

REPURPOSING CAS9 AS A PROGRAMMABLE DNA-BINDING  
PLATFORM FOR GENOME VISUALIZATION, TARGETED  
EPIGENETIC MANIPULATION AND EXPLORATION OF LOCAL  
CHROMATIN COMPOSITION



TOBIAS ANTON

MÜNCHEN 2017





Erstgutachter: Prof. Dr. Heinrich Leonhardt

Zweitgutachter: Prof. Dr. Dirk Eick

Tag der Abgabe: 21.03.2017

Tag der mündlichen Prüfung: 15.05.2017

REPURPOSING CAS9 AS A PROGRAMMABLE DNA-BINDING  
PLATFORM FOR GENOME VISUALIZATION, TARGETED  
EPIGENETIC MANIPULATION AND EXPLORATION OF LOCAL  
CHROMATIN COMPOSITION



DISSERTATION  
AN DER FAKULTÄT FÜR BIOLOGIE  
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

VORGELEGT VON  
TOBIAS ANTON  
AUS DACHAU  
MÜNCHEN, DEN 21.03.2017



---

SUMMARY .....	1
ZUSAMMENFASSUNG .....	3
1 INTRODUCTION.....	5
1.1 Epigenetic information.....	5
1.2 General genome architecture.....	6
1.2.1 Hierarchical genome organization.....	6
1.2.2 Eu- and heterochromatin .....	7
1.3 Epigenetic regulation of chromatin organization.....	10
1.3.1 DNA methylation dynamics .....	10
1.3.2 Histone posttranslational modifications (PTMs).....	13
1.3.3 Crosstalk between epigenetic mechanisms: establishment of eu- and heterochromatin .....	17
1.4 3D-Organization of the genome.....	19
1.4.1 Chromosome territories.....	19
1.4.2 Subchromosomal compartments and chromosomal domains.....	20
1.5 Chromatin dynamics .....	23
1.5.1 Short- and long-range motility of chromatin.....	23
1.5.2 Visualization of chromatin dynamics .....	24
1.5.3 Modular DNA binding proteins.....	25
1.6 The CRISPR/Cas system.....	28
1.6.1 CRISPR/Cas mediated adaptive immunity .....	28
1.6.2 The Cas9 toolbox.....	33
1.8 Aims of this work .....	35
2 RESULTS .....	37
2.1 Visualization of specific DNA sequences with a programmable fluorescent CRISPR/Cas system .....	37

2.2 Visualization of genomic loci in living cells with a fluorescent CRISPR/Cas system .....	49
2.3 Site-specific recruitment of epigenetic factors with a modular CRISPR/Cas system.....	63
2.4 Determination of local chromatin composition by CasID .....	83
3 DISCUSSION .....	99
3.1 The CRISPR/Cas system as a tool to visualize chromatin .....	99
3.1.1 CRISPR imaging in comparison to modular DNA-binding proteins.....	100
3.1.2 Expanding the CRISPR imaging toolkit.....	104
3.2 Editing of epigenetic marks via the CRISPR/Cas system.....	107
3.3 Probing chromatin composition by CasID .....	108
4 ANNEX .....	115
4.1 References .....	115
4.2 Abbreviations .....	138
4.3 Statutory declaration and statement.....	141
4.4 Curriculum vitae.....	142
4.5 Contributions.....	144



## SUMMARY

In eukaryotes, cellular identity is regulated by a complex interplay between epigenetic modifications, transcription factors, and the spatiotemporal organization of chromatin. Dissecting this network requires tools that precisely target desired functions to defined genomic sequences. So far, these tools were predominantly based on polydactyl zinc finger proteins (PZFs) and transcription activator-like effectors (TALEs). Yet, designing sequence-specific PZFs and TALEs for new target sequences remains challenging.

Since 2012, the emergence of the CRISPR/Cas system revolutionized the way of site-specific genome targeting. The endonuclease Cas9, the main component of the type II CRISPR/Cas system, can be recruited to defined sequences by a single guide RNA (sgRNA) and therefore has been extensively used for genome engineering approaches. Importantly, a mutated variant of Cas9 (dCas9) lacks enzymatic activity and can be guided to specific loci without inducing DNA double strand breaks at the underlying sequence.

To test the applicability of the CRISPR/Cas system as tool to visualize chromatin dynamics, we designed sgRNAs to recruit a fluorescently tagged dCas9 to major and minor satellite sequences, as well as telomeric repeats. We demonstrated that dCas9-eGFP labeled nuclear foci coincide with known target associated proteins and expanded this approach to reveal chromatin conformation via super resolution microscopy.

Moreover, we repurposed the CRISPR/Cas system to direct GFP-tagged epigenetic effector proteins to pericentric major satellites. Based on this approach, we showed that prevalent epigenetic modifications, such as cytosine methylation and hydroxymethylation can be erased or set *de novo* by recruiting the catalytic domains of TET1 and DNMT3A, respectively.

Finally, we developed a novel method to determine the protein composition of specific genomic elements. We combined the programmable DNA-recognition of dCas9 with the promiscuous biotin ligase activity of BirA\* to biotinylate target-associated proteins, which were subsequently identified via tandem mass spectrometry. This approach allowed us to identify known and previously uncharacterized proteins, located at the close proximity of telomeres, major satellites, and minor satellites.

In conclusion, by exploiting dCas9 as a programmable DNA-binding platform, we developed novel tools for site-specific chromatin visualization, epigenome manipulation and determination of local protein composition. We anticipate that these methods provide new means to elucidate functional chromatin architecture.



## ZUSAMMENFASSUNG

Die zelluläre Identität eukaryotischer Zellen wird über komplexe Wechselwirkungen zwischen epigenetischen Modifikationen, Transkriptionsfaktoren und dynamischer Chromatinorganisation reguliert. Die Entschlüsselung dieses Netzwerks erfordert präzise molekularbiologische Methoden, die die gezielte Rekrutierung bestimmter Funktionen zu definierten genomischen Sequenzen ermöglichen. Bis vor Kurzem basierten diese Methoden auf Zinkfingerproteinen (PZFs) oder transcription activator-like effectors (TALEs). Es ist jedoch sehr zeit- und kostenintensiv sequenzspezifische PZFs und TALEs für neue Zielsequenzen zu entwerfen und herzustellen.

Im Jahr 2012 revolutionierte die Veröffentlichung des CRISPR/Cas Systems die sequenzspezifische Modifikation des Genoms. Die Endonuklease Cas9, als die Hauptkomponente des Typ II CRISPR/Cas Systems, kann mit Hilfe einer single guide RNA (sgRNA) zu beinahe jeder benutzerdefinierten DNA-Sequenz rekrutiert werden und wurde deshalb exzessiv für gezieltes Genom-Engineering verwendet. Zudem kann eine katalytisch inaktive Form des Cas9-Proteins (dCas9) an definierte Sequenzen rekrutiert werden, ohne dabei DNA-Doppelstrangbrüche zu induzieren.

Um die gezielte Visualisierung bestimmter Chromatinabschnitte mit Hilfe des CRISPR/Cas Systems zu testen, haben wir sgRNAs entworfen, über die ein fluoreszierend markiertes dCas9-Protein zu Telomeren, sowie zu Minor- und Major Satellite Sequenzen rekrutiert werden kann. Wir konnten zeigen, dass mit dCas9-eGFP markierte Foci mit Proteinen kolokalisieren, die an den entsprechenden Sequenzen angereichert sind. Darüber hinaus haben wir diese Methode für hochauflösende Mikroskopie adaptiert und konnten somit die Ultrastruktur von Telomeren und Minor Satellites enthüllen.

Durch weitere Anpassungen des CRISPR/Cas Systems konnten wir GFP-markierte epigenetische Effektoren zu definierten Sequenzen rekrutieren. Hierbei konnten wir zeigen, dass durch die Rekrutierung der katalytischen Domänen von TET1 und DNMT3A zu perizentromerischen Major Satellite Sequenzen bestehende DNA-Modifikationen entfernt bzw. neu gesetzt werden können.

Zudem haben wir eine neue Methode zur Identifikation locus-assoziiierter Proteine entwickelt. Durch die Kombination der sequenzspezifischen DNA-Bindung von dCas9 mit der ungerichteten Biotinligaseaktivität von BirA\* konnten wir Proteine biotinylieren, die mit der Zielsequenz assoziiert sind. Durch die darauffolgende massenspektrometrische Analyse dieser biotinylierten Proteine konnten wir sowohl bekannte, als auch unbekannte Faktoren von Telomeren, Major und Minor Satellites bestimmen.

Zusammenfassend haben wir, basierend auf dem CRISPR/Cas System, neue Methoden entwickelt um definierte DNA-Sequenzen zu visualisieren, deren epigenetische Modifikationen gezielt zu verändern, sowie ihre Proteinzusammensetzung zu entschlüsseln. Wir erwarten, dass diese Methoden zur weiteren Charakterisierung funktioneller Chromatinarchitektur beitragen werden.

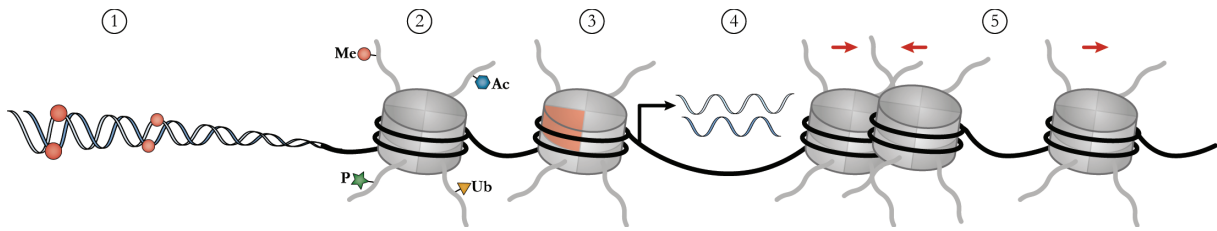
## 1 INTRODUCTION

### 1.1 Epigenetic information

Decades after Friedrich Miescher first isolated DNA from leukocytes in 1869 [Dahm, 2008], it was still widely thought that genetic information is stored and passed on to the next generation by the huge variety of proteins, present in all cells. However, the notion that a molecule that lacks the complexity of proteins cannot be responsible for heredity was proven to be false in the first half of the 20<sup>th</sup> century. With their groundbreaking experiments, Avery and colleagues [Avery et al., 1944] and later Hershey and Chase [Hershey & Chase, 1952] identified DNA as the “transforming principle” which accounts for the propagation of hereditary traits. Building up on the work of Chargaff, Wilkins and Franklin, Watson and Crick then provided the first model, which could adequately explain, how DNA was structured, duplicated and inherited by a daughter cell [Watson & Crick, 1953a; Watson & Crick, 1953b]. Taken together, these findings laid the foundation for the “central dogma of molecular biology”: genetic information is first transcribed from DNA into messenger RNA (mRNA), which in turn is translated into a sequence of amino acids that forms a protein [Crick, 1958; Crick, 1970].

The discovery that DNA is the carrier of genetic information is considered one of the major breakthroughs in modern biology. Yet, the “central dogma” does not explain how differentiated cells of a multicellular organism, a vast majority of them sharing the same genome, differ so fundamentally in shape and function. For this, a second layer of information is necessary, which determines cell type-specific gene expression profiles and does not change the underlying DNA sequence. As this type of information does not alter the genetic background but rather is superimposed to the genetic code, it is termed epigenetic (*epi*: Greek for *over, above*) [Waddington, 1942]. In general, epigenetic mechanisms comprise the chemical modification of cytosines, posttranslational histone modifications, replacement of canonical histones with histone variants, as well as chromatin remodeling and non-coding RNAs (Figure 1). During the course of development, it is the epigenetic machinery, which shapes chromatin structure and thereby promotes a plastic gene expression as a response to interior and exterior cues [Jaenisch & Bird, 2003; Bernstein et al., 2007]. Moreover, the inheritance of stable epigenetic marks defines the so called “epigenetic landscape” i.e. the identity of differentiated cells [Waddington, 1957; Ringrose & Paro, 2004]. It is now clear that individual epigenetic pathways do not act independently, but rather form a complex network, which regulates the

functional outcome of the genetic code [Dobosy & Selker, 2001; Chow & Heard, 2009; Hawkins et al., 2010; Bannister & Kouzarides, 2011].



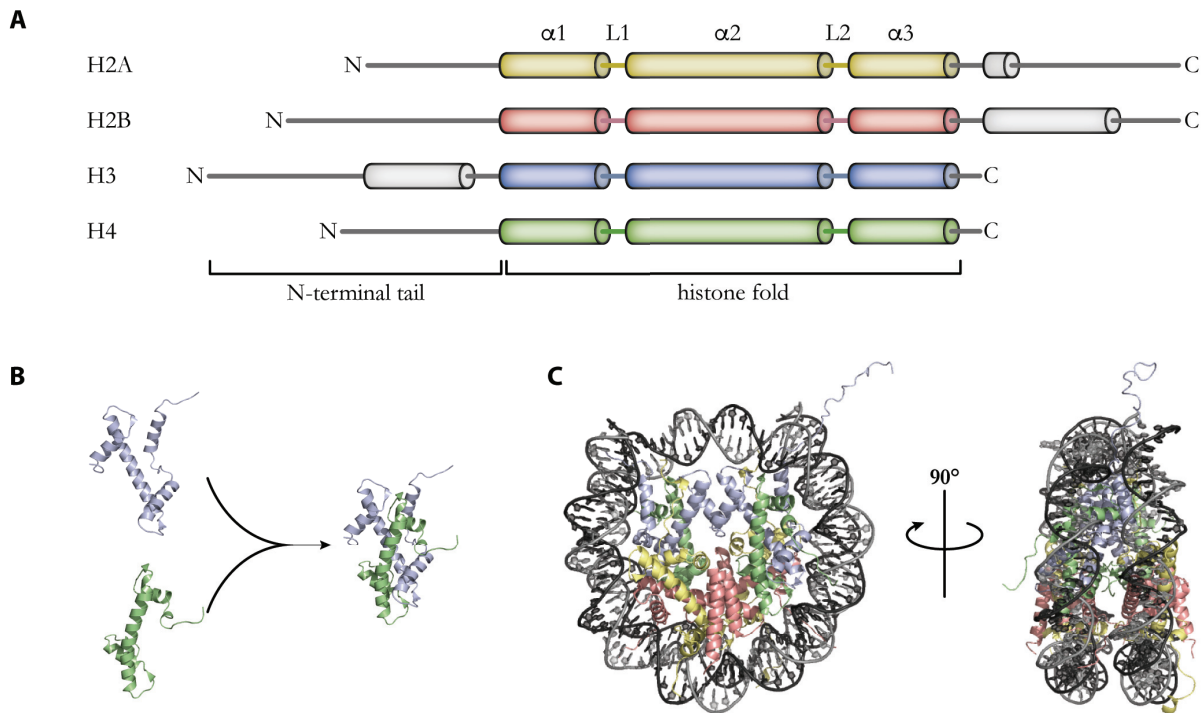
**Figure 1: Overview of epigenetic modifications and mechanisms.** Epigenetic modifications occur either on the DNA itself (DNA methylation (1)) or on histone tails (2). Individual canonical histones can also be exchanged by special histone variants (3). Non-coding RNAs are involved in chromatin modification and transcriptional gene silencing (4). Chromatin remodeling by rearranging nucleosomes can alter the general accessibility of chromatin (5).

## 1.2 General genome architecture

### 1.2.1 Hierarchical genome organization

To meet the spatial confinements of the nucleus and to allow regulated gene expression, eukaryotic genomic DNA is in a complex with proteins and non-coding RNAs to form chromatin. This higher order structure, at its core, is built up by nucleosomes. Each nucleosome is composed by approximately 147 base pairs (bp) of DNA wrapped around a histone octamer [Finch et al., 1977; Luger et al., 1997; Richmond & Davey, 2003]. This octamer comprises two heterodimers of H2A and H2B, which tightly interact with two heterodimers of H3 and H4. Despite their sequence heterogeneity, all core histones share a common tertiary structure motif: three  $\alpha$ -helices ( $\alpha1 - \alpha3$ ) are separated by two disordered linkers (L1 and L2) and form a characteristic histone-fold. By providing a hydrophobic surface, these globular regions are responsible for the pairing of H2A-H2B and H3-H4 heterodimers, respectively (Figure 2) [Arents & Moudrianakis, 1995; Luger et al., 1997]. In addition, each core histone is rich in the basic amino acids lysine and arginine, which are essential for the interaction with the negatively charged phosphodiester backbone of DNA. Individual nucleosomes are linearly connected by 20 – 90 bp of linker DNA, leading to a “beads on a string”-like structure [Olins & Olins, 1974; Oudet et al., 1975]. Interactions between linker DNA, nucleosomes and the linker histone H1 promotes further compaction of this 11-nm fiber into secondary structure chromatin. The nature of this “30-nm fiber” has been under debate for decades [Woodcock, 2005; Tremethick, 2007]. However, a low resolution crystal structure of a tetranucleosome favors a zigzag-like configuration over a solenoidal model [Schalch et al., 2005]. Metaphase chromosomes and, to some extent,

specialized chromatin loops like enhancers and insulators represent a tertiary chromatin structure. This higher-order chromatin organization is formed by long-distance contacts between secondary structure fibers but, similar to the 30-nm fiber, the exact organization and packaging of tertiary chromatin remains elusive [Woodcock & Dimitrov, 2001; Woodcock & Ghosh, 2010].



**Figure 2: Domain architecture of histones and assembly of nucleosomes.** A) The histone fold is a commonly shared motif of all eukaryotic histones. It comprises three  $\alpha$ -helices, which are connected by flexible linkers. Additionally, each histone also harbors a largely unstructured N-terminal tail, which protrudes from the nucleosome. B) Heterodimerization of histones is mediated by their histone fold (only H3 (blue) and H4 (green) are shown as an example). C) Two heterodimers each of H2A-H2B and H3-H4, together with 147 bp of DNA form the barrel-shaped nucleosome particle (PDB 1AOI).

### 1.2.2 *Eu- and heterochromatin*

In addition to its hierarchical order of compaction, chromatin is also categorized in transcriptionally active eu- and inactive heterochromatin. Historically, these two forms of chromatin have been distinguished by their staining properties in interphase nuclei. Whereas euchromatin displays a more open and loosely packed conformation, heterochromatin is highly compacted [Heitz, 1928]. Since then, it became apparent that eu- and heterochromatin also differ in their classes of repetitive genomic elements, gene density and GC-content [Korenberg & Rykowski, 1988; Bickmore & Sumner, 1989; Gardiner, 1995].

In general, euchromatin is defined by harboring predominantly transcriptionally active genes, irregular spaced nucleosomes and short interspersed nuclear elements (SINES) [Caron et al.,

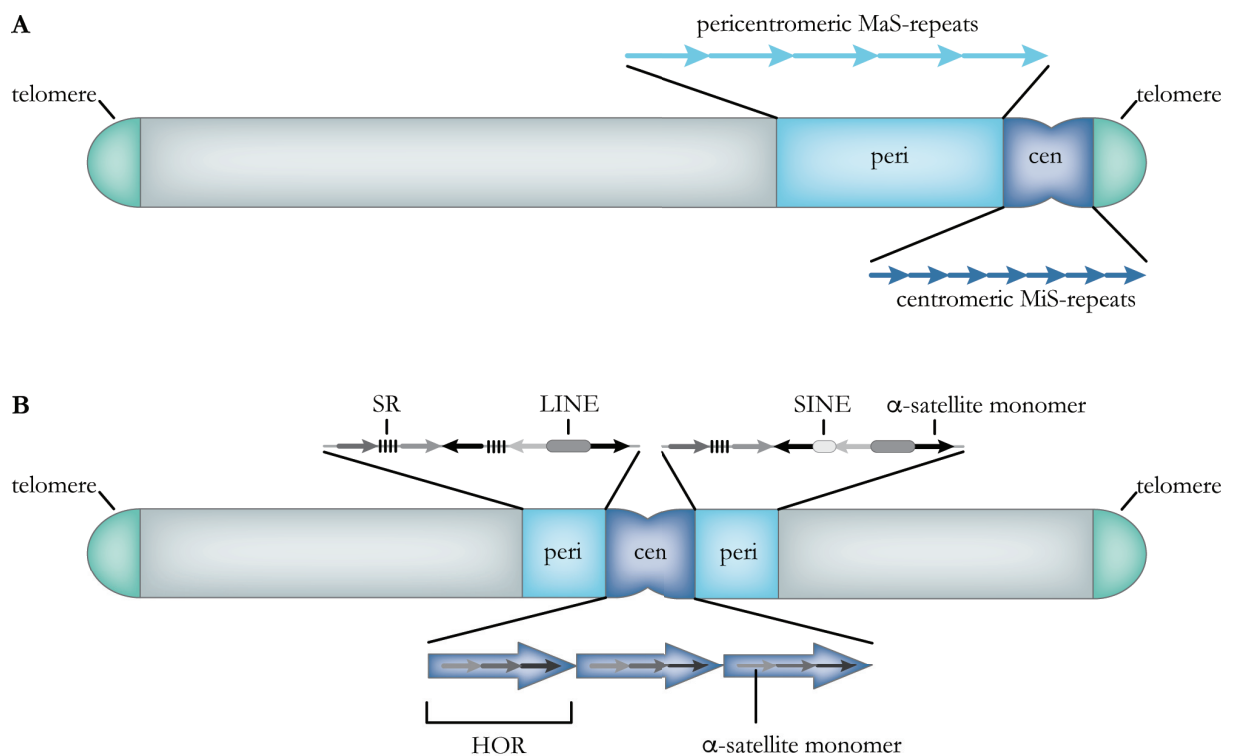
2001; Ozsolak et al., 2007]. In particular, these SINEs belong to the Alu (human, ~ 300 bp) and B1 (mouse, ~140 bp) families and originated from a duplication and fusion of the 7SL RNA gene [Vassetzky et al., 2003; Tsigos & Rigoutsos, 2009]. In metaphase chromosomes, euchromatin coincides with GC-rich R-bands, when reversely stained with Giemsa dye [Bickmore & Sumner, 1989]. Heterochromatin, on the other hand, denotes a transcriptionally inert state, which is characterized by regular nucleosomal arrays, low gene- and GC-content and the presence of long interspersed nuclear elements (LINEs) [Korenberg & Rykowski, 1988; Wallrath & Elgin, 1995]. Moreover, heterochromatin is commonly subdivided into facultative and constitutive heterochromatin (fHC and cHC, respectively). Despite being transcriptionally silent, fHC retains its capability to interconvert between eu- and heterochromatin by reshaping chromatin structure. Formation of fHC is essential in both silencing developmentally regulated homeobox (Hox) genes during the course of differentiation and ensuring dosage compensation in mammalian female cells by establishing an inactive X chromosome. In addition, fHC contains differentially expressed tissue specific genes [Trojer & Reinberg, 2007]. Whereas fHC is critical for regulatory aspects, cHC plays an important role in the structural integrity of chromatin and remains condensed throughout the cell cycle. It is mainly composed of tandemly arranged repeats and located at the pericentromeric, centromeric and telomeric regions (Figure 3) [Grewal & Jia, 2007; Saksouk et al., 2015].

Telomeres are specialized ribonucleoprotein complexes that form at the end of linear chromosomes and protect them from being recognized as DNA double-strand breaks (DSB) and thus subjected to the cellular repair machinery. Inappropriate repairing of chromosomal ends would result in chromosome fusions and unequal distribution of the genetic material during cell division [McClintock, 1938; Blasco et al., 1997]. Vertebrate telomeric DNA is composed of repeating units of a highly conserved motif (5'-TTAGGG3'), ranging from 9 – 15 kilobases (kb) in human and up to 100 kb in mice [Meyne et al., 1989; O'Sullivan & Karlseder, 2010]. These long tracts of telomeric DNA are recognized by a complex of six shelterin proteins (TRF1, TRF2, POP1, TPP1, TIN2 and RAP1), which have been suggested to promote a cap-like structure (T-loop) at chromosomal ends and thereby mask them from DNA damage response [Griffith et al., 1999; Stansel et al., 2001].

Centromeres appear as a constriction in metaphase chromosomes and act as nucleation site for kinetochore assembly during mitosis [Pluta et al., 1995]. Despite their conserved cellular function, centromeres display a varying composition between species. In human cells, centromeric DNA is based on tandemly repeated 171 bp monomers of  $\alpha$ -satellites [Mitchell et al., 1985]. A set number of monomers give rise to a higher order repeat (HOR), which itself is



reiterated so that centromeres span 0.2 – 5 megabases (Mb). Notably, the sequence composition between individual monomers as well as the organization of HORs is highly heterogeneous and chromosome-specific [Willard, 1985; Alexandrov et al., 1988; Aldrup-Macdonald & Sullivan, 2014]. Mouse centromeres, in contrast, are formed by ~120 bp minor satellite (MiS) units, which cluster in 0.6 Mb stretches and are located proximal to the telomere, resulting in acrocentric chromosomes [Pietras et al., 1983; Wong & Rattner, 1988; Kipling et al., 1991]. Similar to centromeric DNA, the pericentromeric region differs in sequence arrangement between species. Murine pericentromeric arrays are formed by tandemly arranged 234 bp major satellite (MaS) repeats and comprise up to ~ 6 Mb of DNA [Vissel & Choo, 1989; Choo, 1997]. Remarkably, MaS-repeats of several chromosomes cluster during interphase in so-called “chromocenters” (CC). MiS-sequences, however, are visible as single entities at the periphery of CCs, demonstrating the distinct three-dimensional organization of major and minor satellites [Guenatri et al., 2004]. Human pericentromeric regions, on the other hand, are predominantly composed of stretches of unordered monomeric  $\alpha$ -satellites. In contrast to alphoid DNA of centromeres, pericentromeric sections of chromosomes do not form higher order repeats and are frequently disrupted by LINES, SINEs and simple repeats [Gosden et al., 1975; Manuelidis, 1978; Schueler et al., 2005; Saksouk et al., 2015].



**Figure 3: Heterochromatic repeat organization in mouse and human metaphase chromosomes.** A) Schematic representation of a typical mouse acrocentric chromosome. Centromeric DNA (dark blue) is composed of tandemly arranged 120 bp minor satellite (MiS) repeats, whereas the pericentromeric region (light blue) comprises 234 bp major satellite repeats.

B) Depiction of a human metacentric chromosome. The centromere (dark blue) is constituted by 171 bp  $\alpha$ -satellite monomers, which are organized in higher order repeats (HOR). Sequence variations of  $\alpha$ -satellites are indicated by different gray values. Pericentromeric DNA contains unordered  $\alpha$ -satellites, which are disrupted by simple repeats (SR), LINEs and SINEs. Peri: pericentromere; cen: centromere.

### 1.3 Epigenetic regulation of chromatin organization

A complex and highly interconnected network of epigenetic mechanisms acts synergistically with transcription factors (TFs) and other non-histone proteins to establish and maintain a differential compaction of chromatin. As a result, chromatin accessibility by the transcription machinery is regulated, shaping the transcriptional activity of the genome and thus cellular identity [Dillon & Festenstein, 2002; Kouzarides, 2007; Lopes Novo & Rugg-Gunn, 2016].

#### 1.3.1 DNA methylation dynamics

The postreplicative addition of a methyl-group at the C5 position of cytosine (5mC) is the first epigenetic modification, which has been shown to directly influence gene expression [McGhee & Ginder, 1979]. In mammals, 5mC is predominantly found within the context of CpG dinucleotides. In human embryonic stem cells (hESCs), however, nearly a quarter of global 5mC has been reported for non-CpG dinucleotides [Lister et al., 2009]. To a lower extent this phenomenon has also been observed in somatic cells, like neurons [Varley et al., 2013].

Spontaneous deamination of cytosines generates uracil and leads to a U:G mismatch, which can be repaired by the base excision repair (BER) machinery [Duncan & Miller, 1980]. Methylated cytosines, however, are particularly prone for hydrolytic or enzymatic deamination, resulting in high rates of cytosine to thymidine transitions. This inherent mutagenic potential is thought to have resulted in an evolutionary depletion of CpG sites [Sved & Bird, 1990; Jurkowska et al., 2011]. Nevertheless, 70 – 80 % of all CpGs in mammalian DNA are methylated [Ehrlich et al., 1982; Bird et al., 1985]. CpG dinucleotides are unevenly distributed throughout the genome. While they are almost absent in inter- and intragenic regions, CpGs show elevated densities in promoter sequences. In fact, approximately 70 % of human promoters are characterized by a strikingly high abundances of CpG dinucleotides [Saxonov et al., 2006]. Such clusters of CpG sites are referred to as CpG islands (CGIs) and defined as patches of ~1 kb in length, within which CpG dinucleotides occur close to their expected frequency [Suzuki & Bird, 2008]. Reflecting the potential mutability of CpG dinucleotides, most CGIs prevail in a hypomethylated state. Hypermethylated CGIs are associated with stable gene repression and play key roles in both development and genome integrity. For example, DNA methylation is required for X inactivation, genomic imprinting and silencing of

transposable elements [Li et al., 1993; Woodcock et al., 1997; Gendrel et al., 2012; Barau et al., 2016].

While DNA methylation in promoter regions is generally associated with transcriptional silencing, methylation of the gene body correlates with gene expression [Wolf et al., 1984; Yang et al., 2014]. The exact functional consequences of this seemingly paradox situation are unknown. However, there is emerging evidence that methylation marks in gene bodies suppress cryptic transcriptional start sites and thereby minimize transcriptional noise [Bird, 1995; Suzuki et al., 2007]. Moreover, it has been shown that introns are less methylated than exons and excessive methylation of introns decreases polymerization speed of RNA polymerase II (RNAPII). As splicing is influenced by RNAPII kinetics, DNA methylation and cotranscriptional splicing might be linked [Laurent et al., 2010; Shukla et al., 2011].

A family of enzymes, termed DNA (cytosine C5) methyltransferases (DNMTs), catalyzes the covalent attachment of a methyl-group to the C5-position of cytosine. In mammals, the DNA methylation machinery is composed of four enzymes (DNMT1, DNMT3A, DNMT3B and DNMT3C) and one cofactor (DNMT3L) [Rottach et al., 2009; Barau et al., 2016]. The process of transferring the methyl-group is highly conserved among different species and has been described in detail for prokaryotic 5mC methyltransferases (MTases), such as M.HhaI [Klimasauskas et al., 1994]. Strikingly, the crystal structure of M.HhaI, bound to its target sequence, revealed that MTases form a covalent complex with their substrate, leading to an increased reactivity of the C5 atom of cytosine. Furthermore, the target base is “flipped” out of the DNA double helix (“base flipping”). The whole catalytic reaction involves the following steps: substrate recognition, base flipping, forming of a covalent bond between the C6-position of cytosine and the sulfhydryl-group of cysteine in a conserved proline-cysteine motif of the DNMT, transfer of the methyl-group from S-adenosyl-L-methionine (SAM) to the activated C5 atom and subsequent release of the enzyme via  $\beta$ -elimination [Cheng & Blumenthal, 2008].

Historically, CpG methylation has been classified into the two different types *de novo* and maintenance methylation. During embryonic development and cell differentiation, methylation marks have to be set up *de novo* by DNMT3 family members in order to establish lineage-specific gene expression patterns [Okano et al., 1999]. As methylation of cytosines is a postreplicative modification, these marks would be lost after several rounds of DNA replication. Therefore, the so-called maintenance DNMT (DNMT1) specifically modifies hemimethylated DNA, thus ensuring that CpG methylation is faithfully transmitted from one generation of cells to the next [Hermann et al., 2004]. This strict separation of *de novo* and maintenance methylation, however, seems to be an oversimplified model. In fact, numerous

studies provide evidence that *de novo* methylation during embryogenesis occurs by the combined activities of DNMT3 and DNMT1 enzymes. DNMT3 enzymes show selective binding preferences for CpG-flanking sequences. Moreover, DNMT3A binds to DNA in a tilted fashion, which allows methylation of two adjacent CpGs but prohibits enzymatic activity on the complementary strand. Sequence specificity, as well as the mode of action result in the accumulation of hemimethylated sites, which in turn represent an ideal substrate for DNMT1 [Fatemi et al., 2002; Kim et al., 2002; Handa & Jeltsch, 2005; Jia et al., 2007]. Furthermore, residual DNA methylation has been observed in DNMT3A/DNMT3B double-knockout mouse embryos, suggesting that DNMT1 possesses *de novo* methylation activity [Okano et al., 1999].

Although 5mC is generally considered a stable epigenetic mark, it has been known for a long time that DNA methylation is a dynamic process and editing of previously set methylation marks is particularly important during mammalian embryogenesis. Upon fertilization, the paternal pronucleus is subject to extensive epigenetic reprogramming via genome-wide demethylation. Since the removal of the methyl-group is independent from DNA replication, this process cannot be explained by a passive demethylation [Mayer et al., 2000]. Mechanisms, however, which underlie active demethylation have remained controversial [Schär & Fritsch, 2011]. It has been known for over two decades that members of the order *Trypanosoma* possess a base J ( $\beta$ -D-glucosyl-hydroxymethyluracil), which is produced by the sequential hydroxylation and glucosylation of the methyl group of thymine [Borst & Sabatini, 2008]. In this context, it has been proposed that JBP1 and JBP2 (J-binding protein 1 and 2), members of the 2-oxoglutarate (2OG)- and Fe(II)- dependent oxygenase superfamily of enzymes, are involved in oxidizing thymine into hydroxymethyluracil [Yu et al., 2007]. Recently, searching for mammalian homologs of JBP1 and JBP2, Tahiliani and colleagues discovered that human TET1 (Ten-eleven translocation 1) is capable of oxidizing 5mC into 5-hydroxymethylcytosine (5hmC) [Tahiliani et al., 2009]. Since then, it has been shown that the other members of the TET family, namely TET2 and TET3, are also capable of performing this reaction. Moreover, it was demonstrated that this class of enzymes is responsible for further oxidization of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [Ito et al., 2011]. To date, two functional roles have been proposed for 5hmC and its further oxidized states 5fC and 5caC. On the one hand, 5hmC and its oxidized derivatives might be intermediates in active demethylation and thus play an important role in DNA methylation dynamics. For instance, 5hmC has been suggested as a target for deamination by the AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) family of cytidine deaminases. According to this, the resulting 5-hydroxymethyluracil is subsequently removed

by the BER pathway [Guo et al., 2011]. Furthermore, 5fC and 5caC are specifically recognized and excised by thymine DNA glycosylase (TDG) *in vitro* and *in vivo* [He et al., 2011; Pfaffeneder et al., 2011; Zhang et al., 2012a; Müller et al., 2014; Weber et al., 2016]. On the other hand, the presence of 5hmC itself may represent an epigenetic mark, as it has been detected in stable amounts in ESCs, which would contradict its role as a sole intermediate in DNA methylation [Dawlaty et al., 2011; Ficz et al., 2011; Pastor et al., 2011].

### 1.3.2 Histone posttranslational modifications (PTMs)

In addition to DNA methylation, epigenetic information is encoded in post-translational modifications (PTMs) of histones. As mentioned above, individual histones comprise a histone fold motif, as well as a largely unstructured 20 – 35 amino acids long N-terminal peptide, which protrudes from the nucleosome core particle [Luger et al., 1997]. These so-called histone-tails, and to a lesser extent the globular histone fold, are subject to a plethora of PTMs, including acetylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, proline isomerization, deimination and methylation [Kouzarides, 2007]. Besides their function in other DNA-based processes such as repair, recombination and replication, histone PTMs play a crucial role in altering chromatin compaction and thereby regulate gene expression [Strahl & Allis, 2000; Taverna et al., 2007]. Whereas PTMs located at the histone fold domain directly affect histone-histone and histone-DNA interactions [Simon et al., 2011], post-translational modifications of the histone tail indirectly act on chromatin compaction by recruiting reader proteins. Notably, distribution and type of PTMs form a signature (often referred to as “histone code”), which is indicative for the chromatin state of a given locus. In the following, acetylation and methylation of histones are presented in more detail.

Active genes are generally enriched in histone acetylation. Histone H3, acetylated on lysine 9 (H3K9ac) and H4K16ac are both found at the promoters and/or enhancers of actively transcribed genes, whereas H3K27ac is located at the transcriptional start site (TSS) [Nishida et al., 2006; Wang et al., 2008; Taylor et al., 2013]. Furthermore, various degrees of methylated lysines on histone H3 also correlate with euchromatin. Monomethylated H3K4 (H3K4me1) shows high levels along enhancer sequences, whereas H3K4me2/3 marks active promoters [Heintzman et al., 2007]. Additionally, H3K36me3 displays a distinctive distribution pattern: it is enriched in the 3'-end of gene bodies and gradually decreases towards the 5'-end [Bannister et al., 2005]. In contrast, cHC is marked by repressive PTMs such as H3K9me3 and H4K20me3 [Schotta et al., 2004], whereas fHC is enriched in H3K27me3 [Trojer & Reinberg, 2007].

### Writers of histone PTMs

Similar to PTMs themselves, the protein machineries that write, read and remove histone modifications have become central figures in studying chromatin organization (Figure 4). Histone PTMs are set by a group of enzymes, termed “writers”. Depending on the type of modification that is catalyzed, they are categorized as histone acetyltransferase (HAT) or methyltransferase (HMT).

HATs utilize the co-factor acetyl-CoA to catalyze the transfer of an acetyl-group to the  $\epsilon$ -amino-group of lysine residues. Based on their subcellular localization, HATs are categorized into two major classes. The highly conserved type-B HATs are predominantly cytoplasmatic and acetylate newly synthesized histones, facilitating their assembly into nucleosomes [Richman et al., 1988; Parthun, 2007]. Type-A HATs, on the other hand, comprise the more diverse families of GNAT (Gcn5-related N-acetyltransferases), MYST (named after their founding members MOZ, Ybf2, Sas2 and Tip60) and p300/CBP (adenoviral E1A-associated protein of 300 kDa/CREB-binding protein) [Yang & Seto, 2007]. Type-A HATs mainly acetylate histone tails within nucleosomes and are often found associated in multiprotein complexes, which regulate substrate recognition and enzyme activity [Bannister & Kouzarides, 2011]. Interestingly, the interaction of p300/CBP HATs with the transcription factor CREB (cAMP response element-binding protein) demonstrates a direct link between histone acetylation and transcriptional activation [Shiama, 1997].

Although methylated glutamines and aspartates have been described [Biterge et al., 2014; Tessarz et al., 2014], histone methylation mainly occurs on lysine (mono-, di- or trimethylated) and arginine (mono-, symmetrically or asymmetrically dimethylated) residues. Depending on their substrate, HMTs are subdivided into histone lysine methyltransferases or histone arginine methyltransferases (HKMTs and PRMTs, respectively). Both classes contain several enzymes, which are characterized by both their substrate-specificity and the extent of methylation. For instance, the HKMT DIM5 specifically trimethylates H3K9 [Tamaru et al., 2003], whereas SET7/9 only monomethylates H3K4 [Xiao et al., 2003]. Similarly, type-I PRMTs generate monomethylated (Rme1) and asymmetrically dimethylated arginines, while type-II enzymes produce Rme1 and symmetrically dimethylated arginines. Yet, HKMTs, as well as PRMTs share SAM as a common methyl-group donor [Bannister & Kouzarides, 2011].

