The FEBS Journal

FEBS PRESS science publishing by scientists

STATE-OF-THE-ART REVIEW

Nuclear speckles: dynamic hubs of gene expression regulation

İbrahim Avşar Ilık and Tuğçe Aktaş

Max Planck Institute for Molecular Genetics, Berlin, Germany

Keywords
nuclear speckles; phase separation; SON; splicing; SRRM2; transcription

Correspondence
T. Aktaş, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195, Berlin Germany
Tel: +493084131550
E-mail: aktas@molgen.mpg.de

(Received 11 March 2021, revised 13 June 2021, accepted 8 July 2021)

doi:10.1111/febs.16117

Complex, multistep biochemical reactions that routinely take place in our cells require high concentrations of enzymes, substrates, and other structural components to proceed efficiently and typically require chemical environments that can inhibit other reactions in their immediate vicinity. Eukaryotic cells solve these problems by restricting such reactions into diffusion-restricted compartments within the cell called organelles that can be separated from their environment by a lipid membrane, or into membrane-less compartments that form through liquid–liquid phase separation (LLPS). One of the most easily noticeable and the earliest discovered organelle is the nucleus, which harbors the genetic material in cells where transcription by RNA polymerases produces most of the messenger RNAs and a plethora of noncoding RNAs, which in turn are required for translation of mRNAs in the cytoplasm. The interior of the nucleus is not a uniform soup of biomolecules and rather consists of a variety of membrane-less bodies, such as the nucleolus, nuclear speckles (NS), paraspeckles, Cajal bodies, histone locus bodies, and more. In this review, we will focus on NS with an emphasis on recent developments including our own findings about the formation of NS by two large IDR-rich proteins SON and SRRM2.

Nuclear speckles: easy to detect, difficult to define

'What are nuclear speckles?' is a fundamental and seemingly simple question, but one that, as of yet, remains without a compelling answer. Since their initial discovery by Ramon y Cajal more than a century ago [1], definitions of nuclear speckles (NS) are descriptive in nature, typically referring to their physical appearance in microscopy experiments such as 'nuclear domains enriched in pre-mRNA splicing factors, located in the interchromatin region of the nucleoplasm of mammalian cells' [2]. Several questions emerge from such definitions: What are NS made of? What functional roles do they play in cells? Are NS functionally essential for cell viability? Are NS restricted to mammalian cells? The dichotomy between the ease of detecting NS (with antibodies, or more recently with live-cell imaging of prominent factors) and their resistance to be functionally defined has led to the accumulation of vast amounts of data around NS and cellular processes that they might be involved in such as splicing, termination, nuclear export of

Abbreviations
CLK1, CDC-like kinase 1; CWC21, complexed with CEF1 protein 21; DYRK3, dual-specificity tyrosine phosphorylation-regulatedkinase 3; IDR, intrinsically disordered region; IGC, interchromatin granule clusters; LLPS, liquid–liquid phase separation; MALAT1, metastasis-associated lung adenocarcinoma transcript 1; NS, nuclear speckles; RBP, RNA-binding protein; SC35, anti-splicing component 35K; SPRITE, split-pool recognition of interactions by tag extension; SRRM2, serine/arginine repetitive matrix protein 2; TSA-seq, tyramide signal amplification.

mRNAs, and transcription [3] but generally without irrefutable evidence. Presumably, controlled disruption of NS could be the key to resolve these issues; however, decades of research, including recent high-throughput screens [4], failed to identify a single factor without which NS do not form. Indeed, due to lack of evidence to the contrary, it has been suggested that NS do not require a core protein or a set of core factors and that they simply form through accumulation of a large number of SR proteins, which could explain their resilience to removing single factors from cells [5]. Nevertheless, NS are highly dynamic structures, which can respond to activation of nearby genes [6], remodeled upon inhibition of splicing [7] and are dissolved during mitosis, and reform after late telophase [3,8] (Fig. 1). Likely mimicking the dynamics of NS during cell cycle, overexpression of CDC-like kinase 1 (Clk1) or dual-specificity tyrosine phosphorylation-regulated kinase 3 (DYRK3) has been shown to dissolve NS in interphase cells [9,10]. However, with these overexpression scenarios, potentially affecting hundreds of proteins involved in transcription and splicing, the utility of such dissolved states in deciphering the role of NS is unclear.

We recently discovered that anti-splicing component 35K (SC35), an antibody raised against human spliceosomal extracts, recognizing a phospho-epitope [11], and identified as an antibody against SRSF2 [12], an SR protein that can be localized at NS, is rather an antibody primarily against serine/arginine repetitive matrix protein 2 (SRRM2) [13], a large, mostly unstructured SR protein that joins spliceosomes at the B^{act} stage [14], that sharply localizes to NS [7,15]. While this information does not affect the usefulness of this antibody in identifying NS in immunofluorescence experiments, it does affect interpretation of certain experiments that have used this reagent. Several groups have independently identified SON, a protein that was first identified as part of NS in a proteomic screen by the Spector group, to be crucial for normal nuclear speckle morphology [16–18]. Curiously, while depletion of SON leads to the dissociation of some SR proteins with NS, SC35 signal does not dissipate and strongly stains what seems to be round-up NS that are also observed in cells treated with transcription or splicing inhibitors [13,17,18]. Codepletion of SRRM2 together with SON, or depleting SON in a cell line where SRRM2's extensive intrinsically disordered region (IDR)s are genetically deleted, results in the dissipation of all nuclear speckle markers that were tested, leading to the idea that SON and SRRM2 are essential for nuclear speckle formation [13] (Fig. 2). Based on these results, and other recent data, we propose to operationally define NS as '*nuclear, phase-separated bodies formed by two large, IDR-rich proteins SON and SRRM2*'. Although more work is needed to

