Title: Visual detection of submicroscopic protein clusters with a phase-separation-based fluorescent reporter.

Authors: Thomas R. Mumford[1], Diarmid Rae[1], Abbas Idris[1], David Gonzalez-Martinez[1], Ayush Aditya Pal[1], Lukasz J. Bugaj[1,2,3]*

Affiliations:
[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, 19104, USA
[2]Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, 19104, USA
[3]Institute of Regenerative Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

*corresponding author

Correspondence: bugaj@seas.upenn.edu

ABSTRACT:
Protein clustering plays numerous roles in cell physiology and disease. However, protein oligomers can be difficult to detect because they are often small and fall below the detection limits of conventional fluorescence microscopy. Existing techniques to visualize such aggregates require specialized microscopy and may require overexpression of the protein of interest, which can introduce clustering artifacts that are not representative of the endogenous protein. Here we describe a fluorescent reporter strategy that detects endogenous protein clustering with high sensitivity, called CluMPS (Clusters Magnified by Phase Separation). A CluMPS reporter detects and visually amplifies even small clusters of a binding partner, generating large, easily quantifiable phase-separated condensates as a readout. We use optogenetic clustering to show that the CluMPS approach can reliably report on target clusters as small as tetramers. Experiments and simulations showed that CluMPS activation depends on the affinity for the target, the target cluster size, and the cluster size distribution. CluMPS detected small aggregates of pathological proteins where the corresponding GFP fusions appeared diffuse. Uniquely, CluMPS permitted visualization of clusters of endogenous proteins, allowing the measurement of drug response kinetics of oncogenic condensates in patient-derived cancer cells. Finally, CluMPS could be multiplexed to report on distinct clustered species in the same cell. CluMPS thus provides a powerful yet straightforward approach to observe higher-order protein assembly in its native cellular context.

Introduction

Protein clustering is a ubiquitous behavior with important roles throughout normal and disease physiology[1–4]. Processes including transmembrane receptor signaling, transcription factor activation, and cellular stress responses all depend on the appropriate formation of protein clusters[5–9]. Aberrant clustering, in turn, has been linked with numerous pathologies, including in neurodegenerative diseases, cancer, and aging[10–19].

Clustering is driven by multivalent interactions whose characteristics generate diverse biophysical, biochemical, and functional outcomes. When interactions are strong, the resulting protein assemblies are stable and show little exchange with the surrounding environment[20]. Weaker multivalent interactions can yield more dynamic protein assemblies, such as those found in mesoscale condensates formed by liquid-liquid phase separation (LLPS)[3,21]. Collectively, such protein assemblies can serve many functions, including enhancing biochemical reaction rates, sequestering proteins, spatially restricting biochemical reactions, and buffering concentrations of cellular components[4,21–23]. Higher-order protein assembly can also play a role in signal processing, able to encode ultra-sensitivity, bistability, and memory into signaling pathways[2,24,25]. A deeper understanding of when/where protein clustering occurs will not only paint a clearer picture of cellular organization, function, and dysfunction, but will also provide novel control nodes where clustering can be modulated to regulate cell behavior for biomedical or biotechnological aims.

However, we have a limited understanding of the extent of protein clustering in cells because we lack straightforward tools to observe it. Physiologically relevant clusters are often too small to easily see using light microscopy[26–30]. The toxic “soluble oligomers" of α-synuclein, amyloid-β, and TDP-43 found in neurodegenerative diseases consist of only 10s of monomers, in assemblies roughly 10-100 nm in size[12–14,31–33]. In signaling, protein nanoclusters transduce signals from activated T cell receptors[34,35], and ~100 nm clusters of scaffold proteins compartmentalize protein kinase A activity[36]. Moreover, it is estimated that most proteins form homo-oligomers[37], and that at least 18% of the human proteome is organized into condensates > ~100 nm in size[38].

Direct observation of such small clusters is often not possible with simple fluorescent protein fusions because the clusters do not concentrate enough fluorescent proteins to exceed the fluorescence background and are smaller than the diffraction limit of light (~200-300 nm)[39]. This principle is demonstrated by *A. thaliana* Cryptochrome 2 (Cry2), which clusters upon blue light stimulation[40]. Cry2 cluster formation is a function of its concentration, and below a certain concentration, clusters of fluorescent fusions of Cry2 can not be observed with conventional fluorescence microscopy[41,42]. Nevertheless, small clusters continue to form in this low concentration regime, evidenced by the fact that certain optogenetic tools based on Cry2 clustering can still be activated despite the absence of visible clusters[40,43]. Additionally, large Cry2 clusters can be generated at low concentrations by co-expression with multivalent binding partners[42] or by appending Cry2 with certain peptides or disordered regions[44,45], suggesting that small clusters are forming in the absence of these additional domains.

To observe such small protein clusters, specialized microscopy methods can be used, including Foerster resonance energy transfer (FRET), fluorescence anisotropy, and super-resolution imaging. However, these methods are subject to drawbacks including complexity in acquisition and analysis, low signal-to-noise ratio, significant optimization for each target, long acquisition times that obscure fast dynamic changes, and the need for protein overexpression[46–55]. Importantly, overexpression of clustering proteins can yield artifacts because clustering is

non-linearly dependent on concentration, thus challenging interpretation of whether the clustering that is observed is representative of the native cellular context[56,57].

We sought to address this gap with a fluorescent reporter of protein clustering that 1) could detect small, endogenous protein clusters, 2) is simple to design and image, 3) is easily adaptable to new targets, and 4) could be imaged in live mammalian cells. Here we introduce such a reporter strategy called CluMPS (Clusters Magnified via Phase Separation). CluMPS is a multivalent fluorescent reporter that detects and amplifies clusters of a target protein, producing large phase-separated condensates that can be easily measured. Combining live cell imaging, optogenetic clustering, and stochastic modeling, we validated the CluMPS approach, demonstrated its sensitivity to small clusters, and determined critical system features required for CluMPS activation. We then demonstrated that CluMPS can visualize diverse pathological clusters that otherwise appear diffuse, including of endogenous proteins, can report on drug-induced cluster dynamics, and can be multiplexed to visualize distinct clustered species in the same cell. Our work thus provides a modular reporter strategy for sensitive and dynamic observation of protein clusters in mammalian cells.

Results

Design of the CluMPS reporter. The CluMPS reporter leverages biomolecular phase separation to detect and visually amplify small protein clusters. We chose this strategy for two main reasons. First, phase separation can form large, micron-scale condensates that are easy to detect through conventional microscopy. Second, phase separations can be seeded by the interaction of small protein clusters[20,24,45,57–60]. This concept was recently harnessed in the SPARK reporter, which forms large phase-separated condensates when triggered by a dimerization event[60]. In this approach, stimulus-induced dimerization of two fluorescent clustered species — a hexamer (HOTag3) and a tetramer (HOTag6)— resulted in large, fluorescent reporter condensates that could be visualized and quantified with a high signal-to-noise ratio.

CluMPS inverts the SPARK reporter to generate a reporter of protein clusters. A CluMPS reporter is multivalent and has affinity for a target protein of interest. If the target is clustered, CluMPS forms large fluorescent condensates of itself and the target. Otherwise, the CluMPS reporter remains diffuse (**Figure 1A**). While CluMPS would necessarily form higher order condensates containing both reporter and target clusters and thus would not be fully orthogonal, its conditional condensation would report on otherwise invisible target protein clusters, could be readily adapted to report on endogenous proteins, and would provide a straightforward and modular method to detect submicroscopic clustering in live cells, thus fulfilling our design criteria.

A CluMPS reporter consists of three domains: 1) a binding domain for a target protein, 2) a fluorescent protein, and 3) an amplification domain that provides multivalency (**Figure 1B**). We selected five candidate amplification domains — two multimeric coiled-coils (HOTag3 and HOTag6 used in the SPARK reporter)[60] and three commonly used intrinsically disordered regions (IDR: FUS(LC), LAF-1 RGG, Xvelo)[61–64] — and we tested their ability to enable CluMPS-amplification of small clusters of GFP. To target GFP, we used the GFP-binding nanobody LaG17 as the binding domain[65].

To benchmark the performance of our candidate CluMPS probes, we used an optogenetic approach to inducibly aggregate GFP. We fused GFP to the photolyase homology region of *A. thaliana* cryptochrome 2 (Cry2), which forms clusters under blue light stimulation in mammalian cells[40](**Figure 1C**). Importantly, the size of light-induced Cry2 clusters is highly dependent on concentration[41,44,66]. While Cry2-GFP forms large, visible clusters under high expression, clusters

in low-expressing cells are too small to observe above the fluorescence background of the cell (**Figure 1D**). Such small unobservable clusters provided an ideal test-case for CluMPS amplification.

We tested the ability of our candidate CluMPS constructs to form protein condensates when co-expressed with Cry2-GFP and stimulated with blue light. When expressed alone or with Cry2-GFP in the dark, all CluMPS variants showed diffuse expression (**Supplementary Figure 1A-C, Supplementary Video 1, Supplementary Figure 2**). Upon blue light stimulation, CluMPS variants with disordered amplification domains showed little or no phase separation, comparable to the negative control (LaG17-mCh with no amplification domain). By contrast, constructs with multimeric amplification domains (HOTag3 or HOTag6) rapidly formed large, easily visible condensates within 10s of seconds of light stimulation (**Figure 1E, Supplementary Figure 1A-C, Supplementary Video 1**). Notably, CluMPS condensation could be observed at all concentrations of Cry2-GFP, including at low levels where Cry2-GFP clustering was otherwise undetectable (**Figure 1F**). Thus, we identified at least two distinct amplification domains that could be used to implement the CluMPS approach. We used HOTag3 (6-mer) as the amplification domain for the majority of the CluMPS experiments presented in this manuscript.

While we intended CluMPS to detect and amplify small protein clusters, we considered whether CluMPS could report false positives, for example by phase separating with a target that was initially monomeric but had a weak intrinsic propensity to self-associate. To test this possibility, we observed whether the LaG17-mCh-HOTag3 construct (LaG17-CluMPS) would phase separate with a fusion of GFP with either FUS(LC) or RGG, two disordered domains with sufficient self-affinity to drive phase separation when multimerized[45,59,61]. In both cases, CluMPS did not induce phase separation at any expression level of the GFP fusions (**Supplementary Figure 3**). Thus, while CluMPS is activated by small clusters, we do not observe activation by monomers, even those with a latent ability to multimerize.

A stochastic model reveals important parameters for CluMPS activation

To understand the key principles that governed the CluMPS strategy, we built an *in silico* model to rapidly explore the experimental parameter space. We modified a previously described kinetic Monte Carlo model of protein clustering[24] and extended it to simulate interactions between two multivalent species, the 'target' and 'reporter' (CluMPS) (**Figure 2A**). Each species moved freely within its 2D grid but could interact with members of the opposite species near its corresponding grid position (**Figure 2B**). The valency of each unit defined the maximum number of interactions that a unit could make with neighbors of the opposite species. A binding energy ΔE defined the strength of a single target:reporter interaction. Reporter units had a valency of six to approximate the valency of CluMPS probes containing HOTag3. During each simulation step, one target or reporter unit could move to an adjacent grid space. The likelihood of movement p_{move} was determined by the enthalpy of cross-plane interactions, a function of ΔE, the number of nearby binding partners, and the valency of the selected target/reporter unit (**Figure 2B,** see **Methods** for more details). Using this framework, we simulated through time the spatial distribution of targets and reporters as a function of three system parameters that we hypothesized were important: 1) binding energy (ΔE), 2) target cluster size, and 3) the fraction of target that was in a cluster, since proteins clusters can exist in a distribution of sizes within the cell (**Figure 2C**). For (3), we approximated cluster size distributions by adding monomeric target units until a predefined percentage of clustered targets was reached (from 25% to 100%). We began each simulation with randomly placed target and reporter units and ran simulations until a steady state

of reporter condensates was reached. We then quantified the fraction of reporter units that were part of a cluster at steady state (see **Methods** for full details).
Simulations of these simple rules of motion and interaction yielded large target/reporter condensates over time (**Figure 2D, Supplementary Video 2**). Both higher ΔE and higher target valency drove stronger condensation, recapitulating intuitive aspects of the CluMPS system (**Figure 2D,E**). Condensation increased sigmoidally with increasing binding affinity, with a more switch-like relationship at higher target valencies (**Figure 2E)**. Condensation was also sensitive to the fraction of target that was in a cluster (multivalent). Increasing the amount of target monomers (decreasing the percentage of target clusters) diminished condensation because monomers that bound reporters lowered the likelihood of multivalent-multivalent interactions required for condensation (**Figure 2F**). Our simulations also suggested that the CluMPS strategy could robustly amplify clusters as small as tetramers, and, under ideal conditions (high affinity, high percent clustered) even dimers and trimers, though to a lesser extent (**Figure 2E,F**). Collectively, these simulations are consistent with the established role of multivalency in driving phase separation and provide predictions for how to optimize and interpret CluMPS behavior[57,60,67].

Optogenetic test systems validate model predictions
We next sought to experimentally test our model predictions that affinity, target valency, and percentage of target clustered can modulate the extent of CluMPS condensation. To test the role of affinity, we measured amplification of Cry2-GFP using CluMPS reporters with distinct GFP-binding nanobodies of different affinities: LaG17-CluMPS (K_d = 50 nM), LaG6-CluMPS (310 nM), and LaG18-CluMPS (3800 nM) (**Figure 3A,B**)[65]. In agreement with predictions, LaG17-CluMPS robustly amplified light-induced Cry2-GFP clusters, whereas LaG6-CluMPS and LaG18-CluMPS were progressively less effective, forming smaller and fewer clusters after light stimulation (**Figure 3C, Supplementary Figure 4, Supplementary Video 3)**.

To test the roles of target cluster size and distribution, we required a method to generate stoichiometrically defined GFP clusters. Cry2 clustering was inadequate for this purpose, since Cry2 cluster size is polydisperse and highly concentration-dependent[66]. We thus adopted a design in which we generated stoichiometrically-defined clusters by concatemerizing the iLid protein (2-6mer “strings”, iLidx2-iLidx6), which forms heterodimers with sspB under blue light[58,68](**Figure 3D**). In this system, blue light transitioned sspB-GFP from a monomer to a cluster, with cluster size defined by the string length. Co-transfection of the iRFP-iLid$_{string}$, sspB-GFP, and LaG17-CluMPS allowed us to observe cluster formation upon blue light stimulation (**Supplementary Figure 5**). As predicted by our simulations, the magnitude of CluMPS amplification depended on string length. Robust condensation was observed in the presence of iLidx6 and was progressively weaker with smaller string lengths (**Figure 3F, Supplementary Figure 5**). Importantly, the strong clustering in cells expressing iLidx6 was still lower compared to that observed in cells expressing Cry2-GFP, which forms clusters with valencies >> 6. Notably, in some cells, condensation could be observed with string sizes as small as 2 and 3, as predicted by our simulations. To further confirm these results, we co-expressed LaG17-CluMPS with tandem GFPs (2X or 3X) on a single peptide chain, and we observed robust clustering in most cells (**Supplementary Figure 6**). These results confirm that, under ideal conditions (stable multimers, high CluMPS:target affinity, 100% multimerized), CluMPS can be activated by target clusters as small as trimers and dimers.

Cell-to-cell variability in CluMPS activation by iLid$_{string}$ was directly correlated with cluster size distribution, i.e. the estimated percentage of target (GFP) that formed clusters. We estimated this percentage by comparing the relative expression levels of iRFP-iLid$_{string}$ and GFP-sspB, the two

components whose dimerization was controlled by light (**Figure 3G, Supplementary Figure 7**). For example, a high sspB/iLid$_{string}$ ratio resulted in a low percentage of targets clustered since there remained an excess of monomeric, unbound GFP-sspB even after light stimulation (**Figure 3G, Supplementary Figure 7**). We poly-transfected GFP-sspB and iLid$_{string}$ plasmids to achieve uncorrelated expression and a wide range of sspB:iLid string ratios in a single pool of cells[69] (**Supplementary Figure 8**). In line with simulations, we found that cells with a high percentage of clustered target (50-100%, 0.5 < sspB/iLid$_{string}$ < 2) showed strong light-induced condensation, whereas cells with 25-50% (2 $\leq$ sspB/iLid$_{string}$ < 4) showed substantially less condensation, and at < 25% (sspB/iLid$_{string}$ $\geq$ 4), condensation was not detectable (**Figure 3H, Supplementary Figure 9**). Collectively, these experiments validate our simulations, set forth design rules for CluMPS reporters, and provide context for interpreting CluMPS activity, for example when targeting endogenous proteins whose cluster status and distribution are unknown.

CluMPS amplifies small clusters of pathological proteins

To determine whether CluMPS enhances visualization of known protein oligomers, we tested whether it could detect and amplify small aggregates of disease-associated proteins. The nucleocapsid (N) protein of SARS-CoV-2 can phase separate due to its self-affinity, disordered regions, and RNA interactions[70–74]. However, a GFP-N fusion appeared diffuse in HEK 293T cells, as previously observed,[70] likely because N-protein condensates are small (**Figure 4A**). By contrast, co-expression with LaG17-CluMPS resulted in large, micron-scale aggregates of both GFP-N and the reporter, observed in 100% of cells (**Figure 4A-C**).

