

## REVIEW

# High-throughput experimental methods for investigating biomolecular condensates

Taoyu Chen<sup>1</sup>, Qi Lei<sup>1</sup>, Minglei Shi<sup>2</sup>, Tingting Li<sup>1,\*</sup>

<sup>1</sup> Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

<sup>2</sup> MOE Key Laboratory of Bioinformatics; Bioinformatics Division and Center for Synthetic & Systems Biology, TNLIST; School of Medicine, Tsinghua University, Beijing 100084, China

\* Correspondence: [litt@bjmu.edu.cn](mailto:litt@bjmu.edu.cn)

Received January 8, 2021; Revised March 4, 2021; Accepted March 12, 2021

**Background:** The concept of biomolecular condensate was put forward recently to emphasize the ability of certain cellular compartments to concentrate molecules and comprise proteins and nucleic acids with specific biological functions, from ribosome genesis to RNA splicing. Due to their unique role in biological processes, it is crucial to investigate their compositions, which is a primary determinant of condensate properties.

**Results:** Since a wide range of macromolecules comprise biomolecular condensates, it is necessary for researchers to investigate them using high-throughput methodologies while low-throughput experiments are not efficient enough. These high-throughput methods usually purify interacting protein libraries from condensates before being scanned in mass spectrometry. It is possible to extract organelles as a whole for specific condensates for further analysis, however, most condensates do not have a distinguishable marker or are sensitive to shear force to be extracted as a whole. Affinity tagging allows a comprehensive view of interacting proteins of target molecule yet only proteins with strong bonds may be pulled down. Proximity labeling serves as a complementary method to label more dynamic proteins with weaker interactions, increasing sensitivity while decreasing specificity. Image-based fluorescent screening takes another path by scanning images automatically to illustrate the condensing state of biomolecules within membraneless organelles, which is a unique feature unlike the previous mass spectrometry-based methods.

**Conclusion:** This review presents a rough glimpse into high-throughput methodologies for biomolecular condensate investigation to encourage usage of bioinformatic tools by researchers in relevant fields.

**Keywords:** biomolecular condensates; high-throughput; phase separation; interaction

**Author summary:** Biomacromolecules often form condensates in cytoplasm or nucleoplasm to keep biomolecular reactions organized and strictly regulated. The past 10 years have witnessed a surge in researches concerning biomolecular condensates and phase-separating molecules, yet most experiments are performed in low-throughput manners. This review summarizes the common high-throughput experimental methods for biomolecular condensate investigation, including organelle purification, affinity purification, proximity labeling, and image-based fluorescence screening and discusses their advantages and shortcomings. We demonstrate that high-throughput screening may provide new insights into related studies and encourage researchers to develop new tools to unravel phase-separated biomolecular condensates.

## INTRODUCTION

Living cells normally pack tens of thousands of biomolecules within its membrane, most of which

participate in a wide range of biochemical reactions that help sustain the cell's overall metabolism and perform specific physiological or pathophysiological functions. Due to the large number of biomolecules one single cell

contains, each living cell has to deal with the problem that the cellular biomolecular reactions need to be highly organized and strictly regulated. Cells adopt various mechanisms to keep the reactions under control, one of which is to condense related proteins and nucleic acids into biomolecular condensates so that different sets of biochemical reactions can be carried out in distinct spatiotemporal environment [1]. Banani *et al.* proposed the concept of “biomolecular condensate” to emphasize the structure within cells that functions to concentrate and comprise biological molecules.

Classical organelles are usually surrounded by lipid bilayer membranes, such as the Golgi apparatus. Some organelles like mitochondria even warp its interior matrix in double lipid bilayer membranes [2]. Lipid bilayer membranes can effectively inhibit molecules from diffusing into the organelle, maintaining microenvironments for specific biochemical reactions. Nevertheless, membrane-bound organelles are not dynamic enough to participate in regulatory events that require temporary condensation of multiple molecules. One example is transcription of active genes [3]. The hyperphosphorylated C-terminal domain (CTD) of RNA polymerase II (Pol II) needs to interact with mediator complex at super-enhancers along chromatin to initiate transcription. As the transcribed RNA elongates, splicing factors like SRSF2, SF3B1 etc. would gather around nascent RNA to splice it simultaneously, which illustrates that a wide range of macromolecules are condensing around super-enhancers and active genes. This is where membraneless organelles shine as molecules have more freedom to diffuse while the boundaries are dynamic. The aforementioned splicing factors are also components of biomolecular condensates like nuclear speckles [4] that might serve in RNA splicing. Nucleolus is also one example of membraneless organelles as it generates ribosome and constantly exports 40S and 60S rRNA assembles [5,6]. These examples indicate the significance membraneless biomolecular condensates play in various physiological processes.

When biomolecular condensates were first defined, a few researchers assume that some of these non-membrane-bound compartments might arise through liquid-liquid phase separation (LLPS), which is a physical chemical phenomenon that occurs when liquid phases demix and form multiple layers or droplets with clearly visible boundaries [7,8]. Physically, LLPS results from over-saturation of specific molecules that have to demix to concentrate excessive molecules. In biological context, proteins might contain multiple folded domains or intrinsically disordered regions that helps concentrating a number of proteins and nucleic acids, leading to high concentration of specific biomacromolecules that separate from matrix, which are

later named biomolecular condensates. Yet there have been controversies since many compartments do not behave like a demixed droplet [9]. Though the mechanism underneath is still being investigated, it's clear that biomolecular condensates can enrich a wide range of biomolecules, hence allowing condensation of specific molecules inside the compartment. However, since many biomolecular condensates are non-membrane-bound, molecules have more freedom to diffuse inside condensates and across the boundaries [10]. This dynamic nature brings challenges to researchers, leading to the application of multiple experimental methods to thoroughly illustrate the properties of biomolecular condensates, like fluorescence recovery after photobleaching (FRAP) [11]. Nevertheless, these low-throughput practices restrain themselves in a limited range of proteins and nucleic acids due to experimental issues. In recent years, high-throughput methodologies and bioinformatic tools rose to fame for their efficiency and convenience in dealing with various biological issues [12]. These methods allow screening of multiple biomolecules, some having weak interactions with or only stays for a short period of time in condensates, like nascent pre-snRNA being spliced in Cajal bodies [13]. In this review, we present some of the commonly-used strategies as well as newly-developed tools when investigating phase-separating biomolecules in a high-throughput fashion.

