**Intrinsically disordered sequences in the Polycomb protein Polyhomeotic regulate condensate formation**

Ibani Kapur[1,2], Elodie L. Boulier[1], and Nicole J. Francis[1,2,3]

[1]Institut de recherches cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, QC, H2W 1R7, Canada
[2]Division of Experimental Medicine, McGill University, 1001 Decarie Boulevard, Montreal, QC H4A 3J1, Canada
[3]Département de biochimie et médecine moléculaire Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, QC H3T 1J4, Canada.
*correspondence to: nicole.francis@ircm.qc.ca

**ABSTRACT:**

The Polycomb group (PcG) complex PRC1 localizes in the nucleus in the form of condensed structures called Polycomb bodies. The PRC1 subunit Polyhomeotic (Ph) contains a polymerizing sterile alpha motif (SAM) that is implicated in both PcG body formation and chromatin organization in *Drosophila* and mammalian cells. A truncated version of Ph containing the SAM (mini-Ph), forms phase separated condensates with DNA or chromatin *in vitro*, suggesting PcG bodies may form by phase separation. In cells, Ph forms multiple condensates, while mini-Ph forms a single large nuclear condensate. We therefore hypothesize that sequences outside of mini-Ph are required for proper condensate formation. We identified three distinct Intrinsically Disordered Regions (IDRs) in Ph based on sequence composition and complexity. We tested the role of each IDR in Ph condensates using live imaging of transfected *Drosophila* S2 cells. We find that each IDR uniquely affects Ph SAM-dependent condensate size, number, and morphology.

## INTRODUCTION:

The role of biomolecular condensates in regulation of cellular processes, including gene expression, is increasingly appreciated. Phase separation, the demixing of molecules in solution to dense and dilute phases, may underlie many condensates that form in cells, providing a means to concentrate and segregate biomolecules[1–3]. However, while the core physical principle of phase separation may explain condensates, phase separation must be tightly regulated in time and space to generate condensates with biological functions. The material properties of condensates, which can include more or less viscous liquids and gel-like or even solid states, are also essential for their function[4,5]. Finally, biological condensates are not simple polymer in solvent systems, but highly complex structures existing in an equally complex solvent (i.e. the cytoplasm or nucleus)[6].

To understand how complex condensates form in cells, identifying key proteins that drive their formation, and determining how the sequences of these proteins contribute is essential. Many examples, as well as theory, have demonstrated how both structured domains and intrinsically disordered sequences can undergo phase separation under biologically relevant conditions[7]. Because weak, multivalent interactions are key to creating liquid condensates, Intrinsically Disordered Regions (IDRs) are especially well suited to undergo phase separation[1,7,8]. However, IDRs may lack the specificity needed to form relevant condensates in the complex cellular environment[9]. There are now several examples of proteins that drive condensate formation through a combination of structured domains and IDRs[10,11].

The Polycomb Group (PcG) proteins are epigenetic regulators of gene expression that function by modifying chromatin structure[12–14]. Two main PcG multiprotein complexes, PRC1 and PRC2 are conserved across evolution[15]. Both complexes modify histones (H2A119Ub and H3K27me3, respectively). PcG proteins also organize chromatin at a large scale, forming large, compacted chromatin domains[13,16]. This activity is distinct from histone modifications. PcG proteins themselves form visible condensates, and these condensates co-localize with PcG-regulated genes[13,16]. This suggests that the formation of PcG condensates and of compacted chromatin domains are linked processes. PRC1 components, particularly Polyhomeotic (Ph) and Polycomb (Pc) (PHC and CBX in mammals) are the most implicated in condensate formation and chromatin organization[17,18].

Ph has a C-terminal Sterile Alpha Motif (SAM) that is essential for its function and implicated in both condensate formation and chromatin organization. The Ph SAM can form head to tail polymers[19], and this polymerization is important for Ph function[20]. Dominant negative Ph SAM mutations that disrupt polymerization of wild-type Ph SAM also disrupt PcG clusters, lead to loss of chromatin compaction, and disrupt gene regulation in both *Drosophila* and mammalian cells, and in mice[21,22]. Loss of Ph leads to decompaction of chromatin in *Drosophila* cells, *Drosophila* embryos, and mouse embryonic stem cells[17,23,24]. Although SAM polymerization is its hallmark activity, genetic experiments indicate that Ph SAM has critical polymerization independent functions because a *ph* transgene with a polymerization interface mutated can rescue some *ph* functions in *Drosophila* embryos, while *phΔSAM* cannot[25].

We recently showed that Ph SAM, in the context of a truncated version of Ph containing its three conserved domains, mini-Ph, undergoes phase separation with DNA or chromatin *in vitro*. Phase separation requires the SAM, but does not strictly require polymerization[26]. PHC1 was

also shown to undergo SAM-dependent phase separation in cells when induced to cluster through optogenetic manipulation[27]. Thus, phase separation could be the critical second Ph SAM activity. However, in cells, Ph forms many condensates, while mini-Ph forms a single condensate; both types of condensates require the SAM[26]. Thus, we hypothesized that sequences N-terminal to mini-Ph regulate condensate formation through the SAM. Indeed, *Drosophila* Ph has ~1200 additional amino acids that have no recognizable domains and are not characterized. Most mammalian PHC isoforms (all of which have the mini-Ph region) also contain uncharacterized sequence N-terminal to the SAM.

Here, we analyze the N-terminal sequence of Ph, which is predicted to be mainly disordered/low complexity. We identified three intrinsically disordered regions (IDRs) based on sequence composition and complexity, and tested their effects on condensate formation in cells. We find that each IDR affects condensates differently. We also found that some of the IDRs can form condensates on their own, or in combination. Both IDR-only condensates and SAM-dependent ones are negatively regulated by an IDR that was previously shown to prevent Ph SAM-dependent aggregation through its O-linked glycosylation. Our results point to a complex network of regulatory interactions mediated by IDRs that govern Ph SAM-driven condensates, with an important role for negative regulation. Since Ph SAM affects chromatin organization, the IDR activities identified here may also affect chromatin.

**RESULTS:**

**The N terminal region of Ph comprises three distinct IDRs.** *Drosophila melanogaster* contains two tandem *ph* genes (*ph-p* and *ph-d*), which are highly similar and largely functionally redundant[28,29]. Our analysis is of Ph-p, referred to hereafter as Ph. Ph is a large protein (1589 amino acids), with three small domains in its C-terminal region namely the HD1, FCS and SAM (**Fig. 1A**). Using the IUPRED disorder predictor[30], the majority of the Ph N-terminal sequence is predicted to be disordered (**Fig. 1A**).Intrinsically disordered protein sequences often also contain low complexity and compositionally biased sequence. To identify such regions in Ph, and potential subregions in the large disordered region, we used PlaToLoCo (Platform of Tools for Low Complexity)[31]. This analysis tool applies an intersection between several algorithms (SEG, CAST[32] fLPS, GBSC and SIMPLE), but we focused on the SEG and CAST analysis.

Although low complexity sequence is present throughout the N-terminal region, the masked (i.e. repetitive) amino acids are different in different regions of the sequence. This allowed us to demarcate Ph1 (S-masked); Ph2 (Q-masked); and Ph3 (S+T masked) (**Fig. 1B**). Ph1 does not have a clear enrichment for a single amino acid, but Ph2 is enriched in Glutamine (36%) and Ph3 in Threonine/Serine (20% each) (**Fig. 1C, D**). PLAAC was used to scan for prion like amino acid composition in the protein sequence[33]. Consistent with its high glutamine content, Ph2 has high scoring predicted prion-like regions; Ph1 also has a predicted prion-like region (**Fig. 1E**).

To understand the contribution of each IDR to the organization of Ph into condensates in cells, we designed a series of constructs consisting of N-terminal fusions of Venus to truncated versions of Ph. Constructs were designed to test: 1) the effect of removing each IDR from Ph; 2) the activity of each IDR when fused to mini-Ph (i.e. deletion of 2 IDRs); 3) the role of the SAM in IDR effects; 4) the condensate forming activity of each IDR alone or in combination. We used the heat shock promoter for inducible expression. These constructs were transiently transfected

into *Drosophila* S2 cells, along with a plasmid encoding H2A-RFP as a nuclear marker. Cells were subjected to mild heat shock (8 min. at 37°C), and allowed to recover overnight before imaging. Venus-containing condensates formed in transfected cells were analyzed by live imaging.