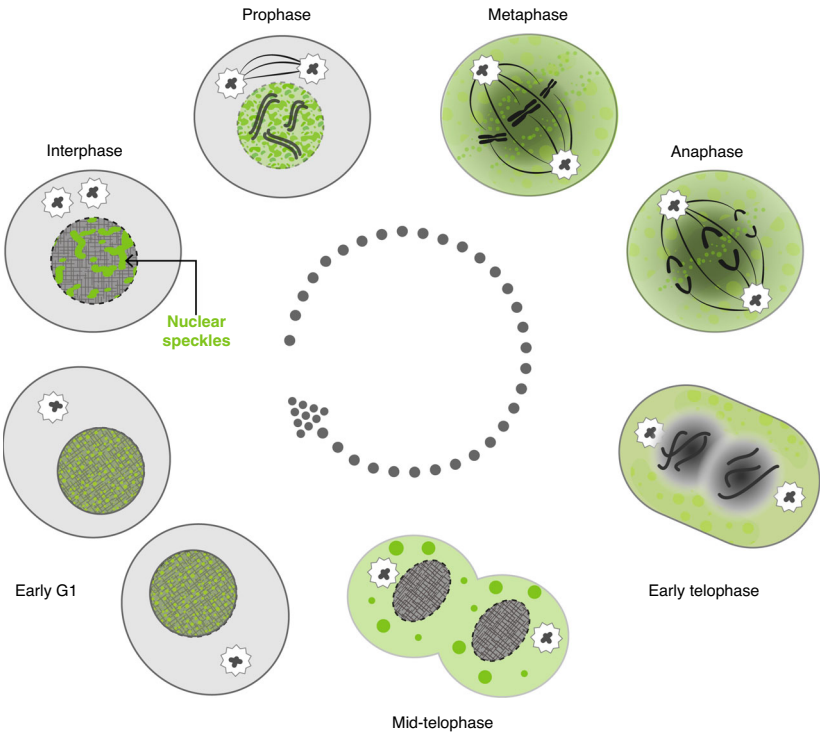

Fig. 1. Dynamic nature of the NS, depicted in green, during cell cycle is shown. NS are dissolved just before cells enter mitosis and remain diffuse until late stages of telophase where many components transiently form spherical bodies called mitotic interchromatin granules (MIGs), in the cytoplasm of newly divided cells. As MIGs disappear, components of NS are imported into the nucleus where they re-establish NS.

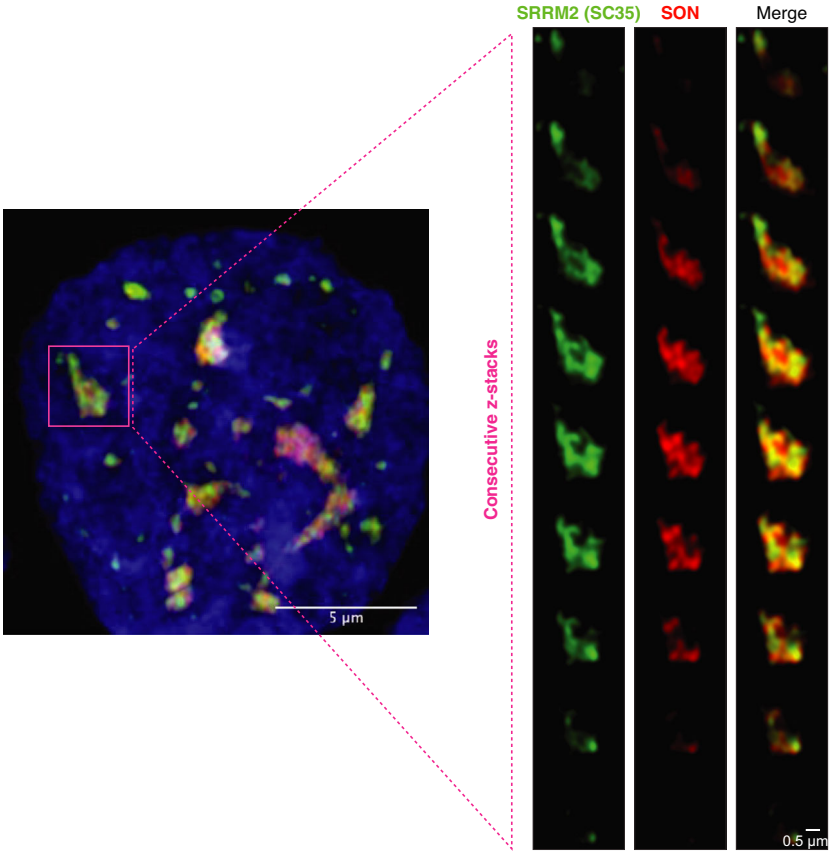

Fig. 2. Maximum intensity projection obtained by confocal microscopy of SC35 (SRRM2) and SON staining is shown. Z-stacks (on right) indicate a subtle difference between the localization of two proteins within NS.

elucidate the exact mechanism of core formation, and the roles these two proteins play in that process, the operational definition we propose has the potential to bring a more comprehensive, functional definition within reach.

What are NS made of?

Understanding the role of NS, or any other cellular body, can immensely benefit from identification of factors that are localized to them. Until 1980s, the only information regarding the composition of NS was derived from the analysis of interchromatin granule clusters (IGCs) through electron microscopy, which revealed that they are ribonucleoprotein particles [19]. Soon thereafter, the invention of monoclonal antibodies led to several key discoveries in the NS field: First, antibodies against snRNPs, essential for nuclear RNA splicing, simultaneously showed that splicing-associated factors are concentrated at NS and established that IGCs in electron microscopy by and large correspond to NS [20,21].

First systematic attempts at identifying NS-associated proteins were carried out by Spector group [22,23]. Using a biochemical fractionation technique, NS were separated from nuclear lamina, nucleoli, and other nuclear compartments and subjected to mass spectrometry. These pioneering studies have uncovered 360 proteins, most of which are RNA-binding proteins (RBP) associated with pre-mRNA splicing with a large overlap with spliceosome-associated proteins [24–26] and other RNA-processing factors involved in 3′-end cleavage/polyadenylation and nuclear export, processes that are tightly connected with each other [27]. These results have validated earlier hypotheses placing NS at the center of nuclear RNA processing and formed the basis of NS research in the following two decades. Curiously, two of the novel NS-associated factors identified were SRRM2 (KIAA0324 in Mintz *et al.* 1999) and SON (KIAA1019 in Saitoh *et al.* 2004), which our group has recently shown both to be essential for NS formation [13]. A more recent work from Belmont group has tackled the same question using tyramide signal amplification mass spectrometry (TSA-

MS) [28], using the monoclonal antibody SC35. While previous work utilized a biochemical purification approach to isolate NS, TSA-MS relies on proximity-labeling, where the target biomolecule and proteins in close proximity to it are labeled by a functional chemical group (usually biotin, but in this case FITC), which can then be pulled down for further analysis. Since this method fixes the cells as the first step followed by proximity-labeling happening *in situ*, it has the potential to uncover more transient interactions compared to biochemical purifications carried on live cells. Reassuringly, the results are by and large in agreement with earlier findings of the Spector group, although novel NS-associated proteins are also identified, such as MFAP1 which was also shown to affect NS morphology when depleted. One striking result of this study is that the two most enriched proteins were found to be SON and SRRM2, underscoring their prominence in the context of NS.