## OVERVIEW OF COMMON STRATEGIES

In order to investigate the composition and potential functions of biomolecular condensates, high-throughput strategies seek to extract all biomacromolecules within the boundary of condensates. Researchers can purify the protein library of condensates from the extract of specific tissues under certain conditions to have a glimpse of the proteome as well as the spatiotemporal locations of proteins and their changes overtime by analyzing the library with mass spectrometry (MS). MS is a central technique that measures mass-to-charge ratio ( $m/z$ ) of ionized fragments derived from the original protein library. After decades of improvement of MS instruments, MS has become a common practice for determining compositions of proteomes with many reviewing on these issues [14,15]. While MS has become a terminal procedure of proteome screening and analysis, strategies obtaining the protein library vary.

For phase-separating biomolecular condensates, common high-throughput strategies investigating condensates consists of the following (Fig. 1): (1) Organelle purification. Some condensates can get fragmented or dissected as a whole. By purifying the condensate as a whole, researchers are able to screen the

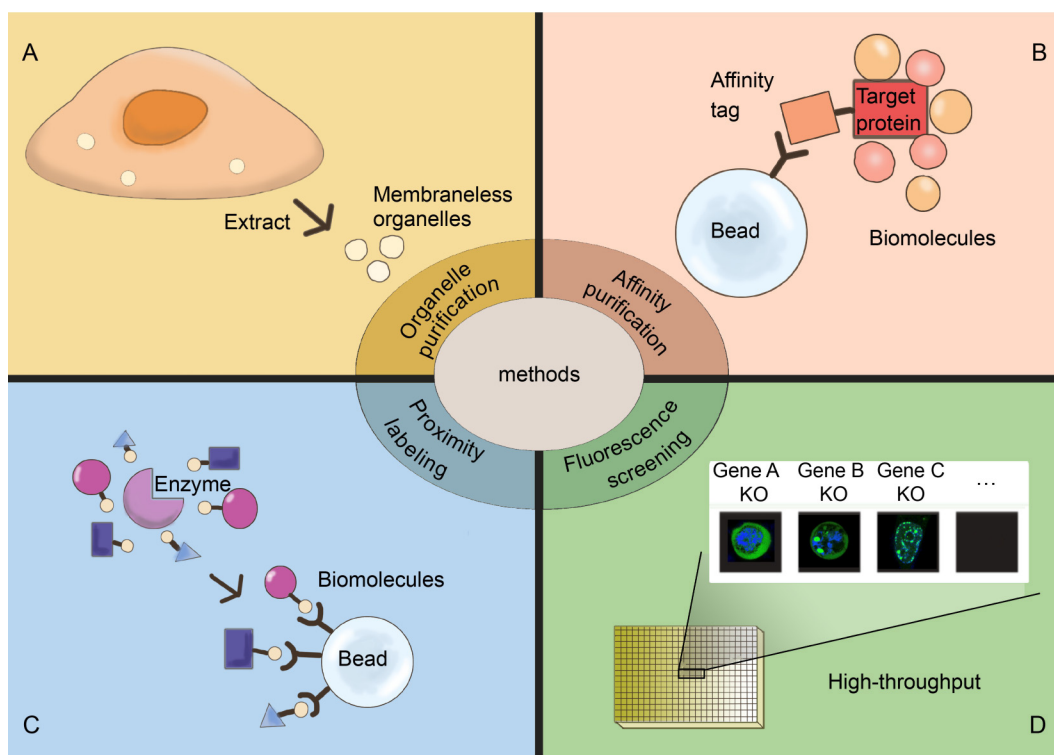
proteins inside the hub. (2) Affinity purification. This method allows precipitation of interacting proteins along with the target. Researchers attach specific tags to the target, which is usually the scaffold of condensates, so as to have a thorough picture of the condensate composition. (3) Proximity labeling. This method complements the previous method since it can crosslink proteins with weak or instant interactions. It can collect all proximal proteins, which expands the library of the condensates. (4) Fluorescence screening. This is a unique method in biomolecular condensate investigation. Researchers ought to screen a large number of fluorescent images with the assistance of computational methods to understand whether specific genes are important to the formation of condensates. This classification is also adapted when we develop PhaSepDB, a database dedicated to phase-separated proteins in biomolecular condensates [16]. Researchers can follow such methods when they wish to explore the composition of biomolecular condensates and discover potential phase-separating molecules.

While these high-throughput experimental methods allow us to better understand biomolecular condensates

in biological term, as bioinformaticians collect proteome data from these experiments, many establish corresponding databases to contain these high-throughput data systematically. These data resources include PhaSepDB [16] and PhaSePro [17], which focuses on proteins separated in individual condensates, as well as LLPSDB collecting *in vitro* phase separation data [18] and DrLLPS featuring *in silico* data [19]. Algorithms like catGRANULE [20] and PScore [21] also help to predict whether a novel protein might phase separate in biological context. However, it should be noted that these computational tools are still far from determining compositions of biomolecular condensates. Our previous review summarized these data resources and computational tools concerning phase-separating proteins and condensates for our readers to refer to [22]. In this review, we would focus on high-throughput experimental methodologies.

## ORGANELLE PURIFICATION

Biomolecular condensates are usually phase-separated from cytoplasm or nucleoplasm and many researchers



**Figure 1. An overview of common high-throughput strategies to investigate biomolecular condensates.** (A) Organelle purification. Some organelles can be extracted through carefully designed protocols. (B) Affinity purification. Affinity tags are attached to target proteins and identified by specific antibodies on beads. (C) Proximity labeling. Specially crafted enzymes linked to target molecule can tag its surrounding molecules, which can be recognized by particular beads. (D) Fluorescence screening. Researchers manage to produce a large set of experimental data (e.g., knocking out different genes of a particular condensate) and process the data with automated fluorescence microscopy. This figure is a rough demonstration of each methodology. Detailed descriptions have been illustrated in length in the review.

would refer to them as “membraneless organelles”. Methods to purify organelles with membrane including manual separation or salting out may also apply to membraneless organelles. Researchers often purify organelles for low-throughput experiments, but high-throughput methods can also be applied after purification. When organelles are purified, mass spectrometry can help to detect protein compositions while sequencing assists in discovering possible RNAs. This is the most basic high-throughput method to study condensate composition.

Nucleolus is among the first condensates to be purified and possibly the most-studied one as well. Anderson *et al.* dug into nucleolar proteome dynamics back in 2005 and extracted nucleolus by density gradient fractionation [23]. Nucleolar proteins were lysed by trypsin or endoproteinase Lys-C and went through mass spectrometry. Nucleolus extraction has since become a common practice in biological experiments. There is also a research that purified Balbiani body by manually dissecting *Xenopus laevis* stage I oocytes with shear force [24]. For bodies that are small and dynamic like P-body, Hubstenberger *et al.* responded with Fluorescence Activated Particle Sorting (FAPS) [25]: they attach green fluorescence protein (GFP) to LSM14A, a canonical P-body marker. After excluding nuclei and supernatant of cytoplasm with differential centrifugation, the remaining particles went through FAPS to collect sorted P-bodies for RNA-seq and proteomic profiling. Upon obtaining proteome from certain condensates, it is possible for researchers to perform enrichment analysis to understand major molecular functions of these proteins thus inferring biological roles of specific condensates. However, a wide range of biomolecular condensates do not have a distinguishable marker and are sensitive to shear force, which limits the use of organelle purification. Researchers developed other methods to investigate more condensates, for instance, affinity purification-mass spectrometry (AP/MS).

