THESIS

DEFINING THE SEQUENCE REQUIREMENTS FOR RECRUITMENT TO THE NAB3 NUCLEAR GRANULE

Submitted by

Samuel Jackson Snyder

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2025

Master's Committee:

Advisor: Eric D. Ross

Santiago Dipietro
Amanda Woerman

ABSTRACT

DEFINING THE SEQUENCE REQUIREMENTS FOR RECRUITMENT TO THE NAB3 NUCLEAR GRANULE

Compartmentalization of cellular components into liquid-liquid phase separated (LLPS) condensates is a process dependent upon intrinsically disordered protein regions (IDRs). Multiple different types of condensates coexist within a cell yet are able to maintain specificity of IDR partitioning. Different models of IDR function have been developed to describe selective recruitment, yet the fundamental properties of IDRs allowing for specificity remain poorly understood. The amino acids within a peptide govern overall biophysical properties, and thus amino acid composition may describe IDR condensate specificity; for this purpose, yeast stress granules (SGs) and the Nab3 nuclear granule were interrogated and computationally quantified. Domains of similar composition to a SG-localizing synthetic domain readily assembled into SGs, but domains of similar composition to the Nab3 prion-like domain (PrLD) had widely variable degrees of assembly incompletely described by composition alone. To further characterize the sequence requirements of the Nab3 granule, the Nab3 PrLD was functionally replaced with a library of PrLDs and synthetic domains and tested for granule assembly. Reduced dispersion of Q residues correlated with increased Nab3 assembly. Further, the C-terminal oligomerization domain of the Nab3 PrLD is important for condensate formation. Based on this, Nab3 granule formation has composition preference and primary sequence dependence, while SGs are compositionally dependent. Together, the mechanism of IDR assembly is not uniform among yeast condensates, with different sequence features being important based on the system.

ACKNOWLEDGEMENTS

Dropping straight into MRB after a year away from school was a bit of a shock. I remember walking into my first graduate school exam realizing I had truly forgotten how to study and learn.

Fortunately, the Master's cohort picked me up, dusted me off, and became lifelong friends. Thank you, Liam, Jillian, Gretchen, and Astrid.

More than anything, I feel incredibly fortunate to have participated in the coolest research happening in MRB, courtesy of the Ross Lab. Eric, thank you for entrusting me with my own project and allowing me to dictate the direction of my research. Your patient mentoring style was of great comfort in moments when nothing was working.

Sean, I learned more from your python sessions than any class I took during my tenure. With your help, computer science is now a tenable part of my future. I love all the random bioinformatic knick-knacks I've inherited from you.

Kacy, Guthrie, Miriam, and Alessandra, I'll miss our random bench discussions and daily interactions.

Family and friends, thank you for the never ending love and support.

TABLE OF CONTENTS

Introduction

Intrinsically Disordered Proteins

Protein folding is the process through which the linear sequence of amino acids in a peptide dictate three-dimensional structure. Unlike globular proteins, which have well defined topology that defines exact function, intrinsically disordered proteins (IDPs) exist within a conformational ensemble and undergo dynamic structural reorganization[1]. Although IDPs may assume a stable structure when interacting with a binding partner or under certain conditions, the lack of a consistent structure allows IDPs to respond rapidly to physiological stimuli and have diverse regulatory roles. A hallmark of many intrinsically disordered regions (IDRs) is their ability to drive assembly of cellular compartments lacking a membrane, termed biomolecular condensates (BMCs) or simply condensates, through demixing of the cellular milieu by liquid-liquid phase separation (LLPS)[2]. Condensates sequester specific proteins and nucleic acids and therefore offer a solution to managing the diverse reactions that occur within the complex cellular biochemical environment. To facilitate this process, roughly half of human proteins are composed of both globular portions and IDRs and have activities complemented by their structured and unstructured regions[3].

Generally, conformational differences of globular regions and IDRs can be understood based on different sequence features. Stable conformations are driven by packing of hydrophobic amino acids away from the aqueous environment, while IDRs have limited hydrophobicity and are instead enriched for hydrophilic and small residues that allow for favorable solvent interactions and conformational elongation[1].

Low Complexity Domains

Low complexity domains (LCDs), characterized by an enrichment of a single or a small subset of amino acids, are compositionally skewed and therefore have highly variable biophysical properties depending on the amino acid(s) that are enriched. Folded LCDs exist, like the structured leucine zipper motif[4] or the poly-proline helix secondary structure[5], but biased compositions generally disfavor stable conformations and make most LCDs intrinsically disordered, as is the case for charged and polar LCD classes (Fig 1 top)[6].

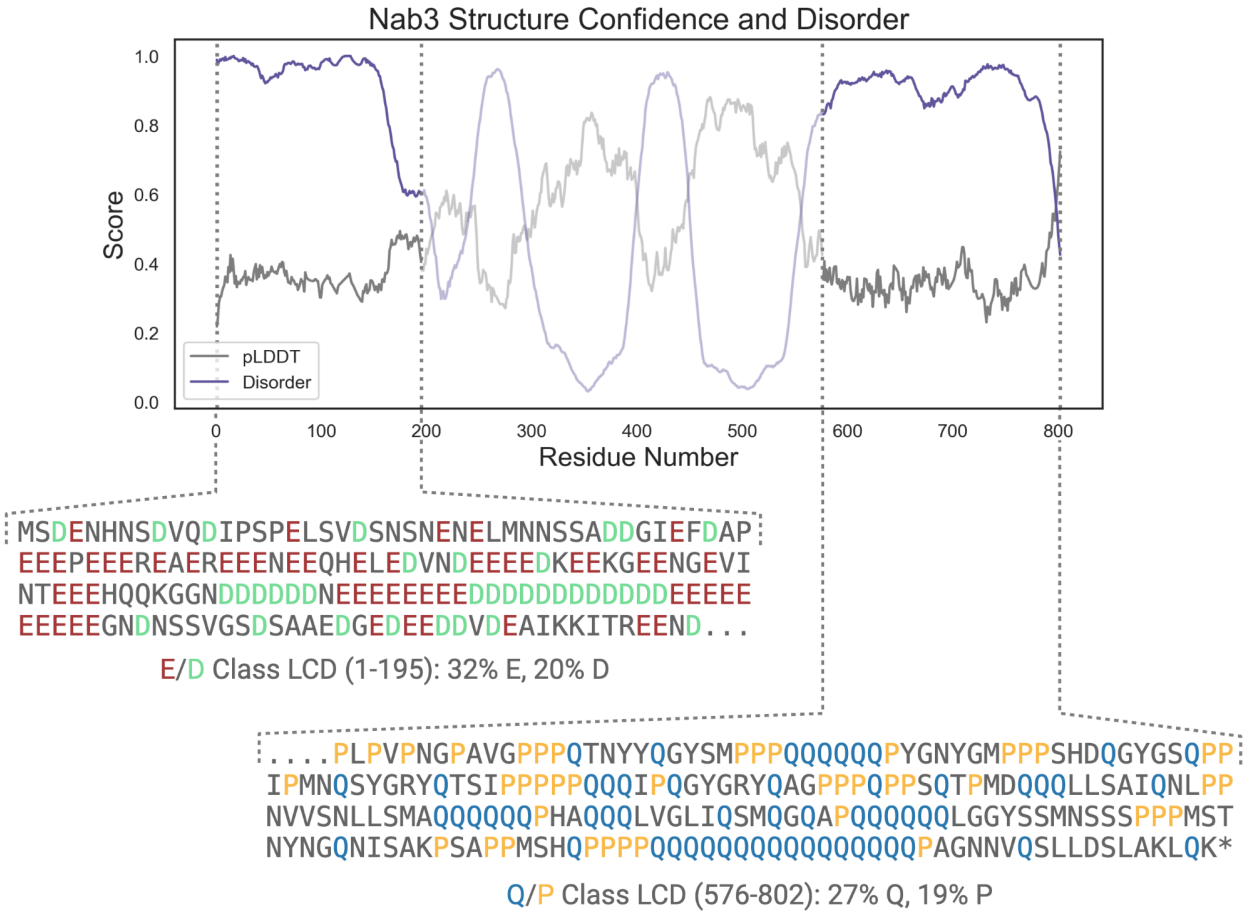

Figure 1: Nab3 disorder and low complexity domains. (top) Nab3 per-residue prediction scores of disorder (Metapredict v3, purple) and structure confidence (pLDDT, AlphaFold2, normalized to 1, grey). (bottom) The two terminal LCDs of Nab3. Residues 1-195, anionic E/D LCD. Residues 576-802, prion-like Q/P LCD. A Nrd1 heterodimerization domain and a RNA recognition motif are semi-structured and exist between the LCDs.

LCD class not only influences folded structure or disorder, but the range of interactions through which an LCD participates. Aromatic residues and polar blocks have a tendency to self-assemble, and LCDs of these classes are accordingly highly associated with LLPS and localization to membraneless assemblies. Gene ontology (GO) has, likewise, shown that LCD classes at large have predictable in-vivo regulatory functions and roles in LLPS[6]. For example: L class LCDs (>20% leucine) are implicated in vesicular transport and packaging pathways, while the S/R class is often found in nuclear condensates where mRNA splicing and processing occur, like nuclear speckles or cajal bodies[7,8].

Because LCDs have predictable regulatory roles, observing domain composition of specific proteins can provide general insights into localization and function. Nab3 is a yeast transcription termination factor and has two terminal LCDs: an N-terminal acidic E/D LCD and a C-terminal Q/P LCD which are separated by other functional domains (Fig 1)[9]. E/D class LCDs are highly associated with nuclear functions, consistent with Nab3's role as a transcriptional regulator, despite Nab3's E/D LCD currently having no known function. Q/P class LCDs frequently regulate the rate and timing of biosynthetic processes, and Nab3's Q/P LCD is indeed involved in granule assembly and RNA sequestration at the nuclear periphery upon glucose starvation[6].

When LCDs assemble through LLPS, the assembly is composed of a small subset of proteins that are not representative of the proteome at large. This results in distinct biochemical environments within each condensate relative to the cytoplasm, nucleoplasm, and other condensates[2]. The local biochemistry within an assembly is dynamic, depends on the ensemble of proteins, bound nucleic acids and small molecules, and offers a convenient platform for spatiotemporal discrimination of processes that must coexist within the cell. As such, cells are

able to achieve specificity of IDR recruitment between membraneless assemblies[10]. Specificity is achieved not through high affinity interactions or motif binding, but transient multivalent interactions of disordered regions. The mechanism of IDR specificity between condensates remains debated, but there is consensus that recruitment is achieved by favorable interactions between amino acid side chains and a lack of repulsive properties to the unique condensate biochemical environment[11].

Models of IDR Assembly

The interplay of forces governing IDR LLPS and recruitment to membraneless assemblies is multifactorial and includes hydrophobic interactions, electrostatic forces, hydrogen bonding, and Van der Waals forces[12]. A popular model of condensate specificity is the sticker-and-spacer hypothesis. Here, attractive interactions between sticker amino acid side chains facilitate multivalent protein-protein interaction (Fig 2a), while inert spacer residues tune flexibility and kinetics[13]. Pi-stacking of aromatic residues, electrostatic interactions of opposite charges, and hydrogen bonding between polar residues are canonical sticker interactions, with alanine, glycine, and proline being commonly viewed as spacer residues. Once a saturation concentration has been reached, and under the influence of physiological stimulus, temperature, and other conditions, sticker interactions locally increase protein concentrations such that 2-phase liquid demixing occurs (Fig 2b,c)[14]. Preferential partitioning of molecules in or out of droplets is a hallmark of LLPS, and can be described by a dense phase of high concentration, and a depleted light phase (Fig 2c).

Amongst condensates, the prevalence of sticker-type interactions is not a monolith; instead, IDR recruitment depends on the constituents of the particular assembly. For example, a

condensate composed predominantly of polar LCD classes may readily recruit additional polar LCDs, but disfavor recruitment of aromatic class LCDs despite enrichment for assembly prone aromatic stickers. In some systems, like with transcription factor IDRs and aromatic residues, sticker valency alone may explain condensate assembly and specificity[15], but in reality, IDR amino acid composition is 20 nearly-independent variables, many of which cannot be neatly categorized as stickers or spacers.

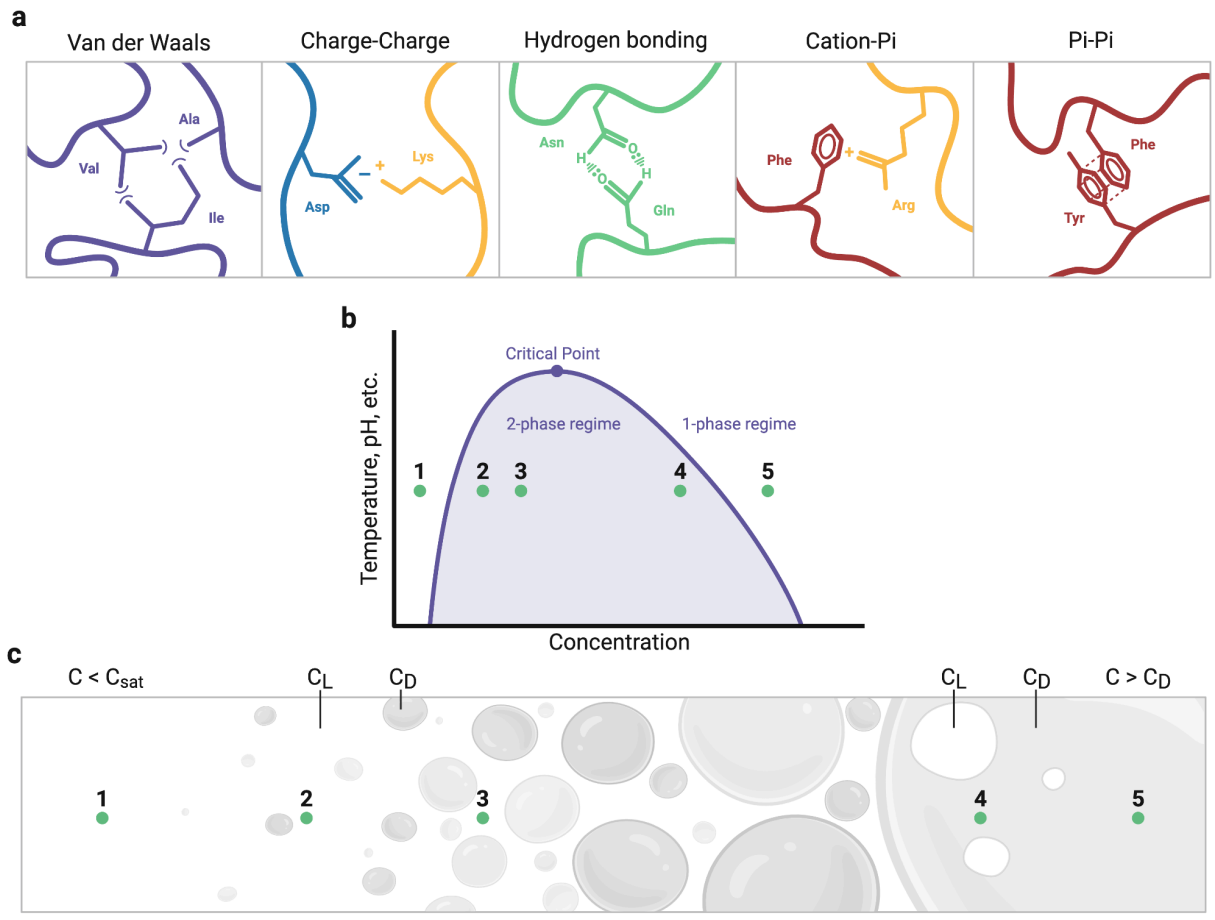

Figure 2: Protein liquid-liquid phase separation. (a) Types of weak interactions that occur between amino acid side chains and facilitate condensation: Van der Waals interactions occur between nonpolar amino acid side chains, oppositely charged side chains participate in electrostatic interactions, polar uncharged side chains donate and/or accept hydrogen bonds, aromatic side chains and cationic side chains participate in cation-pi interactions, two aromatic side chains interact by stacking of their pi orbitals. (b) Phase diagram with line separating the 1-phase and 2-phase regimes. Phase separation is dependent on

environmental factors and does not occur beyond the critical point. Numbered dots correspond to concentration dependent phase behavior shown in part c (Adapted from Alberti et al.). (c) Continuum of condensation based on protein concentration, C. Csat = saturation concentration, CL= light phase, CD = dense phase.

An amino acid composition model of IDR recruitment argues that the overall representation of amino acids within an IDR completely explains localization and function[16]. The composition model can be viewed as an extension of the sticker-and-spacer model, where, instead of a sticker and spacer binary, residues are scored along a gradient of promotion or repulsion from assembly. Scoring is experimentally determined, and must be independently established for each condensate[17]. Proof of concept experiments have shown that composition is important in some systems[18,19], but testing compositional dependence in several types of assemblies remains to be done. Somewhat contentiously, the composition model disallows a role of primary sequence in liquid assembly. Controversy arises due to the biochemical dogma that primary sequence governs function[20], and indeed individual studies have implicated the importance of short primary sequence motifs, sticker residue spacing, and domain blockiness in the organization and functionality of some IDRs[15,21,22].

Like many open questions in biology, the model for IDR assembly may not be all encompassing; some condensates may assemble under strict compositional requirements, while others have primary sequence dependency. Despite much progress in the field of IDR biology, a comprehensive theory of IDR recruitment and function remains incomplete.

