Chromatin domains in space and their functional implications
Frederic Pontvianne, Chang Liu

To cite this version:

HAL Id: hal-02428536
https://hal-univ-perp.archives-ouvertes.fr/hal-02428536
Submitted on 6 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Chromatin domains in space and their functional implications

Frédéric Pontvianne\textsuperscript{1,2} and Chang Liu\textsuperscript{3}

\textsuperscript{1}CNRS, Laboratoire Génome et Développement des Plantes (LGDP); Université de Perpignan Via Domitia, LGDP, UMR 5096, Perpignan 66860, France

\textsuperscript{2}UPVD, Laboratoire Génome et Développement des Plantes (LGDP); Université de Perpignan Via Domitia, LGDP, UMR 5096, Perpignan 66860, France

\textsuperscript{3}Center for Plant Molecular Biology (ZMBP), University of Tübingen, Auf der Morgenstelle 32, Tübingen 72076, Germany

Correspondence:

frederic.pontvianne@univ-perp.fr; ORCID ID: 0000-0002-2913-4104

chang.liu@zmbp.uni-tuebingen.de; ORCID ID: 0000-0003-2859-4288
Abstract

Genome organization exhibits functional compartmentalization. Several factors, including epigenetic modifications, transcription factors, chromatin remodelers, and RNAs shape chromatin domains and the three-dimensional genome organization. Various types of chromatin domains with distinct epigenetic and spatial features exhibit different transcriptional activities. As part of the efforts to better understand plant functional genomics, over the past a few years, spatial distribution patterns of plant chromatin domains have been brought to light. In this review, we discuss chromatin domains associated with the nuclear periphery and the nucleolus, as well as chromatin domains staying in proximity and showing physical interactions. The functional implication of these domains is discussed, with a particular focus on the transcriptional regulation and replication timing. Finally, from a biophysical point of view, we discuss potential roles of liquid-liquid phase separation in plant nuclei in the genesis and maintenance of spatial chromatin domains.

Introduction

In eukaryotes, the nuclear DNA is wrapped around histone octamers to form the chromatin. Chromatin is subject to extensive modifications including DNA methylation and post-translational histone modifications [1]. These modifications, also named epigenetic marks, form the epigenome. To understand the three-dimensional genome organization in relation to local epigenetic states, it is also necessary to consider the subnuclear components that include (i) nuclear bodies such as the nucleolus, nuclear speckles and Cajal bodies, as well as (ii) nuclear pores and the nuclear periphery [2,3]. In mammalian cells, large chromatin regions associate at the nuclear periphery with a network composed of lamin fibers are named Lamina-associated domains (LADs) [4]. Some chromatin domains also associate with the nucleolar periphery, which actually belongs to nucleolus, and are named nucleolus-associated chromatin domains (NADs) [5,6]. Besides, mammal genomes predominantly form thousands of self-organizing chromatin domains known as topologically associated domains (TADs), which are relatively insulated from one another [7]. In plants, chromatin domains comparable to animal LADs, NADs, and TADs have been found. It should be pointed out here that our knowledge of these plant chromatin domains is still preliminary, at the moment, they cannot be deemed fully equivalent to their animal analogues.
Genome organization is also highly dynamic, and is subjected to changes according to the cell cycle progression, developmental transition like commencing photomorphogenesis or flowering, and external cues [8]. For instance, in the presence of light, as a result of progressive compaction of heterochromatin, nuclei in germinating Arabidopsis seedlings produce chromocenters, which appear as large, bright spots upon stained with DAPI [9]. How are chromatin organization patterns, with a certain degree of orderliness in space, formed? For long, affinity between different molecules was thought to be the most important force determining how they are distributed in space. A protein can diffuse through the nucleus and thanks to its affinity and specificity to other factors, this protein might be retained longer in some nuclear compartment than others [10]. Recent advances also revealed the potential role of proteins possessing intrinsically disordered regions (IDRs) in the establishment and maintenance of nuclear compartments [11,12]. In this short review, we refer to plant “chromatin domains” as chromatin regions identified with methods concerning three-dimensional (3D) chromatin organization and positioning. With a focus on the demarcation and functionality of selected plant chromatin domains, we summarize and discuss recent progress in plant three-dimensional (3D) genomics.