### Erasers of histone PTMs

Chromatin is a dynamic environment, which is able to differentially compact as a reaction to external and developmental stimuli. In line with that, histone PTMs are highly dynamic and subject to regulation by the opposing activities of writers and erasers. Histone deacetylases

(HDACs) catalyze the removal of acetyl-groups and are generally considered to be transcriptional repressors. Based on sequence homology to yeast Rpd3 (reduced potassium dependency 3), Hda1 (histone deacetylase 1) and Sir2 (silent information regulator 2), mammalian HDACs are grouped in three different classes [Yang & Seto, 2003]. Besides sequence divergences, HDACs differ in their reaction mechanisms. Rpd3- and Hda1-like deacetylases require  $Zn^{2+}$  for deacetylase activity, whereas Sir2-related HDACs depend on  $NAD^+$  as co-factor [Blander & Guarente, 2004; Lombardi et al., 2011]. In general, HDACs only demonstrate a low substrate specificity and even act on non-histone proteins [Sternier et al., 1979; Luo et al., 2000; Hubbert et al., 2002]. It has been shown, however, that human SIRT1 preferentially deacetylates H4K16 [Vaquero et al., 2004]. Studying substrate specificity is complicated by the fact that often multiple HDACs are part of diverse multiprotein complexes. For example, HDAC1 and HDAC2 are both subunits of CoREST, Sin3 and NuRD complexes [Yang & Seto, 2008].

Histone methylation has long been regarded as a stable modification [Bannister et al., 2002]. Indeed, comparable turnover rates between bulk histones and methylated lysine residues within them, suggested that histone methylation is not reversible [Byvoet et al., 1972; Duerre & Lee, 1974]. Moreover, H3K9me3 is necessary to form permanently silenced chromatin regions such as pericentric heterochromatin, suggesting that it is a static mark [Bannister et al., 2002]. However, with the discovery of the flavin-dependent KDM (lysine demethylase) LSD1 and later LSD2 (lysine-specific demethylase 1 and 2, respectively), this dogma has been changed [Shi et al., 2004; Karytinis et al., 2009]. Due to the formation of an imine intermediate in the oxidation reaction, these flavin-dependent KDMs are only capable to act on mono- or dimethylated lysine residues [Anand & Marmorstein, 2007]. Soon after the description of LSD1, a second class of KDMs has been discovered. This class is characterized by a Jumonji C (JmjC)-domain, catalyzes the removal of methyl-groups in a  $Fe^{2+}/2$ -oxoglutarate (2-OG) dependent manner and thus is capable to also target trimethylated lysines [Whetstine et al., 2006]. In contrast to lysine demethylation, mechanisms that result in arginine demethylation are still controversial [Böttger et al., 2015]. Yet, a recent study suggested that JmjC KDMs oxidize methylated arginines *in vitro* [Walport et al., 2016].

### Readers of histone PTMs

In essence, histone PTMs can regulate chromatin structure and accessibility in two ways. *Cis*-acting PTMs such as acetylation and phosphorylation neutralize the basic charge of histones and thereby weaken the electrostatic interaction of histones with the negatively charged DNA backbone, which is thought to result in a less compact chromatin structure. In fact, genes of

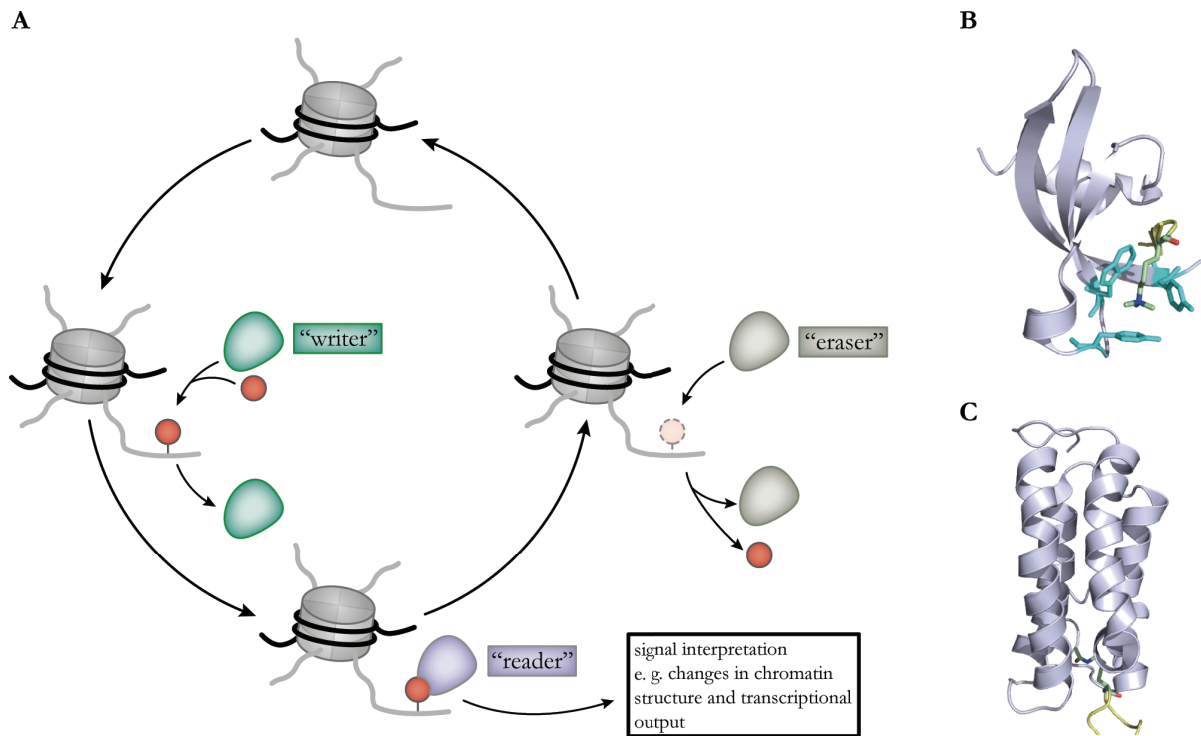
the  $\beta$ -globin locus, as well as many enhancer and promoter elements throughout the genome reside within a hyperacetylated chromatin environment, which presumably facilitates a transcriptionally competent chromatin state [Kiefer et al., 2008; Wang et al., 2008]. Yet, the presence of multiple acetylated sites seems not to be a prerequisite for conformational change, as specific acetylation of H4K16 significantly perturbs the formation of compact higher order chromatin structures [Shogren-Knaak et al., 2006].

The majority of evidence, however, suggests that PTMs act in *trans* by promoting or inhibiting the recruitment of a large variety of regulatory proteins to the nucleosomes. Those effector proteins are characterized by special domains, which confer substrate specificity.

Methylated lysines are recognized by numerous domain types, including PHD (plant homeodomain) fingers and the so-called Tudor “royal” family of domains, which comprise chromodomains (chromatin organization modifier) (Figure 4B), Tudor, PWWP (Pro-Trp-Trp-Pro) and MBT (malignant brain tumor) domains [Maurer-Stroh et al., 2003; Kim et al., 2006; Champagne & Kutateladze, 2009]. Although these binding domains are sensitive to the degree of methylation, they share aromatic cage structures, typically formed by two to four aromatic amino acids, as a common recognition motif [Musselman et al., 2012]. Similar to writers of histone arginine methylation, insights into how this mark is recognized remain sparse. However, a recent study suggests that the Tudor domain of TDRD3 binds to H3R17 and H4R3, when the respective residues are asymmetrically dimethylated [Yang et al., 2010].

Contrary to the plethora of domains that bind to methylated lysines, only two domains have been shown to recognize acetylated lysines. Most prominently, bromodomains are found in HATs and form a conserved four-helix bundle (BRD-fold). Notably, the inter-helical loops constitute a hydrophobic pocket, capable of recognizing acetyl-lysine (Figure 4C) [Sanchez & Zhou, 2009]. More recently, it has been demonstrated that acetylated lysines can also be bound by tandem PHD domain-containing proteins such as DPF3 [Zeng et al., 2010].





**Figure 4: Histone writers, readers and erasers.** A) Histone PTMs are dynamic marks that are set up by writer enzymes (e.g. HATs and HMTs), interpreted by reader proteins (e.g. bromo- or chromodomain-containing proteins) and removed by eraser-proteins (e.g. HDACs and KDMs). B) Crystal structure of *D. melanogaster* HP1 chromodomain in complex with H3K9me3. Aromatic residues of tyrosine and tryptophan form a cage (blue), which is able to recognize K9me3 (green) (PDB 1KNE). C) Solution structure of human CBP bromodomain bound to H4K20ac. The bromodomain is composed of four  $\alpha$ -helices. Inter-helical loops form a hydrophobic pocket, which accommodates acetyl-lysine (green) (PDB 2RNY).

### 1.3.3 Crosstalk between epigenetic mechanisms: establishment of eu- and heterochromatin

Embryonic stem cells are characterized by an open and transcriptional permissive chromatin conformation. During differentiation, DNA methylation and histone PTMs act in an interconnected way to set up differentially compacted chromatin environments and thereby ensure cell type specific gene expression patterns [Meshorer & Misteli, 2006; Cedar & Bergman, 2009].

As mentioned earlier, the paternal genome of pre-implantation embryos is subject to a TET3-dependent wave of active DNA demethylation. Although TET3 is also expressed in the maternal pronucleus, the maternal genome relies on a replication-dependent passive mechanism. Recently, it was demonstrated that PGC7 binds to H3K9me2, set by the HKMT G9a, and thereby blocks TET3-activity at adjacent mCpG-sites. In the paternal pronucleus, on the other hand, histones are widely replaced by protamines, which lack the H3K9me2 mark. Hence, methylated CpGs are not protected from TET3-mediated oxidation [Nakamura et al., 2007; Nakamura et al., 2012].

After implantation, embryos re-establish a bimodal DNA methylation pattern, that is, high methylation levels of interspersed CpG-sites and unmethylated CGIs. In this context, it has been suggested that RNAPII recruits H3K4-specific HKMTs of the trithorax family [Hughes et al., 2004; Guenther et al., 2007]. Since RNAPII is predominantly located at CGI-containing promoters, only these regions are marked by methylated H3K4. In turn, H3K4me3 impairs the binding of DNMT3L, which interacts with DNMT3A/B and is an essential co-factor for *de novo* methylation [Ooi et al., 2007]. Similarly, unmethylated CGIs are bound by CFP1 (CXXC finger protein 1), which in turn recruits H3K4 methyltransferases [Thomson et al., 2010]. These models are supported by the fact that DNA-methylation and H3K4-methylation generally show a strong anti-correlation [Meissner et al., 2008; Mohn et al., 2008].

During the course of embryonic development, targeted silencing of pluripotency-associated factors, such as Oct3/4 is an essential step in the establishment of tissue-specific cell lineages [Gidekel & Bergman, 2002]. Transcriptional silencing of the *Oct3/4* gene is achieved by a multi-step process. First, transcription is reversibly repressed by GCNF (germ cell nuclear factor)-mediated recruitment of transcriptional co-repressors to the promoter [Fuhrmann et al., 2001]. This is followed by the binding of a complex, containing G9a and HDAC. Subsequently, the euchromatic mark H3K9ac is removed and replaced with H3K9me3. This modification, in turn, is recognized by the chromodomain of HP1. Local heterochromatinization is probably facilitated by dimerization of HP1, thus linking adjacent nucleosomes [Ruthenburg et al., 2007]. As a last step, G9a has been shown to recruit DNMT3A and DNMT3B, leading to *de novo* methylation and permanent silencing [Feldman et al., 2006; Epsztejn-Litman et al., 2008].

Formation of constitutive heterochromatin, such as pericentromeric satellite repeats, is dependent on the SET domain-containing HKMTs SUV39H1 and SUV39H2. These enzymes catalyze the formation of H3K9me3, which again is bound by HP1, leading to a further chromatin compaction. Additionally, SUV39 enzymes also recruit DNMT3A/B, which results in *de novo* methylation of these satellite sequences [Cedar & Bergman, 2009; Saksouk et al., 2015]. Interestingly, initial recruitment of HKMTs to satellite repeats seems to be mediated by non-coding RNA-duplexes, which naturally form at satellite sequences. These RNA-duplexes are processed by the RNA endonuclease Dicer, resulting in an RNA-induced silencing complex (RISC), which is targeted back to pericentromeric regions, where it is thought to interact with SUV39H1 and SUV39H2 [Fukagawa et al., 2004; Sugiyama et al., 2005]

Similar to cHC, establishment of an inactive X chromosome (Xi) in mammalian female cells also depends on non-coding RNA. Transcription of the XIC (X inactivation center) results in the 17 kb non-coding RNA *Xist* (X-inactive specific transcript). RepA, a 1.6 kb region within

*Xist*, has been shown to recruit PRC2 (Polycomb repressive complex 2) to Xi in *cis* via its interaction with EZH2, the catalytic subunit of PRC2. EZH2 contains a SET domain and catalyzes trimethylation of H3K27 on surrounding nucleosomes [Zhao et al., 2008]. This mark is bound by HPC (heterochromatin-like chromodomain protein), which is a subunit of PRC1 (Polycomb repressive complex 1) and has been implicated in further heterochromatinization by catalyzing ubiquitination of H2AK119 [Plath et al., 2004; Schwartz & Pirrotta, 2008]. Finally, hypermethylation of X-linked promoter CGIs is thought to be achieved by recruitment of *de novo* methyltransferases by EZH2 [Norris et al., 1991; Vire et al., 2006].

Bivalent domains represent a special case in ESCs and are characterized by large repressing H3K27me3-enriched regions, harboring smaller activating H3K4me3 patches. In ES cells, these domains tend to coincide with promoters of developmentally regulated TF-genes, which are expressed at low levels. This bivalent modification pattern is predicted to keep TF-genes in a poised state that can be resolved to either an activated (H3K4me3) or fully repressed (H3K27me3) chromatin conformation [Azuara et al., 2006; Bernstein et al., 2006].

## 1.4 3D-Organization of the genome

Besides rather local eu- and heterochromatic regions, mammalian genomes in interphase nuclei are organized in a complex topological hierarchy. Historically, methods to decipher structural and topological properties of interphase chromosomes mainly relied on elaborate fluorescence *in situ* hybridization (FISH) techniques. More recently, chromosome conformation capture (3C) and its modifications (4C, 5C and Hi-C) provided local or genome-wide maps of intra- and inter-chromosomal contacts, expanding the understanding of nuclear architecture. These methods are based on mild formaldehyde crosslinking of chromatin, followed by fragmentation, ligation of DNA-fragments and subsequent detection by either PCR or next-generation sequencing [Dekker et al., 2013].

### 1.4.1 Chromosome territories

Utilizing fluorescence *in situ* hybridization (FISH) and microirradiation, early experiments have demonstrated that individual chromosomes are not randomly distributed, but rather occupy distinct territories (chromosome territories, CTs) within interphase nuclei (Figure 5) [Cremer et al., 1982; Lichter et al., 1988]. According to this model, individual CTs are separated by an almost chromatin-free interchromatin domain (ICD), into which actively transcribed genes occasionally loop out and form inter- and intra-chromosomal contacts [Cremer & Cremer, 2001; Scheuermann et al., 2004]. This observation could be confirmed with 3C-based

approaches. Most chromatin contacts are found within individual chromosomes and active loci predominantly form contacts among themselves [Simonis et al., 2006]. In addition, inactive genomic regions are spatially restricted to their own CT, whereas transcribed regions can also form inter-chromosomal contacts [Lieberman-Aiden et al., 2009; Kalhor et al., 2011; Zhang et al., 2012b]. Besides its purpose in functionally clustering transcribed loci, this intermingling between CTs is also thought to be a hot-spot for chromosomal translocations [Branco & Pombo, 2006; Zhang et al., 2012b].

Furthermore, chromosome painting and bleaching approaches demonstrate that the relative spatial distribution of individual CTs is confined during interphase but experiences a high variability between cell-types and from one interphase to the next [Walter et al., 2003; Mayer et al., 2005]. Yet, it has been shown that gene density and, to a lower extent, DNA content of chromosomes determine the spatial distribution of CTs. Whereas small, gene-rich chromosomes predominantly localize towards the nuclear interior, bigger and gene-poor chromosomes are found in vicinity of the nuclear periphery. This radial distribution of CTs is exemplified by human chromosomes 18 and 19. Although both chromosomes are similar in size, CTs of the gene-poor chromosome 18 were typically found at the nuclear periphery, whereas the gene-rich chromosome 19 preferentially localized at the nuclear interior [Croft et al., 1999; Cremer et al., 2001]. Similarly, individual CTs experience a polarized distribution of gene density. While gene-poor segments tend to localize on the surface of CTs, gene-rich segments, on the other hand, preferentially localize in the interior of CTs [Kupper et al., 2007].

### *1.4.2 Subchromosomal compartments and chromosomal domains*

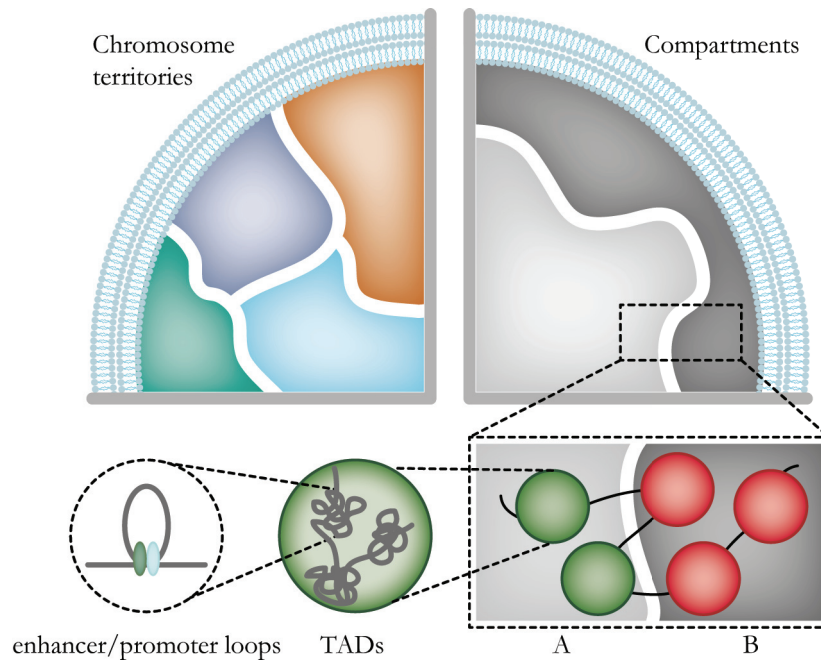
The radial distribution of CTs, as well as the spatial separation of euchromatic and heterochromatic regions within them, indicates that nuclear architecture is based on functional compartmentalization of chromatin [Fraser & Bickmore, 2007; Misteli, 2007]. Indeed, 3C-based methods confirmed the presence of subchromosomal compartments. By mapping the genome-wide contact-frequency of chromosomes in interphase cells, two different subchromosomal compartments (A and B) were detected, which correspond to eu- and heterochromatin, respectively (Figure 5). Loci clustered in A compartments are generally gene-rich, transcriptionally active and experience hypersensitivity to DNaseI, which suggests an open chromatin conformation. Loci that are found in B compartments, however, are relatively gene-poor, transcriptionally inert and insensitive to DNaseI treatment, which indicates a closed chromatin conformation. Hence, these compartments relate to gene expression and, therefore, are cell-type specific [Lieberman-Aiden et al., 2009; Zhang et al., 2012b]. Both A

and B compartments are constituted by groups of chromosomal domains (CDs), which mainly form intrachromosomal contacts. Yet, interchromosomal contacts have been reported, which probably reflect zones of chromosomal intermingling [Branco & Pombo, 2006; Dixon et al., 2012; Gibcus & Dekker, 2013]. The size of CDs varies drastically between species, ranging from tens of kbs to several Mbs (~100 kb in *D. melanogaster*, ~1 Mb in human and ~3 Mb in mouse) [Cavalli & Misteli, 2013; Gibcus & Dekker, 2013]. Notably, compartments are defined by groups of CDs with similar activities, resulting in blocks of alternate transcriptional activity (i.e. active or inactive) on mitotic chromosomes. Yet, these blocks do not represent simple on and off states, but rather a continuum of different transcriptional activities [Imakaev et al., 2012].

A special class of CDs is formed by genomic regions that interact with relatively fixed structures of the nucleus. In all eukaryotes, chromatin is spatially confined by the nuclear envelope (NE), formed by a lipid bilayer. Embedded in the NE are nuclear pores, which are large and highly organized structures that mediate communication between nucleoplasm and cytoplasm [Dingwall & Laskey, 1992]. Interestingly, the inner membrane of the NE is lined with a filamentous mesh, constituted by lamin proteins, termed nuclear lamina (NL) [Goldman et al., 2002; Prokocimer et al., 2009]. Besides ensuring the nuclear integrity, the NL serves as anchor point for interactions between chromatin and the nuclear periphery. Chromosomal domains that specifically interact with the NL have been detected via DamID. This approach utilizes the bacterial DNA adenine methyltransferase (Dam), which is fused to a protein of the NL (e.g. lamin B1). Thereby, interacting chromatin regions are marked with a novel DNA modification. Subsequently, adenine-methylated DNA fragments are isolated and mapped using a microarray [Pickersgill et al., 2006; Guelen et al., 2008; Peric-Hupkes et al., 2010]. These lamina-associated domains (LADs) are generally gene-poor and are highly enriched for the heterochromatic marks H3K27me3 and, to a lesser degree, H3K9me2. In human cells over 1000 LADs were identified, which span between 0.1 and 100 Mb and collectively cover ~40 % of the genome. In addition, LADs are marked by distinct borders, which show an enrichment of binding sites for the insulator protein CTCF (CCCTC-binding factor) and promoters, driving transcription away from LADs [Guelen et al., 2008]. Given the fact that during differentiation some LADs lose their connection to the NL and that these changes correlate with altered gene expression profiles, LADs are further classified in facultative and constitutive LADs (fLADs and cLADs, respectively). Notably, fLADs are cell-type specific, whereas AT-rich cLADs are maintained between a range of cell-types and are conserved among species [Peric-Hupkes et al., 2010; Meuleman et al., 2013]. The nucleolus represents the second subnuclear structure that forms contacts with distinct chromosomal

domains. In addition to the expected ribosomal DNA genes (rDNA), AT-rich, transcriptionally inactive sequences across all chromosomes have been described to associate with nucleoli in a domain-like pattern [Nemeth et al., 2010; van Koningsbruggen et al., 2010]. Interestingly, these sequences, termed nucleolar-associated domains (NADs), experience a substantial overlap with cLAD patterns. Hence, it has been suggested that NADs and cLADs comprise the same type of repressive chromatin, which is randomly distributed between the NL and the nucleolar periphery after mitosis. This model is supported by the observation that some CDs that associate with the nucleolus in a mother cell are repositioned to the NL in daughter cells [van Koningsbruggen et al., 2010].

Recent Hi-C and 5C studies uncovered a third type of chromosomal domains, referred to as topologically associated domains (TADs). These megabase-sized domains are characterized by a high interaction-frequency of loci located within one domain, but less pronounced associations of loci located within two adjacent domains (Figure 5) [Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012]. Interestingly, individual TADs display a high degree of alignment with either active (H3K4me3) or inactive (H3K27me3) histone PTMs and therefore can be assigned to A and B compartments, respectively [Nora et al., 2012]. However, the fact that TADs occupy 91 % of the genome in mice and show a high level of conservation between species and cell-types, suggests that TADs do not represent transcriptional entities per se, but rather form a basic building block for 3D organization of chromosomes [Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012]. This is substantiated by the observation that TADs persist in ESCs, lacking the HKMT Eed [Nora et al., 2012]. Similar to LADs, boundaries between TADs display high levels of the architectural protein CTCF. In addition, enrichment of genomic insulator elements, such as transfer RNA (tRNA) genes and SINES was detected [Lunyak et al., 2007; Raab et al., 2011; Dixon et al., 2012]. Besides its function as boundary protein between TADs, the insulator CTCF, together with cohesin and Mediator, is thought to mediate loop-formation at a sub-TAD level, as well as long-range enhancer/gene interactions [Phillips-Cremins et al., 2013].



**Figure 5: Hierarchical levels of chromatin organization in interphase nuclei.** Top left) Chromosomes occupy distinct territories. Top right) Chromatin is organized in the two compartments A (light-gray) and B (dark-gray), which correspond to eu- and heterochromatin, respectively. Bottom) A and B compartments comprise transcriptionally active (green) and repressed (red) TADs, respectively. TADs are formed by local DNA contacts (e.g. enhancer/promoter loops).

## 1.5 Chromatin dynamics

### 1.5.1 Short- and long-range motility of chromatin

Although being restricted to distinct chromosome territories, chromatin does not represent a static, but rather dynamic structure [Lanctot et al., 2007; Soutoglou & Misteli, 2007]. Chromatin is subject to a plethora of epigenetic marks, which results in differentially compacted regions and thus provides a dynamic balance between genome packing and accessibility to the underlying DNA sequence. Therefore, TFs, in combination with adenosine triphosphate (ATP)-dependent chromatin remodeling complexes, have to induce local chromatin decompaction in order to bind heterochromatic regions [Luo & Dean, 1999; Clapier & Cairns, 2009]

In addition to dynamic chromatin states and fluctuations in composition, chromatin also displays frequent, short-range movements. These movements are locally restrained within approximately  $1\ \mu\text{m}$  by rigid chromatin structures (e.g. the centromere) as well as by the presence or interaction with nuclear substructures (e.g. the NE, nucleoli or nuclear speckles) [Shelby et al., 1996; Hemmerich et al., 2011; Kind et al., 2013]. Moreover, short-range motility occurs in a random walk and depends on both ATP and temperature [Tumbar & Belmont, 2001; Vazquez et al., 2001; Dion & Gasser, 2013]. On a global scale, chromatin motion is

confined to long-range movements, which occur less frequently and require passage through mitosis [Soutoglou & Misteli, 2007]. Except for early G1-phase, at which substantial rearrangements of subchromosomal domains can occur, chromatin motility is restricted to short-range movements during the cell cycle [Walter et al., 2003; Soutoglou & Misteli, 2007; Kind et al., 2013].

Notably, long-range chromatin movement correlates with transcriptional control. According to this model, loci move to either A or B compartments depending to their transcriptional status. Whereas in hematopoietic precursor cells, silent *IgH*- and *IgK*-genes are located at the repressive nuclear periphery, these loci are internalized upon their activation in B-cells [Kosak et al., 2002]. Similarly, targeted transcriptional activation of a transgene results in a long-range movement to the nuclear interior in an actin/myosin dependent manner [Tumbar & Belmont, 2001; Chuang et al., 2006]. In addition, lineage specification of ESCs provides a striking example of large-scale chromatin rearrangement. During differentiation, a large subset of genes switch from A to B compartments and defined heterochromatic regions are formed, resulting in diminished genome plasticity [Hiratani et al., 2008; Dixon et al., 2015].

Yet, there is accumulating evidence, suggesting that radial positioning of genes lacks a direct causality to the transcriptional status. For example, many gene loci remain at the same radial position, even though their expression level changes [Hewitt et al., 2004; Zink et al., 2004; Meaburn & Misteli, 2008]. *Vice versa*, some genes are repositioned without detectable changes of their transcriptional output [Williams et al., 2005; Meaburn & Misteli, 2008].

### *1.5.2 Visualization of chromatin dynamics*

Besides emerging 3C approaches, fluorescence microscopy-based methods are essential for deciphering the dynamics and spatiotemporal organization of chromatin. Direct imaging of nuclear proteins can be achieved by genetically fusing them to fluorescent proteins (FPs; e.g. GFP). Additionally, chromatin-associated proteins are indirectly visualized via FP-tagged camelid antibodies. These heavy-chain antibodies can be raised against virtually any epitope and fold readily in eukaryotic cells [Rothbauer et al., 2006; Romer et al., 2011]. Further, indirect visualization can also be accomplished by fusing the protein of interest to a SunTag (SUPERNOVA tag). This short peptide tag is recognized by a co-expressed single-chain antibody (scFv), conjugated to a FP [Tanenbaum et al., 2014].

*In vivo* dynamics of labeled proteins can then be assessed via diverse live-cell imaging techniques, such as fluorescence recovery after photobleaching (FRAP), fluorescence loss in photobleaching (FLIP) and fluorescence correlation spectroscopy (FCS). For FRAP and FLIP, FP-tagged proteins within a region of interest (ROI) are irreversibly bleached by a short, strong laser pulse. Subsequently, the change of fluorescence intensity at the ROI



(FRAP) or at a defined region around the ROI (FLIP) is measured over time. FCS, in contrast, is a single-molecule detection method, which measures the fluctuations around the mean of fluorescence within a confocal volume [Weiss, 2008]. These measurements provide information about the kinetics of protein diffusion that depend on protein size but also on protein-chromatin interactions [Voss & Hager, 2008].

Sequence-specific visualization of chromatin mainly relies on FISH-based approaches. Here, the target sequence is detected via complementary base pairing with an epitope- or fluorophore-labeled nucleic acid probe after the genomic DNA has been denatured. Using this technique on fixed cells, entire chromosomes, chromosome arms or single loci can be visualized. Moreover, by combining different fluorophores, simultaneous detection of several loci or even all chromosomes can be achieved [Bolzer et al., 2005].

In contrast to analyzing nuclear protein dynamics, *in vivo* visualization of chromatin is rather limited. Replication foci have been observed by incorporation of fluorescently labeled nucleotides, which revealed different replication times for eu- and heterochromatin [Bornfleth et al., 1999; Schermelleh et al., 2001]. Further, tandem insertion of the *lac* operator (LacO) at specific genomic sites and subsequent detection via FP-tagged *lac* repressor (LacI) has been used to study repositioning of chromatin domains and mobility of telomeres [Robinett et al., 1996; Tumber & Belmont, 2001; Chuang et al., 2006; Jegou et al., 2009].

Specific chromatin regions can also be indirectly visualized via associated proteins. For instance, FP-tagged histone variant CENP-B binds to CENP-boxes and specifically localizes to centromeres [Shelby et al., 1996]. Similarly, a FP-fusion of the telomere-specific protein TRF1 has been utilized to study telomere dynamics in living cells [Krawczyk et al., 2012].

Although these methods allow for *in vivo* visualization of chromatin dynamics, target sites are limited.

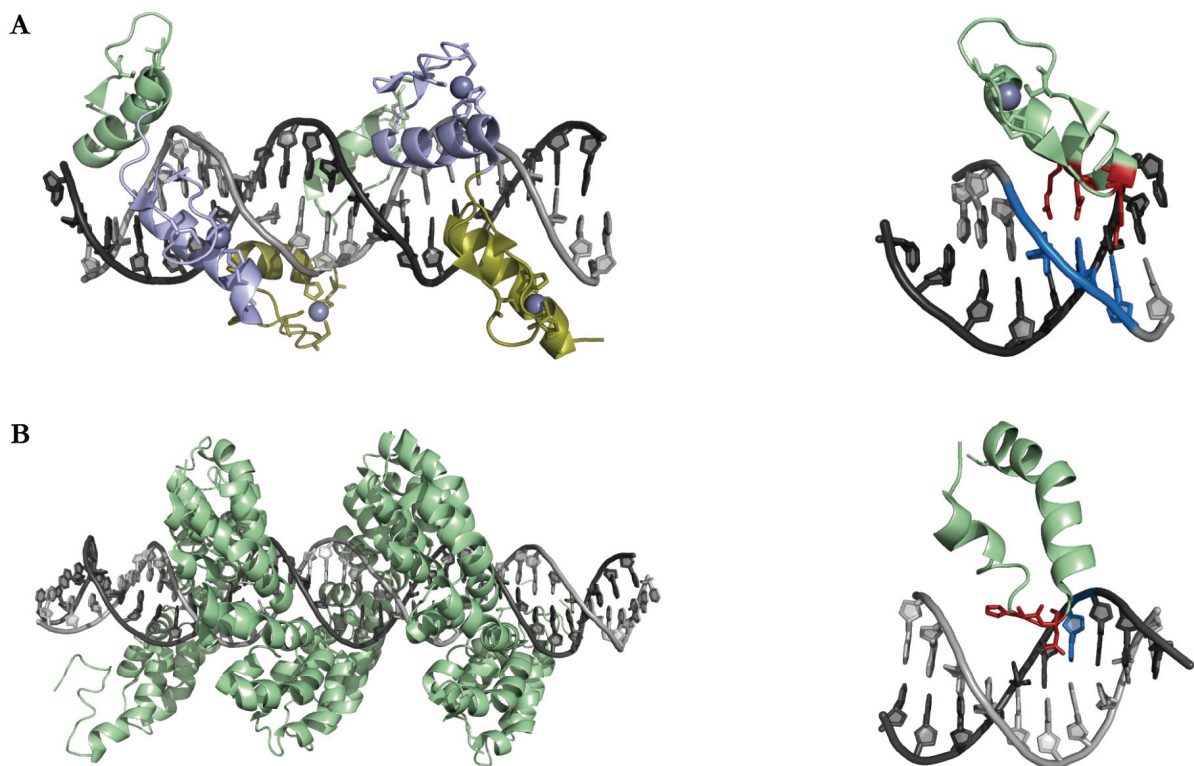
### 1.5.3 Modular DNA binding proteins

#### Zinc finger proteins

Recognition of user-defined sequences was first realized utilizing zinc finger (ZnF) proteins [Gersbach & Perez-Pinera, 2014]. This highly diverse group of proteins serves a large variety of biological functions, including transcriptional activation, protein folding, regulation of apoptosis and nucleic acid binding [Laity et al., 2001]. Notably, the classic Cys<sub>2</sub>His<sub>2</sub> ZnF structural motif, which was first described in the transcription factor IIIA from *Xenopus laevis*, is conserved among higher eukaryotes and represents the predominant DNA-binding domain in humans [Miller et al., 1985; Rubin et al., 2000; Tupler et al., 2001].

Individual Cys<sub>2</sub>His<sub>2</sub> ZnF domains comprise ~30 amino acids, forming a  $\beta\beta\alpha$ -motif, in which two cysteines and histidines coordinate a single Zn<sup>2+</sup> ion (Figure 6A). Target recognition is

predominantly facilitated by the  $\alpha$ -helix, which establishes contact to three bases within the major groove of the DNA via amino acid residues at positions -1, 3 and 6 [Pavletich & Pabo, 1991; Elrod-Erickson et al., 1996]. Since each ZnF motif recognizes a distinct base triplet, tandem arrangement of up to six modules into a polydactyl ZnF protein (PZF) enables the recognition of unique genomic loci [Liu et al., 1997]. Yet, it has been demonstrated that individual zinc fingers display a preference for GC-rich substrates and that neighboring modules affect each other's target specificity. Hence, target sequence prediction is limited and newly designed PZFs have to be subjected to a rigorous selection process, rendering this approach laborious and expensive [Segal et al., 1999; Ramirez et al., 2008; DeFrancesco, 2011].



**Figure 6: Substrate recognition by modular DNA-binding proteins.** A) Left: Crystal structure of two tandem ZnF proteins (Zif268) in complex with DNA. Individual ZnF modules are represented in different colors and the Zn<sup>2+</sup> ion is denoted as a gray sphere. Right: Single ZnF module from Zif268. Alpha-helical amino acid residues -1, 3, 6 and contacted bases in the major groove are highlighted in red and blue, respectively (PDB: 1P47). B) Left: Crystal structure of the TALE PthXo1 bound to its target sequence. The central repeat domain forms a right-handed superhelix, which is wrapped around the major groove. Right: Individual repeat, containing a HD-RVD. Amino acids 12 and 13 are marked in red and the bound base (C) is highlighted in blue (PDB: 3UGM).

### Transcription activator-like effectors (TALEs)

Due to their inherent unfavorable features, PZFs have been widely replaced by designer transcription activator-like effectors (dTALEs).

TALEs represent the major class of virulence factors in the bacterial plant pathogen *Xanthomonas* spp., which is known to cause diseases in more than 200 plant families, among them many agriculturally important crops. Key to the pathogenicity of this bacterial genus is an Hrp-type III secretion system (T3S), enabling the injection of bacterial effector proteins into the plant cell [Buttner & Bonas, 2010]. *Xanthomonas* strains typically express and secrete a mixture of 20 – 40 effector proteins with functions ranging from suppressing the hosts defense mechanisms to modulation of the plants transcriptome. Interestingly, the largest family of effector proteins, termed transcription activator-like (TAL) effectors, was found to mimic plant specific TFs [Kay et al., 2007].

TALEs are characterized by a conserved domain architecture. The N-terminal domain harbors a translocation signal peptide, which, in combination with the T3S system, mediates the translocation in the plant host cell [Szurek et al., 2002]. The C-terminal domain, on the other hand, contains nuclear localization signals (NLSs), as well as an acidic activation domain (AD), which are required for the translocation into the host nucleus and activation of target genes, respectively [Van den Ackerveken et al., 1996; Zhu et al., 1998]. Importantly, target recognition is mediated by a central repeat domain. In general, this central DNA-binding domain is composed of tandemly arranged conserved repeats, each comprising 33 – 35 amino acids. Alignment of various TALEs demonstrated that individual repeats only differ in residues 12 and 13, thus called repeat variable diresidues (RVDs). Furthermore, the length of the central repeat domain of different TAL effectors may vary from 1.5 to 33.5 repeats. However, the majority of analyzed TALEs exhibit 15.5 to 19.5 repeats [Boch & Bonas, 2010]. Crystal structures of TALEs in complex with their target sequence and biochemical studies revealed that DNA binding is mediated by the RVDs, whereas each RVD recognizes one base pair [Boch et al., 2009; Moscou & Bogdanove, 2009]. Strikingly, RVD-DNA interaction follows a simple cipher. Accordingly, NG, NI and HD display a strong preference for T, A and C, respectively. In contrast, other RVDs bind to two different bases (NN binds to G and A) or are non-selective (NS recognizes A, C, G and T) [Scholze & Boch, 2010]. TALEs bind to DNA by forming a right-handed superhelical structure along the major groove (Figure 6B). One repeat of TAL effectors comprises two  $\alpha$ -helices, which are connected by an RVD-containing loop. Interestingly, sequence-specific contacts are solely established by residue 13, which protrudes from the RVD, while residue 12 stabilizes the loop [Deng et al., 2012; Mak et al., 2012].

Due to its modular composition, the central repeat domain of TALEs can be rearranged to create dTALEs, which can be targeted to virtually any genomic sequence. TALEs have been demonstrated to be powerful tools in genome engineering [Miller et al., 2011; Mussolino et al., 2011], transcription modulation [Zhang et al., 2011; Bultmann et al., 2012; Mahfouz et al., 2012] and *in vivo* labeling of specific genomic sequences [Ma et al., 2013; Miyanari et al., 2013; Thanisch et al., 2014]. Yet, this highly repetitive structure also necessitates the use of elaborate and time-consuming cloning techniques [Morbiter et al., 2011].