RNA in Phase Separation

RNA is a potent regulator of LLPS due to RNA's ability to undergo phase separation alone in-vitro[23], RNA's existence within all cellular compartments, and the transcriptome's dramatic reprogramming in response to acute and prolonged stimulus[24]. Like IDR-containing proteins, RNAs also achieve specificity between condensates[25]. The simplest mechanism is through RNA-binding proteins (RBPs) with IDRs, which bind directly to RNA recognition sequences through their RNA-recognition motifs (RRMs) and assemble into membraneless ribonucleoprotein assemblies (RNPs) via their IDRs[26]. The RNA constituents within an RNP assembly then come as a direct result of the RBPs within the condensate. To a lesser extent, the chemical properties of nucleobases within RNA also allow for participation in multivalent interactions that promote condensate transience[27].

Within membraneless assemblies, one predominant role of RNA is as a scaffold[28]. Scaffold RNAs can initiate LLPS and are necessary for the continual stability of condensates, where, as an essential component, removal results in disassembly. Thus, the degradation, modification, or altered transcription of RNA scaffolds is an amenable way to facilitate condensate phase separation or dissolution. Nuclear paraspeckles are an example of RNA scaffold regulation influencing condensate dynamics. Paraspeckle assembly is dependent on the lncRNA NEAT1, where the long isoform is expressed during basal states and promotes paraspeckle formation. Upon stress or immune signaling, isoform switching occurs, preferentially expressing a short NEAT1 isoform that does not support paraspeckle assembly[25]. In this line of thinking, the concentration of scaffold RNAs controls condensate assembly and modulation of availability directly relates to condensate prevalence.

Alternatively, RNAs can be a condensate constituent without necessarily serving a structural role. Termed client or cargo RNAs, these undergo active regulation and are found in numerous different condensates, like nuclear speckles, which target cargo RNAs for splicing and processing, or P-bodies, which degrade cytoplasmic cargo RNAs[29]. The nucleolus, which is a complex multi-layered liquid assembly involved in ribosome biogenesis, contains scaffold pre-rRNA, whose nascent transcription is essential for nucleolar stability, and outputs mature cargo rRNA subsequently exported to the cytoplasm[30]. In this example and others, RNA can be considered scaffolding and cargo, by providing structure and being the endpoint of regulation.

Holistically, RNA plays a crucial role in liquid assemblies by facilitating phase separation and undergoing modification. The convoluted interplay between RNAs, RBPs, and IDRs governs LLPS of condensates with distinct molecular identities and regulatory roles. Many RNA-containing condensates provide the cell with dependable ways of modulating RNA processing and homeostasis[27].

Protein Misfolding

Given the structural freedom and high entropy of IDRs, disordered and low complexity sequences rapidly sample through a massive ensemble of conformations governed by intramolecular forces[31]. Upon stochastic homotypic collision, intermolecular contacts occasionally predominate and a metastable oligomer forms. Normally, these small oligomers are readily handled by the cell's proteostatic machinery (chaperones, degradatory systems, etc.), without consequence to the cell[32]. When not disassembled, some oligomers have the potential to spontaneously adopt a cross-beta structural conformation and extend what is now a nascent protofilament. At this point, the aggregate of misfolded protein is hyper-stable, protease resistant,

and classified as amyloid. Amyloid is not wholly resilient to disaggregation, and attempts at proteostasis result in filament fragmentation and release of small oligomeric species[33]. Endogenous protein expression only exacerbates misfolding by generating additional substrates.

A critical step in the process of aggregation are the nucleation events that generate the nascent protofilament[34]. To explain IDR misfolding, LLPS has been heavily scrutinized as a mechanistic throughline, due to the observations that IDRs are aggregation prone and also responsible for liquid assemblies[35,36]. Indeed, many properties of condensates appear to facilitate seed nucleation and fibrillization. Firstly, protein-protein contacts are unlikely to initiate oligomer misfolding in the mixed environment of the cell simply due to low protein concentrations. As IDRs demix, local concentrations increase and homotypic interactions become more prevalent[14], allowing stable protein-protein structures to form. Evidence of this is droplet aging, where some concentrated phase assemblies suffer decreased internal dynamics and more persistent interactions over time. Although aged droplets are not evidence of aggregation, aging may align misfolding-prone regions through kinetic trapping, where previously unfavorable interactions become forced as the assembly solidifies[37] (Fig 3a). On top of this, condensates have distinct pH, hydration, and ion concentrations, which influence the conformational landscape of aggregation-prone proteins[38]. Lastly, because LLPS structures are selective, a nucleating seed embedded within a droplet is likely shielded from parts of the proteasome, allowing nascent growth and pathogenic progression.

Protein aggregates can be broadly classified as amyloid or amorphous aggregates[39]. Amyloids are inherently fibrilar, thermodynamically hyper stable, and adopt the cross-beta conformation. The fibril poles serve as a template for monomer recruitment, facilitating misfolding and elongation[40]. Two classes of amyloid exist, parallel and antiparallel, which have

identical stacking of monomers or alternating monomer directionality, respectively (Fig 3b). A single filament of amyloid, known as a protofilament, aligns with other protofilaments and forms a mature fibril. Unlike amyloids, which have a well defined crystalline arrangement, amorphous aggregates have no regular spatial repetition and form from generalized hydrophobic collapse; as such, amorphous structures grow isotropically[39] (Fig 3b).

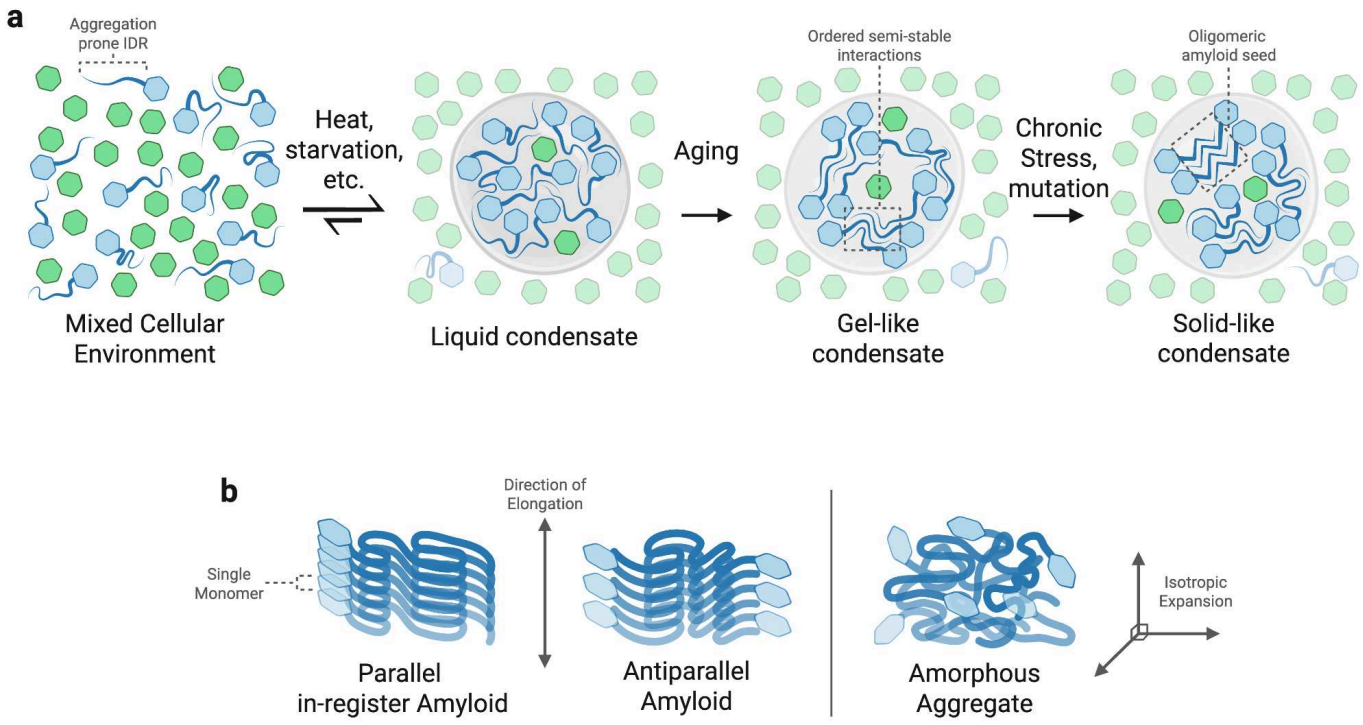

Figure 3: Dysfunctional liquid condensates in protein aggregation. (a) Under stimulus, the cellular environment undergoes demixing, mediated by IDRs, forming a highly dynamic dense-phase liquid assembly. Over time, dense-phase IDRs participate in increasingly stable interactions and droplets enter a gel-like state with reduced fluidity. Rarely, chronic stress or mutation can induce seed nucleation which can progress into irreversible and pathological protein aggregation. (b) Types of misfolded protein aggregates. Parallel and antiparallel amyloid monomers adopt a cross-beta sheet conformation, recruit monomers to the fibril ends, and relegate non-amyloid domains to the periphery of the amyloid core. Amorphous aggregates isotropically expand and are non-amyloid, but share irreversibility and beta sheet enrichment.

A single protein with the capacity for aggregation does not necessarily fall only into the parallel, antiparallel, or amorphous paradigms. For example, amorphous structures can be seen early in pathology which then transition into amyloid-type aggregates[41]. The particular type of aggregate depends on the mechanism of misfolding, environmental conditions, and tissue specific patterns of expression. For the same protein aggregated in different forms, differential pathology is observed, but regardless of type, irreversible aggregates are a massive homeostatic hurdle for the cell[42]. Aggregates are able to facilitate cell to cell spread, result in progressive dysregulation, cell death, and once localized pathogenesis has occurred, misfolded protein deposits are visible by histological staining and are a hallmark of numerous neurodegenerative diseases[43].

Aggregation and Disease

Protein aggregates are cytotoxic for a multitude of reasons. Most simply, aggregation can be understood as a loss of function, where properly functioning endogenous protein undergoes misfolding and cannot function[44]. Aggregates are also physically overbearing, disrupting condensates and blocking shuttling. Additionally, many aggregates acquire new functions that are broadly cytotoxic. One such function is that misfolded proteins often have exposed hydrophobic patches allowing for interaction with membranes, while sticky sequences facilitate aberrant interactions with proteins and RNAs[44]. A cell's attempts to restore proteostasis result in release of toxic oligomers known to disrupt ion channels[45]. As aggregation progresses, proteostatic machinery becomes overwhelmed, preventing proper folding of nascent and denatured proteins, further increasing the demands for proteostasis. This positive feedback loop compounds, leading to complete proteostatic collapse and chronic inflammation[33].

All tissues can be affected by protein misfolding, but specialized cell types are especially vulnerable to issues arising from aggregation due to high dependence on localized protein function and sensitivity to changes in resource allocation. Pancreatic beta-cells[46] and cardiomyocytes[47] are hyper-specialized cell types and susceptible to aggregation, but, more than any other tissue, neuronal cells are asymmetrically affected by the dysregulation and toxicity of protein aggregates. Neurons have extreme polarity, high metabolic demand, extended morphologies, and are non-dividing, all of which contribute to sensitivity to protein aggregation[48]. As such, numerous neurological diseases have been linked to protein aggregation, and mutations in associated genes are hypothesized to facilitate misfolding, aggregation, and pathogenesis. This is true for expansions of alanine or glutamine tracts causing >20 human diseases[49], or single-nucleotide polymorphisms causing Alzheimer's disease, tauopathies, Parkinson's disease, and others[48].

A class of similar but distinct protein misfolding disorders are protein infectious particle (prion) diseases, whereby soluble protein undergoes misfolding to an infectious isoform[50]. The infectious prion has an amyloid fold which facilitates recruitment and conversion of endogenous monomers to a growing insoluble stacked cross-beta sheet parallel protofibril. The growing prion is neurotoxic, induces progressive homeostatic dysregulation, and leads to cell lysis. Through direct cell to cell contacts or once released to the extracellular environment, prions can seed aggregation in proximal cells, with localized aggregation and cell death manifesting as characteristic spongy holes in neuronal cross-sections[43].

The most well-studied mammalian prions result from aggregation of the mammalian prion protein, PrP. These prion diseases are untreatable, rapidly progressive, induce personality and motor changes, and are always deadly. Uniquely, prions are infectious and transmissible,

meaning a prion can translocate from a diseased individual to a new host via diet or medical contamination, allowing prion pathogenesis to continue[51]. Although the initial observations of prions came from mammalian diseases, namely scrapie in sheep, Creutzfeldt-Jakob disease (CJD) in humans, and bovine spongiform encephalopathy (BSE) or "mad cow disease", prions have been found in other organisms as well, including *Saccharomyces cerevisiae* (yeast), where they act as a form of epigenetic inheritance and transmit from one cell to another[52].

Yeast Prions

Multiple different prions exist in yeast. Early discoveries showed Ure2 and Sup35 were the proteinaceous components of the [URE3+] and [PSI+] prions, respectively. Their prion domains (PrDs) are necessary and sufficient for prion activity, and are characterized by >45% Q/N content and depletion of charged and hydrophobic residues compared to the yeast genome[53]. Due to their striking amino acid features and implications in disease, many groups began to search for novel domains that are priogenic. To do so, different computational methods were implemented, including; PLAAC, a hidden markov model (HMM) screens Q/N rich prion candidates from entire proteomes[54], pWaltz identifies amyloidogenic nucleating cores hypothesized to potentiate prion activity[55], and PAPA uses weighted amino acid scores to predict domain prion propensity based on composition[16].

Large proteomic screens have, as in the case of Mot3 (which forms the [MOT3+] prion), identified authentic PrDs, but bioinformatic tools suffer from poor specificity; only a small minority of screened domains are experimentally validated as prions[54]. These domains that compositionally resemble yeast prions but fail to pass prion assays are considered prion-like domains (PrLDs). PrLDs are highly prevalent within the yeast and human genomes, with dozens

to several hundred being encoded in each, depending on the algorithm and parameters used[53]. Despite an inability to prionize, many PrLDs exhibit characteristics common to true yeast PrDs, such as intrinsic disorder, tendency to reversibly assemble into quaternary liquid structures, and, in some cases, have the potential to undergo pathogenic transformation to cause disease[56].

Human Prion-like Domains

In 2006, TDP-43 was shown to be the predominant component of aggregates associated with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[57]. Predisposition to both diseases are closely correlated with mutations in TDP-43's C-terminal domain, which is disordered, essential for mediating protein-protein interactions, and responsible for pathogenic aggregation. Once bioinformatic prion prediction methods were developed, the C-terminal fragment of TDP-43 was shown to be a high-confidence PrLD[58]. A growing number of PrLD-containing proteins have since been implicated in misfolding diseases, namely: FUS, hnRNPA1, TIA1, TAF15, ATXN1 and others[59,60]. Misfolding of a single PrLD species, or in conjunction with additional proteins, are associated with numerous neurodegenerative diseases. ALS, FTD, inclusion body myopathy (IBP), multisystem proteinopathy (MSP), Spinocerebellar Ataxia 1 (SCA1), and other diseases are all associated with pathogenic PrLD misfolding.

Although inherently non-transmissible and therefore distinct from prion diseases, PrLD aggregation is mechanistically similar to that of prions, where PrLDs misfold, propagate between neighboring cells, induce neuronal cell death, and manifest as motor/neurological deficits[48]. Like yeast PrDs, human PrLDs are highly predisposed to localize within condensates, either through passive recruitment or direct facilitation of assembly. As discussed before, the properties of dense-phase liquid assemblies appear to promote stable folding of otherwise disordered regions,

which are then vulnerable to pathogenic conversion. In-vitro study of hnRNPA1 supports a PrLD LLPS aggregation model by linking aged droplets to amyloid formation[61]. Additionally, mutation, dysregulation, and chronic stress predispose PrLD aggregation, consistent with LLPS mediated misfolding[42]. The mechanism of misfolded PrLDs propagating locally like prions can be described mechanistically through a prion-like mechanism.

Some neurological conditions such as Alzheimer's Disease, Huntington's Disease, and Parkinson's Disease, are marked by self-templating protein conversions that resemble the prion-like mechanism, but do not directly implicate a PrLD. The associated misfolded proteins, amyloid-beta and tau, huntingtin, and alpha-synuclein, respectively, do not contain PrLDs, but are sufficient to form misfolded aggregates that spread cell to cell. This has led some to argue for an expanded definition of prions to encompass all proteins that facilitate self-templating and seeded aggregation[62].

Fundamentally different biophysical principles of misfolding between prions and the non-PrLD containing aggregation-prone domains may exist, but an alternative interpretation for why these domains are not considered prion-like is that there are underlying biases in bioinformatic methods of PrLD detection. The yeast PrDs used to develop bioinformatic tools are likely not representative of the range of domains that exhibit prion-like behavior, but rather reflect the serendipitous discovery of yeast prions with extremely biased sequence features. The methods continue to be based on these early discoveries and have undergone minimal revisions since, meaning that what is defined as prion-like must fall in the narrow definition of compositional resemblance rather than demonstrating authentic prion-like misfolding and inheritance. Nonetheless, prion-like aggregation is observed in a growing number of human

diseases; in numerous PrLD-containing protein misfolding diseases, as well as a handful of diseases that implicate proteins without a canonical PrLD.

Nucleic Acid Binding PrLD-Containing Proteins

Among the library of human PrLD-containing proteins, of which PLAAC defines ~240, the collection is disproportionately nuclear with roles biased towards DNA-binding (79 proteins) and RNA-binding (72 proteins)[56]. Generally, the PrLDs of DNA-binding proteins facilitate organization of chromatin associated transcriptional condensates and contain a minimal activation domain embedded within the disordered PrLD. PrLD-containing proteins with a canonical RNA-recognition motif (RRM) are involved in assembly of condensates that regulate RNA homeostasis or use RNA as a structural component. PrLD-containing proteins can regulate both DNA and RNA membraneless processes, as evident by FUS, a PrLD-containing protein that is involved in gene regulation, DNA repair, and mRNA processing[63].