As shown previously[26], transfected wild type (WT) Ph forms several round, bright condensates in cells, while mini-Ph generally forms a single condensate (**Fig. 2**). To quantify condensates in transfected cells, we developed Cell Profiler analysis pipelines to identify condensates, count them, and measure their properties. We counted the number of condensates (referred to as foci) per nucleus (**Fig. 2D**), the median condensate size per nucleus (**Fig. 2E**). We also measured the form factor of condensates, which is a measure of circularity ($4*\pi*\text{Area}/\text{Perimeter}^2$), it equals 1 for a perfectly circular object. However, we obtained measurements >1, likely due to inaccuracies in measuring object perimeter (because of the resolution limit imposed by pixels). We therefore did not use these measurements. Previous reports have characterized the partition ratio of condensates by measuring the intensity inside versus outside[34,35]. However, due to the small size of most of our condensates, we could not measure foci intensities accurately.

**Removal of IDRs affects Ph condensate size, number, and morphology.** Condensates formed in the absence of Ph1 (PhΔ1) are non-round, frequently small, and tend to form interconnected networks and, in some cases, large clumps (**Fig. 2B**). This likely explains the significant increase in foci size, but not number observed for PhΔ1 (**Fig. 2D, E**). Removal of Ph2 (PhΔ2) results in a small number of large condensates. Consistent with this impression, the number of foci per nucleus counted for PhΔ2 is reduced compared to wild type and there is a significant increase in foci size (**Fig. 2D, E**). Condensates formed in the absence of Ph3 (PhΔ3) are bright and numerous, and most similar to WT-Ph condensates (**Fig. 2B**). Removing Ph3 (PhΔ3) results in the largest increase in foci number when compared to WT-Ph, with the number nearly doubling (**Fig. 2D**). Relative to wild type Ph, the size of foci is only slightly higher (**Fig. 2E**). Of the three IDR deletions, PhΔ3 is the most similar to WT.

When we fused a single IDR to mini-Ph (deletion of 2 IDRs), we again noticed changes in the properties of condensates (**Fig.2C**). When Ph1 was fused to mini-Ph (PhΔ2Δ3), both the number and size of foci is slightly increased, when compared with WT-Ph (**Fig. 2D, E**). With Ph2 fused to mini-Ph (PhΔ1Δ3), the foci number was not significantly different from WT-Ph. However, most of the foci formed are much smaller than those formed by WT-Ph (**Fig. 2D, E**). When we have only Ph3 fused to mini-Ph (PhΔ1Δ2), foci formation is reduced (**Fig. 2D**). PhΔ1Δ2 either forms one large round structure in the cell (similar to what is observed with mini-Ph alone) or has diffused Venus signal (as is observed with PhΔSAM).

**Relationship between Venus-Ph expression levels and condensate formation.** If Ph IDRs mediate molecular interactions that promote or inhibit condensate formation, then removing them could change the concentration of Ph required for condensate formation. In the case of a phase separation mechanism, this would reflect a change in the saturation concentration ($C_{sat}$). We found that the fraction of transfected cells that have nuclear condensates is different for different constructs (**Fig. 3A**). This could reflect differences in expression levels for different proteins or changes in the threshold for condensate formation.

To test whether different Venus-Ph fusion proteins are expressed at different levels, and to confirm the integrity of the fusion proteins, we analyzed transfected cells by Western blotting

with antibodies to GFP (to detect Venus-Ph), and RFP (to detect the transfected histone) (**Fig. 3B, C**). Antibodies to tubulin and H2B were used as loading controls. We find that all the proteins are expressed at the expected size (**Fig. 3B**). Quantification of GFP signal indicates an expression range of 1.5 to 3.5 fold for transfected deletion constructs relative to transfected WT-Ph (**Fig. 3E**). Transfection efficiency, determined by analysis of H2A-RFP versus endogenous H2B, was similar for all constructs (**Fig. 3F**). This suggests the larger differences in levels among the Venus-Ph constructs could reflect changes in protein stability, although this was not tested. We also used antibodies to Ph to compare expression of transfected proteins to endogenous Ph (**Fig. 3D**). Total Ph levels (endogenous + transfected) are increased by up to two-fold (**Fig. 3G**).

We also measured the mean nuclear Venus intensity for each cell expressing H2A-RFP for each of the IDR deletion constructs, and compared it to that for transfected WT Venus-Ph. We find that the range of nuclear intensities is more similar for all of the constructs than predicted from Western blots (**Fig. 3H**). There is also no correlation between the Venus-Ph signal calculated from Western blots and that from nuclear intensities (**Fig. 3I**). We noticed that some of the Venus-Ph proteins, particularly PhΔ3, frequently form large cytoplasmic foci when expressed at high levels, in addition to nuclear ones; these were not quantified in our analysis (**Fig. 2B**). An explanation for the discrepancy between Western blot and nuclear intensities is that the level of nuclear Ph is controlled, and excess protein accumulates in the cytoplasm, since Western blotting measures all cellular protein and nuclear intensity only nuclear protein. This suggests the changes in the fraction of cells with nuclear condensates for different proteins could reflect *bona fide* differences in how readily different Ph proteins form condensates in the nucleus.

To determine whether the concentration required for condensate formation is changed by removing Ph IDRs, we took advantage of the wide range of expression levels obtained with transient transfection to assess the relationship between total Venus-Ph levels and condensate formation at the single cell level. **Fig. 4** shows the range of expression levels for each construct, segregated by whether or not nuclear condensates were detected. In all cases, the population of cells with condensates has a higher expression level than that without. However, intensity levels for cells with and without condensates overlap. Thus, although expression levels correlate with condensate formation, expression level is not sufficient to predict condensate formation, since cells can have high levels of expression without forming condensates.

**Removal of the Ph3 IDR allows Ph2-dependent but SAM-independent condensates to form.** We showed previously that the Ph SAM is necessary for condensates to form[26]. To determine if the Ph IDR deletion constructs form condensates in the absence of SAM, we tested all the constructs except PhΔ1Δ2 without SAM (**Fig. 5A**). In the absence of the SAM, most proteins did not form condensates. However, in two cases, deletion of Ph3 and deletions of both Ph1 and Ph3, condensates form without the SAM although the number of cells that form condensates is low (**Fig. 5A, B**). Although the ΔSAM versions are expressed at about 2-fold lower levels than the corresponding proteins with the SAM (**Fig. 4E**), given the wide range of expression levels over which condensates are observed (**Fig. 3B-H**), this is unlikely to explain why most ΔSAM

proteins do not form foci. The ΔSAM versions are also expressed at similar levels as transfected WT-Ph (**Fig. 4E**).

**Ph IDRs alone and in combination can form condensates** To test if the IDRs can form condensates in the absence of mini-Ph, we tested each IDR alone or in combination (**Fig. 7A**). An SV40 NLS was added to ensure that the IDRs localize to the nucleus. Ph3 alone does not form condensates. However, both Ph 1 and Ph2 form condensates in some cells (**Fig. 7B**). When Ph1 and Ph2 are combined (Ph5), many small condensates are formed, significantly more than with WT-Ph (**Fig. 7C**). In contrast, when Ph2 and Ph3 are combined (Ph6), condensates do not form. Finally, when Ph1, Ph2, and Ph3 are combined, condensates are formed, although fewer than with Ph1+Ph2 (Ph5) (**Fig. 7B, C**). **Table 1** shows a comprehensive summary of condensate formation for all of the constructs tested.

In summary, we have identified three IDRs in the N-terminal region of Ph, each of which influence Ph SAM-dependent condensates in cells. The IDRs also functionally interact with each other. **Fig. 8** summarizes these findings.

**DISCUSSION:**

We sought to determine how SAM dependent condensates that may form by phase separation are controlled by the disordered N-terminal region of Ph. We identified three distinct IDRs in this region by sequence analysis and tested their function alone, with the conserved region of Ph, and without the Ph SAM. We find that each IDR affects condensates, and that the IDRs interact functionally with each other. The potential mechanisms underlying their effects are considered below.

A caveat to this work is that we studied behaviour of Ph proteins under conditions of overexpression, and with endogenous Ph present in the background. While this allows us to evaluate how the different IDRs affect Ph condensate formation in a cellular context, we do not know which effects are dependent on endogenous Ph. We think that many of the observed effects are due to intrinsic properties of the IDR sequences because overexpression of the proteins likely saturates their binding partners. However, we cannot rule out that some IDR effects are due to interactions with cellular proteins. These experiments reveal the properties of the IDRs; whether their effects are manifested under normal physiological conditions awaits additional experiments.

Ph1 promotes condensate formation. When Ph1 is fused to mini-Ph (PhΔ2Δ3), wild-type like condensates are formed (although they are more numerous than WT-Ph) (**Fig. 2F**). These condensates are dependent on the Ph SAM. When Ph1 is deleted, condensates are less round, and diffuse granular staining is often observed. Ph1 alone can form condensates in cells, although it does so rarely. Taken together, these data suggest Ph1 promotes the condensate forming ability of Ph SAM, leading to formation of many condensates in a high fraction of cells.