Even though hundreds of proteins have been found to localize to NS, either biochemically [23,28] or microscopically [29], the extent of localization of individual factors to NS is highly variable. While proteins such as SRRM2 and SON are virtually restricted to NS, other factors such as SRSF1, SRSF2, U1-70K, and Magoh [16,30,31] have clear nucleoplasmic pools in addition to their localization in NS. These observations suggest that there might be sequence motifs within proteins that determine the strength of their NS localization. Indeed, polyhistidine stretches were initially identified as sequence elements that can push protein into NS [32]. Similarly, stretches of the dipeptide RS, lending the name to SR protein family [33], are also shown to be NS localization signals. A recent work has also extended these signals to mixed-charged domains (MCDs), specifically to stretches of arginine (but not lysine) and glutamic acid residues in low-complexity regions which turned out to be potent NS localization signals, underscoring the unique role of arginine residues in this process [34]. In addition to proteins, RNAs can also be found to be localized to NS, and NS are known to be enriched with polyadenylated RNA as evidenced by FISH experiments using fluorescently labeled oligo-dT probes [35]. These interactions can be transient, but can also be more stable, as is the case for the abundant, nuclear long noncoding RNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) [36]. Whether certain RNA sequences, structures, or modifications target RNAs to NS or whether they associate with NS through their interactions with NS-resident RBPs remains unknown. Two recent examples show that certain low-complexity RNA sequences that can phase-separate *in vitro* localize to NS *in vivo* [37]. Such RNAs can be generated in repeat expansion disorders and have been shown to be localized to NS in cells isolated from patients [38]; however, it is not yet clear whether the accumulation of these RNAs in NS is harmful due to sequestration of RBPs in the nucleus, or whether it is a protective measure to prevent these RNAs from reaching the cytoplasm where they can sequester RBPs there or possibly produce toxic peptides [39]. These observations strongly suggest that NS are multicomponent systems, which require SON and SRRM2 for their formation. The stickers-and-spacers framework predicts that ligands interacting with a phase-separating core can weaken or strengthen phase-separation behavior, depending on the nature of interactions [40]. The fact that overexpression of arginine-rich MCDs, or inhibition of transcription/splicing can change the morphology of NS supports this idea, which can be explored further to develop tools to manipulate NS in living cells.

Genome organization around NS

In the last decade, significant advances have been made on how our genomes are organized in the interphase nucleus [41]. Leveraging advanced proximity-ligation methods such as Hi-C [42], it has been discovered that the human genome is essentially segregated into two compartments, called A and B, which preferentially engage in interactions within but not across each other. In its simplest form, the A compartment corresponds to transcriptionally active loci, whereas the B compartment corresponds to inactive chromatin, with high-resolution Hi-C further dividing the A compartment into A1 and A2 subcompartments with subtle differences in replication timing, GC content and H3K9me3 enrichment [43]. Even before the era of high-throughput sequencing, NS were linked to transcriptionally active loci through FISH experiments targeting a relatively small number of transcripts [44]. Three recently developed techniques, using different experimental approaches to identify DNA–DNA and/or RNA–DNA contacts in living cells have cemented and expanded this view. Using MARGI (Mapping of RNA-Genome Interactions), Zhong group has mapped the interactions between nuclear speckle-associated RNAs (nsaRNAs) and chromatin. This is achieved by ligating RNA, a linker and proximal DNA to each other, which is then converted into a sequencing library to determine RNA–DNA contacts [45]. This work revealed interactions with A compartments and nsaRNAs in HEK cells, but interestingly not in embryonic stem cells. Split-pool recognition of interactions

by tag extension (SPRITE) uses a unique proximity-ligation independent method which relies on a split-and-pool barcoding approach on formaldehyde cross-linked samples to infer higher order structures in cells. Reminiscent of Hi-C data, SPRITE shows that inter-chromosomal interactions are segregated into two hubs, which were referred to as the 'active hub' which correlates with high gene density and high RNA polymerase II transcription and the 'inactive hub' which correlates with low gene density and low transcriptional activity. Interestingly, the inactive hub was found to be enriched for rRNA–DNA contacts and shown to be proximal to the nucleolus, which is confirmed by FISH experiments. In contrast, the active hub, which is enriched with spliceosomal U1 snRNA and Malat1, was found to be associated with NS which was also corroborated by FISH and SC35 stainings. Finally, tyramide signal amplification (TSA-seq) from Belmont group took a more targeted approach and used antibodies against SON (or SC35) to biotinylate biomolecules in the vicinity of these proteins in formaldehyde cross-linked cells to identify DNA regions that are proximal to NS [46]. Since labeling in TSA-seq depends on the local diffusion of biotin-tyramide free radicals, optimized labeling conditions have been used to translate sequence coverage obtained from TSA-seq into quantitative distance measurements with an estimated accuracy of < 100 nm. While the results of TSA-seq are in general agreement with MARGI and SPRITE, in that transcriptionally active regions tend to be closer to NS, the resolution of the method revealed a striking enrichment of A1 subcompartment with NS. Moreover, top 5% percentile of SON TSA-seq was shown to be deterministically close to NS with a calculated mean distance of < 0.32 μm leading to the terms SPADs (speckle-associated domains) for these regions. Collectively, these studies clearly show that actively transcribed regions tend to be close to NS, but since they all describe the state of the cells at the time of cross linking, they cannot resolve the causality of events: do high levels of transcription summon NS to the vicinity of active loci, or do NS pull actively transcribed loci in toward them? Another related open question is if NS contribute to A/B compartmentalization, or whether they simply follow already established higher order chromatin structures. A recent study [47] has addressed this very question by depleting SRRM2 in mouse AML12 cells and characterizing chromatin interactions using Hi-C. Their analyses show that A/B compartments remain unaffected in SRRM2-depleted cells; however, they report a significant reduction in A-A contacts, which is more severe in A1-A1 contacts, and a significant increase in B-B contacts. Since SRRM2 depletion alone is not sufficient to dissolve NS in human cells [13], Hi-C experiments in cells depleted of SON and SRRM2, or SON alone will likely be highly informative with regard to the contribution of NS to chromatin compartmentalization, or based on results of Hu *et al.* more likely to intracompartmental (A or B) interactions.

