Intrinsically disordered sequences in the Polycomb protein Polyhomeotic regulate condensate formation

Ibani Kapur

Division of Experimental Medicine,

Faculty of Medicine,

McGill University, Montreal

August 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Abstract

The Polycomb group (PcG) complex PRC1 localizes in the nucleus in the form of condensed structures called Polycomb bodies. Polyhomeotic (Ph) which is one of the subunits of PRC1, contains a polymerizing sterile alpha motif (SAM) that is implicated in both PcG body formation and chromatin organization in *Drosophila* and mammalian cells. Liquid Liquid Phase separation is increasingly appreciated as a mechanism by which protein-nucleic acid condensates may form in cells, including to organize chromatin. A truncated version of Ph containing the SAM and two other small, conserved domains, referred to as mini-Ph, forms phase separated condensates with DNA or chromatin *in vitro*. However, in cells, full length Ph forms multiple condensates, while mini-Ph either does not form condensates or forms a single large nuclear condensate. Thus, our hypothesis is that sequences outside of mini-Ph are required in cells for proper condensate formation.

We analyzed Ph sequence and found that regions outside of mini-Ph are predicted to be disordered. Analysis of sequence composition and complexity allowed us to define three distinct intrinsically disordered regions (IDRs). To understand the mechanistic role of these IDRS, the activity of each IDR, and its contribution to regulation of Ph distribution in cells was determined by live imaging of Drosophila S2 cells expressing Venus-tagged truncated Ph proteins. We found that each IDR tunes condensate size, number, and morphology. PcG regulation can be heritable through cell cycles (epigenetic). It is therefore possible that condensate formation driven by Ph SAM is important for epigenetic memory, but little is known about how condensates are regulated through the cell cycle. We show that condensates are present throughout S-phase and tend to increase in size as the cell cycle progresses. In mitosis, they progressively dissociate

from chromatin through anaphase and reform on/reassociate with chromatin in the end stages of mitosis and G1.

Résumé du projet

Au sein des noyaux cellulaires, le complexe Polycomb PRC1 forme des structures condensées appelées corps Polycomb. Parmi les sous-unités de PRC1, Polyhomeotic (Ph) contient le domaine « sterile alpha motif » (SAM) qui est capable de se polymériser et est impliqué dans la formation des corps Polycomb et l'organisation de la chromatine chez de nombreux organismes tels que la Drosophile et les mammifères. La séparation de phase liquide-liquide est de plus en plus considérée comme mécanisme conduisant à la formation de condensats de protéines et d'acides nucléiques, notamment lors de l'organisation de la chromatine. Une version tronquée de Ph, appelée mini-Ph et contenant SAM ainsi que deux autres domaines conservés, est capable de former des condensats avec l'ADN ou la chromatine in vitro. Cependant, en cellules, Ph forme de multiples condensats, alors que mini-Ph ne forme pas de condensats ou un unique condensat de taille supérieure. Ainsi, nous émettons l'hypothèse que la formation de condensats dans les cellules requiert des régions peptidiques autres que mini-Ph.

L'analyse de séquence de Ph prédit que les régions autres que mini-Ph sont désordonnées. La composition et la complexité de ces séquences nous ont permis de définir trois régions intrinsèquement désordonnées (IDRs) distinctes. Pour comprendre leur mécanisme, leur activité respective et leur contribution à la régulation de Ph, leur distribution a été déterminée par microscopie dans des cellules vivantes de Drosophile exprimant différentes troncations de Ph taguées par la protéine fluorescente Venus. Nous avons déterminé que chaque IDR régule la taille, le nombre et la morphologie des condensats.

La régulation des PcG peut être héritée pendant plusieurs cycles cellulaires (épigénétique). Il est donc possible que la formation des condensats formés via le domaine SAM de Ph soit importante pour la mémoire épigénétique, mais peu est connu à propos de la régulation de ces condensats au

cours du cycle cellulaire. Nous avons montré que les condensats sont présents en phase S et ont tendance à augmenter en taille lors de la progression du cycle cellulaire. En mitose, ils se dissocient progressivement de la chromatine lors de l'anaphase et se reforment sur/se réassocient avec la chromatine à la fin de la mitose et la phase G1.

Acknowledgements

This thesis would not have been possible without the help of several people. Firstly, I am extremely grateful to my supervisor, Dr. Nicole Francis, for her guidance, support, and insightful advice throughout my course. Under her mentorship, I had the opportunity to grow as a well-rounded researcher by learning not just the science but also patience, hard work and discipline.

Secondly, all the members of Francis lab played a pivotal role in my journey. I am thankful to Elodie Boulier for teaching me experimental techniques and how to be better organized. Her radiant uplifting energy and positivity made the lab a better place to work at. She was always there for any help or advice (or sometimes reagents when I couldn't find them despite them being right in front of my eyes!) I needed.

I am also thankful to Djamouna Sihou for her valuable advice and suggestions on cloning or any other curiosity driven scientific question. Her perseverance and strength inspired me to grow as a person.

I would also like to thank previous lab member Aurelie Huang Sung who always offered useful technical advice on image analysis.

Lastly, I am forever thankful to my family, Veekay and friends turned family, Dhara and Lynda who supported me throughout my masters. To my family, I owe a debt of gratitude for always believing in me, sacrificing for me, and supporting me in every way they can. I would not be where I am today without them.

Contribution of Authors

Chapter 3

Plasmids were provided by lab member Elodie Boulier except for the ones mentioned in **table 4** that were prepared by me.

Chapter 4

Live imaging of mitotic S2 cells was done in the lab previously. All other experiments were performed by me including the analysis.

All the chapters in the manuscript were written by me including the preparation of figures except for **Figure 14B** which was prepared by Dr. Francis.

The editorial review was done by Dr. Francis.

Table of Contents

List of Abbreviations

3C: Chromosome Conformation Capture

AEBP2: AE Binding Protein 2

ANT-C: Antenapedia Complex

BX-C: Bithorax Complex

Caf1-55: Chromatin assembly factor 1, p55 subunit

CAST: Complexity Analysis of Sequence Tracts

CBX: Chromobox proteins

Csat: saturation concentration

CTCF: CCCTC-Binding Factor

DisoRDPbind: predictor of disorder-mediated RNA, DNA and protein binding regions

Draf: DRING-associated factors

E(z): Enhancer of Zeste

EdU: 5-Ethynyl-2′-deoxyuridine

EH: End Helix

Esc: extra sex comb

EZH1/2: Enhancer of zeste homologue 1/2 (EZH1/2)

FISH: Fluorescent In Situ Hybridization

H2AK119ub: mono-ubiquitinated lysine residue 119 of H2A

H3K27me3: trimethylated lysine residue 27 of histone 3

H3K36me3: trimethylation of histone H3 on lysine K36

H3K4me3: trimethylation of histone H3 on lysine K4

H3K9me3: trimethylated lysine residue 9 of histone 3

H3S10p: phosphorylation of serine 10 on histone H3

HF: High Fidelity

HLB: Histone Locus bodies

HP1: Heterochromatin Protein 1

IDR: Intrinsically Disordered Region

KDM2: Histone lysine demethylase 2

LCDR: Low Complexity Disordered Region

LCR: Low Complexity Region

LLPS: Liquid Liquid Phase Separation

MED1: Mediator complex subunit 1

ML: Mid Loop

MLOs: Membraneless Organelles

MORC3: Microrchidia

Nc: non canonical

NPM1: Nucleophosmin 1

Ogt: protein: Protein O-GlcNAc transferase

PAB1: PolyA Binding Protein

Pc: *Polycomb*

PcG: Polycomb Group

PCGF: Polycomb group RING fingers

PCL1-3: Polycomb-like proteins 1–3

PCR: Polymerase Chain Reaction

Ph: Polyhomeotic

PHC: Polyhomeotic Homolog

ph-d: polyhomeotic distal

Pho: Pleiohomeotic

PhoL: Pho-like

PhoRC: Pho repressive complex

ph-p: polyhomeotic proximal

PlaToLoCo: Platform of Tools for Low Complexity

PML: Promyelocytic leukemia

PRC1: Polycomb Repressive Complex 1

PRC2: Polycomb Repressive Complex 2

PRE: Polycomb Response Elements

Psc: Posterior sex combs

PTM: Post Translational Modification

RAWUL: Ring-finger and WD40 associated Ubiquitin-Like

RbAp46/48: Retinoblastoma protein associated protein 46

RYBP: Ring and YY1 Binding Protein

S2: Schneider 2

SAM: Sterile Alpha Motif

Sce or dRING: Sex combs extra

Sfmbt: Scm-related gene containing four mbt domains

Su(z)12: Suppressor of Zeste 12

SURF6: Surfeit locus protein 6

TADs: Topologically associated domains

TDP-43: TAR DNA-binding protein 43

TrxG: Trithorax Group

YAF1: YY1-associated factor enhancer of zeste homologue 2 (EZH2)

List of Figures

List of Tables

Chapter 1: Literature Review and Introduction

1.1 Epigenetic Inheritance

The earliest definition of the term epigenetics coined by Conrad Waddington, was based on the study of epigenesis: how phenotypes result from genotypes during development.[1,2] In other words, even though an identical genotype is shared among most cells, various cell types arise during differentiation, due to difference in their gene expression patterns. These gene expression patterns are what determine cell function and identity.[3] They are largely governed by an epigenetic landscape as opposed to genetic inheritance. The epigenome relies on chromatin structure. Chromatin is composed of repeating units of histone proteins assembled on DNA, termed nucleosomes. Nucleosomes occupy 147 bp of DNA, but chromatin is further organized at scales from kilobase to megabase.[4] It can be divided into heterochromatin (inactive) and euchromatin (active) and is associated with unique DNA methylation patterns, histone post-translational modifications (PTMs), Polycomb group (PcG) proteins, non-coding RNAs and several chromatin remodelling factors.[5] Heterochromatin has two types, constitutive and facultative. Facultative heterochromatin is repressed by a group of proteins named PcG and characterized by H3K27me3, while constitutive heterochromatin is marked with H3K9me3, and enriched for heterochromatin proteins including Heterochromatin Protein 1 (HP1), which recognizes H3K9me3.[6] Constitutive heterochromatin constitutes repetitive and noncoding genomic sequences and facultative heterochromatin though transcriptionally silent, can switch between heterochromatin and euchromatin regions.[7] In conclusion, a major facet of epigenetics relies on histone modifications, and the plethora of chromatin proteins.[8]

The definition of the term epigenetics has evolved over time.[9–11] Today's widely accepted definition identifies the study of epigenetics as heritable changes that occur on the genome without any changes to the DNA sequence. Also referred to as epigenetic inheritance, it is a means by which epigenetic marks once established, can be transmitted from mother to daughter cells across generations.[11] However, chromatin organization is challenged during cell cycle, as the processes of DNA replication and mitosis lead to disruption of nuclear architecture.[12] Epigenetic signatures must be re-established after DNA replication to propagate chromatin states is for epigenetic regulation.[13] DNA replication requires transient chromatin disassembly of histones. According to one of the models of epigenetic inheritance, following DNA replication when chromatin is reassembled, parental histones, with their modifications, are equally divided between the two replicating DNA.[14] Modified histones may contribute to recruitment of enzyme complexes after being transferred from parent to daughter DNA. This allows the same modifications to be added to new histones.[15] An example is the transmission of the H3K9 methylation mark. HP1 binding to H3K9me3 present on parental, inherited histones can promote recruitment of H3K9 methyltransferases, which may then methylate newly synthesized (unmethylated) histones, thus ensuring transmissibility.[16] The enzymes that deposit H3K9me3 and H3K27me3 do so in a self propagating manner. This means that they not only deposit the PTM, but also recognize it once deposited, thereby behaving like an allosteric switch, causing the mark to spread on the neighboring histone tail. This creates a positive feedback loop.[17]

Understanding mechanisms for chromatin assembly and disassembly could help reveal how epigenetic information is inherited through several cell generations.[18] Recent studies of redeposition of parental histones have suggested that it occurs at or near the same DNA sequence

occupied prior to DNA replication.[18] This strengthens the model that histones and their PTMs may function as epigenetic information.

PcG proteins are key players of epigenetic inheritance. They ensure maintenance of cell identity by repressing developmental genes notably through the repressive mark H3K27me3 and chromatin compaction. PcG proteins have also been shown to be inherited through DNA replication.[19]Studies on PcG proteins during replication, have indicated possible mechanisms of inheritance involving both duplication and segregation of chromatin features.[19] Two general mechanisms were derived. In the first case, PcG dependent chromatin structures and binding could be disrupted by the DNA replication process. They reform completely on the two daughters after DNA replication guided by stable histone marks present on them. The second case is based on segregation of PcG proteins as well as histones to the two daughters which would then be used as a template for complete restoration on both strands.

1.2 History and overview of PcG complexes

Polycomb proteins were initially identified in *Drosophila Melanogaster*.[20] One of the early events in *Drosophila* development is the demarcation of segments along the anterior-posterior axis.[21,22] Each segment is destined to give rise to specific structures in the head, thoracic and abdominal regions of the body (such as legs, wings and eyes) based on its position along the anterior-posterior axis of the animal.[22,23] The genes controlling these identities belong to a class of genes termed Homeotic (*Hox*) genes, which reside in two clusters in the genome, namely the Bithorax Complex (*BX-C*) and the Antennapedia Complex (*ANT-C*).[22,23]

Adult male flies have specialized bristles on the first pair of thoracic legs, termed “sex combs”. Partial development of sex combs on the second and third legs of the adult male flies was identified and named “extra sex comb” (*esc*) in 1942.[20,24] A similar phenotype was observed in organisms

having a mutation in the *Polycomb* (*Pc*) gene.[21,22] This was attributed to a reduction in the repression of homeotic genes due to lower levels of *Pc*. [20] Genes that give rise to the extra sex combs phenotype when mutated were collectively termed the Polycomb Group (PcG). Genetic and later biochemical analysis revealed that the PcG maintains repression of homeotic genes.[20]

PcG proteins assemble into complexes, including PRC1 (Polycomb Repressive Complex 1) and PRC2 (Polycomb Repressive Complex 2). These complexes both have enzymatic activities that can modify histones, through monoubiquitination of H2AK119, and mono-, di-, and trimethylation of histone H3K27 respectively.[25] The homologues of PcG proteins have been found in species from plants to humans suggesting that they have been conserved through evolution.[26] They have broader biological implications in humans and the first mammalian PcG gene Bmi-1 was identified as an oncogene.[27]

1.3 Characterization of Polycomb Repressive Complexes (PRCs)

1.3.1 PRC1

PRC1 is primarily responsible for H2A mono-ubiquitination and chromatin compaction. Polycomb proteins are functionally diverse owing to their structural diversity. Two classes of PRC1 have been described, canonical and noncanonical. *Drosophila* canonical PRC1 (cPRC1) is comprised of four core proteins namely Posterior sex combs (Psc), Sex combs extra (Sce or dRING), Polyhomeotic (Ph) and Polycomb (Pc).[28,29] Polycomb (Pc) has a chromodomain which allows it to bind to H3K27me3. Several proteins associated with chromatin remodeling contain a chromodomain. *In vitro,* Psc binds to chromatin and mediates chromatin compaction, inhibition of transcription and chromatin remodeling through its C-terminal disordered region; this region is essential for *Psc* function *in vivo*.[30] Psc is incorporated into PRC1 via the conserved RING and RAWUL domains in its N-terminus.[31] The third PRC1 subunit, dRING/Sce, is responsible for

mono-ubiquitination of H2A through an E3 ubiquitin ligase activity.[24,32] Although the RING domain of Sce is the actual ligase, its activity depends on the RING domain of Psc.[31] Lastly, Ph contains a conserved domain, the SAM (Sterile Alpha Motif), which self associates to form helical polymers. It also undergoes hetero polymerization with another SAM containing PcG protein, Scm. Ph SAM activity is implicated in clustering of PcG proteins, and in the gene repression activity of PRC1.[33–36]

The expansion of PcG genes in mammals is accompanied by diversification in PRC1 complexes.[36] Several biochemical functions of core PcG complexes are conserved from flies to mammals.[37] In humans, the homologs of Pc are the Chromobox proteins (CBX2, 4, 6, 7, 8). RING1A and RING1B are homologous to dRING, and mediate E3 ligase activity when paired with any of the six Polycomb group RING fingers (PCGF) proteins, which are homologues of Psc.[37,38] Finally, mammals have three homologues of Ph, PHC1-3.[37] cPRC1, defined as containing a CBX protein, a PCGF, RING1A or RING1B, and a PHC, has a repressive function in mammals. [38]

Mammalian PRC1s are classified according to the PCGF subunit, so that cPRC1 is either c.PRC1.2 or c.PRC1.4, containing PCGF2 or PCGF4 respectively.[38]The non canonical (nc) or variant PRC1 complexes occur as PRC1.1, 1.3, 1.5, and 1.6 and contain PCGFs 1, 3, 5, or 6.The mammalian ncPRC1 is also composed of RING1A/B and RYBP or its homologue YAF2.[38] In *Drosophila*, a non-canonical complex, dRAF (dRING associated factors) has been described, which contains a histone demethylase (KDM2), as well as dRING and PSC. [39,40]

1.3.2 PRC2

PRC2 is involved in tri-methylation of H3K27.[24] It is a more well studied complex than PRC1. The core *Drosophila* PRC2 components are Extra sex combs (Esc), Suppressor of Zeste 12 (Su(z)12), Enhancer of Zeste (E(z)) and Caf1-55.[22,24]

E(Z) contains a SET domain which is responsible for tri-methylation of H3K27, a well-known function of PRC2. Su(z)12 also contributes to the methyltransferase activity of E(Z) through its conserved VEFS domain.[41] Both Esc and Caf1 have a unique WD-repeat protein structure. WD-repeat proteins provide sites for interactions with other proteins and have a multitude of cellular functions namely RNA processing, nuclear export and protein trafficking.[42] Apart from these, they are involved in chromatin modification and transcriptional processes as well. Other PRC2 components including Esc, have been shown to bind H3K27me3, the histone core as well as E(z). However, the function of Caf1-55 in PRC2 is not well understood yet.[24]

The core PRC2 components of the mammalian system are: Ezh1/2, Suz12, Eed and RbAp46/48.[24] Three additional polypeptides - AEBP2, PCL1-3, and Jarid2 can associate with PRC2 to form distinct PRC2 variants.[43] These components interact with other PRC2 components and have been shown to influence PRC2 enzymatic activity.

1.3.3 PhoRC

Pho, and its paralogue Pho-like (PhoL), were found to assemble into a third PcG complex in *Drosophila*, the Pho repressive complex (PhoRC). PhoRC is composed of Pho or PhoL and Scm-related gene containing four mbt domains (Sfmbt).[44,45] The MBT domains of Sfmbt can bind methylated residues on histones H3 and H4.[45,46] Sfmbt also has a SAM domain, which can interact with the SAM of SCM.[44,46] The SAM of SCM in turn can interact with that of Ph. Outside of *Drosophila*, the PhoRC is not clearly defined.[44]

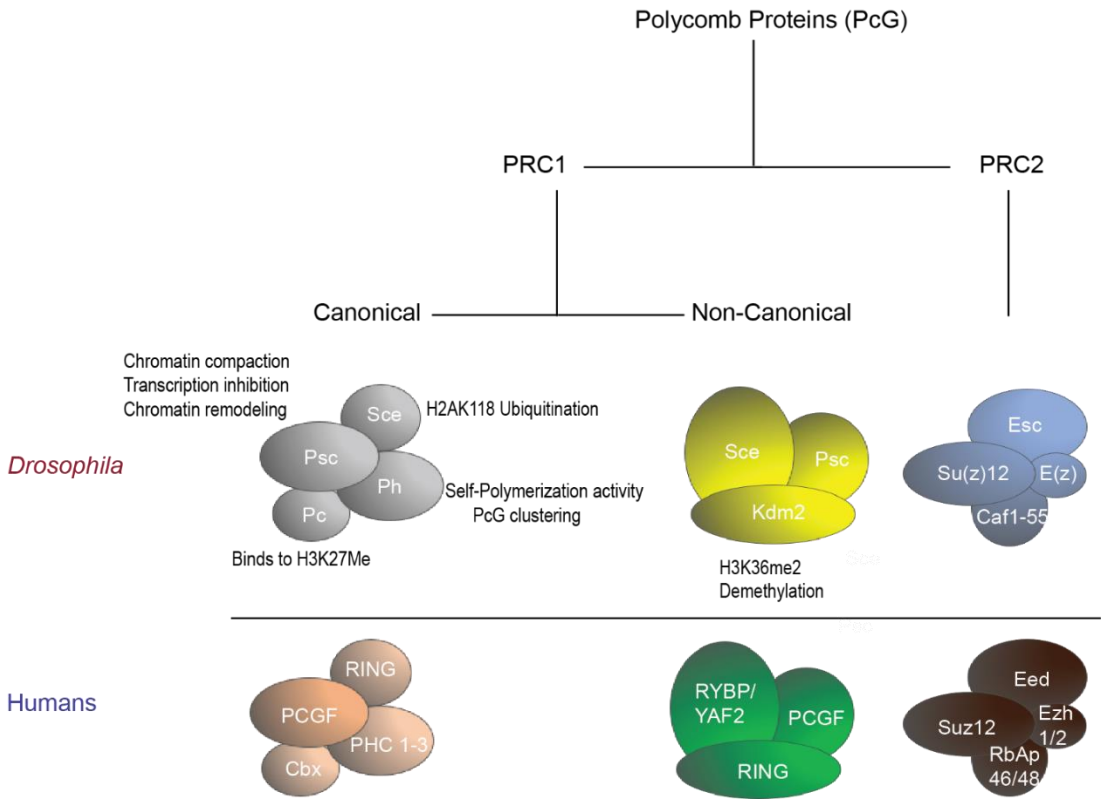


Figure 1. Polycomb complexes in humans and *Drosophila.*

1.4 Function of PcG proteins

Mechanistically, Polycomb proteins are chromatin modifiers that maintain repressed states during development and differentiation through epigenetic (heritable) mechanisms.[26] Chromatin regulation is an important aspect of transcription control. Several factors like gene-specific transcription factors, the transcription state and the chromatin landscape contribute to how PcG proteins are recruited to their gene targets.[47] It is common for Polycomb proteins and H3K27me3 mark to be associated with a repressed transcriptional state. However, studies on Drosophila imaginal disk cells demonstrate binding on genes that have been transcribed. The mechanism by which a repressed state is recognized by PcG is yet to be understood.[48]

