From Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

DEVELOPING MOLECULAR TOOLS FOR PROBING AND MODULATING GENOMIC SPATIAL ADJACENCY

Yunshi Yang

Stockholm 2023

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-018-5
Cover illustration: Cover illustration made with the assistance of DALLE2

Developing molecular tools for probing and modulating genomic spatial adjacency

Thesis for Doctoral Degree (Ph.D.)

By

Yunshi Yang

The thesis will be defended in public at Samuelssonsalen, Tomtebodavägen 6, 17165 Solna, Stockholm
On Friday May 12th 2023, at 13:00

Principal Supervisor:
Björn Högberg
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Biomaterials

Co-supervisor(s):
Ana Teixeira
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Biomaterials

Opponent:
Friedrich Simmel
Technical University of Munich
Department of Physics
Division of Experimental Physics

Examination Board:
Katja Petzold
Karolinska Institutet
Department of Medical
Biochemistry and Biophysics
Division of Molecular structural biology

Camilla Björkegren
Karolinska Institutet
Department of Cell and Molecular Biology
Division of Cell Biology

Johan Elf
Uppsala University
Department of Cell and Molecular Biology
Division of Molecular Systems Biology

...hurry up and play, or soon it's time to go...

Abstract

In addition to the vast information encoded in DNA sequence, the genome has physical features that are also essential for its function, including its organization in three-dimensional space. The development of high-throughput technology has greatly advanced our understanding of the spatial organization of the genome but has also raised more questions.

In this thesis, we developed molecular tools to address the remaining challenges regarding the interplay between genomic organization and function. By breaking down the subject from the global architecture of the genome into an ensemble of spatially adjacent chromatin segments, we came up with different methods covering various aspects.

We demonstrated in Paper I that global spatial information can be transferred in the format of DNA sequence encoding pairwise spatial proximity between two distinct molecular objects. We have shown that by growing network from pairwise relationship encoded in DNA sequence, spatial features at a global scale can be recovered. The results from this work highlighted the potential of using pairwise adjacency as a fundamental unit for recording the spatial organization of complex molecular systems. The high programmability and versatility of nucleic acids make them an ideal medium for encoding this information.

With the aim of studying the pairwise relationship between genomic DNA in cells, we devised a CRISPR-dCas9 system for different purposes by leveraging its high programmability for genome targeting. In Paper III, we have shown that the re-designed guide RNA can direct dCas9 to a pair of genomic loci, inducing DNA contacts. This system can be applied as a modulation tool to introduce pairwise contacts for decoding functional implications in cells. In Paper IV, we developed a method for the direct detection of pairwise interactions between genomic loci at the single-cell level *in situ*. This method is achieved by conjugating oligonucleotide tags to Cas9 and using the tags for probing the spatial adjacency between a pair of genomic loci targeted by Cas9

Meanwhile, we developed an efficient method to fabricate and purify DNA origami with modifications in Paper II. This method makes the production of functionalized nanostructures more time and material-efficient compared to established techniques. The ease of production allows broader applications of functionalized nanostructures, including characterizing the effect of nanoscale distance on biochemical assays, as shown in Paper IV.

List of scientific papers

I. Hoffecker I. T, Yang Y., Bernardinelli G., Orponen P. & Högberg B, A computational framework for DNA sequencing microscopy. Proceedings of the National Academy of Sciences 116, 19282–19287 (2019).

II. Smyrlaki I., Shaw A., Yang Y., Shen B. & Högberg B., Solid Phase Synthesis of DNA Nanostructures in Heavy Liquid. Small 19, 2204513 (2023).

III. Yang, Y., Shen, B., Rocamonde L. I., Berzina I., Zipf J.& Högberg B., Re-engineered guide RNA enables DNA loop and contacts formation in vivo. *Manuscript in revision*

IV. Yang, Y., Rocamonde L. I., & Högberg, B. In situ detection of genomic cis-interaction by proximity ligation assay using oligonucleotide-labelled Cas9

Contents

List of abbreviations

DNA	Deoxyribonucleic acid
RNA	Riboucleic acid
bp	basepair
nt	nucleotide
PAM	Protospacer adjacency motif
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
sgRNA	Single guide RNA
KRAB	Krüppel associated box
TAD	Topologically associating domain
EDTA	Ethylenediaminetetraacetic acid
DTT	Dithiothreitol
PBS	Phosphate buffered saline
dNTPs	Deoxynucleoside triphosphate
PLA	Proximity ligation assay
Polony	Polyerase colony
ROI	Region of interest
ssDNA	Single-stranded DNA
kDa	Kilo-dalton

1 Introduction

1.1 Nucleic Acids as biomolecules

1.1.1. Chemical composition of Nucleic acids

Nucleic acids are a class of biomolecules that play a fundamental role in living organisms for the storage, transmission, and expression of genetic information. They are composed of long chains of nucleotides, which are monomers consisting of a nitrogenous base, a five-carbon sugar, and a phosphate group. The nitrogenous bases can be either purines (adenine and guanine) or pyrimidines (cytosine, thymine, and uracil).

There are two major types of nucleic acids that are found in living organisms including deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Due to their distinct chemical compositions, DNA and RNA exhibit significant structural and functional differences.

DNA is composed of four types of nucleotides, which differ in their nitrogenous base. These bases are adenine (A), thymine (T), guanine (G), and cytosine (C). The sugar component of DNA nucleotides is a deoxyribose sugar, which is a five-carbon sugar molecule similar to ribose but lacks an oxygen atom at the 2' position. The phosphate group is attached to the 5' carbon of the sugar, forming a phosphodiester bond between adjacent nucleotides with hydroxyl (OH) group attached to the 3' carbon.

RNA is also composed of four types of nucleotides, but instead of thymine (T) found in DNA the nitrogenous base uracil (U) is used. The sugar component of RNA nucleotides is a ribose sugar, which is also a five-carbon sugar molecule but has an oxygen atom at the 2' position. RNA monomer is also joined by phosphodiester bond with each other to form long chains of nucleotides. The presence of 2′ hydroxyl group on RNA sugar backbone makes it more reactive compared to DNA, as it can participate in intramolecular or intermolecular hydrogen bonding with other functional groups, including other 2'-OH groups or the phosphate backbone of RNA.

The nitrogenous bases in both DNA and RNA can be classified into purines and pyrimidines. Purines include adenine and guanine, which have a double-ring structure, while pyrimidines include cytosine, thymine, and uracil, which have a single-ring structure. Adenine (A) forms two hydrogen bonds with thymine (T) in DNA or with uracil (U) in RNA, while guanine (G) forms three hydrogen bonds with cytosine (C) in both DNA and RNA. The hydrogen bonding between nitrogenous is fundamental for the structure and function of DNA and RNA.

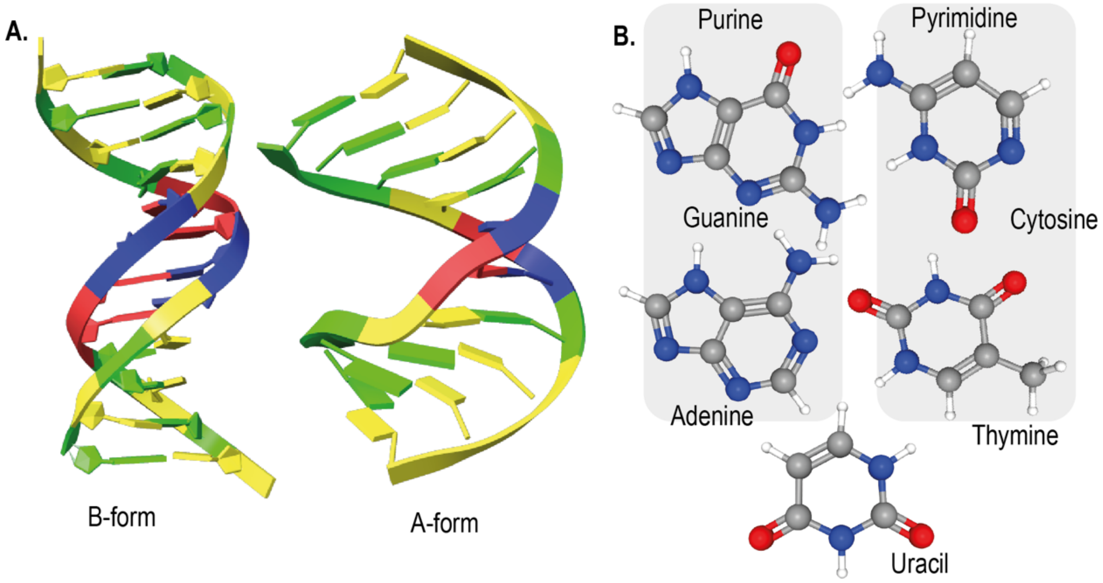

Figure1 . (A) Different forms of DNA double-helix. (B) Five nucleobases for DNA and RNA.

The chemical composition of DNA makes the dominant structure of it in nature to be double-stranded helix, with each strand going to opposite direction to the other. RNA, is generally a single-stranded molecule, although it can form complex secondary and tertiary structures due to base pairing between complementary nucleotides. Compared to DNA, RNA exhibits more flexibility in structure as it can fold into various secondary structures including hairpin loops, bulges, and junctions.

1.1.2. Nucleic acids as information media

The structural differences between RNA and DNA give rise to differences in their functions within cells. DNA is primarily responsible for storing genetic information, while RNA plays a crucial role in transferring that information from the DNA to the protein synthesis machinery of the cell.

The double-stranded helix of DNA provides a mechanism for accurate replication of genetic information. Because the two strands are complementary, each strand can serve as a template for the synthesis of a new strand during DNA replication[1]. This process allows for precise copying of genetic information from one generation to the next. Furthermore, the double-stranded helix provides protection for the genetic information stored within it for long-term storage in cells. The structure of DNA is relatively stable and thus resistant to damage from chemical and physical factors, which helps to protect the genetic information from being degraded or lost.

In contrast, RNA is typically single-stranded and lacks the stability and protection provided by the double-stranded helix structure of DNA. This single-stranded structure makes RNA more flexible and allows it to adopt a wide variety of conformations, which is important for its diverse functions in gene expression and regulation. However, the single-stranded structure also makes RNA more vulnerable to degradation and less stable than DNA. The lower stability of RNA limits its ability to store genetic information over long periods of time but allows it to exert intermediary function to transfer genetic information to ribosome and regulatory role in gene expression with its more rapid turnover rates *in vivo*.

In summary, the double-stranded helix structure of DNA provides key properties for storing genetic information, including accurate replication and chemical and physical stability. RNA, on the other hand, typically adopts single-stranded structure, making it suitable for diverse function in the process of genetic information transfer.

1.2 Nucleic acids as building material

1.2.1 structural details of double-helix

The DNA double helix consists of two complementary strands of nucleotides, which are arranged in an antiparallel fashion and held together by a variety of chemical and physical interactions.

The backbone of each DNA strand is composed of alternating sugar and phosphate groups, which form a repeating pattern of phosphodiester bonds. The sugar in DNA is deoxyribose, and each phosphate group is covalently bonded to both the 3' and 5' carbons of adjacent deoxyribose sugars, giving the DNA strand a directionality from the 5' end to the 3' end.

The B-form helix is the most common conformation of DNA in nature and is characterized by a right-handed double helix structure. In the B-form helix, the base pairs are nearly perpendicular to the helix axis, resulting in a pitch of 3.4 nm per turn and a diameter of about 2 nm. The B-form helix is stabilized by hydrogen bonds between complementary base pairs, hydrophobic interactions between the bases, and electrostatic interactions between the negatively charged phosphate backbone and positively charged cations in solution. The A-form helix is a less common conformation of DNA that is characterized by a shorter and wider helix structure, with a pitch of 2.6 nm per turn and a diameter of about 2.3 nm. In the A-form helix, the base pairs are tilted relative to the helix axis and are shifted towards the major groove, resulting in a more compact structure. The A-form helix is thought to be favored in the presence of high salt concentrations and is stabilized by a combination of hydrogen bonds and base-stacking interactions[2].

Another set of key features of DNA double-helix are the major groove and minor groove formed between two strands of helix. Major groove refers to the feature formed where backbones of two strands are distant to each other, while minor groove formed where the backbones are closer to each other. The major groove and minor groove are functionally important for protein binding or for intercalating small molecules. For instance, sequence recognition of transcription activator-like effector is mediated by hydrogen bond formation between amino acid side chains and specific nucleic acid bases located within major groove of double-helix. Sequence interrogation of Cas9 protein also involves the interaction between the major groove of guanine bases from PAM sequence motif and amino acid side chains.

RNA can form double-stranded secondary structure by complementary base pairing between nucleotides in a single RNA strand. The most common type of RNA double helix structure is the hairpin loop, which is formed when a region of RNA containing complementary base pairs folds back on itself to form a stem and loop structure. The stem of the hairpin loop is composed of base-paired nucleotides, while the loop region is typically unpaired. The base pairing in RNA double helices can involve both canonical Watson-Crick base pairs, such as A-U and G-C, as well as non-canonical base pairs, such as G-U and A-C. The stem of the stem-loop structure is typically composed of around 5-10 base pairs, with a pitch of around 3-4 Å per base pair and a diameter of around 20-30 Å. However, the exact dimensions of RNA double helices can vary depending on factors such as the length and sequence of the stem, the presence of non-canonical base pairs, and the overall structure of the RNA molecule.

1.2.2 Nucleic acids as building material

The specific and predictable base-pairing interactions of DNA double-helix gives it desirable properties such as programmability and self-assembly, to create tailored structures. In the early 1980s, researchers began exploring the use of DNA as a material for creating artificial structures and devices. The first demonstration of this concept came in 1982, when Nadrian Seeman, showed that short DNA sequences could be designed to assemble into a variety of complex shapes and structures, including crosses, triangles, and cubes[3]. Seeman's vision of using DNA as a programmable construction material inspired researchers to explore new ways of harnessing the properties of DNA for creating nanoscale material.

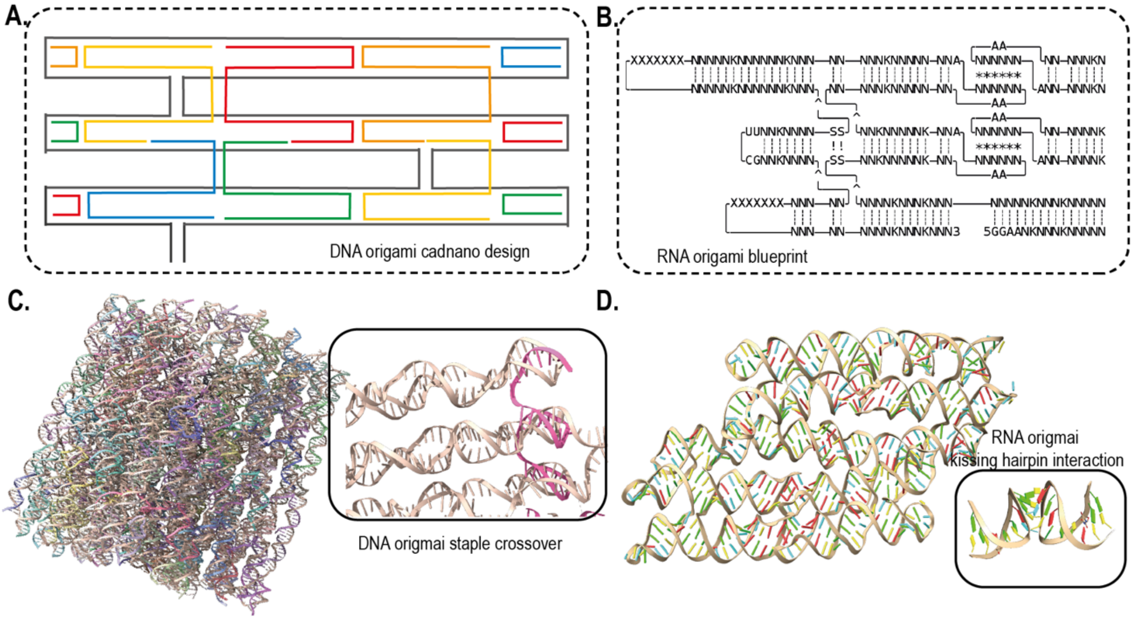
A.
DNA origami cadnano design
B.
XXXXXXX–NNNNNKNNNNNNKNNN—NN—NNNKNNNNN–NNA
NNNNNKNNNNNNKNNN NN—NNNKNNNNN–NN
AA
NNNNNN
******
NNNNNN
AA
NN–NNNKN
ANN–NNNKN
UUNNKNNNN—SS NNKNNNNNK–NNA
CGNNKNNNN SS—NNKNNNNNK–NN
NN–NNNNK
ANN–NNNNK
XXXXXXX–NNN—NN NNNNKNNNKNNN—NNNNNKNNNKNNNNN
NNN—NN—NNNNKNNNKNNN3 5GGAANKNNNKNNNNN
RNA origami blueprint
C.
DNA origmai staple crossover
D.
RNA origmai
kissing hairpin interaction

Figure2 Basics of DNA and RNA origami (A)Routing schemes of DNA origami using Cadnano, scaffold strand in grey, staple strands in color. (B) Blueprint of RNA origami routing using ROAD package, with defined structural motifs.(C) Reconstructed model of DNA origami from Cryo-EM. Inset shows DNA routing with staple crossovers in pink. PDB: 6by7. (D) Reconstructed model of RNA origami from Cryo-EM. Inset shows RNA kissing hairpin. PDB:7ptq.

In 2006, Paul Rothemund introduced the technique of DNA origami, which revolutionized the field of DNA nanotechnology by allowing researchers to fold long DNA strands into arbitrary shapes with unprecedented precision[4]. DNA origami technique is based on the principle of using a long, single-stranded DNA molecule called a scaffold strand, which is folded into a specific shape using short, complementary "staple strands". By carefully assigning the sequence of staple strands to be complementary to specific regions of scaffold strand, the scaffold can be folded into designed structure. The usage of a long single-stranded DNA as the scaffold for DNA self-assembly has greatly improved its yield and reproducibility.

Subsequently, a more modular approach was developed to design DNA origami that extends the principles of DNA origami design to create three-dimensional nanostructures[5]. This design strategy constrains the arrangement of double-helices formed between the scaffold and staples into a honeycomb lattice, thereby confining the relative angle between helices. This design principle enables a modular design strategy to construct arbitrary shapes of DNA origami by reinforcing crossovers of the scaffold and staples only at permitted locations to arrange helices at the correct angles for the honeycomb layout. Based on this strategy, a computer-aided design tool was developed to automate the labor-intensive steps of assigning crossover points on the scaffold and generating complementary staple strands involved in DNA origami design[6]. CaDNAno has greatly facilitated the design and fabrication of DNA origami structures, making the technique more accessible to researchers in a wide range of fields. In addition to the lattice-based DNA origami design, wireframe DNA origami was developed later on, providing complementary properties to DNA origami for broader applications[7].

In parallel to the development of DNA nanostructure, RNA is also explored as an alternative for building nanoscale structures. Similar to DNA nanostructures, RNA nanostructure design strategies also rely largely on the natural base-pairing interaction properties to create two-dimensional or three-dimensional structures. Secondary structures of RNA formed by base-pairing, such as stem-loops, bulges, and internal loops, are used to create more complex tertiary structures. Modular design strategies are developed to create larger structures either from separate RNA tiles or from a single-strand of RNA[8].

In general, by exploiting the natural properties of nucleic acids, strategies to create nanostructures from nucleic acids provide the option to create biocompatible tools with precise control over structure and arrangement of molecules at nanoscale.

1.2.3 Applications of DNA/RNA nanostructures in biomedical research

The ability to design and synthesize complex, three-dimensional DNA and RNA nanostructures with precise control over size, shape has opened up avenues for biomedical research.

DNA origami has emerged as a promising tool for studying nanoclustering in biological systems. The precise control over the size, shape of DNA origami nanostructures has made them suitable for mimicking nanoscale patterns found in cellular environment. For example, DNA origami nanostructures have been used to study the clustering of cell surface receptors, which is a critical process in cell signaling[9]. DNA origami can be functionalized with ligands and control their spatial organization on the nanostructure. This has allowed the investigation of the effect of receptor clustering on signaling pathways and cellular behavior. Additionally, DNA origami has been used to study the process of molecular interactions, particularly the multivalent interactions, including antibody-antigen binding dynamics. By controlling the spatial arrangement of antigens on the surface of DNA origami nanostructures, researchers have been able to study their cooperative activity, specificity[10]. Overall, DNA origami has demonstrated its ability to provide new insights into the nanoscale organization of biological systems and the role of distance and clustering in molecular processes.

Another prominent application of DNA origami is for precise drug delivery. As DNA origami can be assembled into three-dimensional shapes with tunable sizes and surface chemistries, it can be tailored to improve biodistribution and cellular uptake of drugs cargo of the DNA origam[11]. The targeted drug delivery system of DNA origami is made possible either by passive accumulation in tumor tissue given the nanoscale of DNA origami , or by functionalizing DNA origami to sensor the microenvironment[12].In addition to cellular delivery, DNA origami has also been explored as CRISPR-Cas9 gene therapy delivery platform by adjusting the dimension of the origami folded from DNA insertion template below the size of nuclear pore[13].

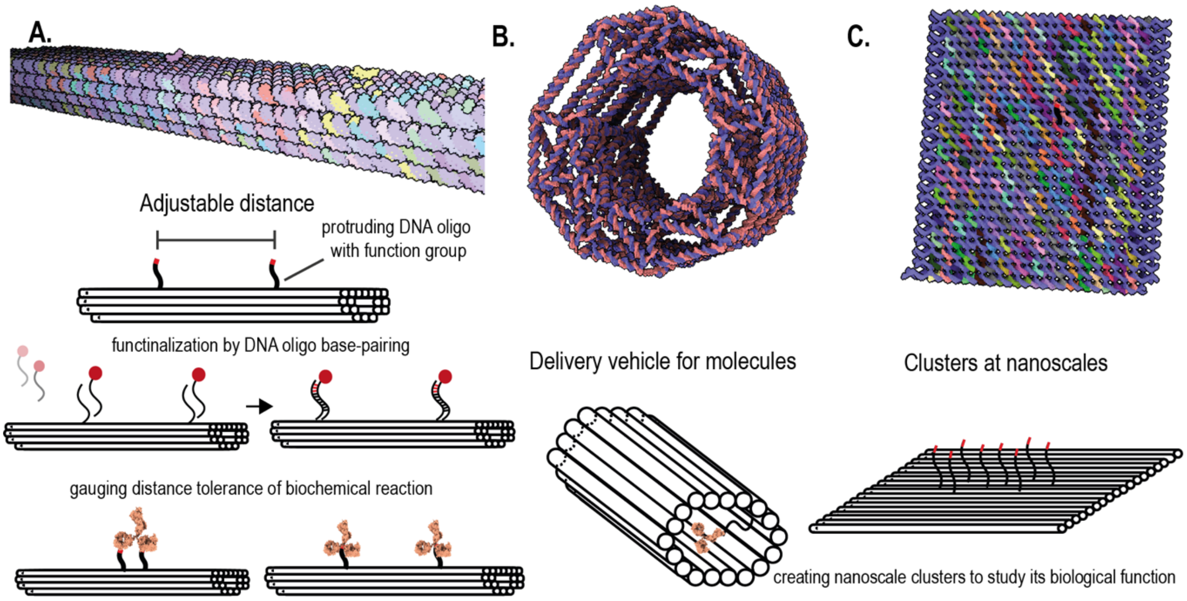
A.
B.
C.
Adjustable distance
protruding DNA oligo with function group
functinalization by DNA oligo base-pairing
gauging distance tolerance of biochemical reaction
Delivery vehicle for molecules
Clusters at nanoscales
creating nanoscale clusters to study its biological function

Figure3. DNA origami as molecular tools for biomedical study. (A) An example of using DNA origami to study distance tolerance of antibody by adjusting the spacing between antigens. (B) Example of using DNA origami barrel for encapsulating functional molecules for *in vivo* delivery. (C) Example of using DNA origami to arrange molecules in patterned clusters with high accuracy to study its functional implication.