Ph2 has complex effects on condensates. In the presence of Ph SAM, it controls condensate size, so that its removal leads to formation of very large condensates, and Ph2 fused to mini-Ph results in many small condensates (although in a small number of cells). Ph2 promotes condensate formation in the absence of the SAM, but only if Ph3 is absent. Ph2 alone has weak condensate forming ability, but when fused to Ph1 forms a large number of small condensates. Thus, Ph2 negatively regulates Ph SAM dependent condensates, but also interacts functionally with the other IDRs in a SAM independent manner—it synergizes with Ph1 to promote

condensates, and is inhibited by Ph3. The most striking sequence feature of Ph2 is its high glutamine (36%) content. Ph2 also has strong predicted coiled-coil forming activity. Alpha helical coiled coil formation has been shown to regulate phase separation behaviour, and could contribute to Ph2 activity. It is also intriguing to note that a frequently used *ph* allele, $ph^{505}$, which was thought to be a null, actually encodes truncated versions of both *ph-p* and *ph-d*. The predicted proteins are truncated midway through the Ph2 region, so would potentially produce proteins similar to Ph5 (i.e. Ph 1+2). $ph^{505}$ effects on gene regulation differ slightly from those of a true *ph* null[36], raising the possibility that $ph^{505}$ effects are due to expression of the Ph5-like protein.

Ph3 has an inhibitory effect on condensates so that more condensates are formed when it is removed. Careful consideration of the effect of removing and having Ph3 in different constructs suggests Ph3 may function through the other IDRs, particularly by restricting the activity of Ph2. First, when Ph3 is fused to mini-Ph (PhΔ1Δ2), few condensates are formed, and the ones that form are quite similar to mini-Ph alone, suggesting Ph3 has little effect in this context. Second, Ph1 fused to mini-Ph (PhΔ2Δ3) forms several condensates, but the number is significantly reduced (and size increased) when Ph3 is present with Ph1 (PhΔ2) (**Fig. 2F**). Third, Ph2 forms condensates in the absence of SAM (PhΔ1Δ3ΔSAM) but with Ph3 present does not (PhΔ1ΔSAM). Fourth, Ph2 alone forms condensates in some cells but again condensate formation is inhibited when Ph2 is fused to Ph3 (Ph6). Thus, Ph3 function may be through the other IDRs, rather than directly on Ph SAM; alternatively, it could inhibit both Ph SAM and Ph2. Ph3 is rich in serine and threonine and is known to be extensively modified with O-linked glycosylation of these residues[37]. Removing glycosylation drives formation of Ph aggregates *in vitro* and *in vivo*, and these aggregates depend on the SAM[25]. Whether these aggregates also depend on Ph2 is not known. However, one hypothesis would be that glycosylation restrains the activity of Ph2 that promotes condensate formation through SAM-dependent and independent mechanisms. Recently, O-GlcNacylation was also shown to reduce both aggregation and phase separation of the N-terminal LCR of EWS[38].

Many mechanisms have been described and hypothesized to explain control of condensate size and number in cells[39–41]. One model that may be relevant for Ph1 and 2 explains how chromatin binding affects condensate number and size. Qi and Zhang used simulations to show that protein-chromatin interactions can promote a multi-droplet state[39]. In this model, protein-chromatin contacts promote nucleation of phase separation, but inhibit their coalescence. Detailed investigation of the thermodynamics underlying these effects indicate that this is due to a kinetic barrier between single and multi-droplet states. This kinetic barrier arises from the chromatin network—when droplets that interact with chromatin coalesce, the chromatin network is constrained, which becomes progressively more energetically costly. Theoretical work from Wingreen and colleagues also finds that chromatin interferes with coarsening of condensates, both by acting as a crowder, and through its crosslinked network[41].Thus, changing protein-chromatin interactions could change droplet number. In the framework of this model, Ph1 may decrease chromatin binding affinity, promoting formation of many round condensates. Loss of Ph1 may lead to tight chromatin binding, explaining the distorted shape of condensates formed by PhD1. Ph2, on the other hand, may promote chromatin binding so that its loss leads to large condensates because fusion is not impaired. Recently, Seydoux and colleagues have demonstrated that clusters of MEG-3 function as Pickering particles, which adsorb to the surface of P-granules and reduce their coarsening by reducing surface tension[42]. As speculated, RNA

may also have this function for some condensates[42]. In this regard it is interesting that the Ph2 region is predicted (using DisoRDPbind[43]) to have a strong RNA binding activity, although this has not been validated experimentally.

In conclusion, we showed that the disordered N-terminal region of Ph, which was entirely uncharacterized, affects Ph SAM dependent condensate formation. Nearly all mammalian Ph homologues have disordered N-terminal IDRs; we hypothesize that IDR regulation of condensate formation by these IDRs may be functionally conserved.

## MATERIALS AND METHODS:

**Cloning:** Ph, and all of the truncations and deletions were cloned into a gateway donor vector using restriction digest and ligation, PCR, or gene blocks spanning deletion junctions. All synthetic and PCR generated sequence was confirmed by Sanger sequencing. The plasmid for Ph truncations (pCR8-ATG-NLS) includes an SV40 nuclear localization sequence (NLS). To transfer Ph truncations into plasmid pHVW (DGRC stock # 1089) for heat-shock inducible expression in *Drosophila* cells, LR-recombination reactions were performed with 75ng of donor and 75ng of acceptor in a 5µl reaction, according to the manufacturer's protocol. Plasmids for transfection were prepared by Qiagen maxiprep.

**Cell Culture:** *Drosophila* Schneider 2 (S2) cells (Expression Systems, 94-005F) were cultured in ESF media (ESF 921 Insect Cell Culture Medium, Expression Systems) with 5% fetal bovine serum (FBS, Weisent) at room temperature on plates. Cells were passaged every 2 to 3 days. night before transfection, 1.5e6 S2 cells were plated per well of a 6-well plate. The next day, the media was changed, and transfection mix was added. Mirus Transit insect Transfection Reagent (Mirus bio) was used for the transfections according to the manufacturer's protocol. To mark nuclei, pAct5C-H2A-RFP (gift of V. Archambault) was co-transfected with Venus constructs. The day after transfection, the media was changed, and the next day, cells were replated on a ConA-coated glass-bottom imaging dish. Cells were heat shocked for 8 min at 37˚C to induce Ph expression, and used the next day for live imaging.

**Western Blotting:** 500,000 S2 cells per well were plated in a 24 well plate (Corning 353047) one day prior to transfection. Fresh media was added the next day and transfection was carried out as above. The day after, the media was changed and the evening of next day, cells were heat shocked for 8 min at 37°C. Post 24 hours, transfected S2 cells were counted and 500,000 cells were centrifuged at 2,500 rpm at 4°C for 5 min. Pellets were re-suspended in 70 µl 2X SDS-PAGE buffer (232µl/ml Tris pH 6.8, 100µl/ml glycerol, 34mg/ml sodium dodecyl sulfate (SDS), 120mg/ml bromophenol blue) and boiled for 5min. Samples were then run on 8% and 16% SDS-PAGE gels for Ph and H2A-RFP respectively for 80 min at 120 Volts, and transferred to nitrocellulose membranes.

Membranes were blocked for 30 min in 5% milk/PBST (1XPBS, 0.3% TWEEN® 20) and incubated overnight at 4°C on a shaker in primary antibody diluted in 5% milk/PBST. Primary antibodies used are as follows: anti-α-tubulin (mouse, 1:3 000, Sigma Aldrich T5168), anti-Ph (rabbit, 1:3 000, Francis lab), anti-GFP (rabbit, 1:3 000, Protein tech 50430-2-AP), anti-RFP (rabbit, 1:3 000, St. Johns Laboratory STJ97083) and anti-H2B (mouse, 1:3 000, Abcam, ab52484). Membranes were washed 3 times for 10 min each in PBST, incubated for 2 hrs in secondary antibody diluted in 5% milk/PBST and washed 3 times again for 10 min in PBST. Secondary antibodies were conjugated to Alexa Fluor 680 (anti-rabbit and anti-mouse,

Invitrogen A21076 and A21057 respectively) or 800 CW (anti-Rabbit, Li-Cor) and used at 1:25 000 in 5% milk/PBST. Blots were scanned on an Odyssey CLx imager.