Transcription factors are known to function in coordination with chromatin modifiers and remodelers thereby linking transcription with chromatin states.[26] Chromatin modifiers are also frequently mutated in pathologies like cancer.[49]

Histone modifications by PcG and modifications at the scale of nucleosomes remain central to regulation of gene expression by the PcG.[50] Post translational modifications on histones can affect its interactions with DNA and serve as a binding site for regulatory proteins.[51] Certain histone modifications may make DNA less accessible by promoting chromatin compaction, while other modifications, notably histone acetylation, may make chromatin more accessible. Histone modifications are also believed to spread on chromatin, in some cases creating large, modified domains.[52] Spreading of histone modifications on chromatin is believed to involve positive feedback between PTM-binding proteins, also called readers and enzymes that deposit PTMs (writers). An example is H3K9me3 in heterochromatin. In this system, HP1 is a "reader" of H3K9me3, via its chromodomain. HP1 can also recruit SUV39h1 (a writer for H3K9me3), which also has a chromodomain.[53] In the PcG system, PRC2 is both a reader and a writer of H3K27me3; reader activity is implicated in spreading H3K27me3. PRC1 is also a reader of H3K27me3, an activity imparted by the chromodomain of Pc; this activity contributes to targeting of PRC1. [43]

Polycomb Proteins act antagonistically with another group of proteins called the Trithorax Group (TrxG).[54] TrXG proteins promote transcription activation by catalyzing trimethylation of histone H3 on lysine K4 (H3K4me3) and lysine K36 (H3K36me3). These "active" modifications antagonize H3K27 methylation by PRC2, and vice versa, suggesting histone PTMs are an important aspect of the functional antagonism between the PcG and TrxG.[51]

In *Drosophila* imaginal disc cells, a K-->R mutation of H3K27 produces a similar phenotype to mutation of PRC2. This is evidence for importance of H3K27 for Polycomb repression in

Drosophila.[42] In contrast, mutations in H2A that prevent ubiquitylation do not fully recapitulate PcG phenotypes. This indicates that ubiquitylation of H2A is not the sole mechanism by which PRC1 represses its target genes.[55] This was confirmed by the mild phenotype of embryos lacking *sce*, relative to that of embryos lacking other PRC1 subunits. Therefore, PRC1 components must also repress transcription through non-enzymatic chromatin-based mechanisms. [55] There is corroborating evidence for the same in mammalian systems from studies on the Cbx2 subunit which is responsible for chromatin compaction, showing that mutating basic residues in its sequence causes upregulation of gene expression.[56]

1.5 Mechanisms of PcG protein recruitment

PREs or Polycomb Response elements are cis-regulatory sequences that can recruit PcG proteins to repress gene expression.[57] PREs can also maintain memory of transcription repression. They are typically ~1kb in length. They contain binding motifs for DNA binding factors that can physically interact with PcG proteins.[58] However, precisely how PRC complexes are targeted to genes (through PREs or other mechanisms) is still not understood.[59]

One mechanism of PcG protein recruitment involves the PcG protein Pleiohomeotic (Pho). Pho is unique in the PcG as it is the only PcG protein that is known to have sequence specific DNA binding activity.[44] PREs typically contain at least one Pho motif, so that Pho recognition of its DNA motif is thought to be an important component of PcG recruitment. One of the earliest models for Polycomb recruitment suggested the binding of Pho to PREs can recruit PRC2.[45] Upon trimethylation of H3K27 by PRC2, PRC1 can be recruited by the interaction between the chromodomain of Pc and H3K27me3.[60] Subsequent genetic and genomic analysis of PRC1 complexes in *Drosophila* and mammals, however, suggest this model does not account for most PRC recruitment.

Another model involves the Pho repressive complex where a tri-SAM interaction between the SAM domain of Sfmbt, SCM and Ph could explain how Pho-RC can recruit PRC1 to DNA. There are now several other models that discuss the various possibilities for how PcG complexes are targeted to the genome, including one suggesting an interaction between PRC1 and PRC2 mediated by SCM.[46]

1.6 Polycomb bodies

Polycomb proteins are associated with a type of repressive Topological Associated Domains (TADs) termed Polycomb domain that contains a high density of H3K27me3 and many binding sites for PcG proteins.[61] For example, both the ANT-C and BX-C *Hox* gene clusters form large PcG domains. Polycomb response elements may act as nucleation sites for Polycomb domains and looping interactions that occur within these domains could drive clustering of PcG target genes forming into macroscopic structures termed Polycomb bodies (**Figure 2**). [62]

Early immunofluorescence studies of transformed mammalian cell lines demonstrated that PcG proteins form highly concentrated foci, which were termed "Polycomb bodies".[63] Subsequent studies using live imaging and immunofluorescence in *Drosophila* cells, intact embryos and larvae, as well as in primary mouse cells, confirm that Polycomb bodies are widespread, if not universal.[63] The number, size, and composition of PcG bodies reported in different cells is different, which is due in part to the use of different methods and components to identify them, but also likely due to cell type and cell cycle phase specific differences.[64]

One of the earliest studies using confocal microscopy to study interphase nuclei in *Drosophila* embryos, revealed PcG proteins forming ~100 foci with Pc and Ph colocalization being over 90%.[65] More recently, live imaging was used to demonstrate accumulation of Pc-GFP and Ph-GFP into PcG bodies, which has been shown to occur in *Drosophila* embryos.[33] These bodies co-

localize with the H3K27me3 mark and *Hox* gene clusters which are known PcG targets that are covered with this histone mark.[33] The functional significance of Pc foci co-localizing with Hox genes was linked to higher order chromatin folding of Hox clusters prior to transcription.[33] Furthermore, in *Drosophila* S2 cells, chromosome conformation capture (3C) studies demonstrated that contacts between PcG bound DNA elements underlies folding of the repressed BX-C Hox gene cluster. Fluorescent In Situ Hybridization (FISH) combined with immunofluorescence revealed co-localization of Polycomb bodies with BX-C.[66] Recent studies using super-resolution microscopy demonstrate that, in addition to a small number of PcG bodies visible by light microscopy, Pc and Ph form hundreds of nanoscale clusters in S2 cells.[34]

The observation of PcG domains and PcG bodies have prompted models for how PcG proteins can create large repressive domains. The 'spreading' or 'structural' model suggests PcG complexes coat the entire repressed region.[34,35] The 'hop-and-skip' or 'organizer' model instead suggests PcG complexes established at PREs form loops with nearby weaker sites, and this iterative looping affects distant regions.[62,67] Protein polymerization could facilitate the spreading of PcG complexes in either model, by polymerizing along the chromosome or linking proteins bound at different sites. Although either the spreading or the organizer models could be consistent with PcG domains and PcG bodies, the findings that most PcG proteins localize to discrete sites (i.e., sharp peaks in ChIP-seq experiments), and that non-repressed genes can exist between two PcG repressed genes, favour some version of the organizer model, although a contribution of local spreading cannot be ruled out.[68]

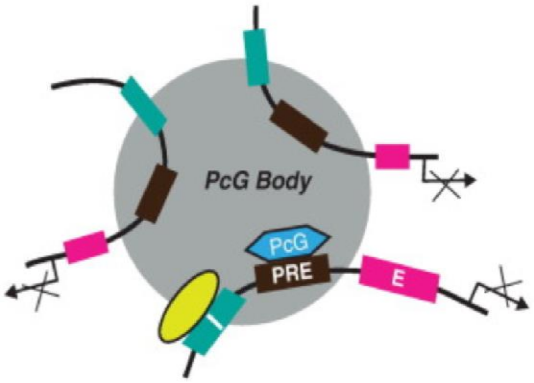


Figure 2. Polycomb body. Depiction of clustering of PcG target genes into Polycomb bodies. Figure is from: Pirrotta, V., & Li, H.-B. (2012). A view of nuclear Polycomb bodies. Current Opinion in Genetics & Development, 22(2), 101. https://doi.org/10.1016/J.GDE.2011.11.004.

1.7 Polyhomeotic

Drosophila has two Polyhomeotic genes, *ph-p* (*polyhomeotic proximal*) and *ph-d* (*polyhomeotic distal*) which arose due to tandem duplication and have largely redundant functions.[24,69] Ph-p is 1589 amino acids long with the N-terminal region being largely disordered.[70] At its C-terminus, Ph and its mammalian homologues (PHCs) have a Sterile Alpha Motif (SAM).[71] The SAM domain has ~ 70 residues and can self-associate or undergo heterotypic interactions with other proteins. The Ph SAM can form head-to-tail polymers, and this polymerization activity is implicated in PcG function.[72]The polymeric structure of SAM is a left-handed helical spiral structure.[67]

Head to tail polymerization is mediated by two binding interfaces namely Mid-Loop (ML) and End-Helix (EH) on the SAM. While the Mid-Loop houses SAM-SAM interacting residues in the middle the End-Helix contains interacting residues on the C-terminal end of the SAM domain.[72] Ph has three human orthologs human Polyhomeotic homolog 1, 2, 3 (PHC1, PHC2, PHC3).[73] PHC1 has been shown to regulate stem cell pluripotency, and a mutation in the SAM of PHC1

causes microcephaly, implicating it in growth control.[74] PHC3 functions as a tumor suppressor in osteosarcomas.[72,75] The SAM of PHC3, like that of Ph was shown to undergo polymerization. It has been shown that the linker region next to SAM which is intrinsically disordered controls SAM polymerization. In *Drosophila* Ph, the linker region restricts SAM polymers, while the linker region of PHC3 facilitates formation of long polymers.[72]

1.7.1 Ph in PcG clustering

Recent observations using stochastic optical reconstruction microscopy (STORM) revealed that PRC1 components Polycomb (Pc) and Polyhomeotic (Ph) form nanoscale clusters. The largest clusters were presumed to be PcG bodies that were observed as foci using conventional microscopy.[34] These clusters are heterogenous in size, and their formation is disrupted by expression of Ph with the ML interface mutated, which functions as a dominant negative to block polymerization.[76] In mammalian cells, PcG clusters visualized with GFP-Ring1B are disrupted by expression of the equivalent dominant negative mutation in Phc2.[34,35]

The effect of PcG clustering on genome organization was tested in S2 cells using a derivative of the chromosome conformation capture methodology (4C) where it was observed that disruption of Ph SAM polymerization activity leads to decreases in chromatin contacts in the BX-C *Hox* gene cluster, as well as in long-range chromosomal interactions.[35] In mammalian cells, disrupting PHC SAM polymerization leads to loss of Hox gene compaction (demonstrated by FISH), and derepression of Hox gene expression. The function of Ph in chromatin compaction has also been addressed by Ph knockdown studies which although not directly implicating SAM, showed decompaction of repressed domains.[4] While SAM polymerization is an essential Ph function, it is not necessary for all Ph functions. Genetic rescue assays demonstrate that the Ph SAM domain is

essential for all Ph functions, while polymerization is required for repressing some PcG target genes.[77]

1.8 Condensate regulation during cell cycle

There has been a recent surge in studies relating to liquid-liquid phase separation (LLPS) as a mechanism for organization of macromolecules and it also underpins chromatin organization. Heterochromatin Protein 1(HP1) was shown to form phase separated condensates and compact chromatin, suggesting a role for phase separation in heterochromatin formation.[78] Cellular processes like transcription are improved due to phase separation. For example, the Mediator complex subunit 1 (MED1) drives transcription by forming phase separated condensates with chromatin and RNA pol II.[79]

Nuclear condensates have been shown to be affected by the local genomic environment which can cause them to largely form in low density euchromatin regions.[80] Likewise, cell-cycle related changes in chromatin can also affect the behavior of condensates. Upon entry into mitosis, several membraneless organelles like Cajal bodies, splicing speckles and PML bodies dissolve and reform upon completion of mitosis; dissolution may be important to ensure even distribution of condensate components among daughter cells.[81,82]

How does PcG repression survive mitosis? The condensation of chromatin during mitosis affects binding of chromatin proteins.[83] This also causes transcriptional silencing as the binding of transcription factors is drastically reduced. This is due in part to the breakdown of nuclear envelope, which releases nuclear proteins into the total cell volume thereby diluting them.[84] In some cases, certain transcription factors and chromatin proteins persist on the chromosome.[84] This phenomenon, called mitotic bookmarking, is believed to be one of the mechanisms by which regulatory information is passed from mother to daughter cells.[83] PcG and TrxG proteins act in a

dynamic manner on their target genes during mitosis. This ensures a capacity to switch between transcription states beyond just "on" and "off".[85] A small fraction of Ph was shown to persist on chromatin during mitosis as observed from immunofluorescence and live imaging, and ChIP-seq of mitotic cells.[121,86]

PML (Promyelocytic leukemia) bodies which play a role in several cellular processes like transcriptional regulation, tumor suppression and apoptosis were found to form nuclear bodies that varied in size and number during DNA replication.[87] Microrchidia 3 (MORC3) localizes to PML bodies by sumoylation activity of its C-terminal domain. The condensates formed by MORC3 exhibited cell cycle regulated behavior as well, with most of them disappearing in mitosis and reforming after cell division.[88]

1.9 LLPS (Liquid Liquid Phase Separation)

LLPS is a process whereby above a certain concentration, macromolecules like proteins de-mix (separate) into a dense phase which is concentrated with the proteins and a surrounding dilute phase.[89,90] Studies governing multi-component mixtures show that the components either undergo condensation to form a single dense phase with composition similar to dilute phase or de-mix to form multiple phases with distinct compositions. Many proteins that form foci *in vivo*, undergo phase separation *in vitro* and phase separation may underlie foci formation in some cases.[89]

Pioneering studies of P-granule formation in Caenorhadbitis elegans germline establishment have greatly improved physical understanding of compartments formed by LLPS by showing that these structures behave like liquids.[91] Several membraneless organelles (MLOs) like Cajal bodies, P bodies, Stress Granules and Nucleoli have been shown to form via LLPS. It provides a mechanism to concentrate molecules like RNA thereby enhancing biochemical reactions. Compartments that form through condensation are also referred to as biomolecular condensates.[89]

This threshold concentration that must be surpassed to allow phase separation to occur, is referred to as saturation concentration (c_{sat}).[92] In a simple phase separated system, the protein concentration of the dilute and dense phases remains constant; increasing the total protein concentration leads to an increase in the volume fraction of the dense phase. Because of this property, phase separation, in principle, can buffer a system (like a cell) against fluctuations in total protein concentration because the excess protein is concentrated in the (expanded) dense phase.[93,94] However, for multicomponent systems where heterotypic interactions drive phase separation, more complex effects of changes in total protein concentration are observed due to saturation of interactions. For example, through a combination of *in vitro* and *in vivo* studies it was shown that the preferential heterotypic interactions of NPM1 with SURF6 versus itself (i.e. homotypic interactions), lead to excess NPM1 accumulating in the dilute phase rather than disrupting SURF6-NPM1 in the dense phase.[94]

1.9.1 Features of proteins undergoing phase separation

Multivalent interactions, which drive molecular interactions without fixed stoichiometry, are central to phase separation. Multivalency can arise from the presence of repeated interaction domains (a common feature in signalling proteins), or multiple different types of interaction domains.[95–97] Some structural domains can be self-interacting via distinct surfaces present on them namely "head" and "tail" allowing polymerization to occur.[98] The presence of oligomerization domains can enhance multivalency as well thereby promoting LLPS.[90]

In vitro experiments with TDP-43 have shown that polymerization mediated by its N-terminal domain enhanced its LLPS activity.[99] Intrinsic strength of these intermolecular interactions is usually weak which facilitates liquid like dynamics. Alternatively, intrinsically disordered regions, which typically can mediate many types of weak interactions, can also impart multivalency.[90]

Many proteins that undergo phase separation use a combination of IDRs and structured domains (**Figure 3**).

1.9.1.1 Intrinsically disordered regions (IDRs)

Proteins that lack a stable secondary or tertiary conformation are referred to as Intrinsically Disordered Proteins (IDP's).[100] Although in isolation they bear a largely undefined structure, upon interaction with other biomolecules, many have the capability to transiently form secondary or tertiary structures. Most proteins undergoing phase separation *in vivo* possess IDRs. They often contain simple repetitive residues such as tandem amino acid repeats also called Low Complexity Regions (LCRs).[101] While LCRs have been reported in proteins that have precise structures or ordered domains, they are also abundant in IDRs.[101,102] IDRs are involved in several cellular functions like transcription, mRNA processing and cell cycle regulation.[103,104]

IDRs are usually deficient in hydrophobic residues which form the core of folded proteins, and abundant in polar and charged amino acids which can mediate several intra-and intermolecular interactions.[103] Non-polar residues participate in hydrophobic interactions.[105] Although self assembly is often driven by aromatic-mediated or electrostatic interactions, purely hydrophobic interactions can also mediate phase separation. An example for this is *S. cerevisiae* PolyA Binding Protein (Pab1), which bears non-polar proline rich domain or P domain in its intrinsically disordered region and undergoes phase separation.[106] It has been speculated that polar residue enrichment could play a role in chain dynamics, providing a soluble scaffold for intermolecular interactions to take place.[107]

Certain PTMs like phosphorylation, which can cause the conversion of a polar residue to a charged residue, can in fact be drivers of intermolecular interactions. The addition of charges through phosphorylation can lead to phase separation through complex coacervation.[108]

Complex coacervation is also defined as an associative phase separation between multivalent oppositely charged molecules. An example where this is evident is between IDRs and nucleic acids. A serine-arginine rich disordered regions was shown to undergo phase separation with RNA.[108] The co-activator MED1 which shows an enrichment in serine residues failed to form droplets upon mutation of serine to alanine suggesting that droplet formation is dependent on the serine bias.[104] Amino acid stretches that are homopolymeric containing other polar residues like glutamine can also mediate intermolecular interactions.[108]

Many Asparagine and Glutamine rich LCRs fall into a specific class defined as prion-like. Prions are misfolded forms of a normal protein with the ability to self propagate. This unique property of prions makes them dominant both epigenetically and conformationally.[89] Prion-like LCRs have also been implicated in LLPS. For example, prion domains in certain RNA binding proteins are essential for their phase separation *in vivo.*[109] A well-known prion like protein that has also been investigated for its phase separation activity is the FUS protein.[109] Whether prion-like domains involved in LLPS can also mediate classical prion function (self-propagation) is not yet clear.

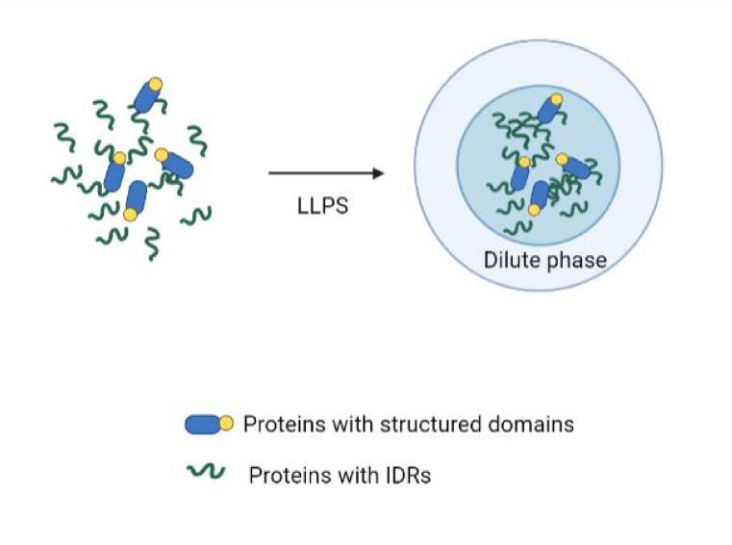


Figure 3. Liquid Liquid Phase Separation in the cell. Multivalent interactions among structural domain proteins and proteins with Intrinsically disordered regions leading to liquid liquid phase separation.

1.9.2 Biological implications of LLPS

LLPS has been suggested to contribute to the fundamental organization of euchromatin and heterochromatin. HP1 for example, contributes to phase separation of constitutive heterochromatin. As is characteristic for proteins that undergo LLPS, HP1 has IDRs which drive its phase separation activity.[110] Although there are three human homologs of HP1 and only one of them, HP1α, (and *Drosophila* HP1a) but not HP1β and HP1γ, undergoes phase separation *in vitro*; phase separation is controlled by phosphorylation of HP1.[110]

This work and other studies investigating liquid-like material properties of HP1 domains *in vivo*, suggest a role for LLPS in heterochromatin formation.[110,111] Some studies, however, have strongly refuted the idea that LLPS is important *in vivo*, pointing out that visible condensates can arise through other biophysical mechanisms.[112] Thus, the role of LLPS in heterochromatin organization is an area of active investigation.

