



## Sequence-specific DNA labelling for fluorescence microscopy

Shalini Pradhan <sup>1</sup>, Sinem Apaydin <sup>1</sup>, Jonas Bucevičius <sup>1</sup>, Rūta Gerasimaitė <sup>1</sup>, Georgij Kostiuk <sup>1</sup>, Gražvydas Lukinavičius <sup>\* 1</sup>

*Chromatin Labelling and Imaging Group, Department of NanoBiophotonics, Max Planck Institute for Multidisciplinary Sciences, Am Fassberg 11, 37077, Göttingen, Germany*



### ARTICLE INFO

**Keywords:**

Chromatin imaging  
Polyamides  
Triplex forming oligonucleotides  
Transcription activator-like effectors  
dCas9  
DNA methyltransferases

### ABSTRACT

The preservation of nucleus structure during microscopy imaging is a top priority for understanding chromatin organization, genome dynamics, and gene expression regulation. In this review, we summarize the sequence-specific DNA labelling methods that can be used for imaging in fixed and/or living cells without harsh treatment and DNA denaturation: (i) hairpin polyamides, (ii) triplex-forming oligonucleotides, (iii) dCas9 proteins, (iv) transcription activator-like effectors (TALEs) and (v) DNA methyltransferases (MTases). All these techniques are capable of identifying repetitive DNA loci and robust probes are available for telomeres and centromeres, but visualizing single-copy sequences is still challenging. In our futuristic vision, we see gradual replacement of the historically important fluorescence *in situ* hybridization (FISH) by less invasive and non-destructive methods compatible with live cell imaging. Combined with super-resolution fluorescence microscopy, these methods will open the possibility to look into unperturbed structure and dynamics of chromatin in living cells, tissues and whole organisms.

### 1. Introduction

More than 300 years ago, the nucleus was first described by Antonie van Leeuwenhoek who constructed light microscopes that could resolve cells and subcellular structures (Lane, 2015). The intricacy of nucleus organization can be appreciated when trying to imagine the almost 2 m long DNA complexed with multiple protein and RNA species, tightly folded into a micrometre-sized nucleus. This highly sophisticated nucleoprotein complex is called chromatin. Historically electron microscopy (EM) was instrumental in revealing several levels of chromatin organization, such as a famous “beads on a string” arrangement of nucleosomes. Multiple levels of chromatin organization are beyond the diffraction-limited light microscopy resolution (~250 nm) (Boopathi et al., 2020). However, EM is of limited utility for characterizing dynamic structures (Maeshima et al., 2019) and most insights into chromatin dynamics are obtained by relatively lower resolution fluorescence microscopy (Heun et al., 2001). Staining DNA with intercalating dyes or labelling high abundance chromatin proteins such as histone H2B visualizes global chromatin structure; nevertheless, true understanding of the role of chromatin organization in the cell life requires visualizing specific genome loci. Numerous methods are available to obtain labelled

synthetic DNA, but sequence-specific labelling of native DNA still represents a challenge. The early studies on chromatin dynamics based on fluorescence microscopy used green fluorescent protein (GFP)-tagged lactose (Lac) repressor (Heun et al., 2001). To achieve high contrast, hundreds of copies of Lac operator sequences had to be integrated into the genome locus under investigation, thus potentially interfering with its natural dynamics. Later a less destructive DNA-labelling method was developed which utilizes a clustered regularly interspaced short palindromic repeats nuclease Cas9 (CRISPR-Cas9) to insert nonrepetitive short (<1 kb) DNA sequence (ANCH). This sequence is specifically detected by bacterial partition protein ParB fused with GFP (Germier et al., 2017). A similar method SHACKTeR (Short Homology and CRISPR/Cas9-mediated Knock-in of a TetO Repeat) allows imaging specific chromosomal regions in living cells without the need for a pre-existing repetitive sequence (Tasan et al., 2018). However, these approaches rely on the genetic engineering of locus to be observed.

Currently the most common methods for studying chromatin structure are a large family of sequencing-based chromosome conformation capture (3C) techniques and microscopy-based fluorescence *in situ* hybridization (FISH) variations (Jerkovic and Cavalli, 2021; Kempfer and Pombo, 2020). Resolution of 3C techniques can reach down to several

\* Corresponding author.

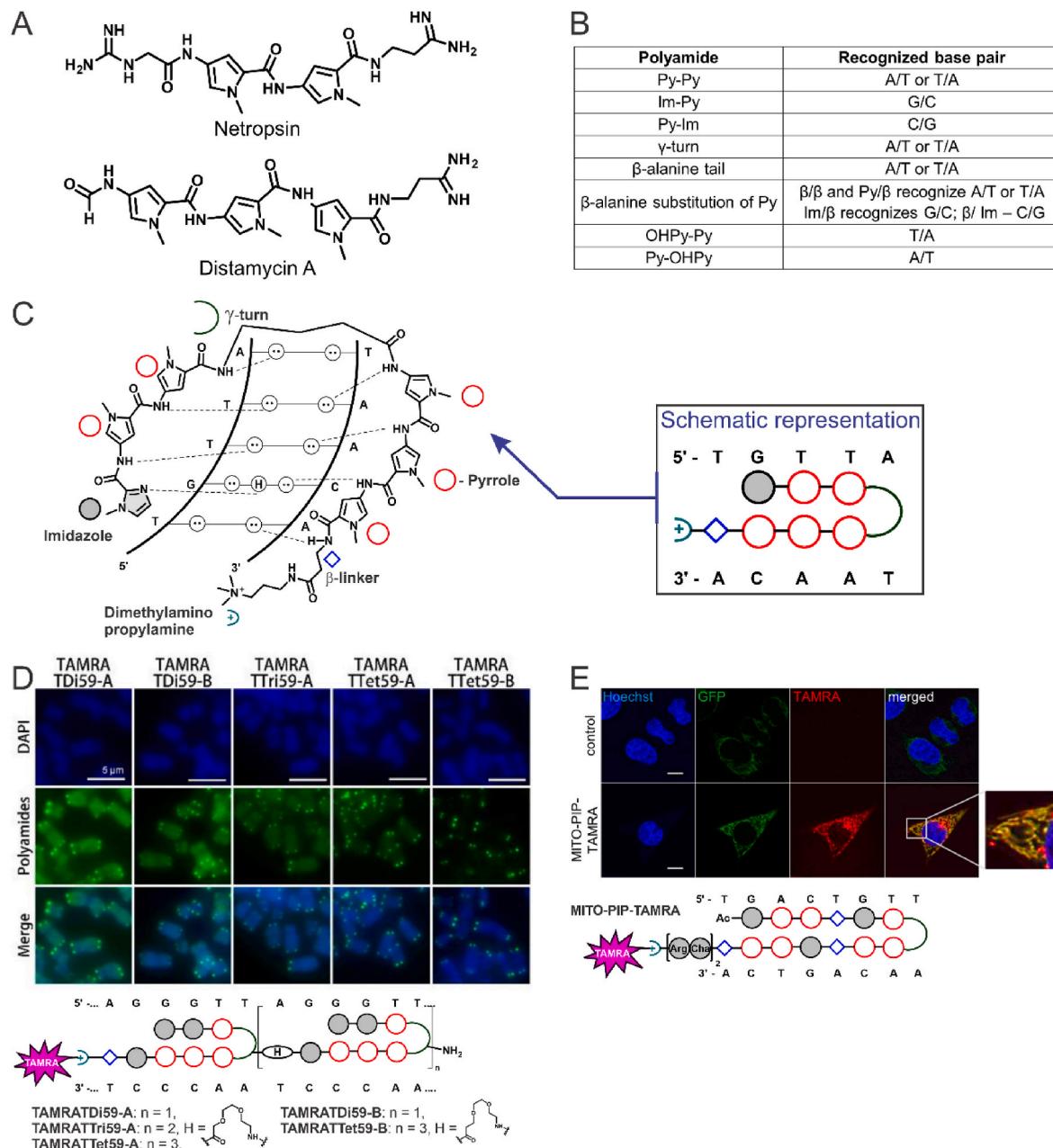
E-mail address: [grazvydas.lukinavicius@mpinat.mpg.de](mailto:grazvydas.lukinavicius@mpinat.mpg.de) (G. Lukinavičius).

<sup>1</sup> These authors contributed equally to this work.

base pairs, but visual model is difficult to build and requires considerable computing power. Recent developments towards FISH automation and combination with super-resolution microscopy allowed researchers to highlight the details of chromatin structure ranging from the whole chromosomes to kilobase-sized interactions between neighbouring cis-regulatory elements (Boettiger and Murphy, 2020). These methods uncovered the relationships between chromatin structure and epigenetics (Boettiger and Murphy, 2020; Wooten et al., 2020; Xu et al., 2018, 2020), the nature of topologically associated domains (TADs) (Barth et al., 2020; Xiang et al., 2018), and the enhancer–promoter interactions underlying transcriptional regulation (Branda et al., 2021; Brown et al., 2018). However, 3C and FISH methods require cell

fixation-permeabilization, which are highly likely to introduce artefacts (Li et al., 2017). Furthermore, harsh conditions of DNA denaturation in FISH inevitably destroy local chromatin structure. Such changes might be largely obscured by light diffraction limit in confocal microscopy, but become visible in super-resolution imaging (Markaki et al., 2012). Consequently, less invasive DNA labelling methods are needed for getting deeper insights into the genome dynamics, such as formation and rearrangement of TADs during cell cycle, pathogenesis, or external stimulation.

In this review, we focus on the methods for highlighting specific sequences in natural DNA that do not require modification of the target locus and could be employed for super-resolution imaging of fixed/



**Fig. 1. DNA labelling with polyamides.** A. Chemical structures of netropsin and distamycin A. B. DNA sequence recognition rules for polyamides. C. Example of sequence-specific dsDNA recognition and schematic representation of Py-Im polyamides. D. Schematic representations of tandem dimer, trimer, and tetramer Py-Im polyamide probes labelled with TAMRA targeting telomeric repeats and imaging of telomeres. Adapted with permission from (Kawamoto et al., 2016). Copyright 2016 American Chemical Society. E. Schematic structure representation and mitochondrial localization of MITO-PIP-TAMRA (red), nuclei and mitochondria were stained with Hoechst 33342 (blue) and CellLight Mitochondria- GFP BacMam 2.0 technology (green). Scale bars - 10  $\mu$ m. Adapted with permission from (Hidaka et al., 2017). Copyright 2017 American Chemical Society.

living cells. We discuss triplex forming oligonucleotides (TFOs) and hairpin polyamides as less destructive alternatives to standard FISH methods and programmable DNA binding proteins (dCas9 and transcription activator-like effectors (TALEs)) that offer many opportunities for monitoring dynamics of the specific chromatin loci in living cells. In addition, we outline the approaches combining bacterial DNA methyltransferases and fluorescent cofactor analogues that are used for optical genome mapping.

## 2. Labelling using synthetic molecules

Small synthetic molecules provide an opportunity to highlight specific DNA sequences in living cells and organism without genetical modification. The most advanced methods involve hairpin polyamides and triplex forming oligonucleotides. These methods became even more attractive after high level of chemical synthesis automatization was reached during the last years.

### 2.1. Sequence-specific DNA labelling using pyrrole-imidazole (Py-Im) polyamides

Natural antibiotic and antiviral agents netropsin and distamycin A are small organic compounds that preferentially bind to the minor groove of adenine/thymine (A/T) - rich double stranded B form-DNA (B-DNA) in 1:1 or 2:1 ligand:DNA ratio (Kopka et al., 1985; Pelton and Wemmer, 1989). Their molecular structures contain two or three pyrrole (Py) moieties, respectively (Fig. 1A). The hydrogen bonding between the amide (NH) and N3-atom on adenine or O2-atom on thymine drive DNA complex formation, while van der Waals interactions serve as supporting forces (Coll et al., 1987; Pelton and Wemmer, 1989). Inspired by these natural molecules, Dervan and co-workers spent a couple of decades grasping the fundamentals of polyamide sequence-specific binding. Introduction of imidazole (Im) unit instead of Py enabled hydrogen bond formation between the exocyclic guanine amino group and N3 of imidazole, thus permitting recognition of guanine/cytosine (G:C) pairs (Dervan, 2001; Lown et al., 1986). This first generation of linear Py-Im polyamides bind targeted DNA sequence preferentially in a 2:1 ratio. Development of *tert*-butyloxycarbonyl (Boc) (Baird and Dervan, 1996) and fluorenylmethoxy carbonyl (Fmoc) (Wurtz et al., 2001) solid-phase synthesis of a wide range of Py-Im polyamide variants facilitated extensive screening of structural modifications and evaluation their binding properties to the double-stranded (ds) DNA. Dervan and co-workers identified a basic DNA binding hairpin motif consisting of two Py-Im polyamides connected via  $\gamma$ -aminobutyric acid ( $\gamma$ -turn) in head-to-tail fusion (Mrksich et al., 1994) and described the principal DNA pairing rules (Fig. 1B).

Hairpin polyamides bind DNA in 1:1 mode and are the most commonly used motif due to their versatility and high affinity to DNA (Fig. 1B). Examples of Py-Im polyamides applications involve up- and downregulation of gene expression (Gottesfeld et al., 1997; Mapp et al., 2000; Oyoshi et al., 2003; Pandian et al., 2012), transcriptional activation (Ren et al., 2017; Richier and Salecker, 2015), bio-sensing (Vaijayanthi et al., 2012), cancer therapeutics (Hargrove et al., 2015; Szablowski et al., 2016) or recently as probes in bio-imaging (Kawamoto et al., 2016; Tsubono et al., 2020). The diversity of applied structural motifs extends far beyond principal pairing rules outlined here and was minutely summed-up in a recent review (Kawamoto et al., 2018). We will focus on the recent examples of fluorescent polyamides applications in fluorescent microscopy.

