



Role of Nuclear Non-Canonical Nucleic Acid Structures in Organismal Development and Adaptation to Stress Conditions

Célia Alecki and Maria Vera*

Department of Biochemistry, McGill University, Montreal, QC, Canada

Over the last decades, numerous examples have involved nuclear non-coding RNAs (ncRNAs) in the regulation of gene expression. ncRNAs can interact with the genome by forming non-canonical nucleic acid structures such as R-loops or DNA:RNA triplexes. They bind chromatin and DNA modifiers and transcription factors and favor or prevent their targeting to specific DNA sequences and regulate gene expression of diverse genes. We review the function of these non-canonical nucleic acid structures in regulating gene expression of multicellular organisms during development and in response to different stress conditions and DNA damage using examples described in several organisms, from plants to humans. We also overview recent techniques developed to study where R-loops or DNA:RNA triplexes are formed in the genome and their interaction with proteins.

Keywords: R-loop, DNA:RNA triplex, ncRNA, gene regulation, development, stress

OPEN ACCESS

Edited by:

Chiara Gamberi,
Coastal Carolina University,
United States

Reviewed by:

Caroline Dean,
John Innes Centre, United Kingdom
Takuya Imamura,
Hiroshima University, Japan
Maria Paola Paronetto,
Foro Italico University of Rome, Italy

*Correspondence:

Maria Vera
maria.veraugalde@mcgill.ca

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 26 November 2021

Accepted: 25 January 2022

Published: 23 February 2022

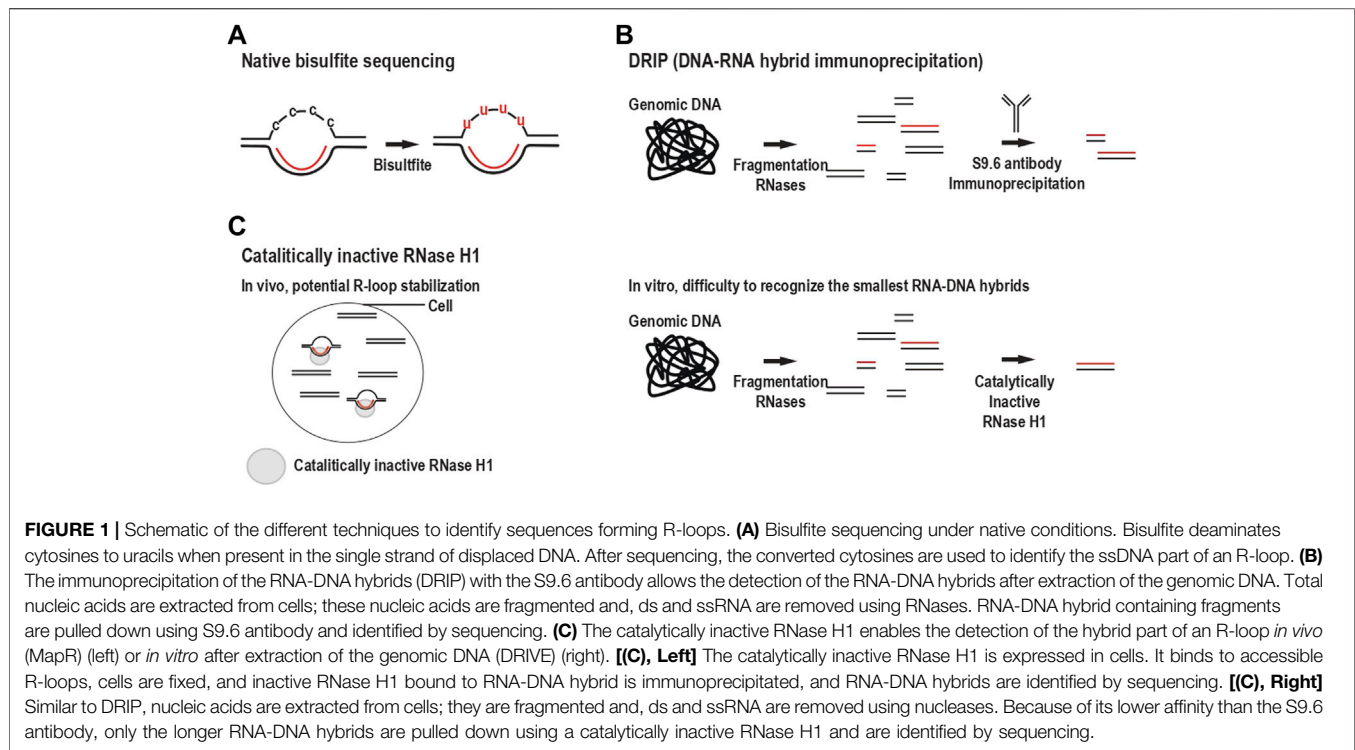
Citation:

Alecki C and Vera M (2022) Role of Nuclear Non-Canonical Nucleic Acid Structures in Organismal Development and Adaptation to Stress Conditions. *Front. Genet.* 13:823241. doi: 10.3389/fgene.2022.823241

1 INTRODUCTION

Non-canonical nucleic acid structures involving RNA can adopt different forms like R-loops and DNA:RNA triplexes and differ structurally from double-stranded DNA (dsDNA) by adopting a non-B-form (Wang et al., 1982; Radhakrishnan and Patel 1994). R-loops are composed of an RNA-DNA hybrid, a displaced single-stranded DNA (ssDNA), and a free RNA part that can be folded or not (Niehrs and Luke 2020). DNA:RNA triplexes are composed of dsDNA and an RNA strand wrapped around and contacting the DNA through Hoogsteen bonds (Li et al., 2016). Morgan and Wells first described non-canonical nucleic acid structures involving RNA in the 1960s (Morgan and Wells 1968). However, their *in vivo* identification and functions as regulatory elements have only been described during the last 2 decades (Ginno et al., 2012; Sentürk Cetin et al., 2019).

R-loops and DNA:RNA triplexes are dynamic and multifunctional structures. They regulate gene expression by inducing the activation or the repression of genes (Niehrs and Luke 2020; Li et al., 2016). The molecular mechanisms underlying their gene expression regulatory activities are diverse and involve different DNA sequences and protein partners, which are the focus of an extensive ongoing investigation. In this review, we discuss studies demonstrating that non-canonical nucleic acid structures act to provide sequence specificity to epigenetic and DNA modifications and transcription factors. They participate in the target of epigenetic modifiers, like the *Polycomb* complex PRC2, which lack sequence-specific binding activity, to specific regions of the genome and can also prevent the targeting of others (Alecki et al., 2020; Skourti-Stathaki et al., 2019; Arab et al., 2019; Grote et al., 2013; Schmitz et al., 2010; Postepska-Igielska et al., 2015; Cristini et al., 2018; Wang et al., 2018). They also favor or prevent the targeting of transcription



factors (Sentürk Cetin et al., 2019; Arab et al., 2019; Cristini et al., 2018; Wang et al., 2018). Consequently, they can either repress or activate gene expression.

Non-canonical nucleic acid structures can act to orchestrate diverse gene expression regulatory mechanisms (Bacolla et al., 2015; Li et al., 2016; Crossley et al., 2019; Niehrs and Luke 2020). Therefore, they are considered important in complex biological processes like development and the response to stress conditions. Development requires a series of specific and synchronous events mediated by ubiquitously expressed transcription factors and epigenetic modifiers (Gilbert, 2000). The binding sites of transcription factors are present on the DNA at any time during development, and many of the epigenetic modifiers lack sequence-specific binding activity. Thus, additional cues must indicate which genes have to be expressed at specific developmental stages or cell types and which genes must be repressed. Non-canonical nucleic acid structures like R-loops and DNA:RNA triplexes can represent one of the missing cues to understand how such a precise regulation of gene expression occurs during development. Likewise, stress conditions impose fast changes in the transcriptional preferences of the cells. While most genes are repressed, specific non-coding and coding RNAs are upregulated to favor cell survival (Vihervaara et al., 2018). Our increased knowledge of the sequence, structure, dynamics, and function of non-canonical nucleic acid structures relies on developing tools that allow their identification *in vivo*. Hence, we review the biochemical and microscopy approaches utilized to study the non-canonical nucleic acid structure as well as the latest advances made to define their function and mechanisms of action in different organisms during development and stress conditions.

2 TECHNIQUES TO STUDY THE EXPRESSION OF NCRNAS AND NON-CANONICAL NUCLEIC ACID STRUCTURES

Over the last 2 decades, several sequencing techniques have been developed to annotate the ncRNA transcriptome, characterize their biogenesis, define their tissue and organism specificity, and elucidate the many fundamental functions they have to regulate gene expression (Sun and Chen 2020). The knowledge acquired by virtue of sequencing techniques is now being complemented with high-resolution microscopy to study the spatiotemporal regulation of ncRNA at the single-cell level (Figure 1) and biochemical approaches to characterize the formation and function of ncRNA engaged in non-canonical structures (Table 1). In this section, we review the principles and advantages that these techniques offer to non-canonical nucleic acid structures.

2.1 Biochemistry

Several techniques to detect non-canonical nucleic acid structures involving coding and non-coding RNA have developed over the past decades. In this section, we focus on R-loops (Table 1) and DNA:RNA triplexes and the main techniques utilized to identify RNA-RNA and RNA-protein interactions.

2.1.1 R-Loops Detection

DNA-RNA hybrid immunoprecipitation (DRIP) is the primary method used to identify genome-wide RNA-DNA hybrids, the main component of an R-loop (Ginno et al., 2012) (Table 1).

TABLE 1 | Techniques allowing the identification of sequences forming R-loops.

Techniques		R-loop region targeted	+	-
Non-denaturing bisulfite-sequencing Yu et al. (2003)	Bisulfite	Displaced ssDNA	High resolution	Sensitive to DNA methylation Other non-canonical DNA structures exposed a ssDNA
DRIP Alecki et al. (2020); Hartono, (2018); Sanz and Chédin. (2019); Sanz et al. (2016); Wahba et al. (2016); Xu et al. (2020)	S9.6 antibody	RNA-DNA hybrid	Most widely used Optimized with multiple organisms Single-stranded version to determine the orientation of the R-loop A near base-pair resolution assay has been developed	Antibody bias against GC-rich and longer sequences Nucleic acid extraction and harsh RNase treatment can affect the detection of the most unstable R-loops
DRIVE Ginno et al. (2012)	Catalytically inactive RNase H1 after nucleic acid extraction	RNA-DNA hybrid		RNase H1 has an affinity for RNA-DNA hybrids 5 times lower compared to S9.6 antibody
R-CHIP (Chen et al. (2017), Chen et al. (2019), Skourti-Stathaki, Proudfoot et al. (2011))	Expression of catalytically inactive RNase H1 in cells	RNA-DNA hybrid	Can detect R-loop unstable R loops which are degraded during nucleic acid extraction	Stabilization of R-loop degraded by RNase H1 RNA-DNA hybrid has to be accessible

DRIP was developed in 2012 and relies on the recognition of RNA-DNA hybrids by the S9.6 antibody. The S9.6 antibody was first thought to recognize RNA-DNA hybrids without sequence specificity. However, a more detailed analysis has shown that it exhibits a bias toward GC-rich sequences and longer RNA-DNA hybrids (Phillips et al., 2013; König et al., 2017). S9.6 antibody also recognizes dsRNA with an affinity five times lower than RNA-DNA hybrids, forcing the removal of single- and double-stranded RNAs (ssRNA and dsRNA) with RNase A and RNase III treatments before nucleic acid fragmentation and immunoprecipitation and precluding the use of the S9.6 antibody for microscopy approaches due to the detection of RNA artifacts (Hartono, 2018; Smolka et al., 2021). DRIP relies on cell lysis and protein removal followed by total nucleic acid extraction with phenol-chloroform and ethanol precipitation. As a quality control to call peaks and identify areas in the genome rich in R-loops by DNA sequencing, a negative control sample is treated with RNase H (RNase H1 and H2) to remove RNAs from RNA-DNA hybrids. This basic DRIP protocol was first developed in Ntera2 mammalian cell lines and optimized to identify R-loop forming sequences in yeast (Wahba et al., 2016; Hartono, 2018), plant tissues (Xu et al., 2020), *Drosophila* embryos, and cell lines (Alecki et al., 2020). Additionally, DRIP has been improved to determine the DNA strand specificity, using either the DNA or RNA moiety of the sample, and to increase the nucleotide resolution in sequencing experiments (Sanz et al., 2016; Sanz and Chédin 2019).

In 2020, a single-molecule assay was developed to obtain a near single-nucleotide resolution of the R-loop forming sequence (Malig and Chedin 2020). Unlike the previously mentioned protocol, this new method relies on non-denaturing bisulfite conversion of cytosine (C) to uracil (U) in the displaced single-stranded DNA followed by sequencing long-read, which identifies the entire R-loop and two mutually exclusive regions forming an R-loop (Malig and Chedin 2020) (Table 1). However, bisulfite converts any C to U in a single-stranded DNA. Therefore, other non-canonical nucleic acid structures that

expose an ssDNA, like G-quadruplex or I-motif, can be detected. Hence, the specificity of bisulfite sequencing to detect R-loops should be validated by DRIP-sequencing. Although C methylation makes ssDNA insensitive to bisulfite (Moye et al., 2015; Miglietta et al., 2020), R-loop and DNA methylation are incompatible in mammals and plants (Ginno et al., 2012; Xu et al., 2017) and R-loops have been involved in DNA demethylation via their recognition by the protein GADD45A, which recruits the demethylase TET1 (Arab et al., 2019).