Condensate formation is also a means for regulation of biochemical reactions by creating high concentration of reactants in the condensate. For example, sumoylation rates increased significantly in condensates compared to the bulk by recruitment of the sumoylation enzyme cascade into engineered condensates formed by liquid-liquid phase separation.[113,114]

Apart from various biochemical functions, phase separated condensates also exhibit unique material properties allowing them to organize structures at a scale larger than individual macromolecules. These functions are referred to as mesoscale functions, include forcefully pushing chromatin away thereby creating low density regions in the chromatin. Another example where condensates can impact cellular architecture is membrane associated condensates mechanically driving membrane invagination.[114]

1.10 Is LLPS involved in PcG body formation?

The finding that PcG proteins form condensates (PcG bodies) in cells, and that these proteins contain many IDRs and protein interaction motifs, suggests that they could form through LLPS. The mammalian Cbx2 protein was shown to undergo LLPS *in vitro*, and Cbx2 forms condensates that colocalize with PRC1 in cells.[115] A positively charged low complexity disordered region (LCDR) in Cbx2 which is required for LLPS *in vitro* is also important for condensate formation in cells.[115,116] Through phase separation activity, Cbx2 assembles CBX-PRC1 condensates leading to gene repression at specific sites on the genome.[117]

In C. elegans, a Polycomb protein Sop-2 acts as a functional counterpart of the PRC1 complex and was observed to form nuclear bodies.[118] Sop-2 is similar to Ph in that it has a SAM, but it lacks other Ph features and is not a clear homologue. Sop-2 undergoes phase separation *in vitro*; this is mediated by an IDR in the protein, rather than by the SAM.[118] The SAM modulates the material

properties of Sop-2 condensates. Sop-2 condensates in animals and *in vitro* are also regulated by sumoylation.[116,118]

The SAM of Ph is implicated in formation of Ph condensates (PcG bodies) *in vivo*.[34,119]To test whether the SAM can drive LLPS, Seif et al. studied a truncated version of Ph, "mini-Ph".[70] Mini-Ph contains the three conserved domains of Ph, the HD1, FCS Zinc finger and the SAM domain, but lacks the large N-terminal IDRs.[70] Mini-Ph forms phase separated condensates with chromatin *in vitro*. Experiments with mutations on the binding interfaces (EH and ML) showed that SAM polymerization was not essential for the condensates to form. However, the presence of SAM was required.[70] Thus, two PRC1 components previously implicated in PcG body formation have domains that can undergo LLPS with chromatin *in vitro*, raising the possibility that LLPS is relevant for organization of PcG proteins and chromatin in cells.

Chapter 2: Rationale and Hypothesis

Mini-Ph forms phase separated condensates with DNA or chromatin *in vitro*. However, in cells, full length Ph forms multiple condensates, while mini-Ph either does not form condensates or forms a single large nuclear condensate. This indicates that mini-Ph lacks sequences required for wild-type (WT)--like condensate formation. The N-terminal region of Ph is predicted to be disordered and to contain LCRs.

Thus, we hypothesize that the disordered sequences outside the conserved region of Ph (mini-Ph) control Ph SAM activity, by affecting phase separation, or through other mechanisms.

PcG regulation is believed to be heritable through cell cycles (epigenetic). Key events in the cell cycle, DNA replication and mitosis, challenge chromatin-based epigenetic information. Ph condensates could be important during DNA replication to maintain high concentrations of PcG proteins near chromatin. If condensates are important for heritable regulation through mitosis, they must either persist on chromatin or be re-established at the end of mitosis. Whether condensates contribute to heritable regulation is not known. We therefore looked at Ph condensates during the cell cycle to understand their potential role in regulation through the cell cycle.

Aim 1: Analyze the IDRs of Ph and how they affect the behavior of Ph SAM in cells

OBJECTIVES:

- Analysis of sequence composition and complexity, and identification of different segments of the disordered region. (IDRs)

- Determination of activity of each IDR and its contribution to regulation of Ph distribution in cells, including dependence on Ph SAM, by live imaging of *Drosophila* S2 cells expressing Venus-tagged Ph proteins with different sequences deleted.

Aim 2: Characterize the cell cycle regulation of Ph condensates, including by the Ph IDRs

- Analysis of Ph condensates during S-phase and mitosis

Chapter 3: Materials and Methods

3.1 Cloning

To express Ph5, Ph6, and Ph7, the sequences were first cloned into a gateway donor vector. The plasmid pCR8, containing the full-length Ph cDNA was prepared previously by lab member Elodie Boulier and was used as a template for PCR. PCR reactions consisted of 1X iProof HF buffer, 2mM of each primer, 10 mM dNTP, 5 ng/ul pCR8-Ph template, 3% DMSO, and 1ul of iProof DNA polymerase (kit from Bio-Rad 1725301). The volume was completed to 100 µl by addition of ddH20. Each reaction was split in 3 to test different annealing temperatures (50°, 55° and 60° C) for Ph6 and Ph7. For Ph 5, higher annealing temperatures of 60°, 65° and 70° C were used. The cycling conditions for iProof were based on the manufacturer's instructions as follows: 35 cycles of 98° C for 5" for denaturing; 15" of annealing at temperatures as mentioned above; extension time was adjusted according to the length of the product, namely 15-30"/kb at 72° C. Final extension was carried out for 5-10 min at 72° C. The amplified DNA fragments were cloned into the donor vector pCR8 containing an SV40-NLS (nuclear localization signal) through restriction digest and ligation reaction. Restriction digests were done with 5 µg of vector DNA, while for the insert the entire volume from PCR was used. The enzymes used for digestion were Acc651 and SpeI and the appropriate buffer was selected depending on the recommendations from NEB. The total volume for the ligation reaction was 10 µl and contained a 3:1 ratio of insert to vector. The ligated products were transformed into NEB5α (NEB) bacteria and selection was done by Spectinomycin resistance on the LB agar (Bioshop LBL406.1) plates.

For transformation, the competent bacterial cells were thawed on ice and 5 µl of ligation reaction was added to it and gently mixed. This was followed by a 30-minute incubation while the products were still on ice. After, the bacteria were heat shocked by holding in the water bath at

37 degrees for 45 seconds. The tubes were put back into ice for at least 2 minutes. 500 µl of SOC media (2% tryptone, 0.5% Yeast, NaCl 10 mM, KCl 6.7 mM, MgCl2 1M, MgSO4 1M, Glucose 2M) was added to each transformation followed by incubation in the shaking incubator at 37 degrees for 1 hour. The mixture was spread on LB agar (32 g LB powder, 1 ml glycerol in milli-Q water filled to 1 L) plates with Spectinomycin (Bioshop SPE201.5) with a glass spreader and incubated overnight in an incubator at 37 degrees. The process was carried out next to a Bunsen burner to maintain a sterile environment.

At least 5 colonies were picked the next day and inoculated in 2 ml cultures of sterile autoclaved LB broth (Bioshop LBL407.500) with spectinomycin (1:1000). The plasmid was extracted using the standard protocol for alkaline lysis. Good clones were confirmed by running a restriction digest on a 1% Agarose gel and Phenol-Chloroform extraction was performed ((Phenol/CHCl3/Isoamyl Alcohol (25:24:1, v/v)) followed by ethanol precipitation with 1:10 volume of sodium acetate (NaOAc) and 3 volumes 100% ethanol incubated at -20°C overnight, to get rid of RNA and other contaminants.

The purified DNA was sent for sequencing to confirm the inserts were correct.

To transfer Ph truncations into plasmid pHVW (DGRC stock # 1089) for heat-shock inducible expression in Drosophila cells, LR-recombination reaction was performed. 75 ng of donor and destination plasmid were mixed in 1.5 ml Eppendorf tube, to which 0.8 µl of Gateway LR Clonase II (Thermo Fisher) enzyme was added. TE buffer (10mM Tris-Cl, pH 8.0, 1mM EDTA) was used to complete the volume to 5 µl. The tubes were left overnight at room temperature. The next day Proteinase K (0.5 µl, 20mg/ml) was added, and reactions were digested for 10 min. at 37°C. Transformation into DH5α bacterial cells was performed as described above. Clones were

checked by restriction digest, and the final DNA plasmids were prepared for transfection by maxi-prep using a Qiagen kit.

3.2 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) added. Cells were passaged every 2 to 3 days, and cells at 80-85% confluency were used for transfections. Cells were grown in 10 cm dishes in an incubator at room temperature.

3.3 Transfection

The night before transfection, 1.5e6 Drosophila S2 cells were plated in 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 minutes at 37˚C to induce Ph expression and used the next day for live imaging on a spinning disc confocal microscope.

3.4 EdU labelling

The EdU labelling kit (Invitrogen #C10340) was used for EdU (5-Ethynyl-2′-deoxyuridine) labelling. Cells were transfected with Venus-Ph constructs two days before labelling, protein expression induced as described above, and cells were replated in a 12 well chamber slide. 2X EdU solution was prewarmed before using. Half of the media was removed and replaced with an

equal volume of 20 uM 2X EdU solution to obtain final concentration of 10uM. Cells were incubated in media complemented with EdU for two hours. Post incubation, the immunofluorescence fixing protocol was followed according to the manufacturer's instructions. Click IT reaction cocktail was freshly prepared, and cells were incubated with it for 30 min in the dark. The wells were washed with 1X PBS (8g/L NaCl, 0.2g/L potassium chloride (KCl), 1.15g/L sodium phosphate dibasic heptahydrate (Na_2HPO_4-$7H_2O$) dibasic, 0.22g/L potassium phosphate (KH_2PO_4) monobasic, ddH2O) before staining with DAPI.

3.5 Immunofluorescence

S2 cells were washed once with 1X PBS, and fixed with 4% Paraformaldehyde (Electron Microscopy Sciences 15710) in PBS for 10 min. They were washed twice with PBS again for 5 min. Permeabilization was carried out in PBS with 0.02% TWEEN® 20 (Bioshop TWN 508.1) and 0.1% triton X-100 (Sigma Aldrich #T8787-250ML) for 15 min. after which the cells were washed 2 times with PBS 0.02% tween (PBST). Blocking with PBST + BSA (5%) was done for 1 hour at room temperature. Primary antibodies Rb anti-Ph (prepared in lab, 75I), mouse anti-H3S10P (Millipore Sigma 05-806), were used (2 ug/ml) in PBST-BSA (1%) and incubated overnight at 4 degrees. Next day, the cells were wash three times for five minutes with PBST and blocked for 5 mins in PBST+2% BSA. Secondary antibodies, Alexa 488(Rb) (Invitrogen A11008), Alexa 647(m) (Invitrogen A32728), were added and incubated for 1-2 hours in the dark. The dilutions for both primary and secondary antibodies were 1:1000. The cells were washed 3x 5 min with PBST and 1x with PBS before mounting in DAPI-containing media.

3.6 Western Blot

S2 cells 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 according to

manufacturer's protocol. The day after, the media was changed and the evening of next day, cells were heat shocked for 8 mins 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 mins. 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 5mins. Samples were then run on 8% and 16% SDS-PAGE gels for Ph and RFP respectively for 80min at 120 Volts in running buffer (14.4g/L glycine, 5.2g/L tris base, 1g/L SDS). The gel was then transferred to a nitrocellulose membrane at 0.3 Ampere for 100mins in transfer buffer (14.4g/L glycine, 5.2g/L tris base, 1g/L SDS, 20% MeOH).

Membranes were blocked for 30 mins 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, made in 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). Membranes were washed 3 times for 10mins each in PBST, incubated for 2 hrs in secondary antibody diluted in 5% milk/PBST and washed 3 times again for 10 mins in PBST. Secondary antibodies were conjugated to Alexa Fluor 680 (anti-rabbit and anti-mouse, Invitrogen A21076 and A21057 respectively), 800 CW (anti-Rabbit, Li-Cor) and used at 1:25 000 in 5% milk/PBST. Blots were scanned on an Odyssey CLx imager.

3.7 Image collection

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 YFP channel, the laser power and exposure were set as follows 2.40% and 77 mS, respectively. For the RFP it was 11% and 500 mS, 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 EYFP (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 638, respectively.

3.8 Image Acquisition

For live imaging, confocal stacks of 1 µm slices were collected. Images were acquired as tiles (3x3). All images were taken for channels YFP and RFP.

3.9 Image Processing and Analysis

Live images were opened using ImageJ (Fiji) in a tiff format. A maximum intensity projection was made for each image. Images were then split into YFP and RFP channels, and were named "foci" and "nuclei", respectively. These were then uploaded to Cell Profiler (3.1.9) where they were processed further.

Cell Profiler possesses modules that can be used to build a pipeline.[127] In the case of our constructs, a few parameters were modified in the pipeline for different constructs. The general modules shared in all pipelines were identification of nuclei from RFP staining for live imaging experiments, or Dapi staining in the case of cells that were fixed, and foci identification from the YFP channel. Cells touching the boundary of images were not considered. The number of foci per nucleus was obtained from relating the parent (nuclei) objects to child (foci) objects. Several measurements like intensity, shape and size of foci were also obtained using available modules (**Figure 4**).

3.9.1 Identification of Nuclei

The module IdentifyPrimaryObjects was used for the identification of the nucleus (from either DAPI or H2A-RFP channels), which was named “nuclei”. A diameter range was specified and objects outside this range as well as touching the border of the image were eliminated. A global thresholding strategy was applied in all cases with either Minimum Cross Entropy or Otsu as the thresholding method.

3.9.2 Identification of EdU labelled cells

EdU labelled cells were identified based on staining with Alxea 647 fluorophore with the IdentifyPrimaryObjects as was done above for nuclei identification. The objects were named “edu”. Early vs late S phase cells were identified by calculating the percentage of labelling by EdU. Cells which had more than 30% of area of nucleus (determined by the DAPI staining) were considered Early S while those with less than 30% staining were categorised as late S. [120] Cells without any Edu label were Non-S or Interphase cells.

An additional module called SplitOrMergeObjects was used for Edu identified cells which merged all Edu per cell. The new name of the objects obtained from this module was “RelabelledEdu”. The MeasureObjectSizeShape was then added to the pipeline to measure this RelabelledEdU size which was used in the calculations for determining early or late S phase.

3.9.3 Identification of foci

The IdentifyPrimaryObjects module was used for foci identification with thresholding using either Otsu or Minimum Cross Entropy and named “foci”. The size of the smoothing filter and the distance between local maxima are parameters that were adjusted for each construct separately to segment clumped objects. In the case of constructs like PhD1 where extensive clumping was observed, EnhanceorSuppressFeatures was used to ignore highly clumped objects.

3.9.4 Foci count per Nucleus

The RelateObjectsModule was used to count the number of foci per nucleus. For determining the number of foci in S phase cells, foci were matched with Dapi using the RelateObjectsModule and these were segregated as S or non-S and early or late S phase based on the presence of RelabelledEdU and the area occupied by it relative to the total nuclear area, respectively.

3.9.5 Intensity and size measurements

MeasureObjectIntensity was used to obtain mean intensity values. For the mean YFP intensity, the nuclei objects corresponding to the nuclei images were selected.

MeasureObjectSizeShape provided the measurements such as size which is the total number of pixels within the object, and it also extracts shape features from measurements of Form Factor. The object selected in this case was foci.

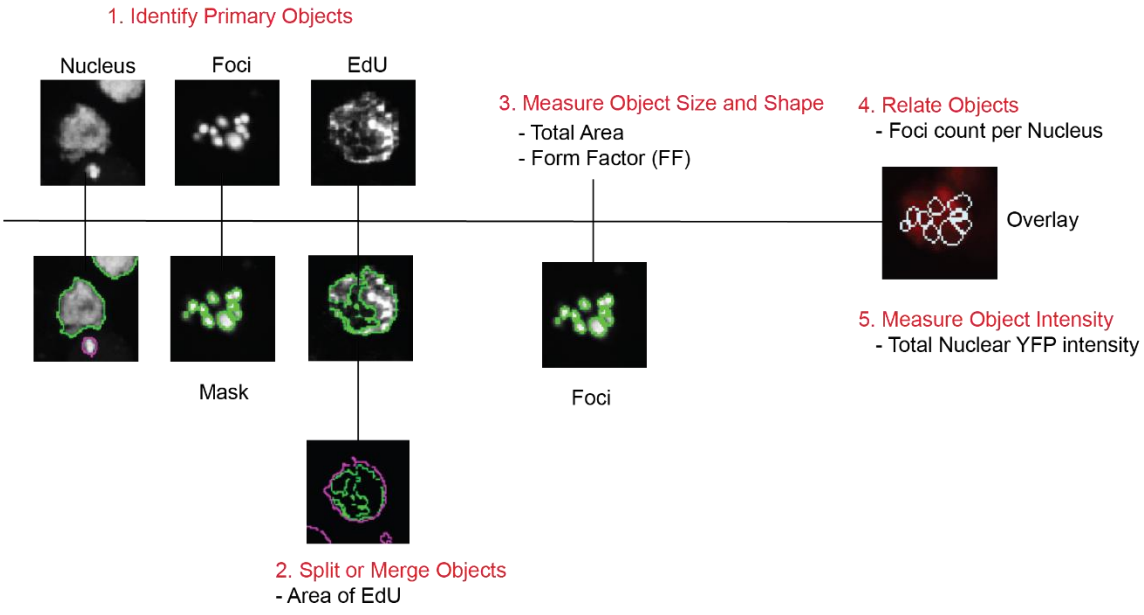


Figure 4. Pipeline for Cell Profiler. Representative images generated from cell profiler showing the image processing pipeline.

3.10 Quantification

For live image analysis, at least 100 cells were analyzed per experiment for each construct. In the case of EdU labelled cells, the minimum number of cells were 8 for S phase cells with foci. Foci count per nucleus was obtained for each construct and compared to wild-type Ph. Intensity threshold graphs were generated from two intensity measurements. One was for nuclei that did not have any foci and the other one for nuclei containing foci. Median foci size per nucleus was calculated for S phase cells expressing WT-Ph or Ph deletion constructs. All observations were made for 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.

Table 1. Materials used

MATERIAL	COMPANY	REFERENCE
Ampicillin	Bioshop	AMP201.100
Acc65I, SpeI	NEB	
NEB® 5-alpha Competent *E. coli*	NEB	C2987I
LB Agar Lennox	Bioshop	LBL406.1
LB Broth Miller	Bioshop	LBL407.500
Spectinomycin	Bioshop	SPE201.5
Terrific Broth Modified	Sigma Aldrich	T0918-1KG

5X iProof HF buffer, DMSO, iProof DNA polymerase, dNTP	Biorad	#1725301
Acetic Acid, Glacial (CH3COOH)	Fisher Scientific Company	351271212
Acrylamide/Bis-acrylamide, 40% Solution 37.5:1	Bioshop	ACR005
Agarose A	Bio Basic	D0012
Beta-Mercaptoethanol	Sigma-Aldrich	M6250
Bovine Serum Albumin	Bioshop	ALB001.100
Glycerol, Biotechnology Grade	BioShop	GLY001
Paraformaldehyde (16%)	Electron Microscopy Sciences	15710
Phenol Chloroform with Isoamyl Alcohol	Bishop	PHE 512.400
Precision Plus Protein™ Standard	Bio-Rad	1610394
2-Propanol (Certified ACS)	Fisher Chemical	A4164
Potassium Chloride (KCl)	BioShop	POC308
Potassium phosphate monobasic (KH2PO4)	American Bioanalytical	AB01660

Sodium phosphate dibasic heptahydrate (Na2HPO4-7H2O)	Sigma-Aldrich	S9390
Triton X-100	Sigma Aldrich	T8787-250ML
Tween 20	Bioshop	TWN 508.1
Tissue culture 6 well plate, surface treated sterile	VWR	734-2323
Tissue culture plate 24 well	Corning	353047
Glass Bottom dish	Ibidi	81218-200
12 well chamber removable glass slide	Ibidi	81201
Vectashield DAPI mounting media	Vector labs	H-1000-10

Table 2. Primers for PCR

Primer 1	Primer 2	Product	Size of insert (kb) (approx)
cgctcaACTAGTatggatcgtcgtg cattgaagtttatgc	agctgGGTACCctactgctgggaa gtctgttgagaagtcgc	Ph5	2853
cgctcaACTAGTatggatcgtcgtg cattgaagtttatgc	ctacgccttggggagatccttcgatccat t	Ph6	3903
ctcttctcaccgatgaacgtcatttcgc	ctacgccttggggagatccttcgatccat t	Ph7	3500

Table 3. Primers for sequencing

Plasmid	Primer	Tm(°C)
pCR8-Ph5	CAGCAGCAGGCGACTTCA	62.546
	ACCTGCATTCCCACAAACC	60.774
	AGTCGACTTTGCCAGTCGGT	62.197
	AGCAAATCCTGCAACAACAA	59.322
pCR8-Ph6	AGCAGCAGCTGTCGGAAG	53
	ACCTGCATTCCCACAAACC	60.45
	AGTCGACTTTGCCAGTCGGT	60.774
	AGCAGCTCTGCAGCGACC	62.197
	AGCAAATCCTGCAACAACAA	59.322
	CAACAGCAACAACAGCAGC	59.18
	TCAGATAGGTCTTCCACGCC	60.218
pCR8-Ph7	AGTCGACTTTGCCAGTCGGT	53

	AGCAGCTCTGCAGCGACC	62.197
	AGCAAATCCTGCAACAACAA	63.084
	CAACAGCAACAACAGCAGC	59.322
	CGGAGGGACTCAACAACAG	59.18
	CCATGTAATACGACTCACTATAGG	59.241
	TTTATTTGATGCCTGGCAGTTCC (REV-COMP)	58