We then tested whether CluMPS could amplify aggregates of TDP-43, a nuclear protein that is found in neuronal inclusions in ~97% of sporadic ALS patients, and whose aggregation is strongly associated with ALS progression[12,32,75]. Because TDP-43 is a nuclear protein, we modified LaG17-CluMPS with a nuclear localization sequence (NLS) at the C-terminus. GFP-TDP-43 was expressed in the nucleus and, like N-GFP, was largely diffuse in HEK 293T cells under 40X confocal microscopy (**Figure 4D**). By contrast, co-expression with nuclear-targeted LaG17-CluMPS increased the detection of TDP-43 condensates from ~2% to ~80% of cells (**Figure 4D-F**), revealing the underlying self-association tendencies of GFP-TDP-43. Notably, phase separation was even more evident in the CluMPS channel, likely an artifact of differential expression levels and background fluorescence of the target and reporter (**Figure 4D**).

Finally, we tested the ability of CluMPS to amplify condensates of EML4-ALK, an oncogenic fusion protein commonly found in lung cancers[10,76]. The EML4-ALK-GFP fusion alone was observed to form condensates in only ~25% of transfected HEK 293T cells and remained diffuse in the remaining 75% of cells (**Figure 4G-I**). However, co-expression with LaG17-CluMPS yielded large condensates in virtually every cell (**Figure 4G-I**). Thus, in the absence of CluMPS, GFP fusions of EML4-ALK form clusters that are too small to see above the diffuse GFP background in most cells, and CluMPS can successfully detect and amplify these small clusters.

Collectively, these results demonstrate how CluMPS allows straightforward detection of pathological oligomers that are not visible as fluorescent fusions, both in the cytoplasm and in the nucleus. Moreover, LaG17-CluMPS can be applied ‘off-the-shelf’ to assess the clustering status of a variety of GFP fusions.

CluMPS visualizes dynamics of endogenous oncogenic aggregates in cancer cells.

A unique and powerful feature of CluMPS is that its design allows visualization of clusters of unmodified endogenous proteins. To test this ability, we designed a CluMPS reporter to amplify condensates of the EML4-ALK oncogene in patient-derived H3122 lung cancer cells (**Figure 5A**).

EML4-ALK drives oncogenic signaling by aggregation and autophosphorylation of the ALK kinase domain, which recruits adapter proteins like Grb2 and Gab1 to trigger downstream growth signals[10]. To allow CluMPS to detect the condensates, we used the proline rich domain (PRD) from Gab1 (amino acids 263-451), which binds Grb2[77], an essential component of EML4-ALK condensation[10] (**Figure 5B**). Upon expression of the resulting construct, Gab1(PRD)-CluMPS, in H3122 cells, we observed large CluMPS condensates in transfected cells, suggesting the presence of a clustered binding target of Gab1(PRD) (**Figure 5C**). By contrast, cells transfected with Gab1(PRD)-mCh (with no amplification domain) showed diffuse mCherry expression, demonstrating the necessity for the multivalent interactions of CluMPS to visualize the underlying clusters. To determine whether the CluMPS condensates were detecting and amplifying EML4-ALK, we performed immunofluorescence. We found that virtually every CluMPS condensate (97 $\pm$ 1%, N = 956 condensates) overlapped with ALK immunostaining, demonstrating a high specificity for Grb2 associated with EML4-ALK condensates (**Figure 5D**). We also found CluMPS condensates in the vast majority of CluMPS-expressing cells (89 $\pm$ 4 % of cells with condensates, N = 224 cells), demonstrating a high sensitivity for CluMPS. EML4-ALK puncta that co-localized with Gab1(PRD)-CluMPS were substantially larger than those that did not (53 ± 2 vs. 11.0 ± 0.1 pixels, roughly 3.2 ± 0.12 vs 0.66 ± 0.01 μm^2). While the absolute sizes may be influenced by fixation artifacts[78], the relative size differences demonstrated that CluMPS successful amplified the endogenous EML4-ALK condensate (**Figure 5D,E, Supplementary Figure 10**).

The ability to visualize endogenous aggregates allowed us to track their dynamics in response to the ALK inhibitor crizotinib (**Figure 5F, Supplementary Figure 11, Supplementary Movie 4**). Although it is known that crizotinib blocks EML4-ALK activity and suppresses its condensation, in part due to dissociation of Grb2[10,79,80], the kinetics and extent of condensate dissolution has not been observed in living cells. We found that, upon drug addition, CluMPS condensate area rapidly decreased within 10 minutes, followed by a slow but measurable decrease thereafter (**Figure 5G, Supplementary Figure 11**). The half-life of the initial decay of the CluMPS signal was ~4 minutes. Interestingly, this rate is ~4X faster than a recent measurement of the rate of Grb2 dissociation from heterologously overexpressed EML4-ALK[79]. We validated this fast decay through a fixed-cell time course experiment and obtained similar fast decay of the EML4-ALK/CluMPS condensates ($t_{1/2}$ = 3.5 minutes) (**Supplementary Figure 10**). This faster measured rate may reflect the differences in measuring endogenous aggregates versus aggregates of overexpressed EML4-ALK, as performed previously[79]. We also observed that, despite extended drug treatment, ALKi did not induce full dissociation of CluMPS condensates, in line with previous reports showing that EML4-ALK (V1) condensates are relatively stable[10,80]. These residual CluMPS condensates were not merely an artifact of CluMPS expression, since residual ALK clusters were also detected in untransfected, drug-treated H3122 cells (**Supplementary Figure 10**). The presence of CluMPS condensates after drug treatment further suggests that these residual EML4-ALK clusters retain kinase activity and signaling potential, which is required for the colocalization of Grb2 and Gab1(PRD). Taken together, these results show that CluMPS can visualize the dynamics of endogenous protein aggregates and showcases potentially important differences between measuring assemblies of endogenous proteins in their naive context as compared to their exogenous overexpression in unrelated cell types.

Multiplexed amplification of distinct target clusters in single cells.

Finally, we asked whether we could multiplex the CluMPS strategy to magnify multiple aggregates in individual cells. We leveraged the fact that both HOTag3 and HOTag6 work as strong amplification domains (**Figure 1**) but do not cross-interact when expressed in the same cells[60]. We thus generated a HOTag6-based CluMPS reporter using miRFP and LaG17 as the

binding domain (LaG17-CluMPS) to multiplex with a HOTag3-based reporter with mRuby and a nanobody for the ALFA tag[81] (nbALFA-CluMPS)(**Figure 6A**). We coexpressed these two constructs in cells with two distinct optogenetic clustering targets: Cry2-ALFAtag or iLID fused to the 24-mer FTH1, which recruits and multimerizes sspB-GFP under blue light (**Figure 6A**)[59,81]. Upon expression, we observed pre-formed aggregates in the iRFP channel, potentially due to basal association between the sspB and the 24-meric iLid. Nevertheless, blue light triggered the appearance of distinct condensates in both the mRuby and iRFP channels, reflecting amplification of the two distinct optogenetic clusters (**Figure 6C, Supplementary Movie 5**). Importantly, condensates in the two channels were non-overlapping and appeared with different kinetics, demonstrating detection of distinct cluster species using orthogonal CluMPS probes (**Figure 6D,E**).

DISCUSSION

Despite the growing appreciation for roles of protein clustering and condensation, we lack accessible methods to observe many such higher-order assemblies because of their small size and the limits of fluorescence microscopy. We developed the CluMPS reporter to overcome these limits through amplification of small protein clusters, with a sensitivity that extends to tetramers and even trimers and dimers under ideal conditions. Through simulations and experiments, we showed that CluMPS amplification is a function of affinity for the target, target cluster size, and the fraction of target that is clustered. We note that although CluMPS was designed to visualize small endogenous clusters, CluMPS works on a large range of cluster sizes and thus will find use to visualize a wide variety of endogenous protein assemblies. Our CluMPS reporters that bind GFP or ALFA-tag can report on fusions of these tags with diverse proteins of interest. Alternatively, CluMPS can visualize the dynamics of endogenous protein clusters and, with an activity-gated binding domain, even downstream signaling (as with Gab1(PRD)-CluMPS, **Figure 5)**. Finally, we showed that the CluMPS reporter can be readily multiplexed with established fluorescence reporters and also with orthogonal CluMPS variants to visualize multiple clusters in single cells.

CluMPS provides several advantages relative to existing methods to visualize aggregates in living cells. Unlike FRET, fluorescence anisotropy, or super-resolution techniques, CluMPS can be imaged using a single channel, has high signal-to-noise ratio, and can be observed without complex microscopy and image processing. Further, CluMPS does not require overexpression of the target protein and thus can avoid altering target concentration levels, providing confidence that the detected clusters were in fact present in their physiological context. A recent study developed a transcriptional reporter of protein aggregation called yTRAP, where the aggregation of a protein of interest sequestered a transcriptional activator and reduced fluorescence from a transcriptional cassette[82]. However, this approach required large, stable aggregates of overexpressed protein and could only report on aggregates over long timescales (~hours to days) in yeast. By contrast, CluMPS provides an easy-to-use method for the dynamic detection for endogenous protein clustering across a broad range of cluster types and sizes within the native cellular context in mammalian cells (**Figure 3**).

Our modeling and experiments clarify the limitations of CluMPS, but also are instructive for its interpretation. When probing an endogenous target with CluMPS, a failure to observe condensation could be interpreted as a lack of clustering, but could equally be attributed to either low binding affinity for the target or a low fraction of clustered target. However, these failure modes can be individually debugged and yield insight, for example using tandem binding domains to increase affinity/avidity, or through testing CluMPS activation against exogenously expressed fusions where binding affinity and stoichiometry can be defined. Another potential limitation is that CluMPS formation necessarily perturbs the clustering state of its targets. Thus, conclusions drawn from the CluMPS reporter should be validated by an orthogonal approach (e.g. immunofluorescence), as is good practice with any live cell reporter. Nevertheless, we do not observe that CluMPS causes cluster formation where none existed (CluMPS expressed alone is diffuse, as is CluMPS targeting GFP, GFP-FusLC and GFP-RGG (**Supplementary Figures 2,3,6**)), suggesting that CluMPS reports on *bona fide* pre-existing clusters. Finally, although we used CluMPS probes that incorporate mCherry, mRuby, and miRFP, the identity of the fluorescent protein can have a strong effect on clustering and phase separation of the construct[44]. Thus, the activity of CluMPS variants of different colors must be determined empirically.

We anticipate that the simplicity and high signal-to-noise ratio of CluMPS will enable numerous unexplored applications, for example for detecting pathogenic aggregates in patient biopsies, for

screening drug libraries for compounds that disrupt small neurodegenerative aggregates, or even as amplifiers embedded in synthetic protein circuits that detect and respond to target protein aggregates. We thus expect CluMPS will enhance our understanding of the prevalence, relevance, and dynamics of protein clusters that have previously been hard to see.

FIGURES

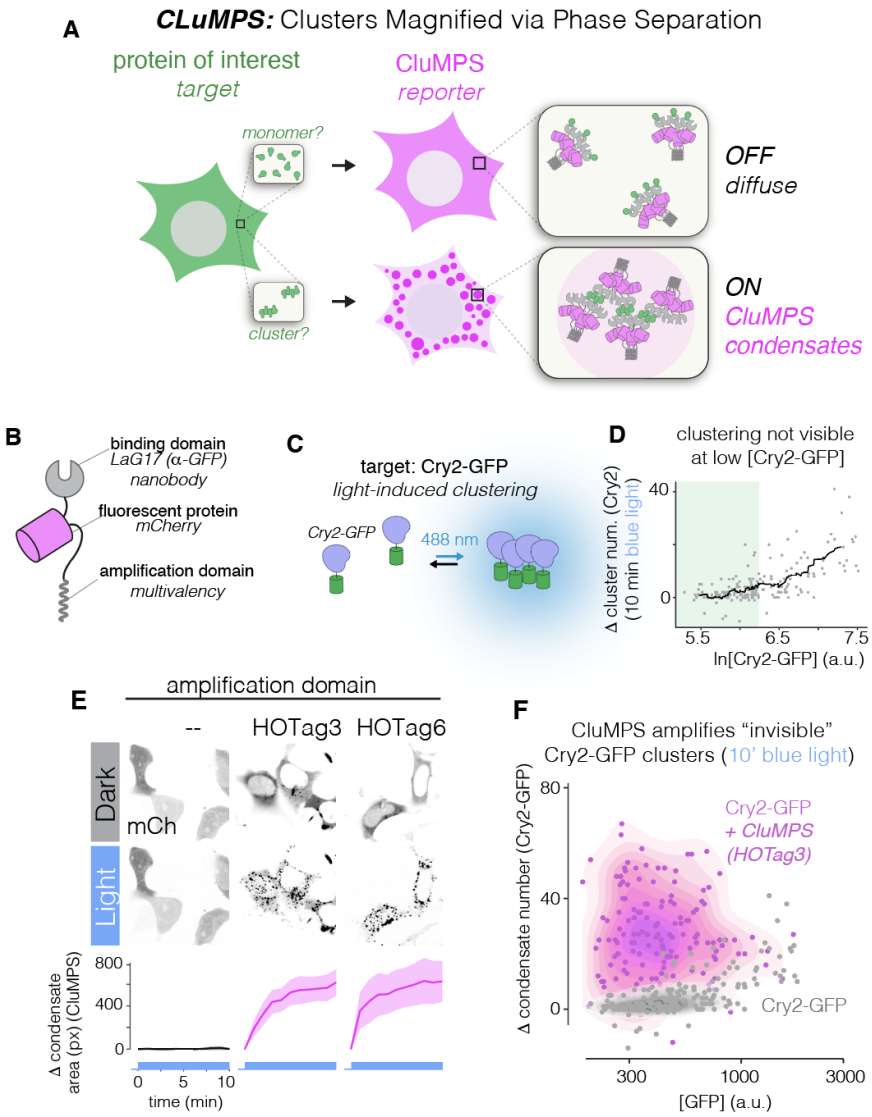


Figure 1: CluMPS reporters amplify small protein oligomers. (A) A CluMPS reporter (magenta) forms large condensates when its target (green) is clustered but remains diffuse when the target is monomeric, providing an easily detectable readout of protein clustering. **(B)** A CluMPS reporter consists of a target binding domain, fluorescent protein, and an amplification domain that facilitates phase separation in the presence of a clustered target. **(C)** Cry2-GFP is diffuse in the dark but clusters when stimulated with blue light. **(D)** Cry2-GFP clustering is concentration dependent, and at low concentrations (green), most light-induced Cry2-GFP clusters are too small to see. **(E)** CluMPS variants with HOTag3 or HOTag6 amplification domains successfully generated large phase-separated condensates upon optogenetic clustering of Cry2-GFP, which was expressed at levels where clusters are not visible. A control reporter construct without an amplification domain (left) shows no condensation. Data represent the means, error bars = 95% CI of approximately 20-50 cells per group. See also (**Supplementary Figure 1**). **(F)** Quantification of light-induced Cry2-GFP clusters, in the presence or absence of Lag17-CluMPS (HOTag3) shows CluMPS-induced amplification of Cry2-GFP clustering including at low concentrations of Cry2-GFP where, in the absence of CluMPS, clustering is undetectable.

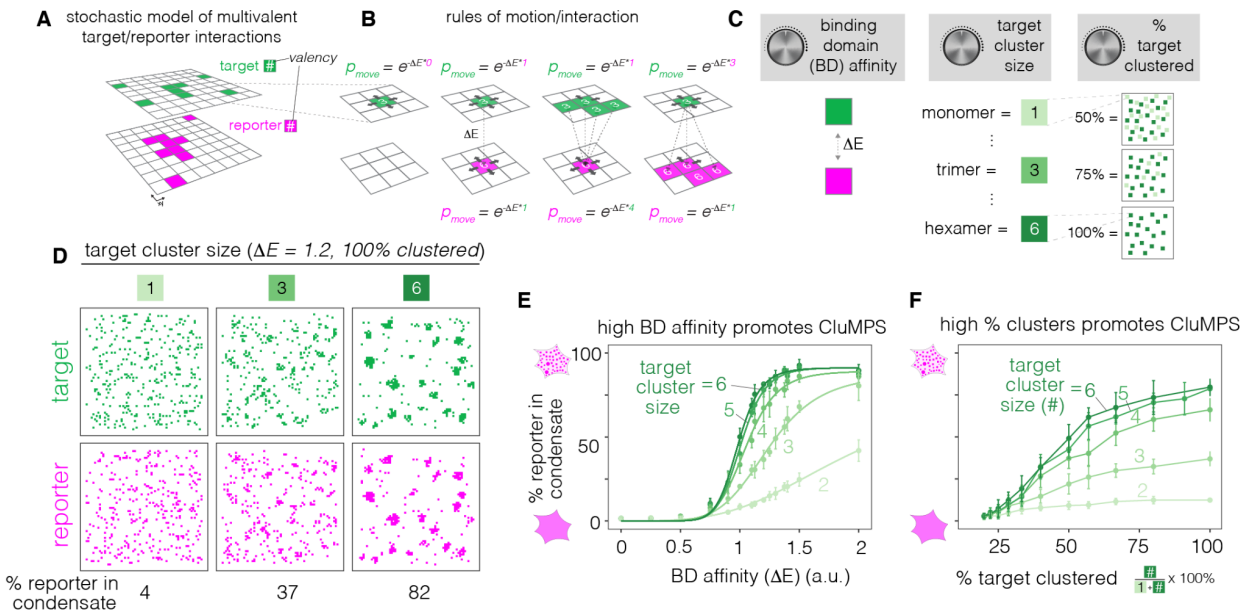


Figure 2. Stochastic model of CluMPS condensation

(A) Schematic of a kinetic Monte Carlo model of CluMPS/target interaction. CluMPS units (magenta) and target units (green) populate separate but interacting 2D grid planes. Both CluMPS and target units have a predetermined valency, which sets the number of adjacent units on the opposite plane with which they can interact. **(B)** During each simulation step a random unit is selected to move. The probability that the move is accepted is set by the enthalpy of cross-plane interactions of the unit's initial state. The number of interactions is limited by the unit's valency; i.e. if a unit set to move has four cross-plane neighbors, but a valency of three, the enthalpy will be calculated as if there are only three neighbors (right-most panel). See **Methods** for more details. **(C)** The model was used to understand how CluMPS parameters (binding affinity ΔE, target cluster size, and percent target clustered) influenced condensation. **(D)** Representative simulation results for different target valencies with constant binding energy and 100% target clustered. Simulated CluMPS/target condensates form more readily with larger target cluster sizes. See also **Supplementary Movie 2**. **(E)** Quantitation of simulations that sampled different binding domain affinities across target cluster sizes. Condensation increased with higher binding energy and larger target cluster sizes. Data points represent means ± 1 s.d. of 10 simulations. Trend lines are sigmoidal curves of best fit for each target cluster size. **(F)** Quantitation of simulations that sampled different distributions of target clustering and target cluster sizes. Addition of monovalent target suppresses CluMPS:target condensation for all target cluster sizes. Data points represent means ± 1 s.d. of 10 simulations.