Nucleic acid binding proteins with a PrLD, particularly those that bind RNA, are strongly correlated with protein aggregation mediated through phase-separated assemblies; of the aggregation prone human PrLDs, a majority are annotated within RBPs which undergo LLPS[56]. RNA's ability to potentiate condensation suggests droplets with RNA-scaffolding are more stable and susceptible to pathogenic conversion compared to condensates assembled solely under the influence of interacting IDRs[27]. Therefore, understanding the mechanism of PrLD LLPS and condensate specificity, especially within RNA-scaffolded contexts, is relevant to human disease. Yeast are a convenient model for study of PrLD behavior due to abundant PrLD encoding within the yeast genome, genetic manipulability, and homologous overlap of condensate structure and function to humans[64]. Here, two yeast RNA-scaffolded condensates, stress granules and the Nab3

granule, will be interrogated to understand how sequence features of PrLDs allow for condensate localization.

Stress Granules

In humans and yeast, stress granules (SGs) are a cytoplasmic membraneless structure composed of proteins and mRNAs that assemble in response to heat, oxidative, viral, and osmotic stress. Current hypotheses of SG assembly suggest that, upon stress, polysomes are partially disassembled and allow mRNAs to be bound by protein complexes enriched for RRMs and low complexity/prion-like fragments[65]. These stalled preinitiation complexes multivalently interact through their disordered protein regions, promoting cytoplasmic phase separation and formation of SGs. Many of the IDRs responsible for SG multivalency are shown to undergo post-translational modification[66] (PTM), and thus altered chemical properties of interacting regions appear to be a fine-tuning mechanism for SG assembly. SGs are thought to reduce basal translation by sequestration and triage of mRNAs without degradation, where, upon loss of stimulus, the SGs disassemble and mRNAs are released back into the cytoplasm[65].

Stress granules are a highly complex assembly composed of a diverse population of mRNAs, translation machinery, and RBPs with PrLDs. The sum biophysical properties of canonical SG components creates a biochemical environment unique from the cytoplasm able to selectively recruit IDRs based on their sequence features[17]. Previous work tested a library of PrLDs for SG localization and highlighted the compositional biases that allow for recruitment of PrLDs: aliphatic residues promote SG localization, polar residues disfavor recruitment, and sequence scrambling does not alter behavior. Based on the compositional features that promote SG assembly and those that do not, synthetic prion-like domains (sPrLDs) and control prion-like

domains (cPrLDs), respectively, were designed and tested for SG localization. sPrLDs localized well to SGs and cPrLDs were not recruited. Together, localization of PrLDs to SGs appears to be compositionally dependent, without significant primary sequence requirements; this allows for engineering of synthetic domains with predictable in-vivo localization[17].

Of the PrLDs within SGs, many are aggregation prone and linked to neurodegenerative diseases. Chronic stress and mutation appear to contribute to pathogenic conversion within SGs through solidification[67]. Therefore, a nuanced understanding of PrLD dynamics within SGs, and RNPs at large, is important to understanding the regulatory framework of IDRs, aggregation, and neurodegeneration.

Nab3 Granule

A diverse set of transcripts are synthesized by RNA polymerase II (Pol II) and can be broadly categorized as coding and noncoding, the latter of which is terminated through a non-polyadenylation mechanism[9,68]. Polyadenylation independent termination occurs, in yeast, via the Nrd1-Nab3-Sen1 (NNS) transcription termination complex, which begins with recruitment of the Nrd1-Nab3 heterodimer to Pol II's disordered C-terminal repeats by binding of Nrd1 to Ser5 phosphorylation sites. RRMs within Nrd1 and Nab3 then recognize specific termination sequences of the nascent transcript and incorporate the helicase, Sen1, which disconnects Nrd1 from Pol II and translocates along the transcript until the elongation complex is dislodged (Fig 4 top). After, Pol II and Sen1 are allowed to dissociate while Nrd1 and Nab3 remain bound to the transcript guiding further processing by the TRAMP complex and nuclear exosome[68].

Perplexingly, biochemical and genetic mapping revealed that, upon glucose starvation, some NNS transcript targets were accumulating instead of being processed, indicating NNS deficiency[69,70]. The mechanism was unknown, but given the conditional behavior, a dynamic stress response was suspected. Using fluorescent microscopy, Nab3 and Nrd1 were observed to be pan-nuclear during basal conditions and, upon prolonged glucose deprivation, a granule would assemble on the periphery of the nucleus[71]. Constituents of the granule have not been fully characterized, but the assembly appears surprisingly simple, composed of: Nab3, Nrd1, and RNAs. Further, the condensate is liquid based on photobleaching and weak aliphatic disruption, and readily disassembles upon glucose reintroduction[71]. Given the observed liquidity, self-assembly of the disordered regions of Nab3 and Nrd1 were suspected to promote LLPS and granule formation. The current model of Nab3 granule assembly relies on many indirect lines of evidence, with the details lacking experimental validation. The model relies on the interplay of Nab3 self assembly, efficiency of proteostasis, and transcriptomic reprogramming.

During rich growth conditions, transcription is expansive and there is high demand on the Nrd1-Nab3 heterodimer for transcription termination, limiting the pool of free heterodimers[69]. Spontaneous heterodimer oligomerization does occur, but the high intracellular concentrations of ATP allow for efficient proteostasis with chaperones rapidly disassembling oligomers (Fig 4 top). During glucose starvation, net transcription decreases and the concentration of free Nrd1-Nab3 increases, favoring spontaneous self-assembly, while decreased respiration limits the energy availability of proteostasis[72], slowing the rate of oligomer disassembly. Global transcriptional changes during starvation may also change the prevalence of scaffold RNAs necessary for the Nab3 granule[70], which when bound, potentiate granule stability and growth.

The Nrd1-Nab3-RNA complex preferentially recruits additional RNAs and heterodimers from solution, leading to phase expansion into a mature Nab3 granule (Fig 4 bottom).

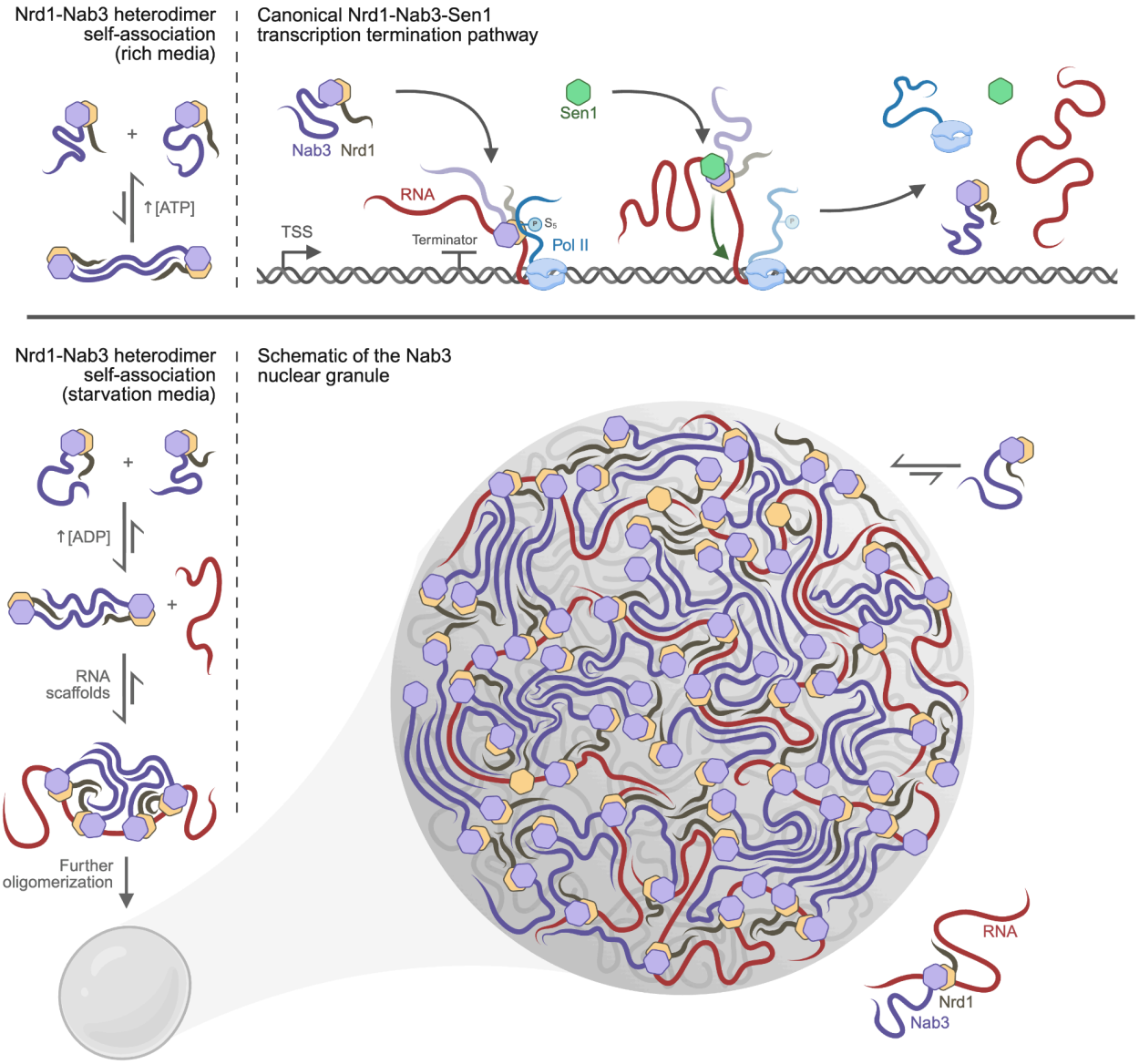

Figure 4: Schematic diagram of the roles of Nab3. (top) Nrd1-Nab3-Sen1 mediated RNA Pol II transcription termination for non-coding RNAs. In rich media, oligomerization of Nrd1-Nab3 heterodimers is unfavorable due to ATP dependent disassembly and involvement in active transcription. (bottom) When glucose starved, transcriptional output reduces, increasing free Nrd1-Nab3. Reduction in ATP prevents oligomer disassembly. Global transcriptional changes increase prevalence of RNA scaffolds, facilitating formation of the Nab3 granule.

The exact function of the Nab3 granule is an open question and has a few proposed roles. Nab3 levels are stable during glucose deprivation, indicating that the granule may sequester termination factors to modulate transcriptional output[69]. Alternatively, the granule could surveil anabolic RNAs, limiting growth during a period of nutrient deprivation. The RNAs may be processed, degraded, or simply sequestered within the condensate. Overall, the Nab3 granule has a role in the starvation stress response.

Important for protein-protein interaction, the C-terminal ~230 amino acids of Nab3 and ~100 amino acids of Nrd1 are low complexity, prion-like, and have skewed overrepresentation of Q and P. Both PrLDs appear important for potentiation of transcription termination, but in the context of granule assembly, Nab3's PrLD is necessary and sufficient for assembly, while the role of Nrd1's PrLD is unknown but is likely complementary/supporting[73]. In-vitro, full length Nab3 and Nab3's PrLD alone are sufficient to form amyloid and undergo condensation[74], acting as the mechanistic underpinning for in-vivo granule formation. Interestingly, Nab3 is human hnRNP-like, a family of nuclear proteins associated with numerous pathologies. Of note, the last 18 residues of Nab3's PrLD are nearly identical to the oligomerization domain of hnRNPC and is implicated in non-liquid oligomerization[75].

Due to the presupposed simplicity of the Nab3 granule and the complete dependency on the endogenous PrLD for nuclear assembly, the Nab3 granule is a convenient platform for study of RNP granule dynamics and the sequence logic of PrLDs. Based on this, studies have demonstrated that certain heterologous domains can functionally replace the Nab3 PrLD and allow for granule formation[71,76]. The types of domains and sequence features that provide functionality when swapped into the Nab3 system, however, remain unclear.

Aims of this Study

PrLDs are implicated in a huge range of physiological processes, yet the sequence grammar that encodes their function is largely unknown. Several observations of PrLDs suggest that their localization can be predicted and indeed certain condensates are well researched. However, a mechanistic throughline of how PrLDs achieve specificity of recruitment to condensates remains enigmatic.

Disordered regions are more tolerant of change and mutation than their folded counterparts[77]. Intuitively, IDR flexibility and entropy contribute to this resistance. Here, the primary sequence of amino acids may be less important than the overall representation of each residue within a sequence. In this study, the properties encoded within PrLDs allowing for condensate specific localization and function are evaluated. Two systems, the Nab3 granule and SGs, will be scrutinized for this purpose.

Firstly, a high-throughput pipeline for Nab3 granule detection will be developed. Able to accurately discriminate between granular and mixed nuclear phenotypes, the algorithm will be used extensively to characterize the Nab3 granule. Secondly, the compositional dependence of SGs and the Nab3 granule will be directly studied by finding and testing new PrLDs with high compositional similarity to PrLDs with known condensate-localizing activities.

Informed by the findings, a library of heterologous PrLDs and synthetic domains will be tested in the Nab3 system. Based on their activity, a hypothesis will be developed regarding the features important for Nab3 granule assembly. Together, the behavior of PrLDs in multiple cellular condensates is characterized, with both systems being directly analogous to condensates and proteins in humans which suffer dysregulation and pathogenesis

Characterization and Algorithmic Detection of Nab3

The Nab3 granule has only recently been discovered and few studies have described the system[71]. For more direct study, the chromosomal copy of Nab3 was knocked out. However, Nab3 is an essential protein as is its PrLD, reportedly. Therefore, a plasmid with a maintenance copy of Nab3 under control of the endogenous promoter was used to support strain viability. The maintenance copy is encoded on a ura3 plasmid, which can be selected against by plating on 5-Fluoroorotic acid (5-FOA) (Fig 5). This is useful for isolating a single copy of Nab3 within a strain, or for observing the viability of Nab3ΔPrLD, a version of Nab3 where the PrLD has been truncated.

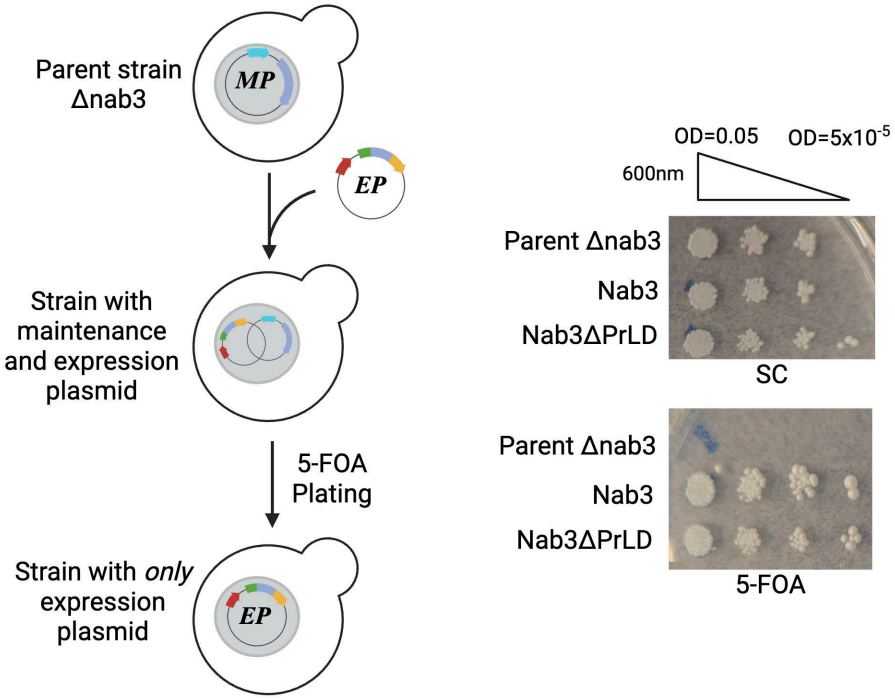

Figure 5: Viability of yeast strains depending on presence or absence of the Nab3 PrLD. (Left) A parent strain knocked out for chromosomal Nab3 is dependent on a ura3 maintenance plasmid (MP). The strain is then transformed with an expression plasmid (EP) encoding GFP-Nab3 or GFP-Nab3ΔPrLD and selected with a different selectable marker. After 5-FOA plating, the MP is lost, making cells dependent on the EP. (Right) 10-fold serial dilutions of different yeast strains on synthetic complete media and 5-FOA. The parent strain is non viable on 5-FOA, indicating dependence on the MP version of Nab3. Strains with EPs containing Nab3 and Nab3ΔPrLD are viable on 5-FOA, indicating that Nab3 is essential while the Nab3 PrLD is non-essential.

Surprisingly, the parent yeast strain (Δ*nab3*) had viability on 5-FOA supported by not only the full length plasmid copy of Nab3, but Nab3ΔPrLD as well (Fig 5, right). Prior work describes the Nab3 PrLD as essential[9,71], however, the PrLD behaves non-essentially for this particular lab strain of yeast. All versions of Nab3 on the expression plasmid are N-terminally GFP tagged and contain the SV40 NLS for nuclear localization. Using the GFP-labelled Nab3 plasmids, the in-vivo behavior of Nab3 will be observed.