Despite being a relatively new field compared to DNA nanostructures, RNA nanostructures exhibit significant potential for diverse applications due to the versatile nature of RNA and its various functions. As RNA nanostructure can be folded from single-strand RNA, it can be folded co-transcriptionally into designed motifs[8], providing alternative delivery route. This prominent feature makes the RNA nanostructure an ideal platform for next-generation therapy, such as gene therapy, gene editing and gene silencing. Such applications have been prototyped in yeast to control the metabolism by using RNA origami to template dCas9-based gene activator[14]. Moreover, the versatility of RNA molecules in natural systems offers opportunities for applications that combine the structural programmability of RNA with its natural functions. One such application is the targeted delivery of small interfering RNA (siRNA) folded into RNA nanostructures, which allows for precise and efficient gene silencing for therapeutic purposes[15].

In general, the significance of DNA and RNA nanostructures lies in their ability to enable precise control over the arrangement and thus the behavior of biological molecules at the nanoscale. The ease of design and fabricate DNA nanostructures with high performance of structural rigidity makes it not only a highly programmable nanomedicine delivery system, but a unique platform to study biophysical properties of interactions between crucial biomolecules at nanoscale. RNA nanostructure, on the hand, with the advantage of integrating the functional moieties into the structural design thanks to the broad catalogues of existing RNA functional motifs, is promising for various in vivo applications.

1.3 CRISPR system

1.3.1 discovery of CRISPR system

CRISPR system is an extraordinary example of natural genome editing system that relies on the base-pairing interaction between nucleic acids to achieve sequence-specific protein function including endonuclease activity. The discovery of the CRISPR system is a significant breakthrough in molecular biology and genetics.

The CRISPR system was initially identified as a mysterious genomic region in the genomes of archaea and bacteria[16]. Bioinformatic studies later revealed that this conserved genomic locus is associated with bacterial immune capacity, as shown by the similarity between the phage genome and the spacer region between conserved repeats[17]. CRISPR-associated proteins, such as Cas proteins, were identified through bioinformatic efforts as essential for maintaining adaptive immunity against invading genomes[18]. The first experimental evidence of CRISPR's function was demonstrated by insertion of a spacer against foreign phage into S. thermophilus CRISPR loci, which provided protection against corresponding phage[19]. The cooperation between processed crRNA transcribed from CRISPR loci and Cas protein was shown through the experiment of the reconstituted expression of CRISPR loci and Cas proteins in *E.Coli* against phage was only preserved by co-expression of both elements[20]. It was also shown by this work that, Cas9, a multidomain Cas protein, can function independently for CRISPR interference.

1.3.2 Molecular mechanisms of CRISPR system

The mechanistical elucidation of Cas9 as a single protein sufficient to perform genome interference guided by RNA has revolutionized the genetic engineering field. The Cas9 protein can form active complex with two RNA, crRNA and tracrRNA, to recognize target DNA sequence encoded in an unpaired region of crRNA[21]. When crRNA forms stable duplex with target strand through base-pairing, the Cas9 undergoes conformational change to become active as endonuclease, making cut on both DNA strands.

A crucial finding in the application of CRISPR-Cas is the protospacer adjacent motif (PAM), a short sequence adjacent to the target sequence that is complementary to the guide RNA. The PAM sequence varies depending on the Cas protein being used, but it is a necessary component across various Cas proteins. In the case of Cas9, the recognition of the PAM sequence occurs prior to the target sequence interrogation step. Recognition of the PAM sequence occurs through hydrogen bonding between amino acids and the major groove of the DNA, a mechanism commonly used by other DNA binding proteins. In the PAM-interacting domain, Arginine residues play a key role in conferring the guanine preference of spCas9, and this feature is conserved among type II-a Cas9 proteins[22–24]. Interesingly, crystal structure analysis of Cas9:RNA:DNA complex has revealed that the minor groove interaction between Cas9 and PAM facilitates the rotation of the phosphate group of the first PAM-proximal base from the target strand. This reorientation may contribute to the destabilization of the DNA duplex and initiation of the hybridization of gRNA-DNA. The remaining unpaired non-complementary strand is hypothesized to be stabilized by the positively charged groove formed between the PAM interacting domain, HNH, and RuvC. Neutralizing the positive charge by replacing positive amino acids has reduced the tolerance to mismatches between DNA and gRNA, indicating the importance of charge balance in maintaining high specificity[25]. Subsequent studies have found that reduced off-target effects are not directly caused by lower affinity for imperfect targets, but rather by trapping the HNH domain in an inactive conformation[26]. Additionally, the speed-limiting step of DNA cleavage has been attributed to conformational changes in the NUC domain. This finding indicates that off-target editing can be delineated from off-target Cas9 binding to DNA, as there is an additional gating step limiting its endonuclease activity after the binding of Cas9 to its compromised targets.

The mechanism of mismatch tolerance of Cas9 has been further studied due to its potential limitations in gene editing. Initially, it was discovered that Cas9 could cleave mismatches at the PAM distal region up to 15 base pairs with reduced efficiency, depending on the specific guide sequence[21]. However, sequencing has revealed heterogeneity in mismatch tolerance [27,28]. Large target libraries spanning sequence space have also revealed the effects of different types of perturbations, such as singly mismatches versus sequence insertions on off-target reactions. The diverse perturbation space made it possible to predict the energetic penalty per perturbation or additive effects involving multiple perturbations. Interestingly, although contiguous mismatches did not follow additive penalties, distant mismatches displayed an approximate additive pattern.

In addition to sequencing-based techniques, cryo-electron microscopy (cryo-EM) has contributed to resolving the structural basis underlying mismatches. A recent study exploited the advantage of cryo-EM to resolve the Cas9 ribonucleoprotein in complex with target DNA containing mismatches at various positions and capture the intermediate conformation at different time courses of the cleavage reaction [29]. The intermediate structure contains a rotation of helix, which agrees with earlier structure of R-loop formation [22]. The study further correlated the insensitivity of Cas9 to mismatches at positions 12-14 to structural details about the missing interaction of REC3 with bases at these sites.

1.3.3 Engineering CRISPR-Cas9 system

The crRNA and tracrRNA interact with each other to form secondary structure that is essential for recruitment of Cas9. The crRNA contains the DNA target information, while tracrRNA can pair with crRNA as structural scaffold to form RNA duplex. This RNA thus performs two essential functions in this system, structural scaffolding for protein, and genetic information encoding. To simplify the CRISPR-Cas9 system for broader application across organisms, these two RNAs were joined together by a hairpin loop into a single guide RNA (sgRNA). The 20-nucleotide unpaired region of the CRISPR RNA is designed to match with the DNA target of interest in such a way that the complementary non-target strand is situated upstream, next to the PAM sequence, to direct Cas9 to perform its function at the correct locations. The 20-nucleotide region can also be designed to include deliberate mismatches at positions 16-20, which can suppress the endonuclease activity of Cas9 while still maintaining the ability to recognize and bind to the DNA target, as mentioned earlier. This approach has been used to co-deliver gene editing donor templates, enabling Cas9 to bind to DNA templates without cutting them[30].

Cas9 protein has also been engineered for various applications by modifying its structure and function through protein engineering techniques such as directed evolution[31], rational design and structural optimization[32]. Directed evolution involves generating a large library of Cas9 variants with random mutations and selecting those that exhibit desirable properties, such as improved specificity, activity, or stability. This technique has been used to create Cas9 variants with altered PAM specificity, reduced off-target effects, and increased efficiency. Rational design involves using knowledge of the Cas9 protein structure and function to design modifications to specific regions of the protein that interact with DNA and RNA. This approach has also been used to engineer Cas9 variants with altered PAM recognition, and improved specificity.

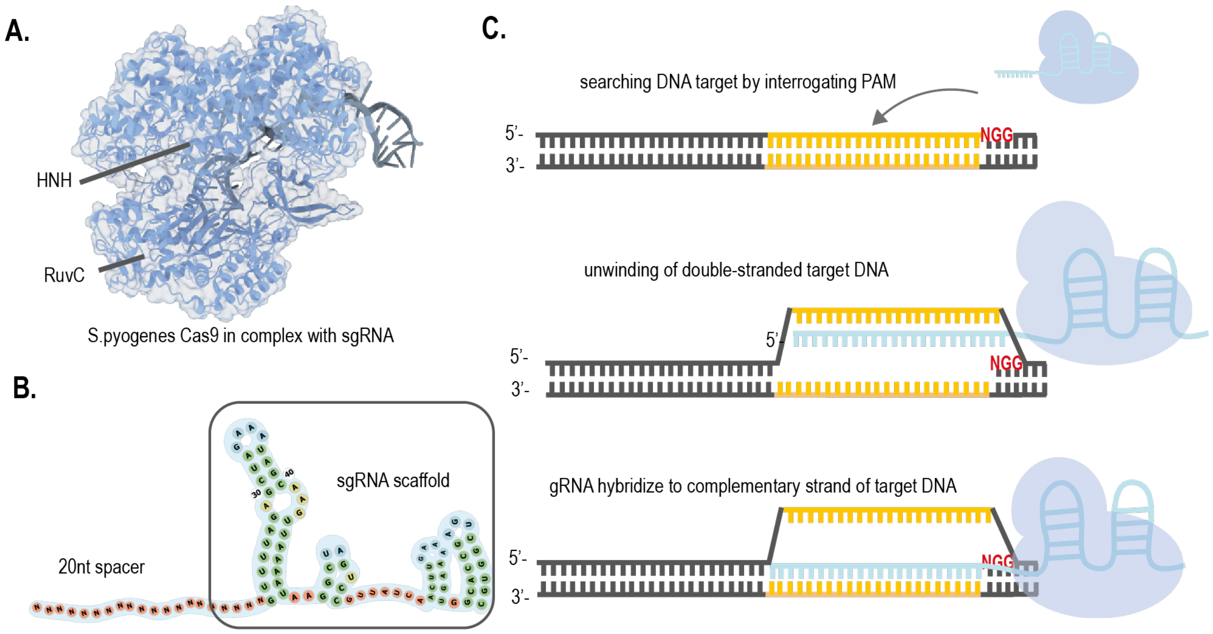
A.
HNH
RuvC
S.pyogenes Cas9 in complex with sgRNA
B.
sgRNA scaffold
20nt spacer
C.
searching DNA target by interrogating PAM
5'-
3'-
NGG
unwinding of double-stranded target DNA
5'
5'-
3'-
NGG
gRNA hybridize to complementary strand of target DNA
5'-
3'-
NGG

Figure 4. Functional mechanism of CRISPR-Cas9 (A) Reconstructed model of Cas9 in complex with sgRNA from Crystallography. HNH domain, responsible for cutting target strand; RuvC domain, responsible for cutting non-target strand. PDB:4zt9 (B) secondary structure of sgRNA. (C) DNA target interrogation mechanisms of Cas9.

Cas9 has a bilobular structure, that are named REC and NUC lobe. NUC lobe consists of two nuclease domains, RuvC and HNH domain.RuvC domain is homologous to RNase H, catalysing the cleavage of non-complementary strand of DNA, while HNH domain cleaves complementary strand from gRNA-DNA duplex. By introducing point mutations, D10A and H814A, to the RuvC and HNH domains, respectively, the nuclease activity of each Cas9 domain can be inactivated. This lead to three modified proteins, with DNA nickase that contains only one of the mutations, and protein containing both mutations referred to as nuclease inactivated Cas9 (dCas9), that is still able to bind to target sites with high stability but lost its cleavage activity.

The PAM interacting domain has garnered significant interest in protein engineering research. Bioinformatic analysis of the protospacer region has revealed that the PAM sequence is crucial for target recognition, and PAM interaction has been linked to the unwinding of the DNA duplex[24]. Modifying key residues in the PAM interacting domain could guide reengineering of PAM specificity and target specificity by influencing the incidence of amino acid and nucleic acid base interactions in major groove recognition[29,32]. For instance, the structure of the Cas9:DNA:RNA complex with mismatches at position 18-20 reveals the role of the tyrosine side chains of RuVC in stabilizing the unpaired nucleic acids. Substitution of corresponding amino acid led to design Cas9 variants with 500-fold slower cleavage rates of mismatches at these positions, while still maintaining efficient cleavage of perfect targets[29].

1.3.4 CRISPR system as tools beyond gene editing

1.3.4.1 CRISPR-dCas9 for targeting genomic loci

In recent years, the CRISPR system has been repurposed for gene regulation through the use of a deactivated Cas9 protein (dCas9) guided to a specific DNA sequence by a single guide RNA (sgRNA). Rather than cleaving the DNA, dCas9 binds to the target sequence and blocks transcription by serving as roadblock for RNA polymerase. This novel application of CRISPR, known as CRISPR interference (CRISPRi), enables precise and targeted gene regulation[33].

The mechanism of CRISPRi is based on steric hindrance, whereby dCas9 physically obstructs RNA polymerase from binding to the DNA or from proceeding transcription upon encountering the dCas9. Previous studies have examined the effect of target strand and distance from the transcription starting site (TSS) on the efficiency of repression. These investigations have revealed that RNA polymerase's helicase activity can remove dCas9-sgRNA RNP hybridizing to the template strand, leading to suboptimal interference. In contrast, when the RNP binds to the non-template strand, the polymerase is blocked, resulting in a more robust repression. The distance from TSS is also a crucial factor that affects repression efficiency, as an inverse correlation has been observed between the distance from TSS and gene expression. This relationship may be explained by the enhanced repression effect resulting from blocking RNAP recognition of the TSS site at its proximity.

Another key attribute of the dCas9 system is its reversibility. Studies conducted in *E.Coli* have shown that removal of the inducer for dCas9 expression leads to recovery of the expression of the targeted protein. However, it has been unclear whether the actual turnover rate of dCas9-RNP affects this recovery or if it is the outcome of cell population expansion. Further investigations using single-molecule tracing within E. coli bacteria have demonstrated that the lifetime of dCas9 at targeted loci correlates well with the generation time of *E.Coli*, suggesting that dCas9 binding to the target locus is irreversible until cell replication. Therefore, controlling the function of dCas9 system requires consideration of the replication rate of targeted cells. It is still unknown whether other mechanisms affect the stability of dCas9-RNP at targeted sites, and further research is needed to elucidate this phenomenon.

Overall, the CRISPR-dCas9 system's high programmability, reversibility, and sequence specificity make it a promising platform for gene regulation with a wide range of potential applications.

1.3.4.2 CRISPR-dCas9 gene regulation toolkits

The development and characterization of CRISPR-dCas9 system for searching and binding to targeted sequence with high programmability has showcased the potential of it as a platform for modulate gene expression. Based on the original dCas9 system, functional domains known to be effective in regulating gene expression were fused to the dCas9 protein to be act in a sequence-specific manner directed by gRNA. In this category are dCas9-based sequence-specific transcription activators and transcription repressors. Though dCas9 alone is capable of repressing targeted genes, fusion of chromatin modifier domains such as KRAB protein to dCas9 allows it to exert higher level of repression at targeted loci by recruiting endogenous chromatin modifier to cause histone methylation and deacetylation[34]. Corresponding to this, fusing histone acetyltransferase domains to dCas9 generates an epigenome editing tool to activate the targeted loci[35].

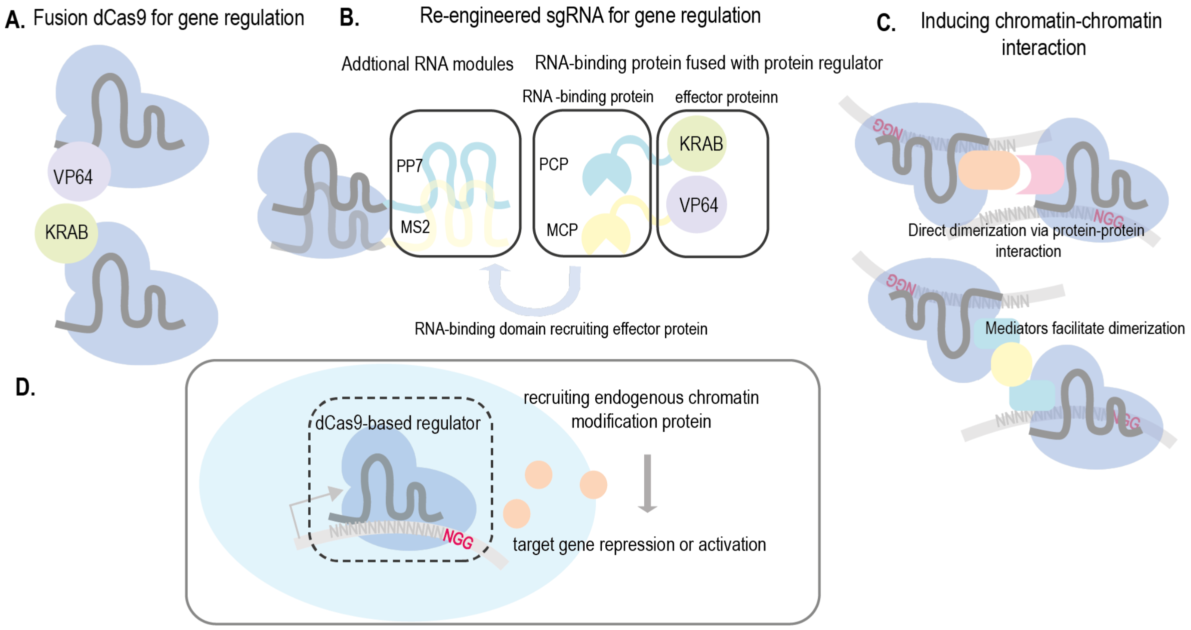
A. Fusion dCas9 for gene regulation
VP64
KRAB
B. Re-engineered sgRNA for gene regulation
Addtional RNA modules
RNA-binding protein fused with protein regulator
RNA -binding protein
effector proteinn
PP7
MS2
PCP
MCP
KRAB
VP64
RNA-binding domain recruiting effector protein
C. Inducing chromatin-chromatin interaction
NGGNNNNNNNNNNNN
NNNNNNNNNNNNNGG
Direct dimerization via protein-protein interaction
Mediators facilitate dimerization
D.
dCas9-based regulator
NNNNNNNNNNNNNGG
recruiting endogenous chromatin modification protein
target gene repression or activation

Figure 5. CRISPR-dCas9 as tools for gene regulation. (A) Fusing dCas9 with gene regulatory protein domain to create sequence-specific regulatory protein. (B) Adding functional RNA motifs to the sgRNA allows specific recognition by corresponding RNA-binding protein. By fusing regulatory protein to these RNA-binding proteins, dCas9-sgRNA can recruit multiple copies of these fusion effector proteins. (C) dCas9 fused with dimerizable, can induce chromatin-chromatin interaction once binding to target genomic loci. Another strategy involves free proteins that facilitate dCas9-dCas9 interaction upon induction.

In addition to strategies modifying chromatin to alter the transcription states of targeted loci, other dCas9-transcriptional activators involve either direct fusion of transcription activation domain to dCas9 or adding multivalent tags to dCas9 or RNA to recruit co-delivered activation domains.[36] dCas9 fused with VP64 and VPR tripartite domains are the commonly used transcriptional activator, as they are known to be able to recruit transcription machinery to the dCas9-targeted loci. Despite the advantage of conciseness for relying on only fusion protein and gRNA, the activation efficiency was not optimal. It was later come up that by recruiting VP64 as a separate effector using either aptamer at sgRNA or poly-peptide tags at the C-terminus of dCas9 the system could perform with much higher activation efficiency[36,37].

A relatively smaller area of research involving the use of dCas9 for controlling gene expression is focused on the potential of restructuring chromatin organization. These techniques involve fusing dCas9 to domains that can induce dimerization either directly or through a molecular mediator. By bringing two dCas9 molecules in close proximity, targeted genomic regions are brought together. One method uses blue light to induce dimerization of dCas9 through light-responsive protein domains attached to dCas9, along with free protein mediator[38]. Another approach employs two classes of dCas9 molecules, each fused with engineered dimerizable domains, which independently bind to target loci and induce DNA loop formation via protein-protein interaction[39]. These techniques have expanded the available methods for interpreting the function of chromatin interaction in various contexts.

In general, dCas9-based transcriptional regulatory tools makes it possible to study arbitrary gene function and cellular processes by simply altering the 20nt gRNA sequence. By controlling the expression of specific genes, the effects of modulating gene of interest on cellular behavior and disease pathways could be investigated. Such tools also enable to identify potential drug targets and develop new therapies by massively parallel screening assay.

1.3.4.3 dCas9 repurposed for genome imaging

In parallel to the application of dCas9 for gene regulation is its application in live cell imaging to visualize targeted genomic loci[40]. This approach involves fusing dCas9 to a fluorescent protein and using it to target specific DNA sequences for imaging. The dCas9-fluorescent protein fusion binds to the target DNA sequence, allowing it to be visualized in live cells under a fluorescence microscope. By designing sgRNAs to target different DNA sequences, it can image specific genomic loci in live cells with high spatial and temporal resolution. Its ability to visualize specific genomic loci in living cells has provided the opportunity to understand of the dynamic nature of the genome and its role in cellular processes.

However, one limitation of live cell imaging with dCas9 fusion protein is the low signal-to-noise ratio, as each dCas9 was originally fused to one fluorescent proteins. Strategies to address this challenge include designing a series of sgRNA targeting a consecutive section of chromatin to introduce multiple fluorophores per diffraction-limited area. The labeling of a genomic region makes it possible to trace the dynamics of targeted chromatin segment.

To be able to visualize a singly targeted non-repetitive locus, a signal-amplification system achieved by protein-tagging system SunTag was introduced to dCas9 by adding multi-copies of tag peptide to dCas9, that can recruit multiple fluorescent proteins co-transfected into the cell[37]. Similar to this, fusing aptamer to gRNA also enables recruitment of multiple fluorescent protein to each dCas9. Additionally, adding recognition sites to 3′ of sgRNA to recruit fluorescent protein fused to RNA recognition domain can also enhance the intensity per dCas9[41].