Recently, R-loops were detected *in vivo* by expressing a catalytically inactive RNase H1 fused to the Micrococcal nuclease (MNase) (Yan and Sarma 2020). The expression of this chimeric protein leads to the cleavage of the DNA at the RNase H1 binding site. R-loops identified by this approach might be a minor fraction because RNase H1-MNase fusion protein would only bind R-loops that are protein-free and should outcompete the endogenous RNase H1. Hence, this technique might be biased to identify stabilized R-loops because they are likely resolved by the endogenous RNase H1. Under physiological conditions, RNase H1 is present in the nucleus, but only a small fraction is bound to the DNA as RNase H1 is essentially targeted to chromatin under stress conditions (Lockhart et al., 2019). RNase H1 has two distinct RNA-DNA hybrid binding sites separated by multiple amino acids that have to bind, meaning that this approach will not be able to detect the smallest R-loops (Nowotny et al., 2008). Additionally, the catalytically inactive RNase H1 has a lower affinity for RNA-DNA hybrid than S9.6 antibody, making this technique less sensitive for detecting R-loops than DRIP (Ginno et al., 2012). In spite of these caveats, combining this technique with bisulfite conversion has made possible to provide strand specificity of the R-loops identified with MapR (Wulfridge and Sarma 2021).

2.1.2 DNA:RNA Triplex Detection

DNA:RNA triplexes are composed of two DNA strands and one RNA in the major groove of the dsDNA formed by a Hoogsteen base pairing between the RNA and the purine-rich DNA strand

(Li et al., 2016). The absence of antibodies or proteins that specifically recognize and bind these structures has made difficult their identification *in vivo*. The first protocol developed to discover DNA:RNA triplexes genome-wide (triplex capture assay) involves the incubation of biotinylated-RNA with isolated nuclei followed by streptavidin pull down to detect the DNA bound to the RNA (Postepska-Igielska et al., 2015). However, triplex capture assay creates artificial DNA:RNA triplexes, prevents the identification of endogenous triplexes, and does not distinguish DNA:RNA triplexes from hybrid G-quadruplexes or RNA-DNA hybrids. In 2019, Sentürk Cetin *et al.* developed other protocol to identify DNA:RNA triplexes *in vivo* in mammalian cells for the first time (Sentürk Cetin et al., 2019). Total nucleic acids are extracted using gentle lysis in the presence of proteinase K and RNA-DNA hybrids and ssRNA are removed by a short treatment with RNase H and RNase A, respectively. DNA- and RNA-associated DNA was immunoprecipitated with an antibody that recognizes DNA, then RNAs were eluted from the beads by digesting the DNA with DNase, and the RNA part of the triplexes were sequenced. In this protocol, samples are treated with a commercial RNase H for 30 min, which is insufficient to remove all RNA-DNA hybrids; thus, a fraction of the RNA-DNA hybrids, the main part of R-loops, might still be present and identified as DNA:RNA triplexes (Alecki and Francis 2021). As a result, they observed that 20% of the peaks of their putative DNA:RNA triplexes overlapped with previously published R-loop forming sequences, suggesting that these RNAs tend to form different non-canonical nucleic acid structures or that their protocol is not fully specific for the identification of DNA:RNA triplexes *in vivo*. Even so, this protocol allowed the identification of DNA:RNA triplexes *in vivo* for the first time, validating that the triplex capture assay using biotinylated RNA can be a tool to identify triplex-forming sequences.

2.1.3 RNA-RNA Interactions

RNA molecules interact with other RNA molecules directly or through proteins. RNA-RNA interactions mediated by proteins are detected using cross-linking reagents, such as UV or formaldehyde. The RNAs are fragmented, and their extremities in proximity are ligated, and the RNA hybrids or chimeras are sequenced to identify RNA-RNA interactions (Gong et al., 2018; Lin and Miles 2019). The direct interaction between two RNA molecules is detected using the compound psoralen. Psoralens were used initially to identify RNA structure because they intercalate into dsRNAs. The subsequent cross-linking of dsRNAs with UV, removal of ssRNA by RNases, and ligation of dsRNA ends afford to sequence the chimeric RNAs (Lu and Chang 2018).

2.1.4 RNA-Protein Interactions

RNA-protein interactions are commonly studied by RNA immunoprecipitation (RIP) and cross-linking immunoprecipitation (CLIP). CLIP relies on UV cross-linking before pulling down the protein of interest, followed by RNA sequencing in order to identify bound RNAs (Ramanathan et al., 2019). CLIP is also used to identify the specific nucleotide sequence

bound by a protein. For this purpose, the unprotected RNA is digested with RNases, and protein-protected RNA analyze by sequencing (Hafner et al., 2021). However, the presence of post-translational modification can decrease the efficiency of UV cross-linking (Vieira-Vieira et al., 2021). Complementary techniques to CLIP have been developed to pull down the RNA and identify its interacting proteome by mass spectrometry. Most of these methods require the insertion of exogenous sequences, like the MS2 stem-loops, into the RNA sequence to facilitate its pull down with the cognate binding protein (Slobodin and Gerst 2010; Hafner et al., 2021). Alternatively, the exogenous expression of a modified CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-Cas) system where the inactive Cas9 fused to a BirA protein is directed to a specific mRNA and interacting proteins get labeled and recognized by BioID (Yi et al., 2020). A recently developed method identifies proteins at the proximity of a particular endogenous RNA without exogenous expression of factors or modification of RNAs (Yap et al., 2021). The RNA of interest is hybridized with digoxigenin-labeled antisense probes, and the digoxigenin is then bound by the so-called hybridization proximity (HyPro) enzyme. The RNAs and proteins in the proximity of the target mRNA are biotinylated by HyPro enzymes, pulled-down using streptavidin beads, and identified by RNA-seq and mass spectrometry, respectively. Even if this technique is limited to the RNA and protein interactome of a specific RNA, it may provide information on low abundant RNAs because it doesn't require the immunoprecipitation step. A limitation of this method is the difficulty to discern between proximal or interacting proteins and RNAs. All techniques aiming to identify RNA-protein interactions lead to the detection of direct and indirect associations. In addition, these techniques are more qualitative than quantitative, and quantification only refers to a relative comparison between experimental conditions (Vieira-Vieira et al., 2021). Their main caveat is the limitation to a single RNA or protein and miss the complete picture of all cellular RNA-protein interactions.

2.2 Microscopy

Visualizing and analyzing the spatial distribution and dynamics of expression of ncRNA inside cells provide the means to relate their nuclear localization and expression regulation to their function. The spatiotemporal resolution obtained with microscopy approaches helps to study ncRNA expression and function in response to internal cues, like the cell-cycle regulation and cell-differentiation, and changes in their environment.

2.2.1 Localization and Quantification of ncRNAs

Single-molecule fluorescence *in situ* hybridization (smFISH) is a robust technique widely used to quantify and localize single RNA molecules inside cells. smFISH was designed initially to detect mRNAs through the multimerization of specific antisense DNA oligonucleotides that are fluorescently labeled (Femino 1998). The localization of several fluorophores on an RNA molecule provides the sufficient signal-to-noise ratio to specifically detect the fluorescence signal from an mRNA, whose localization is later analyzed by specific computational frameworks, like Fishquant (Mueller et al., 2013). Given that an RNA molecule must

accommodate several probes to be detected, smFISH has worked to detect, quantify the abundance, and analyze the distribution of long ncRNAs (Cabili et al., 2015). The distribution of long ncRNAs in the cell is diverse. Still, it shows a predominantly nuclear localization that appears as discrete nuclear aggregates in 20% of the cases, supporting their role in gene expression and chromatin structure. A simultaneous two-color smFISH has distinguished the long ncRNA accumulation at different nuclear locations and transcription sites. In this protocol, the second set of probes, labeled with a different spectral fluorophore, detect the transcription of a proximal divergent gene to the ncRNA or a gene known to be regulated by the ncRNA (Cabili et al., 2015).

Given the multimerization nature of smFISH, the original protocol had to be adapted for the specific detection of small ncRNAs, whose sequence is not long enough to accommodate several antisense probes. The single-molecule resolution for small ncRNAs has been accomplished by increasing the signal provided by fewer antisense oligonucleotides and branching DNA probes. Amplification of the signal provided by a single antisense oligonucleotide with a hapten by tyramide signal amplification (TSA), TSA-FISH. TSA signal amplification utilizes the enzymatic reaction of the horseradish peroxidase and fluorescently labels tyramide substrates that are deposited at the reaction site (Kwon 2013). In branched FISH (bdDNA FISH), two antisense oligonucleotides hybridized the RNA and provided a linker sequence. The linker sequence is hybridized by an amplifier that provides a sequence to multimerized oligonucleotides that are recognized by fluorescently labeled probes (Battich et al., 2013; Sinnamon and Kevin 2014).

2.2.2 Dynamics of ncRNA Expression

Visualizing RNAs in live cells to investigate their expression dynamics relies upon the same fluorescent multimerization principle that smFISH. The MS2 and PP7 are the two most successfully used systems for ncRNAs, like telomeric RNA TERRA (Cusanelli et al., 2013), the X-inactive specific transcript Xist, and the antisense *GAL10* ncRNA (Lenstra et al., 2015). The MS2 and PP7 systems are RNA aptamers derived from bacteriophages that are inserted as an array of 12 to 24 stem-loops in the gene encoding for the ncRNA. Each of the stem-loops is tightly bound by the cognate protein, MCP and PCP, respectively, fused to a fluorescent protein, like the green fluorescent protein (GFP) (Vera et al., 2016). The MS2 and PP7 have been combined to tag in the same cell an ncRNA and its regulated mRNAs to link ncRNA expression dynamics to their regulatory function on mRNA transcription (Lenstra et al., 2015). The caveats of the MS2 and PP7 systems to follow the expression of small ncRNAs is that they preclude their function by adding a long RNA sequence and a bulky protein load.

3 NON-CODING RNAs IN GENE EXPRESSION REGULATION

The expression of ncRNAs is more tissue-specific than the mRNAs; while 50% of ncRNAs are expressed in a maximum

of two different tissues in humans, 65% of mRNAs are detected in the tested sixteen tissues (Derrien et al., 2012). Another characteristic of ncRNAs is their low abundance, as they are expressed at much lower levels than mRNAs. ncRNAs regulate gene expression by mainly four different mechanisms, 1) sequestering microRNAs (miRNAs), 2) targeting proteins to the chromatin, 3) *cis*-transcriptional regulation of a coding gene, and 4) forming non-canonical nucleic acid structures, RNA-loops, and DNA:RNA triplexes, to promote or prevent targeting transcription factors or chromatin and DNA modifiers.

(i) miRNAs have been widely studied over the past 25 years. They are small RNAs of 21 to 25 nucleotides that regulate gene expression at different steps from transcription to translation. They bind mRNAs in their 5' or 3' UTR to prevent translation and can also trigger mRNA degradation by forming an RNA duplex and recruiting the RISC complex (for review O'Brien et al., 2018). miRNAs can also interact with DNA promoting a favorable chromatin context and gene upregulation. The role of ncRNA in the regulation of miRNA availability has been extensively reviewed. It relies on their capacity to bind several miRNAs like a sponge and regulate their activity by competing to bind their target mRNA (for review Fernandes et al., 2019). (ii) Long ncRNAs target chromatin modifiers to the promoter of specific genes, which affect their expression (for review Statello et al., 2021). (iii) There is a set of genic long ncRNAs known as promoted associated ncRNAs (pancRNAs) that regulate in *cis* the transcription of coding genes. pancRNAs are the product of transcription of bidirectional promoters enriched in CpG islands, which drive the expression of two divergent transcripts, an mRNA and a stable ncRNA, the pancRNA (Chellini et al., 2020; An et al., 2021). The expression of pancRNA correlates with the expression of their divergent protein-coding gene, and they have been proposed to activate their transcription by promoting local epigenetic changes (Uesaka et al., 2014; Hamazaki et al., 2015). Before being referred to as pancRNAs, promoter-derived ncRNAs were known as promoter upstream transcripts (PROMPTs), promoter-associated long RNAs (PALRs), promoter-associated small RNAs (PASRs), upstream antisense RNAs (UaRNAs), stable unannotated transcripts (SUTs), and cryptic unstable transcripts (CUTs). They are mostly associated with the activation of transcription of the divergent mRNA (for review An et al., 2021). However, ncRNA originated from the intergenic RNA polymerase I (Pol I) promoter (pRNAs) overlap with the promoter of ribosomal RNAs (rRNAs), forming a DNA:RNA triplex that mediates the silencing of ribosomal genes (Santoro et al., 2010; Wehner et al., 2014; Vydzhak et al., 2020).

In this review, we refer to genome wide studies that have identified the position of R-loops and DNA:RNA triplexes and how they evolved in response to developmental or stress cues. We also provide mechanistic examples that highlight the role of ncRNAs forming DNA:RNA triplex and R-loops in regulating gene expression by targeting proteins to the chromatin during development and stress conditions. Four out of the dozen of mechanistic examples we overview concern Polycomb group proteins targeting by R-loop during development, or triplexes during development or stress. PRC2 complex is an essential

epigenetic regulator conserved from plants to mammals. It lacks sequence-specific DNA binding activity but its targeting to the genome is not random. PRC2 complex binds to thousand of coding and non-coding RNAs (Kaneko et al., 2013; Khalil et al., 2009; Zhao et al., 2010), some forming non-canonical nucleic acid structures (Mondal et al., 2015, 3; Grote and Herrmann 2013; Skourti-Stathaki et al., 2019), and PRC2-target sites form R-loop in *Drosophila melanogaster* (Alecki et al., 2020).