Table 4. Plasmids used

Construct	Purpose	Company Name	Notes
pCR8	Gateway donor plasmid with Ph deletions or truncations		
pCR8-Ph1	Ph1 cloned in pCR8		
pCR8-Ph2	Ph2 cloned in pCR8		
pCR8-Ph3	Ph3 cloned in pCR8		
pCR8-PhD1	Ph with Ph1 deletion cloned in pCR8		
pCR8-PhD2	Ph with Ph2 deletion cloned in pCR8		
pCR8-PhD3	Ph with Ph 3 deletion cloned in pCR8		
pCR8-PhD1D3	Ph with Ph1 and 3 deletions cloned in pCR8		

pCR8-PhD2D3 All DSAMS except Ph8 pCR8 ATG NLS	Ph with Ph2 and 3 deletions cloned in pCR8 Donor plasmid with Nuclear Localization Signal added		PhD1D3DSAM Gateway cloning done by me
pCR8 ATG NLS Ph5	Ph1 and 2 cloned in pCR8 ATG NLS		Restriction cloning done by me
pCR8 ATG NLS Ph6	Ph2 and 3 in pCR8 ATG NLS		Restriction cloning done by me
pCR8 ATG NLS Ph7	Ph1, 2 and 3 in pCR8 ATG NLS		Restriction cloning done by me
pHVW(Venus tagged)	Gateway destination plasmid for heat-shock inducible expression	DGRC stock# 1089	
pHVW-PhD1	Ph with Ph1 deleted in pCR8 and cloned in pHVW by Gateway cloning		
pHVW-PhD2	Ph with Ph2 deleted in pCR8 and cloned in pHVW by Gateway cloning		
pHVW-PhD3	Ph with Ph3 deleted in pCR8 and cloned in pHVW by Gateway cloning		

pHVW-PhD2D3	Ph with Ph2 and 3 deleted in pCR8 and gateway cloned in pHVW		
pHVW-PhD1D3	Ph with Ph1 and 3 deleted in pCR8 and gateway cloned in pHVW		Gateway cloning done by me
pHVW-Ph8	Ph with Ph1 and 2 deleted in pCR8 and gateway cloned in pHVW		
DSAM constructs	Ph with the same deletions as above including deletion of the SAM except for Ph8		
pHVW-Ph1	Ph1 in pCR8 gateway cloned in pHVW		Gateway cloning done by me
pHVW-Ph2	Ph2 in pCR8 gateway cloned in pHVW		Gateway cloning done by me
pHVW-Ph3	Ph3 in pCR8 gateway cloned in pHVW		Gateway cloning done by me
pHVW-Ph5	Ph1 and 2 in pCR8 gateway cloned in pHVW		Gateway cloning done by me
pHVW-Ph6	Ph2 and 3 in pCR8 gateway cloned in pHVW		Gateway cloning done by me
pHVW-Ph7	Ph1, 2 and 3 in pCR8 gateway cloned in pHVW		Gateway cloning done by me

Table 5. Antibodies used:

Antibody	Company/Source	Reference	Conc.
F2F4 c (anti-Cyclin B)	Developmental Studies Hybridoma Bank	AB 2245815	324 µg/mL
Anti-phospho-Histone H3 (Ser 10) clone 3H10	Millipore Sigma	05-806	1.0 mg/mL
Ph75I (WB and IF)	lab made	-	
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 647	Invitrogen	A32728	2.0 mg/mL
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	A11008	2.0 mg/mL
Alexa 555 anti-mouse	E. Lécuyer lab		
Anti-RFP-Tag antibody	St. Johns Laboratory	STJ97083	150 µg/50µL

GFP tag Polyclonal Antibody	Proteintech	50430-2-AP	61 µg/150µL
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 680	Invitrogen	A21057	2 mg/ml
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 680	Invitrogen	A21076	2 mg/ml
Monoclonal mouse anti-α-tubulin, clone B-5-1-2	Sigma-Aldrich	T5168	
IR Dye 800CW anti-Rabbit	Li-Cor	D00923-10	

Chapter 4: Results

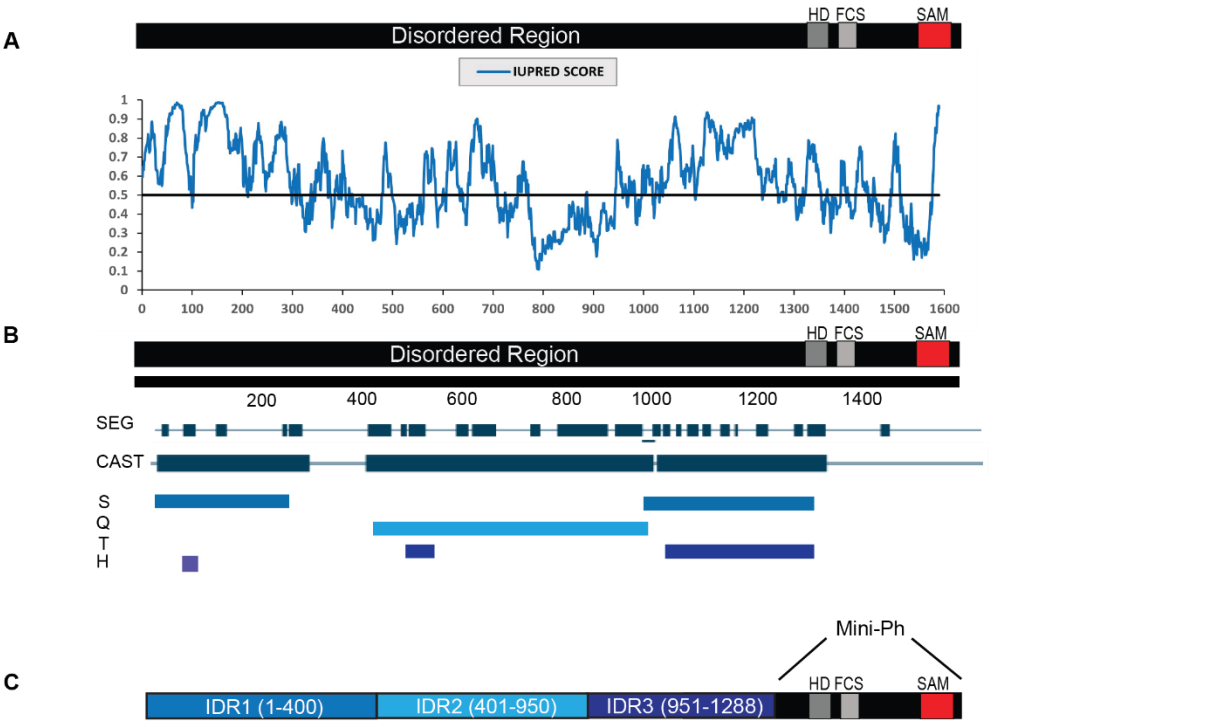
4.1 The N terminal region of Ph comprises three distinct IDRs

Drosophila melanogaster contains 2 tandem Ph genes (*Ph-p* and *Ph-d*), which are highly similar. 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 HD, FCS and SAM (**Figure 5A**). Using the IUPRED disorder predictor[121], the majority of the Ph N-terminal sequence is predicted to be disordered (**Figure 5A**).

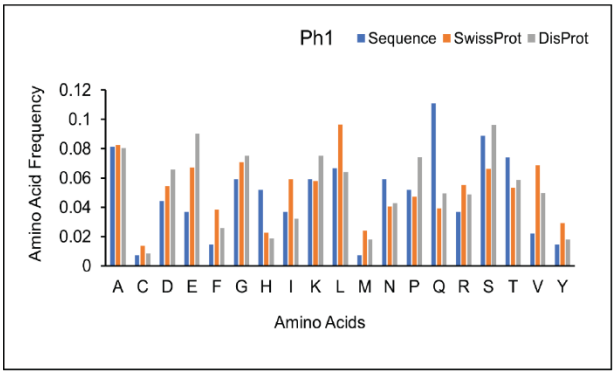
Intrinsically disordered protein sequences often also contain low complexity and compositionally biased sequence. To identify low complexity and compositionally biased regions, and potential subregions in the large disordered region, PlaToLoCo - Platform of Tools for Low Complexity was used.[122] This analysis tool applies an intersection between several algorithms like SEG (shortened from the word "segment"), CAST (complexity analysis of sequence tracts) which detects single residues masked for database searching[123], fLPS (Fast discovery of compositional biases for the protein universe), GBSC (manuscript in preparation by Jarnot et al 2020) and SIMPLE (detection of simple sequences in proteins) for detecting homorepeats like trinucleotides and tetranucleotides of amino acids.

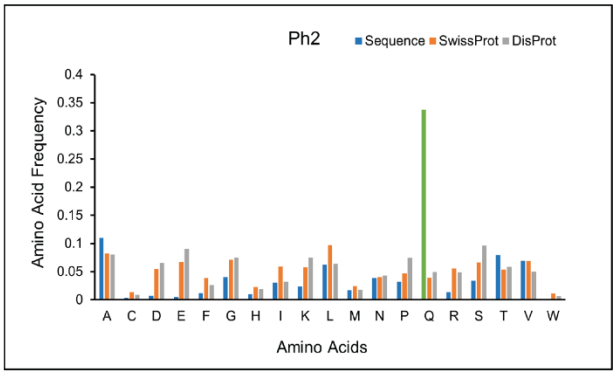
Although low complexity sequence is present throughout the N-terminal region, the masked (i.e., repetitive) amino acids are different in different regions of the sequence (**Figure 5B**). This allowed us to demarcate Ph1 (S-masked); Ph2 (Q-masked), and Ph3 (S+T masked) (**Figure 5B, C**). Ph1 does not have a clear enrichment for a single amino acid, but Ph2 is enriched in Glutamine (36%) and Ph3 in Threonine/Serine (20% each) (**Figure 5D**). PLAAC was used to

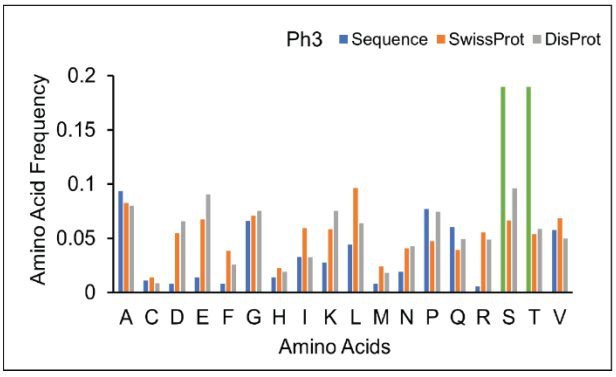
scan for prion like amino acid composition in the protein sequence.[124] Consistent with its high glutamine content, Ph2 has high scoring predicted prion-like regions (**Figure 5E**).


A
Disordered Region
HD FCS
SAM
IUPRED SCORE
B
Disordered Region
HD FCS
SAM
SEG
CAST
S
Q
T
H
C
Mini-Ph
HD FCS
SAM
IDR1 (1-400)
IDR2 (401-950)
IDR3 (951-1288)

D


Ph1
Sequence
SwissProt
DisProt
Amino Acid Frequency
Amino Acids


Ph2
Sequence
SwissProt
DisProt
Amino Acid Frequency
Amino Acids


Ph3
Sequence
SwissProt
DisProt
Amino Acid Frequency
Amino Acids

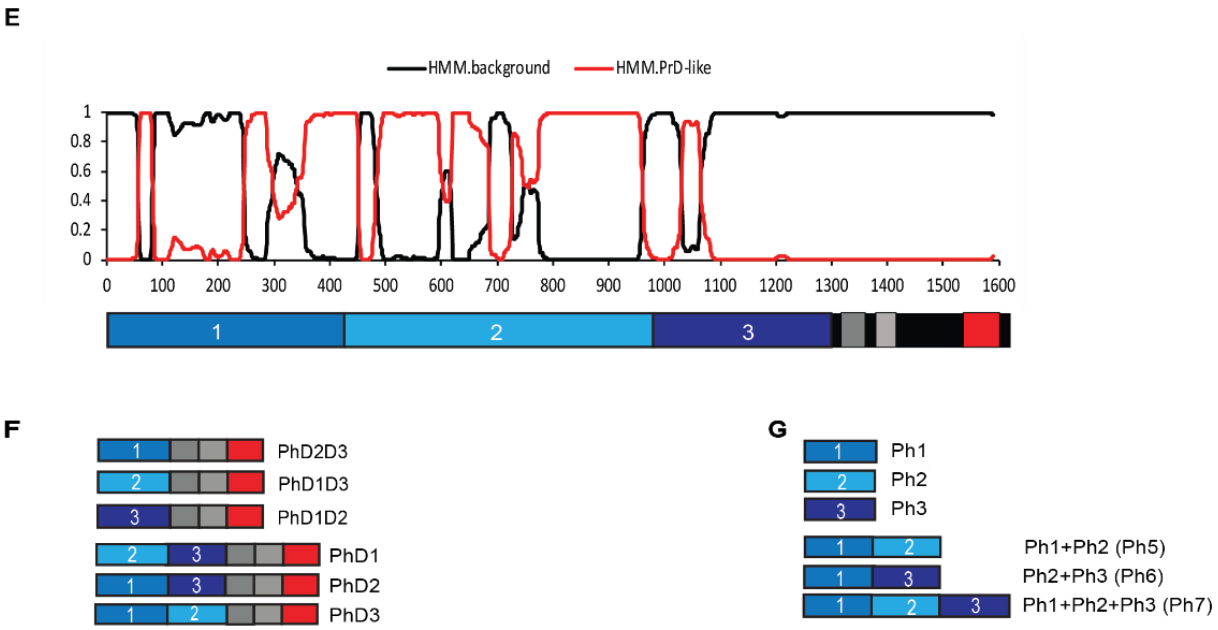


Figure 5. Identification of three IDRs in Ph. (A) Prediction of Ph disordered region using IUPRED. (Mészáros et al., 2018) A large part of Polyhomeotic is predicted to be disordered, as defined by predicted disorder propensity above 0.5 (shown by the black bar). **(B)** SEG/CAST analysis of Ph using PlaToLoCo - Platform of Tools for Low Complexity (Jarnot et al., 2020) 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 with IDR 2 possessing high scoring prion like regions (Lancaster et al., 2014). **(F)** Schematic of constructs with single IDRs removed from Ph fused to mini-Ph or single IDR fused to mini-Ph. **(G)** Schematic of truncation constructs to analyze IDRs alone.

4.2 Ph IDRs affect condensate formation

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 (**Figure 5F, G**). Constructs were designed to test the effect of 1) removing each IDR from Ph; 2) each IDR when fused to mini-Ph; 3) IDR removal in the absence of the SAM;

4) each IDR alone or in combination. These constructs were transiently transfected into *Drosophila* S2 cells, along with a plasmid encoding H2A-RFP as a nuclear marker. Venus-containing condensates formed in transfected cells were analyzed by live imaging. When wild-type (WT) Ph was transfected, it formed several condensates in cells which appear bright and round.

Removal of individual IDRs affects Ph condensate size, number, and morphology differently. Condensates formed in the absence of Ph1 are non-round, frequently small, and tend to form interconnected networks and, in some cases, large clumps. Removal of Ph2 results in a small number of large condensates. Condensates formed in the absence of Ph3 are bright and numerous, and most similar to wild type Ph condensates (**Figure 6A**).

To quantify the effects of each IDR, we developed Cell Profiler analysis pipelines to identify foci, count them, and measure their properties. We counted the number of condensates (foci) per cell, the median condensate size per nucleus, and the form factor (FF) of condensates. Form factor is a measure of the circularity of foci. Calculated as 4*π*Area/Perimeter, it equals 1 for a perfectly circular object.[127] We also measured the total Venus intensity in the nucleus, both for cells with condensates, and those without condensates. All measurements were made for at least 100 cells (with and without condensates) from each of the three independent experiments.

4.3 Removing single IDRs from Ph fused to mini-Ph affects condensate properties

PhD1 (single deletion of Ph1) condensates appear different from wild type. While the foci size is significantly different from WT Ph, the number is not (**Figure 6C, E**). The form factor has a wider range than that exhibited by WT Ph which could suggest variability in morphology (**Figure 6D**).

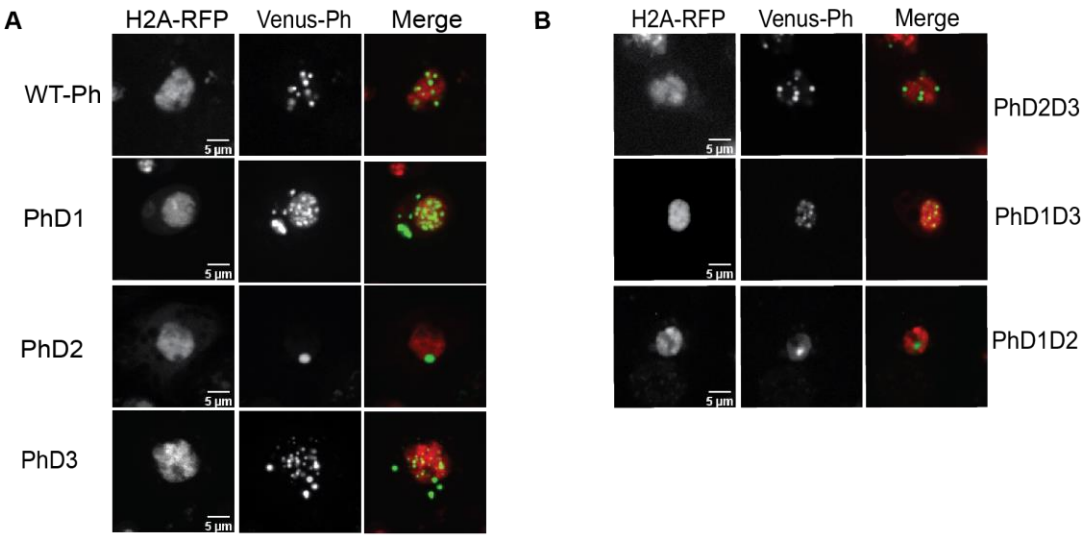
For PhD2 (single deletion of Ph2), while the number of foci is reduced compared to wild type, there is significant increase in the size of the foci (**Figure 6C, E**). A large fraction of foci formed by PhD2 have a form factor close to 1, consistent with the observed large, round condensates (**Figure 6D**).

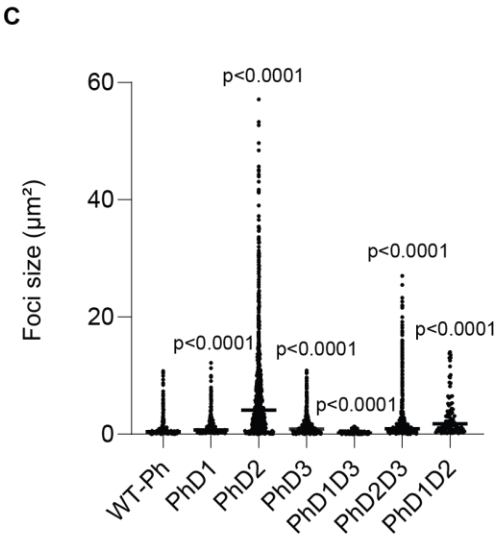
PhD3 (single deletion of Ph3) shows the largest increase in foci number when compared to wild type Ph. Foci number is almost double that of Wild Type (**Figure 6E**). Relative to wild type Ph, the size of foci is only slightly higher (**Figure 6C**). The form factor represented by a violin plot is similar to wild type (**Figure 6D**). Although we did not quantify cytoplasmic foci, we noticed that PhD3 frequently forms large cytoplasmic foci (see **Figure 6A** and **13A**).

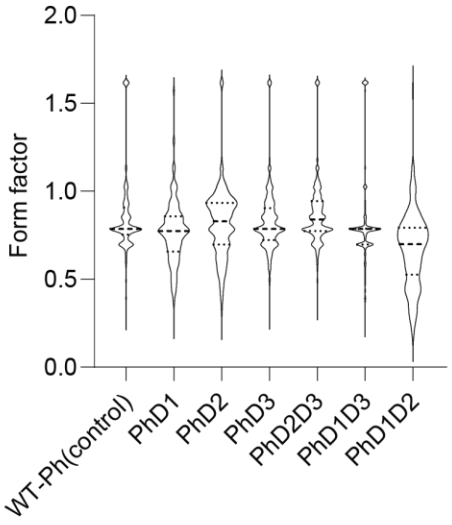
4.4 Single IDRs fused to Mini-Ph alter condensate properties

When we had a single IDR fused to mini-Ph, we again noticed changes in the properties of condensates (**Figure 6B**). When Ph1 was fused to mini-Ph (double deletions of Ph2 and Ph3 or PhD2D3), both the number and size of foci is slightly increased, when compared with WT Ph (**Figure 6C, E**). The form factor of the foci distribution is similar to wild type, except that the population of very round foci (form factor close to 1) is larger (**Figure 6D**).

With Ph2 fused to mini-Ph (double deletions of Ph1 and Ph3 or PhD1D3), 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 (**Figure 6C**). There are two distinct populations of foci morphology, both less than FF=1 (**Figure 6D**). When we have only Ph3 fused to mini-Ph (double deletions of Ph1 and Ph2 or PhD1D2), foci formation is reduced (**Figure 6E**). PhD1D2 either forms one large round structure in the cell (similar to what is observed with mini-Ph alone) or has diffused YFP signal (as is observed with DSAM and mini-Ph alone). The Form Factor looks very different for these foci as compared to the wild type (**Figure 6D**).


A
H2A-RFP
Venus-Ph
Merge
WT-Ph
PhD1
PhD2
PhD3
5 µm
B
H2A-RFP
Venus-Ph
Merge
PhD2D3
PhD1D3
PhD1D2
5 µm


C
Foci size (µm²)
60
40
20
0
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
p<0.0001
WT-Ph
PhD1
PhD2
PhD3
PhD1D3
PhD2D3
PhD1D2


D
Form factor
2.0
1.5
1.0
0.5
0.0
WT-Ph(control)
PhD1
PhD2
PhD3
PhD2D3
PhD1D3
PhD1D2

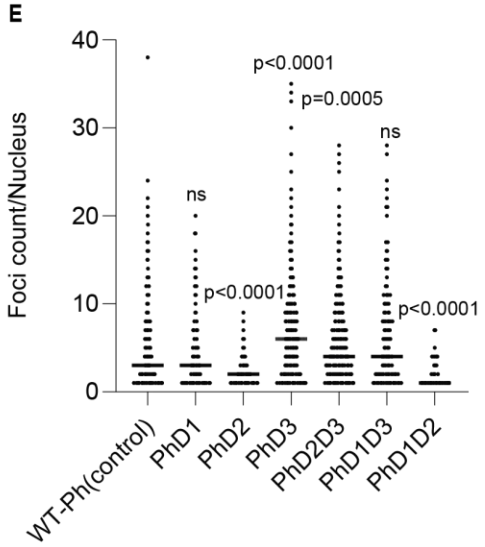

E
Foci count/Nucleus
40
30
20
10
0
p<0.0001
p=0.0005
ns
ns
p<0.0001
p<0.0001
WT-Ph(control)
PhD1
PhD2
PhD3
PhD2D3
PhD1D3
PhD1D2

Figure 6. Effect of Ph IDRs on condensate formation in cells. S2 cells were co-transfected with Venus expressing (green) **(A)** Deletions with single IDRs removed from Ph fused to mini-Ph or **(B)** Deletions with single IDR fused to mini-Ph and H2A-RFP as chromatin marker (red) and imaged by live imaging. **(C-E)** Effect of IDR deletions on foci size **(C)**, circularity **(D)**, and number **(E)**. Each Ph construct was compared to Wild-Type using Kruskal Wallis test with Dunnett's correction for multiple comparisons. Graphs show p-value. Bars show the median value for a total of at least 50 cells that formed condensates, from each of the three independent experiments. The total number of cells with and without foci were as follows for each construct: n=3478, n=2426, n=1539, n=1557, n=1568, n=2911, n=1038 for WT-Ph, PhD1, PhD2, PhD3, PhD1D3, PhD2D3, PhD1D2 respectively.