The dCas9 imaging system has also been redesigned for fluorescence in situ hybridization technique for imaging genomic loci on fixated samples[42]. The advantage of using dCas9 for the purpose of FISH, derives from its ability to bind to double-stranded DNA specifically, without the requirement for DNA denaturation. This feature makes this protocol much more time-efficient and preserves the native architecture of genome. Furthermore, the sequence-specificity of dCas9 system allows it to distinguish single-nucleotide polymorphisms (SNP) in tissue samples, making it an ideal platform for clinical applications. However, the challenge of increasing signal-to-noise ratio persists when targeting non-repetitive genomic locus, thus making signal-amplification essential for these applications. Unlike live cell imaging that relies on co-transfection with mediating effectors for increasing signal intensity, dCas9-based FISH system can be potentially combined with orthogonal biochemical approaches to achieve higher intensity and specificity[43]. For instance, a work aiming to detect SNPs at mitochondria DNA adopted proximity-ligation-assay as a method to amplify the fluorescence signal, which have been widely applied in protein-protein interaction studies[44]. The detection of genomic DNA is still challenging due to the scarcity of non-repetitive locus, requiring detection techniques to be ultra-sensitive and highly specific for the target locus.

1.4 Genome organization

1.4.1 Genome organization features

While DNA can be folded into predesigned structures at near-atomic resolution in test tubes to fabricate DNA origami nanostructure, the resolution of chromatin as genomic material is conventionally determined by the scale of base-pairs. The structure of chromosomes is highly dynamic and accompanied by complicated activities, including replication, transcription, and remodeling of chromosome associated with these procedures. Additionally, organization of chromatin could be further described by radial position and proximity distance. Radial position depicts the location of chromatin relative to the radial center of the nucleus, while proximity distance quantifies the relative distance between genomic segments.

The organization of genomic DNA can be described as a hierarchical structure consisting of multiple levels of compaction. At the smallest level, DNA is wrapped around octameric histone proteins at the scale of nanometers to form nucleosomes, which are the basic units of chromatin. Nucleosomes are further compacted into higher-order structures such as chromatin fibers and loops. These structures are stabilized by various architectural proteins, including insulators and topological associating domain (TAD) boundary proteins and transcription factors, which help to establish and maintain the spatial organization of the genome. The resulting genome architecture plays a critical role in regulating gene expression and other genomic processes. Understanding the principles that govern the organization of genomic DNA is therefore essential for comprehending the functional properties of the genome.

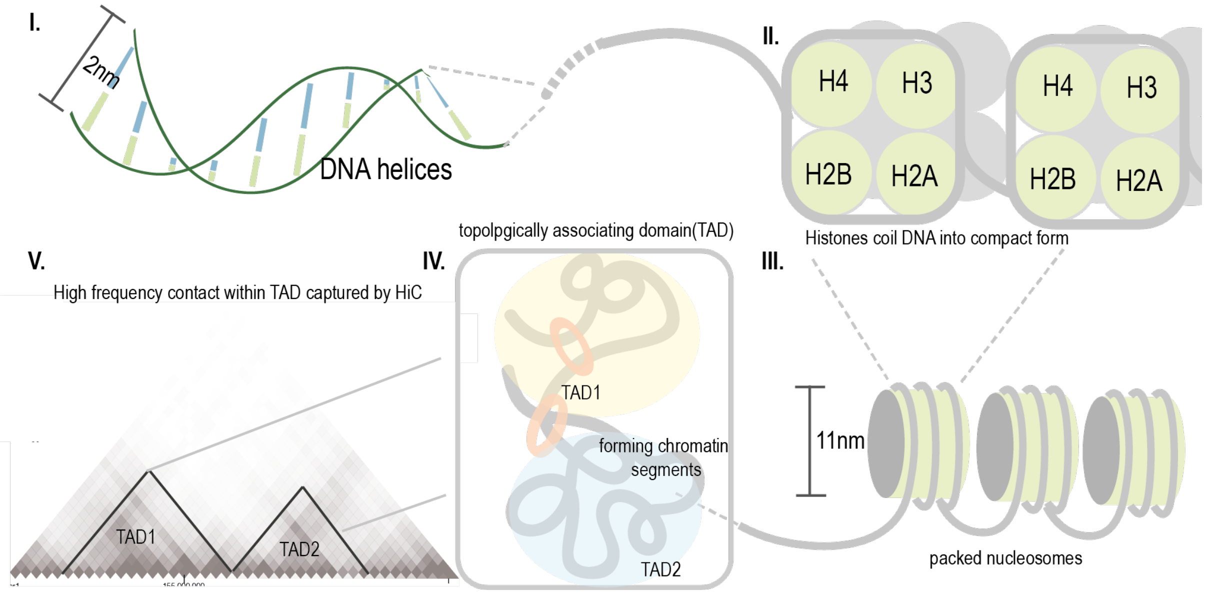
I.
2nm
DNA helices
II.
H4
H3
H2B
H2A
H4
H3
H2B
H2A
Histones coil DNA into compact form
topolpgically associating domain(TAD)
V.
IV.
III.
High frequency contact within TAD captured by HiC
TAD1
forming chromatin segments
TAD2
11nm
TAD1
TAD2
packed nucleosomes

Figure 6. Genome organization DNA helices wrap up around histones to be organized into more compact structure as nucleosomes, that becomes the structural unit of chromatin. Facilitated by proteins such as Cohesin, chromatin can be extruded into loops, potentially leading to the formation of chromatin domains that are relatively insulated from each other. These domains are termed as topologically associating domain (TAD) that are empirically discovered by high throughput chromatin conformational capture assay Hi-C.

One essential feature of chromatin organization is the formation of chromatin loops, as the outcome of chromatin-chromatin interactions in 3D. DNA loop formation involves a loop extrusion mechanism, wherein key effector protein Cohesin promotes the interaction between chromatin segments[45]. Additionally, stochastic loop formation also contributes to the loop structure as the resulted loop could be further stabilized by chromatin proteins.

DNA loop extrusion is also associated with the formation of topologically associated domain (TAD), another key feature of chromatin organization[46]. The extruded DNA loops can be further folded into itself to form local domains, that encompass chromatin segments showing higher interaction frequency with each other. Of note is that TAD is initially discovered by chromosome conformation capture techniques using high-throughput sequencing, which can be challenging to be validated at a global level using complementary techniques.

Phase separation is another prominent feature of chromatin organization that describes the local regions of chromatin forms gel-like condensates[47]. The formation of phase-separation is mediated by protein with self-interaction proteins, in some cases also involves long non-coding RNA. The phase separation also contributes to more stable chromatin-chromatin interaction by restricting the motion of chromatin[48].

1.4.2 Interplay between genome organization and function

Genome organization provides physical premises for dynamic activities, therefore understanding chromosomal organization and its functional implications is crucial for studying cellular activities and related phenotypes in organisms. In lines with the heterogenous expression profile across single cells is the heterogeneity of genome organization across cells. The expression of gene requires transcription factor and transcriptional machinery to be directed to the specific chromatin locations. Given the low copy number of most genes and low concentration of transcription factors, the genome organization could play an important role of modulating the efficiency of their searching and binding procedure[49]. The DNA loops for instance, sustain the interaction between regulatory elements and target genes, thus allowing the recruitment of transcriptional machinery and high expression of the regulated genes. Similarly, the phase separated condensates of chromatin encapsulate effectors and chromatin segments inside, leading to efficient reactions by limiting the motion of these molecules.

The role of TAD in gene expression as a structural feature at larger scale has also been studied, with more heterogenous effects across genome. On the one hand, disrupting cell type-specific TADs identified in motor neuron cells is associated with loss of cell identity, indicating the critical role of TAD formation in this context[50]. On the contrary, the role of gene organization in regulating gene expression has become less clear due to recent findings. For example, disrupting TAD boundaries globally through genomic manipulation has only affected the transcriptional outcomes of a subset of genes, while the rest remained unaffected by this structural perturbation[51]. In agreement of this, DNA enhancer-promoter loop formation involved in fly development was detected before the establishment of the TADs, suggesting indeterministic role of TADs for those gene programs[52].

From the other direction, the genome activity also contributes to shape the genome organization. From a global point of view, heterochromatin forms when chromatin is at transcriptionally repressive stage leading to compartments formation[46]. Moreover, transcription inhibition led to disappearance of chromatin domains but not topology in Drosophila, implying that transcriptional machinery and products cause the establishment of chromatin domains[53] .

One of the primary tools utilized to decipher the functional implications of genomic architecture is genetic perturbation. For instance, knockout of regulatory elements to study the effect of this on the genomic architecture and expression of hypothetical target gene simultaneously is a conventional approach as an indirect way of altering local genomic topology. Targeting the expression of proteins essential for maintaining genomic architecture such as Cohesin allows more direct perturbation of genomic architecture globally, while lacking the option to target specific loci[51].

The recent advances in CRISPR technology have provided an efficient platform for targeting specific genomic loci and studying the function of genomic structure. The CRISPR-Cas9 system allows for massively parallel perturbation of regulatory elements, enabling systematic evaluation of their effects on related genes[54]. Epigenome editing tools derived from CRISPR-dCas9 can alter the epigenetic status of targeted distal elements and assess their influence on regulatory elements, with an indirect linkage to the presence of chromatin-chromatin interaction[55]. In addition, direct techniques for modulating the interaction between pairs of genomic loci have been developed using the CRISPR-dCas9 system, including the introduction of a bivalent domain to dCas9[39] or co-transfection of mediator protein to achieve protein-protein interaction[38], and thus the chromatin-chromatin interactions between genomic loci targeted by dCas9. These techniques offer the unique advantage of facilitating de novo interaction formation for interpreting the function of genomic interaction within a specific context.

Overall, the genome architecture is intrinsically integrated into the gene expression dynamics. However, instead of a determinist effect from the high-order structure of chromatin on gene expression, different features of chromatin structure affect the gene function in a probabilistic manner. The gene expression activity, as feedback for the congregated effects could, in turn, reshape or stabilize the genome structure.

1.4.3 Research tool for studying genome organization

In the early days, the study of genome architecture was limited to microscopic observations of stained chromosomes, with major limitations in terms of resolution and specificity. Recent developments in deep sequencing technology and super-resolution microscopy have greatly facilitated the analysis of chromatin organization.

Sequencing-based methods for studying genome structure include ligation-based methods and barcoding methods. Hi-C is the most widely used ligation-based method to study chromatin conformation[56,57]. Hi-C involves crosslinking, digestion with restriction enzymes, ligation, and sequencing of chimeric DNA fragments generated by ligation of crosslinked fragments. Micro-C is a modified version of Hi-C that uses micrococcal nuclease digestion instead of restriction enzymes to fragment chromatin[58]. Micro-C provides higher resolution than Hi-C, particularly for short-range interactions, and can identify chromatin loops and other higher-order structures. HiChIP combines Hi-C with chromatin immunoprecipitation (ChIP) methods to enrich for specific chromatin-associated proteins, such as cohesin[59]. HiChIP can identify the 3D interactions between chromatin and transcription factors and can reveal the regulatory landscape of specific genomic loci.

SPRITE is a sequencing-based technique that is independent from relegation of interacting chromatin[60]. The SPRITE method involves proximity labeling of interacting chromatin segments followed by tagging of the labeled segments repeatedly resulting in a unique DNA barcode for chromatin in proximity, which can be later read out by sequencing and analysis of the barcoded segments to identify interacting chromatin regions. Other ligation-independent techniques such as ChIA-Drop also involve the barcoding the interacting chromatin to be identified later after sequencing[61]. The major advantage of these methods are that they allow identification of multiway interaction and high sensitivity.

Overall, sequencing-based methods are capable of analyzing chromatin organization at high resolution and with high throughput, allowing for a comprehensive view of the genome efficiently. Furthermore, they are unbiased and do not rely on a priori knowledge about chromatin architecture, allowing for the identification of novel structures and interactions. Additionally, they are capable of capturing the heterogeneity of chromatin organization within a population of cells, revealing the dynamics of chromatin organization. Finally, sequencing-based methods are scalable and can be adapted to a variety of different experimental conditions, making them a versatile tool for studying genome architecture in a wide range of biological contexts.

Microscopy-based techniques have provided valuable insights into the spatial organization of the genome. Various types of microscopy methods have been developed, including fluorescence in situ hybridization (FISH), DNA-based super-resolution microscopy such as Oligo-PAINT, and chromatin immunoprecipitation (ChIP) followed by microscopy.

FISH is a widely used technique for the visualization of specific genomic regions in fixed cells. It uses fluorescently labeled probes that hybridize to complementary DNA sequences to visualize the location of specific genomic regions. FISH is a powerful tool for studying the

3D organization of the genome, as it allows for the visualization of specific chromosomal domains and their spatial relationships. However, FISH has limited resolution typically around 200nm, insufficient for detecting chromatin-chromatin interaction. MERFISH (Multiplexed Error Robust Fluorescence In Situ Hybridization) is a single-molecule imaging technique that allows for the visualization of multiple genomic loci in situ with high spatial resolution. It works by using a large number of oligonucleotide probes, each labeled with a unique combination of fluorophores, to hybridize to the target RNA or DNA molecules. By decoding the combination of fluorophores, MERFISH can distinguish DNA molecules with high specificity and simultaneously image multiple targets[62]. With these advantages, it has been applied to track the 3D organization of long consecutive chromatin.

Similarly, FISH techniques that can be combined with super-resolution microscopy including Oligo-PAINT are also developed. Oligo-PAINT is based on the hybridization of short, fluorescently labeled oligonucleotides to complementary DNA sequences, which are specifically designed to target specific genomic loci. The oligonucleotides can be repeatedly hybridized and removed to achieve a high signal-to-noise ratio and reduce the background fluorescence[63]. The oligonucleotides can be further combined with other techniques, such as single-molecule localization microscopy (SMLM)[64] or stochastic optical reconstruction microscopy (STORM)[65], to achieve high-resolution imaging of multiplexed genomic loci. Super-resolution microscopy techniques have the advantage of providing high-resolution images of genome architecture, allowing for the visualization colocalization of genomic loci. However, these techniques often require specialized instrumentation and labeling techniques, and can be relatively slow compared to other imaging techniques.

In general, Sequencing-based methods, such as Hi-C and ChIA-PET[66], provide a high-throughput, genome-wide view of chromatin interactions. They are well-suited for identifying long-range interactions. These methods also provide a quantitative measure of interaction frequencies and can reveal changes in chromatin architecture under different conditions. However, sequencing-based methods have some limitations, such as the resolution and sensitivity for rare interactions.Microscopy-based methods, such as super-resolution microscopy and fluorescence in situ hybridization (FISH), offer a more direct view of chromatin structure and interactions at the subcellular level. These methods can provide high spatial resolution, allowing visualization of chromatin organization at the level of individual cells or even single molecules. Additionally, microscopy-based methods can be used to visualize multiple chromatins features simultaneously, such as histone modifications and DNA sequences. However, microscopy-based methods are limited by the number of regions that can be visualized simultaneously, making it difficult to obtain a genome-wide view of chromatin interactions.

2 Research aims

The aim of the work included in this thesis is to develop molecular tools for characterizing and modulating the local organization of the genome to decode its function. By exploiting the design space of nucleic acid tools and the Cas9-gRNA system, we developed techniques to address questions about genome organization with complementary strategies to the existing toolkits. The specific aims of the papers presented in this thesis are:

Paper I: To develop computational framework and proof-of-concept for sequencing-based imaging methods, simulating the DNA crosslinking by spatial adjacency and reconstructing image information via several algorithmic strategies.

Paper II: To develop an efficient method for fabricating DNA origami nanostructures functionalized with orthogonal molecules including proteins and fluorophores for nano-applications.

Paper III: To develop a simplistic and flexible system to remodulate local gene organization as gene regulatory tools, by redesigning gRNA of CRISPR-dCas9.

Paper IV: To develop a method directly visualize interaction between targeted genomic loci *in situ*, using oligonucleotide tagged Cas9.

3 Materials and methods

3.1 Production of DNA origami

3.1.1 design of DNA origami

DNA origami in paper II and IV was used as a nano-patterned platform for gauging the efficiency of the reactions. The primary considerations for the design of DNA origami are the location of functional sites on DNA origami. To this end, we used a rod-like DNA origami structure 18-helix bundle (18HB) structure that has shown suitable dimensions for the application in Paper IV with predicted dimensions of nm in length and ideal rigidity to reduce the deviation of folded structure from the predicted dimensions. The structure was designed with CaDNAno to create scaffold crossovers and staple strands for packing 18 parallel DNA double helices in a honeycomb lattices. Breakpoints for staple strands to make it compatible with commercial DNA oligo synthesis were introduced manually to constrain the length of staple within the range of 30-60nt. The sequences of staple strands were generated using the add seq feature from CaDNAno by applying the sequence of selected scaffold DNA to a randomly selected breaking point on scaffold.

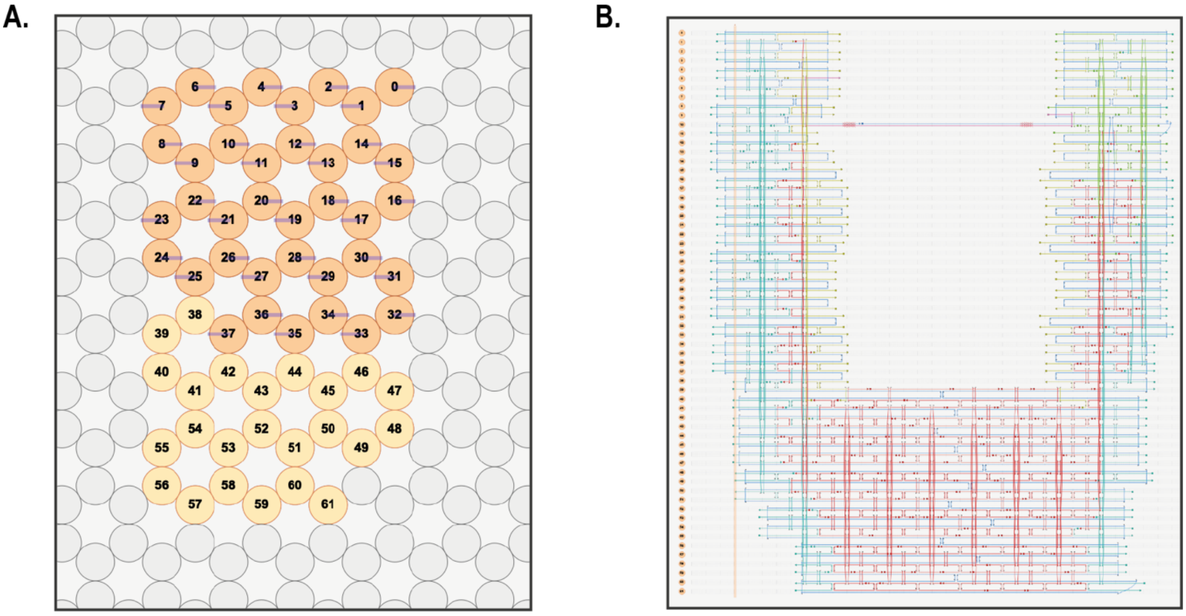

Figure 7. Using CaDNAno for DNA origami design. (A) Screenshot of CaDNAno interface. Cross-sectional view of the structure design on honeycomb lattice (B) Screenshot of CaDNAno interface path panel. DNA routing for the structure shown in (A)

The functional sites were selected from the staple strands, of which the termini were protruding towards the outside of DNA origami. In paper II, to add linker sequence to bridge up the structure and magnetic beads, staples at the edge of the 18HB were chosen for testing by adding extended sequence to the termini of staple oligonucleotides. To add functional groups such as Haptens and fluoreophores, staples at the selected locations were chosen to be either directly synthesized with hapten at 5' end or further extended at termini with extra unpaired sequence. In paper IV, the functional sites were selected in pairs to adjust the distance between protrusions for Cas9 binding. Cas9 recognition sequences were then concatenated to the sequence of staple oligonucleotides at the selected locations. After modifications of the staple sequence, the staples were ordered in 96-well plates from Integrated DNA Technologies.

3.1.2 Folding, purification and characterization of DNA origami

3.1.2.1 scaffold production

A single colony of TOP10 *E.Coli* picked from an agar plate was inoculated into 25 ml of Lysogeny Broth and grown overnight at 37°C in a shaking incubator. The bacterial culture was transferred to 500 ml of 2xYT medium supplemented with 1M MgCl2 and grown until reaching an OD600 of 0.5. The inoculation p7560 phage was added at a multiplicity of infection of 1, and the culture was grown for an additional 5 h. The culture was centrifuged twice at 4000 rcf for 25 min at 4°C to pellet the bacteria. To precipitate the phage, the supernatant was mixed with PEG 8,000 and NaCl and incubated on an ice water bath for 30 min. The mixture was centrifuged at 4000 rcf for 15 min at 25°C, and the phage pellet was resuspended in 10 mM Tris 1 mM EDTA buffer (pH 8.5). The solution was then centrifuged at 15,000 rcf for 15 min at 25°C to remove any bacterial remains. The supernatant was mixed gently with 1% SDS in 0.2 M NaOH and incubated for 3 min at room temperature to remove the coat proteins of the phage. Immediately after, 3 M KOAc (pH 5.5) was added to the solution and mixed gently by swirling. After an incubation of 15 min on ice, the mixture was centrifuged at 16,000 rcf for 10 min at 4°C. The supernatant containing the single-stranded DNA (ssDNA) was mixed gently with 99.5% EtOH and incubated on an ice water bath for 30 min before centrifuging at 16,500 rcf for 30 min at 4°C. The ssDNA was further washed with 75% EtOH and centrifuged again at 16,500 rcf for 10min at 4°C. After air-drying the ssDNA pellet, it was resuspended in 10 mM Tris (pH 8.5). The concentration of p7560 was measured by absorbance at 260nm by nanodrop.

3.1.2.2 Folding and purification DNA origami

The standard folding reaction of the DNA origami structure consists of 20nM ssDNA p7560 scaffold, 100nM of each staple, $MgCl_2$ at concentration ranging from 0 mM to 20 mM, 5mM Tris(pH=8.5), 1mM EDTA. The folding reaction in Paper IV is slightly modified by adding additionally 1.2μM protruding target staples and 4.8μM complementary target oligos. The folding reaction is initiated by rapidly heating the mixture to 65°C and then slowly cooling it to 24°Cover a period of 16 hours. To remove excessive oligos from the folded structure, the reaction mixture is purified by washing repeatedly for 10 cycles of ultracentrifugation with

100kDa cutoff Amicon centrifugal column at 14000rcf for 1 min using 1xPBS,5mM $MgCl_2$. The concentration of purified structure is determined by absorbance at 260nm by nanodrop.

The folding reaction for the solid phase synthesis of DNA origami described in Paper II consists of 10nM ssDNA scaffold, 50nM of each staple, 100nM ssDNA linker, 270mM sodium polytungstate and oligo dT_{25} magnetic beads. The storage buffer of dT_{25} magnetic beads is replaced by equivolume of 5mM Tris(pH=8.5), 1mM EDTA buffer before adding to reaction mixture, keeping the scaffold to beads ratio is commonly 5.83ng/ µl. The folding reaction uses the same program as conventional folding program. After the folding reaction, the magnetic beads are settled on magnetic stand to remove the supernatant. The beads are subsequently washed by 1xPBS, 10mM $MgCl_2$. After washing, the beads are resuspended with 1XPBS, 10mM $MgCl_2$, and 250nM invader oligos for 1h on rotator, before the settlement of magnetic beads and retrieval of supernatant containing eluted DNA origami.

