Published in final edited form as:

Mol Cell. 2023 February 02; 83(3): 442–451. doi:10.1016/j.molcel.2023.01.010.

Reversing the Central Dogma: RNA guided control of DNA in epigenetics and genome editing

Howard Y. Chang[1,2], **Lei S. Qi**[3,4,5]

[1]Center for Personal Dynamic Regulome, Stanford University, Stanford CA 94305

[2]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305

[3]Department of Bioengineering, Stanford University, Stanford CA 94305

[4]Sarafan ChEM-H, Stanford University, Stanford CA 94305

[5]Chan Zuckerberg Biohub, San Francisco, CA 94080

SUMMARY

The central dogma of the flow of genetic information is arguably the crowning achievement of 20th century molecular biology. Reversing the flow of information from RNA to DNA or chromatin has come to the fore in recent years, from the convergence of fundamental discoveries and synthetic biology. Inspired by the example of long noncoding RNAs (lncRNAs) in mammalian genomes that direct chromatin modifications and gene expression, synthetic biologists have repurposed prokaryotic RNA-guided genome defense systems such as CRISPR to edit eukaryotic genomes and epigenome. Here we explore the parallels of these two fields and highlight opportunities for synergy and future breakthroughs.

eTOC blurb

Here, Chang and Qi discuss reversing the flow of information in the Central Dogma, from RNA to DNA or chromatin, highlighting lncRNAs and prokaryotic RNA-guided genome defense systems such as CRISPR to edit eukaryotic genomes and epigenome. They explore the parallels between the two and the future of RNA-based technologies.

INTRODUCTION

The central dogma of gene expression is stunning in its explanatory power and ubiquity in biological systems. As a fundamental theory in molecular biology, the central dogma states that genetic information flows in one direction, from DNA, to RNA, to protein.

Correspondence to: H.Y.C. (howchang@stanford.edu) and L.S.Q. (slqi@stanford.edu).

Declaration of Interests

L.S.Q. is a founder and scientific advisor of Epicrispr Biotechnologies. H.Y.C. is a co founder of Accent Therapeutics, Boundless Bio, Cartography Biosciences, Orbital Therapeutics, and is an advisor of 10x Genomics, Arsenal Biosciences, Chroma Medicine, and Spring Discovery.

From a historical context, the discovery of reverse transcriptase was one the first major revision to the idea of unidirectional information flow from DNA to RNA to protein. In the first part of the 21st century, discovery of small regulatory RNAs was initially focused on post-transcriptional regulation. Subsequent studies of PIWI-interacting RNAs (piRNAs) and long noncoding RNAs (lncRNAs) led to the recognition of the potential of RNA to influence DNA—specifically changing DNA methylation or chromatin in a heritable fashion[1]. In bacteria, retrons are genetic components composed of reverse transcriptase (RT) and non-coding RNA (ncRNA). The RT uses the ncRNA as a template to generate a chimeric RNA/DNA molecule with covalently bonded RNA and DNA components, which work as defenses against phages[2]. The discovery of CRISPR greatly simplified the task of targeting specific DNA sequences[3] and led to a flowering of ideas and new strategies to expand the programmability of RNA, with the aid of protein factors, to control the information flow backwards to DNA (Figure 1). While proteins are considered major players on controlling gene expression, these mounting examples highlight RNA as the often-overlooked and active ingredient in reversing the central dogma.

LONG NONCODING RNAS: INFORMATION FLOW FROM RNA TO CHROMATIN

Eukaryotic genomes are pervasively transcribed to produce numerous lncRNAs. LncRNAs are defined as transcripts that are greater than 200 nucleotides, may be spliced or unspliced, and apparently do not function by encoding for a protein or peptide. This definition of exclusion arose during an era of a protein-centric view of the genome and genome assemblies, despite the fact that just 2% of the human genome is coding. The discovery, characterization, function of lncRNAs have been extensively reviewed elsewhere[4] and we will focus on salient examples that facilitate our discussion. In brief, the human genome encodes over 60,000 lncRNAs, three times the number of protein coding genes. Inherited mutations in lncRNA genes cause Mendelian disorders [5] such as developmental malformations [6].Unbiased genome-wide association studies revealed that inherited variation of lncRNA loci in the human population underlies over one thousand trait associations, 800 of which were not explained by neighboring protein-coding genes[7]. For instance, lncRNA quantitative trait loci are linked to inflammatory bowel disease, type 1 and type 2 diabetes, and coronary artery disease, as well as rare variant associations to body mass index. Systematic functional screens in cells show that approximately 2% of lncRNA gene loci are required for cell growth or survival, the vast majority in a tissue or cell-type selective function[8], which is concordant with more limited examples in genetically engineered animal models.

Many lncRNAs have their functions in the nucleus, recognized archetypes of signal guides, scaffolds, and decoys[9]. First, lncRNAs are produced by the act of transcription, and thus lncRNA marks its chromosomal allele in cis as transcriptionally active. Classic examples of imprinting, where paternal vs. maternally inherited alleles are differentially expressed, are often associated and regulated in part by allele-specific transcription of lncRNA [10,11]. Second, lncRNAs can serve as guides to target chromatin modification machinery. Most if not all chromatin modifying enzymes lack intrinsic DNA binding specificity and rely on RNA or transcription factors to guide them to the target DNA loci. The mechanism of lncRNA targeting is diverse and may be limited to the local genome neighborhood

or chromosome from which it is transcribed (in cis) or broadened to loci throughout the genome (in trans) [5]. Reported target mechanisms include RNA: DNA: DNA triple formation[12,13], lncRNA RNA duplex interaction with nascent RNA [14], and exploitation of 3D spatial proximity[15,16]. Third, lncRNAs are comprised of multiple modular domains that create a scaffold for the assembly of multiple chromatin modification activities. The lncRNA Xist beautifully illustrates each of these archetypes[17]. In each female cell in mammals, one of two X chromosomes is transcriptionally silenced by action of Xist. Xist is only transcribed from the inactive X chromosome (Xi), and Xist RNA spreads across the Xi to nucleate and spread dozens of RNA binding proteins and chromatin regulators to silence gene expression through the lifetime of that cell and all its progenies. This classical example of lncRNA and many other recent entries highlight Nature's heretical invention and elaboration of using RNA to alter and dictate information flow from DNA.

