targeting (Diagle et al., 2013). Second, specific interactions with other proteins can drive protein recruitment. For instance, PMR1 is targeted to stress granules through direct binding to TIA-1, a stress granule protein (Yang et al., 2006). Third, for some proteins, the presence of
an intrinsically disordered region (IDR) is important for targeting to an RNP granule. For example, mutations in the tyrosine residues within the IDRs of FUS and RBM14 disrupt their localization to paraspeckles (Hennig et al., 2015). This is not true for all IDR-containing RNP granule proteins as some proteins are targeted to RNP granules in the absence of their IDRs (Protter et al., 2018; Chapter 4).

RNAs are also enriched in stress granules yet the census of RNAs in stress granules has only recently emerged. Three general principles have become clear. First, only a fraction of total mRNA accumulates in stress granules (Kedersha et al., 1999; Mollet et al., 2008; Anderson and Kedersha, 2008; Zurla et al., 2011). A quantitative description of the stress granule transcriptome, achieved through purification and sequencing of stress granule cores in conjunction with standardization by single molecule fluorescent in situ hybridization (smFISH), revealed that only ~10-15% of cytoplasmic mRNA molecules accumulate within stress granules (Khong et al., 2017). Second, individual RNAs can vary from having 1% to 95% of their molecules in stress granules (Khong et al., 2017). While certain RNAs are enriched in stress granules, including 5’ terminal oligopyrimidine motif (5’ TOP) mRNAs (Damgaard and Lykke-Anderson, 2011), other RNAs are largely excluded, such as β-actin, with only 3% stress granule localization (Zurla et al., 2011), or the mRNAs encoding heat shock proteins, which are excluded from heat shock induced stress granules (Kedersha and Anderson, 2002). Third, the primary determinants for stress granule enrichment are length and translatability, with longer transcripts and non-optimal transcripts localizing most strongly to stress granules (Khong et al., 2017). Additional sequence elements that correlate with increased mRNA accumulation in heavy fractions during stress
have been identified, such as AU-rich elements, but whether these directly target mRNAs into stress granules will require more research (Namkoong et al., 2018). Although features that result in an enrichment of some RNAs over others have been identified, the mechanistic reasoning is still unclear.

An emerging possibility is that long RNAs accumulate in stress granules simply because they can form more non-specific RNA-RNA interactions in trans (Van Treeck et al., 2018; Chapter 2). This possibility is first suggested by the limited correlation of the binding of stress granule proteins and RNA enrichment in stress granules (Khong et al., 2017). For example, on average, long mRNAs found in stress granules are not enriched in binding sites for stress granule proteins based on eCLIP analysis (Khong et al., 2017). Moreover, the NORAD IncRNA, which contains 19 binding sites for the stress granule component pumilio, accumulates in stress granules independent of pumilio (Namkoong et al., 2018). RNA also robustly self-assembles in vitro, even in the absence of proteins (Aumiller et al., 2016; Van Treeck et al., 2018; Chapter 2). Strikingly, the RNAs enriched in assemblies formed in vitro under physiologically relevant conditions, largely recapitulate the stress granule transcriptome (Van Treeck et al., 2018; Chapter 2). Taken together, this implies that a majority of the targeting to stress granules may be an inherent property of RNA. The length and estimated translatability of an RNA can accurately predict whether an mRNA will be enriched in or depleted from stress granules ~75% of the time (Khong et al., 2017). Although this is an impressive feat, it also suggests that other nuanced changes in sequence and structure can impact an RNAs partitioning in the cell.
The most straightforward example of this is through complementary base-pairing between RNAs. Analysis of the stress granule transcriptome has revealed a number of antisense RNAs (asRNAs) that were enriched in stress granules despite their short length, less than 3,000 bases (Khong et al., 2017; Van Treeck et al., 2018; Chapter 2). These enriched asRNAs had complementary regions to stress granule enriched RNAs (Van Treeck et al., 2018; Figure 2.2). Conversely, asRNAs with complementarity to stress granule depleted RNAs were generally depleted from stress granules (Van Treeck et al., 2018; Figure 2.2). This implied that Watson-Crick base-pairing between RNAs might play a role in RNA localization, similar to the specific RNA-RNA interactions important for bicoid or oskar localization to RNP granules (Jambor et al., 2011; Ferrandon et al., 1997).

Here we describe the use of a reporter assay to interrogate sequences responsible for driving RNAs to stress granules. By appending sequence to a luciferase reporter RNA, we begin to ask how the presence of a complementary region to a stress granule enriched RNA can affect targeting. We show that sequences are not equal in their recruitment to stress granules despite being length matched. However, the precise mechanism of recruitment, at this point, remains speculative as knock-down of the hypothesized recruiting RNA has no impact on reporter localization. These experiments serve as a proof of principle that moderate alterations to sequence can change an RNAs localization under stress and that this method will be useful for interrogating the rules of RNA localization to stress granules.
3.3. Results

**smFISH is compromised in transfected cells**

smFISH has emerged as a strong tool for assaying the localization of RNAs to stress granules (Khong et al., 2017). This method allows visualization of a single RNA in a fixed cell by tiling the RNA with fluorescent oligos (Figure 3.1).

To test multiple variations and mutations of a reporter RNA, we aimed to combine plasmid transfection with smFISH. **Luciferase** was chosen as the reporter RNA for two reasons. First, **luciferase** is not endogenous to mammalian cells, thereby eliminating background signal from endogenous RNAs that could complicate the data. This is especially important when mutants and variants of the reporter RNA are analyzed. Second, **luciferase** RNA is ~2 kb and would therefore be predicted to be non-enriched in stress granules (Khong et al., 2017). Assuming this is true, identification of sequences or variations that promote recruitment to stress granules should be possible.

U-2 OS cells were transfected with a plasmid containing luciferase under a Tet-on promoter to test the efficacy of quantifying reporter RNA localization. Following 24 hours of induction with doxycycline, cells were stressed with sodium arsenite for 1 hour and fixed. smFISH for **luciferase** RNA was done in conjunction with IF for G3BP, to demarcate stress granules, and smFISH for **AHNAK**, an endogenous transcript and positive control. Although the **AHNAK** probes showed clear recruitment to stress granules, as previously observed (Khong et al., 2017), the luciferase probes
Figure 3.1. smFISH as a tool for assaying RNA localization to stress granules. Immunofluorescence of G3BP1 (green), a stress granule protein, delineates arsenite-induced stress granules. smFISH using probes to PEG3 (red), a stress granule enriched RNA, shows robust localization as expected. Scale bars, 2 µm
accumulated in large spots outside of the nucleus that were much larger than the typical signal obtained for one RNA (Figure 3.2A). Some cells were identified in which large aggregates were present, but also smaller dots with the size and intensity generally generated by smFISH (Figure 3.2B). This suggested that the probes were able to pick up on luciferase RNA but were overwhelmingly marking something else in the cytoplasm, thereby making smFISH analysis problematic.

Further analysis of the large cytoplasmic clusters showed that the luciferase smFISH probes were hybridizing to the plasmid DNA. Transfection of plasmids, either GFP or luciferase, resulted in multiple cells with extranuclear DAPI signal, which was lacking in untransfected cells (Figure 3.2C, yellow asterix). Importantly, the luciferase signal only colocalized with the DAPI signal in cells transfected with luciferase plasmid (Figure 3.2C, yellow arrows). This suggests that the luciferase probes are hybridizing to the corresponding sequence within the plasmid. This is consistent with previous observations that transfected plasmids often localize predominantly in a few cytoplasmic clusters which contain tens of thousands of molecules (Wang et al., 2016). Taken together, smFISH is largely incompatible with plasmid transfection.

*Reporter constructs integrated into the genome are compatible with smFISH*

Due to the limitations encountered with pairing transfection with smFISH, luciferase was instead integrated into the genome at the adeno-associated virus integration site (AAVS) locus using CRISPR-Cas9. Cells with successful luciferase incorporation had smFISH signal whereas cells without luciferase incorporation lacked distinct foci (Figure 3.3A). This allowed for the identification of smFISH spots using
Figure 3.2. smFISH is not compatible with plasmid transfection. (A) Deconvolved images showing smFISH for *luciferase* (red), an RNA product from the transfected plasmid, and *AHNAK* (gray), an endogenous RNA. As expected, *AHNAK* RNAs are enriched in stress granules, marked with G3BP (green). The *luciferase* smFISH signal localizes in large, bright, cytoplasmic clusters. (B) In some cells, smFISH spots consistent with the size and brightness of single RNA molecules can be observed (circle) in locations of the cell distant from the large cytoplasmic signal (arrows). (C) In the left panels, yellow asterisks mark cells with extranuclear DAPI signal. Untransfected cells lack cells with extranuclear DAPI signal. Cells transfected with either GFP or luciferase expressing plasmids have a much higher fraction of cells with extranuclear DAPI signal. smFISH probes for *luciferase* do not co-localize with extranuclear DAPI signal in GFP transfected cells but do co-localize with extranuclear DAPI signal in *luciferase* transfected cells.
Figure 3.3. smFISH is compatible with genomic integration. (A) Non-deconvolved images of smFISH using luciferase probes on cells with and without genomic integration. Genomic integration allows visualization of luciferase RNAs as there is no longer conflicting plasmid-derived signal. (B) Top panel shows deconvolved luciferase smFISH signal (yellow), stress granules are stained for G3BP (blue). Blue dashed line outlines the nucleus. Purple dashed line outlines a single cell. Bottom panel shows Imaris’s ability to identify individual luciferase RNAs within a masked cytoplasm. (C) Workflow for analyzing images with Imaris. Acquired images are masked to outline a single cell’s cytoplasm. Imaris is used to identify and count smFISH spots within granules and in the entire cytoplasm. This allows quantification of the percent of cytoplasmic luciferase RNA that is localized to stress granules.
Imaris Software (Figure 3.3B). For all subsequent analysis, deconvolved images were masked to outline a single cell’s cytoplasm, Imaris was used to detect smFISH spots, and the percent of the cytoplasmic luciferase in stress granules was tallied after 1 hour sodium arsenite treatment (Figure 3.3C).

Quantification of smFISH signal showed that ~15% of cytoplasmic luciferase RNA (luc) was localized to stress granules (Figure 3.4B). This set the baseline used for comparing the localization of altered luciferase transcripts. Importantly, the modest localization of luciferase will allow for identification of sequences that increase recruitment to stress granules.

RNA sequence can alter localization under stress in a non-length dependent manner

With an assay in place, we first wanted to test the role of direct RNA-RNA interactions in facilitating the recruitment of RNAs to stress granules. Luciferase constructs were created with complementary sequences to either NORAD or AHNAK, two RNAs highly enriched in stress granules (Khong et al., 2017), or a random sequence that lacked perfect complementarity to anything in the human transcriptome. These 300 base inserts were placed in either the ORF or 3’UTR of luciferase (Figure 3.4A). While most of the insertions showed no significant change in reporter RNA localization, asAHNAK insertion into the 3’UTR (luc-A-3UTR) and asNORAD insertion into the ORF (luc-N-ORF) demonstrated increased localization to stress granules (Figure 3.4B). The luc-N-3UTR construct was not successfully cloned at this time and is therefore absent from this analysis.
**Figure 3.4.** Insertion of complementary regions to stress granule enriched RNAs alters the localization of some reporter constructs under stress. **(A)** 300 bases of a random sequence or complementary sequence to either AHNAK or NORAD RNA were inserted into the ORF or 3'UTR of luciferase. The bars under the NORAD and AHNAK antisense inserts show the location of the complementary region on NORAD and AHNAK, respectively. The box shows the 6 constructs tested **(B)** *Luciferase* reporter constructs, on the x-axis, were tallied for the % cytoplasmic RNA within stress granules. Box plots were created in which the center line represents the median and the colored box denotes the interquartile range. Each dot represents a single cell. RS – random sequence, A – antisense to AHNAK, N – antisense to NORAD, ORF – insertion into *luciferase* ORF, 3UTR – insertion into the 3'UTR of *luciferase*. * p<0.001, **p<0.00001.
Importantly, these results illustrate that RNAs of the same length can have different levels of enrichment in stress granules, and that a 300-base change is sufficient for this effect. This is consistent with some splice variants having very different levels of stress granule recruitment despite minor changes in RNA sequence (Khong et al., 2017).

In addition, these results illustrate that the same insert can have differential effects depending on its location within the luciferase transcript, as seen with the asAHNAK sequence inserted into the ORF versus the 3’UTR (Figure 3.4B). This could be due to altered sequence availability due to RNA structure or competitive binding of other RNAs or proteins.

Although the enrichment of luc-N-ORF and luc-A-3UTR is consistent with Watson-Crick base-pairing to their complementary target RNAs, the altered localization could also be the result of increasing interactions by bound proteins, RNA modification, structure, or other RNA-RNA interactions. We reasoned that if reporter RNAs were being targeted by their complementary target RNA, there should be a correlation between the percent localization to stress granules of the target RNA (AHNAK or NORAD) and the reporter RNA (luc-A-3UTR or luc-N-ORF) on a cell by cell basis. For example, we would expect that the recruitment of luc-N-ORF to stress granules would be higher in a cell with 90% of NORAD molecules in stress granules as compared to a cell with 60%. By graphing the % targeting to stress granules of the reporter against the target, we observed that both luc-N-ORF and luc-A-3UTR show increased recruitment to stress granules in cells where their target was more enriched.
(Figure 3.5). This correlation was not seen with the luc RNA control. This suggested that RNA-RNA interactions could be driving the enhanced recruitment of these reporter RNAs.