3.1.2.3 characterization of DNA origami

The folding of the DNA origami is first characterized by 2% agarose gel supplemented with 10mM $MgCl_2$,, 0.5mg/ml Ethidium Bromide(EtBr). For DNA origami bound with Cas9-sgRNA, 1.5% agarose gel supplemented with 10mM $MgCl_2$ is used. The samples are run for 2-3 h at 90V in ice water bath to avoid overheating.

Negative-stain transmission electron microscopy (NS-TEM) is a conventional technique used to validate structural details of DNA origami after the folding reaction. In this process, a folded structure sample of 4ul at a concentration of 10-20nM is applied to glow-discharged Carbon-coated formvar grids for 20 seconds. The excess sample is removed by blotting, and the grids are rinsed by tabbing them to water for 4 seconds, followed by blotting to remove excess water. The grids are then stained with aqueous uranyl formate solution for 20 seconds, after which the staining is blotted off, and the grids are air-dried for at least 45 minutes before imaging.

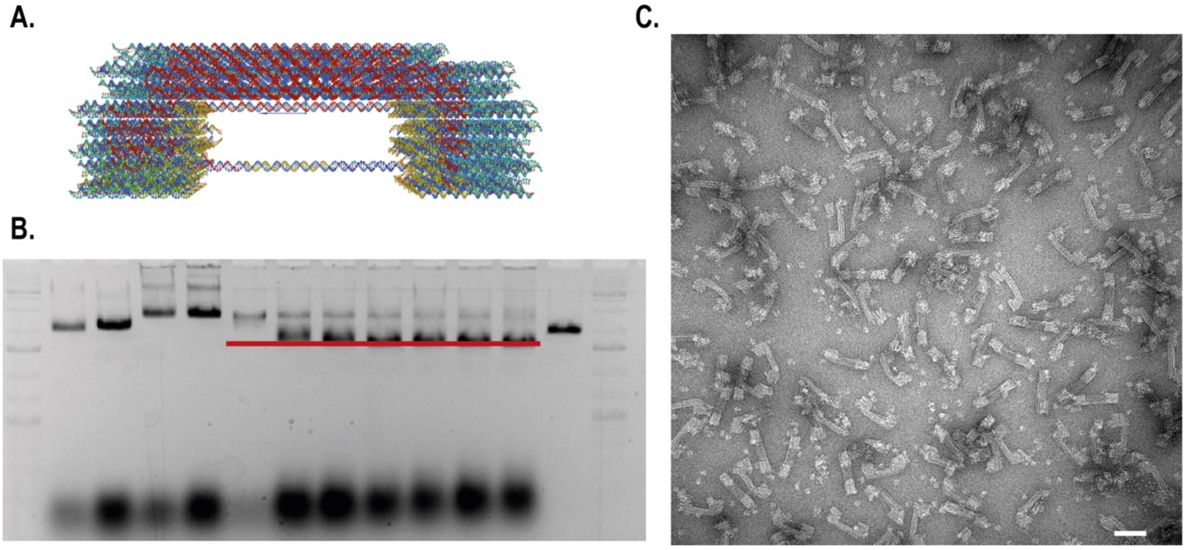

Figure 8. Characterization of DNA origami (A) OxDNA model of a DNA origami (B)Magnesium screening for the folding of DNA origami. (C) Negative-stain transmission electron microscopy images of DNA origami.

Micrographs were obtained using SerialEM on a Talos 120C G2 electron microscope equipped with a Ceta-D detector at a magnification of 73000x. For the purpose of data collection for 2D class average in Paper II, SerialEM was employed to automate the process. Scipion was used for single particle analysis (SPA) to generate 2D classes. In the processing pipeline, the micrographs were down-sampled 5-fold before particle-picking. Particles were picked with a box-size corresponding to the dimensions of each structure, with a 10% margin, using the Autopick option from the Xmipp picker in Scipion. The function was trained using manually picked particles with a minimum number of 15, and auto-picking results were manually corrected to improve performance. The particles were further scrutinized before being used for 2D classification into 16 classes in Paper II, in order to visualize antibody-bound DNA origami.

3.2 Production and characterization of sgRNA/dgRNA

3.2.1 design and production of sgRNA/dgRNA

In papers III and IV for in vitro experiment, the sgRNA and dgRNA were designed targeting pUC19 by identifying the protospacer adjacency motif (PAM) of S.pyogenes Cas9 from the pUC19 sequence. The sgRNA transcription template sequence for each target sequence consists of a T7 promoter, the identified 20nt target sequence located 5' upstream of the PAM, and the sgRNA scaffold sequence in the 5' to 3' order. The dgRNA transcription template sequence includes a T7 promoter, the 20nt first target sequence, the sgRNA scaffold sequence, a poly-adenine linker of selected length, the 20nt second target sequence, and the sgRNA scaffold in the 5' to 3' order. In the CasT-PLA protocol from paper IV, the sgRNAs used contain 4nt mismatches at the 5' end of the sgRNA to avoid cleavage of the target after binding.

For in vivo studies, the dgRNAs were designed to be constitutively expressed in *E.Coli*, that include a synthetic promoter, 20 nt first target sequence, sgRNA scaffold with modifications to remove the transcription terminator, poly-adenine linker, 20 nt second target sequence, full sgRNA scaffold.

To produce sgRNAs, the transcription template was ordered as ssDNA Ultramer with the sequence from the reverse complementary strand of T7 promoter sequence. The ultramer was first annealed to oligo with T7 promoter sequence in a ratio on 1:2 with a final concentration of 5 µM in annealing buffer (20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM MgCl2, 1 mM DTT) by rapidly heating to 95/

3.2.2 Characterization of sgRNA /dgRNA in vitro

Following in vitro transcription, the transcription template DNA was subjected to DNaseI digestion at a concentration of 0.05 U/µl at 37°C for 15 minutes. RNA binding spin column was employed to eliminate any impurities during the reaction. The quality of the RNA products was assessed using 2% agarose gel containing 0.5mg/ml EtBr. The RNA was denatured by adding 100% formamide to the sample at a ratio of 1:1 (v/v) and then heated to 95°C for 10 minutes.

To evaluate the activity of dgRNAs, as described in Paper III, gel electrophoresis mobility shift assay (EMSA) was used. The dgRNPs were assembled at room temperature in a dCas9 buffer comprising of 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM MgCl2, 5% glycerol, and 1 mM DTT, with a ratio of dCas9:dgRNA at 2.4:1, resulting in a final RNA concentration of 250nM. The RNPs were then added to fluorescently labeled DNA substrates at a concentration of 100nM in equal volumes and incubated for one hour at 37°C. The samples were then subjected to 4% native PAGE gel electrophoresis at 300V for 20 minutes in 0.5x TBE buffer supplemented with 5 mM MgCl2 at 4°C.

To further verify the binding of dgRNPs to DNA with sequence-specificity, samples of dgRNP binding to DNA were imaged by NS-TEM as described previously. The data obtained from the experiment was processed using CryoSPARC[67]. Template picking was employed to select particles, and templates were generated using manually picked particles from the dataset. The selected particles were subjected to two rounds of 2D classification, and the final 2D classes were used for subsequent analysis, with the classes from the second round of 2D classification being the ones selected for analysis. The particles were cleaned during the 2D classification rounds to improve data quality.

Atomic force microscopy (AFM) was employed as a complementary technique to verify the interaction between dgRNP and DNA. To prepare the sample for imaging, freshly cleaved mica surface was pre-treated with 5 mM Nickel sulfate for 1 minute at room temperature. The surface was then washed three times with water and dried using nitrogen airflow. The sample was deposited on the mica surface for 1 minute, washed three times with water, and dried with nitrogen airflow.

AFM images were acquired using a JPK instruments Nanowizard 3 ultra equipped with an Olympus Biolever mini cantilever in AC mode. Bruker SCANASYST-AIR probes (Bruker) with a spring constant of 0.4N/m were utilized for imaging at a driving frequency of approximately 80kHz.

3.2.3 functional characterization of dCas9-gRNA in E.Coli

3.2.3.1 construction of Bacteria strains

For the purpose of assessing dgRNA function in vivo in paper III, we generated *E. Coli* strains that expressed inducible dCas9 and constitutively expressed dgRNA. The dgRNA sequence was chemically synthesized as aplasmid by Thermo Fisher, and then subcloned into a pgRNA plasmid containing a synthetic promoter for constitutive expression of the gRNA. The pgRNA vector and dgRNA plasmid were digested using SpeI and HindIII, and the digested products were loaded onto a 2% agarose gel. The desired products were extracted from the gel using the QIAquick Gel Extraction Kit. The gel-extracted vector and dgRNA insert were ligated with T4 DNA ligase at a ratio of 1:2 for 1 hour at room temperature, with a final vector concentration of 5 ng/μL. 4 μL of the ligation mixture was added to 50 μL of DHalpha5 cells on ice for 30 minutes, followed by heat shock for 30 seconds at 42°C. After 5 minutes of recovery on ice, 950 μL of SOC medium was added to the cell mixture, which was then incubated for 1 hour at 37°C in a shaking incubator. The cells were then diluted with SOC medium and plated on pre-warmed ampicillin selection plates.

After overnight culture at 37°C, colonies were selected from the agar plates and cultured in LB broth with ampicillin resistance. The plasmids were then prepared using a miniprep kit from Thermo Fisher, following the manufacturer's protocol. The sequence of the prepared plasmids from selected colonies was verified using Sanger sequencing.

To generate the *E.Coli* strains for functional verification, we performed heat transformation of *E.Coli* NEBExpress cells using pgRNA-constructs for each dgRNA variant and pdCas9-Bacteria plasmids and selected them on agar plates with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. Double positive colonies were then selected after overnight growth and cultured for subsequent experiments.

3.2.3.2 Functional characterization of dgRNA-dCas9 using β-galactosidase assay

For the β-galactosidase assay to measure dgRNA-dCas9 activity, 10 μL of overnight culture of each dgRNA-dCas9 expressing strain was inoculated into 90 μL of Lysogeny broth (LB; VWR) supplemented with 1 mM IPTG (VWR), 100 μg/mL of Ampicillin, and 34 μg/mL of Chloramphenicol. When the bacteria reached mid-log phase, they were induced with or without 2 μM of μTC and grown for an additional 2 hours. A volume of 100 μL from each culture was subjected to the yeast β-galactosidase assay kit (Pierce), following the manufacturer' s instructions. The absorbance at 420 nm and 600 nm was measured to calculate β-galactosidase activity using the following equation:

$$\beta - galatosidase\ activity = \frac{1000 \times A_{420}}{t \times V \times OD_{600}}$$

To conduct serial measurements, 1 µL of overnight culture from each dgRNA-dCas9 expressing strain was inoculated into 99 µL of Lysogeny broth supplemented with iPTG, chloramphenicol, and ampicillin at various time points up to 8 hours. The β -galactosidase activity was then measured using the same procedure as before.

3.3 Production and purification of oligonucleotide-labelled Cas9

To create the oligonucleotide-labelled Cas9, we first produced Cas9 with Sortase A recognition tag at the C-terminus. The sequence encoding Sortase A was synthetized and cloned into the expression vector pET-29b(+) by BioCat GmbH. *E.Coli* BL21 (DE3) T1R pRARE2 cells were cultured overnight at 37°C in Terrific Broth supplemented with 8 g/L glycerol and 34 ug/ml chloramphenicol. The bacterial culture was then inoculated with the Cas9-SRT plasmid-transformed bacteria and further supplemented with 0.4% glucose and 50 ug/ml kanamycin, followed by overnight growth at 30°C. The next day, the culture was transferred to a LEX bioreactor system and grown at 37°C, When the culture reaching OD600 = 2, temperature was reduced to 18°C to grow until OD600=3. The protein expression was induced with 0.5 mM IPTG and maintained overnight at 18°C. The cells were harvested by centrifugation at 4500 g for 10 minutes and the pellets were resuspended in lysis buffer containing 100 mM HEPES pH=8.0, 500 mM NaCl, 10 mM imidazole, 10% glycerol, and 0.5 mM TCEP, and then frozen at -80°C. After thawing, the lysates were sonicated, filtered, and subjected to protein purification by Ni-NTA chromatography (HisTrap HP, GE Healthcare). The protein was bound to nickel Sepharose resin in binding buffer containing 20 mM HEPES pH=7.5, 500 mM NaCl, 10 mM imidazole, 10% glycerol, and 0.5 mM TCEP, washed with the same buffer containing 50 mM imidazole, and eluted with 500 mM imidazole. The eluates from the affinity column were further purified by size exclusion chromatography (SEC) (HiLoad 16/60 Superdex 75, GE Healthcare) using buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, and 10% glycerol, 2mM TCEP. The purity of the CAS9-SRT enzyme was assessed by 12% denaturing polyacrylamide gel electrophoresis (PAGE).

Cas9 was conjugated to oligonucleotide tags using a previously published protein Sortagging method[68]. Prior to the reaction, the Cas9-SRT protein and Sortase A were buffer exchanged into a Sortase A reaction buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl, and 10 mM CaCl2, using Zeba Spin 7 kDa desalting columns. The DBCO-amine was dissolved in DMSO to a final concentration of 1 M, and the Sortagging reaction was carried out at room temperature for 1 hour with a final concentration of 50 mM Cas9, 50 mM Sortase A, and 10 mM DBCO-amine in Sortase A reaction buffer. To remove excess DBCO-amine, the mixture underwent buffer exchange into a click chemistry reaction buffer containing 20 mM HEPES (pH 7.5), 300 mM NaCl, and 10% glycerol using two centrifugation rounds of Zeba Spin 7 kDa desalting columns. The click chemistry reaction was carried out overnight at 4°C with a final concentration of 40 mM for the DBCO-modified proteins and 200 mM for the azide-modified PLA probes. The reaction was then purified using ion-exchange chromatography on a Profire system, and the fractions corresponding to the protein-PLA probes conjugates were pooled using 30 kDa Amicon Ultra centrifugal filter columns and buffer exchanged into Cas9 binding/reaction buffer (20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 0.01% TWEEN20, and 5% glycerol).

The final conjugates were subjected to incubation with a 10-fold molar surplus of complementary oligonucleotides labelled with Cy5. The resultant mixture was then subjected to electrophoresis on a denaturing polyacrylamide gel containing 10% concentration. Additionally, the proteins were stained with SYPRO Orange (Thermo Fisher, #S6650) as per the manufacturer's guidelines, and subsequently visualized using an ImageQuant LAS 4000 gel-imager (GE Healthcare).

3.4 CasT-PLA reaction and analysis

3.4.1 sample preparation

To prepare samples for CasT-PLA, MCF-7 Cells were cultured on Poly-D-Lysine-coated microscopy 8-well chamber slides overnight and rinsed once with 1xPBS. Cells were then fixed and permeabilized with a 1:1 mixture of Acetic Acid and Methanol and incubated at -20°C for 20 minutes. After fixation, the samples were washed in PBS for three times with gentle shaking at room temperature for 5 minutes each. The samples were then blocked in a blocking buffer consisting of 20mM HEPES (pH=7.5), 100mM NaCl, 5mM MgCl2, 0.05% TWEEN20, 0.1mM EDTA, 1mM DTT, 250 μ g/ml BSA, and 25 μ g/ml sonicated Salmon sperm DNA (Thermo Fisher #AM9680) at 37°C for 2 hours in a humidity chamber.

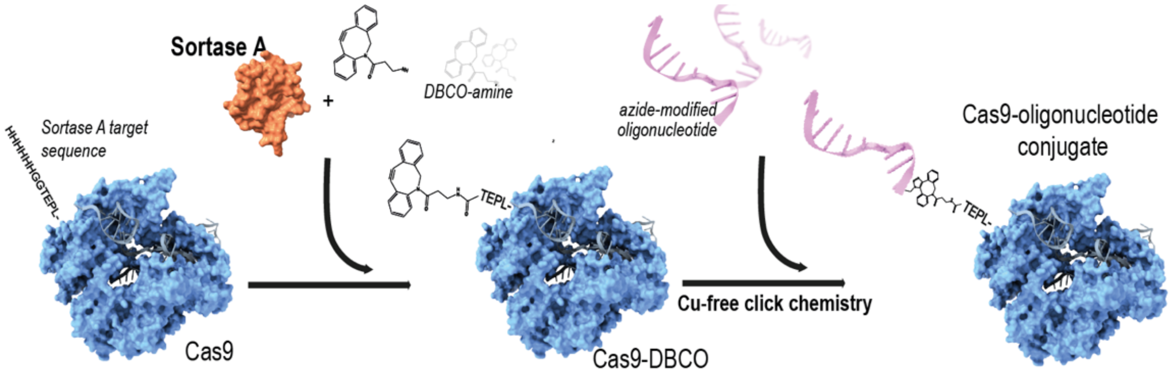
Sortase A
+
DBCO-amine
Sortase A target sequence
HHHHHHGGTEPL-
Cas9
TEPL-
azide-modified oligonucleotide
Cu-free click chemistry
Cas9-DBCO
TEPL-
Cas9-oligonucleotide conjugate

Figure 10. Sortase A-mediated site-specific conjugation of Cas9 with DNA oligos Sortase A directs the modification of Cas9 with DBCO-amine through reaction between SorTag from Cas9 and amine group. DNA oligo is then conjugated to DBCO via copper-free click chemistry.

CasT ribonucleoprotein probes were assembled from CasT protein and sgRNA with a ratio of 1:5 in dCas9 buffer (20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 0.01%TWEEN20, 5% glycerol) and incubated for 10 minutes at room temperature. Each CasT probe was diluted with blocking buffer to desired concentration before adding CasT1-RNP and CasT2-RNP at a 1:1 ratio to the pre-blocked samples, followed by 2-hour incubation at 37°C in a humidity chamber.

To remove the CasT probes after the binding reaction, samples were washed with blocking buffer for 5 minutes with gentle shaking, repeated three times. The ligation reaction mix (1xT4 ligation buffer including 1mM ATP (New England Biolabs), 100nM circularization oligo1, 100nM circularization oligo2, and 0.05U/ μ l T4 DNA ligase (New England Biolabs)) was added to the samples to initiate a 2-hour ligation reaction at 37° C, followed by washing with blocking buffer for three times.

The rolling circle amplification was conducted at 37° C for 2 hours by adding RCA mixture (1x Phi29 reaction buffer (Thermo Fisher), 1mM dNTPs, 100nM Cy5-labelled detection probe, and 0.1U/ μ l Phi29 DNA polymerase). After the reaction, samples were rinsed with blocking buffer for three times, followed by another three times of washing with PBS. Before imaging, samples were stained with 2nM Hoechst 33343 (Thermo Fisher) and 5 μ g/ml Alexa Fluor™ 488 Phalloidin (Thermo Fisher) for 10 minutes in darkness, then rinsed with PBS for three times. Coverslips were removed from the chamber slide before imaging and mounted to glass slides with mounting media.

3.4.2 Image analysis

In the work from Paper IV, the samples were imaged using a home-built inverted microscopy setup equipped with 100x oil immersion objectives and illuminated by OBISTM 405nm, 488nm, and 642nm LS 150mW lasers, along with corresponding emission filters. Images were acquired using an iXon Ultra 888 EMCCD camera (Andor), with a 1024 x 1024 ROI centered on the illuminated sample, and sequential illumination by the lasers. Z-stack images were captured with micromanager software using multi-dimension mode for a series of 1 μ m slices. Stack images from each channel were pre-processed by max intensity Z-projection and merged as multi-color images.

To automate the quantification of puncta per cell nucleus, we used CellProfiler to create automated analysis pipeline[69]. For each multi-channel image, nuclei were identified using the 'IdentifyPrimaryObject' function with an 'Adaptive' threshold strategy, assigned with typical object diameters of 100-230 pixels. The Cy5-channel was pre-processed by applying the 'EnhanceOrSuppressFeatures' function to enhance speckles features. Puncta were identified by another 'IdentifyPrimaryObject' function using a 'Global' threshold strategy, assigned with typical object diameters of 3-15 pixels. The 'RelateObjects' function was applied to assign identified PLA objects as 'Child' objects for identified nuclei objects, and the number of PLA puncta contained within each nucleus identified across images was quantified automatically. All samples were analyzed using the same pipeline and parameters for filtering and quantification.

4 Results

4.1 PaperI

4.1.1 Voronoi tessellation as a model for generating polony

In Paper I we first discussed the use of Voronoi tessellation as a model for polony saturation, a phenomenon where polonies on a bounded two-dimensional (2D) surface are restricted from further growth after encountering neighboring polonies. The resulting planar Voronoi tessellation can be represented by its plane dual Delaunay diagram, whose vertices are the seed points of the Voronoi cells (polonies) and whose edges are the line segments connecting adjacent cells. The untethered graph, defined by its vertices and edges without explicit spatial information, is obtained from the Delaunay diagram by omitting all geometric information, retaining only topological characteristics of the graph.

We the proposed the use of topological metrics as a proxy for Euclidean metrics, by treating pairs of barcodes as unique markers of polony adjacency. By finding a proper straight-line planar embedding of the untethered graph, the metric properties of the underlying Delaunay diagram and the corresponding Voronoi tessellation can be approximated. One constraint on the candidate embedding is that it must be planar, due to the physical assumption that barcode pairings correspond to polony adjacencies and thus cannot bridge non-neighboring polonies.

We next discussed the use of the Tutte or barycentric embedding algorithm for finding a plane embedding of a planar graph, which is applicable to Delaunay diagram-type graphs. Another quality constraint is that the embedding must preserve the adjacency relations of the original graph, as represented by the barcode pairs. We presented empirical evidence that, with a sufficiently dense Poisson-distributed placement of seed points, the topological metric on the untethered graph approximates well the actual Euclidean metric of the points in the placement.

