KMT5C encodes robust heterochromatin retention and liquid-like behavior using limited sequence features

Hilmar Strickfaden,[1] Kristal Missiaen,[1] Justin W. Knechtel,[1] Michael J. Hendzel,[1,2] and D. Alan Underhill[1,3,*]

Departments of [1]Oncology, [2]Cell Biology, and [3]Medical Genetics, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Alberta, Canada

SUMMARY

Cells use multiple strategies to compartmentalize functions through a combination of membrane-bound and membraneless organelles. The latter represent complex assemblies of biomolecules that coalesce into a dense phase through low affinity, multivalent interactions and undergo rapid exchange with the surrounding dilute phase. We describe a liquid-like state for the lysine methyltransferase KMT5C characterized by diffusion within heterochromatin condensates but lacking appreciable nucleoplasmic exchange. Retention was strongly correlated with reduction of condensate surface area, suggesting formation of a liquid droplet with high connectivity. This behavior mapped to a discrete domain whose activity was dependent on multiple short linear motifs. Moreover, it was strikingly resilient to marked phylogenetic differences or targeted changes in intrinsic disorder, charge, sequence, and architecture. Collectively, these findings show that a limited number of sequence features can dominantly encode multivalency, localization, and dynamic behavior within heterochromatin condensates to confer protein retention without progression to a gel or solid.

INTRODUCTION

Cells contain membraneless organelles of varying complexity where constituents partition between a dense liquid demixed phase and a dilute surrounding phase (*1, 2*). Formation of these biological condensates reflects low affinity, multivalent interactions that allow proteins to coalesce into distinct aqueous compartments (*2*). Despite the stable appearance of such assemblies, their constituents rapidly exchange with their surroundings (*3, 4*). Nevertheless, they can adopt a range of dynamic and material states, including gels or solids (*5-7*). This can be exemplified by the nucleolus, which is made up of three biophysically distinct phase separated compartments (*8*), each with specialized functions (*9-11*). At present, however, we have an incomplete understanding of the range of dynamic behaviors exhibited in the context of biomolecular condensates and how this relates to their physiological roles.

Condensate formation can be conceptualized by categorizing components as either scaffolds or clients (*12-14*). Scaffold molecules are essential to initiating phase separation and provide a platform for recruitment of client proteins. The valency or number of interaction sites on scaffold and client molecules defines network connectivity and the material properties of the assembly (*15, 16*). In this regard, proteins with higher valency will phase separate at lower concentration and form condensates with greater stability (*15, 16*). Although the sequence features that sustain these interactions have been defined in many cases, they display considerable diversity (*2, 17-19*). These include complex interaction types mediated by single amino acids, short linear motifs, regions of intrinsic disorder, structured domains, and nucleic acids (*17, 20, 21*). Given the repertoire of proteins that partition to condensates is not complete, there are opportunities to discover new scaffold or client proteins with unique sequence and biophysical properties, as well as those that use potentially novel strategies.

There are multiple biological condensates described within the nucleus (*22, 23*) and phase separation has emerged as a critical organizing principle for chromatin (*24-27*). Notably, analyses of the prototypic heterochromatin protein CBX5 (*aka* HP1α) indicated that phase separation underlies partitioning of pericentromeric heterochromatin into a separate compartment (*28, 29*). This is particularly evident in mouse interphase nuclei where pericentromeres from multiple chromosomes congregate into larger chromocenters (*30, 31*). CBX5 enrichment within pericentromeric heterochromatin requires binding to trimethyllysine-9 on histone H3 (H3K9me3), which is placed by the lysine methyltransferases SUV39H1 and 2 (*32*). CBX5 then recruits the lysine methyltransferase KMT5C to catalyze histone H4 lysine-20 trimethylation (H4K20me3) (*33, 34*). While CBX5 exchanges rapidly between the nucleoplasm and pericentromeric heterochromatin when assessed by fluorescence recovery after photobleaching (FRAP), previous studies of KMT5C have consistently suggested that it forms an immobile scaffold within chromocenters (*35-37*). This conclusion was supported from analyses in mouse embryonic stem cells (*35*) and L929 (*37*), NIH 3T3 (*36*), and mouse embryonic fibroblasts (*35, 36*), supporting its generality. Moreover, it fit with the fact that H4K20me3 catalysis by KMT5C is under strict spatial (*33, 38, 39*) and quantitative (*34, 40, 41*) control in a wide range of cell line and tissue contexts.

In the present study, we assessed inter and intra-chromocenter mobility for KMT5C in comparison to CBX5 (*28, 29*) and MeCP2 (*42-44*), which provided known benchmarks for heterochromatin phase separation. Surprisingly, we found that KMT5C sampled the entire chromocenter volume on the time scale of seconds but underwent only minimal exchange with the surrounding nucleoplasm over the same period. In contrast, both CBX5 and MeCP2 exchanged rapidly, as previously reported (*36, 45-47*). The unique dynamic behavior of KMT5C

was associated with robust chromocenter partitioning and reduction of surface area. Together, this suggests the formation of a liquid-like condensate that effectively sequestered KMT5C within chromocenters without imposing the same constraint on CBX5 or MeCP2. This activity was consolidated within a discrete KMT5C protein segment that displayed considerable sequence divergence amongst tetrapods. Nevertheless, its localization and dynamic behavior were largely conserved and conferred by a unique combination of short linear motifs that functioned in a variety of sequence contexts. Overall, our work provides a model to understand how sparse sequence features can confer efficient protein retention and liquid-like behavior within a biological condensate.

RESULTS

KMT5C defines a novel dynamic state in phase-separated constitutive heterochromatin

In fluorescence recovery after photobleaching (FRAP), CBX5 rapidly exchanged between chromocenters and the nucleoplasm (*36, 37, 46, 47*), whereas SUV39H2 and KMT5C appeared to be immobile when entire chromocenters were bleached (*35-37*). Although concluding the latter two proteins created a stable scaffold, these studies did not assess mobility within chromocenters, which is required to determine if they adopt a liquid or solid-like state. For our analyses, we used the immortalized mouse mammary epithelial cell line NMuMG (*48*), which showed prominent H4K20me3 staining that was largely confined to chromocenters (fig. S1A), indicating the underlying KMT5C methylation pathway was functional. Importantly, this limited H4K20me3 distribution was also observed in epithelial cells of the mouse pubertal mammary gland (fig. S1B), establishing NMuMG cells as a relevant model system to understand how KMT5C dynamics relate to the distribution of its enzymatic product.

To evaluate the possibility of KMT5C mobility within constitutive heterochromatin, we queried KMT5C dynamics in mouse chromocenters in comparison to CBX5 and MeCP2, which exhibit pericentromeric enrichment, nucleoplasmic exchange (*36, 45*), and undergo liquid-liquid phase separation (*28, 29, 42-44*). As previously shown with full chromocenter bleaching (*35-37*), KMT5C displayed long recovery times, indicating minimal exchange with the surrounding nucleoplasm (Fig. 1A, *Total*). Upon partial bleaching, however, fluorescence recovered (movie S1) and progressed from the non-bleached portion of the chromocenter (Fig. 1A, *Partial*). This occurred on a timescale where there was no appreciable recovery of fully bleached chromocenters in the same nucleus (Fig. 1A, *Total*), establishing that KMT5C moved readily within chromocenters but did not efficiently exchange. In other words, KMT5C equilibrated

within individual chromocenters with minimal influx from the surrounding nucleoplasm. Both MeCP2 and CBX5, however, showed near complete recovery in fully bleached chromocenters over the same time, which was readily apparent in both imaging (Fig. 1B) and FRAP analysis (Fig. 1C and movies S2 and S3). KMT5C therefore appeared to be effectively confined to individual chromocenters when compared to MeCP2 or CBX5 (Fig. 1C, schematics) and displayed significantly higher partitioning to chromocenters than either protein (Fig. 1D). Thus, KMT5C, MeCP2, and CBX5 define distinct dynamics states in constitutive heterochromatin defined by progressively increasing intra-chromocenter mobility and nucleoplasmic exchange that is inversely correlated to partitioning (Fig. 1E).

The dynamic behavior of KMT5C was reinforced using two additional imaging approaches. First, photobleaching half of the nucleus established the pan-nuclear nature of KMT5C dynamics when compared to MeCP2 and CBX5 (fig. S2A). As observed for full chromocenter photobleaching, we observed only limited recovery of KMT5C fluorescence, intermediate recovery for MeCP2, and complete recovery for CBX5. Second, the retentive nature of KMT5C localization was particularly evident in inverse FRAP experiments where a single chromocenter was spared from photobleaching and its fluorescence intensity was measured over 30 minutes (fig. S2B). In the absence of fluorescent protein influx, this revealed prolonged maintenance of KMT5C-mEmerald signal intensity and an apparent efflux rate of 0.7%min^{-1} (fig. S2C). By extrapolation, full chromocenter turnover of KMT5C would require ~ 2.4 hours, suggesting a slow rate of exchange that would not be apparent over the shorter 100 second time course used in FRAP analyses (fig. S2C, inset). In the context of phase separation, the behavior of KMT5C signified a remarkable barrier to exit that has not been observed for other proteins without the transition to a solid or gel state, which would render them immobile (*49, 50*).

KMT5C exhibits liquid-like properties in constitutive heterochromatin condensates

The dynamic behavior of KMT5C suggested retention within a liquid droplet. Spot bleaching of a defined volume within individual chromocenters (Fig. 2A) provided clear evidence of internal mobility with a $t_{1/2}$ of ~4.4s and an apparent diffusion (D_{app}) constant of 1.9 x 10^{-2} $\mu m^2 s^{-1}$ (Fig. 2B) (*51*). Internal mixing was also evident in convergence curves (*52*) derived after partial photobleaching (fig. S3A) where KMT5C exhibited fluorescence loss from the unbleached portion and signal gain in the bleached portion of the same chromocenter (fig. S3B). This requires liquid-like equilibration within individual chromocenters given there is minimal influx from the surrounding nucleoplasm over the sampled time. Liquid-like behavior was also apparent upon induction of DNA damage within chromocenters by laser micro-irradiation (Fig. 2C and fig. S3C). This induces rapid chromatin decondensation and formation of a poly(ADP-ribose)-mediated phase separated liquid compartment encompassing damaged DNA (*53*). Bisection of chromocenters with laser micro-irradiation led to rapid KMT5C retraction from the damaged area to create two lobes that preserved the retentive behavior of the original domain by FRAP analysis (Fig. 2C, fig. S3C, and movies S4 and S5), indicating a lack of exchange between lobes. The rounded shape of these compartments provided further evidence of liquidity where KMT5C appeared to drive minimization of surface area (*1*). This was not apparent with either CBX5 or MeCP2, which were characterized by distinct responses. Although CBX5 initially showed depletion from the damage zone, this was transient and accompanied by redistribution throughout the damaged and non-damaged compartments, where it continued to show rapid recovery after photobleaching (Fig. 2C, fig. S3C, and movies S6 and S7). The localization and mobility of MeCP2, however, appeared unperturbed by laser micro-irradiation (Fig. 2C, fig.

S3C, and movies S8 and S9). Lastly, KMT5C was associated with a significantly higher degree of sphericity compared to either MeCP2 or CBX5 in transfected cells (Fig. 2D, $p \leq 0.001$), indicating it was more effective at reducing surface area at the boundary between the chromocenter and nucleoplasm. KMT5C therefore demonstrates that phase separation can achieve robust protein sequestration while nevertheless retaining liquid-like internal mobility in the context of heterochromatin-associated condensates.

