Acta Biochim Biophys Sin 2023, 55(7): 1119–1132
https://doi.org/10.3724/abbs.2023131
Advance Access Publication Date: 18 July 2023
Review

Review

Sequence variations of phase-separating proteins and resources for studying biomolecular condensates

Gaigai Guo[1,†], Xinxin Wang[1,†], Yi Zhang[1], and Tingting Li[1,2,*]

[1]Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China, and [2]Key Laboratory for Neuroscience, Ministry of Education/National Health Commission of China, Peking University, Beijing 100191, China
[†]These authors contributed equally to this work.
*Correspondence address. Tel: +86-10-82801585; E-mail: litt@hsc.pku.edu.cn

Received 13 April 2023 Accepted 6 June 2023

Abstract

Phase separation (PS) is an important mechanism underlying the formation of biomolecular condensates. Physiological condensates are associated with numerous biological processes, such as transcription, immunity, signaling, and synaptic transmission. Changes in particular amino acids or segments can disturb the protein's phase behavior and interactions with other biomolecules in condensates. It is thus presumed that variations in the phase-separating-prone domains can significantly impact the properties and functions of condensates. The dysfunction of condensates contributes to a number of pathological processes. Pharmacological perturbation of these condensates is proposed as a promising way to restore physiological states. In this review, we characterize the variations observed in PS proteins that lead to aberrant biomolecular compartmentalization. We also showcase recent advancements in bioinformatics of membraneless organelles (MLOs), focusing on available databases useful for screening PS proteins and describing endogenous condensates, guiding researchers to seek the underlying pathogenic mechanisms of biomolecular condensates.

Key words phase separation, sequence variation, biomolecular condensate, disease

Introduction

Biomolecular condensates organize cellular components in a temporally and spatially appropriate manner. It has been shown that the formation of some MLOs is initiated by liquid-liquid phase separation (LLPS) of particular proteins [1–3]. There are other cases where proteins switch to highly ordered solid-like phases, such as the fibrillar assemblies of huntingtin exon 1 proteins in Huntington's disease [4]. The condensed state can improve the efficiency of cellular processes, which are involved in a range of cellular processes, including transcription, chromatin organization, RNA processing, protein homeostasis, innate immunity, cell-cell adhesions, signaling, and synaptic transmission [5–12]. For example, an increase in local molecular concentration facilitates the reaction rate of enzymes, such as cyclic GMP-AMP synthase (cGAS), in innate immune signaling [7]. These condensates are usually linked to a subset of phase-separating proteins, which in turn govern their components and material states. Thus, variations observed in such protein-coding sequences can substantially impact the functions of biomolecular condensates, which contribute to various diseases.

Dysfunctional condensation is implicated in a variety of pathological processes linked to liquid- or solid-like condensates. For example, mutations of methyl CpG binding protein 2 (MeCP2) can disrupt heterochromatin assembly, which forms via LLPS in native status and develops into transcriptional dysregulation in Rett syndrome [13]. Gain of LLPS propensity is observed frequently in tumorigenesis, such as in the manner of in-frame fusion with IDRs or LCDs [14–16]. In addition, numerous disorders are caused by liquid-solid transitions [17]. Irreversible mutations of many RNA binding proteins can contribute to degenerative diseases, such as mutations of TDP-43, hnRNPA1, and FUS in amyotrophic lateral sclerosis/frontotemporal dementia (ALS/FTD) [9,18,19].

In this review, we first characterize four types of sequence variations associated with dysfunctional condensates. We then provide a collective assessment of current computational resources for describing proteins driving liquid-like PS or solid-like phase transition, as well as their relations with disease-associated condensates.

© The Author(s) 2023. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).

Interaction Modes and Theories of Phase Separation

Theories that try to explain PS proteins in a universal mechanism often emphasize the importance of multivalency [20,21]. One of the most well-studied theories is the stickers-spacers framework (Figure 1A) [22–24]. Multivalent proteins are generally constructed of diverse attractive groups, including amino acid residues, modular domains, or emerging stickers that function in self-oligomerization (Figure 1B). Spacers are interspersed between stickers and are, on most occasions, considered not to affect PS directly. In this theory, stickers are grouped into three conceptual modes in terms of the molecular identities of stickers (Figure 1C). For folded proteins, the surfaces of folded domains can be considered stickers. For linear multivalent proteins, stickers may appear sporadically along the sequences as regional complexes. For intrinsically disordered regions (IDRs), stickers are likely a particular type of residue (aromatic, charged, or polar amino acids), short linear motifs (SLiMs), or modular domains involved in biomolecular recognition. According to the stickers-spacers theory, we next briefly discuss how sticker valence, interaction strength, and spacer patterning influence the phase behavior of PS proteins.

Driving Force of Phase Separation

Types of protein stickers

Stickers can be distinguished by employing a series of computational [25,26], theoretical [27], and experimental approaches [28–30], most of which are based on the well-known architectures of multivalent proteins or those binding with RNA molecules [29,31,32]. Conceptually, these biological stickers can be categorized into three classes: folded domains, specific linear motifs, and individual residues in IDRs [22,27]. The multivalent condensates can originate from cation-π, π-π, charged-charged, or dipole-dipole interactions (Table 1). Alternatively, emergent stickers can also drive condensate formation through homotypic attractions, namely, oligomerization or polymerization, such as the coiled-coil (CC) domain of ORF1 [34] and the sterile alpha motif (SAM) of Shank3 [35].

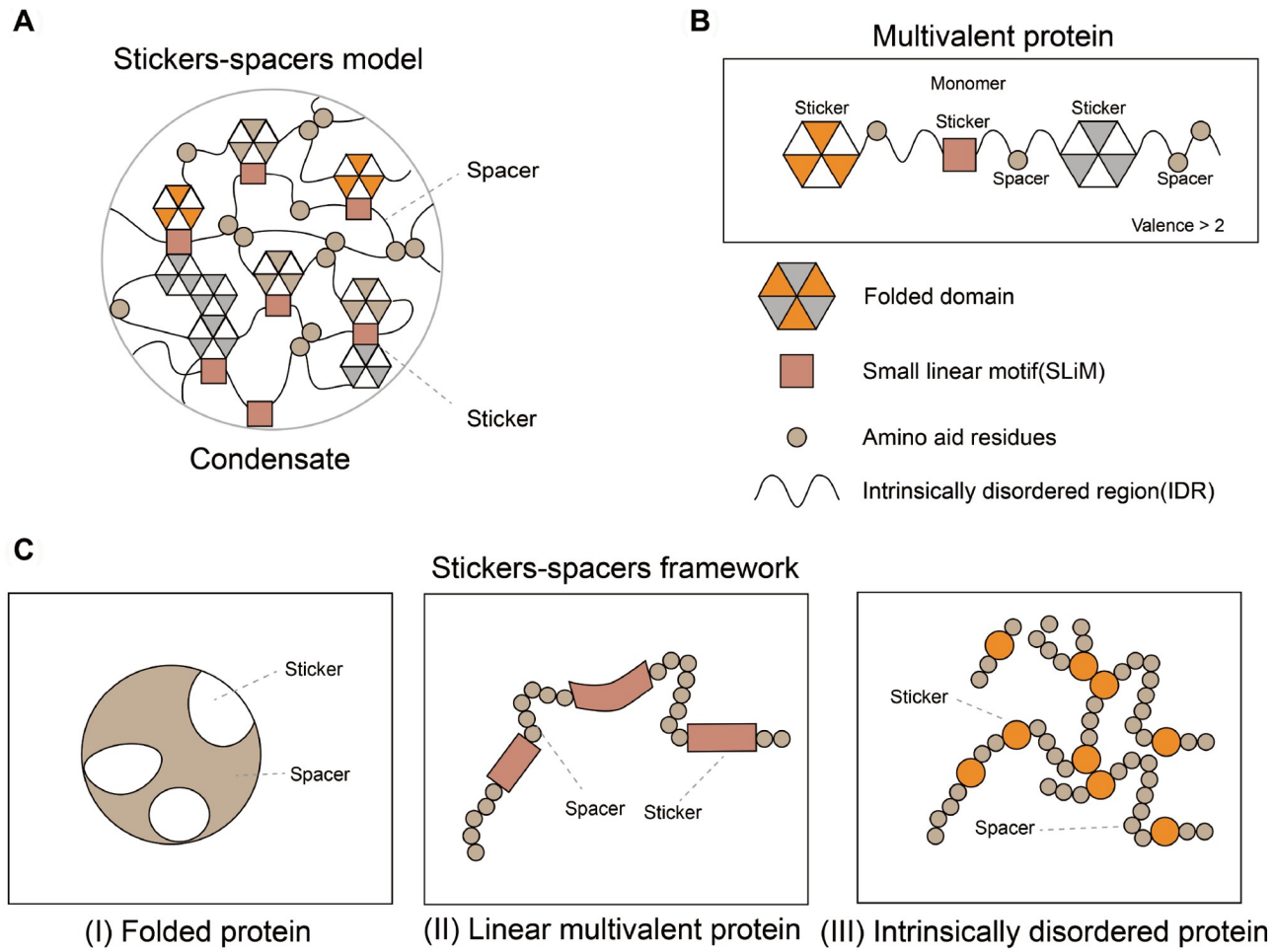


Figure 1. Theories of phase separation (A) Comprehensive composition of membrane-less organelles (MLOs) that consist of phase-separating proteins driven by diverse forces. Biological MLOs are frequently a mixture formed through homotypic and heterotypic attractions between phase-separating proteins. (B) Phase-separating proteins are considered linear polymers that can interact with each other via attractive groups of different types. Polymer proteins are multivalent. Most of these proteins have at least two types of attractive groups, including domains for polymerization, linear motifs, or specific amino acid-rich regions, which are linked by spacers. (C) Three modes of systems with conceptually defined stickers and spacers, including: (I) folded proteins, in which the interaction patches are stickers and the regions on the folded domain are spacers that serve as the surface scaffold for stickers; (II) linear multivalent proteins, in which the folded binding motifs are stickers for molecule recognition. Spacers are linkers connecting those domains; and (III) intrinsically disordered proteins (IDPs), in which stickers are single residues, short linear motifs, or some combination of the two.