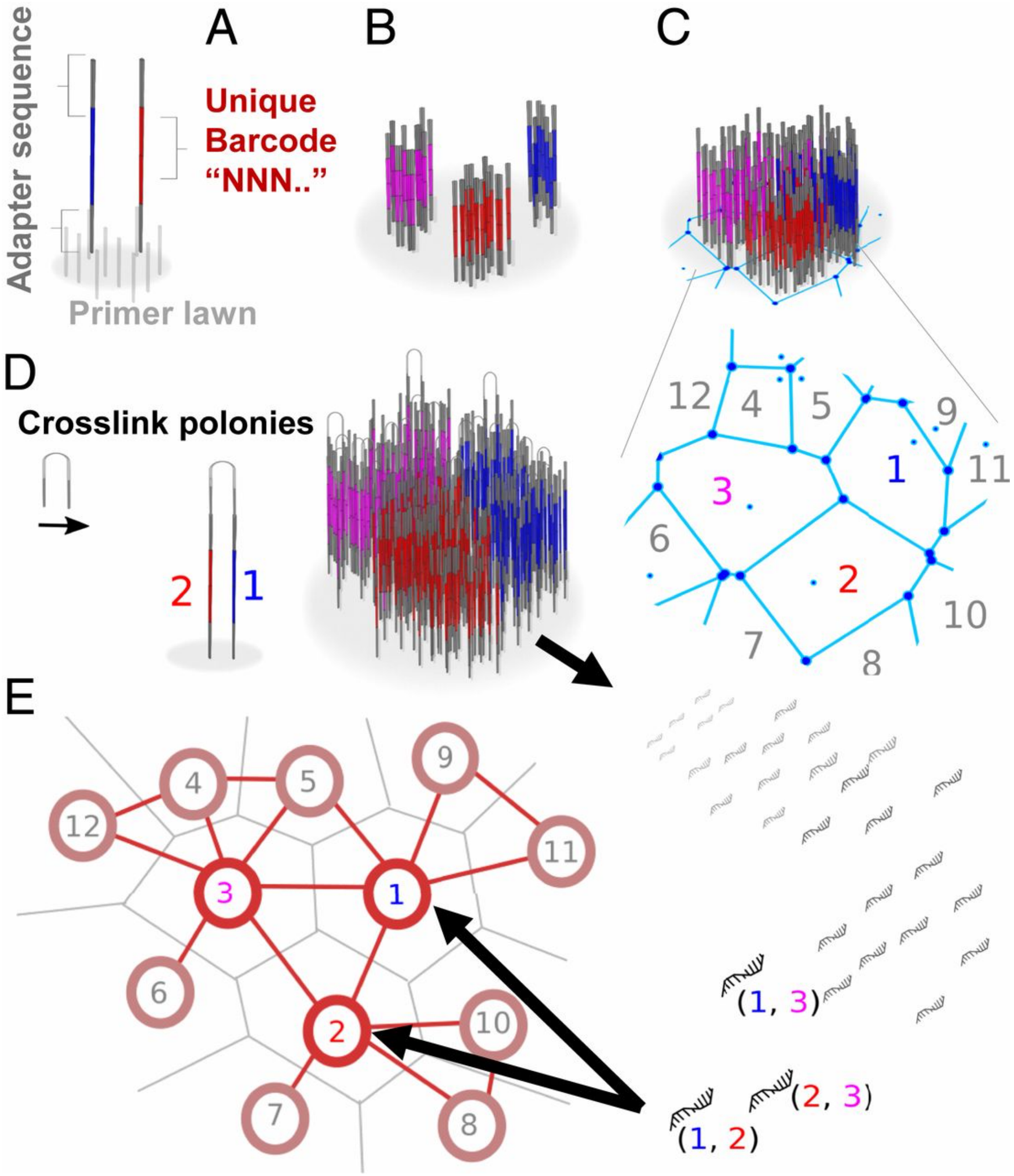

Figure11. The process of encoding and recovering metrics using polony adjacency. (A) seed molecules with unique barcode sequences are randomly placed on a surface of primers. (B)Local amplification of these seed molecules leads to the creation of sequence-distinct polonies.(C) Saturation of polonies occurs when they encounter adjacent polonies, resulting in a tessellated surface.(D) Random cross-linking of adjacent strands leads to pairwise association of nearby barcodes.(E) Finally, barcode pairs are recovered and sequenced, allowing for the reconstruction of a network that closely mirrors the original surface's polony positions.

4.1.2 Simulation of polony adjacency reconstruction

We then carried out the simulation and reconstruction of a primer lawn as a hexagonally packed disk with M primer sites and a random seeding at a polony density of λ. They randomly select N = λ A primer sites and pair adjacent polony primer sites to deduce the presence of a spatial boundary, where the fraction of information-bearing cross-pairs diminishes with the relative site density ρ. We also developed three approaches for approximating Euclidean metrics from the untethered graph and showed that knowledge of polony locations could be exploited to provide spatial information about objects of interest.

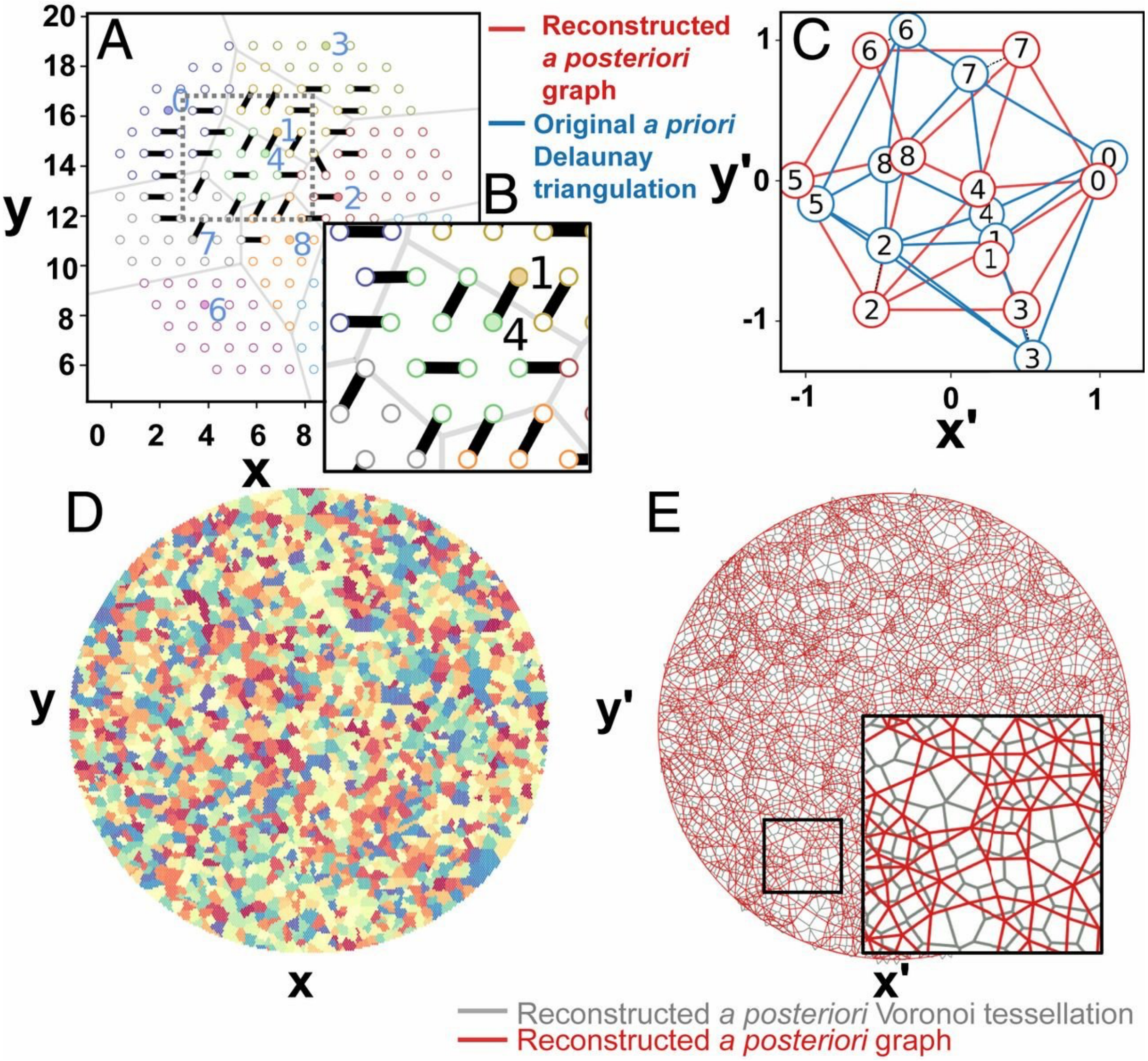
A
B
C
D
E
Reconstructed a posteriori graph
Original a priori Delaunay triangulation
y
x
y'
x'
20
18
16
14
12
10
8
6
0
2
4
6
8
1
0
-1
Reconstructed a posteriori Voronoi tessellation
Reconstructed a posteriori graph

Figure12 Simulation of polony adjacency reconstruction (A) shows a lattice diagram of the primer lawn and polonies, with colors and Voronoi cell boundaries indicating polony locations, and solid circles indicating seed locations. (B) illustrates the random pairing of adjacent primer sites.(C) shows the alignment between a priori and a posteriori points from (A). (D) depicts a larger simulated surface with a polony density of 2,000 polonies per unit

area and a relative site density of 50 sites per polony on average.(E) displays the reconstructed graph (red lines) and corresponding Voronoi tessellation (gray lines) computed using the Tutte embedding approach from scrambled edges derived from the simulated surface in (D).

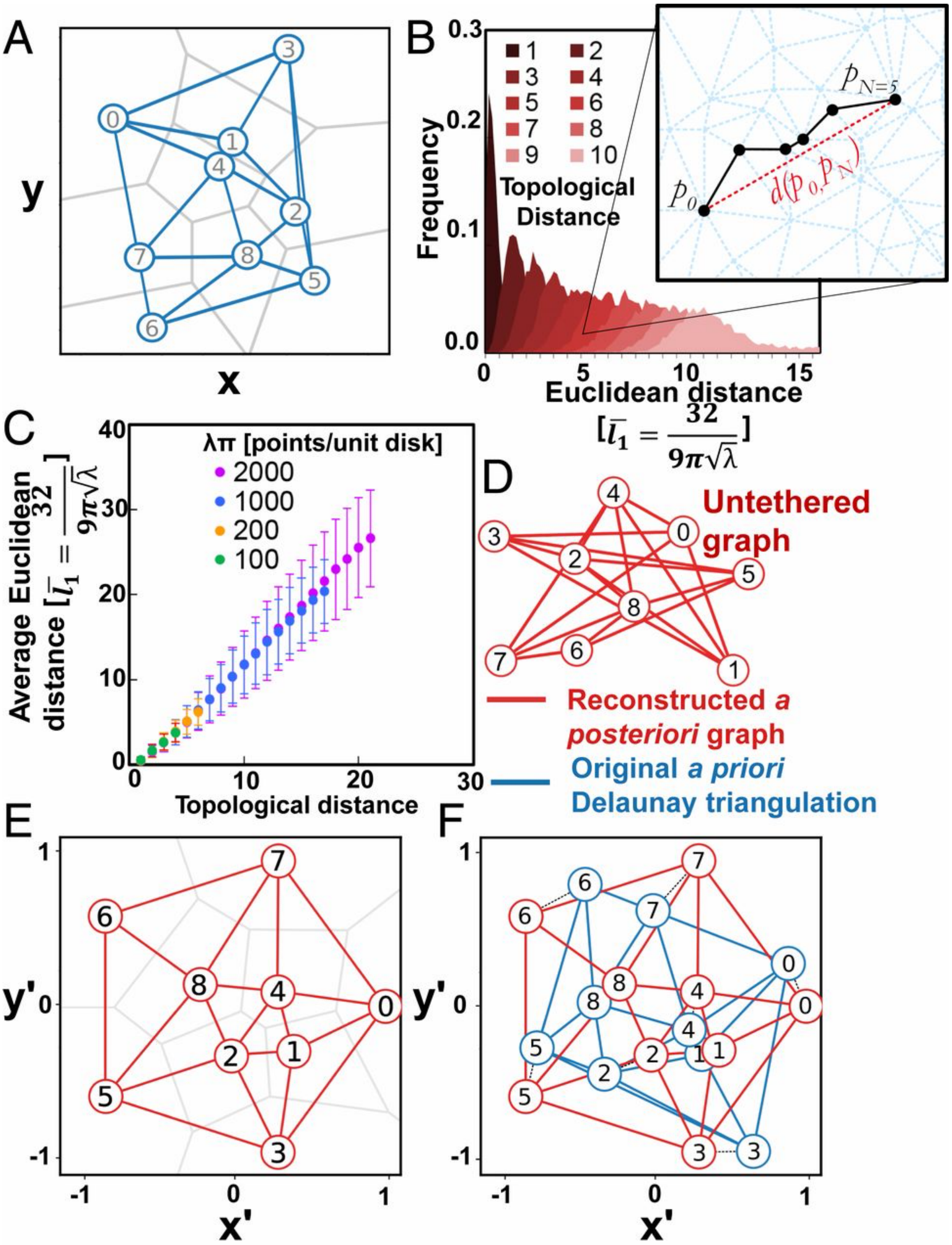

Figure 13 Illustration of the process of encoding and recovering metrics via topology. (A) Nine seed molecule points randomly distributed on a plane and the resulting Voronoi tessellation T (gray lines), Delaunay diagram D (blue lines), and untethered graph G.The distribution of Euclidean distances for a given topological distance (the shortest path between two points) is plotted for random Poisson Delaunay triangulations in (B). In (C), Euclidean distances are normalized to the average length of a typical Poisson Delaunay edge, and a linear relationship between topological and Euclidean distance is observed for different Poisson intensities.The untethered graph is shown in (D), consisting of nodes (black) and edges (red) that preserve the information after dissociation from spatial context. In (E), the Tutte embedding approach is used to reconstruct the planar embedding of the untethered graph (red lines), and the corresponding Voronoi tessellation is shown (gray lines). Finally, (F) demonstrates the alignment of the reconstructed embedding from (E) with the original Delaunay diagram from (A).

4.1.3 Stamping and image formation using polony adjacency information

We then developed a model for image reconstruction based on the transfer of molecules of interest to a mapped surface, using the principle of contact or diffusion-based molecule transfer. The model involves using polonies, or discrete amplification sites, to capture molecular markers labeled with identifying sequences called “red,” “green,” and “blue.” We used a hypothetical probability distribution of three types of molecular markers represented by an image to demonstrate the model's proof of concept. The color of the image indicates the density of markers and the probability of a marker of a particular color being placed on the polony surface. A Voronoi tessellation is produced from the final set of vertex positions, and each cell of the tessellation constitutes a pixel that can be used to form an image. The final RGB value of the cell is determined by tallying the markers that have associated with the primer sites in the polony, as well as the number of unassociated sites. We generated Voronoi images using our algorithm with a scrambling step that removes any spatial information of the original image. The model can be used to provide spatial information about objects of interest, such as cell surfaces covered in oligo-tagged antibodies. The size of the image can be placed in experimentally relevant terms by considering the work of Rodriques et al. (20), who used circular disks of barcoded 10- μ m beads to capture transcriptomic data from tissue slices. The model can be compared to other approximation approaches, as shown in SI Appendix, Figs. S5 and S6.

4.1.4 Assessing distortion and precision of image recontruction

Finally we discussed about how to assess the distortion and precision of th reconstruction method. We defined distortion as the net displacement between the original and inferred positions of corresponding points, which is minimized by applying a linear transform to the set of reconstructed points. We generated 2D histograms of distortion as a function of position in the region of interest, and found that increasing the polony density reduces average distortion, while changes in site density have a negligible effect except at low densities near the point of network disconnection.

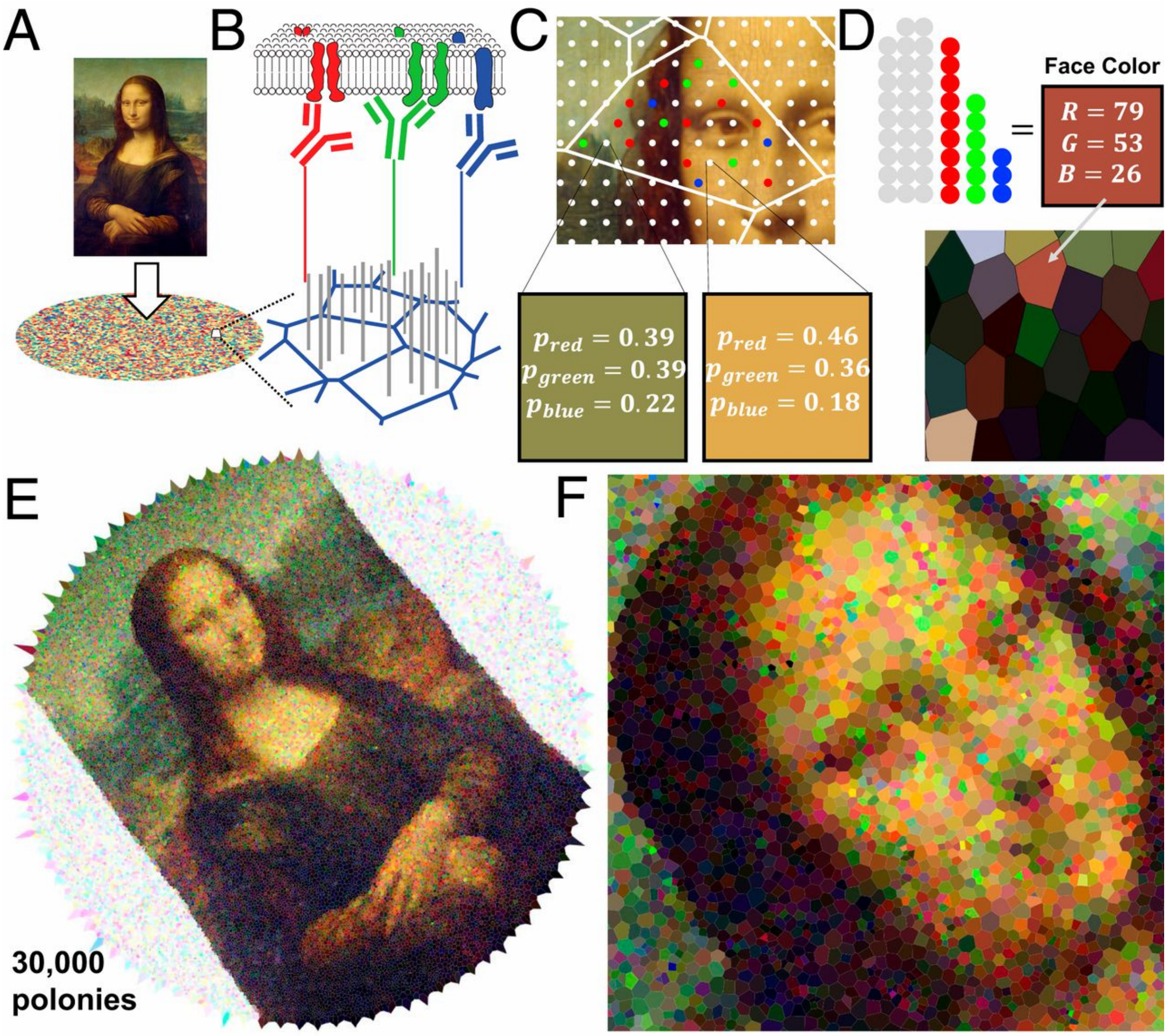

Figure14 Voronoi image formation (A) shows an image overlaid on a surface of primer sites. (B) depicts molecular markers representing different targets (red, green, and blue) being covalently linked to polony barcodes.(C) explains the Monte Carlo sampling method used to determine whether a marker is associated with a given site and if so, which target. This is done by taking the probability from the RGB value normalized to 1 at the corresponding position in the image.(D) shows how the tallying of markers and empty sites within a polony/Voronoi cell determines the color and brightness of that "pixel." A subsequent image is formed by coloring each cell accordingly.(E) displays a larger-scale reconstruction from scrambled edge data using the Tutte embedding approach with 30,000 polonies. Finally, (F) is a close-up of (E) revealing individual Voronoi pixels. This approach is adapted from Leonardo da Vinci's Mona Lisa painting.

We also used Levenshtein distance as a metric to characterize reconstruction quality, which measures the number of edits needed to make two graphs identical. We found that this metric weakly but positively correlates with distortion and grows linearly with polony density, but is relatively constant as a function of site density except at very low densities. Additionally, we measured classical resolution, the full width at half maximum (FWHM) of a point-spread function, and find that it is proportional to $1/\sqrt{\lambda}$, indicating that increasing polony density can improve resolution.

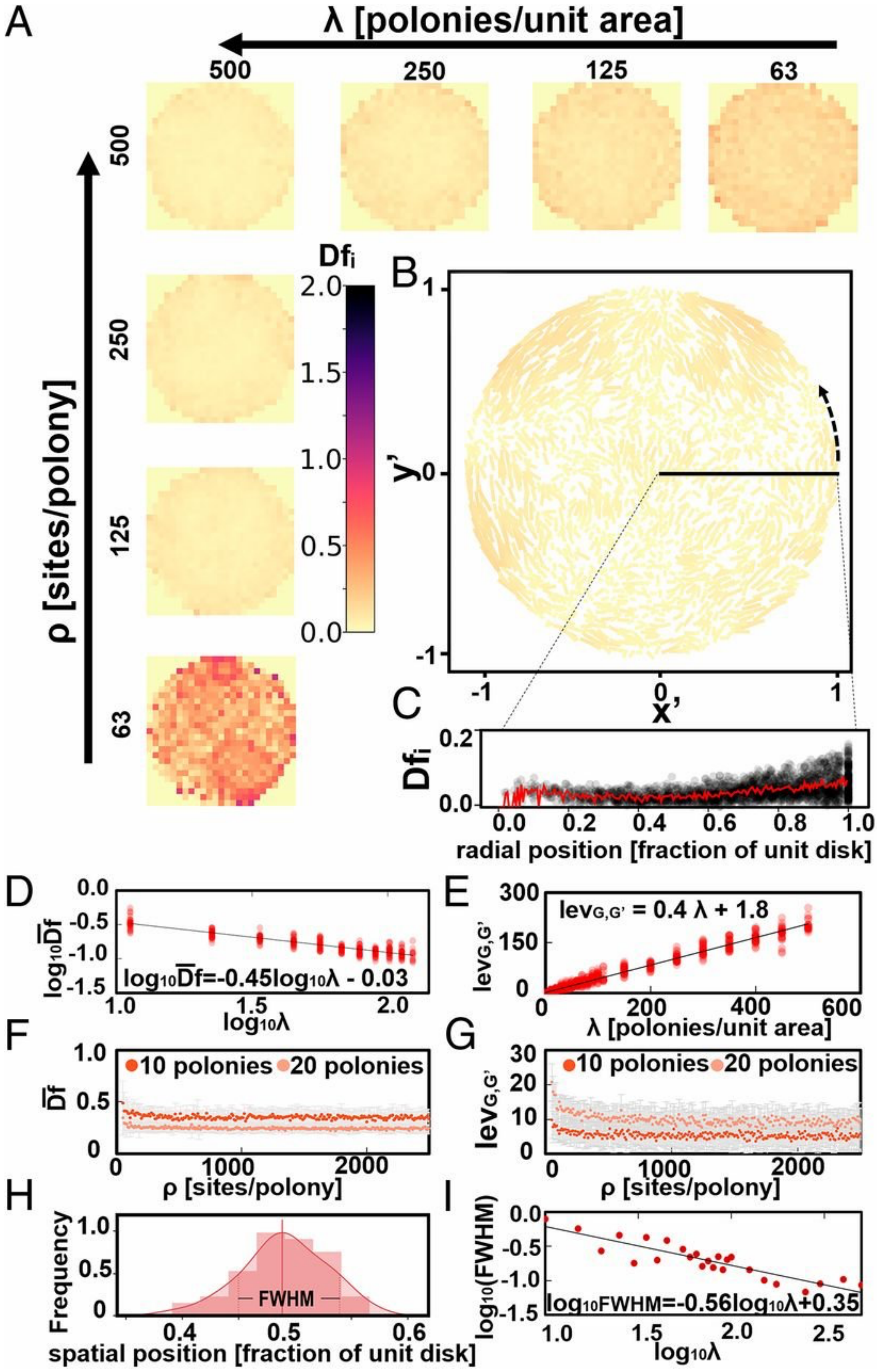

Figure15 Reconstruction quality (A) Two-dimensional histograms of average displacement values binned by relative position in the unit disk (n = 5,000 simulations per

histogram) for various parameter values of polony density (denoted by ρ) and relative site density (denoted by α).(B) Distortion in a single 2,000-polony Tutte embedding, with lines connecting the original and reconstructed vertex locations. The color map indicates the length of each line, with a maximum value of 2.0 (the diameter of the unit disk).(C) The radial profile of the distortion shown in panel B, with a 5-point moving average (in red).(D) A log-log plot of the average displacement as a function of polony density (points represent individual simulation reconstructions), with fixed α = 500 sites per polony. The dashed line indicates displacement proportional to $\sqrt{\rho}$.(E) A linear plot of the Levenshtein distance (lev(G,G0)) between the untethered and reconstructed Delaunay graphs, as a function of polony density.(F) Average displacement as a function of polony density for two values of α. Error bars represent standard deviation (n = 25 simulations per point).(G) Levenshtein distance as a function of polony density for two values of α. Error bars represent standard deviation (n = 25 simulations per point).(H) A single instance of the full width at half maximum (FWHM) of the reconstructed point spread function for a single site.(I) A log-log plot of the FWHM as a function of polony density, scaling approximately according to the negative square root of the polony density.