RNA GUIDED DNA TARGETING

The discovery that prokaryotes possess a form of RNA-targeted immunity, termed clustered regularly interspaced short palindromic repeat (CRISPR), has dramatically expanded the view on the physiological role played by noncoding RNAs [18–20]. Complexed with CRISPR-associated (Cas) proteins, these noncoding RNAs, termed CRISPR RNA (crRNA), serves as guides to direct Cas proteins to specific DNA sequences for DNA cleavage [21,22]. This provides an unprecedented example for reversing the Central Dogma, -- how information encoded in RNA can alter information in DNA. In this review, we focus on how RNAs are leveraged for reversing the central dogma in eukaryotes.

Cas9 and Cas12—There are two distinct classes of CRISPR systems[23], and Class II enzymes have their targeting and nuclease functions encoded in a single protein (including Cas9) which span across a large group of RNA-guided nucleases that have evolved in various prokaryotic species. The most used single Cas effectors are Cas9 and Cas12a (formerly known as Cpf1) [24,25].

Evolution in different host environments have endowed Cas enzymes with unique traits, but they all share a common feature: formation of a ribonucleoprotein (RNP) for DNA or RNA targeting. For example, all Cas9 enzymes require complexation with two non-coding RNAs: one CRISPR-associated RNA (crRNA) encoding the target DNA complementarity and one trans-acting crRNA (tracrRNA) [26]. On the contrary, all Cas12a enzymes utilize a single crRNA without tracrRNA. This can be explained by the fact that Cas12a encodes RNase activity that processes the precursor crRNA array to individual crRNAs for RNP formation, while Cas9 requires tracrRNA and host factors (e.g., RNase III) for the same process [27]. Nevertheless, the evolution of tracrRNA is an interesting event, which may endow unique features, such as offering a versatile evolutionary solution to process long pre-crRNA arrays to individual crRNAs, while maintaining the compactness of the crRNA array. To simplify the use of two RNAs, crRNA and tracrRNA are fused as a single guide RNA (sgRNA).

All DNA-targeting Cas nucleases require a protospacer adjacent motif (PAM), short DNA segments flanking the targeted genomic locus that are not encoded in the corresponding region flanking the crRNA [28]. PAM sites provide an entry point for the Cas-guide RNA

complex for DNA binding and melting. PAMs from different Cas9 or Cas12 species have varied stringency. Examples include that, *Streptococcus pyogenes* Cas9 requires a NGG PAM, *Staphylococcus aureus* Cas9 recognizes an NNGRRT PAM (N is any nucleotide and R is A or G), and some Cas12a molecules use a TTTV PAM (V is A or C or G) [25,29,30]. Protein engineering of the PAMbinding domain of the Cas proteins can alter the PAM preference, which is an important factor for determining its targetable genomic space [31 32 33 34]. Structural and mechanistic insights of the sgRNA have also improved CRISPR-mediated gene editing. For example, varying the full-length of tracrRNA may improve the editing activity[35 36]. Truncated sgRNAs that guide Cas9 binding but not cleavage can shield off target sites and prevent off-target cleavage without affecting on-target sites[37]. Similarly, introduction of RNA hairpins to stabilize sgRNA can increase the specificity of cleavage of Cas9 and Cas12a[38].

Engineered guide RNA with aptamers for instructions—SgRNA can be universally engineered into a scaffold that recruits diverse enzymatic functions. In this case, an sgRNA sequence can be fused to a protein-interacting RNA aptamer that recruits specific RNA-binding proteins (RBPs), termed scaffold RNAs (scRNAs)[39,40]. RBPs fused with activators or repressors can then be used to bring additional instructions onto the scRNAs. When orthogonal aptamer–RBP pairs (e.g., MS2–MCP, PP7–PCP) are used to generate different scRNAs, distinct regulatory modules can be recruited to different DNA sites for distinct regulation. These scRNAs mimic the scaffolding natural function of lncRNAs that can encode both the target location and independent mode of regulatory action. Several systems for enhanced CRISPR-mediated gene regulation leverages this approach, including the broadly used SAM system that fuses MS2 aptamers to the sgRNA structure and recruits multiple activator domains to enhance target gene regulation[41]

Engineering guide RNA with additional RNA domains is a powerful approach to expand CRISPR functionality on the DNA. In addition to DNA regulation, the method of incorporating MS2 or PP7 RNA aptamers into the sgRNA has been utilized for multiple-color chromatin labeling, wherein RNA aptamers recruit MS2 or PP7 coat proteins fused with different fluorescent proteins to various target genomic loci [42]. In a method termed CRISP-Display, large RNA motifs including natural lncRNAs was directly inserted into the CRISPR sgRNA, allowing for targeting functional RNAs and associated ribonucleoprotein (RNP) complexes to desired genomic loci[43]. Young and colleagues tethered a short RNA (60-nt RNA derived from the promoter sequence of Arid1a) to the sgRNA using CRISPR-Cas9 in the vicinity of YY1 binding sites and determined the tethered RNA specifically increased the occupancy of YY1 at nearby enhancers[44]. Thus, engineering guide RNA by coupling with other RNA domains offers a scaffolding system that enables multimodal functionality on target genes using a single Cas protein.

Compact Cas molecules—Recently identified CRISPRs revealed new compact Cas effectors. Specifically, Cas12f (also called Cas14, 400–700 amino acids) and Cas12j (also called CasΦ, 700–800 amino acids) have been characterized from Archaea and bacteriophages[45,46] . In the case of Cas12f, two Cas12f molecules form a homodimer which binds to one crRNA-tracrRNA complex as a RNP [47]. This reduces the ‘genetic

cost' of expression and delivery, as expressing Cas12f and its associated crRNA-tracrRNA takes a smaller fragment size. While the wildtype system showed limited performance in mammalian cells, protein engineering was used to create better performing Cas variants in mammalian systems. One example is our development of a modified Cas12f system (termed CasMINI), which is 1.6 kb and less than half the size of Cas9 [48]. Notably, sgRNA engineering by altering the scaffold sequence played an important role in enhancing the activity of engineered Cas12f systems, which is likely due to better RNP formation and stability.