Table 1. Representative phase-separating proteins related to important biomolecular condensates or diseases

Protein	Material property	Condensate	Attractive sticker	Biological processes	Ref.
Residue-rich IDR-mediated condensate					
NPM1	Liquid	Nucleolus	S/R-rich	Anaplastic large cell lymphoma (NPM1-ALK fusion protein)	[36]
HOXD13	Liquid	Transcriptional condensate	A-rich	Hereditary synpolydactyly (expansion of polyalanine)	[37]
G3BP1	Liquid, solid	Stress granule	E-rich	Activates or modulates innate and adaptive immune responses	[31,38]
RBGD2/4	Liquid	Stress granule	F/Y-rich	Heat resistance regulation	[39]
FXR1	Liquid	mRNA transportation and translation related condensate	S/R-rich	Congenital multi-minicore myopathy (alternative splicing)	[40]
SRSF2	Liquid	Nuclear speckle	S/R-rich	RNA splicing and spliceosome assembly	[41]
UBQLN2	Solid	Stress granule	P-rich	Regulation of autophagosome assembly; ubiquitin-dependent protein catabolic process	[42,43]
hnRNPDL	Liquid, solid	mRNA transportation and translation related condensate	R/Y-rich, G/T-rich	Transcription repression	[44]
FUS	Solid	Stress granule	G/Y/R-rich, polyS/G	Myxoid liposarcoma (FUS-CHOP); ALS/FTD (missense mutation)	[29,45]
SAFB	Liquid	Transcriptional condensate	R/G-rich	Pericentromeric heterochromatin organization	[46]
ARID1A	Liquid	Spliceosome	P/S-rich	Recognition of large constitutive exons	[47]
K-Ras4B	Liquid	P granule	K-rich	GTPase, regulating proliferation, apoptosis and cell growth	[48]
U1-70K	Liquid	Nuclear body	D/E/R/K-rich	RNA splicing	[49]
FBL	Liquid	Nucleolus	G/A-rich	Ribosomal RNA processing	[50]
Prospero	Liquid	Heterochromatin	Aromatic/nonpolar residues	Regulation of terminal neuronal differentiation	[51]
Cavin1	Liquid, hydrogel	Caveolae-related condensate	D/E -rich	Caveolae formation and organization	[52]
CYCT1	Liquid	Nuclear speckle	H-rich	Regulation of cell cycle	[53]
FLOE1	Liquid	Desiccation-related condensate	Q/P/S-rich	Response to desiccation stress	[54]
TDP-43	Liquid, solid	Neuronal granule	Q/N-rich	RNA biogenesis and processing	[55,56]
Motif/domain-mediated condensate					
GRB2-SOS	Hydrogel	Receptor cluster	PRM_SH3	Signal transduction (downstream of RTK signaling pathway)	[57]
Nephrin-NCK	Liquid	Receptor cluster	pT_SH2	Signal transduction (downstream of RTK signaling pathway)	[58]
PML	Liquid	PML nuclear body	SIM_SUMO	Regulation of apoptosis and growth suppression	[1]
MATR3	Liquid	RNA-binding condensate	RRM_RNA	Innate immune response	[59]
cGAS	Liquid, solid	DNA-binding condensate	DNA binding domain	Innate immune response	[7]
TRF1/2	Liquid	DNA-binding condensate	DNA binding domain	Telomere maintenance	[60]
Edc3-Pdc1	Liquid	P-body	LSm_HLM	Deadenylation-independent decapping of nuclear-transcribed mRNA	[61]
SPOP/DAXX	Liquid, solid	Nuclear speckle	BTB_BACK	Prostate cancer (missense mutation)	[8]
Nucleocapsid	Liquid	Stress granule	Zinc-finger	HIV-1 viral nucleocapsid	[62]
RAD23	Liquid	Proteasome containing foci	UBL_UBA	Protein degradation	[63]
SynGAP-PSD95	Liquid	Postsynaptic density	PBM_PDZ	Signal transduction (downstream of glutamate receptor)	[12]
Pon	Liquid	Cell polarization-related condensate	FxNxx[F/L]NP[F/Y]E[V/I]xR	Asymmetric cell division	[64]
TNRC6B-Ago2	Liquid, solid	miRISC	G/W_PIWI	RNA-mediated gene silencing	[65]

(*Continued*)

Protein	Material property	Condensate	Attractive sticker	Biological processes	Ref.
Repeats-mediated condensate					
LAF-1	Liquid	P granule	RGG	ATP-dependent RNA helicase	[66]
NUP98	Solid	Nuclear pore complex	GLFG	Acute myeloid leukemia (NUP98 fusions with HOXA9, KDM5A, or LNP1)	[67]
PGL-3	Liquid	P granule	RGG	Germline development regulation	[68]
EMB1579	Liquid	Transcriptional condensate	RE/RD/RED	Regulation of DNA-templated transcription	[69]
C9orf72	Liquid	Guanine nucleotide exchange-related condensate	PR/GR/GA	ALS/FTD (expansion of polyPR/GR/GA)	[70]
RNA Pol II	Liquid	Transcriptional condensate	YSPTSPS	DNA-dependent RNA polymerase	[71]
MDC1	Liquid	DNA repairing foci	PST	DNA double-strand breaks repairing factors	[72]
tau	Liquid, solid	Microtubule-related condensate	pseudo-repeat region (~31 aa)	Promotion of microtubule assembly and stability	[73]
Polymerization domain-mediated condensate					
Shank3	Liquid	Postsynaptic density	SAM domain	Signal transduction (downstream of glutamate receptor)	[35]
EML4	Solid	Actin-binding condensate	Coiled-coil	Non-small cell lung cancer (ELM4-ALK)	[14,74]
Huntingtin	Liquid	Huntingtin inclusion	polyQ (β-sheet)	Huntington's disease (expansion of polyglutamine tract)	[75,76]
hnRNPA1	Liquid, solid	Stress granule	G-rich (β-sheet)	Multisystem proteinopathy; ALS/FTD (missense mutation)	[9]
PrP	Solid	Aβ oligomers	Polybasic PLD	Prion diseases	[77]
AfrLEA6	Liquid, solid	Stress granule	SMP domain	Desiccation tolerance	[78]
TFEB	Liquid	Autophagy-related condensate	bHLH region	Regulation of autophagy and lysosomal biogenesis	[79]
Osh4	Solid	Stress granule	OSBP-related domain	Lipid transportation and direction of cell polarized growth	[80]
53BP1	Liquid	Heterochromatin	Oligomerization domain	Regulation of DNA double strand break repairment	[81]
Dvl-2	Liquid, solid	Wnt pathway-related condensate	DIX domain	Upstream Wnt signaling effector	[82,83]
EBNA2	Liquid	Chromatin	Dimerization domains	Activation of the host resting B-cell and stimulation of B-cell proliferation	[84]
TMF	Liquid	Transcriptional condensate	Oligomerization domain	Recognition of ROS and regulation of flowering transition	[85]

[a]Abbreviations: ALS/FTD, amyotrophic lateral sclerosis/frontotemporal dementia; PML, promyelocytic leukemia; RTK, receptor of tyrosine kinase; SH2, src homology 2; SIM, SUMO-interacting motif; SUMO, sentrin/small ubiquitin-like modifier; RRM, RNA recognition motif; HLM, helical leucine-rich motif; BTB, Broad-complex, Tramtrack and Bric-à-brac; BACK, BTB and C-terminal Kelch; UBL, ubiquitin-like; UBA, ubiquitin associated; PBM, PDZ binding motif; PDZ, post-synaptic density-95, disks-large and zonula occludens-1; PIWI, P-element-induced wimpy testis; SAM, sterile alpha motif; SMP, synaptotagmin-like mitochondrial-lipid binding protein; bHLH, basic helix-loop-helix; OSBP, oxysterol-binding protein; ROS, reactive oxygen species.[b]Driving factors, effects, and related diseases of phase-separating proteins are listed.[c]The categories and positions of driving forces are distinguished according to the original literature.

Parameters of effective stickers

The driving force for PS is governed by three factors, including the effective valence of stickers, the interaction strengths between stickers, and the patterning of spacers in the protein sequence [22,27]. The multivalence of interaction domains or motifs represents a hallmark of proteins that drive phase transitions [20], such as the heterotypic interactions mediated by poly-SIM/poly-SUMO segments [86] and PRM-binding SH3 domains (GRB2-SOS) [57]. Increasing the sticker valence or interaction strength can lower the saturation concentration (c_{sat}) and thus expand the PS propensity, similar to the effects of modulating the aromatic residues in hnRNPA1-LCD [87]. Additionally, the reduction of valency via destruction of self-oligomerization can also influence phase behavior, such as the CC domain in ORF1 [34] and SAM domains in the Polycomb Repressive Complex 1 (PRC1) subunit Polyhomeotic (Ph) [88]. For example, Newton and coworkers [34] demonstrated a cooperative interaction between the LCD and CC domain of ORF1, of which both mutations of essential residues in the LCD or CC domain can lead to a repressed PS.