4.5 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. We found that the fraction of transfected cells that have foci is different for different constructs (**Figure 7A**). This could reflect differences in expression levels or that removing IDRs changes 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) (**Figure 8A-D**). Antibodies to tubulin and H2B were used as loading controls. We find that all the proteins are expressed at the expected size (**Figure 8A, B**). Quantification of GFP signal indicates an expression range of 1.5 to 3.5 fold relative to transfected wild-type Ph (**Figure 8E,**

F). In some cases, the expression levels correlate with condensate formation (e.g., PhD2D3 has higher expression and a higher fraction of cells with condensates) (**Figure 7A, 8A**). We also used antibodies to Ph to compare expression of transfected proteins to endogenous Ph (**Figure 9A-D**). Total Ph levels (endogenous + transfected) are increased by up to 2 times (**Figure 9C**). Transfection efficiency, determined by analysis of H2A-RFP versus endogenous H2B, was similar for all constructs (**Figure 8 G,H**).

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. For all cells expressing the co-transfected H2A-RFP, we measured total nuclear intensity of Venus, irrespective of whether cells contained condensates. **Figure 7B-H** shows the range of expression levels obtained, and the range over which condensates were observed. 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.

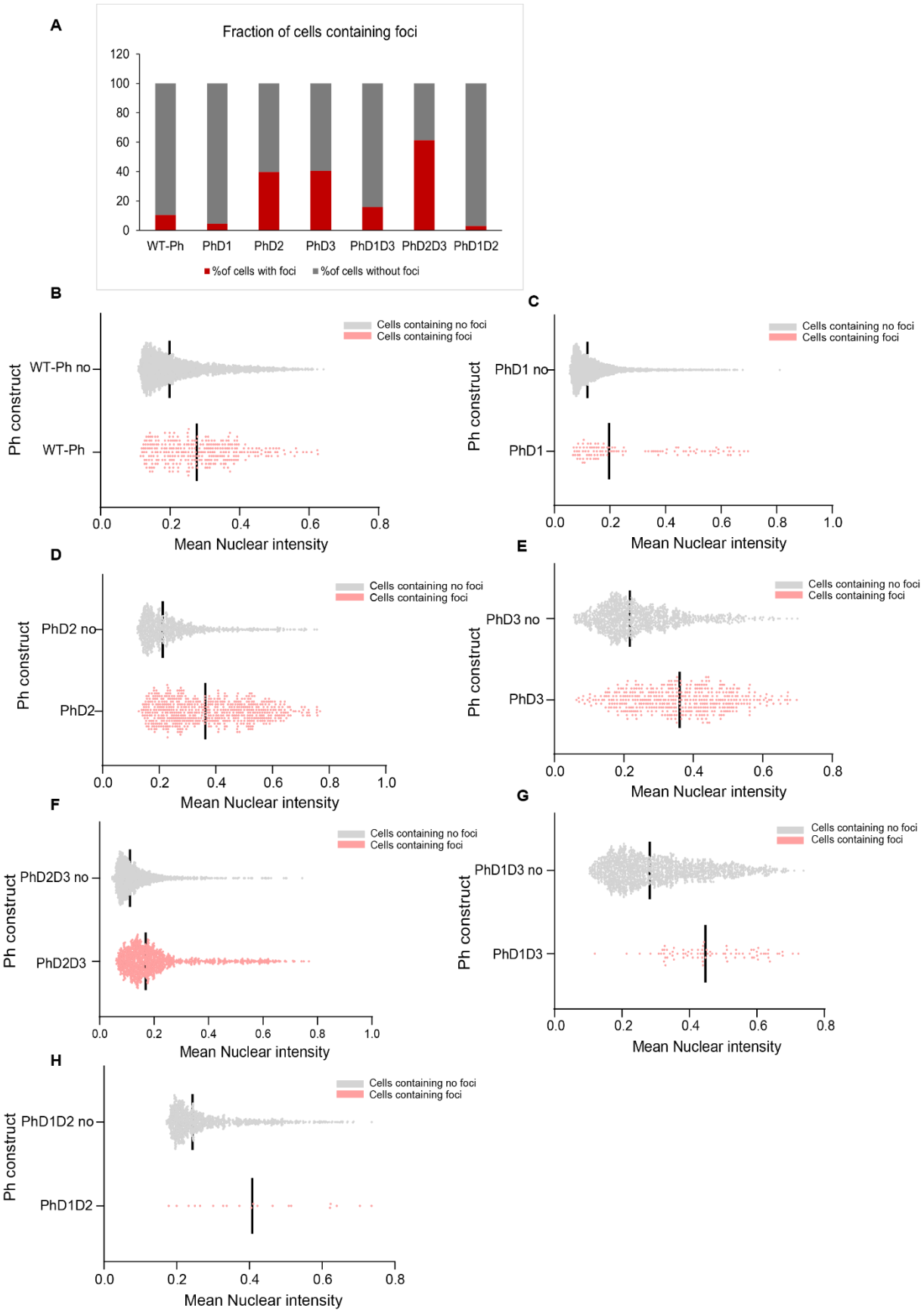

A
Fraction of cells containing foci
120
100
80
60
40
20
0
WT-Ph
PhD1
PhD2
PhD3
PhD1D3
PhD2D3
PhD1D2
%of cells with foci
%of cells without foci
B
Cells containing no foci
Cells containing foci
Ph construct
WT-Ph no
WT-Ph
0.0
0.2
0.4
0.6
0.8
Mean Nuclear intensity
C
Cells containing no foci
Cells containing foci
Ph construct
PhD1 no
PhD1
0.0
0.2
0.4
0.6
0.8
1.0
Mean Nuclear intensity
D
Cells containing no foci
Cells containing foci
Ph construct
PhD2 no
PhD2
0.0
0.2
0.4
0.6
0.8
1.0
Mean Nuclear intensity
E
Cells containing no foci
Cells containing foci
Ph construct
PhD3 no
PhD3
0.0
0.2
0.4
0.6
0.8
Mean Nuclear intensity
F
Cells containing no foci
Cells containing foci
Ph construct
PhD2D3 no
PhD2D3
0.0
0.2
0.4
0.6
0.8
1.0
Mean Nuclear intensity
G
Cells containing no foci
Cells containing foci
Ph construct
PhD1D3 no
PhD1D3
0.0
0.2
0.4
0.6
0.8
Mean Nuclear intensity
H
Cells containing no foci
Cells containing foci
Ph construct
PhD1D2 no
PhD1D2
0.0
0.2
0.4
0.6
0.8
Mean Nuclear intensity

Figure 7. Relationship between expression levels and condensate formation for Ph proteins lacking IDRs. (A) Fraction of transfected cells (identified by H2A-RFP) that formed condensates. **(B-H)** 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, however there is no clear threshold for foci formation. Black bar indicates the median number of cells.

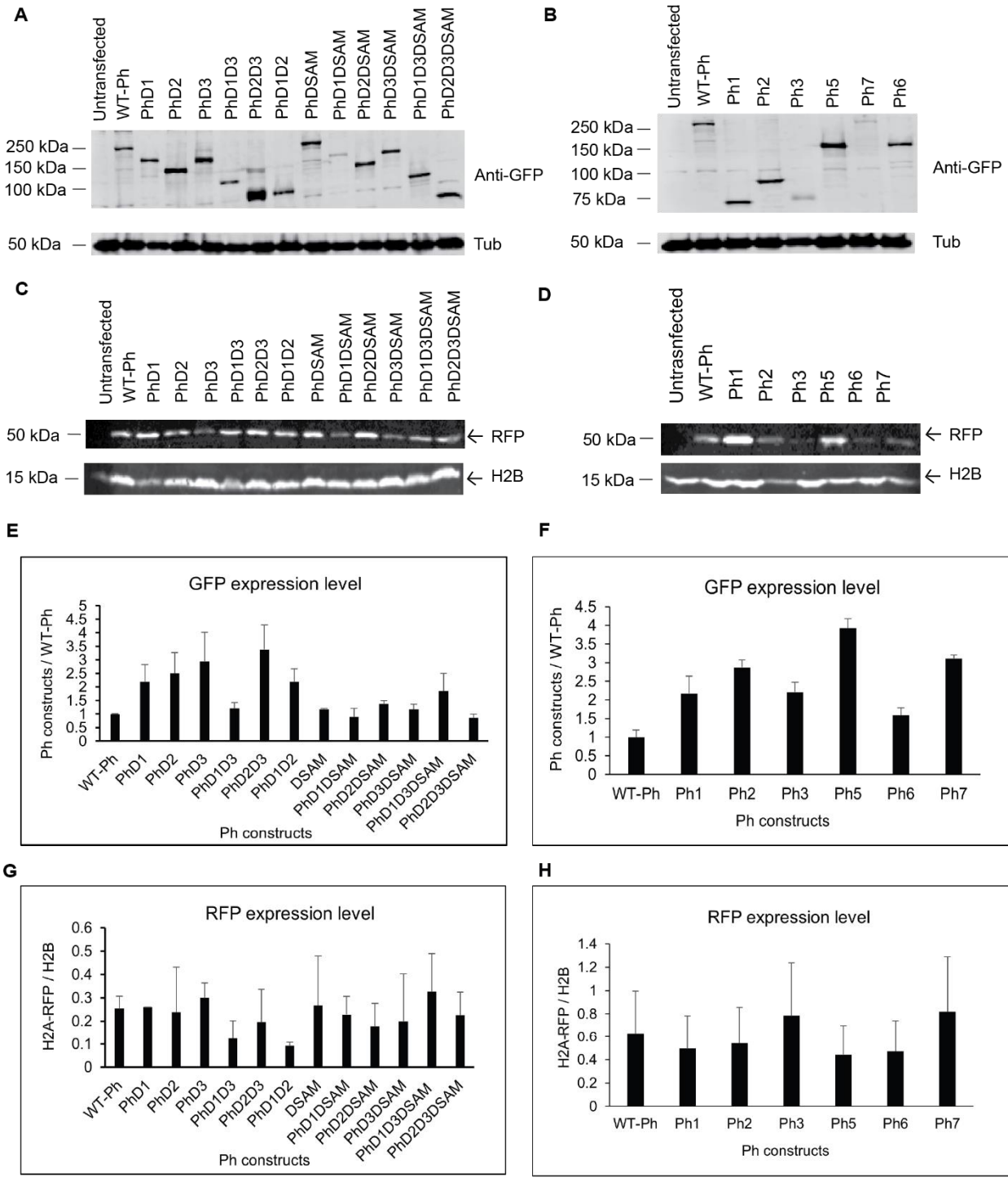

A
Untransfected
WT-Ph
PhD1
PhD2
PhD3
PhD1D3
PhD2D3
PhD1D2
PhDSAM
PhD1DSAM
PhD2DSAM
PhD3DSAM
PhD1D3DSAM
PhD2D3DSAM
250 kDa
150 kDa
100 kDa
Anti-GFP
50 kDa
Tub
B
Untransfected
WT-Ph
Ph1
Ph2
Ph3
Ph5
Ph7
Ph6
250 kDa
150 kDa
100 kDa
75 kDa
Anti-GFP
50 kDa
Tub
C
Untransfected
WT-Ph
PhD1
PhD2
PhD3
PhD1D3
PhD2D3
PhD1D2
PhDSAM
PhD1DSAM
PhD2DSAM
PhD3DSAM
PhD1D3DSAM
PhD2D3DSAM
50 kDa
RFP
15 kDa
H2B
D
Untransfected
WT-Ph
Ph1
Ph2
Ph3
Ph5
Ph6
Ph7
50 kDa
RFP
15 kDa
H2B
E
GFP expression level
Ph constructs / WT-Ph
WT-Ph
PhD1
PhD2
PhD3
PhD1D3
PhD2D3
PhD1D2
DSAM
PhD1DSAM
PhD2DSAM
PhD3DSAM
PhD1D3DSAM
PhD2D3DSAM
Ph constructs
F
GFP expression level
Ph constructs / WT-Ph
WT-Ph
Ph1
Ph2
Ph3
Ph5
Ph6
Ph7
Ph constructs
G
RFP expression level
H2A-RFP / H2B
WT-Ph
PhD1
PhD2
PhD3
PhD1D3
PhD2D3
PhD1D2
DSAM
PhD1DSAM
PhD2DSAM
PhD3DSAM
PhD1D3DSAM
PhD2D3DSAM
Ph constructs
H
RFP expression level
H2A-RFP / H2B
WT-Ph
Ph1
Ph2
Ph3
Ph5
Ph6
Ph7
Ph constructs

Figure 8. Ph proteins are expressed as full-length proteins. Representative Western blots for the transfections analyzing the expression levels of N-terminal tagged GFP for **(A)** the deletions and **(B)** truncations. **(C, D)** RFP tagged H2A and H2B expression levels for deletions and truncations. **(E, F)** Quantification of GFP expression levels by normalizing Ph to tubulin. Ph/Tubulin normalized to WT- Ph for Deletions and Truncations show 1.5-4.5 times expression levels compared to Wild Type. **(G, H)** Transfection efficiency determined by quantifying the expression of H2A-RFP for deletions and truncations, normalized to H2B. The graphs show average values from three replicates. The error bars represent standard error of mean (SEM). Transfection efficiency was similar across all constructs.

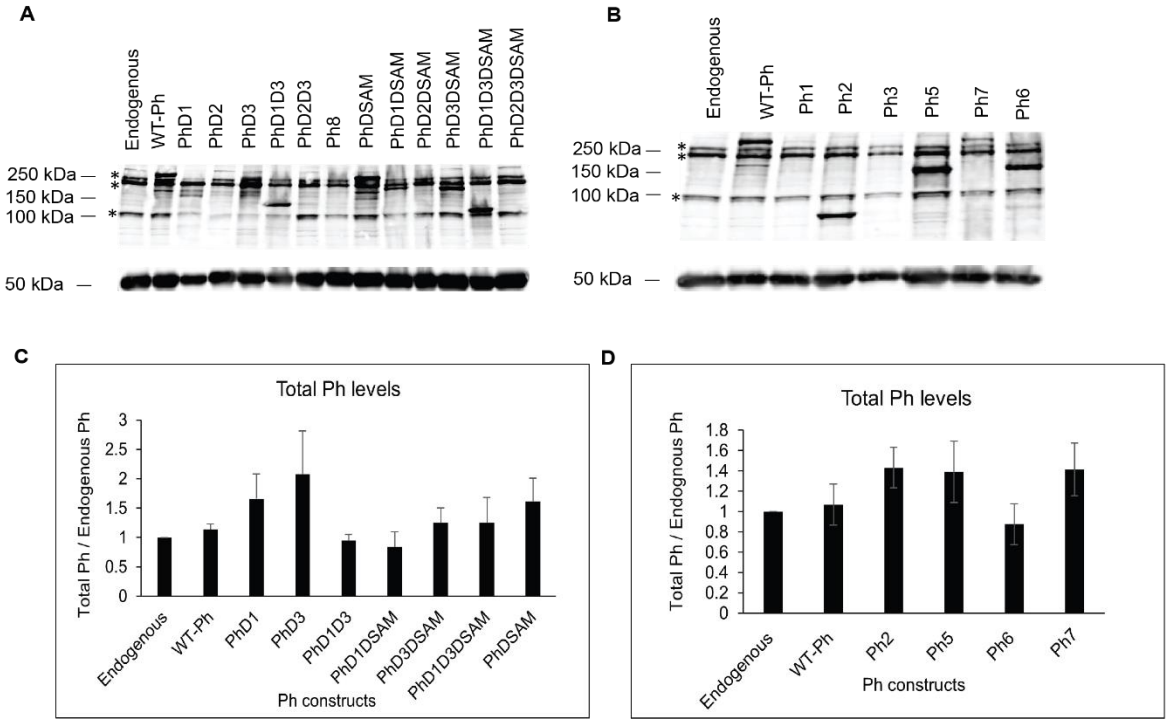


Figure 9. Comparison of transfected and endogenous Ph levels (A, B) Representative Western blots showing total Ph levels for transfected cells from analysis of deletions and truncations. Asterisks next to lane 1 indicate the position of endogenous Ph bands. Note that the antibody epitope is in Ph2 so that proteins lacking this region will not be detected. **(C, D)** Quantification of total Ph levels (endogenous + transfected) for deletions and truncations. Ph was normalized to tubulin, and Ph/tubulin for transfected

cells was normalized to Ph/tubulin for nontransfected cells. The highest expression signal obtained was ~2 times that of endogenous Ph. The graphs show average values from three replicates. The error bars represent standard error of mean (SEM).

4.6 Foci formation is SAM dependent in most cases

It has been shown previously that the Ph SAM is necessary for condensates to form.[70] To determine if the deletions form foci in the absence of SAM, we tested all the deletion constructs except PhD1D2 without SAM (**Figure 10A**). In the absence of the SAM, most proteins did not form foci. 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 do form these condensates is low (**Figure 10 A,B**). Although the DSAM versions of the proteins are expressed at about 2-fold lower levels than the corresponding proteins with the SAM (**Figure 8E**), given the wide range of expression levels over which foci are observed (**Figure 7B-H**), this is unlikely to explain why most DSAM proteins do not form foci. The DSAM versions are also expressed at similar levels as transfected WT-Ph (**Figure 8E**).

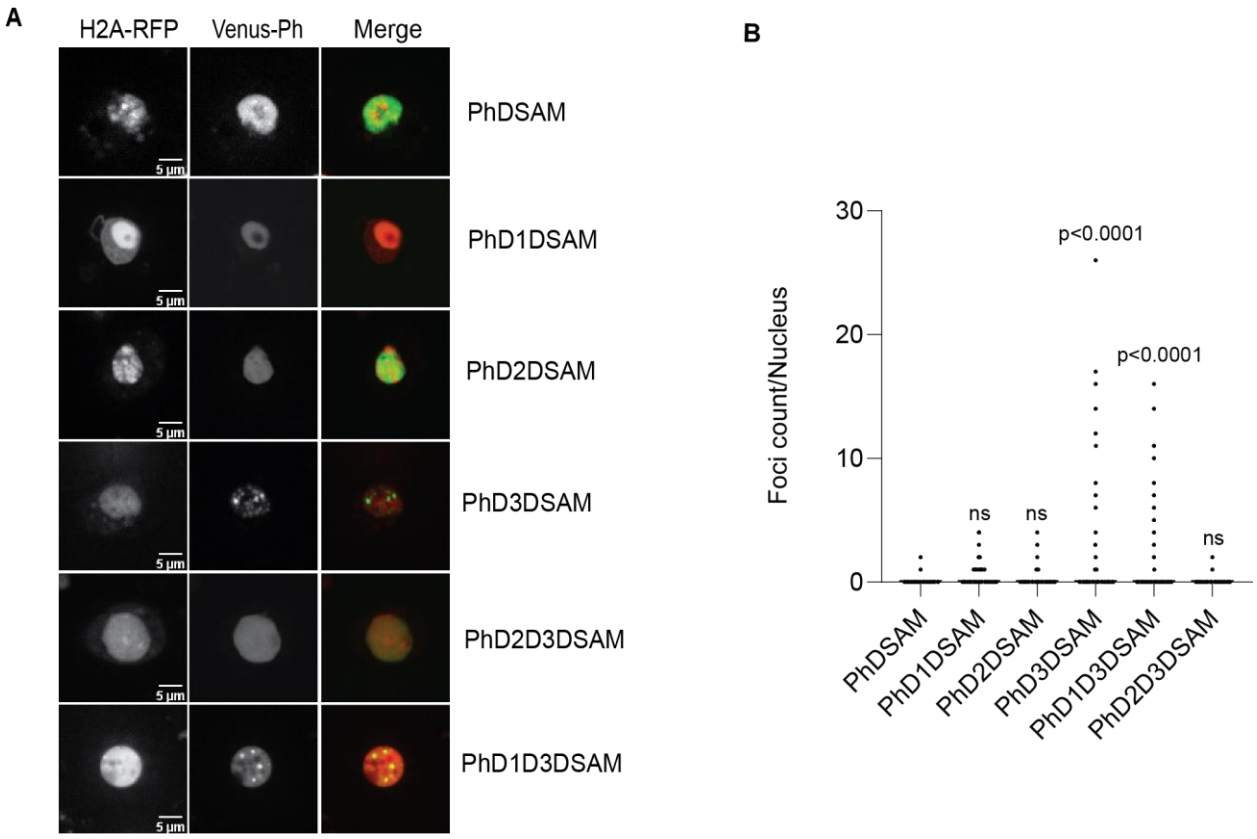


Figure 10. Effect of Ph IDRs on condensate formation in the absence of the SAM. (A) S2 cells were transfected with the Venus-Ph deletion constructs lacking the SAM. **(B)** Quantification of foci count for deletion constructs without SAM and H2A-RFP as a chromatin marker. Graph shows quantification for cells with and without foci. A total of atleast 100 cells from each of the three independent experiments were obtained. Each Ph construct was compared to PhDSAM using Kruskal Wallis test with Dunnett's correction for multiple comparisons. Graph shows the p values. The total number of cells with and without foci were as follows for each construct: n=1348, n=411, n=359, n=214, n=922, n=1413 for PhDSAM, PhD1DSAM, PhD2DSAM, PhD3DSAM, PhD1D3DSAM and PhD2D3DSAM respectively.