Creating an Algorithm for Nab3 Granule Detection and Quantification

Across systems, condensation is commonly viewed by fluorescent microscopy, with high-throughput bioinformatic algorithms proving to be reliable methods of LLPS quantification[78]. In live cell imaging, computational techniques differ in their strategies of defining and characterizing condensates, but they all take advantage of fundamental properties of in-vivo membraneless assemblies. For yeast SGs, computation relies on mCherry tagged Pab1, a scaffold RBP which preferentially partitions into SGs, to define SG coordinates while Rpl1-BFP defines the cytoplasm. Partitioning of GFP-labelled IDRs into SGs allows calculation of a SG enrichment score by comparing SG and cytoplasmic GFP intensities[17,19]. Imaging of cells over several replicates allows for high-power analysis of sequence features that tune recruitment to SGs.

Using similar logic, a recent report describes a MATLAB algorithm for quantifying the existence of the Nab3 granule within individual nuclei[79]. Granule discrimination relies on the dramatic changes in localization of Nab3 following prolonged glucose starvation, from a pan-nuclear distribution during a fed state to high density focus that assembles while fasted. In this section, a similar bioinformatic pipeline has been independently developed, inspired by Hunn et al., featuring several changes and improvements (Fig 6).

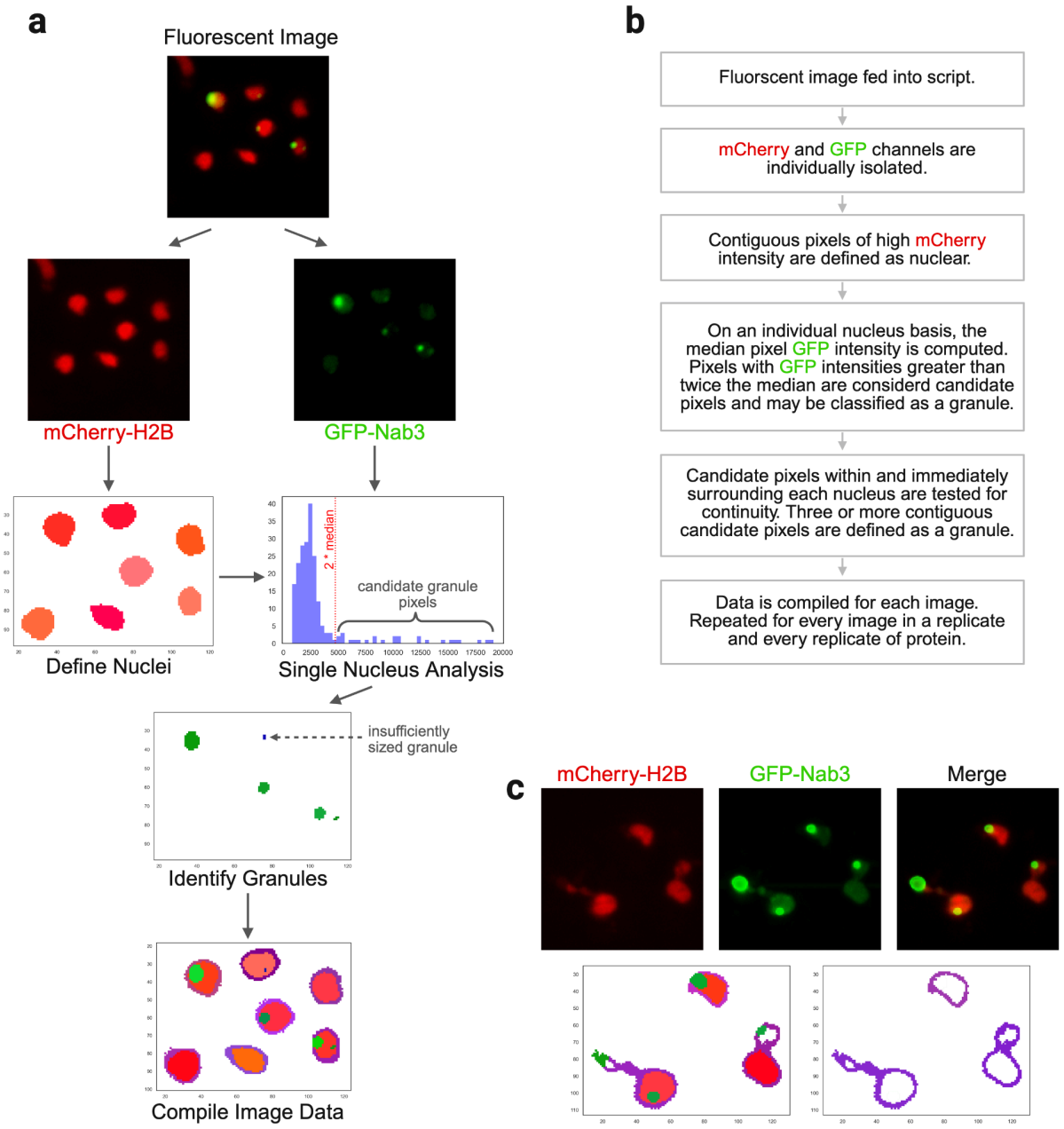

Figure 6: Workflow of the Nab3 granule quantification algorithm. (a) Images are fed into the script and used to define the pixel coordinates of cellular components. Nuclei coordinates are graphed in shades of red, expanded search areas in purple, granules in green, and candidate pixels (greater than twice the median GFP intensity) that are *not* granules in blue. The single nucleus GFP histogram is an example of how pixels are classified. Pixels with GFP intensities greater than twice the median (red vertical line) are classified as candidates and may qualify as a granule should several be adjacent. (b) Flowchart describing the bioinformatic steps in part a (Adapted from Hunn et al.). (c) The necessity of expanded searches for defining granules. The bottom two cells are undergoing nuclear division which results in discontinuous chromatin density. Searching lower density mCherry signal allows detection of granules that would otherwise be lost. Shown bottom right are expanded search areas only.

The algorithm was written in python and relies on two channels of fluorescence: a nuclear marker and GFP tagged Nab3. For this purpose, histone Htb2 (H2B) was endogenously mCherry tagged and the resultant strain was transformed with a plasmid copy of GFP-Nab3 under control of the endogenous promoter. Fluorescent z-stacks of cells are then acquired, maximally projected into a single image, and processed.

Challenges of Algorithmic Condensate Detection

Accurate selection of the Nab3 granule from fluorescent images is challenging for several reasons. Primarily, cells within a population have highly heterogeneous Nab3 signal and blanket thresholding risks missing granules within low expression cells and artificial identification of granules within high expression cells. For this reason, data is processed on a nucleus by nucleus basis, where, for individual pixels to be considered candidates for granule classification, they must be greater than twice the median GFP intensity for that nucleus (Fig 6a,b). By individually processing cells, expression heterogeneity does not confound quantification.

Secondly, Nab3 granules form on the nuclear periphery and partially exclude chromatin, making algorithmic detection dependent on a nuclear counterstain difficult. Lowering the mCherry threshold to qualify as a nucleus is insufficient to solve the problem, as the larger nuclear area results in greater volatility of median nuclear GFP intensities and elevates the false positive granule rate. Building on this, the final algorithm defines nuclei with a stringent threshold and expands the granule search area to include lower density surrounding chromatin. The expanded search areas reliably catch granules in dividing cells with inconsistent chromatin signal (Fig 6c), or simply granules on the nuclear periphery with limited chromatin colocalization.

A tertiary concern is that random signal noise may put certain pixels above the granule threshold without authentically assembling into dense-phase droplets. To test if noise will falsely flag granules and to better characterize the behavior of the Nab3 granule, a strain expressing GFP-Nab3 without the PrLD (Nab3ΔPrLD) and full length GFP-Nab3 were viewed by confocal microscopy after glucose starvation. Upon viewing, Nab3ΔPrLD remained mixed while full length Nab3 assembled into a single bright focus, indicating PrLD dependent granule assembly (Fig 7a).

The dramatic reorganization undergone by full length Nab3 allows for differentiation of granule forming strains from non-formers based on the distribution of their nuclear GFP intensities. Single nucleus histograms of GFP intensities illustrate how 2 fold enrichment over the median GFP intensity is achieved by a nucleus containing a GFP-Nab3 granule and extremely unlikely for a nucleus without a visible granule, as in the case of Nab3ΔPrLD (Fig 7b). Compiling GFP intensities from a set of nuclei within a replicate, a uniform distribution of GFP intensities is observed for cells expressing Nab3ΔPrLD, while granule forming strains (wild type Nab3) have a small subset of high intensity pixels corresponding to sequestration of Nab3 within granules and many pixels of low intensity corresponding to Nab3 depletion throughout the rest of the nucleus (Fig 7d). Similarly, the total number of pixels tested for candidacy (nuclear coordinates plus expanded search areas) can be compared to the number of pixels from that group which qualify as candidates (Fig 7c). In the glucose starved Nab3ΔPrLD strain, out of more than 100,000 pixels tested, only two were defined as candidates. Compared with full length starved Nab3, 6.62% of pixels were candidates.

In the Nab3ΔPrLD strain, which altogether lacks the ability to form nuclear granules, any pixels above the twice the median cutoff are presumably signal noise. Given that only two pixels

met this criteria in the starved PrLD knockout strain, and three candidates must be adjacent for granule classification, algorithmic granule detection is a reflection of real biological assembly occurring in the context of the PrLD. As such, noise plays little to no role in the classification of pixels and reflects the biological reality that loss of the PrLD completely abrogates detectable granules.

The percentage of cells across a population can be algorithmically quantified, before and after glucose starvation (Fig 7e). Full length Nab3 assembles into granules in half of quantified cells, while Nab3ΔPrLD never does. Perplexingly, however, Nab3 washed in glucose rich media and incubated formed condensates, albeit at a lower frequency than while starved, in 25% of cells. Similarly to the Nab3 PrLD being non essential in this yeast strain, observing granules without starvation contradicts current literature[71,79]. There, Nab3 grown in rich media conditions is validated to assemble in single digit percentages of cells.

Extensive time has been spent reconciling and validating this observation: the parent strain was re-transformed several times, varying levels of nutrients were used in media, and alternative carbon sources were tested - however, the efforts reproducibly resulted in some degree of Nab3 condensation without glucose starvation. Despite recreating the exact protocol used by others, there is a distinct possibility that some deficiency of the media used is inducing this prestress phenotype. Alternatively, like the PrLD being nonessential, semi-constitutive assembly of the Nab3 granule may simply exist within the particular yeast strain used for study. Regardless, all strains will be washed with and incubated in glucose free media for granule quantification assays.

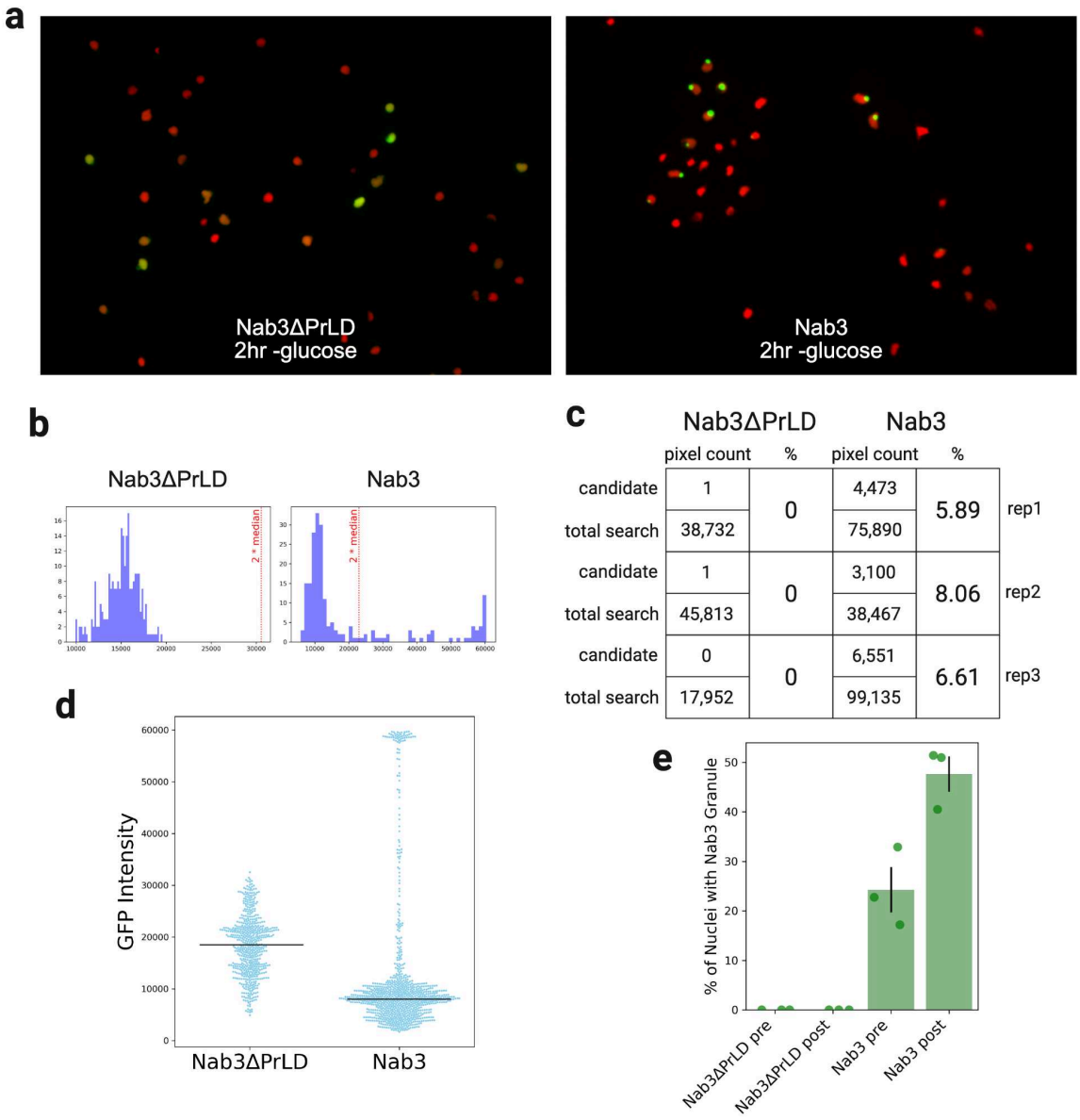

Figure 7: In-vivo behavior of Nab3ΔPrLD and Nab3. (a) Merged fluorescent microscopy of GFP-Nab3 with and without the PrLD. mCherry channel is an H2B nuclear counterstain. (b) Representative single nucleus GFP histograms for the respective strains after 2 hours of glucose deprivation. The Nab3 histogram is a granule containing nucleus while the Nab3ΔPrLD histogram is not. Red lines indicate 2 times the median GFP intensity for the nucleus. (c) The fraction of pixels within the nucleus plus expanded search areas that qualify as candidate pixels across three glucose starved replicates. (d) Distribution of GFP intensities within nuclei for glucose starved Nab3 and Nab3ΔPrLD expressing cells. Horizontal black lines indicate population median intensities. Nuclei with similar GFP intensities were compared for a single replicate of each protein. Each nucleus had GFP intensities greater than an average of 10,000 or less than a median of 29,500. (e) Algorithm quantification of Nab3 and Nab3ΔPrLD pre and post glucose starvation. Large dots are the population prevalence of granules for a single replicate, bar heights are the average of three replicates with standard error shown.

Testing Compositional Dependence of Stress Granules and the Nab3 Granule[1]

The fundamental biophysical characteristics of a protein are governed by the amino acids in its sequence and dictate the protein's binding partners, interactions, and localization within the cell[80]. Folded conformations are entirely dependent on the primary sequence of amino acids within the structure and are highly sensitive to sequence changes which can result in misfolding and dysregulation. Many IDRs, on the other hand, tolerate changes to primary sequence and retain functionality. Consequently, paralogous and orthologous IDRs often diverge considerably with respect to primary sequence[81]. However, IDRs have important biological functions and typically still contain conserved features despite divergence in their primary sequences. For this biological activity, the order of amino acids (i.e., the primary sequence) may be less important than the overall amino acid composition.

Proof-of-concept experiments testing a large library of PrLDs for SG recruitment have established that SGs have compositional preferences. Charged residues and aliphatics are enriched within SG-localized PrLDs, while polar residues and proline appear to strongly inhibit SG recruitment. In previous work, synthetic prion-like domains (sPrLDs) and control prion-like domains (cPrLDs) were designed based on compositional features that promote or obstruct SG recruitment, respectively[17]. The synthetic domains were designed by completely ignoring primary sequence, yet, during heat stress, sPrLDs localize efficiently to Pab1 marked SGs while cPrLDs do not. Furthermore, the degree of localization of PrLDs to SGs could be tuned in a stepwise fashion by gradual changes in hydrophobic content: a key feature of PrLDs contributing to SG recruitment[19].

[1] This chapter is adapted from a preprint by Sean M. Cascarina, Kacy R. Paul, Samuel J. Snyder, Larissa Ford, and Eric D. Ross (2025). Identification of intrinsically disordered regions with predictable in vivo activity based on sequence composition matching. Bioinformatics by Sean M. Cascarina, SG work by Kacy R. Paul.

A less well characterized system is the Nab3 nuclear granule. Nab3's PrLD is essential for condensation, but little is known about the sequence features important to the PrLD's function, including if primary sequence or gross composition governs activity. Some, but not all, heterologous domains have been shown to functionally replace Nab3's PrLD[76], suggesting that composition may be important for granule assembly. Also, replacement of 24 Q residues in Nab3's PrLD with E residues dampened the ability of Nab3 to form granules, suggesting that Q content is an important feature of the PrLD for Nab3 granule formation[71]. Notably, Nab3's PrLD is an LCD of the Q/P class. Q and P are correlated with strong inhibition of SG recruitment, indicating that features which allow for nuclear granule assembly differ strikingly from compositional features which promote SG recruitment. Thus, the types of IDRs that localize well to SGs may struggle in the context of nuclear granule assembly.