Py-Im polyamides by their nature are non-fluorescent compounds and in order to be used as imaging agents a fluorophore has to be attached at a C/N (carboxy-/amino-) terminus or at  $\gamma$ -turn. Fluorescent polyamides were applied to detect specific DNA sequences and to investigate mismatched sequences using their ability to recognize the mismatched base pairs. U.K. Laemmli team demonstrated the first applications of fluorescent polyamides in fluorescence microscopy. In

2000's, they achieved a selective staining of AT-rich satellites, II-associated regions and 5'-(GAGAA)<sub>n</sub>-3' repeats of satellite V in *Drosophila melanogaster* using fluorescein-labelled oligo pyrrole compounds (Janssen et al., 2000). After a year the same group reported specific visualization of insect 5'-(TTAGG)<sub>n</sub>-3' and vertebrate 5'-(TTAGGG)<sub>n</sub>-3' telomeric repeats in fixed cells and on chromosomal spreads with Texas Red labelled tandem hairpin Py-Im polyamides (Maeshima et al., 2001). These pioneering papers outlined the perspectives of imaging DNA tandem repeat sequences as the high local concentration of binding sites attained many reporter molecules and provided a strong signal at a specific genomic locus.

Sugiyama et al. used fluorescent polyamides to label telomeres. They combined solid phase and solution based synthetic approaches to efficiently obtain tandem hairpin dimers, trimers and tetramer polyamides recognizing 12 base pair (bp) (Hirata et al., 2014; Kawamoto et al., 2013), 18 bp (Kawamoto et al., 2015), and a record breaking 24 bp (Kawamoto et al., 2016; Tsubono et al., 2020). It was found that tetra-methylrhodamine (TAMRA) labelled polyamides were superior to cyanine 3 (Cy3) or Texas red conjugates in terms of contrast. The specificity of staining was significantly improved after optimization of hinge length between the hair pins (Hirata et al., 2014) and along with increased number of targeted base pairs (Fig. 1C) (Kawamoto et al., 2013, 2015, 2016). Tandem hairpin polyamide labelled with TAMRA fluorophore was successfully applied in telomere visualization in mouse and human tissue sections (Kawamoto et al., 2016). Recently, an interesting application of TAMRA-linked polypyrrrole for the genomic DNA optical mapping was demonstrated (Lee et al., 2018). The fluorescence increases upon binding 5'-AGGGTT-3' human telomere repeat sequence was demonstrated by substituting TAMRA for the far-red fluorogenic silicon-rhodamine (SiR) fluorophore (Tsubono et al., 2020). The authors visualized telomeres in living human osteosarcoma cells (U-2 OS) using a peptide-based delivery reagent, and hypothesized that the SiR-TTet59B probe can stain telomeres in a length-dependent manner and could be applied measurement of telomere sizes (Tsubono et al., 2020). In addition, the SiR-TTet59B probe should be compatible with stimulated emission depletion (STED) microscopy as the fluorescent marker, silicon-rhodamine, is ideally suited for this purpose, but the authors did not demonstrate that. Boutorine et al. showcased that smaller size fluorescein labelled hairpin polyamides are cell permeable, however their localization was highly dependent on the polyamide labelling approach. For example, a hairpin, which was labelled with fluorescein isothiocyanate, was suitable for the visualization of pericentromere regions (Nozeret et al., 2015). In contrast, labelling with N-hydroxysuccinimide ester or azido group based "click" reaction yielded fluorescent polyamides, which mainly localized in cytosol (Nozeret et al., 2018). Localization of some polyamides in cytosol, instead of nucleus, was also reported by Dervan and co-workers (Nickols et al., 2007) where they used fluorescein isothiocyanate (FITC) conjugation method. Introduction of cyclohexylalanines (Cha) and arginines (Arg) induced mitochondrial localization and allowed selective mitochondrial DNA imaging (Fig. 1D) (Hidaka et al., 2017). Recently, it was demonstrated that fluorescent Py-Im polyamides can stain DNA in a living mouse (Fig. 1E) (Inoue et al., 2018).

Py-Im polyamides can be easily synthesized by machine-assisted solid-phase peptide synthesis and conjugated to a fluorescent dye to yield a fluorescent probe targeting a desired DNA sequence. Machine-assisted synthesis is advantageous to solution phase synthesis as it allows obtaining larger variety of structurally diverse polyamides in a short time scale. Their binding to DNA is relying on the polyamide fitting into the minor groove of dsDNA and establishing a network of hydrogen bonds. It has been demonstrated that polyamides can recognize specific DNA sequences ranging from 4 to 24 base pairs. However, high specificity is still a major problem as smaller polyamides can off-target to other closely similar sequences or even other biomolecules (Lin and Nagase, 2020). In addition, imaging of less abundant genomic DNA sequences yet have not been demonstrated due to off-targeting and low

signal-to-background ratio. Fluorescence signal amplification by attachment of 2 or more fluorophores to a single polyamide usually results in a loss of binding affinity due to increased steric hindrance (Lee et al., 2018). In general, specific DNA sequence labelling with polyamides seems to be a promising technology, but currently it is still in an early “proof of principle” stage.

## 2.2. Triplex forming oligonucleotides (TFOs)

In the native state, double-stranded genome DNA is hardly accessible for hybridization with complementary oligonucleotide probes. Simple solution of this situation is denaturation at high temperatures as for FISH applications. This technique has been routinely used to label specific DNA sequences in interphase and metaphase nuclei for research and medical diagnostics. In short, FISH protocol includes chemical fixation of cells followed by denaturation at 70–95 °C in the presence of chaotropic agents such as formamide. Next, the hybridization of labelled oligonucleotides to denatured single stranded DNA is performed. The denaturation of target DNA is one of the most damaging step in FISH technique that was clearly visualized in super-resolution scanning near-field optical microscopy (SNOM) (Krufczik et al., 2017; Winkler et al., 2003). Therefore, less destructive protocols of FISH are sought. For example, the modified FISH methods, like fast-FISH and low temperature FISH, employ hybridization without the use of formamide or thermal denaturation, which results in a better-preserved morphology of the chromosomes observed by SNOM at the sub-hundred nanometre resolution (Krufczik et al., 2017; Winkler et al., 2003). Recently, exonuclease digestion has been used in a method called Resolution After Single-strand Exonuclease Resection (RASER)-FISH avoiding heat denaturation (Brown et al., 2022). An excellent alternative to FISH is methods based on formation of triple helix that preserve classical Watson-Crick base pairing.

The ability of nucleic acids to form triple helices was discovered in 1957 by Felsenfeld and Rich (Felsenfeld et al., 1957). A biological significance of triplexes in biology was recognized in 1968, when Morgan and Wells showed that a three stranded complex between a dsDNA and a single-stranded RNA (ssRNA) inhibited transcription by *E.coli* RNA polymerase (Morgan and Wells, 1968). Later, a number of studies identified several triplex binding proteins and indirectly proved existence of these structures in multiple species: bacterial Tn7 protein, yeast CDP1 protein, drosophila GAGA transcription factor, murine HMG protein, human Orc4 and XPA-RPA proteins (Jimenez-Garcia et al., 1998; Kusic et al., 2010; Musso et al., 2000; Rahman et al., 2007; Suda et al., 1996; Thoma et al., 2005).

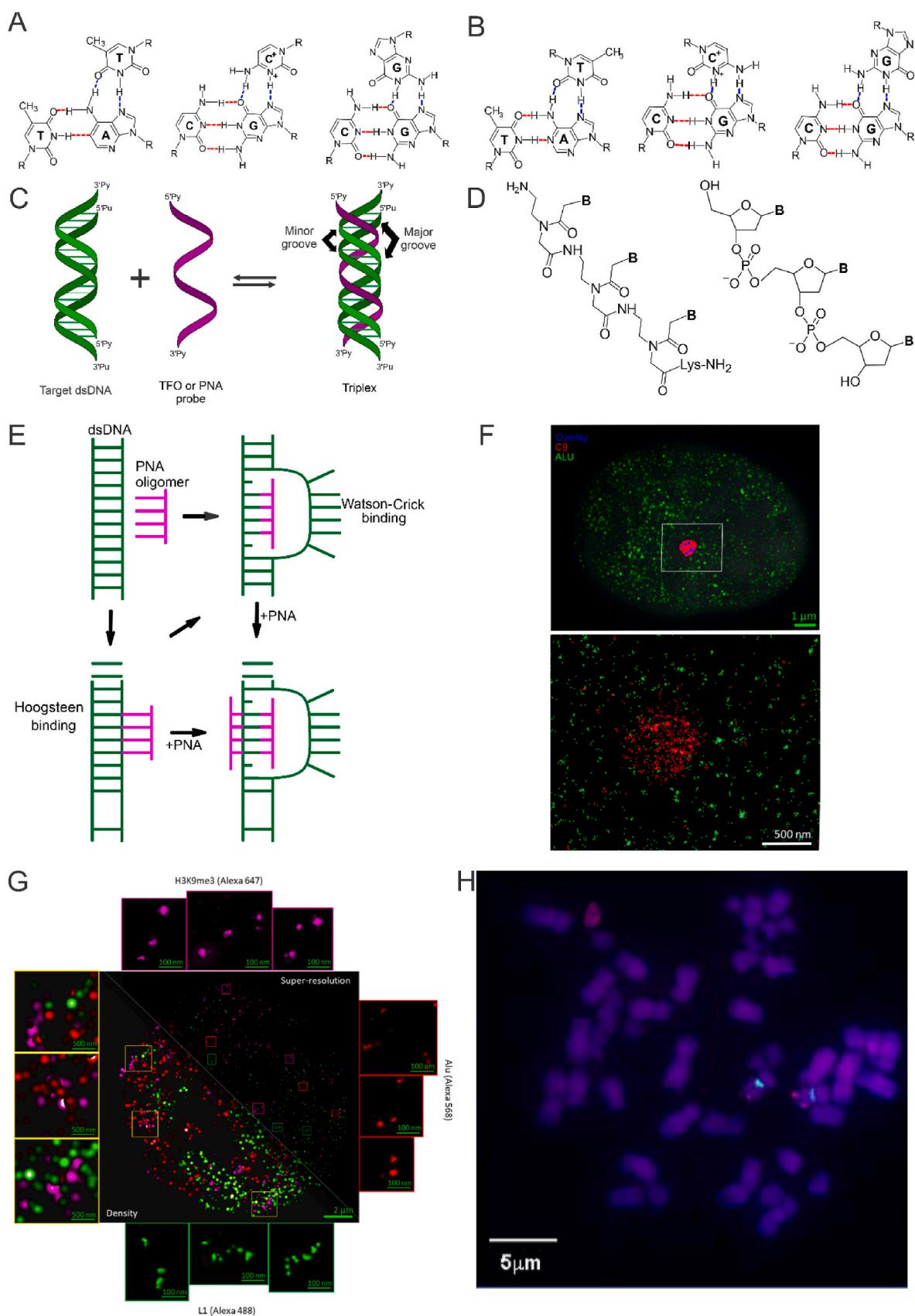
The third strand binds in a major groove of oligopurine or oligopyrimidine tracts of the double-stranded DNA (dsDNA) major groove (Faria and Giovannangeli, 2001). Early studies utilized TFOs composed of ribose or deoxyribose nucleotides, but lower stability of the unmodified ribonucleic acid (RNA) resulted in the introduction of 2'-O-methyl-ribose and other modifications containing probes (Asensio et al., 1999; Cuenoud et al., 1998; Maciaszek et al., 2015; Rahman et al., 2007; Sasaki et al., 2004). In DNA double helix, the two strands are held together by formation of Watson-Crick base pairs (Adenine (A)-Thymine (T) and Guanine (G)-Cytosine (C)). The third strand is added by forming Hoogsteen or reverse Hoogsteen hydrogen bonds with free acceptor or donor groups in dsDNA major groove (Fig. 2A and B). Under normal conditions only purine bases on a target DNA can establish two extra hydrogen bonds with the TFO base without disruption of the Watson-Crick hydrogen bonds (Fig. 2C) (Frank-Kamenetskii and Mirkin, 1995). Typically, TFOs recognize 12–30 bp oligopurine or oligopyrimidine where dsDNA tends to adopt A-form conformation in triplex region and the junction between the duplex and triplex can either be A-form or B-form (Vasquez and Glazer, 2002). The cytosine (C) requires protonation and C:G × CH<sup>+</sup> triplex can form at acidic pH (Boutourine et al., 2013; Lee et al., 1979; Vasquez and Glazer, 2002). In contrast, TFOs containing purine nucleotides (A and G) and pyrimidine, thymine

(T), form pH-independent triplexes (Cooney et al., 1988; Faria and Giovannangeli, 2001). Chemical modification to the backbone of TFOs can enhance the triplex stability. These modifications include N, N-diethylenediamine phosphoramidate (DEED), N3'→P5' phosphoramidate (NP), phosphorothioate (PS) or peptide nucleic acids (PNAs) backbone (Faria and Giovannangeli, 2001). G-rich TFOs containing AG and TG nucleotides and T-rich TFOs containing TC and TG nucleotides are more likely to manifest in a cellular environment. Along with this, some G-rich oligonucleotides are prone to form four stranded G-rich quartets and intermolecular homoduplexes (GA repeats) (Noonberg et al., 1995; Olivas and Maher, 1995). These additional structures compete with the oligonucleotides for triplex formation with dsDNA. This could be overcome by replacement of guanines with synthetic base analogues such as 6-thioguanine or 7-deazaguanine. In addition, 7-deazaxanthine (c<sup>7</sup>X) in place of A or T forming c<sup>7</sup>X:A:T triplex in anti-parallel triplexes are also preferred (Faruqi et al., 1997).

PNAs are uncharged synthetic oligonucleotide analogues where the nucleobases are attached to the pseudopeptide backbone and hybridization follows the traditional base pairing rules: A-T and G-C (Fig. 2D). The uncharged PNAs form more stable triplexes compared to the phosphodiester backbone containing nucleic acids at physiological salt concentrations. However, triplex formation is competes with strand invasion which results in opening up of the duplex (Fig. 2E) (Hansen et al., 2009). Further increase in concentration results in binding of two PNA oligomers to the DNA target strand, leaving the other complimentary strand displaced (Faria and Giovannangeli, 2001; Vasquez and Glazer, 2002). This gives rise to the following formation of PNA/DNA/PNA triplexes (2:1 PNA to DNA ratio): one PNA strand binds to the DNA strand via regular Watson-Crick bonds whereas the other employs Hoogsteen hydrogen bonds (Fig. 2E) (Bentin and Nielsen, 1999). In homopyrimidine PNA oligomers substituting pseudoisocytosine for cytosine, in combination with (oligo)lysine or 9-aminoacridine conjugation results in PNA-dsDNA triplex formation (Bentin and Nielsen, 1999).