**Identification of plant chromatin domains from a 3D perspective**

**Functional annotation of plant long-range cis-regulatory elements**

Besides identifying functional chromatin domains via acquiring a detailed picture of epigenomic and structural features (e.g., by using ChIP-seq and ATAC-seq approaches), investigating 3D chromatin conformation provides complementary structural and functional insights into them. In particular, this information is crucial for identifying gene(s) regulated by a given candidate enhancer element and **vice versa**. In the past decade, Hi-C (Chromosome Conformation Capture coupled with High Throughput Sequencing) has become the most widely used approach to study physical chromatin contact networks in 3D [13, 14]. Hi-C approaches have been applied to a variety of plant species, from which both expected and surprising chromatin organization patterns as opposed to animals have been discovered (reviewed recently in [15-17]). Similar to those in animals, chromatin compartmentalization and local chromatin insulation have been observed in plants, implying that they can prevent
chromatin regions from freely interacting with one another. Such spatial constraints of chromatin contacts are part of how distal cis-regulatory elements regulate expression of their target genes via establishing specific long-range physical interactions. Over the past few years, there have been increasing efforts in systematically identifying cis-regulatory elements and enhancers in various plant species, such as Arabidopsis [18-22], rice [22-24], tomato [22,25], maize [26], and wheat [27]. These approaches are based on searching for chromatin regions with local structural and epigenetic features similar to those in animal genomes. A challenge downstream of this approach is how to correctly annotate these potential regulatory elements by assigning them to their target gene loci.

On the contrary, Hi-C can provide researchers important information regarding chromatin domain interactions; however, Hi-C has limited sensitivity in systematically detecting chromatin loops, as it is financially costly to increase the sequencing depth of a genome-wide Hi-C map to boost the statistical power of loop calling. Nonetheless, Hi-C studies in Arabidopsis [28], rice [29], and cotton [30] show that chromatin regions involved in forming chromatin loops are enriched at gene promoters, reflecting the existence of extensive yet largely uncharted contacts between genes and their regulatory elements in plants. Compared to using Hi-C, one can better resolve spatial organization among chromatin domains with approaches that dedicate sequencing resource to genomic regions of interest. For instance, both the ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing) and HiChIP (Hi-C Chromatin Immunoprecipitation) methods aim to reveal chromatin interaction networks of regions associated with a defined chromatin mark or transcriptional regulator [31,32]. Recently, several studies using ChIA-PET unveiled chromatin interaction patterns associated expressed genes in maize and rice [33-35]. The two maize ChIA-PET studies by Li et al. [33] and Peng et al. [34] focused on chromatin domains with H3K4me3, H3K27ac, and RNA Pol2, which were hallmarks of active promoter, enhancer, and transcribed regions, respectively. Collectively, their work identified unprecedented networks of promoter-enhancer and promoter-promoter interactions in maize, some of which were well known as contributors of important agronomic traits. Likewise, a recent ChIA-PET study of rice revealed physical interactions between many eQTLs (expression Quantitative Trait Loci) and their target genes [35].

In summary, these work demonstrate the advantage of identifying and annotating functional regulatory chromatin regions by integrating both one- and three-dimensional genomic features. In our opinion, a combinatory strategy with two steps can be
considered as a standard practice for functional annotation of regulatory elements in a given plant genome. The first step involves identifying chromatin regions with features of interest (e.g., epigenetic marks), and the second step involves using Hi-C-related methods that explore their chromatin-chromatin interaction network.

Identifying plant LADs and NADs

Another way of annotating 3D chromatin domains is based on their localization in the nucleus. In animals, active and repressed chromatin regions tend to be separated from each other, and some areas in the nucleus, such as nuclear periphery and nucleolar periphery, are enriched with repressed chromatin [36,37]. Recently, chromatin domains preferentially localized at the nuclear and/or nucleolar periphery in Arabidopsis have been identified (Figure 1a).