It has been reported that amino acids with similar chemical properties can demonstrate various sticker strengths. Bremer *et al.* [89] found that Tyr is stronger than Phe as a sticker, while Lys weakens the sticker-sticker interaction, such as in the case of prion-like domain (PLD) proteins. Additionally, a few other studies have

discovered that clustering and segregation of stickers along linear sequences can regulate the biological functions of corresponding proteins by affecting the driving force of PS [90,91]. A prediction-based study from Das *et al.* [91] showed that the strength of sticker-sticker interactions will increase when charged stickers are segregated along the PLD regions. Additionally, it was recently demonstrated that charged blocks are required for specific compartmentalization and that the IDRs of similar patterned charged blocks present similar and selective functions [92]. However, the segregation pattern of aromatic residues can lead to pathological and irreversible aggregation, such as the PLD of hnRNPA1 [87]. In contrast, uniform distribution of aromatic residues (Tyr) will promote LLPS, such as two RNA binding proteins, RBGD2 and RBGD4 [39].

Impacts of spacers

Spacers affect the phase behaviors and material properties of condensates by influencing the effective solvation volume (v_{es}) of PS proteins [22]. The stickers-and-spacers framework can be well applied to PLD proteins. A decrease in v_{es} stabilizes the PS process of A1-LCD, a homolog of hnRNPA1, when the Ser content is increased [89]. It is also possible to increase v_{es} by increasing the charged content [93]. Under specific circumstances, spacers behave as attractive groups and provide multivalent cross-links [94]. Spacers are typically composed of Ser, Gly, and Gln residues, while poly-residue tracts perform their duty as spacers in an alternative way. Kar *et al.* [29] reported that peptides with uniformly distributed Gly do not form fibrils or gels but can promote self-assembly in the form of poly-Gly tract. Some poly tracts can lead to pathological consequences, e.g., phase transitions of some poly-Gln proteins could initiate irreversible pathological aggregation. Peskett and coworkers [4] found that Huntingtin (Htt) exon 1, a Gln- and Pro-rich region, drives the Htt protein to form solid-like assemblies.

Sequence Variations and Regulation of Phase Separation

Dysregulation of phase transitions of biomolecule condensates is linked to a series of pathogenic conditions, including the toxicity induced by the liquid–solid transition of ALS/FTD mutations in FUS (R522G, R524S, P525L, and R495X) [19] and the promotion of tumorigenesis for gain of PS properties, such as NUP98 fusion oncoproteins (NUP98-HOXA9 and NUP98-KDM5A) in leukemic transformation [95]. In particular, many natural mutations in IDRs are associated with diseases [96]. To systematically understand the relationship between sequence variations and condensate dysfunction, we summarize the reports on three type of disease-associated variations (missense mutations, expansions of peptide repeats, and gene fusions) involved in aberrant condensates and one type of post-transcriptional modification process, alternative splicing (AS) (Figure 2). Based on current findings, we attempt to recapitulate the underlying mechanisms that may be implicated in the pathogenic behaviors of dysregulated condensates.

Missense mutation

Missense mutations are a critical type of variation observed in proteins that affect biomolecular assembly. To obtain the overall landscape of functional mutations of PS proteins, we referred to PhaSepDB, which provides a collection of manually curated PS-related proteins [97]. Missense mutations altering the phase behaviors of condensates often occur in diverse conformational regions, such as IDRs, modular domains, or polymerization domains (Figure 2A). Regardless of the complexity and structure of regions harboring mutations, the impacts of mutations on a protein's phase behavior can be two-sided. For example, mutations of aromatic to hydrophobic residues (F291S and Y283S) observed in the IDRs of hnRNPA2 can slightly repress its homotypic assembly, while mutations near the sites D290V and P298L will enhance its condensed status [98]. Interestingly, a number of mutations have been discovered in domains that function in protein polymerization, such as the monomeric SPOP mutant $SPOP^{mutBTB/BACK}$. The mutant tends to diffuse in the nucleus and is unable to localize to SPOP-DAXX bodies in HeLa cells [8] Some mutations affect protein phase behavior by monitoring the valency of condensates. Stortz and coworkers [99] found that GR P481R, a constitutively tetrameric glucocorticoid receptor (GR) mutant, presents a higher capability to form condensates, while most monomeric receptor GR mutants (A465T/I634A) do not form foci in cells. In summary, the perturbation of protein homotypic or heterotypic polymerization is an important manner that affects proteins' capability to form condensates [8,98,99].

In addition to the disorder status and frequency of particular mutations, the physiochemical properties of drivers are relatively preserved in the regulation of PS. Previous study shows that mutating attractive groups into hydrophilic amino acids partially accounts for the loss of PS propensity. For example, the W131G mutation is expected to have an impact on SPOP phase behavior by

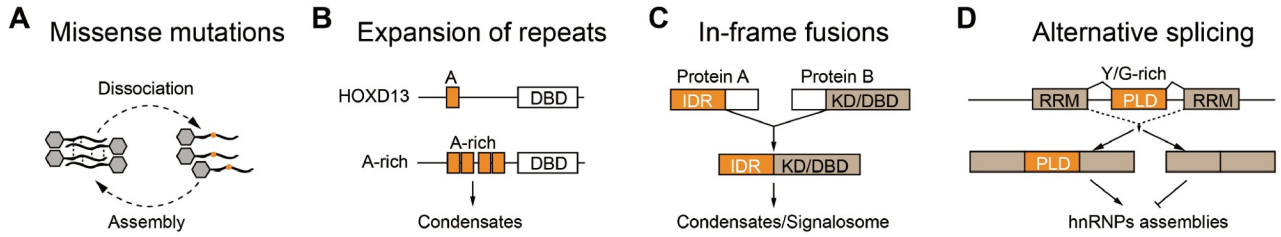


Figure 2. Schematic of four types of gene variations (A) Missense mutations. The impact of missense mutations on phase-separating proteins is two-sided, which is, to some extent, dependent on the structure and general properties of the sequence where mutations occur. (B) Expansion of repeats. An instantiation related to the expansion of Ala-rich repeats. Synpolydactyly associated repeat expansions enhance phase separation of HOXD13-IDR. IDR, intrinsically disordered domain; DBD, DNA binding domain; RRM, RNA recognition motif; PLD, prion-like region. (C) Gene fusions. Functional domains (KD/DBD) fused with phase-separation-prone domains (IDR) can enhance their activities, such as constituent activation of the receptor tyrosine kinase (RTK) pathways in the cellular signalosome. KD, kinase domain; DBD, DNA binding domain. (D) Alternative splicing. The inclusion of exons harboring Y/G-rich PLD promotes multivalent assemblies. For example, the alternative splicing of hnRNPs contributes to the control of global splicing and signaling regulation.

increasing its v_{es}, according to the stickers-spacers model [8]. Conversion to hydrophobic or charged amino acids can accelerate the transition from the liquid state to the solid-state or fibrils. These interacting stickers include aromatic, acidic, or basic amino acids, which are observed frequently in IDRs or low complexity domains (LCDs), such as the FUS G156E mutant, hnRNPA1 D262V mutant, and hnRNPA2 R521C mutant [32,98]. Changes in charged residues often affect the phase behaviors of many posttranslationally modified proteins, such as methylated hnRNPA2 [100] and phosphorylated NPM1 [92]. Nevertheless, the regulation of PS is complicated and context-dependent in cells. These findings demonstrate that appropriate sequence modifications are advisable in *de novo* design for manipulating PS.

Expansion of pathogenic repeats

Expansion of sequence repeats is another key type of disease-associated variation that occurs in dozens of inherited human disorders [101,102]. Previous reports indicated that a majority of sequence expansions are observed in transcription factors (TFs) and proteins associated with neurodegenerative diseases [103]. Expansions of repeated sequences alter their phase behaviors and abilities to recruit regulators, leading to pathogenetic assembly and aggregation. Basu *et al.* [37] found that the expansion of Ala repeats facilitates the homopolymerization and PS of TF HOXD13 (Figure 2B), which in turn excludes the incorporation of MED1, a subunit of the Mediator coactivator. Furthermore, changes in components in the HOXD13-containing assembly due to the global expansion of Ala repeats alter the transcriptional program in a mouse model of synpolydactyly. These findings suggest a framework of how the expanding repeats in IDRs of TFs (HOXA13, RUNX2, and TBP) contribute to human pathologies [37]. Another similar example is the expansion of the GGGGCC repeat in chromosome 9 open reading frame 72 (C9orf72), a key cause of ALS/FTD [104]. The products containing polyGly-Arg (GRn) and polyPro-Arg (PRn) are able to undergo LLPS [70] as toxic dipeptides contribute to pathology. Expanding repeats interact with many other proteins and alter their PS propensity. GRn and PRn repeating sequences perturb the composition of nucleoli, stress granules, and RNA granules, which might play a dominant role in the pathogenesis of ALS/FTD [105]. A well-studied example is the Huntington's disease-associated protein Htt, which contains a disordered Gln-rich region (PolyQ) in its first exon [106]. Mutant Htt protein (mHtt) harboring PolyQ expansion diffuses slowly and forms more stable and solid aggregates than WT Htt [107]. This example indicated that LLPS might mediate solid-like assembly formation [108]. Li *et al.* [109] discovered that the aggregation of mHtt can reduce its targeting sites and impair gene expression programs in neuronal cells.