A rudimentary heterochromatin targeting domain underlies CRD activity

At the primary sequence level, KMT5C comprises an enzymatic SET domain that catalyzes H4K20me3 production (*33, 34*), an RNA-binding domain (*54*), and a region that confers localization to constitutive heterochromatin (*35, 37*) (SET, RBD, and CRD in Fig. 3A). For chromocenter localization, previous FRAP analyses of segments from either human (347-435) (*37*) or mouse (350-412) (*35*) KMT5C suggested they were immobile based on full chromocenter photobleaching. Using phylogenetic comparisons in mammals, we refined the boundaries of this region to amino acids 354-409 in human KMT5C (356-411 in mouse). This segment recapitulated the behavior of full-length KMT5C regarding internal chromocenter mobility, lack of nucleoplasmic exchange (Fig. 3B and movie S10), and partitioning (Fig. 3C). Inclusion of the adjacent RBD, which confers nucleolar localization during quiescence (*54*), did not alter localization or dynamics (not shown). It therefore appears that the interactions underlying the dynamic behavior of full-length KMT5C are largely contained within this discrete region, which we will refer to as the Chromocenter Retention Domain (CRD) to reflect its activity.

We sought to understand the overall architecture and driver(s) of CRD behavior through examination of sequence features. Nevertheless, sequence identity amongst mammalian CRDs was limited to 19 of 57 residues and there was considerable variation within individual orders—for example, 23 positions were variable within Primates. Moreover, manual curation was required to identify non-mammalian CRD counterparts, which were limited to bird and reptile KMT5C orthologs. Notwithstanding marked divergence, we identified three conserved features: a His-x_4-Cys-x_{3-4}-Cys motif and a L-x_{5-6}-V-x_1-L hydrophobic motif that were always separated by a single amino acid (Fig. 4A). These were followed by a related L-x_5-V-x_1-L motif that was embedded within distinct blocks of homology in each taxonomic class and separated by segments of highly variable length and sequence composition (Fig. 4A and table S1). Homology within the CRD therefore occurred as two discrete segments, designated as conserved domains 1 and 2 (CD1 and CD2), separated by a non-conserved linker region, suggesting a bipartite structure. This fit with previous analyses of the mouse KMT5C region spanning residues 350-412 where each half could independently confer chromocenter localization, albeit with lower efficiency and elevated mobility in FRAP experiments (*35*). We confirmed this finding with full and partial chromocenter bleaching of CD1 and CD2-mEmerald fusion proteins, which identified CD1 as the main driver of CRD behavior (fig. S4A and S4B).

To query if differences in sequence, physicochemical properties, or architecture affect the localization and dynamic behavior of the CRD, we assessed versions from *Homo sapiens* (*Hs*), *Mus musculus* (*Mm*), *Cavia porcellus* (*Cp*), *Bubalus bubalis* (*Bb*), *Gallus gallus* (*Gg*), and *Podarcis muralis* (*Pm*). These versions varied in linker length (6-36aa), isoelectric point (7.74-11.83), fraction of charged residues (FCR) (0.13636-0.21053) (Fig. 4B), and multiple other parameters (table S1). Remarkably, each CRD showed efficient chromocenter localization in

mouse cells (fig. S4C), indicating they recognize conserved pericentromeric heterochromatin features despite their limited homology. Moreover, each CRD displayed moderate intra-chromocenter mobility and nucleoplasmic exchange in FRAP experiments, although these measures were elevated for the *Bb*, *Cp*, *Gg*, and *Pm* versions (Fig. 4B). The fact that these behaviors were shared by all orthologs indicates they are driven by a limited number of sequence features that are modulated by species-specific differences.

CRD localization and dynamics are dominantly encoded by Short Linear Motifs (SLiMs)

We examined the role of 4 highly conserved motifs in the CRD, focusing on how $C^{362}C^{366}$, $H^{357}H^{365}$, $L^{368}V^{374}L^{376}$, and $L^{398}V^{404}L^{406}$ (Fig. 4A) influence CRD localization and dynamics. Mutation of individual motifs (Fig. 5A) did not abrogate chromocenter localization but caused varying degrees of nucleoplasmic distribution (Fig. 5B). All mutants were characterized by elevated mobility in both full and partial bleaching conditions (Fig. 5C and movies S11-S15). This was most pronounced for $C^{362}C^{366}$, $H^{357}H^{365}$, and $L^{368}V^{374}L^{376}$ where the partial and full recovery curves were largely superimposable (Fig. 5C), indicating differences between internal and external mobility were abrogated (*49*). In contrast, $L^{398}V^{404}L^{406}$ had a modest effect that was characterized by similar increases in internal mobility and exchange. The nature of differences between $L^{368}V^{374}L^{376}$ and $L^{398}V^{404}L^{406}$ further support CD1 making a greater contribution to CRD localization and dynamics (fig. S4A and S4B). Lastly, we assessed the mammalian-specific motif, $W^{359}W^{390}W^{392}Y^{396}$, because it resembled the aromatic cage motifs found in chromatin reader modules that bind histone methyl-lysine residues (*55-57*). This mutant also ameliorated the difference between partial and full chromocenter fluorescence recovery but had an intermediate effect when compared to $C^{362}C^{366}$, $H^{357}H^{365}$, and $L^{368}V^{374}L^{376}$ (Fig. 5C). Each motif

therefore constituted a distinct contribution to CRD localization and dynamics, and no single motif was absolutely required for these activities. Together, the results indicate that multiple Short Linear Motifs (SLiMs) act in a highly synergistic manner to confer KMT5C chromocenter retention.

Intrinsically disordered regions (IDRs) play an important role in protein liquid-liquid phase separation (*58-60*) but were highly variable for the CRD between and within taxonomic classes (Fig. 6A and table S2). Specifically, while each CRD showed a low propensity for disorder in CD1 that was minimal over the H-C-C segment (Fig. 4A), there was considerable variability in disorder profiles for the linker and CD2 regions (Fig. 6A, top). In fact, the only taxonomic class that displayed a conserved disorder profile was birds, where CD1 and CD2 were separated by a highly disordered linker region regardless of length (Fig. 6A, bottom, and table S2). To assess the potential contributions of intrinsic disorder to the dynamic properties in a controlled manner, we used an IDR series that is known to modulate phase separation behavior when linking globular domains (*61*). These 40 amino acid segments were used to replace the 6 amino acid linker in the mouse CRD to create chimeras with an extended region of disorder (Fig. 6B and table S2) that varied 8-fold in the fraction of charged residues, while net charge per residue varied from -0.213 to 0.085 (Fig. 6C and table S1). Although each CRD-IDR chimera conferred efficient chromocenter localization (fig. S5), all five showed a higher level of fluorescence recovery than the wild-type CRD upon partial bleaching (Fig. 6D). At the same time, the CRD-IDR1 to IDR4 proteins produced recovery curves that were superimposable with the wild-type CRD upon full chromocenter bleaching (Fig. 6D). Only CRD-IDR1, which had the highest charge fraction and net charge per residue (Fig. 6C), showed a modest increase in exchange (Fig. 6D). Importantly, these CRD-IDR chimeras establish that exchange and internal

mobility are separable. Moreover, together with the mutational analyses (Fig. 5), they indicate that SLiMs dominantly encode the retentive behavior of the CRD, while differences in disorder between CRDs may tune material properties within chromocenter condensates.

DISCUSSION

Several studies have assessed KMT5C/SUV420H2 mobility in the context of mouse chromocenters, with all concluding it was immobile (*35-37*). Importantly, this conclusion was also reached using a knockin allele that expressed a KMT5C-GFP fusion protein at endogenous levels (*35*), indicating it was not a consequence of exogenous expression. Revisiting dynamics in the context of heterochromatin condensate formation (*24, 62*), we found that KMT5C and the CRD displayed liquid-like behavior within chromocenters, but underwent limited exchange with the surrounding nucleoplasm (Figs. 1-3). This suggests that the dynamic behavior of KMT5C and the CRD is shaped by strong chromocenter avidity, which could reflect differences in solvation properties between the dilute and dense phases (*63, 64*) that drive robust protein compartmentalization without gelation. This behavior was fundamentally different from that of MeCP2 and CBX5 (Fig. 1), indicating phase separated heterochromatin domains can accommodate a continuum of dynamics states.

Liquid-liquid phase separation has been increasingly evoked to account for the numerous molecular assemblies within the nucleus, involving a diverse range of size scales and regulatory processes (*22, 65, 66*). Whether all are bona fide or can be described by other biophysical processes remains to be determined (*4, 17, 65, 67*). This applies to the heterochromatin assemblies under study here, which have been described as liquid-liquid phase separated (*25, 28, 29, 62*) and alternatively as "collapsed polymer globules" (*24, 52, 68*). In the latter case, the dynamic behavior of proteins would be dictated by percolation through the chromatin polymer (*52*). The preferential internal mobility of KMT5C within chromocenters and reduction of surface area (Figs. 1 and 2), however, are consistent with liquid-liquid phase separation (*52*). Another key consideration in interpreting these dynamic behaviors is the material state of the

underlying chromatin scaffold in terms of being liquid-like or solid-like (*26, 27*). Notably, KMT5C exhibited internal remixing within chromocenter condensates even when the underlying chromatin did not (*26*)—this solid-like chromatin behavior was also seen in studies of CBX5 and MeCP2 chromocenter mobility (*52*). As a result, KMT5C movement within chromocenters and the absence of nucleoplasmic exchange do not reflect stable binding (*35-37*) to mobile, liquid-like chromatin (*27*). Lastly, the positional stability conferred by this chromatin material state likely limits the opportunity for chromocenters to undergo fusion and/or fission, which is observed for condensates that form by liquid-liquid phase separation (*17, 69, 70*). Nevertheless, insomuch as laser micro-irradiation induced a phenomenon akin to fission, KMT5C exhibited the expected behavior for a liquid-like droplet (Fig. 2).

Proteins that undergo phase separation can be classified as either scaffolds that initiate condensate formation or clients that modulate condensate properties through low-affinity interactions (*12, 15, 69*). CBX5 (*71*) and MeCP2 (*72*) are abundant proteins important for heterochromatin architecture that can function as multivalent scaffolds, whereas the much lower endogenous level of KMT5C (https://www.ebi.ac.uk/gxa/home) (*73*) suggests it acts as a client. For a given scaffold, condensate stability can be increased by client proteins that have higher valency or those that add to connectivity without disrupting scaffold-scaffold interactions, both of which support elevated partitioning of the client protein to the condensate (*15*). In contrast, low valency clients reduce condensate stability (*15*) by using capping mechanisms that attenuate or block network connections and thereby oppose phase separation (*74*). Of these two scenarios, the robust partitioning of KMT5C and the CRD are consistent with creating a high level of connectivity. Likewise, that they encompassed dynamic behaviors resembling viscous liquids with respect to internal mobility, or viscoelastic materials with respect to exchange between

phases in FRAP analyses (*50*), requires interactions of sufficient number and strength to modulate emergent properties (*12, 15*). The fact that these features were apparent upon exogenous over-expression indicates connectivity persists over a wide concentration range (*15*) and that the underlying scaffold is not limiting or that KMT5C can undergo client to scaffold switching.

The CRD provides a minimalistic model to understand sequence features that support condensate retention within an otherwise liquid-like state. Whereas CBX5 (*28, 29*) and MeCP2 (*42, 43, 75*) use a combination of structured domains (stickers) and intrinsically disorder regions (spacers) that contribute to specific and non-specific interactions (*2, 61*), the dynamic and partitioning behaviors of KMT5C have been consolidated into the CRD. In comparison to CBX5 and MeCP2, which are highly conserved, it was also clear that the CRD was under very different evolutionary constraints. Comparing this domain across tetrapods revealed limited conservation that was only apparent after manual alignment, but nevertheless conferred efficient chromocenter localization (fig. S4C). It was not until the introduction of discrete mutations in a series of conserved motifs that the CRD showed a gradual shift towards full recovery through nucleoplasmic exchange, indicative of progressive weakening or loss of interactions within the condensate (*15*). Strikingly, the CRD imparted these behaviors in a broad range of sequence contexts and in the presence of multiple heterologous protein sequences that substantially varied physicochemical properties and disorder content. CRD phase separation therefore appears to be dominantly encoded by a collection of SLiMs that can act as strong attraction promoters (*76*) to support robust retention, while sequence differences tune dynamic properties within chromocenters (Fig. 6) (*77*). In the context of protein features or molecular grammar that drive

phase separation (*18-20, 59, 78*), the CRD appears to be an outlier over and above its unique dynamic properties.