To determine if binding to the stress granule enriched RNA is responsible for reporter targeting we focused on luc-N-ORF. If recruitment is driven by RNA-RNA interactions, knocking down NORAD would be expected to reduce luc-N-ORF recruitment to percentages seen with the controls. Upon treatment with siRNAs against NORAD for 24 hours, cytoplasmic NORAD was reduced ~10-fold (Figure 3.6A, B). As expected, knock down of cytoplasmic NORAD had no effect on recruitment of the other enriched construct, luc-A-3UTR, which was predicted to be recruited by binding to AHNAK (Figure 3.6C). Unexpectedly however, knock down of NORAD also had no effect on the recruitment of luc-N-ORF (Figure 3.6C). This is in direct disagreement with the positive correlation between stress granule targeting of reporter RNAs and their targets (Figure 3.5). Instead this suggests that the added sequence is driving the reporter RNA into granules through other means, perhaps through protein binding, RNA structure, or promiscuous binding to other RNAs.

To interrogate whether the reporter and target RNAs were interacting, the distances between stress granule localized luciferase reporter RNAs and the closest AHNAK or NORAD RNAs were compiled in 3D (Figure 3.6D). Although a decrease in the distance would be expected simply due to increased recruitment to stress granules, if two RNAs are interacting we expect the smFISH signals to be within 200 nm. This is based off recent evidence that RNAs are highly compact and become even
Figure 3.5. Correlation between stress granule recruitment of *luc-N-ORF* and *luc-A-3UTR* reporter RNAs and their complementary targets. (A-B) Each dot represents a single cell in which the percent cytoplasmic *luciferase* reporter in stress granules is on the y-axis and the percent cytoplasmic target RNA in stress granules is on the x-axis. (A) *Luc-N-ORF* was counted in conjunction with *NORAD* (blue). *Luc* and *AHNAK* were tallied as the control pair (orange). (B) *luc-A-3UTR* was counted in conjunction with *AHNAK* (dark green). *Luc* is the same as in A (yellow-green).
Figure 3.6. Knock down of NORAD has no effect on recruitment of luc-N-ORF. (A) Wild type and siNORAD cells were analyzed for stress granule formation (G3BP signal) and NORAD expression by smFISH. Dashed line outlines the nucleus. While NORAD transcripts remain in the nucleus, NORAD signal in the cytoplasm is greatly reduced upon siRNA treatment. (B) Quantification of cytoplasmic NORAD in wild type and siNORAD cells. (C) Boxplots showing the distribution of reporter localization to stress granules in wild type and siNORAD cells. Center line represents the median and the outer lines of the colored box represent the first and third quartile. Each dot represents one cell. (D) Frequency plot showing the shortest distance between a stress granule localized reporter RNA and either NORAD or AHNAK, respectively. Inset shows two stress granules (blue) with NORAD RNAs (red) and luciferase RNAs (yellow). The yellow dot between the two stress granules would not be included in the analysis for D.
more compact during stress (Khong et al., 2018). Even extremely large RNAs, e.g., 18kb AHNAK, are physically restricted in the cell to < 200 nm (Khong et al., 2018). Although not conclusive, the fraction of distances in this range is not large enough to account for the increase in reporter recruitment (Figure 3.4B, Figure 3.6D) unless RNA-RNA interactions within stress granules are actively remodeled. The data suggests that most of the reporter RNAs are not actively interacting with the target RNA.

Taken together, these results suggest that either insufficient data points were gathered to robustly claim a correlation between luc-N-ORF and NORAD RNA recruitment (Figure 3.5) or the correlation is the result of some other shared property of the RNAs, e.g., AU-richness or binding sites for RNA-binding proteins. The 300-base antisense NORAD insert is much more AU-rich than either the random sequence or antisense AHNAK inserts (Table 3.1). AU-richness is an RNA property that has been identified in RNAs enriched in large assemblies during stress, especially during heat shock (Namkoong et al., 2018). This could be accounted for by increased base-pairing to other AU-rich RNAs, or to RNA binding proteins that preferentially bind these sequences.
Table 3.1. Sequence composition of 300 base inserts. Count of each nucleotide in the RNA insertion, and the corresponding GC and AU percentages.

<table>
<thead>
<tr>
<th>Insert</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
<th>Tot</th>
<th>%GC</th>
<th>%AU</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS300 NORAD</td>
<td>99</td>
<td>68</td>
<td>43</td>
<td>90</td>
<td>300</td>
<td>37%</td>
<td>63%</td>
</tr>
<tr>
<td>AS300 AHNAK</td>
<td>54</td>
<td>80</td>
<td>76</td>
<td>90</td>
<td>300</td>
<td>52%</td>
<td>48%</td>
</tr>
<tr>
<td>AS300 RANDSEQ</td>
<td>76</td>
<td>81</td>
<td>67</td>
<td>76</td>
<td>300</td>
<td>49.3%</td>
<td>50.7%</td>
</tr>
</tbody>
</table>
3.4. Discussion

In this chapter we developed an RNA localization assay that utilizes smFISH to quantify the stress granule localized fraction of reporter RNAs. Importantly, we show that smFISH is incompatible with transient transfection of plasmids as the fluorescent probes associate with large cytoplasmic clusters of plasmid DNA that contain the hybridizing sequence (Figure 3.2). However, genomic integration of the reporter RNA into the AAVS locus circumvents these problems and allows detailed quantification of reporter RNA variants (Figure 3.3). The luciferase RNA is a good reporter for assaying how sequence alters stress granule recruitment as it is not endogenous to mammalian cells and has a relatively low baseline recruitment to stress granules at ~15%.

Using this system, we determined that the addition of certain 300 base sequences are sufficient to alter the extent of stress granule localization. Two constructs, luc-A-3UTR and luc-N-ORF show significantly increased localization to stress granules (Figure 3.4). This is consistent with examples from the stress granule transcriptome in which splice variants, some of which differ by inclusion/exclusion of a single exon, can differ in stress granule targeting (Khong et al., 2017). Importantly, this shows that certain sequences are capable of overcoming the primary sorting encoded by RNA length.

These experiments also suggest that the recruitment is not simply due to direct antisense-sense base-pairing. The enhanced recruitment of luc-N-ORF to stress granules remained unaffected by knockdown of NORAD (Figure 3.7). This implies that some other property of the altered RNA is affecting recruitment; binding sites for
proteins, RNA structure, nucleotide content, or post-transcriptional RNA modifications could all potentially influence this process.

Notably, these experiments serve as a proof of principle for the use of reporter constructs in conjunction with smFISH for determining the rules governing RNA localization during stress. Classically, hybrid reporter constructs or genetic approaches have been used to find localization elements within RNAs for decades (Chartrand et al., 1999; Kislauskis et al., 1994; Bullock and Ish-Harowicz, 2001; Miyagawa et al 2012; Martin and Ephrussi, 2009). We intend to modify our luciferase reporter to determine a) if certain sequences within NORAD are responsible for its robust targeting to stress granules and b) whether tethering certain RNA-binding proteins to our reporter will alter localization under stress. Recently, global approaches have also become increasingly accessible tools for approaching questions of RNA localization. For example, appending multiple lncRNA derived oligos to a cytoplasmically localized RNA in parallel revealed nuclear enrichment sequences, including a cytosine-rich motif (Shukla et al., 2018). Similar global approaches may be used in conjunction with our luciferase reporter to further characterize stress granule targeting sequences.

RNA localization is important as it often allows function to be enacted quickly and economically and has consequently been studied in a variety of contexts. Distinct localization of RNAs is a widespread phenomenon. For example, in Drosophila embryos ~70% of RNAs have distinct spatial patterns (Lecuyer et al., 2007). The proper localization of RNAs is often reliant upon cis-acting RNA localization elements,
including sequence and structure, RNA-binding proteins, and active transport along microtubules (reviewed in Martin and Ephrussi (2009)). This is likely to be true for RNA localization to various RNP granules. For example, a mutation altering the structure of \textit{CLN3} mRNA in \textit{Ashbya gossypii} promotes its interaction with \textit{SPA2} and \textit{BNI1} mRNAs \textit{in vitro} and in cells and abrogates its localization to distinct RNP granules (Langdon et al., 2018). Other, less direct methods of RNA localization have been also been identified. For example, global analysis of different RNA isoforms in neurons revealed that the use of distal alternative last exons is strongly associated with localization to the neurite projection (Taliaferro et al., 2016). Additionally, in yeast, promoter sequences were found to direct cytoplasmic localization of mRNAs during starvation (Zid et al., 2014).

A greater understanding of the rules governing RNA localization during stress will hopefully lead to the ability to actively program RNA targeting to stress granules. This will allow functional assays to decipher the consequences of stress granule localization on an RNA as well as overall stress granule function.
3.5. Materials and Methods

*Plasmid transient transfection*

Cells were transfected with jetPRIME (Polyplus Transfection) according to manufacturer’s instructions. Briefly, U-2 OS cells were split into 6-well plates at 25% confluency. The next day, 2 µg plasmid with appropriate luciferase reporter construct (pBVT55-pBVT61) were added to 200 µL jetPRIME buffer. The mixture was vortexed and spun down briefly. 4 µL jetPRIME was added to plasmid mixture, vortexed and spun down briefly. The jetPRIME transfection mixture was incubated at room temperature for 10 minutes, then added dropwise onto the cells. Plates were incubated at 37°C for 4 hours and then media was replaced with regular growth medium. Plates were returned to the incubator. Where doxycycline induction was required, 1 µg/mL doxycycline was added to medium when the media was replaced 4 hours post-transfection. Cells were analyzed after 24-48 hours.

*Genomic integration into AAVS locus*

Cells were transfected with 1 µg CRISPR/Cas9 plasmid (pDY59) in conjunction with 1 µg appropriate luciferase reporter construct (pBVT62-pBVT68) as described above. Transfection of pDY59 alone was used as a negative control. 24 hours following transfection, cells were split from a 6-well plate to a 10 cm dish. After another 24 hours, media was replaced with new media containing 1 µg/mL puromycin to begin selection for cells with genomic integration. Following 24 hours with puromycin selection, media was replaced with new puromycin. After all cells were dead in the negative control plate, media was replaced with fresh media lacking puromycin for 48 hours. This is an optional step to help get rid of any residual plasmid. Puromycin was then added for
another 48 hours to finalize the selection. Single colony selection was not done for these experiments.

**Stellaris smFISH probes**

Custom Stellaris FISH probes against PEG3, AHNAK, NORAD and firefly luciferase transcripts ([Table 3.2](#)) were designed with Stellaris RNA FISH Probe Designer (Biosearch Technologies, Petaluma, CA), available online at [http://www.biosearchtech.com/stellaris-designer](http://www.biosearchtech.com/stellaris-designer) (version 4.2). PEG3, AHNAK and NORAD smFISH probes, labeled with Quasar 670 dye, and firefly luciferase probes, labeled with Quasar 570, were ordered from Stellaris (Biosearch Technologies, Petaluma, CA).

**Sequential IF and FISH**

Sequential immunofluorescence and smFISH on fixed U-2 OS cells was performed with Stellaris buffers or homemade buffers (Dunagin et al. 2015) according to the manufacturer’s protocol: ([https://biosearchassets.blob.core.windows.net/assets/bti_custom_stellaris_immunofluorescence_seq_protocol.pdf](https://biosearchassets.blob.core.windows.net/assets/bti_custom_stellaris_immunofluorescence_seq_protocol.pdf)).