## AFFINITY PURIFICATION

As the name suggests, affinity purification requires an

affinity tag to detect interactome of particular biomolecules. Affinity tags were firstly used 30 years ago for protein expression and purification. The very first affinity tags are protein A and LacZ tags. These tags were meant to purify interacting proteins. In order to do so, researchers fuse affinity tags to a target protein and then capture the affinity tags by chromatography or designed affinity beads. After washing the beads multiple times, the tags will be eluted with elution buffer of specific pH, pulling down the target proteins and their interacting biomolecules along with them. After cleaving the tagged protein with protease, researchers are able to scan all the interacting proteins with mass spectrometry to infer possible roles that target proteins play or construct a set of mechanisms that target proteins participate in [26,27]. Some common tags used by many researchers are listed in Table 1.

Since the classic tags became widely used in protein purification, many improved tags appear for better efficiency and efficacy. For example, poly-histidine or His-tag utilizes 6 consecutive histones as an affinity tag to reduce the molecular weight as low as possible and keep the structure of the tag simple, which poses little influence on the target protein and are easily pulled down by nickel beads. However, in high-throughput biomolecular condensate studies, researchers use tandem affinity purification as a common affinity purification method.

**Tandem affinity purification.** Tandem affinity purification (TAP) tag is a frequently used tag in proteomics studies [31]. TAP-tag fuses 2 affinity tags into a dual-affinity tag to allow protein isolation and reduce non-specific background. The most widely used TAP-tag consists of protein A and calmodulin-binding peptide (CBP), separated by tobacco etch virus (TEV) protease cleavage site. Proteins are enriched by IgG Sepharose and cleaved by TEV protease. This allows researchers to capture the interacting biomolecules of around 20%–30% of proteins without applying extreme eluting conditions. TAP-tag is a neat choice for detecting biomolecular condensate composition. For instance, Ma *et al.* utilizes TAP-tags in determining the

**Table 1** Overview of common protein affinity tags

Affinity tags	Chemical composition	Size	Ref.
Protein A	staphylococcal protein A	42 kDa	[27]
LacZ	$\beta$ -galactosidase	130 kDa	[27]
His-tag	2–10 consecutive histones, usually 6	310–1550 Da	[28]
BCCP	Biotin-carboxy carrier protein	120 kDa	[29]
SBP	Straptavidin-biotin Peptide	4.5 kDa	[30]
TAP-tag	Protein A and calmodulin-binding peptide (CBP) separated by tobacco etch virus (TEV) protease cleavage site	21 kDa	[31]
Flag-HA tag	Flag and hemagglutinin glued together	5 kDa	[32]

interaction between Gar1p, Nhp2p and Cbf5p, which are the core proteins of Cajal bodies [33]. However, TAP-tags are still relatively large at 21 kDa, which makes the original TAP-tag difficult to be utilized in experiments. In experimental context, small epitope tags are usually connected in a TAP-tag like manner, rather than the original ones.

**Epitope tags.** Epitope tags are relatively shorter as in Flag or hemagglutinin (HA) [30,34]. For instance, Jonson *et al.* utilizes 3 × Flag tag (roughly 2.73 kDa) attached to insulin-like growth factor II mRNA-binding protein (IMP) to isolate biomacromolecules in IMP1 granules [35]. To isolate IMP1 granules, researchers used FLAG-antibody-coated agarose beads to catch the epitope tags and pelleted by centrifugation. They further collected proteome by SDS-polyacrylamide gel electrophoresis fractionation and identified them by MS, in which RNA-binding proteins were enriched, as well as exon junction complex components. Granule mRNAs were also identified with microarray analysis. Epitope tags are small and does not process complex structural details, thus they may be used in investigating the composition of biomolecular condensates by pulling tagged proteins with immobilized primary antibody. However, elution buffer of extreme pH conditions may irreversibly affect protein properties, which makes epitope tags not the first choice as other affinity tags in many cases.

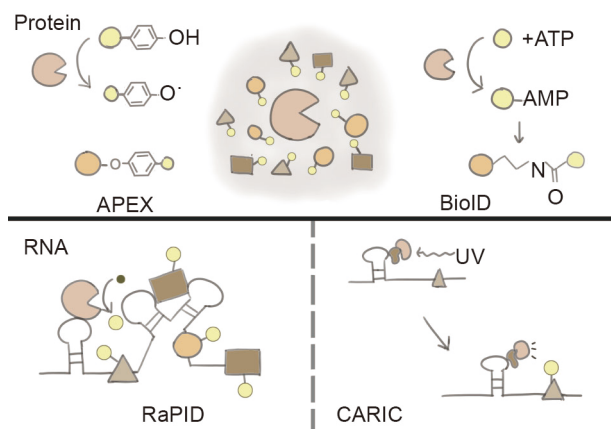
When applying epitope tags in high-throughput methods, TAP-tags are usually considered a good choice. Since TAP-tag is still large, some researchers apply multiple epitope tags combined in TAP-tag-like manner to compensate their shortcomings when investigating biomolecular condensates, like Flag-HA tags, which is what Ayache *et al.* managed to achieve in their investigation into DDX6 complexes [32]. They transfected cells with plasmid encoding DDX6 fused to Flag and HA to enrich P-body proteins. After screening the top DDX6-interacting proteins, they suggested a direct interaction between DDX6 and LSM14A or PAT1B proteins in P-body maintenance, while DDX6-interacting protein 4E-T is essential in P-body de novo assembly. The combination of TAP-tags and epitope tags can be a rational choice in biomolecular condensate composition studies. However, affinity purification requires strong interactions between proteins for them to be co-eluted. If the interaction is weak, related proteins may fail to sustain with the target protein, leading to a decreased sensitivity. Many biomolecules within condensate are loosely connected. To recognize proteins with weak interactions, complementary approaches need to be developed.