NS in evolution

Nuclear speckle biology is almost exclusively studied in mammalian cells, where they have been first identified. In addition to mammalian cells, NS-like structures have been reported in Xenopus oocytes [48] and Drosophila embryos [49]; however, whether or not these structures are functionally analogous to NS of mammals is unclear. The cytological definitions of NS make it practically impossible to look for and verify whether NS exists in other animals or unicellular eukaryotes. We therefore leveraged our operational definition of NS and looked for orthologues of SRRM2 and SON, mainly in Metazoa. Our results show that orthologues of these two proteins exist virtually in every multicellular animal species that was investigated, albeit with dramatic differences in protein lengths. In yeast, where, to our knowledge, NS-like structures have not been reported, no SON orthologue could be found. However, several G-patch-containing proteins in yeast [50], which are involved in splicing and rRNA processing through their modulation of cognate RNA helicases, show some homology at this small stretch (~ 40 amino acids) to the C-terminal G-patch domain of SON, potentially indicating an evolutionary link to splicing. SRRM2 orthologues do exist in yeast, though relatively small in size (150–300 amino acids) corresponding to the N terminus of vertebrate SRRM2 orthologues. Curiously both the yeast protein (Cwf21p) and SRRM2 are located at exactly the same position in cryo-EM structures of spliceosomes and are therefore at least in this context are likely functionally analogous [14]. In fact, human SRRM2 has been shown to rescue a temperature-sensitive phenotype in complexed with CEF1 protein 21 (*CWC21*)/*ISY1* mutants in yeast [51]. Two related frog species *Xenopus laevis* and *Xenopus tropicalis* stand out in this phylogenetic analysis with their exceptionally long SON orthologues (Fig. 3A,B). Amphibians are unique in evolutionary biology as they are the closest animals to the last common ancestors of the first terrestrial vertebrates which needed to go through drastic physiological changes in order to adapt to life on land, and the dramatic length extension of SON and SRRM2 could

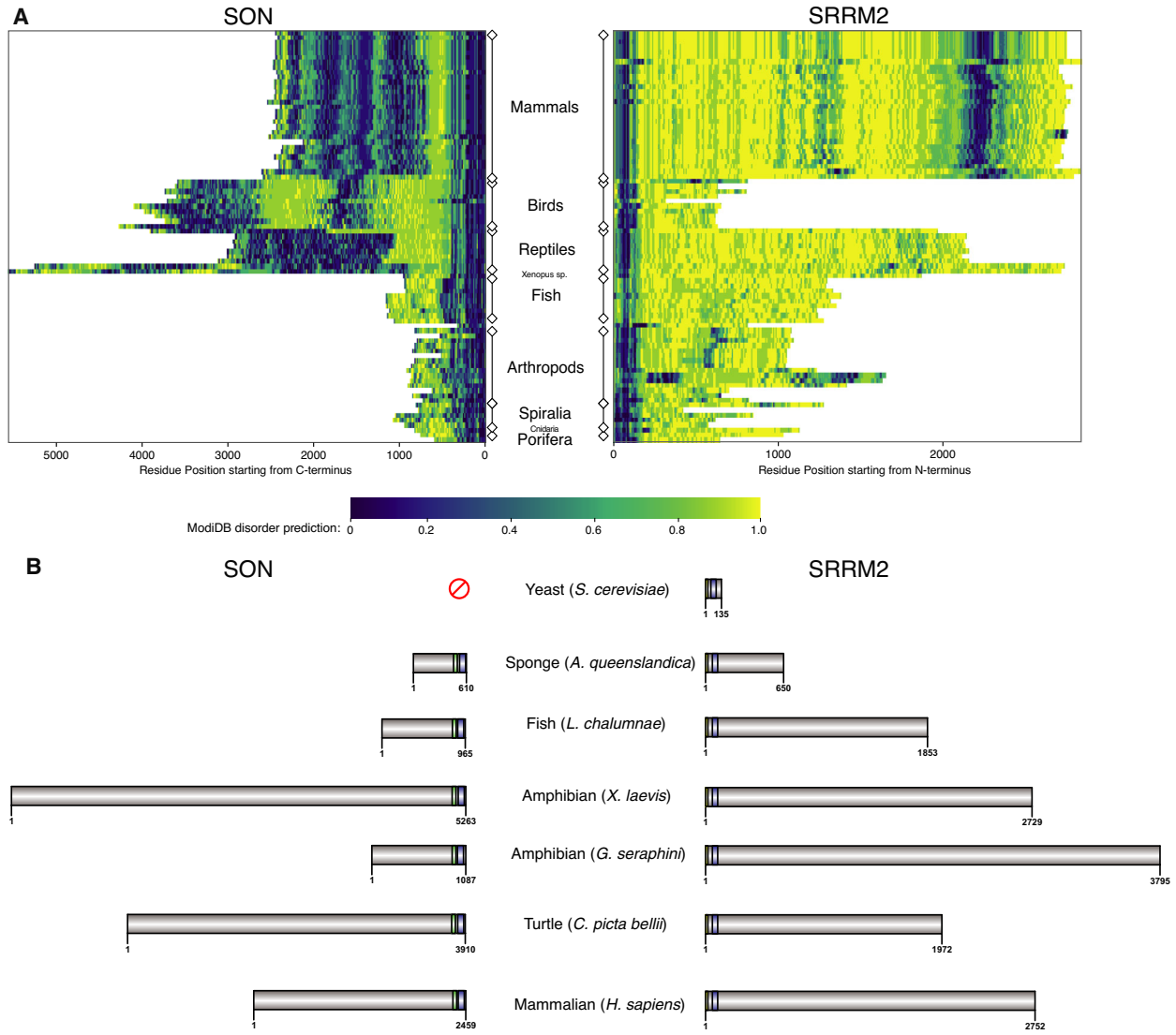

Fig. 3. (A) The disorder probability and the protein length of SON and SRRM2 in metazoa is shown (disorder predictions are obtained from MobiDB-Lite algorithm with the colors scaled from dark blue to yellow. Indicating a decrease in order as the value approaches 1.0 (yellow)). (B) Protein length and domain organization of SON and SRRM2 in yeast (*S. cerevisiae*), sponge (*A. queenslandica*), fish (*L. chalumnae*), amphibians (frog *X. laevis* and caecillian *G. seraphini*), turtle (*C. picta bellii*), and human are shown. Depicted domains for SON are; double stranded RNA-binding domain in blue, G-patch in green. Depicted domains for SRRM2 are; conserved N terminus in yellow, Cwf21 domain in blue.