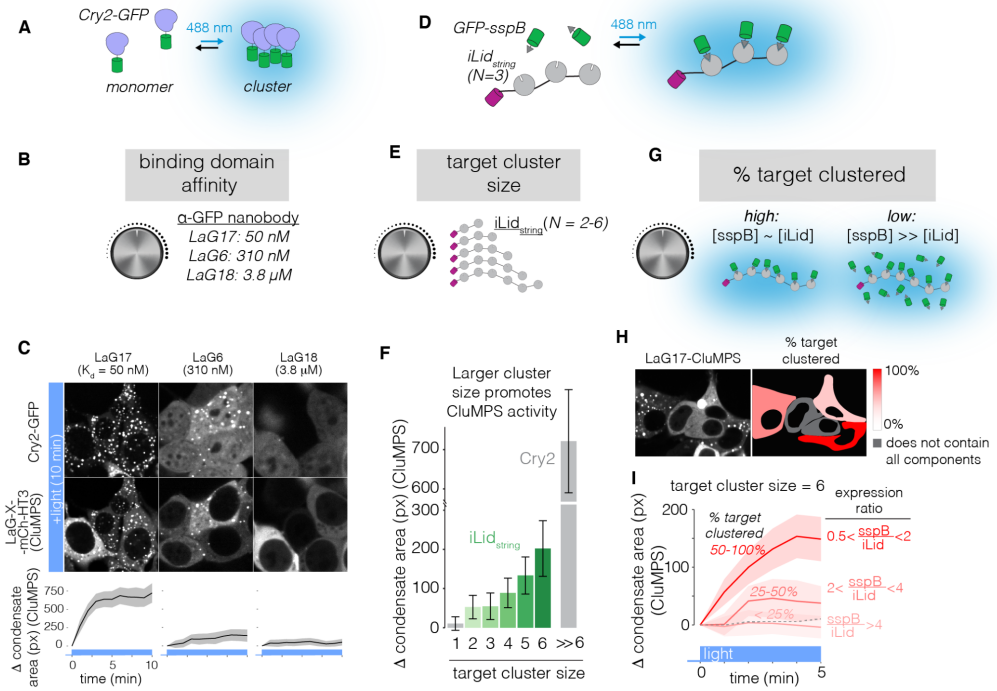


Figure 3. CluMPS activation is determined by binding affinity, target cluster size, and cluster size distribution

(A) To test the effect of CluMPS affinity, we measured amplification of light-induced Cry2-GFP clustering using CluMPS variants with different affinities of GFP-binding nanobody **(B)**. **(C)** LaG17 (K_d = 50 nM, left) yielded robust CluMPS amplification of Cry2-GFP clusters, while LaG6 (K_d = 310 nM, center) showed weaker phase separation. LaG18 (K_d = 3800 nM, right) was largely too weak for CluMPS-based condensation with Cry2-GFP. Traces represent the mean. Ribbons = 95% CI of approximately 100-200 cells per group. **(D)** Stoichiometrically defined multimers of Lid-sspB, an optogenetic dimerization pair, were used to test how CluMPS activity depended on target cluster size and distribution. Multiple iLids were fused on a single peptide ($iLid_{string}$), while sspB was monomeric and fused to GFP. Strings of sizes 2-6 were generated **(E)**. Light-induced dimerization formed small GFP clusters of predefined sizes. **(F)** Light-induced clustering of sspB-GFP with different iLid string sizes in the presence of LaG17-CluMPS. Clustering increases with valency of the iLid string, though condensates induced by iLidx6 are still substantially smaller than those induced by larger clusters of Cry2-GFP. Data represents means. Error bars = 95% CI of approximately 150-200 cells. **(G)** The percentage of light-induced clusters in a given cell can be estimated by examining the relative amounts of iLid string and GFP-sspB (see also **Supplementary Figure 4**). (**H**) Representative image of sspB-GFP/CluMPS condensation in cells that have iLidx6, GFP-sspB, and LaG17-CluMPS. Color coding indicates the estimated percentage of GFP-sspB per cell that is clustered and can thus stimulate phase separation. (**I**) Quantification of relationship between light-induced CluMPS activation and the percentage of clustered target. Robust formation of GFP-sspB/CluMPS condensates is observed when 50-100% of sspB-GFP is clustered. Condensation is diminished when only 25-50% of the target is clustered, and is virtually undetectable when only < 25% of target is clustered. Dashed gray line represents negative control of cells with CluMPS and GFP-sspB but no iLid string. Traces represent means. Error bars = 95% CI of approximately 200 cells for the '50-100% clustered' group and 40-60 cells for other groups.

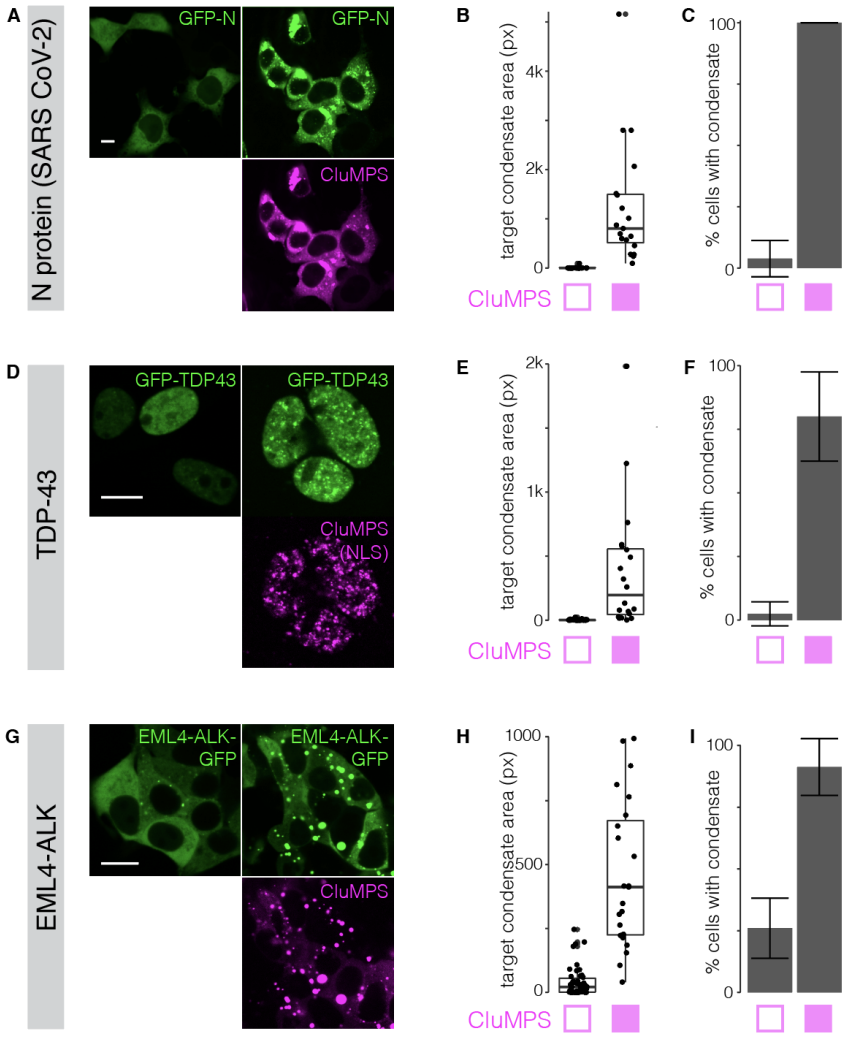


Figure 4. CluMPS detects clustering of pathological proteins.
(A) SARS-CoV-2 nucleocapsid protein (N) with N-terminal GFP fusion is diffuse when expressed alone but aggregates when co-expressed with CluMPS. Quantification of cluster area **(B)** and fraction of cells with detectable clusters **(C)** of GFP-N in the presence or absence of CluMPS. **(D)** TDP-43 is a nuclear protein whose aggregation is associated with amyotrophic lateral sclerosis (ALS). A GFP-TDP43 fusion is diffuse in the nucleus when expressed alone but forms large clusters when co-expressed with LaG-17 CluMPS fused to an NLS. **(E)** Quantification of cluster area **(F)** and fraction of cells with detectable clusters **(G)** of GFP-TDP43 in the presence or absence of LaG17-CluMPS(NLS). EML4-ALK is an oncoprotein that forms aggregates required for its oncogenic signaling. An EML4-ALK-GFP fusion is clustered in few cells when expressed alone but robustly forms droplets when co-expressed with LaG17-CluMPS. Quantification of cluster area **(H)** and fraction of cells with detectable clusters **(I)** of EML4-ALK-GFP in the presence or absence of LaG17-CluMPS. Box-and-whisker plots in (B,E,H) represent median and quartiles from approximately 20-50 cells per group. Vertical lines extend from quartiles ± 1.5*(interquartile range). Bar plots (**C,F,I**) represent frequency. Error bars = 95% confidence interval of approximately 20-50 cells per group

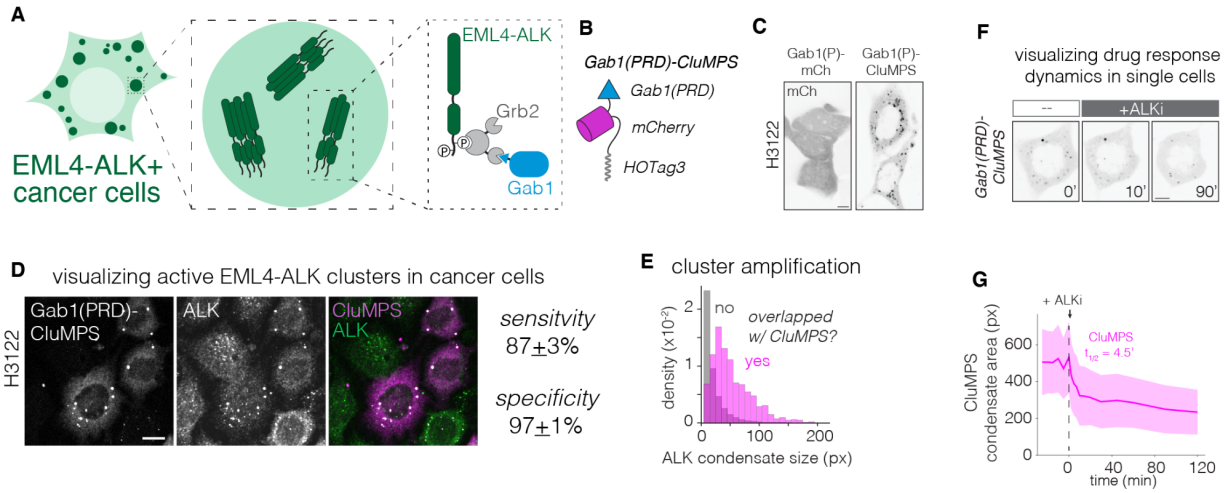


Figure 5. CluMPS tracks dynamics of endogenous protein clusters.
(A) EML4-ALK forms oncogenic protein condensates that drive oncogenic signaling. Condensation and signaling require the downstream adapter Grb2, which recruits additional adapters including Gab1. **(B)** A CluMPS reporter to visualize endogenous EML4-ALK condensates was developed using the proline rich domain (PRD) from Gab1, which binds an SH3 domain of Grb2 (Gab1(PRD)-CluMPS). **(C)** In H3122 cells, which endogenously express EML4-ALK, Gab1(PRD)-mCh appears diffuse (left), whereas Gab1(PRD)-CluMPS forms large condensates (right), indicating detection of protein clusters. Scale bar = 10 µm. **(D)** Immunofluorescence imaging of EML4-ALK and Gab1(PRD)-CluMPS in H3122 cells. Gab1(PRD)-CluMPS showed high sensitivity (89% cells with CluMPS condensate) and specificity (97% of CluMPS condensates that are colocalized with an ALK condensate). **(E)** CluMPS amplifies ALK clusters. Distribution of cluster sizes (pixels) for EML4-ALK condensates that colocalized with a CluMPS condensate (magenta) or not (grey). Data represents approximately 90,000 non-overlapped condensates and 1000 overlapped condensates. **(F)** Live cell imaging of CluMPS condensates before and during ALK inhibition with 1 µm crizotinib (see also **Supplementary Movie 4**). Scale bar = 10 µm. **(G)** Quantification of CluMPS dynamics. CluMPS condensate area remains stable for 30 minutes prior to drug treatment, then rapidly decreases upon treatment with ALKi. Half-lives derived from fitting data between 0 and 90 minutes to a single exponential decay. Trace represents mean. Ribbons = 95% CI of 31 cells.

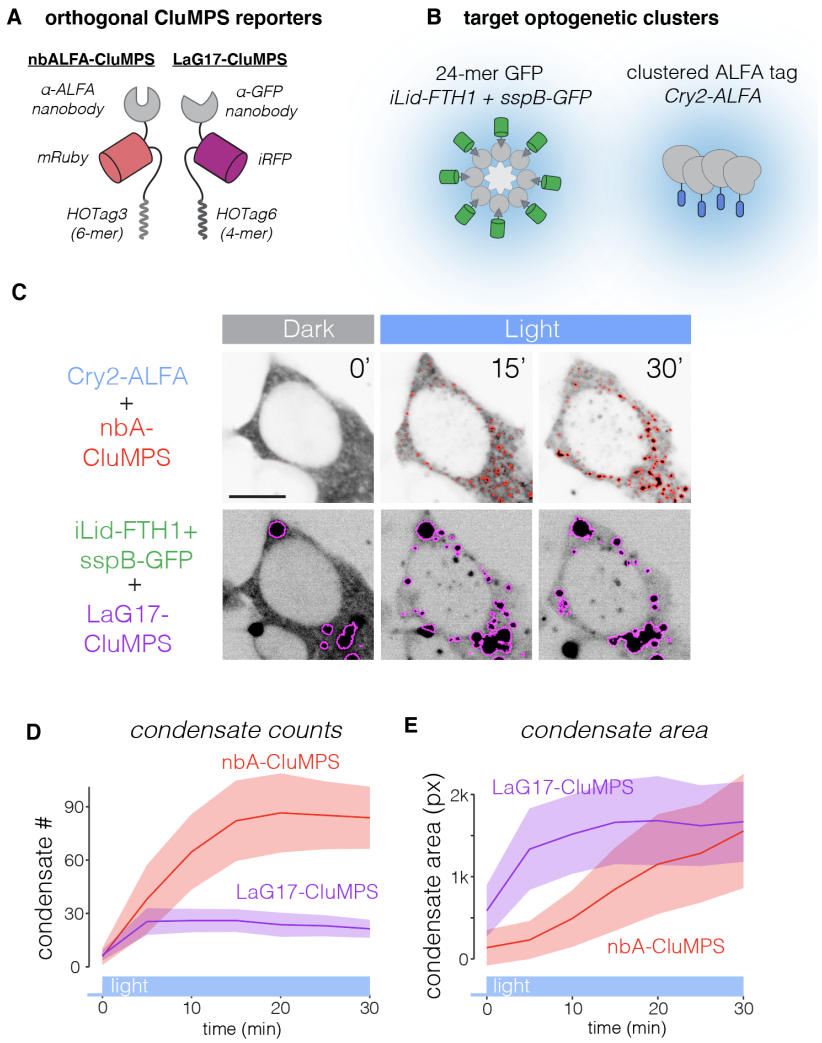
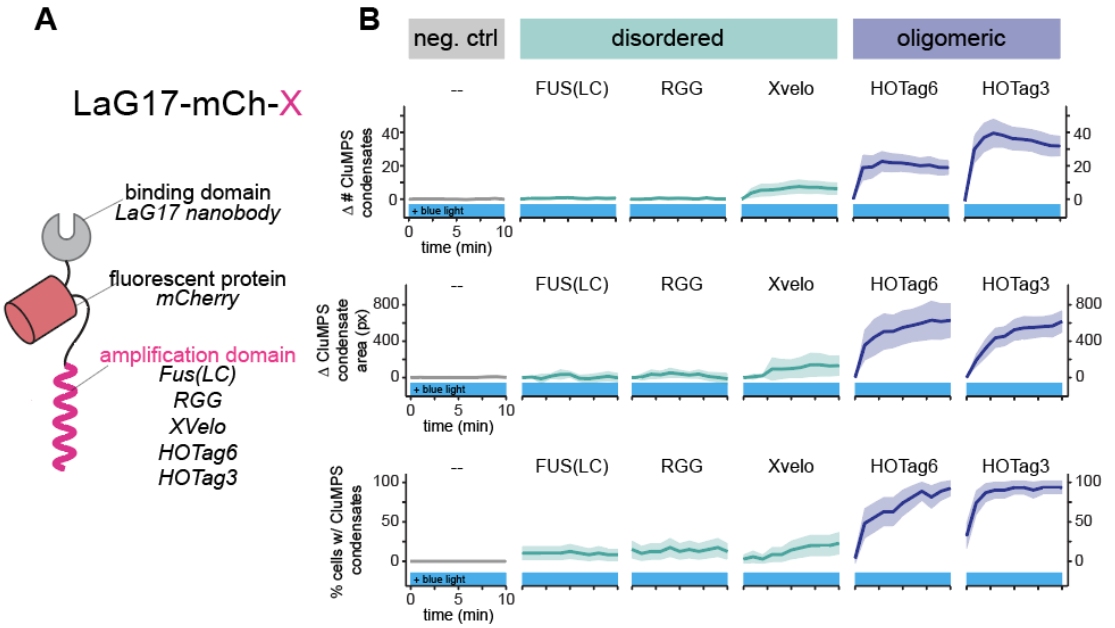