Gene fusions

Fusion genes are a crucial category of cancer drivers that are identified in approximately 20% of cancer morbidity [110]. Among these, in-frame fusions are considered the major events in some cancers, such as BCR-ABL1 in acute myeloid leukemia and EWS-FLI1 in Ewing sarcoma [16,111]. As numerous IDRs or phase-separation-prone regions with the features mentioned above are supposed to be the drivers of PS, it is convincing that fusion proteins with such domains may have a higher potential to undergo PS. In fact, a significant proportion of oncoproteins include such disordered domains, such as PLDs or LCDs. Fusion TFs are an important category of oncoproteins that induce aberrant gene expression [112]. For example, in FET fusion oncoproteins, the disordered PLD is fused with the DNA-binding domains (DBDs) of TFs (CHOP, FLI1, and DDIT3), making PLDs behave as transactivation domains [113,114] (Figure 2C). These aberrant PLD-DBD fusions condense and recruit the transcriptional machinery and regulators, including RNA polymerase II, BRD4, and mSWI/SNF, which activate transcription independent of upstream stimuli [113,115,116]. In addition, another group of oncoproteins is NUP98 fusion proteins that share the FG-repeated IDR of NUP98, a component of the nuclear pore complex (NPC) [15,117]. For example, NUP98-HOXA9 promotes the global expression of leukemogenic genes due to an increase in its LLPS ability [95]. In addition to TFs, receptors of tyrosine kinases (RTKs) are another type of cancer driver [118]. Fusion with LCD from a phase-separating protein mediates the kinase domain (KD) function in cytoplasmic protein granules, constitutively activating the RAS pathway without ligand stimuli [74]. Previous evidence has demonstrated a similar mechanism in ALK fusions, such as ELM4-ALK and NPM1-ALK [14,119,120].

Alternative splicing

The length and amino acid composition of attractive groups in PS proteins can also be modulated pretranslationally through AS, which produces distinct mature mRNAs from a single pre-mRNA by including or skipping different exon segments. Approximately 95% of multiexon genes are expected to undergo AS [121–124], most of which are prone to expression in a tissue-specific manner, such as FXR1 splicing in muscle development [48)] The tissue-specific alternatively spliced exons are such regions where interactions frequently occur [125–127]. AS events are observed as a critical posttranscriptional mechanism that functions on a global scale to rewire cellular responses [128–130]. Interestingly, many studies have revealed that alternatively spliced exons are enriched in IDRs, presumably to maintain functional and regulatory diversity while avoiding the disruption of core protein structure [131–133]. Additionally, IDRs often contain linear motifs that mediate ligand recognition, which regulates the functions of such alternatively spliced IDPs [131,134].

AS of Ser- and Arg-rich IDRs can achieve different morphologies of condensates, *e.g.*, the splicing of FXR1 in the development of Xenopus. Smith *et al.* [40] reported that cells expressing a longer isoform of FXR1 generate larger and more puncta than those expressing the shorter isoform, which lacks exons 15 and 16. Both isoforms bind with RNAs and present concentration-dependent assemblies in cells. Moreover, they also observed that phosphorylation of the Ser residues in its IDRs is critical to regulating the function of FXR1, as phosphorylation by CK2 reduces the aggregation of the FXR1 isoform with the longer IDR. Thus, both the length of IDRs and the potential of PTMs can regulate the behaviors and functions of AS IDPs in a lineage-specific manner. Similarly, it has been suggested that Ser/Arg-rich splicing factors are likely to be subject to PTMs to modulate disorder status [135–137].

Previous studies discovered that mammalian exons with specific AS events are enriched in IDRs containing Gly- and Tyr-rich motifs, such as hnRNP families and other RBPs, which play diverse roles in RNA processing [138]. Such AS events are evolutionally conserved in the control of splicing sites. Aberrant assembly of these proteins

can lead to the formation of pathological aggregates implicated in many degenerative diseases [139]. Additionally, GY-rich IDRs have been proven to function in the formation of higher-order protein assemblies, including RNA granules, hydrogel-like MLOs, and fibrillar-like structures [138,140,141]. Skipping or including different exons regulates the phase behavior of hnRNP assemblies (Figure 2D), which in turn controls global AS events in cells and tissues.

It has been observed that AS regulates the assembly properties of RNA-processing proteins by manipulating the incorporation of IDRs in different isoforms [142]. Batlle *et al.* [44] demonstrated that hnRNPDL isoforms require interactions between Tyr and Arg residues in PLDs to phase separate, which is regulated by AS events. In fact, AS events are frequently observed in PLDs [138] and interfere with the PS property and ability to develop multivalent interactions in proteins, such as in the cases of hnRNP A1 and hnRNP A1B. Compared to the control hnRNP A1 isoform, hnRNP A1B harboring an elongated PLD domain has a greater fibrillization propensity, which is toxic to ALS patients [143].

Resources for Component Screening

Failure in the regulation of material states may lead to aberrant protein assemblies, which could trigger a range of pathological processes, with the sequence variations mentioned above as a possible prominent mechanism [144–147]. To aid the investigation of aberrant condensate-associated diseases, we provide emerging available databases and prediction methods, which are necessary to describe experimental phenomena, focusing on the survey for potential interactions with PS proteins and the regulatory mechanisms underlying the formation of these aberrant condensates.

Databases of proteins and RNAs involved in biological condensates

Several LLPS-associated protein and RNA databases were released until 2023. Since the original publications and web interfaces of these resources have already provided detailed explanations on their usage, we aim to provide a guide for users on the strengths of these resources and explain particular usage where they could be beneficial. The main features of the databases discussed are summarized in Tables 2 and 3. We collected features that highlight the basic principles and characteristic differences.

DrLLPS is an integrative database for the proteins that can undergo LLPS [148]. It classified PS proteins into two categories, *i.e.*, scaffolds and regulators, which were assigned to 40 biomolecular condensates. This database also documents PS-associated proteins in 164 eukaryotes through computational prediction. LLPSDB is a web-accessible database providing curated LLPS proteins and corresponding *in vitro* experimental conditions from published literature [149]. It provides the details for 273 PS proteins associated with 1182 experiments. LLPSDB annotated proteins with biomolecular data, such as protein sequence or protein modifica-

Table 2. Resources for phase-separating proteins and condensates

Name	Last updated	Scope	Annotation	Server	Ref.
DrLLPS	January 2019	150 scaffold proteins and 987 regulators; relations to 40 biomolecular condensates.	(a) Two categories: Scaffolds (drivers of LLPS) and regulators (regulation of MLOs formation: P bodies or stress granule); (b) basic annotations (protein sequence, modification, *etc.*); (c) genetic variations (SNP, cancer mutations, or gene fusions); (d) molecular interactions (proteins, nucleic acid, or relations to drugs).	llps.biocuckoo.cn	[148]
LLPSDB	July 2019	273 PS proteins: 198 natural proteins and 75 designed proteins.	(a) Biomolecular information (protein sequence, modification, *etc.*); (b) experimental conditions (validated in vitro) and phase behavior description; (c) linkage to other open comprehensive databases (Uniprot, MobiDB, or DisProt).	bio-comp.org.cn/llpsdb/	[149]
PhaSePro	September 2019	121 PS proteins: 109 eukaryotic, 5 bacterial and 7 viral proteins.	(a) Proteins/regions: LLPS alone or parts of LLPS systems with co-drivers; (b) experiment type and conditions: *in vivo* or *in vitro*; (c) regulation partnersand connections to MLOs and diseases (SNP).	phasepro.elte.hu	[150]
MloDisDB	April 2020	73 biomolecular condensates; 719 relationships between MLOs and diseases; 52 relationships between LLPS proteins and diseases; diseases: cancer, nervous system, infectious diseases/anemia, and apoptosis/aging.	(a) Basic information (MLO, factor, and disease information); (b) changes of the MLO (relations, organisms, and cell lines); (c) changes of factors (relation to the MLO-diseases or LLPS-diseases); (d) factors include proteins, RNA, chemicals, oligonucleotides, or peptides.	mlodis.phasep.pro/	[3]
PhaSepDB	June 2022	860 PS proteins: 300 PS-self proteins and 560 PS-other proteins; 590 MLO-related proteins (low-throughput) and 5292 MLO-related proteins (high-throughput).	(a) Two categories: PS-self proteins (PS by themselves *in vitro*) and PS-other proteins (PS with partners *in vitro*); (b) PS experiments: droplet state, phase diagram, *in vivo* or *in vitro*; (c) PS regulating macromolecules: co-PS protein, RNA/DNA, and chemicals; (d) PS regulating events on proteins: PTM, mutation, oligomerization, repeat region and alternative splicing.	db.phasep.pro/	[151]

SNP, single nucleotide polymorphism; LLPS, liquid-liquid phase separation; PTM, post-translational modification; MLO, membrane-less organelle.

Table 3. Resources for RNAs related to phase separation and condensates

Name	Last updated	Scope	Annotation	Server	Ref.
RNAPhaSep	June 2021	325 RNAs: validated 22 organisms.	(a) Basic information: ID, sequence, function, modification, and location; (b) binding partner: RNAs or proteins.	rnaphasep.cn	[152]
RPS	June 2021	337 RNAs: validated; 20,153 RNAs: high-throughput analyses; 1358 RNAs: predicted.	(a) RNAs type: protein-coding sequence, lncRNA, pseudogene, or virus RNA; (b) putative sequence features: motif, modification, 2D/3D structures, or binding site with partners; (c) relations to diseases; (d) condensates type: nucleus, cytoplasm, and others.	rps.renlab.org	[153]

tion, and specific PS information, such as experimental conditions and phase behavior descriptions. It is useful for the study of the biophysical background of LLPS. PhaSePro is a manually curated database that provides information on LLPS proteins with structural data. It describes proteins that can drive PS alone or as part of well-defined multicomponent systems [150]. While the database has only 121 proteins, it provides reliable PS drivers. MloDisDB is a curated collection of the relations between MLOs and diseases and the relations between LLPS and diseases [3]. It contains up to 771 relations with the corresponding diseases from 607 publications, which might facilitate the investigation of the pathogenic mechanism underlying PS-related diseases. The LLPS proteins in MloDisDB are the collection of the reviewed proteins from the DrLLPS, LLPSDB, PhaSePro, and PhaSepDB databases. PhaSepDB is a centralized resource that provides experimentally verified PS proteins and MLO-related proteins [97]. The database currently contains 300 PS-self proteins and 560 PS-other proteins. It provides detailed annotations for PS proteins, including the material states of PS droplets, regions used in experiments, phase diagrams, PS partners (protein, RNA, and others), and regulatory factors of PS ability (mutation or modification). The PS proteins in PhaSepDB are annotated with verification experiments based on original publications, such as droplet formation and fluorescence recovery after photobleaching (FRAP) experiments performed *in vivo* or *in vitro*. PhaSepDB is a suitable resource for those who want to gain an overview of *in vitro* and *in vivo* experiments performed in proof of LLPS of a target protein.