be part of this process. It is however important to note that amphibians are extremely diverse animals and have the largest reported variation in genome size among vertebrates [52]. A group of limbless amphibians called Caecilians possess shorter SON orthologues similar to fish, but these amphibians also have the longest SRRM2 orthologues found in vertebrates (> 3500 amino acids). Curiously, these length extensions and possible contractions at the protein level correlate almost perfectly with the length of one long coding exon for both SON and SRRM2 (5.9 and 6.7 kb in humans, respectively, Fig. 4A,B). The mechanism behind these exon length changes is not clear, but splice isoforms where the long exon is not present in the final transcript are reported for both SON and SRRM2 in humans. A single exon encoding most of the IDRs of both SON and SRRM2 can therefore provide regulation at two very different time scales: at the fast, cellular level, biophysical properties of the NS can be modified on-the-fly by alternative splicing, while at evolutionary time scales, bulk changes in NS can be achieved by increasing or shortening of these

critical exons. Interestingly, SRRM2, but not SON, SRRM1, or SRRM3 is mislocalized to cytoplasmic tau aggregates which are a hallmark of several neurodegenerative diseases including frontotemporal dementia and Alzheimer's disease [53]. This mislocalization is only observed with full-length SRRM2 and not with SRRM2 without its C-terminal IDR which is encoded mostly by its longest coding exon (Fig. 4).

It is tempting to speculate that these length changes are adaptive in nature and that fluctuations in IDR lengths could alter biophysical properties of NS to cope with diverse genomic and metabolic challenges faced by these organisms [54]; however, cell culture models of these species or precise genetic manipulation of the mammalian orthologs to mimic these changes will be essential to test these ideas in the laboratory.

The long road to function(s) of NS

Instead of directly tackling the question 'What is the function of NS?', which remains without a compelling answer, we can reformulate it as 'What problem(s) do the NS solve?' which might be easier to unpack with the data currently available. Based on the phylogenetic data present for SON and SRRM2, the evolutionary origins of NS seem to be deeply rooted in splicing, which is supported by earlier and more recent studies on their composition, and a plethora of microscopic evidence locating, in both live and fixed cells, factors that are crucial for splicing in NS. It has been long assumed, and experimentally shown, that there are two distinct mechanisms that delineate exons and introns: one that involves recognition of intronic sequences or 'intron definition' and one that involves recognition of exonic sequences or 'exon definition' [55]. These two mechanisms are not completely mutually exclusive; however simpler eukaryotes rely mostly on the 'intron definition' route, whereas multicellular animals, with an expansion of introns in both number and size, typically rely on 'exon definition' to delineate intron–exon boundaries. As introns get larger through evolution, either via new transposon insertions, repeat expansions or other mechanisms, the distance between exons also expands, making it more difficult for consecutive exons to find each other in the 3D space of the nucleus, which is essential for splicing. Therefore, a mechanism bridging, or bringing exons close together, would be beneficial to organisms that experience rapid intron expansions. NS, forming a cohesive body for SR proteins, are prime candidates for such a mechanism [56], which is supported by SRRM2's direct involvement in catalytic stages of splicing and absence of NS-like structures in yeast. However, this hypothesis does not explain why a dense nuclear compartment is necessary for such a function. It is conceivable that RBPs at high enough concentrations could find each other and their target exons without a phase-separated nuclear domain nearby, and auxiliary factors such as SON and/or SRRM2 could execute their bridging functions without forming a large nuclear body. In this sense, NS could be a consequence of high concentrations of RBPs with low-complexity domains prone to phase separation necessary to faithfully execute splicing programs, rather than a biological innovation meant to solve a transcriptional/RNA-processing problem. In such a scenario, the role of SON and SRRM2 might be to passively sequester aggregation-prone RBPs into molecular communities. The interplay between SON and SRRM2 in human cells illustrates this point in a very interesting way: depletion of SON leads to partial delocalization of some NS-associated proteins (such as SRSF1), but leaves behind spherical NS enriched with SRRM2. At first, these experiments lend support to the idea that NS prevent uncontrolled aggregation of nuclear RBPs, since the removal of a critical component leads to what appears to be nonfunctional NS. However, simultaneous depletion of SON and SRRM2, or depletion of SON and genetical removal of SRRM2's IDR-rich C terminus, leads to the dissipation of all NS-like bodies [13], meaning that SR proteins are able to stay soluble without NS and do not spontaneously form visible phase-separated compartments reminiscent of NS without SON and SRRM2. Collectively, these results suggest that NS formation is a regulated process and not a simple by-product of high concentrations of SR proteins in the nucleus.

In addition to splicing, a direct role of NS in regulating transcription has been elegantly demonstrated by the recent study from Belmont group which shows that spatial proximity to NS directly correlates with a 'gene expression amplification', which was shown by live-cell imaging of heat-shock responsive genes [46], and in their most recent work by comparing distance of endogenously expressed genes to NS in four different cell lines using TSA-Seq [57]. Remarkably, proximity of heat-shock responsive genes correlate, in real time, with their transcriptional output, meaning that the transcriptional boost observed for an allele close to NS is lost if it then moves away from NS. Interestingly, since the model genes used in these experiments are intronless, this observed effect is not necessarily linked to splicing.