4.7 IDRS alone and in combination can form foci

To test if the IDRs can form condensates in the absence of mini-Ph (which includes the SAM), we tested the IDRs alone or in combination with each other (**Figure 11A**). Ph3 alone does not form foci. However, both Ph 1 and Ph2 form condensates in some cells (**Figure 11B**). When Ph1 and Ph2 (Ph5) are combined, many small condensates are formed, significantly higher than with WT-Ph (**Figure 11C**). In contrast, when Ph2 and Ph3 are combined, condensates do not form. Finally, when Ph1, Ph2, and Ph3 are combined, condensates are formed (**Figure 11B, C**). **Table 6** shows a comprehensive summary of foci formation for all the constructs.

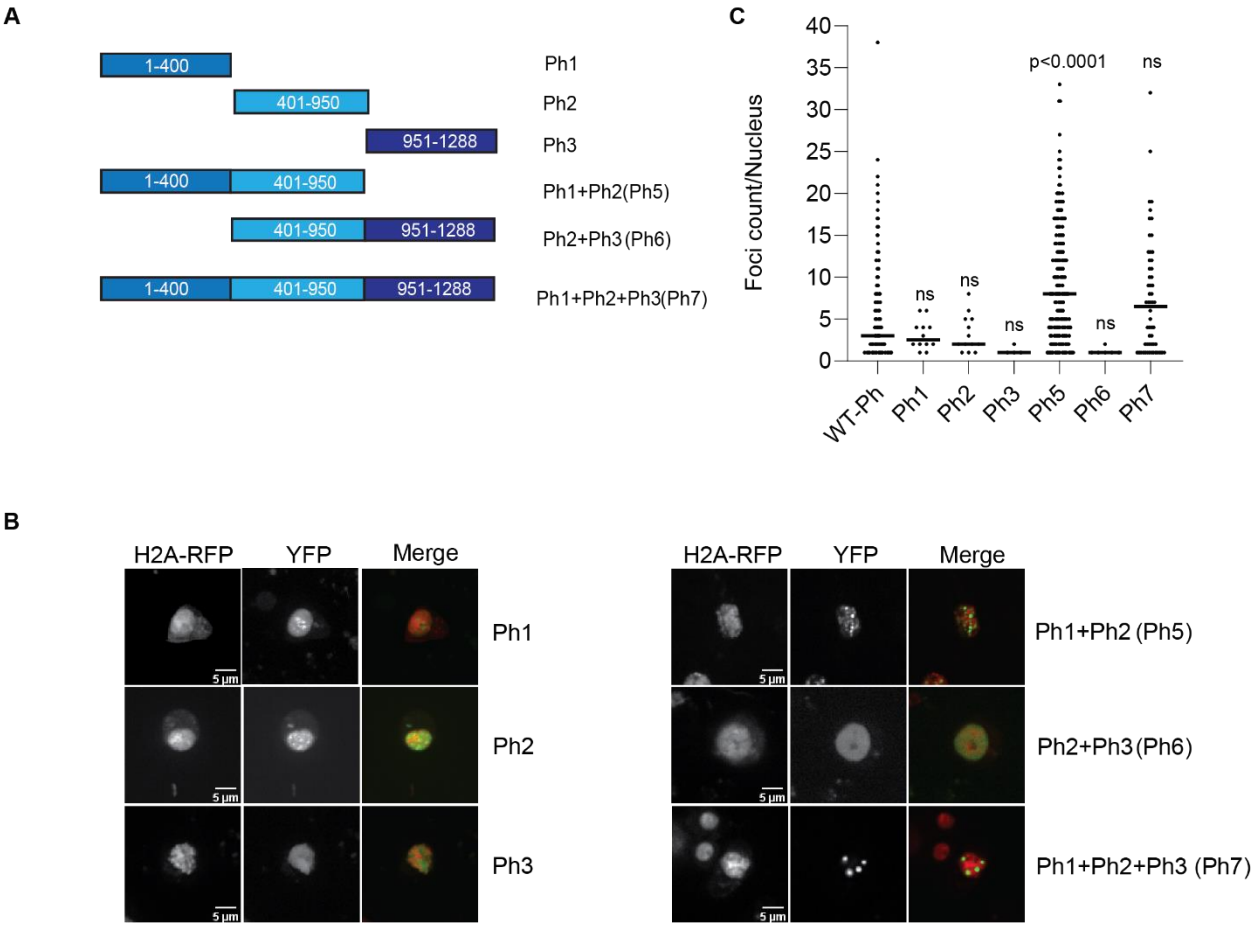


Figure 11. Ph IDRs alone can form foci. (A) Schematic of constructs for Ph truncations containing only IDRs. **(B)** Representative images of S2 cells transfected with Venus-Ph truncations and H2A-RFP as

chromatin marker**. (C)** Quantification of foci count for truncations. Bars show the median value for a total of at least 50 cells that formed foci, from each of the three independent experiments. Each Ph construct was compared to WT-Ph using Kruskal Wallis test with Dunnett's correction for multiple comparisons. Graph shows p value. The total number of cells with and without foci were as follows for each construct: n=268, n=412, n=309, n=857, n=460 and n=175 for Ph1, Ph2, Ph3, Ph5, Ph6 and Ph7 respectively.

Table 6. Summary of foci formation propensity in different constructs.

CONSTRUCT	SAM	FOCI
PhD1	+	+
PhD2	+	+
PhD3	+	+
PhD1D3	+	+
PhD2D3	+	+
PhD1D2	+	-
PhD1DSAM	-	-
PhD2DSAM	-	-
PhD3DSAM	-	+
PhD1D3DSAM	-	+

4.8 Condensates are present throughout S phase and their size increases during the cell cycle

PcG regulation is believed to be heritable through cell cycles (epigenetic). Key events in the cell cycle, DNA replication and mitosis, challenge chromatin-based information. Ph

condensates could be important during DNA replication to maintain high concentrations of PcG proteins near chromatin. If condensates are important for heritable regulation through mitosis, they must either persist on chromatin or be re-established at the end of mitosis. Whether condensates contribute to heritable regulation is not known. We therefore analyzed Ph condensates during S-phase and mitosis.

To determine if Ph condensates are present in S-phase and compare them in early and late S-phase, we labelled replicating DNA in cells transfected with Venus-Ph with EdU and analyzed condensates after fixing and staining (**Figure 12A**). The number of condensates per nucleus was compared for interphase cells (no EdU labeling), early S-phase, and late S-phase. The size of condensates was also measured in early and late S-phase cells and compared with interphase cells. We found that while the median number of foci remains constant between S phase and interphase as well as between early and late S (**Figure 12 B,C**), there is significant increase in the median size of foci per nucleus between early and late S phase. The condensates are larger in interphase (non-S) than S-phase cells (**Figure 12D**).

We also analyzed the IDR deletion constructs during S phase (**Figure 13**). However, the number of cells in S phase for cells transfected with the deletions was too low to obtain conclusive data (**Figure 13A**). Thus, no significant changes in either size or number of foci was detected for any of the IDR deletion constructs between S and non-S or early and late S (**Figure 13B, C**). It is possible that the IDR deletion proteins interfere with cell cycle progression, including DNA replication, which could be tested in future experiments.

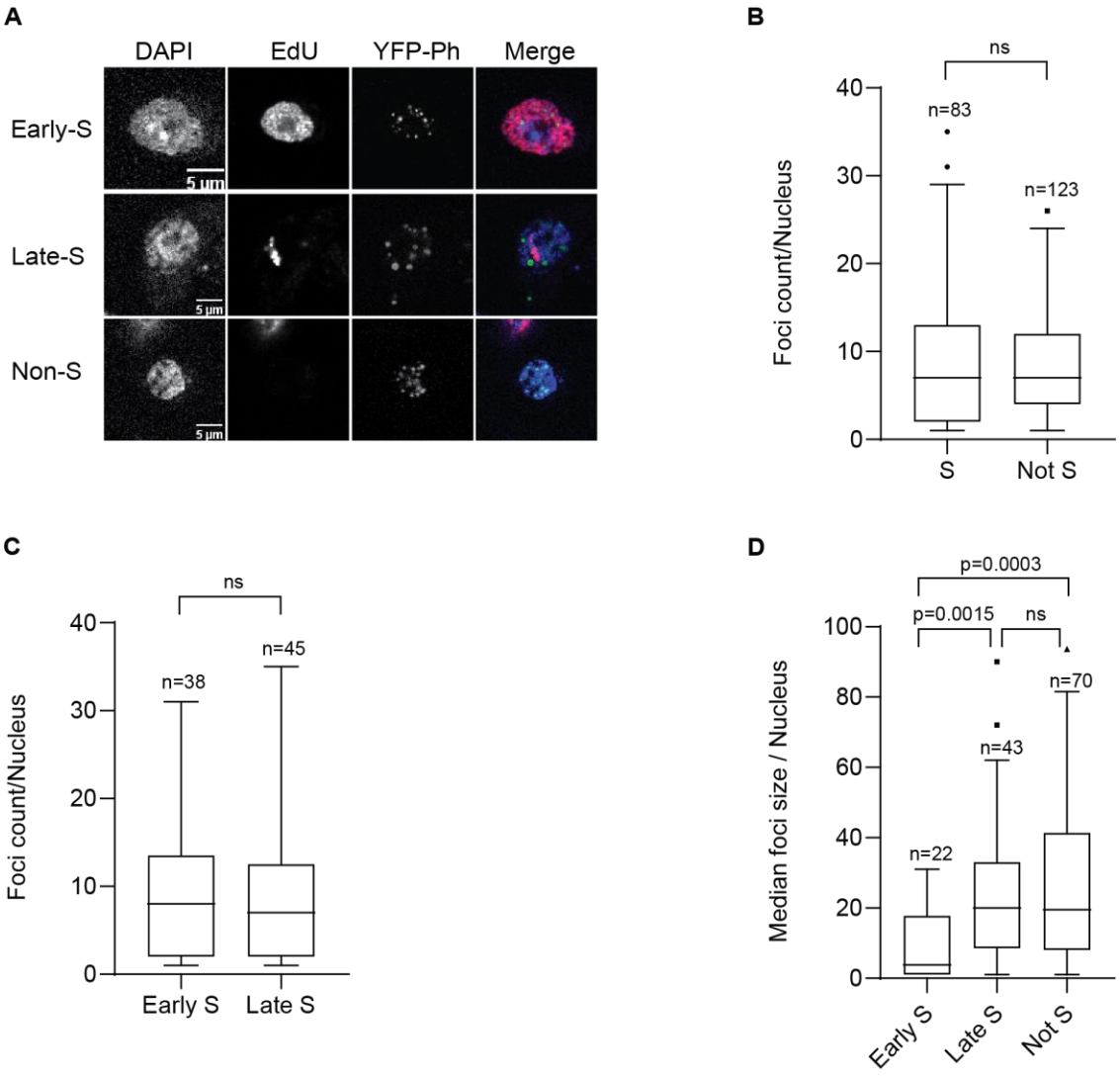


Figure 12. Ph condensate behaviour during S phase. (A) Representative images of Venus-Ph transfected S2 cells labelled with 10 µM EdU for 2 hours. EdU was detected with a Click reaction using Alexa Fluor 647 azide (red) and nuclei were visualized by DAPI staining (blue). **(B)** The number of WT-Ph foci was compared for S phase and non-S phase. **(C)** and Early S versus Late S phase. **(D)** Media foci size per nucleus was compared for Early, Late, and non-S phase cells. Statistics was carried out using Kruskal Wallis test for multiple comparisons from three independent experiments by comparing cells in

early S phase with late S and non-S with Dunnett's correction for multiple comparisons. Data points are shown for three replicates and n-values for number of cells with foci during early, late and non-S phase.

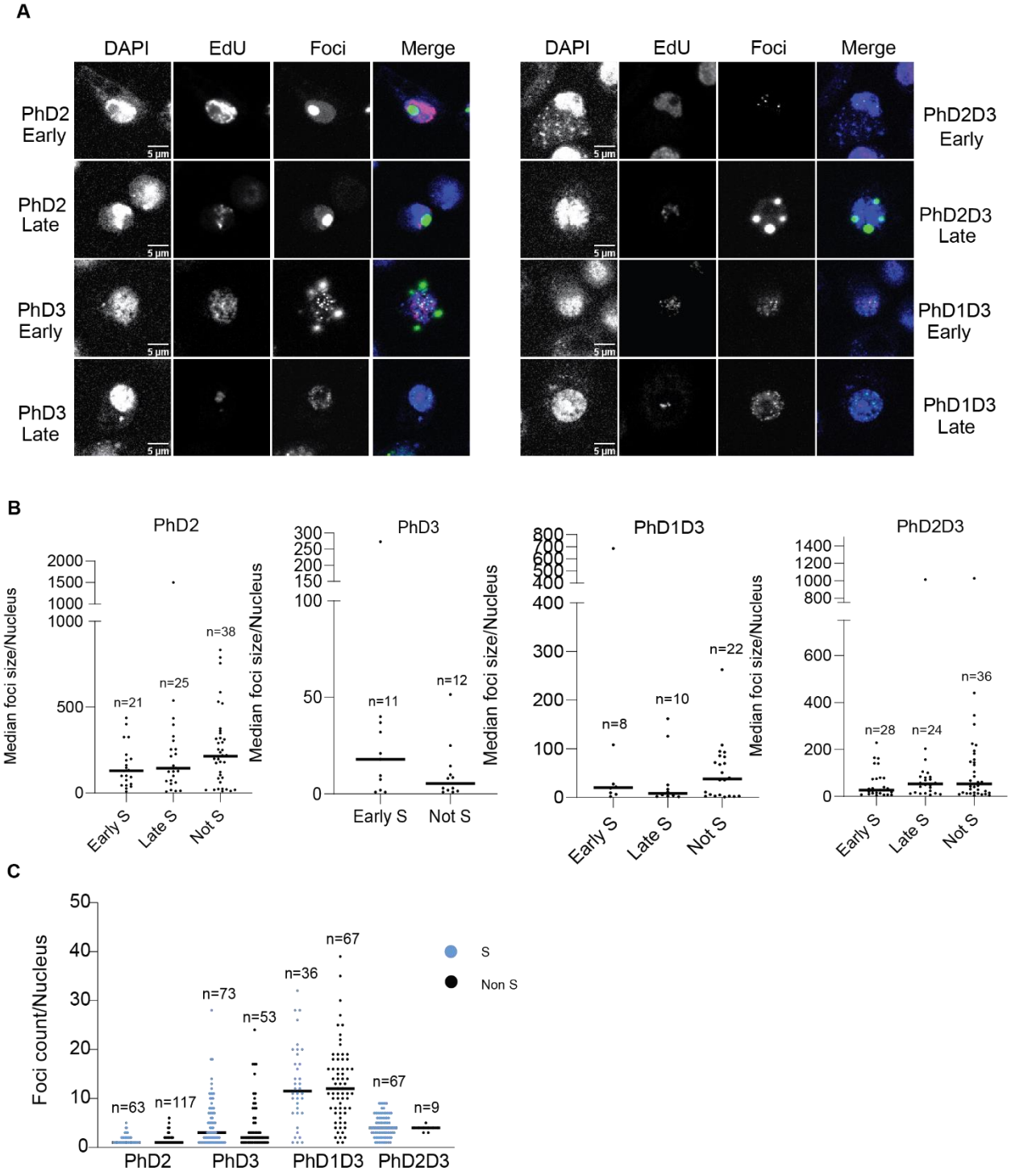


Figure 13. Effect of Ph IDRs on condensate formation deletions during S phase. (A) S2 cells were transfected with Venus-Ph lacking each IDR or combinations of them and stained for DAPI (blue). Replicating DNA was detected by EdU labeling followed by Click reaction using Alexa Fluor 647 (red)**. (B)** Median foci size per nucleus for each construct. Data points are shown for three replicates and n-values for number of cells with foci during early, late and non-S phase. **(C)** Median foci number per nucleus for each construct in S-phase (EdU positive) versus non-S phase cells.

4.9 Dynamic behavior of condensates during Mitosis

We previously showed that a small fraction of Ph is retained on mitotic chromosomes.[125] Staining of fixed cells with anti-Ph (endogenous Ph) and anti-H3S10p to identify mitotic cells indicates that condensates are not visible on metaphase chromosomes (**Figure 14A**). Since immunofluorescence of mitotic chromosomes is prone to technical artifacts, we used time-lapse imaging of Venus-Ph to analyze condensates through mitosis. Condensates formed by transfected Venus-Ph are larger than endogenous ones. Cells transfected with Venus-Ph were imaged every 30 minutes for 12 hours, and a small number of mitotic events were captured. **Figure 14** shows that condensates dissociate from chromatin and re-localize to or reform in the cytosol, decreasing from prophase through to anaphase (**Figure 14B**). As chromosomes segregate and nuclei start to reform, condensates re-appear on chromatin.

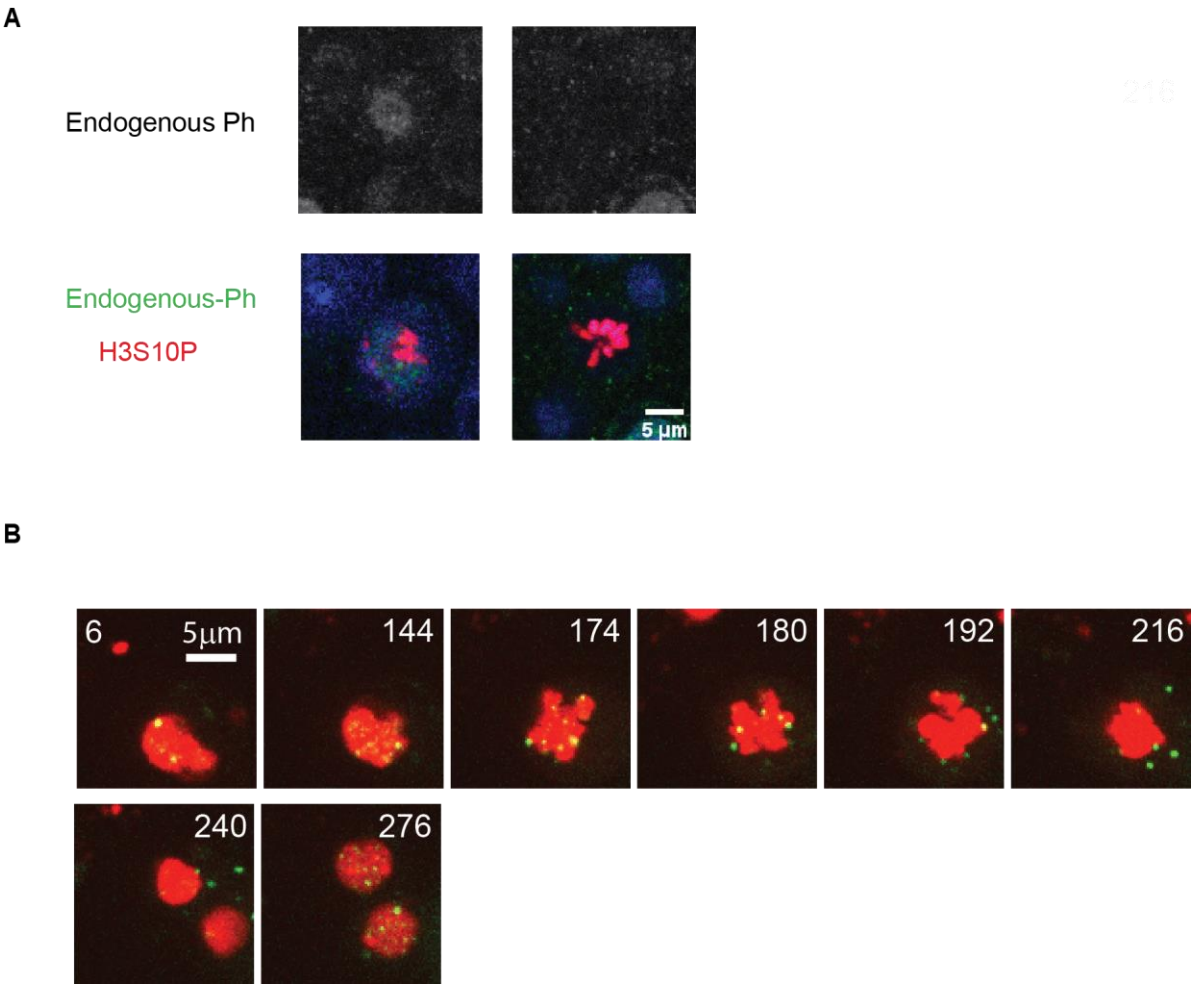


Figure 14. Dynamic behavior of Ph condensates during mitosis (A) Immunofluorescent images of mitotic S2 cells stained for endogenous Ph and H3S10p (Phosphorylation of serine 10 on histone H3). Scale bar is 5µm. **(B)** Live imaging of Venus-Ph in transfected S2 cells reveals dynamic behavior of condensates during Mitosis. Number indicates time in min for a cell starting to enter mitosis up until cytokinesis.

Chapter 5: Discussion and Future Perspective

We sought to determine how SAM dependent condensates that may form by phase separation are controlled by the disordered region in Ph. The identification of three distinct regions in Ph also referred to as Ph1, Ph2 and Ph3 reveals how each affects Ph condensates. Each IDR has a role to play, and the interplay of the IDRs with SAM is a function that has not been investigated before. Our structure function analysis has thus uncovered novel properties of the protein that may have important regulatory functions.