Briefly, U-2 OS cells were seeded on sterilized coverslips in 6-well tissue culture plates. At ~80% confluence, media was exchanged 1 hour before experimentation with fresh media. If stressed, U-2 OS cells were treated with 500 µM NaAsO2 for 1 hour. After stressing cells, the media was aspirated and the cells were washed with pre-warmed 1x PBS. The cells were fixed with 500 µL 4% paraformaldehyde for ten
<table>
<thead>
<tr>
<th>Probe #</th>
<th>PEG3</th>
<th>AHNAK</th>
<th>NORAD</th>
<th>Firefly luciferase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cactgtctgatgccatacaga</td>
<td>tccttcacctccatccttc</td>
<td>tgggaagagaggtgcttg</td>
<td>tcttcgagtgggtgaatgg</td>
</tr>
<tr>
<td>2</td>
<td>ttccccatagatcattcctcc</td>
<td>atgttccggctctgtgctcc</td>
<td>tagagattgaagaccaacgccc</td>
<td>tagcgctctcatggctgttg</td>
</tr>
<tr>
<td>3</td>
<td>tgtaattctccacagcagatg</td>
<td>tgaccttctctgcctctcag</td>
<td>cattctcaccatatctcattt</td>
<td>cgtcggtaaagggcaggg</td>
</tr>
<tr>
<td>4</td>
<td>tccggtgctactgaaagaatg</td>
<td>cacatccacagtctaggc</td>
<td>atacacaggtccctccataaa</td>
<td>gtaatgctccactgtagatg</td>
</tr>
<tr>
<td>5</td>
<td>ttgggtatgatctcagcatcc</td>
<td>atgtctatgtctttggtctc</td>
<td>ccatctgaagggctagatg</td>
<td>cgaacgctctctcgaagta</td>
</tr>
<tr>
<td>6</td>
<td>gcatagtttttagacccctga</td>
<td>gatcctgaacctcggctc</td>
<td>ttatcatacgtggaaccc</td>
<td>atagcgctctagctcctg</td>
</tr>
<tr>
<td>7</td>
<td>cgatttggacgctgtgaca</td>
<td>cgagggcagctctgtacac</td>
<td>gtcatctcagaaagacagttcctcc</td>
<td>gatccgtatgtggttatgttata</td>
</tr>
<tr>
<td>8</td>
<td>tctgacatttctctgggaatct</td>
<td>agaaatggtcccttgctcg</td>
<td>aaacgtagctgtgctcttc</td>
<td>aagctattctcgtgcaacac</td>
</tr>
<tr>
<td>9</td>
<td>gtgtcacaatgtgtagcggctc</td>
<td>tggagcttctggctctgaa</td>
<td>tccgctctcgaagacatcc</td>
<td>ccaacagggcatagaagac</td>
</tr>
<tr>
<td>10</td>
<td>aagggccggcaactcatttaga</td>
<td>cattgcacactgacgctc</td>
<td>cccccccaaatatattcgcttta</td>
<td>acagccacacggcatgaacag</td>
</tr>
<tr>
<td>11</td>
<td>ctagttagttgttcttggca</td>
<td>ccaactctgggaacctgga</td>
<td>aacaggtggtgcatagaggct</td>
<td>ttgtatagtgctgtgalagcttg</td>
</tr>
<tr>
<td>12</td>
<td>tggattttccattgctaat</td>
<td>ctacccccaaagagagatg</td>
<td>gtcaattagagctctgttcctt</td>
<td>cctttctggctcagcatatac</td>
</tr>
<tr>
<td>13</td>
<td>ctttttaaggctgaccagg</td>
<td>accacttcctccagagatgg</td>
<td>atacactggcaaactcttta</td>
<td>ttgcacgtgtgagatctttt</td>
</tr>
<tr>
<td>14</td>
<td>gccatctgaagtcacagaa</td>
<td>tgcgactttcctgtggcttg</td>
<td>atgtgaacattctgctgctag</td>
<td>tggtagctggctgtgtact</td>
</tr>
<tr>
<td>15</td>
<td>gtctcattcctctcagggag</td>
<td>cccgttggagacacaaaaa</td>
<td>catacatcggctctctaaa</td>
<td>aggtgtacatgtttgtaagag</td>
</tr>
<tr>
<td>16</td>
<td>taggggttagagctaatgttt</td>
<td>cgcttttgtgtctggc</td>
<td>tacttgtcacaagttggct</td>
<td>ggtggcaaatgggaagtgcag</td>
</tr>
<tr>
<td>17</td>
<td>agtgggaatcttctggttt</td>
<td>tgtgatgtcagaagctggc</td>
<td>tactttttgcataattttcc</td>
<td>cacgaagtcgtactcgtgta</td>
</tr>
<tr>
<td>18</td>
<td>ctgctctagatattggcctg</td>
<td>aagggcctcaatattggc</td>
<td>atacactggcaacctttta</td>
<td>atgattcgaggccagatgttt</td>
</tr>
<tr>
<td>19</td>
<td>ttgatcctttctggaggtt</td>
<td>cacctcaagtgagttgccc</td>
<td>ctaaatgtggcattttggc</td>
<td>caatccgtatgaccctac</td>
</tr>
<tr>
<td>20</td>
<td>ttggcaggggatactactct</td>
<td>ggccacctttcagggcttag</td>
<td>caaagcagctttcccattcag</td>
<td>tgaatcgagacaacagcggag</td>
</tr>
<tr>
<td>21</td>
<td>catcttcttaaactcacaac</td>
<td>gatgtaccecccaatttg</td>
<td>catacttggggaaccttctt</td>
<td>atgattgtgtgagggagat</td>
</tr>
<tr>
<td>22</td>
<td>gaagtccttacattttgtcc</td>
<td>acatcaatgtcaggggccc</td>
<td>ttcagaagaacagcttttct</td>
<td>tgaatggcaccagacgtgag</td>
</tr>
<tr>
<td>23</td>
<td>tccagtgtaatctctctgt</td>
<td>acagagaaacctgaggtaact</td>
<td>aacattcttgctagaaacc</td>
<td>agcgtggtgaacatgcaag</td>
</tr>
<tr>
<td>24</td>
<td>ttcatggattctctgtgct</td>
<td>cagttttcctttcctttt</td>
<td>gacttaatggttcgccctt</td>
<td>aagcggcagatcaagtagc</td>
</tr>
<tr>
<td>25</td>
<td>ctcctattatgtcattctc</td>
<td>gtaggctgtgtaacattca</td>
<td>ggtactgtctcagagaatttg</td>
<td>aagcggatcatgacacag</td>
</tr>
<tr>
<td>26</td>
<td>atcattctcccatataaagtc</td>
<td>cccaggaacacgtcttcc</td>
<td>cctcgaaccagacaatattat</td>
<td>aagctggtcgaagataagtctc</td>
</tr>
<tr>
<td>27</td>
<td>ggctaagctggaatgtagat</td>
<td>aggctgacatccccagag</td>
<td>aatagaggatctctgtccg</td>
<td>gcaggcagattgaaaccttta</td>
</tr>
<tr>
<td>28</td>
<td>ggggtgcagaaaagatttcctc</td>
<td>ccaccagtagcaaatccta</td>
<td>tgtcatatactggcaaccttt</td>
<td>gaagcattaaatagtgtgag</td>
</tr>
<tr>
<td>29</td>
<td>gaggtgtgattagtagggagga</td>
<td>ctttcattctttctccctc</td>
<td>ctttagggggggtttaaaaa</td>
<td>tactttgtcgtagagagtgc</td>
</tr>
<tr>
<td>30</td>
<td>ctacgatttaagcctgtaatc</td>
<td>ttcaggtcctcttcttt</td>
<td>acagctataagccatctgta</td>
<td>aacgcttttcagttgctttag</td>
</tr>
<tr>
<td>31</td>
<td>atgggagctcagtagttagtt</td>
<td>ttctcagttgagagattttta</td>
<td>aggacattctggcatagaa</td>
<td>ctgtagggtggaagcgtttg</td>
</tr>
<tr>
<td>32</td>
<td>tgagttttcaggtgtttcact</td>
<td>acccaaggctcagatcctca</td>
<td>cctcacttttaacctcact</td>
<td>ttgattttcagggcggactag</td>
</tr>
<tr>
<td>33</td>
<td>tcctaggtaaggttttcta</td>
<td>atctctttcagttaagg</td>
<td>acaaatgctttagaggggtt</td>
<td>cttcgggggtgatcagaatg</td>
</tr>
<tr>
<td>34</td>
<td>gattggtttgattcttgt</td>
<td>ccaggagcagaacccctta</td>
<td>attccatcccaatgcaac</td>
<td>caccttagcctcgaagaag</td>
</tr>
<tr>
<td>35</td>
<td>gtctctactgacattatca</td>
<td>cttaaactcactttgca</td>
<td>tgaatagtctgcatctgct</td>
<td>gttcacaccagctgttctac</td>
</tr>
<tr>
<td>36</td>
<td>agtcaggtgtgtaacacact</td>
<td>ttaagtgccaacttggg</td>
<td>aatctcaggacacagaacc</td>
<td>ttaacgtagcgcctcagat</td>
</tr>
<tr>
<td>37</td>
<td>ttatctatcgctcctacagc</td>
<td>tctatgcactcctggag</td>
<td>ccatgtctagatatatgaga</td>
<td>gtcgatgagcgcgtttag</td>
</tr>
<tr>
<td>38</td>
<td>ggttcacaaaaagccaatcc</td>
<td>ttcccttaggtttttg</td>
<td>tgtgagccctgtttatcat</td>
<td>cggtcacagatcagagaag</td>
</tr>
<tr>
<td>39</td>
<td>tcaaccttcattagccacta</td>
<td>aagccctaggccctgcaaa</td>
<td>cccattctgaagcctgagat</td>
<td>cttgttattgtcaggtctt</td>
</tr>
<tr>
<td>40</td>
<td>cataaggctgatgctgacagc</td>
<td>gttttcccaggaagagtct</td>
<td>catgtcacaagatgtatccta</td>
<td>acatagttcagagccttcct</td>
</tr>
<tr>
<td>41</td>
<td>acatggaggtaggggtcagc</td>
<td>catagagatctctacaggt</td>
<td>tgacatacactgcagaggt</td>
<td>agctctttggcgttgtaaac</td>
</tr>
<tr>
<td>42</td>
<td>tatattaggactccctccttt</td>
<td>tttcacttttaggtggcgc</td>
<td>acacatctgcactatcctct</td>
<td>tgtgcacagcacaacaacac</td>
</tr>
<tr>
<td>43</td>
<td>ttcatctcagaaacctggg</td>
<td>agtggtacacatctaccacc</td>
<td>acaggccaaagagatgcataca</td>
<td>ctttaatgagaatctcgg</td>
</tr>
<tr>
<td>44</td>
<td>acctacaccataccatcaca</td>
<td>tcttccactcactccttc</td>
<td>aatgctgcatctccttaatg</td>
<td>tcttcgagttggttagaatgg</td>
</tr>
<tr>
<td>45</td>
<td>ttttccatgtagtaaacct</td>
<td>aagccattgcagagttcctga</td>
<td>tagcgttcaggcttttgtg</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>ctatcctggagttggaaccgg</td>
<td>aatggccacacgcctctag</td>
<td>cgtgtaaaggccgttgg</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>agtgtcacagataggtttta</td>
<td>gggactgagactgtacagta</td>
<td>qaqtccactcctcagat</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>atatgtctttggatgggttggg</td>
<td>atcctctagatcctgtg</td>
<td>cgaacgtctcatcgaagta</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2. Stellaris single molecule FISH pooled probes. Probe sequences for *AHNAK*, *NORAD*, *PEG3* and *firefly luciferase*.
minutes at room temperature. After fixation, cells were washed twice with 1x PBS, permeabilized in 0.1% Triton X-100 in 1X PBS for five minutes and washed once with 1X PBS.

For IF detection of stress granules, coverslips were incubated in 5 µg/mL mouse α-G3BP primary antibody (Abcam ab10361) for 1 hour. Coverslips were washed three times with 1x PBS for 10 minutes each wash. Then cells were incubated in 1:200 goat anti-mouse IgG (H+L), Alexa Fluor 405 secondary antibody (Thermo Fisher Scientific A-31553). Again, coverslips were washed three times with 1x PBS for 10 minutes each wash. Then, cells were treated with smFISH Buffer A for 5 min. Coverslips were transferred to a humidifying chamber with smFISH probes and placed in the dark at 37C for 16 hours. Coverslips were placed in Buffer A for 30 minutes in the dark, washed with Buffer B for 5 minutes and placed onto a slide with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs, H-1200). For assays requiring quantification of smFISH probes in stress granules, VECTASHIELD Antifade Mounting Medium without DAPI was used (Vector Labs, H-1000). The nucleus is still visible by lack of G3BP staining.

**Microscopy**

Fixed U-2 OS cells stained by immunofluorescence and smFISH, were imaged using a wide field DeltaVision Elite microscope (Applied Biosystems) with a 100x objective and a PCO Edge sCMOS camera. At least five images with 20 Z-sections were taken for each experiment. All images in the manuscript are processed by FIJI (Schindelin et al., 2012) or Imaris (Bitplane).
Construct design

14 plasmids were used in this study (Table 3.3, Table 3.4). Tet-on luciferase was purchased from Addgene (Plasmid #64127). Vector used as backbone for AAVS targeting of Cas9 was an AAVS-TDP43 plasmid in which the TDP43 was excised by restriction digest and replaced with luciferase constructs. 300 base antisense sequences and random sequence were ordered on a G-block.

Luciferase assay

Transfected and untransfected cells were treated with 1 µg/mL doxycycline for 0, 2, 8 and 24 hours. Promega Dual-Luciferase Reporter Assay was done according to manufacturer’s instructions through the firefly luciferase step. Equal amounts of protein were used for each reading.

Imaris identification of smFISH spots

To measure the fraction of smFISH spots in stress granules, deconvolved images were analyzed using Bitplane Imaris image analysis software as described previously (Khong et al., 2018).

siNORAD

Lipofectamine RNAiMAX was used with siRNA against NORAD according to manufacturer’s directions. For our experiments, 1 µg/mL doxycycline to induce our reporter constructs was added at the same time as the siRNAs.
<table>
<thead>
<tr>
<th>Plasmid Name</th>
<th>Description</th>
<th>Selection Marker</th>
<th>Construction Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBVT55</td>
<td>Tet-on luciferase reporter</td>
<td>Amp</td>
<td>Purchased from Addgene. Plasmid # 64127</td>
</tr>
<tr>
<td>pBVT56</td>
<td>Tet-on luciferase with 300 base antisense AHNAK sequence in ORF</td>
<td>Amp</td>
<td>PCR off G-block with oBVT116 and oBVT117. PCR pBVT55 with oBVT114 and oBVT115. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT57</td>
<td>Tet-on luciferase with 300 base random sequence in ORF</td>
<td>Amp</td>
<td>PCR off G-block with oBVT118 and oBVT119. PCR pBVT55 with oBVT114 and oBVT115. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT58</td>
<td>Tet-on luciferase with 300 base antisense NORAD sequence in ORF</td>
<td>Amp</td>
<td>PCR off G-block with oBVT120 and oBVT121. PCR pBVT55 with oBVT114 and oBVT115. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT59</td>
<td>Tet-on luciferase with 300 base antisense AHNAK sequence in 3’UTR</td>
<td>Amp</td>
<td>PCR off G-block with oBVT124 and oBVT125. PCR pBVT55 with oBVT122 and oBVT123. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT60</td>
<td>Tet-on luciferase with 300 base random sequence in 3’UTR</td>
<td>Amp</td>
<td>PCR off G-block with oBVT126 and oBVT127. PCR pBVT55 with oBVT122 and oBVT123. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT61</td>
<td>Tet-on luciferase with 300 base antisense NORAD sequence in 3’UTR</td>
<td>Amp</td>
<td>PCR off G-block with oBVT128 and oBVT129. PCR pBVT55 with oBVT122 and oBVT123. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT62</td>
<td>Tet-on luciferase reporter for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT55 with oBVT164 and oBVT165. Cut Josh’s TDP-43 AAVS plasmid with NotI and BgIII. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT63</td>
<td>Tet-on luciferase with 300 base antisense AHNAK sequence in ORF for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT56 with oBVT164 and oBVT165. Cut Josh’s TDP-43 AAVS plasmid with NotI and BgIII. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT64</td>
<td>Tet-on luciferase with 300 base random sequence in ORF for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT57 with oBVT164 and oBVT165. Cut Josh’s TDP-43 AAVS plasmid with NotI and BgIII. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT65</td>
<td>Tet-on luciferase with 300 base antisense NORAD sequence in ORF for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT58 with oBVT164 and oBVT165. Cut Josh’s TDP-43 AAVS plasmid with NotI and BgIII. Gibson assembly.</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td>Selection Markers</td>
<td>Construction Notes</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>pBVT66</td>
<td>Tet-on luciferase with 300 base antisense AHNAK sequence in 3’UTR for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT59 with oBVT164 and oBVT166. Cut Josh’s TDP-43 AAVS plasmid with Not1 and BglIII. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT67</td>
<td>Tet-on luciferase with 300 base random sequence in 3’UTR for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT60 with oBVT164 and oBVT168. Cut Josh’s TDP-43 AAVS plasmid with Not1 and BglIII. Gibson assembly.</td>
</tr>
<tr>
<td>pBVT68</td>
<td>Tet-on luciferase with 300 base antisense NORAD sequence in 3’UTR for genomic integration into AAVS locus</td>
<td>Amp, Puro</td>
<td>PCR off pBVT61 with oBVT164 and oBVT167. Cut Josh’s TDP-43 AAVS plasmid with Not1 and BglIII. Gibson assembly.</td>
</tr>
</tbody>
</table>