The shortcomings of conventional FISH are addressed in COMBinatorial Oligo-nucleotide FISH (COMBO-FISH) which omits the need of thermal denaturation of target DNA and use of any chaotropic agents. The improved method uses computationally optimized approach for designing of a set of fluorescently labelled triplex forming oligonucleotide probes (15–25mers) that uniquely co-localize at the target site or single oligo stretches that are repetitive but bind exclusively to the target (Schmitt et al., 2012). Schwarz-Finsterle and co-workers found the better-preserved chromatin structures while comparing COMBO-FISH to standard FISH method by quantitative microscopic image analysis of Abelson murine leukemia (ABL) and breakpoint cluster region (BCR) positions in nuclei of lymphocytes and chronic myeloid leukemia (CML) blood cells (Schwarz-Finsterle et al., 2007). In addition, COMBO-FISH can utilize PNA oligonucleotides that form a stable triplex and are metabolically more stable in living cells. These properties of PNA oligonucleotides enables COMBO-FISH to be applied for *in vivo* labelling (Schmitt et al., 2010). Jin-Ho Lee and co-workers applied COMBO-FISH in combination with immunostaining and successfully demonstrated possibility to measure nuclear architecture by 3D confocal microscopy and chromatin nano-architecture by single molecule localization microscopy (SMLM) (Lee et al., 2019). Here they studied the spatial distribution of genomic *Alu* elements around chromosome 9 centromeres (Fig. 2F). Furthermore, the low temperature conditions paved way for the concurrent COMBO-FISH and H3K9me3 antibody immunostaining (Fig. 2G).

One of the early applications of fluorescent PNA in microscopy was described by Molenaar et al. (2003). They used Cy3-labelled PNA to probe telomere structure. PNA's do not penetrate into living cells and requires laborious microinjections, permeabilization by pore-forming toxins (as streptolysin-O) or other invasive cell delivery methods (Flierl et al., 2003). In another application, PNA oligonucleotide probes were used for COMBO-FISH technique in combination with Spatially



(caption on next page)

**Fig. 2. Schematic representation of DNA triplex formation.** A. Examples of Hoogsteen parallel bonding. Base triplets: T:A  $\times$  T, C:G  $\times$  CH<sup>+</sup> and C:G  $\times$  G. B. Triplex formation by Reverse Hoogsteen antiparallel bonding. Base triplets: T:A  $\times$  T, C:G  $\times$  CH<sup>+</sup> and C:G (:) represents Watson-Crick bonding and ( $\times$ ) represents Hoogsteen and Reverse Hoogsteen bonding. Watson-Crick bonds are shown in red, Hoogsteen and Reverse bonds are in blue. R represents deoxyribose. C. The TFO or PNA probe binds to the target duplex through major groove to form a triplex. D. Structure of peptide nucleic acid (PNA) compared to deoxyribonucleic acid (DNA) where B = Nucleobases. E. Modes of PNA-DNA complex formation by strand invasion of duplex DNA. F. Super-resolution overlay image of Alu densities (green), chromosome 9 centromere points cluster (red), and overlaying regions (blue). A magnified point coordinate representation of the white box is shown below. Image reproduced from (Lee et al., 2019) under CC BY 4.0 license. G. Super-resolution SMLM overlay image and the magnifications show regions with different combinations of co-localization (yellow boxes) and characteristic nanostructures of H3K9me3 marks (magenta), genomic Alu (red), and genomic L1 (green). Image reproduced from (Lee et al., 2019) under CC BY 4.0 license. H. Lymphocyte metaphase spread after a two-colour experiment with the PNA COMBO-FISH probe (green signals: OregonGreen 488) and a 9q subtelomere (9qtel) DNA standard FISH probe (red signals: rhodamine). DNA counterstained with TOPRO-3-iodide (blue). Scale bar – 5  $\mu$ m. Image reproduced from (Muller et al., 2010) under CC BY 3.0 license.

Modulated Illumination (SMI) microscopy to measure the size of ABL gene region in 3D conserved blood cell nuclei (Muller et al., 2010). To demonstrate binding specificity and efficiency two colour experiments with OregonGreen 488 and TexasRed were performed. The probes were specific for centromere 9. They were separately hybridized on fixed peripheral blood lymphocyte nuclei and metaphase spreads. This was followed by another experiment to verify whether the labelled chromosome was actually chromosome 9. Here, the lymphocyte samples were labelled with PNA COMBO-FISH probe having OregonGreen 488 and 9q subtelomere (9qtel) DNA standard FISH probe labelled with rhodamine dye. This showed two chromosomes with red and green spots, suggesting that subtelomere ends and centromere region of chromosome 9 were marked by the DNA and PNA probe respectively (Fig. 2H).

In summary, the main advantage of DNA labelling by triplex formation is that it does not require the opening of target DNA duplex and thereby causes less damage to chromatin structure than conventional FISH. However, the background fluorescence of non-bound oligonucleotides is high and intense washing is required before imaging. The possible solution to this problem is generation of fluorogenic TFO probes, which only fluoresce upon DNA binding. For example, this can be achieved by conjugating intercalating agents such as acridine, benzopyridoindole and monomethine cyanines to oligonucleotides (Silver et al., 1997; Sun et al., 1989; Walsh et al., 2018). These modifications were shown to increase the stability of the triple helical complexes significantly, which is an important factor for imaging experiments (Silver et al., 1997). Another important aspect of triplex-forming PNAs is that they show higher affinity for dsRNA and such off-targeting could lead to false signals or background increase (Sato et al., 2016). Furthermore, PNA-dsDNA triplex formation competes with strand invasion that partially destroys local structure of dsDNA. This issue could be addressed by incorporating nucleotide analogues and/or variation of PNA probe length (Hansen et al., 2009). The main drawback of the TFOs and PNAs is their weak intracellular permeability, which limits their applications on living samples.

### 3. Labelling based on protein-DNA interactions

Exploiting proteins interacting with DNA for labelling purposes is a relatively straightforward approach and it is particularly attractive for the observation of dynamic processes in living cells. The use of dCas9 and TALEs offers higher specificity and flexibility compared to small molecule-based approaches. Less specific DNA methyltransferases provide opportunity to covalently link fluorescent label to nucleobase in the double helix with little perturbations. We will discuss these methods in the next chapters.

#### 3.1. Chromatin imaging using catalytically deficient dCas9 nuclease

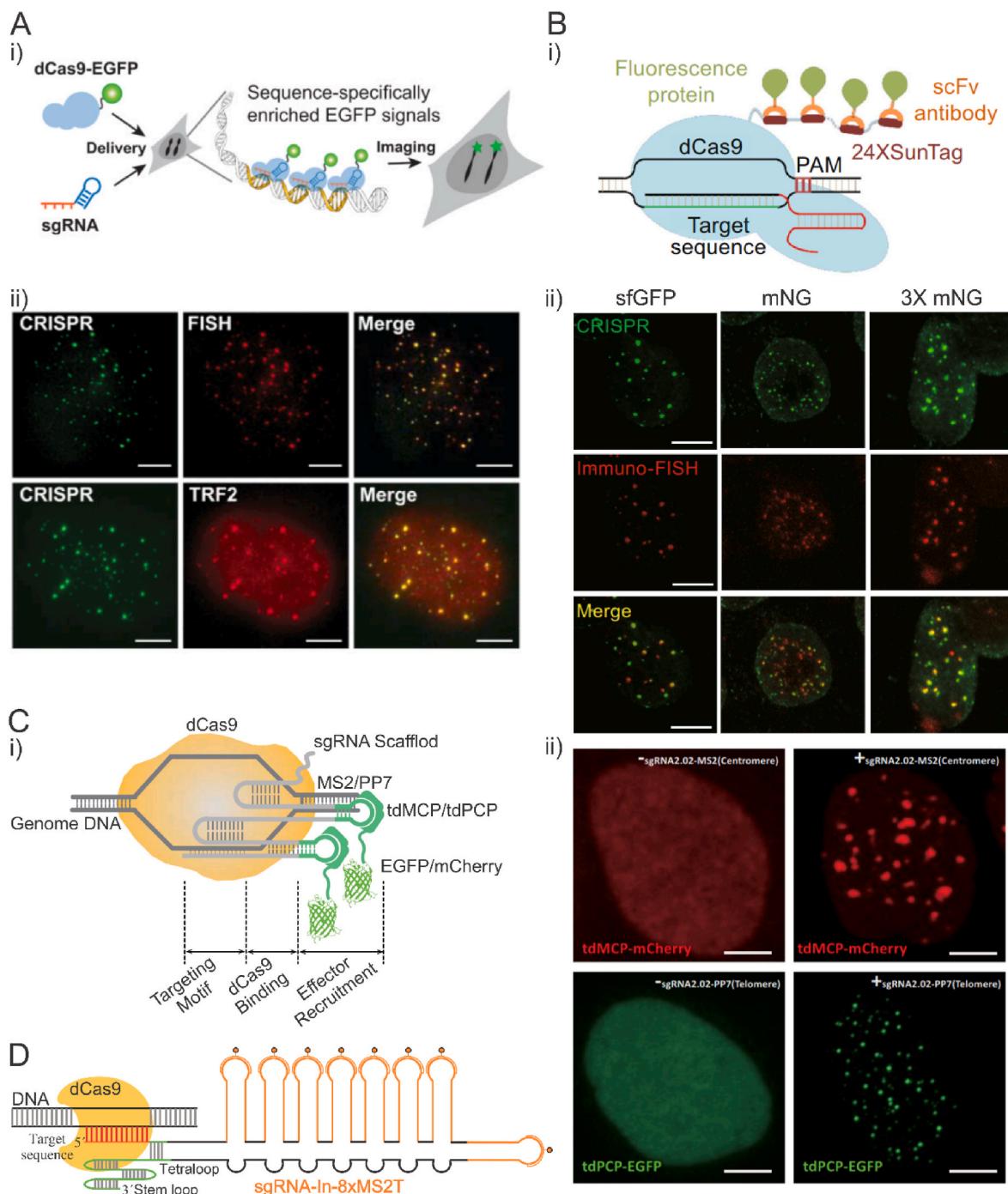
The discovery of the CRISPR antiviral bacterial defence allowed scientists not only to edit target genes in the genomes, but also opened new possibilities for endogenous genome imaging (Doudna and Charpentier, 2014; Ishino et al., 1987; Makarova et al., 2006). The effector protein of the CRISPR system – Cas9, is programmable with a single

guide RNA (sgRNA) that recognizes target DNA sequence by formation of Cas9-sgRNA-dsDNA tertiary complex (Gasiūnas et al., 2012; Jinek et al., 2012). Catalytic activity is unnecessary for DNA binding, thus the endonuclease-dead Cas9 (dCas9) allows generation of RNA-programmable complexes without the ability to cleave the DNA (Chen et al., 2013; Qi et al., 2013). Both, protein and RNA, components of this complex can carry a reporter tag and can be used for chromatin labelling. High contrast is indispensable for locating specific loci in the crowded nucleus environment; therefore, we focus our discussion on the developments aimed at improving signal-to-background ratio (S/B).

The first application of dCas9-sgRNA for imaging specific DNA sequences in living cells was demonstrated by Chen et al. who co-expressed enhanced green fluorescent protein (EGFP)-dCas9 and respective sgRNAs to label telomeres and tandem repeats in *MUC4* gene (Fig. 3A) (Chen et al., 2013). Hundreds of tandem repeats in these targets attracted multiple EGFP-dCas9 molecules and ensured efficient labelling and good contrast. To detect non-repetitive sequences, the authors co-expressed EGFP-dCas9 with multiple sgRNAs tiling a non-repetitive 5 kbp region of *MUC4* gene. They showed that at least 36 distinct sgRNAs were required for sufficient contrast. Later, multicolour imaging scheme was devised by fusing dCas9 orthologues from *S. pyogenes* (Sp), *Neisseria meningitidis* (Nm) and *Streptococcus thermophilus* (St1) to 3 copies of GFP, red fluorescent protein (RFP) and blue fluorescent protein (BFP) (Ma et al., 2015). Pairs of differently coloured 3  $\times$  fluorescent protein (FP)-dCas9/sgRNAs were used to determine distance between loci on different chromosomes. Early on it was noticed that sgRNA is a limiting factor in DNA labelling with CRISPR (Chen et al., 2013). Therefore, signal-to-background ratio (S/B) could be improved not only by adding multiple copies of FPs, but also by stabilizing sgRNA and enhancing its interaction with dCas9 (Chen et al., 2013; Ma et al., 2015).

Introduction of a small multivalent epitope called SunTag (Super-Nova tag) at C-terminus of dCas9 reduces the size of fusion protein and improves S/B. The tag is 24-mer of GCN4 peptide recognized by a highly specific single-chain variable fragment antibody fused to GFP (scFv-GCN4-GFP) resulting in enormous fluorescence signal amplification at the target sequence (Tanenbaum et al., 2014). Similarly, dCas9-SunTag system fused with superfolder GFP (sfGFP), mNeonGreen and 3  $\times$  mNeonGreen was successfully used for imaging telomeres in HEK293 cells (Fig. 3B) (Ye et al., 2017).