### 1.6 The CRISPR/Cas system

Prokaryotes employ a variety of innate defense mechanisms against foreign viral or plasmid DNA, including restriction/modification systems and blocking of phage adsorption [Samson et al., 2013]. In recent years, it became apparent that prokaryotes additionally possess means that confer adaptive immunity against invading genomic elements [Barrangou et al., 2007].

#### 1.6.1 CRISPR/Cas mediated adaptive immunity

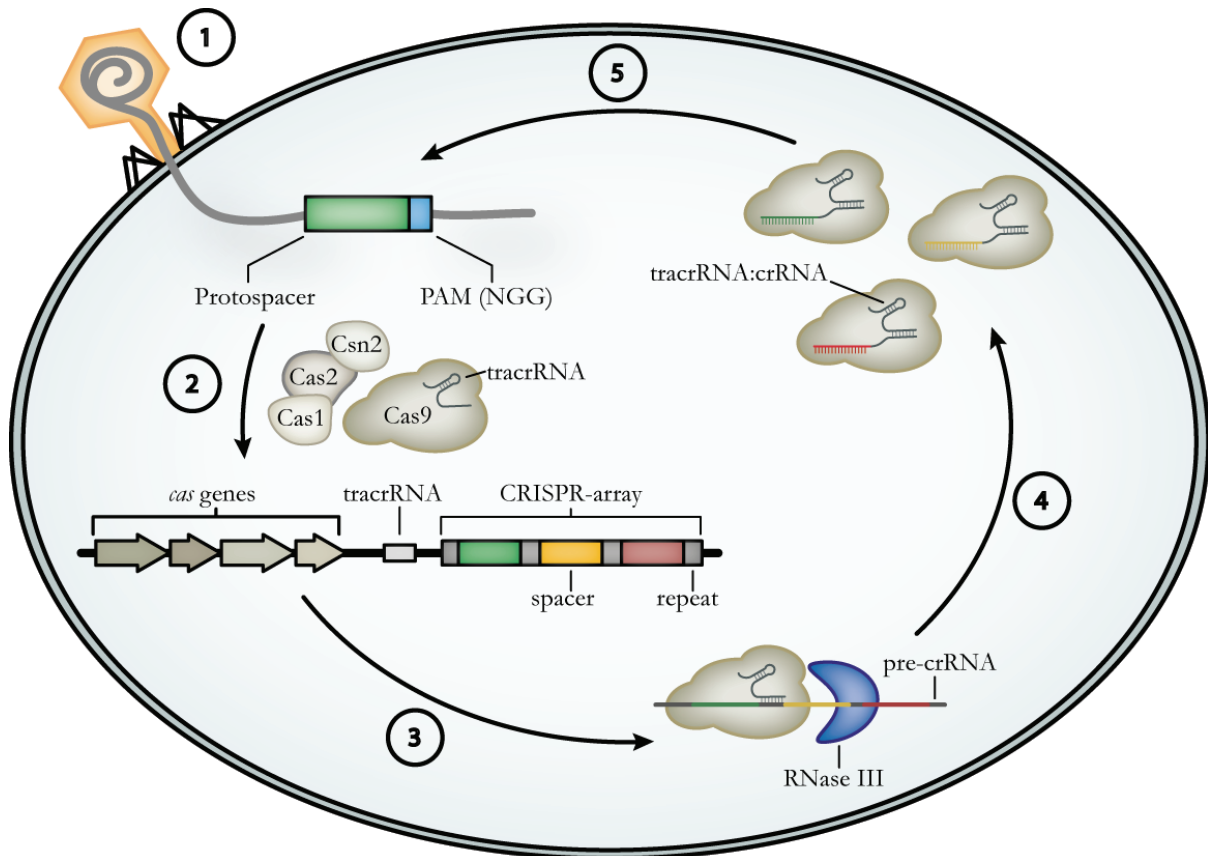
First observed in *Escherichia coli*, approximately half of bacteria (~ 45 %) and nearly all archaea (~ 84 %) are equipped with a sophisticated adaptive defense mechanisms called clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) [Ishino et al., 1987; Grissa et al., 2007; Wiedenheft et al., 2012]. Interestingly, these systems rely on small CRISPR-RNAs (crRNAs) to guide nucleases to foreign nucleic acids. All CRISPR/Cas systems comprise a set of *cas* genes, organized in operons, and a CRISPR-locus, harboring an array of genome-targeting sequences (termed spacers), which are derived from foreign DNA and are flanked by identical direct repeats [Ratner et al., 2016].

In general, CRISPR/Cas mediated adaptive immunity occurs in the following steps. First, upon phage infection or plasmid uptake, short stretches (~ 30 bp) of exogenous DNA (termed protospacer) are recognized and integrated into the CRISPR-array. This array is then transcribed, resulting in a pre-CRISPR-RNA (pre-crRNA), which is subsequently processed into small mature crRNAs. Notably, each crRNA contains a portion of the direct repeat sequence and the spacer. As a last step, the crRNA forms a ribonucleoprotein-complex with Cas protein(s) and, in some cases, a trans-activating crRNA (tracrRNA). Mediated by the complementarity between the spacer of the crRNA and the protospacer of the invading DNA, this complex then binds to and degrades its target. Importantly, target recognition is highly dependent on a short protospacer-adjacent motif (PAM) [Ratner et al., 2016].

Based on the presence of unique Cas proteins, the modes of crRNA maturation and RNA-guided interference, CRISPR/Cas systems are subdivided into three main types (type I, II and III). Type I systems utilize the endonucleases Cas6 or Cas5d to cleave the pre-crRNA within the repeat sequences and thereby facilitate the maturation of crRNA [Carte et al., 2008; Garside et al., 2012; Nam et al., 2012]. Subsequently, crRNA interacts with a complex of five Cas proteins (CasA – E) called Cascade (CRISPR-associated complex for antiviral defense) and mediates target recognition via complementary base pairing [Jore et al., 2011]. Upon target binding, conformational changes (R-loop formation) lead to the recruitment of the nuclease Cas3, which facilitates DNA degradation [Westra et al., 2012].

Similar to type I systems, type III systems also rely on Cas6 for crRNA maturation. However, in these systems, Cascade is replaced by a complex consisting of repeat units of Csm or Cmr proteins and target degradation requires Cas10 [Hrle et al., 2013; Staals et al., 2013; Staals et al., 2014; Samai et al., 2015]. Interestingly, type III systems are capable to target both DNA and RNA. Co-transcriptional RNA degradation is thought to ensure robust immunity against viral infections, when, due to mutations in protospacer sequences, DNA cleavage is abrogated [Jiang et al., 2016].

Contrary to type I and III, type II CRISPR/Cas systems only require a single endonuclease, Cas9, for crRNA maturation and target degradation. In the following the major steps for type II mediated adaptive immunity are presented in more detail (Figure 7).



**Figure 7: Type II CRISPR/Cas mediated adaptive immunity:** Upon phage infection (1), Cas9 in complex with Cas1, Cas2, Csn2 and tracrRNA, scans the invasive genetic element and selects protospacers, which contain an appropriate PAM (5'-NGG-3' for *Staphylococcus pyogenes* Cas9). The newly acquired spacer is then integrated into the CRISPR-array in a Cas1/Cas2-dependent manner (2). Following transcription of the CRISPR-array, the resulting pre-crRNA is bound by tracrRNA, by forming a repeat:anti-repeat duplex and processed into mature crRNA via RNase III in the presence of Cas9 (3). The mature crRNA stays in complex with tracrRNA and Cas9 (4) and mediates target recognition and degradation by complementarity between spacer and protospacer sequences (5).

### Spacer-acquisition in type II CRISPR/Cas systems

Upon infection, CRISPR/Cas systems must select appropriate spacer sequences in a manner that prevents autoimmunity, i.e. the cleavage of spacers after they have been integrated into the hosts CRISPR-array [Stern et al., 2010; Heler et al., 2015]. In type II (and type I) systems autoimmunity is avoided by the fact that Cas-mediated nuclease activity requires a specific nucleotide sequence, located immediately downstream of the protospacer (protospacer adjacent motif; PAM). During spacer-acquisition, Cas9 interacts with Cas1, Cas2, Csn2 and a tracrRNA and is thought to select invasive protospacers via its C-terminally PAM-interacting domain (PI) [Jinek et al., 2014; Nishimasu et al., 2014; Heler et al., 2015; Wei et al., 2015]. Interestingly, mutations in the PI do not abolish spacer-acquisition, but rather result in incorporation of spacers into the CRISPR-array, which are not adjacent to a PAM in the protospacer [Heler et al., 2015]. Moreover, nuclease activity of Cas9 is dispensable during the

acquisition of new spacers. This indicates that Cas9 is solely required for selecting the PAM and binding the protospacer, whereas Cas1 cleaves the adjacent sequence [Heler et al., 2015]. Subsequently, Cas1 and Cas2 facilitate the integration of the newly acquired spacer sequence into the CRISPR-array by interacting with the secondary structure of the CRISPR repeat. This is then followed by nicking the repeat sequence at the 3'-end and ligation of the free hydroxyl-group to the spacer sequence. However, the exact mechanism of this process remains elusive [Nunez et al., 2015].

#### CRISPR-RNA maturation in type II systems

Following transcription of the CRISPR-array, the resulting pre-crRNA is processed, yielding mature crRNAs, each of them specific for one protospacer sequence. Processing of pre-crRNA requires the base pairing of every repeat sequence with a small non-coding tracrRNA, which is encoded in the vicinity of the *cas* genes and the CRISPR-array [Deltcheva et al., 2011]. Once formed, Cas9 interacts with these tracrRNA:pre-crRNA duplexes, probably protecting the spacer sequence from endonucleolytic cleavage by host RNases. Yet, this interaction is thought to be required for the recruitment of RNase III, which subsequently cleaves both strands of RNA within the double stranded repeat region, resulting in intermediate crRNAs composed of repeat-spacer-repeat sequences [Deltcheva et al., 2011; Chylinski et al., 2013]. In an elusive, second maturation event, these intermediate crRNAs are further trimmed, yielding spacer-repeat containing crRNAs. Notably, each mature crRNA remains in complex with the processed tracrRNA and the endonuclease Cas9, forming a ternary silencing complex [Jinek et al., 2012].

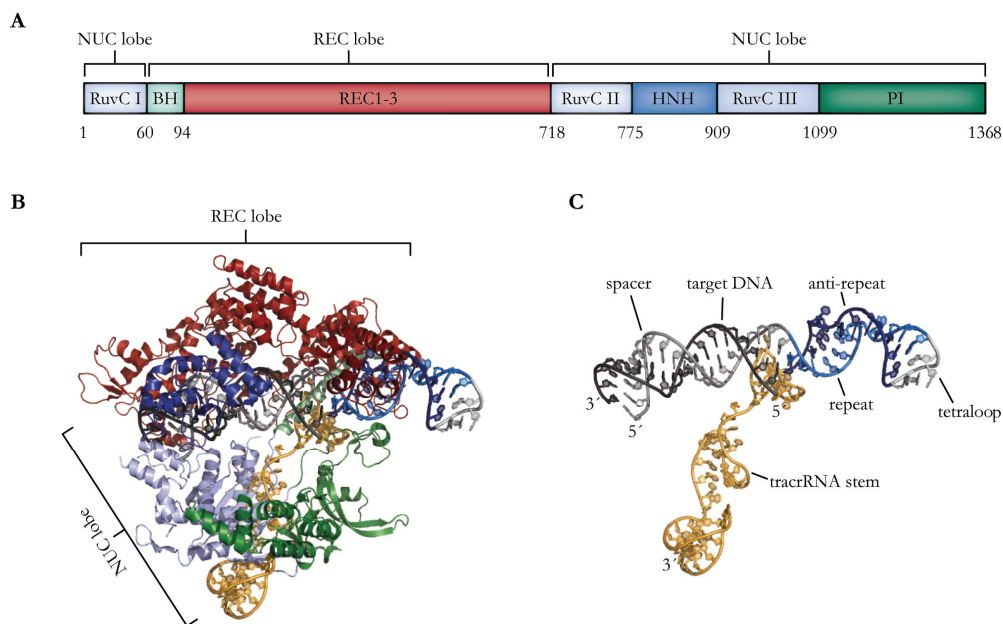
#### Target interference in type II systems

Due to extensive biochemical and crystallographic studies, the mode of target recognition and interference of Cas9 endonucleases is well established. Cas9 forms a bi-lobed structure with a larger recognition lobe (REC lobe) and a smaller nuclease lobe (NUC lobe), which are bridged by a  $\alpha$ -helix (bridge helix; BH). Additionally, the C-terminus contains a PAM-interacting domain (PI) (Figure 8A, B). The two lobes form a clam-like shape with a positively charged central channel, which accommodates both the tracrRNA:crRNA duplex and the target DNA, eventually [Jinek et al., 2014; Nishimasu et al., 2014].

Cas9 first interacts with the repeat:anti-repeat sequences of the tracrRNA:crRNA duplex via a positively charged, arginine-rich motif, located at the inner surface of the REC lobe (Figure 8B, C). RNA-binding is then followed by a conformational change of the NUC lobe, which is reoriented towards the REC lobe. Thus, the central channel is formed and the two nuclease

domains (RuvC and HNH) are positioned in a favorable way for subsequent target cleavage [Jinek et al., 2014; Nishimasu et al., 2014].

Finally, this ribonucleoprotein complex scans DNA to facilitate target cleavage. Notably, Cas9 constantly associates and disassociates from the DNA, until the PI domain encounters an appropriate PAM, which is directly associated with the target sequence [Sternberg et al., 2014; Knight et al., 2015]. In *Staphylococcus pyogenes* Cas9, the PI tightly binds to a GG dinucleotide within the PAM via two conserved tryptophan residues (W476 and W1126) [Jinek et al., 2014]. Interaction with the PAM is thought to lead to destabilization and local unwinding of the adjacent double-stranded DNA [Anders et al., 2014]. Subsequently, Cas9 is fully loaded onto the DNA, whereas the crRNA spacer displaces the non-complementary strand (R-loop) and forms a RNA:DNA heteroduplex. Interestingly, mismatches between crRNA and target DNA at the 5'-region of the spacer seem to be tolerated to some extent [Kabadi & Gersbach, 2014]. The separation of the DNA strands leads to their placement into the active sites of the two nuclease domains. The RuvC domain subsequently cleaves the non-complementary strand, whereas the HNH domain targets the complementary strand 3 bp upstream of the PAM [Jinek et al., 2014; Nishimasu et al., 2014; Sternberg et al., 2015].



**Figure 8: Target recognition in type II CRISPR systems.** A) Domain organization of *Streptococcus pyogenes* Cas9. The endonuclease contains a recognition (REC, red) and a nuclease (NUC, blue) lobe. The nuclease domains RuvC I – III and HNH are highlighted in light- and dark-blue, respectively. The C-terminally located PAM-interacting (PI, dark-green) domain mediates PAM-recognition and prevents autoimmunity. NUC and REC lobe are connected by a bridge helix (BH, light-green). B) Crystal structure of *S. pyogenes* Cas9. Color-coding of Cas9 as in A). The two lobes form a clam-like structure with a central channel, which accommodates the target DNA (black) and the sgRNA (blue and gray). C) Crystal structure of sgRNA, binding its target DNA (black) via Watson-Crick base pairing. The tracrRNA:crRNA duplex is synthetically connected by a tetraloop (light-gray) to form a single guide RNA (sgRNA) (PDB: 4OO8).



### 1.6.2 The Cas9 toolbox

Due to the fact that Cas9 recognizes its genomic target via Watson-Crick base pairing, this system can be reprogrammed to bind to user-defined sequences, by simply exchanging the spacer sequence of the crRNA. CRISPR/Cas9 genome targeting systems can be even further simplified, by replacing the two RNA components, which have to be expressed separately, by a synthetic fusion of tracrRNA and crRNA, named single guide RNA (sgRNA) (Figure 8C). By introducing a tetraloop-linker between the crRNA and tracrRNA, this chimeric sgRNA retains the double stranded repeat:anti-repeat duplex and structural features that are necessary for Cas9 interaction but can be transcribed from a single promoter [Jinek et al., 2012].

The RNA-guided endonuclease Cas9 has been adapted as a tool for genome engineering in a great variety of cell types and organisms, including human, mouse, fly, worm and zebrafish [Mali et al., 2013; Dickinson & Goldstein, 2016; Heno-Mejia et al., 2016; Housden & Perrimon, 2016; Vejnár et al., 2016]. For genome engineering purposes, Cas9 is utilized to introduce a DNA double strand break (DSB) at the desired locus. Mediated by the hosts repair machinery, DSBs are either resolved by non-homologous end joining (NHEJ) or homology-directed repair (HDR). The error-prone NHEJ-pathway results in a high frequency of insertions or deletions (indels) near the break-site, causing premature stop codons or other frame shift mutations and eventually a knock-out of gene function. Yet, in the presence of a homologous donor DNA, DSBs can be repaired, while introducing defined sequences [Lieber, 2010]. To increase the likelihood of HDR, a partially mutated Cas9 has been used. Engineered point mutations in either of nuclease domains (D10A and H840A for RuvC and HNH, respectively), convert Cas9 into a nicking enzyme. This decreases the frequency of NHEJ and single strand breaks are preferentially repaired by HDR [Jinek et al., 2012; Cong et al., 2013].

Importantly, by mutating both nuclease domains, Cas9 can be engineered into a RNA-guided DNA-binding platform [Jinek et al., 2012; Qi et al., 2013]. This catalytically dead Cas9 (dCas9) still binds DNA in a sequence-specific manner but without cleaving the underlying target. Hence, the CRISPR/Cas9 system can be repurposed to manipulate gene expression. In this context, it has been demonstrated that dCas9 bound to a promoter sequence efficiently represses either transcription initiation by blocking TF and RNA polymerase (RNAP) binding or transcription elongation by inhibiting RNAP progression [Qi et al., 2013]. Similarly, transcription of a specific gene can also be repressed by fusing dCas9 to a transcriptional repressor, such as KRAB or SID4X [Gilbert et al., 2013; Konermann et al., 2013]. In addition to transcriptional silencing, dCas9, fused to transcriptional activators has been successfully employed to substantially increase transcription of targeted genes [Bikard et al., 2013; Cheng et al., 2013; Kearns et al., 2013; Perez-Pinera et al., 2013]. Furthermore, the transcriptional

output of specific genes has been manipulated by fusing dCas9 to epigenetic effector proteins. For instance, dCas9 fused to the human HAT p300 is able to specifically target H3K27 acetylation to distal enhancers or promoter regions of the *OCT4* gene, resulting in an increased mRNA level [Hilton et al., 2015]. Targeted *de novo* methylation via a dCas9-DNMT3A fusion, on the other hand, was used to silence the *IL6ST* gene in human cells [Vojta et al., 2016].

Taken together, these results highlight the versatility of CRISPR/Cas9-based approaches in both genome engineering and manipulation of transcription.

## 1.8 Aims of this work

Despite the fact that the genomes of many organisms have been sequenced, labeling and tracing of specific sequences in living cells has been a major challenge in studying the spatiotemporal dynamics of native chromatin. To date, tools to investigate nuclear organization either rely on prior fixation (FISH), or are based on DNA binding properties of fluorescently tagged proteins, such as PZFs or dTALEs. Designing sequence-specific PZFs and dTALEs for new target sequences, however, remains challenging.

Since 2012, the emergence of the CRISPR/Cas system opened new experimental possibilities for sequence-specific DNA recognition. The endonuclease Cas9 plays a vital role in the type II CRISPR/Cas system and can be recruited to endogenous loci by a sgRNA, whereas target specificity is mediated by complementarity between ~20 bp of the sgRNA and the corresponding DNA sequence. Importantly, a catalytic mutant of Cas9 (dCas9) that retains its DNA-binding capability without inducing DSBs at the target sequence, was described.

The first objective of this work was to harness dCas9 as a programmable DNA-binding platform for the visualization of distinct genomic loci *in vivo*. For this, we constructed a dCas9-eGFP protein and co-expressed it with sgRNAs, which were specific for endogenous repetitive sequences. To test the target-specificity of this new approach, we validated the signals, obtained by CRISPR imaging, with FISH and immunofluorescence microscopy.

Besides the spatiotemporal localization, the function of individual genomic loci is determined by their epigenetic status. To elucidate whether the CRISPR/Cas system is a suitable tool for epigenetic studies, we set out to target epigenetic modifications to specific sites. For this aim, we fused dCas9 to a GFP-binding protein (GBP). This setup enabled us to recruit GFP-tagged epigenetic effectors to a genomic region of interest and to investigate their functions in a defined environment.

As a third objective, we aimed to explore the local chromatin composition of defined genomic elements. To this end, we combined the programmable DNA-binding of dCas9 with a promiscuous biotin ligase (BirA\*). By targeting the BirA\*-dCas9 fusion to defined genomic loci, we aimed to achieve biotinylation of target-associated proteins, which can be precipitated and identified via tandem mass spectrometry.



## 2 RESULTS

### 2.1 Visualization of specific DNA sequences with a programmable fluorescent CRISPR/Cas system



# Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system

Tobias Anton, Sebastian Bultmann, Heinrich Leonhardt\*, and Yolanda Markaki\*

Department of Biology II; Center for Integrated Protein Science Munich (CIPSM); Ludwig Maximilians University Munich; Planegg-Martinsried, Germany

**Keywords:** CRISPR/Cas9, embryonic stem cells, DNA labeling, 3D-fluorescent in situ hybridization (3D-FISH), major satellite repeats, minor satellite repeats, telomeres, CENP-B, TRF2, 3D-SIM

**Abbreviations:** CRISPR, Clustered Regulatory Interspaced Short Palindromic Repeats; Cas, CRISPR-associated; eGFP, enhanced-green fluorescent protein; 3D-FISH, 3D-fluorescent in situ hybridization; gRNA, guide RNA; bp, base pair; MaS, major satellite repeats; MiS, minor satellite repeats; CC, chromocenter; 3D-SIM, 3D-structured illumination microscopy

Labeling and tracing of specific sequences in living cells has been a major challenge in studying the spatiotemporal dynamics of native chromatin. Here we repurposed the prokaryotic CRISPR/Cas adaptive immunity system to specifically detect endogenous genomic loci in mouse embryonic stem cells. We constructed a catalytically inactive version of the Cas9 endonuclease, fused it with eGFP (dCas9-eGFP) and co-expressed small guide RNAs (gRNAs) to target pericentric, centric, and telomeric repeats, which are enriched in distinct nuclear structures. With major satellite specific gRNAs we obtained a characteristic chromocenter (CC) pattern, while gRNAs targeting minor satellites and telomeres highlighted smaller foci coinciding with centromere protein B (CENP-B) and telomeric repeat-binding factor 2 (TRF2), respectively. DNA sequence specific labeling by gRNA/dCas9-eGFP complexes was directly shown with 3D-fluorescent in situ hybridization (3D-FISH). Structured illumination microscopy (3D-SIM) of gRNA/dCas9-eGFP expressing cells revealed chromatin ultrastructures and demonstrated the potential of this approach for chromatin conformation studies by super resolution microscopy. This programmable dCas9 labeling system opens new perspectives to study functional nuclear architecture.

## Introduction

Almost one hundred and fifty years since the original description of chromosomes, many genomes have been fully sequenced. While our knowledge of DNA sequences has reached base pair (bp) resolution, detailed information on the positioning, nuclear arrangement and interactions of specific gene loci in living cells is still limited.

Despite the absence of internal membranes, the nucleus is a highly organized organelle. With fluorescent in situ hybridization (FISH) all chromosomes in an interphase nucleus were mapped and shown to occupy distinct territories.<sup>1</sup> Additionally, gene rich chromosomes were found to be preferably located in the center of the nucleus, whereas gene poor chromosomes reside mostly in proximity to the nuclear periphery reflecting a functional nuclear organization.<sup>2,3</sup> Clearly, FISH represents an important tool to label specific DNA sequences and to study nuclear architecture, but it is restricted to fixed specimens.

By now, several methods have been employed to label DNA in vivo<sup>4</sup> and to investigate chromatin spatiotemporal dynamics,

for instance by the incorporation of fluorescently tagged chromatin proteins, like H2B-GFP.<sup>5</sup> However, these methods do not distinguish specific genomic sequences. To overcome these limitations, the *lac* operator and/or repressor recognition system has been developed.<sup>6</sup> This system, however, relies on artificially introduced sequences and does not provide information on endogenous genomic loci.

To date, targeting of specific endogenous genomic loci has been based on the sequence specific binding of Cys<sub>2</sub>His<sub>2</sub> zinc finger modules (ZF),<sup>7-9</sup> where individual ZFs bind to a distinct trinucleotide sequence and are combined into polydactyl zinc finger proteins (PZF). It has been shown, however, that the target specificity of ZFs can be affected by their neighboring modules, which requires evaluation of every newly designed PZF.<sup>10,11</sup> PZFs have been widely replaced by designer transcription activator-like effectors (dTALEs), which have proven to be a powerful tool for genome engineering,<sup>12,13</sup> influencing gene transcription<sup>14-16</sup> and were recently applied for labeling genomic sequences in vivo.<sup>17-19</sup> Yet, the fact that DNA binding of dTALEs is mediated by tandemly arranged repeats, whereby every repeat only differs in two residues, necessitates the use of elaborate cloning techniques.<sup>20,21</sup>

\*Correspondence to: Heinrich Leonhardt; Email: h.leonhardt@lmu.de; Yolanda Markaki; Email: yolanda.markaki@lrz.uni-muenchen.de  
Submitted: 02/17/2014; Revised: 03/04/2014; Accepted: 03/10/2014; Published Online: 03/12/2014  
<http://dx.doi.org/10.4161/nucl.2848>

**Figure 1 (See opposite page).** Labeling genomic loci with specific dCas9-eGFP/gRNA complexes. **(A)** Schematic representation of the dCas9-eGFP expression construct. A chicken  $\beta$ -actin promoter with a CMV enhancer (CAG, blue triangle) drives the expression of dCas9-eGFP. Inactivation of RuvC1 and HNH (red crosses) by amino acid substitutions (D10A and H840A within RuvC1 and HNH, respectively) is indicated. A second nuclear localization signal (NLS) (beige) is introduced upstream of the eGFP coding sequence (green). **(B)** Outline of the experimental design. dCas9-eGFP interacts with a co-expressed gRNA and is thereby guided to the genomic target sequence. Note that the presence of a protospacer adjacent motif (PAM) is a prerequisite for dCas9 binding. **(C)** Schematic representation of a mouse acrocentric chromosome. gRNAs were designed to target 20 bp protospacer sequences of telomeres (Tel, green), major satellites (MaS, yellow) and minor satellites (MiS, red) as indicated. **(D, E, F)** Co-expression of dCas9-eGFP and gRNAs complementary to MaS repeats (MaSgRNA, **D**), MiS (MiSgRNA, **E**) and Tel gRNAs (TelgRNA, **F**) in J1 mouse embryonic stem cells. MaSgRNA recruits dCas9-eGFP to chromocenters (CCs), MiSgRNA/dCas9-eGFP signals are observed in the periphery of CCs, while targeting of TelgRNA/dCas9-eGFP to telomeres results in distinct dCas9-eGFP foci, which can be detected throughout the nucleoplasm. Bar: 5  $\mu$ m.

New experimental options became available with the discovery of the type II CRISPR/Cas system that is composed of clustered regularly interspaced short palindromic repeats (CRISPR) as well as CRISPR-associated (Cas) proteins and plays a vital role in prokaryotic adaptive immunity. Upon viral infection or plasmid uptake, short stretches (~30 bp) of foreign DNA (termed spacers) are incorporated between identical direct repeats into CRISPR arrays. Transcription of these arrays results in pre-CRISPR RNA (pre-crRNA), which subsequently interacts with a transactivating crRNA (tracrRNA). This pre-crRNA/tracrRNA duplex forms a complex with the endonuclease Cas9, followed by further processing of pre-crRNA into crRNA. Endonuclease target-specificity is determined by the complementarity between spacer (crRNA) and protospacer (viral or plasmid) sequences.<sup>22</sup> As an important step toward the applicability of this system, a crRNA:tracrRNA chimera, named guide RNA (gRNA), has been shown to be able to replace the two RNA components and to specifically target Cas9 to user-defined DNA sequences.<sup>23,24</sup> The CRISPR/Cas system has recently been applied as a versatile tool for genome editing in a wide range of organisms.<sup>25-29</sup>

Here, we present an approach for labeling specific endogenous genomic loci in living murine embryonic stem cells based on a modified CRISPR/Cas system.

## Results

### Adapting the CRISPR/Cas system for tracing specific DNA sequences in living cells

The CRISPR/Cas system features easily programmable sequence recognition, but combines it with an endonuclease activity. We, therefore, introduced mutations to inactivate the endonuclease activity of Cas9<sup>30-32</sup> and fused it to the enhanced variant of GFP (eGFP) (Fig. 1A). By co-transfecting a plasmid encoding this eGFP-tagged, nuclease deficient Cas9 (dCas9-eGFP) together with a gRNA expression vector, we aimed to target specific genomic loci in mouse embryonic stem cells. In this way, we expected to achieve specific targeting of dCas9-eGFP without cleavage of the underlying sequences (Fig. 1B).

To test the feasibility of our method, we chose tandemly arranged repetitive DNA sequences, which enabled us to target dCas9-eGFP to extended genomic loci with a single gRNA construct. To this end, gRNAs directed to major (MaSgRNA) and minor satellite (MiSgRNA) repeats, as well as telomeres (TelgRNA) were designed. In mice, major satellite repeats consist of 234 bp repeat units that span within a 6 Mb region, whereas

minor satellite repeats range between 0.6 to 1.2 Mb and consist of 120 bp repeat units.<sup>33-36</sup> Telomeric repeats vary in length reflecting the cell's replicative potential<sup>37</sup> and in mouse amount to approximately 20–30 kb with the 6 bp repeat sequence TTAGGG (Fig. 1C).

### Effective tracing of repetitive DNA sequences using the CRISPR/Cas system

Mouse chromosomes are acrocentric as depicted in Figure 1C. Fluorescent in situ hybridization (FISH) experiments on metaphase chromosomes have shown that minor satellites are centromeric, whereas major satellite repeats occupy the subcentromeric part of the chromosome and are implicated in heterochromatin formation and sister chromatid cohesion.<sup>38-41</sup> In interphase nuclei, centromeres cluster and form distinct chromocenters (CCs),<sup>41,42</sup> which can be readily distinguished by enhanced DAPI-staining intensity due to their AT-richness.

J1 mouse embryonic stem cells were co-transfected with dCas9-eGFP and MaSgRNA encoding plasmids and imaged 48 h post-transfection. As depicted in Figure 1D, the CCs of MaSgRNA/dCas9-eGFP expressing cells show a bright eGFP signal, verifying the successful targeting of genomic DNA. The distribution exhibits remarkable specificity with very low background signals from freely diffusing dCas9-eGFP.

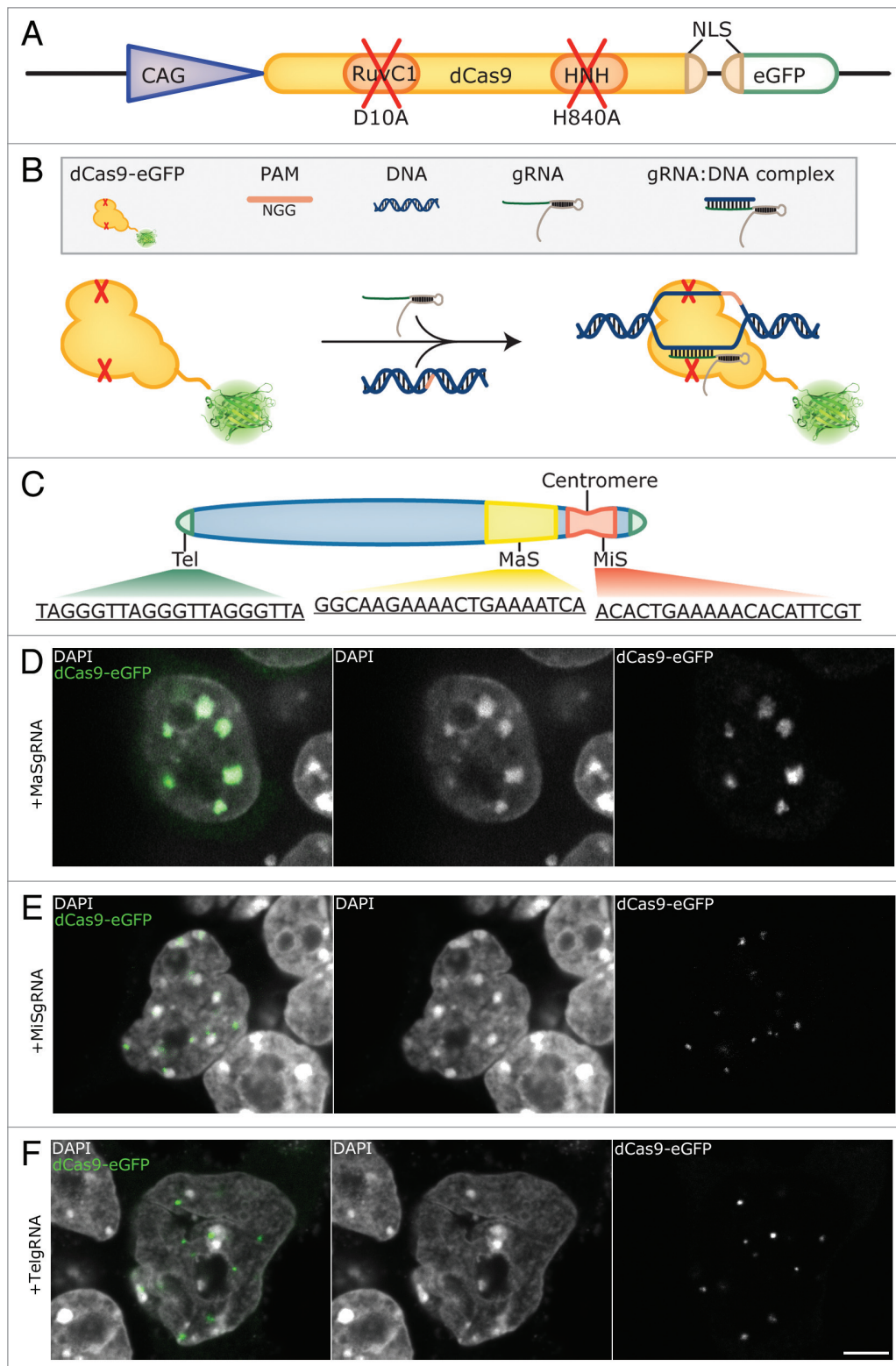
MiS repeats have been observed as individual focal entities at the periphery of CCs.<sup>41</sup> We were able to observe this characteristic distribution as eGFP fluorescent foci in the corresponding regions of MiSgRNA/dCas9-eGFP expressing cells (Fig. 1E).

Telomeres are capping the ends of chromosomes and have been found localized throughout the nucleus.<sup>43</sup> Due to the acrocentric nature of mouse chromosomes, telomeres located at the acrocentric end can also be detected in the direct vicinity of CCs apart from remote nuclear sites. By expressing TelgRNA/dCas9-eGFP in J1 cells, small foci, often in the vicinity of CCs, were visible. The observed variable size of labeled telomeres is consistent with the fact that telomeres may form clusters<sup>44</sup> (Fig. 1F). In some TelgRNA/dCas9-eGFP expressing cells, an elongated and fiber-like telomere-related eGFP signal was apparent, which was not observed in cells expressing MaS- or MiSgRNA/dCas9-eGFP. Taken together our results show that for all three targeted DNA sequences distinct labeling patterns were observed.

### Combined gRNA/dCas9-eGFP labeling and 3D DNA-FISH

To validate the sequence specificity of the dCas9 labeling system, we verified the accurate targeting of all live-cell introduced fluorescent genomic tags by 3D DNA-FISH experiments. For this purpose, FISH probes specific for MaS, MiS and telomeric repeats





**Figure 1.** For figure legend, see page 164.

were generated and immuno-FISH experiments were performed. We used a protocol optimized for eGFP epitope preservation and

efficient probe hybridization,<sup>45</sup> yielding robust and homogeneous signals (Fig. 2). As it can readily be seen in Figure 2A, the probe

for MaS repeats shows confined localization at the CCs, while MaSgRNA/dCas9-eGFP complexes exhibit strict co-localization with the probe. Accordingly, immuno-FISH experiments showed correct localization of MiSgRNA/dCas9-eGFP at MiS repeats (Fig. 2B). Moreover, simultaneous hybridization with MaS and MiS probes showed that the gRNA/dCas9-eGFP complexes specifically labeled their respective target structure and did not affect the spatial integrity of neighboring structures (Fig. 2C). In immuno-FISH of TelgRNA/dCas9-eGFP expressing cells, FISH probe and dCas9-eGFP were fully co-localized and exhibited the same variety of signal intensities with brighter foci possibly representing longer telomeres or telomere clusters (Fig. 2D).

In each case, neighboring non-transfected cells in our specimens served as an internal control for the visualization of normal distribution of the FISH probes and comparison of signal strength. It must be noted here that on some occasions FISH probes did not fully penetrate to the core of well-associated stem cell colonies, while dCas9-eGFP signal was detected. Our observations confirm the specific and restricted localization of dCas9-eGFP to their targeted DNA sequences.

#### **gRNAs/dCas9-eGFP complexes remain associated with mitotic chromosomes**

During cell division, chromatin undergoes dramatic structural changes in order for chromosomes to compact and segregate properly. This chromatin rearrangement could possibly affect binding of recombinant macromolecular complexes. In addition, the targeted MiS repeats, forming the centromeres, are the sites of kinetochore formation and ensure normal chromosome segregation.<sup>46</sup> Likewise, pericentromeric regions, containing MaS repeats, play an important role in mitotic progression, as they have been linked to sister-chromatid cohesion.<sup>47-49</sup> Furthermore, telomeric dysfunction has been shown to promote chromosome fusions, anaphase bridges and genome reorganization.<sup>50-52</sup> Therefore, targeting of dCas9-eGFP at these essential regulatory sequences could potentially perturb chromosome arrangements and mitotic progression. Hence, we investigated, whether dCas9-eGFP stably maintained its association to targeted DNA sequences during mitosis. We were able to observe MaS, MiS, and TelgRNA/dCas9-eGFP expressing cells in different mitotic stages. A metaphase for the MaSgRNA/dCas9-eGFP targeting is depicted in Figure 3A, showing the robust labeling of the MaS domains and its restricted localization to the highly condensed CC regions. Similarly, two optical sections of a metaphase plate (z1, z2) showed that TelgRNA/dCas9-eGFP labeled both ends of mitotic chromosomes and multicolor FISH confirmed the pericentromeric and centromeric arrangement of MaS and MiS repeats, respectively (Fig. 3B).