## PROTEIN-CENTRIC PROXIMITY LABELING

An important strategy often used by researchers investigating biomolecular condensates is proximity labeling, which derives from a complementary approach to classic affinity purification/mass spectrometry (AP/MS) [36]. This method was originally used for interactome mapping, exploring potential interacting proteins by tagging nearby biomolecules and capturing them with affinity matrix for MS analysis. Nevertheless, since phase separating biomolecules may also bind to other “client” proteins, researchers find it useful to determine proximal proteins of molecule of interest to search for potential proteins participating in phase separation and biomolecule condensation. Compared to AP/MS, proximity labeling has better sensitivity for including proteins with weaker interactions. Most interacting proteins having close contact with the target can be purified and analyzed in later MS stages. However, it is important to note that this method tags proximal biomolecules, which means that any nearby molecules might be labelled, regardless of interacting with target protein or not. This feature leads to a decrease in specificity of the method. On the one hand, such evidence becomes less convincing when trying to detect interacting proteins; on the other hand, when investigating biomolecular condensates, proximity labelled proteins may serve as a library of possible composing biomolecules within certain condensates, though relationships with existing compositions may be unclear.

Proximity labeling can be classified according to the target molecule (Fig. 2). When the target molecule is a protein and its surrounding molecules are detected, it is referred to as “protein-centric” methods. While “RNA-centric” methods take RNA as target molecules. Protein-centric methods are more thoroughly developed while RNA-centric methods are still under investigation. Thus when referring to “proximity labeling” we consider it to be protein-centric by default. There are two sets of principles that are often utilized by researchers: one is based on linking biotin with ligase mutants such as BioID and TurboID, the other is driven by peroxidases with APEX being the predominant labeling method.

**Proximity labeling driven by peroxidases.** APEX is one of the most-used methods to label proximal proteins. Before APEX was engineered, horseradish peroxidase (HRP) was used for labeling proximal proteins [28,37]. However, HRP is insensitive to cytosol and can only label surface proteins, restricting its use in proximity labeling. APEX originates from a class I cytosolic plant ascorbate peroxidase referred to as APX.



**Figure 2. Protein-centric proximity labeling and RNA-centric ribonucleoprotein complex mapping.**

These methods are considered complementary to classical affinity purification method and can detect weaker interactions to achieve higher sensitivity. Protein-centric methods include peroxidase-driven tags (e.g., APEX) and biotin ligase-driven tags (e.g., BioID) as two important mechanisms of condensate-related molecule tagging, while RNA-centric methods include RaPID and CARIC etc. which assist in identifying RNA-binding proteins, hence screening condensate-related molecules when RNAs play scaffolding roles. “Protein-centric” methods are more common and systematic while “RNA-centric” methods are still progressing.

APEX is a constitutive homodimer while its variant APEX2 is monomeric. APEX catalyzes oxidation of biotin-phenol to biotin-phenol radical in the presence of  $H_2O_2$ . It was first engineered as a reporter for high-resolution electron microscopy (EM) imaging and later extended to proximal labeling application [38]. Fusion of APEX to bait protein allows APEX to catalyze biotin-phenol to be oxidized by  $H_2O_2$  so that it can attach to nearby electron-rich amino acids (e.g., tyrosine and possibly cysteine, histidine and tryptophan) as radicals [39].

To perform APEX-based proximal labeling, cells should translate APEX-fused bait protein with specifically designed plasmid. Then biotin-phenol can be added to the medium preparing for tagging.  $H_2O_2$  treatment should last less than 1 minute before stopping and purifying tagged proximal proteins with streptavidin beads. Based on the same mechanism, another variant called APEX2 was also developed, with one more mutation than APEX. Lam *et al.* claimed that APEX2 enhances the cellular activity and sensitivity of APEX and improves resistance to high  $H_2O_2$  concentrations [40]. Markmiller *et al.* fused APEX2 to GTPase-activating protein SH3 domain-binding protein 1 (G3BP1) to identify proteins in stress granules (SG),

which proved high specificity [41]. G3BP1 is an essential protein for SG. They exposed the 293T cells containing fused proteins with biotin-phenol and  $H_2O_2$ , with some cells unstressed and some cells challenged by arsenite. Analysis elicited networks of stress-dependent G3BP1 interactors, including many well-characterized SG proteins. While nearly 80% of enriched proteins are SG proteins, proteins like Annexin A11 (ANXA11), a protein with amyotrophic lateral sclerosis (ALS)-associated mutations leading to protein aggregation, Annexin A7 (ANXA7) along with their interactor PEF1 are new to researchers as SG protein compositions.

Another molecule known as tyramide can also label proximal proteins in the same manner as in tyramide signal amplification (TSA) [42,43]. This method had appeared before APEX came out. Its mechanism lies in horseradish peroxidase (HRP) catalyzing generation of biotin-tyramide free radicals. The free radicals will then diffuse and attack nearby macromolecules, predominately tyrosine residues. Dopie *et al.* tagged tyramide to CENPA and SC35, which are key proteins in centromeres and nuclear speckles, to have a thorough picture of their proximal proteome and the composition of respective condensates. By defining centromere as a reference body, researchers are able to pull out lists of abundant proteins enriched in nuclear speckles. Further studies can be done to validate the list, and can move a step forward to prove that MFAP1 and PRPF38A, two proteins in the list, separate into adjacent bodies after transcription inhibition, which appear after mitosis [44]. Since cross-linking and permeabilization is needed to allow HRP-attached antibodies to enter cell membrane, the procedure is often utilized in fixated cells for *in vitro* studies. The composition would not change after crosslinked by formaldehyde so the results would genuinely reflect the proximal composition of target proteins in the very spatiotemporal surrounding, but TSA would not be considered if researchers wish to see a dynamic picture of biomolecular condensates.

APEX has a relatively small tag size (28 kDa) and significantly shorter labeling time, which makes APEX a useful labeling method when we are to investigate proximal proteins at specific spatiotemporal environment. The shortcoming of APEX is also pretty obvious:  $H_2O_2$  is toxic to cells. Due to high oxidizability of peroxides,  $H_2O_2$  treatment may introduce unnatural stresses to living cells, leading to untrusted results, though peroxide treatment time has been reduced to less than a minute. Another disadvantage of APEX lies in animal experiments.  $H_2O_2$  might not penetrate shell or skin of living animal, while it is toxic if injected. Better ways to compromise include knocking out genes to soften the shell or skin [45,46], or simply dissect the main organs and treat with  $H_2O_2$  outside living body [47].

However, scientists noticed that such compromise may not genuinely reflect the physiological status of target proteins and their interactome *in vivo*. We suspect that APEX-like methods could play significant roles in *in vitro* studies rather than *in vivo* animal studies.

**Proximity labeling driven by biotin ligases.** Unlike APEX-based proximity labeling, BioID is a choice of milder proximity labeling process. BioID originates from DNA-protein interaction detection method DamID devised by van Steensel and Henikoff [48], which makes use of the prokaryotic Dam methylase to tag interacting DNA sequences by methylation.

