Nuclear compartmentalization as a mechanism of quantitative control of gene expression

Prashant Bhat[1,2], *Drew Honson*[1] *and Mitchell Guttman*[1] ✉

Abstract | Gene regulation requires the dynamic coordination of hundreds of regulatory factors at precise genomic and RNA targets. Although many regulatory factors have specific affinity for their nucleic acid targets, molecular diffusion and affinity models alone cannot explain many of the quantitative features of gene regulation in the nucleus. One emerging explanation for these quantitative properties is that DNA, RNA and proteins organize within precise, 3D compartments in the nucleus to concentrate groups of functionally related molecules. Recently, nucleic acids and proteins involved in many important nuclear processes have been shown to engage in cooperative interactions, which lead to the formation of condensates that partition the nucleus. In this Review, we discuss an emerging perspective of gene regulation, which moves away from classic models of stoichiometric interactions towards an understanding of how spatial compartmentalization can lead to non-stoichiometric molecular interactions and non-linear regulatory behaviours. We describe key mechanisms of nuclear compartment formation, including emerging roles for non-coding RNAs in facilitating their formation, and discuss the functional role of nuclear compartments in transcription regulation, co-transcriptional and post-transcriptional RNA processing, and higher-order chromatin regulation. More generally, we discuss how compartmentalization may explain important quantitative aspects of gene regulation.

Affinity
The strength of a non-covalent biochemical interaction, defined as the ratio of the association and dissociation rates.

Gene regulation is a highly complex process that requires the dynamic coordination of hundreds of regulatory factors — including chromatin and transcription regulators, splicing factors and other mRNA processing factors — at precise molecular targets (such as DNA sequences and RNA structures) in different cell states. Although many regulatory factors have specific affinity for their targets, molecular diffusion and affinity models alone cannot explain many of the observed quantitative properties of these regulatory processes in the nucleus. For example, the rate at which transcription factors bind to their targets is >1,000-fold faster than is predicted from diffusion and affinity alone[1,2].

Recent advances in our understanding of nuclear organization, driven by new genomics and microscopy methods (BOX 1), suggest a new paradigm that may explain many of these gene regulatory properties. Specifically, DNA, RNA and protein molecules can organize within precise 3D nuclear territories to concentrate groups of functionally related molecules[3–6]. For example, genomic DNA is dynamically organized to promote enhancer–promoter interactions[7], topological association of sets of co-regulated genes[8] and recruitment of DNA and pre-mRNA to different nuclear compartments[9,10]. Moreover, many proteins involved in transcription regulation (such as RNA polymerase II (Pol II)[11] and transcription factors[12]), enhancer–promoter interactions (for example, the Mediator complex)[11], heterochromatin formation and maintenance (for example, heterochromatin protein 1 (HP1)[13,14] and SAFB[15]) and pre-mRNA splicing (for example, FUS)[16] are enriched within high-concentration nuclear territories. A central tenet of this model of gene regulation is that formation of nuclear compartments can act to partition molecules and biochemical functions[17–19] (FIG. 1a).

In this Review, we discuss an emerging paradigm of gene regulation that moves away from classic stoichiometric interaction models, and describe how spatial compartmentalization can lead to non-stoichiometric molecular interactions and non-linear regulatory behaviours. We discuss the molecular mechanisms by which compartmentalization is achieved and the role of compartmentalization in spatially organizing enhancers, promoters and transcription factors to drive transcription initiation. In addition, we explore the role of compartmentalization in controlling higher-order chromatin organization and regulation,

[1]*Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA.*

[2]*David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.*

✉*e-mail: mguttman@caltech.edu*

https://doi.org/10.1038/s41580-021-00387-1

Box 1 | **Methods for exploring compartmentalization**

Several methodologies have been developed to define the molecular components and functional roles of nuclear compartments. We briefly describe some of these key methods below (reviewed in detail elsewhere[174]).

In vitro methods

- Droplet formation assays[175] involve increasing the concentration of a purified recombinant protein or RNA and determining at what overall concentration (if any) the molecules undergo a phase transition.
- Scattering methods[176], such as small-angle X-ray scattering, determine the maximum diameter, approximate shape and radius of gyration of particles in solution in order to infer the level of oligomerization between molecules. Another approach is dynamic light scattering, which detects fluctuations of scattered light to calculate the distribution of particle sizes in a solution.

Microscopy

- In situ imaging of DNA, RNA and protein involves the use of fluorescent probes complementary to target RNA or DNA and antibodies against specific proteins to visualize the position of specific RNA, DNA and proteins.
- Multiplexed imaging. Adaptations of fluorescence in situ hybridization such as sequential fluorescence in situ hybridization (seqFISH)[172,177–180], multiplexed error-robust fluorescence in situ hybridization (MERFISH)[181] and Oligopaint[182–184] involve iterative hybridization of probes to thousands of RNA and DNA molecules to enable multiplexed visualization in 3D space. Recent studies using seqFISH+ enable multiplexed DNA, RNA and protein measurements[185].
- Live imaging of the distribution of fluorescently labelled proteins can be used to observe the localization and dynamics of proteins in live cells.
- Single-molecule photobleaching quantifies the stoichiometry of subunits in macromolecular complexes. Single fluorophore molecules display a quantized intensity of decay when photobleached, which can be converted to a precise number of fluorophores within a complex[186].
- Nanocages consisting of a defined number of molecules can be used as an internal standard to quantify concentration within a territory[187].
- Fluorescence correlation spectroscopy uses live-cell microscopy to infer the molecular concentration based on fluctuations of fluorescence intensity in a given location[188].
- Fluorescence recovery after photobleaching measures the extent of molecule exchange at a location by measuring the rate at which labelled proteins recover from photobleaching within that region.
- Super-resolution microscopy techniques such as structured illumination microscopy (SIM), photoactivated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) can visualize molecules beyond the diffraction limit of light (≈250 nm), which is the limitation of conventional microscopy. These techniques have enabled detection of complexes at finer resolution (≈20–50 nm)[189,190].
- Quantitative phase microscopy is a label-free method that measures the shape and concentration of individual condensates by passing light through the sample using different refractive indices, which causes an optical phase delay that can be quantitatively measured[191].

Genomics methods

- Proximity ligation methods (for example, high-throughput chromosome conformation capture (Hi-C)) enable genome-wide detection of pairwise regions that interact in three dimensions[81,192].
- Split-pool recognition of interactions by tag extension (SPRITE)[10] and genome architecture mapping (GAM)[193] are ligation-independent methods that enable genome-wide detection of multiway DNA interactions. RNA–DNA SPRITE is an adaptation that provides genome-wide maps of RNA and DNA contacts in the nucleus[56].
- TSA-Seq utilizes diffusion of free radicals to map preferential association of specific regions around nuclear bodies[140].

Manipulating compartments

- OptoDroplet induces phase transitions of proteins with intrinsically disordered regions (IDRs) upon photoactivation, thereby enabling studies of various features of phase separated structures[169].
- CasDrop enables the formation of liquid condensates at specific genomic loci by recruiting a modified dCas9 that dimerizes upon photoactivation with IDR-containing proteins[170].
- CRISPR genome organization (CRISPR-GO) is an inducible and reversible method that uses ligand-mediated dimerization to direct DNA regions to specific nuclear compartments[168].
- Light-activated dynamic looping (LADL) brings together two genomic sites through light-induced heterodimerization of cryptochrome 2 and a dCas9–CIBN fusion protein[167].

and co-transcriptional RNA processing. In each of these cases, we discuss the key features of nuclear compartmentalization that enable the non-linear behaviours that quantitatively control gene regulation.

Mechanisms of compartment formation

Formation of macromolecular complexes of protein, DNA and RNA molecules has traditionally been viewed through the lens of stoichiometric molecular interactions. These interactions generally occur between well-structured domains that form high-affinity contacts, resulting in complexes containing fixed ratios of molecules (FIG. 1b). Recently, many important regulatory proteins have been shown to also form low-affinity interactions with other protein, DNA or RNA molecules, primarily through multivalent interactions between unstructured, low-complexity domains or intrinsically disordered regions (IDRs). Although individually these weak associations might not enable specific interactions, cooperative association of molecules at high concentrations can facilitate the formation of biomolecular condensates (reviewed extensively elsewhere[17,19,20]). In this section, we discuss how these two molecular mechanisms — stoichiometric molecular interactions and condensate formation — can lead to nuclear compartmentalization.

Complexes with fixed stoichiometry

Macromolecular complexes are formed through interactions that occur between molecules at precise stoichiometric ratios (FIG. 1b). For example, a single Pol II holoenzyme is formed by high-affinity protein–protein interactions between Pol II and general transcription factors such as general transcription factor IIF (TFIIF)[21]. Transcription factors can also bind with high affinity to specific DNA sequences within the major groove of the double helix through DNA-binding domains such as zinc finger domains and leucine zippers[22]. Similarly, high-affinity protein–RNA and RNA–RNA interactions contribute to various macromolecular complexes in the nucleus. For example, various proteins with RNA-binding domains (for example, RNA-recognition motifs[23] and KH domains[24]) bind directly to specific RNA sequences or structures, and specific RNAs

Diffusion and affinity models
Models describing how a molecule (such as a transcription factor) proceeding on a random walk through the nucleus samples many possible binding partners until it finds its high-affinity cognate target site.

Nuclear territories
A catch-all term for 3D regions contained within the nucleus.

Compartments
Nuclear territories enriched in specific DNA, RNA and/or protein molecules.

Mediator complex
A protein complex that facilitates enhancer–promoter interactions, RNA polymerase II (Pol II) loading and transcription initiation.

directly hybridize with other RNA molecules to form macromolecular complexes. A single U4/U6.U5 triple small nuclear ribonuclear protein particle (tri-snRNP) spliceosomal complex is assembled through extensive base pairing between U4 and U6 small nuclear RNAs (snRNAs) and high-affinity interactions between more than 30 proteins, which interact with each other and with snRNAs[25].