Many observations are consistent with the importance of composition for the functionality of IDRs: some sequence scrambled domains retain activity, heterologous domains can rescue protein function, and amino acid composition correlates with localization. In this section, the hypothesis that amino acid composition predicts IDR functionality is evaluated. Specifically, does composition-matching allow identification and prediction of new PrLDs which localize to SGs or form nuclear granules?

To do this, Match IDR, a nearest-neighbor composition matching algorithm, was employed. Developed by Sean Cascarina, Match IDR requires a user to input a query sequence, a proteome, and optionally window sizes. The tool outputs the fragment of each protein within the sequence file that has the highest compositional similarity to the query. Compositional similarity or identity (CI) derives from the total Manhattan distance of the difference in percent composition of the 20 canonical amino acids between the query and the subject fragment. The

tool is flexible and can use Euclidean or Manhattan distance, although Manhattan is preferred because the magnitude of compositional differences more intuitively relates to compositional similarity. Window sizes determine the length of the resultant matched IDR (Fig 8).

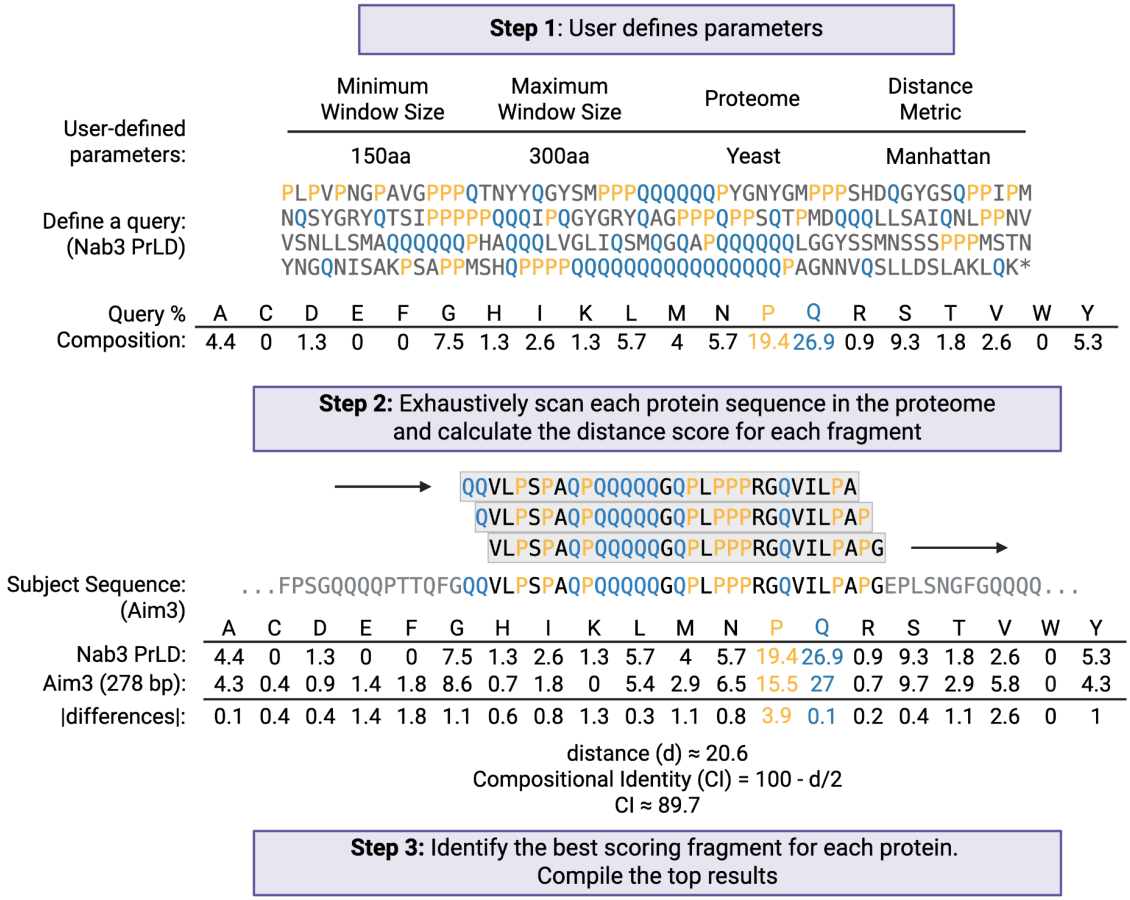

Figure 8: The Match IDR algorithm. The user inputs a query to compare against and a proteome to search for high CI matches. CI is calculated exhaustively starting at the minimum window size. At the given window sizes of 150-300, the algorithm will begin by calculating the subject CI for amino acids 1-150, 2-151, 3-152… until each fragment of the subject sequence has been screened with the minimum window. Then, the window size increases by 1, and the fragments are again screened. This process repeats until the maximum window size is evaluated. The fragment with the highest CI across all lengths is stored and each protein sequence within the provided file is evaluated.

Stress Granule Localization of Compositionally Matched Domains

If sPrLD and cPrLD have SG localizing and non-localizing functions, respectively, and were designed on compositional observations alone, then will compositionally similar new PrLDs, identified by Match IDR, retain this activity? Furthermore, if composition predicates IDR localization, then the protein origin should hold no sway over functionality. As such, the algorithm makes no organism specific assumptions, and identifies the best-matching regions regardless of whether the query IDR is derived from the same organism, a different organism, or is entirely synthetic. Therefore, MatchIDR was used with sPrLD and cPrLD query sequences to identify the top three matches for each sequence from the yeast, human, and *Dictyostelium discoideum* (slime mold or DCDI) proteomes. The same search parameters were applied across all proteomes using window sizes of 90-120 amino acids, used to detect fragments of similar length to the 100 amino acid sPrLD and cPrLD queries. All matches for the sPrLD or cPrLD sequences share strong compositional identity (CI) with each other and with the original query sequence yet very low primary sequence identity in pairwise alignments (Fig 9), indicating that the PrLDs identified by MatchIDR are not simply homologs or paralogs with conserved primary sequences.

All sPrLD and cPrLD matched sequences were GFP tagged and tested for colocalization with mCherry-Pab1, a SG marker, before and after heat stress. All constructs remained diffuse throughout the cytoplasm during basal states, with sPrLD matches forming discrete foci following heat stress. On the other hand, all nine cPrLD matched domains failed to form foci after heat stress and were unable to partition to Pab1 marked SGs. cPrLD matches had enrichment levels indiscernible from GFP alone while sPrLD matched domains preferentially localized to SGs and were fold-enriched within condensates greatly above GFP alone and the

cPrLD matched domains (Fig 10). SHOC1, notably, had limited SG partitioning, but nonetheless appears to favor condensate localization over the cPrLD matches.

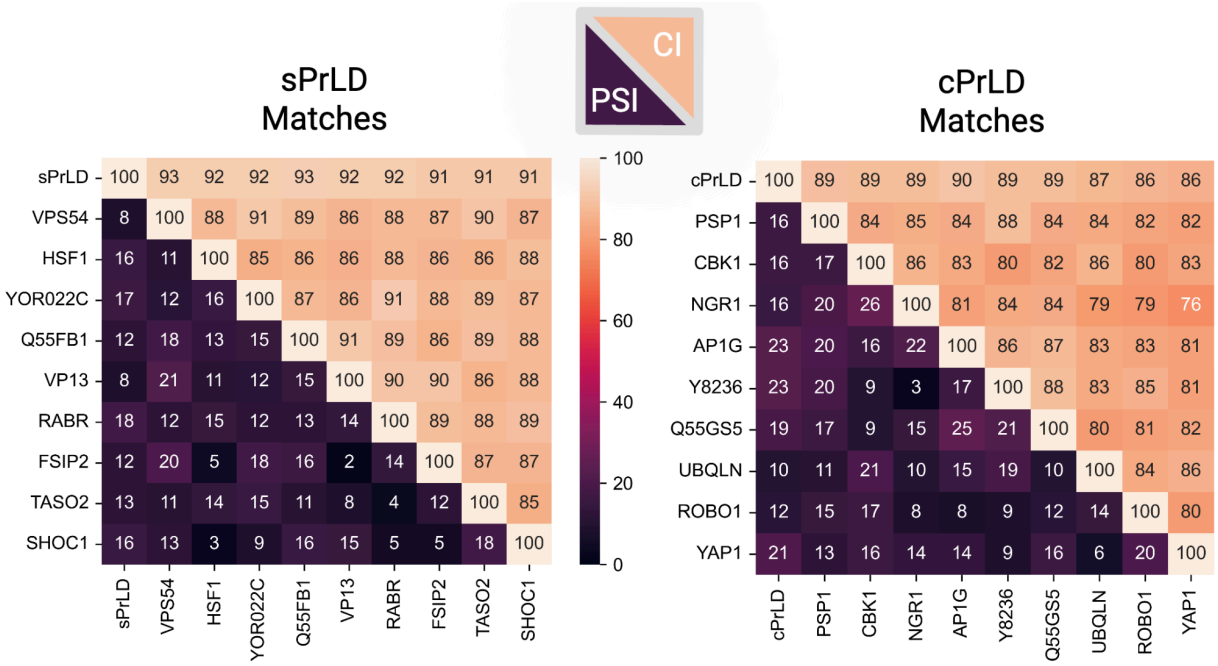

Figure 9: Primary sequence identity vs compositional identity alignment of 90-120 amino acid sPrLD and cPrLD matched domains. The left-lower halves are pairwise Smith-Waterman primary sequence alignments. The top-right halves are pairwise Manhattan compositional identities. The central diagonal indicates perfect primary and compositional alignment of sequences aligned against themselves.

Overall, the degree of SG localization does not correlate well with the magnitude of composition similarity of each sPrLD match compared to the original sPrLD sequence. This suggests that the composition-matching strategy is excellent at classifying candidates with or without the ability to localize to SGs, but does not perform well at the task of predicting the degree of localization. Furthermore, new PrLDs with or without the ability to localize to SGs were identified from all eukaryotic organisms screened and the proteomic origin of the PrLDs had no obvious effect on the degree to which SG partitioning occurred.

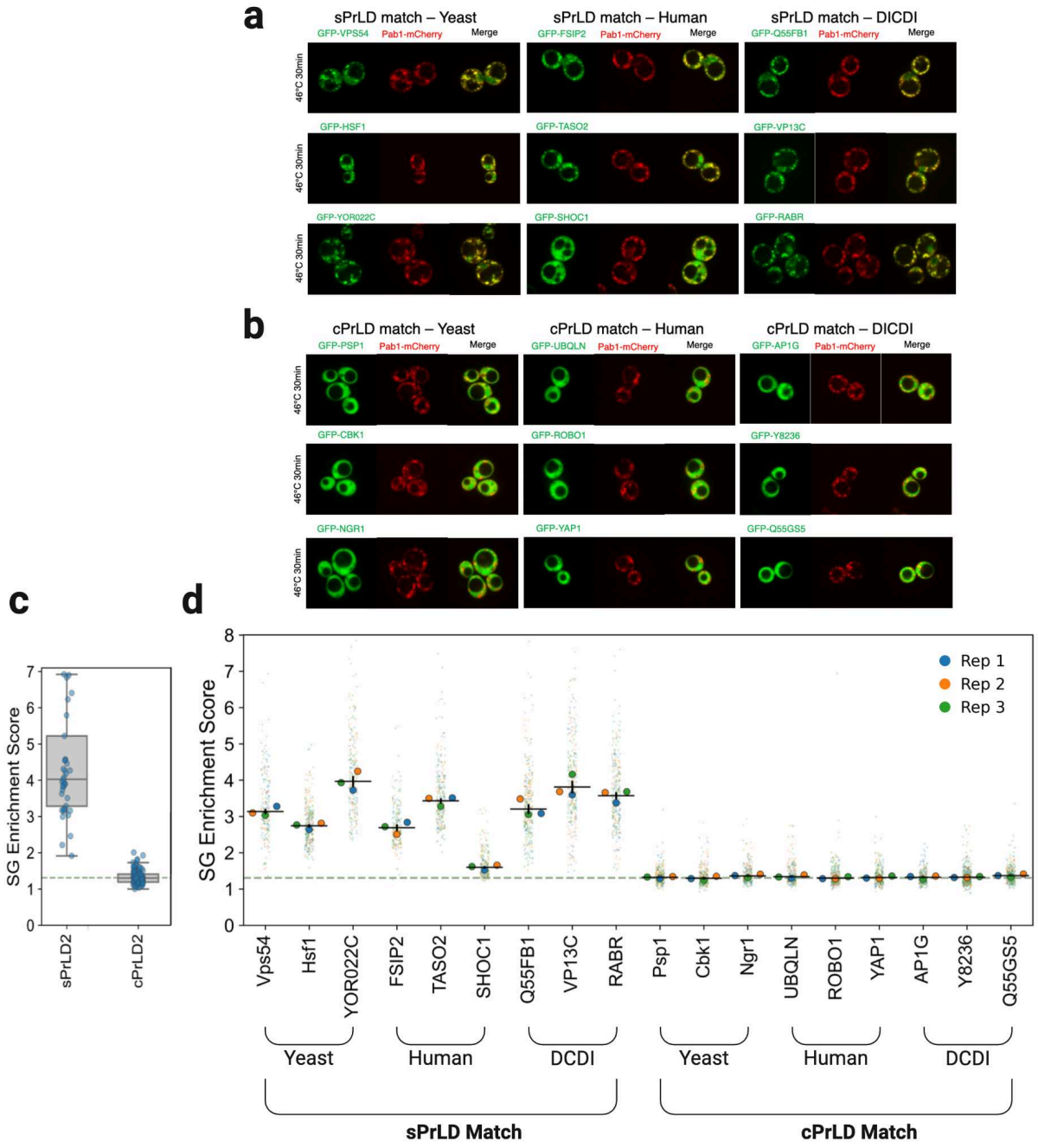

Figure 10: SG localization activity of composition-matched PrLDs from eukaryotic organisms. (a) Representative images of sPrLD-matched domains (GFP channel) with Pab1 foci (mCherry channel) that form in response to heat shock. (b) Representative images of cPrLD-matched domains. (c/d) Quantification of SG enrichment scores for each PrLD. (d) Dots are colored according to experimental replicate. Small dots represent individual cell SG enrichment scores. Large dots represent the median SG enrichment score for each experimental replicate, along with the mean and standard error of these medians. The horizontal dotted line represents the average of the median SG enrichment score for GFP.

Nab3 Granulation Employing Matched Domains

To examine whether compositional similarity to the Nab3 PrLD is sufficient to identify additional PrLDs capable of forming nuclear glucose-deprivation granules, a similar search was performed using the Nab3 PrLD as a query sequence. In order to minimize length as a variable, window sizes of 150-300 amino acids were used due to the comparable length to Nab3's 227 amino acid PrLD query. Unlike the SG-localization experiments, the sequence features that disfavor nuclear granule formation are unknown. However, sPrLD might serve as an appropriate negative control query sequence since it is capable of localizing to other types of condensates (SGs) yet is depleted in the most prominent features of the Nab3 PrLD – Q and P. Accordingly, the top three CI matches from the yeast, human, and slime mold proteomes were found using the Nab3 PrLD and sPrLD as query sequences. The matches have high within-group CI and relatively low within-group primary sequence identity (Fig 11). Additionally, the between-group CIs hover around 50 and are substantially lower than the within-group CIs and typically have single-digit primary sequence identities between groups (alignments not shown). Interestingly, despite the use of sPrLD as a query with a 90-120 window (SG experiments) and a 150-300 window, only one protein, SHOC1, was a top hit in both instances (Fig 9, Fig 11)

The same parent strain of yeast which endogenously expresses mCherry-H2B (histone Htb2) to define the nucleus was used for study of the Nab3 granule. Using the domains compositionally matched to Nab3's PrLD and the sPrLD, plasmids engineered with each of the GFP-Nab3 variants were transformed into the parent strain (Fig 12a). The matched PrLD sequences were used to heterologously replace the endogenous Nab3 PrLD, resulting in hybrid or chimera proteins. After glucose starvation, the wild-type Nab3 protein forms foci in roughly half of cells, while deletion of the Nab3 PrLD abolishes granule formation completely (Fig 12b).

When domains with high CI to Nab3's PrLD were observed after glucose starvation, a wide range of granule forming propensities were seen (Fig 12c). Four of the nine Nab3 PrLD matches, AIM3, NAF1, Q54KM5, and Q86AY8, formed obvious granules across all of the replicates while the other five Nab3 matches had very limited ability to support nuclear granulation (Fig 12e). Surprisingly, the NAF1 PrLD match from yeast actually formed nuclear granules in a higher percentage of cells than the native Nab3 PrLD. In contrast, all of the sPrLD matched PrLDs remained diffuse within the nucleus after glucose starvation (Fig 12d), and had little or no detectable granule formation (Fig 12e). Of note, STB4 had poor expression across replicates and preferentially localized to the cytoplasm, and was, therefore, unquantifiable.