Increasing evidence has demonstrated the essential roles of RNAs in the functions and maintenance of MLOs such as nucleoli, P-bodies, and stress granules. At present, two resources have addressed and collected RNAs related to PS and condensates (Table 3). RNAPhaSep is a curated database that contains self-assembling RNAs or RNAs and proteins, which can undergo cophase separation in validated experiments [152]. RPS integrates 21,613 LLPS-related RNAs with validated, high-throughput, and predicted evidence [153]. It provides RNAs with informative annotations, such as putative motifs, structures, modification sites, or related diseases. Both databases serve as valuable platforms for the investigation of RNA-related PS and biological processes.

Prediction tools for phase-separating proteins

A range of computational algorithms have been developed to predict PS proteins, which are to some extent applicable for proteins with higher contents of IDRs or LCDs. PScore predicts long-range π-π interactions based on statistically expected interactions between π-orbital-containing residues [154]. π-π interactions can occur between residues containing aromatic rings (Tyr, Phe, Trp, and His) as well as amino acids with π-side chains (Gln, Asp, Glu, and Arg). PLAAC predicts disordered PLDs using a hidden Markov model (HMM) [155]. PLAAC was trained on the yeast proteome and extended to screen human proteins [156]. ZipperDB collects protein regions predicted to form fibrils using segments from more than 20,000 putative amyloid-forming sequences. It supports the analysis of individual prediction from user input [157]. catGRANULE is a more generic prediction algorithm and excels at predicting dosage-sensitive proteins. It predicts the tendency of proteins to localize in cytoplasmic foci by combining nucleic acid binding propensities, structural disorder, sequence length, and amino acid composition [158]. The PSPer approach predicts and screens proteins with similar characteristics to the FUS-like phase-separating regions, including PLDs, RNA binding motifs, and disordered arginine-rich regions [159]. FuzDrop is a method to predict droplet-promoting regions and proteins, which can spontaneously undergo LLPS, providing a sequence-based profile of propensities [160].

Most current predictors only perform well in proteins that can self-assemble to undergo PS (SaPS). PhaSePred is a machine-learning predictor that can predict the potential of proteins to assemble through partners, considered partner-dependent PS (PdPS) proteins [151] and SaPS proteins. By incorporating multi-modal features such as phosphorylation level and immunofluorescence (IF) images, PhaSePred can discriminate proteins located in human MLOs, such as the components recorded in the OpenCell nuclear puncta set, the DACT1-particulate proteome set, the G3BP1 proximity labelling set, and the PhaSepDB high-throughput set. It provides a web server that integrates other well-used PS predictors, such as PLAAC, Pscore, and catGRANULE.

The discovery and prediction of protein constituents in biological condensates present enormous opportunities for uncovering underlying mechanisms. The DeepPhase method provides a global view of the human PS proteome, according to the morphology of their condensates, instead of using sequence-dependent traits such as other predicting tools [161]. Based on the shape of targeted proteins in IF images from the Human Protein Atlas (HPA), DeepPhase scores the PS proteins in a cell-type specific manner, which can inform users with appropriate cell line types, as well as antibody choices for the proteins they are interested in. DeepPhase also suggests that a subset of kinases and TFs with higher predicting scores tend to display significantly reduced substrate specificity.

Informative reviews have comprehensively summarized the route performance and primary principle, which suggests a limited overlap between existing predictors that may generate a significant rate of false negatives [162–164]. The molecular mechanism underlying PS is complicated and involves diverse molecular components with distinct roles, *e.g.*, partner-dependent PS proteins

and regulators. Prediction efficiency depends on standard negative and positive phase separating protein datasets, which are thus far limited. Indeed, these predictors successfully predict numerous phase-separating proteins, confirming the importance of phase-separation-prone features and providing the foundation to establish more integrative and accurate predictors or tools for protein science in the future [26,165].

Conclusions and Perspectives

It is clear that PS is presumed to play ubiquitous roles in fundamental biological processes. The question is thus raised as to whether PS is a requisite in a given function. A detailed experimental pipeline for LLPS-associated studies has been provided by Gao and coworkers, addressing how we can relate the experimental phenomena to biological functions [166]. The morphological and topological characteristics of target proteins are dependent on image information. Therefore, the robustness is limited to optical resolution. In addition, PS can be driven and regulated by protein-protein, protein-ion, and protein-nucleic acid interactions, which suggests the essentiality of PS-related networks when judging the roles of target proteins in biological processes.

Increasing numbers of useful bioinformatics methods have been developed to predict PS proteins. However, most of these methods are designed for screening such proteins with higher IDR content [163], such as PScore for π-π enriched segments [154], PLAAC for PLDs [155], and FuzDrop for droplet-promoting regions and proteins [160]. Proteins with ordered structures or modular domains account for a large subset of PS proteins as well, such as HP1A [167], UTX [168], and DVL2 [82,83]. The computational tools to identify such proteins with lower IDR content are insufficient. Nevertheless, it may not be so challenging to identify such proteins. For example, the DeepPhase method can distinguish and predict some condensed proteins enriched in the WD40 motif [161], which is considered a feature of the PS protein Gro [169]. The PhaSePred method can also predict a set of proteins with structured domains, such as SH3 and PDZ domains [151]. Thus, further developments in computational methods should consider more comprehensive features in addition to sequence composition.

A range of diseases are associated with gene mutations, which can lead to the gain or loss of the ability to condense during the development of pathogenesis [6,8,147,170]. For example, the typical prostate cancer mutations in SPOP, W131G and F133V, disrupt SPOP and DAXX colocalization, resulting in decreased substrate ubiquitination [8]. However, under many other pathologic conditions, condensation of particular proteins underlies disease progression, such as KEAP1 (R320Q) in lung cancer [6] and PTPN (Y279C) in liver cancer [170]. It is promising to model the precise effects of sequence mutations on the phase behavior of proteins, especially those related to diseases. Therefore, future bioinformatics approaches should provide more insights into the effects of sequence alterations, such as missense mutations, aiding in providing a mechanistic understanding of disease mutations. Therapeutic approaches targeting such aberrant condensates might restore their physiological states, as in the case of myricetin inhibiting droplet formation of the tau protein [171], inspiring further interest in the therapeutic area.

Funding

This work was supported by the grants from the National Key Research and Development Program of China (Nos. 2021YFF1200900 and 2018YFA0507504), the National Natural Science Foundation of China (No. 32070666), and the High-Performance Computing Platform of Peking University.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Banani SF, Rice AM, Peeples WB, Lin Y, Jain S, Parker R, Rosen MK. Compositional control of phase-separated cellular bodies. *Cell* 2016, 166: 651–663
2. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, *et al.* Coexisting liquid phases underlie nucleolar subcompartments. *Cell* 2016, 165: 1686–1697
3. Hou C, Xie H, Fu Y, Ma Y, Li T. MloDisDB: a manually curated database of the relations between membraneless organelles and diseases. *Briefings BioInf* 2021, 22: bbaa271
4. Peskett TR, Rau F, O'Driscoll J, Patani R, Lowe AR, Saibil HR. A liquid to solid phase transition underlying pathological huntingtin exon1 aggregation. *Mol Cell* 2018, 70: 588–601.e6
5. Hirose T, Ninomiya K, Nakagawa S, Yamazaki T. A guide to membraneless organelles and their various roles in gene regulation. *Nat Rev Mol Cell Biol* 2022, 24: 288–304
6. Cloer EW, Siesser PF, Cousins EM, Goldfarb D, Mowrey DD, Harrison JS, Weir SJ, *et al.* p62-dependent phase separation of patient-derived KEAP1 mutations and NRF2. *Mol Cell Biol* 2018, 38: e00644–17
7. Du M, Chen ZJ. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* 2018, 361: 704–709
8. Bouchard JJ, Otero JH, Scott DC, Szulc E, Martin EW, Sabri N, Granata D, *et al.* Cancer Mutations of the Tumor Suppressor SPOP disrupt the formation of active, phase-separated compartments. *Mol Cell* 2018, 72: 19–36.e8
9. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 2013, 495: 467–473
10. Beutel O, Maraspini R, Pombo-García K, Martin-Lemaitre C, Honigmann A. Phase separation of zonula occludens proteins drives formation of tight junctions. *Cell* 2019, 179: 923–936.e11
11. Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, Gerlich DW, *et al.* Organization of chromatin by intrinsic and regulated phase separation. *Cell* 2019, 179: 470–484.e21
12. Zeng M, Shang Y, Araki Y, Guo T, Huganir RL, Zhang M. Phase transition in postsynaptic densities underlies formation of synaptic complexes and synaptic plasticity. *Cell* 2016, 166: 1163–1175.e12
13. Jiang Y, Fu X, Zhang Y, Wang SF, Zhu H, Wang WK, Zhang L, *et al.* Rett syndrome linked to defects in forming the MeCP2/Rbfox/LASR complex in mouse models. *Nat Commun* 2021, 12: 5767
14. Sampson J, Richards MW, Choi J, Fry AM, Bayliss R. Phase-separated foci of EML4-ALK facilitate signalling and depend upon an active kinase conformation. *EMBO Rep* 2021, 22: e53693
15. Terlecki-Zaniewicz S, Humer T, Eder T, Schmoellerl J, Heyes E, Manhart G, Kuchynka N, *et al.* Biomolecular condensation of NUP98 fusion proteins drives leukemogenic gene expression. *Nat Struct Mol Biol* 2021, 28: 190–201
16. Latysheva NS, Babu MM. Molecular signatures of fusion proteins in cancer. *ACS Pharmacol Transl Sci* 2019, 2: 122–133
17. Ramaswami M, Taylor JP, Parker R. Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell* 2013, 154: 727–736