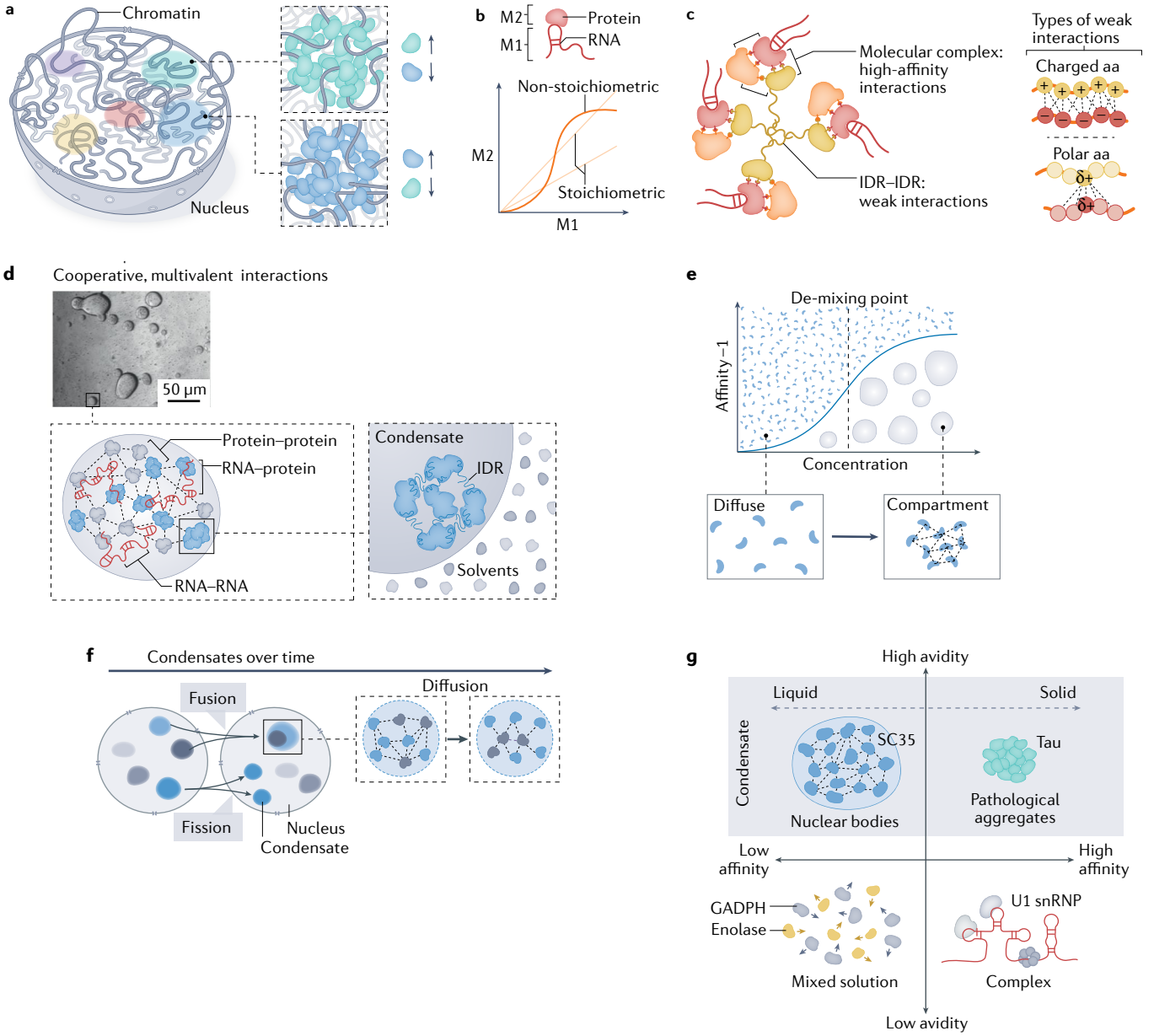

Fig. 1 | **Mechanisms of nuclear compartments formation. a** | Nuclear compartments contain high local concentration of specific molecules in 3D space. Upward arrows represent higher concentration and downward arrows represent lower concentration. **b** | Stoichiometric interactions between two molecules, M1 (an RNA) and M2 (a protein), show a linear increase in the number of M2 molecules bound per unit increase in M1. By contrast, non-stoichiometric interactions lead to an exponential increase in binding of the target molecule (M2) as M1 increases. **c** | Stoichiometric interactions between molecules and cooperative interactions between multivalent intrinsically disordered regions (IDRs) of proteins can be found within a sample of complexes. The low-affinity interactions between IDRs are often mediated by stretches of charged and polar amino acids (aa). **d** | Image of polypyrimidine tract-binding protein 1 (PTBP1) molecules undergoing concentration-dependent condensate formation with RNA in vitro (top). When protein concentration is high, multivalent interactions can promote 'de-mixing' (bottom). **e** | Phase diagram describing whether molecules will be present in a diffusive state or a compartmentalized state. Formation of these structures is controlled by local concentration leading to sharp transition behaviours (blue line). **f** | Condensates can exhibit properties of liquids, including the ability to split (fission), merge (fusion) and undergo rapid internal rearrangement (diffusion). **g** | Cartesian plane highlighting the relationship between affinity, avidity and physical manifestations of different types of molecular associations. Well-mixed solution of two soluble proteins with little to no affinity for one another, such as GAPDH and enolase (bottom left). U1 small nuclear ribonucleoprotein (U1 snRNP) is an example of a macromolecular complex composed of high-affinity interactions between a fixed stoichiometric ratio of interacting component molecules (bottom right). Nuclear bodies are composed of self-interacting proteins, for example, SC35 (also known as SRSF2) proteins in nuclear speckles (top left). Solid aggregates of tau are observed in neurodegenerative diseases (top right). Image courtesy of Inna-Marie Strazhnik/Caltech. Part **d** adapted from REF.[47], Springer Nature Limited.

Stoichiometric interaction
Biochemical interaction that occurs with defined ratios of components, generally with high affinity and specificity, including in protein complexes or the binding of a transcription factor to its DNA motif.

Non-linear regulatory behaviours
Responses to alteration in the reaction rate or efficiency that exceed those expected under a linear model for the underlying changes in the amounts of reactants or catalysts.

Phase transitions
The biophysical process that leads to phase separation, referring to a discontinuous change in the thermodynamic equilibrium state of a system in response to a change in a parameter such as temperature, pressure or molecular concentration.

Multivalent interactions
Molecular associations between multiple binding sites on the interacting molecules; can result in variable stoichiometric ratios.

Intrinsically disordered regions
(IDRs). Protein regions that do not have a single preferred structural conformation.

Biomolecular condensates
Concentration-dependent assemblies of molecules of variable stoichiometry, usually driven by multivalent and cooperative interactions that can form with or without phase separation.

Avidity
(Also known as 'functional affinity'). The collective strength of multiple non-covalent molecular interactions. Avidity represents the overall force conferred by multiple affinities in concert, which exceeds the sum of the strength of those interactions.

Phase separation
Thermodynamically driven partitioning of a homogeneous mixture into locally distinct chemical sub-mixtures (phases) with distinct properties.

Homotypic interactions
Interactions occurring between two or more copies of the same type of molecule.

To date, the vast majority of studied transcriptional and post-transcriptional regulators represent high-affinity macromolecular complexes (FIG. 1c). This is, in large part, because traditional methods for exploring the structure of macromolecular complexes (such as X-ray crystallography) require formation of stable, high-affinity conformations. Despite the great importance of such macromolecular complexes in nuclear functions, many proteins involved in transcription regulation (for example, the Mediator, Pol II and cell type-specific transcription factors), chromatin regulation (Polycomb group proteins, HP1) and RNA processing (serine/arginine-rich splicing factor 1 (SRSF1), FUS and polypyrimidine tract-binding protein 1 (PTBP1)) contain large IDRs, which do not readily form structures and, thus, have traditionally been excluded from attempts to solve the structure of macromolecular complexes[26–28].

Multivalent-interaction condensates

In contrast to macromolecular complexes that occur at fixed stoichiometries, condensates contain molecules that can associate at variable stoichiometries and are spatially enriched relative to the surrounding cellular environment[17–19] (FIG. 1d). Condensate formation is a concentration-dependent process that is often driven by molecules that engage in cooperative, multivalent, low-affinity interactions[29].

Chemistry and soft-matter physics have provided insights into the thermodynamics underlying condensate formation in cells[30] (FIG. 1e). A useful way to conceptualize this is to consider a set of molecules — for example, multiple copies of a particular protein ('A') — mixed with a collection of other molecules — for example, the nucleoplasm ('B'). If the 'A' and 'B' molecules are attracted to one another, a well-mixed solution will form to maximize the entropy of the system. However, if the 'A' molecules exhibit preferential molecular attraction to other 'A' molecules compared with 'B' molecules, formation of these preferential 'A' molecular interactions will be more energetically favourable than random mixing of 'A' and 'B' molecules. Although molecules with weak-affinity individual interactions (for example, forming at micromolar concentrations)[31] might not be able to promote energetically favourable association, interactions that occur at multiple independent sites (multivalency) lead to a large increase in the overall affinity — referred to as avidity[32] — between molecules. When the concentration of 'A' molecules achieves a critical threshold, the self-interacting 'A' molecules will 'de-mix' from the 'B' molecules (FIG. 1e), a process referred to as concentration-dependent phase separation. The thermodynamics of multivalency[19,32], the physics of phase separation[18,30] and their features and limitations in biological processes are explored in more detail in other reviews[18–20,33,34]. In this Review, we use the term condensate to refer to molecular assemblies that are formed through concentration-dependent, multivalent associations regardless of whether they undergo phase separation.

Many of the proteins that are important for promoting condensate formation contain large IDRs that are often composed of charged and polar amino acids, which can facilitate low-affinity interactions[35] (FIG. 1c). The charge properties of these IDRs combined with their length allow them to engage in multivalent, low-affinity homotypic interactions and heterotypic interactions. Because multivalent binding can induce a multiplicative increase in the overall avidity between two molecules, systematically increasing the number of possible interactions that a given molecule can form (valency) enables association and de-mixing to occur at lower overall molecular concentrations[18,31] (FIG. 1d,e). Importantly, such concentration-dependent assemblies can form through homotypic or heterotypic multivalent associations of RNA, DNA and protein molecules[36,37].

Condensates and phase separation

Molecules can undergo phase separation to form condensates that exist in different physical states including liquid, solid and intermediate ('gel-like') states (FIG. 1f). The key to formation of the liquid state involves rapid molecular motion, resulting in the breaking and re-formation of individually weak interactions within the condensate. Owing to the low affinity of individual interactions, proteins can associate with and disassociate from other molecules in the condensate. By contrast, if there is a strong affinity between individual interacting molecules, high avidity can still be achieved, but it will restrict diffusion within the condensate and result in a more rigid, solid-like structure. Because affinity is a continuous property, the precise physical states of a condensate can be thought of as a continuum across the liquid to solid range (FIG. 1g).

The specific phenomenon of phase separation that forms liquid-like condensates is known as liquid–liquid phase separation (LLPS)[18]. One of the first discovered examples of LLPS-mediated condensate formation was P granules in *Caenorhabditis elegans*[38]. P granules form condensates that exhibit several liquid-like properties, including the ability to undergo fusion, fission and diffusion[38] (FIG. 1f) and displaying a spherical shape that reflects its surface tension with the surrounding cytoplasm. By virtue of its liquid-like properties, the molecules contained within an LLPS compartment are expected to be well mixed and uniformly distributed[33].

Several such membraneless compartments in the nucleus have been shown to form liquid-like structures, including the nucleolus[39]. In addition, many proteins involved in gene regulation have been proposed to undergo LLPS. These include chromatin regulatory proteins (HP1 (REFS[13,14]) and Polycomb proteins[40–42]), transcription machinery proteins (Mediator[11,43], Pol II[11,43], TAZ[44] and bromodomain-containing protein 4 (BRD4)[45]), transcription factors (OCT4 (REF.[12]) and TAF15 (REF.[46])), splicing factors (SRSF1 and SRSF2 (REF.[43])) and RNA processing factors (such as PTBP1)[31,47].

Seeding nuclear compartmentalization

In order to form a nuclear compartment, molecules need to achieve high concentrations within a spatially constrained territory of the nucleus. Because nuclear proteins are translated in the cytoplasm and then trafficked into the nucleus, they are intrinsically diffusible — they

Heterotypic interactions
Interactions occurring between at least two molecules of different types.

need to diffuse through the nucleus to associate with their targets. In theory, proteins that form condensates could stochastically come together through molecular diffusion at the concentrations needed to undergo condensate formation. However, this is likely a rare event because the general concentrations of individual nuclear proteins are often too low to allow for the simultaneous associations of the required multiple proteins[31]. Instead, many nuclear compartments are 'seeded' through high-affinity interactions with spatially constrained molecules within the nucleus (a process known as nucleation). These can include interactions with nucleic acids, histone modifications or existing nuclear structures such as the nuclear lamina[48,49]. These spatially constrained molecules can bind to various diffusible molecules and recruit them to specific nuclear territories to seed the formation of nuclear compartments (FIG. 2a).

Many nuclear structures can form through interactions with non-coding RNAs (ncRNAs): the nucleolus is seeded by transcription of pre-ribosomal RNA[50,51], paraspeckles are seeded by the long ncRNA nuclear paraspeckle assembly transcript 1 (*NEAT1*)[52], histone locus bodies are seeded around transcription of histone pre-mRNAs[53] and the Barr body (inactive X chromosome (Xi)) is seeded by the long ncRNA X-inactive specific transcript (*XIST*)[8,54]. Moreover, many chromatin regulators form spatial compartments in the nucleus that are dependent on RNA, for example, HP1 (REFS[55–57]), SHARP[56] and Polycomb proteins[57,58]. Reported in a recent preprint[56], live cell imaging of SHARP — an RNA-binding protein that recruits histone deacetylase 3 (HDAC3) to chromatin[59] — revealed dozens of condensate-like structures throughout the nucleus, which become diffuse upon deletion of its RNA-binding domain.

ncRNAs are especially well suited to seed nuclear compartment formation for two reasons. First, transcription creates high concentrations of spatially constrained RNA. Once transcribed, a ncRNA can be retained at high concentration on chromatin near its site of transcription, thereby achieving higher local concentrations (relative to DNA) at these specific nuclear locations. Thus, a nuclear compartment can be dynamically created by controlling the expression of a specific ncRNA (FIG. 2a). Second, these spatially constrained ncRNAs contain sequence motifs and secondary structures that can bind diffusible RNA and protein molecules, thereby forming stoichiometric interactions that drive high local concentrations of these diffusible molecules. In specific cases, formation of these territories with high local concentrations can further promote concentration-dependent condensate formation through homotypic and heterotypic interactions[51,56,60] (FIG. 2a). Interestingly, several ncRNAs contain multiple binding sites for the same protein, which can further increase the valency of these interactions and local molecule concentrations[47,61–63] (FIG. 2b).

The above-mentioned study also showed that hundreds of ncRNAs can form high-concentration and stable territories near their site of transcription, and therefore might similarly serve to seed the formation of various nuclear compartments of different sizes[56] (FIG. 2c).