Pericentromeric heterochromatin is the principal target for KMT5C-mediated catalysis of H4K20me3 (*33, 34, 38, 39*) (fig. S1). As a result, the retention of KMT5C within mouse chromocenters coincides with its product distribution and provides a means for the strict spatial control of this epigenetic mark. Its requirement for pre-existing H3K9me3 and HP1 chromobox proteins (*33, 37*) resembles the dependence of H2A lysine-119 ubiquitylation by RNF2 on formation of a phase separated compartment driven by binding of the CBX2 chromobox protein to H3K27me3 (*79*). Here, condensate formation is thought to provide a "reaction crucible" for localized activity of RNF2 (*79*). Consistent with this, perturbation of KMT5C chromocenter localization in *Suv39h1/2* knockout cells is associated with spreading of H4K20me3 outside of constitutive heterochromatin (*33*). Retention also provides a means to limit KMT5C access to substrates, which can account for why the trimethyl modification represents only a small fraction of H4K20 methylation states in mouse (*35*) and human (*40, 41*) cell lines. This differs from KMT5B catalysis of H4K20me2 (*34*), which accounts for the majority of H4K20 methylation states in mouse (*35*) and human (*40, 41*) cell lines and is therefore broadly distributed across the genome. At the DNA level, this reflects the association of H4K20me3 with arrays of tandem satellite repeats that make up the bulk of constitutive heterochromatin at pericentromeres (*39, 80-82*). Although these sequences are highly divergent between species (*81, 83-85*), they are thought to share a core set of epigenetic features (*86, 87*). Our findings with divergent CRD orthologs support a model wherein they recognize constitutive heterochromatin features that are conserved across distant evolutionary timescales. At the same time, the CRD may accommodate differences in rapidly evolving pericentromeric repeat sequences (*84, 88*) and it will be interesting to

determine if and how orthologs contribute to differences in constitutive heterochromatin between and within species (*30, 89-91*).

Methods

Cell Culture and transfection

Cells were cultured at 37°C and 5% CO_2 in a humidified incubator. All cell lines were grown in DMEM containing 10% FBS. D5 (*Suv39h1/2* knockout) and W8 (*Suv39h1/2* wild-type) mouse embryonic fibroblast cells lines (*32*) were obtained from Dr. Thomas Jenuwein. All other analyses were carried out using the mouse NMuMG breast cell line. Cells were transfected by lipofection using Effectene (Qiagen) 1 day prior to experiments. Expression plasmids were synthesized (www.biomatik.com), obtained from the addgene repository (www.addgene.org), or previously described (*92*). Pubertal mouse mammary gland sections were provided by Dr. Ing Swie Goping in accordance with the guidelines and regulations set forth by the Canadian Council on Animal Care and approved by the University of Alberta Health Sciences 2 Animal Care and Use Committee (Protocol #AUP00000386).

Live-cell imaging

Live cell imaging was carried out using Zeiss Axiovert 200M inverted microscopes attached to either an LSM510 NLO laser scanning system with a 25 mW argon laser line, a Zeiss LSM 770 confocal microscope attached to an Axio Observer Z3 equipped with 405, 488, 561, and 633 nm diode lasers, or a PerkinElmer Ultra*VIEW* ® spinning-disk confocal microscope equipped with 405, 488, and 561 nm diode lasers and a FRAP-unit. For all platforms, a 40 x 1.3 NA oil immersion lens was used. Long-term live-cell observations were conducted on the spinning disk microscope at 37°C with humidification and 5% CO_2. In cases where Z-stacks were acquired, spacing was set at 400 nm. Fluorescence recovery after photobleaching was performed on transiently transfected cells using the 488 nm solid state (spinning disk confocal) or 488 nm

argon laser line (LSM 510). Circular (chromocenter) or linear (nucleoplasm) regions were demarcated and subsequently bleached by intense light from the 488 nm laser. Fluorescence recovery of the bleached regions was quantified over multiple time scales (seconds to minutes). FRAP data was extracted using Zeiss LSM 5 Zen or ImageJ software (*93*) by measuring fluorescence intensity of the background, the whole nucleus, and the bleached area in each of the recorded time-lapse pictures for a minimum of 30 cells. Normalized relative intensity (including standard deviation) was calculated in Microsoft Excel and plotted using GraphPad Prism software. For half nucleus bleaching, a mask dividing the nuclei of interest into two halves was used, while a mask that covered all chromocenters except one was used for the iFRAP experiments. Convergence curves were obtained after partial chromocenter bleaching by measuring fluorescence intensity in the bleached and unbleached regions of interest using ImageJ and normalized as stated above.

Chromocenter spot-bleaching was performed on a PerkinElmer Ultra*VIEW*® spinning disc confocal microscope with Volocity® 6.3 software. Data was acquired using a 100x 1.4 NA Zeiss Plan-APOCHROMAT objective and a 1.6x optovar at 37°C with humidification and 5% CO_2. The center of each chromocenter was selected using the Volocity® Spot Tool and bleached using 120 cycles of the 488 nm laser at maximum power. Image acquisition was set to record 5 pre-bleach frames followed by 40 post-bleach frames at 1-second intervals. The recorded time series were imported to ImageJ for registration. Mean intensities were acquired for the bleached chromocenter, a reference chromocenter, and background. Individual curves were normalized in Microsoft Excel, averaged, and plotted with GraphPad Prism using a mono-exponential model ($f(t) = A(1-e^{(-t/t)})$ for curve fitting. The radius of bleach spots was determined using single images or stacks of paraformaldehyde-fixed KMT5C-mEmerald expressing cells by measuring the full

width at half maximum intensity and dividing the average by two. The apparent diffusion constant (0.019 $\mu m^2 s^{-1}$) was calculated from the measured radius (~0.32 ±0.08 μm) of the bleached spot and the time value (*t*) that indicated full fluorescence recovery using the following formula: $D_{app} = r^2/t$ (*51*).

Immunofluorescence

Cells grown on adherent coverslips were fixed in 4% paraformaldehyde in PBS for 10 min, permeabilized in 0.5 % Triton X-100 in PBS for 5min, and then incubated with primary antibodies diluted in PBS. After 30 min at room temperature, the cells were washed once with 0.1 % Triton X-100 in PBS for 1 min then rinsed 3 times with PBS prior to addition of secondary antibodies diluted in PBS. Cells were incubated for 30 min at room temperature, washed with 0.1 % Triton X-100 in PBS for 1 min and rinsed three times with PBS. Coverslips were mounted onto microscope slides with in-house-made polyvinyl alcohol mounting media containing 1 μg/mL 4′,6-diamidino-2-phenylindole (DAPI). Z-Stacks were obtained on a Zeiss Imager.Z1 equipped with a Photometrics Prime BSI camera and Metamorph® software version 7.10.2.240 (Molecular Devices, Sunnyvale, CA) using a Zeiss 63 x 1.3 NA oil lens. The step size used was 0.2 μm. Primary antibody for α-H4K20me3 (Active Motif 39672) was used at a dilution of 1:500 and detected using goat α-mouse Alexa-488 at 1:500.

Image analysis

Images were deconvolved with Huygens Professional version 19.04 (Scientific Volume Imaging, http://svi.nl). Z-stacks were imported into Imaris 9.3 (Oxford Instruments) and cropped to generate 3D images of single nuclei. The Imaris surfaces function was used to generate masks

based on intensity to encompass all chromocenters and the statistics function was used to measure the volume, sphericity, and total number of chromocenters for output to a Microsoft Excel spreadsheet. For partition coefficients, line scans through nuclei of undeconvolved 3D images were recorded using ImageJ. For each cell, the partition coefficient was calculated by subtracting the background level from the maxima of the brightest chromocenter and dividing it by the background corrected fluorescence intensity of the nucleoplasm. For each protein analyzed, measurements were taken from n>10 nuclei. Kymographs were acquired for regions of interest from laser micro-irradiation time series using the ImageJ Multi Kymograph tool.

Sequence analyses

In addition to recovery of all tetrapod KMT5C orthologs annotated in the NCBI database (https://www.ncbi.nlm.nih.gov/gene/84787/ortholog/?scope=32523&term=KMT5C), the Basic Local Alignment Search Tool (BLAST) (*94*) was used to expand this list to include unannotated orthologs in birds and reptiles using tblastn and a range of genome and transcript level sequence databases. Multiple alignments were carried out independently for CD1 and CD2 using ClustalΩ (*95*). Disorder profiles were derived using PONDR® (http://www.pondr.com/) and VL-XT mode. Per residue charge properties were determined using EMBOSS (https://www.bioinformatics.nl/cgi-bin/emboss/charge) with a window size of 1 and rendered as heatmaps using Heatmapper (http://www1.heatmapper.ca/) (*96*). Protein sequence features were obtained using the ProtPram tool (https://web.expasy.org/protparam/) (*97*) and CIDER (http://pappulab.wustl.edu/CIDER/) (*98*).

Statistical analysis

For partition coefficients, significance was evaluated using the Kruskal-Wallis one-way analysis test and individual comparisons between proteins were done using the Wilcoxon rank sum test. Statistical significance of sphericity data was evaluated using the embedded stats function within PlotsOfDifferences, which calculates p-values using a randomization test (*99*).

References

1. A. A. Hyman, C. A. Weber, F. Julicher, Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39-58 (2014).
2. J. M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu Rev Biophys* **49**, 107-133 (2020).
3. N. O. Taylor, M. T. Wei, H. A. Stone, C. P. Brangwynne, Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophysical journal* **117**, 1285-1300 (2019).
4. D. T. McSwiggen, M. Mir, X. Darzacq, R. Tjian, Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev* **33**, 1619-1634 (2019).
5. S. Boeynaems *et al.*, Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420-435 (2018).
6. S. Alberti, The wisdom of crowds: regulating cell function through condensed states of living matter. *J Cell Sci* **130**, 2789-2796 (2017).
7. J. B. Woodruff, A. A. Hyman, E. Boke, Organization and Function of Non-dynamic Biomolecular Condensates. *Trends Biochem Sci* **10.1016/j.tibs.2017.11.005**, (2017).
8. M. Feric *et al.*, Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, (2016).
9. O. V. Iarovaia *et al.*, Nucleolus: A Central Hub for Nuclear Functions. *Trends Cell Biol* **29**, 647-659 (2019).
10. C. C. Correll, J. Bartek, M. Dundr, The Nucleolus: A Multiphase Condensate Balancing Ribosome Synthesis and Translational Capacity in Health, Aging and Ribosomopathies. *Cells* **8**, (2019).
11. D. L. J. Lafontaine, J. A. Riback, R. Bascetin, C. P. Brangwynne, The nucleolus as a multiphase liquid condensate. *Nat Rev Mol Cell Biol* **22**, 165-182 (2021).
12. S. F. Banani *et al.*, Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663 (2016).
13. J. A. Ditlev, L. B. Case, M. K. Rosen, Who's In and Who's Out-Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666-4684 (2018).
14. A. S. Holehouse, R. V. Pappu, Functional Implications of Intracellular Phase Transitions. *Biochemistry* **57**, 2415-2423 (2018).
15. J. R. Espinosa *et al.*, Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proc Natl Acad Sci U S A* **117**, 13238-13247 (2020).
16. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340 (2012).
17. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).
18. J. Wang *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e616 (2018).
19. E. W. Martin *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694-699 (2020).
20. R. M. Vernon, J. D. Forman-Kay, First-generation predictors of biological protein phase separation. *Curr Opin Struct Biol* **58**, 88-96 (2019).