**Table 3.3. Plasmids used in Chapter 3.** Table includes the names, description, selection marker and construction notes for all plasmids used in this section.
<table>
<thead>
<tr>
<th><strong>Oligo</strong> (fwd/rev)</th>
<th><strong>Sequence</strong></th>
<th><strong>Priming Tm (red sequence)</strong></th>
<th><strong>Full length Tm</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>oBVT114 (fwd)</td>
<td>aagattcaatctgcccctg</td>
<td>---</td>
<td>58.8C</td>
</tr>
<tr>
<td>oBVT115 (rev)</td>
<td>atagcttgcaagctgcg</td>
<td>---</td>
<td>62.5C</td>
</tr>
<tr>
<td>oBVT116 (fwd)</td>
<td>tgcgcagcttgcaagactatgggaccttgatattttcttcaggcatc</td>
<td>56.3C</td>
<td>67.2C</td>
</tr>
<tr>
<td>oBVT117 (rev)</td>
<td>aacaggccagattgaacttggtgacctcaagggcccc</td>
<td>60.1C</td>
<td>70.2C</td>
</tr>
<tr>
<td>oBVT118 (fwd)</td>
<td>tgcgcagcttgcaagacttagaacagaagttacctagg</td>
<td>47.6C</td>
<td>65.2C</td>
</tr>
<tr>
<td>oBVT119 (rev)</td>
<td>aacaggccagattgaacggcagttggttgcag</td>
<td>48.1C</td>
<td>64.7C</td>
</tr>
<tr>
<td>oBVT120 (fwd)</td>
<td>tgcgcagcttgcaagacttatagtcagccgcaagagatg</td>
<td>48.3C</td>
<td>64.9C</td>
</tr>
<tr>
<td>oBVT121 (rev)</td>
<td>aacaggccagattgaacttgtggacaaagttattattttgc</td>
<td>48.4C</td>
<td>63.7C</td>
</tr>
<tr>
<td>oBVT122 (fwd)</td>
<td>taaagcagaacttggttcctgag</td>
<td>---</td>
<td>61C</td>
</tr>
<tr>
<td>oBVT123 (rev)</td>
<td>ataagactttacttttgtggc</td>
<td>---</td>
<td>56.3C</td>
</tr>
<tr>
<td>oBVT124 (fwd)</td>
<td>ccaagaagtaaagatcttatgggagctttgatattttcttcaggcatc</td>
<td>56.3C</td>
<td>62.5C</td>
</tr>
<tr>
<td>oBVT125 (rev)</td>
<td>aaaaaaagttctgctttaggtgacctcaagggcccc</td>
<td>60.1C</td>
<td>65.5C</td>
</tr>
<tr>
<td>oBVT126 (fwd)</td>
<td>ccaagaagtaaagatcttatgaaaaagtaagggcccttag</td>
<td>47.6C</td>
<td>59.3C</td>
</tr>
<tr>
<td>oBVT127 (rev)</td>
<td>aaaaaaagttctgctttagtcatgtgtaaacaagacg</td>
<td>48.1C</td>
<td>60.2C</td>
</tr>
<tr>
<td>oBVT128 (fwd)</td>
<td>ccaagaagtaaagatcttatagtcagccgcaagagatg</td>
<td>48.3C</td>
<td>59C</td>
</tr>
<tr>
<td>oBVT129 (rev)</td>
<td>aaaaaaagttctgctttagtggacaagaagttattattttgc</td>
<td>48.4C</td>
<td>59.7C</td>
</tr>
<tr>
<td>oBVT164 (fwd)</td>
<td>gagaattggctagcagacttggtaccgaattcgcaccat</td>
<td>57.7C</td>
<td>67.1C</td>
</tr>
<tr>
<td>oBVT165 (rev)</td>
<td>gatccctgaggagcggccggcttgttaataagagctttactttcttggc</td>
<td>54.5C</td>
<td>69.4C</td>
</tr>
<tr>
<td>oBVT166 (rev)</td>
<td>gatccctgaggagcggccggctctctttactgtgctggacctcaag</td>
<td>52.9C</td>
<td>72C</td>
</tr>
<tr>
<td>oBVT167 (rev)</td>
<td>gatccctgaggagcggccggctctctttactgtgctggacctcaag</td>
<td>53.5C</td>
<td>68.9C</td>
</tr>
<tr>
<td>oBVT168 (rev)</td>
<td>gatccctgaggagcggccggctctctttactgtgctggacctcaag</td>
<td>54.1C</td>
<td>70.4C</td>
</tr>
</tbody>
</table>

Table 3.4. Oligos used for constructing plasmids. Table includes the oligo names, sequence, and Tm values.
Distance analysis

Distances were determined with the help of Bitplane Imaris Imaging Analysis software. Imaris reassembled the Z-stacks into a 3-D image automatically. The smFISH signal coming from the nuclei was masked. The spot creation wizard identified reporter and target RNA spots using the following parameters: 1) a fixed xy diameter spot size of 200 nm and 2) a manually determined fluorescent quality threshold. Upon identification of smFISH spots, the spot creation wizard provides all x, y, z coordinates for the center of each smFISH spot in an excel spreadsheet. The coordinates of all luciferase smFISH spots within stress granules were exported and the distances between luciferase spots and all smFISH spots of the target RNAs were calculated by applying the distance formula between two points in 3D-space. The distance from a given luciferase RNA to its nearest target RNA were recorded.
CHAPTER 4: INTRINSICALLY DISORDERED REGIONS OF PROTEINS: ROLE IN PROTEIN TARGETING TO GRANULES AND TRANSITIONS TO AMYLOID-LIKE STRUCTURES

4.1. Abstract

Cytoplasmic mRNP granules have emerged as ubiquitous features of eukaryotic cells. mRNP granules are enriched in RNA-binding proteins that contain intrinsically disordered regions (IDRs). Two common models in the literature are that 1) the IDRs are responsible for targeting their respective proteins into RNP granules, and 2) that a high local concentration of specific IDRs in RNP granules can enhance the rate of formation of stable, amyloid-like assemblies, possibly explaining the connection of stress granules to degenerative diseases. In this work both of these models are tested. Generally, IDRs are neither sufficient nor necessary for the targeting of proteins to P-bodies, which is supportive of a model wherein most RNP granule components are recruited by RNA binding or specific protein-protein interactions. Moreover, in our model system, concentrating proteins with amyloid prone IDRs to RNP granules does not increase the rate of stable assembly formation. This later result suggests that either the model system is not mimicking natural RNP granules, and/or the concentration of IDRs within RNP granules does not increase the rate of formation of amyloid-like assemblies.
4.2. Introduction

P-bodies and stress granules are cytoplasmic ribonucleoprotein (RNP) granules that are most prevalent in times of cellular stress. They are cell biological markers for different mRNP states as they require RNA unengaged in translation to form (Liu et al., 2005; Sheth and Parker, 2003; Teixeira et al., 2005; Pillai et al., 2005). While P-bodies are presumed sites of mRNA degradation (Sheth and Parker, 2003) and/or repression (Hubstenberger et al., 2017), stress granules accumulate primarily long, non-translating RNAs (Buchan and Parker, 2009; Khong et al., 2017).

Several observations have led the field of RNP granules to focus heavily on intrinsically disordered regions (IDRs) of proteins found within RNP granules. First, IDRs are enriched in proteins involved in RNA binding and processing as well as in RNP granules (Decker et al., 2007; Reijns et al., 2008; King et al., 2012; Li et al., 2013, Jain et al., 2016). While approximately 1% of the known human protein-coding genes contain a canonical RNA recognition motif (RRM), nearly 12% of protein-coding genes that have a predicted IDR also contain an RRM (King et al., 2012). Second, IDRs have been shown to promote RNP granule assembly and dynamics. For example, in S. cerevisiae the deletion of Dhh1 or Lsm4’s IDRs reduce P-body assembly in certain genetic backgrounds (Decker et al., 2007; Protter et al., 2018). This assembly defect can be rescued by substitution with other IDRs (Decker et al., 2007; Protter et al., 2018). Similarly, in some cases the IDR is important for regulating the localization of RNP granule components, as is the case with TIA-1 and RBM14 in stress granules and paraspeckles, respectively (Gilks et al., 2004; Hennig et al., 2015). Third, many IDRs are capable of self-assembling in vitro (Appendix Table A.1). Weak multivalent
interactions of IDRs in isolation have been shown to facilitate liquid-liquid phase separation in vitro which has been suggested to be a model system for RNP granule assembly (Nott et al., 2015; Lin et al., 2015; Molliex et al., 2015; Elbaum-Garfinkle et al., 2015; Zhang et al., 2015).

Three observations connect the IDR of RNP granule components to several degenerative diseases. First, mutations that cause inclusion body myopathy (IBM), amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD) and other degenerative disorders are often in the IDRs of several granule components (Buchan et al., 2013; Kim et al., 2013; Taylor et al., 2016; Mackenzie et al., 2017) Often these mutations are predicted to increase the ability of these proteins to form amyloid structures (Kato et al., 2012; Kim et al., 2013). Moreover, since the self-assembly of IDRs into a LLPS can increase the rate of amyloid initiation, a model has been proposed wherein the high local concentration of IDRs in RNP granules might create an environment that triggers initiation of amyloid structures, leading to a toxic and pathological state (Ramaswami et al., 2013).

Here, in contradiction to the model above, we show that IDRs are dispensable for P-body targeting. In conjunction with data showing that a) IDR assembly in vitro can be disrupted by promiscuous interactions with other proteins, and b) IDRs can promote phase separation and RNP granule assembly when paired with specific interactions by folded domains, we propose a more balanced model of assembly in which IDRs can contribute promiscuous interactions to RNP granules. In addition, we illustrate that many IDRs will form SDS-resistant assemblies upon their robust
overexpression. By placing IDRs on an inducible promoter, the formation of large stable assemblies can be tracked over time. Using this assay, we are unable to detect a change in assembly formation as a result of RNA binding; however, we note this may be a consequence of the IDR tested. Finally, we show that the presence of an RNA-binding domain on an assembly-prone IDR can alter assembly composition by recruiting target RNAs to the assembly.
4.3. Results

*IDRs are neither sufficient nor necessary for P-body localization.*

If IDRs are forming amyloid structures or undergoing phase separations within a P-body or stress granule, one would expect the isolated IDR to be recruited to a pre-existing granule. To examine how IDRs of yeast proteins affect their targeting to P-bodies, we examined if IDRs within Lsm4, Dhh1, Pop2, and Ccr4 (Figure 4.1A) were necessary and/or sufficient for their recruitment to P-bodies. The IDRs of Lsm4, Dhh1, Pop2, and Ccr4 were fused separately to either GFP or mCherry. IDR-fusion proteins were expressed in yeast co-expressing a chromosomally GFP-tagged P-body component or containing a secondary plasmid containing an mCherry tagged P-body component. P-bodies were then induced by glucose deprivation for 15 minutes, and the percentage of P-bodies containing the IDR fusion protein was recorded.