The compact size of miniature Cas proteins offers useful solutions for better delivery, using adeno-associated virus (AAV) or lipid nanoparticles (LNP), and promise better in vivo applications. Interestingly, large Cas effectors such as Cas9 or Cas12a (3~4kb) often utilizes smaller crRNA (or crRNA-tracrRNA)[29], and smaller Cas effectors such as Cas12f (1.2~1.8kb) uses larger crRNA complexes[49]. This makes sense structurally as smaller proteins require larger noncoding RNAs, serving as 'ligand', to form a sufficiently stable ribonucleoprotein for interaction with the stable double-stranded DNA.

Base editors—RNA-guided Cas nucleases induce double-strand break (DSBs) in DNA and DSB-mediated random indels (insertion/deletion), which is a 'primary' form of RNA->DNA information writing using CRISPR. Beyond this, Cas molecules can be fused to nucleotide modifiers to enable other forms of DNA editing. DNA base editing enzymes have been fused to the nuclease-dead dCas or the nickase variant of Cas that generates a single-strand break, to generate precise DNA mutations at a targeted location. Noncoding RNA transcripts guide activation-induced cytidine deaminase to mediate somatic hypermutation in B cells [50], providing a precedent for using RNA to guide base editing.

The cytidine deaminase enzyme APOBEC1, which converts cytosine (C) to uracil (U), was fused to nCas9. This base editor, termed cytosine base editors (CBEs), catalyzes a targeted C-to-U on the DNA strand not bound by sgRNA, which is read as a thymine (T) upon DNA replication and creates a precise C to T mutation [51–53]. Fusing uracil DNA glycosylase (UGI) and mutating APOBEC1, enhances the C-to-T reaction. In a similar manner, nCas9 fused with an engineered *Escherichia coli* TadA, termed adenosine base editors (ABEs), created an adenosine (A) to guanine (G) conversion [54–56]. ABE was also engineered into an efficient cytosine base editor [57]. Fusing both cytosine and adenosine deaminases to nCas9 generates both C-to-T and A-to-G conversions at a single site [58,59]. Base editors for nucleotide transversion, conversion between a purine and a pyrimidine, have been developed for C-to-G and C-to-A [60–62]. In addition to nCas9, the nuclease-dead Cas12a can efficiently modulate base editing [63]. Base editors are great examples for how RNAs are used as instructions to introduce specific mutations at DNA regions.

RNA AS GENOME TEMPLATE

Beyond RNA as a guide, RNA can directly template DNA sequence addition and connectivity. In yeast *Saccharomyces cerevisiae*, endogenous transcript RNA is used to mediate homologous recombination with chromosomal DNA [64 65]. In eukaryotic cells, telomerase RNA is the scaffold for the assembly of the multiprotein telomerase enzyme

complex that protects the ends of linear chromosomes (reviewed by [66]). Moreover, telomerase RNA provides the template for the addition of hundreds to thousands of a precise repeat sequence to chromosome ends by the reverse transcriptase activity of TERT. Ciliates, like the unicellular organism *Oxytrichia,* undergo a highly unusual life cycle that involves fragmenting its germline genome into numerous small pieces, eliminating ~95% of its genome, and then assembling the remaining fragments into ~16,000 genesized chromosomes. The assembly into chromosomes is not random but has been shown to be guided by RNA [67,68]—that is the order of genes on the RNA specifies the order of the DNA pieces that are stitched together [69]. Current model suggests base pairing between the RNA and regions of DNA sequences. However, because point mutations in the RNA are not copied into the assembled chromosomal DNA, the *Oxytrichia* lncRNA is likely not a template that is reverse transcribed into the DNA. Nonetheless, these examples illustrate the possibility of RNA to program long sequences into DNA by either direct or indirect means.

Prime editors—To further expand the mode of DNA writing, Liu and colleagues developed a method termed prime editing to insert longer stretches of DNA [70–72] Click or tap here to enter text.. In this method, nCas9 is fused to a reverse transcriptase and the sgRNA is engineered with an elongated 3’ sequence that encodes the desired mutations and a priming region complementary to the DNA. This altered sgRNA is a critical development, serving as a template encoding multiple instructions (both target location and desired mutations), which is collectively called prime editing guide RNA (pegRNA). In this RNA-directed system, the pegRNA primes the nicked DNA strand and the reverse transcriptase converts the template sequence from RNA to DNA, creating a targeted insertion or deletion in place. It was utilized to make all 12 base-to-base conversions, as well as small-fragment insertions and deletions (<100 bp). Combining with Bxb1 recombinase, the pegRNA system enables larger insertion or deletions (up to several kb) [74,75].

Notably, several improvements in prime editing have taken place via iterative improvement of the pegRNA design. Nelson et al. found that degradation of the 3’ template end of pegRNA appears to be a major bottleneck [73]. Such truncated pegRNAs cannot serve as template for RT, and moreover may block productive full length pegRNA-CasRT complexes from occupying target sites in a dominant negative manner. Enhancing greater stability of linear pegRNAs has improved the prime editing. Liu et al. further found that pegRNA can be split into a guide and a template RNA, and greatly stabilized by circularization, which makes the template RNA impervious to exonucleases [76]. Circular template RNA appears to enable greater duration of productive enzyme occupancy of the target site DNA, because dCas9 and RT can be expressed as separate proteins and still support prime editing. This split design solves the size problem of the Cas9-RT fusion construct, a current limitation for AAV delivery of genetic medicines in vivo [77].

RNA-guided DNA transposition—More recently, diverse Cas proteins were discovered to exhibit targeted genomic transposition or insertions. For example, A Type I-F and a Type II-K CRISPR system can insert DNA fragments to a genomic site via RNA-guided CRISPR-mediated recruitment of transposition machinery [78,79]. This ability for targeted transposition has been demonstrated *in vitro* and in prokaryotic hosts [78]. While targeted transposon

needs to be developed in human cells, these offer excellent examples on how CRISPR can use RNA instructions to enable diverse forms of DNAlevel information rewriting. Beyond natural Cas-transposase systems, Durrant et al. repurposed and fused bacterial serine integrase with dCas9 to enable programmable insertion with multikilobase DNA inserts [80] .