Robust but dynamic partitioning of the genome into active and inactive domains (loosely corresponding to euchromatin and heterochromatin) is crucial not only for executing developmental programs but for general

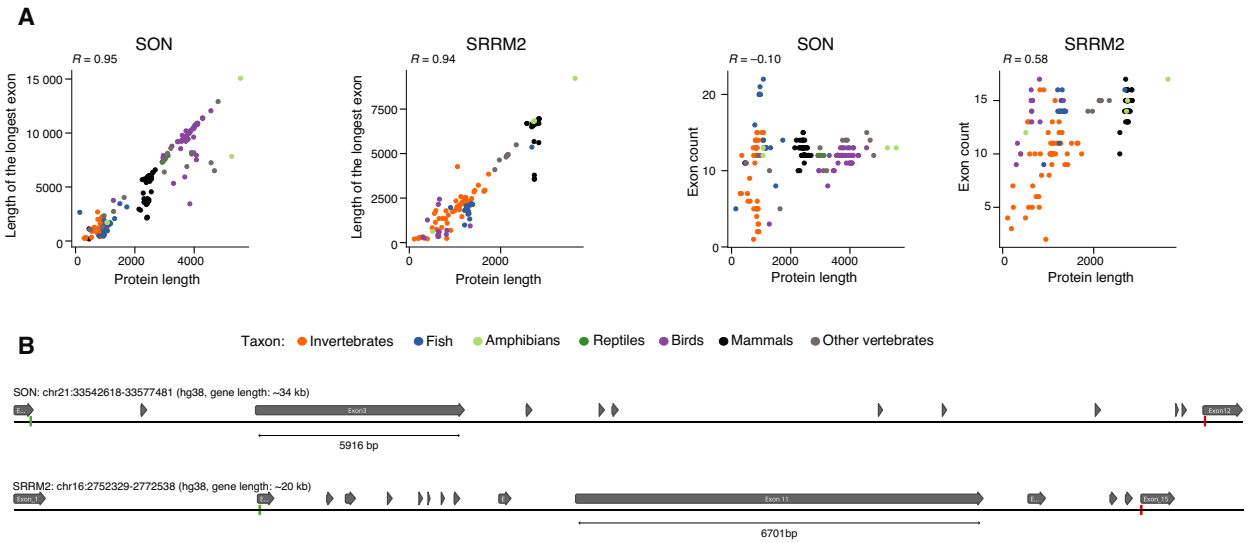

Fig. 4. (A) The length of protein in relation to the length of the longest exon or to the exon count is plotted. Both for SON and SRRM2, increase in the length of the longest exon is correlated with the increase in protein size (with R = 0.95 and R = 0.94, respectively). (B) The human gene locus for SON and SRRM2 are shown. The longest exon of the 34kb long *SON* gene is exon 3 and is 5916bp. The longest exon of the 20kb long *SRRM2* gene is exon 11 and is 6701bp. Green lines depict the start codon and the red lines depict the stop codon.

homeostasis within the cell. Spreading of the active chromatin environment into heterochromatin regions can be detrimental to cells, for example, by activating transposable elements. Similarly, spreading of heterochromatin into actively transcribed regions can lead to suppression of essential genes leading to cell death. Exact mechanisms that prevent mixing of these states are not yet elucidated; however, the role of local RNA (coding or noncoding) together with transcription factors and RBPs with low-complexity domains are becoming more and more clear. In addition to specific RNAs, such as roX RNAs of fruit flies activating the paternal X chromosome [58], and XIST of mammals repressing one of the two maternal X chromosomes [59], it is proposed that nascent RNAs can serve as a platform that recruit and locally concentrate transcription/RNA-processing factors, creating a feed-forward loop augmenting transcriptional activity and creating a physical barrier to spreading of heterochromatin. Based on the transcriptional boosting by NS, initially pioneered by the Lawrence group and demonstrated genome-wide by Belmont group, together with genome-wide association of NS with active chromatin (work of Belmont and Guttman groups), the highest level role of NS in mammalian nuclei could to be to create a positive-feedback loop around transcriptionally active loci, dependent on the nascent RNA acting as an attractant for NS-associated factors that are also involved in transcription elongation such as SRSF1 and SRSF2 [60], and keeping these factors at high local concentrations from forming insoluble aggregates. The underlying biological problem necessitating such a solution could be the need to concentrate transcription and RNA-processing factors on islands of transcriptionally active loci within the ever expanding Metazoan genomes, particularly driven by bursts of transposon activity which creates two problems for cells: when new insertions land on intronic regions, distance between exons increase, making cells more reliant on 'exon definition' which is dependent on SR and SR-like proteins. New insertions landing outside of genes must be silenced through heterochromatin formation, which needs to be kept in check and prevented from spreading over euchromatic regions. Unicellular eukaryotes such as yeast do not tolerate transposons to the extent higher eukaryotes do and thus do not have to deal with these problems, which is in line with the lack NS-like nuclear domains, absence of a SON orthologue, lack of A/B-type compartmentalization [61] and reliance on 'intron definition'. Pending experimental validation, phylogenetic analysis of SON also suggests that NS have originated during the transition into multicellularity, which is accompanied by intron expansion events [62].

In this model, transcriptional activation of genetic loci, primarily dictated by DNA-binding transcription factors, is stabilized and boosted by NS, which simultaneously bring RNA-processing machinery, just-in-time, where they are needed. In addition to the studies discussed above, a recent study from Young group

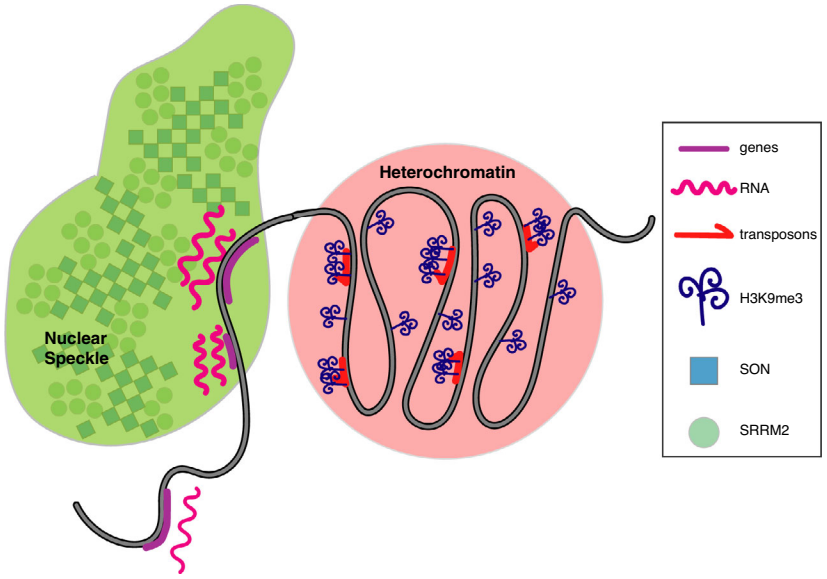

Fig. 5. A model of NS-driven gene activation and the separation of active vs. inactive domains within the nucleus are shown. According to this model, genes coming in close proximity with NS produce higher quantities of RNA than the ones that are further away. Nascent RNA emerging from RNA polymerase II can attract RBPs that are closely associated with NS, creating a positive-feedback loop that supports high levels of transcription and efficient RNA processing.