4 NON-CANONICAL RNA STRUCTURES IN ORGANISM DEVELOPMENT

4.1 R-Loops and Development

4.1.1 Where They Are Formed and Their Implications for the Organism

Genome-wide studies have identified R-loops during the development of several organisms. R-loops have been observed in *Drosophila melanogaster* embryos at two different stages of development: at two to 6 h (h) that corresponds to the steps of cellularization and gastrulation, and at 10–14 h during the embryo segmentation (Alecki et al., 2020). The study was restricted to detect R-loops in DNA elements called *Polycomb* response elements (PREs). PREs can be transcribed into ncRNAs and form R-loops. A subset of R-loops formed in PREs in 2–6 h old embryos was absent in 10–14 h embryos. PREs covered by R-loops had a higher level of the repressive transcription mark H3K27me3 than PREs lacking R-loops, and PREs which have formed an R-loop only in 2–6 h old embryos are more likely to be repressed by *Polycomb* group proteins in 10–14 h old embryos. *Polycomb* group trimethylated the lysine 27 of histone 3 (H3K27me3) and bound to it to compact the chromatin and inhibit its remodeling. Hence, this result suggested that forming an R-loop at PRE sequences during an early developmental stage may lead to *Polycomb*-dependent gene repression at a later stage. Their data shows that only a small number of R-loops are exclusively formed in 2–6 h embryos, but a higher number of them formed in early embryos and showed a decrease in the R-loop level in 10–14 h embryos, suggesting that upon development and cell differentiation some cells form R-loop and some other don't.

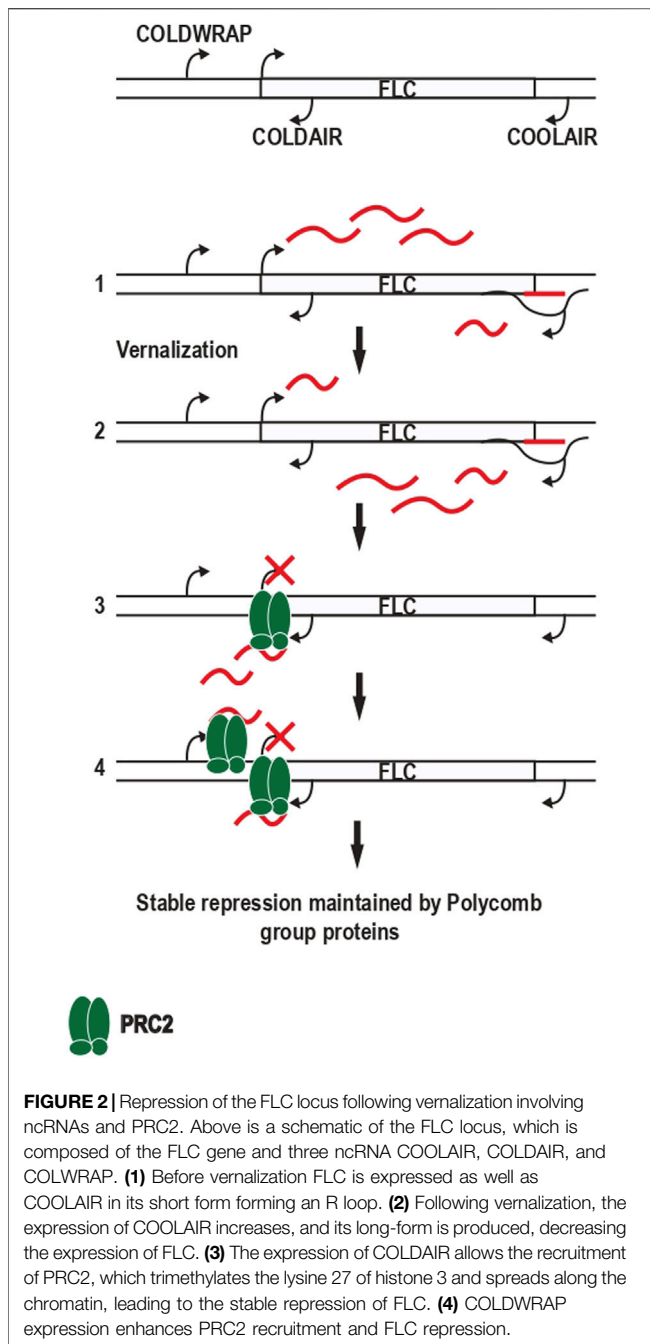
R-loops forming at PREs are not the only developmentally regulated R-loops in *Drosophila*. *Drosophila* embryos also form R-loops in retrotransposon sequences known as Long interspersed nuclear elements (LINE). R-loops are enriched in LINEs and retrotransposons in embryos but S2 cells (derived from 20 to 24 h old embryos) have them enriched at low complexity sequences and simple repeats (Zeng et al., 2021). The formation of R-loops over LINE sequences was previously reported in the human LINE-1 sequence, and it is related to the Aicardi-Goutières syndrome, an inherited encephalopathy that affects newborn infants. The Aicardi-Goutières syndrome is associated with mutations in the exonucleases TREX1 or SAMHD1, the RNA deaminase ADAR1R, and the RNase H2 (Benitez-Guijarro et al., 2018). All these mutations lead to an increase of R-loop levels and DNA hypomethylation, a decrease of LINE-1 retrotransposition, which result in an induction of the innate immune response and neurological dysfunction (Lim et al., 2015).

RNase H2 activity is also required to degrade the RNA part of the RNA-DNA hybrid intermediate and complete the retrotransposition, suggesting that R-loop formation and resolution are two essential steps of transposon insertion (Benitez-Guijarro et al., 2018). R-loops misregulation has also been linked to myelodysplastic syndrome, a blood cell cancer in which blood cells don't mature (Sperling et al., 2017). The myelodysplastic syndrome can be caused by a decrease of the DEAD-box protein DDX41, a putative RNA helicase that was shown to interact with RNA-DNA hybrids (Wang et al., 2018). The reduction of DDX41 leads to an increase of R-loops in hematopoietic stem and progenitor cells and increases inflammatory signaling, leading to a rise in the number of hematopoietic stem and progenitor cells (Weinreb et al., 2021).

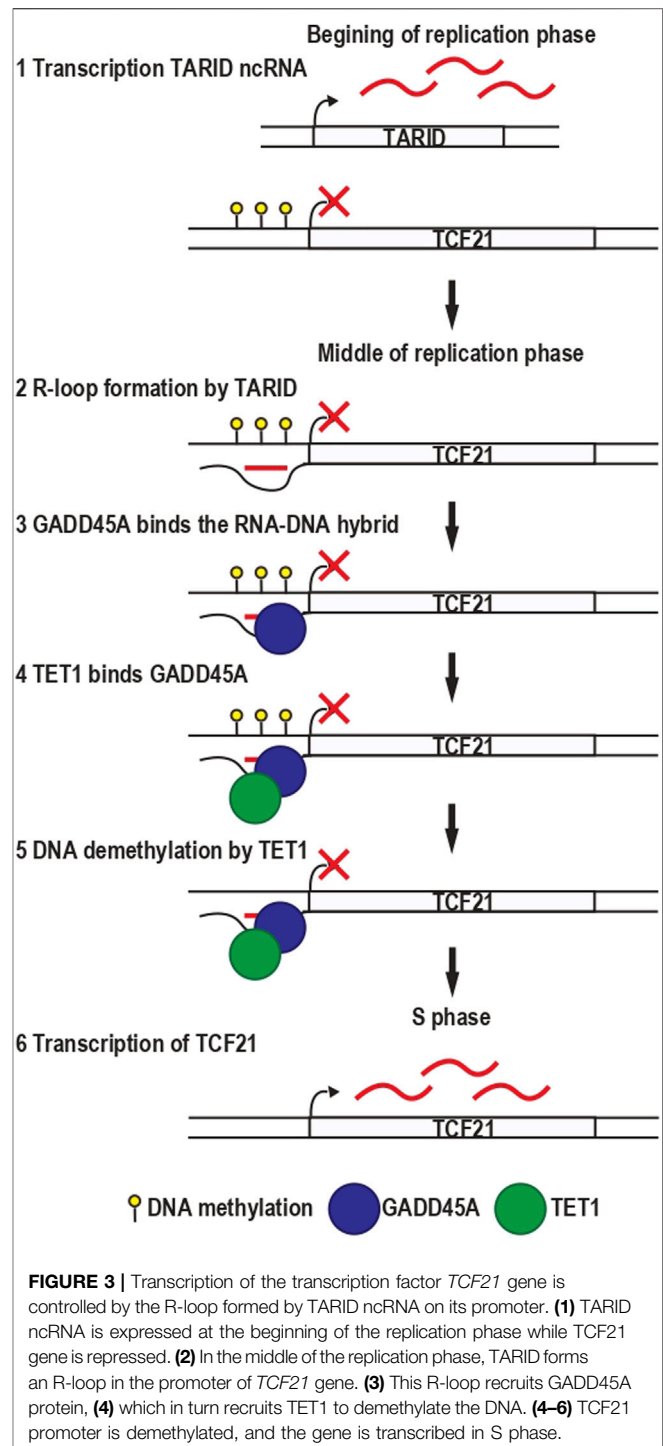
These studies, together with the fact that R-loops are detected at repetitive Satellite sequences in mammalian cells (Zeng et al., 2021), suggest that R-loop dynamics during mammalian development might support gene expression changes during this process. Mouse embryonic stem cells (mESC) have been used to identify R-loop interactors using immunoprecipitation with the S9.6 antibody followed by mass spectrometry (Wu et al., 2021). RBP1, a subunit of RNA polymerase II, and DHX9, RNA helicase, were among the already known R-loop interactors and regulators identified in this study. Additionally, nucleolar proteins and proteins involved in rRNA transcription and processing were particularly enriched. This result was expected because the nucleolus is one of the hotspots for R-loop formation (Sollier et al., 2014). Some proteins of the DEAD-box family, which are known to be involved in R-loop regulation and bind RNAs were also enriched (Cargill et al., 2021). Interestingly, the knockdown of several members of the DEAD-box family leads to the overexpression of genes involved in cell fate commitment and neuronal differentiation (Wu et al., 2021), suggesting that the R-loops regulated by these proteins may be important for the pluripotency of mESC.

The idea that R-loop regulation is essential for the pluripotency of stem cells is further supported by experiments done to reprogram Mouse Embryonic Fibroblast (MEFs) into pluripotent stem cells (PSC) (Li et al., 2020). During the dedifferentiation of MEFs using the Yamanaka cocktail, R-loop forming sequence changes preceded gene expression changes (Li et al., 2020). Interestingly, the stabilization of R-loops by knocking down RNase H1 using shRNA or expressing a catalytically inactive RNase H1 decreased or prevented MEF reprogramming, respectively. Genes involved in somatic cell state had increased R-loop levels upon the expression of the catalytically inactive RNase H1 in MEFs, which required a higher amount of the Yamanaka factor Sox2 to restore pluripotency. Surprisingly, Sox2 did not reduce R-loops' total level but induced R-loops formation at other genes, including pluripotency genes. Sox2 interacts with Ddx5, a member of the DEAD-box family, and Dhx9, an RNA helicase, and inhibits their R-loop resolution activity. Overall, (Yaoyi et al., 2020) have highlighted the relevance of the tight regulation of R-loops to maintain the cell somatic state or reprogram pluripotency.

R-loop formation also coordinates different aspects of plant development. One study in *Arabidopsis thaliana* and another in rice have looked at R-loop formation genome-wide in various tissues and under different growth conditions (Fang et al., 2019;



Xu et al., 2020). Differences in R-loop formation were found between seeds and flowers in *Arabidopsis thaliana* and seed and leaf callus in rice, where 30% of the R-loop forming sequences were different. Although the expression of ncRNA was not investigated, differences in R-loop formation between seeds and developed plants suggested that R-loops are developmentally regulated and, therefore, differentially expressed ncRNAs are probably involved in the formation of these non-canonical nucleic acid structures. Hence, efforts done over the last decade have identified R-loop forming sequences genome-wide in multiple organisms.



4.1.2 R-Loops and the Regulation of Gene Expression: Mechanistic Examples

Recent studies have started to highlight the role of R-loops in regulating the expression of ncRNAs and transcription factors with a role in development. There are two examples, one in plants and another in mammalian cell lines, that have been well documented. Among R-loops whose expression changed between the vegetative and reproductive state in plants, one of

the best studied examples is in *Flowering locus C* (FLC) gene. In order to flower only after winter, plants like *Arabidopsis thaliana* need to repress the *FLC* gene (Costa and Dean 2019). Long cold exposure during vernalization leads to the progressive epigenetic silencing of the *FLC* gene. Three ncRNAs are transcribed from the *FLC* locus: COOLAIR, COLDAIR, and COLDWRAP. COOLAIR is a set of ncRNA antisense transcripts that originate close to the termination region of *FLC*. COOLAIR forms an extensive R-loop covering the *FLC* 3' region (Sun et al., 2013), which in warm conditions triggers chromatin modifications that decrease transcriptional output from the whole locus. The R-loop is resolved by 3' processing factors terminating COOLAIR transcription, in a process involving m⁶A methylation (Xu et al., 2021) (Figure 2). The R-loop is also important after 1–2 weeks of cold exposure where COOLAIR expression increases and clouds of nascent transcript form around the locus. This leads to a decrease of *FLC* expression because of their mutually exclusive production (Rosa et al., 2016). After 20 days of vernalization, COLDAIR, sense ncRNA transcribed from the first intron of *FLC*, is expressed. It is reported to interact with Polycomb complex PRC2 at the *FLC* locus (Kim et al., 2017). There is an initial nucleation site of PRC2 which tri-methylates the lysine 27 of histone 3, spread along the chromatin leading to stable repression of *FLC* (Yang et al., 2017). Finally, COLDWRAP, sense ncRNA transcribed in the promoter region of *FLC*, is expressed after COLDAIR and can serve as a backup mechanism in the absence of COLDAIR, allowing the Polycomb-dependent repression of *FLC* following transfer back to warm (Kim and Sung 2017). The promoter of *FLC* and its first intron form a gene loop identified by 3C depending on ncRNA COLDAIR and COLDWRAP (Kim and Sung 2017). The loop formation is thought to reinforce recruitment of PRC2 and its epigenetic mark H3K27me₃. A second loop between the *FLC* promoter and termination site associates with *FLC* gene activity and this is lost as plants are transferred to cold (Crevillén et al., 2013).