For example, clear enrichment in P-bodies was detectable for full length Dhh1 (Figure 4.1B). However, the Dhh1 IDR was not sufficient for P-body localization (Figure 4.1B). Similarly, the IDRs of Lsm4, Pop2, and Ccr4 were insufficient for recruitment to P-bodies (Figure 4.1C). The removal of these IDRs from their full-length proteins had little to no effect on their recruitment to P-bodies (Figure 4.1B, C). However, for the already poorly localized Pop2, deletion of the IDR did have a noticeable impact on localization (Figure 4.1C). Thus, the IDRs of Lsm4, Dhh1, Ccr4, and Pop2 are not sufficient their recruitment into P-bodies but may contribute in cases where recruitment is already poor, such as Pop2.
Figure 4.1. IDR are neither sufficient nor necessary for P-body localization. (A) Domain structures of the yeast proteins Dhh1, Lsm4, Ccr4, and Pop2. (B) Dhh1-GFP variant fusions were expressed in Edc3-mCherry expressing yeast, to mark P-bodies. Images were taken by fluorescence microscopy after 10 minutes glucose deprivation. (C) Quantification of the percentage of P-bodies that exhibited colocalization with the expressed fusion protein. GFP was fused to the N-terminus of Dhh1, Ccr4, and Pop2 variants. These variants were co-transformed with Edc3-mCherry. mCherry was fused to the C-terminus of the Lsm4 variants, which were expressed in cells with genomically-tagged Dcp2-GFP, another P-body protein. >100 foci were counted per condition.
A balanced model of RNP granule assembly

Concurrent with this work, other experiments from our lab provided evidence that at least some IDRs function to promote RNP granule assembly, both in cells and model biochemical systems, through weak interactions that require coupling with specific interactions from other protein domains. First, in vitro assembly of FUS, hnRNPA1, and eIF4GII IDRs was disrupted by the addition of generic proteins and yeast lysates, arguing that IDRs interact nonspecifically with generic proteins (Protter et al., 2018). However, when tethered to the PTB RNA-binding protein, which phase separates in the presence of its target RNA, promiscuous IDRs can promote LLPS, even in the presence of competitor proteins (Protter et al., 2018). Finally, the C-terminal P/Q rich IDR of Dhh1, which is required for P-body assembly in S. cerevisiae when P-body assembly is impaired (Rao and Parker, 2017), can be replaced by the IDRs of human Lsm4, hnRNPA1, or FUS, or by specific LEA proteins from brine shrimp or nematodes (Protter et al., 2018). This argues that even though IDRs are insufficient in themselves for RNP granule localization or LLPS in the presence of competing interactions, IDRs can decrease the critical concentration for phase separation primarily driven by more specific interactions (Figure 4.2). This highlights that in a phase diagram describing an assembly based on specific and promiscuous interactions, the addition of promiscuous interactions can shift the system from an unassembled state to an assembled state (Figure 4.2B)

An assembly mechanism for RNP granules driven by specific interactions and aided by promiscuous interactions of IDRs has predictions for how components would be recruited to granules. Specifically, one would predict that generally IDRs would
Figure 4.2. Model of RNP Granule Assembly and Contributions of IDRs. (A) RNP granules assemble by a wide variety of specific and nonspecific interactions. (B) A theoretical phase diagram depicting how the addition of nonspecific, IDR-driven interactions could decrease the critical concentration of assembly for higher-order structures. This figure was created by David Protter.
not be sufficient to target a protein to an RNP granule, unless they contained a short linear motif (SLiM) with a specific binding partner. Moreover, IDRs would not be required for recruitment to a granule, although they could affect the partition coefficient (the concentration of a component within versus outside of a granule). This is consistent with our observations regarding various proteins and their recruitment to P-bodies (Figure 4.1).

Various IDRs form stable assemblies upon their overexpression

A current model is that the formation of stress granules can promote amyloid initiation by forming a high local concentration of IDRs (Ramaswami et al., 2013). In order to examine this possibility, we used two assays to monitor the formation of stable aggregates within the cell. First, constructs containing an IDR of interest were fused to GFP allowing quantification of the formation of microscopically visible assemblies (Figure 4.3A). Second, the detection of large, SDS-resistant assemblies can be visualized using semi-denaturing detergent agarose gel electrophoresis (SDD-AGE; Figure 4.3B). In this assay, lysates are added to a sample buffer with a final concentration of 2% SDS, which disrupts most protein-protein interactions but leaves highly stable interactions, e.g., from amyloids (Kryndushkin et al., 2003; Halfmann and Lindquist, 2008).

To validate our assays, we utilized hnRNPA1, a human RNA-binding protein with an ALS-causing mutation within the IDR that also increases the prion nature of this protein (Kim et al., 2013). This protein was chosen for two main reasons. First, three hnRNPA1-IDR variants had already been constructed: wild-type, a mutant with
Figure 4.3. Overexpression of various IDRs can drive the assembly of SDS-resistant aggregates \textit{in vivo}. Two assays can be used to visualize the formation of large, stable assemblies: (A) foci formation visible under a microscope and (B) Semi-Denaturing Detergent Agarose Gel Electrophoresis (SDD-AGE). Monomer bands can be seen on the bottom and SDS resistant higher order assemblies above the monomers. (C) Foci form in cells in a manner consistent with their predicted ability to form amyloids. Left panels show representative images of yeast transfected with a GFP negative control, hnRNPA1$_{\text{hypo}}$ deficient in amyloid formation, hnRNPA1$_{\text{WT}}$ and hnRNPA1$_{\text{hyper}}$-prone to assemble. Graph on the right shows the quantification of the \%GFP expressing cells that contain foci. Inset is SDD-AGE gel of all hnRNPA1 variants as well as the GFP control. (D) Various IDRs from proteins important in RNA biology were tagged with GFP and overexpressed with a constitutive GPD promoter. Lsm4, Puf1, Puf2, Sup35N, Sup35NM, Pub1, and Ccr4 exhibit foci formation in a fraction of the expressing cells. (E) Quantification of the \%GFP expressing cells that contain foci. Over 100 cells counted for each IDR tested. (F) A handful of these IDRs were tested for their ability to form SDS resistant aggregates. IDRs that form SDS resistant assemblies correlate with IDRs that form foci upon their overexpression.
increased propensity for amyloid assembly (hnRNPA1_{hyper}), and a mutant that eradicates the ability of this protein to form amyloid fibers \textit{in vitro} (hnRNPA1_{hypo}; Kim et al., 2013). Second, all three variants are capable of undergoing phase separation \textit{in vitro}, although their ability to form amyloid-like fibers \textit{in vitro} is affected by the specific mutations (Lin et al., 2015). The varying abilities to form these two different structures \textit{in vitro} will allow us to correlate the structures these domains are forming \textit{in vivo} and whether our assays are specific to stable assembly formation.

To validate our assays, we utilized hnRNPA1, a human RNA-binding protein with an ALS-causing mutation within the IDR that also increases the prion nature of this protein (Kim et al., 2013). This protein was chosen for two main reasons. First, three hnRNPA1-IDR variants had already been constructed: wild-type, a mutant with increased propensity for amyloid assembly (hnRNPA1_{hyper}), and a mutant that eradicates the ability of this protein to form amyloid fibers \textit{in vitro} (hnRNPA1_{hypo}; Kim et al., 2013). Second, all three variants are capable of undergoing phase separation \textit{in vitro}, although their ability to form amyloid-like fibers \textit{in vitro} is affected by the specific mutations (Lin et al., 2015). The varying abilities to form these two different structures \textit{in vitro} allows us to correlate the structures these domains are forming \textit{in vivo} and whether our assays are specific to stable assembly formation.

The overexpression of hnRNPA1 IDR variants that have been shown to be form amyloid structures robustly \textit{in vitro} were much more likely to aggregate into microscopically visible foci \textit{in vivo} (Figure 4.3C). hnRNPA1_{hypo} IDRs were diffuse in the cytoplasm, hnRNPA1_{WT} IDRs were mostly diffuse with foci visible in only a few
cells, and hnRNPA1_{hyper} IDRs were found in foci significantly more often (Figure 4.3C). In addition, only the overexpression of hnRNPA1_{hyper} IDR resulted in enough SDS resistant assemblies to be visualized by SDD-AGE (Figure 4.3C inset). Thus, in the case of hnRNPA1, the ability of the IDR to form an assembly upon its overexpression correlates with the amyloid forming properties of the domain and not with its ability to undergo phase separation \textit{in vitro}.

To test the propensity of other IDRs to form stable assemblies in cells, we extended our analysis to various IDRs from yeast proteins found in either P-bodies or stress granules (Figure 4.3D). Specifically, we analyzed the IDRs from Lsm4, Ccr4, Pop2, Ngr1, eIF4G2, Pub1, Puf1, Puf2, Puf3, Puf4 and Puf5. IDRs were fused to a GFP marker and placed under a GPD promoter to facilitate their robust overexpression. As a positive control we included the IDR domain of Sup35, a known prion forming protein in yeast, and as a negative control we expressed GFP alone with the same promoter. Yeast were collected during log phase and analyzed under the microscope for the presence of microscopically visible foci. IDRs were scored by the percentage of GFP expressing cells that contained at least one focus (Figure 4.3E).

It was determined that some, but not all, of the IDRs tested were capable of forming microscopically visible foci. (Figure 4.3D, E). IDRs that formed foci in $>15\%$ of cells expressing GFP included Lsm4, Ccr4, Puf1, Puf2, Pub1, and our positive control Sup35. Domains that seldom, if ever, formed foci included Ngr1, eIF4G2, Puf3, Puf4, and Puf5 or GFP alone. We also observed that detection of SDS-resistant assemblies by SDD-AGE correlated with aggregates observed by microscopy, with
only the IDRs that formed foci having detectable stable assemblies by SDD-AGE analysis (Figure 4.3F). This data supports the notion that IDRs may assemble into stable amyloid-like aggregates when at a sufficient concentration within the cell.

Importantly, not all IDRs will undergo this transition. Running the proteins through PAPA, an algorithm designed to find potential amyloid forming regions (Toombs et al., 2012), revealed that only the proteins that were bioinformatically predicted to have amyloid forming capabilities in the region we overexpressed formed foci and stable assemblies upon overexpression (Table 4.1).

**Does a high local concentration of IDRs, facilitated by RNA binding, lead to increased rates of stable aggregate formation?**

A leading hypothesis in the field is that constitutive stress granules, either through loss of clearance or persistent stress, allow IDRs the time and proximity necessary to nucleate the formation of stable structures (Ramaswami et al., 2013; Wolozin et al., 2014).

To test these predictions, we first aimed to find a more quantifiably-reliable assay. Although certain constructs always formed SDS-resistant structures, there was a lot of heterogeneity in the extent of assembly as quantified by the signal in the SDS-resistant higher-order assemblies over the total signal for a lane. By expressing the constructs under a galactose inducible promoter allowed for the visualization of foci formation over time (Figure 4.4A, B) that was mirrored by the consistent accumulation of SDS resistant assemblies (Figure 4.4C, D). This provided us with a kinetic assay
Table 4.1. Foci formation from overexpression of IDRs correlates with stable, amyloid-like assemblies, and not with phase separation. Each IDR is included in the table with available data for % of cells with foci upon overexpression, the presence of SDS-resistant assemblies, the ability to phase separate \textit{in vitro}, and it’s predicted ability to form amyloids based off PAPA algorithm. n.d.= no data.

<table>
<thead>
<tr>
<th>IDR</th>
<th>% cells with foci</th>
<th>SDS resistant assemblies</th>
<th>Phase separation (Lin et al., 2015)</th>
<th>Predicted amyloid formation (Toombs et al., 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP control</td>
<td>0</td>
<td>n.d.</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>Lsm4</td>
<td>45</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ccr4</td>
<td>33</td>
<td>++</td>
<td>n.d.</td>
<td>+</td>
</tr>
<tr>
<td>Pop2</td>
<td>3</td>
<td>n.d.</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>Pub1</td>
<td>51</td>
<td>+++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ngr1</td>
<td>0</td>
<td>n.d.</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>eIF4G2</td>
<td>0</td>
<td>n.d.</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Puf1</td>
<td>45</td>
<td>n.d.</td>
<td>n.d.</td>
<td>+</td>
</tr>
<tr>
<td>Puf2</td>
<td>15</td>
<td>n.d.</td>
<td>n.d.</td>
<td>+</td>
</tr>
<tr>
<td>Puf3</td>
<td>0</td>
<td>n.d.</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>Puf4</td>
<td>0</td>
<td>-</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>Puf5</td>
<td>0</td>
<td>n.d.</td>
<td>n.d.</td>
<td>-</td>
</tr>
<tr>
<td>Sup35NM</td>
<td>30</td>
<td>n.d.</td>
<td>n.d.</td>
<td>+</td>
</tr>
<tr>
<td>Sup35N</td>
<td>37</td>
<td>n.d.</td>
<td>n.d.</td>
<td>+</td>
</tr>
<tr>
<td>hnRNPA1\textsubscript{WT}</td>
<td>2</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>hnRNPA1\textsubscript{hyper}</td>
<td>17</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>hnRNPA1\textsubscript{hypo}</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 4.4. Galactose induction shows increased SDS resistant aggregates over time that correlate with microscopically visible foci. (A) Foci formation is dependent on the expression of nucleating IDR. Images show representative images of yeast at 4, 8, 12, 16 and 24 hours post galactose induction. (B) Quantification of the percent of cells with foci at 4, 8, 12, 16 and 24 hours post galactose induction. Over 100 cells counted. (C) SDD-AGE gel of lysates from GFP induced cells over time. Signal blown out to show lack of SDS-resistant assemblies. (D) SDD-AGE gel of lysates from Lsm4IDR-GFP induced cells over time.
for the formation of SDS resistant assemblies, which we then used to determine if creating high local concentration of protein within the cell changed the rate of accumulation.

To mimic the high local concentration of IDRs and RNA found within granules, we fused Lsm4IDR-GFP and Pub1IDR-GFP to MS2, an RNA binding protein with high affinity for specific stem loops in RNA (Peabody, 1993). These constructs were then expressed in cells with and without RNA containing 12xMS2 stem loops. Unfortunately, fusion of MS2 to Lsm4IDR disrupted its ability to form SDS-resistant assemblies in vivo and therefore was not subject to further analysis (Figure 4.5A, B). Pub1IDR-GFP-MS2 formed SDS-resistant assemblies upon galactose induction as expected (Figure 4.5C).

An important observation was that the rate of SDS resistant assemblies formed from the Pub1IDR-GFP-MS2 protein was similar in cells with and without 12xMS2-RNA (Figure 4.5D). This indicates that creating a high local concentration of this protein did not change its rate of initiation of amyloid-like assemblies.