RNA-GUIDE GENE REGULATION AND EPIGENETIC EDITING

CRISPRa—What best illustrates the concept of reversed RNA->DNA information flow is probably the development of CRISPR activation (CRISPRa) [81,82]. In this method, a nuclease-dead dCas (e.g., dCas9, dCas12a, dCasMINI) is fused to a transcriptional activator [83–85]. Guided by its cognate sgRNA bearing a sequence indicating the target DNA locus, the RNP can specifically bind to an endogenous genomic locus to switch on the gene that is nearby. In mammalian cells, various transactivator domains such as VP64, tripartite VPR (VP64, p65, and RTA) are fused to dCas9 [86], tandem repeats of peptides [87], or indirect recruitment via engineered RNA scaffold, have been adopted for gene activation [88]. In bacterial cells, fusing synthetic transcriptional activators or the ω subunit of RNA polymerase to dCas9 enabled gene activation [89,90]. Unlike ectopic transgene expression, CRISPRa can be used to precisely tune the magnitude of gene upregulation. Furthermore, CRISPRa can be utilized to upregulate multiple gene targets by simply using multiple sgRNAs. In some cases, via a dCas12a-activator system, a compact crRNA array that can encode multiple gene-site instructions to activate many genes (7~10) can be achieved [91,92].

CRISPRi and CRISPRoff—CRISPR interference (CRISPRi) allows programmable gene silencing and is the conceptual opposite of CRISPRa. In CRISPRi, dCas9 is fused to a protein domain, such as KRAB, that recruits the endogenous SETDB1 histone modification enzyme and additional factors that places the repressive chromatin mark histone H3 lysine 9 trimethylation (H3K9me3) [93]. CRISPRi deposits H3K9me3 on ~1 kb window surrounding the target locus and can be targeted to promoters and distal enhancers in a multiplexed fashion to study gene or DNA element function [94,95]. One fruitful application of CRISPRi was the systematic functional interrogation of human lncRNA genes [96], and the factors required for maintenance of XIST-mediated gene silencing in adult somatic cells [97].

CRISPRoff is a new generation of programmable gene silencing with epigenetic memory. CRISPRoff builds on CRISPRi and contains dCas9, KRAB domain, and is further fused to DNMT3A and DNMT3L, components of the de novo DNA methyltransferase complex [98]. This combination of DNA element deactivation (KRAB) and maintenance (DNMT3A/3L) mimics the natural mechanism of epigenetic genetic silencing, such as X inactivation. CRISPRoff silences gene expression from the targeted promoters or enhancers, and further adds DNA methylation to the target locus. DNA methylation is copied and propagated by endogenous DNMT1 during cell division, thus ensuring memory of gene silencing over multiple somatic cell generations or even cell differentiation [99]. CRISPRoff allows a “hit-and-run” approach to epigenetic editing. Unlike CRISPRa or CRISPRi that needs to be continually expressed to exert their effects, transient expression of CRISPRoff construct and sgRNA allows durable gene silencing in cultured cells. The application of CRISPRoff in vivo and primary cells is an active area of investigation.

RNA guided epigenetic editing—Fusion of dCas with epigenetic domains with different enzymatic activities can be used to write or erase a broad spectrum of chromatin chemical modifications. To erase DNA methylation, TET1 was fused with dCas9 to remove DNA methyl groups and upregulate gene expression at specific chromatin sites [99,100]. In addition, to specifically add acetyl groups to local histones, the catalytic core of the acetyltransferase p300 domain was fused to dCas9 [101–103]. When directed to enhancers, this fusion adds an acetyl group to Lysine 27 of Histone H3 (H3K27ac), leading to activation of gene expression. Other modes of writing methylation at H3K4, H3K9, and H3K27 have been achieved by engineering the fusion effectors to dCas proteins [104]. All of these expand the modes that RNA->DNA epigenetic changes.

IN THE HORIZON

Here we have discussed several parallel themes in lncRNA biology and CRISPR technologies. What are potential areas of cross pollination and advances in the future? We offer several suggestions that are by no means exhaustive. Indeed, we look forward to being surprised by the richness of biology and the ingenuity of investigators in the field.

Gene activation with memory—While CRISPRoff has achieved gene silencing with epigenetic memory, no comparable technology exists for durable gene activation. CRISPRa requires continual expression of the dCas9-VP64 construct. The ability to durably induce gene expression after a transient expression of a synthetic regulator will offer many advantages for experimental biology and medicine. For instance, genetic haploinsufficiency can be treated by durable gene activation. LncRNA biology offers case studies for how durable gene activation is achieved in nature. Dosage compensation in Drosophila involves the transcriptional activation of the single male X chromosome by two-fold via two lncRNAs named roX1 and roX2 [38,105].The roX RNAs are also transcribed from the X chromosome and guide a protein complex that hyperacetylates histones and promotes transcriptional elongation on the X, creating a positive feedback loop. In the mammalian *HOXA* locus, the lncRNA HOTTIP is located on the 5’ end of the locus, associates with the MLL H3K4 methylase complex and promotes expression of the 5’ *HOXA* genes, including itself [106]. Both dosage compensation and positional identity are both remarkably stable once established and can even persist despite extensive ex vivo culture [107,108]. Accessing endogenous systems of positive feedback loops with cis-acting lncRNAs may unlock programmable gene activation with greater durability.

RNA control 3D genome and nuclear architecture—LncRNAs exert powerful roles in the compartmentalization of nuclear neighborhoods (reviewed by [109,110]) (Figure 2). The nucleolus (rRNA repeats), nuclear paraspeckle (Neat1), and the Barr body (Xist) are several prominent examples. In X inactivation, Xist not only spreads across the inactive X chromosome (Xi), Xist also compacts and pulls in the Xi and localizes the Xi to the nuclear periphery [111], creating a chromosomal structure and nuclear territory that are distinct from that of the homologous active X chromosome. Synthetic biologists have started to use RNA to program the 3D genome. CLOUD9 is a system that targets multiple dCas9 proteins fused to inducible dimerization domains to DNA elements spaced kilobases apart [112]. Upon sgRNA expression and dimerization, distal DNA elements such as enhancers and promoters

can be pulled into spatial proximity, leading to gene activation [112]. CRISPR-GO is a method to direct chromosomal loci position within the nucleus. Via a dimerization system that repositions the target DNA to desired nuclear compartment, CRISPR-GO mediates RNA-guided spatial relocalization of specific chromatin to a different location including nuclear periphery, Cajal bodies, or heterochromatin condensates [113–115]. The apparent difference is the scale and complexity of the 3D genome architecture being programmed by endogenous lncRNAs. One potential mechanistic distinction is that endogenous lncRNAs have much more elaborate scaffold function, and each unit of the scaffold is repeated multiple times. Xist is an 81 kb lncRNA in the mouse and consists of multiple repeats, named A-F, which have a division of labor in spreading across the chromosome, gene deactivation, and epigenetic memory [116,117].