**Image Acquisition:** Live images were acquired on a Zeiss microscope, equipped with a Yokogawa CSU-1 spinning-disk confocal head. A 63x oil objective was used and the software for image acquisition was Zen 2012. The excitation wavelengths for Venus and RFP were 488, 561 nm, respectively. For the VENUS channel, the laser power and exposure were set as follows 2.40% and 77mS, respectively. For the RFP it was 11% and 500mS, respectively. For fixed cell imaging, a Leica SP8 confocal microscope was used. The 63x/1.40 oil objective was used in all cases for channels EVENUS (gain 900 laser power 13%), DAPI (gain 930 laser power 6%) and Alexa 647 (gain 25 laser power 2 percent) with excitation wavelengths as follows 488, 405 and 638nm, respectively. For live imaging, 3*3 tiles of confocal stacks of 1µm slices were collected.

**Image Processing and Analysis:** Live images were opened using ImageJ (Fiji) in tiff format. A maximum intensity projection was made for each image. Images were then split into VENUS and RFP channels, which were named as foci and nuclei, respectively. These were then uploaded to Cell Profiler (3.1.9). Cell Profiler modules were used to build the basic analysis pipeline but a few parameters were modified for different constructs. The basic pipeline modules were identification of nuclei from RFP staining and foci identification from the VENUS channel. The module IdentifyPrimaryObjects was used for the identification of the nucleus, which was named "nuclei". Objects outside a diameter range and touching the image border were excluded. Global thresholding was applied with either Minimum Cross Entropy or Otsu algorithms. Identification of foci was done using the IdentifyPrimaryObjects module with thresholding using Otsu or Minimum Cross Entropy, and named "foci". The size of the smoothing filter and the distance between local maxima parameters were adjusted for each construct separately to segment clumped objects. In the case of constructs like PhΔ1 where extensive clumping was observed, EnhanceorSuppressFeatures was used to ignore highly clumped objects. The RelateObjectsModule was used to count the number of foci per nucleus. MeasureObjectIntensity was used to obtain mean intensity values. For the mean VENUS intensity, the nuclei objects corresponding to the nuclei images were selected. MeasureObjectSizeShape provided the measurements of size which is the total number of pixels within the object and circularity. The object selected in this case was foci. For condensates that have complex morphologies, are small, or are faint relative to surrounding levels (such as those formed by PhΔ1), the software does not detect them well. To confirm that the number counted by Cell Profiler is not vastly different from reality, manual foci counts were compared to Cell Profiler generated ones for several images. In most cases, the counts differed by less than 3 foci per cell.

For live image analysis, at least 100 cells were analyzed per experiment for each construct. All observations were made from at least three experiments. Statistics were calculated using GraphPad Prism v8.4.3, using recommended settings, including for correction for multiple comparisons. p values were calculated for Kruskal–Wallis tests with Dunn's multiple-comparison correction.

**DATA AVAILABILITY**: The datasets generated during the current study are available from the corresponding author on reasonable request.

**REFERENCES:**


1. 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).
2. Larson, A. G. & Narlikar, G. J. The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. *Biochemistry* **57**, 2540–2548 (2018).
3. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
4. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420–435 (2018).
5. Lasker, K. *et al.* A modular platform for engineering function of natural and synthetic biomolecular condensates. *bioRxiv* 2021.02.03.429226 (2021) doi:10.1101/2021.02.03.429226.
6. Swain, P. & Weber, S. C. Dissecting the complexity of biomolecular condensates. *Biochem. Soc. Trans.* **48**, 2591–2602 (2020).
7. 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).
8. Martin, E. W. & Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **57**, 2478–2487 (2018).
9. Protter, D. S. W. *et al.* Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep.* **22**, 1401–1412 (2018).
10. Mittag, T. & Parker, R. Multiple Modes of Protein–Protein Interactions Promote RNP Granule Assembly. *J. Mol. Biol.* **430**, 4636–4649 (2018).
11. Yang, P. *et al.* G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345 e28 (2020).
12. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147–55 (2013).
13. Loubiere, V., Martinez, A. M. & Cavalli, G. Cell Fate and Developmental Regulation Dynamics by Polycomb Proteins and 3D Genome Architecture. *Bioessays* **41**, e1800222 (2019).
14. Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**, 808–24 (2013).
15. Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).
16. Cheutin, T. & Cavalli, G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genet Dev* **25**, 30–7 (2014).
17. Cheutin, T. & Cavalli, G. Loss of PRC1 induces higher-order opening of Hox loci independently of transcription during Drosophila embryogenesis. *Nat Commun* **9**, 3898 (2018).
18. Kim, J. & Kingston, R. E. The CBX family of proteins in transcriptional repression and memory. *J. Biosci.* **45**, (2020).
19. Kim, C. A., Gingery, M., Pilpa, R. M. & Bowie, J. U. The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* **9**, 453–7 (2002).
20. Robinson, A. K. *et al.* The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. *J Biol Chem* **287**, 8702–13 (2012).

21. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* **26**, 565–77 (2013).
22. Wani, A. H. *et al.* Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat Commun* **7**, 10291 (2016).
23. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–22 (2016).
24. Kundu, S. *et al.* Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell* **65**, 432-446 e5 (2017).
25. Gambetta, M. C. & Muller, J. O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. *Dev Cell* **31**, 629–39 (2014).
26. Seif, E. *et al.* Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nat. Commun.* **11**, 5609 (2020).
27. Eeftens, J. M., Kapoor, M. & Brangwynne, C. P. Epigenetic memory as a time integral over prior history of Polycomb phase separation. *bioRxiv* 2020.08.19.254706 (2020) doi:10.1101/2020.08.19.254706.
28. Dura, J.-M. *et al.* A complex genetic locus, polyhomeotic, is required for segmental specification and epidermal development in D. melanogaster. *Cell* **51**, 829–839 (1987).
29. Hodgson, J. W. *et al.* The polyhomeotic locus of Drosophila melanogaster is transcriptionally and post-transcriptionally regulated during embryogenesis. *Mech Dev* **66**, 69–81 (1997).
30. Mészáros, B., Erdos, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
31. Jarnot, P. *et al.* PlaToLoCo: the first web meta-server for visualization and annotation of low complexity regions in proteins. *Nucleic Acids Res.* **48**, W77–W84 (2020).
32. Promponas, V. J. *et al.* CAST: an iterative algorithm for the complexity analysis of sequence tracts. *Bioinformatics* **16**, 915–922 (2000).
33. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinforma. Oxf. Engl.* **30**, 2501–2502 (2014).
34. Riback, J. A. *et al.* Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214 (2020).
35. Xing, W., Muhlrad, D., Parker, R. & Rosen, M. K. A quantitative inventory of yeast P body proteins reveals principles of composition and specificity. *eLife* **9**, (2020).
36. Feng, S., Huang, J. & Wang, J. Loss of the Polycomb group gene polyhomeotic induces non-autonomous cell overproliferation. *EMBO Rep* **12**, 157–63 (2011).
37. Gambetta, M. C., Oktaba, K. & Muller, J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* **325**, 93–6 (2009).
38. Nosella, M. L. *et al.* O-Linked-N-Acetylglucosaminylation of the RNA-Binding Protein EWS N-Terminal Low Complexity Region Reduces Phase Separation and Enhances Condensate Dynamics. *J. Am. Chem. Soc.* **143**, 11520–11534 (2021).
39. Qi, Y. & Zhang, B. Chromatin Network Retards Droplet Coalescence. *bioRxiv* 2021.03.02.433564 (2021) doi:10.1101/2021.03.02.433564.
40. Söding, J., Zwicker, D., Sohrabi-Jahromi, S., Boehning, M. & Kirschbaum, J. Mechanisms for Active Regulation of Biomolecular Condensates. *Trends Cell Biol.* **30**, 4–14 (2020).

41. Zhang, Y., Lee, D. S. W., Meir, Y., Brangwynne, C. P. & Wingreen, N. S. Mechanical Frustration of Phase Separation in the Cell Nucleus by Chromatin. *Phys. Rev. Lett.* **126**, 258102 (2021).
42. Folkmann Andrew W., Putnam Andrea, Lee Chiu Fan, & Seydoux Geraldine. Regulation of biomolecular condensates by interfacial protein clusters. *Science* **373**, 1218–1224 (2021).
43. Peng, Z., Wang, C., Uversky, V. N. & Kurgan, L. Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. *Methods Mol. Biol. Clifton NJ* **1484**, 187–203 (2017).

**ACKNOWLEDGEMENTS:** This work was funded by a grant from the Canadian Institutes for Health Research (CIHR) to NJF. IK thanks members of the Francis lab, particularly Djamouna Sihou, for critical input on the project.

**AUTHOR CONTRIBUTIONS:** IK prepared constructs, carried out transfections, imaging, all data analysis, prepared figures and contributed to writing the manuscript. ELB prepared and validated constructs, and provided input on the manuscript. NJF supervised the project, collected data, wrote the manuscript and provided funding. IK and NJF conceived the project.