Alternative to dCas9 modification with FPs is sgRNA modification with phage-derived RNA aptamers MS2 (*Emesivirus zinderi*) or PP7 (*Pseudomonas* phage). The fluorescent label is brought-in by the respective binding proteins fused to FPs: MS2 coat protein (MCP) and PP7 coat protein (PCP). MS2 and PP7 systems are orthogonal and allow multiplexing. 2-fold enhancement of S/B was achieved by modifying sgRNA tetraloop and stem loop 2 with MS2 or PP7 aptamers compared to sgRNA extension with corresponding aptamers (10 versus 5) or two unmodified sgRNA (80 versus 40) (Fig. 3C) (Shao et al., 2016; Wang et al., 2016). S/B can be enhanced further by modifying sgRNA with RNA aptamers (MS2/PP7) that bind FP-tagged tandem dimer (td) MS2 coat (tdMCP) and PP7 coat (tdPCP) proteins. EGFP-tdPCP and mCherry-tdMCP were used for labelling telomeres and  $\alpha$ -satellites of human adenocarcinoma cells MDA-MB-23 (Fu et al., 2016; Shao et al.,



**Fig. 3. Labelling of DNA sequences with dCas9.** A. Labelling of genomic loci with specific dCas9-EGFP/sgRNA complexes. i) Experimental design ii) Labelling of telomeres with dCas9-EGFP using improved sgRNA. Validation with PNA FISH (top row), and validation with antibody to telomeric repeat-binding factor 2 (TRF2) (bottom row). Scale bars - 5  $\mu$ m. Figure adapted with permission (License Number 5323650983454) from Elsevier (Chen et al., 2013). B. i) Schematic of dCas9-SunTag system combined with fluorescent proteins. ii) Visualizing human telomeres with CRISPR/Cas9 system using different fluorescent proteins. HEK293T cells were transfected with sgRNA and dCas9-SunTag system combined with sfGFP, mNeonGreen and 3 × mNeonGreen. Scale bars - 5  $\mu$ m. Figure adapted from (Ye et al., 2017) under CC BY 4.0 license. C. i) Schematic representation of dCas9-sgRNA-aptamers complex with DNA. Tetraloop (lower hairpin) and stem loop 2 (upper hairpin) are extended by aptamers. DNA depicted in dark grey, sgRNA – in light grey. ii) MDA-MB-231 cells; labelled centromeres with tdMCP-mCherry or labelled telomeres with tdPCP-EGFP. Specific labelling pattern is observed with high signal-to-background ratios for both cases (~80) (lower row). Experiments without specific sgRNA shows unspecific uniform cell labelling (Left column, upper row). Scale bars - 5  $\mu$ m. Figure adapted from (Shao et al., 2016) under CC BY 4.0 license. D. Diagrams of sgRNAs-In-8 × MS2T bound with MCP-HaloTag used in DNA labelling (Ma et al., 2018).

2016). Much faster tdMCP and tdPCP exchange rates than dCas9 compensates photobleaching and makes it suitable for long-term tracking of telomeres and centromeres (Shao et al., 2016). To improve brightness of labelling with MS2, thermostable octets were rationally designed (Fig. 3D) (Ma et al., 2018). A careful optimization of sgRNA structure

and inserting 16 × MS2 allowed imaging a non-repetitive genome locus with as few as 4 different sgRNAs. This approach allowed targeted loci tracking over the course of the cell cycle (Qin et al., 2017).

A recently described CRISPR-mediated fluorescence *in situ* hybridization amplifier (CRISPR FISHer) system employs a very interesting

principle of signal amplification at the targeted locus (Lyu et al., 2022). It consists of untagged dCas9, sgRNA labelled with at least 2 copies of PP7 aptamer and T4 fibrin trimeric motif foldon fused to GFP-PCP. Foldon-GFP-PCP forms stable trimers. When recruited to dCas9-sgRNA-2 × PP7 bound to the target DNA, foldon-GFP-PCP seeds rapid aggregation of multiple sgRNA-2 × PP7 and foldon-GFP-PCP molecules resulting in exponential signal amplification, which is sufficient for detecting non-repetitive sequences with a single sgRNA.

The number of colours was increased by introducing the third aptamer – binding protein pair boxB – affinity-enhanced λN22. Thereby, CRISPRainbow system was created that can detect 6 loci simultaneously. It consists of sgRNA that is modified by a pair of hairpins from MS2/PP7/boxB set and respective binding proteins fused to FPs of different colours (RFP, GFP or BFP). In such a system, each of the six targeted loci recruits a different combination of FPs creating red, green, blue, magenta, yellow, and cyan spots (Ma et al., 2016).

An important aspect in improving S/B is suppression of background originating from unbound fluorescent component(s). One such example is the bimolecular fluorescence complementation (BIFC) based on split Venus FP assembly on multivalent SunTag peptide. The system consists of three components: single-chain antibody fused to C-terminal part of Venus (scFv-Venus C), MCP fusion to complementary N-terminal part of Venus (MCP-Venus N) and 24 copies of GCN4 peptide fused to dCas9. Fluorescence is switched ON only after a correct binding of all components resulting in up to 3-fold S/B improvement compared to SunTag system (6 versus 2) (Hong et al., 2018). Another example is self-complementing tripartite split sfGFP system. Here, sfGFP regains fluorescence only after GFP<sub>1-9</sub> (the first component) is complemented with dCas9-SunTag, which is fused to scFv-GFP<sub>10</sub> (the second component) and sgRNA-MS2 is fused to MCP-mCherry-GFP<sub>11</sub> (the third component) (Chaudhary et al., 2020).

Alternative to aptamer – binding protein pairs, the RNA binding domains of Pumilio/fem-3 binding factor (PUF) were employed in Casilio system (Cheng et al., 2016). These protein domains can be programmed to bind specific linear 8-mer RNA sequences, called PUF binding sites (PBS). Multiple copies of PBS can be attached to sgRNAs without negative effects on their expression, allowing for high signal amplification. The programmable specificity of PUF domains offers ample opportunities for multiplexing. In all-in-one Casilio system all the components (dCas9, PUF-mClover and sgRNA-25 × PBS) are encoded in just one plasmid for telomere and major satellite imaging (Zhang and Song, 2017). A recent study demonstrated a capability of Casilio system to simultaneously image up to three non-repetitive loci with a single sgRNA for each (Clow et al., 2022). In this study, sgRNAs were tagged with 15–25 copies of PBS. To increase number of colours, PBS and aptamers can be used together (Maass et al., 2018).

Most of the discussed dCas9-based methods rely on lentivirus-based delivery systems: a) delivery by lentiviral plasmids that allow simultaneous transfection of all components; b) delivery by lentivirus infection, which requires lentivirus production with every plasmid of imaging system and sequential infection-selection cycles. To simplify these procedures, dCas9-sgRNA complexes can be preassembled *in vitro* and used as imaging reagents *in vitro* or *in vivo*. For example, *in vitro* assembled dCas9-mNeonGreen/sgRNA targeted to RHM2 and RHM3 genes and telomere repeats was used to monitor mitotic chromatid formation of sperm nuclei in metaphase-arrested *Xenopus* egg extract (Lane et al., 2015). High S/B was ensured by targeting 100 separate loci within 3.4 megabase region on *Xenopus* chromosome 4 with multiple sgRNAs (Lane et al., 2015). In addition, chromatin labelling in fixed cells was performed by CasFISH method, applying dCas9-C-Halo-tag labelled with fluorescent JF646/JF549 substrates and sgRNA-DY547/Cy5 targeting major satellites, minor satellites, telomeres or coding gene in fixed mouse embryonic fibroblasts (Deng et al., 2015). Another method for labelling with dCas9 system is RNA-guided endonuclease-*in situ* labelling (RGEN-ISL) which employs *S. pyogenes* dCas9, crRNA and a dye (5'-Alexa488 or 5'-ATTO 550) on tracrRNA to label telomeres and

centromeres in fixed animal and plant cells (Ishii et al., 2019; Nemecova et al., 2019). Mild conditions of sample preparation allow easy combination of these methods with immunostaining. Preassembled dCas9-sgRNA complexes can be delivered into the living cells by electroporation or microinjection in a method called CRISPR LiveFISH (Wang et al., 2019). Either sgRNAs or dCas9 can be labelled with fluorophores (Geng and Pertsinidis, 2021; Wang et al., 2019). Interestingly, free sgRNAs are quickly degraded inside the cell, while those in complex with dCas9 and complementary DNA are protected, which ensures a good contrast when using labelled sgRNAs. While CRISPR LiveFISH greatly simplifies delivery of multiple sgRNAs, electroporation is an invasive procedure prone to disturbing cell physiology and generating artefacts.

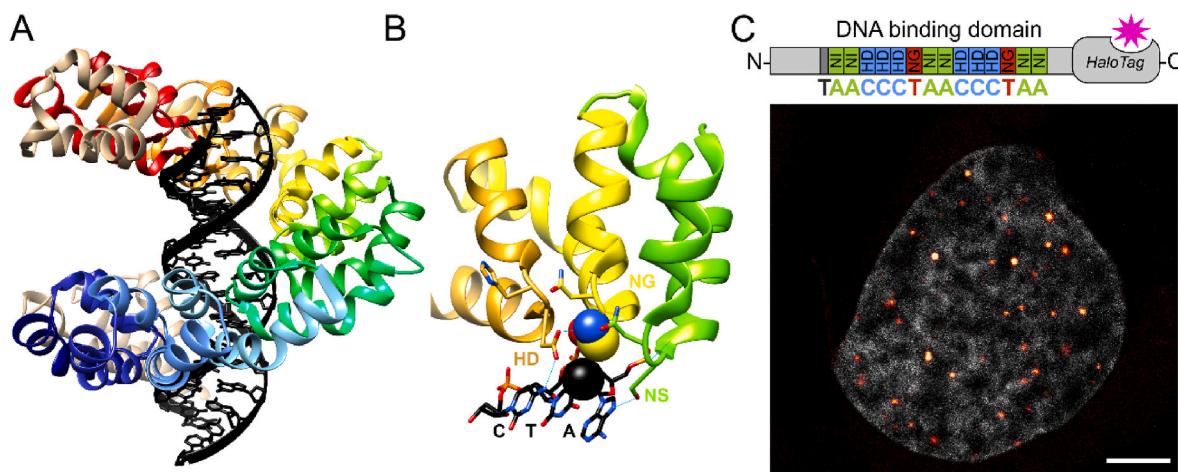
In summary, CRISPR-based chromatin labelling is a versatile method that works in living and fixed cells. Importantly, elaborate signal amplification and background suppression methods allow labelling low-repeat containing loci and truly unique sequences. On the downside, it is a multicomponent system, and a lot of effort must be put into tool development before the real biological question can be addressed. This is particularly the case when labelling non-repetitive sequences, where efficient delivery and then expression of dCas9, multiple sgRNAs and fluorescent reporters within a single cell must be ensured. Fast pace at which this technology is developing leads to expect more user-friendly reagents and protocols in the near future.

### 3.2. Applications of TALEs for fluorescence microscopy

TALEs from plant-pathogenic bacteria *Xanthomonas* are bacteria-encoded transcriptional activators (Bogdanove et al., 2010). During the infection, they are injected into cytoplasm through type III secretory system and then are imported into nucleus, to upregulate host susceptibility genes (Bogdanove et al., 2010). TALEs recognize specific sequences via their central domain, which is composed of 11–20 nearly perfect repeats with length of 33–35-amino acids. Inside the repeats, the most variable are 12 and 13 positions, which are called the repeat-variable diresidues (RVD). Each repeat folds into helix-loop-helix motif, and the whole domain forms a right-handed superhelix that wraps around DNA following the sense strand (Deng et al., 2012) (Fig. 4A). RVD is located inside the loop facing DNA and residue 13 makes specific contacts with a base in the major groove, while residue 12 stabilizes the local conformation. RVDs bind to DNA bases in one-to-one ratio and encode the specificity of TALE (Fig. 4B) (Boch et al., 2009; Bogdanove et al., 2010). The most common RVDs are HD (histidine-aspartate), NG (asparagine-glycine), NI (asparagine-isoleucine), and NN (asparagine-asparagine) (Boch et al., 2009; Moscou and Bogdanove, 2009). They recognize C, T, A, and G bases in DNA, respectively. Importantly, the specificity of RVD – base interaction is largely context-independent, which allows programming TALEs to recognize almost any sequence of interest (Boch et al., 2009). Multiple bioinformatics tools are available for TALE specificity design and off-targeting prediction. The programmable code has led to adoption of TALEs for a wide range of genome manipulations: control of gene expression, genome editing, depositing epigenetic code and DNA imaging (Becker and Boch, 2021).

In microscopy applications, engineered TALEs targeting the sequences of interest are fused to FPs or self-labelling tags on N- or C-terminus. Multiplexing with FPs of different colours is readily possible. Such reporters can be expressed *in vivo* from the transfected plasmid(s) or delivered into the cells via mRNA or protein microinjection (Boutorine et al., 2013; Miyanari et al., 2013). TALEs also bind DNA in fixed cells, which allows specificity validation of the engineered constructs via co-localization with FISH probes (Boutorine et al., 2013; Miyanari et al., 2013).

TALE-FPs were used to study structure, localization and dynamics of repetitive sequences in human, mouse, drosophila and plant cells (Boutorine et al., 2013; Fujimoto et al., 2016; Miyanari et al., 2013; Thanisch et al., 2014; Yuan et al., 2014). TALE-mClover targeted to



**Fig. 4.** DNA labelling with TALEs. **A.** Overall structure of a complex between DNA and synthetic TALE dHax3 (pdb code 3v6t (Deng et al., 2012)). 34-aminoacid long RVDs are coloured from N- to C-end from blue to red. **B.** Interaction of three RVDs with DNA bases. A possible van de Waals interaction between methyl group of thymine and Cα-atom of glycine is shown. Serine and aspartate form H-bonds (cyan) to respective DNA bases. Note, that residues at position 12 from the shown RVDs (two asparagines and histidine) point away from DNA and are important for stabilization of the local conformation. Figures were prepared with UCSF Chimera (Pettersen et al., 2004). **C.** Schematic representation of telomere-specific TALE-HaloTag fusion and STED microscopy image of telomere staining in U-2 OS cells expressing this construct (Bucevičius et al., 2019). Note, that the first T of the target is recognized by the TALE N-terminus, and not by RVD. HaloTag signal is shown in red-hot, DNA stained with 5-580CP-Hoechst is shown in grayscale. Scale bar – 5  $\mu$ m.

major satellite repeats allowed investigation of phase separation in heterochromatin in mouse embryonic stem cells (Novo et al., 2022). Establishing stable *Arabidopsis* cell lines expressing TALE-FPs targeted to centromeres, telomeres and rRNA genes allowed imaging of these sequences in multiple plant tissues – roots, hypocotyls, leaves, and flowers (Fujimoto et al., 2016). TALE-FPs turned out to be sensitive indicators of size and structure of the repetitive loci: staining intensity correlated with telomere length in a series of human cell lines (Boutorine et al., 2013); expansion of compact signals was observed during replication of satellite DNA in *drosophila* cells (Yuan et al., 2014) and as little as one polymorphism in 15 bp target was sufficient to discriminate against TALE-FP binding (Miyanari et al., 2013). Thereby, TALE-FPs helped to confirm previous knowledge and provided new insights into dynamics of repetitive loci.