In addition to ncRNAs, genomic DNA can also drive high spatial concentrations of protein complexes in the nucleus. For example, genomic regions containing a high density of enhancers, each individually binding to Mediator, can drive high local concentrations and condensate formation[34]. Similarly, multiple DNA and histone modifications can efficiently recruit and drive

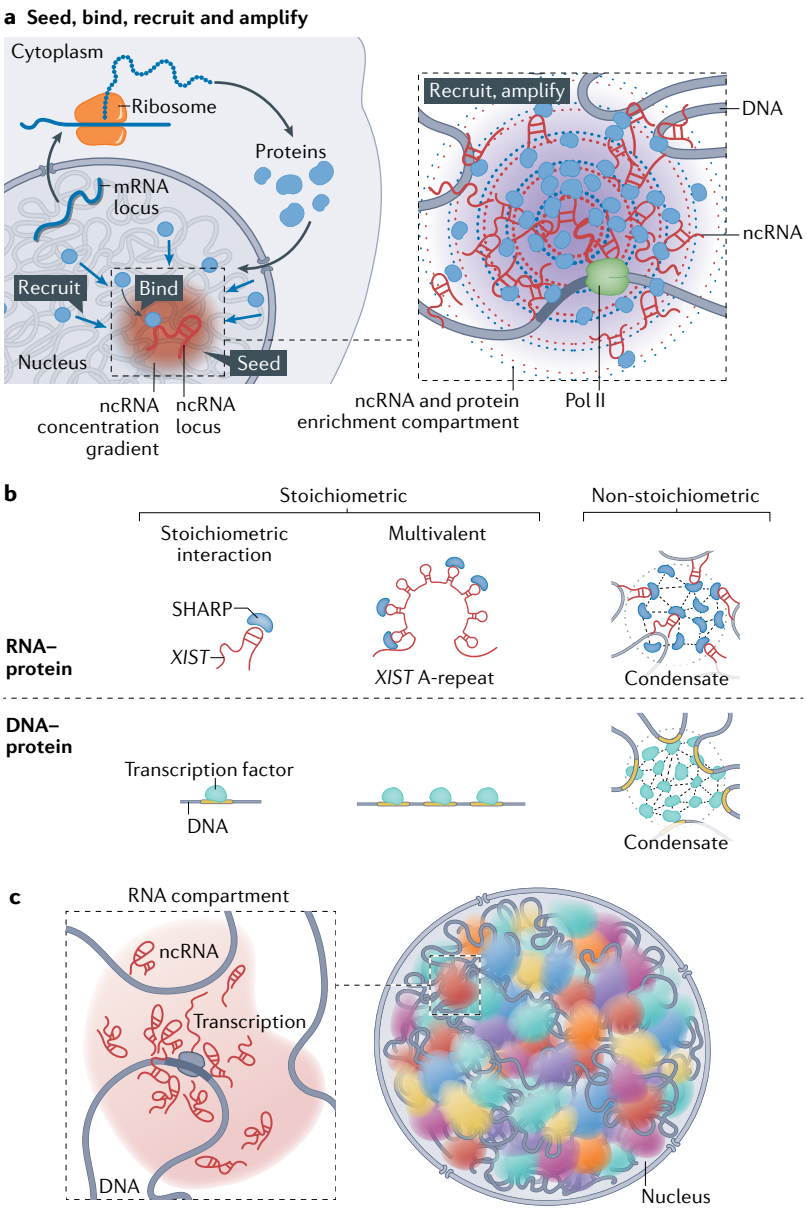

Fig. 2 | **Spatially constrained non-coding RNAs can drive compartmentalization in the nucleus. a** | Model of how non-coding RNAs (ncRNAs) can drive compartmentalization. Sites of ncRNA transcription can support high ncRNA concentrations and, thus, seed the formation of high-concentration territories (seed); ncRNAs can bind to diffusible proteins or ncRNAs through stoichiometric interactions (bind) and, in this manner, lead to enrichment of diffusible factors within a spatially defined region (recruit). In some cases, the recruited proteins can recruit other proteins and/or form condensates through homotypic and heterotypic interactions (not shown). **b** | Stoichiometric RNA–protein or DNA–protein interactions can involve single or multiple binding events, but the number of proteins recruited is limited to the number of available sites on the nucleic acid. By contrast, non-stoichiometric interactions enable binding of more proteins than available binding sites. **c** | Space filling model of localization of hundreds of ncRNAs (coloured areas) across the nucleus in mouse embryonic stem cells. Pol II, polymerase II; *XIST*, X-inactive specific transcript. Image courtesy of Inna-Marie Strazhnik/Caltech.

Box 2 | **How common is 'phase separation' as a functional mechanism in cells?**

Phase separation is a physical process by which molecules can separate from their environment. For example, crystallization of soluble proteins and RNA molecules is a type of phase separation used by structural biologists to solve structures of biomolecules. Recently, numerous examples of another type of phase separation, referred to as liquid–liquid phase separation (LLPS), have been described in the context of cell biology, with important implications for how we think about the formation of membraneless structures. Phase separation has become the focus of intense research, leading to the description of an increasingly large number of structures as 'phase-separated' compartments. Although this physical process is useful for thinking about how molecules can behave in cells, whether many of these compartments form through phase separation or through other mechanisms of cooperative multivalent association is not clear in most studies.

A recent perspective article describes experimental issues in current studies and the challenges of defining whether a compartment indeed forms through phase separation within a cell[33]. The main problem is that most studies of phase separation rely on in vitro experiments or on analysis of tagged proteins that are overexpressed in cells. Given that phase separation is a concentration-dependent process, such experimental systems may lead to behaviours that do not accurately reflect how these molecules assemble in their endogenous contexts within a cell. The authors argue that many of the properties often ascribed to liquid-like condensates (spherical shape, fusion and fission, and rapid diffusion) are not specific to liquid-like structures. In support of their opinion, they highlight an example of a compartmentalized structure formed upon infection of cells with the herpes simplex virus, which displays the three key features ascribed to liquid-like condensates despite the fact that it does not form through phase separation[70]. The authors propose two additional criteria for defining LLPS within a cell: that the molecules are well mixed within the compartment, and demonstrating that there is a change in the diffusion rate of molecules across the boundary of the compartment. These two features are often challenging to measure with current methods, especially for small compartments such as transcriptional condensates. Given these technical challenges, the precise biophysical properties of most nuclear compartments in vivo remain largely uncharacterized.

As more details are revealed about the various biophysical properties of nuclear compartments and their mechanisms of formation, we expect that the range of terminology will expand dramatically in order to capture the true range of these processes and their importance in biology. Despite the imperfect nature of our current terminology, these different considerations do not directly affect the central defining feature of a nuclear compartment, namely the enrichment of specific molecules within a specific territory.

high concentrations of modification 'reader' proteins. Interestingly, several different readers can form condensates in both normal and disease contexts, including chromobox protein homologue 2 (CBX2)[41,42], HP1 (REFS[13,14,64]) and methyl-CpG-binding domain protein 2 (MBD2)[65–67]. The presence of numerous protein binding sites on DNA in close proximity ('spatial valency') may promote protein concentrations that are needed to drive condensate formation (FIG. 2b).

These shared features — spatial anchoring in 3D space and the ability to bind and recruit diffusible molecules into high-concentration territories — appear to be essential for seeding the formation of many nuclear structures.

Compartmentalization mechanisms in vivo

In this section, we present various requirements for formation of nuclear compartments, including the ability of molecules to form stoichiometric interactions, the ability to form non-stoichiometric assemblies (condensates) and the ability of anchored molecules to recruit diffusible molecules into precise nuclear territories. Importantly, these mechanisms are not mutually exclusive and assembly of nuclear compartments in cells likely involves combinations of these and, possibly, other mechanisms.

We note that the terminology can be confusing and used differently by different research groups (BOX 2). For example, the term 'phase separation' is sometimes used to refer specifically to LLPS[33] and sometimes more broadly to describe a process by which condensates are formed[18–20]. As described above, LLPS, phase separation and condensate formation are distinct terms representing distinct chemical and physical behaviours. Moreover, it is still quite challenging to specifically demonstrate phase separation in vivo and most descriptions are based on in vitro studies. Accordingly, the precise biophysical properties of most nuclear compartments are still unknown. To avoid ambiguity, we use the more general term 'condensate', except where the more specific biophysical properties of a compartment are well characterized and crucial for describing its function.

Importantly, condensate formation is not the only mechanism by which nuclear compartments can be formed. Alternative mechanisms include the ability for diffusible proteins to form stoichiometric molecular interactions with spatially anchored molecules such as ncRNA, DNA and histones. For example, initiation of X-chromosome inactivation (XCI) requires the recruitment of the RNA-binding protein SHARP to the Xi. This occurs through expression and retention of *XIST* on the Xi[68], which is followed by recruitment of SHARP through a stoichiometric interaction between its RNA-recognition motifs and *XIST*[69]. Similarly, upon herpes simplex virus infection, cells form a 'replication compartment' in the nucleus that shows strong enrichment of Pol II. This enrichment is driven by preferential binding of Pol II to the viral DNA contained within the compartment[70].

Facilitating transcription regulation

Biological processes as diverse as cell differentiation, response to environmental cues and innate immunity all depend on the ability of a cell to rapidly and specifically activate or repress transcription of specific genes[71–73]. Transcription regulation involves numerous cell type-specific transcription factors that bind to DNA regulatory elements, including promoters and enhancers, through high-affinity (stoichiometric) interactions with defined DNA sequences. The Mediator complex binds to enhancers, brings them near promoters and facilitates Pol II loading during transcription initiation[74]. Despite important progress in deciphering the molecular mechanisms of transcription regulation, current models largely rely on stoichiometric molecular interactions and cannot fully explain many of the quantitative features of transcription regulation. Here, we discuss recent data that establish an important role for nuclear compartmentalization in transcription initiation, and describe how compartmentalization may enable many of the important quantitative features of gene regulation.

Transcriptional condensates

The discovery of enhancers in animal genomes initially posed a challenge for classical genetics, because it was not obvious how a DNA element tens or hundreds of kilobases away from a promoter could influence the activity of its target[75]. Early studies of the mouse Sonic

Valency
The number of non-covalent interactions with other molecules that a single molecule or domain can support.

Liquid–liquid phase separation
(LLPS). A specific form of phase separation defined by the formation of a liquid compartment within a larger liquid environment.

Granules
Small (<1 μm) condensates that generally have a simple composition compared with the larger nuclear bodies.

Diffusible
Refers to molecules that proceed on a random walk throughout the volume that contains them.

Constrained molecules
Molecules that proceed on a random walk preferentially within a sub-volume of their overall environment, often owing to having high affinity to other molecules in that sub-volume.

Bodies
In the nucleus, large (≥1 μm), functionally distinct territories, often involved in molecular biogenesis, such as the nucleolus (ribosome biogenesis) and Cajal bodies (biogenesis of small nuclear RNAs (snRNAs)).

hedgehog locus (*Shh*) using 3D imaging provided initial evidence that enhancers come into close physical proximity with their target genes[76]. This led to a prevailing model, in which enhancers and promoters form direct interactions through chromatin looping. Yet this looping model cannot fully explain long-range enhancer function, because a promoter can be simultaneously regulated by multiple enhancers[7] and, in specific cases, the spatial distance between enhancers and their promoters can increase rather than decrease upon transcription activation[77,78]. Recent observations that Pol II, Mediator and multiple enhancer regions can associate within 'transcriptional condensates' provide a possible explanation for many of these previous observations[11,45,79,80].

Enhancers and their targets coexist in topologically associating domains. Enhancer–promoter interactions generally occur within chromosomal structures known as topologically associating domains (TADs)[8,81,82]. A TAD is a 3D structure consisting of a large genomic region (generally on the order of hundreds of kilobases of DNA), in which DNA sequences interact more frequently with each other than with linearly neighbouring DNA outside the TAD, often owing to chromatin loop formation and the presence of insulators (FIG. 3a). TADs may represent the functional unit of promoter–enhancer interactions, because altering the linear distance between *Shh* and its enhancer within a TAD has only a modest effect on *Shh* expression, but disrupting the TAD boundaries substantially decreases *Shh* expression[7,83]. Many pathologies are thought to occur as a result of aberrant enhancer–promoter contacts due to loss of TAD boundaries. For example, TAD boundary disruption is thought to allow enhancers to aberrantly activate oncogenes in certain cancers[84,85]. Although some TADs are essential facilitators of specific