The transcription factor TCF21 plays a critical role in developing several cell types during embryogenesis (Lotfi et al., 2021). The expression of the *TCF21* gene in human embryonic cells is controlled by an R-loop formed by the lncRNA TARID in the promoter of *TCF21* gene (Figure 3). The RNA-DNA hybrid part of this R-loop is recognized by the growth arrest and DNA-damage-inducible protein GADD45 alpha (GADD45A) protein, which interacts with the enzyme Ten-eleven translocation methylcytosine dioxygenase 1 (TET1). TET1 mediates the DNA demethylation of *TCF21* promoter, leading to its expression (Arab et al., 2019). Interestingly, all the previously mentioned events occur following DNA replication in a specific order during the cell cycle. DNA replication leads to the eviction of nucleosomes and their distribution between the two new DNA molecules. Likewise, the level of DNA methylation is divided by two and must be restored to maintain gene expression. Transcription of TARID starts at the beginning of the replication phase, and it is followed by R-loop formation, which recruits GADD45A and DNA demethylation to the *TCF21* promoter in the middle of the replication phase and leads to the transcription of *TCF21* during the S phase. This

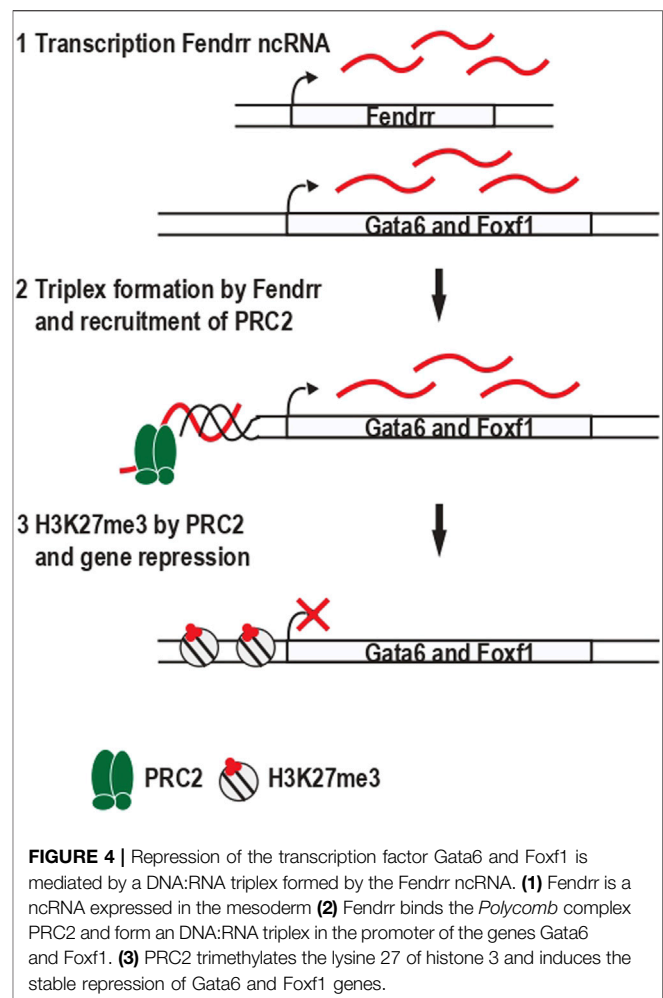


FIGURE 4 | Repression of the transcription factor *Gata6* and *Foxf1* is mediated by a DNA:RNA triplex formed by the *Fendrr* ncRNA. **(1)** *Fendrr* is a ncRNA expressed in the mesoderm **(2)** *Fendrr* binds the Polycomb complex PRC2 and form an DNA:RNA triplex in the promoter of the genes *Gata6* and *Foxf1*. **(3)** PRC2 trimethylates the lysine 27 of histone 3 and induces the stable repression of *Gata6* and *Foxf1* genes.

R-loop dependent targeting of TET1 is not restricted to the *TCF21* gene because the overexpression of RNase H1, enzyme degrading the RNA moiety of an RNA-DNA hybrid, leads to a decreased binding of GADD45A and TET1 at other binding sites controlling the expression of transcriptional coactivators and repressors, and chromatin binding proteins. On the contrary, knockdown of RNase H1 induces the stabilization of the R-loop and an increase in the general DNA methylation level (Arab et al., 2019).

4.2 DNA:RNA Triplexes and Development

4.2.1 Where They Are Formed and Their Implications for the Organism

Both RNA and DNA involved in DNA:RNA triplexes have been identified in human cells independently (Sentürk Cetin et al., 2019). As expected, more than 50% of the RNAs involved in DNA:RNA triplex formation come from genes, and 8% of these structures are composed of ncRNAs. Likewise, 50% of the DNA sequences involved in DNA:RNA triplexes correspond to coding genes and 8% to ncRNAs. DNA and RNA involved in triplex are present in transcription start sites and enriched with active epigenetic marks H3K4 methyl. Surprisingly, the RNA part

but not the DNA of DNA:RNA triplexes are enriched at super-enhancer, which are transcribed bidirectionally into ncRNAs (Xiao et al., 2021), suggesting that this RNA transcribed from enhancer can form triplexes *in trans*. Hence, it was proposed that these ncRNAs produced at enhancers might contact distal regions via the formation of DNA:RNA triplexes and induce transcription (Sentürk Cetin et al., 2019). However, the examples for which mechanistic details are available involved gene repression by DNA:RNA triplex.

4.2.2 DNA:RNA Triplex and the Regulation of Gene Expression

Two examples highlight the relevance of ncRNAs forming DNA:RNA triplexes during development. Fendrr is an essential ncRNA expressed in the mesoderm, whose depletion leads to embryonic death (Grote et al., 2013). The depletion of Fendrr leads to a decrease of the repressive mark H3K27me3 and an increase of the active mark H3K4me3 in the promoter region of transcription factor like Foxf1 (Grote et al., 2013). Fendrr interacts with the Polycomb complex PRC2, which trimethylates the lysine 27 of histone 3 promoting stable gene repression (Khalil et al., 2009). Using Chromatin Isolation by RNA purification (ChIRP)-seq analysis, it was shown the interaction of Fendrr with the promoter region of Foxf1 and Gata6, in a manner non-sensitive to RNase H, suggesting that Fendrr form an DNA:RNA triplex (Grote and Herrmann 2013) (Figure 4). The authors proposed a model in which Fendrr ncRNA bound by PRC2 includes an DNA:RNA triplex in the promoter region of its target genes, allowing the sequence-specific recruitment of PRC2, the deposition of H3K27me3, and the stable gene repression necessary for the proper embryonic development.

Promoter-associated RNAs, pRNAs, of ribosomal genes (pRNAs) are ncRNAs transcribed from the PolII promoter located in the intergenic region of ribosomal DNA (rDNA) (Vydzhak et al., 2020; Wehner et al., 2014; Santoro et al., 2010). These ncRNAs are degraded to ~250 nucleotides long fragments that are bound by the chromatin remodeling complex NoRC (Vydzhak et al., 2020). pRNAs can bind the promoter sequence of rDNA by sequence complementation. The structure formed by pRNA and its target DNA is not sensitive to RNase H, nor RNase VI indicating that it is not composed of RNA-DNA hybrid nor dsRNA, respectively, and is not mediated by protein. Thus it is likely an DNA:RNA triplex as the EMSA (electrophoretic mobility shift assay) with dsDNA and pRNAs suggested (Schmitz et al., 2010). The formation of the DNA:RNA triplex between pRNAs and the promoter of rDNA leads to the targeting of the methyltransferase DNMT3b, which methylates the DNA (Schmitz et al., 2010) (Figure 5). NoRC binds pRNAs by recognizing a stem-loop structure, and this binding is essential for NoRC targeting to the nucleolus (Mayer et al., 2008). NoRC complex is then implicated in the deacetylation of histone H4 and the trimethylation of H3K20 an epigenetic repressive mark, allowing the stable repression of rDNA genes (Guett et al., 2010).

More examples for DNA:RNA triplexes in development are expected, given recent findings in cell lines in culture. Taking advantage of the technique developed by Sentürk Cetin that identifies both the DNA and RNA part of DNA:RNA

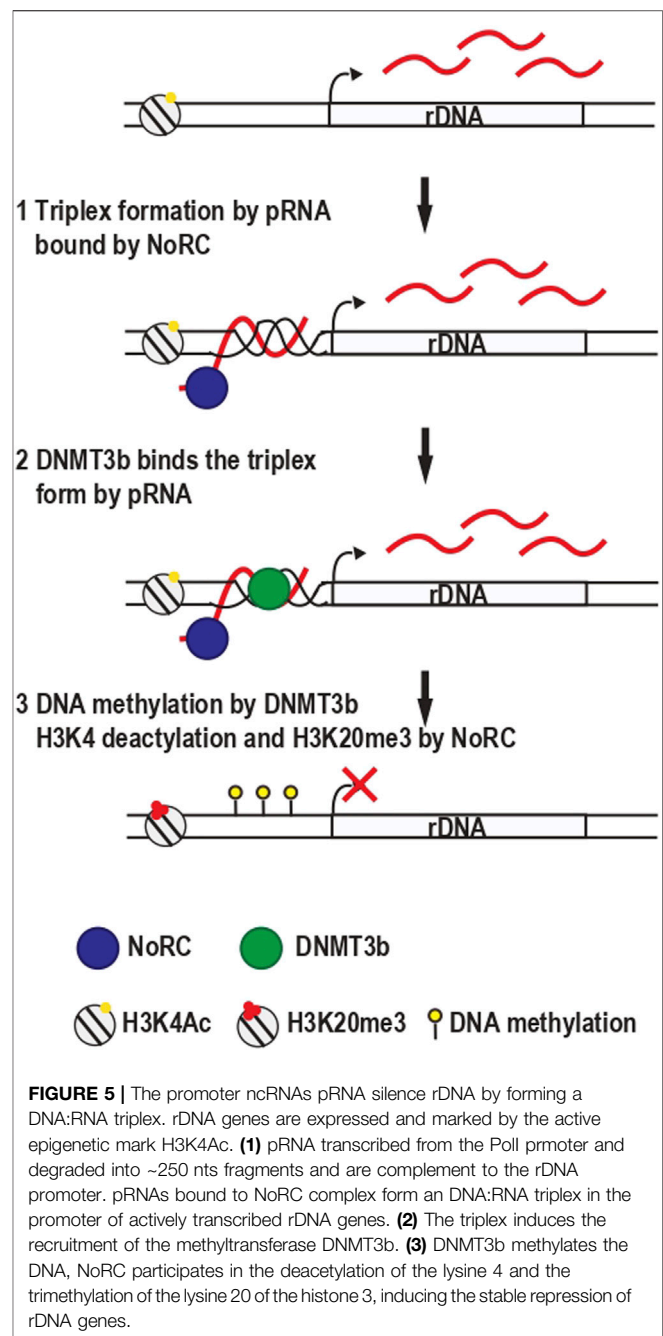


FIGURE 5 | The promoter ncRNAs pRNA silence rDNA by forming a DNA:RNA triplex. rDNA genes are expressed and marked by the active epigenetic mark H3K4Ac. (1) pRNA transcribed from the PolII promoter and degraded into ~250 nts fragments and are complementary to the rDNA promoter. pRNAs bound to NoRC complex form an DNA:RNA triplex in the promoter of actively transcribed rDNA genes. (2) The triplex induces the recruitment of the methyltransferase DNMT3b. (3) DNMT3b methylates the DNA, NoRC participates in the deacetylation of the lysine 4 and the trimethylation of the lysine 20 of the histone 3, inducing the stable repression of rDNA genes.

triplexes, it has been found that the ncRNA NEAT1 has two regions able to form DNA:RNA triplex *in vivo*. One of the triplex-forming regions was identified in HeLa S3 cells and the other in U2OS, suggesting that multiple domains from a single RNA can form triplex and their formation can be developmentally regulated or, at least, cell type-specific (Sentürk Cetin et al., 2019). ncRNA NEAT1 was shown to bind DNA and regulate gene expression by binding transcription factors and preventing them from being targeted to their binding sites. It can also bind and recruit PRC2 to repress specific genes. Finally, NEAT1 is also involved in RNA splicing and RNA stability (Wang et al., 2020,

1). The pull-down of NEAT1 RNA after incubation with protein-free DNA reveals that this ncRNA can bind multiple DNA sequences suggesting that NEAT1 can act *in trans* to target proteins by forming DNA:RNA triplexes (Sentürk Cetin et al., 2019). The research done so far to investigate the mechanism by which DNA:RNA triplexes operate during development shows that they mediate the modification of the DNA and chromatin that lead to the repression of protein-encoding and ribosomal genes.