Figure 6. Multiplexed imaging of orthogonal CluMPS reporters.
(A) Orthogonal CluMPS reporters can be generated to visualize multiple targets in the same cell. **(B)** We generated two orthogonal optogenetic clustering modules as inducible targets for the multiplexable CluMPS reporters. Light induces GFP aggregation based on either optogenetic Cry2 clustering or induced dimerization of sspB-GFP to iLid fused to the 24-mer FTH1. (**C**) In cells expressing both targets and reporters, nbALFA-CluMPS is diffuse in the dark (top), while LaG17-CluMPS (bottom) exhibits a few pre-formed large droplets prior to light stimulation. Blue light stimulation of target clusters triggers non-overlapping CluMPS condensation in both channels, demonstrating independent detection of distinct clustered species. Quantification of number (**D**) and area (**E**) of nbALFA-CluMPS (red) and LaG17-CluMPS (magenta) clusters per cell over upon blue light stimulation. Data represents mean, ribbons = 95% CI of 18 cells that expressed all components. The nbALFA-CluMPS forms more clusters but over a longer period of time, whereas LaG17-CluMPS forms fewer clusters but faster, further demonstrating that the reporters are amplifying distinct underlying clusters.

SUPPLEMENTARY FIGURES



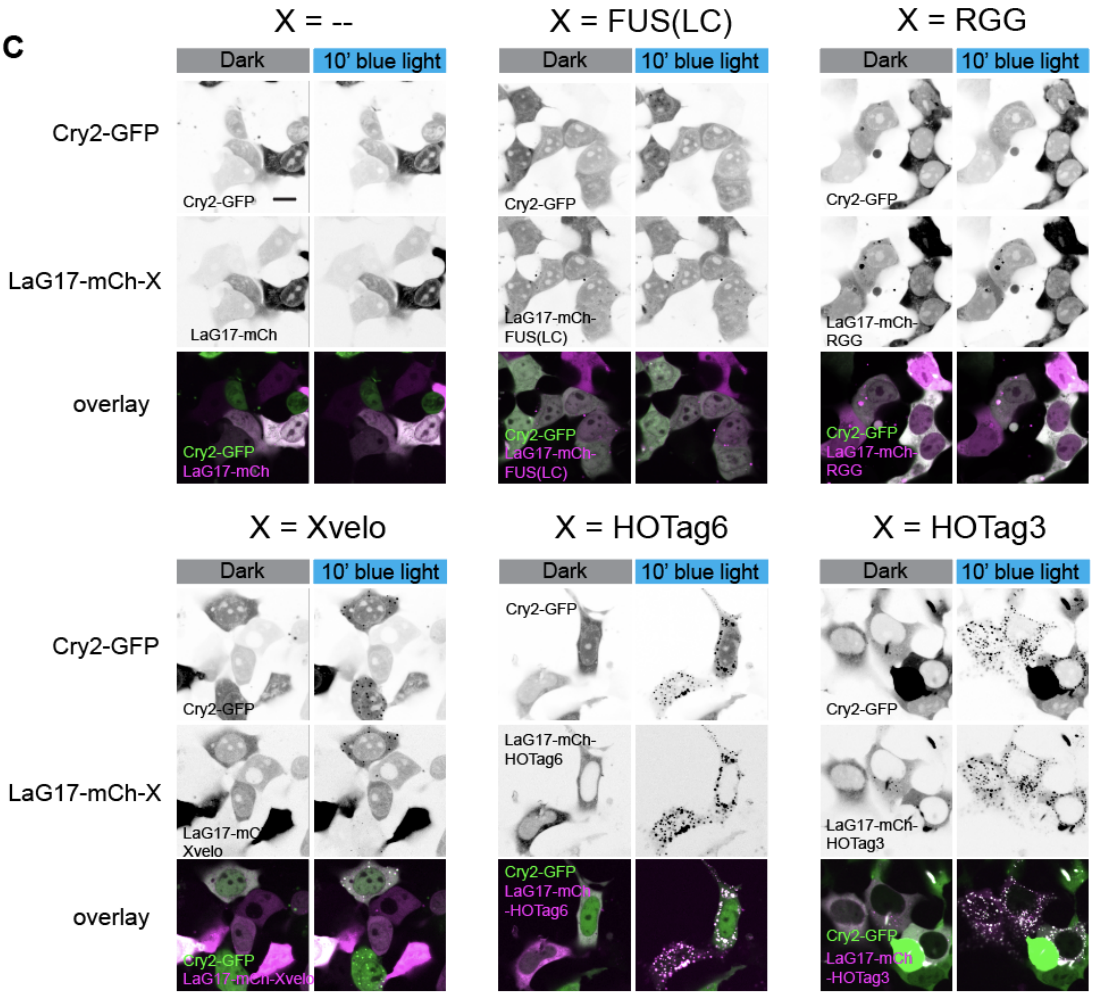


Figure S1. Comparing candidate CluMPS amplification domains. A) Design of candidate CluMPS probes. CluMPS constructs comprise an N-terminal target-binding domain, a fluorescent protein, and a C-terminal “amplification domain” that can drive phase separation when clustered through binding to a clustered protein. Tested domains include the intrinsically disordered regions (IDRs) FUS(LC) (1-163), RGG, and Xvelo, and the homo-oligomeric tags HOTags 6 (tetrameric) and 3 (hexameric). A construct with no amplification domain was tested as a negative control. **B)** Quantification of CluMPS condensate formation in cells cotransfected with Cry2-GFP and a CluMPS variant and stimulated with blue light over 10 minutes to trigger oligomerization of Cry2-GFP. CluMPS constructs (left-right) are LaG17-mCh, LaG17-mCh-FUS-LC, LaG17-mCh-RGG, LaG17-mCh-Xvelo, LaG17-mCh-HOTag6, and LaG17-mCh-HOTag3. Change in number of condensates per cell (top), change in area of clusters per cell (px, middle), and percentage of cells with clusters (defined as 5+ clusters AND 200+ pixels of cluster area (bottom) are shown for all constructs. Data represents mean ± 95% c.i. for approximately 20-50 cells per group. **C)** Representative images of all CluMPS variants co-expressed with Cry2-GFP in the dark and after 10 minutes of blue light stimulation. Composite images overlay Cry2-GFP (green) and CluMPS variant (magenta) images. Scale bar = 20 μm. In sum, CluMPS constructs with oligomeric amplification domains (HOTags) phase separate strongly with clustered Cry2-GFP, while constructs with (IDRs) do not.

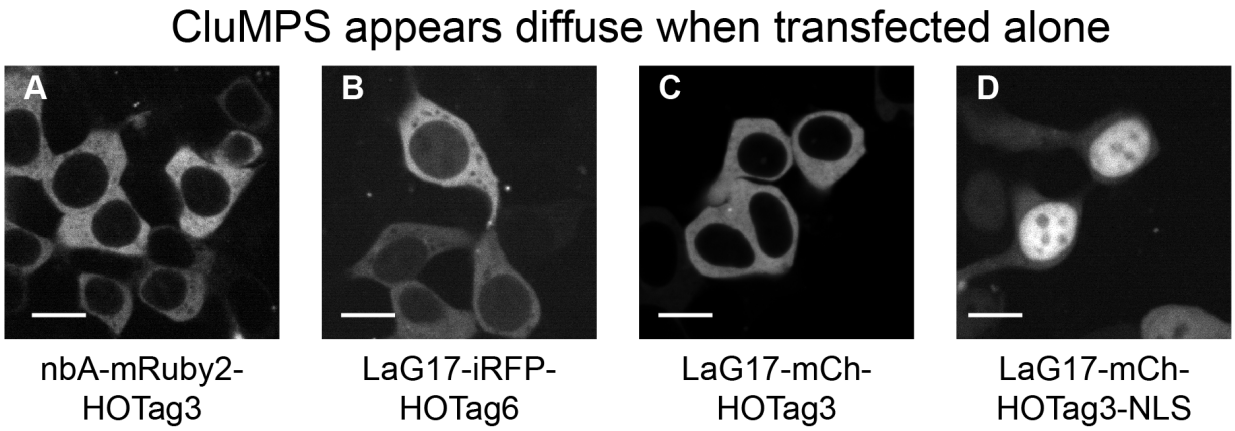


Figure S2. CluMPS appears diffuse when transfected alone. The multiple GFP- and ALFAtag-targeting CluMPS reporters generated for this work always appear diffuse when expressed alone in cell

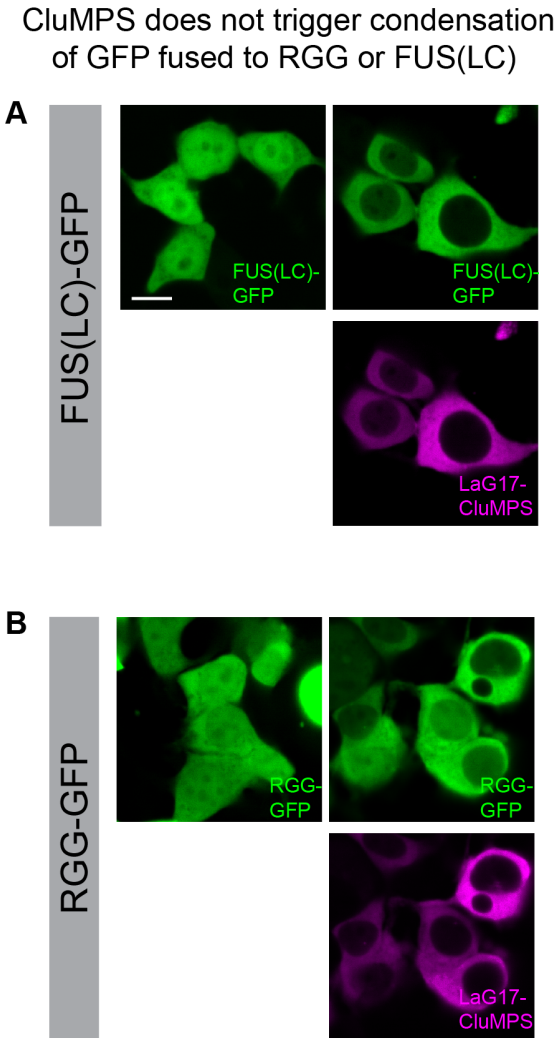


Figure S3. LaG17-CluMPS does not trigger condensation of GFP fused to the IDRs FUS(LC) or RGG. The lack of condensation of CluMPS when co-expressed with its target FUS(LC)-GFP or RGG-GFP suggests that CluMPS activation is specific to *bona fide* pre-existing clusters and does not give 'false positive' results by inducing clusters/condensates of a target where none existed. Note the altered nuclear localization of FUS(LC)-GFP and RGG-GFP in the presence of CluMPS, which confirms binding of CluMPS to its target despite lack of condensation. Scale bar = 20 μm.

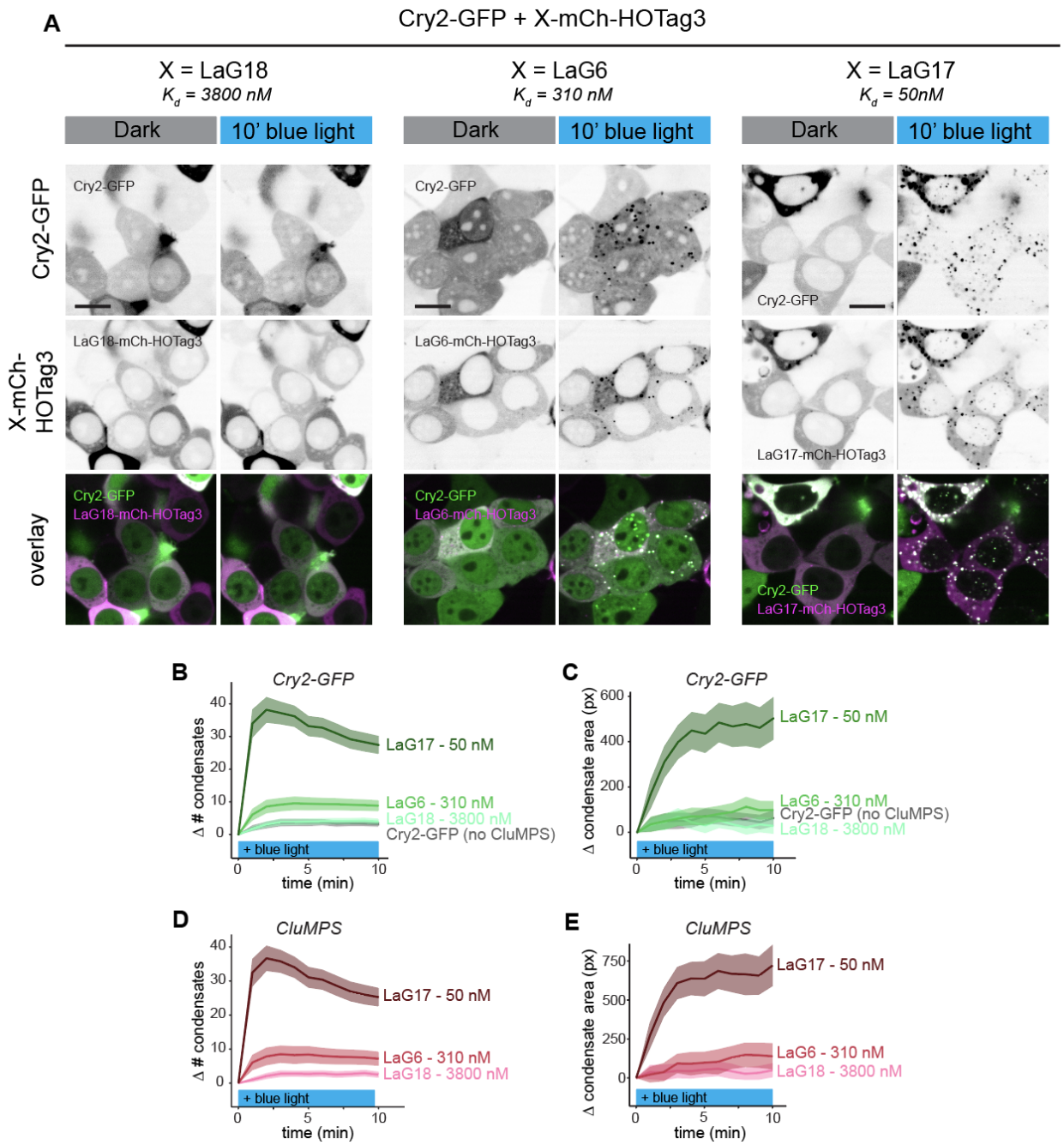


Figure S4. CluMPS activity depends on the affinity of the target binding domain. To test the effect of binding domain affinity on CluMPS activation, we co-expressed Cry2-GFP with CluMPS variants with binding domains with differing affinities for GFP. Binding domains used were LaG18 (K_d = 3800 nM), LaG6(K_d = 310 nM) and LaG17 (K_d = 50 nM). **A)** Representative images of condensation of both Cry2-GFP and CluMPS both before and after 10 min of stimulation with blue light. Scale bars = 20 µm. Quantification of condensate number (**B,D**) and area (**C,E**) in both the GFP (**B,C)** and miRFP (**D,E**) shows that stronger binding between CluMPS and target yields stronger condensation. Each trace represents mean. Ribbons = 95% CI for approximately 100-200 cells per group.