RNA induced condensates—Nuclear RNAs are powerful nucleators and participants in multivalent RNA-protein interactions. Studies of condensates to create membraneless compartments have relied extensively on in vitro biochemistry of purified component proteins and imaging in cells. Tools to program condensate formation and dissolution are starting to be developed[118]. LncRNAs offer some potentially useful lessons. The lncRNA NORAD contains 18 copies of binding sites for the RNA binding protein Pumilio and titrates Pumilio activity in the cell. NORAD can compete Pumilio away from the multitude of Pum binding sites in the mRNA transcriptome, even though the latter collectively exists in ~40-fold excess [119]. Mendell and colleagues showed that the repeated nature of Pum binding sites in NORAD creates multivalency that facilitates Pumilio condensate formation in a super-stochiometric manner [119]. A synthetic circular RNA composed of Pum binding sites phenocopied NORAD function. These emerging studies suggest the possibility that elaborating and engineering scaffolds on sgRNAs targeting DNA or RNA may be a fruitful path to new technologies to dissect and manipulate condensates and complex multivalent molecular interactions.

The 21th century has been envisioned as the century of biology. Rewiring the the central dogma with RNA may play major roles in the next phase of exciting biological advances.

ACKNOWLEDGEMENT

We thank Dr. M.C. Tsai for assistance and comments. L.S.Q. acknowledges support from National Science Foundation CAREER award (Award #2046650), National Institutes of Health (Grant # 1R01CA266470), Stanford Innovative Medicines Accelerator. L.S.Q. is a Chan Zuckerberg Biohub investigator. H.Y.C. is supported by NIH RM1-HG007735, R35-CA209919 and is an Investigator of the Howard Hughes Medical Institute.

References

1. Sabin LR, Delás MJ & Hannon GJ Dogma derailed: the many influences of RNA on the genome. Mol Cell 49, 783–794 (2013). [PubMed: 23473599]
2. Millman A. et al. Bacterial Retrons Function In Anti-Phage Defense. Cell 183, 1551–1561.e12 (2020). [PubMed: 33157039]
3. Hsu PD, Lander ES & Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278 (2014). [PubMed: 24906146]
4. Herman AB, Tsitsipatis D. & Gorospe M. Integrated lncRNA function upon genomic and epigenomic regulation. Mol Cell 82, 2252–2266 (2022). [PubMed: 35714586]

5. Rinn JL, Chang HY & Chang HY Long Noncoding RNAs: Molecular Modalities to Organismal Functions. Annu Rev Biochem 89, 283–308 (2020). [PubMed: 32569523]
6. Allou L. et al. Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. Nature 592, 93–98 (2021). [PubMed: 33568816]
7. de Goede OM et al. Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. Cell 184, 2633–2648.e19 (2021). [PubMed: 33864768]
8. Liu SJ et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science 355, (2017).
9. Wang KC & Chang HY Molecular mechanisms of long noncoding RNAs. Mol Cell 43, 904–914 (2011). [PubMed: 21925379]
10. Sleutels F, Zwart R. & Barlow DP The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415, 810–813 (2002). [PubMed: 11845212]
11. Latos PA et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338, 1469–1472 (2012). [PubMed: 23239737]
12. Zhang X. et al. . KCNQ1OT1 promotes genome-wide transposon repression by guiding RNA-DNA triplexes and HP1 binding. Nat Cell Biol (2022) doi:10.1038/S41556-022-01008-5.
13. Postepska-Igielska A. et al. LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Mol Cell 60, 626–636 (2015). [PubMed: 26590717]
14. Lee N, Moss WN, Yario TA & Steitz JA EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. Cell 160, 607–618 (2015). [PubMed: 25662012]
15. Wang KC et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–126 (2011). [PubMed: 21423168]
16. Engreitz JM et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341, (2013).
17. Loda A, Collombet S. & Heard E. Gene regulation in time and space during X-chromosome inactivation. Nat Rev Mol Cell Biol 23, 231–249 (2022). [PubMed: 35013589]
18. Barrangou R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712 (2007). [PubMed: 17379808]
19. Mojica FJM, Díez-Villaseñor C, Soria E. & Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol 36, 244–246 (2000). [PubMed: 10760181]
20. Ishino Y, Shinagawa H, Makino K, Amemura M. & Nakatura A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol 169, 5429–5433 (1987). [PubMed: 3316184]
21. Jansen R, van Embden JDA, Gaastra W. & Schouls LM Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43, 1565–1575 (2002). [PubMed: 11952905]
22. Haft DH, Selengut J, Mongodin EF & Nelson KE A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol 1, 0474–0483 (2005).
23. Makarova KS et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18, 67–83 (2020). [PubMed: 31857715]
24. Zetsche B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015). [PubMed: 26422227]
25. Jinek M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012). [PubMed: 22745249]
26. Deltcheva E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607 (2011). [PubMed: 21455174]
27. Fonfara I, Richter H, BratoviÄ M, le Rhun A. & Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature 532, 517–521 (2016). [PubMed: 27096362]
28. Marraffini LA & Sontheimer EJ Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463, 568–571 (2010). [PubMed: 20072129]