Arabidopsis perinuclear chromatin domains were initially identified with an artificial system, which did not reveal direct interactions between the nuclear envelope and these domains [38]. Nevertheless, these plant perinuclear chromatin domains were enriched with various repressive marks (e.g., H3K27me3 and DNA methylation), suggesting that the plant nuclear periphery was a compartment in favor of holding repressed genes [38]. Later on, it was shown that some plant-specific nuclear lamin candidate proteins, CROWDED NUCLEI (CRWN), were required to tether chromatin to the nuclear periphery in Arabidopsis [39,40]. By using CRWN1 as bait, chromatin domains bound by CRWN1 at the nuclear periphery (named plant LADs) were identified with chromatin immunoprecipitation [39]. Pattern analyses of plant LADs confirmed the previous conclusion that the plant nuclear periphery is a repressive environment [39]. On the other hand, the identification of NADs was achieved by isolating intact nucleoli [41,42]. In addition to ribosomal RNA loci, NADs are clearly enriched with lowly expressed protein-coding genes, as well as inactive chromatin marks and transposons [41]. Thus, plant LADs and NADs are both transcriptionally inactive; however, as they are located in different nuclear compartments, the respective silencing mechanisms might be different to a certain extent. For instance, the silencing of NAD-genes might be due to preventing RNA polymerase II from being associated with the nucleolus [43]. On the other hand, the plant lamin protein CRWN1 was shown to interact with PWWP INTERACTOR OF POLYCOMBS 1 (PWO1), which associated with Polycomb-group proteins, suggesting the involvement of H3K27me3-mediated transcriptional repression in LADs [44]. At the moment, research of plant
LADs and NADs are still at an infant stage, as the knowledge of proteins required for forming these chromatin domains is extremely limited. Also, it is not known how variable plant LADs and NADs demarcations are across different cell types and growth conditions. Given the highly dynamic nature of plant nuclei [45,46], we envisage that these plant chromatin domains possess a certain degree of flexibility, participating in modulating 3D genome organization and transcriptional regulations.

A comparison between Arabidopsis LADs and NADs revealed that a tiny fraction of the genome is enriched both at the nuclear periphery and the nucleolus [47]; notably, most of these domains overlap with pericentromeric regions at chromosome 4, and to a less extent with those at chromosome 2 (Figure 1b). The occurrence of interchangeable perinuclear and nucleolar chromatin domains has been found in animals before [5,48]. A recent study of NADs identification in mouse embryonic fibroblast cells reported that a small subset of NADs were also frequently associated with the nuclear lamina [49]. These chromatin domains, shared by LADs and NADs (named “type I NADs”), appeared to be more heterochromatic; while the other type of NADs (“type II NADs”) tend to be relatively promoting gene expression and enriched with developmentally regulated genes [49]. We speculate that the chromatin domains shuffling between the nuclear periphery and the nucleolus in plants might be functionally distinct from the domains without such dual localization. For the Arabidopsis genome, it would also be interesting to investigate whether these LAD/NAD interchangeable regions are involved in modulating dynamics of chromocenter (specifically chromosomes 2 and 4) structures during plants’ growth and development [50].

Functions of plant chromatin domains in 3D

In this section we discuss functional implications of the abovementioned chromatin domains.

co-expression of genes

In an earlier Hi-C work by Dong and colleagues, tomato and maize genomes were shown to form a large number of long-range chromatin loops linking interstitial active chromatin regions [51], suggesting spatial clustering of expressed genes. Later on, the interaction networks of maize active chromatin were revealed by two research groups using the ChIA-PET method, and suggested a role for these physical interactions on gene expression [33,34]. Albeit the datasets from these two teams are difficult to
compare due to the use of different growth conditions, tissues types, and antibodies
(for ChIP) [33,34], three consensus patterns can be extracted. Firstly, a substantial
fraction of the identified chromatin loops connects gene loci; secondly, genes forming
long-range chromatin interactions tend to show higher expression levels than those
without; thirdly, gene pairs linked with chromatin loops tend to show co-expression.
Based on a recent rice ChIA-PET study, coordinated expression of active genes can
also be found among those connected by chromatin loops [35]. Together, these results
strongly suggest that active chromatin domains in plant nuclei can form extensive
physical contacts via chromatin interactions.