For a more detailed investigation of the MiS signals, we applied 3D-SIM microscopy and examined metaphases in MiSgRNA/dCas9-eGFP expressing cells (Fig. 3C). We were able to distinguish distinct dCas9-eGFP foci in every chromosome and strict co-localization with the MiS probe. Using conventional microscopy, MiS clusters appear as diffraction-limited foci. With 3D-SIM, these clusters displayed ultrastructural organization, visible by FISH probe, as well as dCas9-eGFP labeling (Fig. 3C, magnification insets).

#### **Application of CRISPR/Cas labeling in super resolution studies**

Next, we investigated whether the coverage of dCas9-eGFP labeling correlated with known proteins bound at these genomic loci, perturbed the canonical distribution of these proteins or prevented antibody binding. For this purpose, J1 cells expressing MiSgRNA/dCas9-eGFP were immunolabeled with anti-centromere protein-B (CENP-B) antibodies. CENP-B is a constitutive centromere protein located at the primary constriction by direct binding to MiS repeats at the CENP-B box.<sup>53-55</sup> In mitosis it forms a link between the kinetochore and the underlying centromeric repeats<sup>56,57</sup> and regulates centromere formation.<sup>58,59</sup> In interphase it has been visualized in association with MiS repeat clusters at the periphery of the CCs.<sup>41</sup>

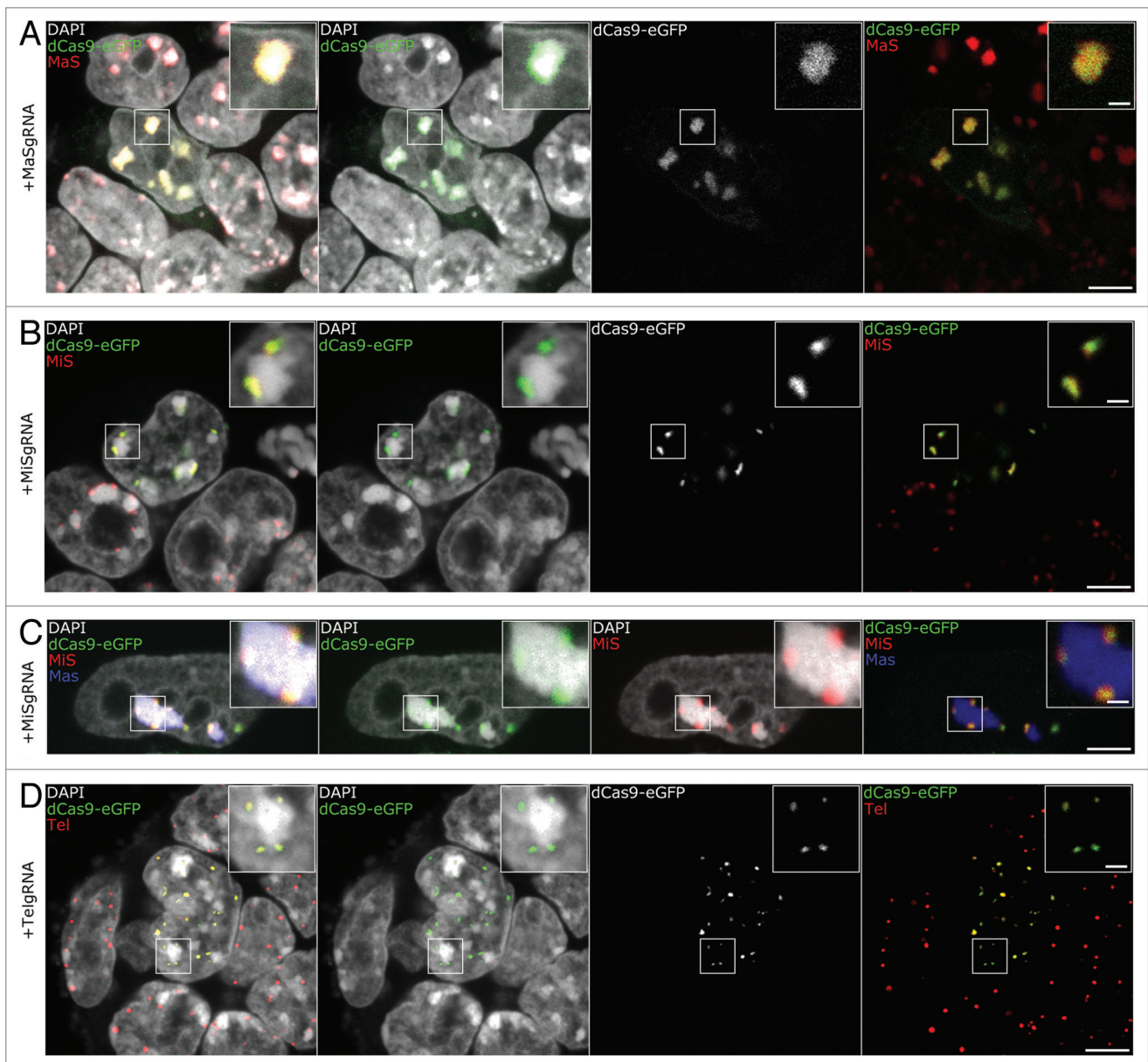
In MiSgRNA/dCas9-eGFP expressing cells, distinct CENP-B foci were consistently associated with dCas9-eGFP clusters (Fig. 4A), suggesting that incorporation of the gRNA/dCas9-eGFP does neither interfere with the recruitment of CENP-B to MiS repeats, nor antibody binding upon fixation. Compared with wide-field deconvolution (wf), where MiS domains appeared as diffraction-limited foci, 3D-SIM revealed a sub-structural organization of the targeted repeats (Fig. 4A, blowups 3D-SIM compared with wf).

Consistent with the above findings, in TelgRNA/dCas9-eGFP expressing cells, the incorporation of the complexes to telomeric repeats allowed binding of anti-telomeric repeat-binding factor 2 (TRF2) antibodies (Fig. 4B). Telomeric DNA sequences have been shown to form a T-loop structure, which is thought to protect the 3'-overhangs and regulate telomerase activity.<sup>60,61</sup> By conventional fluorescence microscopy, telomeres are detected as diffraction-limited foci. In a recent study applying 3D-STORM, FISH detected telomeres appeared as ovoid clusters with an average diameter of 180 nm.<sup>62</sup> Although SIM does not offer the localization accuracy of STORM,<sup>63</sup> high-resolution microscopy of telomeres using the dCas9 labeling system revealed a consistent morphology to the above-mentioned study (Fig. 4B).

With dCas9 labeling, we were able to visualize repetitive DNA sequences on fixed cells without the need of FISH probes and thus avoided flattening or destruction of chromatin due to sample denaturation. Furthermore, combination of 3D-SIM with immunostaining and dCas9 labeling highlighted ultrastructural properties of targeted MiS and Tel repeats.

## **Discussion**

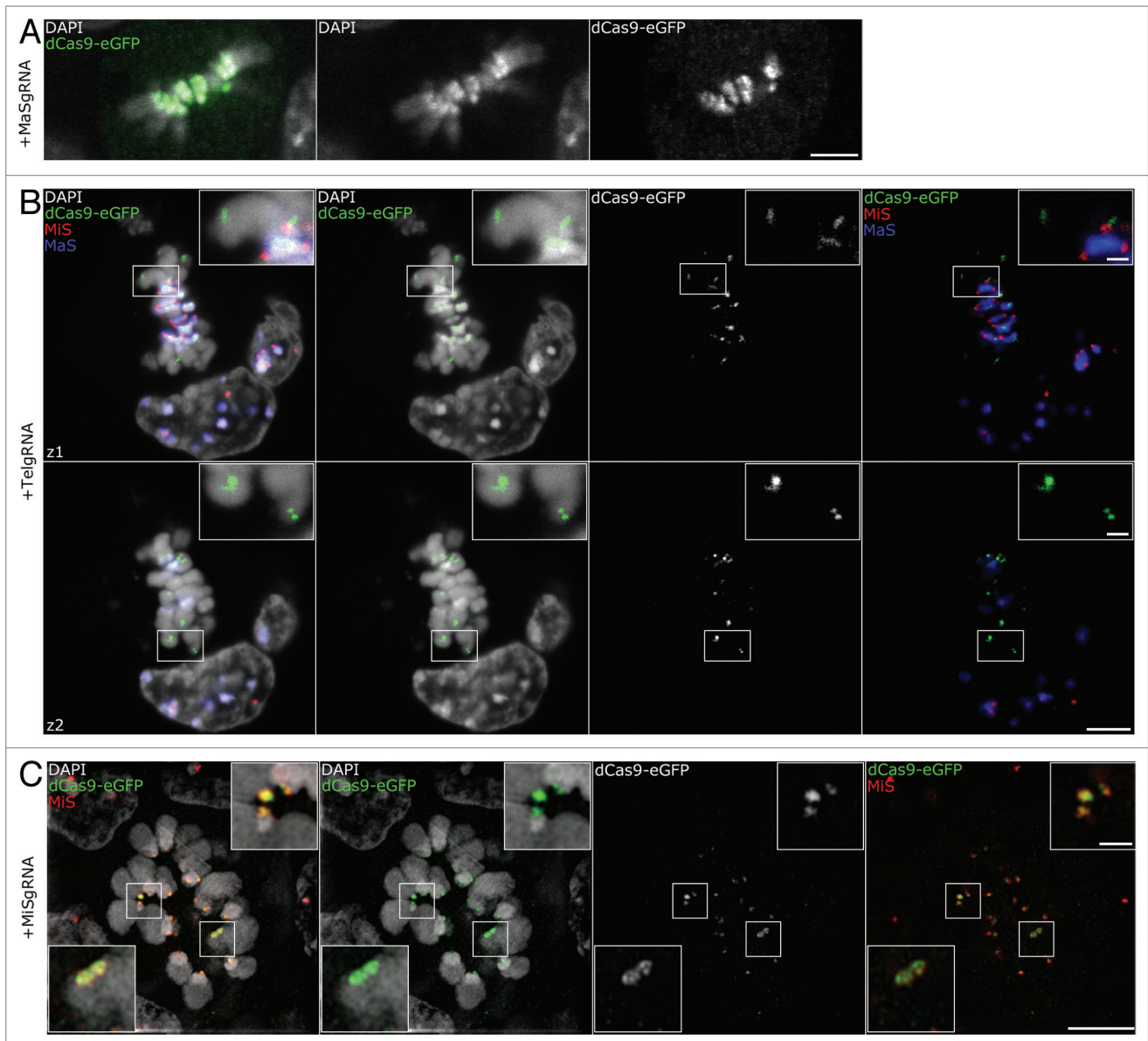
For modern cytogenetics and diagnostics, fluorescent in situ hybridization (FISH) has proven to be an indispensable method, but its application is still limited, due to its complexity and variability. Since its original introduction, numerous scientific publications have dealt with the optimization of probe and sample preparation. Although improvements have been made, harsh treatments, such as heat denaturation are required for probe hybridization that may compromise sample integrity.<sup>64,65</sup> Furthermore, combining 3D-FISH with protein



**Figure 2.** 3D-FISH shows precise targeting of dCas9-eGFP. **(A, B, C, D)** Immuno-FISH experiments in gRNA/dCas9-eGFP expressing cells. Hybridization of probes designed to target MaS **(A, red and C, blue)**, MiS **(B-C, red)**, Tel **(D, red)** demonstrate that dCas9-eGFP co-localizes with the respective sequences. Multicolor immuno-FISH in **C** (MiS, red; MaS, blue) highlights the restricted targeting of MiSgRNA/dCas9-eGFP and demonstrates that non-targeted neighboring nuclear structures maintain their integrity. Bars, 5  $\mu\text{m}$ ; insets, 1  $\mu\text{m}$ .

immunofluorescent detection remains a challenge and on occasion is not feasible due to detrimental effects on epitopes.<sup>66</sup> Therefore, even at the fixed cell level, live-cell genomic labeling systems offer superior sample preservation and their use is far less laborious compared with FISH probe creation and long hybridization and/or detection procedures. Noteworthy is the fact that at the dense stem cell colony level, we observed on occasion that FISH probes did not penetrate efficiently the core of the colony. In these cells, dCas9 labeling was uniform thereby rendering this method more efficient than FISH.

Here we report the successful labeling of endogenous centric, pericentric and telomeric chromatin loci in living mouse embryonic stem cells by repurposing the prokaryotic CRISPR/Cas system. During the preparation of this manuscript, a similar study has been published,<sup>67</sup> which showed that, besides repetitive sequences, also individual genes can be labeled by a catalytically inactive Cas9 endonuclease, confirming the potential of the dCas9 system to label genomic DNA in vivo. While their study focused on detection sensitivity, we performed a thorough comparison with 3D-FISH and immunolabeling methods.



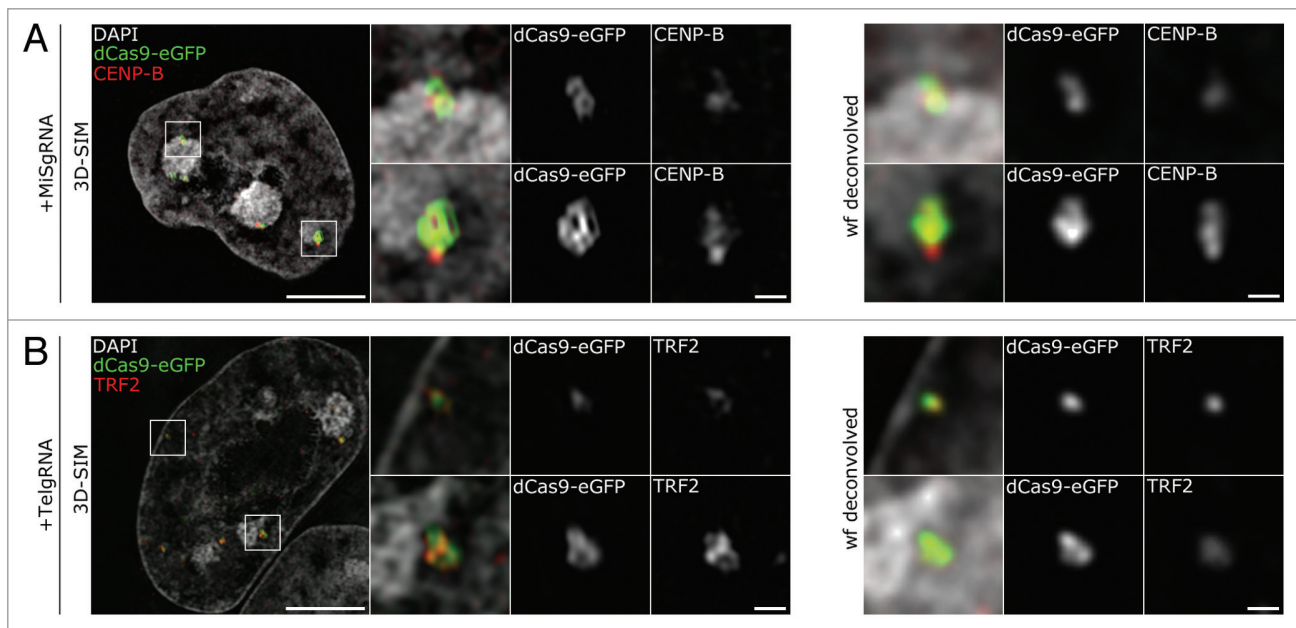
**Figure 3.** Association of gRNA/dCas9-eGFP to chromatin in mitotic cells. **(A)** Confocal optical section of a metaphase plate shows successful targeting of MaSgRNA/dCas9-eGFP. Note the robust eGFP signals at the CCs (DAPI bright regions, middle panel). Bar, 5  $\mu\text{m}$ . **(B)** Two confocal optical sections (z1, z2) of a multicolor immuno-FISH stained metaphase show TelgRNA/dCas9-eGFP signals at the ends of chromosomes. The Integrity of (peri-) centromeric chromatin (MaS, MiS) is not compromised. Bar, 5  $\mu\text{m}$ ; insets, 1  $\mu\text{m}$ . **(C)** Metaphase plate of a MiSgRNA/dCas9-eGFP expressing cell after immuno-FISH with MiS probe (red) acquired via 3D-SIM. Note the overlap between MiS probe and dCas9-eGFP. Bar, 5  $\mu\text{m}$ ; inset, 1  $\mu\text{m}$ .

In addition we demonstrated potential applications in high resolution microscopy studies.

So far, dTALE based approaches are commonly applied as powerful tools for genome manipulation and live-cell tracing of chromatin. However, large-scale studies of nuclear organization require the design of dTALEs specific for many different target sequences including single copy genes. Due to laborious cloning procedures, large-scale production of these DNA binding proteins still remains a challenge. In contrast, the target specificity of the dCas9 labeling system relies on small and easily exchangeable

gRNAs, which greatly expands the range of possible targets and is the basis for the convenience of dCas9 labeling. Moreover, combining dCas9 labeling with 3D-SIM microscopy enabled us to visualize the ultrastructure of diffraction limited chromatin clusters and their spatial relationship with known associated proteins.

Clearly, this programmable dCas9 DNA labeling system represents a powerful tool to monitor the spatiotemporal dynamics of endogenous genomic loci during cell cycle progression and differentiation and opens new perspectives to study functional nuclear architecture.



**Figure 4.** 3D-SIM highlights the ultrastructure of MiS and Tel repeats. **(A)** Left panel depicts a mid z-section of a DAPI stained nucleus (gray) in MiSgRNA/dCas9-eGFP (green) expressing cells immunolabeled with anti-CENP-B antibodies (red). Four  $\times$  magnifications of boxed areas (mid gallery) show the spatial association of dCas9-eGFP decorated domains to CENP-B assemblies. Wide-field (wf) deconvolved simulations of the corresponding 3D-SIM magnifications are shown for comparison (far right gallery). Bar, 5  $\mu$ m; magnifications, 500 nm. **(B)** Left panel depicts a mid z-section of a DAPI stained nucleus (gray) in TelgRNA/dCas9-eGFP (green) expressing cells immunolabeled with anti-TRF2 antibodies (red). Four  $\times$  magnifications of boxed areas (mid gallery) show overlapping pattern of dCas9-eGFP signals with TRF2. Wide-field (wf) deconvolved simulations of the corresponding 3D-SIM magnifications are shown for comparison (far right gallery). Note the elucidation of interconnected ovoid intensities within the telomere-cluster (mid-gallery, lower panel, and dCas9-eGFP). Bar, 5  $\mu$ m; magnifications, 500 nm.

## Materials and Methods

### Cell culture and transfection

J1 embryonic stem cells (ESCs)<sup>68</sup> were cultivated at 37 °C and 5% CO<sub>2</sub> on gelatin-coated petri dishes in Dulbecco's modified Eagle's medium supplemented with 16% fetal bovine serum (Biochrom), 0.1 mM  $\beta$ -mercaptoethanol (Invitrogen), 2 mM L-glutamine, 1x MEM non-essential amino acids, 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin (PAA Laboratories GmbH), 1000 U/ml recombinant mouse LIF (Millipore), 1  $\mu$ M PD032501, and 3  $\mu$ M CHIR99021 (Axon Medchem). Transfection of J1 cells was performed using Lipofectamin 2000 (Invitrogen) according to the manufacturer's instructions. For FISH experiments  $2.5 \times 10^5$ , for all other experiments  $5 \times 10^5$  cells were transfected. Cells were analyzed 48 h post-transfection.

### Plasmid generation

For generating the dCas9-eGFP construct, plasmid hCas9\_D10A was purchased from Addgene (addgene ID: 41816<sup>24</sup>). Inactivation of the second nuclease domain (H840A) was performed by site-directed mutagenesis using primers dCas1-F, dCas2-R, dCas2-F, and dCas3-R. The resulting PCR-product (dCas9) was digested with BsrGI and XbaI and ligated into pCAG.<sup>69,70</sup> The NLS-eGFP-sequence was amplified by PCR using primers eGFP1 and eGFP2, digested with AsiSI and NotI and ligated downstream of dCas9. pEX-A-U6-gRNA was synthesized at Eurofins MWG Operon according to Mali et al.<sup>24</sup>

gRNA-expression vectors were generated by amplifying pEX-A-U6-gRNA with forward and reverse primers, which introduced the protospacer sequence for minor satellites repeats (MiS), major satellites repeats (MaS) and telomeres (Tel), respectively.

Nucleotide sequences:

dCas1-F: 5'-aaagegatcg ctctagaatg gacaagaagt actccattgg g-3'  
dCas2-R: 5'-ctggggcagc atagatcca cgctg-3'  
dCas2-F: 5'-cgacgtggat gctatcgtgc cccag-3'  
dCas3-R: 5'-tttggcggc ctcattgtac aatcacctc ctctcttct tggggtc-3'  
eGFP1: 5'-aaagcgatcg catccaaaga agaagagaaa ggcatgggtg agcaagggcg agg-3'  
eGFP2: 5'-tttggcggc cttacttga cagctgtcc atgcc-3'  
MaSgRNA-F: 5'-ggcaagaaa ctgaaatca gtttagagc tagaaatagc aag-3'  
MaSgRNA-R: 5'-tgatttcag tttcttgc cggtgttgc tccttccac -3'  
TelgRNA-F: 5'-tagggtagg gtaggggta gtttagagc tagaaatagc aag-3'  
TelgRNA-R: 5'-taaccctaac cctaacccta cggtgttgc tccttccac -3'  
MiSgRNA-F: 5'-acactgaaaa cacattcgtg ttttagagc agaaatagc aag-3'  
MiSgRNA-R: 5'-acgaatgtg tttcagtgt cggtgttgc tccttccac -3'  
pEX-A-U6-gRNA: Sequence is available upon request.

### Immunofluorescence staining (IF) and fluorescent in situ hybridization (FISH)

Immunostaining and FISH experiments were performed as described previously.<sup>45,64</sup> Briefly, J1 ESCs grown on coverslips were washed 48 h post-transfection with phosphate buffered

saline (PBS) and fixed with 4% formaldehyde for 10 min. After permeabilization with 0.5% Triton X-100 in PBS, cells were incubated with PBST (PBS, 0.02% Tween) supplemented with 2% BSA and 0.5% fish skin gelatin (blocking buffer) for 1 h. Both primary and secondary antibodies were diluted in blocking buffer and cells were incubated with the respective antibodies for 1 h in a dark humidified chamber at room temperature. Washings were performed with PBST. For immuno-FISH detection, cells were first incubated with primary (anti-GFP, Roche) and secondary antibodies, followed by postfixation and pre-treatment for hybridization. Hybridization was performed overnight at 37 °C. Following post-hybridization washings were performed with 2xSaline Sodium Citrate (SSC) at 37 °C and 0.1xSSC at 61 °C.<sup>64</sup> FISH probes for MiS and MaS were generated by PCR using mouse Cot1 DNA (Invitrogen) as a template (MaS-primers: 5'-GCG AGA AAA CTG AAA ATC AC-3' and 5'-TCA AGT CGT CAA GTG GAT G-3'; MiS-primers: 5'-CAT GGA AAA TGA TAA AAA CC-3' and 5'-CAT CTA ATA TGT TCT ACA GTG TGG-3'). The probe for telomeric repeats was produced by using self-annealing primers in the PCR (5'-TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG TTA GGG-3' and 5'-CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA-3'). All probes were directly labeled by nick translation with Texas Red-dUTP or Cy3-dUTP and dissolved in hybridization mixture (50% formamide, 10% dextran sulfate, 1xSSC) at a concentration of 10–20 ng/ml. Cells were counterstained with 4',6-diamidino-2-phenylindole (DAPI), mounted with antifade medium (Vectashield, Vector Laboratories) and sealed with Covergrip sealant (Biotium). Primary antibodies used in this study were: anti-GFP (1:400; Roche 11814460001), anti-TRF2 (1:250, Abcam ab13579), anti-CENP-B (1:500, Abcam ab84489). The secondary antibodies were anti-rabbit conjugated to DyLight 594 (Jackson ImmunoResearch, 711-505-152), anti-mouse conjugated to Alexa 488 (Invitrogen, A21202) and anti-mouse

conjugated to Alexa 594 (Invitrogen, A11032). Cells for 3D-SIM were grown on precision cover glass, thickness no. 1.5H (170 µm ± 5 µm; Marienfeld Superior) using immersion oil with a refractive index of 1.514 to minimize spherical aberration.

#### Microscopy and image acquisition

Optical sections for **Figure 1D** were acquired with an UltraVIEW VoX spinning disk confocal microscope (PerkinElmer), which was operated with the Volocity® software. For **Figure 1E-F, 2, and 3A-B**, single optical sections or stacks of optical sections were collected with a Leica TCS SP5 confocal microscope using a Plan Apo 63×/1.4 NA oil immersion objective. High-resolution images (**Fig. 3C and 4**) were obtained with a DeltaVision OMX V3 3D-SIM microscope (Applied Precision Imaging, GE Healthcare), equipped with a 60×/1.42 NA PlanApo oil objective and sCMOS cameras (Olympus). Images were acquired with a z-step size of 125 nm. Reconstruction and image deconvolution was applied to the SI raw data using the SoftWorX 4.0 software package (Applied Precision). Image processing and assembly was performed with FIJI and Photoshop CS5.1 (Adobe), respectively. Stacks of confocal optical sections were corrected for chromatic shifts with ImageJ plugins.<sup>71</sup>

#### Disclosure of Potential Conflicts of Interest

No potential conflict of interest was disclosed.

#### Acknowledgments

The authors thank Katharina Thanisch (LMU Munich) for kindly providing the pEX-U6-gRNA cloning vector, George Church (Harvard Medical School) for the hCas9\_D10A construct and En Li (China Novartis Institutes for BioMedical Research) for the J1 cell line. We also thank Irina Solovei (LMU Munich) for advice with FISH protocols. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, SFB 1064, Nanosystems Initiative Munich, NIM) and T.A. is a fellow of the Graduiertenkolleg GRK1721.

#### References

- Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 2005; 3:e157; PMID:15839726; <http://dx.doi.org/10.1371/journal.pbio.0030157>
- Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 1999; 145:1119-31; PMID:10366586; <http://dx.doi.org/10.1083/jcb.145.6.1119>
- Cremer M, von Hase J, Volm T, Brero A, Kreth G, Walter J, Fischer C, Solovei I, Cremer C, Cremer T. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* 2001; 9:541-67; PMID:11721953; <http://dx.doi.org/10.1023/A:1012495201697>
- Martin RM, Leonhardt H, Cardoso MC. DNA labeling in living cells. *Cytometry Part A: the journal of the International Society for Analytical Cytology* 2005; 67:45-52.
- Kimura H, Cook PR. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol* 2001; 153:1341-53; PMID:11425866; <http://dx.doi.org/10.1083/jcb.153.7.1341>
- Robinett CC, Straight A, Li G, Wilhelm C, Sudlow G, Murray A, Belmont AS. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol* 1996; 135:1685-700; PMID:8991083; <http://dx.doi.org/10.1083/jcb.135.6.1685>
- Klug A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* 2010; 43:1-21; PMID:20478078; <http://dx.doi.org/10.1017/S0033583510000089>
- Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem* 2001; 70:313-40; PMID:11395410; <http://dx.doi.org/10.1146/annurev.biochem.70.1.313>
- Segal DJ, Barbas CF 3<sup>rd</sup>. Design of novel sequence-specific DNA-binding proteins. *Curr Opin Chem Biol* 2000; 4:34-9; PMID:10679372; [http://dx.doi.org/10.1016/S1367-5931\(99\)00048-4](http://dx.doi.org/10.1016/S1367-5931(99)00048-4)
- Segal DJ, Dreier B, Beerli RR, Barbas CF 3<sup>rd</sup>. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* 1999; 96:2758-63; PMID:10077584; <http://dx.doi.org/10.1073/pnas.96.6.2758>
- DeFrancesco L. Move over ZFNs. *Nat Biotechnol* 2011; 29:681-4; PMID:21822235; <http://dx.doi.org/10.1038/nbt.1935>
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 2011; 29:143-8; PMID:21179091; <http://dx.doi.org/10.1038/nbt.1755>
- Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 2011; 39:9283-93; PMID:21813459; <http://dx.doi.org/10.1093/nar/gkr597>
- Mahfouz MM, Li L, Piatek M, Fang X, Mansour H, Bangarusamy DK, Zhu JK. Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol* 2012; 78:311-21; PMID:22167390; <http://dx.doi.org/10.1007/s11103-011-9866-x>
- Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsasser J, Lahaye T, Leonhardt H. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* 2012; 40:5368-77; PMID:22387464; <http://dx.doi.org/10.1093/nar/gks199>

16. Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 2011; 29:149-53; PMID:21248753; <http://dx.doi.org/10.1038/nbt.1775>
17. Miyazari Y, Ziegler-Birling C, Torres-Padilla ME. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* 2013; 20:1321-4; PMID:24096363; <http://dx.doi.org/10.1038/nsmb.2680>
18. Ma H, Reyes-Gutierrez P, Pederson T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc Natl Acad Sci U S A* 2013; 110:21048-53; PMID:24324157; <http://dx.doi.org/10.1073/pnas.1319097110>
19. Thanisch K, Schneider K, Morbitzer R, Solovei I, Lahaye T, Bultmann S, Leonhardt H. Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic Acids Res* 2013; PMID:24371265; <http://dx.doi.org/10.1093/nar/gkr1348>
20. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanov AJ, Voytas DF. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 2011; 39:e82; PMID:21493687; <http://dx.doi.org/10.1093/nar/gkr218>
21. Morbitzer R, Elsaesser J, Hausner J, Lahaye T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res* 2011; 39:5790-9; PMID:21421566; <http://dx.doi.org/10.1093/nar/gkr151>
22. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J. The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* 2012; 46:311-39; PMID:23145983; <http://dx.doi.org/10.1146/annurev-genet-110711-155447>
23. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; 339:819-23; PMID:23287718; <http://dx.doi.org/10.1126/science.1231143>
24. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science* 2013; 339:823-6; PMID:23287722; <http://dx.doi.org/10.1126/science.1232033>
25. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 2013; 31:227-9; PMID:23360964; <http://dx.doi.org/10.1038/nbt.2501>
26. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013; 31:233-9; PMID:23360965; <http://dx.doi.org/10.1038/nbt.2508>
27. Shen B, Zhang J, Wu H, Wang J, Ma K, Li Z, Zhang X, Zhang P, Huang X. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* 2013; 23:720-3; PMID:23545779; <http://dx.doi.org/10.1038/cr.2013.46>
28. Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods* 2013; 10:741-3; PMID:23817069; <http://dx.doi.org/10.1038/nmeth.2532>
29. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 2013; 31:230-2; PMID:23360966; <http://dx.doi.org/10.1038/nbt.2507>
30. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivalila CS, Dadon DB, Jaenisch R. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res* 2013; 23:1163-71; PMID:23979020; <http://dx.doi.org/10.1038/cr.2013.122>
31. Kearns NA, Genga RM, Enuameh MS, Garber M, Wolfe SA, Maehr R. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* 2014; 141:219-23; PMID:24346702; <http://dx.doi.org/10.1242/dev.103341>
32. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013; 152:1173-83; PMID:23452860; <http://dx.doi.org/10.1016/j.cell.2013.02.022>
33. Kipling D, Ackford HE, Taylor BA, Cooke HJ. Mouse minor satellite DNA genetically maps to the centromere and is physically linked to the proximal telomere. *Genomics* 1991; 11:235-41; PMID:1685135; [http://dx.doi.org/10.1016/0888-7543\(91\)90128-2](http://dx.doi.org/10.1016/0888-7543(91)90128-2)
34. Kipling D, Wilson HE, Mitchell AR, Taylor BA, Cooke HJ. Mouse centromere mapping using oligonucleotide probes that detect variants of the minor satellite. *Chromosoma* 1994; 103:46-55; PMID:8013255; <http://dx.doi.org/10.1007/BF00364725>
35. Hörz W, Altrenburger W. Nucleotide sequence of mouse satellite DNA. *Nucleic Acids Res* 1981; 9:683-96; PMID:6261227; <http://dx.doi.org/10.1093/nar/9.3.683>
36. Kuznetsova IS, Prusov AN, Erukashvily NI, Podgornaya OI. New types of mouse centromeric satellite DNAs. *Chromosome Res* 2005; 13:9-25; PMID:15791408; <http://dx.doi.org/10.1007/s10577-005-2346-x>
37. Shay JW, Wright WE. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* 2005; 26:867-74; PMID:15471900; <http://dx.doi.org/10.1093/carcin/bgh296>
38. Pidoux AL, Allshire RC. The role of heterochromatin in centromere function. *Philos Trans R Soc Lond B Biol Sci* 2005; 360:569-79; PMID:15905142; <http://dx.doi.org/10.1098/rstb.2004.1611>
39. Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC. Requirement of heterochromatin for cohesion at centromeres. *Science* 2001; 294:2539-42; PMID:11598266; <http://dx.doi.org/10.1126/science.1064027>
40. Wong AK, Rattner JB. Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Res* 1988; 16:11645-61; PMID:3211746; <http://dx.doi.org/10.1093/nar/16.24.11645>
41. Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* 2004; 166:493-505; PMID:15302854; <http://dx.doi.org/10.1083/jcb.200403109>
42. Joseph A, Mitchell AR, Miller OJ. The organization of the mouse satellite DNA at centromeres. *Exp Cell Res* 1989; 183:494-500; PMID:2767161; [http://dx.doi.org/10.1016/0014-4827\(89\)90408-4](http://dx.doi.org/10.1016/0014-4827(89)90408-4)
43. Weierich C, Brero A, Stein S, von Hase J, Cremer C, Cremer T, Solovei I. Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome Res* 2003; 11:485-502; PMID:12971724; <http://dx.doi.org/10.1023/A:1025016828544>
44. Nagele RG, Velasco AQ, Anderson WJ, McMahon DJ, Thomson Z, Fazekas J, Wind K, Lee H. Telomere associations in interphase nuclei: possible role in maintenance of interphase chromosome topology. *J Cell Sci* 2001; 114:377-88; PMID:11148139
45. Markaki Y, Smeets D, Cremer M, Schermelleh L. Fluorescence in situ hybridization applications for super-resolution 3D structured illumination microscopy. *Methods Mol Biol* 2013; 950:43-64; PMID:23086869
46. Pidoux AL, Allshire RC. Centromeres: getting a grip of chromosomes. *Curr Opin Cell Biol* 2000; 12:308-19; PMID:10801468; [http://dx.doi.org/10.1016/S0955-0674\(00\)00094-6](http://dx.doi.org/10.1016/S0955-0674(00)00094-6)
47. Stephens AD, Snider CE, Haase J, Haggerty RA, Vasquez PA, Forest MG, Bloom K. Individual pericentromeres display coordinated motion and stretching in the yeast spindle. *J Cell Biol* 2013; 203:407-16; PMID:24189271; <http://dx.doi.org/10.1083/jcb.201307104>
48. Allshire RC, Karpen GH. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet* 2008; 9:923-37; PMID:19002142; <http://dx.doi.org/10.1038/nrg2466>
49. Gartenberg M. Heterochromatin and the cohesion of sister chromatids. *Chromosome Res* 2009; 17:229-38; PMID:19308703; <http://dx.doi.org/10.1007/s10577-008-9012-z>
50. De Lange T. Telomere-related genome instability in cancer. *Cold Spring Harb Symp Quant Biol* 2005; 70:197-204; PMID:16869754; <http://dx.doi.org/10.1101/sqb.2005.70.032>
51. Lovejoy CA, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, De S, Petrini JH, Sung PA, Jasin M, et al.; ALT Starr Cancer Consortium. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet* 2012; 8:e1002772; PMID:22829774; <http://dx.doi.org/10.1371/journal.pgen.1002772>
52. Blackburn EH. Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Lett* 2005; 579:859-62; PMID:15680963; <http://dx.doi.org/10.1016/j.febslet.2004.11.036>
53. Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. A human centromere antigen (CENP-B) interacts with a short specific sequence in aliphoid DNA, a human centromeric satellite. *J Cell Biol* 1989; 109:1963-73; PMID:2808515; <http://dx.doi.org/10.1083/jcb.109.5.1963>
54. Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T. Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol* 1992; 116:585-96; PMID:1730770; <http://dx.doi.org/10.1083/jcb.116.3.585>
55. Pluta AF, Saitoh N, Goldberg I, Earnshaw WC. Identification of a subdomain of CENP-B that is necessary and sufficient for localization to the human centromere. *J Cell Biol* 1992; 116:1081-93; PMID:1740467; <http://dx.doi.org/10.1083/jcb.116.5.1081>
56. Suzuki N, Nakano M, Nozaki N, Egashira S, Okazaki T, Masumoto H. CENP-B interacts with CENP-C domains containing Mif2 regions responsible for centromere localization. *J Biol Chem* 2004; 279:5934-46; PMID:14612452; <http://dx.doi.org/10.1074/jbc.M306477200>
57. Foltz DR, Jansen LE, Black BE, Bailey AO, Yates JR 3<sup>rd</sup>, Cleveland DW. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 2006; 8:458-69; PMID:16622419; <http://dx.doi.org/10.1038/ncb1397>
58. Kitagawa K, Masumoto H, Ikeda M, Okazaki T. Analysis of protein-DNA and protein-protein interactions of centromere protein B (CENP-B) and properties of the DNA-CENP-B complex in the cell cycle. *Mol Cell Biol* 1995; 15:1602-12; PMID:7862152
59. Okada T, Ohzeki J, Nakano M, Yoda K, Brinkley WR, Larionov V, Masumoto H. CENP-B controls centromere formation depending on the chromatin context. *Cell* 2007; 131:1287-300; PMID:18160038; <http://dx.doi.org/10.1016/j.cell.2007.10.045>

60. Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T. Mammalian telomeres end in a large duplex loop. *Cell* 1999; 97:503-14; PMID:10338214; [http://dx.doi.org/10.1016/S0092-8674\(00\)80760-6](http://dx.doi.org/10.1016/S0092-8674(00)80760-6)
61. Nikitina T, Woodcock CL. Closed chromatin loops at the ends of chromosomes. *J Cell Biol* 2004; 166:161-5; PMID:15249582; <http://dx.doi.org/10.1083/jcb.200403118>
62. Doksani Y, Wu JY, de Lange T, Zhuang X. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* 2013; 155:345-56; PMID:24120135; <http://dx.doi.org/10.1016/j.cell.2013.09.048>
63. Schermelleh L, Heintzmann R, Leonhardt H. A guide to super-resolution fluorescence microscopy. *J Cell Biol* 2010; 190:165-75; PMID:20643879; <http://dx.doi.org/10.1083/jcb.201002018>
64. Solovei I, Cremer M. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol Biol* 2010; 659:117-26; PMID:20809307; [http://dx.doi.org/10.1007/978-1-60761-789-1\\_8](http://dx.doi.org/10.1007/978-1-60761-789-1_8)
65. Chaumeil J, Micsinai M, Skok JA. Combined immunofluorescence and DNA FISH on 3D-preserved interphase nuclei to study changes in 3D nuclear organization. *J Vis Exp* 2013; pii:e50087; PMID:23407477
66. Molenaar C, Wiesmeijer K, Verwoerd NP, Khazen S, Eils R, Tanke HJ, Dirks RW. Visualizing telomere dynamics in living mammalian cells using PNA probes. *EMBO J* 2003; 22:6631-41; PMID:14657034; <http://dx.doi.org/10.1093/emboj/cdg633>
67. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 2013; 155:1479-91; PMID:24360272; <http://dx.doi.org/10.1016/j.cell.2013.12.001>
68. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; 69:915-26; PMID:1606615; [http://dx.doi.org/10.1016/0092-8674\(92\)90611-F](http://dx.doi.org/10.1016/0092-8674(92)90611-F)
69. Niwa H, Yamamura K, Miyazaki J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 1991; 108:193-9; PMID:1660837; [http://dx.doi.org/10.1016/0378-1119\(91\)90434-D](http://dx.doi.org/10.1016/0378-1119(91)90434-D)
70. Meilinger D, Fellingner K, Bultmann S, Rothbauer U, Bonapace IM, Klinkert WE, Spada F, Leonhardt H. Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 2009; 10:1259-64; PMID:19798101; <http://dx.doi.org/10.1038/embor.2009.201>
71. Ronneberger O, Baddeley D, Scheipl F, Verweir PJ, Burkhardt H, Cremer C, Fahrmeir L, Cremer T, Joffe B. Spatial quantitative analysis of fluorescently labeled nuclear structures: problems, methods, pitfalls. *Chromosome Res* 2008; 16:523-62; PMID:18461488; <http://dx.doi.org/10.1007/s10577-008-1236-4>



## 2.2 Visualization of genomic loci in living cells with a fluorescent CRISPR/Cas system



## Visualization of Genomic Loci in Living Cells with a Fluorescent CRISPR/Cas9 System

Tobias Anton, Heinrich Leonhardt, and Yolanda Markaki

### Abstract

The discovery that the RNA guided bacterial endonuclease Cas9 can be harnessed to target and manipulate user-defined genomic sequences has greatly influenced the field of genome engineering. Interestingly, a catalytically dead Cas9 (dCas9) can be employed as a targeted DNA-binding platform to alter gene expression. By fusing this dCas9 to eGFP, we and others could show that the CRISPR/Cas9 system can be further expanded to label and trace genomic loci in living cells. We demonstrated that by exchanging the sgRNA, dCas9-eGFP could be specifically directed to various heterochromatic sequences within the nucleus. Here, we provide a basic protocol for this versatile tool and describe how to verify new dCas9-eGFP targets.