29. Zetsche B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015). [PubMed: 26422227]
30. Ran FA et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186–191 (2015). [PubMed: 25830891]
31. Hu JH et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57–63 (2018). [PubMed: 29512652]
32. Gao L. et al. Engineered Cpf1 variants with altered PAM specificities. Nat Biotechnol 35, 789–792 (2017). [PubMed: 28581492]
33. Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481–485 (2015). [PubMed: 26098369]
34. Nishimasu H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259–1262 (2018). [PubMed: 30166441]
35. Bikard D. et al. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nat Biotechnol 32, 1146–1150 (2014). [PubMed: 25282355]
36. Chen B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155, 1479–1491 (2013). [PubMed: 24360272]
37. Rose JC et al. Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. Nat Commun 11, (2020).
38. Kocak DD et al. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. Nat Biotechnol 37, 657–666 (2019). [PubMed: 30988504]
39. Dong C, Fontana J, Patel A, Carothers JM & Zalatan JG Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. Nat Commun 9, (2018).
40. Zalatan JG et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160, 339–350 (2015). [PubMed: 25533786]
41. Konermann S. et al. Genome-scale transcriptional activation by an engineered CRISPRCas9 complex. Nature 517, 583–588 (2015). [PubMed: 25494202]
42. Wang S, Su JH, Zhang F. & Zhuang X. An RNA-aptamer-based two-color CRISPR labeling system. Sci Rep 6, (2016).
43. Shechner DM, Hacisuleyman E, Younger ST & Rinn JL Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat Methods 12, 664–670 (2015). [PubMed: 26030444]
44. Sigova AA et al. Transcription factor trapping by RNA in gene regulatory elements. Science 350, 978–991 (2015). [PubMed: 26516199]
45. Pausch P. et al. CRISPR-CasΦ from huge phages is a hypercompact genome editor. Science 369, 333–337 (2020). [PubMed: 32675376]
46. Harrington LB et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science 362, 839–842 (2018). [PubMed: 30337455]
47. Takeda SN et al. Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol Cell 81, 558–570.e3 (2021). [PubMed: 33333018]
48. Xu X. et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol Cell 81, 4333–4345.e4 (2021). [PubMed: 34480847]
49. Harrington LB et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science 362, 839–842 (2018). [PubMed: 30337455]
50. Pefanis E. et al. Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. Nature 514, 389–393 (2014). [PubMed: 25119026]
51. Arbab M. et al. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. Cell 182, 463–480.e30 (2020). [PubMed: 32533916]
52. Kim YB et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol 35, 371–376 (2017). [PubMed: 28191901]
53. Komor AC, Kim YB, Packer MS, Zuris JA & Liu DR Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424 (2016). [PubMed: 27096365]
54. Gaudelli NM et al. Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017). [PubMed: 29160308]

55. Richter MF et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat Biotechnol 38, 883–891 (2020). [PubMed: 32433547]
56. Gaudelli NM et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nat Biotechnol 38, 892–900 (2020). [PubMed: 32284586]
57. Neugebauer ME et al. Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. Nat Biotechnol (2022) doi:10.1038/S41587-022-01533-6.
58. Zhang X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. Nat Biotechnol 38, 856–860 (2020). [PubMed: 32483363]
59. Grünewald J. et al. A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. Nat Biotechnol 38, 861–864 (2020). [PubMed: 32483364]
60. Zhao D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. Nat Biotechnol 39, 35–40 (2021). [PubMed: 32690970]
61. Kurt IC et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat Biotechnol 39, 41–46 (2021). [PubMed: 32690971]
62. Koblan LW et al. Efficient C·G-to-G·C base editors developed using CRISPRi screens, target-library analysis, and machine learning. Nat Biotechnol 39, 1414–1425 (2021). [PubMed: 34183861]
63. Kempton HR, Love KS, Guo LY & Qi LS Scalable biological signal recording in mammalian cells using Cas12a base editors. Nat Chem Biol 18, 742–750 (2022). [PubMed: 35637351]
64. Keskin H. et al. Transcript-RNA-templated DNA recombination and repair. Nature 515, 436–439 (2014). [PubMed: 25186730]
65. Mazina OM, Keskin H, Hanamshet K, Storici F. & Mazin A. v. Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. Mol Cell 67, 19–29.e3 (2017). [PubMed: 28602639]
66. Roake CM & Artandi SE Regulation of human telomerase in homeostasis and disease. Nat Rev Mol Cell Biol 21, 384–397 (2020). [PubMed: 32242127]
67. Lindblad KA, Bracht JR, Williams AE & Landweber LF Thousands of RNA-cached copies of whole chromosomes are present in the ciliate Oxytricha during development. RNA 23, 1200–1208 (2017). [PubMed: 28450531]
68. Nowacki M. et al. RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature 451, 153–158 (2008). [PubMed: 18046331]
69. Bracht JR et al. Chromosome fusions triggered by noncoding RNA. RNA Biol 14, 620–631 (2017). [PubMed: 27267579]
70. Anzalone A. v. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149–157 (2019). [PubMed: 31634902]
71. Chen PJ et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell 184, 5635–5652.e29 (2021). [PubMed: 34653350]
72. Nelson JW et al. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol 40, 402–410 (2022). [PubMed: 34608327]
73. Nelson JW et al. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol 40, 402–410 (2022). [PubMed: 34608327]
74. Anzalone A. v. et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat Biotechnol 40, 731–740 (2022). [PubMed: 34887556]
75. Yarnall MTN et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nat Biotechnol (2022) doi:10.1038/S41587-022-01527-4.
76. Liu B. et al. A split prime editor with untethered reverse transcriptase and circular RNA template. Nat Biotechnol 40, 1388–1393 (2022). [PubMed: 35379962]
77. Liu B. et al. A split prime editor with untethered reverse transcriptase and circular RNA template. Nat Biotechnol 40, 1388–1393 (2022). [PubMed: 35379962]
78. Strecker J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48–53 (2019). [PubMed: 31171706]