The original method utilizes a mutated variation of the 35-kDa biotin-protein ligase called BirA, which is found in *Escherichia coli* as a regulator of acetyl-CoA carbonxylase subunit biotinylation [49]. BirA-driven biotinylation requires two steps: First, biotin and ATP are combined and become biotinoyl-5'-AMP (bioAMP); Then bioAMP is kept at the BirA active site until a specific lysine residue of biotin acceptor tag (BAT) reacts with it. A number of BirA variants allow bioAMP to be released from BirA active site and react with nearby primary amines, among which BirA-R118G attracts researchers' attention, later named BirA\*. BirA\* biotinylates endogenous proteins in a proximity-dependent manner, after which biotinylated proteins can be precipitated using streptavidin-coated beads and identified by mass spectrometry. This is the original BioID method.

Youn *et al.* take the advantage of BioID to illustrate the proximity map of stress granules and processing bodies (PB) [50]. They performed proximity mapping on 119 proteins concerning different aspects of mRNA biology. The baits cover a wide range from mRNA capping, splicing, 3' end processing through cleavage and/or polyadenylation, nuclear export, miRNA induced silencing, deadenylation, exosome-mediated degradation, nonsense-mediated decay, decapping, translation initiation and translation control etc. These 119 proteins revealed 9,054 pairwise and 7,424 unique interactions, which leads to the prediction of SG and PB proteome. 11 unique genes from SG were also knocked out to explore essential factors of SG assembly. The conclusion says that PRRC2C, UBAP2L and CDSE1 are required for SG assembly.

As mentioned above, BioID does not utilize toxic chemicals like peroxides, which makes the reaction milder and can be performed on live animals by feeding transgenic animals with biotin-enriched food. Nevertheless, BioID needs 18 hours' reaction time, making it unsuitable for detecting proximal proteins that may only emerge for a fraction of time. Besides, BioID is also relatively larger than APEX (35 kDa vs 28 kDa), which may greatly influence the physical and chemical

properties of proteins [36].

Kim *et al.* identified the smallest known biotin ligase possible from *Aquifex aeolicus* and developed BioID2 by mutating a conserved residue within biotin catalytic domain (R40G), reducing the molecular size to 27 kDa [51]. BioID2 also requires less biotin than BioID. However, it still takes hours for BioID2 to tag proximal proteins with biotin. BASU is also a variant of BirA that came from *Bacillus subtilis*, but with its N-terminal depleted [52]. BASU reduces the amount of reaction time to 0.5–1 hour. Another optimized version of BioID was put forward by Branon *et al.* which shortens reaction time to around 10 minutes, which is later known as TurboID [53]. TurboID introduces 15 mutations to wild-type BirA. They also developed an N-terminal-depleted version called miniTurbo that can help with precise temporal control [53]. TurboID and miniTurbo exhibit high activity at 30°C, which is a bit lower than mammalian animals' body temperature, since they were evolved in yeast. Yet they provide fast proximal labeling that enables scientists to tag proteins that interact with bait protein for a limited amount of time. A recent tool developed by Ting *et al.* called Split-TurboID moves a step forward: Split-TurboID cut TurboID into two halves attached to two different motifs FRB and FKBP. When rapamycin is added, FRB and FKBP combines as one while the split halves reconstituted into a complete TurboID to perform its proximity labeling functions. This allowed improved signal-to-noise ratio and greater control over experiments when detecting proximal compositions of contact sites [54].

Summing up, proximity labeling is a great leap forward in biomolecular condensate formation since it pulls down proteins with weaker interactions or are just merely in proximal with target proteins. Proximal proteome serves as a representation of biomolecular condensate composition, though not exactly accurate. However, RNAs also comprise many condensates and their proximal proteins should also be carefully studied. These methods have not been widely used yet have great potentials in future studies.

## RNA-CENTRIC RIBONUCLEOPROTEIN COMPLEX MAPPING

Previous proximity labeling methods are usually referred to as “protein-centric” methods, which suggests that these methods detect biomolecules that are close to target proteins. It is clear that RNAs are also of great significance in composing biomolecular condensates. Many RNAs serve as scaffolds or clients in some condensates, with nuclear speckles and paraspeckles being two examples. These condensates play a big role in RNA splicing [55,56]. Paraspeckles are stemmed with

nuclear enriched abundant transcript 1 (NEAT1), which is a long-noncoding RNA (lncRNA) of over 22,000 base pairs that serve as a scaffold. Structured illumination microscopy indicated that the distal ends of NEAT1 are situated at the peripheral of paraspeckles while the middle parts are sequestered in the core. Client proteins like NONO and SFPQ interact with NEAT1, composing a fully-functioned paraspeckle [57]. The structure of nuclear speckles is still unclear since it is larger and have more complex compositions, but some lncRNAs still present within these speckles, including MALAT1 [58]. However, RNA-centric methods detecting interacting biomolecules are not frequently used than protein-centric methods.

One of the most well-known RNA-centric method is called “RNA-protein interaction detection” (RaPID) [52]. For this method, researchers flanked two bacteriophage lambda BoxB stem loops upstream and downstream of target RNA motifs. A 22-amino-acid peptide called  $\lambda$ N has high affinity to lambda-BoxB stem loop region, thus researchers attach  $\lambda$ N to BASU to biotinylate proximal proteins, which can later be captured by streptavidin. There are also variations of such methods, including incorporating MS2 stem loops into target RNA for MS2 coat protein (MCP) fused with BirA\* to recognize and label proximal proteins [59]. Such methods alike make use of known RNA-protein interactions to recruit proximity labeling enzymes to RNA motif of interest, which is efficient in investigating RNA proximal proteins. Nevertheless, flanking stem loops attached to the sides of RNA motifs might have an influence on their physiochemical states and may not represent that interactions still exist in physiological conditions.

There are also straightforward methods that detect target RNAs directly. RNA interactome capture (RIC) is one such method [60]. Instead of searching for an interacting protein attached to biotinylating enzymes, RIC makes use of ultraviolet (UV) radiation to crosslink RNA-interacting proteins. When exposed to UV of 254 nm wavelength, short-lived radicals will be induced on nucleotide bases, attacking proximal proteins and form covalent bonds, after which poly(dT) affinity resin is utilized to capture all polyadenylated RNAs. A modified version called RICK, which combines click reactions with RNA interactome capturing, features 5-ethynyluridine (EU) as RNA-labels [61]. Researchers apply click reactions to biotinylate EU so that streptavidin-coated beads can precipitate labelled RNAs along with all UV-crosslinked proteins. CARIC (click chemistry-associated RNA interactome capture) moves a step forward and promotes more efficient UV crosslinking. CARIC incorporates the photo-sensitive uridine analog 4-thiouridine (4-SU) (Fig. 2). 4-SU is labelled on RNA

along with EU during transcription. After radiated under UV light with wavelength of 365 nm, photoactive 4-SU will crosslink interacting proteins to make them stick to the RNA backbone. Then click chemistry helps to biotinylate EU for efficient capture of streptavidin beads and MS analysis later. RIC is a suitable method to capture nascent polyadenylated RNAs, which mostly comprises of mRNAs while ignoring non-coding RNAs (ncRNAs) or mature RNAs.