Earlier studies of gene expression in several plant species have pointed out that it is
common to observe co-expression between neighboring genes [52-55]. An
explanation of this phenomenon is that neighboring genes (especially those with
overlapping divergent promoters) share some common cis-regulatory elements. The
promoters of neighboring gene can also contact with one another via forming
chromatin loops. Most of the reported plant promoter-promoter interactions are
between physically linked loci in the genome [33,34]. Considering chromatin as a
polymer, due to distance-dependent stochastic contacts, it is known that nearby
loci have much stronger contacts than do loci separated by large genomic
distances [56]. This correlates well with the fact that Hi-C maps, regardless of species
and cell types, always display strong contacts around their diagonal lines (indicative of
interactions over short genomic distances). We speculate that stochastic contacts
among loci along the chromatin fibre, as a function of genomic distance, contribute
significantly to interactions between promoters and cis-elements. In addition, we also
speculate that transcriptional regulators are involved in forming these chromatin
contacts (Figure 1a) (see discussion in the next section). Together, the cooperative
interactions among multiple transcribed loci form a spatial domain of “transcriptional
ecosystem equilibrium” in the nucleus that fosters co-expression patterns [57]. Such
physical interactions among active chromatin could be a mechanism underlying co-
expression of metabolic genes residing close to each other (i.e., members belonging
to a gene cluster annotated in the linear genome) [58,59].

**DNA replication timing**

Spatial chromatin domain distribution is not only associated with gene transcription
regulation, but also with other essential chromatin activities. As part of the cell cycle,
DNA replication is a process by which genomic content is duplicated before a cell enters mitosis. Interestingly, DNA replication timing across the genome is not homogeneous, rather, it displays a correlation to local histone marks and 3D chromosome structures [60]. In animals, euchromatin, which is localized in the nuclear interior, is replicated earlier than perinuclear localized heterochromatin [60]. Similarly, studies comparing chromatin regions with different replication timing patterns in maize root tip nuclei showed that open chromatin and densely packed heterochromatin domains tend to be duplicated in early and late S phases, respectively [61,62]. The same correlation was seen in Arabidopsis suspension cells, that repressed chromatin were enriched in late replicated loci [63,64]. Further, live imaging of Arabidopsis replisomes revealed their dynamic distribution in early and late S phase [65]. All these observations suggest that plant DNA duplication happens in accordance with different chromatin features (e.g., heterochromatin tends to be replicated late). As mentioned above, Arabidopsis nuclei show enrichment of repressed chromatin regions at the nuclear periphery and the nucleolus; therefore, it is expected that such chromatin compartmentalization correlate with late/early replication patterns. Indeed, for chromatin loci belonging to either LADs or NADs in Arabidopsis, they clearly show a preference for being replicated in the last S phase (Figure 2a, b). Interestingly, further analyses on DNA replication origins with these chromatin domains reveal that the distribution of leading nascent strands over LADs and NADs are different (Figure 2c, d). Overall, among the nascent strands identified in a recent study [66], LADs overlap more with those pointing inward; while NADs overlap more with those pointing outward, suggesting that the replication of these two types of repressed chromatin domains are regulated by different mechanisms. Recent work in mammals has led to the identification of Early Replicating Control Elements (ERCEs) that play roles in regulating both DNA replication timing and 3D chromatin organization [67]. Certainly, it would be interesting to study if plants also have such a mechanism that integrates DNA replication and chromatin organization. Many plant species can carry on endoreduplication, which is a process doubling the nuclear genome in the absence of mitosis [68]. As an extreme example, during tomato fruit development, the endopolyploidy level of pericarp cells can reach 512C (C is the haploid DNA content; and 512C means nine rounds of endoreduplication) [69]. So far, it is not known whether the recurring DNA replication during endoreduplication cycle is accompanied with changes in chromatin organization and epigenetic landscape.
Liquid-liquid phase separation (LLPS) as a prominent biophysical process implicated in arranging chromatin domains in 3D