79. Klompe SE, Vo PLH, Halpin-Healy TS & Sternberg SH Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature 571, 219–225 (2019). [PubMed: 31189177]
80. Durrant MG et al. Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. Nat Biotechnol (2022) doi:10.1038/S41587-022-01494-W.
81. Guo LY et al. Multiplexed genome regulation in vivo with hyper-efficient Cas12a. Nat Cell Biol 24, 590–600 (2022). [PubMed: 35414015]
82. Gilbert LA et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442 (2013). [PubMed: 23849981]
83. Xu X. et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol Cell 81, 4333–4345.e4 (2021). [PubMed: 34480847]
84. Kempton HR, Goudy LE, Love KS & Qi LS Multiple Input Sensing and Signal Integration Using a Split Cas12a System. Mol Cell 78, 184–191.e3 (2020). [PubMed: 32027839]
85. Qi LS et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183 (2013). [PubMed: 23452860]
86. Chavez A. et al. Comparison of Cas9 activators in multiple species. Nat Methods 13, 563–567 (2016). [PubMed: 27214048]
87. Liu Y. et al. CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. Cell Stem Cell 23, 758–771.e8 (2018). [PubMed: 30318302]
88. Konermann S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588 (2015). [PubMed: 25494202]
89. Bikard D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41, 7429–7437 (2013). [PubMed: 23761437]
90. Dong C, Fontana J, Patel A, Carothers JM & Zalatan JG Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. Nat Commun 9, (2018).
91. Campa CC, Weisbach NR, Santinha AJ, Incarnato D. & Platt RJ Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nat Methods 16, 887–893 (2019). [PubMed: 31406383]
92. Magnusson JP, Rios AR, Wu L. & Qi LS Enhanced Cas12a multi-gene regulation using a CRISPR array separator. Elife 10, (2021).
93. Gilbert LA et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442 (2013). [PubMed: 23849981]
94. Thakore PI et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods 12, 1143–1149 (2015). [PubMed: 26501517]
95. Lin X. et al. Nested epistasis enhancer networks for robust genome regulation. Science 377, 1077–1085 (2022). [PubMed: 35951677]
96. Liu SJ et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science **355**, (2017).**355**
97. Yu B. et al. B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells. Cell 184, 1790–1803.e17 (2021). [PubMed: 33735607]
98. Amabile A. et al. Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell 167, 219–232.e14 (2016). [PubMed: 27662090]
99. Nuñez JK et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell 184, 2503–2519.e17 (2021). [PubMed: 33838111]
100. Liu XS et al. Editing DNA Methylation in the Mammalian Genome. Cell 167, 233–247.e17 (2016). [PubMed: 27662091]
101. Dominguez AA et al. CRISPR-Mediated Synergistic Epigenetic and Transcriptional Control. CRISPR J 5, 264–275 (2022). [PubMed: 35271371]
102. Hilton IB et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33, 510–517 (2015). [PubMed: 25849900]
103. Gemberling MP et al. Transgenic mice for in vivo epigenome editing with CRISPR-based systems. Nat Methods 18, 965–974 (2021). [PubMed: 34341582]
104. Nakamura M, Gao Y, Dominguez AA & Qi LS CRISPR technologies for precise epigenome editing. Nat Cell Biol 23, 11–22 (2021). [PubMed: 33420494]

105. Samata M. & Akhtar A. Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. Annu Rev Biochem 87, 323–350 (2018). [PubMed: 29668306]
106. Wang KC et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–126 (2011). [PubMed: 21423168]
107. Rinn JL et al. A dermal HOX transcriptional program regulates site-specific epidermal fate. Genes Dev 22, 303–307 (2008). [PubMed: 18245445]
108. Quinn JJ et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat Biotechnol 32, 933–940 (2014). [PubMed: 24997788]
109. Quinodoz SA & Guttman M. Essential Roles for RNA in Shaping Nuclear Organization. Cold Spring Harb Perspect Biol 14, (2022).
110. Batista PJ & Chang HY Long noncoding RNAs: cellular address codes in development and disease. Cell 152, 1298–1307 (2013). [PubMed: 23498938]
111. Chaumeil J, le Baccon P, Wutz A. & Heard E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev 20, 2223–2237 (2006). [PubMed: 16912274]
112. Morgan SL et al. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat Commun **8**, (2017).**8**
113. Wang H, Han M. & Qi LS Engineering 3D genome organization. Nat Rev Genet 22, 343–360 (2021). [PubMed: 33558716]
114. Gao Y, Han M, Shang S, Wang H. & Qi LS Interrogation of the dynamic properties of higher-order heterochromatin using CRISPR-dCas9. Mol Cell 81, 4287–4299.e5 (2021). [PubMed: 34428454]
115. Wang H. et al. CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. Cell 175, 1405–1417.e14 (2018). [PubMed: 30318144]
116. Lu Z. et al. Structural modularity of the XIST ribonucleoprotein complex. Nat Commun 11, (2020).
117. Chu C. et al. Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416 (2015). [PubMed: 25843628]
118. Bracha D, Walls MT & Brangwynne CP Probing and engineering liquid-phase organelles. Nat Biotechnol 37, 1435–1445 (2019). [PubMed: 31792412]
119. Elguindy MM & Mendell JT NORAD-induced Pumilio phase separation is required for genome stability. Nature 595, 303–308 (2021). [PubMed: 34108682]