Probes are also an alternative way to pulldown target RNAs. Methods including CHART and RAP take this path: CHART utilizes short biotinylated complementary oligodeoxyribonucleotides while RAP tries long antisense probes across entire RNAs for higher specificity [62,63].

Since biologists make use of CRISPR systems to alter genes *in vivo*, researchers apply similar mechanisms on RNA-centric protein mapping, which is another potential means to investigate nucleic-acid-scaffolding condensates [64]. For example, Han *et al.* fuses APEX2 to catalytically inactive Cas13d whose sgRNA binds to human telomerase RNA and discovered its novel interaction with N6-methyladenosine demethylase ALKBH5 [65]. Similar studies attaching dCas13d to BioID2 [66] and BASU [67] are also possible. CRISPR methods allow direct biotinylation of RNA-interacting proteins *in vivo* and avoid crosslinking. However, it should be noted that Cas13 is relatively large in size and might interfere with RNA-protein binding.

These aforementioned methods help to map RNA-binding proteins, which might serve as possible complementary methods to protein-centric proximity labeling. Considering the role RNAs play in the formation and functions of biomolecular condensates, RNA-centric interactome mapping should be utilized when investigating the composition of biomolecular condensate with RNAs. Unfortunately, these methods are scarcely used in studying condensate composition, therefore RNA-centric ribonucleoprotein complex mapping has great potential in inspiring researchers to investigate RNA-scaffolding condensates.

## IMAGE-BASED FLUORESCENCE SCREENING

Previous methods focus on extracting protein or nucleic acid libraries of targeted biomolecular condensates for mass spectrometry or sequencing analysis. Recent advances in machine learning led to a surge in development of computational method to screen fluorescent images. Fluorescent images help to identify spatiotemporal distribution of specific biomolecules in a straightforward way. As a dominant measurement method in high-throughput screening, fluorescence screening has

high sensitivity and versatility, fast signaling speed, and does not require destruction of targeted condensates. It has been used in various researches benefiting disease diagnosis and marker/drug discovery, and many researchers are starting to apply such method in investigating biomolecular condensates [68].

Berchtold *et al.* provided a neat example of how image-based fluorescence screening can be applied to biomolecular condensate researches [69]. In their study, they targeted 1,354 human genes, each with 3 pooled siRNAs. Researchers select one marker for each biomolecular condensate of interest. Images were acquired from an automated spinning disc microscope and processed with set pipelines. Through pixel classification researchers were able to segment different biomolecular condensates, quantify important properties including organelle numbers, area and fluorescence intensity etc. for clustering, which helps to assess how different molecules influence the formation and maintenance of biomolecular condensates. Selected molecules will then move on for further confirmation. Another study done by Wheeler *et al.* developed CRAFT-ID as a method to couple image-based phenotyping with pooled CRISPR screening workflow. Researchers prepared a set of gRNA library with over 12,000 sgRNAs targeting over 1,000 RNA-binding proteins. HEK293T cells were transduced at low multiplicity of infection and cultured for 7 days to eliminate lethal guides. Then the cells were stressed and fixed for automating confocal screening. Through this pipeline they managed to reveal new stress granule modifiers and regulators [70]. These two researches demonstrated how fluorescent images can be applied to high-throughput investigations of biomolecular condensates.

These researches indicated that image-based fluorescent screening is often utilized in determining regulators and inferring potential functions rather than recognizing specific components of condensates accurately. Bioinformaticians can provide useful pipelines for analyzing these images automatically or even train programs to learn different properties of condensates. Compared to low-throughput image analysis, image-based fluorescence screening serves as an automatic “eye” built through pattern recognition that scans through thousands of pictures. Training programs to read these pictures allows results of large amounts of experiments to be processed in a glimpse, yet the accuracy of such programs could be improved. In a nutshell, image-based high-throughput fluorescence screening is a prospering means to indicate properties of biomolecular condensates.

## CONCLUSION AND PERSPECTIVES

As bioinformatics approaches continue to prosper, high-throughput screening as well as computational tools are also utilized to discover or predict possible phase-separating proteins, which greatly helps researchers to narrow down their predictions to a limited number of biomolecules involved in condensation. In this review, we managed to include the major high-throughput methods that might be useful when studying biomolecular condensates.

Different methods have different specificity and sensitivity. Proximity labeling have better sensitivity than AP/MS since it captures interacting proteins with weaker interactions, yet its specificity is not guaranteed as proteins occasionally passing by could get biotinylated. However, both methods can obtain more proteins than organelle purification since the latter one can only identify fixed components of limited condensates. None of the aforementioned methods are guaranteed to collect every potential biomolecules of targeted condensates, while it is possible for researchers to apply multiple methodologies to verify condensate compositions. Besides, some details during experiment may also alter the results: proteins may fail to be tagged; affinity-tagged biomolecules may get lost during precipitation; some immunofluorescence-binding antibodies may produce large background noise or bind to non-specific biomolecules which disturbs observation; etc. To overcome these problems, researchers are constantly improving the methods by discovering new tags or adjusting the experimental conditions. Nevertheless, these approaches do provide new insights into phase separation in biological context. More screening tools and bioinformatic tools may ignite novel methods to unravel biomolecules participating phase separation faster and more precisely.

The past decade has seen a surge of researches in biomolecular condensates. Researchers developed a variety of approaches to validate the physical properties of condensed membraneless organelles, as well as a wide range of screening methods to speculate potential biomolecules that might participate in biomolecular condensation. These studies highlighted that biomolecular condensates participate in various biological processes by concentrating molecules to increase reaction kinetics or sequestering or excluding molecules to inhibit reactions. Although many validation experiments have been developed reflecting different aspects of properties that condensation may possess, there are still a number of questions to be answered, including compositions, functions, structures as well as forming mechanisms and physical properties [71]. For

example, we still do not have enough evidence to confirm phase separation of a specific molecule *in vivo*. Paraspeckle, which is an important nuclear body participating many cellular processes, does not take spherical shapes and looks elongated, yet molecules inside still moves around freely [57]. This is against the traditional model of liquid-liquid phase separation and may include multiple mechanisms for the phenomenon to be fully understood. Further researchers need to provide more evidence to determine various properties of biomolecular condensates unbeknownst to us.

## ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China (2018YFA0507504), the National Natural Science Foundation of China (61773025, 32070666), Clinical Medicine Plus X - Young Scholars Project of Peking University (PKU2021LCXQ012) and the Fundamental Research Funds for the Central Universities.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Taoyu Chen, Qi Lei, Minglei Shi and Tingting Li declare that they have no conflict of interests.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

## OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Banani, S. F., Lee, H. O., Hyman, A. A. and Rosen, M. K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, 18, 285–298
- Friedman, J. R. and Nunnari, J. (2014) Mitochondrial form and function. *Nature*, 505, 335–343
- Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnesse, A., Hannett, N. M., Spille, J. H., Afeyan, L. K., Zamudio, A. V., Shrinivas, K., *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, 572, 543–548
- Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A., *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell*, 39, 925–938
- Pederson, T. (2011) The nucleolus. *Cold Spring Harb. Perspect. Biol.*, 3, a000638
- Boisvert, F. M., van Koningsbruggen, S., Navascués, J. and Lamond, A. I. (2007) The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.*, 8, 574–585
- Alberti, S., Gladfelter, A. and Mittag, T. (2019) Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, 176, 419–434
- Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F. and Hyman, A. A. (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324, 1729–1732
- Peng, A. and Weber, S. C. (2019) Evidence for and against liquid-liquid phase separation in the nucleus. *Noncoding RNA*, 5, 50
- Hyman, A. A., Weber, C. A. and Jülicher, F. (2014) Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.*, 30, 39–58
- McSwiggen, D. T., Hansen, A. S., Teves, S. S., Marie-Nelly, H., Hao, Y., Heckert, A. B., Umemoto, K. K., Dugast-Darzacq, C., Tjian, R. and Darzacq, X. (2019) Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife*, 8, e47098
- Li, Q., Wang, X., Dou, Z., Yang, W., Huang, B., Lou, J. and Zhang, Z. (2020) Protein databases related to liquid-liquid phase separation. *Int J Mol Sci*, 21, 1–16
- Sawyer, I. A., Sturgill, D., Sung, M. H., Hager, G. L. and Dundr, M. (2016) Cajal body function in genome organization and transcriptome diversity. *BioEssays*, 38, 1197–1208
- Domon, B. and Aebersold, R. (2006) Mass spectrometry and protein analysis. *Science*, 312, 212–217
- Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, 422, 198–207
- You, K., Huang, Q., Yu, C., Shen, B., Sevilla, C., Shi, M., Hermjakob, H., Chen, Y. and Li, T. (2019) PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.*, 48, 354–359
- Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O. P., Kovács, M., *et al.* (2020) PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.*, 48, D360–D367
- Li, Q., Peng, X., Li, Y., Tang, W., Zhu, J., Huang, J., Qi, Y. and Zhang, Z. (2019) LLPSDB: a database of proteins undergoing liquid-liquid phase separation *in vitro*. *Nucleic Acids Res.*, 48, 320–327
- Ning, W., Guo, Y., Lin, S., Mei, B., Wu, Y., Jiang, P., Tan, X., Zhang, W., Chen, G., Peng, D., *et al.* (2020) DrLLPS: a data resource of liquid-liquid phase separation in eukaryotes. *Nucleic Acids Res.*, 48, D288–D295
- Bolognesi, B., Lorenzo Gotor, N., Dhar, R., Cirillo, D., Baldrighi,

- M., Tartaglia, G. G. and Lehner, B. (2016) A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.*, 16, 222–231
21. Vernon, R. M. C., Chong, P. A., Tsang, B., Kim, T. H., Bah, A., Farber, P., Lin, H. and Forman-Kay, J. D. (2018) Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife*, 7, e31486
22. Shen, B., Chen, Z., Yu, C., Chen, T., Shi, M. and Li, T. (2021) Computational screening of biological phase-separating proteins. *Genom. Proteom. Bioinform.*, S1672-0229(21)00022-X
23. Andersen, J. S., Lam, Y. W., Leung, A. K. L., Ong, S. E., Lyon, C. E., Lamond, A. I. and Mann, M. (2005) Nucleolar proteome dynamics. *Nature*, 433, 77–83
24. Boke, E., Ruer, M., Wühr, M., Coughlin, M., Lemaitre, R., Gygi, S. P., Alberti, S., Drechsel, D., Hyman, A. A. and Mitchison, T. J. (2016) Amyloid-like self-assembly of a cellular compartment. *Cell*, 166, 637–650
25. Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.-B., Munier, A., Fradet, M., *et al.* (2017) P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell*, 68, 144–157
26. Pina, A. S., Batalha, I. L. and Roque, A. C. (2014) Affinity tags in protein purification and peptide enrichment: an overview. *Methods Mol. Biol.*, 1129, 147–168
27. Green, M. R. and Sambrook, J. (2012) *Molecular Cloning: A Laboratory Manual*, 4th Ed., New York City: Cold Spring Harbor Laboratory Press
28. Li, X. W., Rees, J. S., Xue, P., Zhang, H., Hamaia, S. W., Sanderson, B., Funk, P. E., Fardale, R. W., Lilley, K. S., Perrett, S., *et al.* (2014) New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. *J. Biol. Chem.*, 289, 14434–14447
29. Cronan, J. E. J. Jr. (1990) Biotinylation of proteins *in vivo*. A post-translational modification to label, purify, and study proteins. *J. Biol. Chem.*, 265, 10327–10333
30. Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.*, 60, 523–533
31. Li, Y. (2011) The tandem affinity purification technology: an overview. *Biotechnol. Lett.*, 33, 1487–1499
32. Ayache, J., Bénard, M., Ernoult-Lange, M., Minshall, N., Standart, N., Kress, M. and Weil, D. (2015) P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol. Biol. Cell*, 26, 2579–2595
33. Ma, X., Yang, C., Alexandrov, A., Grayhack, E. J., Behm-Ansmant, I. and Yu, Y. T. (2005) Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J.*, 24, 2403–2413
34. Tai, T. N., Havelka, W. A. and Kaplan, S. (1988) A broad-host-range vector system for cloning and translational lacZ fusion analysis. *Plasmid*, 19, 175–188
35. Jønson, L., Vikesaa, J., Krogh, A., Nielsen, L. K., Hansen, T., Borup, R., Johnsen, A. H., Christiansen, J. and Nielsen, F. C. (2007) Molecular composition of IMP1 ribonucleoprotein granules. *Mol. Cell. Proteomics*, 6, 798–811
36. Trinkle-Mulcahy, L. (2019) Recent advances in proximity-based labeling methods for interactome mapping. *F1000 Res.*, 8, F1000 Faculty Rev-135
37. Honke, K. and Kotani, N. (2011) The enzyme-mediated activation of radical source reaction: a new approach to identify partners of a given molecule in membrane microdomains. *J. Neurochem.*, 116, 690–695
38. Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H. and Ting, A. Y. (2012) Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.*, 30, 1143–1148
39. Han, S., Li, J. and Ting, A. Y. (2018) Proximity labeling: spatially resolved proteomic mapping for neurobiology. *Curr. Opin. Neurobiol.*, 50, 17–23
40. Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K. and Ting, A. Y. (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods*, 12, 51–54
41. Markmiller, S., Soltanieh, S., Server, K. L., Mak, R., Jin, W., Fang, M. Y., Luo, E. C., Krach, F., Yang, D., Sen, A., *et al.* (2018) Context-dependent and disease-specific diversity in protein interactions within stress granules. *Cell*, 172, 590–604
42. Bobrow, M. N., Harris, T. D., Shaughnessy, K. J. and Litt, G. J. (1989) Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays. *J. Immunol. Methods*, 125, 279–285
43. Bobrow, M. N., Shaughnessy, K. J. and Litt, G. J. (1991) Catalyzed reporter deposition, a novel method of signal amplification. II. Application to membrane immunoassays. *J. Immunol. Methods*, 137, 103–112
44. Dopie, J., Sweredoski, M. J., Moradian, A. and Belmont, A. S. (2020) Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *J. Cell Biol.*, 219, e201910207
45. Reinke, A. W., Balla, K. M., Bennett, E. J. and Troemel, E. R. (2017) Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. *Nat. Commun.*, 8, 14023
46. Reinke, A. W., Mak, R., Troemel, E. R. and Bennett, E. J. (2017) *In vivo* mapping of tissue- and subcellular-specific proteomes in *Caenorhabditis elegans*. *Sci. Adv.*, 3, e1602426
47. Chen, C. L., Hu, Y., Udeshi, N. D., Lau, T. Y., Wirtz-Peitz, F., He, L., Ting, A. Y., Carr, S. A. and Perrimon, N. (2015) Proteomic mapping in live *Drosophila* tissues using an engineered ascorbate peroxidase. *Proc. Natl. Acad. Sci. USA*, 112, 12093–12098
48. van Steensel, B. and Henikoff, S. (2000) Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.*, 18, 424–428
49. Roux, K. J., Kim, D. I., Raida, M. and Burke, B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and