18. Gopal PP, Nirschl JJ, Klinman E, Holzbaur ELF. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proc Natl Acad Sci USA* 2017, 114: E2466
19. Murakami T, Qamar S, Lin JQ, Schierle GSK, Rees E, Miyashita A, Costa AR, *et al.* ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* 2015, 88: 678–690
20. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 2017, 18: 285–298
21. Meier SM, Farcas AM, Kumar A, Ijavi M, Bill RT, Stelling J, Dufresne ER, *et al.* Multivalency ensures persistence of a +TIP body at specialized microtubule ends. *Nat Cell Biol* 2023, 25: 56–67
22. Choi JM, Holehouse AS, Pappu RV. Physical principles underlying the complex biology of intracellular phase transitions. *Annu Rev Biophys* 2020, 49: 107–133
23. Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 2012, 109: 16155–16160
24. Ranganathan S, Shakhnovich EI. Dynamic metastable long-living droplets formed by sticker-spacer proteins. *Elife* 2020, 9: e56159
25. Joseph JA, Reinhardt A, Aguirre A, Chew PY, Russell KO, Espinosa JR, Garaizar A, *et al.* Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nat Comput Sci* 2021, 1: 732–743
26. Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu RV. CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys J* 2017, 112: 16–21
27. Borcherds W, Bremer A, Borgia MB, Mittag T. How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? *Curr Opin Struct Biol* 2021, 67: 41–50
28. Qamar S, Wang GZ, Randle SJ, Ruggeri FS, Varela JA, Lin JQ, Phillips EC, *et al.* FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell* 2018, 173: 720–734.e15
29. Kar M, Posey AE, Dar F, Hyman AA, Pappu RV. Glycine-rich peptides from FUS have an intrinsic ability to self-assemble into fibers and networked fibrils. *Biochemistry* 2021, 60: 3213–3222
30. Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* 2018, 174: 688–699.e16
31. Guillén-Boixet J, Kopach A, Holehouse AS, Wittmann S, Jahnel M, Schlüßler R, Kim K, *et al.* RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. *Cell* 2020, 181: 346–361.e17
32. Niaki AG, Sarkar J, Cai X, Rhine K, Vidaurre V, Guy B, Hurst M, *et al.* Loss of dynamic RNA interaction and aberrant phase separation induced by two distinct types of ALS/FTD-linked FUS mutations. *Mol Cell* 2020, 77: 82–94.e4
33. Choi JM, Hyman AA, Pappu RV. Generalized models for bond percolation transitions of associative polymers. *Phys Rev E* 2020, 102: 042403
34. Newton JC, Naik MT, Li GY, Murphy EL, Fawzi NL, Sedivy JM, Jogl G. Phase separation of the LINE-1 ORF1 protein is mediated by the N-terminus and coiled-coil domain. *Biophys J* 2021, 120: 2181–2191
35. Zeng M, Chen X, Guan D, Xu J, Wu H, Tong P, Zhang M. Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell* 2018, 174: 1172–1187.e16
36. Kuravi S, Cheng J, Fangman G, Polireddy K, McCormick S, Lin TL, Singh AK, *et al.* Preclinical evaluation of gilteritinib on NPM1-ALK-driven anaplastic large cell lymphoma cells. *Mol Cancer Res* 2021, 19: 913–920
37. Basu S, Mackowiak SD, Niskanen H, Knezevic D, Asimi V, Grosswendt S, Geertsema H, *et al.* Unblending of transcriptional condensates in human repeat expansion disease. *Cell* 2020, 181: 1062–1079.e30
38. Yang P, Mathieu C, Kolaitis RM, Zhang P, Messing J, Yurtsever U, Yang Z, *et al.* G3BP1 is a tunable switch that triggers phase separation to assemble stress granules. *Cell* 2020, 181: 325–345.e28
39. Zhu S, Gu J, Yao J, Li Y, Zhang Z, Xia W, Wang Z, *et al.* Liquid-liquid phase separation of RBGD2/4 is required for heat stress resistance in Arabidopsis. *Dev Cell* 2022, 57: 583–597.e6
40. Smith JA, Curry EG, Blue RE, Roden C, Dundon SER, Rodríguez-Vargas A, Jordan DC, *et al.* FXR1 splicing is important for muscle development and biomolecular condensates in muscle cells. *J Cell Biol* 2020, 219: e201911129
41. Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, Spille JH, *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 2019, 572: 543–548
42. Dao TP, Martyniak B, Canning AJ, Lei Y, Colicino EG, Cosgrove MS, Hehnly H, *et al.* ALS-linked mutations affect UBQLN2 oligomerization and phase separation in a position- and amino acid-dependent manner. *Structure* 2019, 27: 937–951.e5
43. Riley JF, Fioramonti PJ, Rusnock AK, Hehnly H, Castañeda CA. ALS-linked mutations impair UBQLN2 stress-induced biomolecular condensate assembly in cells. *J Neurochem* 2021, 159: 145–155
44. Batlle C, Yang P, Coughlin M, Messing J, Pesarrodona M, Szulc E, Salvatella X, *et al.* hnRNPDL phase separation is regulated by alternative splicing and disease-causing mutations accelerate its aggregation. *Cell Rep* 2020, 30: 1117–1128.e5
45. Murthy AC, Tang WS, Jovic N, Janke AM, Seo DH, Perdikari TM, Mittal J, *et al.* Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. *Nat Struct Mol Biol* 2021, 28: 923–935
46. Huo X, Ji L, Zhang Y, Lv P, Cao X, Wang Q, Yan Z, *et al.* The nuclear matrix protein SAFB cooperates with major satellite RNAs to stabilize heterochromatin architecture partially through phase separation. *Mol Cell* 2020, 77: 368–383.e7
47. Kawachi T, Masuda A, Yamashita Y, Takeda J, Ohkawara B, Ito M, Ohno K. Regulated splicing of large exons is linked to phase-separation of vertebrate transcription factors. *EMBO J* 2021, 40: e107485
48. Li L, Herzog M, Möbitz S, Winter R. Liquid droplets of protein LAF1 provide a vehicle to regulate storage of the signaling protein K-Ras4B and its transport to the lipid membrane. *Phys Chem Chem Phys* 2021, 23: 5370–5375
49. Xue S, Gong R, He F, Li Y, Wang Y, Tan T, Luo SZ. Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Sci Adv* 2019, 5: eaax5349
50. Yao RW, Xu G, Wang Y, Shan L, Luan PF, Wang Y, Wu M, *et al.* Nascent pre-rRNA sorting via phase separation drives the assembly of dense fibrillar components in the human nucleolus. *Mol Cell* 2019, 76: 767–783.e11
51. Liu X, Shen J, Xie L, Wei Z, Wong C, Li Y, Zheng X, *et al.* Mitotic implantation of the transcription factor prospero via phase separation drives terminal neuronal differentiation. *Dev Cell* 2020, 52: 277–293.e8
52. Tillu VA, Rae J, Gao Y, Ariotti N, Floetenmeyer M, Kovtun O, McMahon KA, *et al.* Cavin1 intrinsically disordered domains are essential for fuzzy electrostatic interactions and caveola formation. *Nat Commun* 2021, 12: 931