Role of liquid-liquid phase separation in the nucleus

Nuclear sub-compartments are membrane-less organelles or condensates and are characterized by liquid-phase properties. In that case, they are liquid-phase compartments and remain separated from each other through liquid-liquid phase separation (LLPS). LLPS form nuclear condensate or droplets, and are generated by spontaneous nucleation of a given molecules resent at a high concentration. These phenomenon participate in the creation of functional hubs that allow the enrichment of factors required in a specific biological process such as mRNA biosynthesis or ribosome biogenesis. Recent advances clearly demonstrated the implication of LLPS in the establishment of non-membrane organelles in the nucleus [12]. Proteins with intrinsically disordered regions (IDRs) play a crucial role in the genesis and maintenance of phase-separated bodies. Recent work demonstrated that LLPS could act at the scale of large chromatin domains (i.e TAD or NAD), at the scale of a chromatin loop to participate in transcriptional regulation and also at the scale of the nucleosome [11,70,71]. For example, plant-specific Agenet Domain Containing Protein 1 (ADCP1) has recently been shown to drive the phase separation of H3K9me3-marked nucleosome arrays to form condensates [72]. Such a mechanism might be employed in the rice nucleoplasm to create physical contacts between multiple heterochromatin loci, which was revealed by a ChIA-PET study [35]. For example, nucleosome arrays behave like LLPS, with histone tail and linker histone H1 playing a substantial role in their level of compaction [73]. Interestingly, histone tail acetylation seems to be able to regulate LLPS of nucleosome arrays [74]. Another example is with the clustering of RNA polymerase II, which is due to LLPS mediated by the presence of IDRs in its C-terminal domain [75]. Furthermore, droplets generated via LLPS can potentially act as mechano-active chromatin filters. Most IDR-containing proteins indeed exclude chromatin, which explain why nuclear bodies usually display a low chromatin density. This mechanism facilitate chromatin factors to target genomic loci by changing their concentration in a given compartment [76] (Wei et al. 2019). For example, the MEDIATOR complex subunit MED1 was shown to form nuclear puncta at enhancers, concentrating RNA polymerase II to achieve desired expression levels at target loci [77].

IDRs are usually composed of Arginine/Glycine (R/G) rich and/or Glutamine/Asparagine (Q/N) rich domains [78,79]. One of the best-studied cases of
IDRs-mediated LLPS is the nucleolus [80]. In mammal cells, the R/G-rich domains of the nucleolar proteins nucleoplasmin and fibrillarin were both shown to be required for the formation of the nucleolus, as well as for the sub-nucleolar compartments [81]. In plants, there is no homolog of nucleoplasmin, but FIBRILLARIN 2 (FIB2), NUCLEOLIN 1 (NUC1), and many other nucleolar proteins possess strong IDRs (Figure 3). Thus, potential LLPS driven by these nucleolar proteins might be crucial for forming functional plant nucleoli. This hypothesis is supported by the fact that NUC1 disruption leads to the nucleolus disorganization [82,83].

Arabidopsis thaliana proteins with IDRs

Although the plant science community is aware of the potential importance of LLPS in shaping chromatin domains, there are few examples described in plants so far [15,84]. We therefore attempted to search for A. thaliana proteins containing IDRs with R/G-rich and/or Q/N-rich stretches (Figure 3a). Amongst the 27416 proteins encoded by the A. thaliana genome, 1234 R/G-rich and/or Q/N-rich IDR-containing proteins were identified (Supplemental Table 1). Interestingly, there are only 4 proteins that have both types of IDR motifs (Figure 3b). The 51 proteins containing at least 4 GGRG motifs are implicated in the RNA metabolism (GO:0016070; p value 3.5E-3) and are found in nuclear bodies like the nucleolus, nuclear speckles, photobodies or Cajal bodies (Figure 3a) (Love et al. 2017; Zhu and Brangwynne 2015; Montacie et al. 2017; Li et al. 2019). This list is also composed of proteins known to localize in cytoplasmic bodies like processing bodies and stress granules (e.g., DCP5 and AGO1).