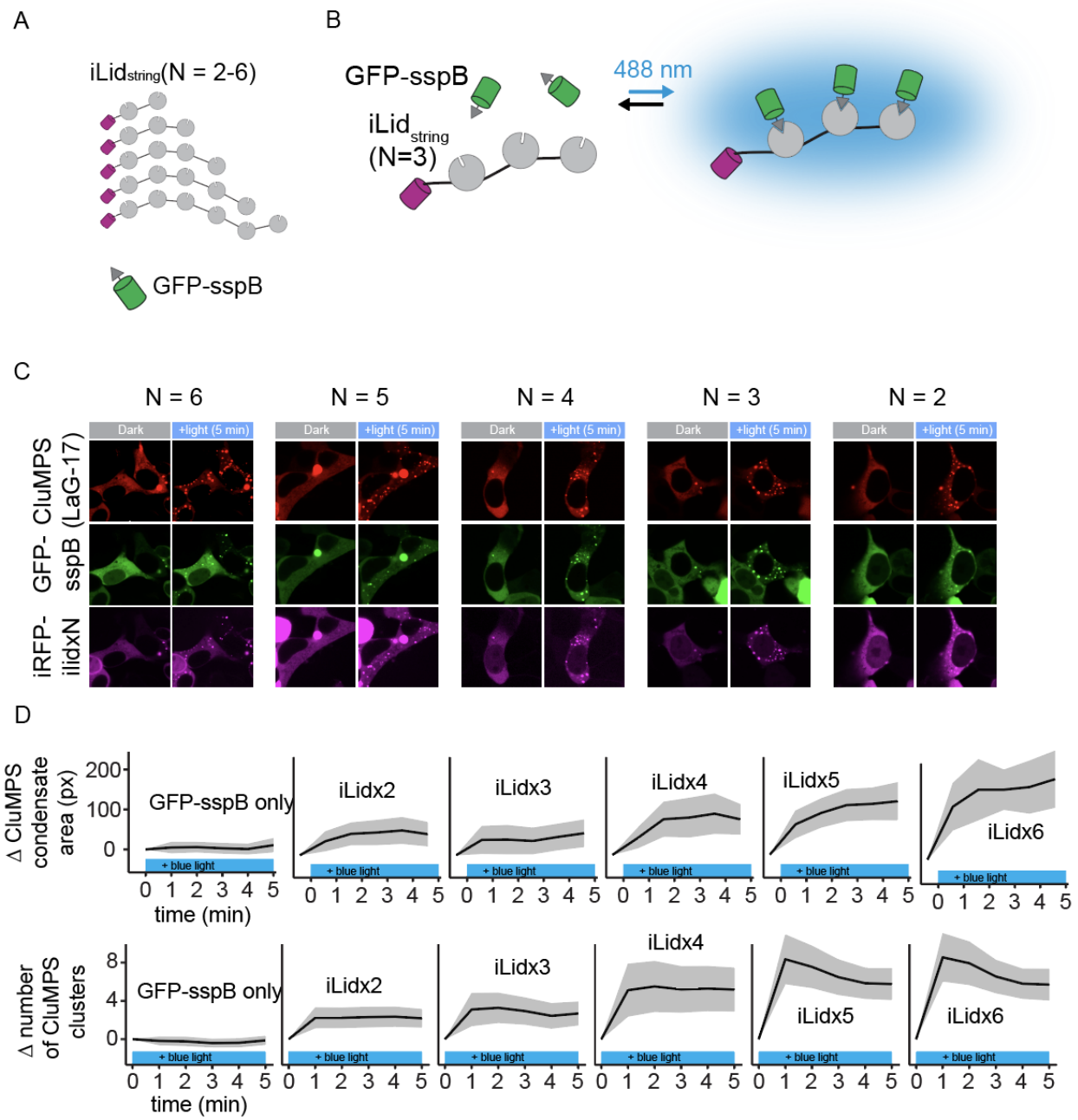


Figure S5. iLid$_{strings}$ and GFP-sspB form defined multimers upon blue light stimulation, providing a platform for characterizing CluMPS sensitivity to small multimers. A) Multiple iLids on a single peptide, termed iLid$_{strings}$, span between two and six iLids, determining the number of GFP-sspB monomers which are recruited to form a small oligomer. **B)** GFP-sspB is recruited to iLid$_{string}$ upon blue light exposure. **C)** CluMPS condensation is possible with iLid$_{string}$ (length 2-6) and GFP-sspB upon blue light stimulation under ideal conditions (Figure S7, S9). **D)** Change in CluMPS condensate area (pixels, top) and number of CluMPS condensates after blue light stimulation for GFP-sspB only, and GFP-sspB expressed with iLid$_{string}$ length 2-6. Data represents mean traces of 150-200 cells per condition, shading shows 95% CI.

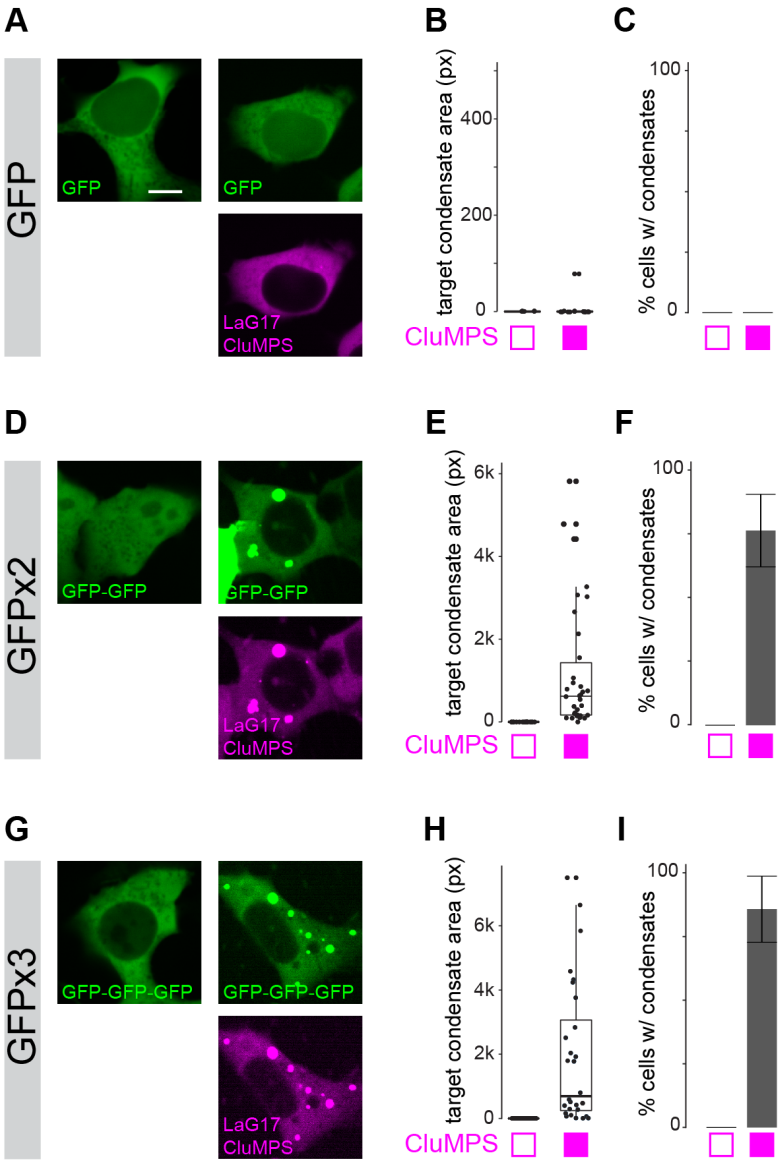


Figure S6. Tandem GFP trimer and dimer, but not GFP monomer, triggers CluMPS condensation. Representative images of GFP(**A**), GFP-GFP (**D**), or GFP-GFP-GFP (**G**) transfected in the presence or absence of LaG17-mCh-HOTag3 (LaG17-CluMPS). Scale bar = 20 µm. Quantification of GFP condensate area (**B,E,H**) and % of cells with GFP condensates (**C,F,I**) shows CluMPS-induced condensation with the multimeric GFPs, but not with monomeric GFP. Box-and-whisker plots in (B,E,H) represent median and quartiles from approximately 20-30 cells per group. Vertical lines extend from quartiles ± 1.5*(interquartile range). Bar plots (**C,F,I**) represent frequency. Error bars = 95% confidence interval of approximately 20-30 cells per group

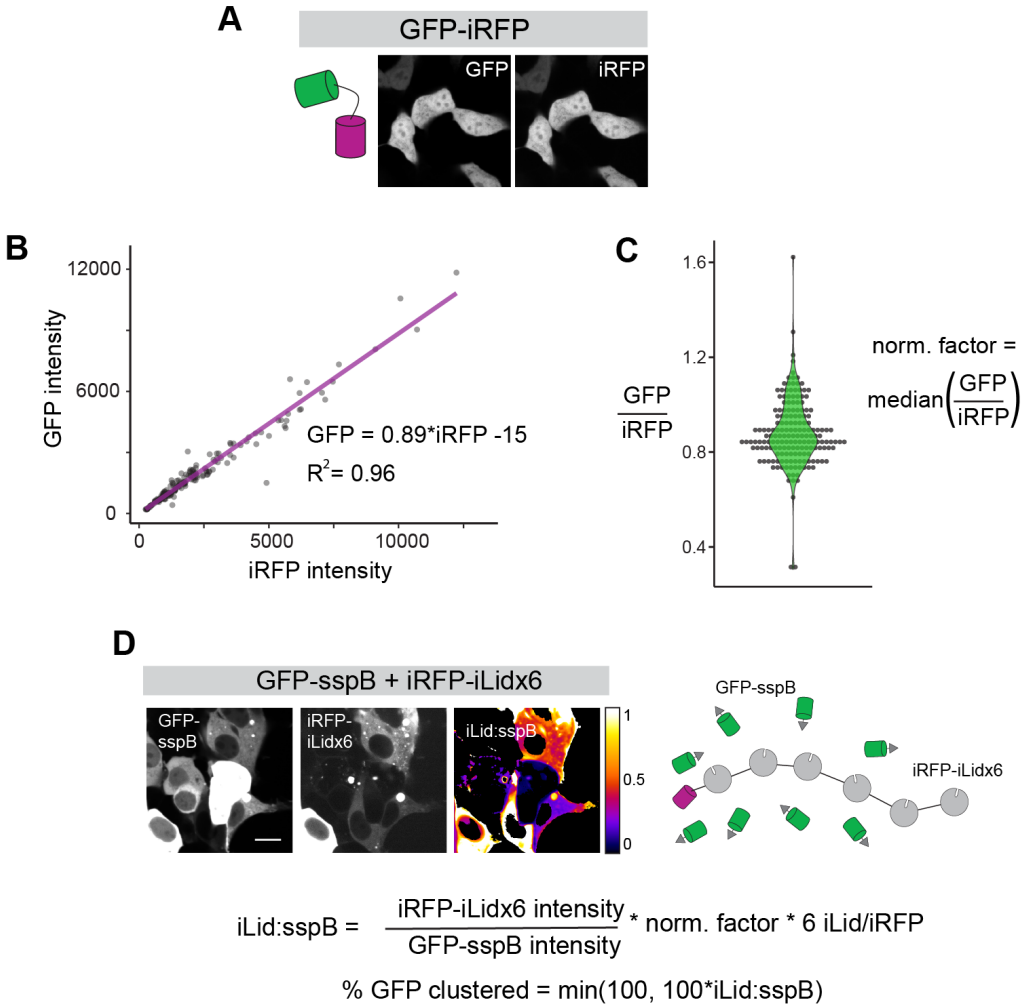


Figure S7. Using GFP-iRFP as a control for determining stoichiometric ratio of GFP-sspB and iRFP-iLid$_{string}$. The relative stoichiometries of GFP-sspB and iRFP-iLid$_{string}$ were determined by separately imaging a direct fusion of GFP-iRFP, which expressed both fluorophores in a 1:1 ratio. **A**) Representative images of EGFP-miRFP670 (GFP-iRFP) in HEK 293T. Scale bar 20 µm. **B**) Plot of GFP intensity vs iRFP intensity shows a linear relationship (R^2 = 0.96) between expression of the two fluorophores. Note that linear agreement was obtained only when supplementing cells with exogenous iRFP cofactor biliverdin (see **Methods** for details). **C**) Ratio of GFP to iRFP intensity for individual cells with overlaid violin plot. The normalization factor used in subsequent calculations is the median GFP:iRFP ratio. **D**) Representative image of GFP-sspB (left), iRFP-iLidx6 (middle), and iLid:sspB ratio pseudo-image (right), and corresponding cartoon showing GFP-sspB and iRFP-iLidx6. Pseudo-image generated according to the equation shown: iRFP intensities are divided by GFP intensities, normalized with the GFP-iRFP normalization factor from (**C**), and multiplied by the number of iLids per iRFP on each string (in this case, 6). The percentage of GFP clustered is the percentage of GFP-sspB molecules that have iLid sites to bind on the multimer and is calculated by multiplying the iLid:sspB ratio by 100 (calculated values over 100 are considered 100% clustered). Note that although the pseudoimage shows pixel-level iLid:sspB values, the iLid:sspB ratio and corresponding % target clustered in the manuscript are cell-level metrics that were calculated from the median values of the GFP and iRFP channels.

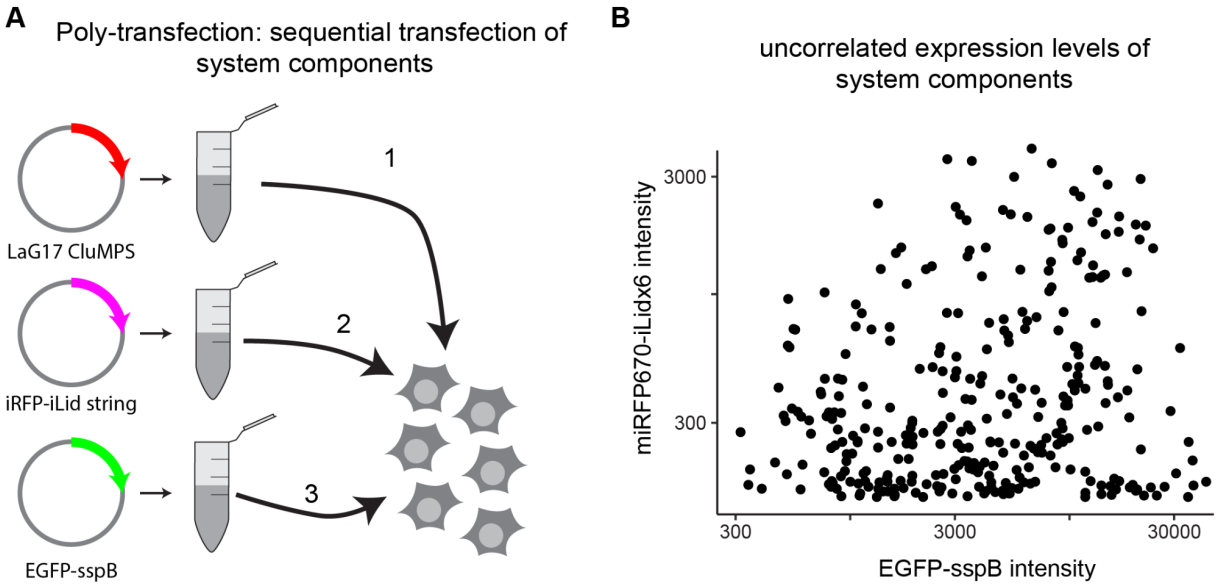


Figure S8. Poly-transfection allows transfection of multiple components with low expression correlation and allows sampling of a wide range of iLid:sspB ratios. A) Cartoon depicting poly-transfection[69]. Plasmids were separately mixed with transfection reagent and sequentially added to cells so that their subsequent expression would be largely uncorrelated. **B**) Plot of miRFP670 intensity vs EGFP intensity after poly-transfection shows no correlation between iRFP and GFP levels.