**Key words** CRISPR/Cas9, sgRNA, In vivo labeling, Repetitive sequences

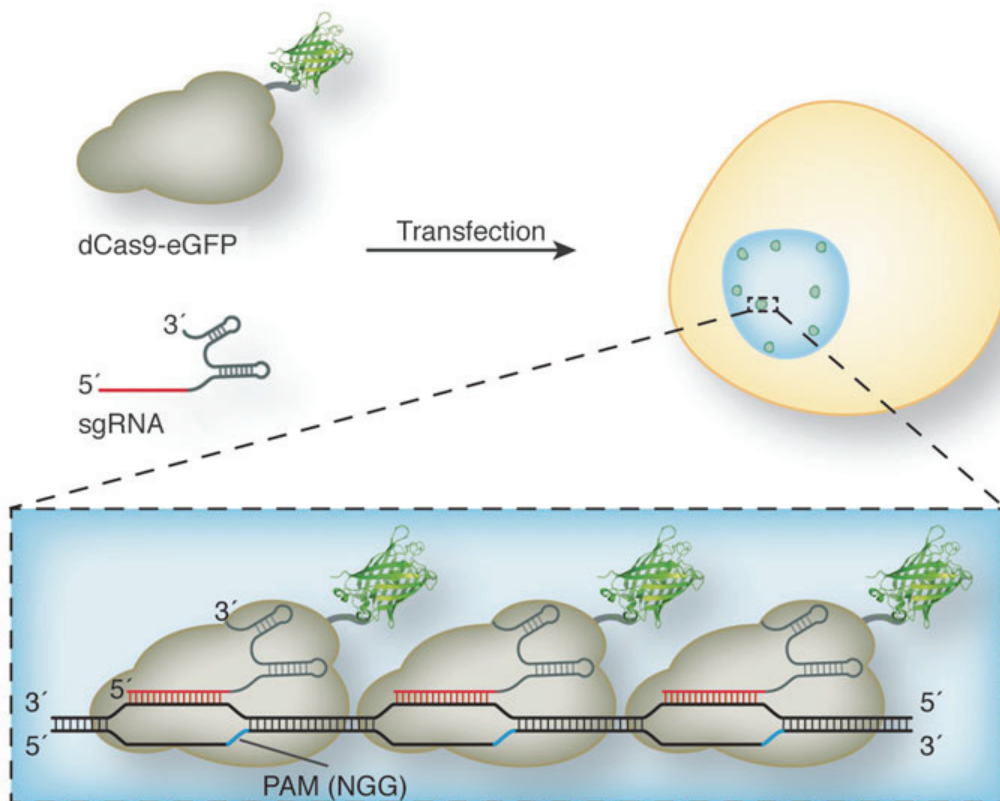
---

### 1 Introduction

To study nuclear architecture and the spatiotemporal organization of chromatin, it is essential to have tools that allow visualization of proteins as well as specific genomic loci. Whereas nuclear proteins can be readily imaged in living cells by fusing the target protein to a fluorescent tag, in vivo visualization of DNA sequences has been challenging. Fluorescent in situ hybridization (FISH) represents a very reliable and prominent tool to label specific loci. However, this method relies on fixed samples and thereby only represents a “snapshot” in cellular development. Moreover, the relatively harsh fixation and denaturation of DNA may potentially alter the native context of the targeted locus [1]. To overcome these limitations, fluorescently tagged proteins can be employed, which recognize DNA in a sequence specific manner. For example, it has been shown that individual Cys<sub>2</sub>His<sub>2</sub> zinc finger modules, each recognizing 3 bp, can be tandemly arranged to form a polydactyl zinc finger protein (PZF), which in turn recognizes a user-defined DNA sequence [2–4]. Yet testing PZFs for their target specificity

renders this method very time-consuming [5, 6]. Similar to PZFs, designer TAL (transcription activator-like) effectors (dTALEs) have been used to label distinct sequences in live cells [7–9]. Here, repeat variable diresidues (RVDs), located in the central repeat domain of the protein, confer target specificity [10]. However, due to the tandemly arrangement of these repeats, designing new dTALEs can be tedious and often requires multistep cloning approaches [11, 12].

In this chapter, we describe a novel method to label and trace endogenous repetitive sequences (Fig. 1) [13–15], which is based on the bacterial CRISPR/Cas9 adaptive immune system [16]. Whereas in dTALE or PZF based approaches the protein itself confers target specificity, in this case an inactive version of the endonuclease Cas9 (dCas9) is directed to its target by homologous base pairing between a single guide RNA (sgRNA) and the respective genomic sequence [17, 18]. The fact that the sgRNA, and with it the sequence to be targeted, is easily interchangeable, greatly enhances the practicality of this system by minimizing cloning efforts.



**Fig. 1** Schematic overview of the CRISPR labeling approach. dCas9-eGFP is guided to the desired locus by a co-transfected sgRNA. Note that target specificity is mediated by homologous base pairing between DNA (*black*) and RNA (*red*). Due to the repetitive organization of this locus, one specific sgRNA is sufficient to ensure bright signals

---

## 2 Materials

### 2.1 Cell Culture

1. Dulbecco's modified Eagle's medium (DMEM).
2. Fetal bovine serum (FBS).
3. 200 mM L-glutamine.
4. 100× penicillin–streptomycin (Pen–Strep).
5. Trypsin (0.25 % trypsin–EDTA solution).
6. Phosphate-buffered saline (PBS).
7. Blasticidin S (10 mg/mL) (optional).

### 2.2 Medium

C2C12 medium: 390 mL DMEM supplemented with 100 mL FBS (final concentration: 20 %), 5 mL L-glutamine (final concentration: 2 mM), and 5 mL Pen–Strep (final concentration: 1×).

### 2.3 Transfection (See Note 1)

1. Lipofectamine® 3000 reagent.
2. Opti-MEM serum-free medium.
3. CAG-dCas9-eGFP plasmid [14].
4. U6-MaSgRNA plasmid [14] (optional).
5. U6-MiSgRNA plasmid [14] (optional).
6. U6-TelgRNA plasmid [14] (optional).

### 2.4 Buffers and Solutions for Immuno-FISH

1. 20× PBS: 2.74 M NaCl, 53.7 mM KCl, 130 mM Na<sub>2</sub>HPO<sub>4</sub>, 35 mM KH<sub>2</sub>PO<sub>4</sub>. For working dilution (1×): dilute 1:20 in ddH<sub>2</sub>O.
2. Fixation solution: 4 % formaldehyde in 1× PBS (4 % FA).
3. Permeabilization solution: 0.5 % Triton X-100 in 1× PBS.
4. PBST: 0.02 % Tween 20 in 1× PBS.
5. Immunofluorescence blocking solution: 2 % bovine serum albumin (BSA), 0.5 % fish skin gelatin (FSG) in PBST.
6. 20 % glycerol solution in 1× PBS.
7. 0.1 N HCl.
8. 20× saline-sodium citrate (20× SSC) pH 7: 3 M NaCl, 0.3 M sodium citrate, pH 7.
9. 4× SSCT: dilute 20× SSC 1:5 in ddH<sub>2</sub>O and supplement with 0.02 % Tween 20.
10. 2× SSC: dilute 20× SSC 1:10 with ddH<sub>2</sub>O.
11. 0.1× SSC: dilute 20× SSC 1:200 with ddH<sub>2</sub>O.
12. 50 % formamide in 2× SSC: 50 mL 20× SSC, 250 mL formamide, 200 mL ddH<sub>2</sub>O. Adjust pH to 7 and store at –20 °C.
13. DAPI counterstaining solution: 2 µg/mL DAPI in 1× PBST.

**2.5 Microscopy**

While many imaging setups would work, a standard fluorescence microscope should suffice. Ideally, the microscope should be equipped with a CCD camera, a 63×/1.4 NA oil immersion objective and filters for eGFP excitation/emission. When CRISPR imaging is combined with Immuno-FISH or any other immunofluorescence detection method, appropriate filter sets should be installed. For live cell experiments the microscope should additionally be equipped with a live cell chamber, which guarantees stable conditions (37 °C, 5 % CO<sub>2</sub>) (*see Note 2*).

1. Image analysis software.

**2.6 Cloning**

1. U6-gRNA plasmid [14].
2. Restriction enzymes: DpnI, SacI, and NsiI.
3. dNTPs (2,5 mM each).
4. 100 % dimethyl sulfoxide (DMSO).
5. Phusion High Fidelity polymerase (*see Note 3*).
6. DNA purification kit.

**2.7 General  
Laboratory Equipment**

1. p100 cell culture plates.
2. Six-well cell culture plates.
3. 12-well cell culture plates.
4. FACS machine (optional).
5. PCR cycler.
6. Fine-tip forceps.
7. Borosilicate coverslips (# 1.5).
8. Glass microscope slides.
9. Dewar container for liquid N<sub>2</sub> transport.
10. Heating block with flat surface for microscope slides.
11. Water bath.
12. Humidified chamber.
13. Heated incubation chamber.
14. Shaking platform.
15. Floating tin box.
16. Soft tissue paper.
17. Rubber cement.
18. Non-hardening antifade mounting medium (e.g., Vectashield).
19. Transparent nail polish.
20. Parafilm.

### 3 Methods

#### 3.1 Cloning

##### 3.1.1 Primer Design

1. Choose appropriate dCas9-eGFP target sequence (N<sub>17-28</sub>NGG) (see **Notes 4** and **5**; Fig. 2).

2. Order HPSF purified DNA oligonucleotides (for comparison, see **Note 6**; Fig. 2):

Forward: 5'-N<sub>17-28</sub> GTTTTAGAGCTAGAAATAGCAAG-3'

Reverse: 5'-N<sub>17-28</sub>(rev.comp.) CGGTGTTCGTCCTTCCAC-3'.

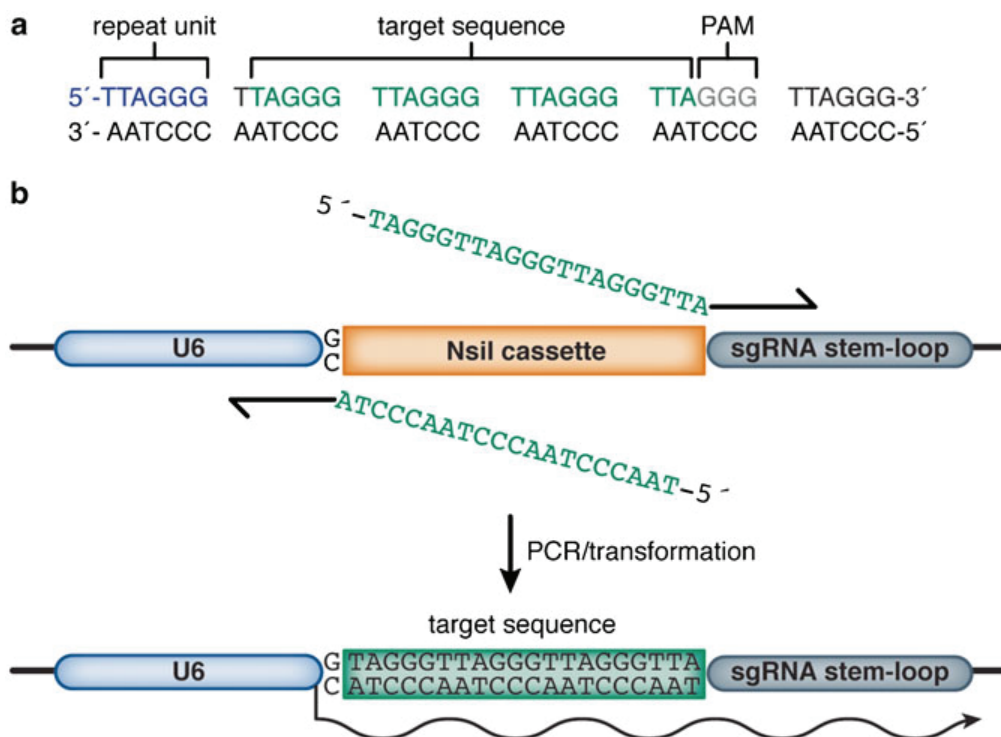
##### 3.1.2 PCR and Transformation

1. Set up PCR mix for sgRNA plasmid with new target sequence, as shown in Table 1:

2. Run rolling cycle PCR with the following program: 1 cycle of 98 °C for 2 min; 30 cycles of: 98 °C for 15 s, 62 °C for 15 s and 72° for 1.5 min; 1 cycle of 72 °C for 2 min; 4 °C hold.

3. Purify PCR product to remove DMSO.

4. Digest up to 2 µg of purified PCR product with *DpnI* for at least 1 h at 37 °C. This step ensures the removal of any residual PCR template, which would lead to a smaller number of positive clones after transformation.



**Fig. 2** Overview of gRNA design. **(a)** Example of a repetitive genomic locus (telomeres). Indicated are the basic repeat unit (*blue*), the target sequence (*green*), and the PAM (*gray*, NGG). **(b)** Schematic representation of the U6-gRNA plasmid. Primers are designed in a way that the target sequence is included in non-annealing regions. After PCR and transformation, the *NsiI* cassette is exchanged by the target sequence

**Table 1**  
**Set up of PCR mix**

Component	Volume in $\mu\text{L}$ (final conc.)
H <sub>2</sub> O	32.5
5 $\times$ buffer	10 (1 $\times$ )
DMSO	2.5 (5 %)
Forward primer	1 (0.2 $\mu\text{M}$ )
Reverse primer	1 (0.2 $\mu\text{M}$ )
U6-gRNA plasmid	1 (1 ng)
dNTPs	1 (50 $\mu\text{M}$ )
Polymerase	1 (1 $\mu$ )

5. Heat-inactivate restriction enzyme according to manufacturer's instructions.
6. Use up to 5  $\mu\text{L}$  of heat-inactivated reaction to transform 50  $\mu\text{L}$  chemically competent *E. coli* culture. Refer to standard transformation protocols and plate bacteria on LB agar containing ampicillin (100  $\mu\text{g}/\text{mL}$ ). Incubate the bacteria over night at 37  $^{\circ}\text{C}$ .
7. Pick 3–12 colonies and inoculate 2 mL LB-medium supplemented with 100  $\mu\text{g}/\text{mL}$  ampicillin. Incubate cultures over night at 37  $^{\circ}\text{C}$  while shaking at 200 rpm.
8. Perform plasmid preparation according to standard protocols and digest plasmid DNA with *SaeI* and *NsiI*. Run digested DNA on a 1 % agarose gel (*see Note 7*).

### 3.2 Cell Culture

1. Culture C2C12 cells at 37  $^{\circ}\text{C}$  and 5 % CO<sub>2</sub> on p100 plates in 10 mL C2C12 medium.
2. Split cells every 2–3 days, typically in ratio of 1:6 to 1:8. For this, aspirate medium and wash cells with 10 mL PBS. Add 1 mL of trypsin solution and incubate for 5 min at 37  $^{\circ}\text{C}$ . After trypsinization, resuspend cells in fresh medium and transfer an appropriate amount to a new culture dish. Add medium to a final volume of 10 mL.

### 3.3 Transfection

In general, any transfection reagent may be used. However, we recommend testing the transfection efficiency beforehand. According to our experience, cells grown to a confluency of ~60–80 % are most suitable for transfection and result in > 20 % transfected cells.



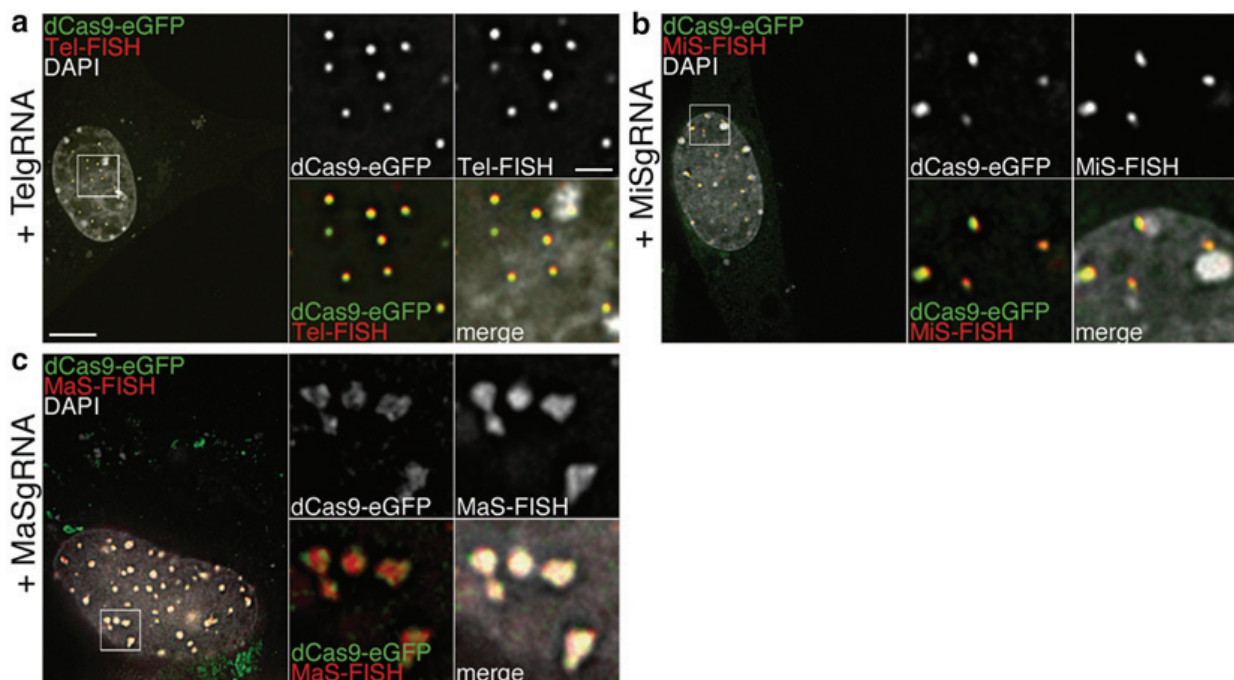
### 3.3.1 Transient Transfection

1. Seed cells 1 day prior to transfection in a six-well plate. For fixed-cell experiments, seed the cells on coverslips.
2. Co-transfect cells according to manufacturer's instructions with the CAG-dCas9-eGFP plasmid and a U6-sgRNA plasmid, which encodes the desired target sequence.
3. Collect images 24–48 h post-transfection.

### 3.3.2 Stably dCas9-eGFP Expressing Cells (Optional)

As the dCas9-eGFP target is easily exchanged by switching the sgRNA, it might be preferable to establish a cell line, which stably expresses dCas9-eGFP (Fig. 3).

1. Seed cells 1 day prior to transfection in a p100 plate.
2. Transfect cells according to manufacturer's instructions with the CAG-dCas9-eGFP plasmid.
3. After 24 h, supplement medium with 10  $\mu\text{g}/\text{mL}$  blasticidin S.
4. FACS sort GFP-positive cells 2–3 weeks after transfection. Time of sorting depends on recovery of blasticidin resistant cells, after an initial wave of cell death.
5. Culture sorted cells in C2C12 medium supplemented with 5  $\mu\text{g}/\text{mL}$  blasticidin S.



**Fig. 3** Example data set for verification of CRISPR labeling via Immuno-FISH. C2C12 cells, stably expressing dCas9-eGFP (green) were transfected with (a) TelgRNA, (b) MiSgRNA or (c) MaSgRNA. Cells were FISH-treated with the corresponding probes (red) and counterstained with DAPI (gray). Galleries show 4 $\times$  magnifications of boxed areas. Scale bar, 5  $\mu\text{m}$ ; scale bar in magnifications, 1  $\mu\text{m}$

### 3.4 Immuno-FISH

When dCas9-eGFP targeting has to be confirmed by FISH, we recommend seeding and transfect the cells in a 12-well cell culture plate on coverslips (Fig. 3). The cells should be fixed 24–48 h after transfection. For FISH-probe generation, *see* refs. [19, 20]. A modified Immuno-FISH protocol is presented below [1]. Unless otherwise noted, all washing steps are performed with 1 mL of buffer/solution.

1. Wash cells three times with 1× PBS.
2. Fix cells with 4 % FA for 10 min at room temperature (RT).
3. Stepwise exchange fixative with PBST: add an equal amount of PBST directly to the fixative and gradually remove the solution from the well. Do not remove all the solution as fixed samples are prone to drying out, which may lead to a deformed cell morphology. Repeat this step three times, until the fixative has been completely exchanged with PBST.
4. Permeabilize cells with permeabilization solution for 15 min at RT.
5. Incubate coverslip for at least 1 h in 20 % glycerol solution at RT.
6. Snap-freeze coverslip in liquid nitrogen. Dip the coverslip into a Styrofoam-box filled with liquid N<sub>2</sub> using the fine-tip forceps. Place the coverslip (cell-side up) for a few seconds on a soft tissue paper to defreeze. Return the coverslip to the well containing the 20 % glycerol solution. Repeat this step three times.
7. Wash three times with PBST.
8. Incubate coverslips in IF blocking buffer for 1 h at RT to block unspecific binding sites.
9. Incubate the coverslip in a dark humidified chamber with primary antibody (anti-GFP) for 1 h at RT. For this, firmly attach Parafilm on a smooth surface (e.g., lid of a six-well cell culture plate), place a drop (~50 µL) of diluted antibody onto the Parafilm and carefully place the coverslip (cell-side facing downwards) onto the drop.
10. Wash four times with PBST.
11. Incubate the coverslip in humidified chamber with appropriate secondary antibody (e.g., coupled to Alexa 488) as described in **step 9**.
12. Wash four times with PBST.
13. Post-fix the sample with fixation solution for 10 min at RT.
14. Wash twice with 1× PBS.
15. Denature samples with 0.1 N HCl for 5 min at RT.
16. Wash twice with 1× PBS.

17. Wash three times with 2× SSC for 3 min.
18. Incubate coverslips for 1 h at 37 °C with pre-warmed 50 % formamide in 2× SSC.
19. Denature FISH-probe at 94 °C for 3 min and place probe on ice.
20. Place a drop of FISH-probe on a clean microscope slide and mount coverslip.
21. Seal coverslip with rubber cement.
22. Denature the slide for 2 min at 76 °C.
23. Transfer the slide to a floating tin box and hybridize the probe over night at 37 °C in a water bath.
24. Carefully unmount coverslip by rehydrating with 2× SSC dropped around the edges of the coverslip.
25. Wash three times with 2× SSC for 15 min at RT.
26. Wash with preheated 0.1× SSC (60 °C). Repeat this step three times.
27. Wash three times with 4× SSCT.
28. Wash twice with PBST.
29. Counterstain nuclei with 2 µg/mL DAPI for 5–8 min at RT.
30. Wash two times with PBST
31. Mount coverslips on a drop of Vectashield (~10 µL) placed on a clean microscope slide.
32. Seal with see-through nail polish.

---

## 4 Notes

1. sgRNAs targeting MaS, MiS, and Tel may be used as positive controls to test labeling and transfection efficiency. For designing new sgRNAs, *see* Subheading 3.1.
2. Turn live cell chamber on at least 4 h prior to imaging session. This is done to ensure that CO<sub>2</sub> flux and temperature are stable.
3. We recommend using a High Fidelity polymerase to minimize the risk of incorrect incorporated nucleotides.
4. Preferentially, the genomic target sequence is highly repetitive (e.g., telomeric repeats or LINEs) to ensure bright signals. However, it has been shown that ~30 consecutive repeats are sufficient to visualize a genomic locus [13].
5. Although it is necessary that a transcript driven by the U6 promoter starts with a G, this base does not have to be part of the target sequence, as it is already included in the reverse primer sequence.

6. Forward and reverse primer for TelgRNA. The telomere target sequence is underlined. The part of the primer which anneals to the U6-gRNA PCR template is written in *italic* letters (Fig. 2).  
 Forward: 5'- TAGGGTTAGGGTTAGGGTTA *GTTTTAGAGCTAGAAATAGCAAG* -3'  
 Reverse: 5'- TAACCCTAACCCCTAACCCCTA *CGGTGTTTCGTCCTTCCAC* -3'
7. Since the *Nsi*I cassette of the U6-gRNA plasmid is exchanged by the target sequence, positive clones show a linearized band at ~2.9 kb. In contrast, negative clones show two bands at ~2.6 kb and ~0.35 kb.

---

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, SFB 1064 and Nanosystems Initiative Munich, NIM), and T.A. thankfully acknowledges the Graduiertenkolleg GRK1721.

## References

1. Markaki Y, Smeets D, Cremer M, Schermelleh L (2013) Fluorescence in situ hybridization applications for super-resolution 3D structured illumination microscopy. *Methods Mol Biol* 950:43–64
2. Klug A (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* 43:1–21
3. Pabo CO, Peisach E, Grant RA (2001) Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem* 70:313–340
4. Segal DJ, Barbas CF 3rd (2000) Design of novel sequence-specific DNA-binding proteins. *Curr Opin Chem Biol* 4:34–39
5. DeFrancesco L (2011) Move over ZFNs. *Nat Biotech* 29:681–684
6. Segal DJ, Dreier B, Beerli RR, Barbas CF 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* 96:2758–2763
7. Miyanari Y, Ziegler-Birling C, Torres-Padilla ME (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* 20:1321–1324
8. Ma H, Reyes-Gutierrez P, Pederson T (2013) Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc Natl Acad Sci U S A* 110:21048–21053
9. Thanisch K, Schneider K, Morbitzer R, Solovei I, Lahaye T, Bultmann S, Leonhardt H (2014) Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic Acids Res* 42, e38
10. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509–1512
11. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39, e82
12. Morbitzer R, Elsaesser J, Hausner J, Lahaye T (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res* 39:5790–5799
13. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS et al (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155:1479–1491
14. Anton T, Bultmann S, Leonhardt H, Markaki Y (2014) Visualization of specific DNA

- sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 5:163–172
15. Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* 112:3002–3007
  16. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* 46:311–339
  17. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823
  18. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* 339:823–826
  19. Cremer M, Grasser F, Lanctot C, Muller S, Neusser M, Zinner R, Solovei I, Cremer T (2008) Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol* 463:205–239
  20. Byron M, Hall LL, Lawrence JB (2013) A multifaceted FISH approach to study endogenous RNAs and DNAs in native nuclear and cell structures. *Curr Protoc Hum Genet* Unit 4:15



### 2.3 Site-specific recruitment of epigenetic factors with a modular CRISPR/Cas system





**Site-specific recruitment of epigenetic factors with a modular CRISPR/Cas system**

Tobias Anton<sup>1</sup> and Sebastian Bultmann<sup>1\*</sup>

<sup>1</sup>Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), LMU Munich, 82152 Martinsried, Germany

\*To whom correspondence should be addressed. Tel: +49 (0)89 2180 74233; FAX +49 (0)89 2180-74236; Email: bultmann@bio.lmu.de

**Conflict of interest:**

None declared.

**Financial disclosure:**

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant SPP1623 and by the GRK1721.

**Keywords:**

CRISPR/Cas, DNA *de novo* methylation, DNA demethylation, GBP, Nanobody, GFP

**Abbreviations:**

CC: Chromocenter; 5mC: 5-methylcytosine; 5hmC: 5-hydroxymethylcytosine; TET1: ten-eleven translocation methylcytosine dioxygenase 1, DNMT3A: DNA methyltransferase 3A; GBP: GFP-binding protein

## Abstract

Dissecting the complex network of epigenetic modifications requires tools that combine precise recognition of DNA sequences with the capability to modify epigenetic marks. The CRISPR/Cas system has been proven to be a valuable addition to existing methodologies that fulfill these tasks. So far, sequence-specific editing of epigenetic modifications such as DNA methylation and histone posttranslational modifications relied on direct fusions of enzymatically inactivated Cas9 (dCas9) with epigenetic effectors. Here, we report a novel, modular system that facilitates the recruitment of any GFP-tagged protein to desired genomic loci. By fusing dCas9 to a GFP-binding nanobody (GBP) we demonstrate that prevalent epigenetic modifications at mouse major satellite repeats can be erased or set *de novo* by recruiting GFP-coupled catalytic domains of TET1 and DNMT3A, respectively. Furthermore, we construct an inducible expression system that enables a temporally controlled expression of both GBP-dCas9 and the effector protein. Thus, our approach further expands the CRISPR/Cas toolbox for site-specific manipulation of epigenetic modifications with a modular and easy-to-use system.

## Introduction

Eukaryotic gene expression is controlled by a complex network of epigenetic mechanisms that include the posttranslational modification of histones as well as covalent DNA modifications<sup>1</sup>. Dissection of this network using knockout or overexpression studies have greatly advanced our understanding of how epigenetic modifications contribute to transcriptional regulation. However, using traditional techniques, the complex relationships and feedback circuits that interconnect epigenetic pathways make it difficult to differentiate direct consequences of epigenetic modifications on transcription from secondary effects. Site-specific manipulation of epigenetic marks therefore represents a highly desirable tool to study and understand their direct functional relevance on gene expression and genome organization.

Previously, tools that direct the enzymatic activity of epigenetic effectors to specific loci were based on zinc finger proteins (ZFPs)<sup>2-6</sup> or transcription activator-like effectors (TALEs)<sup>6-8</sup>. However, custom design and engineering of ZFPs and TALEs is based on the rearrangement of their modular DNA-binding domains, requiring elaborate cloning techniques and rigorous testing<sup>9-12</sup>. In contrast, the RNA-guided endonuclease Cas9 of the type II CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated) system recognizes specific loci via Watson-Crick base pairing between a readily exchangeable 20 bp sequence of the single guide RNA (sgRNA) and the target DNA in the direct vicinity of a PAM (protospacer adjacent motif)<sup>13-15</sup>. Due to this ease of use, Cas9-based approaches have

been rapidly adopted for genome engineering strategies in a wide variety of cell types and organisms<sup>16-20</sup>.

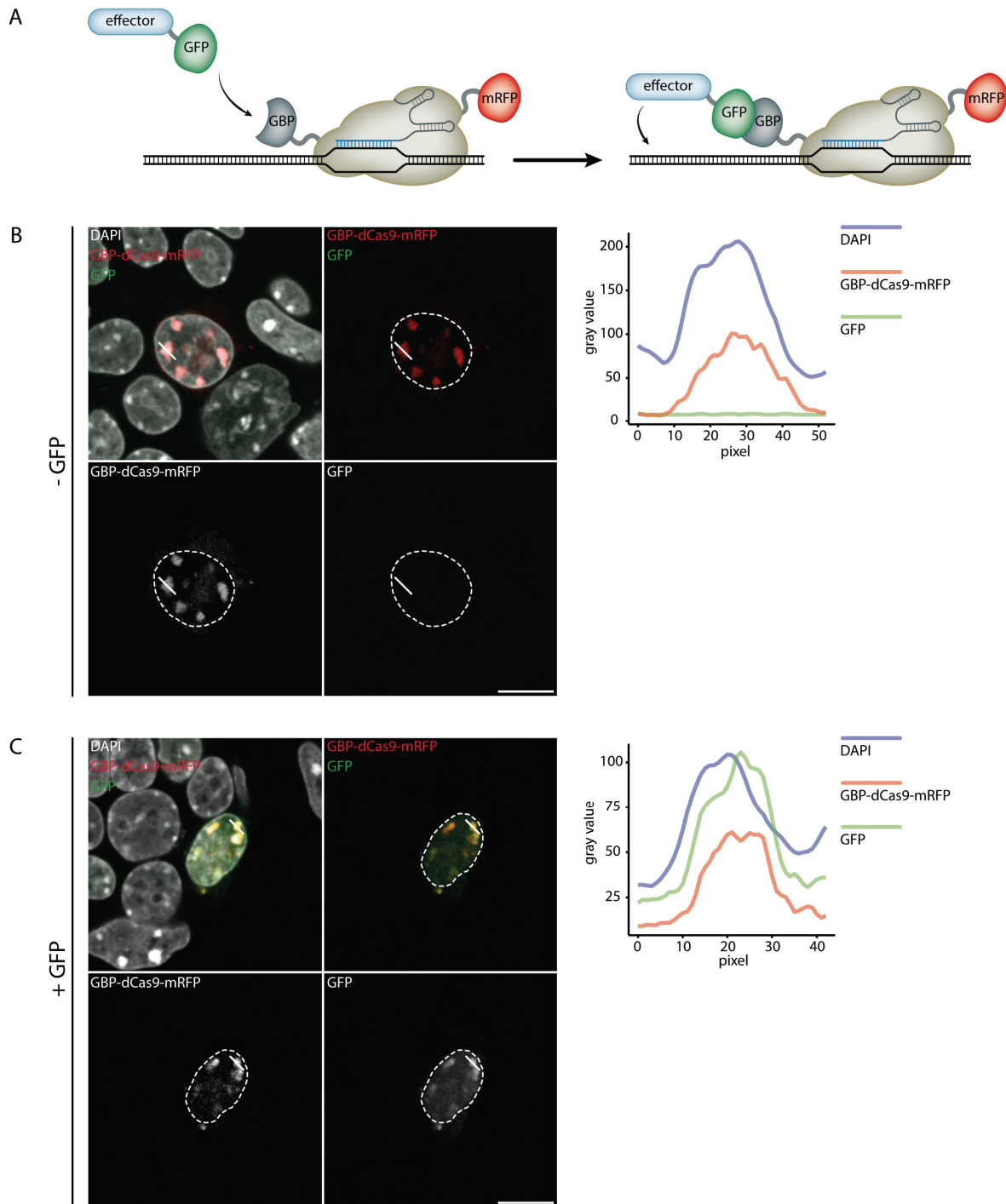
Importantly, engineering of a catalytically inactive variant of Cas9 (dCas9) facilitates RNA-guided genome targeting in a sequence specific manner, without cleaving the underlying DNA<sup>13,21</sup>. We and others have previously demonstrated that this programmable DNA-binding platform can be harnessed for *in vivo* visualization of specific genomic regions<sup>22-24</sup> as well as for determination of local chromatin composition<sup>25-26</sup>. In addition, fusion of dCas9 with the methyltransferase DNMT3A enables targeted transcriptional repression by catalyzing *de novo* methylation at gene regulatory regions<sup>27-28</sup>. *Vice versa*, dCas9 mediated targeting of ten-eleven translocation methylcytosine dioxygenase 1 (TET1) to regulatory elements results in up-regulation of silenced genes via active DNA demethylation<sup>29-31</sup>. Similarly, transcriptional modulation has also been reported by targeting histone acetylase<sup>32</sup> and histone demethylase activities<sup>33</sup> via dCas9, highlighting the versatility of this approach.

Here, we introduce a modular CRISPR/Cas9 system, which combines the sequence specificity of dCas9 with stringent recruitment of GFP-coupled epigenetic effectors via a GFP-binding nanobody (GBP)<sup>34</sup>. We show that this versatile setup can be exploited to control the levels of DNA modifications at target loci using GFP fusions of DNA methyltransferases and methylcytosine dioxygenases. Furthermore, using a bidirectional doxycycline-inducible promoter we develop a single vector system that allows the timed expression and targeted recruitment of GBP-Cas9 and GFP-fusion proteins.

## Results and Discussion

We first set out to assess, whether GFP can be recruited to a defined genomic locus via dCas9. To this end, we constructed a GBP-dCas9-mRFP construct, which enabled us to simultaneously visualize dCas9 and GFP localization (Figure 1A). Since heterochromatic chromocenters (CCs) are distinct subnuclear regions, which can be readily distinguished and are characterized by well-defined epigenetic marks<sup>35-36</sup>, we decided to tether GBP-dCas9-mRFP to these loci via a major satellite specific sgRNA (MaSgRNA).

Transient co-transfection of mouse embryonic stem cells ESCs (wt J1) with GBP-dCas9-mRFP and MaSgRNA resulted in a specific enrichment of GBP-dCas9-mRFP at CCs. Importantly, when we additionally transfected a GFP-encoding plasmid, we observed co-localization of GFP with GBP-dCas9-mRFP at CCs (Figure 1B and C). This experiment confirms the functionality of the GBP-dCas9-mRFP construct in facilitating the recruitment of GFP to target loci.