In *Arabidopsis* and human cells, the observed number of telomeres were consistently lower than expected from the respective karyotypes (Boutorine et al., 2013; Fujimoto et al., 2016), suggesting telomere clustering that was also observed by other methods (Adam et al., 2019). Super-resolution fluorescence microscopy imaging with TALE-based probes would provide a good tool for investigating telomere-clustering mechanisms *in vivo*. Proof-of-principle experiment of telomere imaging with TALE-HaloTag by STED microscopy already has been demonstrated (Fig. 4C) (Bucevičius et al., 2019). A very extensive study targeted TALEs to telomeric, centromeric, and ribosomal DNA loci in a series of tumor cell lines, stem and differentiated cells and even in the living mice (Ren et al., 2017). FP-TALEs efficiently detected a well-known telomere attrition upon aging and in the premature aging models. For the first time this work demonstrated attrition of ribosomal DNA repeats as a molecular marker for human aging. This study has some interesting technical aspects. It rigorously validated TALE specificity by FISH, expressed FP-TALEs as fusions with thioredoxin to prevent their aggregation *in vivo* and compared telomere staining in Henrietta Lacks (HeLa) cells by EGFP-TALE and dCas9-sgRNA. The latter experiment found EGFP-TALE signal stronger and with lower background. The structure of nucleolar organizer region-related ribosomal DNAs was investigated with structured illumination microscopy-transmission electron microscopy (SIM-TEM) (Ren et al., 2017).

5-methylcytosine (5 mC) is a prominent epigenetic mark in higher eukaryotes, with >70% of CpG dinucleotides methylated (Jaenisch and Bird, 2003). The canonical RVD recognizing cytosine (HD), is sensitive

to 5 mC, which may compromise TALE binding to CpG containing sequences (Bultmann et al., 2012). This can be overcome by substituting HD with a truncated artificial RVD of a lower specificity that can accommodate 5 mC (Giess et al., 2018; Valton et al., 2012). Furthermore, comparison of staining by TALEs with canonical and lower specificity RVD at the position of interest can be used to probe methylation status of a particular CpG. This approach was validated in a model system – the cells expressing DNA methyltransferase targeted to SATIII sequences (Munoz-Lopez et al., 2020). Interestingly, RVD with a clear preference to 5 mC has been engineered recently and, in theory, could be used to report directly the presence of 5 mC at the target CpG (Munoz-Lopez et al., 2021). It remains to be determined whether such tools can be useful outside the model experiments.

Most of the studies using TALEs focus on imaging the repetitive sequences, where attracting multiple copies of the reporter ensures high signal-to-background ratio. However, even this might not be sufficient: when labelling plant telomere sequences, one of the TALEs employed had to be fused to three copies of GFP to provide sufficient signal (Fujimoto et al., 2016). To our best knowledge, so far only one study succeeded in visualizing a non-repetitive locus – a single integrated HIV provirus (Ma et al., 2017). In order to discriminate between bound and unbound reporters, a pair of TALEs targeting two separate HIV loci were constructed, expressed with two different tags in cells and conjugated with quantum dots of different colour *in vivo* via biorthogonal ligation reactions. Co-localization of both colours in 3D images indicated the presence of HIV provirus. Consistent with FISH data,  $1.5 \pm 0.5$  viruses per cell were found in chronically infected HIV-1 promonocytic (U1) cells.

In summary, TALEs are single-component genetically encoded constructs that work in living and fixed cells with little restrictions on target sequence. Programming TALEs is rather straightforward and they can be delivered to the cell by transfection, mRNA or protein injection and imaging can be multiplexed easily. Using TALE-FPs has already provided novel insights into biology of repetitive genomic sequences, however popularity of this tool in imaging community remains limited. Due to a highly repetitive sequence, constructing TALEs is not amenable to conventional cloning techniques and relies on sets of preassembled building blocks that are combined together by a series of restriction-ligation reactions, Golden Gate or ligation-independent cloning (see (Sakuma and Yamamoto, 2017) for overview). Several such sets are available from the

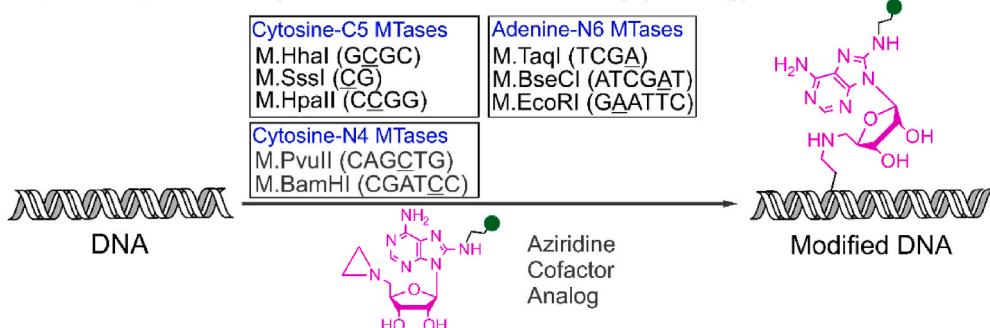
non-profit plasmid repository Addgene (Kamens, 2015). Existence of such kits allow to streamline TALE assembly, however, establishing and maintaining collections of building blocks is a labour-intense task with little appeal for those who prefer focusing on biological questions rather than on tool development. Achieving good image contrast and visualizing non-repetitive genome loci is still an unsolved challenge. Signal amplification methods, similar to those applied to dCas9 protein, need to be established in order for TALEs to be adopted by a wider community.

### 3.3. DNA labelling with methyltransferases

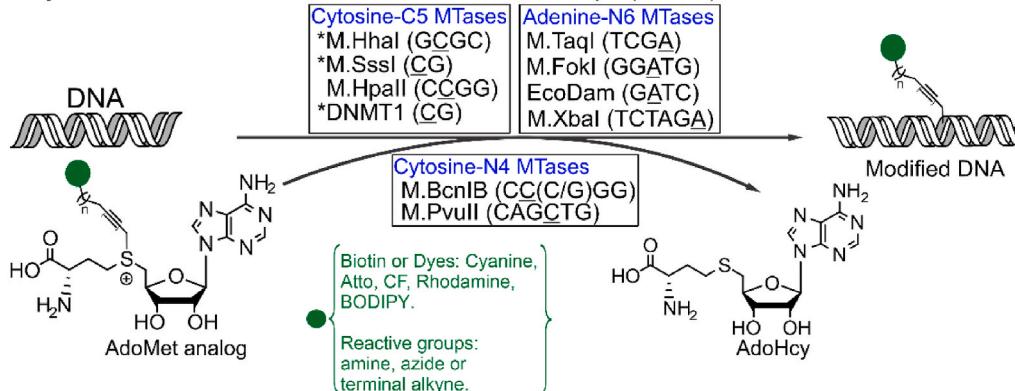
Methylation of DNA is an essential reaction for many biological

processes in both mammals and prokaryotes (Jeltsch et al., 1999). This reaction is performed by the DNA methyltransferases (MTases), which recognize specific 2–8 bp target sequences and use a cofactor S-adenosylmethionine (AdoMet) as the methyl donor to modify a cytosine or adenine therein (Fauman et al., 1999; Katz et al., 2003; Szulik et al., 2020). Based on their target nucleobase DNA MTases are divided into three classes: adenine-N<sup>6</sup>, cytosine-N<sup>4</sup>, and cytosine-C<sup>5</sup> (Jeltsch and Jurkowska, 2016). The reaction in the active site proceeds via bimolecular nucleophilic substitution ( $S_N2$ ) mechanism (Wu and Santi, 1987). This reaction can be utilized to transfer groups larger than methyl to functionalize DNA (Liutkevičiūtė et al., 2009; Lukinavičius et al., 2007; Pljevaljčić et al., 2004) (Fig. 5A).

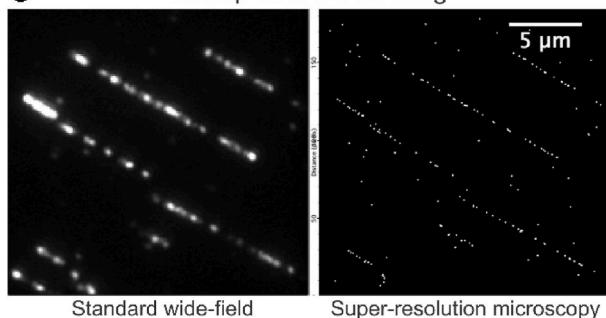
#### A Sequence-specific Methyltransferase-Induced Labelling (SMILing)



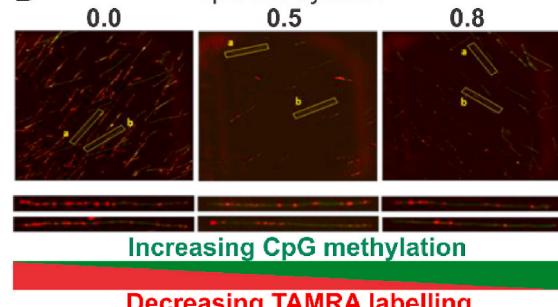
#### B Methyltransferase-directed Transfer of Activated Groups (mTAG)



#### C M.TaqI-directed labelling



#### D CpG methylation



**Fig. 5. DNA labelling with DNA Methyltransferases.** A. DNA labelling with aziridine cofactors - the whole cofactor containing a fluorophore (reporter group) is attached to the target site. The product complex containing labelled DNA and MTase cannot dissociate resulting in only a single turnover. Modified base is underlined. B. DNA labelling with mTAG – MTases catalyse transfer of a side chain containing reporter groups (fluorophore or biotin) or reactive moiety (azide or terminal alkyne). The product cofactor S-adenosyl-L-homocysteine (AdoHcy) is not attached to DNA resulting in multiple turnover possibility. Modified base is underlined. \* refers to engineered MTase. C. Standard and single molecule localization super-resolution microscopy images for M. TaqI-directed labelling of T7 bacteriophage genome. Scale bar – 5  $\mu$ m. Image adapted from (Vranken et al., 2014) under CC BY 3.0 license. D. Fluorescence microscopy images of DNA from c. DNA were combed using mixed hydrophilic/hydrophobic surface-modified cover slips. The green channel shows the unspecific DNA intercalator YOYO-1 (a cyanine dye with excitation: 488 nm) and the red channel shows the TAMRA labelling (a rhodamine dye with excitation: 561 nm). The highest number of labels was obtained for non-methylated DNA (left panel) while the number of labels decreases with increasing CpG methylation levels (middle to right panel). Figure adapted from (Sharim et al., 2019) under CC BY-NC 4.0 license.

The first successful attempt to use MTases for DNA functionalization used aziridine cofactors (Pljevaljcic et al., 2003). In this method called Sequence-Specific Methyltransferase-induced Labelling (SMILing) MTase facilitates covalent coupling of the entire cofactor to the DNA target site via aziridine-ring opening (Fig. 5A) (Pljevaljcic et al., 2003, 2004; Zhang et al., 2006). However, this approach produces the locked DNA-MTase-cofactor complex and results in a single turnover reaction (Fig. 5A). Several years later a collaborative work between Weinhold and Klimašauskas groups showed that a double or a triple carbon-carbon bond at  $\beta$ -position to sulfonium centre facilitates the transfer of groups larger than methyl to specific DNA sequence sites by the MTases (Dalhoff et al., 2006). This method is known as “MTase-directed Transfer of Activated Groups (mTAG)” (Fig. 5B) (Lukinavičius et al., 2007). In contrast to SMILING, mTAG produces labile DNA-MTase-cofactor complex and allows multiple enzyme turnovers.

Cofactor binding pocket of MTase TaqI (M. TaqI) is extremely promiscuous, thus it is highly active with a wide range of AdoMet analogues and remains the enzyme of choice in multiple studies (Dalhoff et al., 2006; Weinhold and Chakraborty, 2021). However, the activity of other wild-type (wt) enzymes with AdoMet analogues is low to moderate (Kriukienė et al., 2013; Lukinavičius et al., 2012; Vranken et al., 2014). In some cases, engineering of the cofactor binding pocket can be employed to enable utilization of AdoMet analogues, as demonstrated for the enzymes M.HhaI, M.SspI, M.TaqI and eukaryotic Dnmt1 (Deen et al., 2017; Heimes et al., 2018; Kriukienė et al., 2013; Lukinavičius et al., 2007; Stankevičius et al., 2022; Staševskij et al., 2017). These engineering efforts are facilitated by a highly conserved structure of DNA MTases and extensive understanding of their reaction mechanism (Gerasimaitė et al., 2009; Lukinavičius et al., 2012; Merkienė and Kli-  
mašauskas, 2005).

MTase-directed labelling can have various applications, including imaging of DNA via attachment of fluorophores to target sites, functionalization of plasmid DNA for gene delivery, mapping epigenetic modifications and genotyping (Pljevaljcic et al., 2004). Most common application of DNA MTases is in the field of DNA optical mapping (OM). Established by Schwartz et al., in 1993, it is a microscope-based technique where MTases are used to fluorescently label specific DNA sequences and their images are used to create one dimensional maps of stretched DNA molecules (Schwartz et al., 1993; Yuan et al., 2020). This part of the method is well developed and commercialized thanks to the advancements in nano- and microfluidics (Yuan et al., 2020). This is a powerful tool complementing whole genome sequencing as it aids genome assembly by providing a scaffold for aligning repetitive regions (Bogas et al., 2017). It is also shown to be helpful in discovering and characterizing rare structural variants in heterogeneous tumor samples (Bocklandt et al., 2019; Lam et al., 2012). In addition, possibility to use optical mapping for virus identification has been explored (Wand et al., 2019) and its utility for taxonomic analysis of bacterial strains has been demonstrated (D'Huys et al., 2021). Most recently, combining OM and 523-gene next generation sequencing panel (NGS) was proposed by Sahajpal et al. to provide a cost-effective way of profiling myeloid cancers to improve clinical outcomes (Sahajpal et al., 2022). Although the resolution of OM is still low, making it unsuitable for maps shorter than 100–150 kb, it is a valuable complementary method in genomic studies (Yuan et al., 2020). mTAG can be applied to epigenome studies, because activity of bacterial MTases is blocked by the presence of 5-methylcytosine in their target sequences. Consequently, they attach fluorophores only to the sites lacking 5 mC epigenetic mark. Multiple studies exploit mTAG for unmethylated DNA enrichment in sequencing-based epigenome analysis (Kriukienė et al., 2013; Stankevičius et al., 2022; Staševskij et al., 2017). The same principles can be applied also for OM. Thereby, profiling the unmethylome of human cells using M.TaqI was demonstrated (Fig. 5D) (Sharim et al., 2019).