We have shown that Ph1 is important for condensates to form. This is evident in the context of SAM, where only Ph1 fused to mini-Ph (PhD2D3), shows an increase in the number of condensates (**Figure 6E**). The role of Ph1 in the absence of SAM is not clear and it may not be sufficient to induce condensate formation since Ph1 alone can form condensates, but they are observed in a small number of cells (**Figure 11C**).

Among the three IDRs, Ph2 shows an interesting behavior since it not only encourages condensate formation in the presence of SAM but also in its absence (**Figure 10A**). This function of Ph2 reveals that the disordered region could have a function or activity independent of the SAM. When both Ph1 and Ph2 are fused to mini-Ph (PhD3), condensate formation increases significantly (**Figure 6E)**. Ph1 fused to Ph2 alone also has a high propensity for condensate formation (**Figure 11C**). This evidence strongly points at Ph1 and Ph2 being important in inducing the formation of condensates. Ph2 has a high predicted prion like region predominantly due to homo-repeats of glutamine in its sequence. Many RNA-binding proteins similarly contain prion-like domains, that can affect RNA-protein condensates.[126] In this study RNA concentration was shown to affect the phase separation behaviour of prion like RNA binding proteins.[126] RNA

recruitment can affect the size of condensates formed by RNA-binding proteins,[127] although the contribution of prion like domains was not investigated in this study.

Our results show that Ph2 is not only involved in regulating the number but also the size of condensates. One possibility is that Ph2 effects on condensates is related to RNA, through direct RNA binding or interactions with RNA-binding proteins. Intriguingly, the DisoRDPbind algorithm, which predicts protein, DNA and RNA binding propensity in IDRs, predicts RNA binding activity for Ph2.[128] Another possibility is that the homo-repeats of polyglutamine in Ph2 encourage a coiled coil configuration. This alpha helical coiled coil formation has been shown to regulate phase separation behaviour. [129]

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 (PhD1D2), 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 (PhD2D3) forms several condensates, but the number is significantly reduced (and size increased) when Ph3 is present with Ph1 (PhD2) (**Figure 6E**). Third, Ph2 forms condensates in the absence of SAM (PhD1D3DSAM) but with Ph3 present does not (PhD1DSAM). 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. Ph3 is rich in serine and threonine and is known to be extensively modified with O-linked glycosylation of these residues. Removing glycosylation drives formation of Ph aggregates *in vitro* and *in vivo,* and these aggregates depend on the SAM.[77] 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. [130]

While our studies in the cellular context describe each IDR and their role, future studies could be done *in vitro* with purified proteins. Preliminary data from our lab indicates that Ph1 can form condensates with chromatin or DNA and can join mini-Ph condensates. It will be interesting to compare the *in vitro* activities of each IDR, alone and together *in vitro*. It would be interesting to test Ph2 interactions with RNA, whether the IDRs can form condensates with RNA, and whether RNA affects condensate size in a Ph2-dependent manner. Lastly, the possible control of Ph2 and Ph1 activity by Ph3 is an exciting observation that could also be investigated further using *in vitro* assays to test for direct interactions.

We studied the 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, use of an endogenous promoter (i.e. knock-in of a tagged Ph to the endogenous locus) would be a way to extend our findings to a more physiological setting. While we showed how each IDR regulates condensate formation, we did not account for interactions with endogenous Ph, or other cellular proteins. Immunoprecipitations could be used to test the extent to which the constructs used here interact with endogenous Ph under the overexpression conditions used. More broadly, effects of the Ph IDRs could involve interactions with other cellular proteins; affinity purification and mass spectrometry of Ph lacking each IDR, and of the IDRs alone, could identify these interactors.

We used automated image analysis with Cell Profiler to measure condensate properties in a large number of cells, and to compare them across all of the Ph constructs.[131] While Cell Profiler is a

powerful image analysis tool, it is nevertheless challenging to quantify condensate parameters. For condensates that have complex morphologies, are small, or are faint relative to surrounding levels (such as those formed by PhD1), 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.

Models for condensate number and size regulation

The significance of regulation of condensate size and number has been considered in other systems. The size and number regulation could be a mechanism by which condensates regulate several genes at once. [132] One example where this applies is the large size of condensates formed by Mediator as a possible mechanism allowing it to be in contact with transcription sites at multiple gene promoters. [133] For Polycomb bodies that are formed by clustering of PcG proteins, contacts with multiple repressive domains could underlie how larger sized bodies are formed. It is possible that the regulation of Ph condensate size by the Ph IDRs contributes to this process.

In a passive phase separating system (in the simplest case consisting of a single protein), multiple condensates will not exist at equilibrium because of a mechanism termed coarsening, which causes large condensates to grow at the expense of small ones, ultimately resulting in a single large condensate. Thus, for multiple condensates to co-exist, additional mechanisms must be involved (which could include not allowing the system to reach equilibrium). Many different mechanisms have been proposed, based on theoretical and/or experimental data.

Simulation studies have shown that protein-chromatin interactions can promote a multi-droplet state.[134] 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, this constrains the chromatin network. Thus, changing protein-chromatin interactions could change droplet number. [134]

Soding et al. proposed two mechanisms that can regulate condensate size based on compositional control of condensates.[81] The first suggested mechanism is termed the "enrichment-inhibition" model. This model posits that enzymes that reduce the phase separation propensity of the protein(s) that forms the condensate co-exist in the condensate. The activity of these enzymes counterbalances condensate growth, and can maintain condensates at a certain size. This model can also allow dynamic regulation of condensate size in response to changes in enzyme activity or abundance (for example across the cell cycle). Key predictions of the model are that there are PTMs (or possibly other mechanisms such as ligand binding) that can regulate phase separation, and that the enzymes/molecules involved are concentrated in condensates. P granule and stress granule size regulation by the DYRK3 kinase, nuclear speckles, Cajal bodies, and synaptic vesicles are all suggested to use this mechanism.

The second mechanism is termed "localization-induction". This model posits that a localized enzyme that modifies its protein substrate(s) in a manner that promotes condensate formation leads to formation of condensates with the enzyme at their core. Condensate size is limited by the radius around the enzyme cluster where modified proteins are sufficiently concentrated to phase separate. Key predictions of this model are localization of the enzyme, and its ability to trigger phase separation through modification of its substrates. Condensates that form in a highly localized manner, including condensates at DNA double-strand breaks, clusters formed by

transmembrane signaling molecules, and assembled centrosomal microtubules are believed to be examples of this mechanism.

Other mechanisms of size control that do not depend on post-translational modifications include formation of condensates around a scaffold (such as an RNA or chromatin region), where the binding sites on the scaffold limit the condensate size. Mechanical restriction of condensate size (for example by the actin cytoskeleton or chromatin) can also restrict condensate size. There is also no reason why a single mechanism would control condensate size.

Which of these mechanisms is likely to apply to Ph? Mini-Ph forms a single condensate in cells; one interpretation of this is that it reflects passive phase separation and coarsening. It could also mean that activities like post-translational modifications are not occurring, and protein-chromatin interactions are limited. Addition of Ph1 or 2 (or both) to mini-Ph results in formation of multiple condensates. This could indicate that information for size/number regulation is encoded in the sequences of Ph1 and 2. Ph3 restricts condensate formation, in a manner that depends on the other IDRs (especially Ph2). In the case of Ph3, it is known to be glycosylated. It is therefore possible that size/number control through Ph3 follows the enrichment-inhibition model. A clear prediction would be that changing glycosylation by changing levels of Ogt (the enzyme that adds glycosylation) or OGA (the enzyme that removes it) would alter condensate size/number in a Ph3-dependent manner. A second prediction is that Ogt localizes to Ph condensates. Ph3 can also be phosphorylated, so that its activity may not be due solely to glycosylation. In the case of Ph1 and 2, nothing is known about post-translational modification of these regions. Analyzing modifications in these regions (using mass spectrometry) would begin to address possible mechanisms by which these IDRs regulate condensate size. A scaffold model, in which specific chromatin regions scaffold condensates could also explain how the IDRs regulate condensate

number and size. This model could be tested by determining whether condensates depend on a chromatin scaffold (for example by digesting the chromatin in permeabilized cells and monitoring condensates by imaging). In a chromatin scaffold model, Ph with different IDRs deleted may also be recruited to different chromatin regions, which could be measured by ChIP.[127] Protein-chromatin interactions mediated by the IDRs could also promote a multi-droplet state as in the model of Qi and Zhang.[129] A simple hypothesis that arises from this model with respect to Ph2 is that Ph2 interacts with chromatin (directly or indirectly)—deleting Ph2 may release Ph from chromatin, allowing condensates to fuse, and explaining the large condensates observed in PhD2.

The shape or circularity has been used as a determining factor for liquid like behaviour. However, additional assessment using FRAP would be required to confirm it. Our analysis of the shape from form factor measurements reveals considerable differences in circularity for each construct. Loss of circularity has been attributed in some cases, to increased association with chromatin which was reported for HP1a condensates. [135]

Among the models discussed for condensate size and number, protein-chromatin interactions which encourage multi-droplet state could also be affecting circularity. This is evident for all constructs apart from PhD2 which have a high population of non-circular condensates. Another potential factor affecting circularity could be post translational modifications. It was shown in that acetylation of histone tails by tuning the density and material properties of chromatin droplets led to a loss of droplet circularity with time. However, further studies depicting the effect of PTMs on circularity are required.[136]

The form factor closest to 1 is observed for PhD2. It is speculated that circularity is retained in phase separated systems containing less DNA and more RNA as well as other proteins that can

freely intermix. This would not explain why removing Ph2 which is RNA binding would cause more circular looking condensates. Another explanation could be attributed to the physical nature of surface tension driving the droplets to a more spherical state.[137] Thus, the mechanisms affecting size and number of condensates could also subsequently contribute to and explain the variability in morphology.

Behaviour of condensates during the cell cycle

We found that the size, but not the number, of condensates increases from early to late S-phase. Contrary to this, other studies report a link between condensate size and number. For example, inhibition of transcription was shown to increase the size of nuclear speckles, making them rounder while a decrease in their number was observed.[138] Repeated fusion events led to a decrease in the speckle number and increase in the mean speckle size. One explanation for our observations could be that condensate fusion events increase in S-phase (leading to increased size) and are counterbalanced by new condensate formation, or condensate splitting, in S-phase, thus maintaining condensate number.

Alternatively, there could be another mechanism at play, which prevents coalescence between droplets and instead leads to increase in individual condensate size through recruitment of soluble protein (monomers or small aggregates). This could follow the classical nucleation model where nucleation occurs from monomers to form oligomers. [139] It is also possible that micro-condensates, not visible with our methods, undergo fusion events with large condensates.[140] Mechanisms that block condensate coalescence could be the same as those discussed above.

Regardless of what mechanism is involved, our results indicate that condensates are dynamic through S-phase, so that at least one parameter controlling their size must be altered. Post-translational modification (perhaps of the IDRs) is a candidate mechanism. O-GlcNac is globally decreased towards G2/M in cultured mammalian cells, if this is also true in *Drosophila* cells, it might contribute to the size increase (by decreasing Ph glycosylation). [141] Another possibility is suggested by analysis of Histone Locus bodies (HLB), which increase in size during S phase before entry into mitosis. [142] HLB size is correlated with histone transcription. It is also correlated with the number of histone genes present—HLBs on paired chromosomes are twice as large as those on unpaired chromosomes. Finally, HLB size is also regulated by Cdk activity.[142] In the case of Ph, it is possible that the increase in size across S-phase is related to the doubling of chromatin that occurs when the DNA is replicated. A prediction of this model would be that condensates double in size as cells progress through S-phase. This could be tested in our system by live imaging across S-phase, or in a system where endogenous Ph was tagged so that Ph levels are the same in all cells.

We observe dynamic behavior of Ph condensates during mitosis. This is consistent with several studies in the literature. PML bodies demonstrate continuous change in number, biochemistry, and integrity over the course of the cell cycle.[87] The number of PML bodies dropped during mitosis due to mitotic aggregation of PML bodies into larger bodies. MORC-3 nuclear bodies also exhibit a drastic change in number from interphase to metaphase reducing from 40 to 2 or 3 in metaphase, suggesting cell-cycle dependence. [143] This decrease was correlated with chromatin compaction occurring during mitosis. Although we do not know which aspects of Ph condensates change in mitosis, cytoplasmic condensates are present in mitotic cells, indicating that mitotic condensates differ fundamentally from interphase ones in that they do not include chromatin.

We attempted to analyze the effect of the Ph IDRs across the cell cycle. However, the number of cells obtained during S phase for most of the deletions was very low. We also were not able to observe mitotic cells expressing Ph lacking IDRs in preliminary live imaging experiments. It is therefore possible that Ph lacking IDRs interferes with cell cycle progression, particularly at the high expression levels obtained with transfection. Future experiments using proteins expressed at physiological levels can be used to test whether the Ph IDRs contribute to condensate regulation across the cell cycle.

Chapter 6: Conclusion

We showed that the disordered region in Ph affects SAM dependent condensate formation (**Figure 15**). Three disordered regions in the sequence of Ph were identified. These regions namely Ph1, Ph2 and Ph3 affect condensation formation differently. While Ph1 and Ph2 are condensate promoting, Ph3 plays an inhibitory role. Ph2 also controls the size of condensates. Furthermore, we investigated the behaviour of Ph condensates during the cell cycle. During DNA replication Ph condensates are present and do not change in number from early to late S phase. However, there is an increase in the size of condensates as they go from early to late S. Additionally, we also reported dynamic behaviour of condensates during Mitosis. Since we have shown the regulatory behavior of the disordered regions, it would be interesting to see their interactions with other proteins or molecules like RNA as well as their effect on Ph function.

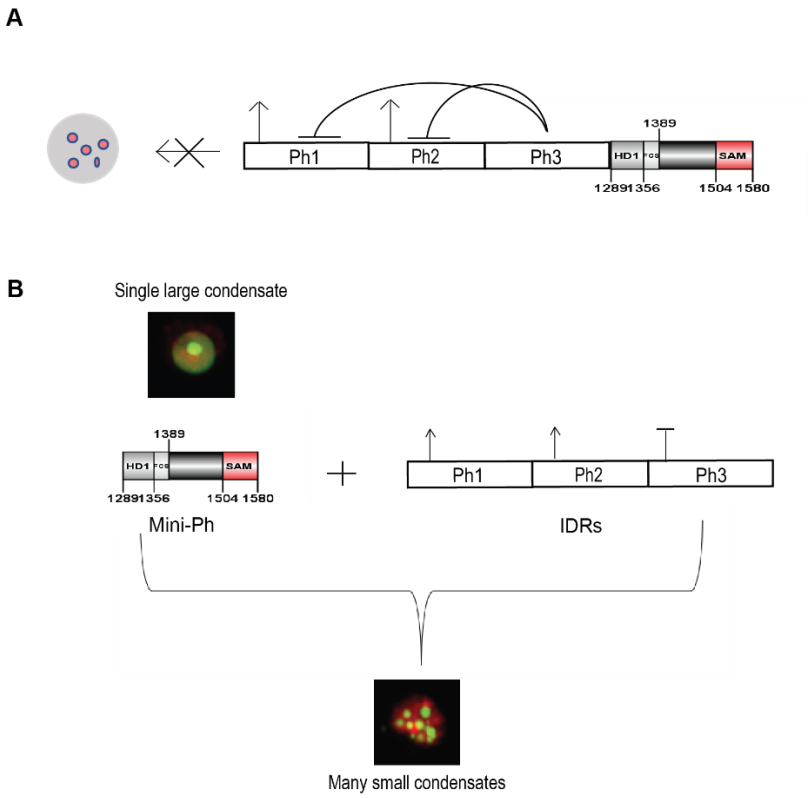


Figure 15. Model for role of Ph IDRs in SAM dependent condensate formation. (A) Working model for regulatory function of IDRs. Ph3 functions through inhibition of Ph1 and Ph2, both of which promote condensate formation. **(B)** Depiction of requirement of IDRs for condensate formation to occur with mini-Ph.

REFERENCES

1. Goldberg AD, Allis CD, Bernstein E. Epigenetics: A Landscape Takes Shape. *Cell*. 2007;128(4):635-638. doi:10.1016/J.CELL.2007.02.006

2. Bird A. Perceptions of epigenetics. *Nature 2007 447:7143*. 2007;447(7143):396-398. doi:10.1038/nature05913

3. Mohn F, Schübeler D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends in Genetics*. 2009;25(3):129-136. doi:10.1016/J.TIG.2008.12.005

4. Boettiger AN, Bintu B, Moffitt JR, et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*. 2016;529(7586):418. doi:10.1038/NATURE16496

5. Thiagalingam S. Epigenetic memory in development and disease: Unraveling the mechanism. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2020;1873(2):188349. doi:10.1016/J.BBCAN.2020.188349

6. Zhang H, Xiong J, Li P, et al. Liquid-liquid phase separation in biology: mechanisms, physiological functions and human diseases SCIENCE CHINA Life Sciences. 2020;63(7):953-985. doi:10.1007/s11427-020-1702-x

7. Trojer P, Reinberg D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell*. 2007;28(1):1-13. doi:10.1016/J.MOLCEL.2007.09.011

8. Jarome TJ, Thomas JS, Lubin FD. The Epigenetic Basis of Memory Formation and Storage. *Progress in Molecular Biology and Translational Science*. 2014;128(C):1-27. doi:10.1016/B978-0-12-800977-2.00001-2

9. S H, JM G. Epigenetics, cellular memory and gene regulation. *Current biology : CB*. 2016;26(14):R644-R648. doi:10.1016/J.CUB.2016.06.011

10. E J, MJ L. The changing concept of epigenetics. *Annals of the New York Academy of Sciences*. 2002;981:82-96. doi:10.1111/J.1749-6632.2002.TB04913.X

11. Probst A v., Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. *Nature Reviews Molecular Cell Biology 2009 10:3*. 2009;10(3):192-206. doi:10.1038/nrm2640

12. Groth A, Rocha W, Verreault A, Almouzni G. Chromatin Challenges during DNA Replication and Repair. *Cell*. 2007;128(4):721-733. doi:10.1016/J.CELL.2007.01.030

13. Lanzuolo C, Sardo F lo, Diamantini A, Orlando V. PcG Complexes Set the Stage for Epigenetic Inheritance of Gene Silencing in Early S Phase before Replication. *PLOS Genetics*. 2011;7(11):e1002370. doi:10.1371/JOURNAL.PGEN.1002370

14. Budhavarapu VN, Chavez M, Tyler JK. How is epigenetic information maintained through DNA replication? *Epigenetics & Chromatin 2013 6:1*. 2013;6(1):1-7. doi:10.1186/1756-8935-6-32

15. Bantignies F, Grimaud C, Lavrov S, Gabut M, Cavalli G. Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. *Genes & Development*. 2003;17(19):2406-2420. doi:10.1101/GAD.269503

16. Kouzarides T. Leading Edge Chromatin Modifications and Their Function. doi:10.1016/j.cell.2007.02.005

17. Villaseñor R, Baubec T. Regulatory mechanisms governing chromatin organization and function. *Current Opinion in Cell Biology*. 2021;70:10-17. doi:10.1016/J.CEB.2020.10.015

18. Francis NJ, Sihou D. Inheritance of Histone (H3/H4): A Binary Choice? Published online 2021. doi:10.1016/j.tibs.2020.08.009

19. Francis NJ. Mechanisms of epigenetic inheritance: Copying of Polycomb repressed chromatin. *Cell Cycle*. 2009;8(21):3521-3526. doi:10.4161/cc.8.21.9876

20. Pirrotta V. Introduction to Polycomb Group Mechanisms. In: *Polycomb Group Proteins*. Elsevier Inc.; 2017:1-3. doi:10.1016/B978-0-12-809737-3.00001-5

21. Karch F, Weiffenbach B, Peifer M, et al. The abdominal region of the bithorax complex. *Cell*. 1985;43(1):81-96. doi:10.1016/0092-8674(85)90014-5

22. Dura J-M, Brock HW, Santamaria P. *Polyhomeotic: A Gene of Drosophila Melanogaster Required for Correct Expression of Segmental Identity*. Vol 198.; 1985.