**COMPETING INTERESTS:** The authors declare no competing interests.

**FIGURE LEGENDS:**

**Figure 1 Identification of three IDRs in Ph.** (**A**) Prediction of Ph disordered region using IUPRED[30], and PONDR-VSL2 A large part of Polyhomeotic is predicted to be disordered, as defined by predicted disorder propensity above 0.5 (shown by the red bar). (**B**) SEG/CAST analysis of Ph using PlaToLoCo [31] to determine boundaries in the disordered region by looking at the position of masked (repetitive) amino acids in the sequence. (**C**) Determination of different regions in the Ph disordered region defined as Ph1(IDR1), Ph2(IDR2) and Ph3(IDR3). (**D**) Frequency of each amino acid in the disordered region. (**E**) Prion like domain prediction using PLAAC[33]. IDR 2 is dominated by high scoring prion-like regions.

**Figure 2 Effect of Ph IDRs on condensate formation in cells. (A)** Schematic of Polyhomeotic. **(B-C)** Representative images of live S2 cells that were co-transfected with Venus-Ph with a single **(B)** or two **(C)** IDRs deleted. Ph (wild-type) is show at the top of B, and mini-Ph at the top of C for comparison. H2A-RFP was co-transfected as a nuclear marker. Images are maximum intensity projections of confocal stacks. **(D-E)** Effect of IDR deletions on foci number per nucleus **(D)** and size **(E).** Bars show the median value for a total of at least 50 cells that formed condensates, pooled from each of the three independent experiments. The total number of transfected cells analyzed (with and without foci) are as follows: WT-Ph, n=3478; PhΔ1, n=2426; PhΔ2, n=1539; PhΔ3, n=1557: PhΔ1Δ3, n=1568; PhΔ2Δ3, n=2911; PhΔ1Δ2, n=1038. p-values are for comparison with WT using a Kruskal-Wallis test with Dunnett's correction for multiple comparisons.

**Figure 3 Ph proteins lacking IDRs are (over)expressed as full-length proteins. (A)** Fraction of transfected cells (identified by H2A-RFP) that formed condensates (pooled data from three experiments). **(B-D)** Representative Western blots of cells transfected with Venus-Ph proteins. **(B)** Venus-Ph proteins, detected with anti-GFP. **(C)** H2A-RFP, detected with anti-RFP, to measure transfection efficiency. **(D)** Ph, detected with anti-Ph to measure total Ph levels (endogenous + transfected). Asterisks next to indicate the position of endogenous Ph bands. Note that the antibody epitope is in Ph2 so that proteins lacking this region are not detected. Venus-PhΔ1 and Venus-PhΔ3 migrate similarly to Ph-d. **(E-G)** Quantification of Western blots. GFP and Ph were normalized to tubulin, and H2A-RFP to H2B. Each sample was then normalized to the signal for WT-Ph in the same experiment, or to untransfected cells (G) (to measure endogenous Ph levels). Bars show the mean, and error bars the SEM. n=3 (E, G); n=2 (F). **(H)** Mean nuclear intensities (from analysis in Fig. 4) for Venus-Ph proteins lacking one or two IDRs. Error bars show 95% CI. Nuclear intensities are much less variable than total protein levels (compare with E). **(I)** Total Venus-Ph levels determined by Western blot are not correlated with mean nuclear intensity. Median nuclear intensities were normalized to WT.

**Figure 4 Relationship between expression levels and condensate formation for Ph proteins lacking IDRs. (A-F)** Nuclear intensity (in AU) was measured in S2 cells with and without foci for each construct. Cells containing condensates have higher expression levels than those without, but there is no clear threshold for foci formation. Black bars indicates the median number of cells. Data are pooled from three experiments.

**Figure 5 Effect of Ph IDRs on condensate formation in the absence of the SAM. (A)** Representative live images of S2 cells transfected with the Venus-Ph deletion constructs lacking

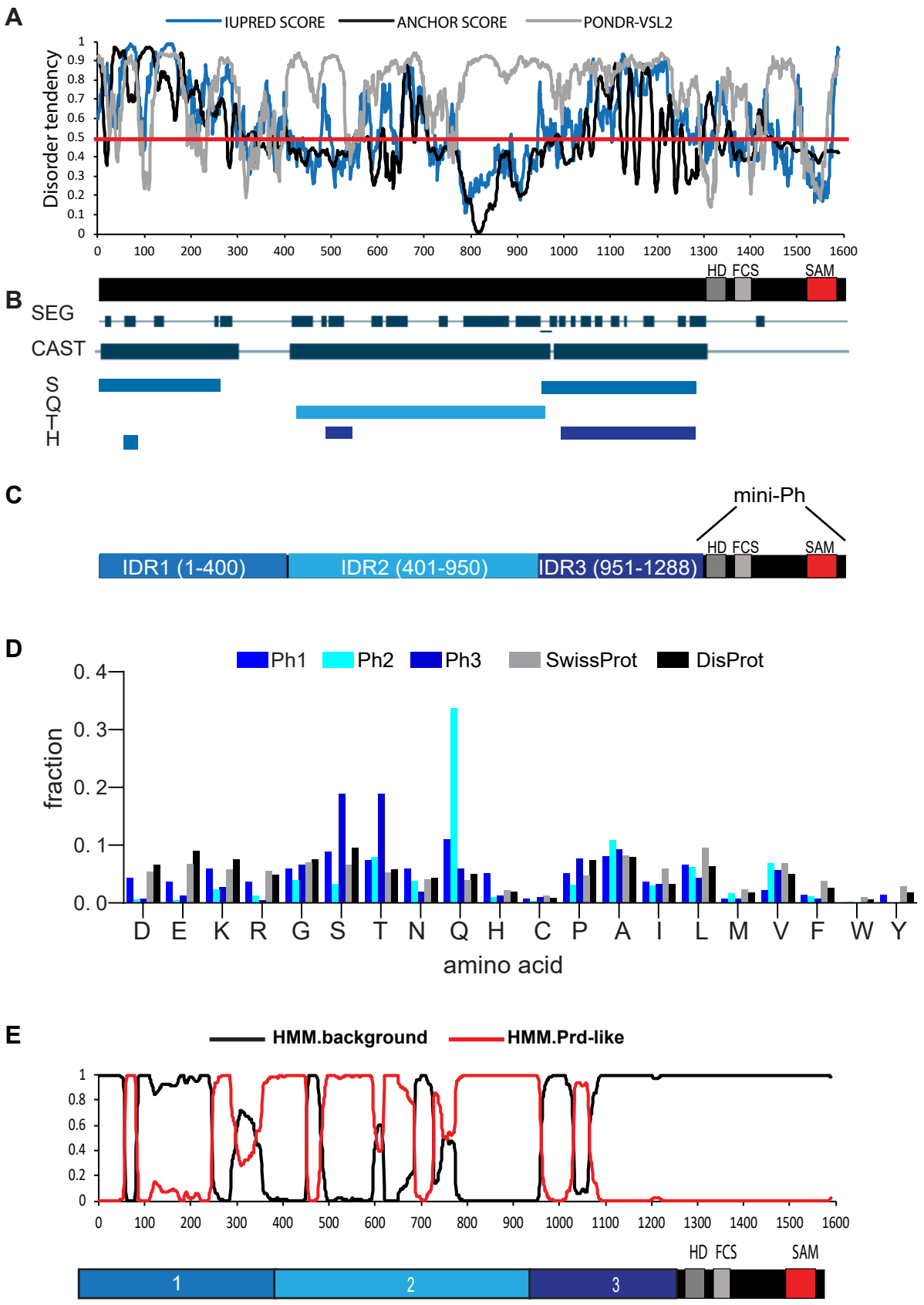
the SAM with H2A-RFP as a nuclear marker. Images show maximum intensity projections of confocal stacks. **(B)** Quantification of foci count for H2A-RFP positive cells. Graph shows quantification for cells with and without foci. A total of at least 100 cells from each of the three independent experiments were analyzed. The total number of cells with and without foci were as follows: PhΔSAM, n=1348; PhΔ1ΔSAM, n=411; PhΔ2ΔSAM, n=359; PhΔ3ΔSAM, n=214; PhΔ1Δ3ΔSAM, n=922; PhΔ2Δ3ΔSAM, n=1413. p-values are for comparison with PhΔSAM using a Kruskal-Wallis test with Dunnett's correction for multiple comparisons.