21. H. X. Zhou, V. Nguemaha, K. Mazarakos, S. Qin, Why Do Disordered and Structured Proteins Behave Differently in Phase Separation? *Trends Biochem Sci* **43**, 499-516 (2018).
22. A. R. Strom, C. P. Brangwynne, The liquid nucleome - phase transitions in the nucleus at a glance. *J Cell Sci* **132**, (2019).
23. L. Zhu, C. P. Brangwynne, Nuclear bodies: the emerging biophysics of nucleoplasmic phases. *Curr Opin Cell Biol* **34**, 23-30 (2015).
24. F. Erdel, K. Rippe, Formation of Chromatin Subcompartments by Phase Separation. *Biophysical journal* **114**, 2262-2270 (2018).
25. A. G. Larson, G. J. Narlikar, The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. *Biochemistry* **57**, 2540-2548 (2018).
26. H. Strickfaden *et al.*, Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell* **183**, 1772-1784 e1713 (2020).
27. B. A. Gibson *et al.*, Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484 e421 (2019).
28. A. G. Larson *et al.*, Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240 (2017).
29. A. R. Strom *et al.*, Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-245 (2017).
30. T. Haaf, M. Schmid, Chromosome topology in mammalian interphase nuclei. *Exp Cell Res* **192**, 325-332 (1991).
31. M. Guenatri, D. Bailly, C. Maison, G. Almouzni, Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* **166**, 493-505 (2004).
32. A. H. Peters *et al.*, Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-337 (2001).
33. G. Schotta *et al.*, A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**, 1251-1262 (2004).
34. G. Schotta *et al.*, A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev* **22**, 2048-2061 (2008).
35. M. Hahn *et al.*, Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. *Genes Dev* **27**, 859-872 (2013).
36. K. Muller-Ott *et al.*, Specificity, propagation, and memory of pericentric heterochromatin. *Molecular systems biology* **10**, 746 (2014).
37. P. P. Souza *et al.*, The histone methyltransferase SUV420H2 and Heterochromatin Proteins HP1 interact but show different dynamic behaviours. *BMC Cell Biol* **10**, 41 (2009).
38. V. L. Biron, K. J. McManus, N. Hu, M. J. Hendzel, D. A. Underhill, Distinct dynamics and distribution of histone methyl-lysine derivatives in mouse development. *Dev Biol* **276**, 337-351 (2004).
39. J. H. Martens *et al.*, The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *Embo J* **24**, 800-812 (2005).
40. G. Leroy *et al.*, A quantitative atlas of histone modification signatures from human cancer cells. *Epigenetics & chromatin* **6**, 20 (2013).

41. J. J. Pesavento, H. Yang, N. L. Kelleher, C. A. Mizzen, Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. *Mol Cell Biol* **28**, 468-486 (2008).
42. C. Fan *et al.*, Rett mutations attenuate phase separation of MeCP2. *Cell Discov* **6**, 38 (2020).
43. L. Wang *et al.*, Rett syndrome-causing mutations compromise MeCP2-mediated liquid-liquid phase separation of chromatin. *Cell Res* **30**, 393-407 (2020).
44. C. H. Li *et al.*, MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* **586**, 440-444 (2020).
45. A. Kumar *et al.*, Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J Cell Sci* **121**, 1128-1137 (2008).
46. T. Cheutin *et al.*, Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**, 721-725 (2003).
47. R. Festenstein *et al.*, Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells. *Science* **299**, 719-721 (2003).
48. N. E. Hynes *et al.*, New acceptor cell for transfected genomic DNA: oncogene transfer into a mouse mammary epithelial cell line. *Mol Cell Biol* **5**, 268-272 (1985).
49. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
50. S. C. Weber, Sequence-encoded material properties dictate the structure and function of nuclear bodies. *Curr Opin Cell Biol* **46**, 62-71 (2017).
51. R. S. Fisher, S. Elbaum-Garfinkle, Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nature communications* **11**, 4628 (2020).
52. F. Erdel *et al.*, Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. *Mol Cell* **78**, 236-249 e237 (2020).
53. M. Altmeyer *et al.*, Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nature communications* **6**, 8088 (2015).
54. H. Bierhoff *et al.*, Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. *Mol Cell* **54**, 675-682 (2014).
55. W. Fischle *et al.*, Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**, 1870-1881 (2003).
56. S. A. Jacobs, S. Khorasanizadeh, Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080-2083 (2002).
57. C. A. Musselman, M. E. Lalonde, J. Cote, T. G. Kutateladze, Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**, 1218-1227 (2012).
58. V. N. Uversky, Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr Opin Struct Biol* **44**, 18-30 (2017).
59. E. W. Martin, A. S. Holehouse, Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerg Top Life Sci* **4**, 307-329 (2020).
60. A. E. Posey, A. S. Holehouse, R. V. Pappu, Phase Separation of Intrinsically Disordered Proteins. *Methods Enzymol* **611**, 1-30 (2018).

61. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, (2017).
62. S. Sanulli, J. N. G, Liquid-like interactions in heterochromatin: Implications for mechanism and regulation. *Curr Opin Cell Biol* **64**, 90-96 (2020).
63. I. Peran, T. Mittag, Molecular structure in biomolecular condensates. *Curr Opin Struct Biol* **60**, 17-26 (2020).
64. D. M. Mitrea *et al.*, Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *eLife* **5**, (2016).
65. A. Peng, S. C. Weber, Evidence for and against Liquid-Liquid Phase Separation in the Nucleus. *Noncoding RNA* **5**, (2019).
66. A. Zidovska, The rich inner life of the cell nucleus: dynamic organization, active flows, and emergent rheology. *Biophys Rev* **12**, 1093-1106 (2020).
67. M. Mir, W. Bickmore, E. E. M. Furlong, G. Narlikar, Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development* **146**, (2019).
68. K. Rippe, Liquid-Liquid Phase Separation in Chromatin. *Cold Spring Harbor perspectives in biology*, (2021).
69. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
70. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).
71. C. Maison, G. Almouzni, HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* **5**, 296-304 (2004).
72. P. J. Skene *et al.*, Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell* **37**, 457-468 (2010).
73. D. Wang *et al.*, A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular systems biology* **15**, e8503 (2019).
74. D. W. Sanders *et al.*, Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306-324 e328 (2020).
75. C. H. Li *et al.*, MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature*, (2020).
76. A. Ghosh, K. Mazarakos, H. X. Zhou, Three archetypical classes of macromolecular regulators of protein liquid-liquid phase separation. *Proc Natl Acad Sci U S A* **116**, 19474-19483 (2019).
77. S. Elbaum-Garfinkle *et al.*, The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A* **112**, 7189-7194 (2015).
78. G. van Mierlo *et al.*, Predicting protein condensate formation using machine learning. *Cell reports* **34**, 108705 (2021).
79. J. M. Eeftens, M. Kapoor, D. Michieletto, C. P. Brangwynne, Polycomb condensates can promote epigenetic marks but are not required for sustained chromatin compaction. *Nature communications* **12**, 5888 (2021).
80. N. Saksouk, E. Simboeck, J. Dejardin, Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & chromatin* **8**, 3 (2015).

81. J. J. Yunis, W. G. Yasmineh, Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. *Science* **174**, 1200-1209 (1971).
82. D. I. Ostromyshenskii, E. N. Chernyaeva, I. S. Kuznetsova, O. I. Podgornaya, Mouse chromocenters DNA content: sequencing and in silico analysis. *BMC genomics* **19**, 151 (2018).
83. M. A. Garrido-Ramos, Satellite DNA: An Evolving Topic. *Genes* **8**, (2017).
84. S. S. Lower, M. P. McGurk, A. G. Clark, D. A. Barbash, Satellite DNA evolution: old ideas, new approaches. *Curr Opin Genet Dev* **49**, 70-78 (2018).
85. M. Plohl, N. Mestrovic, B. Mravinac, Satellite DNA evolution. *Genome Dyn* **7**, 126-152 (2012).
86. G. Almouzni, A. V. Probst, Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* **2**, 332-338 (2011).
87. J. Dejardin, Switching between Epigenetic States at Pericentromeric Heterochromatin. *Trends Genet* **31**, 661-672 (2015).
88. M. Plohl, A. Luchetti, N. Mestrovic, B. Mantovani, Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* **409**, 72-82 (2008).
89. Z. V. Avramova, Heterochromatin in animals and plants. Similarities and differences. *Plant Physiol* **129**, 40-49 (2002).
90. L. Manuelidis, A view of interphase chromosomes. *Science* **250**, 1533-1540 (1990).
91. H. A. Foster, J. M. Bridger, The genome and the nucleus: a marriage made by evolution. Genome organisation and nuclear architecture. *Chromosoma* **114**, 212-229 (2005).
92. L. W. Tsang, N. Hu, D. A. Underhill, Comparative analyses of SUV420H1 isoforms and SUV420H2 reveal differences in their cellular localization and effects on myogenic differentiation. *PLoS ONE* **5**, e14447 (2010).
93. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
94. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
95. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).
96. S. Babicki *et al.*, Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res* **44**, W147-153 (2016).
97. E. Gasteiger *et al.*, in *The Proteomics Protocols Handbook,* J. M. Walker, Ed. (Humana Press, Totowa, NJ, 2005), pp. 571-607.
98. R. K. Das, R. V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A* **110**, 13392-13397 (2013).
99. J. Goedhart, PlotsOfDifferences – a web app for the quantitative comparison of unpaired data. *bioRxiv*, 578575 (2019).
100. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394 (2015).

Supplementary Data

Supplementary figures

S1A. H4K20me3/DAPI in NMuMG

S1B. H4K20me3/DAPI in pubertal mammary gland

S2A. Half nucleus bleach for KMT5C, MeCP2, and CBX5

S2B. Inverse FRAP (iFRAP) image series for KMT5C

S2C. Inverse FRAP (iFRAP) graph for KMT5C

S3A. Convergence image series for KMT5C

S3B. Convergence graph for KMT5C

S3C. Extended laser micro-irradiation and FRAP analyses of KMT5C, MeCP2, and CBX5

S4A. CD1 and CD2 localization

S4B. CD1 and CD2 FRAP

S4C. Localization of *Hs*, *Mm*, *Cp*, *Bb*, *Gg*, and *Pm* CRDs

S5. CRD-IDP1, 2, 3, 4, and 5 localization

Supplementary movies

S1. KMT5C FRAP timelapse

S2. MeCP2 FRAP timelapse

S3. CBX5 FRAP timelapse

S4. KMT5C laser micro-irradiation

S5. KMT5C FRAP post laser micro-irradiation

S6. MeCP2 laser micro-irradiation

S7. MeCP2 FRAP post laser micro-irradiation

S8. CBX5 laser micro-irradiation

S9. CBX5 FRAP post laser micro-irradiation

S10. CRD FRAP timelapse

S11. CRD Cys mutant FRAP

S12. CRD His mutant FRAP

S13. CRD LVL1 mutant FRAP

S14. CRD LVL2 mutant FRAP

S15. CRD Trp mutant FRAP

Supplementary tables

S1. Physicochemical properties of CRD orthologs and chimeras assessed using CIDER and ProtParam tools

S2. Disorder profile data for CRD orthologs and chimeras assessed using PONDR (VL-XT mode)

Acknowledgements

The authors acknowledge funding support from the Canadian Breast Cancer Foundation (D.A.U., grant no. 300073), Cancer Research Society (D.A.U. and M.J.H., grant no. CRSDI 2018 OG 23446), and Canadian Institutes of Health Research (M.J.H., grant no. PJT-148753). The authors thank Dr. Xuejun Sun, Gerry Barron, and Dr. Guobin Sun of the Cross Cancer Institute Cell Imaging Facility for support, and Darryl Glubrecht for immunofluorescence staining of mouse mammary gland tissue.