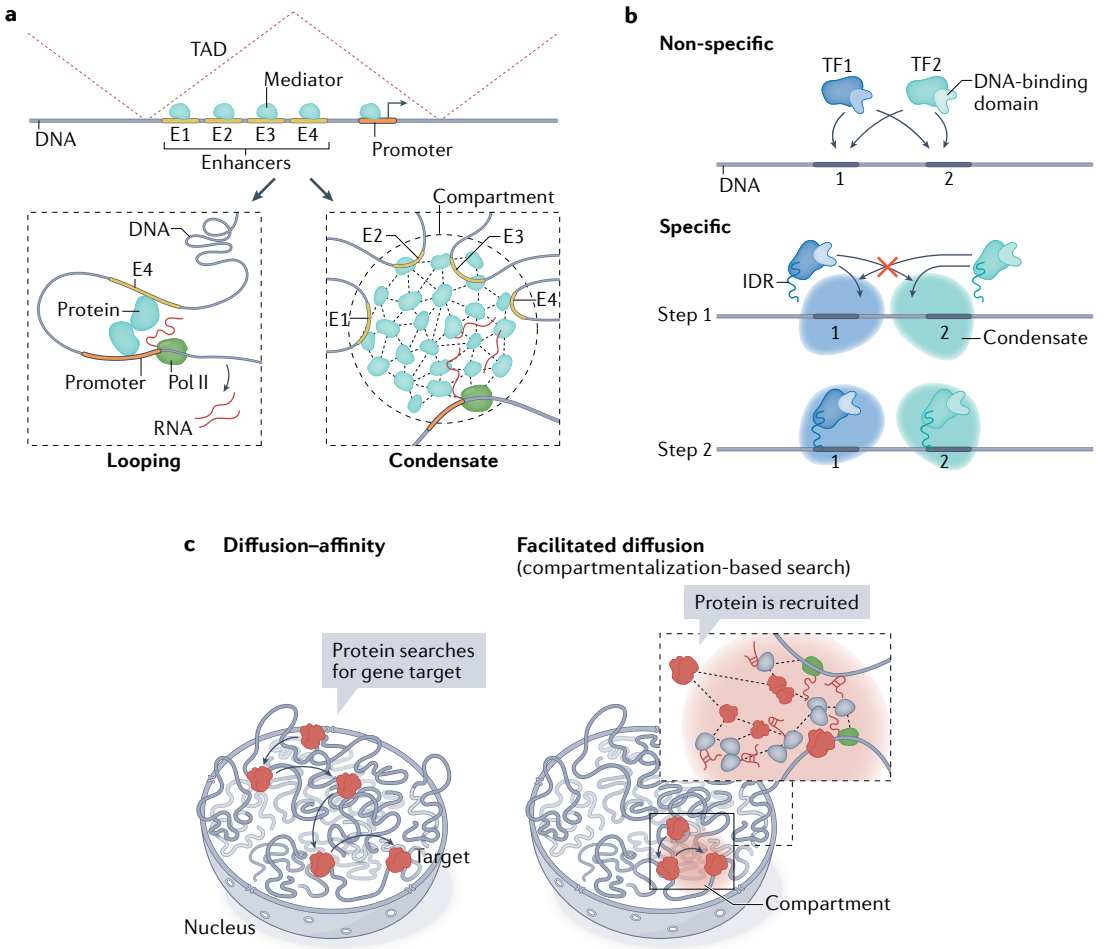

Fig. 3 | **Enhancers, promoters and transcription factors can form condensates that may facilitate rapid target search in the genome. a** | Example of multiple enhancers (super-enhancer) confined to a topologically associating domain (TAD; dotted red line) that contains the promoter they regulate. Comparison of a simple chromatin looping model of enhancers and promoters (left) with a condensate model (right) shows how multiple enhancers can occupy the same territory by forming cooperative interactions. **b** | Two distinct transcription factors (TF1 and TF2) can have the same affinity for the same DNA sequences, but occupy different genomic locations. Transcription factors lacking intrinsically disordered regions (IDRs) cannot distinguish between these high-affinity sites ('non-specific'), but transcription factors containing IDRs are targeted to specific genomic regions ('specific'). **c** | Comparison of diffusion–affinity models that involve stochastic movement of proteins to target sites (left) with the facilitated diffusion model, which involves a combination of 3D movement with sliding (right). Pol II, polymerase II. Image courtesy of Inna-Marie Strazhnik/Caltech. Part **b** adapted with permission from REF.[11], AAAS.

Transcriptional condensates
Assemblies of general transcription factors and Mediator complexes around enhancers and promoters that facilitate transcription activation.

Facilitated diffusion
In the context of compartmentalization, the process by which compartments restrict the random walks of diffusible molecules to a smaller volume. For example, constraining the diffusion of a transcription factor to a small nuclear volume around its target.

enhancer–promoter interactions, others appear to be dispensable for gene regulation, indicating that not all TADs work in this way[86].

Mediator and Pol II cluster at super-enhancers. The expression of many essential genes — such as those encoding pluripotency factors in embryonic stem cells, the gene encoding the transcription factor PU.1 in B cell progenitors and *MYC* in multiple myeloma cells — is controlled by many enhancers[87–89]. Such groups of enhancers that control individual target genes (often involved in specifying cell identity) are referred to as 'super-enhancers', and are defined as genomic regions with high concentration of Mediator complexes and a large fraction of enhancer-associated Pol II within the cell[88].

Early imaging studies suggested that multiple molecules of Pol II localize in specific regions of the nucleus ('Pol II clusters')[90–94], but because this work relied on diffraction-limited microscopy, the precise spatial organization of Pol II was difficult to assess. The advent of super-resolution microscopy dramatically improved the ability to directly visualize individual molecules of Pol II in the nucleus[95]. Super-resolution studies show that Pol II and Mediator co-occupy genomic regions, forming both large and small clusters in the nucleus[11]. The large Pol II–Mediator clusters (which make up <10% of all Pol II–Mediator clusters) were proposed to be the previously described super-enhancer-containing DNA loci[96]. Indeed, linking these Pol II–Mediator clusters with specific DNA loci (using immunofluorescence and DNA fluorescence in situ hybridization) demonstrated that these large clusters tend to overlap with super-enhancer-containing loci[43]. These Pol II and Mediator clusters were shown to form through interactions mediated by their IDRs in a concentration-dependent manner[12,80], and as such are referred to as transcriptional condensates.

Condensates can increase the concentration of transcription regulation proteins on DNA. The high concentration of Pol II–Mediator at super-enhancer-containing genes suggests that transcriptional condensates act to increase the concentration of Pol II at highly regulated genes (FIG. 3a). Indeed, live cell super-resolution imaging at specific loci has shown that the level of nascent transcription is directly correlated with the size of the Pol II condensate[97]. These data suggest that super-enhancers may act to compartmentalize Mediator around target promoters and may facilitate Pol II loading. Similarly, transcription factors containing IDRs can be recruited to genomic regions with a high density of their binding sites, to form a compartment that achieves higher transcription factor concentrations than could be supported by the number of binding sites present. For example, the EWS–FLI1 fusion transcription factor can bind to regions of repetitive DNA and cluster in the nucleus at concentrations that exceed the number of available cognate DNA binding sites. In this way, EWS–FLI1 can promote robust expression of genes at these sites[79].

The ability of Mediator, Pol II and specific transcription factors to undergo concentration-dependent condensate formation may increase their concentration beyond the stoichiometric concentration that can be achieved by binding to individual DNA binding sites.

Condensates may enable target search

Cell type-specific transcription factors diffuse through the nucleus and form high-affinity interactions with their cognate DNA binding sequence. Although this diffusion and affinity model can explain many of the qualitative aspects of gene regulation, that is, where on the DNA transcription factors bind, measurements in bacteria demonstrate that the observed rates of association of the Lac repressor to its target DNA site are considerably faster than would be predicted by molecular diffusion and DNA binding affinity only[98]. This discrepancy is even larger in eukaryotes, where transcription factor binding to chromatin is vastly more complex[1]. For example, it would take several hundred hours for a specific transcription factor to identify a single binding site in the nucleus using diffusion and affinity alone[2]; yet transcription factors can dynamically localize and induce transcription within minutes of stimulation in many distinct contexts[72]. In addition, not all high-affinity DNA binding sites are occupied by a transcription factor, indicating that there are parameters beyond DNA binding affinity that are important for controlling target recognition.

To address this quantitative challenge, a biophysical model referred to as facilitated diffusion was proposed[1,98]. This model suggests that transcription factors identify their targets in two steps: rapid 3D diffusion to identify genomic neighbourhoods, and slow scanning along the DNA within a neighbourhood to identify high-affinity targets[99,100]. Recent studies suggest that the ability of transcription factors — and other DNA binding proteins — to form condensates might be a crucial component of this facilitated search[2].

Specifically, the IDRs of transcription factors are often required for specific DNA localization within the genome[101]. For example, two distinct transcription factors in yeast, Msn2 and Yap1, have DNA-binding domains with affinity for the same DNA sequences but occupy different genomic locations. This genomic specificity is encoded not by the DNA-binding domains of these proteins but by their IDRs (FIG. 3b). Indeed, following deletion of their IDRs, Msn2 and Yap1 can still bind to DNA, but they no longer achieve specificity to their correct genomic targets. Instead, mutant Msn2 and Yap1 bind equally well to both sets of binding sites. This example demonstrates that specificity of transcription factor localization can be achieved through the same unstructured regions required for condensate formation.

Gene dosage compensation provides an example where condensate formation is required to define the specificity of a DNA-binding protein[102]. In fruit flies, X chromosome dosage compensation entails recruitment of the gene activating complex male sex lethal (MSL) exclusively to the promoters of genes on the single male X chromosome, even though this DNA-binding complex also has high-affinity DNA-binding sites on autosomes. How MSL achieves specific localization at DNA

regions on the X chromosome had been a long-standing question. A recent study demonstrated that the IDR within MSL2 is required for X chromosome-specific targeting of this complex[102]. Importantly, specific localization on the X chromosome is mediated by an interaction with the roX ncRNA (which is transcribed from the X chromosome and is enriched on it), whose function is to seed a concentration-dependent condensate through homotypic interactions between the IDRs of MSL2.

Together, these results indicate that DNA-binding proteins can use their IDRs to achieve localization specificity by creating high local concentration in specific nuclear regions. In this way, compartmentalization may reduce the range of possible targets in the nucleus, thereby accelerating target search (FIG. 3c). In addition to the specific examples described above, many DNA-binding proteins contain IDRs that can form condensates in the nucleus[12,45], suggesting that this may be a more general mechanism by which DNA-binding proteins achieve targeting specificity.

Higher-order chromatin regulation

Chromatin is spatially organized within the nucleus into active and inactive territories[4]. Post-translational modifications of histone proteins (chromatin modifications) can change the accessibility and spatial organization of large stretches of chromatin rather than of individual genes. These chromatin landscapes can form higher-order 3D structures that share transcriptional states. For example, large, actively transcribed regions form higher-order interactions with other transcribed regions whereas heterochromatin forms 3D structures that can contain DNA from different chromosomes[10,103–105]. In this section, we discuss emerging principles of compartmentalization in higher-order chromatin packaging and coordinated regulation of multiple genes.

Condensates on the Xi

One example of chromatin-mediated multigene regulation is XCI. XCI is an essential developmental process, in which one of the two X chromosomes in the cells of female mammals is silenced to ensure dosage balance of X-linked gene expression between males and females[106]. XCI has emerged as a paradigm for understanding the relationship between nuclear organization, epigenetics and gene regulation[106,107].

XIST initiates a cascade of events that leads to XCI[59,108,109]; it binds to the SHARP–SMRT–HDAC3 complex[59,110,111] to evict Pol II from the X chromosome[68]. Although ensemble measurements across a cell population show that *XIST* localizes over the entire X chromosome[68], its expression levels are not high enough for it to occupy the entire Xi within an individual cell. Specifically, there are ~200 copies of *XIST* in each cell[60], which would need to bind to >167 million base pairs to cover the entire Xi (~1 *XIST* molecule per 1 million base pairs of DNA). Thus, a direct stoichiometric interaction alone cannot explain how the *XIST*–SHARP complex leads to irreversible silencing of the entire X chromosome during mammalian development.

Spatial amplification of regulatory proteins. Recently, a super-resolution microscopy study reported in a preprint[60] showed that the concentration of SHARP on the Xi exceeds the concentration of *XIST*, suggesting that SHARP is recruited to the Xi in a super-stoichiometric manner. *XIST* can form direct, high-affinity stoichiometric interactions with the SHARP RNA-recognition motifs, and these interactions are required to recruit SHARP to the Xi. SHARP also contains a large IDR that is dispensable for binding to Xist and to SMRT–HDAC3 (REF.[69]), but is essential for super-stoichiometric recruitment of SHARP to the Xi[60].

One mechanism to achieve such super-stoichiometric recruitment in an IDR-dependent manner is through the establishment of a *XIST*-mediated condensate that leads to high spatial concentration of the SHARP–SMRT–HDAC3 silencing complex. Specifically, once SHARP achieves a sufficiently high concentration on the Xi through stoichiometric interaction with *XIST*, it can undergo a concentration-dependent transition to enable multiple SHARP molecules to form multivalent interactions with each other, independently of direct binding to *XIST*[60]. In this way, condensate formation may act to amplify the number of repressive SHARP proteins on the Xi relative to *XIST* molecules and may enable transcriptional silencing independently of a direct molecular interaction with *XIST* (FIG. 4a).

Concentration thresholds in gene expression. The principle that local concentrations of RNA and protein determine the activity state of a gene may also explain why transcriptional gene silencing is restricted to a single X chromosome during XCI. Although *XIST* is predominantly retained on the X chromosome from which it is expressed, it can also diffuse to autosomal genes and accumulate there at a lower concentration[47,68]. Despite its occasional presence on autosomes, *XIST* does not appear to repress autosomal gene expression. However, when the entire X-inactivation centre is moved to an autosome, *XIST* is capable of silencing that chromosome[112]. This indicates that autosomes are not intrinsically resistant to repression by *XIST*, but rather that the concentration of *XIST* required to repress transcription is not achieved on autosomes in wild-type cells. These long-observed phenotypes may be explained by a requirement for a critical concentration of *XIST* on a chromosomal region in order to recruit silencing-competent SHARP–HDAC3 complexes (FIG. 4b). In this way, it can achieve silencing of the X chromosome (where it is present at its highest levels), but avoid silencing autosomal regions (where *XIST* never reaches this critical threshold).