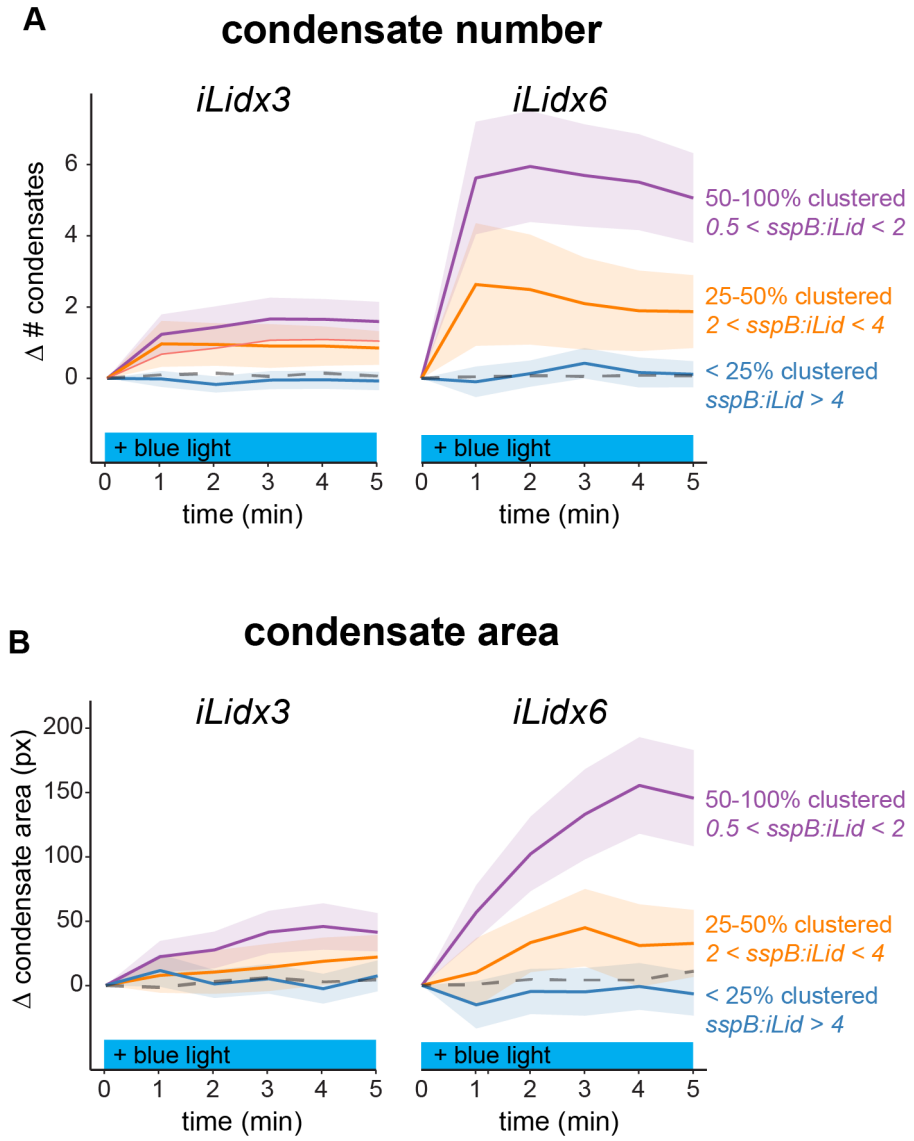


Figure S9. CluMPS activation depends on both cluster size and the fraction of target that exists in a cluster. Condensation was quantified for HEK 293T cells transfected with LaG17-CluMPS, GFP-sspB, and iLid strings (of size 3 or 6). The number (**A**) and size (**B**) of condensates was measured during 10 min of blue light stimulation. Condensation was stronger for cells co-transfected with iLidx6, demonstrating how larger cluster sizes promote stronger CluMPS activation. However, for both target cluster sizes, condensation also depended on the estimated number of sspB monomers in a cluster, which was estimated based on the stoichiometry of sspB and iLid in single cells. Dotted gray line shows change in number of clusters after blue light stimulation in cells from the same pool that were transfected with CluMPS and GFP-sspB but not with iLid string, and therefore cannot form light-induced clusters. Each trace represents mean. Ribbons = 95% CI for approximately 50-200 cells per group.

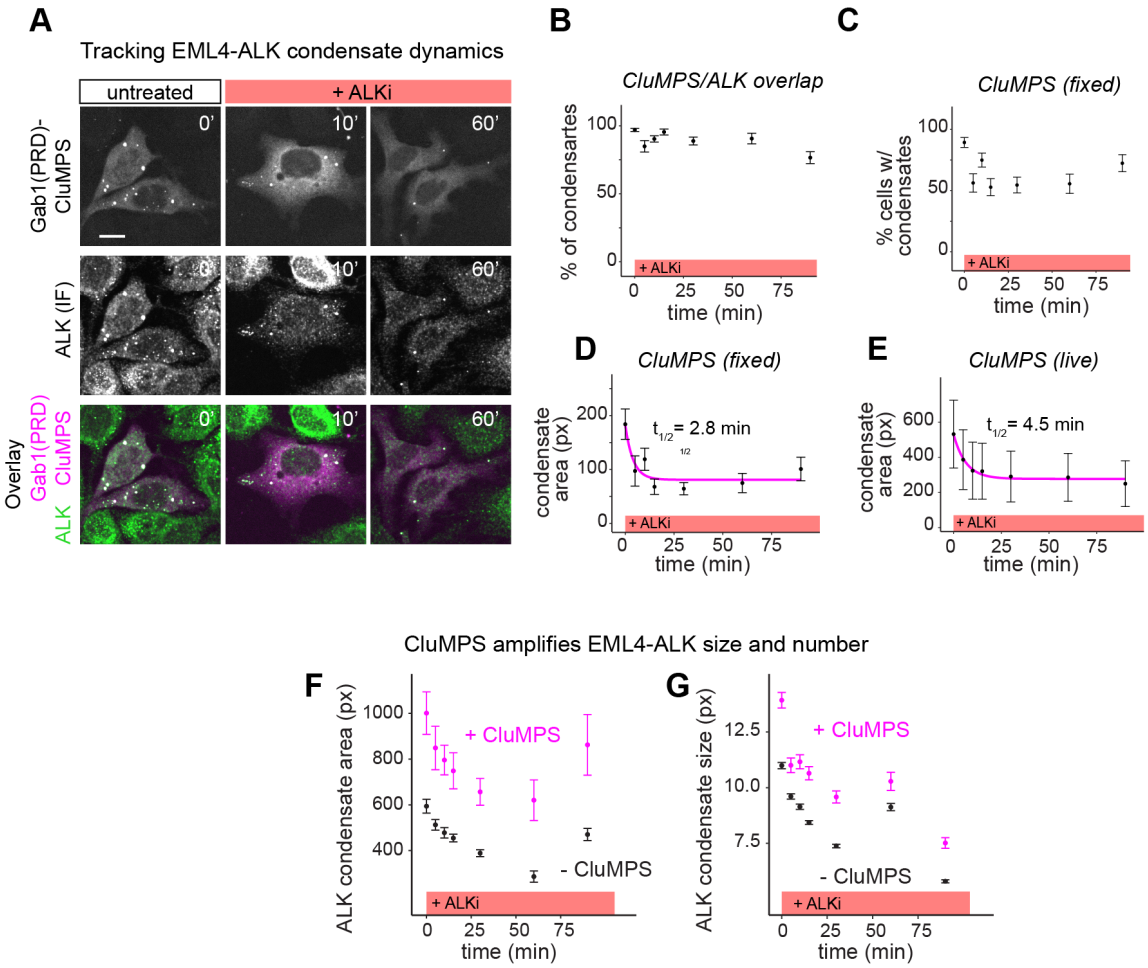


Figure S10. Gab1(PRD)-CluMPS amplifies EML4-ALK condensates and visualizes their dynamics in H3122 cancer cells. A) Representative immunofluorescence images of H3122 cells expressing Gab1(PRD)-CluMPS before and during treatment with 1 μM crizotinib. Scale bar = 20 μm. **B)** Percentage of all CluMPS condensates that overlapped with ALK (defined as having >10% overlap of pixels identified as both a CluMPS and ALK condensate) at each time point of drug treatment. Data represents percentage, error bars = 95% CI of between 200-1000 condensates per time point. **C)** Percentage of cells at each time point containing CluMPS condensates (defined as having >1 cluster and >25 total pixels of CluMPS condensate area). Data represents percentage, error bars = 95% CI of between 150-250 cells per time point. **D)** Area of CluMPS condensates (pixels per cell) in fixed cells after different durations of drug treatment. Exponential decay fit overlayed in magenta. Data represents mean, error bars = 95% CI of between 150-250 cells per time point. E) Area of CluMPS condensates (pixels per cell) from live cell imaging of H3122 cells treated with 1 μM crizotinib, reproduced from **Figure 5E** and presented here for side-by-side comparison with fixed cells. Exponential decay fit overlayed in magenta. Data represents mean, error bars = 95% CI of 31 cell traces. **F,G)** Comparison of ALK condensate area per cell and ALK condensate mean size in cells with and without CluMPS demonstrates that CluMPS amplifies EML4-ALK condensates. Data in **(F)** represents mean, error bars = 95% CI of between 150-250 cells with CluMPS and 1200-1800 cells without CluMPS. Data in **(G)** represent mean, error bars = 95% CI for approximately 40,000-110,000 clusters per group in cells without CluMPS, and 9000-18000 clusters per group in cells transfected with CluMPS.

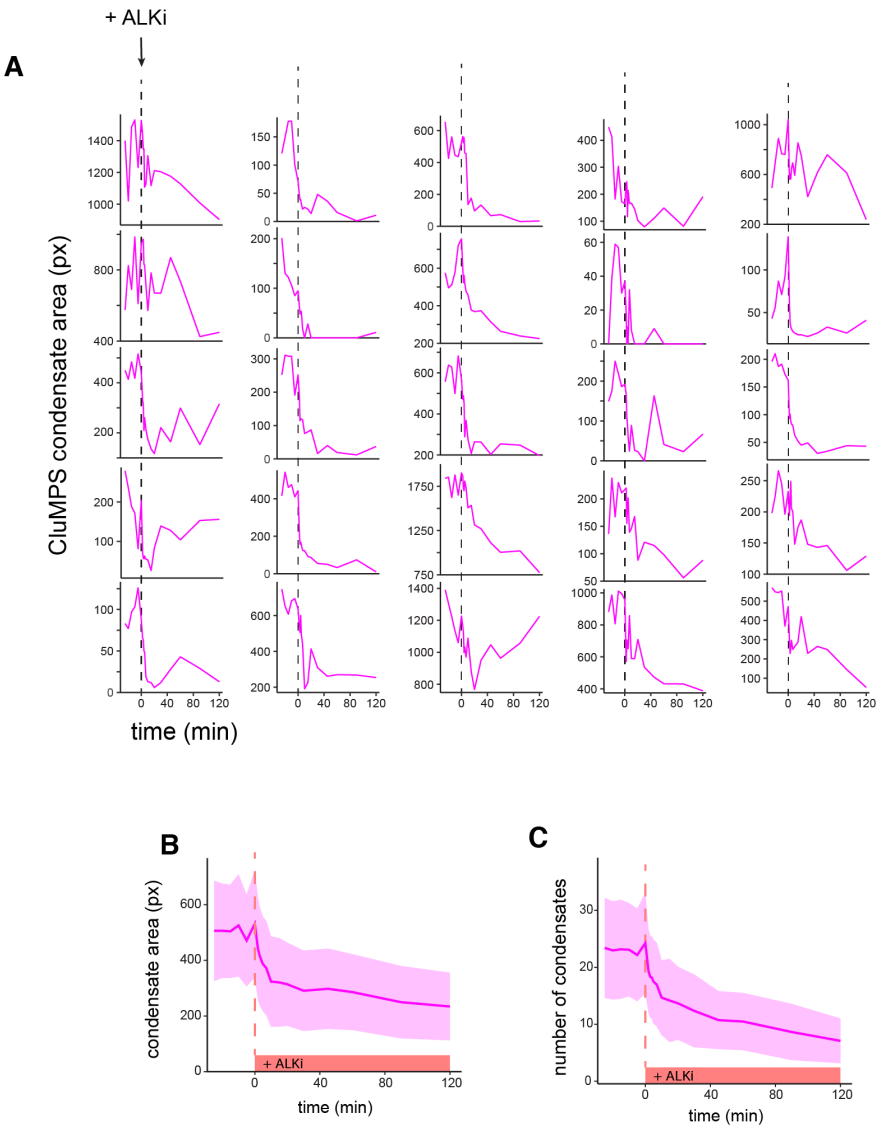


Figure S11. Single-cell visualization of Gab1(PRD)-CluMPS dynamics through live-cell imaging of H3122 cells. A) Single-cell traces of multiplexed CluMPS in response to ALK inhibitor (1 μM crizotinib). Dotted line indicates when drug was added. **B,C**) Average CluMPS condensate area and number in H3122 cells after crizotinib treatment. **(B)** is reproduced from **Figure 5E**.Traces represents mean, ribbon = 95% CI of 31 cell traces.

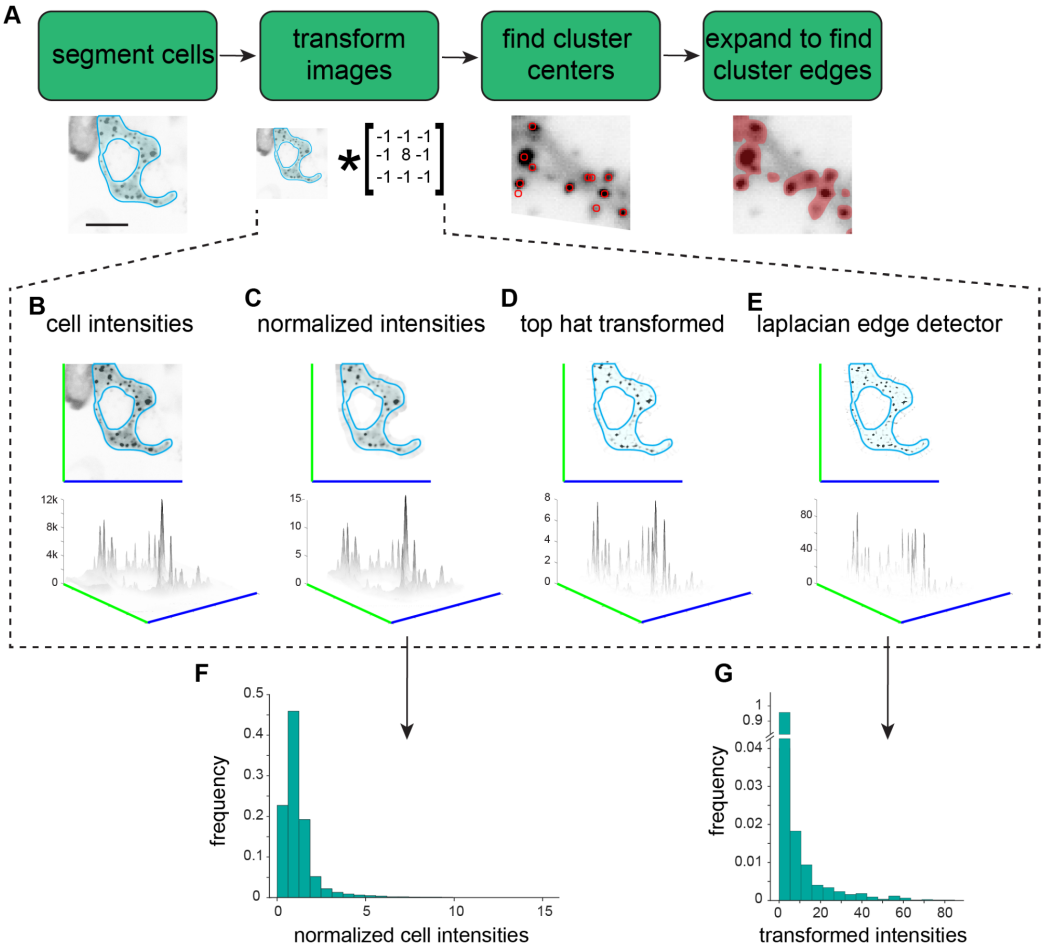


Figure S12. Fitting clusters in single cells with a custom MATLAB script: flow chart and image transformation. A custom MATLAB algorithm was developed to segment cells and identify clusters (see **Methods** and associated code files for more details). **A)** Flow chart showing basic steps of cluster fitting. First, single cells were segmented, and cell intensities were transformed to normalize them, suppress background, and enhance clusters (see **C-E** below). This transformed image was used to identify points at the center of clusters, and finally a stepwise algorithm expanded from each cluster center to determine the pixels within each cluster. Further details on image transformation: After segmentation **(B)**, image intensities were normalized within a cell mask by subtracting background and dividing by the median. **C)** A Top hat transform suppressed all features on longer length scales than desired, effectively removing background fluorescence within the cell (non-clusters go to zero). **D)** Next, a laplacian edge detector transformation enhanced contrast and further amplified the intensities of areas of high contrast (clusters). **E)** A histogram of intensities after the normalization stage shows most pixels around 1, and the brightest pixels between 3 and 15. **F)** By contrast, a histogram of final transformed values shows >95% of pixels ~0, and the brightest pixels are 15-80, thus enhancing cluster contrast. In sum, this procedure successfully suppressed background fluorescence while enhancing clusters, enabling robust, semi-automated detection of fluorescent clusters.