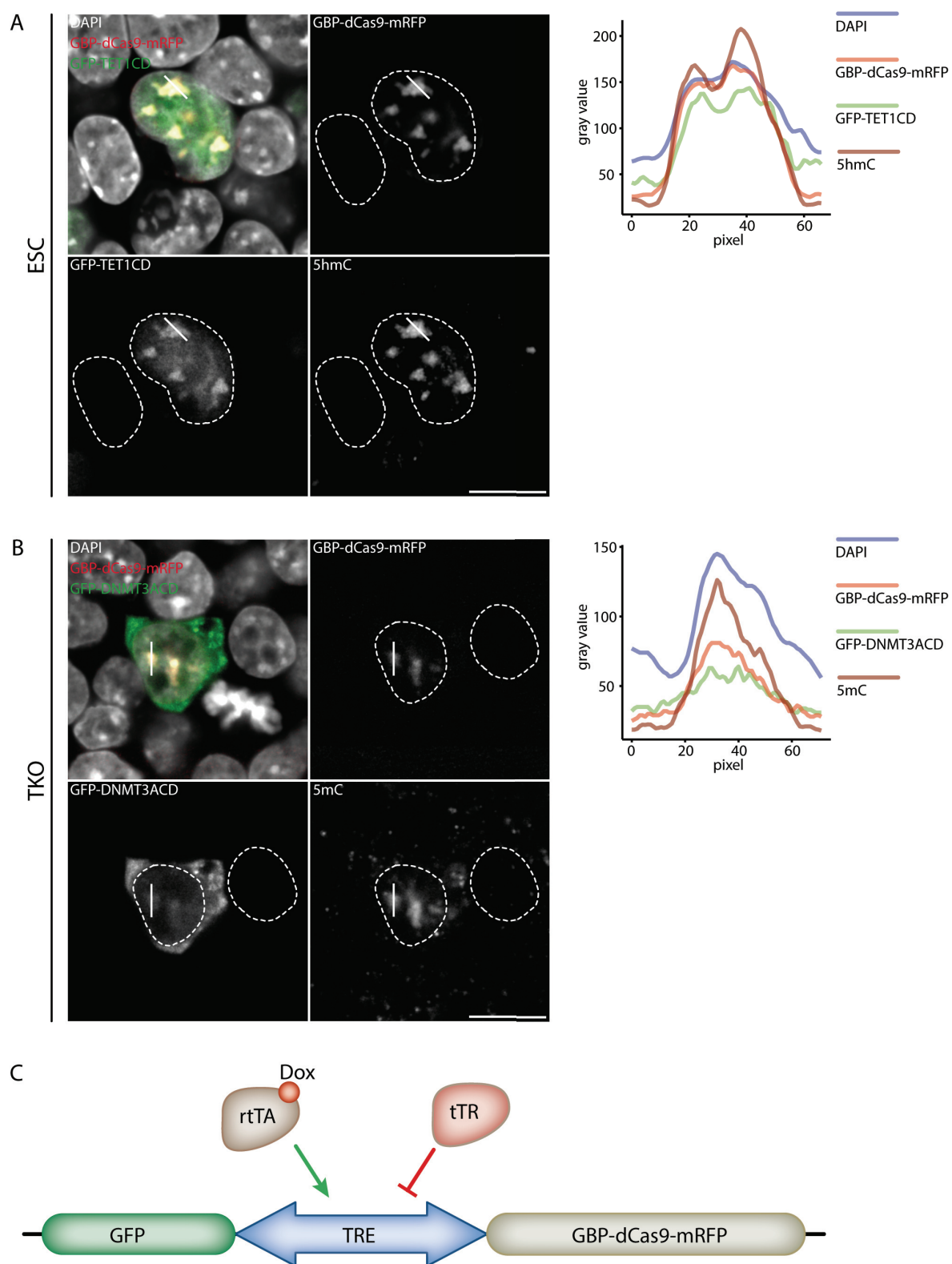


**Figure 1: Targeted recruitment of GFP to major satellites.** A) Schematic outline of dCas9-mediated effector recruitment. GBP-dCas9-mRFP is guided to a desired locus by a sgRNA and interacts with a GFP-coupled epigenetic effector via GBP. Subsequently, the effector modifies the underlying DNA. B-C) Representative confocal images of ESCs, co-transfected with GBP-dCas9-mRFP and major satellite specific sgRNA. GBP-dCas9-mRFP specifically localizes at CCs and recruits GFP, when it is additionally co-transfected (C). Line plots represent the signal intensity of the different channels along the indicated chromocenter (solid white line). White dashed lines indicate the nuclear border. Scale bar: 10  $\mu$ m.

Next, we aimed to target GFP-tagged epigenetic effector proteins to chromocenters via GBP-dCas9-mRFP. To test the feasibility of such an approach we used the catalytic domains of the methylcytosine dioxygenase TET1 and the *de novo* methyltransferase DNMT3A coupled to 68

GFP (GFP-TET1CD and GFP-DNMT3ACD, respectively). DNMT3A catalyzes the methylation of cytosine generating 5-methylcytosine (5mC), a repressive epigenetic mark enriched at CCs<sup>36</sup>. In contrast, TET1 oxidizes 5mC to 5-hydroxymethylcytosine (5hmC), a DNA modification generally found in euchromatin and depleted at the heterochromatic CCs<sup>37</sup>.

Similar to GFP alone, GFP-TET1CD was successfully recruited to CCs in cells, which co-expressed GBP-dCas9-mRFP and MaSgRNA. Notably, TET1CD recruitment to the highly methylated CCs in wt ESCs, resulted in an ectopic enrichment of 5hmC at these sites (Figure 2A). To test the feasibility of GFP-DNMT3ACD recruitment to CCs we used DNMT triple knock-out (TKO) cells<sup>38</sup>, which are virtually devoid of genomic DNA methylation. In TKO cells, which co-expressed GBP-dCas9-mRFP and MaSgRNA GFP-DNMT3ACD was successfully recruited to CCs, leading to a dramatic increase of 5mC at CCs (Figure 2B). Taken together, these data show that GBP-dCas9-mRFP is capable of directing the enzymatic activity of epigenetic factors to targeted genomic regions.

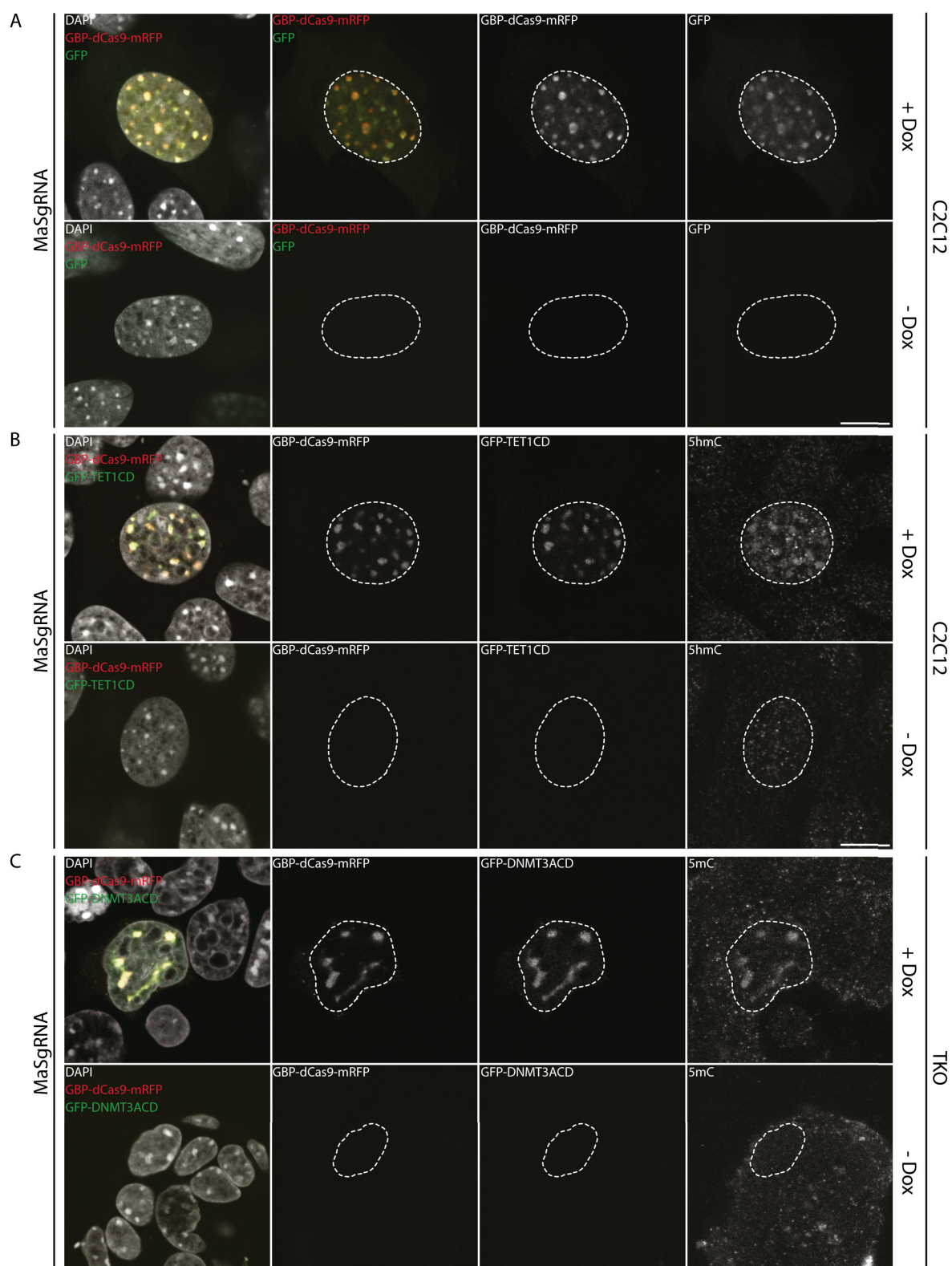


**Figure 2: Targeted recruitment of GFP-tagged effector proteins.** A) GBP-dCas9-mRFP recruits GFP-TET1CD to chromocenters. Recruited GFP-TET1CD oxidizes 5mC to 5hmC at CCs in transfected ESCs. In untransfected cells, no 5hmC signal was detected. B) When targeted to CCs, GFP-DNMT3ACD mediates *de novo* DNA methylation in TKO cells, which was not observed in untransfected control cells. Line plots represent the signal intensity of the different channels along the indicated chromocenter (solid white line). White dashed lines indicate the nuclear border. Scale bar: 10  $\mu$ m. C) Schematic representation of the inducible vector system. A bi-directional promoter drives the expression of GBP-dCas9-mRFP as well as GFP. The vector additionally encodes a transcriptional repressor (tTR) and a transcriptional activator (rtTA). In the

---

absence of doxycycline (Dox), tTR binds to a tetracycline response element (TRE) within the promoter sequence and represses transcription. Upon addition of Dox to the culture medium, rtTA replaces tTR and induces gene transcription.

Besides targeted recruitment, control over the timing of site-specific epigenetic editing is crucial for the dissection of direct functional consequences resulting from local epigenetic perturbations. To this end, we constructed an inducible system for GBP-dCas9-mRFP and GFP/GFP-effector expression. We used an Epstein-Barr virus (EBV) derived episomal expression plasmid (pRTS), which harbors a doxycycline-inducible bi-directional promoter and additionally encodes a tet-transcriptional (tTR) repressor as well as a tet-transcriptional activator (rtTA)<sup>39-40</sup>. Conditional transcriptional activation is achieved in the presence of doxycycline (Dox) by the rtTA, whereas tTR mediates active repression in Dox-free conditions. This single vector system allows tight control over the timing and expression level of GBP-dCas9-mRFP as well as a GFP-fusion protein (pRTS-GBP-dCas9-mRFP; Figure 2C). Moreover, using a bidirectional promoter ensures that both proteins are expressed at comparable levels reducing the amount of freely diffusing GFP-fusion protein, thereby minimizing off-target effects. We first tested the inducible system in mouse myoblast (C2C12) using a pRTS-GBP-dCas9-mRFP, additionally harboring either GFP alone or GFP-TET1CD. Both GFP and GFP-TET1CD were expressed and specifically recruited to CCs in a strictly sgRNA- and Dox-dependent manner (Figure 3A and B; Figure S1A; Figure S2A and B). Similar to co-transfections performed in wt ESCs, GFP-TET1CD recruitment resulted in a drastic increase of hmC at myoblast CCs (Figure 3B). Comparably, transfection of pRTS-GBP-dCas9-mRFP harboring GFP-DNMT3ACD into TKO cells led to a specific, Dox-dependent enrichment of 5mC at CCs (Figure 3C; Figure S1B; Figure S2C).



**Figure 3: Doxycycline-dependent, coordinated expression of GBP-dCas9-mRFP and GFP-tagged effectors.** Cells were co-transfected with MaSgRNA and pRTS-GBP-dCas9-mRFP, additionally encoding GFP (A), GFP-TET1CD (B) or GFP-DNMT3ACD (C). Upon induction with doxycycline both GBP-dCas9-mRFP and the corresponding GFP-tagged effector are expressed, resulting in the oxidation of 5mC (5hmC; B) or *de novo* methylation of CCs (5mC; C). White dashed lines indicate the nuclear border. Scale bar: 10  $\mu$ m.



In summary, we demonstrated that our modular system can be used to edit prevalent epigenetic marks such as DNA methylation at heterochromatic chromocenters in a timely controlled manner. While we used repetitive target sequences, which allow a fast and simple readout of efficient effector recruitment by microscopy, it is in principle possible to apply our system for the epigenetic modification of single target loci. Using multiple gene/target specific sgRNAs in parallel will also enable the modification of multiple loci at once or the visualization of a single copy locus<sup>22</sup>. The bidirectional, inducible system offers the added advantage to titrate of the amount of epigenetic modifier and Cas9 protein simultaneously, greatly reducing the possibility of off-target effects. Furthermore, our GBP-based approach for dCas9-assisted targeted recruitment can be combined with virtually any GFP-tagged protein. Considering the widespread use of GFP-fusions in cell lines and animal models, this system will help to facilitate the systematic dissection of biological processes in basic and biomedical research.

## **Material and methods**

### *Cell culture and transient transfections*

J1 ESCs and TKO cells were cultivated at 37 °C and 5% CO<sub>2</sub> on gelatin-coated dishes in Dulbecco's modified Eagle's medium (DMEM, Sigma), supplemented with 16 % fetal bovine serum (FBS, Biochrom), 0.1 mM β-mercaptoethanol (Invitrogen), 2 mM L-glutamine, 1x MEM non-essential amino acids, 100 U/ml penicillin, 100 µg/ml streptomycin (PAA Laboratories GmbH), 1000 U/ml recombinant mouse LIF (Millipore), 1 µM PD032501 and 3 µM CHIR99021 (Axon Medchem). C2C12 cells were cultured at 37 °C and 5 % CO<sub>2</sub> in DMEM, supplemented with 20 % FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. For conditional transcription activation, the culture medium was additionally supplemented with 1.5 µg/ml doxycycline for 24 hours. Transient transfections were performed using Lipofectamine<sup>®</sup> 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions and cells were analyzed 24 – 48 hrs post-transfection.

### *Plasmid generation*

For generating the GBP-dCas9-mRFP expression plasmid, the GBP and mRFP coding sequences were amplified from pGFPbinderImR<sup>41</sup> and cloned into pCAG-dCas9<sup>23</sup> via XbaI and AsiSI/NotI, respectively. The expression constructs for GFP-TET1CD and MaSgRNA were described previously<sup>23,42</sup>. GFP-DNMT3ACD was constructed by amplifying the catalytic domain of DNMT3A from pCAG-GMT3a<sup>43</sup>. Subsequently TET1CD was exchanged by DNMT3ACD using AsiSI and NotI restriction enzymes. The doxycycline-inducible

expression system is based on the pRTS plasmid described previously<sup>39</sup>. The GBP-dCas9-mRFP coding sequence was cloned into pRTS via SfiI, whereas TET1CD and DNMT3ACD, respectively, were inserted upstream of GFP via SmaI.

Constructs generated for this study are available via Addgene.

#### *Immunofluorescence staining and microscopy*

Immunofluorescence staining was performed as described previously<sup>44</sup>. Briefly, cells were grown on coverslips (thickness 1.5H, 170  $\mu\text{m} \pm 5 \mu\text{m}$ ; Marienfeld Superior) and transfected with the respective expression plasmids. 24 – 48 hours after transfection, cells were washed with phosphate buffered saline (PBS), fixed with 3.7 % formaldehyde for 10 min and permeabilized with 0.5 % Triton X-100 in PBS. For 5mC and 5hmC detection, DNA was first denatured with 1 N HCl for 15 min and then neutralized with 150 mM TRIS-HCl (pH 8.5). Subsequently, cells were transferred into blocking buffer (0.02 % Tween and 2 % bovine serum albumin in PBS) for 1 hour. Both primary and secondary antibodies were diluted in blocking buffer and cells were incubated in a dark, humidified chamber for 1 hour at room temperature. Nuclei were counterstained with 0.2  $\mu\text{g}/\text{ml}$  DAPI in PBS or SiR-DNA (Spirochrome). Coverslips were mounted with antifade medium (Vectashield, Vector Laboratories) and sealed with colorless nail polish. Primary antibodies used in this study were: GFP- and RFP-booster conjugated to Atto 488 and Atto 593, respectively (1:200; Chromotek), anti-5hmC (1:250; Active Motif) and anti-5mC (1:500, Diagenode). The secondary antibodies were: anti-rabbit IgG conjugated to Alexa 647, anti-mouse IgG conjugated to Alexa 647 (1:400; Thermo Fisher Scientific) and anti-mouse IgG conjugated to Alexa 405 (1:400; Invitrogen).

Confocal images were acquired with a Leica TCS SP5 microscope equipped with a Plan Apo 63x/1.4 NA oil immersion objective. Image processing and assembly of figures was performed with FIJI and Photoshop CS5.1 (Adobe), respectively.

#### **Acknowledgements**

First and foremost, we would like to thank Heinrich Leonhardt (LMU Munich) for helpful discussions and advice on the manuscript. We further would like to thank Dirk Eick (Helmholtz Zentrum München) for providing the pRTS vector, Masaki Okano (RIKEN) for providing TKO cells and Kerry Tucker (Ruprecht-Karls-University, Heidelberg) for providing wt J1 ESCs. We also thank Sebastian Kaluscha, Tsothe Chitiashvili and Orsi Wade for help with cloning the pRTS constructs. T.A. thankfully acknowledges the Graduiertenkolleg GRK1721.

---

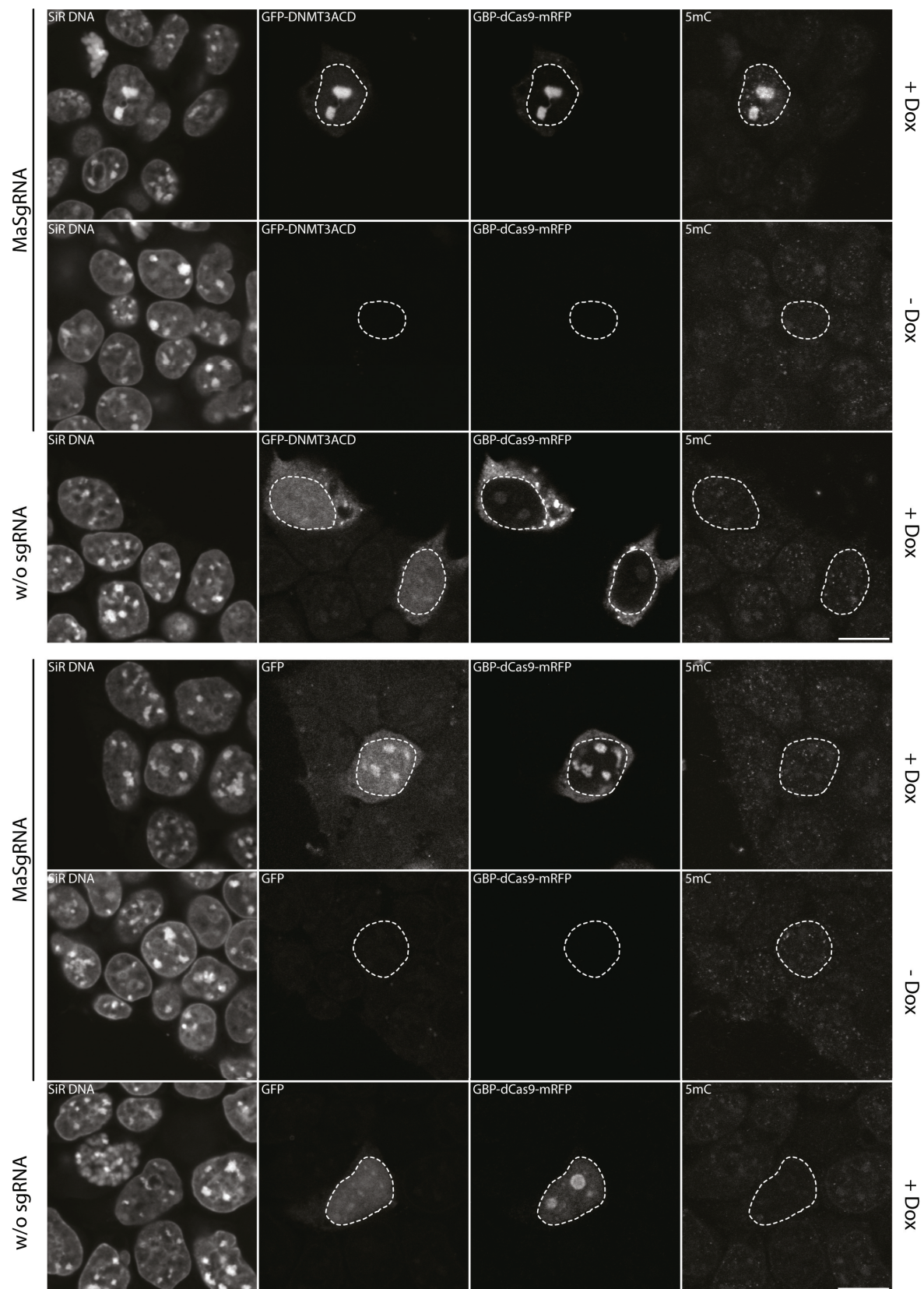
## References

1. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* 2016; 17:487-500.
2. Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD, et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* 2012; 7:350-60.
3. Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, et al. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. *J Mol Biol* 2013; 425:479-91.
4. Falahi F, Huisman C, Kazemier HG, van der Vlies P, Kok K, Hospers GA, et al. Towards sustained silencing of HER2/neu in cancer by epigenetic editing. *Mol Cancer Res* 2013; 11:1029-39.
5. Chen H, Kazemier HG, de Groote ML, Ruiters MH, Xu GL, Rots MG. Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. *Nucleic Acids Res* 2014; 42:1563-74.
6. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* 2013; 31:1137-42.
7. Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, et al. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* 2013; 31:1133-6.
8. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 2013; 500:472-6.
9. DeFrancesco L. Move over ZFNs. *Nat Biotech* 2011; 29:681-4.
10. Segal DJ, Dreier B, Beerli RR, Barbas CF, 3rd. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* 1999; 96:2758-63.
11. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 2011; 39:e82.
12. Morbitzer R, Elsaesser J, Hausner J, Lahaye T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res* 2011; 39:5790-9.
13. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; 337:816-21.
14. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; 339:819-23.
15. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. *Science* 2013; 339:823-6.
16. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 2013; 31:227-9.
17. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013; 31:233-9.

18. Shen B, Zhang J, Wu H, Wang J, Ma K, Li Z, et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* 2013; 23:720-3.
19. Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods* 2013; 10:741-3.
20. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 2013; 31:230-2.
21. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013; 152:1173-83.
22. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 2013; 155:1479-91.
23. Anton T, Bultmann S, Leonhardt H, Markaki Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* 2014; 5:163-72.
24. Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc Natl Acad Sci U S A* 2015; 112:3002-7.
25. Schmidtman E, Anton T, Rombaut P, Herzog F, Leonhardt H. Determination of local chromatin composition by CasID. *Nucleus* 2016; 7:476-84.
26. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun* 2013; 439:132-6.
27. McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, et al. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open* 2016; 5:866-74.
28. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* 2016; 44:5615-28.
29. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, et al. Editing DNA Methylation in the Mammalian Genome. *Cell* 2016; 167:233-47 e17.
30. Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, et al. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov* 2016; 2:16009.
31. Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* 2016; 7:46545-56.
32. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 2015; 33:510-7.
33. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* 2015; 12:401-3.
34. Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H. A versatile nanotrapp for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* 2008; 7:282-9.

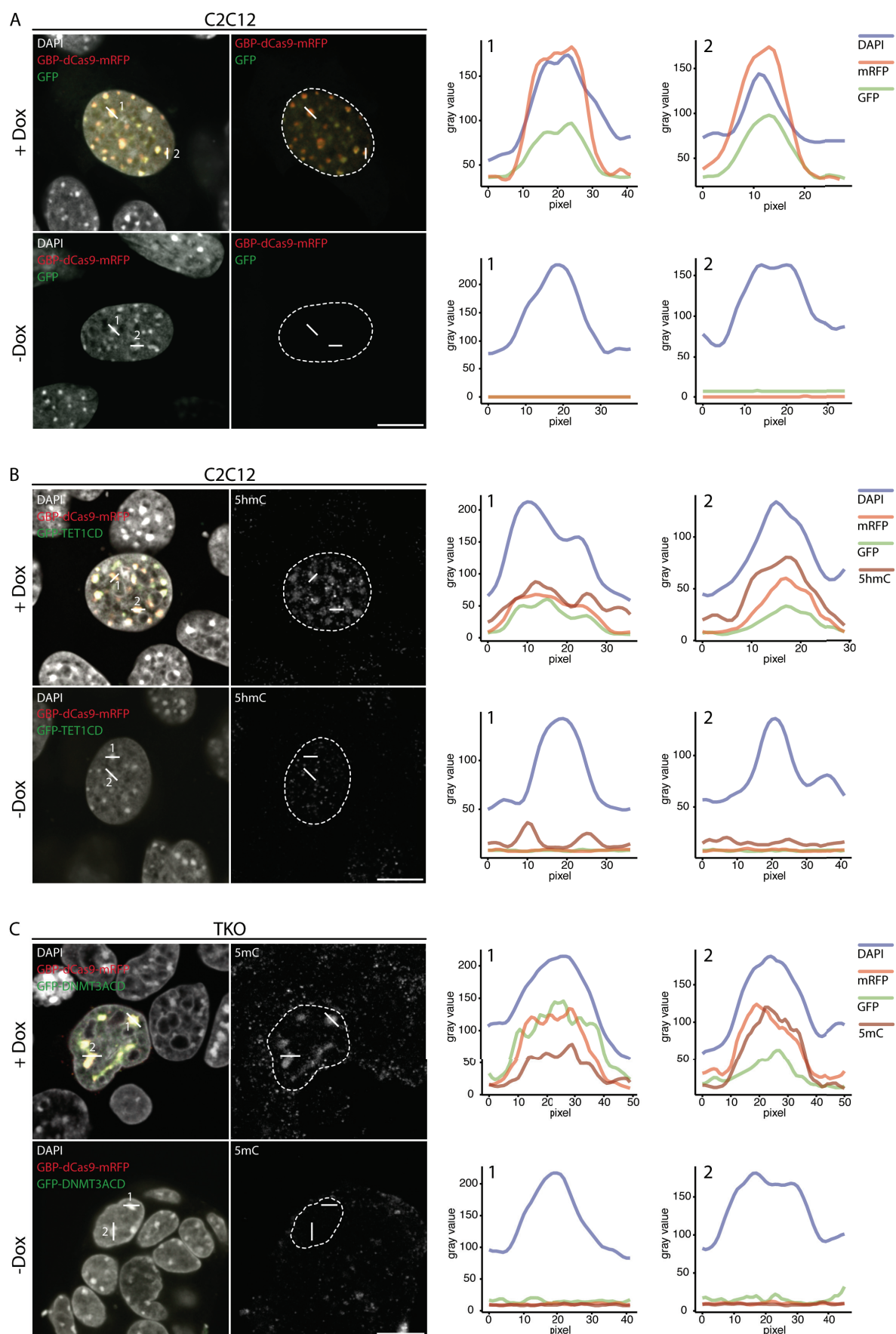
- 
35. Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* 2004; 166:493-505.
36. Gilbert N, Thomson I, Boyle S, Allan J, Ramsahoye B, Bickmore WA. DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J Cell Biol* 2007; 177:401-11.
37. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* 2011; 42:451-64.
38. Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 2006; 11:805-14.
39. Bornkamm GW, Berens C, Kuklik-Roos C, Bechet JM, Laux G, Bachl J, et al. Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res* 2005; 33:e137.
40. Hölzel M, Rohrmoser M, Orban M, Homig C, Harasim T, Malamoussi A, et al. Rapid conditional knock-down-knock-in system for mammalian cells. *Nucleic Acids Res* 2007; 35:e17.
41. Zolghadr K, Mortusewicz O, Rothbauer U, Kleinhans R, Goehler H, Wanker EE, et al. A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. *Mol Cell Proteomics* 2008; 7:2279-87.
42. Müller U, Bauer C, Siegl M, Rottach A, Leonhardt H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* 2014; 42:8592-604.
43. Meilinger D, Fellinger K, Bultmann S, Rothbauer U, Bonapace IM, Klinkert WE, et al. Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 2009; 10:1259-64.
44. Solovei I, Cremer M. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol Biol* 2010; 659:117-26.

Supplementary information



**Figure S1: Site-specific recruitment of GFP-tagged proteins is Dox- and sgRNA-dependent.** TKO cells were transfected with pRTS-GBP-dCas9-mRFP harboring either GFP (A) or GFP-DNMT3A (B). Where indicated, MaSgRNA was co-transfected and culture medium was supplemented with Dox. While Dox is required for gene expression, lack of a

sequence-specific sgRNA results in disperse nuclear and cytoplasmic localization of GBP-dCas9-mRFP and GFP/GFP-DNMT3ACD. White dashed lines represent nuclear border. Scale bar: 10  $\mu$ m.



**Figure S2: Signal intensities of GBP-dCas9-mRFP and effector proteins at chromocenters.** A-C) Quantification of signal intensities at CCs for images shown in Figure 3. In each cell, two CCs were analyzed and line plots represent signal



intensity along the indicated CCs (white solid lines). Expression of GBP-dCas9-mRFP and effector proteins was only observed after induction with Dox. White dashed lines represent nuclear border. Scale bar: 10  $\mu\text{m}$ .



## 2.4 Determination of local chromatin composition by CasID



## Determination of local chromatin composition by CasID

Elisabeth Schmidtmann<sup>a,†</sup>, Tobias Anton<sup>a,†</sup>, Pascaline Rombaut<sup>b</sup>, Franz Herzog<sup>b</sup>, and Heinrich Leonhardt<sup>a</sup> 

<sup>a</sup>Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), LMU Munich, Martinsried, Germany; <sup>b</sup>Gene Center and Department of Biochemistry, LMU Munich, Munich, Germany

### ABSTRACT

Chromatin structure and function are determined by a plethora of proteins whose genome-wide distribution is typically assessed by immunoprecipitation (ChIP). Here, we developed a novel tool to investigate the local chromatin environment at specific DNA sequences. We combined the programmable DNA binding of dCas9 with the promiscuous biotin ligase BirA\* (CasID) to biotinylate proteins in the direct vicinity of specific loci. Subsequent streptavidin-mediated precipitation and mass spectrometry identified both known and previously unknown chromatin factors associated with repetitive telomeric, major satellite and minor satellite DNA. With super-resolution microscopy, we confirmed the localization of the putative transcription factor ZNF512 at chromocenters. The versatility of CasID facilitates the systematic elucidation of functional protein complexes and locus-specific chromatin composition.

### ARTICLE HISTORY

Received 17 August 2016  
Revised 12 September 2016  
Accepted 13 September 2016

### KEYWORDS

biotinylation; CRISPR/Cas; CasID; chromatin composition; repetitive elements

### Introduction



Regulation of gene expression involves a yet undetermined number of nuclear proteins ranging from tightly bound histones to loosely attached or transiently interacting factors that directly and indirectly bind DNA sequences along the genome. Establishment, maintenance and alteration of functional DNA states during development and disease requires dynamic changes in local enrichment and posttranslational modification of chromatin proteins. The genome-wide distribution of a given protein is traditionally determined by chromatin immunoprecipitation (ChIP) and subsequent sequencing of co-precipitated DNA fragments. However, ChIP experiments rely on the availability of suitable antibodies and provide data on global antigen distribution rather than local chromatin composition.

Previously described strategies to directly analyze chromatin complexes such as HyCCaPP (Hybridization Capture of Chromatin Associated Proteins for Proteomics)<sup>1</sup> and PICh (Proteomic Isolation of Chromatin fragments)<sup>2</sup> were based on chemical crosslinking and precipitation with complementary DNA


probes. Alternatively, DNA binding proteins were used for chromatin precipitation and subsequent analysis by mass spectrometry.<sup>3–5</sup>

For visualization and manipulation, specific genomic loci can be targeted by different recombinant DNA binding proteins such as engineered polydactyl zinc finger proteins (PZFs),<sup>6</sup> designer transcription activator-like effectors (dTALEs)<sup>7,8</sup> or an enzymatically dead Cas9 (dCas9).<sup>9–11</sup> Whereas target specificity of PZFs and dTALEs is determined by their amino acid sequence, DNA binding of dCas9 is programmed by an easily exchangeable single guide RNA (sgRNA).<sup>12</sup>

Here, we exploited the RNA-programmable DNA binding of dCas9 to direct a biotin ligase to specific genomic sites and mark adjacent chromatin proteins for subsequent identification by mass spectrometry. Proximity-dependent biotin identification (BioID) employs a promiscuous biotin ligase (BirA\*) fused to a target protein for biotinylation of proteins within a 10 nm range.<sup>13,14</sup> Biotinylated proteins can then be identified by robust streptavidin-mediated capture

**CONTACT** Heinrich Leonhardt  [h.leonhardt@lmu.de](mailto:h.leonhardt@lmu.de)  Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), LMU Munich, Martinsried, Germany.

<sup>†</sup>The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

 Supplemental data for this article can be accessed on the publisher's website.

© 2016 Elisabeth Schmidtmann, Tobias Anton, Pascaline Rombaut, Franz Herzog, and Heinrich Leonhardt. Published with license by Taylor & Francis.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

and subsequent mass spectrometry. Based on BirA\* and dCas9 we developed a hybrid approach (CasID) to elucidate chromatin composition at specific DNA sequences.

## Results and discussion

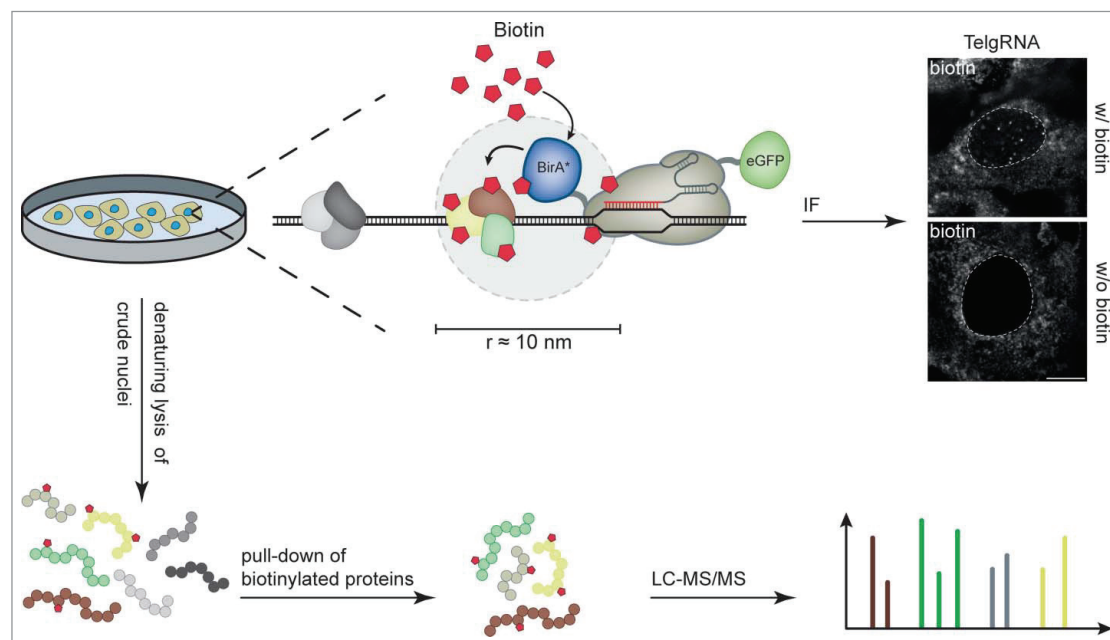
### Immunofluorescence microscopy reveals protein biotinylation at targeted loci

To evaluate whether the CasID approach is suited to biotinylate proteins at specific genomic loci we constructed a BirA\*-dCas9-eGFP fusion (Fig. 1). We co-transfected C2C12 myoblasts with this BirA\*-dCas9-eGFP construct and a sgRNA plasmid, targeting dCas9 to either telomeres, major or minor satellite sequences. We previously showed that all sgRNAs used in this study successfully target dCas9-eGFP to the desired loci.<sup>10</sup> Although here dCas9 is tagged on both N- (BirA\*) and C-terminus (eGFP), we observed specific recruitment to the designated sequences (Supplementary Fig. 1). In control cells without sgRNA expression, BirA\*-dCas9-eGFP shows a diffuse localization throughout the cell and a nucleolar enrichment (Supplementary Fig. 1). Importantly, in the presence

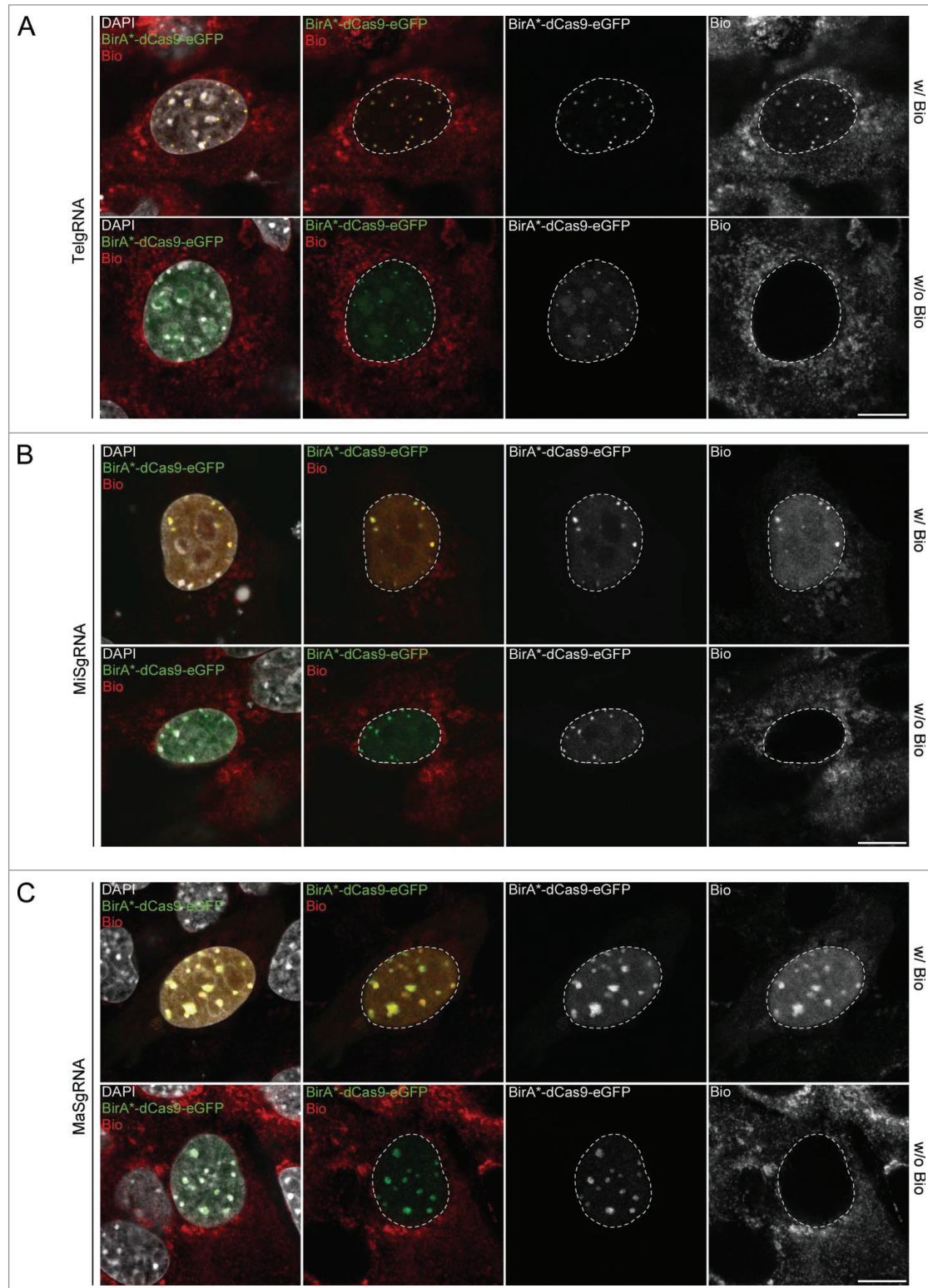
of functional sgRNAs, BirA\*-dCas9-eGFP was targeted to the respective loci and co-localized with a strong biotin signal, when the growth medium was supplemented with exogenous biotin (Fig. 2). These results demonstrate that the promiscuous biotin ligase BirA\* can be directed to endogenous loci via dCas9.