showing partitioning of RNA polymerase II from transcriptional condensates to splicing condensates through CTD phosphorylation [63] are also in agreement with this model. Nascent RNA is the most likely bridge between chromatin and NS-associated factors, since inhibition of transcription by small molecules leads to retraction of NS into spherical bodies [64], which can provide additional opportunities for transcriptional regulation since the nucleotide composition of nascent RNAs will dictate the RBPs they will interact with which can attract or repel NS (Fig. 5). Testing the fundamentals of this model is now possible by quantifying changes in RNA polymerase II occupancy and chromatin contacts by methods such as Hi-C or SPRITE in cells depleted of SON and SRRM2. The most important challenge remaining will be to uncouple mitotic defects caused by SON depletion, which can be overcome by either focusing on shorter time scales or engineering a system where SON can be depleted in nondividing cells.

Future perspectives

Research on NS closely follows technological advances in studying biological systems, beginning with acid-aniline staining techniques perfected by Cajal which revealed hitherto unknown cellular structures such as NS, to electron microscopy techniques that allowed visualization of the ultrastructural organization of NS, followed by advances in live-cell imaging techniques that allowed investigation of biophysical properties of NS in living cells, and finally techniques fueled by high-throughput sequencing that brought unprecedented insights into interactions between NS and chromatin. Thanks to large-scale genome sequencing efforts, we are now in the midst of a new era in biology where we can trace the evolutionary history of genes and proteins *in silico*, infer evolutionary origins of biological phenomena and test hypotheses in cellular, organismal, and organoid models. Deployment of these innovative techniques on nonmodel or understudied organisms, informed by evolutionary analyses, will further demystify NS and their roles in human cells.

Acknowledgements

The authors would like to thank Aktas group members, especially Olga Jasnovidova and Michal Malszycki for their help with the figures and anonymous reviewers for their helpful suggestions. The authors would also like to acknowledge Max Planck Society for supporting their work.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

IAI and TA planned the outline together and IAI wrote the paper with input from TA. TA edited the paper and reviewed the final version.

References

1 y Cajal SR (1910) El núcleo de las células piramidales del cerebro humano y de algunos mamíferos Travaux du Laboratoire de Recherches Biologiques de l'Universite de Madrid.

2 Spector DL & Lamond AI (2011) Nuclear speckles. *Cold Spring Harb Perspect Biol* **3**, a000646.

3 Galganski L, Urbanek MO & Krzyzosiak WJ (2017) Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* **45**, 10350–10368.

4 Berchtold D, Battich N & Pelkmans L (2018) A systems-level study reveals regulators of membrane-less organelles in human cells. *Mol Cell* **72**, 1035–1049.e5.

5 Dundr M & Misteli T (2010) Biogenesis of nuclear bodies. *Cold Spring Harb Perspect Biol* **2**, a000711.

6 Misteli T, Cáceres JF & Spector DL (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523–527.

7 O'Keefe RT, Mayeda A, Sadowski CL, Krainer AR & Spector DL (1994) Disruption of pre-mRNA splicing in vivo results in reorganization of splicing factors. *J Cell Biol* **124**, 249–260.

8 Prasanth KV, Sacco-Bubulya PA, Prasanth SG & Spector DL (2003) Sequential entry of components of the gene expression machinery into daughter nuclei. *Mol Biol Cell* **14**, 1043–1057.

9 Sacco-Bubulya P & Spector DL (2002) Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. *J Cell Biol* **156**, 425–436.

10 Rai AK, Chen J-X, Selbach M & Pelkmans L (2018) Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216.

11 Fu XD & Maniatis T (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437–441.

12 Fu XD & Maniatis T (1992) Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science* **256**, 535–538.

13 Ilik İA, Malszycki M, Lübke AK, Schade C, Meierhofer D & Aktaş T (2020) SON and SRRM2 are essential for nuclear speckle formation. *Elife* **9**, e60579.

14 Jia X & Sun C (2018) Structural dynamics of the N-terminal domain and the Switch loop of Prp8 during spliceosome assembly and activation. *Nucleic Acids Res* **46**, 3833–3840.

15 Blencowe BJ, Nickerson JA, Issner R, Penman S & Sharp PA (1994) Association of nuclear matrix antigens with exon-containing splicing complexes. *J Cell Biol* **127**, 593–607.

16 Sharma A, Takata H, Shibahara K-I, Bubulya A & Bubulya PA (2010) Son is essential for nuclear speckle organization and cell cycle progression. *Mol Biol Cell* **21**, 650–663.

17 Ahn E-Y, DeKelver RC, Lo M-C, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu X-D & Zhang D-E (2011) SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol Cell* **42**, 185–198.

18 Fei J, Jadaliha M, Harmon TS, Li ITS, Hua B, Hao Q, Holehouse AS, Reyer M, Sun Q, Freier SM *et al.* (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J Cell Sci* **130**, 4180–4192.

19 Monneron A & Bernhard W (1969) Fine structural organization of the interphase nucleus in some mammalian cells. *J Ultrastruct Res* **27**, 266–288.

20 Spector DL (1984) Colocalization of U1 and U2 small nuclear RNPs by immunocytochemistry. *Biol Cell* **51**, 109–112.

21 Reuter R, Appel B, Bringmann P, Rinke J & Lührmann R (1984) 5'-Terminal caps of snRNAs are reactive with antibodies specific for 2,2,7-trimethylguanosine in whole cells and nuclear matrices. Double-label immunofluorescent studies with anti-m3G antibodies and with anti-RNP and anti-Sm autoantibodies. *Exp Cell Res* **154**, 548–560.