4.2 Paper II

4.2.1 Using SPT for folding ad purification of DNA origami

In Paper II, we described a method for folding DNA origami on magnetic beads using solid phase synthesis. The reaction mixture consists of scaffold, staple strands, magnetic beads, and a salt, SPT, which helps keep the magnetic beads in suspension. After thermal annealing, the DNA origami is formed on the magnetic beads and excess staples and unbound DNA origami are removed by washing. The folded DNA origami is then eluted from the magnetic beads using toehold mediated strand displacement. The efficiency of the folding process is tested under different SPT concentrations, and it is found that the folding occurs in all concentrations without the need for magnesium salts, but the most efficient folding occurs between 250-280 mM SPT. The purity of the final product is confirmed by running agarose gels and comparing the relative intensity of each fraction to a reference structure. The method offers a simple and efficient way to fold DNA origami with high purity, which can be used in a variety of applications.

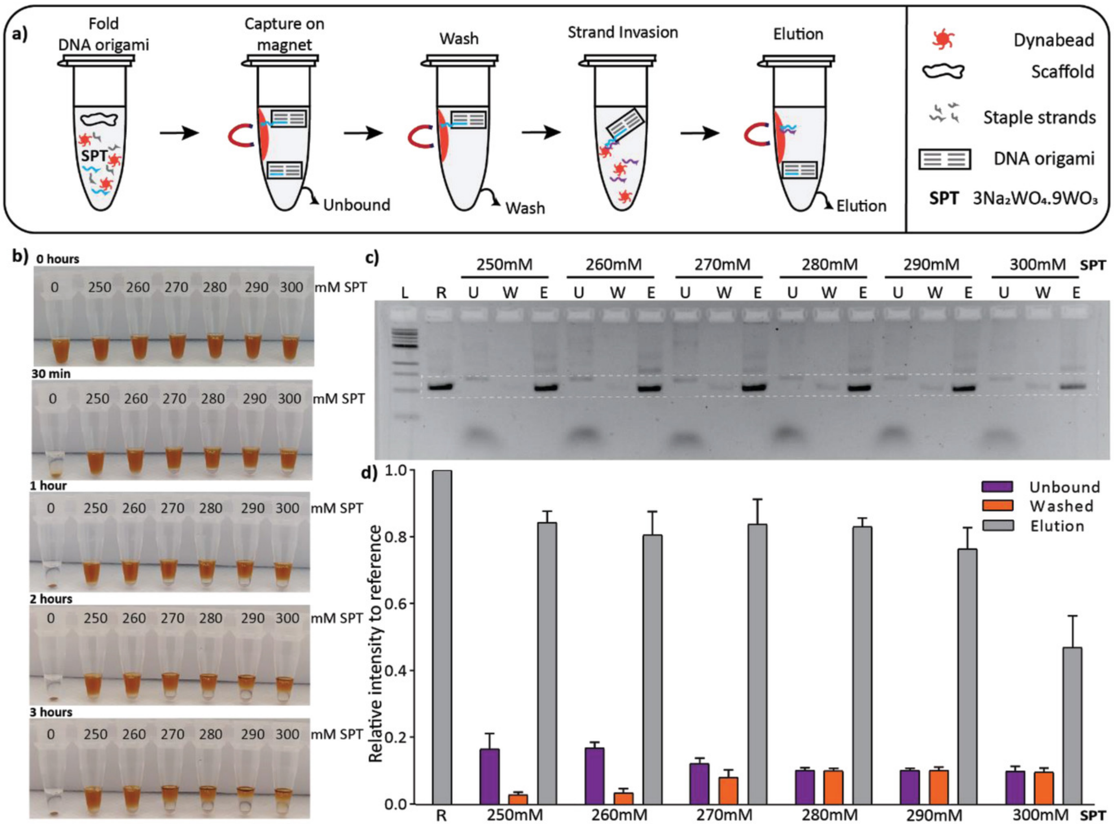

Figure16 (a)The process of folding DNA origami on magnetic beads in suspension using SPT. To initiate folding, magnetic beads, scaffold, staple strands, and SPT were mixed in Tris-EDTA buffer. Once folded, a magnet was used to capture the magnetic beads with the origami, and any unbound origami and excess staples were removed (Unbound). The beads were then washed with buffer to ensure the removal of any remaining excess material (Wash). Next, the DNA origami captured on the magnetic beads was resuspended in elution buffer containing invasion strands. By using a magnet, only the beads were captured, leaving the eluted DNA origami in solution (Elution). In panel (b), images show magnetic beads re-suspended in 0 mm and 250-300 mm SPT on folding buffer at different time points up to 3 h. Panel (c) displays an agarose gel of an 18HB DNA origami folded in various concentrations of SPT (250-300 mm). The ladder 1 kb (L), reference structures (R), and fractions of unbound (U), wash (W), and elution (E) steps ran on agarose gel. The intensity of each band inside the white dashed box in Figure 2c was measured and compared to a reference 18HB structure folded with a standard procedure. Standard deviations from 4 individual experiments are shown in the bar plots in panel (d).

4.2.2 Optimization of the solid phase synthesis protocol for DNA origami

We next optimized the solid-phase synthesis technique for high folding yield in the shortest possible time. Five different factors were investigated, including the addition of MgCl2, linker concentration, invader concentration, invader time, and the ratio between the amount of scaffold and the amount of magnetic beads. The addition of MgCl2 did not result in the highest folding yield, and 5 to 10 times excess of linker concentration to scaffold concentration gave the highest folding yield. The invader concentration was optimized to 25 times excess to scaffold concentration, and a 1-hour invader time was found to be sufficient. The ratio between the amount of scaffold and the amount of magnetic beads was optimized, and a scaffold concentration of 5.83 ng/μL bead was found to result in up to 90% folding yield. The use of magnetic beads for both folding and subsequent multimerization could facilitate the production of DNA origami superstructures. The regeneration of magnetic beads after elution was also explored. Finally, the successful production of the correct shape was confirmed by negative stain transmission electron microscopy for the reference and elution from the solid-phase technique.

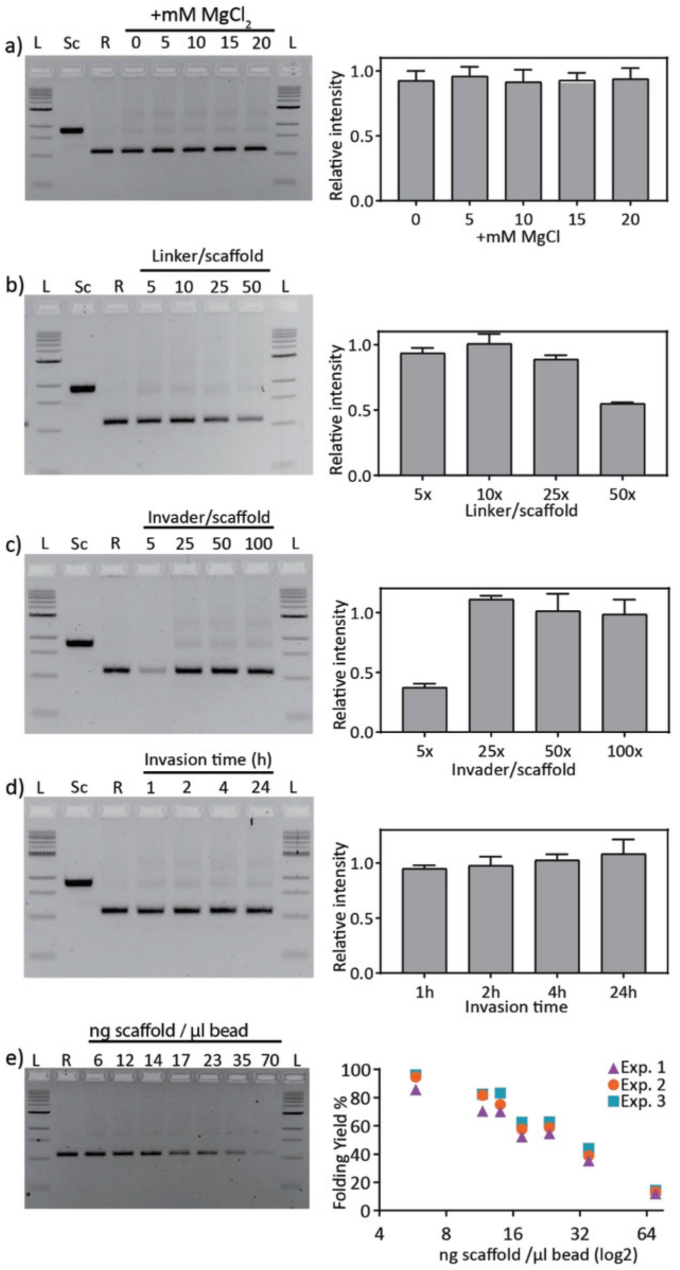

Figure 17. The solid-phase synthesis of DNA origami was optimized by testing five factors on the 18HB structure, and fractions of elution samples were run on agarose gel. The intensity of the bands was measured and plotted relative to the reference structure. In the agarose gels, the following were run: 1 kb ladder (L), scaffold (Sc), reference structure (R), and fractions of solid-phase synthesis samples for each experiment. The experiments included: a) SPT in folding buffer supplemented with 5–20 mm MgCl2 during solid-phase synthesis. b) Testing 5–50 times excess of Linker to scaffold concentration during folding. c) Using 5–100 times excess of invader to scaffold concentration during strand invasion step. d) Strand invasion was performed for 1–24 h after solid-phase synthesis. Standard deviations were shown in bar plots from three individual experiments. e) Testing 6 - 70 ng of scaffold per μL of magnetic beads during folding and yield was calculated from the intensity of the bands compared to the intensity of the reference product. Purple triangles, blue squares, and red dot symbols correspond to individual experiments.

The successful production of the correct shape was confirmed by negative stain transmission electron microscopy for the reference and elution from the solid-phase technique, and the regeneration of magnetic beads after elution was also explored.

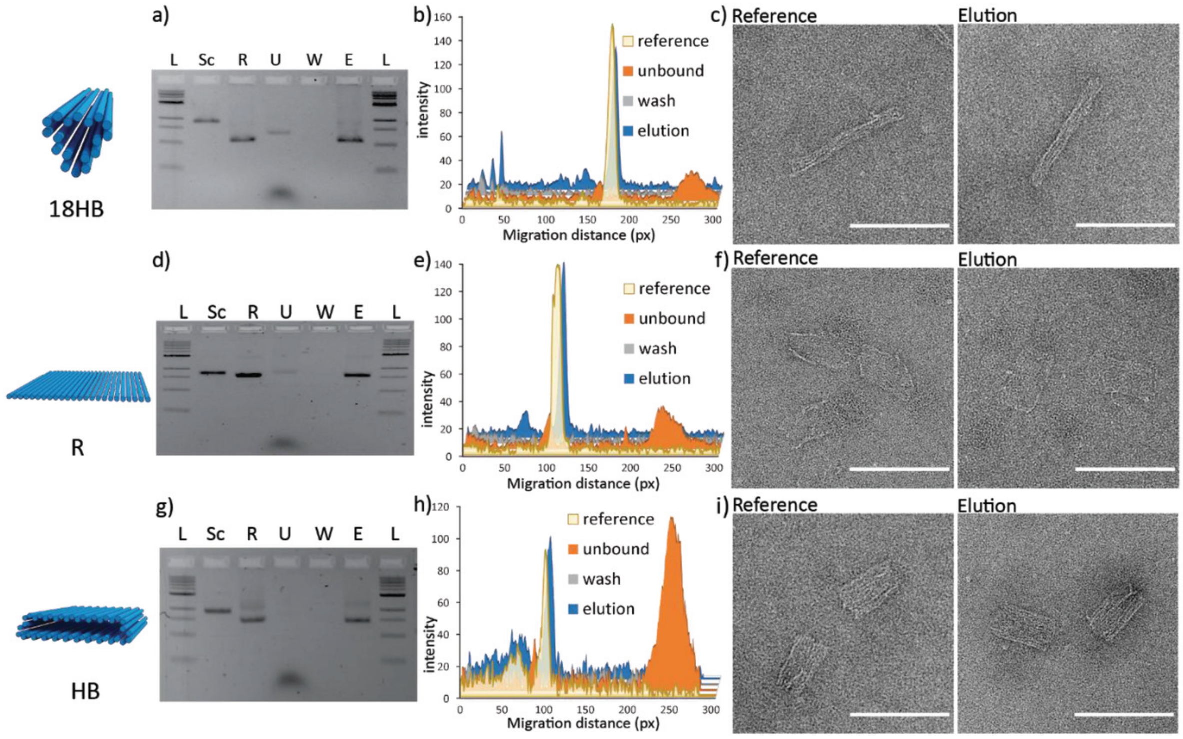

Figure 18. The experiment involved folding different structures using the solid-phase synthesis technique. The resulting agarose gels for 18HB, R, and HB structures displayed the migration of various samples, including 1 kb ladder (L), scaffold (Sc), reference (R), unbound (U), wash (W), and elution (E) samples. The migration distance of each lane from the agarose

gels was plotted in b, e, and h. Additionally, sample TEM pictures of reference structures and structures from the solid phase synthesis for each case were included in c, f, and i, with a scale bar of 100 nm.

4.2.3 using solid-phase synthesis for post-folding modification of DNA origami

We next investigated the use of solid-phase synthesis for post-folding modification of DNA-origami structures. We folded 18HB conjugated with two haptens (digoxigenin, DIG) using solid-phase technique and added a high-affinity rabbit anti-DIG IgG antibody to the solution in excess for 1 hour. After removing the excess unbound antibody, we observed similar mobilities of the same samples between the two different folding methods on an agarose gel. Using TEM, we confirmed the successful production of antibody-modified structures for both normal folded and solid-phase synthesized samples. Gel electrophoresis and TEM imaging also confirmed the successful removal of excess antibodies. We further demonstrated the ability to sequentially modify DNA origami with different functionalized molecules and remove excess modifications in a one-pot reaction. This technique has the potential to impact applications that study co-localization of different types of proteins on the same DNA origami.

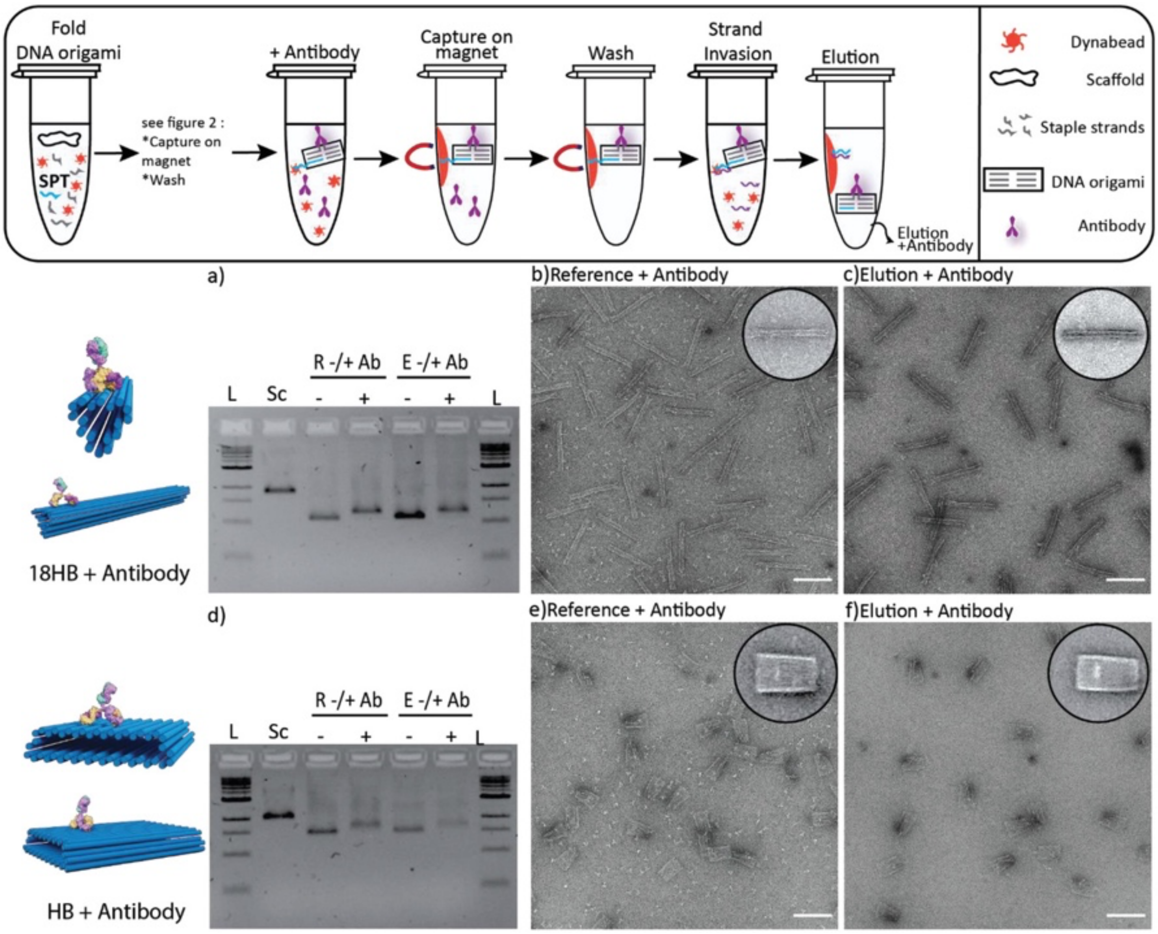

Figure 19. Solid-phase synthesis for post-folding production of DNA origami (a) shows agarose gel electrophoresis results for several samples, including a 1 kb ladder, scaffold (Sc), reference structure (R), reference structure with antibody (R+Ab), elution sample (E), and elution sample with antibody (E+Ab) for the 18HB structure. (b) and (c) show transmission electron microscopy (TEM) images of reference nanostructures and eluted samples modified with antibodies, respectively, for the 18HB structure. (d) shows the same agarose gel electrophoresis results as panel (a), but for HB structures. (e) and (f) show TEM images of reference nanostructures and eluted samples modified with antibodies, respectively, for HB structures. The upper right corner of all TEM images includes a class average from 100 TEM micrographs, and the scale bar in the HB images is 100 nm.

4.3 Paper III

4.3.1 Validate the function of chimeric guide RNA

We report on the use of the CRISPR-Cas9 system as a powerful genome editing tool that utilizes a single guide RNA (sgRNA) to direct the Cas9 endonuclease to a specific DNA target. In our study, we engineered a dual guide RNA (dgRNA) with two guide sequences linked by a poly-adenine sequence to facilitate binding to two distinct DNA targets. We have demonstrated that the dgRNA can recruit dCas9 to form an active ribonucleoprotein (RNP) complex and bind to both DNA targets simultaneously, as shown by electrophoresis mobility shift assay (EMSA) and negative-stain transmission electron microscopy (NS-TEM).

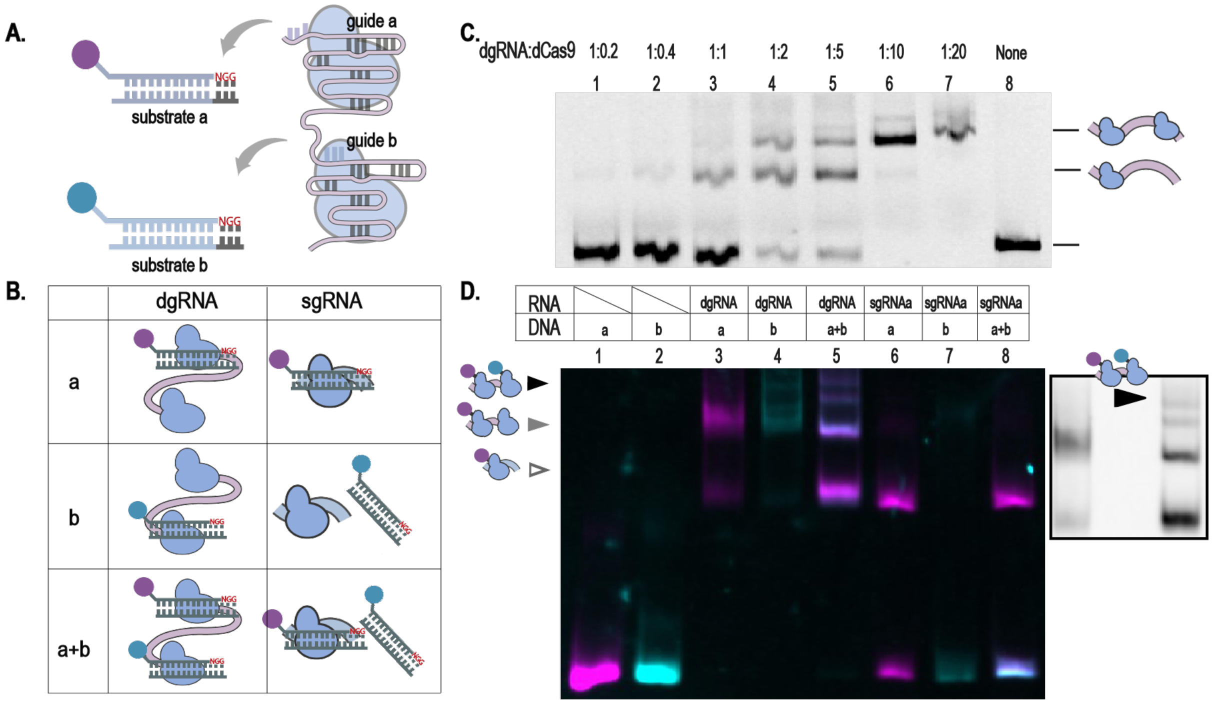

Figure 20. dgRNA demonstrates sequence bi-specificity for DNA targets Illustration of dgRNA guiding dCas9 to recognize separate DNA targets.(A)dgRNA consists of two sgRNA, with 5′ end-proximal guide sequence referred as 'guide a' and 3′ end-proximal guide referred as 'guide b'. (B)Two separate DNA substrates substrate 'a' and substrate 'b' were used to keep track of dgRNP-DNA complex formation.(C)Electrophoretic mobility shift assay measuring the titration of dCas9 over dgRNP while binding to a Cy5-labled single DNA substrate.(D)Electrophoretic mobility shift assay visualizing dgRNP binding to two separate DNA targets simultaneously.(E)dgRNP formed complex with each DNA substrate 'a' or 'b' and when incubated with both DNA substrates, additional gel bands formed showingli further reduced mobility with overlapping fluorescence from Cy5 and Cy3. Products were visualized using native PAGE gel supplemented with 5mM Mg2+.