Further important mTAG developments turned into a reversible tagging direction, which allows installation and on-demand removal of various functionalities on the natural DNA. For example, the photo-

reversible labelling, termed re-mTAG, was demonstrated using photocaged cofactor analogues and M.TaqI (Anhauser et al., 2018; Heimes et al., 2018). These groups can be used to photoregulate gene expression on the plasmid DNA target sites (Heimes et al., 2018). Recently, AdoMet analogue equipped with an acyl hydrazone linker and a terminal azide was used to tag, untag, and permanently tag DNA. It allows to sequentially introduce two fluorescent dyes or affinity tags into natural DNA (Wilkinson et al., 2020).

One of the main limiting factors for use of DNA MTases for DNA labelling and imaging in living cells is cell membrane impermeability of cofactor analogues (Sohtome et al., 2021). This issue can be overcome by employing chemo-enzymatic synthesis *in situ*. The most popular example exploits methionine adenosyltransferase (MAT) which uses adenosine triphosphate (ATP) and methionine as substrates (Fig. 5C) (Park et al., 1996; Singh et al., 2014; Vranken et al., 2016; Wang et al., 2013). Michailidou et al. demonstrated that MATs can synthesize AdoMet analogues with photocaging groups which can be used to detect the labelled DNA sites (Michailidou et al., 2021). Furthermore, propargyl AdoMet analogue can be generated from methionine analogue and ATP inside the cells (Huber et al., 2020; Sohtome et al., 2021). Alternative possibility to deliver cofactor analogues into the cell via electroporation has been demonstrated for epigenetic profiling and imaging of protein methylation sites (Doll et al., 2019; Gade et al., 2021; Stankevičius et al., 2022).

Poor stability of cofactor analogues is another obstacle hampering their use for DNA labelling and microscopy applications. There are two most common degradation pathways for these analogues: intramolecular cyclization and depurination (Huber et al., 2016). To overcome the latter, 7-deaza AdoMet analogues were created by removing nitrogen from the 7 position of the adenine to prevent depurination, and aziridine-based versions of these cofactors were used in sequence-specific labelling of DNA with M.HhaI for electron microscopy analysis (Huber et al., 2016).

In summary, MTase-based labelling of DNA is a promising technique for generating a predictable labelling pattern along large natural DNA molecules. Small label size opens the way to fully exploit the potential of super-resolution fluorescence microscopy, where localization precision reaches down to nanometres and structure distortion by a label becomes a potential issue. However, there is still a long way to go before it becomes a benchmark technique for microscopy.

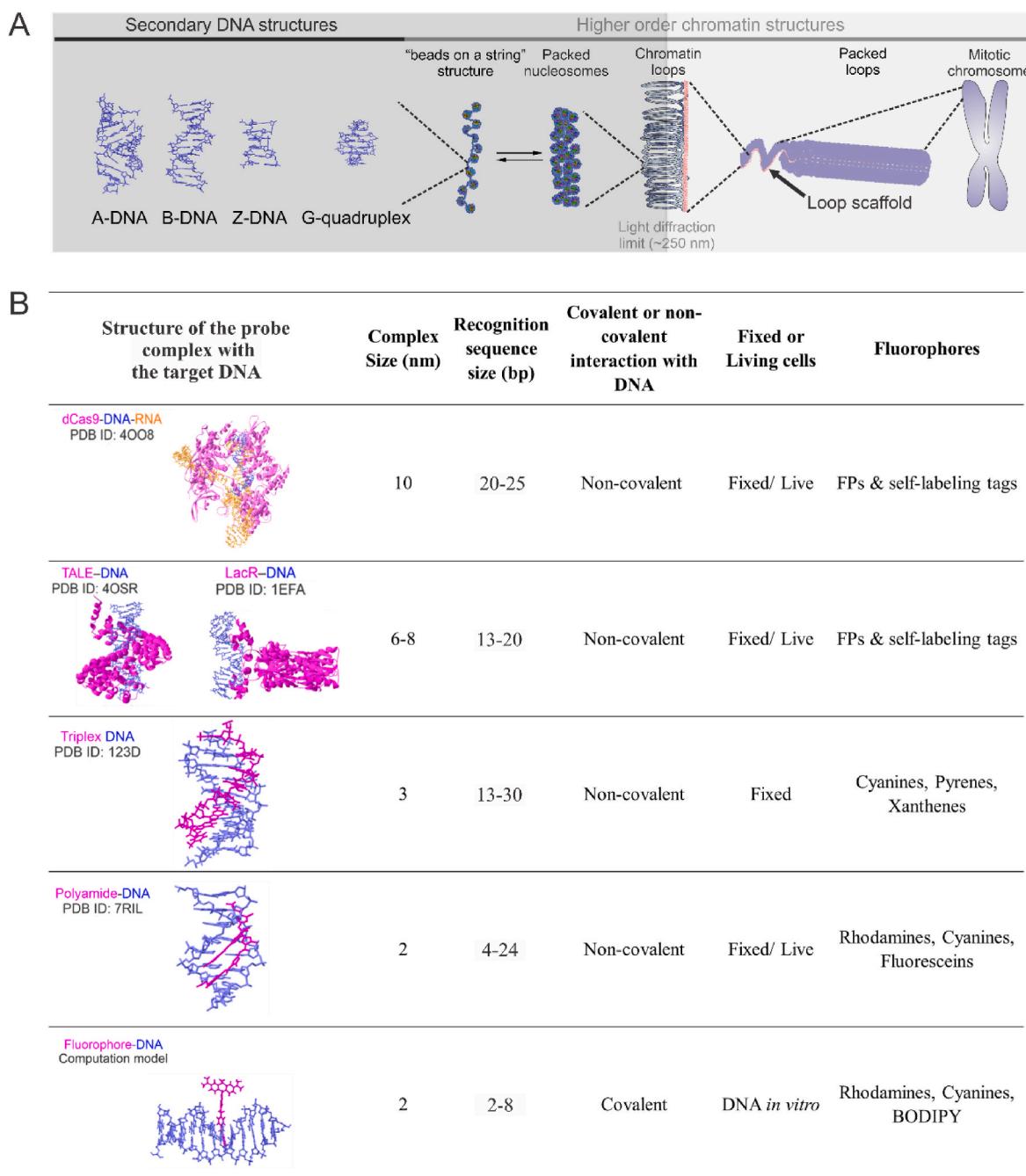
#### 4. Conclusions and future perspectives

Site-specific DNA labelling which introduces minimal perturbations to the DNA double helix is indispensable tool for chromatin research (Boettiger and Murphy, 2020). The advantages of these techniques become apparent in super-resolution fluorescence microscopy images. Long standing denaturation-based DNA labelling methods are slowly being substituted by minimally disruptive methods based on triplex forming polyamides and TFOs. Recent advances in genome editing provided researchers with programmable proteins dCas9 and TALEs, which offer a lot of flexibility for both *in vivo* and *in vitro* imaging. Despite multiple developments, sequence-specific DNA labelling is far from ideal. Limitations shared among most labelling methods include off-targeting, probe aggregation *in vivo*, as well as the risk of disrupting the local chromatin structure and affecting the process under study. The methods discussed in this review can readily detect repetitive DNA loci, such as telomeres and centromeres, but they often struggle with visualization of single-copy sequences. It is worth noting, that dCas9-based methods are the most advanced towards this direction. Thanks to the elaborate signal amplification approaches, they demonstrate potential to replace classical FISH with mild highly specific CasFISH.

All the aforementioned developments are resonating with current advances of the of super-resolution fluorescence microscopy methods which are achieving astonishing sub-nanometre (nm) localization precision of fluorophores in complex biological samples (Balzarotti et al.,

2017; Schmidt et al., 2021; Weber et al., 2021, 2022). This could allow identification of DNA conformations and structural forms (A- and B-DNA, left-handed Z-DNA, cruciforms, intramolecular triplexes, quadruplex DNA, slipped-strand DNA, parallel-stranded DNA, and unpaired DNA) provided that appropriate labelling scheme is available (Fig. 6A). However, the Nyquist criterion has to be satisfied to take full advantage of the resolving power, meaning that the labelling density should be at least 2- or 3-fold higher than the resolution of microscope. This imposes physical limitations on the label size which must allow sufficient spacing between separate reporter molecules (Liu et al., 2022). For example, probes with sizes of few nanometres are required for MINFLUX (minimal photon fluxes) or MINSTED (minimal photon fluxes – stimulated emission depletion) microscopes that attain resolution in the range of 1–2 nm (Balzarotti et al., 2017; Weber et al., 2021, 2022). One could think that

labelling each base pair on the DNA would be beneficial because of high labelling density (every 0.34 nm), but this would certainly destroy the structure of interest by distorting DNA geometry and interfering with protein-DNA interactions. On other hand, these imaging techniques allow extremely fast tracking of sparsely labelled structures. This can be achieved using all sequence-specific DNA labelling methods discussed in this review (Fig. 6B). Furthermore, adapting optical mapping principles to high-resolution 3D images of the entire nucleus is capable of endowing the microscopy images with the sequence context. However, this is a truly challenging task that will require optimization of labelling and super-resolution imaging regimes to be able to reliably resolve and register individual fluorophores or fluorophore groups in a very densely packed nucleus, as well as sophisticated machine-learning approaches for data processing.



**Fig. 6. Chromatin organization levels and sequence-specific DNA labelling approaches.** A. Observation of multiple DNA forms and chromatin structures is hindered by the light diffraction limit. B. Comparison of the probes for sequence-specific DNA labelling. FP = fluorescent protein; BODIPY = boron-dipyrromethene.

For the future, we anticipate that the research community seeking to unravel molecular mechanisms underlying chromatin functions will adopt the highlighted methods more and more. Particularly powerful approach would be combination of super-resolution microscopy and sequencing methods which allows precise positioning of specific DNA sequences in the spatial context of the nucleus. We envision that synergistic approaches would allow visualization of the fine details of chromatin structure not only in single cells seeded on glass coverslips, but also in tissues or even whole organs.

### CRediT authorship contribution statement

**Shalini Pradhan:** Writing – review & editing. **Sinem Apaydin:** Writing – review & editing. **Jonas Bucevičius:** Writing – review & editing, Visualization. **Rūta Gerasimaitė:** Writing – original draft, Conceptualization. **Georgij Kostiuk:** Writing – review & editing. **Gražvydas Lukinavičius:** Writing – original draft, Conceptualization, Supervision.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

No data was used for the research described in the article.

### Acknowledgements

The funding is provided by the Max Planck Society. SP and SA were supported by the International Max Planck Research School for Genome Science, part of the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences. The authors are grateful to Van-Thang Nguyen for critical reading of the manuscript.