Author Contributions

D.A.U. and M.J.H. conceived the project. D.A.U., M.J.H, and H.S. designed all experiments and interpreted data. K.M. made the initial observation of KMT5C mobility and performed FRAP and fluorescence imaging experiments. H.S. performed live-cell imaging, laser micro-irradiation, and partition coefficient analyses. J.W.K. assisted with imaging, FRAP, and phylogenetic analyses. D.A.U. carried out sequence analysis and expression plasmid design. D.A.U. wrote the manuscript with contributions from M.J.H. and H.S.. D.A.U. prepared figures with contributions from H.S., J.W.K., and K.M.

Containing data deposition statement

N/A

Competing interests

The authors declare no competing interests.

Corresponding author line

D. A. Underhill Department of Oncology, 2328 Cross Cancer Institute, 11560 University Avenue, Edmonton, Alberta, Canada, T6G 1Z2. Phone: (780) 432-8903, Fax: (780) 432-8892, E-mail: **alan.underhill@ualberta.ca**

Figure 1. KMT5C is mobile within chromocenters but undergoes limited nucleoplasmic exchange. A, Time-lapse series of Total and Partial KMT5C-mEmerald photobleaching (Movie S1) from 0.3 to 60s in mouse NMuMG breast cancer cells (n=30) (see Methods). **B,** Time-lapse series of Total and Partial MeCP2 and CBX5-mEmerald photobleaching (Movies S2 and S3) from 0.3 to 60s in mouse NMuMG breast cancer cells (see Methods). **C,** FRAP curves for KMT5C, MeCP2 and CBX5-mEmerald fusion proteins under Total (closed circles) and Partial (open circles). Symbols represent mean fluorescence intensity, while vertical lines indicate standard deviation (n=30). The schematics summarize KMT5C, MeCP2, and CBX5 movement within chromocenters (arrows of increasing weight) and the presence (dashed line) or absence (solid line) of exchange with the surrounding nucleoplasm. **D,** Scatter plots display corresponding partition coefficients (chromocenter vs. nucleoplasm) for KMT5C (n=26), MeCP2 (n=30), and CBX5 (n=30). **E,** Summary schematic indicating relationship between internal mobility, nucleoplasmic exchange, and partitioning for KMT5C, MeCP2, and CBX5.

Figure 2. KMT5C displays liquid-like behavior in phase-separated heterochromatin compartments. A, Representative images for spot bleach analysis of KMT5C. Individual chromocenters underwent photobleaching of a defined volume and fluorescence recovery was measured over time. **B,** Corresponding recovery curve depicting the $t_{1/2}$ for recovery and apparent diffusion rate (D_{app}) of KMT5C within chromocenters. **C,** Response of KMT5C, MeCP2, and CBX5 to laser micro-irradiation induced DNA damage within chromocenters. For each protein, the response to laser micro-irradiation (red arrow with dashed line in top left panel) was monitored over time (right hand panels showing pre-bleach, 2, and 60s post-bleach). Kymographs over the same time (0-60s) are shown for each protein in the lower left-hand panel.

FRAP performed post laser micro-irradiation is shown in figure S3C. **D,** Comparison of chromocenter sphericity induced by expression of KMT5C (*n=775*), MeCP2 (*n=641*), and CBX5 (*n=644*).

Figure 3. A rudimentary heterochromatin targeting domain underlies CRD activity. **A,** Primary structure schematic for human KMT5C (total length 462 aa). Depicted are the Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain, which catalyzes H4K20me3; the RNA-binding domain (RBD); and the chromocenter retention domain (CRD), which underlies binding of KMT5C to constitutive heterochromatin. **B,** Time-lapse series of *Total* and *Partial* CRD-mEmerald photobleaching (movie S10) from 0.3 to 60s in mouse NMuMG breast cancer cells (*n=30*) (see Methods). **C,** Scatter plots display corresponding partition coefficients (chromocenter vs. nucleoplasm) for KMT5C (n=26) (from figure 1D) and human (n=39) and mouse (n=30) CRDs.

Figure 4. Divergent tetrapod CRDs retain localization and dynamic behavior. A, Sequence alignments for representative CRD orthologs from mammals (human icon), bird (chicken icon), and reptiles (lizard icon) are shown in JPred (*100*) format. Conservation within each species class is depicted graphically using ClustalΩ (*95*) as follows: '*', fully conserved; ":", conservation of strongly similar properties; or ".", conservation of weakly similar properties. Both CD1 and CD2 regions are shown. Conservation of hydrophobic motifs is shown in gray, while all other conserved residues are indicated in red. The divergent linker sequence has been omitted but varies from 5-191 amino acids as noted. CD1 contains a highly conserved triad of a histidine and 2 cysteine residues in addition to 5 conserved hydrophobic residues. The only conserved feature within CD2 is the hydrophobic motif, with all other sequences being class

specific. Species key: *Homo sapiens* (*Hs*), *Pan troglodytes* (*Pt*), *Ochotona princeps* (*Op*), *Mus musculus* (*Mm*), *Jaculus jaculus* (*Jj*), *Echinops telfairi* (*Et*), *Sorex araneus* (*Sa*), *Orycteropus afer afer* (*Oaa*), *Canis lupus familiaris* (*Clf*), *Bubalus bubalis* (*Bb*), *Cavia porcellus* (*Cp*), *Coturnix coturnix* (*Cc*), *Gallus gallus* (*Gg*), *Phasianus colchicus* (*Pc*), *Numida meleagris* (*Nm*), *Catharus ustulatus* (*Cu*), *Anolis carolinensis* (*Ac*), *Pogona vitticeps* (*Pv*), *Lacerta agilis* (*La*), *Python bivittatus* (*Pb*), and *Podarcis muralis* (*Pm*). CRD orthologs shown in red where used in FRAP analyses. **B,** FRAP analyses of CRD-mEmerald orthologs (expressed as mEmerald fusion proteins) from mammals (*Hs*, *Mm*, *Cp*, and *Bb*), birds (*Gg*), and reptiles (*Pm*) under Total (closed circles) and Partial (open circles). Symbols represent mean fluorescence intensity, while vertical lines indicate standard deviation (n=30). For each CRD, linker length (L), isoelectric point (pI), and fraction of charged residues (FCR) is shown. Corresponding localization data is shown in fig. S4C

Figure 5. CRD localization and dynamics are conferred by Short Linear Motifs (SLiMs). A, Schematic summary of CRD mutants (formatted as in figure 4A). Mutations indicated above the CRD sequence correspond to residues conserved in mammals, birds, and reptiles, while those below the sequence were unique to mammals (indicated by species icons). Mutations targeted $C^{362}C^{366}$ (Cys), $H^{357}H^{365}$ (His), $L^{368}V^{374}L^{376}$ (LVL1), $L^{398}V^{404}L^{406}$ (LVL2), and $W^{359}W^{390}W^{392}Y^{396}$ (Trp). In each case, the corresponding amino acid substitutions are indicated. **B,** Representative images for the wild-type, $C^{362}C^{366}$ (Cys), $H^{357}H^{365}$ (His), $L^{368}V^{374}L^{376}$ (LVL1), $L^{398}V^{404}L^{406}$ (LVL2), and $W^{359}W^{390}W^{392}Y^{396}$ (Trp) mutant CRDs expressed as mEmerald fusion proteins. **C,** FRAP analyses of CRD-mEmerald mutant proteins under Total (closed circles) and Partial (open circles) conditions. The wild-type CRD (from figure 4) is shown for reference.

Symbols represent mean fluorescence intensity, while vertical lines indicate standard deviation (n=30). **D,** For the wild-type and mutant proteins, schematics summarize CRD movement within chromocenters (arrows of increasing weight) and the presence (dashed line) or absence (solid line) of exchange with the surrounding nucleoplasm.

Figure 6. CRD localization and dynamics are not affected by substantive changes in intrinsic disorder content and charge properties. A, *Top panel*, predicted disorder profiles (PONDR in VL-XT mode) for CRD orthologs from mammals (table S2). Derivatives from human (*Hs*), mouse (*Mm*), guinea pig (*Cp*), and water buffalo (*Bb*) are noted, while all other species are shown in faded gray. *Lower panel*, predicted disorder profiles (PONDR in VL-XT mode) for representative bird (blue) and reptile (gray) CRD orthologs (table S2). **B,** Predicted disorder profiles (PONDR in VL-XT mode) for CRD-IDP (1-5) chimeras (table S2). **C,** Per residue charge (EMBOSS) of heterologous intrinsically disorder linkers used to generate CRD-IDR chimeras. Fraction of charged residues (FCR) and net charge per residue (NCPR) for each linker is shown above the heatmap (FCR/NCPR), while the values for the corresponding CRD-IDR chimera are provided below (FCR of the wild-type CRD is 0.183 and NCPR is 0.083). **D,** FRAP analyses of CRD-IDR-mEmerald chimeric proteins under Total (closed circles) and Partial (open circles) conditions. Symbols represent mean fluorescence intensity, while vertical lines indicate standard deviation (n=30). For reference, the dashed lines represent the maximal recovery values for the wild-type CRD for total (lower) or partial (upper) bleaching.

FIGURE 1

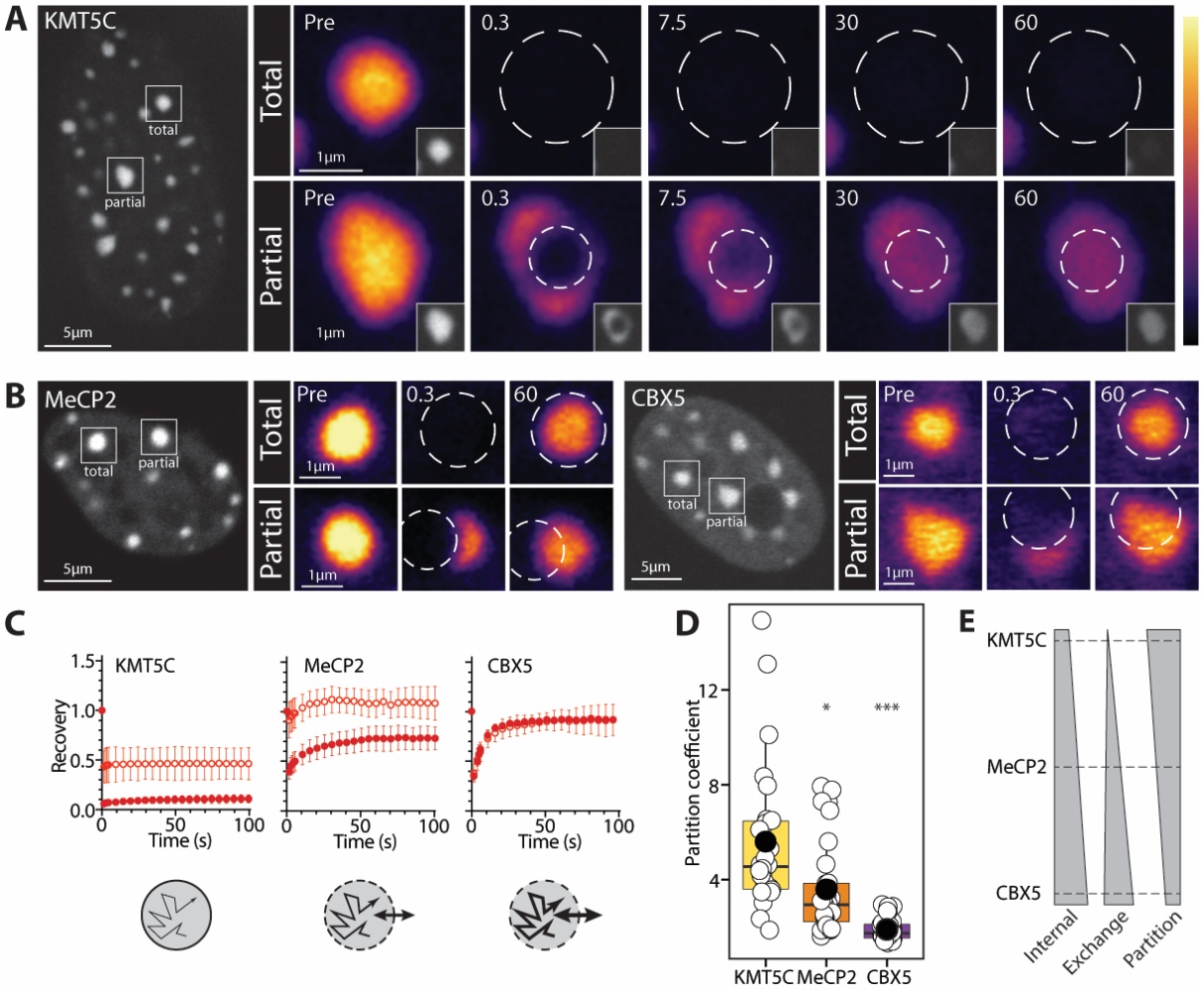

A
KMT5C
total
partial
5µm
Total
Partial
Pre
0.3
7.5
30
60
1µm
B
MeCP2
CBX5
C
KMT5C
MeCP2
CBX5
Recovery
Time (s)
D
Partition coefficient
KMT5C MeCP2 CBX5
E
Internal
Exchange
Partition