Furthermore, we have shown that the dgRNA can induce DNA loop formation at targeted sites, as observed by atomic force microscopy (AFM). These results indicate the programmability of dgRNA to induce DNA loop and demonstrate its potential as a versatile tool for genome editing and manipulation.

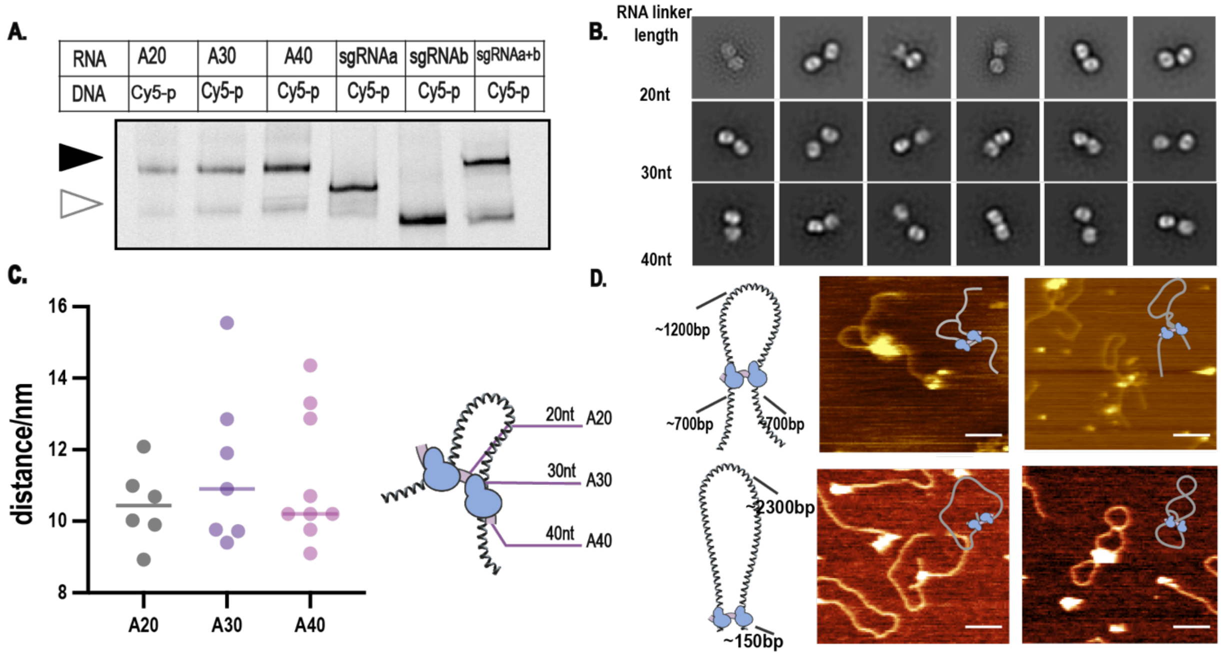

Figure 21. (A)dgRNP binding to DNA substrates with varied linker length and formation of RNP-bound complex. sgRNA-a and sgRNA-b also bind to DNA substrates.(B)Purified dgRNP-bound to DNA observed through 2D class averages from negative stain-TEM. Each class shows density from two dCas9, with undetectable signal from RNA and DNA.(C)Distribution of distance between two dCas9 in formed complex shown, with mean value indicated by horizontal line.(D)dgRNP targeting at different pairs of sites on the same

DNA substrate, forming DNA loops as observed through AFM images. White triangles indicate dgRNP, yellow arrows indicate formed DNA loop. Scale bars, 100nm.

4.3.2 Using dgRNA in *E.Coli* to recapitulate DNA loop induction mediated by Lac repressor

We next tested whether we could induce loop formation and regulate gene expression in a cellular environment by using the Lac operon in *E.Coli* as a canonical DNA loop model. We co-expressed dCas9 and dgRNA targeting neighbour sites of two auxiliary operators, O2 and O3, on either the template or non-template strand of LacZ gene. We also constructed *E.Coli* strains expressing sgRNA targeting the same neighbour sites of O2 or O3 operators. Our results showed that dgRNP could recapitulate the Lac repressor-mediated LacZ repression by inducing DNA loop formation between targeted sites. The repression rate of LacZ was sustained for up to 8 hours by dgRNP compared to sgRNP, which showed a synchronized fold change of LacZ expression with or without dCas9 induction. We hypothesize that the potent repression by dgRNP is due to the formation of a more stable complex with genomic DNA via a loop-mediated mechanism, enabling higher resistance to destabilization factors such as RNA polymerase by modifying local genomic topology.

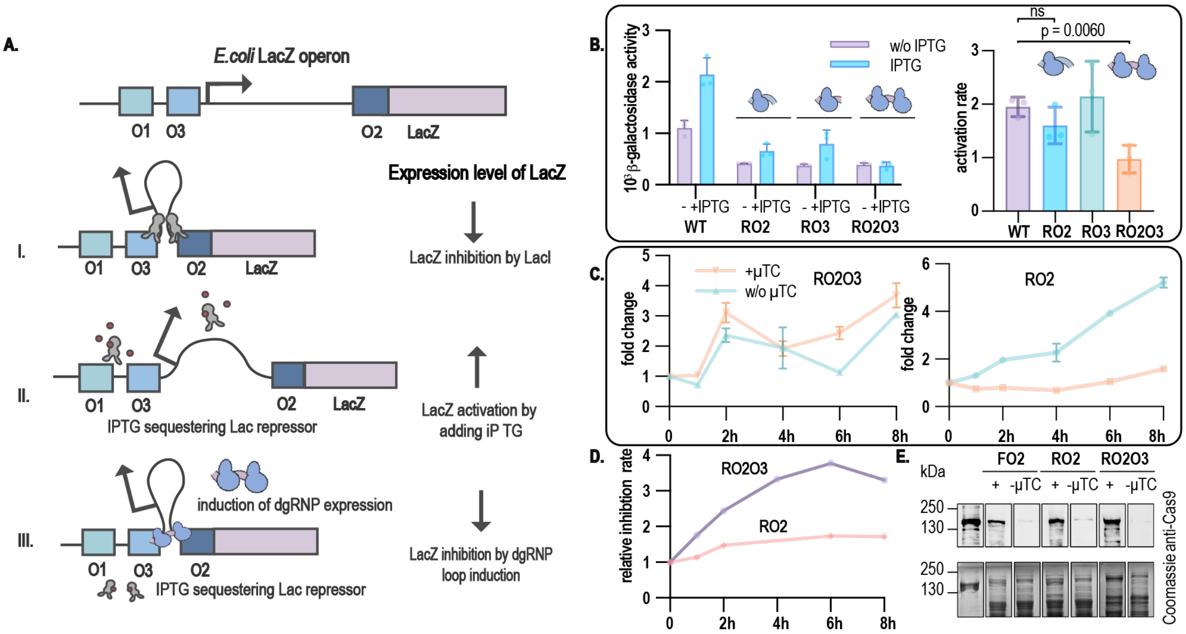

Figure 22. (A)Schematic illustration of LacZ regulation in *E.Coli* and by dgRNP. (B) β -galactosidase activity measurement in *E.Coli* expressing guide RNA targeting operator O2 and O3.(C)Time-course measurement of β -galactosidase activity fold change in constructed *E.Coli* strains after adding IPTG.(D)Calculated inhibition rate by RNP during 8h-time course in constructed *E.Coli* strains after adding IPTG.(E)Western blot of dCas9 in constructed strains

4.3.3 Effect of distance between target sites on dgRNA-mediated loop induction

Continuing the functional test using LacZ as reporter, we tested the effect of distance between target sites on DNA looping efficiency mediated by dgRNP. We designed dgRNAs targeting O2 and five distal target sites (O4-O8) located at increasing distances from O2. We observed that all dgRNP constructs showed higher repression rates of LacZ compared to the sgRNP control. However, we found that the repression rate was not determined solely by the distance between the target sites. For instance, the repression rate of LacZ was higher for the dgRNA-RO2O8 construct with target sites spaced more than 4500 bp apart than for the dgRNA-RO2O3 construct, with target sites located less than 500 bp apart. Additionally, the repression rate of LacZ for dgRNA-RO2O7, which had a target site located 4000 bp away from O2 and less than 500 bp away from O8, was significantly lower than for dgRNP-RO2O8. These observations suggest that factors other than distance, such as altered local topology of the genomic DNA at target sites, can affect the accessibility of targets to dgRNP and thus affect inhibition rates. Therefore, we conclude that the efficiency of loop formation mediated by dgRNP in vivo was confounded by the altered local genomic context when varying the distance between two target sites.

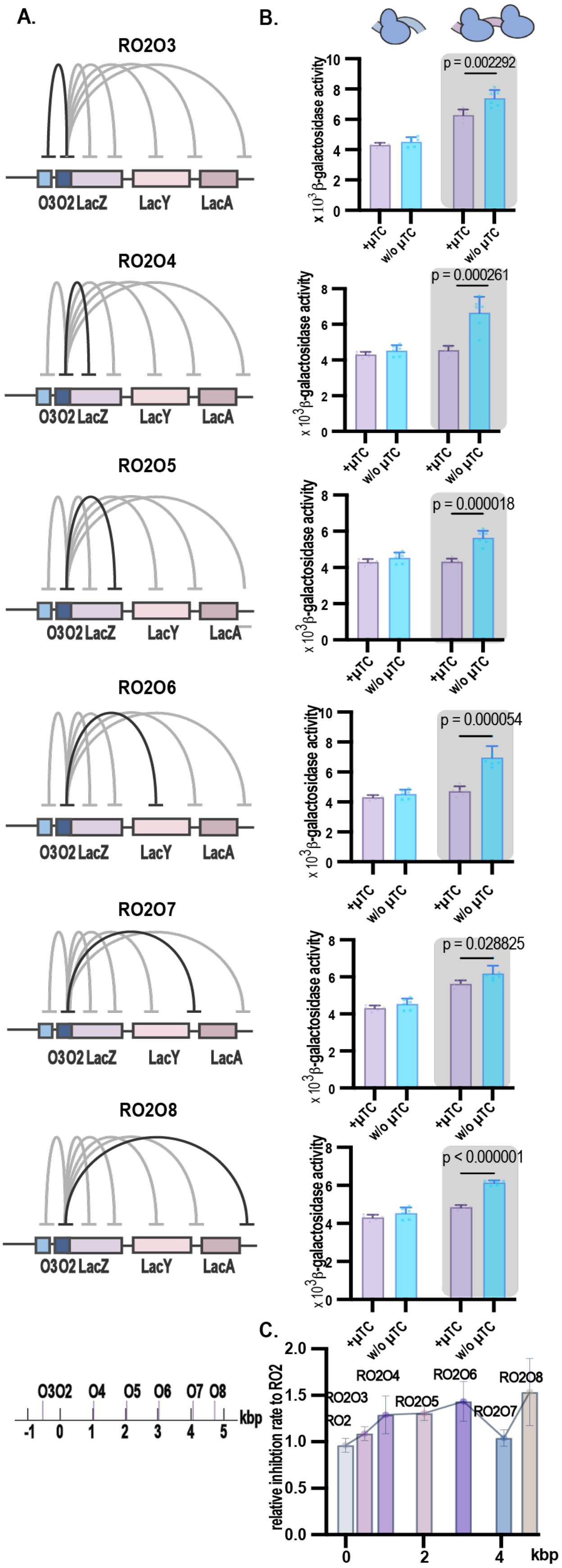

Figure 23. (A)Schematic representations of DNA target sites location at LacZYA locus in each *E.coli* strain. Lowest panel shows the distance of each target site related to O2. Target

sites are termed as O3-O8, with ascending number suggesting longer distance from O2. The pair of targeted sites were connected by black curve.. (B)Measurements of β - galactosidase activity showing varied repression level of LacZ from strains with varied targeting sites. Each plot shows measurements of strain expressing corresponding dgRNA illustrated in (A) against strain expressing sgRNP-RO2. Mean value and SD of biological replicates are plotted(n=6).(C)Calculated inhibition rate of dgRNP plotted against the genomic distance between target sites. Mean value and SD of biological triplicates were included.

4.4 Paper IV

4.4.1 Production and validation of oligonucleotide-conjugated Cas9

Our study aimed to create Cas9-DNA tags for proximity ligation assays. We produced Cas9 with Sortase A recognition tag at the C-terminus of Cas9 and then conjugated it with azide-modified oligonucleotide tags using copper-free click chemistry. The resulting CasT proteins were purified using anion-exchange chromatography and validated using Cy5-labelled oligos complementary to the oligonucleotide tag. We then demonstrated the sequence-specific nuclease activity of CasT conjugates using sgRNA-directed digestion of pUC19.

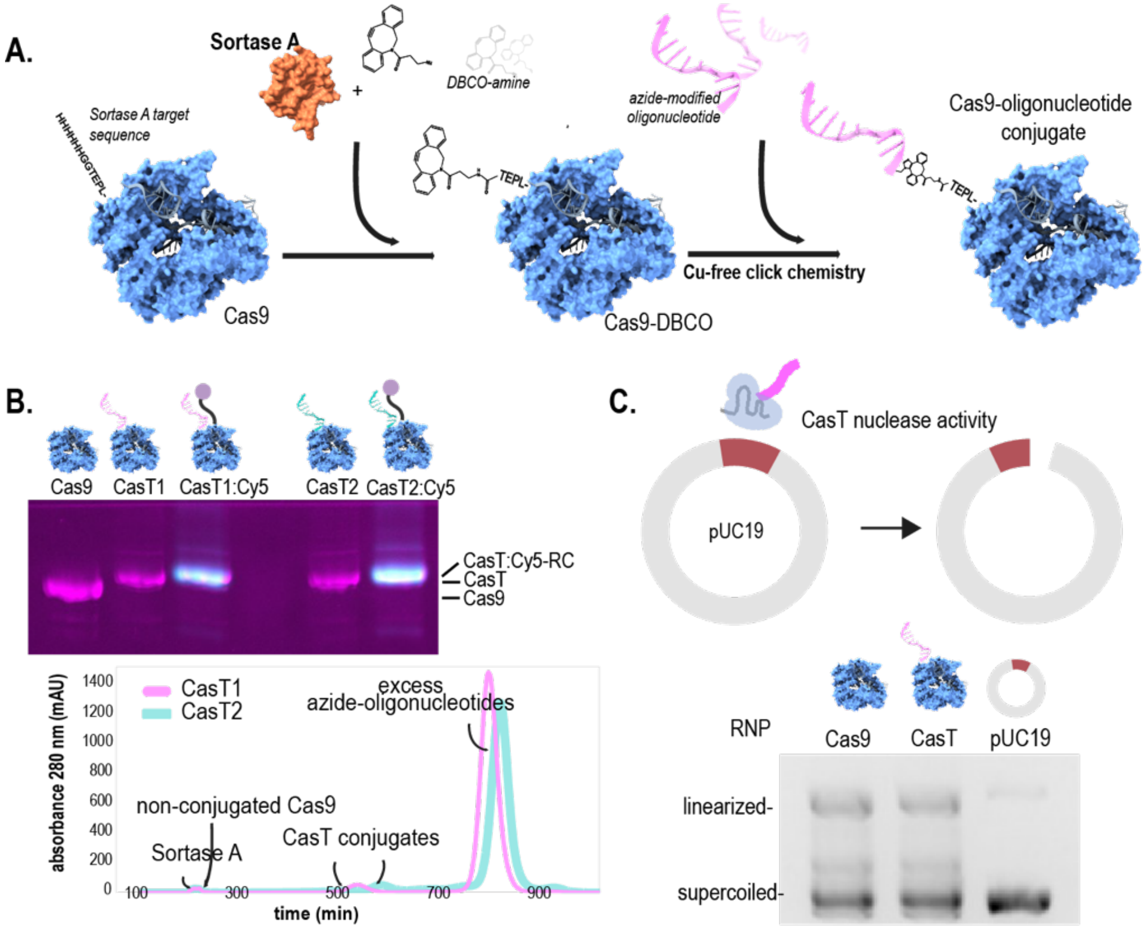

Figure 24. (A) Cas9 is produced with a Sortase A recognition tag at its C-terminus, which enables Sortase A to catalyze the formation of a peptide bond between threonine and amine-modified DBCO. The DBCO moiety on Cas9 can then react with an azide-modified oligonucleotide through copper-free click chemistry, resulting in the formation of the conjugated Cas9, referred to as CasT protein.(B) The lower panel shows the chromatogram of the Sortase A-mediated Cas9-oligonucleotide conjugation reaction illustrated in (A), separated by anion-exchange chromatography. The upper panel shows Cas9 and purified CasT conjugates separated by 10% SDS-PAGE. CasT1 represents Cas9 conjugated with oligo tag 1, while CasT2 represents Cas9 conjugated with oligo tag 2. CasT: Cy5-RC represents CasT conjugates hybridized to Cy5-labelled complementary oligo. The magenta color shows Sypro Orange staining, while the cyan color shows the Cy5 signal.(C) CasT nuclease activity was characterized by digesting pUC19. Equal amounts of Cas9 and CasT were directed to cut pUC19 using the same sgRNA, and the resulting products were resolved by 2% agarose gel.

To validate the feasibility of CasT-directed proximity ligation assay, we used DNA origami as a reaction substrate to place a pair of DNA targets with varying distances. CasT1 and CasT2 were assembled with truncated sgRNA recognizing each of the two DNA duplex, and two bridge oligos were added followed by T4 DNA ligase-mediated DNA ligation reaction. We then initiated the RCA reaction and added Cy5 probes complementary to RCA products. Gel electrophoresis results showed Cy5-labelled RCA products only from samples with both probes, validating the general feasibility of CasT-PLA in solution. Overall, we successfully demonstrated the creation of Cas9-DNA tags for proximity ligation assays and validated its sequence-specific nuclease activity and feasibility for proximity ligation assays.

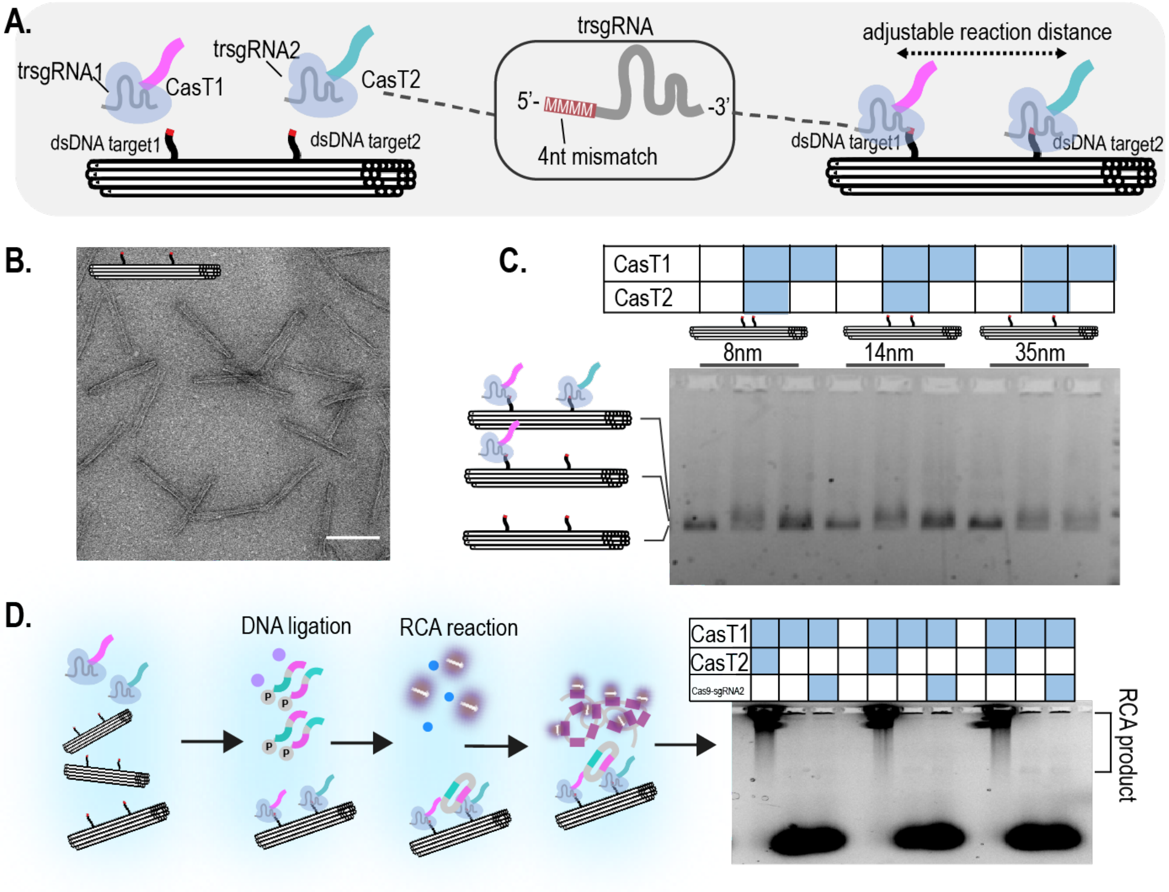

Figure 25. (A)Schematic illustration of arranging CasT probes in solution with DNA origami placing a pair of double-stranded DNA targets protruding at designed locations. CasT1 and CasT2 are paired with trsgRNA, each targeting one protrusion site; trsgRNA is sgRNA with 4nt mismatch to target DNA at 5′ end of sgRNA. The distance between CasT proteins is adjusted by designing target sites at different locations of DNA origami, which recruits CasT by specific target binding.(B)Characterization of the DNA origami used for placing CasT probes with negative-stain TEM. Scale bar, 100 μm.(C)Gel electrophoresis assay showing the binding of both CasT probes to DNA origami with varied distance between target sites.(D)Visualizing CasT-PLA products labelled by complementary Cy5-oligos with gel electrophoresis. Left panel shows the illustration of the stepwise CasT-PLA reaction in solution: Both CasT probes binding to DNA origami at target sites, followed by ligation of bridging oligos hybridizing to both tags to form circular single-stranded DNA; RCA reaction is subsequently initiated by adding phi29 polymerase and labeled by complementary cy5-labelled oligo. Right panel shows cy5 signal from reaction products resolved by 1.5% agarose gel. Three different distances (8nm,14nm,35nm) between target DNA were tested for the illustrated reactions. Lane 1,4,7 show products formed with both CasT probes added to three different DNA origami holder; lane 2,5,8 show products formed by adding only CasT1 probe; lane 3,6,9 show products formed by adding CasT1 probe and unconjugated Cas9 paired with trsgRNA recognizing target2.