23. Bender W, Akam M, Karch F, et al. Molecular genetics of the bithorax complex in Drosophila melanogaster. *Science*. 1983;221(4605):23-29. doi:10.1126/science.221.4605.23

24. Kassis JA, Kennison JA, Tamkun JW. Polycomb and trithorax group genes in drosophila. *Genetics*. 2017;206(4):1699-1725. doi:10.1534/genetics.115.185116

25. Lanzuolo C, Orlando V. Memories from the Polycomb Group Proteins. Published online 2012. doi:10.1146/annurev-genet-110711-155603

26. Piunti A, Shilatifard A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nature Reviews Molecular Cell Biology*. 2021;22(5):326-345. doi:10.1038/s41580-021-00341-1

27. Pasini D, di Croce L. Emerging roles for Polycomb proteins in cancer. *Current Opinion in Genetics and Development*. 2016;36:50-58. doi:10.1016/j.gde.2016.03.013

28. Conway EM, Bracken AP. Unraveling the Roles of Canonical and Noncanonical PRC1 Complexes. In: *Polycomb Group Proteins*. Elsevier Inc.; 2017:57-80. doi:10.1016/B978-0-12-809737-3.00004-0

29. Chittock EC, Latwiel S, Miller TCR, Müller CW. Molecular architecture of polycomb repressive complexes. *Biochemical Society Transactions*. 2017;45(1):193-205. doi:10.1042/BST20160173

30. G King IF, Emmons RB, Francis NJ, et al. Analysis of a Polycomb Group Protein Defines Regions That Link Repressive Activity on Nucleosomal Templates to In Vivo Function. *MOLECULAR AND CELLULAR BIOLOGY*. 2005;25(15):6578-6591. doi:10.1128/MCB.25.15.6578-6591.2005

31. Taherbhoy AM, Huang OW, Cochran AG. BMI1–RING1B is an autoinhibited RING E3 ubiquitin ligase. *Nature Communications 2015 6:1*. 2015;6(1):1-13. doi:10.1038/ncomms8621

32. Gutiérrez L, Oktaba K, Scheuermann JC, Gambetta MC, Ly-Hartig N, Müller J. The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development*. 2012;139(1):117-127. doi:10.1242/dev.074450

33. Cheutin T, Cavalli G. Loss of PRC1 induces higher-order opening of Hox loci independently of transcription during Drosophila embryogenesis. *Nature Communications 2018 9:1*. 2018;9(1):1-11. doi:10.1038/s41467-018-05945-4

34. Wani AH, Boettiger AN, Schorderet P, et al. Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nature Communications 2016 7:1*. 2016;7(1):1-13. doi:10.1038/ncomms10291

35. Isono K, Endo TA, Ku M, et al. SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Developmental Cell*. 2013;26(6):565-577. doi:10.1016/J.DEVCEL.2013.08.016

36. Simon JA, Kingston RE. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell*. 2013;49(5):808-824. doi:10.1016/j.molcel.2013.02.013

37. Gahan JM, Rentzsch F, Schnitzler CE. The genetic basis for PRC1 complex diversity emerged early in animal evolution. *bioRxiv*. Published online March 20, 2020:2020.03.18.997064. doi:10.1101/2020.03.18.997064

38. Gao Z, Zhang J, Bonasio R, et al. PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Molecular Cell*. 2012;45(3):344-356. doi:10.1016/j.molcel.2012.01.002

39. Schuettengruber B, Bourbon HM, di Croce L, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*. 2017;171(1):34-57. doi:10.1016/j.cell.2017.08.002

40. Lagarou A, Mohd-Sarip A, Moshkin YM, et al. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes and Development*. 2008;22(20):2799-2810. doi:10.1101/gad.484208

41. Nekrasov M, Wild B, Müller J. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO Reports*. 2005;6(4):348-353. doi:10.1038/sj.embor.7400376

42. Pengelly AR, Copur Ö, Jäckle H, Herzig A, Müller J. A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. *Science*. 2013;339(6120):698-699. doi:10.1126/SCIENCE.1231382

43. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011;469(7330):343-349. doi:10.1038/nature09784

44. Frey F, Sheahan T, Finkl K, et al. Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. *Genes & Development*. 2016;30(9):1116. doi:10.1101/GAD.279141.116

45. JL B, JA K. Architectural and functional diversity of polycomb group response elements in Drosophila. *Genetics*. 2013;195(2):407-419. doi:10.1534/GENETICS.113.153247

46. Klymenko T, Papp B, Fischle W, et al. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes & Development*. 2006;20(9):1110. doi:10.1101/GAD.377406

47. di Croce L, Helin K. Transcriptional regulation by Polycomb group proteins. *Nature Structural and Molecular Biology*. 2013;20(10):1147-1155. doi:10.1038/nsmb.2669

48. Langlais KK, Brown JL, Kassis JA. Polycomb Group Proteins Bind an engrailed PRE in Both the "'ON'" and " 'OFF'" Transcriptional States of engrailed. doi:10.1371/journal.pone.0048765

49. Tamburri S, Lavarone E, Ferná Ndez-Pé Rez D, et al. Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. Published online 2020. doi:10.1016/j.molcel.2019.11.021

50. Golbabapour S, Majid NA, Hassandarvish P, Hajrezaie M, Abdulla MA, Hadi AHA. Gene Silencing and Polycomb Group Proteins: An Overview of their Structure, Mechanisms and Phylogenetics. *OMICS : a Journal of Integrative Biology*. 2013;17(6):283. doi:10.1089/OMI.2012.0105

51. van Kruijsbergen I, Hontelez S, Veenstra GJC. Recruiting polycomb to chromatin. *The International Journal of Biochemistry & Cell Biology*. 2015;67:177-187. doi:10.1016/J.BIOCEL.2015.05.006

52. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications – writers that read. *EMBO Reports*. 2015;16(11):1467. doi:10.15252/EMBR.201540945

53. Hathaway NA, Bell O, Hodges C, Miller EL, Neel DS, Crabtree GR. Dynamics and Memory of Heterochromatin in Living Cells. *Cell*. 2012;149(7):1447. doi:10.1016/J.CELL.2012.03.052

54. F T, R B, AR S, et al. Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development (Cambridge, England)*. 2014;141(5):1129-1139. doi:10.1242/DEV.102392

55. Pengelly AR, Kalb R, Finkl K, Müller J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes & Development*. 2015;29(14):1487-1492. doi:10.1101/GAD.265439.115

56. Cohen I, Bar C, Ezhkova E. Activity of PRC1 and histone H2AK119 monoubiquitination: Revising popular misconceptions. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2020;42(5):e1900192. doi:10.1002/BIES.201900192

57. Srinivasan A, Mishra RK. Genomic organization of Polycomb Response Elements and its functional implication in Drosophila and other insects. *Journal of Biosciences*. 2038;45. doi:10.1007/s12038-019-9975-2

58. Bredesen BA, Rehmsmeier M. DNA sequence models of genome-wide Drosophila melanogaster Polycomb binding sites improve generalization to independent Polycomb Response Elements. *Nucleic Acids Research*. 2019;47(15):7781-7797. doi:10.1093/NAR/GKZ617

59. Grossniklaus U, Paro R. Transcriptional Silencing by Polycomb-Group Proteins. *Cold Spring Harbor Perspectives in Biology*. 2014;6(11):1-26. doi:10.1101/CSHPERSPECT.A019331

60. Dorafshan E, Kahn TG, Schwartz YB. Hierarchical recruitment of Polycomb complexes revisited. *https://doi.org/101080/1949103420171363136*. 2017;8(5):496-505. doi:10.1080/19491034.2017.1363136

61. Tolhuis B, Blom M, Kerkhoven RM, et al. Interactions among Polycomb Domains Are Guided by Chromosome Architecture. *PLoS Genetics*. 2011;7(3):1001343. doi:10.1371/JOURNAL.PGEN.1001343

62. Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang JM, Cavalli G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Molecular Cell*. 2018;71(1):73-88.e5. doi:10.1016/J.MOLCEL.2018.05.032

63. Matheson L, Elderkin S. Polycomb Bodies. *Nuclear Architecture and Dynamics*. Published online January 1, 2018:297-320. doi:10.1016/B978-0-12-803480-4.00013-2

64. Šmigová J, Juda P, Krejčí J, Raška I, Šmigová J. Structural Basis of Polycomb Bodies (Polycomb group proteins / Polycomb body / post-translational chromatin modifications / nuclear subcompartments / nuclear factories / chromatin domains). *Folia Biologica (Praha)*. 2014;60:13-20.

65. Buchenau P, Hodgson J, Strutt H, Arndt-Jovin DJ. The Distribution of Polycomb-Group Proteins During Cell Division and Development in Drosophila Embryos: Impact on

Models for Silencing. *Journal of Cell Biology*. 1998;141(2):469-481. doi:10.1083/JCB.141.2.469

66. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nature Cell Biology 2007 9:10*. 2007;9(10):1167-1174. doi:10.1038/ncb1637

67. Kim CA, Gingery M, Pilpa RM, Bowie JU. The SAM domain of polyhomeotic forms a helical polymer. *Nature Structural Biology 2002 9:6*. 2002;9(6):453-457. doi:10.1038/nsb802

68. Comet I, Savitskaya E, Schuettengruber B, et al. PRE-Mediated Bypass of Two Su(Hw) Insulators Targets PcG Proteins to a Downstream Promoter. *Developmental Cell*. 2006;11(1):117-124. doi:10.1016/J.DEVCEL.2006.05.009

69. Beisswanger S, Stephan W. Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in Drosophila. *Proceedings of the National Academy of Sciences*. 2008;105(14):5447-5452. doi:10.1073/PNAS.0710892105

70. Seif E, Kang JJ, Sasseville C, et al. Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nature Communications 2020 11:1*. 2020;11(1):1-19. doi:10.1038/s41467-020-19435-z

71. Nanyes DR, Junco SE, Taylor AB, et al. Multiple polymer architectures of human Polyhomeotic homolog 3 (PHC3) SAM. *Proteins*. 2014;82(10):2823. doi:10.1002/PROT.24645

72. Robinson AK, Leal BZ, Chadwell L v., et al. The Growth-Suppressive Function of the Polycomb Group Protein Polyhomeotic Is Mediated by Polymerization of Its Sterile Alpha Motif (SAM) Domain. *The Journal of Biological Chemistry*. 2012;287(12):8702. doi:10.1074/JBC.M111.336115

73. Isono K, Fujimura Y, Shinga J, et al. Mammalian Polyhomeotic Homologues Phc2 and Phc1 Act in Synergy To Mediate Polycomb Repression of Hox Genes. *Molecular and Cellular Biology*. 2005;25(15):6694. doi:10.1128/MCB.25.15.6694-6706.2005

74. Awad S, Al-Dosari MS, Al-Yacoub N, et al. Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. *Human Molecular Genetics*. 2013;22(11):2200-2213. doi:10.1093/HMG/DDT072

75. Deshpande AM, Akunowicz JD, Reveles XT, et al. PHC3, a component of the hPRC-H complex, associates with 2A7E during G 0 and is lost in osteosarcoma tumors. *Oncogene 2007 26:12*. 2006;26(12):1714-1722. doi:10.1038/sj.onc.1209988

76. CA K, NJ F. Chromatin: Polycomb Group SAMs Unite. *Current biology : CB*. 2016;26(15):R710-R712. doi:10.1016/J.CUB.2016.06.001

77. Gambetta MC, Müller J. O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. *Developmental Cell*. 2014;31(5):629-639. doi:10.1016/J.DEVCEL.2014.10.020

78. Sanulli S, Trnka MJ, Dharmarajan V, et al. HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature 2019 575:7782*. 2019;575(7782):390-394. doi:10.1038/s41586-019-1669-2

79. Harrison RES, Weng K, Wang Y, Peng Q. Phase separation and histone epigenetics in genome regulation. *Current Opinion in Solid State and Materials Science*. 2021;25(1):100892. doi:10.1016/J.COSSMS.2020.100892

80. Shin Y, Chang Y-C, Lee DSW, et al. Liquid nuclear condensates mechanically sense and restructure the genome. *Cell*. 2018;175(6):1481. doi:10.1016/J.CELL.2018.10.057

81. Söding J, Zwicker D, Sohrabi-Jahromi S, Boehning M, Kirschbaum J. Mechanisms of active regulation of biomolecular condensates. *bioRxiv*. Published online July 6, 2019:694406. doi:10.1101/694406

82. Kumar Rai A, Chen J-X, Selbach M, Pelkmans L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature*. Published online 2018. doi:10.1038/s41586-018-0279-8

83. Festuccia N, Owens N, Papadopoulou T, et al. Transcription factor activity and nucleosome organisation in mitosis. *bioRxiv*. 2018;29(2):392241. doi:10.1101/392241

84. Steffen PA, Ringrose L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nature Reviews Molecular Cell Biology 2014 15:5*. 2014;15(5):340-356. doi:10.1038/nrm3789

85. L R, R P. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual review of genetics*. 2004;38:413-443. doi:10.1146/ANNUREV.GENET.38.072902.091907

86. Fonseca JP, Steffen PA, Müller S, et al. In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. *Genes & Development*. 2012;26(8):857. doi:10.1101/GAD.184648.111

87. G D, RW C, H D, Y R, DP B-J. The number of PML nuclear bodies increases in early S phase by a fission mechanism. *Journal of cell science*. 2006;119(Pt 6):1026-1033. doi:10.1242/JCS.02816

88. Zhang Y, Bertulat B, Tencer AH, et al. MORC3 Forms Nuclear Condensates through Phase Separation. *iScience*. 2019;17:182-189. doi:10.1016/J.ISCI.2019.06.030

89. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*. 2019;176(3):419-434. doi:10.1016/J.CELL.2018.12.035

90. E G, J S. The molecular language of membraneless organelles. *The Journal of biological chemistry*. 2019;294(18):7115-7127. doi:10.1074/JBC.TM118.001192

91. CP B, CR E, DS C, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, NY)*. 2009;324(5935):1729-1732. doi:10.1126/SCIENCE.1172046

92. O'Flynn BG, Mittag T. The role of liquid–liquid phase separation in regulating enzyme activity. *Current Opinion in Cell Biology*. 2021;69:70-79. doi:10.1016/J.CEB.2020.12.012

93. Klosin A, Oltsch F, Harmon T, et al. PHASE SEPARATION Phase separation provides a mechanism to reduce noise in cells. Accessed August 1, 2021. http://science.sciencemag.org/

94. P S, SC W. Dissecting the complexity of biomolecular condensates. *Biochemical Society transactions*. 2020;48(6):2591-2602. doi:10.1042/BST20200351

95. Zumbro E, Alexander-Katz A. Multivalent polymers can control phase boundary, dynamics, and organization of liquid-liquid phase separation. *bioRxiv*. Published online December 31, 2020:2020.12.31.424934. doi:10.1101/2020.12.31.424934

96. Banani SF, Rice AM, Peeples WB, et al. Compositional Control of Phase-Separated Cellular Bodies. *Cell*. 2016;166(3):651-663. doi:10.1016/J.CELL.2016.06.010

97. Boeynaems S, Alberti S, Fawzi NL, et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends in cell biology*. 2018;28(6):420. doi:10.1016/J.TCB.2018.02.004

98. Bienz M. Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. *Cell*. 2020;182(4):799-811. doi:10.1016/J.CELL.2020.07.037

99. HR L, WC C, PC C, WJ W, JR H. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *The Journal of biological chemistry*. 2018;293(16):6090-6098. doi:10.1074/JBC.AC117.001037

100. Voehler M, Ashoka MA, Meiler J, Bock PE. Carbon and amide detect backbone assignment methods of a novel repeat protein from the staphylocoagulase in S. aureus. *Biomolecular NMR assignments*. 2017;11(2):243. doi:10.1007/S12104-017-9757-4

101. Coletta A, Pinney JW, Solís DYW, Marsh J, Pettifer SR, Attwood TK. Low-complexity regions within protein sequences have position-dependent roles. *BMC Systems Biology*. 2010;4:43. doi:10.1186/1752-0509-4-43

102. Tatham AS, Shewry PR. Elastomeric proteins: biological roles, structures and mechanisms. *Trends in Biochemical Sciences*. 2000;25(11):567-571. doi:10.1016/S0968-0004(00)01670-4

103. Müller-Späth S, Soranno A, Hirschfeld V, et al. From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National*

Academy of Sciences of the United States of America. 2010;107(33):14609. doi:10.1073/PNAS.1001743107

104. Yang J, Zeng Y, Liu Y, et al. Electrostatic interactions in molecular recognition of intrinsically disordered proteins. *https://doi.org/101080/0739110220191692073*. 2019;38(16):4883-4894. doi:10.1080/07391102.2019.1692073

105. Dyson HJ, Wright PE, Scheraga HA. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(35):13057. doi:10.1073/PNAS.0605504103

106. JA R, CD K, JL K-S, et al. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell*. 2017;168(6):1028-1040.e19. doi:10.1016/J.CELL.2017.02.027

107. Wang J, Choi JM, Holehouse AS, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*. 2018;174(3):688-699.e16. doi:10.1016/J.CELL.2018.06.006

108. Salvi N. Intrinsically disordered proteins : dynamics, binding, and function.

109. A P, HO L, L J, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. 2015;162(5):1066-1077. doi:10.1016/J.CELL.2015.07.047

110. Larson AG, Elnatan D, Keenen MM, et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature Publishing Group*. 2017;547. doi:10.1038/nature22822

111. Tatarakis A, Behrouzi R, Moazed D. Molecular Cell Previews Evolving Models of Heterochromatin: From Foci to Liquid Droplets. Published online 2017. doi:10.1016/j.molcel.2017.08.022

112. Mcswiggen DT, Mir M, Darzacq X, Tjian R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. Published online 2019. doi:10.1101/gad.331520.119

113. Lyon AS, Peeples WB, Rosen MK. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology 2020 22:3*. 2020;22(3):215-235. doi:10.1038/s41580-020-00303-z

114. Peeples W, Rosen MK. Phase Separation Can Increase Enzyme Activity by Concentration and Molecular Organization. doi:10.1101/2020.09.15.299115

115. Tatavosian R, Kent S, Brown K, et al. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *Journal of Biological Chemistry*. 2019;294(5):1451-1463. doi:10.1074/JBC.RA118.006620

116. Plys AJ, Davis CP, Kim J, et al. Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes & Development*. 2019;33(13-14):799. doi:10.1101/GAD.326488.119

117. Kent S, Brown K, Yang C hsun, et al. Phase-Separated Transcriptional Condensates Accelerate Target-Search Process Revealed by Live-Cell Single-Molecule Imaging. *Cell Reports*. 2020;33(2):108248. doi:10.1016/J.CELREP.2020.108248

118. Qu W, Wang Z, Zhang H. Phase separation of the C . elegans Polycomb protein SOP-2 is modulated by RNA and sumoylation. *Protein & Cell 2019 11:3*. 2020;11(3):202-207. doi:10.1007/S13238-019-00680-Y

119. Isono K, Endo TA, Ku M, et al. SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Developmental Cell*. 2013;26(6):565-577. doi:10.1016/J.DEVCEL.2013.08.016

120. PV L, F S, CF L. Distinct modes of centromere protein dynamics during cell cycle progression in Drosophila S2R+ cells. *Journal of cell science*. 2013;126(Pt 20):4782-4793. doi:10.1242/JCS.134122

121. Mészáros B, Erdös G, Dosztányi Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*. 2018;46(W1):W329-W337. doi:10.1093/nar/gky384

122. Jarnot P, Ziemska-Legiecka J, Dobson L, et al. PlaToLoCo: The first web meta-server for visualization and annotation of low complexity regions in proteins. *Nucleic Acids Research*. 2020;48(W1):W77-W84. doi:10.1093/NAR/GKAA339

123. Promponas VJ, Enright AJ, Tsoka S, et al. *CAST: An Iterative Algorithm for the Complexity Analysis of Sequence Tracts*. Vol 16.; 2000. Accessed June 30, 2021. www.ebi.ac.uk/

124. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics*. 2014;30(17):2501-2502. doi:10.1093/bioinformatics/btu310

125. Follmer NE, Wani AH, Francis NJ. A Polycomb Group Protein Is Retained at Specific Sites on Chromatin in Mitosis. *PLOS Genetics*. 2012;8(12):e1003135. doi:10.1371/JOURNAL.PGEN.1003135

126. Maharana S, Wang J, Papadopoulos DK, et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*. 2018;360(6391):918-921. doi:10.1126/SCIENCE.AAR7366

127. Navarro MG-J, Kashida S, Chouaib R, et al. RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nature Communications 2019 10:1*. 2019;10(1):1-13. doi:10.1038/s41467-019-11241-6

128. CJ O, Z P, L K. Disordered RNA-Binding Region Prediction with DisoRDPbind. *Methods in molecular biology (Clifton, NJ)*. 2020;2106:225-239. doi:10.1007/978-1-0716-0231-7_14

129. Seim I, Posey AE, Snead WT, et al. Negative regulation of a ribonucleoprotein condensate driven by dilute phase oligomerization. doi:10.1101/2021.04.19.440511

130. Nosella ML, Tereshchenko M, Pritišanac I, et al. O-GlcNAcylation reduces phase separation and aggregation of the EWS N-terminal low complexity region. doi:10.1101/2021.05.11.443654

131. Carpenter AE, Jones TR, Lamprecht MR, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology 2006 7:10*. 2006;7(10):1-11. doi:10.1186/GB-2006-7-10-R100

132. Sö J, Zwicker D, Sohrabi-Jahromi S, Boehning M, Kirschbaum J. Mechanisms for Active Regulation of Biomolecular Condensates Trends in Cell Biology. *Trends in Cell Biology*. 2020;30(1):4-14. doi:10.1016/j.tcb.2019.10.006

133. Cho W-K, Spille J-H, Hecht M, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*. 2018;361(6400):412-415. doi:10.1126/SCIENCE.AAR4199

134. Qi Y, Zhang B. Chromatin Network Retards Droplet Coalescence. doi:10.1101/2021.03.02.433564

135. Strom AR, Emelyanov A v., Mir M, Fyodorov D v., Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. *Nature*. 2017;547(7662):241. doi:10.1038/NATURE22989

136. Gibson BA, Doolittle LK, Schneider MWG, et al. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*. 2019;179(2):470-484.e21. doi:10.1016/J.CELL.2019.08.037

137. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. doi:10.1073/pnas.1017150108

138. Spector DL, Lamond AI. Nuclear Speckles. *Cold Spring Harbor Perspectives in Biology*. 2011;3(2):1-12. doi:10.1101/CSHPERSPECT.A000646

139. Martin EW, Harmon TS, Hopkins JB, et al. A multi-step nucleation process determines the kinetics of prion-like domain phase separation. *Nature Communications 2021 12:1*. 2021;12(1):1-12. doi:10.1038/s41467-021-24727-z

140. Kim J, Han KY, Khanna N, Ha T, Belmont AS. Nuclear speckle fusion via long-range directional motion regulates speckle morphology after transcriptional inhibition. *Journal of Cell Science*. 2019;132(8). doi:10.1242/JCS.226563

141. Fardini Y, Dehennaut V, Lefebvre T, Issad T. O-GlcNAcylation: A new cancer hallmark? *Frontiers in Endocrinology*. 2013;4(AUG). doi:10.3389/FENDO.2013.00099

142. Hur W, Kemp JP, Tarzia M, et al. CDK-Regulated Phase Separation Seeded by Histone Genes Ensures Precise Growth and Function of Histone Locus Bodies. *Developmental Cell*. 2020;54(3):379-394.e6. doi:10.1016/J.DEVCEL.2020.06.003

143. Zhang Y, Bertulat B, Tencer AH, et al. MORC3 Forms Nuclear Condensates through Phase Separation. *iScience*. 2019;17:182-189. doi:10.1016/J.ISCI.2019.06.030