Among the 80 proteins containing a long Q/N-rich stretch (at least 40 Q/N residues), half of them are implicated in transcriptional regulation (GO: 0006355; p value 1.05E-14). Notably, a member of this list, FCA, is involved in LLPS and required for proper transcriptional termination [85]. Our screen also identified a strong Q/N-rich IDR in NERD, a nuclear protein implicated in the 3' end formation of another subset of mRNAs [86]. In this case, proper mRNA termination requires both NERD and FIP37-dependent N6-adenosine mRNA methylation [86]. The fact that NERD forms nuclear foci through LLPS remains to be investigated. Additionally, we observed many transcription factors (e.g., MADS-box proteins) and mediators in this list (Figure 3a). MADS-box proteins have been long known for mediating chromatin looping via forming protein complexes [87,88]. MED25, which contains an extraordinarily strong Q/N-rich stretch, has been recently shown to be required for establishing chromatin contacts between enhancers
and target genes in the jasmonate signaling pathway in plants [89]. Although our IDR-containing protein list implies an existing 3D interaction network functioning in transcriptional regulation (e.g., promoter-promoter interactions and cooperative transcription), the functional implication of these Q/N-rich stretches remains to be evidenced in plants. A systematic analysis of the nuclear localization of plant IDR should lead to the discovery of proteins implicated in LLPS-dependent nuclear puncta formation.

**Perspectives**

Recent advances have greatly helped scientists better understand the mechanisms by which chromatin domains are brought together in 3D. It is noteworthy that the list of chromatin organization regulators is expanding rapidly. Recent work from Xiang-Dong Fu and colleagues revealed the presence of a large, unexpected subset of RNA-binding proteins at numerous chromatin sites [90]. As many RNA-interacting proteins are implicated in LLPS, it is reasonable to speculate that some of them may regulate chromatin looping and compartmentalization. This study also gives us a hint that the interactions between RNA-binding proteins and chromatin in plants might have been unsuspected.

As discussed earlier, there is a correlation between chromatin domains and their local epigenetic signatures. In addition, to assess the transcriptional regulation of a given gene, it is essential to identify all the direct and indirect interacting-factors of the gene, i.e., the other genomic regions, RNA, and proteins [57]. Moreover, the supra-molecular arrangements created by LLPS also seem to play a role in protein regulation via their retention as demonstrated for the DNA methyltransferase DNMT1, retained in the nucleolus during acidosis in human cells (Audas et al 2012). With the growing evidence in animal models (Zhu and Brangwynne, 2015), LLPS processes are likely to play an equally important role in the organization of plant chromatin and its partitioning into functional, spatially separated domains. In animal cells, high-to-super resolution techniques have led to a better understanding of the 3D chromatin domain analyses (Shin Y et al, 2018; Wei et al, 2019; (Szabo et al. 2018). Although important progress have been made, plant cells specificity makes the use of these techniques more challenging (Dumur et al 2019).
Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

We apologize to authors whose work could not be cited due to space constraint. The authors thank Moaine El baidouri and Ariadna Picart-Picolo for their input. F.P. was supported by the French Laboratory of Excellence project TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02), the ANR NucleoReg (ANR-15-CE12-0013-01). C.L. was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 757600). C.L. and F.P. are part of the COST ACTION CA16212 INDEPTH.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- - of outstanding interest
References:


from 3D image analysis of developing maize (Zea mays L.) root tip nuclei. *Plant Mol Biol* 2015, **89**:339-351.