4.4.2 Validating the CasT-PLA reaction with tandem repeats in human genome

The CasT-PLA assay was applied to target repetitive genomic loci in the human genome, specifically the exon 2 of the human mucin 4 gene (MUC4-E2) on chromosome 3, as a proof-of-principle for in situ reaction targeting of colocalization of genomic targets. Each MUC4-E2 locus consists of a 46bp sequence repeat with varied copy numbers ranging from 100 to 4008. CasT1 and CasT2 probes were programmed to target the repeats in tandem directed by the same trsgRNA, allowing pairs of probes to be colocalized within spatial proximity. The MUC4-E2 repeat region was further labeled with Cy5 puncta by CasT-PLA if the distance between two probes was below the distance tolerance of the assay. After CasT-PLA reaction, Cy5 puncta from RCA products were observed located within the nucleus of MCF-7 cells. The number of puncta per cell varied from 0 to 6, with the majority of cells containing 2-3 puncta. The specificity of CasT-PLA was demonstrated by using a nonsense trsgRNA (sg1a) for CasT2 while CasT1 still paired with trsgRNA targeting MUC4-E2 loci, which led to no RCA signal in cells. The assay was also applied to visualize another well-characterized variable number of tandem repeats (VNTR) on chromosome 1 encoding for mucin 1 protein. The cytoplasmic distribution of MUC1-PLA dots could be attributed to detection of MUC1 transcripts by simultaneous binding of both CasT probes to tandem repeats on MUC1 transcripts.

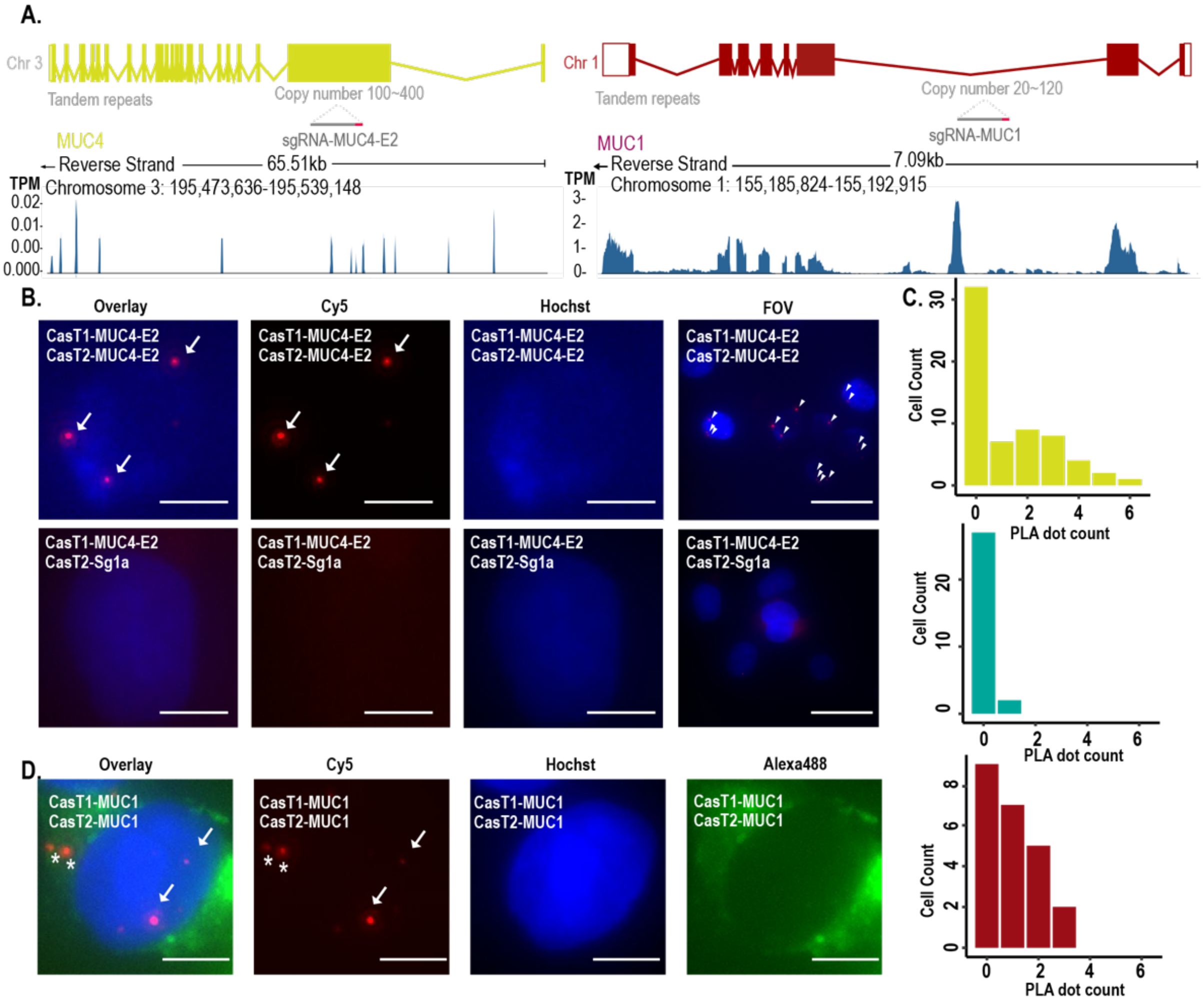

Figure 26. (A) The genomic region of MUC4 and MUC1 genes are schematically depicted, with highlighted sgRNA-targeted repeats. The RNA-seq data from MCF-7 from ENCODE project9 mapped to the same genomic region is shown as a reference.(B) The top panel shows CasT-PLA puncta labelling MUC4-E2 loci, while the lower panel shows the results from CasT2 targeted to a nonsense sequence. Arrows indicate PLA puncta, with a scale bar of 10 μ m for the inset and 50 μ m for the field-of-view.(C) Histogram of puncta per cell corresponds to the samples on the left in (B).(D) The VNTR region of MUC1 is targeted by CasT-PLA, with arrows showing puncta overlapping with nuclei and asterisks indicating puncta in the cytoplasm overlapping with cell membrane staining in the Alexa488 channel. A scale bar of 10 μ m is shown.

4.4.3 Directly targeting cis-regulatory element with target gene

We used a novel imaging technique, CasT-PLA, to visualize the interaction between cis-regulatory elements at the locus control region (LCR) and β -globin locus in our study. Our goal was to explore the role of the long-range chromatin interaction region LCR in downstream gene expression. Specifically, we focused on the enhancer HS2 and its regulation of downstream hemoglobin genes, including HBE and HBG genes. Using CasT-

PLA, we were able to visualize the interaction between HS2 and HBE genes, but not between HS2 and HBG genes, in MCF-7 cells that do not express these genes. Our findings suggest that the larger genomic distance between HS2 and HBG genes compared to the 11kb distance between HS2 and HBE genes may explain the even lower possibility of detecting CasT-PLA puncta between HS2 and HBG. Overall, our results demonstrate that the CasT-PLA assay has higher spatial sensitivity for distance between genomic loci in situ and is capable of resolving the interaction frequency across single cells, which can help us better understand the regulatory mechanisms of gene expression.

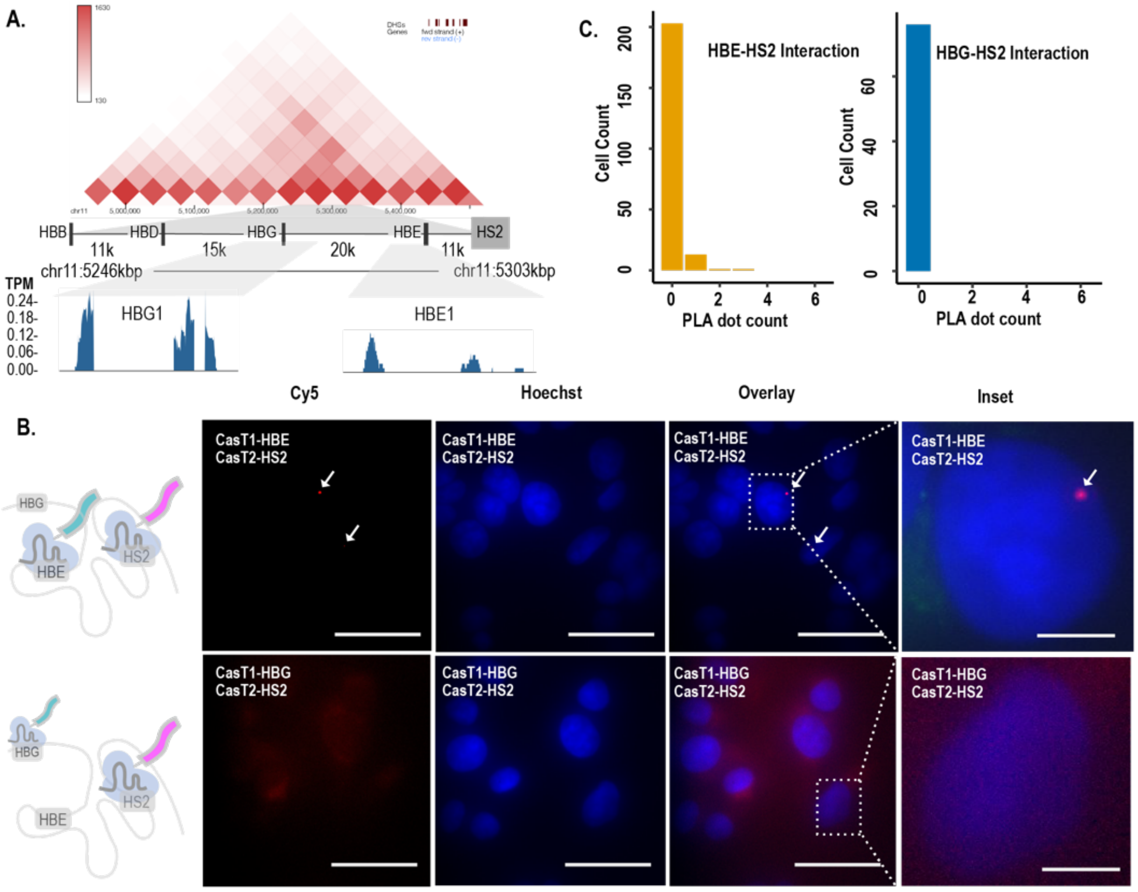

Figure 27. (A) The genomic regions of the β-globin gene are shown in schematics. The top panel displays Hi-C data from IMR90 cells with 40kb resolution at β-globin flanking regions16, while the lower panel shows RNA-seq reads mapped to the highlighted genomic regions9. Hemoglobin beta (HBB), hemoglobin delta (HBD), hemoglobin gamma (HBG), hemoglobin epsilon 1 (HBE), and hypersensitive region 2 (HS2) are indicated.(B) Simultaneous targeting of HS2 and HBE or HBG by CasT-PLA. The left panel displays schematics of hypothesized local chromatin organization. The top panel shows images of CasT-PLA targeting HS2 and HBE, and the lower panel shows CasT-PLA targeting HS2 and HBG. Arrows indicate PLA puncta, and the scale bar for the inset is 10 μm, while the scale bar for the field-of-view is 50 μm.(C) Histogram of CasT-PLA puncta per cell.

5 Conclusions

This work encompasses four projects that collectively showcase the bidirectional transfer and reshaping of spatial information by nucleic acids through various techniques. The projects cover a wide range of subjects, including DNA polonies on surfaces, customized DNA nanostructures, and genomic loci within cells.

Conclusion from Paper I: This work presents a comprehensive method for modeling polony saturation and reconstructing spatial information from sequencing data. The effectiveness of this method was validated by characterizing the distortion and precision of reconstruction quality.

Conclusion from PaperII: We have devised a method that is both efficient in terms of time and materials to produce functionalized DNA nanostructures. Our approach has enabled us to create DNA origami structures that are either bound with Antibody at specific locations or labeled with fluorophores. These structures have been successfully validated to demonstrate the efficacy of our method.

Conclusion from Paper. III: We redesigned the sgRNA in the CRISPR-dCas9 system to target two specific DNA sequences simultaneously, resulting in a new system capable of inducing DNA loop formation both in vitro and in vivo. To demonstrate its potential for modulating gene expression, we applied this system as a proof-of-concept and induced DNA loop formation in *E.Coli* at the LacZ operon, which recapitulated the repression effect from the Lac repressor.

Conclusion form Paper IV: We developed a method combining the specificity of dCas9 in searching genomic loci and high spatial sensitivity from proximity ligation assay for the detection of chromatin-chromatin interactions. To validate the efficacy of this technique, we labeled repetitive genomic loci in MCF-7 cells and subsequently expanded it to visualize cis-interactions in situ.

6 Points of perspective

High-throughput sequencing technique has greatly advanced our understanding in cell biology by transcriptomic profiling single-cell and other high-throughput functional genomic studies. However, traditional library preparation methods for sequencing do not retain spatial information related to cells and tissues. To overcome this limitation, spatial transcriptomic techniques have emerged, which enable transcriptomic profiling of single cells while preserving their spatial information. In our study (Paper I), we proposed a framework that addresses the current challenges associated with spatial transcriptomic techniques, including resolution and instrumentation dependency. By registering the spatial information in neighboring DNA barcodes, we were able to maintain the native spatial information and read it out in the same sequencing pipeline used for RNA sequencing. Our approach eliminates the need for advanced microscopy techniques but following standard biochemical library preparation procedure, making it ideal for diagnostic purposes in clinical settings.

For method development aiming to resolve spatial information, an efficient platform to gauging the distance could greatly facilitate the evaluation with high accuracy. DNA nanostructures with functionalized molecules have shown promise in this area, but limitations in their production and purification processes have restricted their application in broader research field. Our study, outlined in Paper II, presents a solution to this issue with the added advantages of high recovery rates, low input requirements, and a simple protocol that does not rely on specialized instruments. This newly developed technique enables researchers from diverse fields to utilize it for a range of applications that demand nanoscale spatial accuracy.

The advent of high-throughput sequencing techniques, such as Hic, has allowed for the mapping of genomic architecture with high resolution. However, understanding the functional implications of local genomic features remains limited due to a lack of available toolkits. In our work (Paper III), we developed a technique using a minimal CRISPR-dCas9 system without additional mediators, by introducing a bivalent gRNA. The simplicity of this technique makes it universally applicable to organisms compatible with canonical CRISPR-Cas9 technology, with the potential for in vitro delivery. By introducing de novo DNA interactions, this technique can be flexibly applied to study genomic function in various contexts across organisms, thereby deepening our understanding of genomic function and potentially leading to the discovery of novel drug targets and therapeutic options.

During the development of the technique presented in Paper III, it was necessary to directly assess the efficiency of DNA contact formation in cells. However, conventional chromatin conformational capture techniques require a lengthy protocol and a large number of input cells. Moreover, resolving chromatin-chromatin interactions within a single cell while preserving cellular context remains a significant challenge. In Paper IV, we addressed these challenges by introducing a time-efficient and cost-effective approach that is independent of super-resolution microscopy. Our flexible technique allows for the study of the correlation between DNA interactions and functional outcomes within a single cell, making it useful for both research and diagnostic purposes. The use of DNA oligos as information carriers also provides broader design space for microscopy-independent methods based on the current version of our technique.

7 Acknowledgements

So now I have come to the end of my PhD that began before the pandemic. The learning and researching experience became a bit fragmented during those days. But in the end with all the efforts dedicated, all the help I received, pieces came together. All these experiences and works, whether finished or unfinished, meant a lot to me and I will continue the exploration in the future, whether that be in academia or elsewhere.

At this point, I would like to express my sincere gratitude to my supervisor Björn, thank you for giving me the opportunity to join the lab as a master student and for providing this explorative environment for research and innovation. Your passion and curiosity in science and technology are always inspiring. If were not given the chance to join this lab, I would not imagine there is this fun space for biomedical research, and probably would not see myself working and trying to control these molecules all day long.

I would like to express my gratitude to Ian, my master's thesis supervisor, for showing me all the tricks needed to test my ideas with minimal resources, and for your insightful thinking on tech in general. Your guidance exceeded my expectations for a master's project. I would also like to thank Iris for all the help and discussion with my presumptuous ideas and questions over the years, as well as technical support in both work and life. I want to thank Ioanna for all the discussions about research and PhD study and beyond; Yang and Boxuan for help in the lab and for the relaxing chit-chat in Mandarin; Ferenc for engaging in fast-paced discussions that are always super rational and entertaining at the same time, and Anurupa for her help with administrative tasks and her extra energy in setting up all the lab events. My gratitude extends to all the other brilliant colleagues, including Ieva, Marco, Igor, Alexander, and former members Erik, Giulio, and Mino, for sharing their insights, passion, creativity, and occasional frustrations in our co-working environment.

And I want to thank 9B colleagues, Joel, Ting, Chris, Miina, Helene, Magda, Björn Reinius, Natalie, Anneke, Bernard and all other kind colleagues I don't immediately write down that I got help from during PhD study and work.

Then it comes to my family, my mom and dad. 这三年都没有回家，我们各自都经历了一些波折。感谢你们总是提醒我坦然面生活的起伏而不是急于否定一切过程。你们的支持与关心让我在面对实验室内外的困难时更加坦荡。

To all my friends outside of lab here in Sweden and remotely, I can't tell what's the concentration of PhD and work in our conversations. But all the struggling and inspiring life stories we have shared did become elements of my work presented here. Always grateful for the experiences we shared over these years.

8 References

1. WATSON, J. D. & CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

2. Waters, J. T. *et al.* Transitions of Double-Stranded DNA Between the A- and B-Forms. *J Phys Chem B* **120**, 8449–8456 (2016).

3. Seeman, N. C. Nucleic acid junctions and lattices. *J Theor Biol* **99**, 237–247 (1982).

4. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).

5. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).

6. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res* **37**, 5001–5006 (2009).

7. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).

8. Geary, C., Grossi, G., McRae, E. K. S., Rothemund, P. W. K. & Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nat Chem* **13**, 549–558 (2021).

9. Shaw, A. *et al.* Spatial control of membrane receptor function using ligand nanocalipers. *Nat Methods* **11**, 841–846 (2014).

10. Shaw, A. *et al.* Binding to nanopatterned antigens is dominated by the spatial tolerance of antibodies. *Nat Nanotechnol* **14**, 184–190 (2019).

11. Zhao, Y.-X. *et al.* DNA Origami Delivery System for Cancer Therapy with Tunable Release Properties. *ACS Nano* **6**, 8684–8691 (2012).

12. Li, S. *et al.* A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger in vivo. *Nat Biotechnol* **36**, 258–264 (2018).

13. Lin-Shiao, E. *et al.* CRISPR–Cas9-mediated nuclear transport and genomic integration of nanostructured genes in human primary cells. *Nucleic Acids Res* **50**, 1256–1268 (2022).

14. Pothoulakis, G., Nguyen, M. T. A. & Andersen, E. S. Utilizing RNA origami scaffolds in Saccharomyces cerevisiae for dCas9-mediated transcriptional control. *Nucleic Acids Res* **50**, 7176–7187 (2022).

15. Wu, X. *et al.* An RNA/DNA hybrid origami-based nanoplatform for efficient gene therapy. *Nanoscale* **13**, 12848–12853 (2021).

16. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, 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).

17. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**, 174–182 (2005).

18. Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim. & Schouls, Leo. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565–1575 (2002).

19. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (1979)* **315**, 1709–1712 (2007).

20. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* **39**, 9275–9282 (2011).

21. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)* **337**, 816–821 (2012).

22. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science (1979)* **351**, 867–871 (2016).

23. Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science (1979)* **343**, 1247997 (2014).

24. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).

25. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (1979)* **351**, 84–88 (2016).

26. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

27. Jones, S. K. *et al.* Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat Biotechnol* **39**, 84–93 (2021).

28. Boyle, E. A. *et al.* High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proceedings of the National Academy of Sciences* **114**, 5461–5466 (2017).

29. Bravo, J. P. K. *et al.* Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **603**, 343–347 (2022).

30. Nguyen, D. N. *et al.* Polymer-stabilized Cas9 nanoparticles and modified repair templates increase genome editing efficiency. *Nat Biotechnol* **38**, 44–49 (2020).

31. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

32. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (1979)* **368**, 290–296 (2020).

33. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

34. Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143–1149 (2015).

35. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).

36. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160**, 339–350 (2015).

37. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635–646 (2014).

38. Kim, J. H. *et al.* LADL: light-activated dynamic looping for endogenous gene expression control. *Nat Methods* **16**, 633–639 (2019).

39. Hao, N., Shearwin, K. E. & Dodd, I. B. Programmable DNA looping using engineered bivalent dCas9 complexes. *Nat Commun* **8**, 1628 (2017).

40. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491 (2013).

41. Hong, Y., Lu, G., Duan, J., Liu, W. & Zhang, Y. Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. *Genome Biol* **19**, 39 (2018).

42. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proceedings of the National Academy of Sciences* **112**, 11870–11875 (2015).

43. Wang, M. *et al.* RCasFISH: CRISPR/dCas9-Mediated in Situ Imaging of mRNA Transcripts in Fixed Cells and Tissues. *Anal Chem* **92**, 2468–2475 (2020).

44. Zhang, K. *et al.* Direct Visualization of Single-Nucleotide Variation in mtDNA Using a CRISPR/Cas9-Mediated Proximity Ligation Assay. *J Am Chem Soc* **140**, 11293–11301 (2018).

45. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science (1979)* **366**, 1338–1345 (2019).

46. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).

47. Ahn, J. H. *et al.* Phase separation drives aberrant chromatin looping and cancer development. *Nature* **595**, 591–595 (2021).

48. Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* **114**, 2262–2270 (2018).

49. Kim, S. & Shendure, J. Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol Cell* **76**, 306–319 (2019).

50. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science (1979)* **347**, 1017–1021 (2015).

51. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

52. Espinola, S. M. *et al.* Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nat Genet* **53**, 477–486 (2021).

53. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216-228.e19 (2017).

54. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377-390.e19 (2019).

55. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510–517 (2015).

56. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science (1979)* **295**, 1306–1311 (2002).

57. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science (1979)* **326**, 289–293 (2009).

58. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119 (2015).

59. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919–922 (2016).

60. Quinodoz, S. A. *et al.* SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nat Protoc* **17**, 36–75 (2022).

61. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).

62. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659.e26 (2020).

63. Beliveau, B. J. *et al.* Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat Commun* **6**, 7147 (2015).

64. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophys J* **91**, 4258–4272 (2006).

65. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793–796 (2006).

66. Fullwood, M. J. *et al.* An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).

67. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

68. Mastronarde, D. N. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microscopy and Microanalysis* **9**, 1182–1183 (2003).

69. Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).