These results imply, at least for the case of Pub1, that binding RNA does not enhance the transition to a pathological state. Since we are expressing our Pub1IDR-GFP-MS2 construct to very high levels, it is also possible that the nucleation of amyloid-like assemblies is saturated. In this case, detecting a change in stable assembly formation would require more tempered expression or an IDR with decreased assembly propensity.
Figure 4.5. RNA-binding has no effect on the formation of SDS-resistant assemblies formed by Pub1IDR-GFP-MS2. (A) Cells that contained or lacked RNA tagged with 12 MS2 binding sites were transformed with plasmids containing Lsm4IDR-GFP-MS2. Lsm4IDR-GFP-MS2 failed to form stable higher order assemblies. (B) Lsm4IDR-GFP will form SDS resistant assemblies upon induction. Addition of MS2 is inhibiting this effect. (C) Pub1IDR-GFP-MS2 can form SDS resistant assemblies. (D) There is no large change in the percent Pub1IDR-GFP signal in SDS resistant assemblies in the presence of MS2 containing RNA. (D) Pub1IDR-GFP-MS2 foci colocalize with MS2-2xdsRed, a marker for the 12xMS2 RNA in happy and stressed cells. (E) IDR driven foci formed in the absence of the MS2 fusion do not colocalize with 12xMS2 containing RNA.
To verify that the Pub1IDR-GFP-MS2 was binding our RNA of interest we co-transformed cells with MS2-2xdsRed. Importantly, MS2-2xdsRed colocalized with Pub1IDR-GFP-MS2 assemblies in stressed and unstressed conditions (Figure 4.5E). In contrast, Pub1IDR-GFP assemblies did not localize 12xMS2-RNA (Figure 4.5F). This implies that the aggregation of RNA binding proteins can also accumulate their corresponding RNAs. This could be a contributing factor in disease as aggregates contain not only mis-folded protein, but also sequester RNAs, leading to loss of function at both levels.
4.4. Discussion

In this chapter we propose a new model for the role of IDRs in RNP granules in which IDRs are not the sole drivers of assembly, but rather contribute to assembly through the addition of multivalent interactions (Figure 4.2). This is supported by the observations that in many cases IDRs are neither sufficient or necessary for the recruitment to P-bodies (Figure 4.1), which is contradictory to an IDR-centric assembly model. In addition, concurrent experiments illustrated even when IDRs are insufficient to form assemblies alone, promiscuous IDRs can decrease the critical concentration for phase separation driven by more specific interactions (Protter et al., 2018).

Although it is well documented that IDRs play crucial roles in granule assembly in some cases, there is also evidence consistent with our findings. For example, although mutations of tyrosines within the IDR of FUS abrogate FUS recruitment to stress granules (Kato et al., 2012), deletion of the entire IDR allows proper stress granule localization (Bentmann et al., 2012). Similarly, the IDR of Pat1 is dispensable for its localization to P-bodies (Pilkington and Parker, 2008). In addition, there are many examples of specific interactions important for RNP granule assembly that cannot be overlooked. For example, Edc3 and G3BP dimerization are critical for P-body and stress granule assembly, respectively (Decker et al., 2007; Kedersha et al., 2016). It should be noted that IDRs can also be part of specific interactions, through the interactions of short linear motifs (SLiMs) with wellfolded domains of other proteins (reviewed in Jonas and Izaurralde (2013), Fromm et al., 2014; Harigaya et al., 2010).
IDRs are also of interest to the RNP granule field as they have been implicated in pathology. A favored hypothesis in the field is that the increased local concentration of IDRs within RNP granules can stimulate the formation of stable, amyloid-like assemblies. Through the use of fluorescently tagged IDRs and SDD-AGE, we show that some, but not all, IDRs are capable of forming large stable assemblies in vivo upon robust overexpression (Figure 4.3). Despite the ability of many IDRs to form very stable assemblies when in high concentrations, it is intriguing that RNP granules maintain high levels of protein dynamicity. It is likely that there are active processes that ensure stable amyloid-like structures are not forming in the context of normal granule assembly, as suggested by chaperones playing roles in disassembly and multiple ATP-driven machines present in the stress granule proteome (Walters et al., 2015; Jain et al., 2016). In addition, the ability of IDRs to interact promiscuously with RNAs and other proteins may serve as competitors for interaction sites, thereby preventing nucleation of amyloid-like interactions in most cases.

By inducing the expression of our IDRs, we created an assay in which we were able to track the formation of stable assemblies over time (Figure 4.4). Targeting our constructs to high local concentrations on RNA showed no significant difference in the formation of stable assemblies (Figure 4.5), however we note that this could be an artifact of our assay and IDR tested. Although we observed no change in our system, others have shown that in vitro assemblies of IDRs are capable of maturing over time (Han et al., 2012; Kato et al., 2012; Lin et al., 2015; Molliex et al., 2015; Patel et al., 2015; Zhang et al., 2015; Peskett et al., 2018). In addition, the presence of RNA has
been observed to promote protein assembly in many cases (Burke et al., 2015; Zhang et al., 2015; Schwartz et al., 2013, Lin et al., 2015). Many IDRs are capable of promiscuous interactions with RNAs, suggested by observations that some IDRs cross-link to RNA in vivo (Castello et al., 2016) and some IDRs bind RNA in vitro (Mayeda et al., 1994; Lin et al., 2015; Molliex et al., 2015). This may facilitate an interesting balance between IDRs, RNAs, and the physical properties of the assemblies they form within the cell.
4.5. Materials and Methods

Microscopy and Quantification for P-body Colocalization

Cells were grown at 30°C to OD600 of 0.3-0.5 in minimal media with 2% glucose as a carbon source and with necessary amino acid dropout to maintain plasmids (Table 4.2) and express constructs. Cells were stressed by glucose deprivation for 15 minutes before cells were concentrated for immediate microscopic examination at room temperature. All images were acquired on a DeltaVision epi-fluorescence microscope, equipped with an SCMOS camera. All images underwent deconvolution using DeltaVision’s algorithm and were quantified using FIJI. To optimize yeast colocalization accuracy, single plane images were used and analysis were done in a blind manner. P-bodies were identified using protein markers (either Dcp2-GFP or Edc3-mCherry). Corresponding enrichment of the construct within the P bodies was then assessed manually. Manual assessment was required due to differential strengths of cytoplasmic signals between cells arising from stochastic variation and/or potentially different copy numbers of plasmids between cells.

IDR galactose induction assays

IDR constructs from plasmids were induced by switching growth media to 2% galactose. Cells were grown in galactose for 24 hours and monitored microscopically at 8, 12, 16 and 24hrs. At these time points, pellets were also spun down and flash frozen for SDD-AGE analysis.

Semi-denaturing detergent agarose gel electrophoresis (SDD-AGE)
SDD-AGE was done as previously described (Halfmann and Lindquist, 2008). Briefly, a 1.5% agarose gel was created with 1x TAE. SDS was added to the melted agarose to 0.1%. The gel was submerged in 1x TAE containing 0.1% SDS.

Samples were prepared from 10 mL of culture. Cells were harvested by centrifugation at 2000 RCF for 5 min at room temperature. Cells were harvested in speroplasting solution and incubated for approximately 30 min at 30°C. Cells were centrifuged again and the supernatant was removed. Pelleted spheroplasts were resuspended in 1 mL lysis buffer and vortexed on high speed for 2 minutes. Cellular debris was pelleted by centrifugation at 4000 rcf for 2 minutes. Supernatant was transferred to a new container. 4X sample buffer was added to the lysates to generate lysates containing 1X sample buffer and incubated at room temperature for 5 minutes. The gel was loaded and run at 30V for 7-9 hours.

Protein in the gel was then transferred to a nitrocellulose membrane using capillary transfer. Nitrocellulose membrane was placed on top of a stack of blotting paper. The gel was placed on top of the membrane followed by three pieces of pre-wetted blotting paper. A pre-wet wick was placed across the stack and into two elevated trays of TBS. The assembled transfer stack was covered with an additional plastic tray bearing extra weight and the transfer proceeded overnight.

Following the transfer, the membrane can be processed by standard Western blotting techniques. To detect IDR constructs we used α-GFP Mouse IgG1 primary antibody
(Biolegend Cat#902605) and anti-mouse IgG, HRP-linked secondary (Cell Signaling Technology #7076).

**Solutions and Buffers for SDD-AGE**

Spheroplasting solution: 1.2 M D-sorbitol, 0.5 mM MgCl$_2$, 20 mM Tris pH 7.5, 50 mM BME, 0.5 mg/mL Zymolase 100T. Zymolase and BME should be added fresh. Lysis buffer: 100 mM Tris pH7.5, 50 mM NaCl, 10 mM BME, protease inhibitors. Add protease inhibitor tab and BME fresh. 4X Sample Buffer: 2X TAE, 20% glycerol, 8% SDS, bromophenol blue to preference.

**Quantification of SDD-AGE gels**

Gel images were opened in FIJI (Schindelin et al., 2012). Each lane was selected. After putting a box around the first lane, hit command + 1. Command +2 will allow the selection of subsequent lanes of the same size. Once all lanes are selected, command + 3 will plot the lanes. Lines were drawn across the bottom of each peak. The wand tool was selected to highlight the area under the curve. In this way, the signal from just the SDS-resistant fraction can be calculated as well as the total signal (SDS-resistant fraction + monomer). Divide the SDS-resistant fraction signal by the total to get the % of signal in the SDS-resistant fraction.
<table>
<thead>
<tr>
<th>Plasmid Name</th>
<th>Description</th>
<th>Selection Marker</th>
<th>Construction Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRP2778 (pMR203)</td>
<td>Pop2- IDR. pGFP-C-POP2(1-156): N-terminal 156 residues of Pop2p in frame with GFP-tag in pGFP-C-FUS (CEN6, URA3 shuttle vector with MCS for C-terminal GFP tagging). Under control of MET25 promoter.</td>
<td>Amp, Ura</td>
<td>Generously provided by Jean Begg's lab (Reijns et al., 2008)</td>
</tr>
<tr>
<td>pRP2781 (pMR211)</td>
<td>Dhh1-IDA, pGFP-N-DHH1ΔC: Dhh1p aa 1-427 in frame with GFP-tag in pGFP-N-FUS (CEN6, URA3 shuttle vector with MCS for N-terminal GFP tagging). Under control of MET25 promoter.</td>
<td>Amp, Ura</td>
<td>Generously provided by Jean Begg's lab (Reijns et al., 2008)</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td>Selection Markers</td>
<td>Source</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------------</td>
<td>--------</td>
</tr>
<tr>
<td>pRP2783 (pMR214)</td>
<td>Pop2 (full length). POP2 coding sequence in frame with GFP-tag in pGFP-N-FUS (CEN6, URA3 shuttle vector with MCS for N-terminal GFP tagging). Under control of MET25 promoter.</td>
<td>Amp, Ura</td>
<td>Generously provided by Jean Begg’s lab (Reijns et al., 2008)</td>
</tr>
<tr>
<td>pRP2785 (pMR218)</td>
<td>Ccr4-IDRA. Ccr4p aa 148-837 in frame with GFP-tag in pGFP-C-FUS (CEN6, URA3 shuttle vector with MCS for C-terminal GFP tagging). Under control of MET25 promoter.</td>
<td>Amp, Ura</td>
<td>Generously provided by Jean Begg’s lab (Reijns et al., 2008)</td>
</tr>
<tr>
<td>pBVT24</td>
<td>Full length Lsm4-mcherry under control of Lsm4 promoter</td>
<td>Amp, Ura</td>
<td>PCR genomic DNA from BY4741 with oBVT15 and oBVT16 to amplify Lsm4 and its endogenous promoter. Cut pRP1699 with SacI and BamHI. Homologous recombination in yeast.</td>
</tr>
<tr>
<td>pBVT25</td>
<td>Lsm4IDRA-mcherry under control of Lsm4 promoter</td>
<td>Amp, Ura</td>
<td>PCR genomic DNA from BY4741 with oBVT15 and oBVT17. Homologous recombination into pRP1686, then cut this plasmid and pRP1699 with SacI and BamHI. Pieces were ligated together.</td>
</tr>
<tr>
<td>pBVT26</td>
<td>Lsm4IDR-mcherry under control of Lsm4 promoter</td>
<td>Amp, Ura</td>
<td>PCR genomic DNA from BY4741 with oBVT78 and oBVT15 to amplify the Lsm4 promoter. PCR genomic DNA with oBVT79 and oBVT16 to amplify Lsm4’s IDR. Cut pRP1699 with SacI and BamHI. Three part homologous recombination in yeast.</td>
</tr>
<tr>
<td>pRP1686</td>
<td>Lsm4-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Constructed by Carolyn Decker</td>
</tr>
<tr>
<td>pRP1687</td>
<td>Ccr4-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Constructed by Carolyn Decker</td>
</tr>
<tr>
<td>Plasmid ID</td>
<td>Description</td>
<td>Selection Markers</td>
<td>Constructor</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>pRP1693</td>
<td>Pop2-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1689</td>
<td>Pub1-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1690</td>
<td>Ngr1-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1688</td>
<td>eIF4G2-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1694</td>
<td>Puf1-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1695</td>
<td>Puf2-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1696</td>
<td>Puf3-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1697</td>
<td>Puf4-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1698</td>
<td>Puf5-IDR-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1691</td>
<td>Sup35N-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pRP1692</td>
<td>Sup35NM-GFP expressed under constitutive GPD promoter</td>
<td>Amp, Leu</td>
<td>Carolyn Decker</td>
</tr>
<tr>
<td>pBVT08</td>
<td>GFP-hnRNPA1&lt;sub&gt;hyper&lt;/sub&gt; expressed under the constitutive GPD promoter in a 2µ plasmid</td>
<td>Amp, Ura</td>
<td></td>
</tr>
<tr>
<td>pBVT09</td>
<td>GFP-hnRNPA1&lt;sub&gt;wt&lt;/sub&gt; expressed under the constitutive GPD promoter in a 2µ plasmid</td>
<td>Amp, Ura</td>
<td></td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td>Selection Marker</td>
<td>Notes</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>pBVT10</td>
<td>GFP-hnRNPA1&lt;sub&gt;hypo&lt;/sub&gt; expressed under the constitutive GPD promoter in a 2µ plasmid</td>
<td>Amp, Ura</td>
<td>Original plasmid containing hnRNPA1 variants were received from Paul Taylor’s lab (Kim et al., 2013). Variants were removed by digestion with NruI and HindIII. pRP1699 was cut with NruI and Hind III as new vector backbone. Pieces were ligated together.</td>
</tr>
<tr>
<td>pBVT29</td>
<td>Lsm4IDR-GFP under a galactose inducible promoter</td>
<td>Amp, Ura</td>
<td>PCR amplify LSM4-IDR with oBVT88 and oBVT91. Cut PCR and backbone vector with galactose promoter with XbaI and HindII. Ligate.</td>
</tr>
<tr>
<td>pBVT30</td>
<td>Pub1IDR-GFP under a galactose inducible promoter</td>
<td>Amp, Ura</td>
<td>PCR amplify Pub1-IDR with oBVT89 and oBVT91. Cut PCR and backbone vector with galactose promoter with XbaI and HindII. Ligate.</td>
</tr>
<tr>
<td>pBVT48</td>
<td>Lsm4IDR-GFP-MS2 under galactose inducible promoter</td>
<td>Amp, Ura</td>
<td>PCR amplify pBVT43 with oBVT88 and oBVT91. Cut PCR product and pBVT28 with XbaI and XhoI. Ligate together.</td>
</tr>
<tr>
<td>pBVT49</td>
<td>Pub1IDR-GFP-MS2 under a galactose inducible promoter</td>
<td>Amp, Ura</td>
<td>PCR amplify pBVT44 with oBVT89 and oBVT91. Cut PCR product and pBVT28 with XbaI and XhoI. Ligate together.</td>
</tr>
<tr>
<td>pBVT50</td>
<td>GFP-MS2 under a galactose inducible promoter</td>
<td>Amp, Ura</td>
<td>PCR amplify pBVT44 with oBVT92 and oBVT91. Cut PCR product and pBVT28 with XbaI and XhoI. Ligate together.</td>
</tr>
<tr>
<td>pRP1614</td>
<td>Edc3 mCherry</td>
<td>Amp, Leu</td>
<td>Obtained from Carolyn Decker</td>
</tr>
<tr>
<td>pMS2-CP-2xdsRED</td>
<td>MS2-2xdsRed</td>
<td>Amp, His</td>
<td>Obtained from Jennifer Garcia</td>
</tr>
</tbody>
</table>