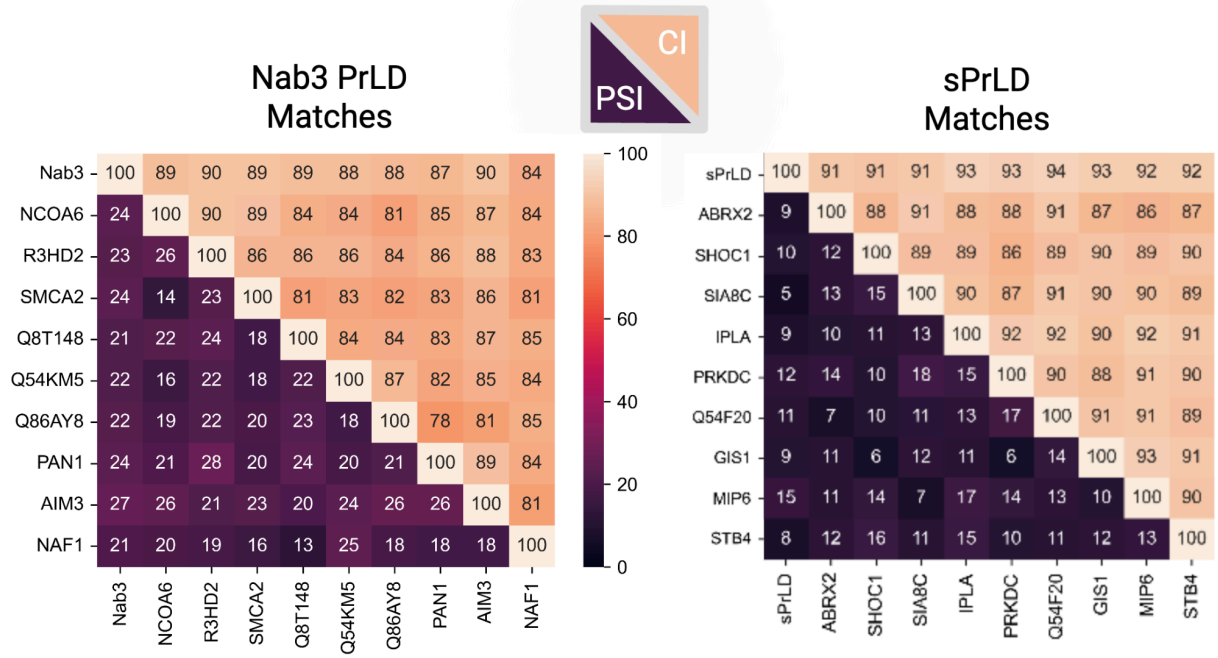

Figure 11: Primary sequence identity vs compositional identity alignment of 150-300 amino acid Nab3 PrLD and sPrLD matched domains. The left-lower halves are pairwise Smith-Waterman primary sequence alignments. The top-right halves are pairwise Manhattan compositional identities. The central diagonal indicates perfect primary sequence and compositional identity of sequences aligned against themselves. Domains matched to the Nab3 PrLD have comparatively higher primary sequence alignment scores due to their low complexity nature.

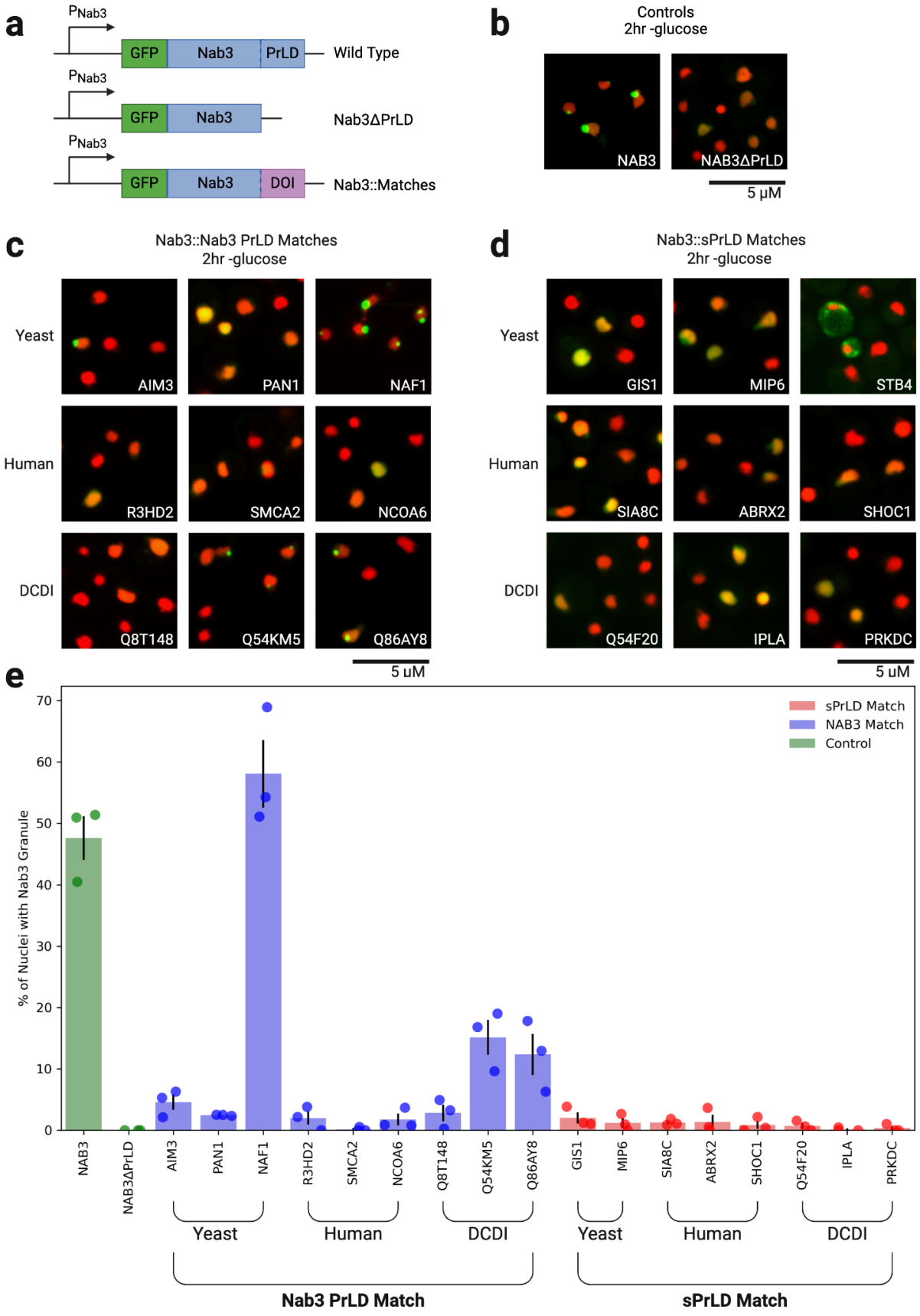
a
PNab3
GFP
Nab3
PrLD
Wild Type
PNab3
GFP
Nab3
Nab3ΔPrLD
PNab3
GFP
Nab3
DOI
Nab3::Matches
b
Controls
2hr -glucose
NAB3
NAB3ΔPrLD
5 μM
c
Nab3::Nab3 PrLD Matches
2hr -glucose
Yeast
AIM3
PAN1
NAF1
Human
R3HD2
SMCA2
NCOA6
DCDI
Q8T148
Q54KM5
Q86AY8
5 μM
d
Nab3::sPrLD Matches
2hr -glucose
Yeast
GIS1
MIP6
STB4
Human
SIA8C
ABRX2
SHOC1
DCDI
Q54F20
IPLA
PRKDC
5 μM
e
% of Nuclei with Nab3 Granule
sPrLD Match
NAB3 Match
Control
NAB3
NAB3ΔPrLD
AIM3
PAN1
NAF1
R3HD2
SMCA2
NCOA6
Q8T148
Q54KM5
Q86AY8
GIS1
MIP6
SIA8C
ABRX2
SHOC1
Q54F20
IPLA
PRKDC
Yeast
Human
DCDI
Yeast
Human
DCDI
Nab3 PrLD Match
sPrLD Match

Figure 12: Nuclear granule formation activity of composition-matched PrLDs from eukaryotic organisms. (a) Design scheme of plasmids expressing Nab3 variants. The endogenous Nab3 PrLD has been replaced with the top three compositionally similar fragments from each of the organisms, for both the Nab3 PrLD query and the sPrLD query. (b) Merged fluorescent micrographs of mCherry-H2B and GFP-Nab3 or GFP-Nab3ΔPrLD after glucose starvation. Nuclear foci are readily observed in the full length construct while PrLD loss completely negates this activity. (c) Merged fluorescent micrographs after glucose starvation of mCherry-H2B and GFP-Nab3 where the endogenous PrLD has been swapped for domains with high compositional similarity to the Nab3 PrLD. (d) Merged fluorescent micrographs after glucose starvation of mCherry-H2B and GFP-Nab3 where the endogenous PrLD has been replaced with domains compositionally similar to the sPrLD. STB4 had poor expression and cytoplasmic localization and was omitted from quantification. (e) Algorithm quantification of all the matched constructs. Large dots are the population prevalence of granules for a single replicate, bar heights are the average of three replicates with standard error shown. The variants are colored by Match IDR query origin.

Interpreting the behavior of matched PrLDs in the context of Nab3 is nuanced. Firstly, the sPrLD matches, which are expected to support localization to SGs, clearly lack the ability to form robust nuclear granules. These PrLDs are compositionally distant from Nab3's PrLD and suggest that composition can accurately predict domains that do not allow for Nab3 assembly. However, composition alone is insufficient to explain the type of domains that can functionally replace Nab3's endogenous PrLD. Only a subset of the positively matched domains formed granules upon glucose starvation and to drastically varying degrees. Further, the NAF1 PrLD, which outperforms the wild type domain, had the lowest CI of the entire dataset and by far the greatest tendency to assemble. This raises a profound insight: the Nab3 PrLD may be suboptimal for maximizing granule formation, either by primary sequence organization or the prevalence of inhibitory residues within the domain, both of which may explain why some positively matched domains lack formation. Regardless, the data is clear that compositionally similar PrLDs are more likely to form nuclear granules than compositionally distant PrLDs. Yet, other seemingly important and unknown features appear to exist.

Undoubtedly, Match IDR allowed for identification of new PrLDs with and without the ability to localize to SGs during heat stress or to form nuclear granules upon glucose starvation. With cPrLD matched domains being unable to localize to SGs and sPrLD domains being unable to form Nab3 granules, the ability to predict new PrLDs that do *not* have an activity is reliably predicted in both condensate systems. However, the degree to which new PrLDs recapitulated condensate localization/formation from the "positive" datasets were inconsistent between the condensates.

Multiple lines of evidence support a composition-driven activity model for SGs[17,65]. At present, much less is known about the sequence or compositional requirements for localization to Nab3 granules compared to SG localization, although this work and others have suggested that Nab3 granule formation is compositionally influenced, but by no means definitive. Even in the absence of a clear understanding of the compositional/sequence features driving Nab3 granule formation, MatchIDR identified new PrLDs with Nab3 localization activity, albeit with a lower success rate compared to SG localization activity. Collectively, these results provide further evidence that certain protein activities are driven by amino acid composition. When applied to these types of proteins and activities, the MatchIDR tool is a powerful tool to uncover additional protein regions with shared compositional features and encoded activities.

Replacing the Nab3 PrLD with Synthetic and Heterologous PrLDs

Compiling all of the evidence thus far, the activity of Nab3's PrLD is influenced by primary sequence and compositional factors. Previous work established that loss of Q residues disfavored granule formation and showed how some heterologous domains could functionally replace the endogenous PrLD[71,76]. Additionally, the matched PrLD dataset demonstrates that compositional features allowing for condensation in one system do not necessarily provide the same activity for other condensates. Together these data support a composition-driven model of Nab3 assembly. However, positively matched PrLDs were unreliable Nab3 granule formers and the magnitude of compositional similarity had no relation to granule propensity. The validation of NAF1 as a PrLD which outperforms the wild-type in granule assembly suggests that presence of inhibitory residues or patterning of the native PrLD may be suboptimal for maximal granule formation.

Synthetic Scrambles of the Endogenous PrLD

Artificial versions of Nab3's PrLD were designed to try and identify sequence features that govern granule condensation. Using the exact 227 amino acids of the Nab3 PrLD, two synthetic domains were designed by assigning each residue a random position within the sequence. The scrambled sequences, Nab3::SC2 and Nab3::SC3, have significantly increased linear dispersion of Qs compared to the original PrLD (Fig 13c). The scrambles were swapped in for the endogenous PrLD and tested for granule formation following glucose depletion. Both domains had limited or no ability to form granules and remained diffuse within the nucleus during glucose deprivation (Fig 13c,d).

The failure of full-length scrambles to form granules is near-conclusive evidence against a composition-only model of Nab3 assembly. Instead, specific patterning and features of the wild-type PrLD appear to provide functionality. An obvious feature of the full-length domain are the several poly-Q tracts, the longest being a 16 Q run that abuts the very C-terminal oligomerization domain[75] (Fig 13a). Interestingly, the Nab3 oligomerization domain is highly conserved and orthologous to the oligomerization domain of hnRNPC. In-vitro crosslinking experiments have demonstrated that appending the Nab3 oligomerization domain onto thioredoxin is sufficient for oligomerization, with further potentiation when the adjacent poly-Q run is included[75]. Therefore, the poly-Q tract and the oligomerization domain may play a role in facilitating LLPS of the PrLD. To test this, random scrambles were again designed. One variant, Nab3::scramble-oligomerization (Nab3::SCO), has the entire PrLD scrambled except for the last 18 residues, and the other variant, Nab3::scramble-polyQ-oligomerization (Nab3::SCPO), has the poly-Q run and oligomerization domain intact and the rest of the PrLD scrambled.

Upon glucose starvation, both constructs provided partial recovery of nuclear assembly compared to the full-length scrambles. Granules were detected in roughly 11% of Nab3::SCO expressing cells and Nab3::SCPO formed granules in 22% of cells - nearly half the frequency of the wild-type domain (Fig 13b,d). In the case of Nab3::SCPO, scrambling of 193 residues only resulted in a moderate loss of granules despite the scrambled residues accounting for 85% of the domain length. Based on this, the oligomerization domain and the poly Q stretch are important for LLPS of the native PrLD. However, characterizing the poly Q and oligomerization domain as responsible for half of granule assembly would be a tempting, but unfounded, way to interpret the data. Instead, these features, particularly the oligomerization domain, likely promote small-scale oligomerization that in turn facilitates valency amongst the rest of the PrLD, even

when scrambled. Obvious follow-up experiments would involve deletion of the PrLD's 193 N-terminal residues to test assembly of the poly-Q tract and the oligomerization domain alone, or would use the poly-Q tract and oligomerization domain appended onto other PrLDs to test how granule assembly is modulated.

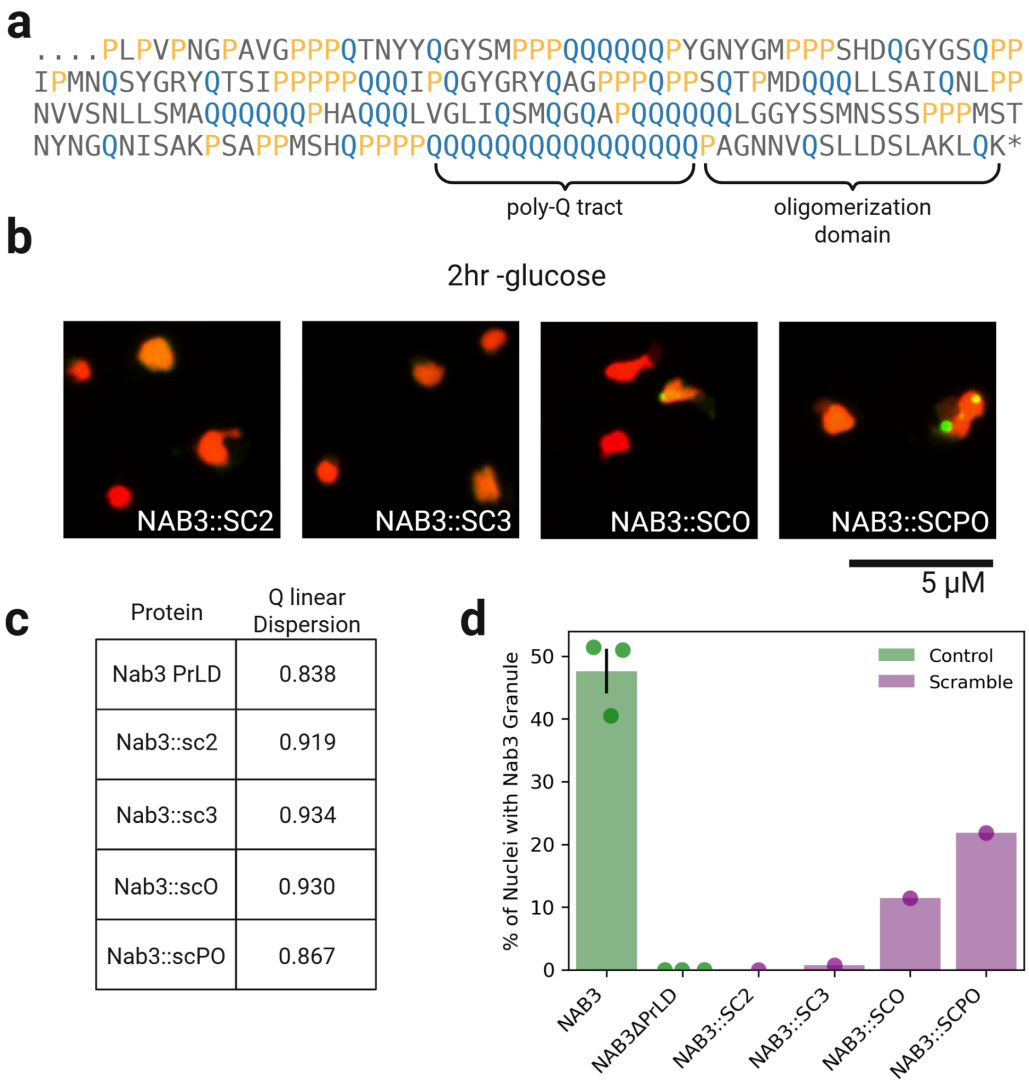

Figure 13: Nab3 granule propensities of scrambled domains. (a) Schematic of the 227 amino acid Nab3 PrLD. C-terminally, the 16 poly-Q tract and 18 residue oligomerization domains are bracketed. (b) Merged fluorescent micrographs after glucose starvation of mCherry-H2B and PrLD scrambled GFP-Nab3 variants. (c) Linear dispersion of Q residues within the PrLD of wild-type Nab3 and PrLD scrambled variants. Computed with coefficient of variation. A value of 1 indicates perfect dispersion and 0 is complete segregation. (d) Algorithm quantification of scrambled proteins. Nab3 and Nab3ΔPrLD are the same data as before. Large dots are the population prevalence of granules for a single replicate, bar heights are the average of replicates with standard error shown.