FIGURE 2

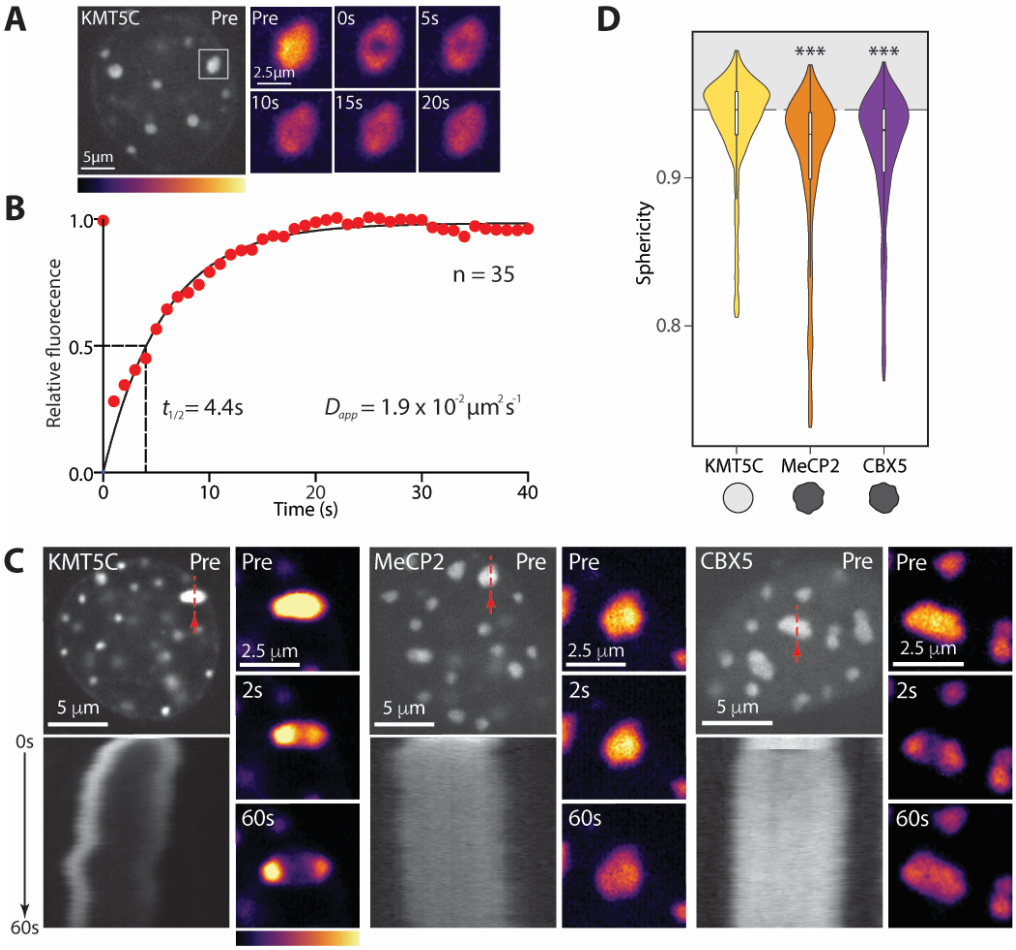

A
KMT5C
Pre
5µm
Pre
0s
5s
2.5µm
10s
15s
20s
B
Relative fluorescence
1.0
0.5
0.0
n = 35
$t_{1/2}$= 4.4s
D_{app}= 1.9 x 10^{-2} µm^{2}s^{-1}
0
10
20
30
40
Time (s)
D
Sphericity
0.9
0.8
***
***
KMT5C
MeCP2
CBX5
C
KMT5C
Pre
5 µm
Pre
2.5 µm
2s
60s
0s
60s
MeCP2
Pre
5 µm
Pre
2.5 µm
2s
60s
CBX5
Pre
5 µm
Pre
2.5 µm
2s
60s

FIGURE 3

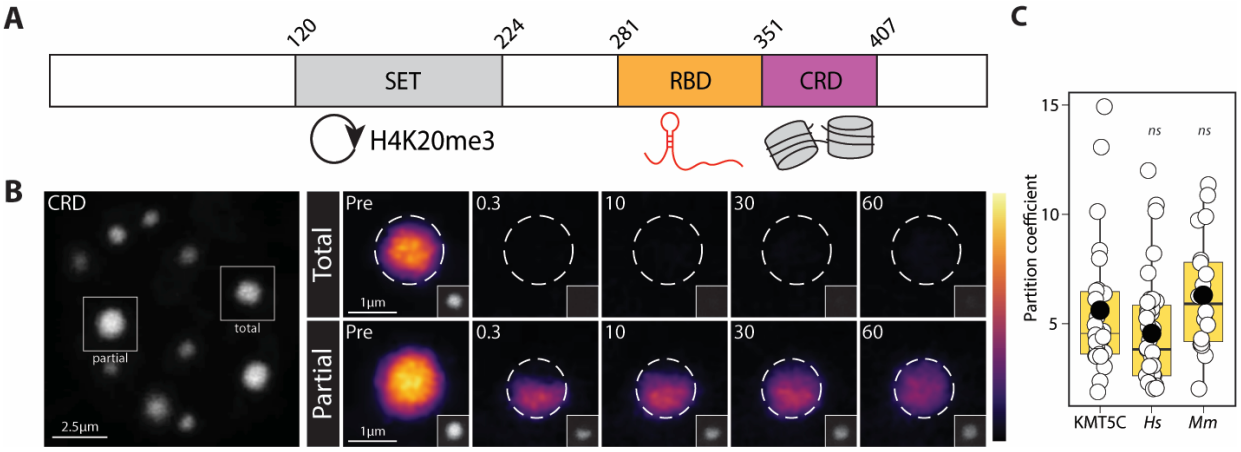

A
120
224
281
351
407
SET
RBD
CRD
H4K20me3
B
CRD
partial
total
2.5µm
Total
Partial
Pre
0.3
10
30
60
1µm
C
Partition coefficient
15
10
5
ns
ns
KMT5C
Hs
Mm

FIGURE 4

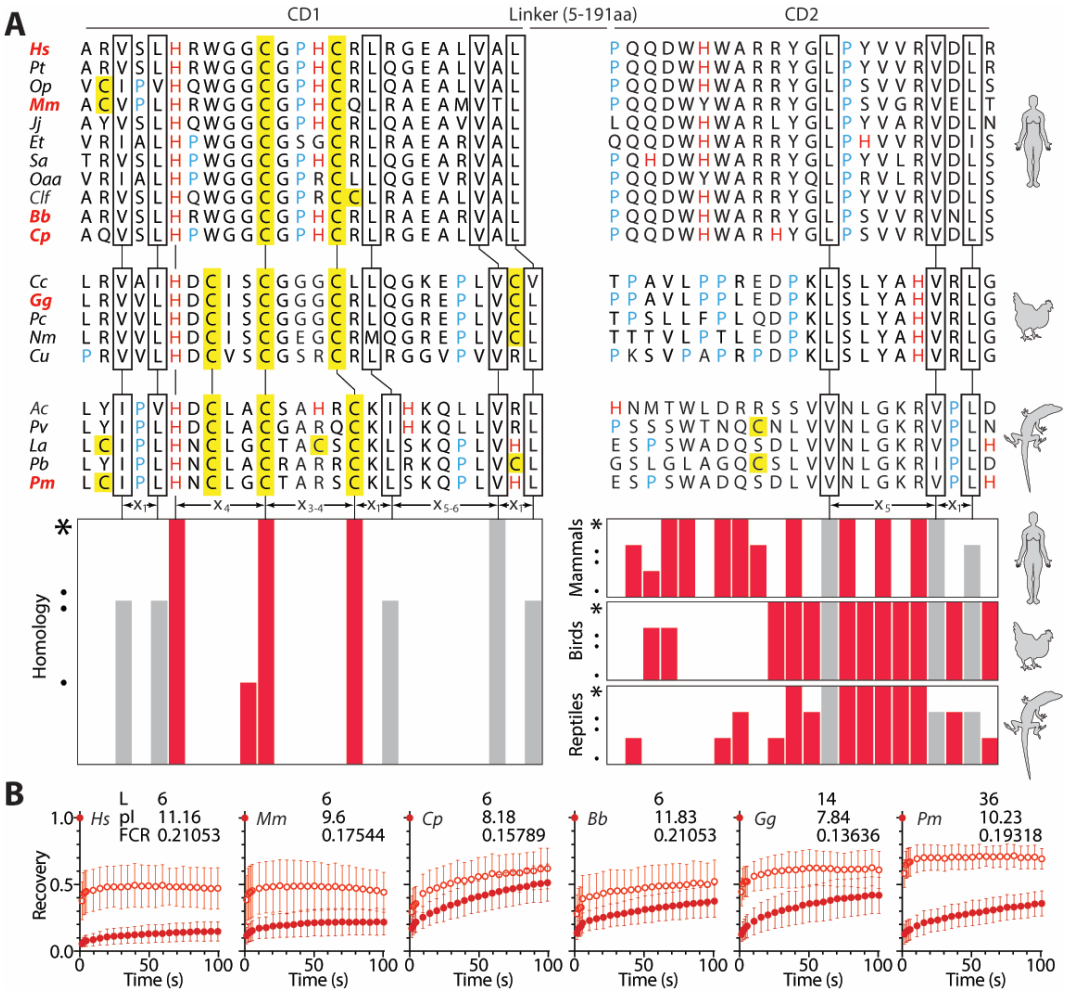

A
CD1
Linker (5-191aa)
CD2
Hs
Pt
Op
Mm
Jj
Et
Sa
Oaa
Clf
Bb
Cp
Cc
Gg
Pc
Nm
Cu
Ac
Pv
La
Pb
Pm
Homology
Mammals
Birds
Reptiles
B
Hs L 6 pI 11.16 FCR 0.21053
Mm 6 9.6 0.17544
Cp 6 8.18 0.15789
Bb 6 11.83 0.21053
Gg 14 7.84 0.13636
Pm 36 10.23 0.19318
Recovery
Time (s)