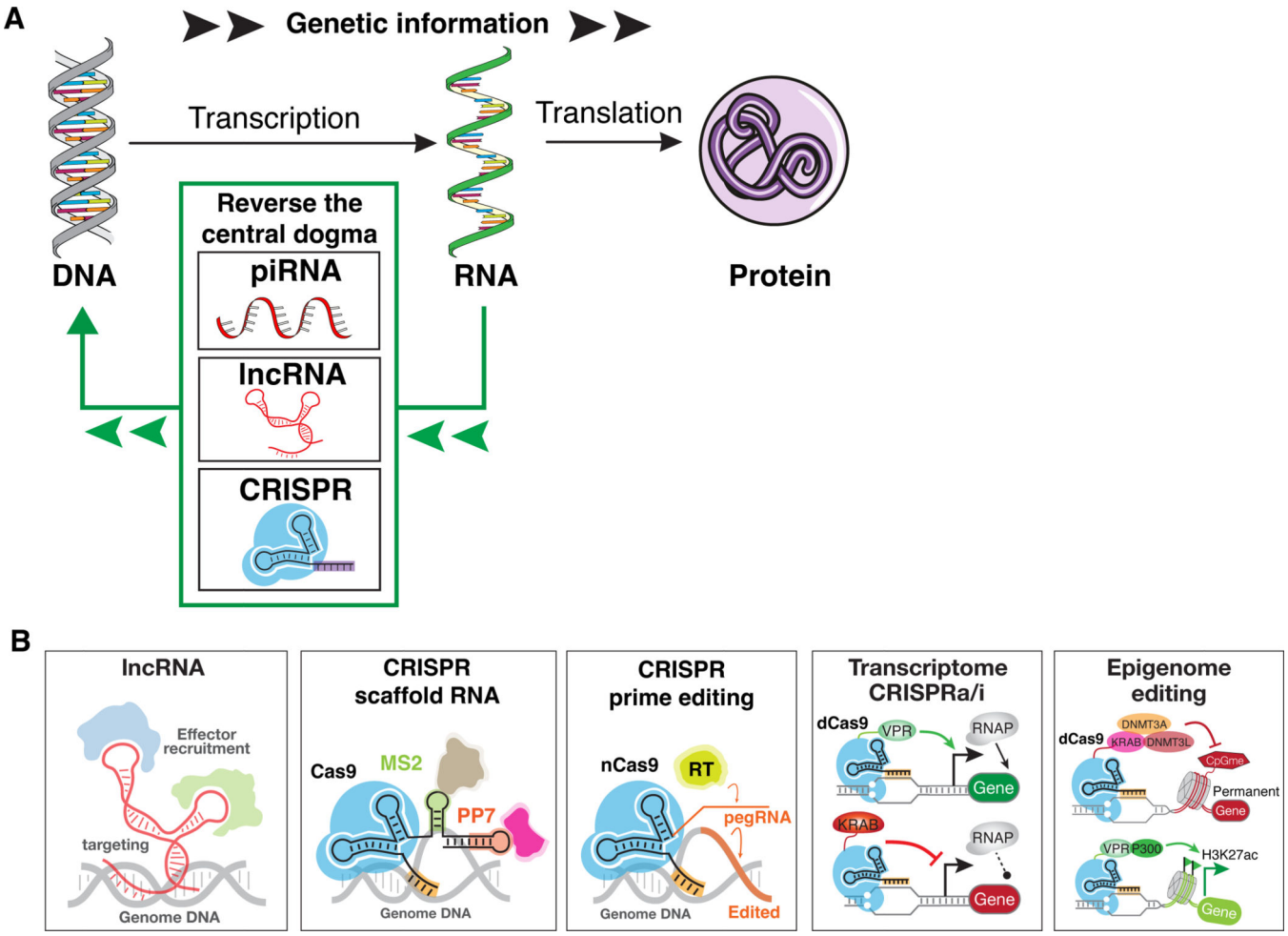

Figure 1.

Diagram showing scheme for reversing the Central Dogma. (**A**) Genetic information flows from DNA to RNA to protein. Natural and synthetic systems allow reversing this flow direction by modifying DNA based on information encoded in RNA, with examples including piRNA, lncRNA, and CRISPR. (**B**) Examples for lncRNA and CRISPR for targeting genome regions and interacting with host factors or synthetic regulators to modulate gene expression, chromatin status, etc.

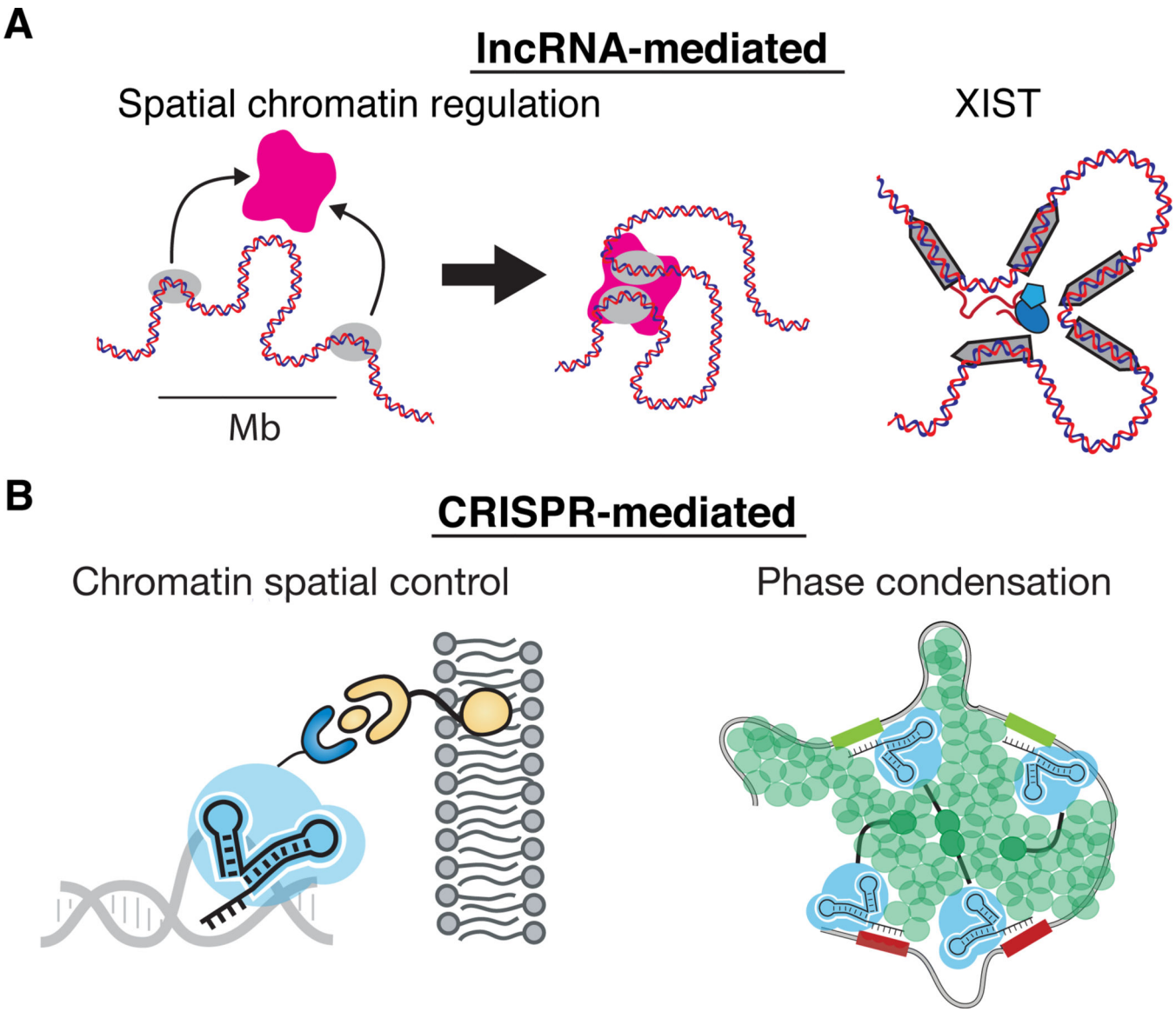

Figure 2.

Future direction for using RNA to modulate DNA 3D configuration and phase condensation for using lncRNA (**A**) or CRISPR (**B**).