### References

- Adam, N., Degelman, E., Briggs, S., Wazen, R.M., Colarusso, P., Riabowol, K., Beattie, T., 2019. Commun. Biol. 2, 1–12.
- Anhauser, L., Muttach, F., Rentmeister, A., 2018. Chem. Commun. 54, 449–451.
- Asensio, J.L., Carr, R., Brown, T., Lane, A.N., 1999. J. Am. Chem. Soc. 121, 11063–11070.
- Baird, E.E., Dervan, P.B., 1996. J. Am. Chem. Soc. 118, 6141–6146.
- Balzarotti, F., Eilers, Y., Gwosch, K.C., Gynna, A.H., Westphal, V., Stefani, F.D., Elf, J., Hell, S.W., 2017. Science 355, 606–612.
- Barth, R., Bystricky, K., Shaban, H.A., 2020. Sci. Adv. 6, eaaz2196.
- Becker, S., Boch, J., 2021. Gene and Genome Editing 2, 100007.
- Bentin, T., Nielsen, P.E., 1999. In: Malvy, C., Harel-Bellan, A., Pritchard, L.L. (Eds.), Triple Helix Forming Oligonucleotides. Springer US, Boston, MA, pp. 245–255.
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U., 2009. Science 326, 1509–1512.
- Bocklandt, S., Hastie, A., Cao, H., 2019. In: Suzuki, Y. (Ed.), Single Molecule and Single Cell Sequencing. Springer Singapore, Singapore, pp. 97–118.
- Boettiger, A., Murphy, S., 2020. Trends Genet. 36, 273–287.
- Bogas, D., Nyberg, L., Pacheco, R., Azevedo, N.F., Beech, J.P., Gomila, M., Lalucat, J., Manaia, C.M., Nunes, O.C., Tegenfeldt, J.O., Westerlund, F., 2017. Biotechniques 62, 255–267.
- Bogdanov, A.J., Schornack, S., Lahaye, T., 2010. Curr. Opin. Plant Biol. 13, 394–401.
- Boopathi, R., Dimitrov, S., Hamiche, A., Petosa, C., Bednar, J., 2020. Curr. Opin. Struct. Biol. 64, 97–103.
- Boutorine, A.S., Novopashina, D.S., Krasheninnikova, O.A., Nozeret, K., Venyaminova, A.G., 2013. Molecules 18, 15357–15397.
- Branda, H.B., Gabriele, M., Hansen, A.S., 2021. Curr. Opin. Cell Biol. 70, 18–26.
- Brown, J.M., De Ornellas, S., Parisi, E., Schermelleh, L., Buckle, V.J., 2022. Nat. Protoc. 17, 1306–1331.
- Brown, J.M., Roberts, N.A., Graham, B., Waithe, D., Lagerholm, C., Telenius, J.M., De Ornellas, S., Oudelaar, A.M., Scott, C., Szczerbala, I., Babbs, C., Kassouf, M.T., Hughes, J.R., Higgs, D.R., Buckle, V.J., 2018. Nat. Commun. 9, 1–15.
- Bucevičius, J., Keller-Findeisen, J., Gilat, T., Hell, S.W., Lukinavičius, G., 2019. Chem. Sci. 10, 1962–1970.
- Bultmann, S., Morbitzer, R., Schmidt, C.S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T., Leonhardt, H., 2012. Nucleic Acids Res. 40, 5368–5377.
- Chaudhary, N., Nho, S.H., Cho, H., Gantumur, N., Ra, J.S., Myung, K., Kim, H., 2020. Genome Res. 30, 1306–1316.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., Huang, B., 2013. Cell 155, 1479–1491.
- Cheng, A.W., Jillette, N., Lee, P., Plaskon, D., Fujiwara, Y., Wang, W., Taghbalout, A., Wang, H., 2016. Cell Res. 26, 254–257.
- Clow, P.A., Du, M., Jillette, N., Taghbalout, A., Zhu, J.J., Cheng, A.W., 2022. Nat. Commun. 13, 1871.
- Coll, M., Frederick, C.A., Wang, A.H., Rich, A., 1987. Proc. Natl. Acad. Sci. U. S. A 84, 8385–8389.
- Cooney, M., Czernuszewicz, G., Postel, E.H., Flint, S.J., Hogan, M.E., 1988. Science 241, 456–459.
- Cuenoud, B., Casset, F., Hüsken, D., Natt, F., Wolf, R.M., Altmann, K.-H., Martin, P., Moser, H.E., 1998. Angew. Chem. Int. Ed. Engl. 37, 1288–1291.
- Dalhoff, C., Lukinavičius, G., Klimasauskas, S., Weinhold, E., 2006. Nat. Chem. Biol. 2, 31–32.
- Deen, J., Vranken, C., Leen, V., Neely, R.K., Janssen, K.P.F., Hofkens, J., 2017. Angew. Chem. Int. Ed. Engl. 56, 5182–5200.
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., Yan, N., 2012. Science 335, 720–723.
- Deng, W., Shi, X., Tjian, R., Lionnet, T., Singer, R.H., 2015. Proc. Natl. Acad. Sci. U. S. A 112, 11870–11875.
- Dervan, P.B., 2001. Bioorg. Med. Chem. 9, 2215–2235.
- D'Huys, L., Vitale, R., Ruppeka-Rupeika, E., Goovaerts, V., Ruckebusch, C., Hofkens, J., 2021. ACS Omega 6, 21276–21283.
- Doll, F., Steimbach, R.R., Zumbusch, A., 2019. Chembiochem 20, 1315–1325.
- Doudna, J.A., Charpentier, E., 2014. Science 346, 1077–1087.
- Faria, M., Giovannangeli, C., 2001. J. Gene Med. 3, 299–310.
- Faruqi, A.F., Krawczyk, S.H., Matteucci, M.D., Glazer, P.M., 1997. Nucleic Acids Res. 25, 633–640.
- Fauman, E.B., Blumenthal, R.M., Cheng, X., 1999. In: Cheng, X., Blumenthal, R.M. (Eds.), S-Adenosylmethionine-Dependent Methyltransferases. World Scientific, Singapore, pp. 1–38.
- Felsenfeld, G., Davies, D.R., Rich, A., 1957. J. Am. Chem. Soc. 79, 2023–2024.
- Flierl, A., Jackson, C., Cottrell, B., Murdock, D., Seibel, P., Wallace, D.C., 2003. Mol. Ther. 7, 550–557.
- Frank-Kamenetskii, M.D., Mirkin, S.M., 1995. Annu. Rev. Biochem. 64, 65–95.
- Fu, Y., Rocha, P.P., Luo, V.M., Raviram, R., Deng, Y., Mazzoni, E.O., Skok, J.A., 2016. Nat. Commun. 7, 1–18.
- Fujimoto, S., Sugano, S.S., Kuwata, K., Osakabe, K., Matsunaga, S., 2016. J. Exp. Bot. 67, 6101–6110.
- Gade, M., Tan, L.L., Damry, A.M., Sandhu, M., Brock, J.S., Delaney, A., Villar-Briones, A., Jackson, C.J., Laurino, P., 2021. JACS Au 1, 2349–2360.
- Gasiūnas, G., Barrangou, R., Horvath, P., Siksnys, V., 2012. Proc. Natl. Acad. Sci. U. S. A 109, E2579–E2586.
- Geng, Y., Pertisnidis, A., 2021. Sci. Rep. 11, 12220.
- Gerasimaitė, R., Vilkauskaitė, G., Klimasauskas, S., 2009. Nucleic Acids Res. 37, 7332–7341.
- Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H.A., Sellou, H., Politis, A.Z., Ellenberg, J., Gallardo, F., Bystricky, K., 2017. Biophys. J. 113, 1383–1394.
- Giess, M., Witte, A., Jasper, J., Koch, O., Summerer, D., 2018. J. Am. Chem. Soc. 140, 5904–5908.
- Gottesfeld, J.M., Neely, L., Trauger, J.W., Baird, E.E., Dervan, P.B., 1997. Nature 387, 202–205.
- Hansen, M.E., Bentin, T., Nielsen, P.E., 2009. Nucleic Acids Res. 37, 4498–4507.
- Hargrove, A.E., Martinez, T.F., Hare, A.A., Kurmis, A.A., Phillips, J.W., Sud, S., Pienta, K.J., Dervan, P.B., 2015. PLoS One 10, e0143161.
- Heimes, M., Kolmar, L., Briuke, C., 2018. Chem. Commun. 54, 12718–12721.
- Heun, P., Laroche, T., Shimada, K., Furrer, P., Gasser, S.M., 2001. Science 294, 2181–2186.
- Hidaka, T., Pandian, G.N., Taniguchi, J., Nobeyama, T., Hashiya, K., Bando, T., Sugiyama, H., 2017. J. Am. Chem. Soc. 139, 8444–8447.
- Hirata, A., Nohikura, K., Kawamoto, Y., Bando, T., Sasaki, A., Ide, S., Maeshima, K., Kasama, T., Sugiyama, H., 2014. J. Am. Chem. Soc. 136, 11546–11554.
- Hong, Y., Lu, G., Duan, J., Liu, W., Zhang, Y., 2018. Genome Biol. 19, 1–10.
- Huber, T.D., Clinger, J.A., Liu, Y., Xu, W., Miller, M.D., Phillips Jr., G.N., Thorson, J.S., 2020. ACS Chem. Biol. 15, 695–705.
- Huber, T.D., Johnson, B.R., Zhang, J., Thorson, J.S., 2016. Curr. Opin. Biotechnol. 42, 189–197.
- Inoue, T., Shimozato, O., Matsuo, N., Mori, Y., Shinozaki, Y., Lin, J., Watanabe, T., Takatori, A., Koshikawa, N., Ozaki, T., Nagase, H., 2018. Bioorg. Med. Chem. 26, 2337–2344.
- Ishii, T., Schubert, V., Khosravi, S., Dreissig, S., Metje-Sprink, J., Sprink, T., Fuchs, J., Meister, A., Houben, A., 2019. New Phytol. 222, 1652–1661.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. J. Bacteriol. 169, 5429–5433.
- Jaeisch, R., Bird, A., 2003. Nat. Genet. 33 (Suppl. 1), 245–254.
- Janssen, S., Durussel, T., Laemmli, U.K., 2000. Mol. Cell 6, 999–1011.
- Jeltsch, A., Christ, F., Fatemi, M., Roth, M., 1999. J. Biol. Chem. 274, 19538–19544.
- Jeltsch, A., Jurkowska, R.Z., 2016. In: Jeltsch, A., Jurkowska, R.Z. (Eds.), DNA Methyltransferases - Role and Function. Springer International Publishing, Cham, Switzerland, pp. 1–17.
- Jerkovic, I., Cavalli, G., 2021. Nat. Rev. Mol. Cell Biol. 22, 511–528.
- Jimenez-Garcia, E., Vaquer, A., Espinas, M.L., Soliva, R., Orozco, M., Bernues, J., Azorin, F., 1998. J. Biol. Chem. 273, 24640–24648.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. Science 337, 816–821.