4.3 ncRNA-mRNA Interaction

Interaction between one ncRNA and an mRNA has also been demonstrated to target proteins to specific loci. Hox genes are a family of conserved and essential transcription factors that are master regulators of the development of the anteroposterior axis (Hajirnis and Mishra 2021). Discovered in *Drosophila melanogaster*, they are organized in clusters, and their mutation can lead to homeotic transformation: the replacement of one segment by another. These genes have been duplicated in mammals and are present in four clusters on four different chromosomes. Their expression is controlled in a spatiotemporal manner and is essential for proper development (Mallo et al., 2010). In order to express Hox genes only in the appropriate compartments, these genes are repressed by epigenetic modifiers the *Polycomb* group proteins. The *Polycomb* complex PRC2 is a methyltransferase that trimethylates the lysine 27 of histone 3 (H3K27me3), and this epigenetic mark participates in the targeting of another *Polycomb* complex PRC1 which can compact the chromatin and stably repress gene expression (Steffen and Ringrose 2014). HOTAIR, ncRNA transcribed between the Hox genes HoxC11 and HoxC12 in mammals, is essential for the targeting of PRC2 *in trans* to the locus HoxD (Rinn et al., 2007; Gupta et al., 2010). In the absence of HOTAIR, PRC2 is not recruited to HoxD, leading to a derepression of this cluster and homeotic transformations (Gupta et al., 2010; Li et al., 2013; Tsai et al., 2010). Recently, Balas et al. showed that HOTAIR ncRNA hybridizes with Junctional Adhesion Molecule 2 (JAM2) mRNA. The formation of this dsRNA structure leads to a conformational change of HOTAIR RNA, decreasing the affinity of PRC2 for this RNA (Balas et al., 2020). They proposed that PRC2 binds HOTAIR and targets it to the HoxD cluster; the change of conformation upon hybridization with JAM2 leads to the release of PRC2 at its target site (Balas et al., 2020). Future investigations will determine whether JAM2, transcribed from a different chromosome than HoxD, forms a non-canonical nucleic acid structure at the locus HoxD, which will also explain the specific targeting of PRC2.

The results obtained over the past decade allow us to better-understanding how ncRNAs act during development to regulate gene expression directly or indirectly via the recruitment of proteins.

5 NON-CANONICAL NUCLEIC ACID STRUCTURES AND THEIR ROLE IN STRESSED CELLS

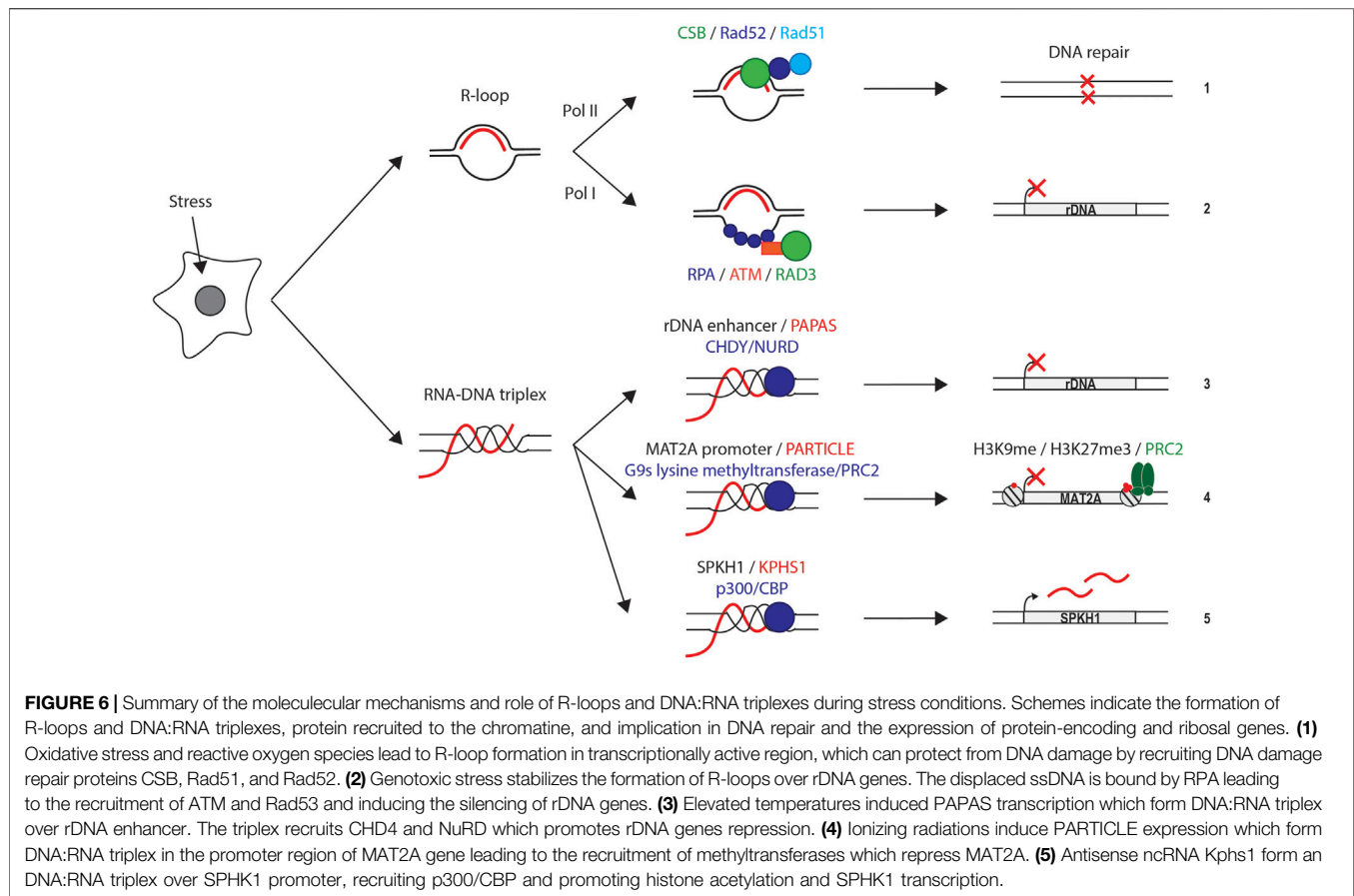
Like development, stress conditions represent a challenge to the cell because it has to adjust several gene expression pathways

simultaneously. A common hallmark of stress conditions is the general shutdown of transcription of ubiquitously expressed protein-encoding and ribosomal biogenesis-related genes (Vihervaara et al., 2018). This response enables cells to decrease the load of damaged molecules downstream of the gene expression pathway and provides time to repair DNA damage. Many of the protein players and molecular mechanisms orchestrating transcription and DNA repair have been well studied. It is also well established that several stresses induce the transcription of diverse ncRNAs in different organisms (McKechnie et al., 1998; Jolly et al., 2004; Audas et al., 2012; Place and Noonan 2014; Goenka et al., 2016; Koskas et al., 2017; Schreiner et al., 2019). However, their mechanism of action and relevance for cell survival to stress have not been yet established. In the last few years, increasing examples suggest that the mode of actions of these ncRNAs is through the formation of non-canonical nucleic acid structures. Hence, stress-related ncRNAs might orchestrate the recruitment of factors to specific transcriptional active regions of the genome to modulate transcription and favor DNA repair (Fig. 6).

5.1 R-Loops

There are two recent examples supporting the role of stress-induced R-loop formation in protecting DNA from damage in mammalian cells. Oxidative stress increases the concentration of Reactive Oxygen Species (ROS) that damages the DNA. Hence, ROS can cause mutations that are deleterious for the cell when affecting protein-encoding and ribosomal genes. It has been recently shown that ROS induces the formation of R-loops in transcriptionally active regions of the mammalian genome. Their formation protects DNA from ROS-induced damage by a mechanism that involves transcription-coupled homologous recombination (TC-HR). R-loops recruit other proteins than the canonical BRC1 and BRC2 to the site of DNA damage in order to mediate TC-HR. The RNA-DNA hybrid part of the R-loop is recognized by the Cockayne Syndrome Protein B (CSB) that targets Rad52 to the damaged-transcribed region. Rad52 recruits the recombinase Rad51 that assists in repairing DNA double-strand breaks. (Schieber and Chandel 2014). The relationship between R-loops, Rad52, and DNA break repair has been previously reported in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* when a double-strand break was induced (Keskin et al., 2014; Ohle et al., 2016; Mazina et al., 2017). Genotoxic stresses also induce transcriptional silencing of ribosomal genes (rDNAs) in the nucleolus. The stabilization of R-loops on rDNAs during hyperosmotic stress conditions leads to a functional cascade of events that repressed Pol-I-dependent transcription. In this case, the ssDNA displaced by the R-loop is coated by the Replication Protein A (RPA) that leads to the recruitment of the ATM- and Rad3-related (ATR) kinase. Once the ATR kinase is activated, it shuts down the transcription of rRNAs (Velichko et al., 2019). Hence, stress-induced R-loops use their RNA-DNA hybrid or the displaced ssDNA to target DNA repair proteins and transcriptional repressors to actively transcribed genes in order to protect the cell from stress-induced damage.

Organismal consequences of R-loop formation in the response to DNA damage have been recently published (Chatzidoukaki



et al., 2021). One of the consequences of persistent DNA damage is premature aging. Using a knockout *ERCC* mouse model deficient for DNA repair, Chatzidoukaki and others demonstrated that pancreatic cells of these animals accumulated ssDNA in the cytoplasm, which activates a viral-like immune response. The accumulation of cytoplasmic ssDNA depends on R-loop formation in response to DNA damage. Co-transcriptional formed R-loops cannot be solved because they do not bind the nuclease XPF efficiently in the absence of ERCC. Therefore, DNA damage leads to the accumulation of ssDNA in the nucleus that is further released to the cytoplasm (Chatzidoukaki et al., 2021). The position and sequence of the R-loops detected in this study remain unknown since most of the experiments done to detect them rely on immunofluorescence using the S9.6 antibody.

Localization of R-loops over the genome and its changes in response to environmental stimuli in *Arabidopsis Thaliana* have recently provided an overview of their landscape and dynamics (Xu et al., 2020). In general, R-loops remain stable to changes in environmental conditions except at 84 h of recovery from a long heat shock. Both heat shock and recovery change the patterns and levels of R-loop localized in gene bodies, and antisense R-loops, usually found at the promoter of genes, also increased upon heat shock and recovery. These results suggest that R-loops might regulate the responses to changes in temperature. However, genome-wide changes in R-loop did not reflect changes in

RNA levels during recovery from stress. Hence, either R-loops have a limited role during stress or longer recovery, and recurrent stress should be considered to analyze the biological relevance of stress-induced R-loops.

5.2 DNA:RNA Triplexes

Similar to R-loops, we have just started discovering that ncRNAs also participate in the stress response by forming DNA:RNA triplexes. Three examples of triplex-forming ncRNAs, *PAPAS*, *PARTICLE*, and *KHPS1*, have been described to regulate gene expression activities during stress conditions (Greifenstein et al., 2021). Elevated temperatures (42°C) upregulate the ncRNA *PAPAS* transcription (promoter and pre-rRNA antisense) in mammalian cells. Transcription of >10 kb *PAPAS* occurs in an antisense orientation to pre-rRNAs, which favors the *cis*-interaction of *PAPAS* with different regions of the rDNA and the formation of an DNA:RNA triplex at the enhancer of rDNA. The DNA:RNA triplex structure mediates the repression of rDNA transcription through the recruitment of the CHD4 subunit of the nucleosome remodeling and deacetylation NuRD complex to the rDNA promoter. Stress-induced dephosphorylation of CHD4 favors its interaction with the A-rich region of *PAPAS*, further promoting the targeting of chromatin remodeling complexes like the histone methyltransferase Suv4-20h2 to repress the transcription of ribosomal genes upon stress (Zhao et al., 2018). The

recruitment of CHD4/NuRD to rDNA by *PAPAS* during heat and hypoosmotic stress also represses transcription by shifting the promoter-bound nucleosomes. Targeting of the adenosine triphosphatase subunit of NuRD leads to the deacetylation of histones and movement of the promoter-bound nucleosome into a position that is refractory to transcription initiation (Zhao et al., 2016b; Zhao et al., 2016a).

Ionizing radiation also represents stress for the cells because it causes macromolecular damage. Low doses of ionizing radiation induce the expression of the ncRNA *PARTICLE* (promoter of *MAT2A*-antisense radiation-induced circulating ncRNA) in mammalian cells in culture. *PARTICLE* forms a DNA:RNA triplex within the CpG island localized in the promoter of the *MAT2A* gene (methionine adenosyltransferase) and suppresses its expression (O'Leary et al., 2015). *PARTICLE* also interacts with the G9s lysine methyltransferase and the SUZ12 subunit of the PRC2 complex targeting the methylation of H3K9 and trimethylation of H3K27 to the *MAT2A* promoter, respectively, which also mediates its transcriptional silencing (O'Leary et al., 2015). In contrast with these repressive transcriptional functions of *PAPAS* and *PARTICLE*, the human ncRNA *Khps1* favors the transcription of the *SPHK1* proto-oncogene to protect cells from apoptosis. The transcription factor E2F1, which plays a role in cell-cycle progression, DNA-damage response, and apoptosis, induces the transcription of *Khps1* in the antisense orientation of *SPHK1* gene. *Khps1* forms an DNA:RNA triplex in the *SPHK1* promoter that recruits the histone acetyltransferase p300/CBP. Acetylation of histones favors the open of chromatin and the activation of *SPHK1* transcription by E2F1, which offsets E2F1 proapoptotic activity (Postepska-Igielska et al., 2015).