The abundance of Q residues within the Nab3 PrLD paired with low Q dispersion suggests that polar blocks are important facilitators of LLPS in the Nab3 system. A possible reason the full-length scrambles do not retain activity is the loss of large poly Q regions. This is evident in their increased Q dispersion relative to the wild type (Fig 13c) with neither domain containing more than three Qs in succession. Of note, Nab3's heterodimerization binding partner, Nrd1, also contains a C-terminal PrLD. Composed of ~100 amino acids, the absolute C-terminus of Nrd1's PrLD is 10 polar residues, 8 of which are Qs in series. The similar organization of Nrd1's and Nab3's PrLDs are unlikely to be coincidental. Essentially nothing is known of the Nrd1 PrLD, but it may serve as an unappreciated link between organizational features of Nab3's PrLD and granule assembly.

Replacing the Nab3 PrLD with Heterologous Library of PrLDs

Much insight was gained into the compositional features that favor SG localization when a large library of PrLDs and PrDs were tested for SG assembly in response to heat stress[17]. The domains within the library were identified by PLAAC and PAPA, two tools for PrLD prediction [16,54]. The protein library encompasses a wide range of dissimilar sequences, with variable lengths and distinct compositional architectures. Some of the domains reversibly assembled into SGs upon heat stress while most did not, allowing elucidation of the compositional biases of condensate localizing PrLDs.

For several reasons, the Nab3 system is amenable to testing of the same library. Firstly, the Nab3 PrLD is unusually long but appears to have flexible length requirements. The comparatively short 123 amino acid SUP35 PrD has previously been shown to support nuclear assembly when swapped in for the endogenous PrLD, and restoration of the primary sequence of

the last 34 residues allowed for meaningful granule formation. Based on these observations, PrLD sequence length may be of secondary importance compared to the overall biophysical properties that provide for activity. Heterologous PrLD length, therefore, should not confound interpretation. Additionally, the matched PrLD dataset indicates that, while the Nab3 granule has compositional preference, unknown organizational features predominate activity. Testing a wide variety of Nab3 PrLD replacements, therefore, may allow discovery of important functional characteristics.

In total, thirty three PrLDs were swapped into the Nab3 system from the PrLD sequence library and viewed for granule assembly upon glucose starvation (Fig 14a,b). PrLD grammar not only dictates condensate recruitment, but cellular compartment localization as well; five of the Nab3 variants preferentially localized to the cytoplasm and were omitted from quantification. Of the twenty-eight quantified variants, variable degrees of nuclear LLPS were observed (Fig 14). Of these, six formed granules more than 10% of the time. For later analysis, these domains are "positive" Nab3 assemblers. Four positive Nab3-assembling variants were non-localizers to SGs while two were efficiently recruited to SGs. Wild type Nab3 forms granules in roughly half of cells and none of the domains tested near this level of assembly.

Some of the Nab3 variants were modest supporters of assembly, with granules detected in 2-10% of cells. On gross inspection, the condensates formed by this group were smaller and weaker than the positive variants. Low partitioning of protein from the nucleoplasm into condensates is seen in the selected microscopy of YCK2, SRO9, and SIZ1 variants (Fig 14a,b).

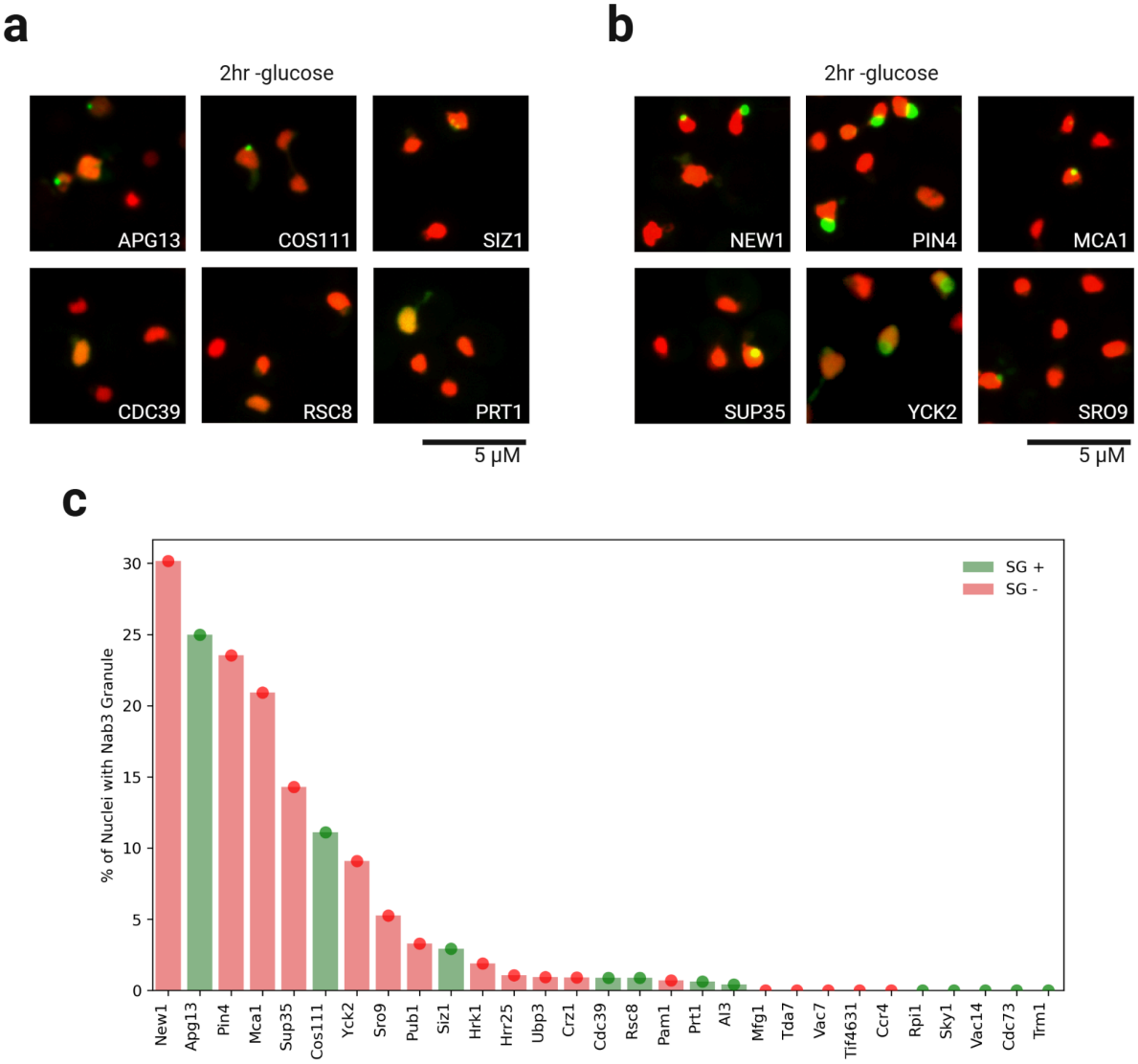

Figure 14: Nuclear granule formation of heterologous PrLDs swapped into Nab3. (a) Merged fluorescent micrographs of the top six SG-localizing PrLDs in the Nab3 system. mCherry channel corresponds to a H2B chromatin and GFP channels are Nab3 variants. (b) Merged fluorescent micrographs of the top six non SG-localizing PrLDs in the Nab3 system. (c) Algorithm quantification of nuclear granules for all PrLDs from the library, colored by ability of the PrLD alone to localize, or not, to SGs.

Nineteen of the Nab3 variants formed foci in less than 2% of cells after glucose depletion. These “negative” formers were equally composed of PrLDs that supported SG localization and PrLDs that did not. Broadly, whether or not a PrLD localized to SGs had no predictive effect on formation of the Nab3 granule.

Something that does correlate with nuclear assembly is the distinction between PrD and PrLD. Three authentic PrDs, NEW1, PIN4, and SUP35[52], were included in the sequence library and all three supported high levels of granule assembly when swapped into Nab3. Consistent with this, the endogenous Nab3 PrLD forms amyloid in-vitro[74]. PrDs may allow for nuclear assembly but not localization to SGs due to inherent differences in the molecular composition of the condensates. SGs are highly complex and PrLD recruitment therefore is dependent on net biophysical characteristics. The Nab3 granule, on the other hand, is more simple and valency arises predominantly from Nab3-homotypic interactions. It stands to reason, then, that highly self-assembly prone PrDs would be able to functionally replace the Nab3 PrLD. However, the Nab3::APG13 variant forms granules at a high level despite the APG13 PrLD not being classified as amyloidogenic.

Evaluating groups positive or negative for nuclear assembly based on composition must be approached carefully, given the nuanced role of composition in Nab3 condensation. However, the differences between the positive and negative groups are nonetheless informative in discerning the types of domains expected to allow or disallow Nab3 assembly. When comparing sequences of the positive (>10%) and negative (<2%) groups, certain compositional differences were noticed. Traditional disorder-promoting residues (G, A, P), were biased toward Nab3 positive variants (Fig 15). Additionally, Q, the primary compositional feature of the Nab3 PrLD, was found preferentially within positive forming PrLDs. Aliphatic residues and all charged residues were biased toward the non-assembling variants. Aromatic residues were mixed between the positive and negative groups. NEW1, SUP35, and MCA1 have multiple tandem repeats of Y which explains the strong bias of this residue towards the positive group. Very few C, M, and W residues exist between both groups. Notably, only six residues are probable within

the positive group (Fig 15, right). A plausible explanation is that, across these six residues, the positive group is of lower complexity while the negative group has greater sequence complexity and heterogeneity of all amino acids.

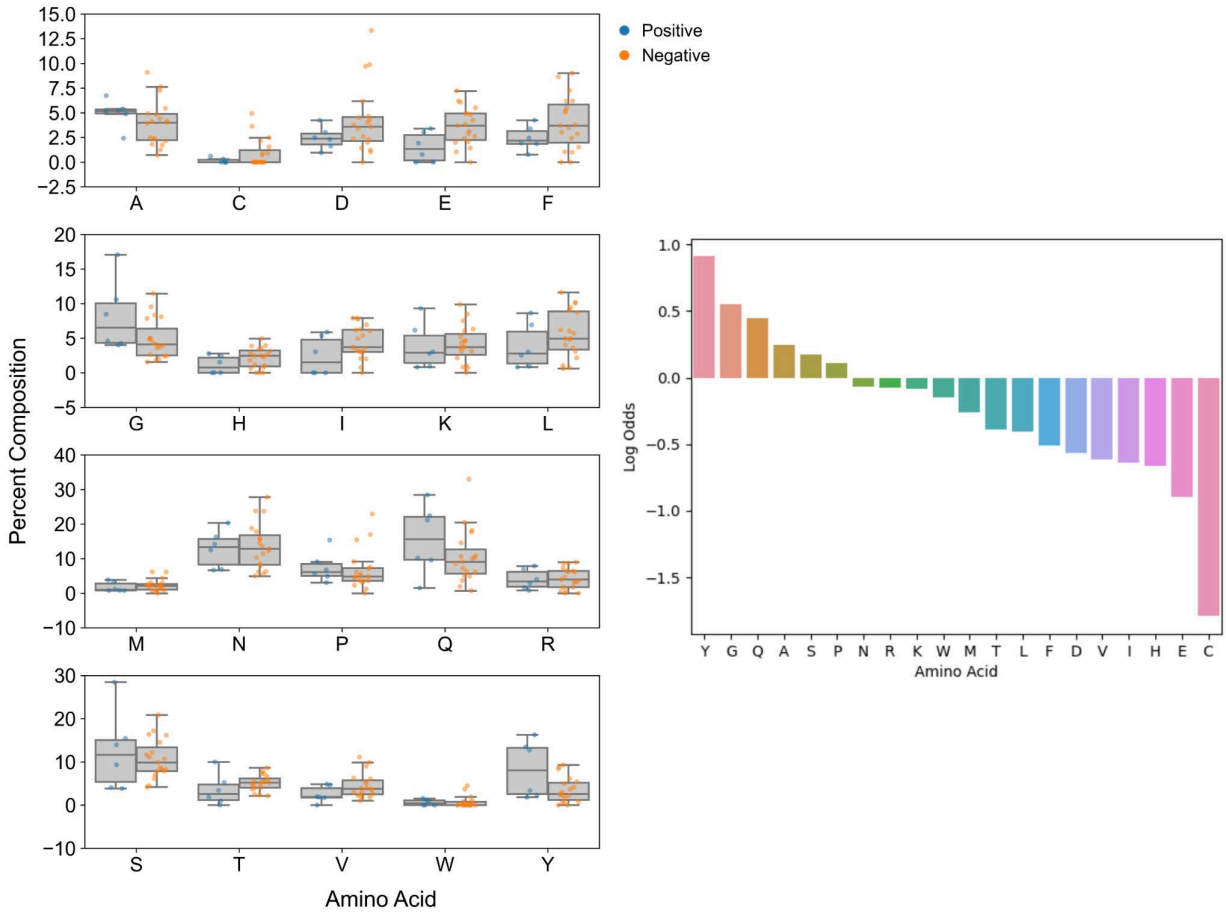

Figure 15: Amino acid comparison of Nab3 variants scored as positive assemblers or negative assemblers. (Left) Percent compositions of each of the 20 canonical amino acids for each quantified PrLD/PrD. Individual dots indicate the percent composition of a single sequence. Positive or negative Nab3 granule propensity groups are compared against each other. (Right) Log odds of finding each residue within the positive or negative Nab3 forming groups. Positive values indicate bias toward the positive group.

Discussion

PrLDs are enriched among RBPs and are involved in recruitment to many different types of condensates. Despite this general trend, the specific features that govern targeting to distinct condensates is incompletely understood.

The Nab3 granule in yeast is an interesting case, as early observations suggested a role of composition for PrLD recruitment[71,76], yet scrambles of the Nab3 PrLD were non-functional for granule assembly. Perplexingly, new PrLDs matched to the Nab3 PrLD were sometimes functional while others were not. Further, PrLDs matched to a SG localizing PrLD were non-viable for nuclear assembly. Some overlap for PrLD recruitment exists between the condensates, as a library of PrDs and PrLDs validated two PrLDs with SG localizing and Nab3 assembling properties. The features of these two domains may provide hints to the fundamental properties allowing recruitment to RNP granules.

Based on the compositional features of heterologous PrLDs, the Nab3 granule appears biased towards, unsurprisingly, its low complexity features of Q and P. Small disorder promoting residues, and S were also preferentially found within Nab3 assembling PrLDs. Paired with the depletion of aliphatic residues, the Nab3 compositional biases run nearly opposite to that of SGs, supporting a compositional model of condensate specificity. Another difference from SGs is the partial primary sequence reliance of the Nab3 PrLD. A data point implicating a role of primary sequence for the endogenous PrLD is the potentiation of condensation by the C-terminal fragment when the rest of the domain is scrambled. For the scrambled domains, the linear dispersion of Q correlates with activity, although specific interrogation of Q dispersion is yet to be done.

Perhaps the single best predictor of an ability to form nuclear granules is identity as a yeast PrD. The three yeast PrDs tested were previously shown to altogether lack SG localization. All three provided robust condensation in the context of Nab3. Here, the strong biophysical tendency towards self assembly of PrDs appears to correlate with Nab3 formation. A possible explanation is that the relatively simple chemistry of the Nab3 granule allows strong homotypic interactions to dominate, while the complex SG environment forces heterotypic PrD interactions that disfavor SG localization. Activity of three domains is not conclusive, however, and further experimentation testing other yeast PrDs is necessary.

Together, the features allowing for Nab3 granule condensation are more thoroughly characterized. Composition does not predict positive activity in the Nab3 system, but can accurately predict PrLDs without Nab3 formation activity. Some compositional features matter, including low complexity enrichment of polar residues and disorder promoting residues while retaining polar block organization.

Logical design of fully synthetic sequences analogous to sPrLD and cPrLD are a logical next step. These domains will have compositions that resemble Nab3-positive or Nab3-negative PrLDs, allowing direct testing of which features are necessary and sufficient for granule incorporation. By systematically varying properties such as Q dispersion and charge, synthetic domains can further elucidate biophysical rules or specific sequence motifs. Ultimately, this approach will help clarify why the Nab3 granule represents a specialized subclass of RNP condensates with distinct recruitment logic from stress granules and other PrLD-driven condensates. Further dissection of the sequence features and interaction networks underlying Nab3 granule formation will not only improve our understanding of condensate specificity but may also offer broader insights into functionality of human condensates.