Compartmentalization may help establish physical boundaries for biological processes that require high concentrations of effectors. For example, in mice, the long ncRNA *Kcnq1ot1* diffuses continuously from its locus and exhibits volumetric decay, with the highest concentration around its own locus and decreased concentration at regions farther away[56] (FIG. 4b). Similar to *XIST*, *Kcnq1ot1* also binds directly to SHARP[56] and represses transcription through the HDAC3 complex[56,113] (FIG. 4c). However, despite its

Super-stoichiometric
Refers to interactions that occur without fixed ratios of components and can exceed the binding capacity of any individual molecule, often involving high-avidity, multivalent interactions.

X-inactivation centre
A region on the X chromosome containing X-inactive specific transcript (*XIST*) and its *cis*-regulators; necessary and sufficient for initiation of X-chromosome inactivation (XCI) and protected from it to allow continual expression of *XIST*.

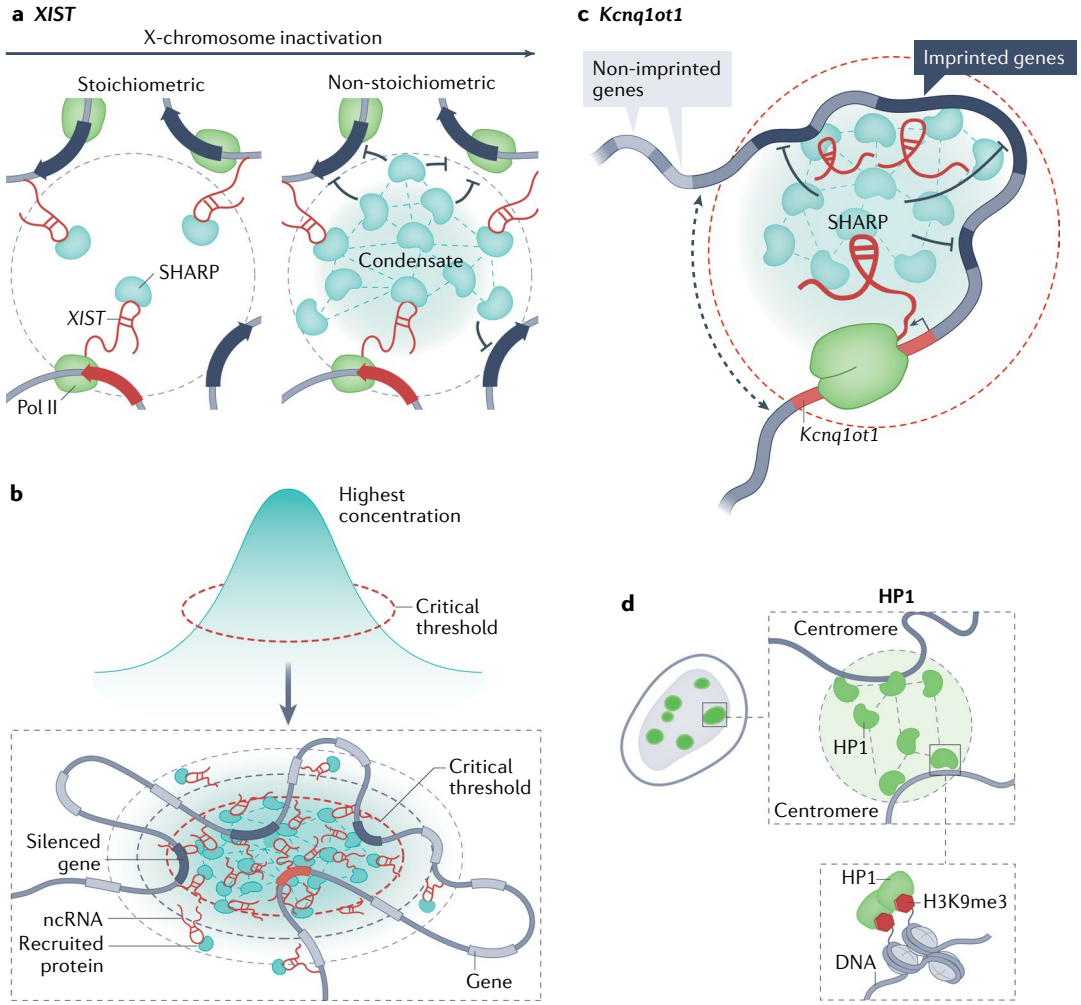

Fig. 4 | **Compartmentalization and chromatin regulation. a** | Stoichiometric interactions occur at defined ratios of components, whereas non-stoichiometric interactions occur without fixed ratios and can exceed the binding capacity of any individual molecule. Increasing the concentration of the long non-coding RNA (ncRNA) X-inactive specific transcript (*XIST*) along the X chromosome increases in the concentration of the *XIST*-binding protein SHARP, which eventually results is silencing of most genes on the chromosome. Spatial amplification of SHARP (which exceeds *XIST* concentration) requires its intrinsically disordered regions (IDRs) and may be achieved through concentration-dependent condensate formation. **b** | Concentration thresholds may explain the 'switch-like' control of gene repression within and outside ncRNA compartments. ncRNAs can achieve their highest concentration at the centre of a spatial territory. If the ncRNA concentration at the central area reaches a critical threshold, enough proteins to mediate compartment formation and to achieve gene repression can be recruited. In this model, ncRNA molecules can still diffuse outside the silencing compartment, but not at levels high enough to recruit sufficient regulatory molecules to change gene expression. As a result, the observed effect on gene expression is binary: 'on' within the compartment and 'off' outside it. **c** | Schematic of the *Kcnq1ot1* compartment, which also recruits SHARP to silence imprinted genes on the paternal allele. Dashed arrow represents a cohesin complex promoting the formation of the *Kcnq1ot1* compartment. **d** | Schematic showing heterochromatin protein 1 (HP1)-mediated compartmentalization of centromeric heterochromatin, which is marked by trimethylated histone H3 Lys9 (H3K9me3). Pol II, polymerase II. Image courtesy of Inna-Marie Strazhnik/Caltech.

diffusive properties, the repressive effect of *Kcnq1ot1* is confined to a topological domain that includes its target imprinted genes (where *Kcnq1ot1* is present at the highest concentration)[56,114,115]. Rather than displaying continuous silencing effects, that is, genes being proportionally repressed relative to the *Kcnq1ot1* concentration, the effects of *Kcnq1ot1* appear to be more deterministic and switch-like. This might be explained if a minimum concentration of the *Kcnq1ot1*–SHARP complex is required to induce phase separation to silence transcription (FIG. 4b).

Compartmentalization and epigenetics

Chromatin regulation is often characterized by long-lasting regulatory effects. For example, after initiation of XCI, the Xi remains silenced even upon loss of *XIST*[112], and the expression of chromatin regulators such as HP1, which mediates heterochromatin formation and maintenance, tends to remain stable across many cell divisions[116,117]. Here, we discuss emerging evidence that spatial organization can have an important role in establishing epigenetic memory in specific cases.

Epigenetic memory
The set of DNA and histone modifications that are heritable either from parents to offspring or from a mother cell to daughter cells.

XIST ***establishes a nuclear condensate that is required for persistent and stable silencing.*** After a critical development time point, *XIST* induces stable and epigenetically heritable chromosome-wide transcriptional gene silencing on the Xi, which is maintained even in the absence of *XIST*[112]. *XIST* expression leads to the recruitment of numerous chromatin regulatory proteins, including Polycomb repressive complex 2 (PRC2)[59,118,119], and deposition of DNA methylation (of CpG cytosines)[120]. In addition, the Xi undergoes large-scale structural changes, and several of the proteins that are recruited to the Xi can undergo concentration-dependent condensate formation[106,121].

Recent evidence suggests that compartmentalization may be essential for the maintenance of XCI. The RNA-binding protein PTBP1 interacts with the E-repeat region of *XIST*, which contains more than 50 PTBP1 binding sites[47]. PTBP1 has been studied largely because of its role as a splicing factor and was one of the first nuclear proteins shown to undergo concentration-dependent condensate formation[31]. Consistent with this observation, PTBP1 forms a condensate structure on the X chromosome that is dependent on its interaction with *XIST*. Intriguingly, deletion of PTBP1 (or of its binding sites) does not affect the initiation of transcriptional gene silencing on the X chromosome (in contrast to deletion of crucial initiation proteins such as SHARP)[47,59]. However, over time, the frequency at which PTBP1 mutant cells 'reactivate' expression from the Xi is considerably higher than in wild-type cells[47], suggesting that PTBP1 is important for maintaining silencing on the Xi. In this model, once initiation of silencing is complete and a condensate is formed, the silenced state is maintained through self-interactions between proteins within the condensate and, therefore, no longer requires *XIST* expression.

Chromatin regulators may seed the re-establishment of nuclear condensates. To have a role in the maintenance of gene silencing, the *XIST*-initiated condensate must be propagated across cell divisions. Although it remains unknown how this propagation might occur, DNA methylation and various histone modifications are stably associated with the Xi through cell division[107,119,122]. These modifications may seed the re-establishment of the silencing compartment following mitosis. For example, trimethylated histone H3 Lys27 (H3K27me3; a gene-repressive histone modification) recruits the PRC1 complex[123–125], which was shown to form concentration-dependent condensates. Specifically, the CBX2 component of PRC1 contains a chromodomain that binds with high affinity to modified histones, an IDR that can form condensates, and domains that are required for binding and recruiting other PRC1 or PRC2 regulatory proteins that reinforce gene silencing[41]. In this way, PRC1 may be able to re-establish a nuclear compartment following cell division.

In addition to the Xi, Polycomb proteins can form eponymous high-concentration territories in the nucleus. These 'Polycomb bodies' contain loci that are located at distinct linear locations on a chromosome or on different chromosomes[126]. For example, two Hox gene clusters in *Drosophila melanogaster*, ANT-C and BX-C, co-localize in Polycomb bodies despite being separated by 10 Mb[127]. Recently, CBX2 was shown to form condensates and is required for the formation of Polycomb bodies[41,128]. Other Polycomb complex proteins are also contained within these condensates. Interestingly, the IDR of CBX2 that is required for condensate formation is also required for PRC1-mediated chromatin compaction, suggesting that chromatin compaction might be a product of condensate formation[41].

Other chromatin proteins can also form condensates in the nucleus following similar principles. Constitutive heterochromatin is an important feature of the eukaryotic genome and is associated with H3K9me3, which is recognized by HP1 (REF.[129]). Similar to CBX2, several HP1 homologues contain a chromodomain (required for recognizing H3K9me3) and IDRs, and can form compartmentalized structures in the nucleus[13]. For example, HP1 forms high-concentration territories of genomic regions, including centromeric DNA regions from multiple chromosomes[55,56] (FIG. 4d). In summary, spatial organization of chromatin proteins and their regulatory targets might act to increase their concentration and reinitiate transcriptional gene silencing following cell division.

Nuclear bodies and RNA processing

Gene regulation involves both transcription modulation and various forms of co-transcriptional and post-transcriptional RNA processing, such as base modification, polyadenylation and splicing. Several RNA processing steps are known to occur co-transcriptionally. For example, the majority of mRNA splicing in mammals occurs co-transcriptionally: the nascent pre-mRNA is spliced as it is being transcribed by Pol II[130].

Pol II has a unique, highly conserved, carboxy-terminal domain (CTD), which contains 52 heptad repeats that undergo extensive post-translational modifications during distinct stages of transcription[131,132]. These modifications act as binding sites for distinct mRNA processing factors. As an example, phosphorylation of Ser2 (Ser2P) is associated with transcription elongation and has been reported to bind several proteins required for mRNA splicing and 3′ cleavage and polyadenylation[133]. Deletion of the Pol II CTD leads to a defect in the rate of co-transcriptional splicing of pre-mRNAs, suggesting that the CTD is important for coordination of transcription and splicing[15].

In this section, we discuss emerging evidence relating spatial organization in classical nuclear bodies with kinetic coupling of co-transcriptional RNA processing.

Coupling of transcription and splicing

Based on these observations, the traditional view of co-transcriptional splicing is that various proteins form a complex with Pol II, which brings the splicing machinery into proximity with a pre-mRNA immediately upon its transcription[134–136] (FIG. 5a). However, this direct interaction model cannot fully explain many of the observed properties of this system. Specifically, although multiple mRNA splicing and processing factors have been reported to bind to the CTD[137], a single Pol II molecule

Splicing condensates
High concentrations of splicing factors localized around nascent RNA transcripts, often in proximity to (but distinct from) nuclear speckles.

is not big enough to accommodate a direct interaction with all of these proteins simultaneously[138]. Nevertheless, splicing and processing factors are present at each nascent transcript, which could be facilitated by condensate formation (FIG. 5). Furthermore, the relationship between transcription and splicing kinetics is non-linear — genes that are highly transcribed exhibit a non-linear increase in their splicing rate[139] (FIG. 5b).