53. Lu H, Yu D, Hansen AS, Ganguly S, Liu R, Heckert A, Darzacq X, *et al.* Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 2018, 558: 318–323
54. Dorone Y, Boeynaems S, Flores E, Jin B, Hateley S, Bossi F, Lazarus E, *et al.* A prion-like protein regulator of seed germination undergoes hydration-dependent phase separation. *Cell* 2021, 184: 4284–4298.e27
55. Jiang LL, Che MX, Zhao J, Zhou CJ, Xie MY, Li HY, He JH, *et al.* Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J Biol Chem* 2013, 288: 19614–19624
56. Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* 2016, 24: 1537–1549
57. Huang WYC, Alvarez S, Kondo Y, Lee YK, Chung JK, Lam HYM, Biswas KH, *et al.* A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. *Science* 2019, 363: 1098–1103
58. Case LB, Zhang X, Ditlev JA, Rosen MK. Stoichiometry controls activity of phase-separated clusters of actin signaling proteins. *Science* 2019, 363: 1093–1097
59. Gallego-Iradi MC, Strunk H, Crown AM, Davila R, Brown H, Rodriguez-Lebron E, Borchelt DR. N-terminal sequences in matrin 3 mediate phase separation into droplet-like structures that recruit TDP43 variants lacking RNA binding elements. *Lab Invest* 2019, 99: 1030–1040
60. Jack A, Kim Y, Strom AR, Lee DSW, Williams B, Schaub JM, Kellogg EH, *et al.* Compartmentalization of telomeres through DNA-scaffolded phase separation. *Dev Cell* 2022, 57: 277–290.e9
61. Fromm SA, Kamenz J, Nöldeke ER, Neu A, Zocher G, Sprangers R. In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery. *Angew Chem Int Ed* 2014, 53: 7354–7359
62. Monette A, Niu M, Chen L, Rao S, Gorelick RJ, Mouland AJ. Pan-retroviral nucleocapsid-mediated phase separation regulates genomic RNA positioning and trafficking. *Cell Rep* 2020, 31: 107520
63. Uriarte M, Sen Nkwe N, Tremblay R, Ahmed O, Messmer C, Mashtalir N, Barbour H, *et al.* Starvation-induced proteasome assemblies in the nucleus link amino acid supply to apoptosis. *Nat Commun* 2021, 12: 6984
64. Shan Z, Tu Y, Yang Y, Liu Z, Zeng M, Xu H, Long J, *et al.* Basal condensation of Numb and Pon complex via phase transition during Drosophila neuroblast asymmetric division. *Nat Commun* 2018, 9: 737
65. Sheu-Gruttadauria J, MacRae IJ. Phase transitions in the assembly and function of human miRISC. *Cell* 2018, 173: 946–957.e16
66. Wei MT, Elbaum-Garfinkle S, Holehouse AS, Chen CCH, Feric M, Arnold CB, Priestley RD, *et al.* Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat Chem* 2017, 9: 1118–1125
67. Ng SC, Güttler T, Görlich D. Recapitulation of selective nuclear import and export with a perfectly repeated 12mer GLFG peptide. *Nat Commun* 2021, 12: 4047
68. Zhang G, Wang Z, Du Z, Zhang H. mTOR regulates phase separation of PGL granules to modulate their autophagic degradation. *Cell* 2018, 174: 1492–1506.e22
69. Zhang Y, Li Z, Chen N, Huang Y, Huang S. Phase separation of Arabidopsis EMB1579 controls transcription, mRNA splicing, and development. *PLoS Biol* 2020, 18: e3000782
70. Boeynaems S, Bogaert E, Kovacs D, Konijnenberg A, Timmerman E, Volkov A, Guharoy M, *et al.* Phase Separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol Cell* 2017, 65: 1044–1055.e5
71. Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 2013, 155: 1049–1060
72. Gai X, Xin D, Wu D, Wang X, Chen L, Wang Y, Ma K, *et al.* Pre-ribosomal RNA reorganizes DNA damage repair factors in nucleus during meiotic prophase and DNA damage response. *Cell Res* 2022, 32: 254–268
73. Singh V, Xu L, Boyko S, Surewicz K, Surewicz WK. Zinc promotes liquid–liquid phase separation of tau protein. *J Biol Chem* 2020, 295: 5850–5856
74. Tulpule A, Guan J, Neel DS, Allegakoen HR, Lin YP, Brown D, Chou YT, *et al.* Kinase-mediated RAS signaling via membraneless cytoplasmic protein granules. *Cell* 2021, 184: 2649–2664.e18
75. Ratovitski T, Jiang M, O'Meally RN, Rauniyar P, Chighladze E, Faragó A, Kamath SV, *et al.* Interaction of huntingtin with PRMTs and its subsequent arginine methylation affects HTT solubility, phase transition behavior and neuronal toxicity. *Hum Mol Genet* 2022, 31: 1651–1672
76. Posey AE, Ruff KM, Harmon TS, Crick SL, Li A, Diamond MI, Pappu RV. Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers. *J Biol Chem* 2018, 293: 3734–3746
77. Kamps J, Lin YH, Oliva R, Bader V, Winter R, Winklhofer KF, Tatzelt J. The N-terminal domain of the prion protein is required and sufficient for liquid–liquid phase separation: a crucial role of the Aβ-binding domain. *J Biol Chem* 2021, 297: 100860
78. Belott C, Janis B, Menze MA. Liquid-liquid phase separation promotes animal desiccation tolerance. *Proc Natl Acad Sci USA* 2020, 117: 27676–27684
79. Chen D, Wang Z, Zhao YG, Zheng H, Zhao H, Liu N, Zhang H. Inositol polyphosphate multikinase inhibits liquid-liquid phase separation of TFEB to negatively regulate autophagy activity. *Dev Cell* 2020, 55: 588–602.e7
80. Omnus DJ, Cadou A, Thomas FB, Bader JM, Soh N, Chung GHC, Vaughan AN, *et al.* A heat-sensitive Osh protein controls PI4P polarity. *BMC Biol* 2020, 18: 28
81. Ghodke I, Remisova M, Furst A, Kilic S, Reina-San-Martin B, Poetsch AR, Altmeyer M, *et al.* AHNAK controls 53BP1-mediated p53 response by restraining 53BP1 oligomerization and phase separation. *Mol Cell* 2021, 81: 2596–2610.e7
82. Schwarz-Romond T, Merrifield C, Nichols BJ, Bienz M. The Wnt signalling effector Dishevelled forms dynamic protein assemblies rather than stable associations with cytoplasmic vesicles. *J Cell Sci* 2005, 118: 5269–5277
83. Schwarz-Romond T, Fiedler M, Shibata N, Butler PJG, Kikuchi A, Higuchi Y, Bienz M. The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. *Nat Struct Mol Biol* 2007, 14: 484–492
84. Yang Y, Ye X, Dai R, Li Z, Zhang Y, Xue W, Zhu Y, *et al.* Phase separation of Epstein-Barr virus EBNA2 protein reorganizes chromatin topology for epigenetic regulation. *Commun Biol* 2021, 4: 967
85. Huang X, Chen S, Li W, Tang L, Zhang Y, Yang N, Zou Y, *et al.* ROS regulated reversible protein phase separation synchronizes plant flowering. *Nat Chem Biol* 2021, 17: 549–557
86. Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, Llaguno M, *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* 2012, 483: 336–340
87. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* 2020, 367: 694–699
88. Seif E, Kang JJ, Sasseville C, Senkovich O, Kaltashov A, Boulier EL, Kapur I, *et al.* Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their

activity. *Nat Commun* 2020, 11: 5609
89. Bremer A, Farag M, Borcherds WM, Peran I, Martin EW, Pappu RV, Mittag T. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat Chem* 2022, 14: 196–207
90. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, ***et al.*** Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* 2015, 57: 936–947
91. Das S, Eisen A, Lin YH, Chan HS. A lattice model of charge-pattern-dependent polyampholyte phase separation. *J Phys Chem B* 2018, 122: 5418–5431
92. Lyons H, Veettil RT, Pradhan P, Fornero C, De La Cruz N, Ito K, Eppert M, ***et al.*** Functional partitioning of transcriptional regulators by patterned charge blocks. *Cell* 2023, 186: 327–345.e28
93. Harmon TS, Holehouse AS, Rosen MK, Pappu RV. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife* 2017, 6: e30294
94. Rauscher S, Baud S, Miao M, Keeley FW, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure* 2006, 14: 1667–1676
95. Chandra B, Michmerhuizen NL, Shirnekhi HK, Tripathi S, Pioso BJ, Baggett DW, Mitrea DM, ***et al.*** Phase separation mediates NUP98 fusion oncoprotein leukemic transformation. *Cancer Discov* 2022, 12: 1152–1169
96. Vacic V, Markwick PR, Oldfield CJ, Zhao X, Haynes C, Uversky VN, Iakoucheva LM. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput Biol* 2012, 8: e1002709
97. Hou C, Wang X, Xie H, Chen T, Zhu P, Xu X, You K, ***et al.*** PhaSepDB in 2022: annotating phase separation-related proteins with droplet states, co-phase separation partners and other experimental information. *Nucleic Acids Res* 2023, 51: D460–D465
98. Xiang S, Kato M, Wu LC, Lin Y, Ding M, Zhang Y, Yu Y, ***et al.*** The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* 2015, 163: 829–839
99. Stortz M, Pecci A, Presman DM, Levi V. Unraveling the molecular interactions involved in phase separation of glucocorticoid receptor. *BMC Biol* 2020, 18: 59
100. Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, Amaya J, ***et al.*** Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol Cell* 2018, 69: 465–479.e7
101. Albrecht A, Mundlos S. The other trinucleotide repeat: polyalanine expansion disorders. *Curr Opin Genet Dev* 2005, 15: 285–293
102. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci* 2007, 30: 575–621
103. Darling A, Uversky V. Intrinsic disorder in proteins with pathogenic repeat expansions. *Molecules* 2017, 22: 2027
104. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, ***et al.*** A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011, 72: 257–268
105. Lee KH, Zhang P, Kim HJ, Mitrea DM, Sarkar M, Freibaum BD, Cika J, ***et al.*** C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell* 2016, 167: 774–788.e17
106. Schulte J, Littleton JT. The biological function of the Huntingtin protein and its relevance to Huntington's disease pathology. *Curr Trends Neurol* 2011, 5: 65–78
107. Kim S, Nollen EAA, Kitagawa K, Bindokas VP, Morimoto RI. Polyglutamine protein aggregates are dynamic. *Nat Cell Biol* 2002, 4: 826–831
108. Aktar F, Burudpakdee C, Polanco M, Pei S, Swayne TC, Lipke PN, Emtage L. The huntingtin inclusion is a dynamic phase-separated compartment. *Life Sci Alliance* 2019, 2: e201900489
109. Li L, Liu H, Dong P, Li D, Legant WR, Grimm JB, Lavis LD, ***et al.*** Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. *Elife* 2016, 5: e17056
110. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007, 7: 233–245
111. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* 2015, 15: 371–381
112. Wang Y, Chunyu Yu C, Pei G, Jia W, Li T, Li P. Dissolution of oncofusion transcription factor condensates for cancer therapy. *Nat Chem Biol* 2023, Jul 3. doi: 10.1038/s41589-023-01376-5.
113. Owen I, Yee D, Wyne H, Perdikari TM, Johnson V, Smyth J, Kortum R, ***et al.*** The oncogenic transcription factor FUS-CHOP can undergo nuclear liquid-liquid phase separation. *J Cell Sci* 2021, 134: jcs258578
114. Zuo L, Zhang G, Massett M, Cheng J, Guo Z, Wang L, Gao Y, ***et al.*** Loci-specific phase separation of FET fusion oncoproteins promotes gene transcription. *Nat Commun* 2021, 12: 1491
115. Davis RB, Kaur T, Moosa MM, Banerjee PR. FUS oncofusion protein condensates recruit mSWI/SNF chromatin remodeler via heterotypic interactions between prion-like domains. *Protein Sci* 2021, 30: 1454–1466
116. Lindén M, Vannas C, Österlund T, Andersson L, Osman A, Escobar M, Fagman H, ***et al.*** FET fusion oncoproteins interact with BRD4 and SWI/SNF chromatin remodelling complex subtypes in sarcoma. *Mol Oncol* 2022, 16: 2470–2495
117. Ahn JH, Davis ES, Daugird TA, Zhao S, Quiroga IY, Uryu H, Li J, ***et al.*** Phase separation drives aberrant chromatin looping and cancer development. *Nature* 2021, 595: 591–595
118. Lin CC, Suen KM, Jeffrey PA, Wieteska L, Lidster JA, Bao P, Curd AP, ***et al.*** Receptor tyrosine kinases regulate signal transduction through a liquid-liquid phase separated state. *Mol Cell* 2022, 82: 1089–1106.e12
119. Qin Z, Sun H, Yue M, Pan X, Chen L, Feng X, Yan X, ***et al.*** Phase separation of EML4–ALK in firing downstream signaling and promoting lung tumorigenesis. *Cell Discov* 2021, 7: 33
120. Fawal M, Espinos E, Jean-Jean O, Morello D. Looking for the functions of RNA granules in ALK-transformed cells. *Bioarchitecture* 2011, 1: 91–95
121. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, ***et al.*** Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008, 456: 470–476
122. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008, 40: 1413–1415
123. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, ***et al.*** The evolutionary landscape of alternative splicing in vertebrate species. *Science* 2012, 338: 1587–1593
124. de Klerk E, 't Hoen PAC. Alternative mRNA transcription, processing, and translation: insig hts from RNA sequencing. *Trends Genet* 2015, 31: 128–139
125. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim TH, Calarco JA, Wang X, ***et al.*** Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell* 2012, 46: 884–892
126. Davis MJ, Shin CJ, Jing N, Ragan MA. Rewiring the dynamic interactome. *Mol BioSyst* 2012, 8: 2054
127. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, Babu MM. Tissue-specific splicing of disordered segments