22 Mintz PJ, Patterson SD, Neuwald AF, Spahr CS & Spector DL (1999) Purification and biochemical characterization of interchromatin granule clusters. *EMBO J* **18**, 4308–4320.

23 Saitoh N, Spahr CS, Patterson SD, Bubulya P, Neuwald AF & Spector DL (2004) Proteomic analysis of interchromatin granule clusters. *Mol Biol Cell* **15**, 3876–3890.

24 Rappsilber J, Ryder U, Lamond AI & Mann M (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**, 1231–1245.

25 Zhou Z, Licklider LJ, Gygi SP & Reed R (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182–185.

26 Wahl MC, Will CL & Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718.

27 Neugebauer KM (2019) Nascent RNA and the Coordination of Splicing with Transcription. *Cold Spring Harb Perspect Biol* **11**, a032227.

28 Dopie J, Sweredoski MJ, Moradian A & Belmont AS (2020) Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *J Cell Biol* **219**, e201910207.

29 Fong K-W, Li Y, Wang W, Ma W, Li K, Qi RZ, Liu D, Songyang Z & Chen J (2013) Whole-genome screening identifies proteins localized to distinct nuclear bodies. *J Cell Biol* **203**, 149–164.

30 Politz JCR, Tuft RA, Prasanth KV, Baudendistel N, Fogarty KE, Lifshitz LM, Langowski J, Spector DL &

Pederson T (2006) Rapid, diffusional shuttling of poly (A) RNA between nuclear speckles and the nucleoplasm. *Mol Biol Cell* **17**, 1239–1249.
31 Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA *et al*. (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**, 925–938.
32 Salichs E, Ledda A, Mularoni L, Albà MM & de la Luna S (2009) Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet* **5**, e1000397.
33 Cáceres JF, Misteli T, Screaton GR, Spector DL & Krainer AR (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* **138**, 225–238.
34 Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV & Jedd G (2020) Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *Mol Cell* **77**, 1237–1250.e4.
35 Hall LL, Smith KP, Byron M & Lawrence JB (2006) Molecular anatomy of a speckle. *Anat Rec A Discov Mol Cell Evol Biol* **288**, 664–675.
36 Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A & Lawrence JB (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* **33**, 717–726.
37 Jain A & Vale RD (2017) RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247.
38 Urbanek MO, Jazurek M, Switonski PM, Figura G & Krzyzosiak WJ (2016) Nuclear speckles are detention centers for transcripts containing expanded CAG repeats. *Biochim Biophys Acta* **1862**, 1513–1520.
39 Mori K, Weng S-M, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C *et al*. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335–1338.
40 Ruff KM, Dar F & Pappu RV (2021) Ligand effects on phase separation of multivalent macromolecules. *Proc Natl Acad Sci USA* **118**, e2017184118.
41 Kempfer R & Pombo A (2020) Methods for mapping 3D chromosome architecture. *Nat Rev Genet* **21**, 207–226.
42 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO *et al*. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293.
43 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al*. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680.
44 Shopland LS, Johnson CV, Byron M, McNeil J & Lawrence JB (2003) 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.
45 Chen W, Yan Z, Li S, Huang N, Huang X, Zhang J & Zhong S (2018) RNAs as proximity-labeling media for identifying nuclear speckle positions relative to the genome. *iScience* **4**, 204–215.
46 Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, van Steensel B, Ma J & Belmont AS (2018) Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol* **217**, 4025–4048.
47 Hu S, Lv P, Yan Z & Wen B (2019) Disruption of nuclear speckles reduces chromatin interactions in active compartments. *Epigenetics Chromatin* **12**, 43.
48 Gall JG, Bellini M, Wu Z & Murphy C (1999) Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol Biol Cell* **10**, 4385–4402.
49 Ségalat L & Lepesant JA (1992) Spatial distribution of the Sm antigen in Drosophila early embryos. *Biol Cell* **75**, 181–185.
50 Aravind L & Koonin EV (1999) G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem Sci* **24**, 342–344.
51 Grainger RJ, Barrass JD, Jacquier A, Rain J-C & Beggs JD (2009) Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15**, 2161–2173.
52 Liedtke HC, Gower DJ, Wilkinson M & Gomez-Mestre I (2018) Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat Ecol Evol* **2**, 1792–1799.
53 Lester E, Parker R, Ooi FK, Bakkar N, Ayers J, Woerman AL, Wheeler JR, Bowser R, Carlson GA & Prusiner SB. (2021) Tau aggregates are RNA-protein assemblies that mis-localize multiple nuclear speckle components. *Cold Spring Harb Lab* **109**, 1675-1691.e9.
54 Haltenhof T, Kotte A, De Bortoli F, Schiefer S, Meinke S, Emmerichs A-K, Petermann KK, Timmermann B, Imhof P, Franz A *et al*. (2020) A conserved kinase-based body-temperature sensor globally controls alternative splicing and gene expression. *Mol Cell* **78**, 57–69.e4.
55 De Conti L, Baralle M & Buratti E (2013) Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* **4**, 49–60.

56 Liao SE & Regev O (2020) Splicing at the phase-separated nuclear speckle interface: a model. *Nucleic Acids Res* **49**, 636–645.
57 Zhang L, Zhang Y, Chen Y, Gholamalamdari O, Wang Y, Ma J & Belmont AS (2020) TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Res* **31**, 251–264.
58 Valsecchi CIK, Basilicata MF, Georgiev P, Gaub A, Seyfferth J, Kulkarni T, Panhale A, Semplicio G, Manjunath V, Holz H *et al.* (2021) RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature* **589**, 137–142.
59 McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A *et al.* (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236.
60 Lin S, Coutinho-Mansfield G, Wang D, Pandit S & Fu X-D (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* **15**, 819–826.
61 Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA & Noble WS (2010) A three-dimensional model of the yeast genome. *Nature* **465**, 363–367.
62 Lozada-Chávez I, Stadler PF & Prohaska SJ (2018) Genome-wide features of introns are evolutionary decoupled among themselves and from genome size throughout Eukarya. *bioRxiv*, 283549. [PREPRINT].
63 Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, Spille J-H, Afeyan LK, Zamudio AV & Shrinivas K *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548.
64 Kim J, Han KY, Khanna N, Ha T & Belmont AS (2019) Nuclear speckle fusion via long-range directional motion regulates speckle morphology after transcriptional inhibition. *J Cell Sci* **132**, jcs226563.