- interacting proteins in mammalian cells. *J. Cell Biol.*, 196, 801–810
50. Youn, J. Y., Dunham, W. H., Hong, S. J., Knight, J. D. R., Bashkurov, M., Chen, G. I., Bagci, H., Rathod, B., MacLeod, G., Eng, S. W. M., *et al.* (2018) High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell*, 69, 517–532
  51. Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K. and Roux, K. J. (2016) An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell*, 27, 1188–1196
  52. Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., Roth, J. G., Gai, H., Kovalski, J. R., Siprashvili, Z., *et al.* (2018) RNA-protein interaction detection in living cells. *Nat. Methods*, 15, 207–212
  53. Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N. and Ting, A. Y. (2018) Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.*, 36, 880–887
  54. Cho, K. F., Branon, T. C., Rajeev, S., Svinkina, T., Udeshi, N. D., Thoudam, T., Kwak, C., Rhee, H. W., Lee, I. K., Carr, S. A., *et al.* (2020) Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl. Acad. Sci. USA*, 117, 12143–12154
  55. Spector, D. L. and Lamond, A. I. (2011) Nuclear speckles. *Cold Spring Harb. Perspect. Biol.*, 3, a000646
  56. Fox, A. H. and Lamond, A. I. (2010) Paraspeckles. *Cold Spring Harb. Perspect. Biol.*, 2, a000687
  57. West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R. E., Hirose, T., Bond, C., *et al.* (2016) Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.*, 214, 817–830
  58. Fei, J., Jadalaha, M., Harmon, T. S., Li, I. T. S., Hua, B., Hao, Q., Holehouse, A. S., Reyer, M., Sun, Q., Freier, S. M., *et al.* (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.*, 130, 4180–4192
  59. Mukherjee, J., Hermesh, O., Eliscovich, C., Nalpas, N., Franz-Wachtel, M., Maček, B. and Jansen, R. P. (2019)  $\beta$ -Actin mRNA interactome mapping by proximity biotinylation. *Proc. Natl. Acad. Sci. USA*, 116, 12863–12872
  60. Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L. M., Krijgsveld, J. and Hentze, M. W. (2016) Comprehensive identification of RNA-binding proteins by RNA interactome capture. *Methods Mol. Biol.*, 131–139
  61. Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., *et al.* (2018) Capturing the interactome of newly transcribed RNA. *Nat. Methods*, 15, 213–220
  62. Simon, M. D. (2013) Capture hybridization analysis of RNA targets (CHART). *Curr. Protoc. Mol. Biol.*, 101, 21.25.1–21.25.16
  63. Engreitz, J., Lander, E. S. and Guttman, M. (2015) RNA antisense purification (RAP) for mapping RNA interactions with chromatin. *Methods Mol. Biol.*, 1262, 183–197
  64. Qin, W., Cho, K. F., Cavanagh, P. E. and Ting, A. Y. (2021) Deciphering molecular interactions by proximity labeling. *Nat. Methods*, 18, 133–143
  65. Han, S., Simen, B., Myers, S. A., Carr, S. A., He, C. and Ting, A. Y. (2020) RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proc. Natl. Acad. Sci. USA*, 17, 22068–22079
  66. Li, Y., Liu, S., Cao, L., Luo, Y., Du, H., Li, S. and You, F. (2021) CBRPP: a new RNA-centric method to study RNA-protein interactions. *RNA Biol.*, doi: [10.1080/15476286.2021.1873620](https://doi.org/10.1080/15476286.2021.1873620)
  67. Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., Liao, L., Zhang, J., Li, X., Ye, J., *et al.* (2020) CRISPR-assisted detection of RNA-protein interactions in living cells. *Nat. Methods*, 17, 685–688
  68. Fang, X., Zheng, Y., Duan, Y., Liu, Y. and Zhong, W. (2019) Recent advances in design of fluorescence-based assays for high-throughput screening. *Anal. Chem.*, 91, 482–504
  69. Berchtold, D., Battich, N. and Pelkmans, L. (2018) A systems-level study reveals regulators of membrane-less organelles in human cells. *Mol. Cell*, 72, 1035–1049. e5
  70. Wheeler, E. C., Vu, A. Q., Einstein, J. M., DiSalvo, M., Ahmed, N., Van Nostrand, E. L., Shishkin, A. A., Jin, W., Allbritton, N. L. and Yeo, G. W. (2020) Pooled CRISPR screens with imaging on microrraft arrays reveals stress granule-regulatory factors. *Nat. Methods*, 17, 636–642
  71. Lyon, A. S., Peeples, W. B. and Rosen, M. K. (2021) A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.*, doi: [10.1038/s41580-020-00303-z](https://doi.org/10.1038/s41580-020-00303-z)