that embed binding motifs rewires protein interaction networks. *Mol Cell* 2012, 46: 871–883
128. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D^2 concept. *Annu Rev Biophys* 2008, 37: 215–246
129. Conrad RJ, Fozouni P, Thomas S, Sy H, Zhang Q, Zhou MM, Ott M. The short isoform of BRD4 promotes HIV-1 latency by engaging repressive SWI/SNF chromatin-remodeling complexes. *Mol Cell* 2017, 67: 1001–1012.e6
130. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem Soc Trans* 2016, 44: 1185–1200
131. Weatheritt RJ, Luck K, Petsalaki E, Davey NE, Gibson TJ. The identification of short linear motif-mediated interfaces within the human interactome. *Bioinformatics* 2012, 28: 976–982
132. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, *et al.* Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci USA* 2006, 103: 8390–8395
133. Pentony MM, Jones DT. Modularity of intrinsic disorder in the human proteome. *Proteins* 2010, 78: 212–221
134. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays* 2009, 31: 328–335
135. Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res* 2006, 34: 305–312
136. Hofweber M, Dormann D. Friend or foe—post-translational modifications as regulators of phase separation and RNP granule dynamics. *J Biol Chem* 2019, 294: 7137–7150
137. Kim TH, Tsang B, Vernon RM, Sonenberg N, Kay LE, Forman-Kay JD. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* 2019, 365: 825–829
138. Gueroussov S, Weatheritt RJ, O'Hanlon D, Lin ZY, Narula A, Gingras AC, Blencowe BJ. Regulatory expansion in mammals of multivalent hnRNP assemblies that globally control alternative splicing. *Cell* 2017, 170: 324–339.e23
139. Taylor JP, Brown Jr RH, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature* 2016, 539: 197–206
140. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 2012, 149: 753–767
141. Weber SC, Brangwynne CP. Getting RNA and protein in phase. *Cell* 2012, 149: 1188–1191
142. Feng H, Bao S, Rahman MA, Weyn-Vanhentenryck SM, Khan A, Wong J, Shah A, *et al.* Modeling RNA-binding protein specificity *in vivo* by precisely registering protein-RNA crosslink sites. *Mol Cell* 2019, 74: 1189–1204.e6
143. Deshaies JE, Shkreta L, Moszczynski AJ, Sidibé H, Semmler S, Fouillen A, Bennett ER, *et al.* TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain* 2018, 141: 1320–1333
144. Alberti S, Hyman AA. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat Rev Mol Cell Biol* 2021, 22: 196–213
145. Alberti S, Dormann D. Liquid-liquid phase separation in disease. *Annu Rev Genet* 2019, 53: 171–194
146. Banani SF, Afeyan LK, Hawken SW, Henninger JE, Dall'Agnese A, Clark VE, Platt JM, *et al.* Genetic variation associated with condensate dysregulation in disease. *Dev Cell* 2022, 57: 1776–1788.e8
147. Vendruscolo M, Fuxreiter M. Protein condensation diseases: therapeutic opportunities. *Nat Commun* 2022, 13: 5550
148. Ning W, Guo Y, Lin S, Mei B, Wu Y, Jiang P, Tan X, *et al.* DrLLPS: a data resource of liquid–liquid phase separation in eukaryotes. *Nucleic Acids Res* 2020, 48: D288–D295
149. Li Q, Peng X, Li Y, Tang W, Zhu J, Huang J, Qi Y, *et al.* LLPSDB: a database of proteins undergoing liquid–liquid phase separation in vitro. *Nucleic Acids Res* 2020, 48: D320–D327
150. Mészáros B, Erdős G, Szabó B, Schád É, Tantos Á, Abukhairan R, Horváth T, *et al.* PhaSePro: the database of proteins driving liquid–liquid phase separation. *Nucleic Acids Res* 2020, 48: D360–D367
151. Chen Z, Hou C, Wang L, Yu C, Chen T, Shen B, Hou Y, *et al.* Screening membraneless organelle participants with machine-learning models that integrate multimodal features. *Proc Natl Acad Sci USA* 2022, 119: e2115369119
152. Zhu H, Fu H, Cui T, Ning L, Shao H, Guo Y, Ke Y, *et al.* RNAPhaSep: a resource of RNAs undergoing phase separation. *Nucleic Acids Res* 2022, 50: D340–D346
153. Liu M, Li H, Luo X, Cai J, Chen T, Xie Y, Ren J, *et al.* RPS: a comprehensive database of RNAs involved in liquid–liquid phase separation. *Nucleic Acids Res* 2022, 50: D347–D355
154. Vernon RMC, Chong PA, Tsang B, Kim TH, Bah A, Farber P, Lin H, *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* 2018, 7: e31486
155. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* 2014, 30: 2501–2502
156. Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. *Biochem J* 2017, 474: 1417–1438
157. Castillo V, Graña-Montes R, Sabate R, Ventura S. Prediction of the aggregation propensity of proteins from the primary sequence: aggregation properties of proteomes. *Biotechnol J* 2011, 6: 674–685
158. Bolognesi B, Lorenzo Gotor N, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, Lehner B. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep* 2016, 16: 222–231
159. Orlando G, Raimondi D, Tabaro F, Codice F, Moreau Y, Vranken WF. Computational identification of prion-like RNA-binding proteins that form liquid phase-separated condensates. *Bioinformatics* 2019, 35: 4617–4623
160. Hardenberg M, Horvath A, Ambrus V, Fuxreiter M, Vendruscolo M. Widespread occurrence of the droplet state of proteins in the human proteome. *Proc Natl Acad Sci USA* 2020, 117: 33254–33262
161. Yu C, Shen B, You K, Huang Q, Shi M, Wu C, Chen Y, *et al.* Proteome-scale analysis of phase-separated proteins in immunofluorescence images. *Brief Bioinform* 2021, 22: bbaa187
162. Chong S, Mir M. Towards decoding the sequence-based grammar governing the functions of intrinsically disordered protein regions. *J Mol Biol* 2021, 433: 166724
163. Shen B, Chen Z, Yu C, Chen T, Shi M, Li T. Computational screening of phase-separating proteins. *Genomics Proteomics Bioinformatics* 2021, 19: 13–24
164. Vernon RM, Forman-Kay JD. First-generation predictors of biological protein phase separation. *Curr Opin Struct Biol* 2019, 58: 88–96
165. Cai H, Vernon RM, Forman-Kay JD. An interpretable machine-learning algorithm to predict disordered protein phase separation based on biophysical interactions. *Biomolecules* 2022, 12: 1131

166. Gao Y, Li X, Li P, Lin Y. A brief guideline for studies of phase-separated biomolecular condensates. *Nat Chem Biol* 2022, 18: 1307–1318
167. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* 2017, 547: 236–240
168. Shi B, Li W, Song Y, Wang Z, Ju R, Ulman A, Hu J, *et al.* UTX condensation underlies its tumour-suppressive activity. *Nature* 2021, 597: 726–731
169. Treen N, Shimobayashi SF, Eeftens J, Brangwynne CP, Levine M. Properties of repression condensates in living Ciona embryos. *Nat Commun* 2021, 12: 1561
170. Zhu G, Xie J, Kong W, Xie J, Li Y, Du L, Zheng Q, *et al.* Phase separation of disease-associated SHP2 mutants underlies MAPK hyperactivation. *Cell* 2020, 183: 490–502.e18
171. Dai B, Zhong T, Chen ZX, Chen W, Zhang N, Liu XL, Wang LQ, *et al.* Myricetin slows liquid–liquid phase separation of Tau and activates ATG5-dependent autophagy to suppress Tau toxicity. *J Biol Chem* 2021, 297: 101222