Table 4.2. Plasmids used in Chapter 4. Table includes the names, description, selection marker and construction notes for all plasmids used in this section.
<table>
<thead>
<tr>
<th>Oligo (fwd/rev)</th>
<th>Sequence</th>
<th>Priming Tm (red seq)</th>
<th>Full length Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td>oBVT15 (fwd)</td>
<td>ttaaccctcactaaaggaacagaagctggagctccagagtaatgctgcgttc</td>
<td>48.5C</td>
<td>67.9C</td>
</tr>
<tr>
<td>oBVT16 (rev)</td>
<td>cctgctacgtaattaaccggagatccgaattcgacctttgtgg</td>
<td>48C</td>
<td>68.5C</td>
</tr>
<tr>
<td>oBVT17 (rev)</td>
<td>tactgtaattaaccgggatccgtttgacttgaatcttactgtaatattatcttgcaatt</td>
<td>54.8C</td>
<td>66.2C</td>
</tr>
<tr>
<td>oBVT79 (fwd)</td>
<td>agaaacacaatagaatatattttactcagcagaaattaactcacaacataact</td>
<td>52.2C</td>
<td>63.0C</td>
</tr>
<tr>
<td>oBVT88 (fwd)</td>
<td>atccatattcagatcggaatgcagcaaattaactcacaacaa</td>
<td>53.7C</td>
<td>65C</td>
</tr>
<tr>
<td>oBVT89 (fwd)</td>
<td>atccatattcagatcggaatgcaacaaattaactcacaacagc</td>
<td>51.9C</td>
<td>62.7C</td>
</tr>
<tr>
<td>oBVT91 (rev)</td>
<td>gccccccccctcgaggtcgta</td>
<td>---</td>
<td>68.6C</td>
</tr>
<tr>
<td>oBVT92 (fwd)</td>
<td>atccatattcagatcggaatgcaacaaattaactcacaacagc</td>
<td>60.8C</td>
<td>67.5C</td>
</tr>
</tbody>
</table>

**Table 4.3. Oligos used for constructing plasmids.** Table includes the oligo names, sequence, and Tm values.
CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS
5.1. Summary

In this thesis, we examined multiple aspects of ribonucleoprotein (RNP) granules including domains and sequences that drive component recruitment, mechanisms of granule assembly, and potential pathways for disease.

Mechanisms of granule assembly and component recruitment

RNP granules are non-membranous assemblies that are ubiquitous in eukaryotic cells. Importantly, despite their dynamicity, they are enriched for specific proteins and RNAs (Jain et al., 2016; Khong et al., 2017) creating an environment within the granule that is compositionally different from the surrounding milieu.

The RNP granule field has primarily focused on the assembly and recruitment roles of intrinsically disordered domains (IDRs) for three main reasons. First, IDRs are enriched in RNP granules, and, more generally, in RNA-binding proteins (Decker et al., 2007; Reijns et al., 2008; King et al., 2012; Li et al., 2013, Jain et al., 2016). Second, IDRs have been shown in some cases to be important for RNP granule formation or the targeting of specific components (Decker et al., 2007; Gilks et al., 2004; Hennig et al., 2015; Protter et al., 2018). Finally, multiple studies have shown that IDRs are capable of self-assembling in vitro (Appendix Table A.1).

We show that in multiple cases, IDRs are neither necessary nor sufficient for protein targeting to P-bodies (Chapter 4). One prediction of an IDR-centric model is that IDRs would be strongly recruited to RNP granules by homotypic or heterotypic interactions with other IDRs, and we show that this is not the case. In combination with
observations showing that IDRs promote assembly when paired with underlying specific interactions but are disrupted by generic proteins when in isolation (Protter et al., 2018), we propose a model in which IDRs contribute to assembly by adding promiscuous interactions that strengthen the underlying network (Chapter 4).

The role of RNA in the formation of RNP granules has been largely overlooked until recently. Similar to IDRs, we show that RNAs are also capable of assembling in vitro (Chapter 2). Notably, the RNAs that are enriched in assemblies formed in vitro under conditions that mimic the cellular environment largely recapitulate the stress granule transcriptome (Chapter 2). The major determinant for the enrichment of RNAs in assemblies in vitro or in stress granules is length. The simplest explanation for this observation is that longer RNAs have more potential sites of RNA-RNA interactions with other RNAs and are therefore more likely to assemble. We propose a model, based on our data as well as supporting data from the literature in which RNA-RNA interactions, in conjunction with protein-protein and protein-RNA interactions, are important for the assembly of RNP granules (Chapter 1).

If RNA-RNA interactions are helping to drive assembly, it reasons that RNA-RNA interactions can directly recruit RNAs to RNP granules. Analysis of the stress granule transcriptome revealed a number of short antisense RNAs that were recruited to stress granules despite their short length. Importantly, their corresponding sense RNAs were overwhelmingly RNAs enriched in stress granules (Chapter 2). To test if complementary base-pairing can recruit an RNA to stress granules we created reporter RNAs with complementary sequence to enriched RNAs. We demonstrate that
relatively short sequences can alter an RNAs localization during stress in a manner that overrides the primary length determinant (Chapter 3). However, we also show that, at least in the case of our reporter RNA, antisense-sense base-pairing does not seem to be the driver as the target RNA is dispensable for the enrichment of our reporter. This suggests that some other mechanism is responsible for the targeting, which could include RNA-binding proteins, RNA structure, RNA modification, or promiscuous RNA-RNA interactions.

*RNP granules as a stepping stone for disease*

Stress granules are one of the most studied RNP granules because of their connections to disease. Again, the primary focus of the field has been on the contributions of IDRs. Multiple mutations have been identified in the IDRs of RNA-binding proteins that are causative in degenerative disorders (Buchan et al., 2013; Kim et al., 2013; Taylor et al., 2016; Mackenzie et al., 2017). Often these mutations are predicted to increase the ability of these proteins to interact with each other and form stable, amyloid-like assemblies (Kato et al., 2012; Kim et al., 2013). Moreover, the biophysical properties of self-assembled IDRs can drastically change over time, often resulting in the formation of amyloid-like fibers (Han et al., 2012; Kato et al., 2012; Lin et al., 2015; Molliex et al., 2015; Patel et al., 2015; Zhang et al., 2015; Peskett et al., 2018). This compelling evidence has been combined into a model wherein the high local concentration of IDRs in RNP granules may create an environment that triggers initiation of stable and pathogenic assemblies (Ramaswami et al., 2013).
We show that many, but not all, IDRs are capable of forming large, SDS-resistant aggregates upon their robust overexpression (Chapter 4). We created a kinetic assay for the assembly of amyloid-like structures and used it to test the effect of a high local concentration of IDRs facilitated through RNA-binding. In our system, increasing the concentration of Pub1IDR had no effect on the rates of forming SDS-resistant aggregates (Chapter 4). This suggests that either our model system fails to accurately portray the environment of an RNP granule, or that increased localization of IDRs within RNP granules does not increase the nucleation of amyloid-like structures.

In this thesis I also propose that RNA assembly, driven by GR and PR dipeptide repeats, could also contribute to disease (Chapter 2). Expansions of C9orf72, causative in ALS, result not only in extended RNA transcripts, but also in the production of several dipeptides (GR, PR, GA, PA and GP) produced from repeat-associated non-AUG (RAN) translation (Zu et al., 2013). Intriguingly, only GR and PR dipeptides have been found to be toxic and to trigger stress granule formation (Ash et al., 2013; Zu et al., 2013; Boeynaems et al., 2017; Lee et al., 2013). We determined that the self-assembly of RNAs was promoted by the presence of pathogenic dipeptides GR and PR, but not GP. The production of large, positively charged peptide sequences may trigger RNAs to aberrantly assemble, resulting in disease.

In all, this work provides new perspectives regarding both the mechanisms of RNP granule assembly and of disease progression. The observations and models presented here will serve as a platform for exciting future studies.
5.2. Looking Forward

The work presented here lays important groundwork for future studies regarding the roles of RNA-RNA interactions in RNP granule assembly. Combining the luciferase reporter assay with global methods will elucidate RNA features important for targeting RNAs to stress granules. Conversely, a different reporter with a baseline enrichment in stress granules may be used to determine elements important for exclusion. Reporter RNAs are versatile and will also allow us to interrogate the roles of RNA-binding proteins, length, membrane targeting and RNA processing events on stress granule localization. Importantly, a greater understanding of RNA targeting under normal biological contexts will provide greater insight into how these processes may be mis-regulated in disease.

The future applications of these RNA targeting features, once found, are truly exciting. Targeting sequences can be used in conjunction with live single-molecule tracking to determine how the kinetics of RNA interactions with stress granules changes with the addition of these sequences. A deliberate control of RNA localization under stress will also allow for experiments aimed at determining the function of stress granules. Differentially localized reporters will be useful in determining the consequences of being in a stress granule on an RNA. There is a strong possibility that being trafficked to a stress granule alters an RNA’s fate, perhaps by altering its stability or ability to re-enter translation once stress is removed.

Importantly, the methods and insights gained here will also shed light on the mechanisms in place that regulate all RNP granules. Despite sharing the same space
within the cell (e.g., the cytoplasm), RNP granules can be distinct from each other. The rules that regulate the selective partitioning of RNAs and protein components into one granule and not another may shed light on how these structures evolved.
REFERENCES