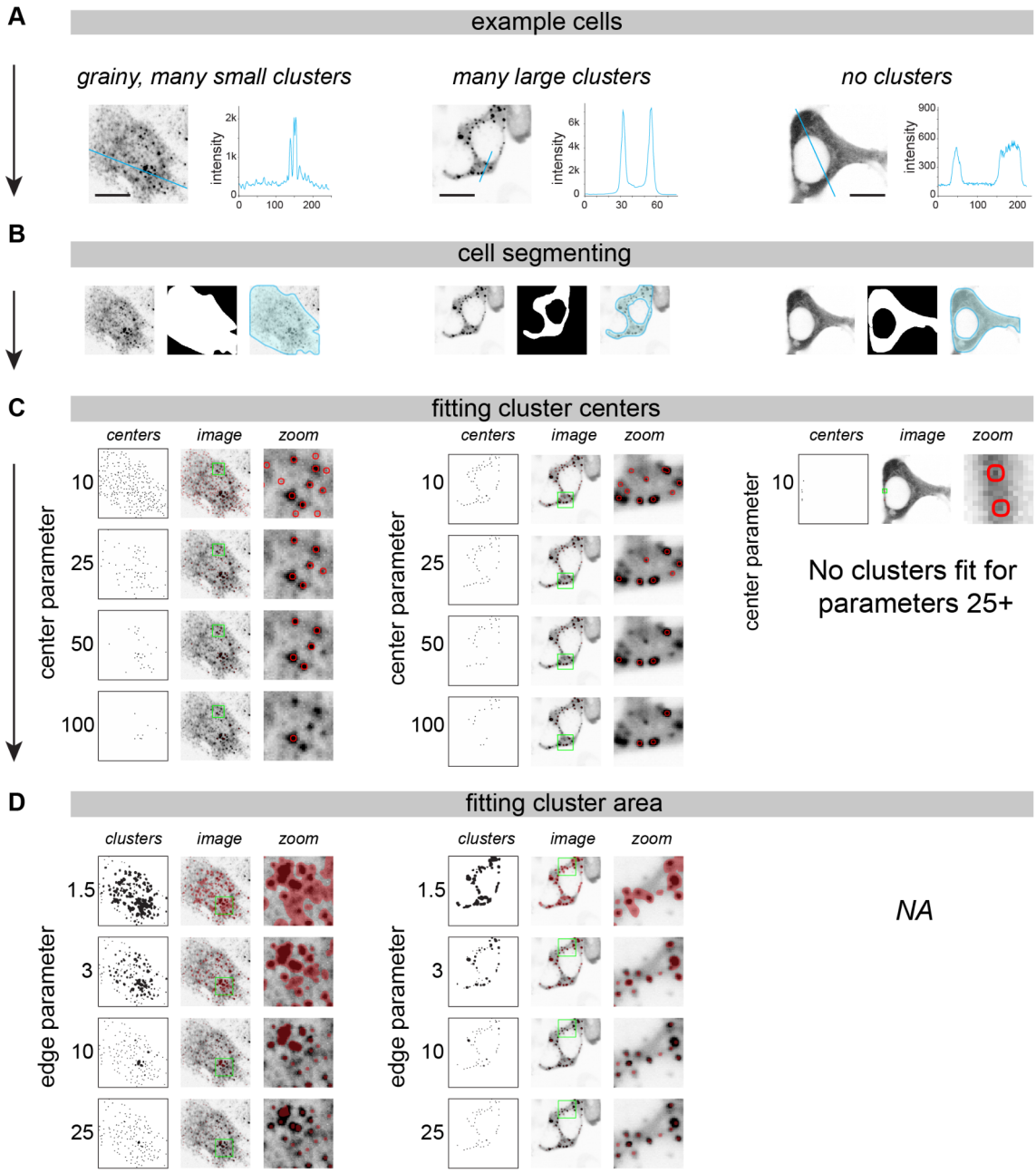


Figure S13. Fitting clusters in single cells with a custom MATLAB script: two parameters determine how clusters are fit. A) Cluster fitting is depicted for three classes of clustered cells. Left, a cell with grainy fluorescence and small clusters. Middle, a cell with many clearly-defined clusters. Right, a cell with no clusters. **B)** Cell masks are generated from initial cell segmentation (see **Methods** for more details). **C)** The first fitting parameter (center parameter) determines how strictly cluster centers are fit (see **Methods**), with higher values being more strict and resulting in fewer identified clusters. **D)** The second (edge) parameter determines how strictly the edges of clusters are selected, with higher values being more strict and resulting in smaller cluster areas.

METHODS

Cell lines and cell culture

All cell lines were maintained in standard cell culture incubators at 37°C and 5% CO_2. Lenti X HEK 293T cells were cultured in DMEM containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S). H3122 cells were cultured in RPMI-1640 media containing L-glutamine with 10% FBS and 1% P/S. For experiments, cells were seeded with P/S-free media in 96 or 384 well plates coated with 10 µg/mL MilliporeSigma™ 597 Chemicon™ Human Plasma Fibronectin Purified Protein in PBS. For 384 well plate experiments, 3000 HEK 293T cells or H3122 cells were seeded in 50 µL of cell culture media in each well. For 96 well plate experiments, 20,000 HEK 293T or H3122 cells were seeded in 200 µL of cell culture media in each well.

Plasmid design and assembly

Constructs were assembled using Gibson assembly. DNA fragments and backbones were generated via PCR and inserted into the backbone via HiFi cloning mix (New England Biolabs). DNA sequences encoding GFP binding nanobodies (LaGs) were a kind gift from Dr. Michael P. Rout [65]. HOTag6 and HOTag3 were obtained from pcDNA3-ERK-SPARK, which was a gift from Xiaokun Shu (Addgene plasmid # 106921; http://n2t.net/addgene:106921 ; RRID:Addgene_106921) [60]. FUS(LC) (1-163), Xvelo, and LAF-1 RGG were kindly provided by Matthew C. Good. iLid strings were adapted from mCherry-iLidx6, which was a gift from Takanari Inoue (Addgene plasmid # 103779; http://n2t.net/addgene:103779 ; RRID:Addgene_103779) [58]. GFP-TDP-43 was kindly provided by Ophir Shalem. SARS-CoV2-N protein sequence was sourced from pLVX-EF1alpha-SARS-CoV-2-N-2xStrep-IRES-Puro, which was a kind gift from Nevan Krogan (Addgene plasmid # 141391 ; http://n2t.net/addgene:141391 ; RRID:Addgene_141391). EML4-ALK-GFP was made through fusing EML4-ALK(V1) to GFP as previously described [79]. FTH1 was sourced from pCMV-SPORT6_FTH1, obtained from the High-Throughput Screening (HTS) Core at Penn Medicine. The sequence encoding ALFA tag was synthesized as a single strand DNA oligomer (Genewiz from Azenta Life Sciences), and sequence encoding the ALFA binding nanobody was synthesized as a gBlock gene fragment (Integrated DNA Technologies) [81]. Gab1 (PRD) was sourced from the human GAB1 cDNA sequence obtained from the High-Throughput Screening (HTS) Core at Penn Medicine. PCSDest vectors for RNA synthesis were obtained from the Zebrafish Core at Children's Hospital of Philadelphia. miRFP670, used in $iLid_{strings}$ and LaG17-miRFP670-HOTag3, was adapted from pCSII-EF-miRFP670v1-hGem(1/110), which was a gift from Vladislav Verkhusha (Addgene plasmid # 80006 ; http://n2t.net/addgene:80006 ; RRID:Addgene_80006).

RNA synthesis

RNA transfection was used instead of DNA transfection for cell types that showed low efficiency of plasmid DNA transfection. To generate RNA from the PCSDest expression plasmid, the plasmid DNA was linearized with NotI and used as a template for RNA synthesis. RNA was synthesized using the mMESSAGE mMACHINE™ (ThermoFisher Scientific) transcription kit with SP6 RNA polymerase. RNA was spin-column purified using RNA Clean and Concentrator kit (Zymo Research).

Plasmid and RNA transfection

HEK 293 T cells were transfected with DNA plasmids using Lipofectamine™ 3000 Transfection Reagent (ThermoFisher Scientific) per manufacturer's protocol. Final transfection mixture was 1-30 ng μL^{-1} DNA, 2% Lipofectamine™ reagent, 2% P3000 reagent and was brought to final volume with Opti-MEM™ (ThermoFisher Scientific). Cells seeded in 384 well plates received between 1 and 5 µL of transfection mixture, and cells in 96 well plates received 10 µL of transfection mixture. Cells were imaged 24-60 hr after transfection. H3122 cells were transfected

with RNA using Lipofectamine™ MessengerMAX™ Transfection Reagent (ThermoFisher Scientific), as RNA transfection was found to be more efficient than DNA transfection (described above) in this cell type. Final transfection mixture was 10 ng μL^{-1} RNA, 2% Lipofectamine™ MessengerMAX™ reagent, and was brought to final volume with Opti-MEM™. Cells were imaged 12-36 hr after RNA transfection.

Live cell imaging

Live-cell imaging was done using a Nikon Ti2E microscope equipped with a Yokagawa CSU-W1 spinning disk, 405/488/561/640nm laser lines, an sCMOS camera (Photometrics), a motorized stage and an environmental chamber (Okolabs). HEK 293T cells were imaged with a 40X oil immersion objective at 37°C and 5% CO_2. In addition to exposure for imaging, experiments requiring optogenetic stimulation of Cry2 and/or iLid-sspB received an additional 500 ms per field of view per minute blue light exposure using the 488 nm laser line. Experiments where HEK293T cells were transfected with a construct containing miRFP670 had the media supplemented with 25 μM biliverdin hydrochloride (Sigma-Aldrich, product number 30891) 24 hours prior to imaging. H3122 cells were imaged with a 40X oil immersion objective. H3122 cells were starved overnight with RPMI-1640 media containing L-glutamine and 1% P/S using a BioTEK 405 plate washer. For live nuclear staining, H3122 cells were treated with 5 μg/ml Hoechst 33342 for 15 min before imaging. Cells were imaged for 30 minutes, then treated with 1μM crizotinib (Sigma-Aldrich, PZ0191) and imaged for 2 hours.

Immunofluorescence staining

Immediately after treatment, cells were fixed and stained as previously described [84]. Briefly, cells were fixed for 10 min in 4% paraformaldehyde. Cells were permeabilized using PBS with 0.5% Triton-X100 for 10 min. Cells were further permeabilized using 100% methanol at -20°C for 10 min and samples were blocked with 1% bovine serum albumin (Fisher, BP9706100) in PBS + 1% BSA for 1 hour at room temperature. Cells were then incubated in primary antibody diluted in PBS + 1% BSA for 2 hours at room temperature or overnight at 4 °C (anti-ALK (D5F3), Cell Signaling, catalog number 3633, 1:400). Primary antibody was then washed with 0.1% Tween-20 in PBS (PBS-T). Cells were then incubated in secondary antibody diluted in PBS + 1%BSA as recommended (IgG (H+L) Cross-Adsorbed Goat anti-Rabbit, DyLight™ 488, Invitrogen, catalog number 35553, 1:500) and 4,6-diamidino-2-phenylindole (DAPI; ThermoFisher Scientific, catalog number D1306, 300 nM) for 1 hour at room temperature and washed with PBS-T. Cells were imaged in 50 μL of PBS-T.

iLid strings experimental procedures

For testing CluMPS sensitivity with small multimers using $iLid_{strings}$ (multiple iLid domains fused on a single peptide, iLidxN: N = 2:6 iLids fused to miRFP670) transfected with GFP-sspB, it was important to achieve a broad range of transfection conditions in order to sample a range of sspB:iLid ratios across cells (**Figure 3G-I**). For these transfections, plasmids were 'poly-transfected'; each plasmid was mixed with transfection reagents in a separate tube from the other plasmids and added to cells separately (**Supplementary Figure 8A**)[69]. This resulted in uncorrelated expression between components, permitting sampling of a wide range of sspB:iLid ratio across cells (**Supplementary Figure 8B**).

$iLid_{string}$ analysis

To determine the stoichiometric ratio of sspB:iLid, which sets the fraction of EGFP-sspB that could bind sites on iLid multimers vs the fraction that remains monomeric, we used fusion of EGFP-miRFP670 as a control (**Supplementary Figure 4**). For each $iLid_{string}$ experiment, a separate well on the same plate was transfected with this EGFP-miRFP670 control, and imaged with the same microscope parameters as the wells containing EGFP-sspB, miRFP670-$iLid_{string}$,

and LaG17-CluMPS. With this control, the ratio EGFP:miRFP670 intensity in cells transfected with EGFP-miRFP670, in which the number of EGFP and miRFP670 proteins are assumed to be equal, could be calculated for many cells (**Supplementary Figure 4**). The median ratio is then used as a normalization factor to determine the stoichiometric ratio of EGFP:miRFP670 in cells transfected with EGFP-sspB and miRFP670-iLid$_{string}$ (**Supplementary Figure 4**). With this, it is trivial to calculate the sspB:iLid ratio given that EGFP-sspB are 1:1 and miRFP670:iLid is known and dependent on the valency of the specific iLid$_{string}$ that was transfected. The percentage of GFP-sspB that was able to participate in clustering can then be estimated as min(100, 100*(iLid/sspB)).

Cell segmentation

Cells were segmented using either CellProfiler for automated segmentation or by hand with a custom MATLAB script. For CellProfiler segmentation, nuclei were identified. Then, cytoplasmic fluorescence channels were gaussian and median filtered to prevent artifacts from segmenting into clustered cells, and nuclear masks were expanded using the watershed method to segment cell cytoplasms. For custom segmentation in MATLAB, a script was developed that allowed a user to manually draw and modify cell masks. This script was used in cases where highly accurate, low-throughput cell segmentation was needed. For quantifying clusters, segmentation masks were fed forward into the cluster fitting algorithm (**see below**).

Fitting and quantifying clusters

Clusters were quantified from images of live and fixed cells using a custom MATLAB script. After cell segmentation, clusters were quantified in each cell at each timepoint (**Supplementary Figure 12**). The algorithm processes a single cell at a time, first normalizing the intensities within the cell by subtracting the minimum intensity within the cell, then dividing by the median background-subtracted intensity within the cell. Pixel intensities at the edge of the cell mask are then propagated to the surrounding pixels outside of the cell mask to reduce contrast at the edges of the cell (**Supplementary Figure 12C**)).

Selecting which pixels within the cell are included in a cluster is a two step process. First, single high-intensity pixels at the center points of clusters are identified. To select these pixels, a series of transforms is applied to the normalized cell image to remove background, suppress noise, and enhance medium-high frequency objects (clusters). These are as follows:

- Gentle gaussian low-pass filter to suppress the highest frequency noise. This aims to prevent single pixels that are randomly of high intensity from being enhanced by the subsequent contrast enhancing steps.
- A top hat filter to suppress background; areas with low and uniform (longer than length scales of ~5 pixels) values.
- Another gentle gaussian low-pass filter is used again to suppress the highest frequency contrast, which tends to be noise rather than clusters
- Convolution with a 5x5 grid with the following values
    -

-1	-1	-1	-1	-1
-1	-1	-1	-1	-1
-1	-1	24	-1	-1
-1	-1	-1	-1	-1
-1	-1	-1	-1	-1

    - This is a laplacian edge detector, and is typically used to enhance edges/areas of contrast

- Resultant negative values are set to zero
    - Values less than zero occur when a pixel has low intensity compared to its surroundings. Suppression of such features enhances cluster-finding.

With this enhanced contrast image (**Supplementary Figure 12E,G**), points above a user-defined threshold are identified. Quality control is performed to delete pixels from this list until only the highest intensity pixel (in the raw fluorescent image, rather than the transformed image) from each cluster is kept. This is defined as the cluster center (**Supplementary Figure 13C**).

Next, an expansion algorithm determines which pixels are included in the cluster. From each identified cluster center, adjacent pixels are checked to be included or excluded from the cluster as determined by an intensity cutoff. The cutoff is set by the local background in the area surrounding the cluster or by the median of the cell (whichever is lower) multiplied by a scaling factor set by the user. Pixels above this cutoff are included in the cluster, while pixels below the cutoff are discarded (**Supplementary Figure 13D**). This occurs recursively, where adjacent pixels to each included pixel are also checked to be included or excluded until all new neighbors are below the threshold.

With this two step algorithm, we were able to extract information from cells on the single cluster level, where the location, size, and intensity of each cluster in a cell is known.

<u>Quality control in analysis of cluster fitting from automated fits</u>

When clusters were fit on cells that were segmented in CellProfiler, rare errors occurred where either a large fraction of a cell mask would be background or a mask would contain a cell, and also a portion of a nearby and far brighter cell. In both cases, the algorithm would fit the entire bright portion (either cell in a mask of mostly background, or the portion of the very bright cell segmented with a dimmer cell) as a cluster. To address this, clusters fit from CellProfiler-generated masks were discarded if they were greater than 500 pixels in area, as such clusters were nearly exclusively caused by erroneous fits.

<u>Analysis of Cry2-GFP + CluMPS variants</u>

In all experiments where analysis is presented of Cry2-GFP is expressed alone or with a CluMPS variant, cells are gated for low expression of Cry2-GFP that do not form visible clusters after blue light stimulation. This is demonstrated in **Figure 1.**

<u>Data visualization</u>

Clustering data were exported from MATLAB. Data were organized and visualized in Rstudio using the tidyR package [85].

<u>Curve fitting</u>

Gab1(PRD)-CluMPS cluster area quantification during treatment with crizotinib (**Figure 5**) were fit to exponential decay functions of the form

$$A*\exp(-B*t)+C$$

where t is time in minutes, and with the moment of crizotinib addition being t = 0 minutes. A MATLAB function was written where the output was the sum of the squared difference between an exponential decay of the form above and experimental data. Minimization of this function generates the parameters for an exponential decay which most closely match the experimental data. This process required ‘guess’ parameters to act as a starting point for the minimization process. Guess parameters were set as A = the first experimental datapoint minus the lowest experimental data point, B = log(2)/3.5, and C = the lowest experimental data point.