CTD and splicing factors form nuclear condensates. Pol II and various mRNA splicing and processing factors have been shown to form nuclear condensates[11,43,80]. These splicing condensates are formed through low-affinity, heterotypic interactions between the IDRs of various splicing factors and the CTD of Pol II, itself also an IDR. In contrast to Mediator condensates, which form with the hypo-phosphorylated CTD during transcription initiation, condensates of splicing factors form specifically with the hyper-phosphorylated CTD, which is associated with transcription elongation[43]. As such, phosphorylation of the CTD enables a transition from initiation to elongation and from the initiation condensate (containing enhancers and Mediator) to an elongation condensate (containing pre-mRNAs and splicing and processing factors) (FIG. 5c). This elongation condensate would enable simultaneous association of a large number of splicing and processing factors with the elongating Pol II.

Actively transcribed genes are organized on the periphery of nuclear speckles. Recent studies have shown that the genomic regions of highly transcribed Pol II genes — and their corresponding nascent pre-mRNAs — are organized around nuclear speckles[10,56,140]. A nuclear speckle is a nuclear body that contains numerous splicing and processing proteins[11]. Specifically, the inner structural core of nuclear speckles is composed of the serine/arginine-rich mRNA splicing factors SRSF1 and SRSF2 and other mRNA processing factors (such as the cleavage and polyadenylation factors CPSF, CSTF and CFIm (also known as CPSF6))[141,142], whereas the periphery consists of chromatin and nascent pre-mRNAs[143].

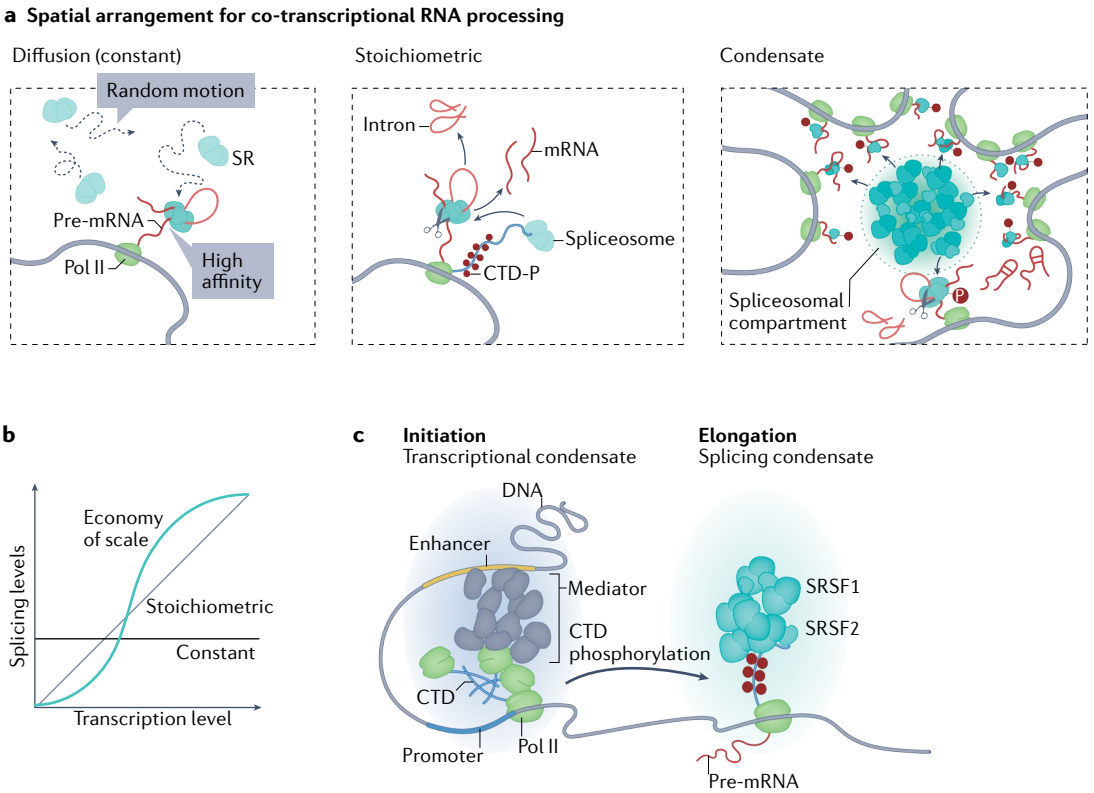

Fig. 5 | **Spatial and kinetic coupling of RNA polymerase II transcription and mRNA splicing. a** | Different models of co-transcriptional splicing. Random diffusion of splicing factors (for example, SR proteins) that have high affinity for target pre-mRNAs (left). In this model, splicing levels are independent of transcription and constant (part **b**). Stoichiometric model of co-transcriptional splicing, showing molecular interactions between the phosphorylated carboxy-terminal domain (CTD-P) 'tail' of RNA polymerase II (Pol II) and the spliceosome, which facilitates splicing as nascent pre-mRNAs are extruded from the polymerase (middle). In this model, splicing increases linearly with transcription rate (part **b**). Condensate model of co-transcriptional splicing, showing cooperative interactions between the CTD of Pol II, phosphorylated (P) splicing factors and a reservoir of unphosphorylated splicing factors in close proximity to nascent pre-mRNAs (right). **b** | A condensate model of co-transcriptional splicing would lead to an 'economy of scale' dynamic, in which splicing increases non-linearly with increased transcription. **c** | Pol II CTD phosphorylation mediates a switch from transcriptional condensates of the Mediator complex during transcription initiation, to condensates of splicing factors during transcription elongation. SRSF1, serine/arginine-rich splicing factor 1. Image courtesy of Inna-Marie Strazhnik/Caltech.

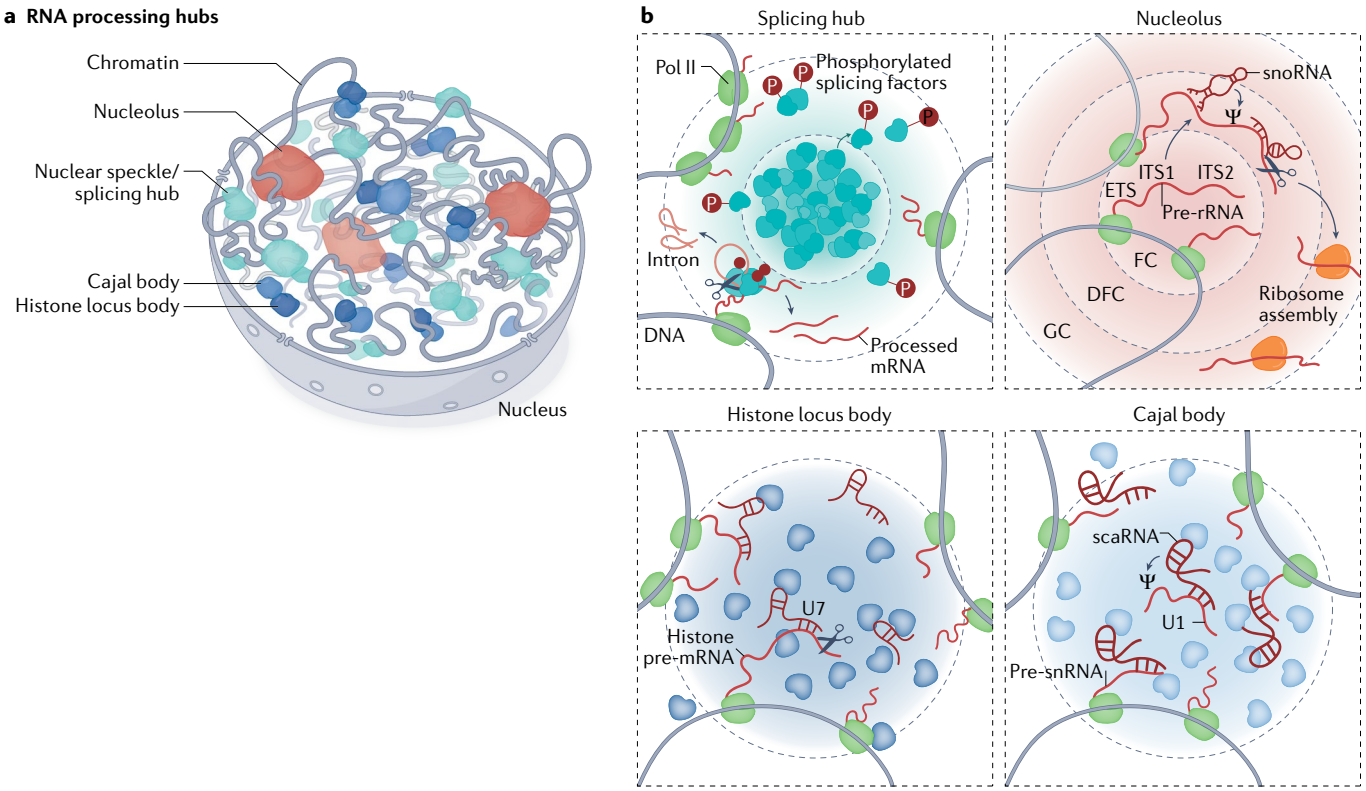

Fig. 6 | **Involvement of nuclear bodies in RNA processing. a** | Schematic of a cell nucleus highlighting many RNA processing hubs, including nuclear speckles (compartments of mRNA splicing factors), nucleoli (ribosome biogenesis), histone locus bodies (processing histone pre-mRNAs) and Cajal bodies (biogenesis of small nuclear RNAs (snRNAs)). **b** | Close-up schematics of different RNA processing hubs. A splicing hub is generally composed of two layers: an inner layer of splicing factors and an outer layer of genomic regions with active RNA polymerase II (Pol II), phosphorylated (P) splicing factors and nascent pre-mRNAs. The nucleolus is composed of three layers: inner fibrillar centre (FC), which is the site of Pol I-mediated pre-rRNA transcription; middle dense fibrillar component (DFC), where rRNA processing occurs by small nucleolar RNAs (snoRNAs) and RNase MRP (scissors); and outer granular component (GC), which is the site of ribosome assembly. Histone locus bodies are sites of histone pre-mRNA maturation, including by the U7 snRNA (U7). Cajal bodies are sites of snRNA maturation, including by small Cajal body RNAs (scaRNAs). Ψ, pseudouridylation. Image courtesy of Inna-Marie Strazhnik/Caltech.

Interestingly, the distance between a gene and a nuclear speckle is inversely correlated with its transcription level[139] and genes in highly transcribed neighbourhoods are preferentially proximal to speckles[10,140]. Although proximity to nuclear speckles is associated with increased gene expression[10,144,145], such genomic regions also include intergenic regions and inactive genes interspersed with highly transcribed Pol II neighbourhoods[10]. These observations suggest a model in which the induction of transcription by Pol II leads to dynamic reorganization of DNA into 3D proximity to nuclear speckles.

Spatial organization may coordinate co-transcriptional processing. Although splicing does not appear to occur within nuclear speckles, the splicing factors that are contained within them were shown to diffuse from the speckle to the nascent pre-mRNA[146]. A recent study showed that genes with increased splicing are also closer to a nuclear speckle[139]. Compartmentalization of regulatory components around speckles increases the spatial concentration of splicing factors near nascent pre-mRNAs, and in this way may increase the rate of co-transcriptional splicing. This would lead to a non-linear relationship between Pol II concentration (and transcription rate) and splicing, because the localization of a gene closer to a speckle would increase the concentration of spliceosomal components non-linearly relative to its pre-mRNA targets. Because rates of a biochemical reaction increase with increased enzyme or substrate concentrations, concentrating the relevant molecules within a specific territory would increase the kinetic rate of the specific reaction without changing the overall concentration of the molecules within the cell. This could explain why splicing efficiency behaves in an 'economy of scale'[139] manner, increasing non-linearly as transcription levels increase.

Nuclear bodies and gene regulation

This relationship between spatial organization of genomic regions, their associated nascent RNAs and the diffusible regulators required for their co-transcriptional processing is shared across different functions of RNA processing, suggesting that spatial organization might be a general feature of co-transcriptional RNA processing (FIG. 6a). In addition to the connection between nuclear

speckles and pre-mRNAs, the nucleus contains several nuclear bodies that are organized around transcription and processing of specialized RNA molecules[56,143,147–150].

The nucleolus is organized around transcription and processing of ribosomal RNA. The nucleolus is the site of ribosome biogenesis and consists of ribosomal DNA, ncRNAs (including small nucleolar RNAs) and proteins involved in ribosomal RNA (rRNA) processing. Importantly, transcription of the 45S pre-rRNA is essential for formation and maintenance of the nucleolus[50,51,151,152] and direct (stoichiometric) binding of diffusible factors to the 45S pre-rRNA is required for their recruitment into the nucleolus.