http://doi.org/10.1016/j.molcel.2017.02.013


http://doi.org/10.1146/annurev.bi.54.070185.005443


APPENDIX
### Table A.1. Self-Assembly Conditions of Proteins *in vitro*

<table>
<thead>
<tr>
<th>Protein Used</th>
<th>Concentration</th>
<th>Conditions</th>
<th>Comments</th>
<th>Reference</th>
</tr>
</thead>
</table>
| FG repeat domain from Nsp1p (nuclear pore) | 26 mg/mL (400 μM) | 50 mM KOH, 1 mM EDTA followed by quick neutralization with ¼ volume (250 mM K₂HPO₄/KH₂PO₄ pH 7.5; 50 mM KOH; 250 mM acetic acid; 1 mM EDTA) | • Hydrogel  
• F to S mutated repeat domains fail to form hydrogels                                     | (Frey and Görlich, 2007)                      |
| FUS-LC 2-214                          | 60 mg/mL      | 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride | • Hydrogel                                                                                      | (Han et al., 2012)                            |
| FUS-GST, FUS-LC 2-214, hnRNPA2        | 10-50 mg/mL   | Proteins stored in cold                                                    | • Hydrogel  
• Amyloid-like fibers  
• Enhanced formation at low temperatures                                                              | (Kato et al., 2012)                            |
| SH3 and PRM domains (interacting signaling molecules) | ~0.7 mg/mL PRM₄ and ~2.6 mg/mL SH3₄ (molecular concentrations 75 μM. module concentrations 300 μM.) | 150 mM KCl, 1 mM MgCl₂, 1 mM EGTA and 10 mM imidazole pH 7 | • Droplets  
• Can vary the number of modules; with PRM₄ and SH3₄ 25 μM of each is sufficient for assembly.  
• Specific interactions | (Li et al., 2012)                             |
| LAF-1 (P-granule component)           | ~60 μg/mL (800 nM) | 125 mM NaCl, 20 mM Tris pH 7.4, 1 mM DTT                                 | • Droplets  
• Assembly promoted in low salt  
• Same order of magnitude as estimated cytoplasmic [LAF-1] *in vivo* (as determined by fluorescence intensity) | (Elbaum-Garfinkle et al., 2015)                 |
<table>
<thead>
<tr>
<th>Protein/Complex</th>
<th>Concentration</th>
<th>Buffer</th>
<th>Conditions</th>
<th>Properties/Assays</th>
</tr>
</thead>
</table>
| Ddx4 and Ddx4-N-terminus | ~11 mg/mL (140 μM) | 20 mM Tris pH 8.0, 150 mM NaCl, 5 mM TCEP, 37°C | • Droplets  
• Assembly promoted in low salt and low temperature  
• Hypotonic (~150 μM ionic strength) shock disrupts Ddx4 granules in nucleus, but high salt disrupts formation *in vitro*  
• Arginine methylation destabilizes Ddx4 organelles | (Nott et al., 2015) |
| RBM14 PLD, FUS PLD | 40 mg/mL | 20 mM NaCl, 100 mM DTT, 0.5 mM EDTA, 0.1 mM PMSF, 4°C | • Hydrogel | (Hennig et al., 2015) |
| SNAP-IDRs (Pub1, Lsm4, eIF4GII, Tia1, hnRNPA1, Fus), SNAP-hnRNPA1 | ~0.5–2 mg/mL (6-33 μM) | 37.5 mM NaCl, 20 mM Tris pH 7.4, 1 mM DTT | • Droplets  
• Phase separation promoted by low salt  
• Maturation over time | (Lin et al., 2015) |
| His-SUMO-hnRNPA1 | ~ 30 mg/mL (300 μM hnRNPA1) | 100-300 mM NaCl, 75-150 mg/mL Ficoll, 50 mM HEPES, 5 mM DTT | • Droplets  
• Assembly promoted at low temperature  
• Dependent on low complexity C terminal region of hnRNPA1  
• Phase separation promotes fibrillization  
• Formation inhibited by increased salt | (Molliex et al., 2015) |
| FUS-GFP | Gel: ~50 mg/mL (500 μM); Droplet: ~1 mg/mL (10 μM) | Gel: 20 mM Tris-HCl, 200 mM KCl, 0.5 mM EDTA, 1 mM DTT, pH 7.5, 4°C for 48 hrs.; Droplet: 10% dextran, 50 mM Tris-HCl, 500 mM KCl, 2.5% glycerol, 1 mM DTT, pH 7.4 | • Hydrogel and droplets  
• Physical properties of FUS droplets change over time  
• Accelerated by mutation in prion-like domain | (Patel et al., 2015) |
<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
<th>Buffers / Conditions</th>
<th>Processes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GST:FIB-1</td>
<td>~30-650 µg/mL (0.5-10 µM)</td>
<td>150 mM NaCl</td>
<td>• Droplets&lt;br&gt;• Assemblies promoted by low salt</td>
<td>(Berry et al., 2015)</td>
</tr>
<tr>
<td>SUMO/SIM modules</td>
<td>~100-500 µg/mL module (50-90 µM module)</td>
<td>150 mM KCl, 1 mM MgCl₂, 10 mM Imidazole pH 7.0, 1 mM EGTA, 100 mg/mL BSA</td>
<td>• Droplets&lt;br&gt;• Phase diagram position dictates client recruitment</td>
<td>(Banani et al., 2016)</td>
</tr>
<tr>
<td>FUS WT and FUS 12E mutants</td>
<td>0.5 mg/mL (5 µM)</td>
<td>20 mM Tris pH 7.4, 150 mM NaCl</td>
<td>• Droplets and fibers&lt;br&gt;• Phosphorylation impedes phase separation of FUS</td>
<td>(Monahan et al., 2017)</td>
</tr>
<tr>
<td>Arginine containing dipeptides (GR₃₀, PR₃₀ and PR₂₀ as well as RGG domains)</td>
<td>~1-2 mg/mL (250 µM)</td>
<td>100 mM K₂HPO₄/KH₂PO₄ buffer at pH 7, 30% PEG 300 or 0.2-1 mg/mL Poly-L-tyrosine</td>
<td>• Droplets&lt;br&gt;• Arginine-rich peptides undergo LLPS dependent on counterions or polyaromates</td>
<td>(Boeynaems et al., 2017)</td>
</tr>
<tr>
<td>HP1α</td>
<td>~5 mg/mL (200-300 µM)</td>
<td>4C, 75 mM KCl, 20 mM HEPES pH 7.2, 1 mM DTT</td>
<td>• Droplets&lt;br&gt;• Phosphorylation or DNA binding promotes phase separation</td>
<td>(Larson et al., 2017)</td>
</tr>
<tr>
<td>RNA Used</td>
<td>Concentration</td>
<td>Conditions</td>
<td>Comments</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>12-14 nt RNA and DNA oligos</td>
<td>&gt;300 mg/mL</td>
<td>Oligomers dissolved in pure water</td>
<td>• Liquid crystals</td>
<td>(Zanchetta et al., 2008)</td>
</tr>
<tr>
<td>2Luc mRNA (3000 nts)</td>
<td>~2 μg/mL (2 nM)</td>
<td>20% PEG 35K, KCl, 20 mM Tris-HCl pH 7.5; Deposited onto NiCl₂ pretreated mica for 1 min, rinse with 0.02% uranyl acetate</td>
<td>• PEG 35K induced mRNA aggregation whereas PEG 1000 was insufficient • KCl increases formation • Microtubule bundling occurs at very low PEG 35K concentrations (1% w/v) compared to mRNA (~20%)</td>
<td>(Bounedjah et al., 2012)</td>
</tr>
<tr>
<td>polyU</td>
<td>500 μg/mL (0.05 wt%, ~0.5-0.6 μM)</td>
<td>0.5 wt % spermine or spermidine (16 mM spermine, or 22 mM spermidine), 5 mM HEPES, pH 7.6, 1 mM MgCl₂</td>
<td>• Droplets • Assembly promoted at higher temperatures • Assembly disrupted by NaCl (after 50-150 mM)</td>
<td>(Aumiller et al., 2016)</td>
</tr>
<tr>
<td>Various lengths of CUG and CAG repeat RNAs</td>
<td>500 μg/mL (Assemblies as low as 25 nM or ~2.1 μg/mL)</td>
<td>10 mM Tris pH 7.0, 10 mM MgCl₂, 25 mM NaCl</td>
<td>• Gels • Doxorubicin and NH₄OAc disrupt gelation in vitro and nuclear foci (speckles) in vivo</td>
<td>(Jain and Vale, 2017)</td>
</tr>
<tr>
<td>(G4C2)₄</td>
<td>~2-4 μg/mL (0.25-0.5 μM)</td>
<td>U-2 OS lysate, 75, 150, 200 mM NaCl, +/-2.5% Ficoll</td>
<td>• Small assemblies of unknown dynamics • Salt reduces assembly</td>
<td>(Fay et al., 2017)</td>
</tr>
<tr>
<td>Homopolymer RNAs (polyU, polyA, polyC, polyG)</td>
<td>500 μg/mL</td>
<td>10% PEG, 750 mM NaCl, 1 mM MgCl₂</td>
<td>• polyU and polyC droplets, polyA viscous droplets, polyG static assemblies • Lower concentrations not tested</td>
<td>(Van Treeck et al., 2018)</td>
</tr>
</tbody>
</table>
| Total yeast RNA | 150 μg/mL | 10% PEG, 150 mM NaCl, 1 mM MgCl$_2$ (for sequencing); also in 223 μM spermine, 1,339 μM spermidine, 150 mM NaCl, 1mM MgCl$_2$ | • Droplets or tangles  
• Range of NaCl for assembly at 10% PEG, even droplets without NaCl  
• NaCl and crowding promote assembly | (Van Treeck et al., 2018) |
<table>
<thead>
<tr>
<th>Protein Concentration</th>
<th>RNA Concentration</th>
<th>Conditions</th>
<th>Comments</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FUS</strong></td>
<td>~2-10 mg/mL (25-60 μM)</td>
<td>prD RNA ~50 μg/mL (3 μM)</td>
<td>PBS pH 7.4, temperature change 37°C to -20°C and back • RNA promotes assembly, but too much RNA is prohibitive</td>
<td>(Schwartz et al., 2013)</td>
</tr>
<tr>
<td><strong>YB-1</strong></td>
<td>5 μg/mL (100 nM); TIA-1 0.6 μg/mL 15 nM</td>
<td>2Luc mRNA 2 μg/mL 20 mM 2-(N-morpholino) ethanesulfonic acid (MES)-KOH, pH 6.8, 1 mM ethylene glycol tetraacetic acid (EGTA), 50 mM KCl, 2 mM MgCl₂, 10 μM spermidine • AFM detectable oligomers • YB-1 dissociates RNA granules in vitro, while TIA-1 promotes assembly</td>
<td>(Bounedjah et al., 2014)</td>
<td></td>
</tr>
<tr>
<td><strong>Whi3</strong></td>
<td>20-500 μg/mL (0.3-7.5 μM)</td>
<td>CLN3 and BNI1 RNA ~2-600 μg/mL (5-300 nM)</td>
<td>50 mM Tris pH 8.0, 1 mM DTT, 150 mM KCl • RNA alters droplet viscosity and dynamics in RNA dependent manner</td>
<td>(Zhang et al., 2015)</td>
</tr>
<tr>
<td><strong>SNAP-IDRs</strong> (Pub1, Lsm4, eIF4GII, Tia1, hnRNPA1, Fus) and SNAP hnRNPA1</td>
<td>~75-110 μg/mL (1.25 μM, but as low as 100 nM)</td>
<td>PTB RNA ligand or hnRNPA1 RNA ligand ~7.5 μg/mL 0.4-0.8 μM (As low as 20 nM) 100-175 mM NaCl, 20 mM imidazole (pH 7.0), 1 mM DTT and 10% glycerol • Droplets • Phase separation of IDRs and hnRNPA1 promoted by RNA • Maturation over time</td>
<td>(Lin et al., 2015)</td>
<td></td>
</tr>
<tr>
<td><strong>hnRNPA1</strong></td>
<td>~10 mg/mL (100 μM)</td>
<td>44nt RNA ~0-1.4 mg/mL (0-100 μM)</td>
<td>50 mM HEPES, 150 mM NaCl, 5 mM DTT, 150 mg/mL Ficoll at 10°C • RNA increases assembly of hnRNPA1</td>
<td>(Molliex et al., 2015)</td>
</tr>
<tr>
<td><strong>GST:FIB-1</strong></td>
<td>~30-650 μg/mL (0.5-10 μM)</td>
<td>Total yeast RNA or a 50-mer of in 150 mM NaCl • RNA lowers protein concentration required for assembly</td>
<td>(Berry et al., 2015)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Concentration</td>
<td>Buffer Conditions</td>
<td>RNA Function</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>LAF-1</td>
<td>~60 µg/mL (800 nM)</td>
<td>125 mM NaCl, 20 mM Tris pH 7.4, 1 mM DTT</td>
<td>Addition of RNA increases fluidity of in vitro assemblies</td>
<td>(Elbaum-Garfinkle et al., 2015)</td>
</tr>
<tr>
<td>PTB</td>
<td>~500 µg/mL (30-70 µM)</td>
<td>100 mM NaCl, 1 mM MgCl₂, 20 mM Imidazole pH 7.0, 50 mM sodium phosphate pH 6.0, 100 mg/mL PEG3350</td>
<td>Phase diagram position dictates client recruitment</td>
<td>(Banani et al., 2016)</td>
</tr>
<tr>
<td>RP₃ or SR₈</td>
<td>200-700 µg/mL; FUS</td>
<td>10 mM Tris-HCl, pH 7.9, 50-225 mM NaCl</td>
<td>RNA promotes droplet assembly at low concentration; charge inversion and disassembly as RNA concentration increases</td>
<td>(Banerjee et al., 2017)</td>
</tr>
<tr>
<td>PR₂₀</td>
<td>~1-2 mg/mL (250 µM)</td>
<td>100 mM K₂HPO₄/KH₂PO₄ buffer at pH 7, no molecular crowder</td>
<td>RNA induces PR₂₀ LLPS in the absence of molecular crowder and in a concentration dependent manner</td>
<td>(Boeynaems et al., 2017)</td>
</tr>
<tr>
<td>GR₁₀ or PR₁₀</td>
<td>~50-60 µg/mL (25 µM)</td>
<td>150 mM NaCl, 1 mM MgCl₂</td>
<td>Addition of R containing dipeptides increases coacervate formation</td>
<td>(Van Treeck et al., 2018)</td>
</tr>
<tr>
<td>Whi3</td>
<td>~500 µg/mL (8 µM)</td>
<td>150 mM KCl, 20 mM Tris pH 8.0, 1 mM DTT</td>
<td>RNA can denote specificity and immiscibility in assembly</td>
<td>(Langdon et al., 2018)</td>
</tr>
<tr>
<td>FUS</td>
<td>365 µg/mL (5 µM), TAF15 248 µg/mL (4 µM), hnRNPA1 513 µg/mL (12.5)</td>
<td>25 mM Tris/HCl buffer (pH 7.4), 75 mM KCl, 1 mM MgCl₂, 1 mM TCEP; for TDP43, 10% Dextran also used</td>
<td>RNA promotes assembly at low concentrations but inhibits assembly at high concentrations</td>
<td>(Maharana et al., 2018)</td>
</tr>
<tr>
<td>μM), EWSR1 136 μg/mL (2 μM), TDP43 430 μg/mL (10 μM)</td>
<td>ABD-Ago2 ~130 μg/mL (1.35 μM), TNRC6B-FL ~60 μg/mL (300 nM)</td>
<td>tRNA (20 μg/mL), 1 nM 8xlet-7 cap labeled target RNA (1 nM)</td>
<td>0.01% CHAPS, 0.5 mM Desthiobiotin, 42.5 mM HEPES pH 7.5, 30 mM Imidazole pH 8, 30 mM NaCl, 85 mM Potassium Acetate, 5 mM 0.5 mM TCEP, Tris pH 8; conditions varied slightly depending on experiment</td>
<td>• Target RNAs are sequestered into miRISC droplets (Sheu-Gruttadauria and MacRae, 2018)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>p-tau441 ~500 μg/mL (2 μM)</td>
<td>Total human brain RNA (0.5 μg/mL)</td>
<td>10 mM HEPES pH 7.4, 10 mM citrate pH 3.0 or 10 mM Tris Base pH 9.5. Varying PEG-800, Ficoll-400 or dextran-70</td>
<td>• Phosphorylation of tau important for droplet formation and tau aggregation; RNA and heparin can trigger assembly (Wegmann et al., 2018)</td>
<td></td>
</tr>
</tbody>
</table>