**Figure 6 Ph IDRs alone can form foci. (A)** Representative live images of S2 cells transfected with the Venus-Ph truncation constructs (IDRs only) with H2A-RFP as a nuclear marker. Images show maximum intensity projections of confocal stacks. **(B)** Quantification of foci count for truncations. Bars show the median value for a total of at least 50 cells that formed foci, pooled from each of the three independent experiments. The total number of transfected cells (with and without foci) were as follows: Ph1, n=268; Ph2, n=412; Ph3, n=309; Ph5, n=857; Ph6, n=460; Ph7, n=175. p-values are for comparison with WT-Ph using the Kruskal-Wallis test with Dunnett's correction for multiple comparisons.

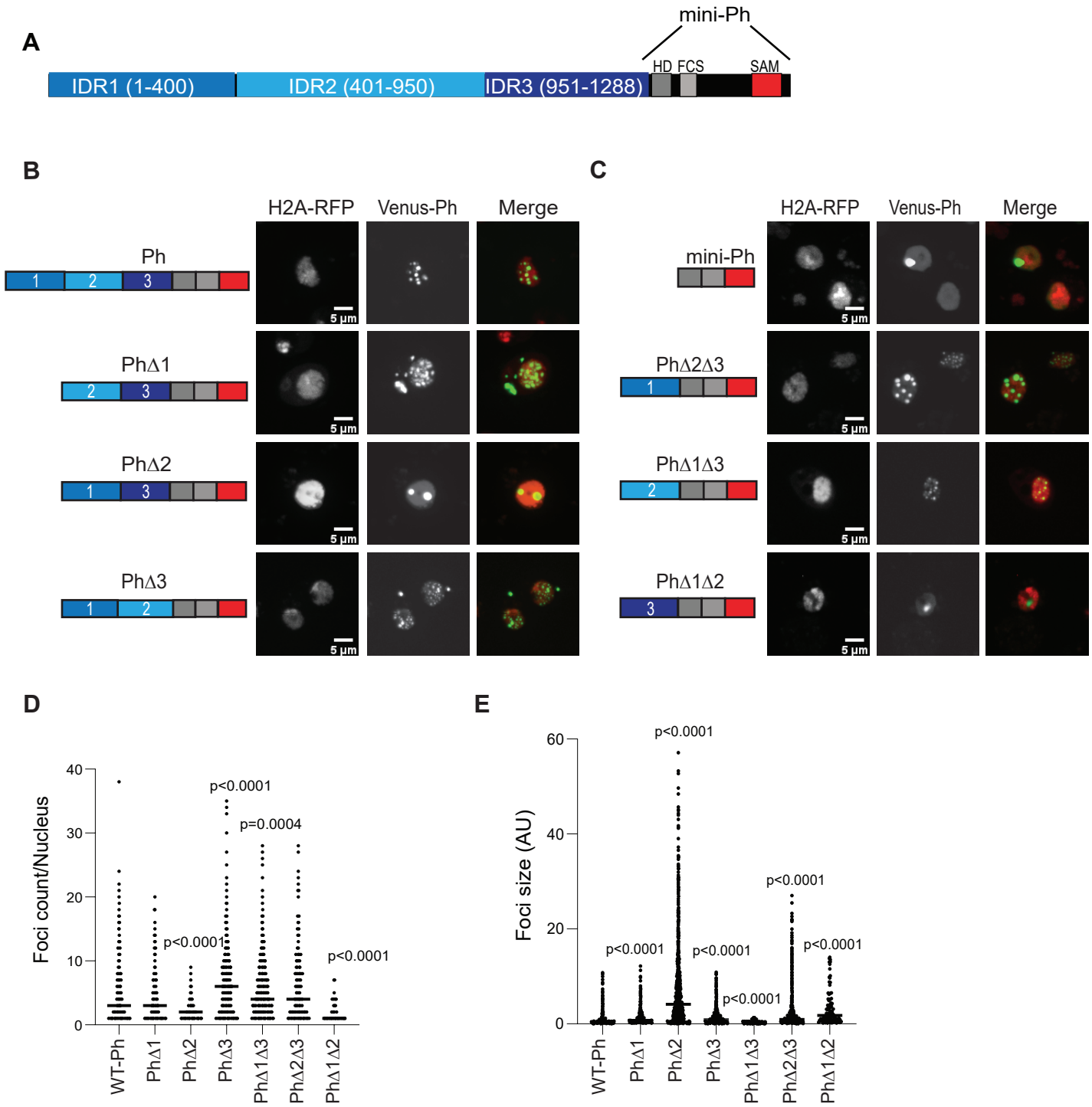
**Figure 7 Quantification of Ph IDR expression levels in transfected cells. (A-C)** Representative Western blots of cells transfected with Venus-Ph IDR constructs. anti-GFP is used to quantify Venus (A) and anti-RFP to quantify H2A-RFP (B) (to measure transfection efficiency). anti-Ph (C) indicates the amount of IDR+ Ph in cells. Because the antigen for the anti-Ph antibody is in Ph2, only proteins containing this region can be measured (Ph2, Ph5, Ph6, Ph7). **(D-E)** Quantification of Western blot signals. D and F are normalized to tubulin, and E to H2B. Signals are then normalized to WT-Ph from the same experiment (D, E), or untransfected (F) (endogenous Ph levels). Bars are mean and error bars are SEM. n=3 (D); n=2 (E, F).

**Figure 8 Model for role of Ph IDRs in SAM dependent condensate formation. (A)** Ph IDRs are required for normal condensate formation in cells. **(B)** Working model for regulatory network among IDRs. Ph2 (gray arrow) restricts both the size of condensates and propensity for formation when Ph SAM is present, but promotes condensates independent of the SAM. Ph3 functions through inhibition of Ph1 and Ph2, which together promote condensate formation; it may also inhibit the SAM.

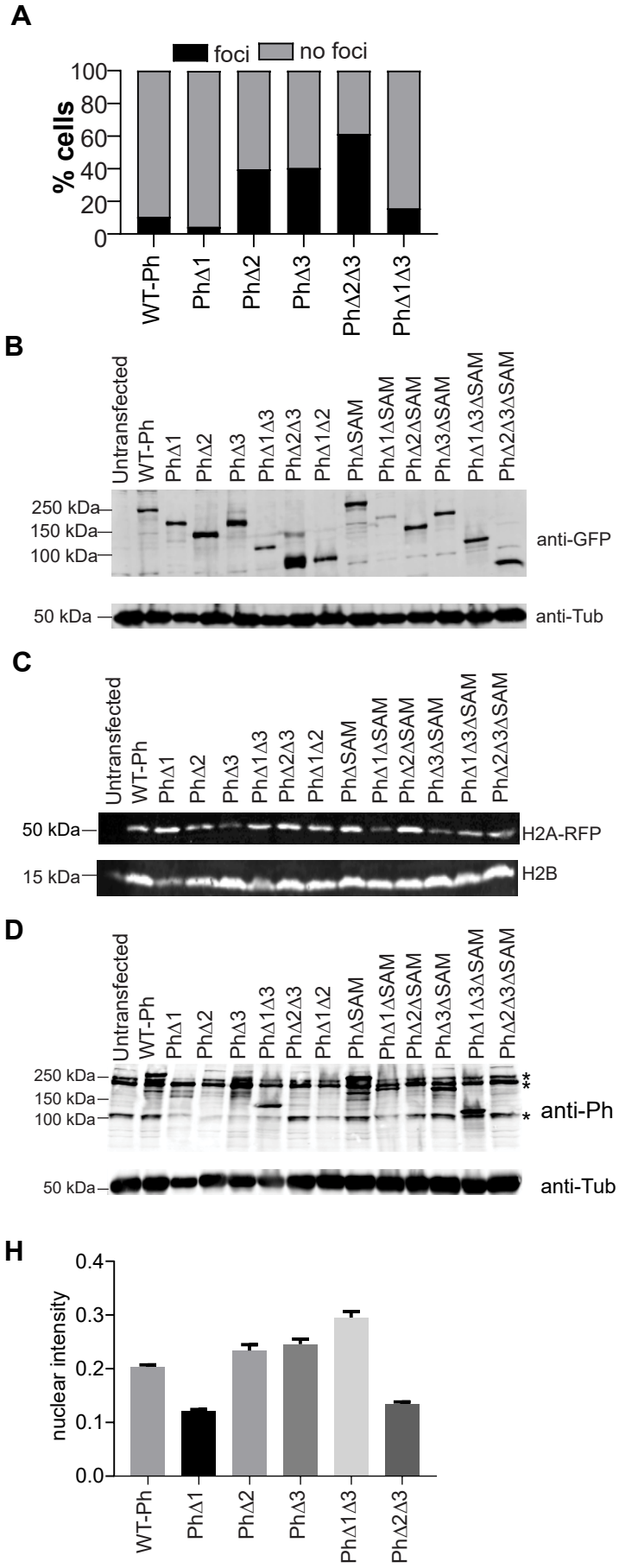
**Table 1 Summary of condensate formation by different Ph proteins.**