The nucleolus organizes rRNA biogenesis in three liquid-like phases: the fibrillar centre, which corresponds with molecules required for rRNA transcription; the dense fibrillar component (rRNA processing); and the granular component (ribosome assembly)[39]. Several ncRNAs and proteins important for rRNA biogenesis are contained in the first two, inner layers, including small nucleolar RNAs (which bind to and modify pre-rRNAs) and RNase MRP (RMRP; which cleaves pre-rRNAs into mature rRNAs)[153–155] (FIG. 6b). In this way, compartmentalization may help to physically separate steps of RNA biogenesis, thereby ensuring that one step is completed before the intermediate rRNA form is shuttled to the next compartment. A similar mechanism might be used to ensure completion of splicing prior to diffusion and export to the cytoplasm and for other forms of co-transcriptional processing[143].

Histone locus bodies are organized around transcription and processing of histone mRNAs. The biogenesis of histone mRNAs occurs in histone locus bodies, which are nuclear compartments that contain multiple histone DNA loci, U7 snRNAs and various regulatory proteins (NPAT, FLASH, NELF and others)[53]. Unlike other pre-mRNAs, histone genes do not contain introns and are not polyadenylated[150]. Instead, the U7 snRNP binds to the 3′ end of histone mRNAs and is required for their cleavage before they are exported as mature transcripts to the cytoplasm[150,156–158]. Nascent histone mRNA is required for the formation of histone locus bodies[149] and for the recruitment of various diffusible ncRNAs and proteins to this specialized processing compartment (FIG. 6b).

Cajal bodies are organized around transcription and processing of snRNAs. Cajal bodies are the sites of snRNA biogenesis and contain snRNA genes (U1, U2, U4, U5 and U6), small Cajal body RNAs (scaRNAs) and proteins such as coilin and SMN[56,159]. Similar to small nucleolar RNAs in the nucleolus, scaRNAs bind directly to pre-snRNAs and mediate methylation and pseudouridylation within Cajal bodies[160,161]. Concentrating regulatory components (scaRNAs) and targets (snRNAs) may increase the efficiency of snRNA processing. This example also highlights another potentially important role of spatial compartmentalization in RNA processing: compartmentalization can increase the rate by which regulators identify and engage targets, which may be particularly important in cases where the regulators (for example, scaRNAs) are expressed at lower levels than their substrates (for example, snRNAs) (FIG. 6b).

Conclusions and future directions

In this Review, we aimed to highlight the functions of specific nuclear compartments in various aspects of gene regulation, including transcription initiation, higher-order chromatin regulation, epigenetics and RNA processing. We have described several emerging principles by which compartmentalization can have important quantitative effects on gene regulation. These principles include increasing overall affinity and kinetic rates, increasing specificity of target recognition, memory and persistence of transcriptional states, time-dependent and concentration-dependent transitions in responses to cues, and the ability to achieve non-stoichiometric responses. We anticipate that this emerging perspective will promote a deeper understanding of how changes in spatial concentration of molecules in the nucleus drive quantitative and kinetic properties of gene regulation.

Nevertheless, nuclear compartmentalization remains a new field of research with many open questions and challenges, which we discuss next.

Functions of nuclear compartments

Although nuclear compartments can have key functions and roles in specific cases, not all nuclear compartments may be functionally important. For example, although HP1α can form condensates, it was recently shown that heterochromatin maintenance can occur independently of this property[13,14,162]. Moreover, transcriptionally repressed chromatin is known to be localized at the nuclear lamina — a nuclear compartment enriched in repressive chromatin regulators and heterochromatin factors[163] — but several studies have shown that experimentally anchoring DNA to the nuclear lamina has little to no effect on transcription regulation[164–166]. Thus, in some cases, nuclear compartments may simply represent physical properties of functional states (for example, repressed chromatin tends to associate with the nuclear lamina) rather than mechanisms for controlling function.

To date, what functions and roles, if any, most nuclear compartments have remains largely unknown. One of the main challenges in addressing this question is that studying nuclear compartments, their formation and their functional effects is often confounded. For example, it is challenging to disrupt the formation of transcription factor condensates without also affecting their activation domains[79], or to disrupt the ability of PRC1 proteins to form a condensate without affecting their ability to condense chromatin[41]. A key example of this challenge is illustrated by the nucleolus, which is one of the most well-characterized nuclear compartments. Although it is now well established that the molecular components associated with ribosome biogenesis are contained within this compartment, whether this spatial organization is essential for ribosome biogenesis or whether ribosome biogenesis can occur even when these components are not spatially assembled remains untested, because there is no simple way to disrupt nucleolar assembly without also directly affecting ribosome biogenesis. Thus, it is often difficult to disentangle the cause

and effect relationship between compartmentalization of molecules and their functions.

Recent technological innovations in organizing condensates and DNA around nuclear bodies have begun to enable compartment manipulation and exploration of these questions[167–170] (BOX 1). Yet most of these methods and tools still rely on creating compartments and condensates through overexpression of molecular components, and as such they do not perfectly represent the normal cellular context. Defining the functions and roles of different nuclear compartments will require creative new genetic approaches to test the cause and effect of compartmentalization and gene regulation.

Mechanisms of compartment formation

Although specific compartments have been well mapped out, there are still many open questions about the molecular mechanisms by which most nuclear compartments are formed. Specifically, which molecules act as seeds that nucleate formation of a nuclear compartment and whether spatial enrichment of molecules is mediated by stoichiometric interactions with a seed, concentration-dependent condensate formation or a combination of multiple events is largely unknown. For example, the Xi is well characterized as forming a nuclear compartment — the Barr body — yet the precise mechanism of its formation is not fully resolved. Although there is a clear indication that individual *XIST*-binding proteins can form condensates[47,56], whether or not the entire Barr body itself forms a condensate remains unknown.

One of the key challenges is that it remains difficult to measure phase separation and molecular dynamics and assembly in living cells. Most studies to date have characterized phase separation in vitro using purified proteins, or by imaging tagged and overexpressed proteins in vivo. Both of these approaches have limitations because they do not represent endogenous concentrations within the cell. Moreover, nuclear compartments are often too small to allow accurate measurement of their internal diffusion rates or spherical shapes, or to observe individual molecules and their stoichiometries. Improvements in super-resolution microscopy have begun to address some of these challenges and we anticipate that further advances in microscopy-based approaches will be needed to resolve many of these questions.

Components within nuclear compartments

It remains largely unknown which molecular components are contained within individual nuclear compartments. For example, whereas we know that Pol II can form spatially enriched compartments within the nucleus, we still do not know the full ensemble of proteins that are contained within these condensates, or at which genomic regions (if any) these condensates localize.

The main challenge here is that nuclear compartments can contain dozens to thousands of distinct molecular components, including DNA, RNA and protein; occupy small (nanometre) to large (micrometre) distances in the nucleus; and undergo precise concentration-dependent transitions. Current methods for measuring molecular organization, including microscopy and genomic methods, are limited in their ability to measure these features because they cannot yet map the combinatorial and spatial organization of these complex assemblies. Recent advances in highly multiplexed super-resolution microscopy methods, such as sequential fluorescence in situ hybridization (seqFISH+)[171,172], and in genomic methods for mapping multiway spatial interactions (for example, SPRITE)[10,56] have begun to shed light on some of these questions. We expect that further technological developments that increase scale, resolution and detection modalities will be crucial for fully defining the molecular and spatial composition of nuclear compartments.

Biophysical properties and function

Finally, it remains largely unclear what the functional implications are for nuclear compartments that have different biophysical properties. For example, are there particular functional implications for compartments that are liquid-like and are they functionally distinct from compartments that may display gel or solid properties? One proposal is that the ability to form a liquid-like state would lead to a barrier that would limit molecular diffusion across compartment boundaries[33,173]. Another proposal is that liquid-like condensates may enable persistence of molecular assembly even upon the loss of the initiating molecule[47]. Although attractive, these possibilities are yet to be experimentally demonstrated for biological compartments. Testing the relationship between the biophysical properties of a compartment and its functional roles will require the capacity to accurately define their physical properties in vivo, specifically disrupt or transform these biophysical states and measure their functional roles.

We expect that experimental advances will enable further advances in studying the quantitative relationships between compartmentalization and gene regulation, and will likely uncover additional connections that have yet to be defined.

Published online: 02 August 2021

1. Mirny, L. et al. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.* **42**, 434013 (2009).
2. Jana, T., Brodsky, S. & Barkai, N. Speed–specificity trade-offs in the transcription factors search for their genomic binding sites. *Trends Genet.* **37**, 421–432 (2021).
3. Strom, A. R. & Brangwynne, C. P. The liquid nucleome—phase transitions in the nucleus at a glance. *J. Cell Sci.* **132**, jcs235093 (2019).
4. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).
5. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 772–772 (2016).
6. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
7. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
8. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
9. Dundr, M. & Misteli, T. Biogenesis of nuclear bodies. *Cold Spring Harb. Perspect. Biol.* **2**, a000711 (2010).
10. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018).
11. Cho, W. K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
12. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
13. Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).

14. Larson, A. G. et al. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
15. Huo, X. et al. The nuclear matrix protein SAFB cooperates with major satellite RNAs to stabilize heterochromatin architecture partially through phase separation. *Mol. Cell* **77**, 368–383.e7 (2020).
16. Murray, D. T. et al. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615–627.e16 (2017).
17. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular condensates in the nucleus. *Trends Biochem. Sci.* **45**, 961–977 (2020).
18. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid–liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
19. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
20. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
21. Zhang, Z. et al. Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.* **30**, 2106–2118 (2016).
22. Struhl, K. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* **14**, 137–140 (1989).
23. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131 (2005).
24. García-Mayoral, M. F. et al. The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure* **15**, 485–498 (2007).
25. Stevens, S. W. et al. Biochemical and genetic analyses of the U5, U6, and U4/U6•U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543–1553 (2001).
26. Derewenda, Z. S. Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallogr. D* **66**, 604–615 (2010).
27. Derewenda, Z. S. Rational protein crystallization by mutational surface engineering. *Structure* **12**, 529–535 (2004).
28. Deller, M. C., Kong, L. & Rupp, B. Protein stability: a crystallographer's perspective. *Acta Crystallogr. F* **72**, 72–95 (2016).
29. Boeynaems, S. et al. Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
30. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
31. Li, P. et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
32. Kitov, P. I. & Bundle, D. R. On the nature of the multivalency effect: a thermodynamic model. *J. Am. Chem. Soc.* **125**, 16271–16284 (2003).
33. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).
34. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
35. Van Der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
36. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
37. Van Treeck, B. & Parker, R. Emerging roles for intermolecular RNA–RNA interactions in RNP assemblies. *Cell* **174**, 791–802 (2018).
38. Brangwynne, C. P. et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
39. Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).
40. Seif, E. et al. Phase separation by the polyhomeotic sterile α motif compartmentalizes Polycomb group proteins and enhances their activity. *Nat. Commun.* **11**, 5609 (2020).
41. Plys, A. J. et al. Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).
42. Tatavosian, R. et al. Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. *J. Biol. Chem.* **294**, 1451–1463 (2019).
43. Guo, Y. E. et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
44. Lu, Y. et al. Phase separation of TAZ compartmentalizes the transcription machinery to promote gene expression. *Nat. Cell Biol.* **22**, 453–464 (2020).
45. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
46. Altmeyer, M. et al. Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088 (2015).
47. Pandya-Jones, A. et al. A protein assembly mediates Xist localization and gene silencing. *Nature* **587**, 145–151 (2020).
48. Padeken, J. & Heun, P. Nucleolus and nuclear periphery: Velcro for heterochromatin. *Curr. Opin. Cell Biol.* **28**, 54–60 (2014).
49. Tzur, Y. B., Wilson, K. L. & Gruenbaum, Y. SUN-domain proteins: 'Velcro' that links the nucleoskeleton to the cytoskeleton. *Nat. Rev. Mol. Cell Biol.* **7**, 782–788 (2006).
50. Falahati, H., Pelham-Webb, B., Blythe, S. & Wieschaus, E. Nucleation by rRNA dictates the precision of nucleolus assembly. *Curr. Biol.* **26**, 277–285 (2016).
51. Berry, J. et al. RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl Acad. Sci. USA* **112**, E5237–E5245 (2015).
52. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
53. Nizami, Z., Deryusheva, S. & Gall, J. G. The Cajal body and histone locus body. *Cold Spring Harb. Perspect. Biol.* **2**, a000653 (2010).
54. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20**, 2223–2237 (2006).
55. Maison, C. et al. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* **30**, 329–334 (2002).
56. Quinodoz, S. A. et al. RNA promotes the formation of spatial compartments in the nucleus. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.25.267435 (2020).
57. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**, 1635–1655 (2005).
58. Bernstein, E. et al. Mouse Polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* **26**, 2560–2569 (2006).
59. McHugh, C. A. C. A. et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
60. Markaki, Y. et al. Xist-seeded nucleation sites form local concentration gradients of silencing proteins to inactivate the X-chromosome. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.22.393546 (2020).
61. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167–174 (2002).
62. Hacisuleyman, E. et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).
63. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
64. Wang, L. et al. Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. *Mol. Cell* **76**, 646–659.e6 (2019).
65. Fan, C. et al. Rett mutations attenuate phase separation of MeCP2. *Cell Discov.* **6**, 646 (2020).
66. Li, C. H. et al. MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* **586**, 440–444 (2020).
67. Wang, L. et al. Rett syndrome-causing mutations compromise MeCP2-mediated liquid–liquid phase separation of chromatin. *Cell Res.* **30**, 393–407 (2020).
68. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
69. Dossin, F. et al. SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455–460 (2020).
70. McSwiggen, D. T. et al. Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife* **8**, e47098 (2019).
71. Gasch, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
72. Chapal, M., Mintzer, S., Brodsky, S., Carmi, M. & Barkai, N. Resolving noise-control conflict by gene duplication. *PLoS Biol.* **17**, e3000289 (2019).
73. Zobeck, K. L., Buckley, M. S., Zipfel, W. R. & Lis, J. T. Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells. *Mol. Cell* **40**, 965–975 (2010).
74. Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018).
75. Banerji, J., Olson, L. & Schaffner, W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**, 729–740 (1983).
76. Amano, T. et al. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57 (2009).
77. Benabdallah, N. S. et al. Decreased enhancer–promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484.e7 (2019).
78. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
79. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
80. Zamudio, A. V. et al. Mediator condensates localize signaling factors to key cell identity genes. *Mol. Cell* **76**, 753–766.e6 (2019).
81. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
82. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
83. Symmons, O. et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529–543 (2016).
84. Flavahan, W. A. et al. Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229–233 (2019).
85. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
86. Williamson, I. et al. Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**, dev179523 (2019).
87. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934 (2013).
88. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
89. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
90. Van Steensel, B. et al. Localization of the glucocorticoid receptor in discrete clusters in the cell nucleus. *J. Cell Sci.* **108**, 3003–3011 (1995).
91. Grande, M. A., Van Der Kraan, I., De Jong, L. & Van Driel, R. Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J. Cell Sci.* **110**, 1781–1791 (1997).
92. Bregman, D. B., Du, L., Van Der Zee, S. & Warren, S. L. Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. *J. Cell Biol.* **129**, 287–298 (1995).
93. Papantonis, A. & Cook, P. R. Transcription factories: genome organization and gene regulation. *Chem. Rev.* **113**, 8683–8705 (2013).
94. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109**, 1427–1436 (1996).
95. Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).