### Determination of local chromatin composition at distinct genomic loci by mass spectrometry

To identify proteins associated with distinct genomic regions, cells stably expressing BirA\*-dCas9-eGFP targeted to either telomeric regions, minor satellite repeats or major satellite repeats were supplemented with 50  $\mu$ M biotin for 24 h, representing standard BioID conditions.<sup>13</sup> We enriched for biotinylated proteins from crude nuclear extract with streptavidin-coated magnetic beads and analyzed them via tandem mass spectrometry (LC-MS/MS, Fig. 1). With label free quantification, we compared protein levels in pull-downs from cells expressing both BirA\*-dCas9-eGFP and a sgRNA with control samples of cells stably expressing untargeted BirA\*-dCas9-eGFP (without any sgRNA). Common BioID contaminants,<sup>15</sup> like



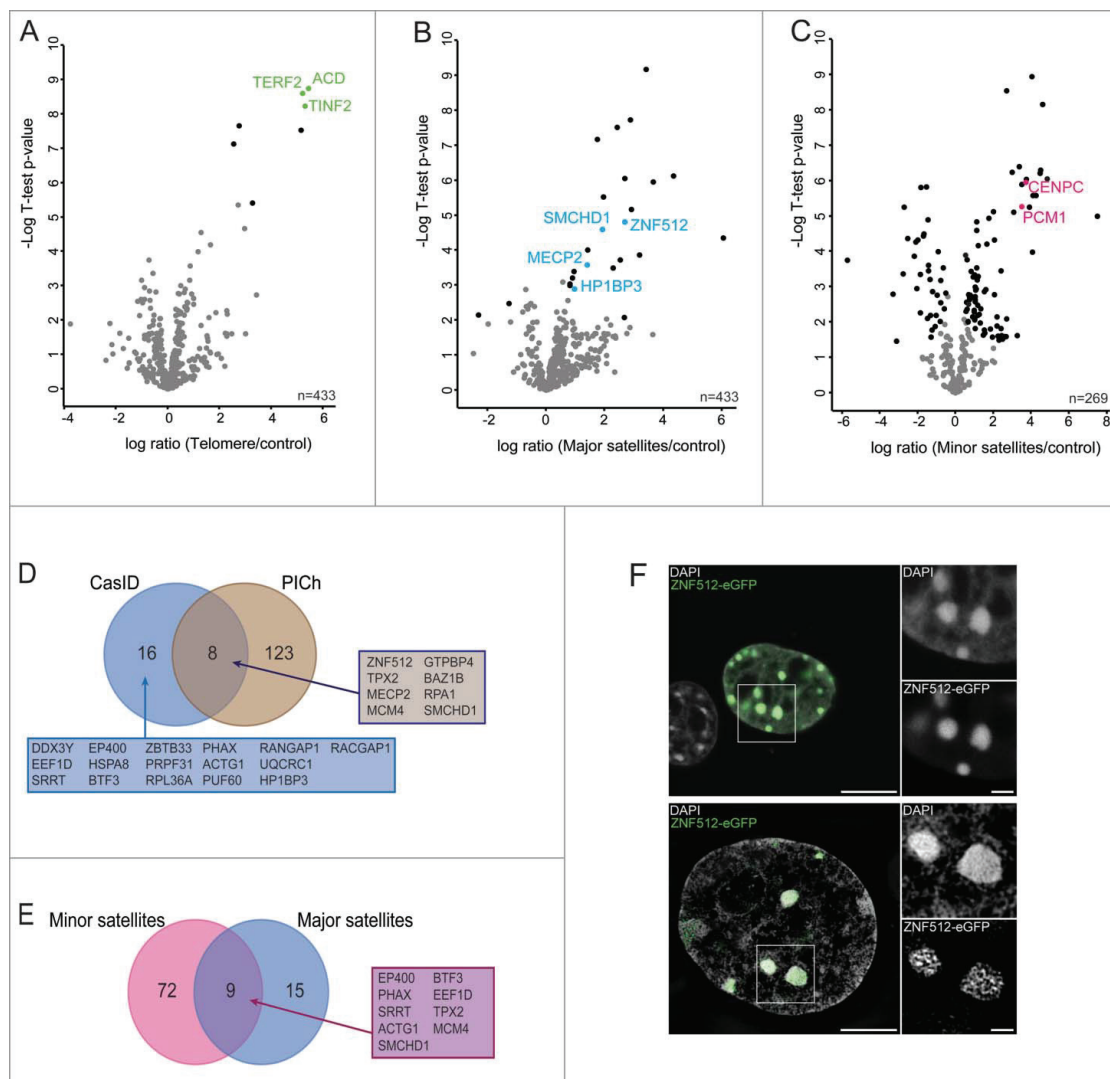
**Figure 1.** Workflow for CasID. BirA\*-dCas9-eGFP/sgRNA expressing cells are cultured in growth medium without exogenous biotin. The BirA\*-dCas9-eGFP fusion is directed to the desired target by sequence complementarity between sgRNA and the genomic locus. Upon addition of biotin to the medium, BirA\* ligates biotin to lysine residues of proteins in close proximity. Successful biotinylation of locus-associated proteins can directly be visualized via immunofluorescence microscopy. For mass-spectrometric analysis, cells are harvested, followed by isolation of crude nuclei. After a denaturing lysis, biotinylated proteins can be pulled from the lysate with streptavidin and subjected to mass spectrometry. White dashed lines indicate the border between nucleus and cytoplasm. Scale bar: 10  $\mu$ m.



**Figure 2.** Targeted biotinylation of telomeres, major and minor satellites. Representative confocal images of C2C12 cells, co-transfected with CAG-BirA\*-dCas9-eGFP and a plasmid encoding either telomere- (A, TelgRNA), minor satellite- (B, MiSgRNA) or major satellite-specific sgRNA (C, MaSgRNA). Nuclear enrichment of biotin at targeted sequences is only detectable after addition of exogenous biotin. White dashed lines indicate the border between nucleus and cytoplasm. Scale bar: 10  $\mu$ m.

endogenously biotinylated mitochondrial carboxylases were found in all pull-downs including the negative control (Supplementary Table S1). Besides proteins predicted to associate with DNA, we also detected numerous unexpected proteins in our dataset (Supplementary Table S1) providing a basis for the identification of new chromatin factors and their future comprehensive characterization. For statistical analysis in a two-sided Student's T-test, only proteins present in at least 3 out of 4 biological replicates were included.

First, we targeted telomeric regions and observed a strong enrichment of several proteins when compared to pull-downs from control cells (Fig. 3A). Most prominent among these significantly enriched proteins were TERF2, TINF2 and ACD which are components of the shelterin complex known to directly bind telomeric DNA.<sup>16</sup> We did not identify additional shelterin components which could be explained by sterical hindrances leading to a selective labeling of complex subunits. Altogether, these data show that CasID is suitable to investigate the



**Figure 3.** Chromatin composition of distinct genomic loci determined by mass spectrometry. Volcano plots of proteins enriched at telomeric regions (A), major satellites (B) and minor satellites (C), respectively. Black: significantly enriched/de-enriched proteins relative to BirA\*-dCas9-eGFP control cells without sgRNA. FDR = 0.01, S0 = 0.1, n = 4. (See Table S1.) (D) Overlap between proteins identified at major satellites by CasID and candidates from PICCh analysis.<sup>17</sup> (E) Overlap between proteins significantly enriched at minor and major satellite repeats. (F) Localization of ZNF512-eGFP at major satellite repeats in transiently transfected C2C12 cells. Blow-ups depict DAPI and eGFP signal of boxed regions. Conventional confocal microscopy (upper panel) shows a homogeneous and strong association of ZNF512 at heterochromatin and high-resolution microscopy (3D-SIM, lower panel) reveals a network-like structure. Scale bars: 10 μm (confocal) and 5 μm (3D-SIM).



native protein environment at specific genomic loci in mammalian cells.

Second, we investigated the local protein environment at major satellite repeats. Here, we find not only known heterochromatic proteins such as MECP2, SMCHD1 and HP1BP3 but also previously uncharacterized proteins like ZNF512 (Fig. 3B). We validated the localization of ZNF512 by recombinantly expressing a GFP fusion (ZNF512-eGFP) which showed a distinct signal at heterochromatic loci in C2C12 cells (Fig. 3F). ZNF512 strongly associates with the major satellites also during mitosis (Supplementary Fig. 2), hinting at a structural or regulatory role for this protein throughout the cell cycle. One third of the proteins significantly enriched at major satellite repeats were also found in a data set obtained by PiCh in mouse embryonic stem cells<sup>17</sup> (Fig. 3D). Proteins found in both studies as well as those exclusively detected by CasID are categorized as DNA and RNA binding proteins or repressors (Supplementary Fig. 3A). In contrast to PiCh, CasID requires BirA\*-dCas9 to be introduced in target cells, yet it can be performed with considerably smaller sample sizes ( $\sim 4 \times 10^7$  vs.  $\sim 8 \times 10^8$  cells per sample<sup>17</sup>) rendering CasID feasible and cost-effective. In total, fewer proteins were considered significant with CasID, which may be caused by a stringent statistical cutoff (FDR = 0.01) as well as the proximity-dependent nature of the CasID strategy. Collectively, these results validate CasID as a novel method to study local chromatin composition.

Third, we explored proteins in close proximity to minor satellite repeats and obtained both enriched and de-enriched proteins (Fig. 3C). To our knowledge, this is the first data set describing the protein environment of this genomic element. Among the significantly enriched proteins 12 annotated repressors or chromatin regulators and 25 DNA binding- or zinc finger motif containing-proteins were identified (Supplementary Fig. 3B). Furthermore, we find the known centromere-associated proteins CENPC<sup>18</sup> and PCM1,<sup>19</sup> which may reflect the close proximity of minor satellite repeats and centromeric regions or functions of these factors outside centromeres. Notably, the overlap between minor satellites and major satellite-associated proteins comprises only 9 out of 96 proteins (Fig. 3E), suggesting a distinct protein landscape of these two heterochromatic regions.

In summary, with CasID we developed a simple and robust workflow for *in vivo* labeling and systematic elucidation of locus specific chromatin composition that does not require prior cell fixation or protein cross-linking. We validated CasID for repetitive sequences where multiple Cas9 molecules are recruited to one target site. This approach could be extended to single copy loci by either using multiple sgRNAs, larger sample sizes and/or adapted pulldown conditions. In general, CasID experiments could be further fine-tuned by varying concentration and duration of biotin pulses and the use of a smaller biotin ligase (BioID2)<sup>20</sup> with various linker lengths. While traditional ChIP techniques produce data on genome-wide distribution of specific antigens, CasID allows to study local chromatin composition including the identification of new factors. Therefore, ChIP and CasID are complementary approaches that bring together global and local views of dynamic and functional chromatin complexes and thus help to reveal how these complexes control structure and function of the genome and how they change during development and disease.

## Material and methods

### Cell culture and transfection

C2C12 cells<sup>21</sup> were cultured at 37°C and 5 % CO<sub>2</sub> in Dulbecco's modified Eagle's medium (DMEM, Sigma), supplemented with 20 % fetal bovine serum (FBS, Biochrom), 2 mM L-glutamine (Sigma), 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma). For the CasID assay the culture medium was additionally supplemented with 50 µM biotin (Sigma) one day prior to analysis. For transfections,  $\sim 5 \times 10^5$  cells were seeded in a p35 plate one day prior of transfection and transfections were performed with Lipofectamine<sup>®</sup> 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions.

### Plasmid generation

All plasmid and primer sequences can be found in Supplementary Tables S2 and S3, respectively. To generate the BirA\*-dCas9-eGFP construct, BirA\* was amplified from pcDNA3.1-mycBioID<sup>13</sup> (Addgene plasmid #35700) with primers BirA\*-F and BirA\*-R. The resulting PCR product was ligated into the XbaI

site of pCAG-dCas9-eGFP<sup>10</sup> via Gibson Assembly (New England Biolabs). To generate the pEX-A-U6-sgRNA-PuroR plasmid, the PGK-PuroR cassette was amplified from pPthc-Oct3/4<sup>22</sup> and ligated into the SacI site of pEX-A-sgRNA<sup>10</sup> via Gibson Assembly. sgRNA protospacer sequences were introduced into pEX-A-U6-sgRNA-PuroR by circular amplification as described previously.<sup>10</sup> The Znf512-sequence was amplified from wt E14 cDNA with gene specific primers and cloned between the AsiSI/NotI sites of pCAG-eGFP<sup>23</sup> via Gibson Assembly. The H2B-mRFP expression plasmid was described previously.<sup>24</sup>

### Generation of stable cell lines

C2C12 cells were transfected with pCAG-BirA\*-dCas9-eGFP using Lipofectamine<sup>®</sup> 3000 according to the manufacturer's instructions. Twenty-four h after transfection, the culture medium was supplemented with 10  $\mu$ g/ml blasticidin S (Thermo Fisher Scientific). After two weeks of selection, eGFP-positive cells were single-cell sorted with a FACS Aria II (Becton Dickinson). A clonal cell line, stably expressing BirA\*-dCas9-eGFP was used as entry cell line for transfections with sgRNA plasmids. Twenty-four h after transfection, the medium was supplemented with 2  $\mu$ g/ml puromycin (Applichem). Two weeks after the start of selection, puromycin resistant cells were single-cell sorted. Individual clones (C2C12<sup>BirA\*-dCas9-eGFP/sgRNA</sup>) were checked for correct BirA\*-dCas9-eGFP localization by epifluorescence microscopy.

### Immunofluorescence staining and image acquisition

Immunofluorescence staining was performed as described previously.<sup>25</sup> Briefly, C2C12 cells transfected with pCAG-BirA\*-dCas9-eGFP and the respective sgRNA were grown on coverslips (thickness 1.5H, 170  $\mu$ m  $\pm$  5  $\mu$ m; Marienfeld Superior), washed with phosphate buffered saline (PBS) 24 h after addition of 50  $\mu$ M biotin and fixed with 3.7 % formaldehyde for 10 min. After permeabilization with 0.5 % Triton X-100 in PBS, cells were transferred into blocking buffer (0.02 % Tween, 2 % bovine serum albumin and 0.5 % fish skin gelatin in PBS) and incubated for 1 h. Antibodies were diluted in blocking buffer and cells were incubated with antibodies in a dark, humidified chamber for 1 h at room temperature (RT). Nuclei were counterstained with DAPI (200 ng/ml in PBS, 1  $\mu$ g/ml in PBS for 3D-SIM). Coverslips were mounted with

antifade medium (Vectashield, Vector Laboratories) and sealed with nail polish. Immuno-fluorescence in situ hybridization (FISH) detection of telomeres was performed as described previously.<sup>10</sup> Primary antibodies used in this study were: anti-GFP (1:400, Roche), anti-H3K9me3 (1:500, Active Motif), anti-CENP-B (1:500, Abcam), Streptavidin conjugated to Alexa 594 (1:800, Dianova) and GFP-booster conjugated to Atto 488 (1:200, Chromotek). Secondary antibodies used in this study were: anti-rabbit IgG conjugated to Alexa 594 (1:400, Thermo Fisher Scientific) and anti-mouse IgG conjugated to Alexa 488 (1:300, Invitrogen).

Single optical sections or stacks of optical sections were acquired with a Leica TCS SP5 confocal microscope using a Plan Apo 63x/1.4 NA oil immersion objective. Super-resolution images were acquired with a DeltaVision OMX V3 3D-SIM microscope (Applied Precision Imaging, GE Healthcare), equipped with a 100x/1.4 Plan Apo oil immersion objective and Cascade II EMCCD cameras (Photometrics). Optical sections were acquired with a z-step size of 125 nm using 405 and 488 nm laser lines and SI raw data were reconstructed using the SoftWorX 4.0 software (Applied Precision). For long-term imaging experiments, C2C12 cells were seeded on 8-well chamber slides (ibidi) and transfected with ZNF512-eGFP and H2B-mRFP. Images were obtained with an UltraVIEW VoX spinning disc microscope (PerkinElmer), equipped with a 63x/1.4 NA Plan-Apochromat oil immersion objective and a heated environmental chamber set to 37°C and 5 % CO<sub>2</sub>. Confocal z-stacks of 12  $\mu$ m with a step size of 2  $\mu$ m were recorded every 30 min for ~20 h. Image processing and assembly of the figures was performed with FIJI<sup>26</sup> and Photoshop CS5.1 (Adobe), respectively.

### Denaturing pulldown of biotinylated proteins and sample preparation for mass spectrometry

Four  $\times$  10<sup>7</sup> C2C12<sup>BirA\*-dCas9-eGFP/sgRNA</sup> cells incubated for 24 h with 50  $\mu$ M biotin were processed as described previously.<sup>27-29</sup> In brief, cells were washed once in buffer A (10 mM HEPES/KOH pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.15 % NP-40, 1 $\times$  protease inhibitor (SERVA)), then lysed in buffer A and homogenized using a pellet pestle. After centrifugation (15 min, 3200 rcf, 4°C), the pellet was washed once with PBS. Crude nuclei were resuspended in BioID lysis buffer (0.2 % SDS, 50 mM Tris/HCl pH 7.4,

500 mM NaCl, 1 mM DTT, 1 × protease inhibitor), 0.2 % Triton-X100 was added and proteins were solubilized via sonication (Diagenode Bioruptor<sup>®</sup>, 200 W, 15 min, 30 s “on,” 1 min “off”). Lysates were 2-fold diluted with 50 mM Tris/HCl pH 7.4, centrifuged (10 min, 16000 rcf, 4°C) and the supernatant was incubated with 50  $\mu$ l M-280 Streptavidin Dynabeads (Life Technologies) overnight at 4°C with rotation. A total of 5 washing steps were performed: once with wash buffer 1 (2 % SDS), wash buffer 2 (0.1 % desoxycholic acid, 1 % Triton X-100, 1 mM EDTA, 500 mM NaCl, 50 mM HEPES/KOH pH 7.5), wash buffer 3 (0.5 % desoxycholic acid, 0.5 % NP-40, 1 mM EDTA, 500 mM NaCl, 10 mM Tris/HCl pH 7.4) and twice with 50 mM Tris/HCl pH 7.4. Proteins bound to the streptavidin beads were digested as previously described.<sup>29</sup> Beads were resuspended in digestion buffer (2 M Urea in Tris/HCl pH 7.5), reduced with 10 mM DTT and subsequently alkylated with 50 mM chloroacetamide. A total of 0.35  $\mu$ g trypsin (Pierce, Thermo Scientific) was used for overnight digestion at RT. Desalting of peptides prior to LC-MS/MS analysis was performed via StageTips.<sup>30</sup>

### LC-MS/MS analysis

Tandem mass spectrometry analysis was performed as described previously.<sup>27</sup> In brief, reconstituted peptides (20  $\mu$ l mobile phase A: 2% v/v acetonitrile, 0.1% v/v formic acid) were analyzed using a EASY-nLC 1000 nano-HPLC system connected to a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). For peptide separation, a PepMap RSLC column (75  $\mu$ m ID, 150 mm length, C18 stationary phase with 2  $\mu$ m particle size and 100 Å pore size, Thermo Fisher Scientific) was used, running a gradient from 5% to 35% mobile phase B (98% v/v acetonitrile, 0.1% v/v formic acid) at a flow rate of 300 nl/min. For data-dependent acquisition, up to 10 precursors from a MS1 scan (resolution = 60,000) in the range of m/z 250-1800 were selected for collision-induced dissociation (CID: 10 ms, 35% normalized collision energy, activation q of 0.25).

### Computational analysis

Raw data files were searched against the UniprotKB mouse proteome database (Swissprot)<sup>31</sup> using MaxQuant (Version 1.5.2.8)<sup>32</sup> with the MaxLFQ label free quantification algorithm.<sup>33</sup> Additionally to common

contaminants specified in the MaxQuant “contaminants.fasta” file, a custom-made file containing sequences of BirA\*-dCas9 and fluorescence proteins was included in the database search. Trypsin/P derived peptides with a maximum of 3 missed cleavages and a protein false discovery rate of 1 % were set as analysis parameters. Carbamidomethylation of cysteine residues was considered a fixed modification, while oxidation of methionine, protein N-terminal acetylation and biotinylation were defined as variable modifications.

For evaluation of the identified protein groups, Perseus (Version 1.5.2.6) was used.<sup>32</sup> The data set was filtered for common contaminants classified by the MaxQuant algorithm and only proteins quantified in at least 3 out of 4 replicates per cell line were subjected to statistical analysis. For minor satellite repeats, the dataset was further filtered to exclude proteins only detected in the control sample. Missing values were replaced by a constant value of 17 for significance testing with a two-sided Student’s T-test and a permutation based FDR calculation. Venn diagrams were obtained using the Webtool of the University of Gent (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

### Abbreviations

BioID	proximity dependent biotin identification
ChIP	chromatin immuno precipitation
dCas9	enzymatically dead Cas9
eGFP	enhanced green fluorescent protein
FDR	false discovery rate
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
MaS	major satellite repeats
MiS	minor satellite repeats
PiCh	Proteomic Isolation of Chromatin fragments
sgRNA	single guide RNA
Tel	telomere

### Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

### Acknowledgments

The authors would like to thank Joel Ryan (LMU Munich) and Susanne Leidescher (LMU Munich) for help with live cell and super-resolution imaging, suggestions on the manuscript and FISH staining. T.A., E.S. and P.R. are members of the DFG

Graduiertenkolleg GRK1721. E.S. gratefully acknowledges the International Max Planck Research School for Molecular and Cellular Life Sciences.

## Funding

This work was supported by the Deutsche Forschungsgemeinschaft [SFB1064 and SFB646 to H.L.] and the European Research Council [MolStruKT StG no. 638218 to F.H.]. Funding for open access charge: Deutsche Forschungsgemeinschaft.

## ORCID

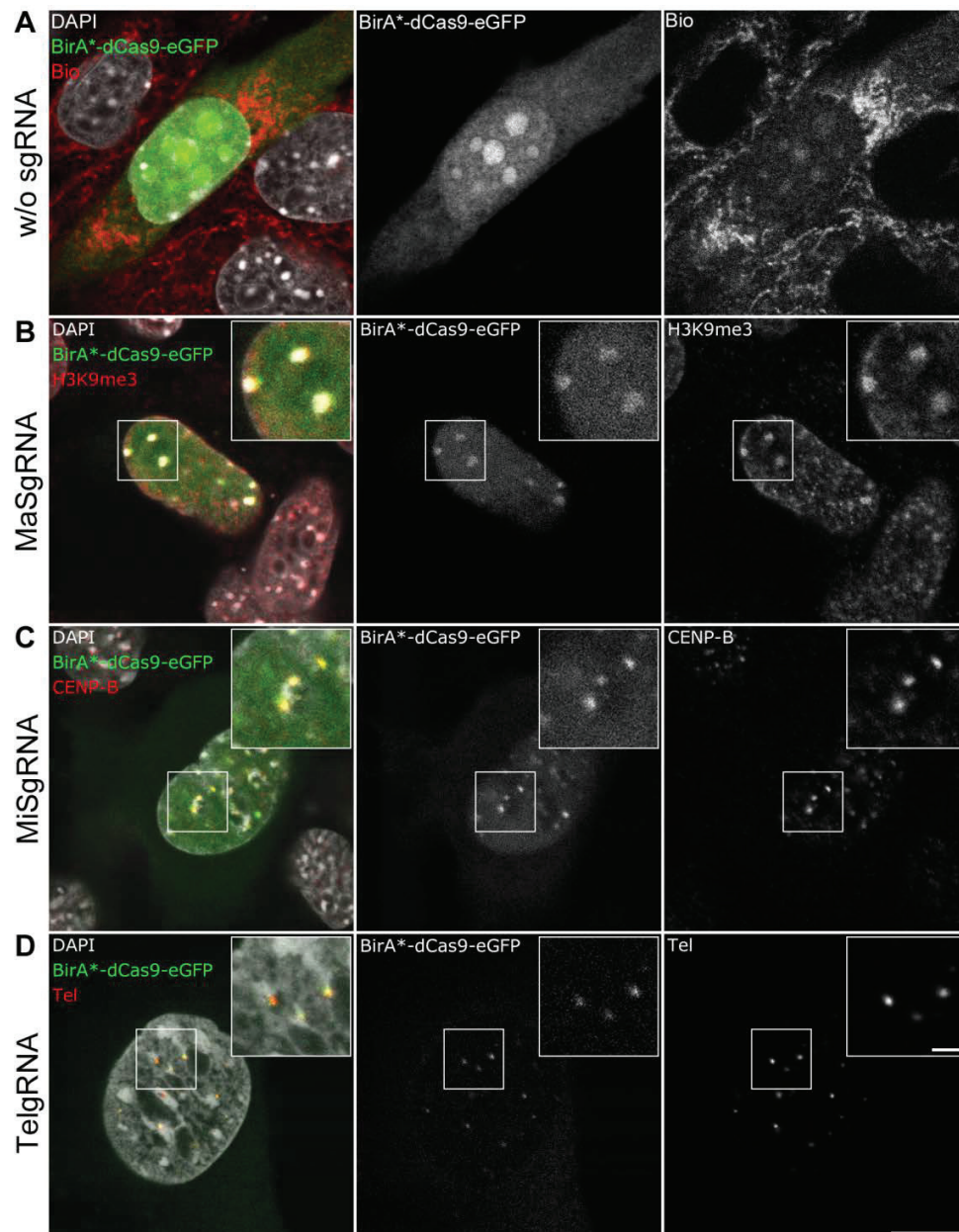
Heinrich Leonhardt  <http://orcid.org/0000-0002-5086-6449>

## References

- [1] Kennedy-Darling J, Guillen-Ahlers H, Shortreed MR, Scalf M, Frey BL, Kendziorski C, Olivier M, Gasch AP, Smith LM. Discovery of Chromatin-Associated proteins via sequence-specific capture and mass spectrometric protein identification in *saccharomyces cerevisiae*. *J Proteome Res* 2014; 13:3810-25; PMID:24999558; <http://dx.doi.org/10.1021/pr5004938>
- [2] Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. *Cell* 2009; 136:175-86; PMID:19135898; <http://dx.doi.org/10.1016/j.cell.2008.11.045>
- [3] Fujita T, Asano Y, Ohtsuka J, Takada Y, Saito K, Ohki R, Fujii H. Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Scientific Reports* 2013; 3:3171; PMID:24201379
- [4] Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, Wahls WP, Taverna SD, Raney KD, Tackett AJ. A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* 2014; 9:1207-11; PMID:25147920; <http://dx.doi.org/10.4161/epi.29919>
- [5] Grolimund L, Aeby E, Hamelin R, Armand F, Chiappe D, Moniatte M, Lingner J. A quantitative telomeric chromatin isolation protocol identifies different telomeric states. *Nat Commun* 2013; 4:2848; PMID:24270157; <http://dx.doi.org/10.1038/ncomms3848>
- [6] Klug A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Quarterly Rev Biophys* 2010; 43:1-21; PMID:20478078; <http://dx.doi.org/10.1017/S0033583510000089>
- [7] Miyanari Y, Ziegler-Birling C, Torres-Padilla ME. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Structural Mol Biol* 2013; 20:1321-4; PMID:24096363; <http://dx.doi.org/10.1038/nsmb.2680>
- [8] Thanisch K, Schneider K, Morbitzer R, Solovei I, Lahaye T, Bultmann S, Leonhardt H. Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic Acids Res* 2013; PMID:24371265
- [9] Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 2013; 155:1479-91; PMID:24360272; <http://dx.doi.org/10.1016/j.cell.2013.12.001>
- [10] Anton T, Bultmann S, Leonhardt H, Markaki Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus (Austin, Tex)* 2014; 5:163-72; PMID:24637835
- [11] Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 2013; 152:1173-83; PMID:23452860; <http://dx.doi.org/10.1016/j.cell.2013.02.022>
- [12] Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science* 2013; 339:823-6; PMID:23287722; <http://dx.doi.org/10.1126/science.1232033>
- [13] Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 2012; 196:801-10; PMID:22412018; <http://dx.doi.org/10.1083/jcb.201112098>
- [14] Kim DI, Birendra KC, Zhu W, Motamedchaboki K, Doye V, Roux KJ. Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc Natl Acad Sci U S A* 2014; 111:E2453-61; PMID:24927568; <http://dx.doi.org/10.1073/pnas.1406459111>
- [15] Lambert JP, Tucholska M, Go C, Knight JD, Gingras AC. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J Proteomics* 2015; 118:81-94; PMID:25281560; <http://dx.doi.org/10.1016/j.jprot.2014.09.011>
- [16] Liu D, O'Connor MS, Qin J, Songyang Z. Telosome, a Mammalian Telomere-associated complex formed by multiple telomeric proteins. *J Biol Chem* 2004; 279:51338-42; PMID:15383534; <http://dx.doi.org/10.1074/jbc.M409293200>
- [17] Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla ME, et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol Cell* 2014; 56:580-94; PMID:25457167; <http://dx.doi.org/10.1016/j.molcel.2014.10.001>
- [18] Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* 2004; 166:493-505; PMID:15302854; <http://dx.doi.org/10.1083/jcb.200403109>
- [19] Balczon R, Bao L, Zimmer W. PCM-1, A 228-kD centrosome autoantigen with a distinct cell cycle distribution. *J Cell Biol* 1994; 124:783-93; PMID:8120099; <http://dx.doi.org/10.1083/jcb.124.5.783>
- [20] Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, Roux KJ. An improved smaller biotin

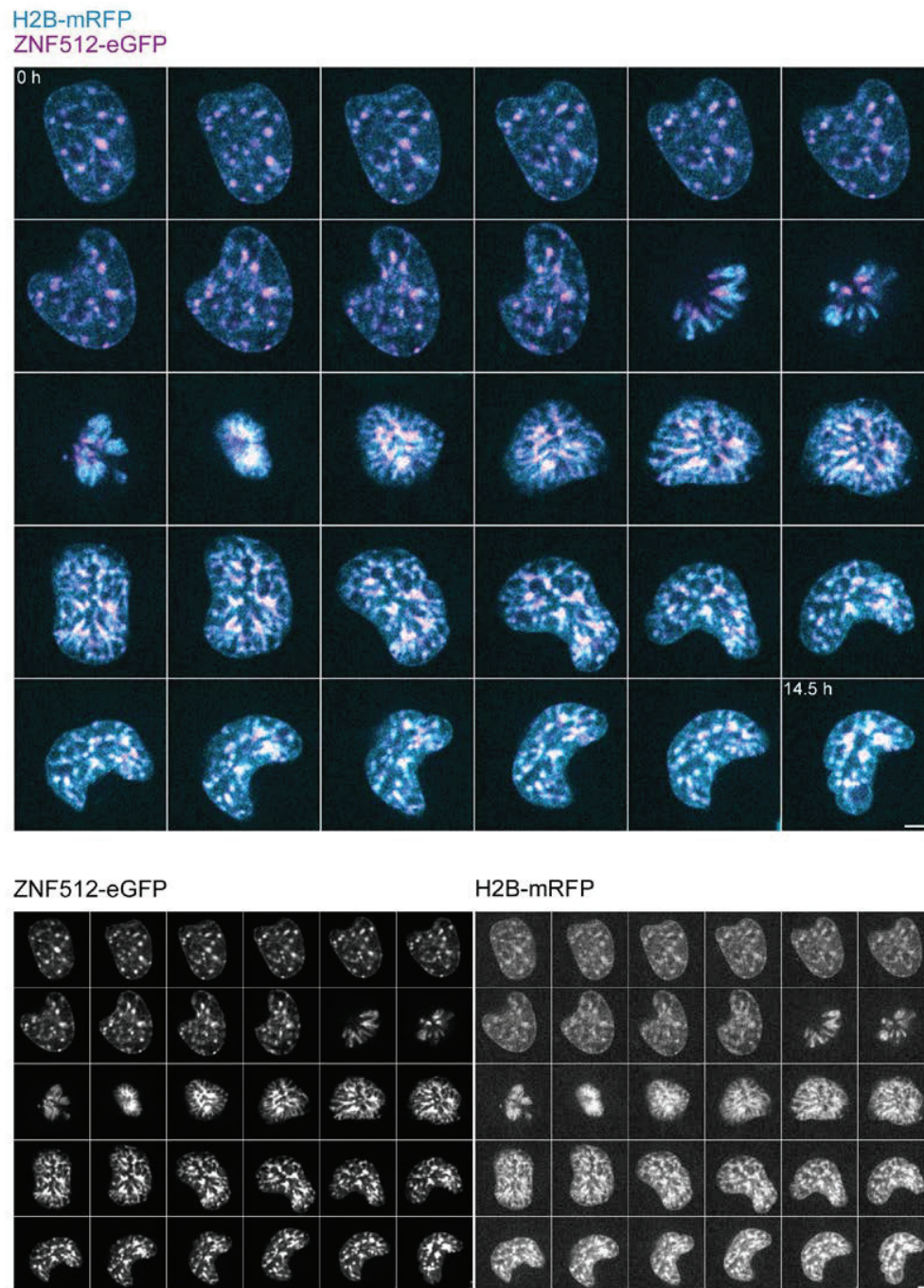
- ligase for BioID proximity labeling. *Mol Biol Cell* 2016; 27(8):1188-96
- [21] Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* 1977; 270:725-7; PMID:563524; <http://dx.doi.org/10.1038/270725a0>
- [22] Masui S, Shimosato D, Toyooka Y, Yagi R, Takahashi K, Niwa H. An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res* 2005; 33:e43; PMID:15741176; <http://dx.doi.org/10.1093/nar/gni043>
- [23] Meilinger D, Fellinger K, Bultmann S, Rothbauer U, Bonapace IM, Klinkert WE, Spada F, Leonhardt H. Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 2009; 10:1259-64; PMID:19798101; <http://dx.doi.org/10.1038/embor.2009.201>
- [24] Martin RM, Cardoso MC. Chromatin condensation modulates access and binding of nuclear proteins. *FASEB J* 2010; 24:1066-72; PMID:19897663; <http://dx.doi.org/10.1096/fj.08-128959>
- [25] Solovei I, Cremer M. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol Biol* 2010; 659:117-26; [http://dx.doi.org/10.1007/978-1-60761-789-1\\_8](http://dx.doi.org/10.1007/978-1-60761-789-1_8)
- [26] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; 9:676-82; PMID:22743772; <http://dx.doi.org/10.1038/nmeth.2019>
- [27] Mulholland CB, Smets M, Schmidtmann E, Leidescher S, Markaki Y, Hofweber M, Qin W, Manzo M, Kremmer E, Thanisch K, et al. A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res* 2015; PMID:26007658
- [28] Roux KJ, Kim DI, Burke B. BioID: A screen for protein-protein interactions. *Curr Protoc Protein Sci* 2013; 19(23):1-19.23.14; PMID:24510646
- [29] Baymaz HI, Spruijt CG, Vermeulen M. Identifying nuclear protein-protein interactions using GFP affinity purification and SILAC-based quantitative mass spectrometry. *Meth Mol Biol* 2014; 1188:207-26; [http://dx.doi.org/10.1007/978-1-4939-1142-4\\_15](http://dx.doi.org/10.1007/978-1-4939-1142-4_15)
- [30] Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protocols* 2007; 2:1896-906; PMID:17703201; <http://dx.doi.org/10.1038/nprot.2007.261>
- [31] Consortium TU. UniProt: a hub for protein information. *Nucleic Acids Res* 2015; 43:D204-D12; PMID:25348405; <http://dx.doi.org/10.1093/nar/gku989>
- [32] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech* 2008; 26:1367-72; <http://dx.doi.org/10.1038/nbt.1511>
- [33] Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate Proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 2014; 13:2513-26; <http://dx.doi.org/10.1074/mcp.M113.031591>

Fig. S1



**Supplementary Figure 1 | Sub-cellular localization of BirA\*-dCas9-eGFP.** (A) Without a sgRNA, BirA\*-dCas9-eGFP shows a disperse localization throughout the cell and an enrichment at nucleoli. Cells were incubated with 50  $\mu$ M biotin (Bio) for 24 hours. (B-C) When co-expressed with a sequence-specific sgRNA, BirA\*-dCas9-eGFP is recruited to distinct loci. Correct localization is confirmed by either immunofluorescence of H3K9me3 (B), CENP-B (C) or fluorescence in situ hybridization with a telomere-specific probe (D). Scale bar: 10  $\mu$ m. Scale bar in blow-ups: 2  $\mu$ m.

Fig. S2



**Supplementary Figure 2 | Sub-nuclear localization of ZNF512-eGFP during the cell cycle.** Time lapse imaging of C2C12 cells, transfected with H2B-mRFP and ZNF512-eGFP. Images were acquired every 30 min. Scale bar: 5  $\mu$ m.

**Fig. S3**



**Supplementary Figure 3 | Uniprot (Keyword) annotations of proteins. (A)** major satellite associated proteins. bold: proteins exclusively identified in CasID **(B)** minor satellite associated proteins.



**Table S1 | Proteins identified in CasID pulldowns.** Significantly enriched proteins (Student's T-test, FDR = 0.01) are highlighted in color. Common BioID contaminants are marked in grey.