- Kamens, J., 2015. Nucleic Acids Res. 43, D1152–D1157.
- Katz, J.E., Dlakic, M., Clarke, S., 2003. Mol. Cell. Proteomics 2, 525–540.
- Kawamoto, Y., Bando, T., Kamada, F., Li, Y., Hashiya, K., Maeshima, K., Sugiyama, H., 2013. J. Am. Chem. Soc. 135, 16468–16477.
- Kawamoto, Y., Bando, T., Sugiyama, H., 2018. Bioorg. Med. Chem. 26, 1393–1411.
- Kawamoto, Y., Sasaki, A., Chandran, A., Hashiya, K., Ide, S., Bando, T., Maeshima, K., Sugiyama, H., 2016. J. Am. Chem. Soc. 138, 14100–14107.
- Kawamoto, Y., Sasaki, A., Hashiya, K., Ide, S., Bando, T., Maeshima, K., Sugiyama, H., 2015. Chem. Sci. 6, 2307–2312.
- Kempfer, R., Pombo, A., 2020. Nat. Rev. Genet. 21, 207–226.
- Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R.E., 1985. Proc. Natl. Acad. Sci. U. S. A 82, 1376–1380.
- Kriukienė, E., Labrie, V., Khare, T., Urbanavičiūtė, G., Lapinaitė, A., Koncevičius, K., Li, D., Wang, T., Pai, S., Ptak, C., Gordevičius, J., Wang, S.C., Petronis, A., Klimašauskas, S., 2013. Nat. Commun. 4, 1–10.
- Krufczik, M., Sievers, A., Hausmann, A., Lee, J.H., Hildenbrand, G., Schaufler, W., Hausmann, M., 2017. Int. J. Mol. Sci. 18, 1–16.
- Kusic, J., Tomic, B., Divac, A., Kojic, S., 2010. Mol. Biol. Rep. 37, 2317–2322.
- Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarejan, N., Xiao, M., Kwok, P.Y., 2012. Nat. Biotechnol. 30, 771–776.
- Lane, A.B., Strzelecka, M., Ettinger, A., Grenfell, A.W., Wittmann, T., Heald, R., 2015. Dev. Cell 34, 373–378.
- Lane, N., 2015. Philos. Trans. R. Soc. Lond. Ser. B. 370, 1–10.
- Lee, J.S., Johnson, D.A., Morgan, A.R., 1979. Nucleic Acids Res. 6, 3073–3091.
- Lee, S., Kawamoto, Y., Vaijayanthi, T., Park, J., Bae, J., Kim-Ha, J., Sugiyama, H., Jo, K., 2018. Nucleic Acids Res. 46, e108.
- Lee, J.-H., Tchetgna, F.L.D., Krufczik, M., Schmitt, E., Cremer, C., Bestvater, F., Hausmann, M., 2019. OEMB Genetics 3, 64.
- Li, Y., Almassalha, L.M., Chandler, J.E., Zhou, X., Stypula-Cyrus, Y.E., Hujšak, K.A., Roth, E.W., Bleher, R., Subramanian, H., Szleifer, I., Dravid, V.P., Backman, V., 2017. Exp. Cell Res. 358, 253–259.
- Lin, J., Nagase, H., 2020. Biomolecules 10, 544–563.
- Liu, S., Hoess, P., Ries, J., 2022. Annu. Rev. Biophys. 51, 301–326.
- Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D., Klimašauskas, S., 2009. Nat. Chem. Biol. 5, 400–402.
- Lown, J.W., Krowicki, K., Bhat, U.G., Skorobogaty, A., Ward, B., Dabrowski, J.C., 1986. Biochemistry 25, 7408–7416.
- Lukinavičius, G., Lapinié, V., Staševskij, Z., Dalhoff, C., Weinhold, E., Klimašauskas, S., 2007. J. Am. Chem. Soc. 129, 2758–2759.
- Lukinavičius, G., Lapinié, A., Urbanavičiūtė, G., Gerasimaitė, R., Klimašauskas, S., 2012. Nucleic Acids Res. 40, 11594–11602.
- Lyu, X.Y., Deng, Y., Huang, X.Y., Li, Z.Z., Fang, G.Q., Yang, D., Wang, F.L., Kang, W., Shen, E.Z., Song, C.Q., 2022. Cell Res. 32, 969–981.
- Ma, H., Naseri, A., Reyes-Gutiérrez, P., Wolfe, S.A., Zhang, S., Pederson, T., 2015. Proc. Natl. Acad. Sci. U. S. A 112, 3002–3007.
- Ma, H., Tu, L.C., Naseri, A., Chung, Y.C., Grunwald, D., Zhang, S., Pederson, T., 2018. Nat. Methods 15, 928–931.
- Ma, H., Tu, L.C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., Pederson, T., 2016. Nat. Biotechnol. 34, 528–530.
- Ma, Y., Wang, M., Li, W., Zhang, Z., Zhang, X., Tan, T., Zhang, X.E., Cui, Z., 2017. Nat. Commun. 8, 1–8.
- Maass, P.G., Barutcu, A.R., Shechner, D.M., Weiner, C.L., Mele, M., Rinn, J.L., 2018. Nat. Struct. Mol. Biol. 25, 176–184.
- Maciak, A., Krakowiak, A., Janicka, M., Tomaszewska-Antczak, A., Sobczak, M., Mikolajczyk, B., Guga, P., 2015. Org. Biomol. Chem. 13, 2375–2384.
- Maeshima, K., Ide, S., Babkov, M., 2019. Curr. Opin. Cell Biol. 58, 95–104.
- Maeshima, K., Janssen, S., Laemmli, U.K., 2001. EMBO J. 20, 3218–3228.
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V., 2006. Biol. Direct 1, 1–26.
- Mapp, A.K., Ansari, A.Z., Ptashne, M., Dervan, P.B., 2000. Proc. Natl. Acad. Sci. U. S. A 97, 3930–3935.
- Markaki, Y., Smets, D., Fiedler, S., Schmid, V.J., Schermelleh, L., Cremer, T., Cremer, M., 2012. Bioessays 34, 412–426.
- Merkinė, E., Klimašauskas, S., 2005. Nucleic Acids Res. 33, 307–315.
- Michaillidou, F., Klocker, N., Cornelissen, N.V., Singh, R.K., Peters, A., Ovcharenko, A., Kummel, D., Rentmeister, A., 2021. Angew. Chem. Int. Ed. Engl. 60, 480–485.
- Miyanari, Y., Ziegler-Birling, C., Torres-Padilla, M.E., 2013. Nat. Struct. Mol. Biol. 20, 1321–1324.
- Molenaar, C., Wiesmeijer, K., Verwoerd, N.P., Khazen, S., Eils, R., Tanke, H.J., Dirks, R. W., 2003. EMBO J. 22, 6631–6641.
- Morgan, A.R., Wells, R.D., 1968. J. Mol. Biol. 37, 63–80.
- Moscou, M.J., Bogdanov, A.J., 2009. Science 326, 1501.
- Mrksich, M., Parks, M.E., Dervan, P.B., 1994. J. Am. Chem. Soc. 116, 7983–7988.
- Müller, P., Schmitt, E., Jacob, A., Hoheisel, J., Kaufmann, R., Cremer, C., Hausmann, M., 2010. Int. J. Mol. Sci. 11, 4094–4105.
- Munoz-Lopez, A., Buchmuller, B., Wolffgramm, J., Jung, A., Hussong, M., Kanne, J., Schweiger, M.R., Summerer, D., 2020. Angew. Chem. Int. Ed. Engl. 59, 8927–8931.
- Munoz-Lopez, A., Jung, A., Buchmuller, B., Wolffgramm, J., Maurer, S., Witte, A., Summerer, D., 2021. Chembiochem 22, 645–651.
- Musso, M., Bianchi-Scarrà, G., Van Dyke, M.W., 2000. Nucleic Acids Res. 28, 4090–4096.
- Nemeckova, A., Wasch, C., Schubert, V., Ishii, T., Hribova, E., Houben, A., 2019. Cytogenet. Genome Res. 159, 48–53.
- Nickols, N.G., Jacobs, C.S., Farkas, M.E., Dervan, P.B., 2007. Nucleic Acids Res. 35, 363–370.
- Noonberg, S.B., Francois, J.C., Garestier, T., Helene, C., 1995. Nucleic Acids Res. 23, 1956–1963.
- Novo, C.L., Wong, E.V., Hockings, C., Poudel, C., Sheekey, E., Wiese, M., Okkenhaug, H., Boulton, S.J., Basu, S., Walker, S., Kaminski Schierle, G.S., Narlikar, G.J., Rugg-Gunn, P.J., 2022. Nat. Commun. 13, 3525.
- Nozeret, K., Loll, F., Escude, C., Boutorine, A.S., 2018. Biochimie 149, 122–134.
- Nozeret, K., Loll, F., Escude, C., Boutorine, A.S., 2015. Chembiochem 16, 549–554.
- Olivas, W.M., Maher 3rd, L.J., 1995. Biochemistry 34, 278–284.
- Oyoshi, T., Kawakami, W., Narita, A., Bando, T., Sugiyama, H., 2003. J. Am. Chem. Soc. 125, 4752–4754.
- Pandian, G.N., Nakano, Y., Sato, S., Morinaga, H., Bando, T., Nagase, H., Sugiyama, H., 2012. Sci. Rep. 2, 1–8.
- Park, J., Tai, J., Roessner, C.A., Scott, A.I., 1996. Bioorg. Med. Chem. 4, 2179–2185.
- Pelton, J.G., Wemmer, D.E., 1989. Proc. Natl. Acad. Sci. U. S. A 86, 5723–5727.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. J. Comput. Chem. 25, 1605–1612.
- Plejvalcic, G., Pignot, M., Weinhold, E., 2003. J. Am. Chem. Soc. 125, 3486–3492.
- Plejvalcic, G., Schmidt, F., Weinhold, E., 2004. Chembiochem 5, 265–269.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Cell 152, 1173–1183.
- Qin, P., Parlak, M., Kuscu, C., Bandaria, J., Mir, M., Szlachta, K., Singh, R., Darzaz, X., Yıldız, A., Adli, M., 2017. Nat. Commun. 8, 1–10.
- Rahman, S.M.A., Seki, S., Obika, S., Haitani, S., Miyashita, K., Imanishi, T., 2007. Angew. Chem. Int. Ed. Engl. 46, 4306–4309.
- Ren, R., Deng, L., Xue, Y., Suzuki, K., Zhang, W., Yu, Y., Wu, J., Sun, L., Gong, X., Luan, H., Yang, F., Ju, Z., Ren, X., Wang, S., Tang, H., Geng, L., Zhang, W., Li, J., Qiao, J., Xu, T., Qu, J., Liu, G.H., 2017. Cell Res. 27, 483–504.
- Richier, B., Salecker, I., 2015. Wiley Interdiscip. Rev. Dev. Biol. 4, 161–180.
- Sahajpal, N.S., Mondal, A.K., Ananth, S., Saul, D., Shams, S., Hastie, A.R., Savage, N.M., Kota, V., Chaubey, A., Kolhe, R., 2022. medRxiv, pp. 1–22.
- Sakuma, T., Yamamoto, T., 2017. Methods Mol. Biol. 1630, 25–36.
- Sasaki, S., Taniguchi, Y., Takahashi, R., Senko, Y., Kodama, K., Nagatsugi, F., Maeda, M., 2004. J. Am. Chem. Soc. 126, 516–528.
- Sato, T., Sato, Y., Nishizawa, S., 2016. J. Am. Chem. Soc. 138, 9397–9400.
- Schmidt, R., Weihls, T., Wurm, C.A., Jansen, I., Rehman, J., Sahl, S.J., Hell, S.W., 2021. Nat. Commun. 12, 1–12.
- Schmitt, E., Schwarz-Finstreiter, J., Stein, S., Boxler, C., Müller, P., Mokhir, A., Krämer, R., Cremer, C., Hausmann, M., 2010. In: Bridger, J.M., Volpi, E.V. (Eds.), Fluorescence in Situ Hybridization (FISH): Protocols and Applications. Humana Press, Totowa, NJ, pp. 185–202.
- Schmitt, E., Wagner, J., Hausmann, M., 2012. J. Comput. Sci. 3, 328–334.
- Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J., Wang, Y.K., 1993. Science 262, 110–114.
- Schwarz-Finstreiter, J., Stein, S., Grossmann, C., Schmitt, E., Trakhtenbrot, L., Rechavi, G., Amariglio, N., Cremer, C., Hausmann, M., 2007. J. Biochem. Biophys. Methods 70, 397–406.
- Shao, S., Zhang, W., Hu, H., Xue, B., Qin, J., Sun, C., Sun, Y., Wei, W., Sun, Y., 2016. Nucleic Acids Res. 44, e86.
- Sharim, H., Grunwald, A., Gabrieli, T., Michaeli, Y., Margalit, S., Torchinsky, D., Arielli, R., Nifker, G., Juhasz, M., Gularek, F., Almalvez, M., Dufault, B., Chandra, S., Liu, A., Bhattacharya, S., Chen, Y.W., Vilain, E., Wagner, K.R., Pevsner, J., Reifenberger, J., Lam, E.T., Hastie, A.R., Cao, H., Barseghyan, H., Weinhold, E., Ebenstein, Y., 2019. Genome Res. 29, 646–656.
- Silver, G.C., Sun, J.-S., Nguyen, C.H., Boutorine, A.S., Bisagni, E., Hélène, C., 1997. J. Am. Chem. Soc. 119, 263–268.
- Singh, S., Zhang, J., Huber, T.D., Sunkara, M., Hurley, K., Goff, R.D., Wang, G., Zhang, W., Liu, C., Rohr, J., Van Lanen, S.G., Morris, A.J., Thorson, J.S., 2014. Angew. Chem. Int. Ed. Engl. 53, 3965–3969.
- Sohtome, Y., Shimazu, T., Shinkai, Y., Sodeoka, M., 2021. Acc. Chem. Res. 54, 3818–3827.
- Stankevičius, V., Gibas, P., Masiulionytė, B., Gasiulė, L., Masevičius, V., Klimašauskas, S., Vilkaitys, G., 2022. Mol. Cell 82, 1053–1065.
- Stašeckij, Z., Gibas, P., Gordevičius, J., Kriukienė, E., Klimašauskas, S., 2017. Mol. Cell 65, 554–564.
- Suda, T., Mishima, Y., Takayanagi, K., Asakura, H., Odani, S., Kominami, R., 1996. Nucleic Acids Res. 24, 4733–4740.
- Sun, J.S., Francois, J.C., Montenay-Garestier, T., Saison-Behmoaras, T., Roig, V., Thuong, N.T., Helene, C., 1989. Proc. Natl. Acad. Sci. U. S. A 86, 9198–9202.
- Szablowski, J.O., Raskatov, J.A., Dervan, P.B., 2016. Mol. Cancer Therapeut. 15, 608–617.
- Szulik, M.W., Davis, K., Bakhtina, A., Azarcon, P., Bia, R., Horiuchi, E., Franklin, S., 2020. Am. J. Physiol. Heart Circ. Physiol. 319, H847–H865.
- Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., Vale, R.D., 2014. Cell 159, 635–646.
- Tasan, I., Sustackova, G., Zhang, L., Kim, J., Sivaguru, M., Hamedirad, M., Wang, Y., Genova, J., Ma, J., Belmont, A.S., Zhao, H., 2018. Nucleic Acids Res. 46, e100.
- Thanisch, K., Schneider, K., Morbitzer, R., Solovei, I., Lahaye, T., Bultmann, S., Leonhardt, H., 2014. Nucleic Acids Res. 42, e38.
- Thoma, B.S., Wakasugi, M., Christensen, J., Reddy, M.C., Vasquez, K.M., 2005. Nucleic Acids Res. 33, 2993–3001.
- Tsubono, Y., Kawamoto, Y., Hidaka, T., Pandian, G.N., Hashiya, K., Bando, T., Sugiyama, H., 2020. J. Am. Chem. Soc. 142, 17356–17363.
- Vaijayanthi, T., Bando, T., Pandian, G.N., Sugiyama, H., 2012. Chembiochem 13, 2170–2185.
- Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliani, K., Juillerat, A., Duchateau, P., 2012. J. Biol. Chem. 287, 38427–38432.
- Vasquez, K.M., Glazer, P.M., 2002. Q. Rev. Biophys. 35, 89–107.

- Vranken, C., Deen, J., Dirix, L., Stakenborg, T., Dehaen, W., Leen, V., Hofkens, J., Neely, R.K., 2014. Nucleic Acids Res. 42, e50.
- Vranken, C., Fin, A., Tufar, P., Hofkens, J., Burkart, M.D., Tor, Y., 2016. Org. Biomol. Chem. 14, 6189–6192.
- Walsh, S., El-Sagheer, A.H., Brown, T., 2018. Chem. Sci. 9, 7681–7687.
- Wand, N.O., Smith, D.A., Wilkinson, A.A., Rushton, A.E., Busby, S.J.W., Styles, I.B., Neely, R.K., 2019. Nucleic Acids Res. 47, e68.
- Wang, H., Nakamura, M., Abbott, T.R., Zhao, D., Luo, K., Yu, C., Nguyen, C.M., Lo, A., Daley, T.P., La Russa, M., Liu, Y., Qi, L.S., 2019. Science 365, 1301–1305.
- Wang, R., Islam, K., Liu, Y., Zheng, W., Tang, H., Lailler, N., Blum, G., Deng, H., Luo, M., 2013. J. Am. Chem. Soc. 135, 1048–1056.
- Wang, S., Su, J.H., Zhang, F., Zhuang, X., 2016. Sci. Rep. 6, 1–7.
- Weber, M., Leutenegger, M., Stoldt, S., Jakobs, S., Mihaila, T.S., Butkevich, A.N., Hell, S. W., 2021. Nat. Photonics 15, 361–366.
- Weber, M., Von Der Emde, H., Leutenegger, M., Gunkel, P., Sambandan, S., Khan, T.A., Keller-Findeisen, J., Cordes, V.C., Hell, S.W., 2022. Nat. Biotechnol. 1–12.
- Weinhold, E., Chakraborty, B., 2021. Nanoscale 13, 2465–2471.
- Wilkinson, A.A., Jagu, E., Ubych, K., Coulthard, S., Rushton, A.E., Kennefick, J., Su, Q., Neely, R.K., Fernandez-Trillo, P., 2020. ACS Cent. Sci. 6, 525–534.
- Winkler, R., Perner, B., Rapp, A., Durm, M., Cremer, C., Greulich, K.O., Hausmann, M., 2003. J. Microsc. 209, 23–33.
- Wootton, M., Li, Y., Snedeker, J., Nizami, Z.F., Gall, J.G., Chen, X., 2020. Nat. Protoc. 15, 1188–1208.
- Wu, J.C., Santi, D.V., 1987. J. Biol. Chem. 262, 4778–4786.
- Wurtz, N.R., Turner, J.M., Baird, E.E., Dervan, P.B., 2001. Org. Lett. 3, 1201–1203.
- Xiang, W., Roberti, M.J., Heriche, J.K., Huet, S., Alexander, S., Ellenberg, J., 2018. J. Cell Biol. 217, 1973–1984.
- Xu, J., Ma, H., Jin, J., Uttam, S., Fu, R., Huang, Y., Liu, Y., 2018. Cell Rep. 24, 873–882.
- Xu, J., Ma, H., Ma, H., Jiang, W., Mela, C.A., Duan, M., Zhao, S., Gao, C., Hahm, E.R., Lardo, S.M., Troy, K., Sun, M., Pai, R., Stoltz, D.B., Zhang, L., Singh, S., Brand, R.E., Hartman, D.J., Hu, J., Hainer, S.J., Liu, Y., 2020. Nat. Commun. 11, 1–17.
- Ye, H., Rong, Z., Lin, Y., 2017. Protein Cell 8, 853–855.
- Yuan, K., Shermoen, A.W., O'farrell, P.H., 2014. Curr. Biol. 24, R144–R145.
- Yuan, Y., Chung, C.Y., Chan, T.F., 2020. Comput. Struct. Biotechnol. J. 18, 2051–2062.
- Zhang, C., Weller, R.L., Thorson, J.S., Rajski, S.R., 2006. J. Am. Chem. Soc. 128, 2760–2761.
- Zhang, S., Song, Z., 2017. J. Mol. Histol. 48, 293–299.