96. Wang, X., Cairns, M. J. & Yan, J. Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Res.* **47**, 11481–11496 (2019).
97. Cho, W. K. et al. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, e13617 (2016).
98. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory. *Biochemistry* **20**, 6929–6948 (1981).
99. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
100. Li, G. W. & Xie, X. S. Central dogma at the single-molecule level in living cells. *Nature* **475**, 308–315 (2011).
101. Brodsky, S. et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* **79**, 459–471.e4 (2020).
102. Valsecchi, C. I. K. et al. RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature* **589**, 137–142 (2021).
103. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, 780–788 (2006).
104. Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J. & Lawrence, J. B. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* **162**, 981–990 (2003).
105. Lomvardas, S. et al. Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
106. Strehle, M. & Guttman, M. Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Curr. Opin. Cell Biol.* **64**, 139–147 (2020).
107. Loda, A. & Heard, E. Xist RNA in action: past, present, and future. *PLoS Genet.* **15**, e1008333 (2019).
108. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
109. Minajigi, A. et al. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
110. Moindrot, B. et al. A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* **12**, 562–572 (2015).
111. Monfort, A. et al. Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).
112. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5**, 695–705 (2000).
113. Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
114. Kanduri, C. Kcnq1ot1: a chromatin regulatory RNA. *Semin. Cell Dev.Biol.* **22**, 343–350 (2011).
115. Nagano, T. & Fraser, P. Emerging similarities in epigenetic gene silencing by long noncoding RNAs. *Mamm. Genome* **20**, 557–562 (2009).
116. Cheutin, T. et al. Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**, 721–725 (2003).
117. Maison, C. & Almouzni, G. HP1 and the dynamics of heterochromatin maintenance. *Nat. Rev. Mol. Cell Biol.* **5**, 296–304 (2004).
118. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
119. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
120. Mira-Bontenbal, H. & Gribnau, J. New Xist-interacting proteins in X-chromosome inactivation. *Curr. Biol.* **26**, R338–R342 (2016).
121. Cerase, A. et al. Phase separation drives X-chromosome inactivation: a hypothesis. *Nat. Struct. Mol. Biol.* **26**, 331–334 (2019).
122. Wutz, A. Xist function: bridging chromatin and stem cells. *Trends Genet.* **23**, 457–464 (2007).
123. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
124. Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
125. Fischle, W. 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).
126. Illingworth, R. S. Chromatin folding and nuclear architecture: PRC1 function in 3D. *Curr. Opin. Genet. Dev.* **55**, 82–90 (2019).
127. Pirrotta, V. & Li, H. B. A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* **22**, 101–109 (2012).
128. Eeftens, J. M., Kapoor, M. & Brangwynne, C. P. Epigenetic memory as a time integral over prior history of Polycomb phase separation. *bioRxiv* https://doi.org/10.1101/2020.08.19.254706 (2020).
129. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
130. Girard, C., Will, C., Peng, J. et al. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994 (2012).
131. Napolitano, G., Lania, L. & Majello, B. RNA polymerase II CTD modifications: how many tales from a single tail. *J. Cell. Physiol.* **229**, 538–544 (2014).
132. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).
133. Ahn, S. H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* **13**, 67–76 (2004).
134. Merkhofer, E. C., Hu, P. & Johnson, T. L. Introduction to cotranscriptional RNA splicing. *Methods Mol. Biol.* **1126**, 83–96 (2014).
135. Han, J., Xiong, J., Wang, D. & Fu, X. D. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336–343 (2011).
136. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).
137. McCracken, S. et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–360 (1997).
138. Zhang, S. et al. Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* **371**, 305–309 (2021).
139. Ding, F. & Elowitz, M. B. Constitutive splicing and economies of scale in gene expression. *Nat. Struct. Mol. Biol.* **26**, 424–432 (2019).
140. Chen, Y. et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
141. Dopie, J., Sweredoski, M. J., Moradian, A. & Belmont, A. S. Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *J. Cell Biol.* **219**, e201910207 (2020).
142. Saitoh, N. et al. Proteomic analysis of interchromatin granule clusters. *Mol. Biol. Cell* **15**, 3876–3890 (2004).
143. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605–612 (2003).
144. Kim, J., Venkata, N. C., Gonzalez, G. A. H., Khanna, N. & Belmont, A. S. Gene expression amplification by nuclear speckle association. *J. Cell Biol.* **219**, e201904046 (2020).
145. Alexander, K. A. et al. p53 mediates target gene association with nuclear speckles for amplified RNA expression. *Mol. Cell* **81**, 1666–1681 (2021).
146. Misteli, T. & Spector, D. L. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell* **3**, 697–705 (1999).
147. Pederson, T. The nucleolus. *Cold Spring Harb. Perspect. Biol.* **3**, 1–15 (2011).
148. Carmo-Fonseca, M. & Rino, J. RNA seeds nuclear bodies. *Nat. Cell Biol.* **13**, 110–112 (2011).
149. Shevtsov, S. P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* **13**, 167–173 (2011).
150. Marzluff, W. F. & Koreski, K. P. Birth and death of histone mRNAs. *Trends Genet.* **33**, 745–759 (2017).
151. Scheer, U. & Benavente, R. Functional and dynamic aspects of the mammalian nucleolus. *BioEssays* **12**, 14–21 (1990).
152. Schlegel, R. A., Miller, L. S. & Rose, K. M. Reduction in RNA synthesis following red cell-mediated microinjection of antibodies to RNA polymerase I. *Cell Biol. Int. Rep.* **9**, 341–350 (1985).
153. Watkins, N. J. & Bohnsack, M. T. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip. Rev. RNA* **3**, 397–414 (2012).
154. Kiss-László, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. & Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077–1088 (1996).
155. Ni, J., Tien, A. L. & Fournier, M. J. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565–573 (1997).
156. Spycher, C. et al. 3′ end processing of mouse histone pre-mRNA: evidence for additional base-pairing between U7 snRNA and pre-mRNA. *Nucleic Acids Res.* **22**, 4023–4030 (1994).
157. Kolev, N. G. & Steitz, J. A. Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs. *Genes Dev.* **19**, 2583–2592 (2005).
158. Mowry, K. L. & Steitz, J. A. Identification of the human U7 snRNP as one of several factors involved in the 3′ end maturation of histone premessenger RNA's. *Science* **238**, 1682–1687 (1987).
159. Machyna, M. et al. The coilin interactome identifies hundreds of small noncoding RNAs that traffic through cajal bodies. *Mol. Cell* **56**, 389–399 (2014).
160. Darzacq, X. et al. Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746–2756 (2002).
161. Kiss, A. M. et al. A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. *Nucleic Acids Res.* **30**, 4643–4649 (2002).
162. Erdel, F. et al. Mouse heterochromatin adopts digital compaction states without showing hallmarks of HP1-driven liquid-liquid phase separation. *Mol. Cell* **78**, 236–249.e7 (2020).
163. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
164. Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243–247 (2008).
165. Kumaran, R. I. & Spector, D. L. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.* **180**, 51–65 (2008).
166. Finlan, L. E. et al. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* **4**, e1000039 (2008).
167. Kim, J. H. et al. LADL: light-activated dynamic looping for endogenous gene expression control. *Nat. Methods* **16**, 633–639 (2019).
168. Wang, H. et al. CRISPR-mediated programmable 3D genome positioning and nuclear organization. *Cell* **175**, 1405–1417.e14 (2018).
169. Shin, Y. et al. Spatiotemporal control of intracellular phase transitions using light-activated optodroplets. *Cell* **168**, 159–171.e14 (2017).
170. Shin, Y. et al. Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* **175**, 1481–1491.e13 (2018).
171. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
172. Eng, C. H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
173. Collombet, S. et al. RNA polymerase II depletion from the inactive X chromosome territory is not mediated by physical compartmentalization. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.26.437188 (2021).
174. Jerkovic', I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* https://doi.org/10.1038/s41580-021-00362-w (2021).
175. Ganser, L. R. & Myong, S. Methods to study phase-separated condensates and the underlying molecular interactions. *Trends Biochem. Sci.* **45**, 1004–1005 (2020).
176. Robinson, C. V., Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450**, 973–982 (2007).

177. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
178. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376.e16 (2018).
179. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
180. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
181. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl Acad. Sci. USA* **116**, 19490–19499 (2019).
182. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
183. Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun.* **6**, 7147 (2015).
184. Beliveau, B. J., Apostolopoulos, N. & Wu, C. T. Visualizing genomes with Oligopaint FISH probes. *Curr. Protoc. Mol. Biol.* **105**, 14.23.1–14.23.20 (2014).
185. Takei, Y. et al. Global architecture of the nucleus in single cells by DNA seqFISH+ and multiplexed immunofluorescence. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.29.403055 (2020).
186. Dey, S. & Maiti, S. Single-molecule photobleaching: instrumentation and applications. *J. Biosci.* **43**, 447–454 (2018).
187. Hsia, Y. et al. Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
188. Elson, E. L. Introduction to fluorescence correlation spectroscopy — brief and simple. *Methods* **140–141**, 3–9 (2018).
189. Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* **143**, 1047–1058 (2010).
190. Godin, A. G., Lounis, B. & Cognet, L. Super-resolution microscopy approaches for live cell imaging. *Biophys. J.* **107**, 1777–1784 (2014).
191. McCall, P. M. et al. Quantitative phase microscopy enables precise and efficient determination of biomolecular condensate composition. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.25.352823 (2020).
192. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
193. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

Acknowledgements
The authors thank members of the Guttman laboratory, especially M. Strehle, J. Jachowicz, S. Quinodoz and I. Goronzy for helpful comments and discussions, I.-M. Strazhnik for figures and S. Hiley for editing. P.B. is supported by the University of California, Los Angeles (UCLA)-Caltech Medical Scientist Training Program (MSTP), National Institutes of Health (NIH) F30CA247447 and a Chen Graduate Innovator Grant. M.G. is a New York Stem Cell Foundation Robertson Investigator and an investigator at the Heritage Medical Research Institute. Research in the Guttman laboratory is funded by the NIH 4DN programme, an NIH Director's Transformative R01 Award, the Chan-Zuckerberg Initiative and funds from Caltech.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Molecular Cell Biology thanks Sergey Razin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.