FIGURE 5

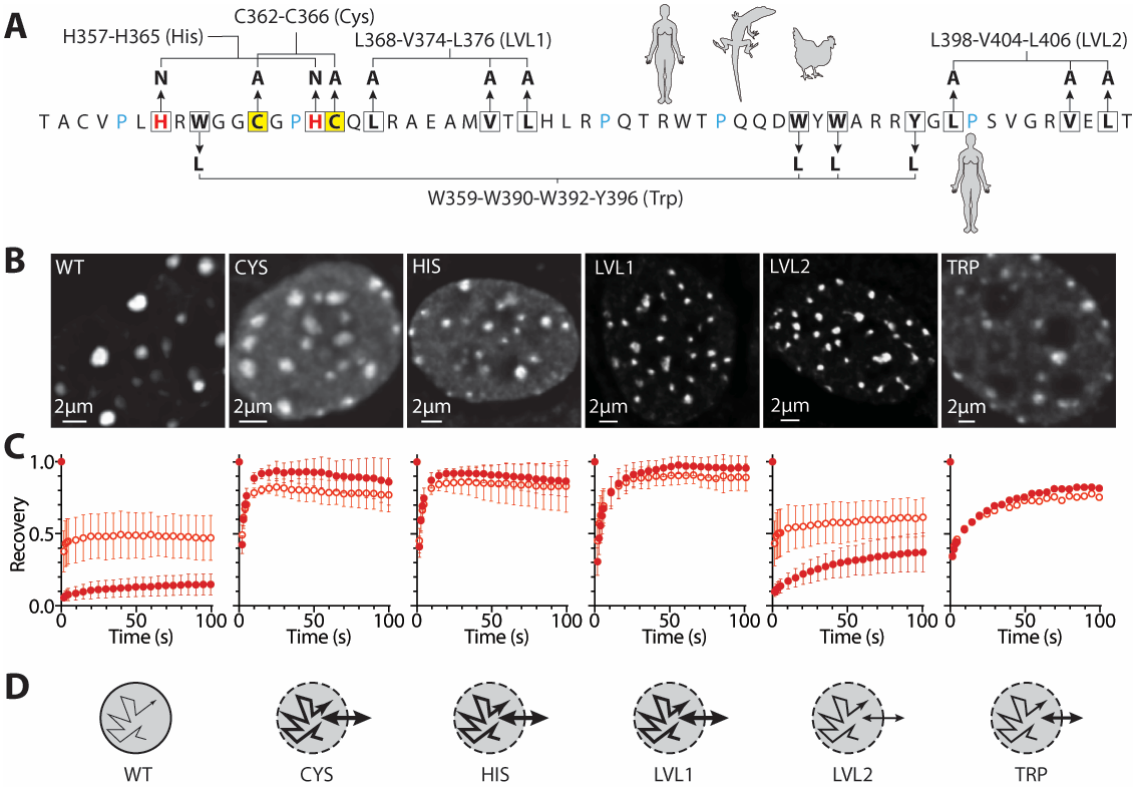

A
H357-H365 (His)
C362-C366 (Cys)
L368-V374-L376 (LVL1)
L398-V404-L406 (LVL2)
W359-W390-W392-Y396 (Trp)
B
WT
CYS
HIS
LVL1
LVL2
TRP
2µm
C
Recovery
Time (s)
D
WT
CYS
HIS
LVL1
LVL2
TRP

FIGURE 6

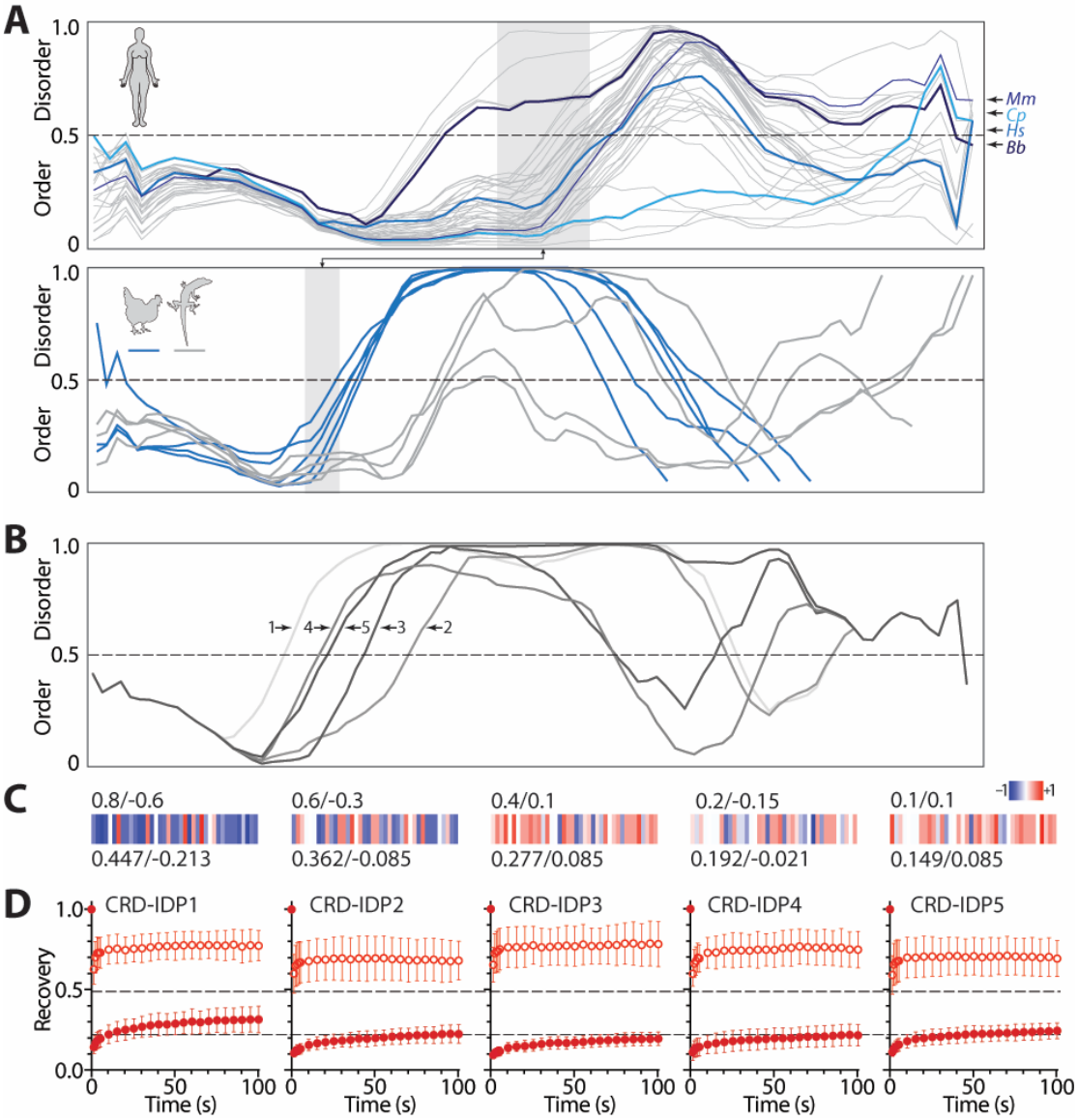

A
1.0
Disorder
0.5
Order
0
Mm
Cp
Hs
Bb
B
1
4
5
3
2
C
0.8/-0.6
0.447/-0.213
0.6/-0.3
0.362/-0.085
0.4/0.1
0.277/0.085
0.2/-0.15
0.192/-0.021
0.1/0.1
0.149/0.085
-1
+1
D
CRD-IDP1
CRD-IDP2
CRD-IDP3
CRD-IDP4
CRD-IDP5
Recovery
0.0
Time (s)
0
50
100

Figure S1. H4K20me3 shows strict spatial control in situ and in vivo. A, H4K20me3 immunofluorescence analysis in the NMuMG immortalized mouse mammary epithelial cell line. **B,** H4K20me3 immunofluorescence analysis in the 8-week mouse pubertal mammary gland. DAPI panels are shown on the right

Figure S2. The retentive behavior of KMT5C is pan-nuclear and reflects minimal efflux from chromocenter to the surrounding nucleoplasm. A, Half-nucleus photobleaching of KMT5C, MeCP2, and CBX5. Image panels correspond to pre-bleach (Pre), immediately after photobleaching (Post), or 60 seconds after photobleaching (60s). The vertical dashed line demarcates the boundary between the bleached and unbleached portions of the nucleus. The dynamic behaviors are summarized schematically below the image panels as limited remixing (KMT5C), partial remixing (MeCP2), or complete remixing (CBX5) when comparing the pre-bleach and 60 second time points. **B,** Inverse FRAP (iFRAP) timelapse of KMT5C-mEmerald. Image panels are shown using mpl-inferno to illustrate intensity differences pre-bleach (Pre), immediately after photobleaching (Post), or 30 minutes after photobleaching (30min). The signal has been enhanced in the bottom right panel to indicate residual fluorescence of unbleached chromocenters at 30 minutes. **C,** Corresponding iFRAP curve depicts relative fluorescence intensity over a 30-minute period. The inset shows fluorescence intensity over the first 100 seconds, which was the standard recovery period for conventional FRAP analyses in this study. The apparent Efflux Rate (ER_{app}) of 0.7%min^{-1} was calculated from the slope of the iFRAP curve.

Figure S3. KMT5C exhibits like-like behavior within chromocenters. A, Representative time lapse images of half chromocenter photobleaching showing pre-bleach (Pre), immediately after

photobleaching (Post), and at the 50 (50s) and 100 second (100s) time points. The same time series is shown below using the mpl-inferno LUT to emphasize intensity differences. **B,** Convergence curves for KMT5C-mEmerald following half chromocenter bleaching illustrate relative fluorescence over time in the bleached (red symbols) and unbleached (black symbols) portions of the chromocenter (n=10). The schematic indicates internal remixing of fluorescent molecules in the absence of substantive influx from the surrounding nucleoplasm. **C,** Response of KMT5C, MeCP2, and CBX5 to laser micro-irradiation induced DNA damage within chromocenters (expanded version of Fig. 2B). For each protein, the response to laser micro-irradiation (red arrow and dashed line) was monitored over time (upper panel, micro-IR). The lower panel (Photobleach) depicts FRAP analysis that was carried out post-laser micro-irradiation on the same chromocenter. For KMT5C, one of the two lobes generated following laser micro-irradiation was fully bleached. For MeCP2 and CBX5, a portion of the damaged chromocenter was partially photobleached. The region selected for photobleaching is indicated by a dashed white circle. Fluorescence recovery is shown over a 60 second period for KMT5C and 20 seconds for both MeCP2 and CBX5.

Figure S4. CRD orthologs from mammals, birds, and reptiles show efficient chromocenter localization that is driven by the CD1 subdomain. A, Representative images for the CD1 and CD2 subdomains of the CRD expressed as mEmerald fusion proteins. **B,** FRAP analyses of CD1 and CD2-mEmerald fusions under Total (closed circles) and Partial (open circles). Symbols represent mean fluorescence intensity, while vertical lines indicate standard deviation (n=30). **C,** Representative images of CRD orthologs from human (*Hs*; *Homo sapiens*), mouse (*Mm*; *Mus*

musculus), guinea pig (*Cp*; *Cavia porcellus*), water buffalo (*Bb*; *Bubalus bubalis*), chicken (*Gg*; Gallus gallus), and wall lizard (*Pm*; *Podarcis muralis*) expressed as mEmerald fusion proteins.