Materials and Methods

Chromosomal Modification of Yeast

Chromosomal manipulations were carried out using a high-efficiency lithium acetate transformation protocol for yeast. The KanMX cassette was amplified with tails containing homology to regions directly flanking the *nab3* gene body. The KanMX amplicon was added to lithium acetate permeabilized yeast cells. Colonies which have undergone homologous recombination were identified by selection on G418 plates. Positive clones were further validated by Sanger sequencing of the genomic regions flanking the integration site. Only strains with verified, clean knockouts were used for subsequent experiments.

Two sets of primers were designed to amplify a Met1-mCherry cassette. The amplicon was designed such that efficient incorporation of mCherry onto the N-terminal tail of histone Htb2. Using a similar protocol, the Δ*nab3* strain was permeabilized with the high-efficiency lithium acetate protocol and transformed with the Met1-mCherry cassette. Colonies were then screened for growth on SC -met plates. Positive growth colonies were sequence validated and then viewed fluorescently.

Plasmid Generation

Plasmid constructs were designed in SnapGene and assembled via infusion assembly methods. All Nab3 proteins were designed to contain GFP, followed by the SV40NLS, followed by the engineered Nab3 variant. Nab3 and Nab3ΔPrLD were cloned within e. Coli using simply single fragment infusion cloning. All other Nab3 plasmids were generated using two-piece infusion into competent e. coli cells. Positive transformants were PCR screened and sanger sequencing prior to yeast transformation.

Yeast Transformations

Yeast strains were transformed using the lithium acetate carrier salmon sperm DNA and PEG method. Maintenance plasmid transformants were selected on SC -ura, while expression plasmid transformants were selected on SC -leu.

Growth Conditions

The day before imaging, cells expressing the desired Nab3 protein were grown to saturation overnight at 30 degrees C. The next morning, colonies were reseeded. Once cells entered mid-log phase growth in SC -leu, cells were washed three times with SC -leu -ura - glucose media for starvation assays, or SC -leu -ura. Cells were then incubated for 2 hours at 30 degrees C.

Viability Assay

Cells were grown to mid log phase growth, and diluted down to a starting OD 600 of 0.05. Serial 10-fold dilutions were prepared down to OD 600 5 × 10^-5. Dilutions were spotted on synthetic complete media, or 5-FOA, and incubated at 30 degrees C for 2 days before imaging.

Nab3 Confocal Imaging

All confocal images processed and shown are maximal z-stacks, derived from 8 slices taken over 2 micrometers. Laser powers were altered between proteins, to minimze overexposure artifacts. For starvation or rich media imaging assays cells were washed three times with SC -leu -ura - glucose media for starvation assays or SC -leu -ura for prestress assays. A minimum of 50 quantifiable cells per replicate were used. Strains were not selected on FOA before imaging.

Stress Granule Imaging

Each PrLD was cloned into a vector to express the PrLD with an N-terminal GFP tag from the SUP35 promoter. These vectors were transformed into a strain expressing the cytoplasmic marker (Rpl1b-BFP) and SG marker (Pab1-mCherry) from their endogenous loci. Cells were grown overnight in SC-Leu media (synthetic complete media lacking leucine) then diluted into fresh media, grown to mid-log phase, and subjected to a 30-minute heat shock at 46 degrees C immediately prior to imaging. At least three experimental replicates were performed for each PrLD, with a minimum of 50 quantifiable cells per replicate.

References

1. Holehouse, A. S. & Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell Biol.* 25, 187–211 (2024).
2. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298 (2017).
3. van der Lee, R. *et al.* Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* 114, 6589–6631 (2014).
4. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. Design of DNA-binding peptides based on the leucine zipper motif. *Science* 249, 774–778 (1990).
5. Adzhubei, A. A., Sternberg, M. J. E. & Makarov, A. A. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* 425, 2100–2132 (2013).
6. Cascarina, S. M. & Ross, E. D. Identification of Low-Complexity Domains by Compositional Signatures Reveals Class-Specific Frequencies and Functions Across the Domains of Life. *PLOS Comput. Biol.* 20, e1011372 (2024).
7. Motta, M. *et al.* SHOC2 subcellular shuttling requires the KEKE motif-rich region and N-terminal leucine-rich repeat domain and impacts on ERK signalling. *Hum. Mol. Genet.* 25, 3824–3835 (2016).
8. Haynes, C. & Iakoucheva, L. M. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.* 34, 305–312 (2006).
9. Chaves-Arquero, B. & Pérez-Cañadillas, J. M. The Nrd1–Nab3–Sen1 transcription termination complex from a structural perspective. *Biochem. Soc. Trans.* 51, 1257–1269 (2023).
10. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein

structure-function paradigm. *J. Mol. Biol.* 293, 321–331 (1999).

11. Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* 49, 107–133 (2020).
12. Poudyal, M. *et al.* Intermolecular interactions underlie protein/peptide phase separation irrespective of sequence and structure at crowded milieu. *Nat. Commun.* 14, 6199 (2023).
13. Borcherds, W., Bremer, A., Borgia, M. B. & Mittag, T. How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Curr. Opin. Struct. Biol.* 67, 41–50 (2021).
14. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* 176, 419–434 (2019).
15. Naderi, J. *et al.* An activity-specificity trade-off encoded in human transcription factors. *Nat. Cell Biol.* 26, 1309–1321 (2024).
16. Toombs, J. A. *et al.* De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci.* 109, 6519–6524 (2012).
17. Boncella, A. E. *et al.* Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proc. Natl. Acad. Sci.* 117, 5826–5835 (2020).
18. Cascarina, S. M., Paul, K. R., Machihara, S. & Ross, E. D. Sequence features governing aggregation or degradation of prion-like proteins. *PLoS Genet.* 14, e1007517 (2018).
19. Baer, M. H., Cascarina, S. M., Paul, K. R. & Ross, E. D. Rational Tuning of the Concentration-independent Enrichment of Prion-like Domains in Stress Granules. *J. Mol. Biol.* 436, 168703 (2024).
20. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The Protein Folding Problem. *Annu. Rev. Biophys.* 37, 289–316 (2008).

21. Yamazaki, H., Takagi, M., Kosako, H., Hirano, T. & Yoshimura, S. H. Cell cycle-specific phase separation regulated by protein charge blockiness. *Nat. Cell Biol.* 24, 625–632 (2022).
22. Sologova, S. S., Zavadskiy, S. P., Mokhosoev, I. M. & Moldogazieva, N. T. Short Linear Motifs Orchestrate Functioning of Human Proteins during Embryonic Development, Redox Regulation, and Cancer. *Metabolites* 12, 464 (2022).
23. Yu, C.-L., Chuang, T.-W., Samuel, S. Y., Lou, Y.-C. & Tarn, W.-Y. Co-phase separation of Y14 and RNA in vitro and its implication for DNA repair. *RNA N. Y. N* 29, 1007–1019 (2023).
24. Girgenti, M. J., Pothula, S. & Newton, S. S. Stress and its impact on the transcriptome. *Biol. Psychiatry* 90, 102–108 (2021).
25. Hirose, T., Yamazaki, T. & Nakagawa, S. Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles. *Wiley Interdiscip. Rev. RNA* 10, e1545 (2019).
26. Hofmann, S., Kedersha, N., Anderson, P. & Ivanov, P. Molecular mechanisms of stress granule assembly and disassembly. *Biochim. Biophys. Acta Mol. Cell Res.* 1868, 118876 (2021).
27. Wadsworth, G. M. *et al.* RNA-driven phase transitions in biomolecular condensates. *Mol. Cell* 84, 3692–3705 (2024).
28. Fernandes, N. & Buchan, J. R. RNAs as Regulators of Cellular Matchmaking. *Front. Mol. Biosci.* 8, (2021).
29. Luo, Y., Na, Z. & Slavoff, S. A. P-Bodies: Composition, Properties, and Functions. *Biochemistry* 57, 2424–2431 (2018).
30. Caudron-Herger, M., Pankert, T. & Rippe, K. Regulation of nucleolus assembly by

non-coding RNA polymerase II transcripts. *Nucleus* 7, 308–318 (2016).

31. Braselmann, E., Chaney, J. L. & Clark, P. L. Folding the proteome. *Trends Biochem. Sci.* 38, 337–344 (2013).
32. Rinauro, D. J., Chiti, F., Vendruscolo, M. & Limbocker, R. Misfolded protein oligomers: mechanisms of formation, cytotoxic effects, and pharmacological approaches against protein misfolding diseases. *Mol. Neurodegener.* 19, 20 (2024).
33. Emin, D. *et al.* Small soluble α-synuclein aggregates are the toxic species in Parkinson's disease. *Nat. Commun.* 13, 5512 (2022).
34. Lee, C.-T. & Terentjev, E. M. Mechanisms and rates of nucleation of amyloid fibrils. *J. Chem. Phys.* 147, 105103 (2017).
35. Wittmer, Y. *et al.* Liquid Droplet Aging and Seeded Fibril Formation of the Cytotoxic Granule Associated RNA Binding Protein TIA1 Low Complexity Domain. *J. Am. Chem. Soc.* 145, 1580–1592 (2023).
36. Chiti, F. *et al.* Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci.* 99, 16419–16426 (2002).
37. Tsoi, P. S. *et al.* Electrostatic modulation of hnRNPA1 low-complexity domain liquid-liquid phase separation and aggregation. *Protein Sci. Publ. Protein Soc.* 30, 1408–1417 (2021).
38. Mukherjee, S. & Schäfer, L. V. Thermodynamic forces from protein and water govern condensate formation of an intrinsically disordered protein domain. *Nat. Commun.* 14, 5892 (2023).
39. Juković, M., Ratkaj, I., Kalafatovic, D. & Bradshaw, N. J. Amyloids, amorphous aggregates and assemblies of peptides – Assessing aggregation. *Biophys. Chem.* 308, 107202 (2024).
40. L. Almeida, Z. & M. M. Brito, R. Structure and Aggregation Mechanisms in Amyloids.

Molecules 25, 1195 (2020).

41. Adachi, M. *et al.* Aggregation-phase diagrams of β2-microglobulin reveal temperature and salt effects on competitive formation of amyloids versus amorphous aggregates. *J. Biol. Chem.* 293, 14775–14785 (2018).
42. Majid, N. & Khan, R. H. Protein aggregation: Consequences, mechanism, characterization and inhibitory strategies. *Int. J. Biol. Macromol.* 242, 125123 (2023).
43. Sabate, R., Rousseau, F., Schymkowitz, J., Batlle, C. & Ventura, S. Amyloids or prions? That is the question. *Prion* 9, 200–206 (2015).
44. Winklhofer, K. F., Tatzelt, J. & Haass, C. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. *EMBO J.* 27, 336–349 (2008).
45. Verma, M., Vats, A. & Taneja, V. Toxic species in amyloid disorders: Oligomers or mature fibrils. *Ann. Indian Acad. Neurol.* 18, 138–145 (2015).
46. Mukherjee, A., Morales-Scheihing, D., Butler, P. C. & Soto, C. Type 2 Diabetes as a Protein Misfolding Disease. *Trends Mol. Med.* 21, 439–449 (2015).
47. Kittleson, M. M. *et al.* Cardiac Amyloidosis: Evolving Diagnosis and Management: A Scientific Statement From the American Heart Association. *Circulation* 142, e7–e22 (2020).
48. Lim, J. & Yue, Z. Neuronal aggregates: formation, clearance and spreading. *Dev. Cell* 32, 491–501 (2015).
49. Basu, S. *et al.* Unblending of transcriptional condensates in human repeat expansion disease. *Cell* 181, 1062-1079.e30 (2020).
50. Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U. S. A.* 95, 13363–13383 (1998).
51. Prusiner, S. B. & Hsiao, K. K. Human prion diseases. *Ann. Neurol.* 35, 385–395 (1994).
52. Wickner, R. B. *et al.* Yeast prions: structure, biology, and prion-handling systems. *Microbiol.*

Mol. Biol. Rev. MMBR 79, 1–17 (2015).

53. Cascarina, S. M. & Ross, E. D. Yeast prions and human prion-like proteins: sequence features and prediction methods. *Cell. Mol. Life Sci. CMLS* 71, 2047–2063 (2014).
54. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* 137, 146–158 (2009).
55. Sabate, R., Rousseau, F., Schymkowitz, J. & Ventura, S. What Makes a Protein Sequence a Prion? *PLoS Comput. Biol.* 11, e1004013 (2015).
56. Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438 (2017).
57. Mackenzie, I. R. A. & Rademakers, R. The role of TDP-43 in amyotrophic lateral sclerosis and frontotemporal dementia. *Curr. Opin. Neurol.* 21, 693–700 (2008).
58. Guo, L. & Shorter, J. Biology and Pathobiology of TDP-43 and Emergent Therapeutic Strategies. *Cold Spring Harb. Perspect. Med.* 7, a024554 (2017).
59. Maziuk, B., Ballance, H. I. & Wolozin, B. Dysregulation of RNA Binding Protein Aggregation in Neurodegenerative Disorders. *Front. Mol. Neurosci.* 10, 89 (2017).
60. Petrakis, S. *et al.* Identification of human proteins that modify misfolding and proteotoxicity of pathogenic ataxin-1. *PLoS Genet.* 8, e1002897 (2012).
61. Morelli, C. *et al.* RNA modulates hnRNPA1A amyloid formation mediated by biomolecular condensates. *Nat. Chem.* 16, 1052–1061 (2024).
62. Ayers, J. I. *et al.* Different α-synuclein prion strains cause dementia with Lewy bodies and multiple system atrophy. *Proc. Natl. Acad. Sci.* 119, e2113489119 (2022).
63. Shorter, J. Liquidizing FUS via prion‐like domain phosphorylation. *EMBO J.* 36, 2925–2927

(2017).

64. Rojas, M. *et al.* Yeast Gis2 and its human ortholog CNBP are novel components of stress-induced RNP granules. *PloS One* 7, e52824 (2012).
65. Hofmann, S., Kedersha, N., Anderson, P. & Ivanov, P. Molecular mechanisms of stress granule assembly and disassembly. *Biochim. Biophys. Acta Mol. Cell Res.* 1868, 118876 (2021).
66. Wang, Z., Zhang, C., Fan, C. & Liu, Y. Post-translational modifications in stress granule and their implications in neurodegenerative diseases. *Biochim. Biophys. Acta Gene Regul. Mech.* 1866, 194989 (2023).
67. Marcelo, A., Koppenol, R., de Almeida, L. P., Matos, C. A. & Nóbrega, C. Stress granules, RNA-binding proteins and polyglutamine diseases: too much aggregation? *Cell Death Dis.* 12, 592 (2021).
68. Tudek, A. *et al.* Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Mol. Cell* 55, 467–481 (2014).
69. Darby, M. M., Serebreni, L., Pan, X., Boeke, J. D. & Corden, J. L. The Saccharomyces cerevisiae Nrd1-Nab3 Transcription Termination Pathway Acts in Opposition to Ras Signaling and Mediates Response to Nutrient Depletion. *Mol. Cell. Biol.* 32, 1762–1775 (2012).
70. Webb, S., Hector, R. D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* 15, R8 (2014).
71. Loya, T. J., O'Rourke, T. W., Simke, W. C., Kelley, J. B. & Reines, D. Nab3's localization to a nuclear granule in response to nutrient deprivation is determined by its essential prion-like

domain. *PLoS ONE* 13, e0209195 (2018).

72. Hutchinson, K. M., Hunn, J. C. & Reines, D. Nab3 nuclear granule accumulation is driven by respiratory capacity. *Curr. Genet.* 68, 581–591 (2022).
73. Loya, T. J., O'Rourke, T. W., Degtyareva, N. & Reines, D. A Network of Interdependent Molecular Interactions Describes a Higher Order Nrd1-Nab3 Complex Involved in Yeast Transcription Termination *. *J. Biol. Chem.* 288, 34158–34167 (2013).
74. O'Rourke, T. W., Loya, T. J., Head, P. E., Horton, J. R. & Reines, D. Amyloid-like assembly of the low complexity domain of yeast Nab3. *Prion* 9, 34–47 (2015).
75. Loya, T. J., O'Rourke, T. W. & Reines, D. Yeast Nab3 Protein Contains a Self-assembly Domain Found in Human Heterogeneous Nuclear Ribonucleoprotein-C (hnRNP-C) That Is Necessary for Transcription Termination. *J. Biol. Chem.* 288, 2111–2117 (2013).
76. Loya, T. J., O'Rourke, T. W. & Reines, D. The hnRNP-like Nab3 termination factor can employ heterologous prion-like domains in place of its own essential low complexity domain. *PloS One* 12, e0186187 (2017).
77. Lee, K. E., Procopio, R., Pulido, J. S. & Gunton, K. B. Initial Investigations of Intrinsically Disordered Regions in Inherited Retinal Diseases. *Int. J. Mol. Sci.* 24, 1060 (2023).
78. Fritzsche, M. & Charras, G. Dissecting protein reaction dynamics in living cells by fluorescence recovery after photobleaching. *Nat. Protoc.* 10, 660–680 (2015).
79. Hunn, J. C., Hutchinson, K. M., Kelley, J. B. & Reines, D. Variable penetrance of Nab3 granule accumulation quantified by a new tool for high-throughput single-cell granule analysis. *Curr. Genet.* 68, 467–480 (2022).
80. Protter, D. S. W. *et al.* Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep.* 22, 1401–1412 (2018).

81. Bremer, A. *et al.* Deciphering how naturally occurring sequence features impact the phase behaviors of disordered prion-like domains. *Nat. Chem.* 14, 196–207 (2022).