Modeling

We generated a stochastic model of CluMPS amplification of target clustering in Python, using the NumPy[86], Math, Statistics[87], and matplotlib[88] libraries. XX packages. Our model modifies a previously reported model by Dine et al that describes clustering of a single protein species[24]. We extended this model to simulate the kinetics of condensation of two multivalent species, the key principle that drives CluMPS reporter activation upon clustering of a protein target, using a rejection kinetic Monte Carlo process. The simulation contains two species, the target and reporter. In each simulation there are 350 reporter units and 350 target units. At the simulation's initialization, targets and reporters are randomly arranged on a 70 x 70 square grid. Each square on the grid can contain up to one target and one reporter, but cannot contain two units of the same type. Thus, the simulation can be conceptualized as two superimposed planes: one containing targets and one containing reporters (**Figure 2A**). The model simulates kinetics through accepting or rejecting movements of units on the grids (a random unit of either type and direction for it to move are chosen at each step of the simulation). Whether a movement trajectory is accepted or rejected is determined by a calculated probability for the move to happen, and a random number roll (e.g. a move with a calculated probability of 60% is accepted if a randomly generated number between 0 and 1 is less than 0.6, while the move is rejected if the number generated is greater than 0.6). The probability of a move being accepted is calculated as a function of the enthalpy of the interactions a unit experiences at the initial state when a move is proposed. Units can only interact with units of the opposite species that are at the same or adjacent coordinates (**Figure 2A,B**). Both reporter and target units have a user-defined valency, which determines the maximum number of cross-species interactions a unit is able to participate in. The reporter has a valency of 6 in all simulations, while the target valency varies between simulations (between 2-6) as indicated (**Figure 2C,D,E,F**). The valency of all units stays constant throughout the duration of a simulation. Interactions between units have a user-defined interaction energy, ΔE. The probability for a trajectory that moves a protein to a position not occupied by another of its type is e^{-T} where T is the sum of the interaction energies from all the interactions that the protein currently experiences (the minimum of the number of interactions and the units valency, multiplied by the energy of a single interaction), $T = min(valency, \# of\ neighbors) * \Delta E$. The following metrics are calculated every 2000 proposed steps: the total interaction strength of the system (the sum of the enthalpies that each protein is currently experiencing), the percent clustering (percent of reporter units that are in a cluster of six or more reporters), and finally the rate of change of the total interaction strength from the previous interval to the current one. After 200 intervals have passed (200 intervals*200 steps per interval = 40,000 steps) the model also sums the last 200 rates of change over their corresponding total interaction strengths. When this sum is <= 0.01, the simulation has reached steady state and ends, and the average percent clustering over the last 100 intervals is reported.

For simulations in **Figure 2F**, in addition to multivalent target units, additional monovalent target units were added to the simulation. For example, to run a simulation with 50% of clustered target, 350 monovalent target units would be added to the simulation in addition to the 350 multivalent target units. Monovalent targets have the same interaction strength, ΔE, with reporter units, can be randomly selected to move like any other unit, but can interact with a maximum of one reporter unit. When a monovalent target is selected to move to a grid square containing a multivalent target or vice versa, the probability that the two units switch places on the grid is set as $e^{-(T_1+T_2)}$, where T_1 is the enthalpy of interactions for the first unit, $T = min(valency, \# of\ neighbors) * \Delta E$, and T_2 is the enthalpy of interactions for the second unit, calculated in the same way.

Construct table

Construct	Backbone
LaG17-mCh-HOTag3	pHR PGK
LaG6-mCh-HOTag3	pHR PGK
LaG18-mC-HOTag3	pHR PGK
LaG17-mCh-HOTag6	pHR PGK
LaG17-mCh-FUS(LC)	pHR PGK
LaG17-mCh-RGG	pHR PGK
LaG17-mCh-Xvelo	pHR PGK
LaG17-mCh	pHR PGK
LaG17-mCh-HOTag3-NLS(c-MYC)	pHR PGK
LaG17-miRFP670-HOTag6	pHR PGK
Cry2-EGFP	pCMV
EGFP-sspB	pCMV
EGFP-miRFP670	pCMV
miRFP670-iLid-iLid	pCMV
miRFP670-iLid-iLid-iLid	pCMV
miRFP670-iLid-iLid-iLid-iLid	pCMV
miRFP670-iLid-iLid-iLid-iLid-iLid	pCMV
miRFP670-iLid-iLid-iLid-iLid-iLid-iLid	pCMV
EGFP-TDP43	pCMV
SARS-CoV-2 Nucleocapsid-EGFP	pCMV
EML4-ALK-EGFP	pCMV
iLid-FTH1-2A-EGFP-sspB	pCMV
Cry2-ALFA-2A-BFP	pCMV
LaG17-mRuby2-HOTag3	pCMV
Gab1(263-451)-mCh-HOTag3	PCSDest

ERK KTR-mClover	PCSDest

Supplementary Movie Captions

Supplementary Movie 1. Screening amplification domains for CluMPS reporter. Representative time lapse videos of candidate CluMPS variants (magenta) co-transfected with Cry2-GFP and stimulated with blue light. CluMPS variants harbor different amplification domains as indicated. Time is mm:ss.

Supplementary Movie 2. Representative simulations of stochastic model of CluMPS amplification. Each row shows motion of the target and reporter units under the indicated simulation parameters. All rows has identical simulation parameters except for the valency of the target, which is indicated to the left of each row. The right-most column shows the extent of reporter clustering and the total interaction strength of the simulation. Simulations were stopped once the interaction strength reached steady state. See **Methods** for more model details.

Supplementary Movie 3. CluMPS:target affinity regulates CluMPS amplification. LaG-CluMPS variants with differing affinities were each co-transfected with Cry2-GFP and stimulated with blue light. Higher-affinity CluMPS provided greater cluster amplification. Time is mm:ss.

Supplementary Movie 4. Gab1(PRD)-CluMPS visualizes endogenous EML4-ALK condensates and their shrinkage in response to ALK inhibitor. Gab1(PRD)-CluMPS was expressed in H3122 cancer cells, which harbor the EML4-ALK (V1) oncogene. ALK inhibitor crizotinib (1 μM) is added at 30'. Time is mm:ss.

Supplementary Movie 5. Multiplexed imaging of orthogonal CluMPS reporters in single cells. Orthogonal CluMPS variants (LaG17-CluMPS, nbALFA-CluMPS) were co-expressed with orthogonal blue-light-induced clustering systems that clustered GFP or ALFA-tag, respectively, in the single cells. Blue light is applied at the beginning of the video to induce clustering of both GFP and ALFA-tag. Distinct clusters are observed in both the miRFP and mRuby2 channels, suggesting independent amplification of the distinct clusters. TIme is mm:ss.

REFERENCES

1. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84**, 551–575 (2015).
2. Chiesa, G., Kiriakov, S. & Khalil, A. S. Protein assembly systems in natural and synthetic biology. *BMC Biol.* **18**, 35 (2020).
3. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
4. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
5. Zamudio, A. V. *et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cell* **76**, 753–766.e6 (2019).
6. Hartman, N. C. & Groves, J. T. Signaling clusters in the cell membrane. *Curr. Opin. Cell Biol.* **23**, 370–376 (2011).
7. Riback, J. A. *et al.* Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028–1040.e19 (2017).
8. Guillén-Boixet, J. *et al.* RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346–361.e17 (2020).
9. Yang, P. *et al.* G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325–345.e28 (2020).
10. Tulpule, A. *et al.* Kinase-mediated RAS signaling via membraneless cytoplasmic protein granules. *Cell* **184**, 2649–2664.e18 (2021).
11. Mehta, S. & Zhang, J. Liquid-liquid phase separation drives cellular function and dysfunction in cancer. *Nat. Rev. Cancer* **22**, 239–252 (2022).
12. Johnson, B. S. *et al.* TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J. Biol. Chem.* **284**, 20329–20339 (2009).

13. Haass, C. & Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101–112 (2007).
14. Fusco, G. *et al.* Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science* **358**, 1440–1443 (2017).
15. Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1332–1340 (2018).
16. Zhang, J. Z. *et al.* Phase Separation of a PKA Regulatory Subunit Controls cAMP Compartmentation and Oncogenic Signaling. *Cell* **0**, (2020).
17. Kaushik, S. & Cuervo, A. M. Proteostasis and aging. *Nat. Med.* **21**, 1406–1415 (2015).
18. David, D. C. *et al.* Widespread protein aggregation as an inherent part of aging in C. elegans. *PLoS Biol.* **8**, e1000450 (2010).
19. Ben-Zvi, A., Miller, E. A. & Morimoto, R. I. Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14914–14919 (2009).
20. Korkmazhan, E., Tompa, P. & Dunn, A. R. The role of ordered cooperative assembly in biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **22**, 647–648 (2021).
21. Su, Q., Mehta, S. & Zhang, J. Liquid-liquid phase separation: Orchestrating cell signaling through time and space. *Mol. Cell* **81**, 4137–4146 (2021).
22. Castellana, M. *et al.* Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* **32**, 1011–1018 (2014).
23. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
24. Dine, E., Gil, A. A., Uribe, G., Brangwynne, C. P. & Toettcher, J. E. Protein Phase Separation Provides Long-Term Memory of Transient Spatial Stimuli. *Cell Syst* **6**, 655–663.e5 (2018).
25. Koch, D. Homo-Oligomerisation in Signal Transduction: Dynamics, Homeostasis, Ultrasensitivity, Bistability. *J. Theor. Biol.* **499**, 110305 (2020).

26. Ali, M. H. & Imperiali, B. Protein oligomerization: how and why. *Bioorg. Med. Chem.* **13**, 5013–5020 (2005).

27. Kumari, N. & Yadav, S. Modulation of protein oligomerization: An overview. *Prog. Biophys. Mol. Biol.* **149**, 99–113 (2019).

28. Farhadi, S. A., Liu, R., Becker, M. W., Phelps, E. A. & Hudalla, G. A. Physical tuning of galectin-3 signaling. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

29. Choi, M. L. & Gandhi, S. Crucial role of protein oligomerization in the pathogenesis of Alzheimer's and Parkinson's diseases. *FEBS J.* **285**, 3631–3644 (2018).

30. de Oliveira, G. A. P. *et al.* The Status of p53 Oligomeric and Aggregation States in Cancer. *Biomolecules* **10**, (2020).

31. Chen, S. W. *et al.* Structural characterization of toxic oligomers that are kinetically trapped during α-synuclein fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1994–2003 (2015).

32. Fang, Y.-S. *et al.* Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun.* **5**, 4824 (2014).

33. Li, S. *et al.* Soluble oligomers of amyloid Beta protein facilitate hippocampal long-term depression by disrupting neuronal glutamate uptake. *Neuron* **62**, 788–801 (2009).

34. Douglass, A. D. & Vale, R. D. Single-molecule microscopy reveals plasma membrane microdomains created by protein-protein networks that exclude or trap signaling molecules in T cells. *Cell* **121**, 937–950 (2005).

35. Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity* **25**, 117–127 (2006).

36. Mo, G. C. H. *et al.* Genetically encoded biosensors for visualizing live-cell biochemical activity at super-resolution. *Nat. Methods* **14**, 427–434 (2017).

37. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).

38. Keber, F. C., Nguyen, T., Brangwynne, C. P. & Wühr, M. Evidence for widespread cytoplasmic structuring into mesoscopic condensates. *bioRxiv* 2021.12.17.473234 (2021) doi:10.1101/2021.12.17.473234.

39. Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für Mikroskopische Anatomie* **9**, 413–468 (1873).

40. Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S. & Schaffer, D. V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nat. Methods* **10**, 249–252 (2013).

41. Taslimi, A. *et al.* An optimized optogenetic clustering tool for probing protein interaction and function. *Nat. Commun.* **5**, 4925 (2014).

42. Lee, S. *et al.* Reversible protein inactivation by optogenetic trapping in cells. *Nat. Methods* **11**, 633–636 (2014).

43. Bugaj, L. J. *et al.* Regulation of endogenous transmembrane receptors through optogenetic Cry2 clustering. *Nat. Commun.* **6**, 6898 (2015).

44. Park, H. *et al.* Optogenetic protein clustering through fluorescent protein tagging and extension of CRY2. *Nat. Commun.* **8**, 30 (2017).

45. Shin, Y. *et al.* Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* vol. 168 159–171.e14 (2017).

46. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).

47. Bader, A. N., Hofman, E. G., Voortman, J., en Henegouwen, P. M. P. van B. & Gerritsen, H. C. Homo-FRET imaging enables quantification of protein cluster sizes with subcellular resolution. *Biophys. J.* **97**, 2613–2622 (2009).

48. Breedijk, R. M. P. *et al.* A live-cell super-resolution technique demonstrated by imaging germinosomes in wild-type bacterial spores. *Sci. Rep.* **10**, 5312 (2020).

49. Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **10**, 4756–4761

(2010).

50. Ma, Y. *et al.* An intermolecular FRET sensor detects the dynamics of T cell receptor clustering. *Nat. Commun.* **8**, 15100 (2017).
51. Sharma, P. *et al.* Nanoscale organization of multiple GPI-anchored proteins in living cell membranes. *Cell* **116**, 577–589 (2004).
52. Digman, M. A., Dalal, R., Horwitz, A. F. & Gratton, E. Mapping the number of molecules and brightness in the laser scanning microscope. *Biophys. J.* **94**, 2320–2332 (2008).
53. Ojosnegros, S. *et al.* Eph-ephrin signaling modulated by polymerization and condensation of receptors. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13188–13193 (2017).
54. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
55. Huang, B. Super-resolution optical microscopy: multiple choices. *Curr. Opin. Chem. Biol.* **14**, 10–14 (2010).
56. Mammen, M., Choi, S.-K. & Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem. Int. Ed Engl.* **37**, 2754–2794 (1998).
57. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
58. Nakamura, H. *et al.* Intracellular production of hydrogels and synthetic RNA granules by multivalent molecular interactions. *Nat. Mater.* **17**, 79–89 (2018).
59. Bracha, D. *et al.* Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467–1480.e13 (2018).
60. Zhang, Q. *et al.* Visualizing Dynamics of Cell Signaling In Vivo with a Phase Separation-Based Kinase Reporter. *Mol. Cell* **69**, 334–346.e4 (2018).
61. Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**, 2985 (2018).

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

63. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).

64. Murthy, A. C. *et al.* Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).

65. Fridy, P. C. *et al.* A robust pipeline for rapid production of versatile nanobody repertoires. *Nat. Methods* **11**, 1253–1260 (2014).

66. Che, D. L., Duan, L., Zhang, K. & Cui, B. The Dual Characteristics of Light-Induced Cryptochrome 2, Homo-oligomerization and Heterodimerization, for Optogenetic Manipulation in Mammalian Cells. *ACS Synth. Biol.* **4**, 1124–1135 (2015).

67. Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).

68. Guntas, G. *et al.* Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 112–117 (2015).

69. Gam, J. J., DiAndreth, B., Jones, R. D., Huh, J. & Weiss, R. A 'poly-transfection' method for rapid, one-pot characterization and optimization of genetic systems. *Nucleic Acids Res.* **47**, e106 (2019).

70. Iserman, C. *et al.* Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* **80**, 1078–1091.e6 (2020).

71. Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **12**, 1936 (2021).

72. Perdikari, T. M. *et al.* SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J.* **39**, e106478 (2020).

73. Savastano, A., Ibáñez de Opakua, A., Rankovic, M. & Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.* **11**, 6041 (2020).

74. Carlson, C. R. *et al.* Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Mol. Cell* **80**, 1092–1103.e4 (2020).

75. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).

76. Hrustanovic, G. *et al.* RAS-MAPK dependence underlies a rational polytherapy strategy in EML4-ALK–positive lung cancer. *Nat. Med.* **21**, 1038 (2015).

77. McDonald, C. B., Seldeen, K. L., Deegan, B. J., Bhat, V. & Farooq, A. Binding of the cSH3 domain of Grb2 adaptor to two distinct RXXK motifs within Gab1 docker employs differential mechanisms. *J. Mol. Recognit.* **24**, 585–596 (2011).

78. Irgen-Gioro, S., Walling, V. & Chong, S. Fixation Can Change the Appearance of Phase Separation in Living Cells. *bioRxiv* 2022.05.06.490956 (2022) doi:10.1101/2022.05.06.490956.

79. Gonzalez-Martinez, D. *et al.* Oncogenic protein condensates modulate cell signal perception and drug tolerance. *bioRxiv* 2022.02.02.478845 (2022) doi:10.1101/2022.02.02.478845.

80. Sampson, J., Richards, M. W., Choi, J., Fry, A. M. & Bayliss, R. Phase-separated foci of EML4-ALK facilitate signalling and depend upon an active kinase conformation. *EMBO Rep.* **22**, e53693 (2021).

81. Götzke, H. *et al.* The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nat. Commun.* **10**, 4403 (2019).

82. Newby, G. A. *et al.* A Genetic Tool to Track Protein Aggregates and Control Prion Inheritance. *Cell* **171**, 966–979.e18 (2017).

83. Regot, S., Hughey, J. J., Bajar, B. T., Carrasco, S. & Covert, M. W. High-Sensitivity Measurements of Multiple Kinase Activities in Live Single Cells. *Cell* **157**, 1724–1734 (2014).

84. Bugaj, L. J. & Lim, W. A. High-throughput multicolor optogenetics in microwell plates. *Nat. Protoc.* **14**, 2205–2228 (2019).

85. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

86. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

87. Van Rossum, G. & Python Development Team. *The Python Library Reference: Release 3.6.4 - Book 1 of 2*. (12th Media Services, 2018).

88. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).