Overall, the specificity provided by the nucleotide sequence of ncRNAs and their capacity to form DNA:RNA triplexes provides the means to target chromatin remodelers to specific genes and regulate their transcriptional status.

DISCUSSION

R-loops and DNA:RNA triplexes involving coding and ncRNA are present throughout the genome of various organisms ranging from yeast to humans. As presented in this review, these non-canonical nucleic acid structures involving ncRNA participate in the transcriptional activation and repression of protein-coding genes by targeting transcription factors or epigenetic and DNA modifiers lacking sequence-specific DNA binding activity. We have highlighted examples that provide detailed information about the molecular mechanisms underlying gene expression regulation. However, few mechanistic examples have been the object of in-depth analysis. Non-canonical nucleic acids might rather provide a structure instead of a sequence to be recognized by a protein. Therefore, these structures may constitute an additional general regulator of gene expression, which might be essential for development and survival to stress.

To this day, all the techniques available to study non-canonical nucleic acid structures have limitations. The common caveat to all of them is their lack of capacity to identify the RNA and the DNA

part simultaneously. In mammals (Sanz et al., 2016), *Arabidopsis thaliana* (Xu et al., 2017), and *Drosophila melanogaster* (Alecki et al., 2020), 8 to 15% of the R-loops are formed at intergenic regions. In *Drosophila melanogaster*, 20% of the R-loops don't overlap with any annotated transcript. Hence, R-loops can be formed in trans between DNA and RNA with potentially microhomology regions of ten to thirty nucleotides. Developing specific antibodies or markers to identify DNA:RNA triplexes *in vivo* and improving the specificity of S9.6 for imaging purposes (Smolka et al., 2021) will allow the identification of these structures within individual cells. Consequently, the number of RNAs that can act *in trans* to form non-canonical nucleic acid structures might be underestimated. Although few studies have looked at RNA-DNA hybrid, the main part of an R-loop, to identify its interacting proteins by mass spectrometry, no attention has been given to the remaining part of the R-loop, the dsDNA-ssDNA-RNA-DNA hybrid junction, nor the DNA:RNA triplex. Additionally, these techniques lack the temporal resolution needed to study the dynamics of formation and resolution of R-loops in the context of the cell and the organisms during development.

Proteins responsible for the resolution of R-loops have been identified, but it is still unknown which proteins participate in the formation or resolution of DNA:RNA triplexes. The main R-loop resolving proteins are RNase H1 and H2, RNA-DNA helicases, and topoisomerases (Hyjek et al., 2019; Chon et al., 2009; Kim et al., 1999; Sollier et al., 2014; Chakraborty and Grosse 2011; Manzo et al., 2018; Zhang et al., 2019). The knockdown or overexpression of RNase H1 has limited and controversial effects on R-loops genome-wide, and RNase H2 and RNA-DNA helicases affect many R-loops, which makes it difficult to conclude on a general function of R-loops because these approaches stress cells. Topoisomerases cannot be knockdown in higher eukaryotes because they are essential as they reduce the level of supercoiling DNA during DNA replication and transcription (Lee et al., 1993; Morham et al., 1996; McClendon et al., 2005; Plank et al., 2005; Baranello et al., 2016; Pommier et al., 2016; Capranico et al., 2017; Achar et al., 2020). Hence, their role on R-loop resolution has not been studied in higher eukaryotes. Our understanding of the function of specific R-loop in gene regulation requires tools to remove specific R-loops without affecting the remaining others instead of removing the RNA and the DNA that composed them. These approaches complemented with high-resolution imaging will provide the spatiotemporal resolution required to study their nuclear localization and dynamics over development and stress (Sato et al., 2020). We foresee that these technologies will widen our knowledge on their formation and resolution and the capacity to investigate their specific functions in higher organisms.

AUTHOR CONTRIBUTIONS

CA wrote the sections on biochemistry techniques and development, reviewed, and edited the manuscript. MV wrote the sections on microscopy and stress, reviewed, and edited the manuscript.

FUNDING

FRQS postdoctoral fellowship 300232, NERC grant RGPIN-2019-04767.

REFERENCES

- Achar, Y. J., Adhil, M., Choudhary, R., Gilbert, N., Foiani, M., and Foiani, Marco. (2020). Negative Supercoil at Gene Boundaries Modulates Gene Topology. *Nature* 577 (7792), 701–705. doi:10.1038/s41586-020-1934-4
- Alecki, C., and Francis, N. J. (2021). Identification of R-Loop-Forming Sequences in *Drosophila Melanogaster* Embryos and Tissue Culture Cells Using DRIP-Seq. *Bio Protoc.* 11 (9), e4011. doi:10.21769/BioProtoc.4011
- Alecki, C., Chiwara, V., Sanz, L. A., Grau, D., Arias Pérez, O., Boulier, E. L., et al. (2020). RNA-DNA Strand Exchange by the *Drosophila* Polycomb Complex PRC2. *Nat. Commun.* 11 (1), 1–14. doi:10.1038/s41467-020-15609-x
- An, B., Kameda, T., and Imamura, T. (2021). The Evolutionary Acquisition and Mode of Functions of Promoter-Associated Non-coding RNAs (PancRNAs) for Mammalian Development. *Essays Biochem.* 65 (4), 697–708. doi:10.1042/EBC20200143
- Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., et al. (2019). GADD45A Binds R-Loops and Recruits TET1 to CpG Island Promoters. *Nat. Genet.* 51 (2), 217–223. doi:10.1038/s41588-018-0306-6
- Audas, T. E., Jacob, M. D., and Lee, S. (2012). Immobilization of Proteins in the Nucleolus by Ribosomal Intergenic Spacer Noncoding RNA. *Mol. Cell* 45 (2), 147–157. doi:10.1016/j.molcel.2011.12.012
- Bacolla, A., Wang, G., and Vasquez, K. M. (2015). New Perspectives on DNA and RNA Triplexes as Effectors of Biological Activity. *Plos Genet.* 11 (12), e1005696. doi:10.1371/journal.pgen.1005696
- Balas, Maggie. M., Hartwick, Erik. W., Barrington, Chloe., Roberts, Justin. T., Wu, Stephen. K., Ryan, Bettcher., et al. (2020). RNA Matchmaking Remodels LncRNA Structure and Promotes PRC2 Activity. April, 2020.04.13.040071. BioRxiv. doi:10.1101/2020.04.13.040071
- Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B. N., Chung, H.-J., Chan-Salis, K. Y., et al. (2016). RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* 165 (2), 357–371. doi:10.1016/j.cell.2016.02.036
- Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-Based Transcriptomics in Thousands of Single Human Cells at Single-Molecule Resolution. *Nat. Methods* 10 (11), 1127–1133. doi:10.1038/nmeth.2657
- Benitez-Guijarro, M., Lopez-Ruiz, C., Tarnauskaite, Z., Murina, O., Mohammad, M. M., Williams, T. C., et al. (2018). RNase H₂, Mutated in Aicardi-Goutieres Syndrome, Promotes LINE-1 Retrotransposition. *EMBO J.* 37 (15), e98506. doi:10.15252/embj.201798506
- Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Bialesch, A., Padovan-Merhar, O., Regev, A., et al. (2015). Localization and Abundance Analysis of Human LncRNAs at Single-Cell and Single-Molecule Resolution. *Genome Biol.* 16 (1), 20. doi:10.1186/s13059-015-0586-4
- Capranico, G., Marinello, J., and Chillemi, G. (2017). Type I DNA Topoisomerases. *J. Med. Chem.* 60 (6), 2169–2192. doi:10.1021/acs.jmedchem.6b00966
- Cargill, M., Venkataraman, R., and Lee, S. (2021). DEAD-box RNA Helicases and Genome Stability. *Genes* 12 (10), 1471. doi:10.3390/genes12101471
- Chakraborty, P., and Grosse, F. (2011). Human DHX9 Helicase Preferentially Unwinds RNA-Containing Displacement Loops (R-Loops) and G-Quadruplexes. *DNA Repair* 10 (6), 654–665. doi:10.1016/j.dnarep.2011.04.013
- Chatzidoukaki, O., Stratigi, K., Goulielmaki, E., Niotis, G., Akalestou-Clocher, A., Gkirtzimanaki, K., et al. (2021). R-loops Trigger the Release of Cytoplasmic SsDNAs Leading to Chronic Inflammation upon DNA Damage. *Sci. Adv.* 7 (47), eabj5769. doi:10.1126/sciadv.abj5769
- Chellini, L., Frezza, V., and Paronetto, M. P. (2020). Dissecting the Transcriptional Regulatory Networks of Promoter-Associated Noncoding RNAs in Development and Cancer. *J. Exp. Clin. Cancer Res.* 39 (March), 51. doi:10.1186/s13046-020-01552-8

ACKNOWLEDGMENTS

We would like to thank Rafael Cuesta-Sanchez for critically reading the manuscript.

- Chen, L., Chen, J.-Y., Zhang, X., Ying, G., Xiao, R., Shao, C., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell.* 68 (4), 745–757.e5. doi:10.1016/j.molcel.2017.10.008
- Chen, J.-Y., Zhang, X., Fu, X.-D., and Chen, L. (2019). R-ChIP for Genome-Wide Mapping of R-Loops by Using Catalytically Inactive RNASEH1. *Nat Protoc.* 14 (5), 1661–1685. doi:10.1038/s41596-019-0154-6
- Chon, H., Vassilev, A., DePamphilis, M. L., Zhao, Y., Zhang, J., Burgers, P. M., et al. (2009). Contributions of the Two Accessory Subunits, RNASEH2B and RNASEH2C, to the Activity and Properties of the Human RNase H2 Complex. *Nucleic Acids Res.* 37 (1), 96–110. doi:10.1093/nar/gkn913
- Costa, S., and Dean, C. (2019). Storing Memories: The Distinct Phases of Polycomb-Mediated Silencing of Arabidopsis FLC. *Biochem. Soc. Trans.* 47 (4), 1187–1196. doi:10.1042/BST20190255
- Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A Gene Loop Containing the Floral Repressor FLC Is Disrupted in the Early Phase of Vernalization. *Embo J.* 32 (1), 140–148. doi:10.1038/emboj.2012.324
- Cristini, A., Groh, M., Kristiansen, M. S., and Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cel Rep.* 23 (6), 1891–1905. doi:10.1016/j.celrep.2018.04.025
- Crossley, M. P., Bocek, M., and Cimprich, K. A. (2019). R-loops as Cellular Regulators and Genomic Threats. *Mol. Cell* 73 (3), 398–411. doi:10.1016/j.molcel.2019.01.024
- Cusanelli, E., Romero, C. A. P., and Chartrand, P. (2013). Telomeric Noncoding RNA TERRA Is Induced by Telomere Shortening to Nucleate Telomerase Molecules at Short Telomeres. *Mol. Cell* 51 (6), 780–791. doi:10.1016/j.molcel.2013.08.029
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE V7 Catalog of Human Long Noncoding RNAs: Analysis of Their Gene Structure, Evolution, and Expression. *Genome Res.* 22 (9), 1775–1789. doi:10.1101/gr.132159.111
- Fang, Y., Chen, L., Lin, K., Feng, Y., Zhang, P., Pan, X., et al. (2019). Characterization of Functional Relationships of R-Loops with Gene Transcription and Epigenetic Modifications in rice. *Genome Res.* 29, 1287–1297. gr.246009. doi:10.1101/gr.246009.118
- Femino, A. M., Fay, F. S., Fogarty, K., and Singer, R. H. (1998). Visualization of Single RNA Transcripts *In Situ*. *Science* 280 (5363), 585–590. doi:10.1126/science.280.5363.585
- Fernandes, J., Acuña, S., Juliana, I., Aoki, Lucile., Floeter-Winter, M., Aoki, J., et al. (2019). Long Non-coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *ncRNA* 5 (1), 17. doi:10.3390/ncrna5010017
- Gilbert, Scott. F. (2000). *The Origins of Anterior-Posterior Polarity* *Developmental Biology*. 6th Edition. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK10039/>.
- Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., and Chédin, F. (2012). R-loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Mol. Cell* 45 (6), 814–825. doi:10.1016/j.molcel.2012.01.017
- Goenka, A., Sengupta, S., Pandey, R., Parihar, R., Mohanta, G. C., Mukerji, M., et al. (2016). Human Satellite-III Non-coding RNAs Modulate Heat Shock-Induced Transcriptional Repression. *J. Cell Sci.* 129 (19), 3541–3552. doi:10.1242/jcs.189803
- Gong, J., Ju, Y., Shao, D., and Zhang, Q. C. (2018). Advances and Challenges towards the Study of RNA-RNA Interactions in a Transcriptome-wide Scale. *Quant Biol.* 6 (3), 239–252. doi:10.1007/s40484-018-0146-5
- Greifenstein, A. A., Jo, S., and Bierhoff, H. (2021). RNA:DNA Triple Helices: from peculiar Structures to Pervasive Chromatin Regulators. *Essays Biochem.* 65 (4), 731–740. doi:10.1042/EBC20200089
- Grote, P., and Herrmann, B. G. (2013). The Long Non-coding RNA Fendrr Links Epigenetic Control Mechanisms to Gene Regulatory Networks in Mammalian Embryogenesis. *RNA Biol.* 10 (10), 1579–1585. doi:10.4161/rna.26165

- Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., et al. (2013). The Tissue-specific LncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Dev. Cel* 24 (2), 206–214. doi:10.1016/j.devcel.2012.12.012
- Guett, C., Lienemann, P., Sirri, V., Grummt, I., Hernandez-Verdun, D., Hottiger, M. O., et al. (2010). The NoRC Complex Mediates the Heterochromatin Formation and Stability of Silent RRNA Genes and Centromeric Repeats. *Embo J.* 29 (13), 2135–2146. doi:10.1038/emboj.2010.17
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Wong, David, J., Horlings, H. M., et al. (2010). Long Non-coding RNA HOTAIR Reprograms Chromatin State to Promote Cancer Metastasis. *Nature* 464 (7291), 1071–1076. doi:10.1038/nature08975
- Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., et al. (2021). CLIP and Complementary Methods. *Nat. Rev. Methods Primers* 1 (1), 1–23. doi:10.1038/s43586-021-00018-1
- Hajirnis, N., and Mishra, R. K. (2021). Homeotic Genes: Clustering, Modularity, and Diversity. *Front. Cel Dev. Biol.* 9 (August), 718308. doi:10.3389/fcell.2021.718308
- Hamazaki, N., Uesaka, M., Nakashima, K., Agata, K., and Imamura, T. (2015). Gene Activation-Associated Long Noncoding RNAs Function in Mouse Preimplantation Development. *Development (Cambridge, England)* 142 (5), 910–920. doi:10.1242/dev.116996
- Hartono, S. R., Malapert, A., Legros, P., Bernard, P., Chédin, F., and Vanoosthuysse, V. (2018). Amélie Malapert, Pénélope Legros, Pascal Bernard, Frédéric Chédin, and Vincent Vanoosthuysse The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *J. Mol. Biol.* 430 (3), 272–284. doi:10.1016/j.jmb.2017.12.016
- Hyjek, M., Figiel, M., and Nowotny, M. (2019). RNases H: Structure and Mechanism. *DNA Repair* 84 (December), 102672. doi:10.1016/j.dnarep.2019.102672
- Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B. M., Khochbin, S., et al. (2004). Stress-Induced Transcription of Satellite III Repeats. *J. Cel Biol.* 164 (1), 25–33. doi:10.1083/jcb.200306104
- Kaneko, S., Son, J., Shen, S. S., Reinberg, D., and Bonasio, R. (2013). PRC2 Binds Active Promoters and Contacts Nascent RNAs in Embryonic Stem Cells. *Nat. Struct. Mol. Biol.* 20 (11), 1258–1264. doi:10.1038/nsmb.2700
- Keskin, H., Shen, Y., Huang, F., Patel, M., Yang, T., Ashley, K., et al. (2014). Transcript-RNA-templated DNA Recombination and Repair. *Nature* 515 (7527), 436–439. doi:10.1038/nature13682
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., et al. (2009). Many Human Large Intergenic Noncoding RNAs Associate with Chromatin-Modifying Complexes and Affect Gene Expression. *Pnas* 106 (28), 11667–11672. doi:10.1073/pnas.0904715106
- Kim, D.-H., and Sung, S. (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Dev. Cel* 40, 302–312. doi:10.1016/j.devcel.2016.12.021
- Kim, D.-H., Xi, Y., and Sung, S. (2017). Modular Function of Long Noncoding RNA, COLDAIR, in the Vernalization Response. *Plos Genet.* 13 (7), e1006939. doi:10.1371/journal.pgen.1006939
- Kim, H.-D., Choe, J., and Seo, Y.-S. (1999). The Sen1+ Gene of Schizosaccharomyces Pombe, a Homologue of Budding Yeast SEN1, Encodes an RNA and DNA Helicase. *Biochemistry* 38 (44), 14697–14710. doi:10.1021/bi991470c
- König, F., Schubert, T., and Längst, G. (2017). The Monoclonal S9.6 Antibody Exhibits Highly Variable Binding Affinities towards Different R-Loop Sequences. *PLOS ONE* 12 (6), e0178875. doi:10.1371/journal.pone.0178875
- Koskas, S., Decottignies, A., Dufour, S., Pezet, M., Verdel, A., Vourc'h, C., et al. (2017). Heat Shock Factor 1 Promotes TERRA Transcription and Telomere Protection upon Heat Stress. *Nucleic Acids Res.* 45 (11), 6321–6333. doi:10.1093/nar/gkx208
- Kwon, S. (2013). Single-Molecule Fluorescence *In Situ* Hybridization: Quantitative Imaging of Single RNA Molecules. *BMB Rep.* 46 (2), 65–72. doi:10.5483/BMBRep.2013.46.2.016
- Lee, M. P., Brown, S. D., Chen, A., and Hsieh, T. S. (1993). DNA Topoisomerase I Is Essential in *Drosophila* Melanogaster. *Proc. Natl. Acad. Sci.* 90 (14), 6656–6660. doi:10.1073/pnas.90.14.6656
- Lenstra, T. L., Coulon, A., Chow, C. C., and Larson, D. R. (2015). Single-Molecule Imaging Reveals a Switch between Spurious and Functional NcRNA Transcription. *Mol. Cel* 60 (4), 597–610. doi:10.1016/j.molcel.2015.09.028
- Li, L., Liu, B., Wapinski, O. L., Tsai, M.-C., Qu, K., Zhang, J., et al. (2013). Targeted Disruption of Hotair Leads to Homeotic Transformation and Gene Derepression. *Cel Rep.* 5 (1), 3–12. doi:10.1016/j.celrep.2013.09.003
- Li, Y., Song, Y., Xu, W., Li, Q., Wang, X., Li, K., et al. (2020). R-loops Coordinate with SOX2 in Regulating Reprogramming to Pluripotency. *Sci. Adv.* 6 (24), eaba0777. doi:10.1126/sciadv.aba0777
- Li, Y., Syed, J., and Sugiyama, H. (2016). RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cel Chem. Biol.* 23 (11), 1325–1333. doi:10.1016/j.chembiol.2016.09.011
- Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R., and Chédin, F. (2015). Genome-wide DNA Hypomethylation and RNA:DNA Hybrid Accumulation in Aicardi-Goutières Syndrome. *ELife* 4. doi:10.7554/eLife.08007
- Lin, C., and Miles, W. O. (2019). Beyond CLIP: Advances and Opportunities to Measure RBP-RNA and RNA-RNA Interactions. *Nucleic Acids Res.* 47 (11), 5490–5501. doi:10.1093/nar/gkz295
- Lockhart, A., Pires, V. B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., et al. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cel Rep.* 29 (9), 2890–2900.e5. doi:10.1016/j.celrep.2019.10.108
- Lotfi, C. F. P., Passaia, B. S., and Kremer, J. L. (2021). Role of the BHLH Transcription Factor TCF21 in Development and Tumorigenesis. *Braz. J. Med. Biol. Res.* 54 (5), e10637. doi:10.1590/1414-431X202010637
- Lu, Z., and Chang, H. Y. (2018). The RNA Base-Pairing Problem and Base-Pairing Solutions. *Cold Spring Harb Perspect. Biol.* 10 (12), a034926. doi:10.1101/cshperspect.a034926
- Malig, M., and Chedin, F. (2020). Characterization of R-Loop Structures Using Single-Molecule R-Loop Footprinting and Sequencing. *Methods Mol. Biol. (Clifton, N.J.)* 2161, 209–228. doi:10.1007/978-1-0716-0680-3_15
- Mallo, M., Wellik, D. M., and Deschamps, J. (2010). Hox Genes and Regional Patterning of the Vertebrate Body Plan. *Dev. Biol.* 344 (1), 7–15. doi:10.1016/j.ydbio.2010.04.024
- Manzo, S. G., Hartono, S. R., Sanz, L. A., Marinello, J., De Biasi, S., Cossarizza, A., et al. (2018). DNA Topoisomerase I Differentially Modulates R-Loops across the Human Genome. *Genome Biol.* 19 (July). doi:10.1186/s13059-018-1478-1
- Mayer, C., Neubert, M., and Grummt, I. (2008). The Structure of NoRC-associated RNA Is Crucial for Targeting the Chromatin Remodelling Complex NoRC to the Nucleolus. *EMBO Rep.* 9 (8), 774–780. doi:10.1038/embo.2008.109
- Mazina, O. M., Keskin, H., Hanamshet, K., Storic, F., and Mazin, A. V. (2017). Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. *Mol. Cel* 67 (1), 19–29.e3. doi:10.1016/j.molcel.2017.05.019
- McClendon, A. K., Rodriguez, A. C., and Osheroff, N. (2005). Human Topoisomerase II α Rapidly Relaxes Positively Supercoiled DNA. *J. Biol. Chem.* 280 (47), 39337–39345. doi:10.1074/jbc.M503320200
- McKechnie, S. W., Halford, M. M., McColl, G., and Hoffmann, A. A. (1998). Both Allelic Variation and Expression of Nuclear and Cytoplasmic Transcripts of Hsr-Omega Are Closely Associated with Thermal Phenotype in *Drosophila*. *Proc. Natl. Acad. Sci.* 95 (5), 2423–2428. doi:10.1073/pnas.95.5.2423
- Miglietta, G., Russo, M., and Capranico, G. (2020). G-quadruplex-R-loop Interactions and the Mechanism of Anticancer G-Quadruplex Binders. *Nucleic Acids Res.* 48 (21), 11942–11957. doi:10.1093/nar/gkaa944
- Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., et al. (2015). MEG3 Long Noncoding RNA Regulates the TGF- β Pathway Genes through Formation of RNA-DNA Triplex Structures. *Nat. Commun.* 6 (1), 7743. doi:10.1038/ncomms8743
- Morgan, A. R., and Wells, R. D. (1968). Specificity of the Three-Stranded Complex Formation between Double-Stranded DNA and Single-Stranded RNA Containing Repeating Nucleotide Sequences. *J. Mol. Biol.* 37 (1), 63–80. doi:10.1016/0022-2836(68)90073-9
- Morham, S. G., Kluckman, K. D., Vouliamos, N., and Smithies, O. (1996). Targeted Disruption of the Mouse Topoisomerase I Gene by Camptothecin Selection. *Mol. Cel Biol* 16 (12), 6804–6809. doi:10.1128/MCB.16.12.6804
- Moye, A. L., Porter, K. C., Cohen, Scott, B., Phan, Tram., Cohen, S. B., Phan, T., et al. (2015). Telomeric G-Quadruplexes Are a Substrate and Site of Localization for Human Telomerase. *Nat. Commun.* 6 (July). doi:10.1038/ncomms8643

- Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., et al. (2013). FISH-quant: Automatic Counting of Transcripts in 3D FISH Images. *Nat. Methods* 10 (4), 277–278. doi:10.1038/nmeth.2406
- Niehrs, C., and Luke, B. (2020). Regulatory R-Loops as Facilitators of Gene Expression and Genome Stability. *Nat. Rev. Mol. Cell Biol* 21, 167–178. doi:10.1038/s41580-019-0206-3
- Nowotny, M., Cerritelli, S. M., Ghirlando, R., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2008). Specific Recognition of RNA/DNA Hybrid and Enhancement of Human RNase H1 Activity by HBD. *Embo J.* 27 (7), 1172–1181. doi:10.1038/emboj.2008.44
- O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* 9, 402. doi:10.3389/fendo.2018.00402
- Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* 167 (4), 1001–1013.e7. doi:10.1016/j.cell.2016.10.001
- O'Leary, V. B., Ovsepian, S. V., Carrascosa, L. G., Buske, F. A., Radulovic, V., Niyazi, M., et al. (2015). PARTICLE, a Triplex-Forming Long NcRNA, Regulates Locus-specific Methylation in Response to Low-Dose Irradiation. *Cel Rep.* 11 (3), 474–485. doi:10.1016/j.celrep.2015.03.043
- Phillips, D. D., Garboczi, D. N., Singh, K., Hu, Z., Leppla, S. H., and Leysath, C. E. (2013). The Sub-nanomolar Binding of DNA-RNA Hybrids by the Single-Chain Fv Fragment of Antibody S9.6. *J. Mol. Recognit.* 26 (8), 376–381. doi:10.1002/jmr.2284
- Place, R. F., and Noonan, E. J. (2014). Non-Coding RNAs Turn up the Heat: An Emerging Layer of Novel Regulators in the Mammalian Heat Shock Response. *Cell Stress and Chaperones* 19 (2), 159–172. doi:10.1007/s12192-013-0456-5
- Plank, J. L., Chu, S. H., Pohlhaus, J. R., Wilson-Sali, T., and Hsieh, T.-s. (2005). *Drosophila melanogaster* Topoisomerase III α Preferentially Relaxes a Positively or Negatively Supercoiled Bubble Substrate and Is Essential during Development. *J. Biol. Chem.* 280 (5), 3564–3573. doi:10.1074/jbc.M411337200
- Pommier, Y., Sun, Y., Huang, S.-y. N., and Nitiss, J. L. (2016). Roles of Eukaryotic Topoisomerases in Transcription, Replication and Genomic Stability. *Nat. Rev. Mol. Cell Biol* 17 (11), 703–721. doi:10.1038/nrm.2016.111
- Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., et al. (2015). LncRNA Khps1 Regulates Expression of the Proto-Oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. *Mol. Cell* 60 (4), 626–636. doi:10.1016/j.molcel.2015.10.001
- Radhakrishnan, I., and Patel, D. J. (1994). Solution Structure and Hydration Patterns of a Pyrimidine-Purine-Pyrimidine DNA Triplex Containing a Novel T⁺CG Base-Triple. *J. Mol. Biol.* 241 (4), 600–619. doi:10.1006/jmbi.1994.1534
- Ramanathan, M., Porter, D. F., and Khavari, P. A. (2019). Methods to Study RNA-Protein Interactions. *Nat. Methods* 16 (3), 225–234. doi:10.1038/s41592-019-0330-1
- Rinn, J. L., Kertesz, M., Wang, J. K., Wang, Sharon, Squazzo, L., Squazzo, S. L., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* 129 (7), 1311–1323. doi:10.1016/j.cell.2007.05.022
- Rosa, S., Duncan, S., and Dean, C. (2016). Mutually Exclusive Sense-Antisense Transcription at FLC Facilitates Environmentally Induced Gene Repression. *Nat. Commun.* 7 (1), 1–7. doi:10.1038/ncomms13031
- Santoro, R., Schmitz, K. M., Sandoval, J., and Grummt, I. (2010). Intergenic Transcripts Originating from a Subclass of Ribosomal DNA Repeats Silence Ribosomal RNA Genes in *Trans. EMBO Rep.* 11 (1), 52–58. doi:10.1038/embor.2009.254
- Sanz, L. A., and Chédin, F. (2019). High-resolution, Strand-specific R-Loop Mapping via S9.6-based DNA-RNA Immunoprecipitation and High-Throughput Sequencing. *Nat. Protoc.* 14 (6), 1734–1755. doi:10.1038/s41596-019-0159-1
- Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., et al. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol. Cell* 63 (1), 167–178. doi:10.1016/j.molcel.2016.05.032
- Sato, H., Das, S., Singer, R. H., and Vera, M. (2020). Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression. *Annu. Rev. Biochem.* 89 (1), 159–187. doi:10.1146/annurev-biochem-011520-104955
- Schieber, M., and Chandel, N. S. (2014). ROS Function in Redox Signaling and Oxidative Stress. *Curr. Biol.* 24 (10), R453–R462. doi:10.1016/j.cub.2014.03.034
- Schmitz, K.-M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of Noncoding RNA with the RDNA Promoter Mediates Recruitment of DNMT3b and Silencing of RRNA Genes. *Genes Dev.* 24 (20), 2264–2269. doi:10.1101/gad.590910
- Schreiner, W. P., Pagliuso, D. C., Garrigues, J. M., Chen, J. S., Aalto, A. P., and Pasquinelli, A. E. (2019). Remodeling of the *Caenorhabditis Elegans* Non-coding RNA Transcriptome by Heat Shock. *Nucleic Acids Res.* 47 (18), 9829–9841. doi:10.1093/nar/gkz693
- Sentürk Cetin, N., Kuo, C.-C., Ribarska, T., Li, R., Costa, I. G., and Grummt, I. (2019). Isolation and Genome-wide Characterization of Cellular DNA:RNA Triplex Structures. *Nucleic Acids Res.* 47 (5), 2306–2321. doi:10.1093/nar/gky1305
- Sinnamon, J. R., and Czaplinski, K. (2014). RNA Detection *In Situ* with FISH-STICs. *RNA* 20 (2), 260–266. doi:10.1261/rna.041905.113
- Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019). R-loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol. Cell* 73 (5), 930–945.e4. doi:10.1016/j.molcel.2018.12.016
- Slobodin, B., and Gerst, J. E. (2010). A Novel mRNA Affinity Purification Technique for the Identification of Interacting Proteins and Transcripts in Ribonucleoprotein Complexes. *RNA* 16 (11), 2277–2290. doi:10.1261/rna.2091710
- Smolka, J. A., Sanz, L. A., Hartono, S. R., and Chédin, F. (2021). Recognition of RNA by the S9.6 Antibody Creates Pervasive Artifacts when Imaging RNA:DNA Hybrids. *J. Cell Biol.* 220 (6). doi:10.1083/jcb.202004079
- Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., and Cimprich, K. A. (2014). Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol. Cell* 56 (6), 777–785. doi:10.1016/j.molcel.2014.10.020
- Sperling, A. S., Gibson, C. J., and Ebert, B. L. (2017). The Genetics of Myelodysplastic Syndrome: From Clonal Haematopoiesis to Secondary Leukaemia. *Nat. Rev. Cancer* 17 (1), 5–19. doi:10.1038/nrc.2016.112
- Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene Regulation by Long Non-Coding RNAs and its Biological Functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9
- Steffen, P. A., and Ringrose, L. (2014). What Are Memories Made of? How Polycomb and Trithorax Proteins Mediate Epigenetic Memory. *Nat. Rev. Mol. Cell Biol* 15 (5), 340–356. doi:10.1038/nrm3789
- Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., and Dean, C. (2013). R-loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. *Science* 340 (6132), 619–621. doi:10.1126/science.1234848
- Sun, Y.-M., and Chen, Y.-Q. (2020). Principles and Innovative Technologies for Decrypting Noncoding RNAs: From Discovery and Functional Prediction to Clinical Application. *J. Hematol. Oncol.* 13 (1), 109. doi:10.1186/s13045-020-00945-8
- Tsai, M.-C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J. K., Lan, F., et al. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* 329 (5992), 689–693. doi:10.1126/science.1192002
- Uesaka, M., Nishimura, O., Go, Y., Nakashima, K., Agata, K., and Imamura, T. (2014). Bidirectional Promoters Are the Major Source of Gene Activation-Associated Non-coding RNAs in Mammals. *BMC Genomics* 15 (1), 35. doi:10.1186/1471-2164-15-35
- Velichko, A. K., Petrova, N. V., Luzhin, A. V., Strelkova, O. S., Ovsyannikova, N., Kireev, I. I., et al. (2019). Hypoosmotic Stress Induces R Loop Formation in Nucleoli and ATR/ATM-Dependent Silencing of Nucleolar Transcription. *Nucleic Acids Res.* 47 (13), 6811–6825. doi:10.1093/nar/gkz436
- Vera, M., Biswas, J., Senecal, A., Singer, R. H., and Park, H. Y. (2016). Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. *Annu. Rev. Genet.* 50 (November), 267–291. doi:10.1146/annurev-genet-120215-034854
- Vieira-Vieira, C. H., Selbach, M., and Selbach, Matthias. (2021). Opportunities and Challenges in Global Quantification of RNA-Protein Interaction via UV Cross-Linking. *Front. Mol. Biosci.* 8 (May). doi:10.3389/fmolb.2021.669939
- Vihervaara, A., Duarte, F. M., and Lis, J. T. (2018). Molecular Mechanisms Driving Transcriptional Stress Responses. *Nat. Rev. Genet.* 19 (6), 385–397. doi:10.1038/s41576-018-0001-6

- Vydzhak, O., Luke, B., and Schindler, N. (2020). Non-coding RNAs at the Eukaryotic rDNA Locus: RNA-DNA Hybrids and beyond. *J. Mol. Biol.* 432 (15), 4287–4304. doi:10.1016/j.jmb.2020.05.011
- Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., and Koshland, D. (2016). S1-DRIP-Seq Identifies High Expression and PolyA Tracts as Major Contributors to R-Loop Formation. *Genes Dev.* 30 (11), 1327–1338. doi:10.1101/gad.280834.116
- Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., Rich, A., et al. (1982). Molecular Structure of r(CGCG)D(TATACGC): a DNA-RNA Hybrid helix Joined to Double Helical DNA. *Nature* 299 (5884), 601–604. doi:10.1038/299601a0
- Wang, I. X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., et al. (2018). Human Proteins that Interact with RNA/DNA Hybrids. *Genome Res.* 28 (9), 1405–1414. doi:10.1101/gr.237362.118
- Wang, Z., Li, K., and Huang, W. (2020). Long Non-coding RNA NEAT1-Centric Gene Regulation. *Cell. Mol. Life Sci.* 77 (19), 3769–3779. doi:10.1007/s00018-020-03503-0
- Wehner, S., Dörrich, A. K., Ciba, P., Wilde, A., and Marz, M. (2014). PRNA. *RNA Biol.* 11 (1), 3–9. doi:10.4161/rna.27448
- Weinreb, J. T., Ghazale, N., Pradhan, K., Gupta, V., Potts, K. S., Tricomi, B., et al. (2021). Excessive R-Loops Trigger an Inflammatory Cascade Leading to Increased HSPC Production. *Dev. Cell* 56 (5), 627–640.e5. doi:10.1016/j.devcel.2021.02.006
- Wu, T., Nance, J., Chu, F., and Fazzio, T. G. (2021). Characterization of R-Loop-Interacting Proteins in Embryonic Stem Cells Reveals Roles in rRNA Processing and Gene Expression. *Mol. Cell Proteomics* 20 (August), 100142. doi:10.1016/j.mcpro.2021.100142
- Wulfridge, P., and Sarma, K. (2021). A Nuclease- and Bisulfite-Based Strategy Captures Strand-specific R-Loops Genome-wide. *ELife* 10 (February), e65146. doi:10.7554/eLife.65146
- Xiao, S., Huang, Q., Ren, H., and Yang, M. (2021). The Mechanism and Function of Super Enhancer RNA. *Genesis* 59 (5–6), e23422. doi:10.1002/dvg.23422
- Xu, C., Wu, Z., Duan, H.-C., Fang, X., Jia, G., and Dean, C. (2021). R-loop Resolution Promotes Co-transcriptional Chromatin Silencing. *Nat. Commun.* 12 (1), 1790. doi:10.1038/s41467-021-22083-6
- Xu, W., Li, K., Li, S., Hou, Q., Zhang, Y., Liu, K., et al. (2020). The R-Loop Atlas of Arabidopsis Development and Responses to Environmental Stimuli. *Plant Cell* 32 (4), 888–903. doi:10.1105/tpc.19.00802
- Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., et al. (2017). The R-Loop Is a Common Chromatin Feature of the Arabidopsis Genome. *Nat. Plants* 3 (9), 704–714. doi:10.1038/s41477-017-0004-x
- Yan, Q., and Sarma, K. (2020). MapR: A Method for Identifying Native R-Loops Genome Wide. *Curr. Protoc. Mol. Biol.* 130 (1), e113. doi:10.1002/cpmb.113
- Yang, H., Berry, S., Olsson, T. S. G., Hartley, M., Howard, M., Dean, C., et al. (2017). Distinct Phases of Polycomb Silencing to Hold Epigenetic Memory of Cold in Arabidopsis. *Science* 357 (6356), 1142–1145. doi:10.1126/science.aan1121
- Yaoyi, L., Yawei, S., Wei, X., Qin, L., Xinxiu, W., Kuan, L., et al. (2020). R-loops coordinate with SOX2 in regulating reprogramming to pluripotency. *Sci Adv.* 2020 Jun 10; 6(24): eaba0777. doi:10.1126/sciadv.aba0777
- Yap, K., Chung, T. H., and Makeyev, E. V. (2022). Hybridization-Proximity Labeling Reveals Spatially Ordered Interactions of Nuclear RNA Compartments. *Mol. Cell* 82 (0), 463–478. doi:10.1016/j.molcel.2021.10.009
- Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., et al. (2020). CRISPR-assisted Detection of RNA-Protein Interactions in Living Cells. *Nat. Methods* 17 (7), 685–688. doi:10.1038/s41592-020-0866-0
- Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T. E., and Lieber, M. R. (2003). R-Loops At Immunoglobulin Class Switch Regions In The Chromosomes Of Stimulated B Cells. *Nat. Immunol.* 4 (5), 442–451. doi:10.1038/ni919
- Zeng, C., Onoguchi, M., and Hamada, M. (2021). Association Analysis of Repetitive Elements and R-Loop Formation across Species. *Mobile DNA* 12 (January), 3. doi:10.1186/s13100-021-00231-5
- Zhang, T., Wallis, M., Petrovic, V., Challis, J., Kalitsis, P., and Hudson, D. F. (2019). Loss of TOP3B Leads to Increased R-Loop Formation and Genome Instability. *Open Biol.* 9 (12), 190222. doi:10.1098/rsob.190222
- Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., et al. (2010). Genome-Wide Identification of Polycomb-Associated RNAs by RIP-Seq. *Mol. Cell* 40 (6), 939–953. doi:10.1016/j.molcel.2010.12.011
- Zhao, Z., Dammert, M. A., Grummt, I., and Bierhoff, H. (2016b). LncRNA-Induced Nucleosome Repositioning Reinforces Transcriptional Repression of RRNA Genes upon Hypotonic Stress. *Cel Rep.* 14 (8), 1876–1882. doi:10.1016/j.celrep.2016.01.073
- Zhao, Z., Dammert, M. A., Hoppe, S., Bierhoff, H., and Grummt, I. (2016a). Heat Shock Represses RRNA Synthesis by Inactivation of TIF-IA and LncRNA-dependent Changes in Nucleosome Positioning. *Nucleic Acids Res.* 44 (17), 8144–8152. doi:10.1093/nar/gkw496
- Zhao, Z., Sentürk, N., Song, C., and Grummt, I. (2018). Lncrna Papas Tethered to the Rdna Enhancer Recruits Hypophosphorylated Chd4/Nurd to Repress Rrna Synthesis at Elevated Temperatures. *Genes Dev.* 32 (11–12), 836–848. doi:10.1101/gad.311688.118

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Alecki and Vera. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.