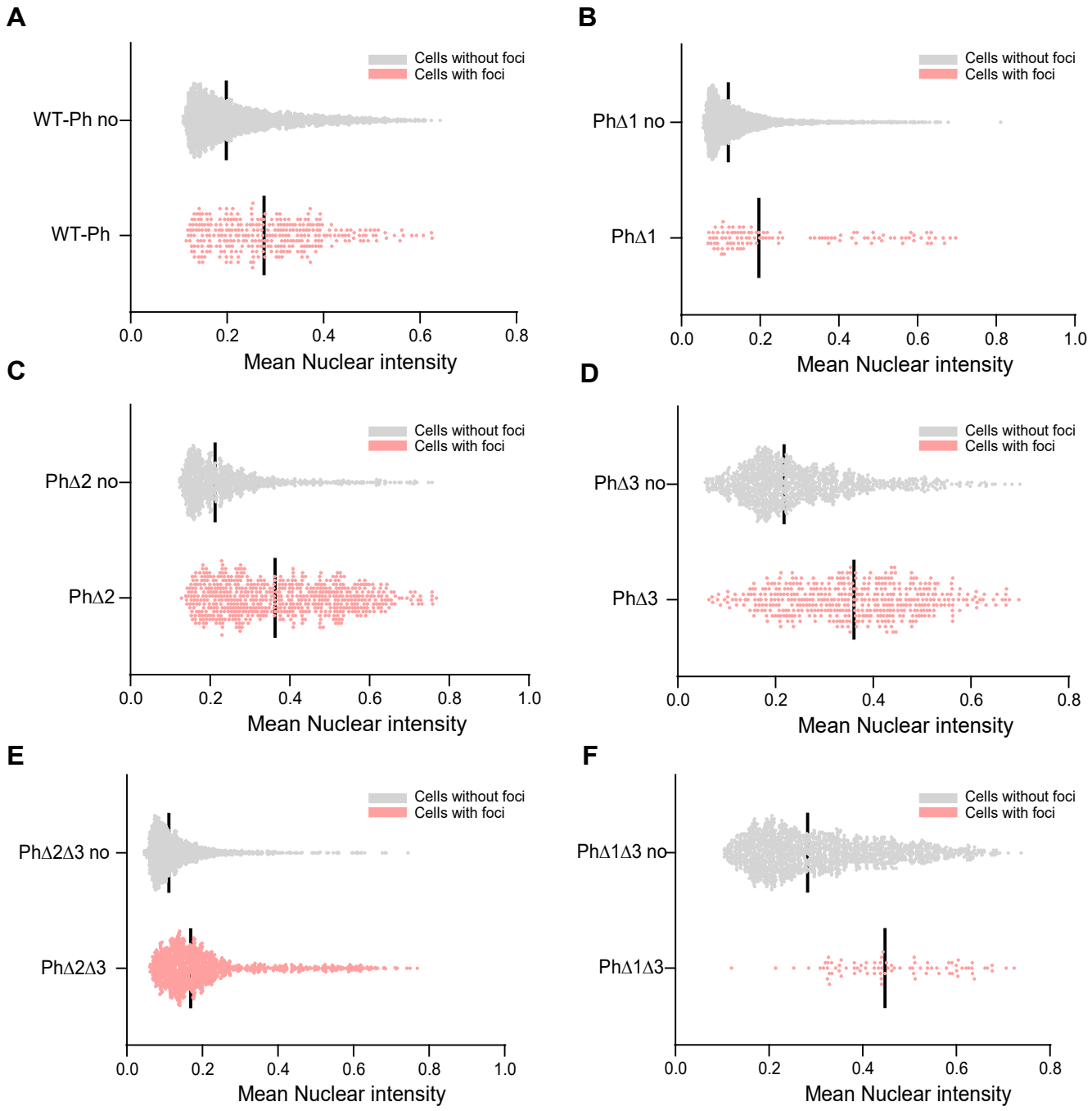

Kapur et al., Figure 1



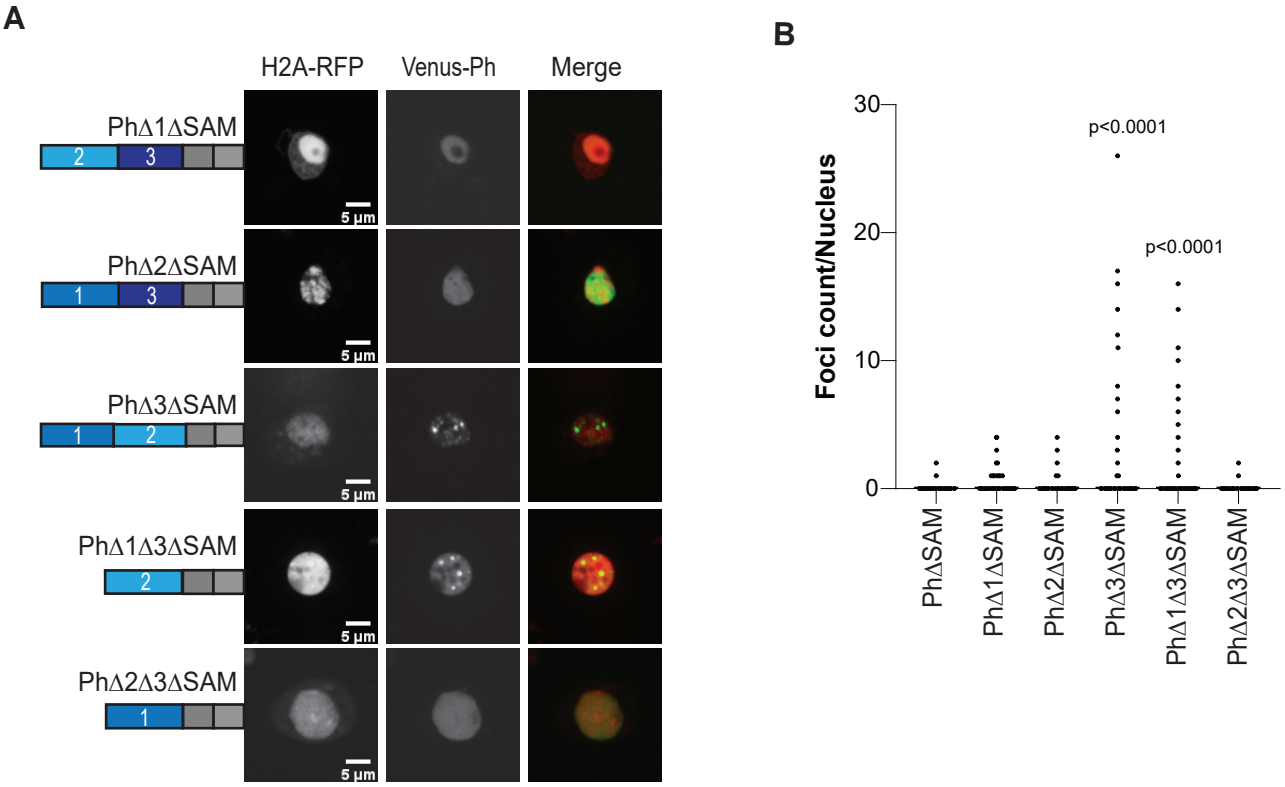

Kapur et al., Figure 2



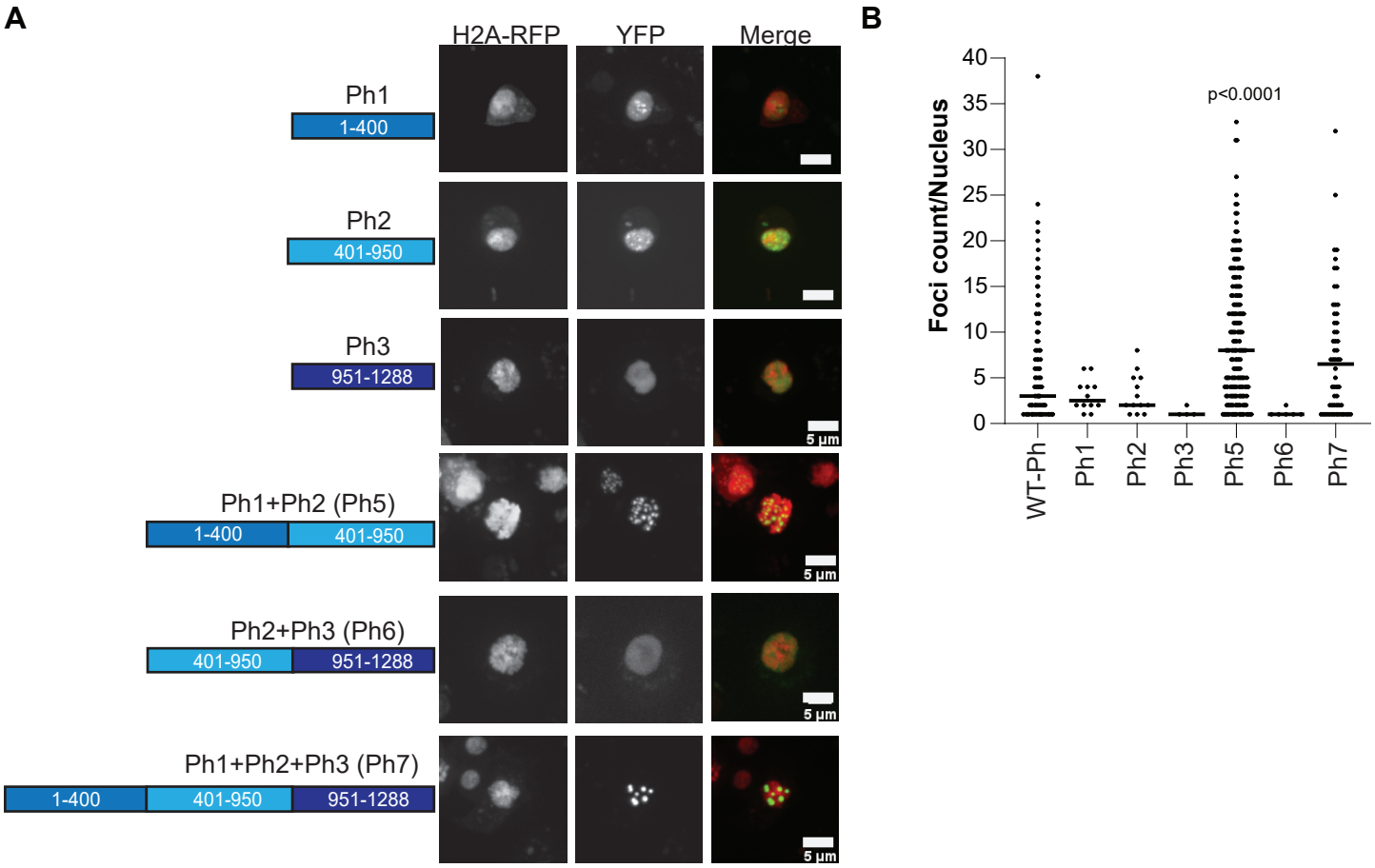

Kapur et al., Figure 3



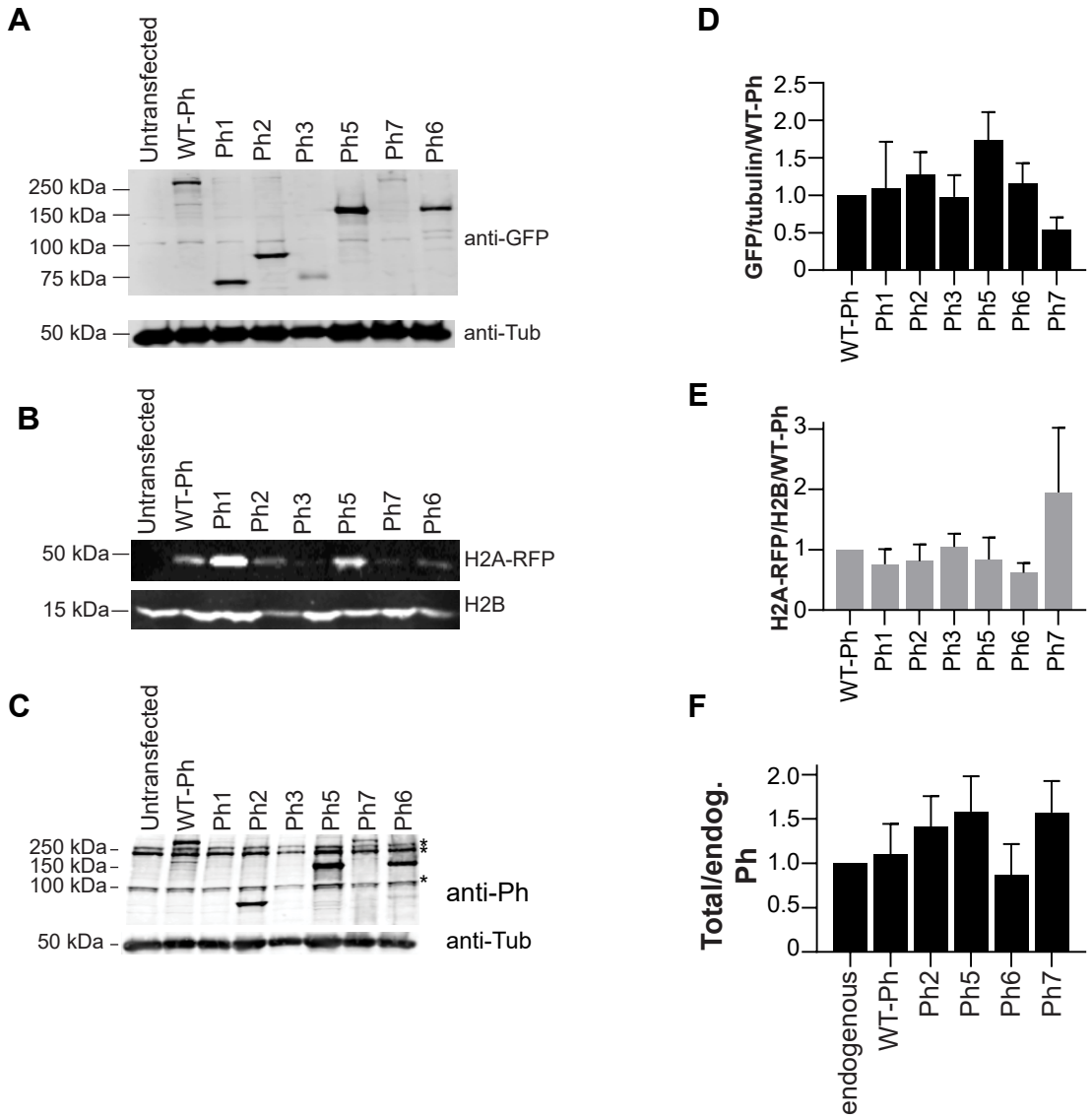

Kapur et al., Figure 4



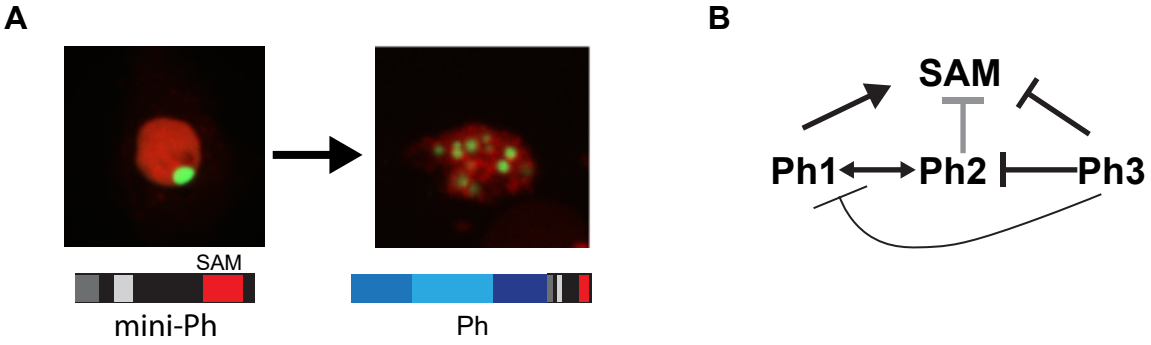

Kapur et al., Figure 5




Kapur et al., Figure 6




Kapur et al., Figure 7




Kapur et al., Figure 8

| CONSTRUCT | SAM | CONDENSATES |
|---|---|---|
| PhΔ1 | + | + |
| PhΔ2 | + | + |
| PhΔ3 | + | + |
| PhΔ1Δ3 | + | + |
| PhΔ2Δ3 | + | + |
| PhΔ1Δ2 | + | - |
| PhΔ1ΔSAM | - | - |
| PhΔ2ΔSAM | - | - |
| PhΔ3ΔSAM | - | + |
| PhΔ1Δ3ΔSAM | - | + |

Kapur et al., Table I