70. Sawyer IA, Bartek J, Dundr M: Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. *Seminars in Cell & Developmental Biology* 2019, **90**:94-103.


clustering through carboxy-terminal domain phase separation.


Figure 1. Spatial distribution of chromatin regions.

(a) A sketch illustrating spatial patterns of plant chromatin in the nucleus and their association with gene expression. Note that plants do not encode lamin proteins. The term “plant nuclear lamina” refers to filamentous protein structure that underlies the inner nuclear membrane [91]. Plant nuclear lamina very likely consists of plant-specific Nuclear Matrix Constituent Proteins (NMCP, also known as CRWN in Arabidopsis) [92]. In general, chromatin regions located at the nuclear periphery (i) and at the nucleolus (ii) tend to be inactive. Recent studies have revealed a large number of chromatin contacts linking actively expressed gene with one another (iii), as well as with distal regions having potential roles in transcriptional regulation (iv). The question marks besides these chromatin contacts depict the current situation that little is known about how the interactions are established. These chromatin contacts are established by factor yet unknown (question marks), which we speculate to be combinatorial activities of stochastic chromatin movements, specific bridging interactions of proteins and RNAs, and liquid-liquid phase separation. The contacts among expressed genes foster the formation of sub-compartments and coordinated transcription. Besides, multiple H3K9me-marked loci can form puncta in the nucleoplasm (v), which is likely driven by liquid-liquid phase separation mediated by plant-specific ADCP1 proteins [72]. LAD, lamina-associate domains; NAD, nucleolus-associated domains. (b) Location of LADs and NADs loci across the Arabidopsis genome. This circos plot is generated based on domain coordinates described in [39] and [41].
Figure 2. Association between DNA replication timing and chromatin localization in *Arabidopsis*.

(a and b) Comparisons of *Arabidopsis* DNA replication activities (measured with Repli-seq by Concia et al. [63]) in early and late S phase stages in LADs (a) and NADs (b).

(c and d) Distribution of leading nascent DNA strands in ORIs (DNA replication origin) across LADs (c) and NADs (d). Note that the dataset describing nascent DNA strand is from a study by Sequeira-Mendes et al. [66], in which a size cutoff (0.3 to 2 kb) was used so that the recovered nascent strands were primarily leading strands in ORIs. This information, in turn, can be used to infer whether ORIs occur across a given genomic region evenly. For instance, the curves in (c) imply that around LAD boundary regions, ORIs fire more often outside LADs than inside. Plant materials used for generating these datasets are partly comparable: Repli-seq, 7-day-old seedlings; nascent DNA strands, 4-day-old and 10-day-old seedlings; LADs, 10-day-old seedlings; and NADs, 3-week-old seedlings. Datasets and scripts for reproducing plots in panels (a-d) are available from figshare repository with DOI: 10.6084/m9.figshare.8953235. Before publication, these datasets and scripts are accessible with this private link: https://figshare.com/s/e9ec0a926a1840e2455b
Figure 3. Identification of *A. thaliana* proteins containing an R/G-rich and/or a Q/N-rich IDR.

(a) Among the 27416 referenced proteins in the *A. thaliana* genome (TAIR10), 1234 possess at least a GGRG motif or a stretch of Q/N (at least 12 Q or N in 30 continuous amino acids). The numbers of proteins with more GGRG motifs or stronger Q/N stretch are presented. The key proteins are listed, and their respective subnuclear compartment is specified in brackets. (b) Venn diagram demonstrating the lack of overlap between proteins containing at least two R/G-rich and a strong Q/N-rich IDR. Protein sequences were downloaded from TAIR10 [https://www.arabidopsis.org/](https://www.arabidopsis.org/). The identification of IDR-containing proteins was based on text mining with following criteria: R/G motifs were called if they exactly matched text string “GGRG”; Q/N motifs were called when at least 12 Q/N residues were found in a window of 30 amino acids. Overlapping Q/N motifs were further merged. For each IDR-containing protein, the number of R/G motifs and/or the number of Q/N residues in IDR can be found in Supplemental Table 1.