Figure S5. CRD-IDR chimeras retain efficient chromocenter localization. Representative images for CRD-IDR1/2/3/4/5 chimeras expressed as mEmerald fusion proteins.

figure S1

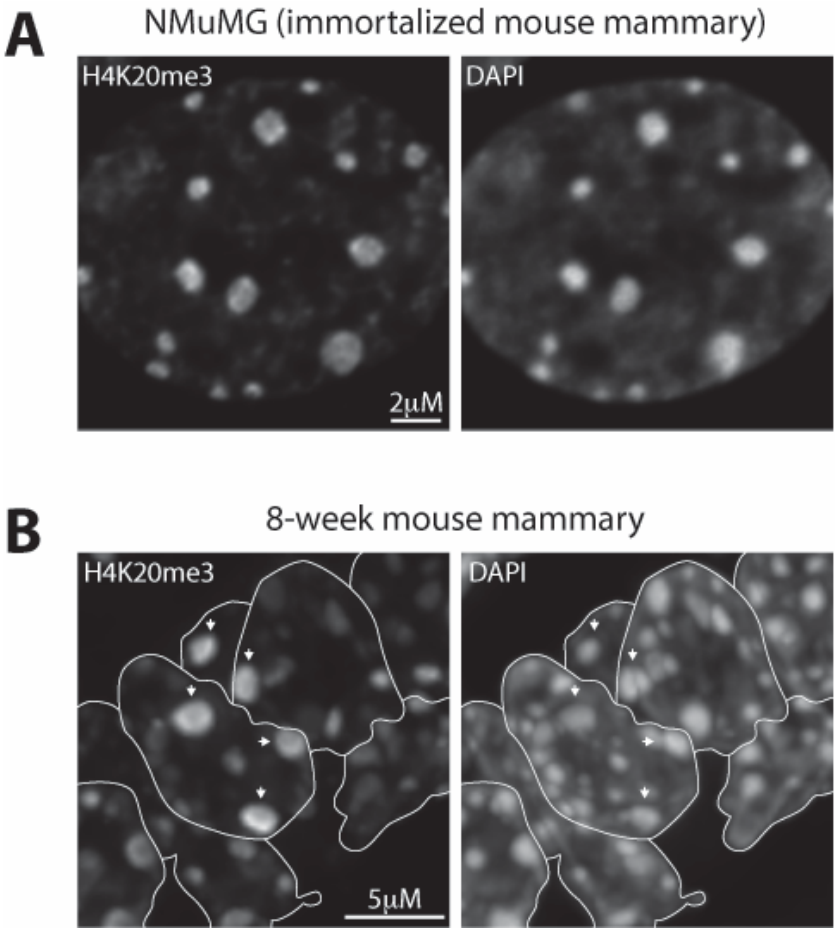

A
NMuMG (immortalized mouse mammary)
H4K20me3
DAPI
2μM
B
8-week mouse mammary
H4K20me3
DAPI
5μM

figure S2

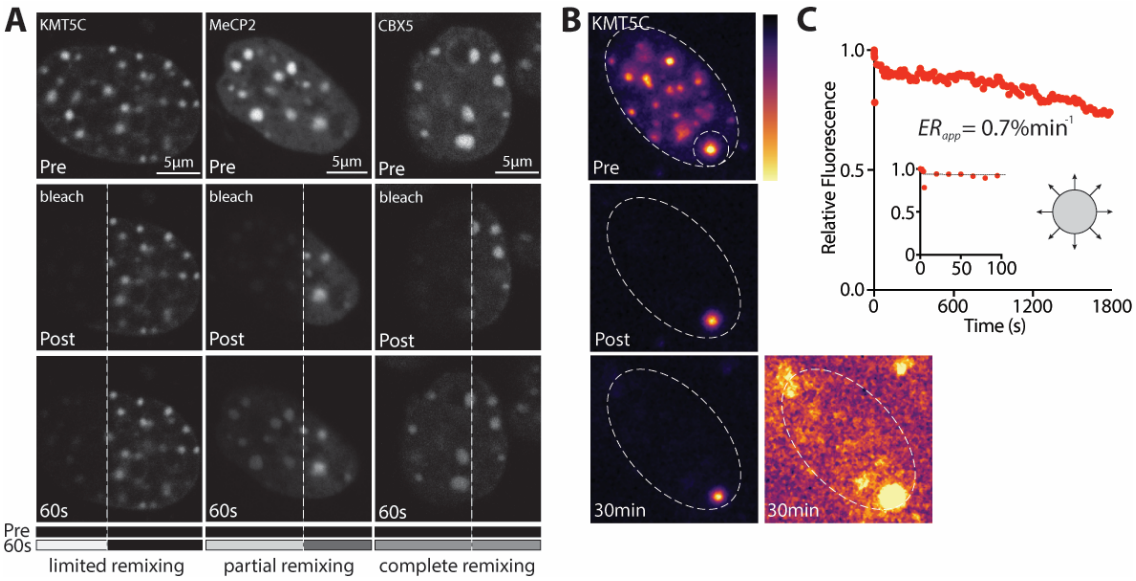

A
KMT5C
MeCP2
CBX5
Pre
5µm
bleach
Post
60s
Pre
60s
limited remixing
partial remixing
complete remixing
B
KMT5C
Pre
Post
30min
C
1.0
0.5
0.0
Relative Fluorescence
ER_{app}= 0.7%min^{-1}
0
50
100
600
1200
1800
Time (s)

figure S3

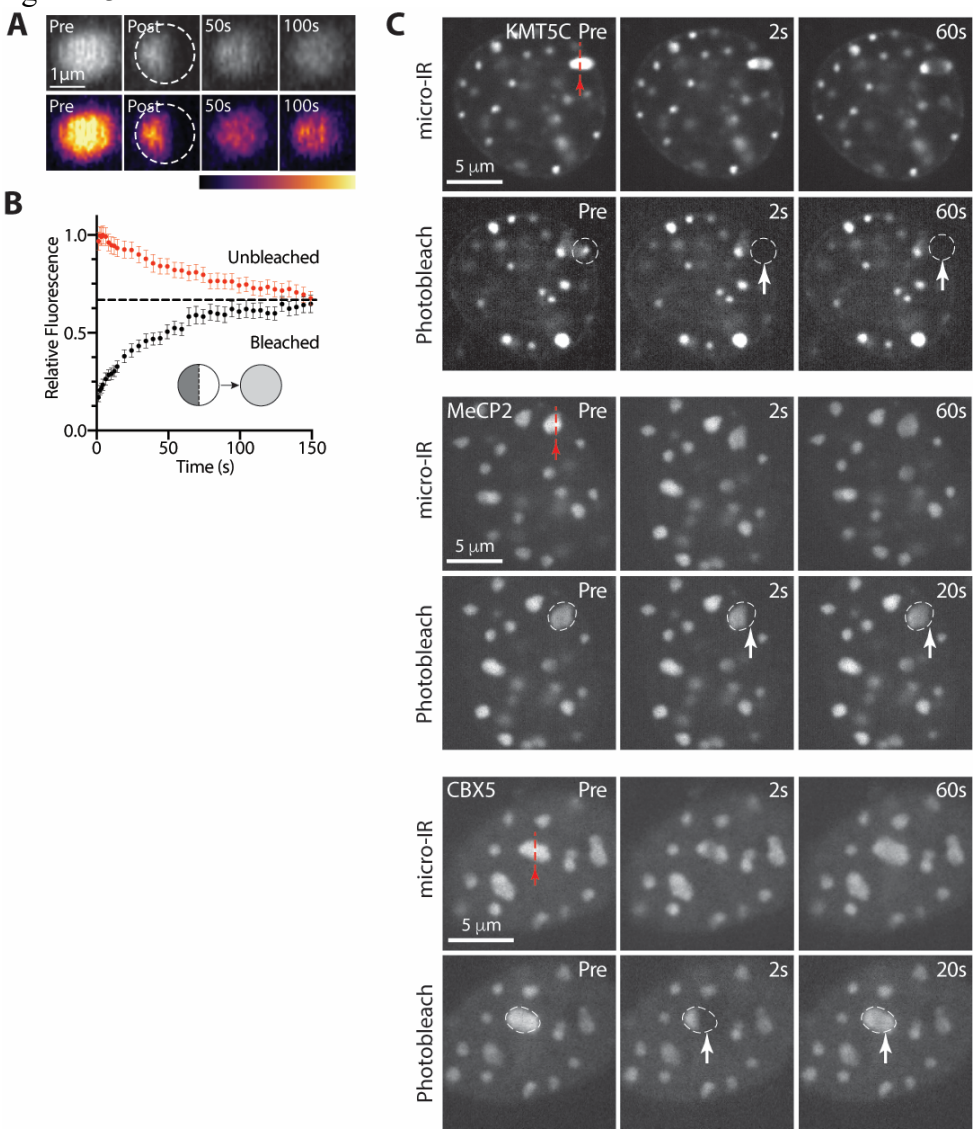

A
Pre
Post
50s
100s
1µm
B
Relative Fluorescence
1.0
0.5
0.0
Unbleached
Bleached
0
50
100
150
Time (s)
C
KMT5C
MeCP2
CBX5
micro-IR
Photobleach
Pre
2s
20s
60s
5 µm

figure S4

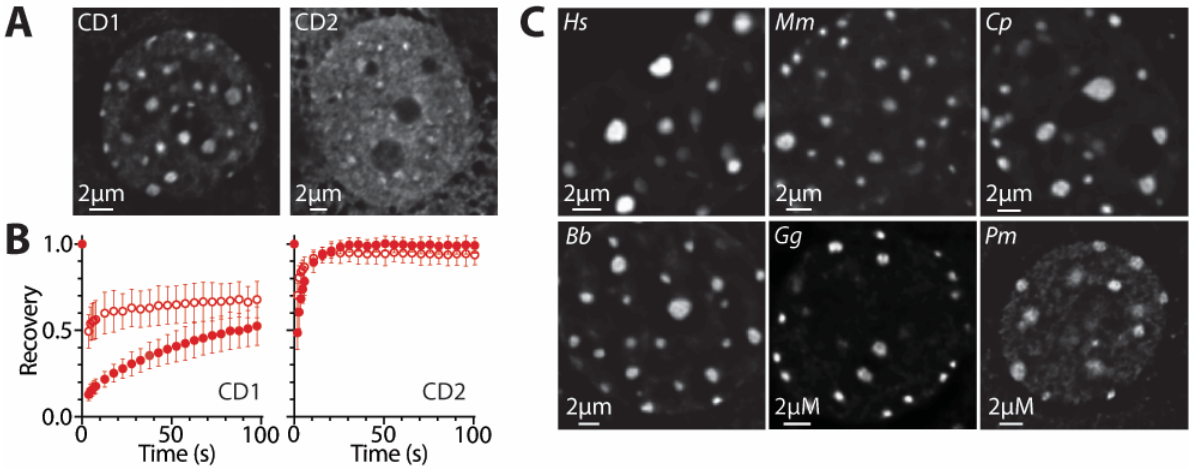
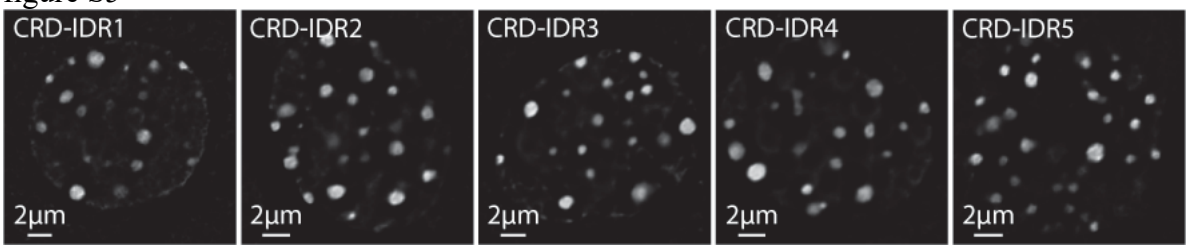

A
CD1
CD2
2µm
2µm
B
1.0
0.5
0.0
Recovery
CD1
CD2
0
50
100
Time (s)
C
Hs
Mm
Cp
Bb
Gg
Pm
2µm
2µM

figure S5


CRD-IDR1
2μm
CRD-IDR2
2μm
CRD-IDR3
2μm
CRD-IDR4
2μm
CRD-IDR5
2μm