Table\_S1.xlsx

**Table S2 | Plasmid sequences of constructs used in this study.**

Table\_S2.docx

**Table S3 | Sequences of oligonucleotides used in this study.**

Primer	Sequence 5'-3'
BirA*-F	GGCGTGTGACCGGCGGCTatggaacaaaaactcatc
BirA*-R	GAGTACTTCTTGTCCATTCCgctaccgctgccgctaccGCGGTTTAAACTTAAGC
PuroR-F	catatgggtaccgagcttaCCGGGTAGGGGAGGCG
PuroR-R	gcttgccggccgcgagctggtCCGCCCTCAGAAGCCATAG
MaSgRNA-F	GGCAAGAAAACACTGAAAATCAgtttttagagctagaaatagcaag
MaSgRNA-R	TGATTTTCAGTTTTCTTGCCcggtgtttcgtcctttccac
MiSgRNA-F	ACACTGAAAACACATTCGTgtttttagagctagaaatagcaag
MiSgRNA-R	ACGAATGTGTTTTTCAGTGTcggtgtttcgtcctttccac
TelgRNA-F	TAGGGTTAGGGTTAGGGTTAgtttttagagctagaaatagcaag
TelgRNA-R	TAACCCTAACCCTAACCCTAcggtgtttcgtcctttccac
Znf512-F	CGCCACCATGGgcatATGTCTTCCAGACTCGGTG
Znf512-R	GGAATTCGTAACTgcCTACTTCCTCCCTCGTTTGTG



### 3 DISCUSSION

#### 3.1 The CRISPR/Cas system as a tool to visualize chromatin

The spatiotemporal organization of the genome within the nucleus, as well as its interactions with epigenetic factors and other cellular components play a fundamental role in regulating cellular functions such as gene expression and genome integrity. Yet, the exact mechanisms of how chromatin is packaged in the physical space of the nucleus remain largely elusive. To investigate the spatial relationship between chromatin segments, 3C and derived methods such as Hi-C have been successfully employed. These approaches have provided new insights into folding principles of chromatin and uncovered, among others, the existence of TADs as one of the major building blocks of higher chromatin organization. On a larger scale, FISH-based approaches allowed to examine morphological characteristics of chromosomes. For instance, FISH-probes, designed to label whole chromosomes, revealed that individual chromosomes are not randomly distributed within the nucleus, but rather occupy distinct chromosome territories. Moreover, it was demonstrated that in many cases the radial positioning of chromosome territories reflects chromosomal gene density, that is gene-poor chromosomes are predominantly located at the nuclear periphery, whereas gene-rich chromosomes cluster towards the center of the nucleus. However, both methodologies rely on fixed specimens and thus only present a snapshot of nuclear organization. In addition, chromatin conformation capture is performed on bulk chromatin, hence single cell resolution is limited. FISH on the other hand, allows to directly visualize the spatial organization of chromatin in individual cells, yet probe hybridization requires harsh DNA denaturation and can obstruct the ultra-structure of native chromatin arrangements. Notably, although data acquired by FISH and 3C-based approaches are generally considered to be concordant, a study, comparing both methods to investigate the *HoxD* locus in mouse ESCs at different developmental stages, demonstrated that high resolution 3C interaction maps do not always reflect the chromatin topography as it is visualized by FISH [Williamson et al., 2014]. Taken together, these drawbacks highlight the necessity to test hypotheses about genome organization with independent methods. In this context, live-cell imaging approaches are essential to visualize spatiotemporal chromatin dynamics and to corroborate the static snapshots provided by FISH and 3C-based methods.

The discovery that the bacterial protein Cas9 can be harnessed as a RNA-guided endonuclease in prokaryotic, as well as eukaryotic cells enabled CRISPR/Cas-based genome engineering approaches and widely replaced previous types of sequence-specific nucleases, such as zinc finger- or TALE nucleases [Haeussler & Concordet, 2016]. In line with that, also the nuclease

deficient mutant of Cas9, dCas9, has been utilized to target specific loci without cleaving the underlying genomic sequence. By fusing dCas9 to regulatory domains, it has been shown that the CRISPR/Cas system is a versatile tool to manipulate endogenous gene expression. To further expand the CRISPR/Cas toolkit, we and others aimed to utilize dCas9 as an RNA-guided DNA-binding platform to visualize defined genomic regions in living cells.

In our proof-of-principle study, we decided to target an eGFP-tagged dCas9 to the heterochromatic sequences of telomeres, minor and major satellites in mouse ESCs. Since these genomic elements are predominantly composed of repetitive sequences, this approach enabled us to recruit dCas9-eGFP to large stretches of DNA and thereby enrich the eGFP-signal with only one sgRNA per target. By combining dCas9-based labeling (CRISPR imaging) of these loci with FISH and immunofluorescence detection, we confirmed that dCas9-eGFP efficiently and specifically labels repetitive sequences in living cells. Moreover, we demonstrated that CRISPR imaging can be adopted for high-resolution microscopy, enabling us to resolve the ultrastructure of diffraction limited chromatin clusters such as telomeres.

Notably, our approach could be implemented in future correlative microscopy studies. This would include to first trace the dynamics of a targeted chromatin region *in vivo* at a lower resolution via live-cell microscopy. Subsequently, the cells could be fixed and processed for high-resolution microscopy, allowing to correlate insights of spatiotemporal and structural characteristics of the same locus.

Although to date CRISPR imaging is mainly limited to repetitive sequences, one study could demonstrate that by co-expressing ~30 different sgRNAs, tracing of a single-copy locus is feasible [Chen et al., 2013]. The prospect of visualizing the 3D organization of two distinct loci in living cells is particularly interesting as it offers means to substantiate results, which were obtained by 3C and related methods. During development, for instance, gene expression profiles have been shown to be tightly regulated by the formation of distal enhancer/promoter loops [Phillips-Cremins et al., 2013]. In this context, CRISPR imaging could be used to investigate, when and under what circumstances these distally located regions come into contact. Similarly, it is known that TADs are absent in mitotic cells. However, it remains elusive, how these chromatin domains are reestablished after cells enter G1-phase [Dekker, 2014]. Here, specific labeling of sequences at TAD-boundaries is likely to shed light on this question.

### *3.1.1 CRISPR imaging in comparison to modular DNA-binding proteins*

In living cells, bulk chromatin can be visualized by cell-permeable DNA-dyes (e.g. DRAQ5), fluorescently labeled nucleotides (e.g. Cy5-dUTP) or histones fused to fluorescent proteins (e.g. H2B-GFP) [Kanda et al., 1998; Manders et al., 1999; Martin et al., 2005]. Furthermore,

---

specific genomic sequences that are characterized by a defined protein composition, such as centromeres or telomeres can be traced by fluorescently tagging their associated binding proteins [Shelby et al., 1996; Krawczyk et al., 2012]. Since loci, which meet this requirement are limited, *lac* or *tet* operator repeats were introduced at specific genomic regions via genome engineering. These sequences can then be recognized by Lac- or Tet-repressor proteins, respectively, fused to fluorescent proteins [Robinett et al., 1996; Roukos et al., 2013]. However, since these large artificial sequences might interfere with chromatin dynamics, this method only provides indirect information about the native locus. Hence, live-cell approaches to examine the spatiotemporal dynamics of endogenous loci rely on DNA-interacting proteins, which bind their target in a sequence-specific manner (Figure 9).

Fused to GFP, PZFs were the first modular DNA-binding proteins, which were used to visualize repetitive genomic sequences *in vivo* [Lindhout et al., 2007]. It was shown that PZFs can be harnessed to trace major satellite sequences in mouse cells, as well as centromeric repeats in *Arabidopsis*, demonstrating that this method can be applied in different organisms. Although PZFs have additionally been successfully employed for genome editing and transcriptional regulation, cross-talk between adjacent zinc finger modules can alter the binding preference towards the target sequence. Hence, newly designed PZFs have to be tested thoroughly, rendering this approach time-consuming and expensive [Ramirez et al., 2008].

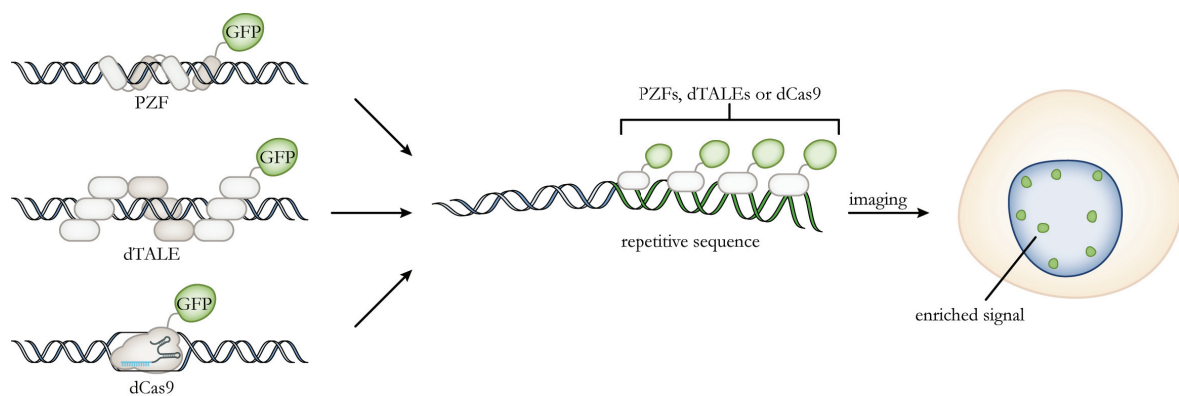
To overcome these limitations, fluorescently tagged dTALEs have been employed to substitute PZFs [Ma et al., 2013; Miyanari et al., 2013; Thanisch et al., 2014; Yuan et al., 2014; Ren et al., 2017]. Contrary to PZFs, target recognition via dTALEs is facilitated via a simple RVD-based code. The fact that one RVD specifically recognizes one nucleotide greatly simplifies the process of designing new dTALEs. Importantly, dTALEs are characterized by a high target specificity and in fact, were shown to distinguish 1 – 2 nt differences on the target sequence [Boch et al., 2009]. Hence, fluorescent dTALEs allow to detect individual chromosomes by single nucleotide polymorphisms (SNPs) [Miyanari et al., 2013]. Despite these evident advantages over PZFs, the highly repetitive central domain of dTALEs has been suggested to cause self-assembly and formation of protein aggregates, preventing an effective binding of the cognate target DNA. To this end, a recent study demonstrated that fusion of dTALEs to the chaperone-like protein thioredoxin (TRX) enhances their solubility in human cells [Ren et al., 2017]. However, the repetitive nature of the DNA-recognition domain renders it difficult to reassemble dTALEs for different sequences and necessitates the use of laborious cloning techniques [Morbiter et al., 2011; Reyon et al., 2012; Schmid-Burgk et al., 2013]. Since the target specificity of dTALEs is mediated by the protein itself, simultaneous

tracing of multiple genomic elements can be easily achieved by tagging different site-specific dTALEs with different fluorescent proteins. Using this approach, it was shown that centromeres and telomeres can be visualized independently within the same nucleus [Ma et al., 2013]. Importantly, also sequences, which are located in close proximity to one another, such as minor and major satellites can be resolved [Miyanari et al., 2013]. Elucidating the spatiotemporal organization of chromatin is of key importance in understanding how gene expression is regulated. To this end, multicolor imaging via dTALEs is likely to play a crucial role.

CRISPR imaging can be regarded as the next generation of tools to study chromatin dynamics in living cells. Contrary to PZF- or dTALE-based approaches, target specificity is solely mediated by the sgRNA. Therefore, CRISPR imaging can easily be customized to visualize new sequences by simply exchanging the sgRNA without the need to replace the protein itself. By employing CRISPR imaging, we and others successfully labeled repetitive sequences. Although such tandem repeats are present in virtually all eukaryotes, it might prove difficult to find one near a locus of interest. Tracing of single-copy loci, therefore, requires targeting of many different consecutive sequences in order to enrich the signal over background levels. In this context, CRISPR imaging is likely to be superior, as it only requires the introduction of small RNAs, whereas for dTALE- or PZF-based approaches, different proteins for each target sequence have to be expressed. This is illustrated by the fact that labeling of a non-repetitive sequence via dTALEs has not been realized, so far. Moreover, a recent study presented a method that employs a cocktail of restriction enzymes to generate a genome-wide library of sgRNAs [Lane et al., 2015]. Further refinement of this method could allow to produce sgRNA-libraries, which are specific for whole chromosomes or non-repetitive subsets of chromosomes. Together with suitable means to deliver these libraries into living cells (e.g. via lentiviral transduction), CRISPR imaging could facilitate to monitor the spatiotemporal dynamics of chromosomes in real time.

To study the spatiotemporal dynamics of a distinct locus, continuous labeling throughout the cell-cycle is crucial. To this end, we assessed, whether dCas9-eGFP associates with telomeres, minor and major satellites during mitosis. We found that the dramatic structural changes, to which mitotic chromosomes are subjected to, do not perturb dCas9-eGFP binding. This is particularly striking, since most endogenous DNA-binding proteins, such as sequence-specific transcription factors and co-factors are removed from mitotic chromatin [Kadauke & Blobel, 2013]. Interestingly, to date only two out of five studies, in which dTALEs were employed to visualize repetitive sequences could demonstrate that dTALEs remain associated to their cognate target sequence during mitosis [Miyanari et al., 2013; Ren et al., 2017].

An important aspect of labeling DNA in living cells is whether the employed system disturbs the function and integrity of the targeted locus. Since the sgRNA forms a heteroduplex with the targeted DNA-strand, binding of dCas9 results in the local unwinding of the DNA double helix. This could potentially interfere with nucleosome positioning and the recruitment of other DNA-binding proteins [Costantino & Koshland, 2015]. In line with that, it was observed that the presence of nucleosomes at the target sequence impedes the DNA recognition of Cas9 *in vitro* and *in vivo* [Horlbeck et al., 2016; Isaac et al., 2016]. Interestingly, in our experiments the compact, heterochromatic conformation of repetitive elements did not impair dCas9 binding. However, we detected rare events of elongated and fiber-like structures, when we visualized telomeres by CRISPR imaging. This suggests that, in some cases, dCas9-eGFP recruitment to telomeric tandem repeats might forestall T-loop formation. However, the vast majority of observed telomeres displayed characteristic spot-like structural features and an independent study confirmed unaltered dynamics of dCas9-labeled telomeres [Chen et al., 2013]. Moreover, the structural integrity of both minor and major satellite repeats was not affected by CRISPR imaging. Contrary to dCas9, dTALEs do not induce DNA-unwinding, but rather bind directly along the major groove. Conceivably, this mode of DNA-detection could lead to a sterical clash between abundant nucleosomes and the dTALE. Yet, expression of a major satellite-specific dTALE resulted in unchanged histone H3 and H3K9me3 occupancies at the pericentromere, suggesting that dTALEs do not affect chromatin configuration [Miyanari et al., 2013].



**Figure 9: Sequence-specific *in vivo* visualization of distinct loci via DNA-binding proteins.** Modular DNA-binding proteins (PZFs and dTALEs), as well as a catalytically inactive Cas9 (dCas9) bind to genomic DNA in a sequence-specific manner. When fused to a fluorescent protein (GFP, green), these systems can guide the fluorescence signal to pre-defined genomic elements. Targeting heterochromatic tandem repeats or repetitive sequences within protein-coding genes, leads to local enrichment of the signal, which can be readily imaged in living cells. Modified from [Chen et al., 2016a].

Collectively, CRISPR imaging offers high flexibility and represents a time- and cost-effective alternative to visualization techniques, based on modular DNA-binding proteins. Especially the potential to trace non-repetitive sequences renders CRISPR imaging a versatile tool to study chromatin dynamics. Yet, considering recent advances such as TRX-coupled dTALEs and the fact that dTALEs are independent from a PAM-sequence, indicates that dTALE-based genome visualization and CRISPR imaging are complementary approaches in labeling repetitive genomic elements.

### *3.1.2 Expanding the CRISPR imaging toolkit*

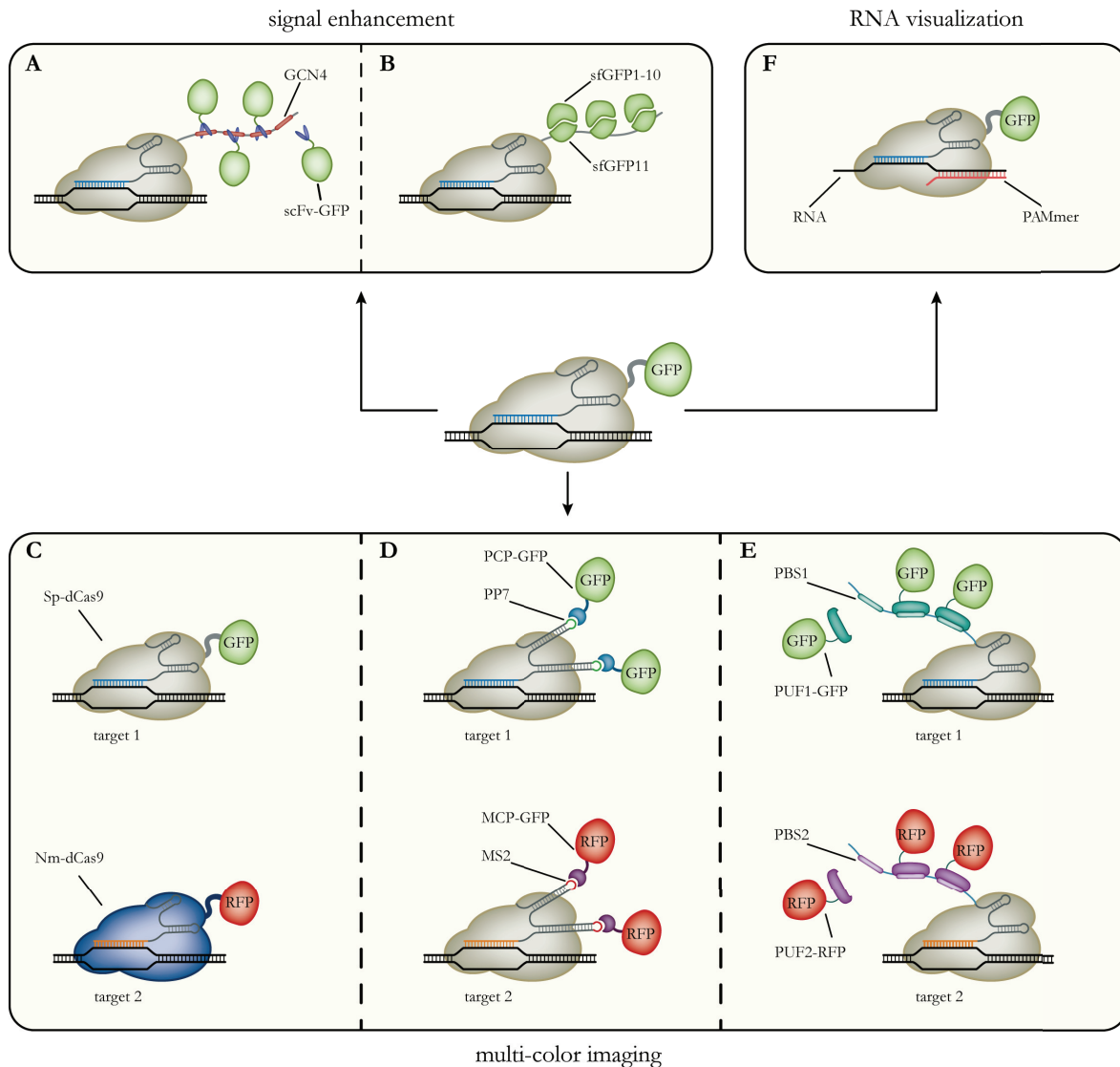
Since the first demonstration of CRISPR imaging, great effort has been made to further optimize this system (Figure 10). One issue of visualizing single-copy loci via CRISPR imaging is a rather low signal-to-noise ratio, caused by freely diffusing dCas9-eGFP molecules. To this end, it has been recently demonstrated that the SunTag (SUPERNOVA tag) is suitable to significantly amplify the fluorescence signal at targeted loci (Figure 10A) [Tanenbaum et al., 2014]. This system comprises a peptide, which contains up to 24 tandemly arranged copies of a short peptide epitope (GCN4), and a GFP-tagged cognate single-chain antibody (scFv). By combining the SunTag with CRISPR imaging of telomeres, it was shown that a ~20-fold signal enhancement is possible. Due to this bright signal, it would be feasible to visualize genomic loci with lower light illumination, thus reducing phototoxic effects during long-term imaging. As this system has only been tested on repetitive sequences, however, it would be interesting to assess its functionality on single-copy loci.

Similar to the SunTag, a split version of super-folder GFP (sfGFP) has been adopted as an epitope tag for signal enhancement in targeted gene activation and fluorescence imaging (Figure 10B) [Kamiyama et al., 2016]. For this approach, sfGFP is split between the 10<sup>th</sup> and 11<sup>th</sup>  $\beta$ -strand, resulting in the non-fluorescent sfGFP1–10 and sfGFP11, which serves as a short epitope. A tandem array, comprising up to seven repeats of the sfGFP11 epitope, can be fused to a protein of interest (POI) and subsequently recognized by the co-expressed sfGFP1–10 fragment. Upon self-complementation, the chromophore matures and sfGFP regains its fluorescence [Cabantous et al., 2005]. By expressing dCas9-sfGFP11<sub>x7</sub> together with VP64-sfGFP1–10, it was demonstrated that this system is capable to drastically increase the expression level of the targeted *CXCR4*-locus. Hence, it would be interesting to evaluate, whether the sfGFP11-tag can be adopted for CRISPR imaging. Whereas the SunTag system still might cause background fluorescence due to unbound scFv-GFP molecules, this GFP-derived epitope tag would be particularly promising, since sfGFP1–10 can be overexpressed without causing background signals.



In comparison to dTALE-based visualization of genomic elements, one inherent drawback of the original CRISPR imaging technique is the fact that simultaneous multi-color labeling of several loci is not feasible. To overcome this limitation, orthogonal dCas9 proteins from the three different bacterial species *S. pyogenes* (Sp-dCas9), *Neisseria meningitidis* (Nm-dCas9) and *Streptococcus thermophilus* (St1-dCas9) were harnessed (Figure 10C) [Ma et al., 2015]. Importantly, target recognition of these dCas9 variants is constrained by different PAM sequences, enabling multiplexed sgRNA-guided recruitment to multiple genomic sites [Esvelt et al., 2013]. Hence, by simultaneously co-expressing differently tagged dCas9 orthologs with their cognate sgRNAs, each specific for a distinct locus, it was possible to resolve the inter- and intrachromosomal distances between two repetitive sequences in living human cells. In addition, another study repurposed the small Cas9 ortholog from *Staphylococcus aureus* (Sa-dCas9) for multiplexed CRISPR imaging and demonstrated that a combination of Sp-dCas9-eGFP and Sa-dCas9-mCherry is capable to resolve two different loci, which are separated by less than 300 kb [Chen et al., 2016b].

Although representing a significant improvement in CRISPR imaging, orthogonal CRISPR/Cas systems are characterized by more complex PAM requirements. For instance, whereas Sp-dCas9 recognizes a 5'-NGG-3'-sequence, Nm-dCas9 is only targeted to sequences followed by a 5'-NNNNGGTT-3' motif, thus restricting the flexibility towards target sequences. To circumvent this issue, a second strategy for multi-color CRISPR imaging has been developed (Figure 10D). Here, the *S. pyogenes* sgRNA is fused to the RNA aptamers MS2 or PP7 that are specifically bound by the bacteriophage coat proteins MCP (MS2 coat protein) and PCP (PP7 coat protein), respectively [Fu et al., 2016; Wang et al., 2016]. The resulting scaffold RNA (scRNA) is still capable to guide dCas9 to the desired target and can be detected by fluorescently tagged coat proteins. A similar approach combined dCas9-mediated DNA-recognition with Pumilio-assisted RNA-binding (Figure 10E) [Cheng et al., 2016]. This technique, termed Casilio, utilizes the fact that the RNA-binding domain of the *Drosophila* protein Pumilio (PUF) can be reprogrammed to bind specific 8-mer RNA sequences (PUF-binding site, PBS) [Chen & Varani, 2013]. Therefore, appending different sgRNAs with PUF-binding sites, specific for differently tagged PUF domains, facilitates simultaneous tracing of multiple genomic loci.



**Figure 10: Expanding the CRISPR imaging toolkit.** A-B) To enhance the signal-to-noise ratio of CRISPR imaging, dCas9 is fused to arrays of small peptide epitopes (GCN4 and sfGFP11). These epitopes then either recruit fluorescent molecules (scFv-GFP) or are complemented (sfGFP1–10), reconstituting fluorescence. C-E) Multi-color CRISPR imaging can be achieved by either co-expressing differentially labeled orthogonal dCas9 proteins (Sp-dCas9-GFP and Nm-dCas9-RFP) or by fusion of RNA aptamers (PP7 or MS2) to the sgRNA and co-expressing the cognate, fluorescently tagged binding proteins (PCP-GFP and MCP-GFP, respectively). Moreover, sgRNAs can be appended by PUF-binding sites (PBS1 or PBS2, respectively). These sites are then recognized by differentially tagged PUF proteins (PUF1-GFP and PUF2-RFP). F) By substituting the PAM sequence in the form of an oligonucleotide (PAMmer), dCas9 can be targeted to single stranded RNA molecules.

Recently, the scope of CRISPR imaging has been further expanded by demonstrating that dCas9 can be harnessed as a tool to visualize endogenous, unmodified RNAs (Figure 10F) [Nelles et al., 2016]. Although the CRISPR/Cas system has evolved to target double-stranded DNA, Cas9 can be guided to RNA, when the PAM is provided in *trans* by an oligomer (PAMmer) that is partially complementary to the targeted RNA [O'Connell et al., 2014]. However, due to high background signals, so far RNA could only be visualized, when large

---

quantities accumulated within stress granules. In this case, fusing dCas9 to the aforementioned sfGFP11-tag is likely to enhance the signal-to-noise ratio, thus facilitating the tracking of less abundant RNAs.

### 3.2 Editing of epigenetic marks via the CRISPR/Cas system

The eukaryotic transcriptional landscape is regulated by a complex network of epigenetic modifications, transcription factors and chromatin organization. To dissect cause and consequence of any of these factors, it is crucial to have tools at hand that allow their precise manipulation. Since the first description of Cas9-mediated genome engineering, the CRISPR/Cas system has been in parallel refined for site-specific DNA visualization and manipulation of epigenetic marks and thus represents a powerful tool to interrogate the mechanistic relationship between chromatin state and regulation of gene expression.

Direct fusions of dCas9 with epigenome editing factors have been successfully employed to activate transcription of previously silenced genes. For instance, recruiting the catalytic core of the human acetyltransferase p300 to promoters or enhancers leads to robust transcriptional activation and directly implicates H3K27ac in this process [Hilton et al., 2015]. Additionally, targeted DNA demethylation, mediated by dCas9-TET1 fusions results in gene reactivation [Choudhury et al., 2016; Liu et al., 2016; Xu et al., 2016]. Analogous to targeted activation, the CRISPR/Cas system has been utilized to recruit DNA *de novo* methyltransferases or histone demethylases to specific gene regulatory regions, resulting in a local repressive epigenetic state and thus gene silencing [Kearns et al., 2015; McDonald et al., 2016; Vojta et al., 2016].

Based on these results, we set out to further optimize the CRISPR/Cas system as a precise and flexible tool for epigenetic modulation. To this end, we constructed a fusion of dCas9 with a GFP-binding nanobody (GBP) [Helma et al., 2015], which enabled us to direct virtually any GFP-tagged effector to distinct target sequences. In our proof-of-concept study, we demonstrated that this modular system facilitates the recruitment of GFP-coupled catalytic domains of TET1 and DNMT3A to major satellite repeats. Furthermore, we confirmed preceding results, suggesting that by recruiting the catalytic activities of TET1 and DNMT3A, prevalent epigenetic marks can be erased or set *de novo*, respectively. In addition to site-specific targeting, control over the timing of epigenetic editing is crucial to distinguish direct consequences from local epigenetic perturbations. Therefore, we utilized an inducible system, which facilitated expression of both GBP-dCas9 and the GFP-tagged effector in a timely controlled manner.

For future studies it would be interesting to assess, whether our modular recruitment system is applicable to target epigenetic effector proteins, which are expressed at physiological levels. For this, the multifunctional integrase (MIN) tag could be used to label an effector of interest [Mulholland et al., 2015]. This strategy requires to first insert a phage attachment (*attP*) site directly downstream of the start codon or upstream of the stop codon via CRISPR/Cas-mediated genome engineering. This *attP* sequence then serves as an entry site for Bxb1-assisted integration of a GFP-coding sequence. Notably, this approach would ensure that expression of the GFP-tagged effector is driven by the endogenous promoter, which would minimize potential artificial effects, caused by overexpression. Moreover, it would be possible to knock-in a cassette, which encodes GFP-tagged, mutated variants of the effector of interest, enabling to directly determine their consequences at targeted loci. To further expand our approach, multiplexed recruitment of fluorescently tagged effectors could be tested. By fusing a dCas9 ortholog (e.g. Nm-dCas9) with an RFP-specific nanobody and co-expressing it with GBP-dCas9, two differently tagged effector proteins could be simultaneously recruited to distinct genomic elements.

### 3.3 Probing chromatin composition by CasID

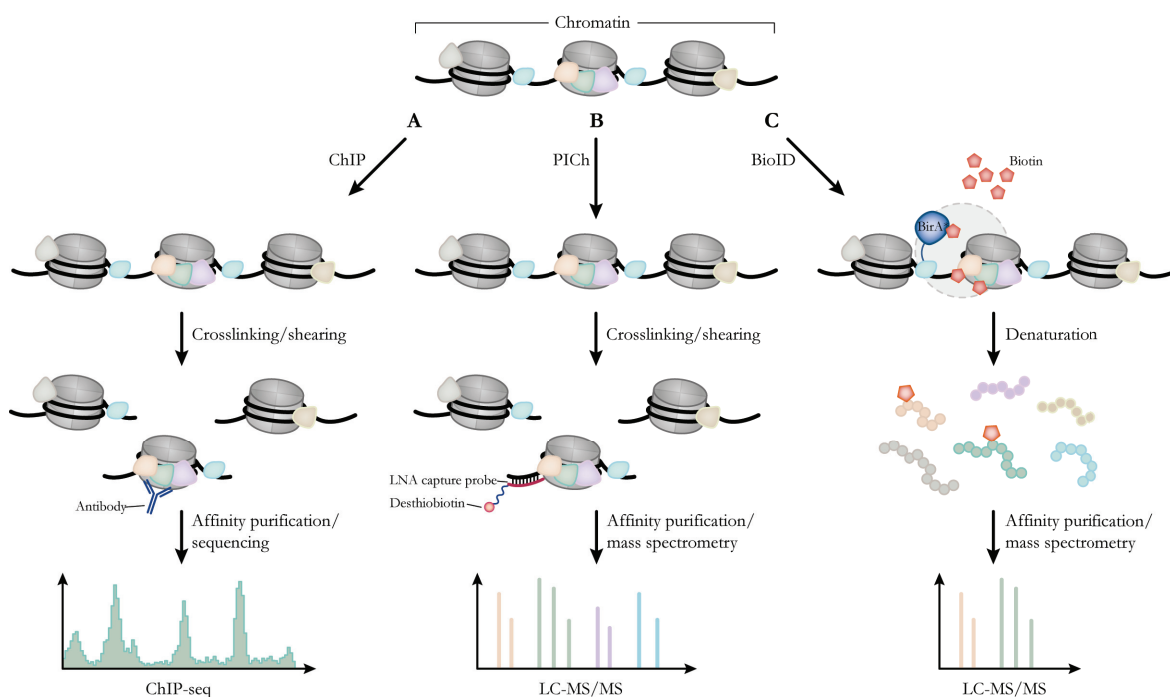
Besides visualizing the spatiotemporal dynamics of chromatin, it is crucial to also assess its protein composition to fully decipher the structure and function of chromatin. To this end, numerous methods have been developed that facilitate the profiling of genome-wide or local binding sites of chromatin proteins, as well as protein-protein associations (Figure 11).

#### Chromatin immunoprecipitation (ChIP)

ChIP represents a well established and powerful method to study the distribution of a DNA-binding POI along the genome. For this, DNA-protein complexes are crosslinked *in vivo* by treating the cells with formaldehyde [Jackson, 1978]. Subsequently, the chromatin is fragmented by either sonication or digestion with micrococcal nuclease (MNase). An antibody, which is specific for the POI is used to immunoprecipitate chromatin fragments. Finally, DNA-POI crosslinks are reversed and the released DNA is assayed by next-generation sequencing (ChIP-seq), PCR (ChIP-qPCR) or microarray (ChIP-chip) to determine the sequences bound by the POI [Collas, 2010]. Using refined bioinformatics algorithms, these sequences are aligned to a reference genome, which allows to call enrichments of POI-DNA interactions (Figure 11A) [Wilbanks & Facciotti, 2010].

### Proteomics of isolated chromatin segments (PICh)

Contrary to ChIP-based approaches, PICh enables the analysis of the overall protein composition of a specific locus without prior knowledge of the identity of bound proteins (Figure 11B) [Déjardin & Kingston, 2009]. Here, chromatin is thoroughly crosslinked and subsequently sheared via sonication. Purification of specific loci is based on ~25 nt long nucleic acid probes, which are complementary to the targeted locus and contain approximately 50 % locked nucleic acid (LNA) residues. Due to their altered backbone, these LNA residues favor base stacking and thereby increase the melting temperature of DNA-LNA hybrids, resulting in increased hybrid stability [Vester & Wengel, 2004]. Moreover, these probes are linked to desthiobiotin at their 5'-end, which facilitates the capture of LNA-chromatin hybrids via avidin-coated magnetic beads [Hirsch et al., 2002; Morocho et al., 2005]. Finally, locus-specific proteins are identified via mass spectrometry (MS). Initially developed to study the protein composition of human telomeres, PICh has since been successfully applied to determine associated proteins of other repetitive genomic elements (MaS), as well as more complex targets, such as ribosomal RNA gene promoters and telomere-associated sequences (TAS) [Déjardin & Kingston, 2009; Antão et al., 2012; Saksouk et al., 2014; Ide & Dejardin, 2015].

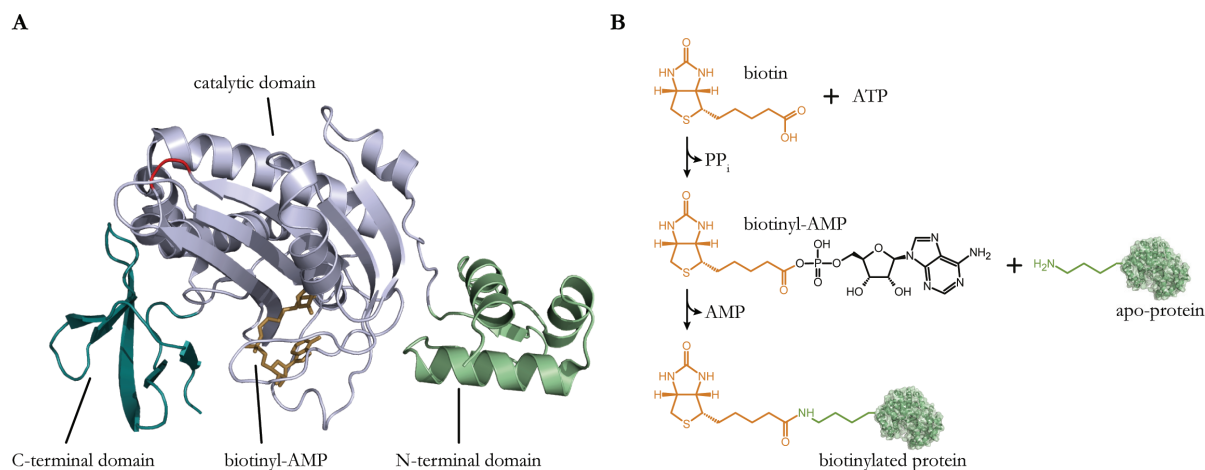


**Figure 11: Schematic representation of methods to study chromatin composition.** A) With ChIP, the DNA-binding profile of a POI at a given time or condition can be assessed. For this, the chromatin is first crosslinked and fragmented, followed by affinity purification via a POI-specific antibody. The crosslinks are then reversed and the released DNA is sequenced. B) PICh is a method that aims to determine the total protein composition at a specific locus. Here, crosslinked chromatin fragments are recognized by short LNA capture probes with a defined sequence. These oligos are linked to

desthiobiotin, which is used to specifically capture LNA-chromatin hybrids. Finally, affinity purified proteins are identified via tandem MS. C) Contrary to ChIP and PiCh, BioID requires the expression of a POI fused to the promiscuous biotin ligase BirA\*. After the culture medium has been supplemented with exogenous biotin, BirA\* catalyzes the addition of biotin to proteins, which are in close proximity to the POI-BirA\* fusion protein. Lysis of the cells and denaturation of proteins is then followed by affinity purification of biotinylated peptides, which are analyzed via tandem MS.

### Proximity-dependent biotin identification (BioID)

Whereas both ChIP and PiCh rely on harsh crosslinking of chromatin to preserve protein-DNA and protein-protein interactions, BioID takes advantage of the biotin ligase BirA to mark proteins, which interact with, or are located in the vicinity of a POI (Figure 11C) [Roux et al., 2012]. In *Escherichia coli*, this DNA-binding biotin protein ligase regulates the biotinylation of a biotin acceptor tag (BAT) of an acetyl-CoA carboxylase and thereby acts as a transcriptional repressor for the biotin biosynthetic (*bio*) operon [Chapman-Smith & Cronan Jr, 1999]. The crystal structure of BirA revealed that the protein comprises three distinct domains (Figure 12A): upon binding to biotin, the C-terminal domain mediates the dimerization of BirA, whereas the N-terminal domain facilitates binding to the 40 bp *bio*-operator sequence [Streaker & Beckett, 1998; Chakravartty & Cronan, 2013]. The central domain is essential for the two-step biotinylation reaction (Figure 12B): first BirA binds biotin and ATP to synthesize biotinyl-AMP with the release of pyrophosphate (PP<sub>i</sub>). In a second step, a conserved lysine residue within the BAT acts as a nucleophile and attacks the mixed anhydride bond, resulting in AMP and a biotinylated protein. To circumvent the stringent target selectivity of BirA, a promiscuous biotin ligase is used for BioID. This R118G mutant BirA (BirA\*), which is defective in both self-association and DNA-binding, displays a lower affinity for biotinyl-AMP [Kwon & Beckett, 2000]. Importantly, BirA\* promiscuously biotinylates proteins at primary amines in a proximity dependent manner [Choi-Rhee et al., 2004; Cronan, 2005]. For BioID, a POI is fused to BirA\* and ectopically expressed. Addition of excess biotin to the culture medium leads to the selective biotinylation of proteins in close proximity of this fusion protein. Subsequently, the cells are lysed and proteins are stringently denatured. Finally, biotinylated proteins are subjected to affinity purification via streptavidin- or avidin-coupled beads and identified by MS.



**Figure 12: Crystal structure of BirA and biotin ligase reaction.** A) Crystal structure of *E. coli* BirA monomer. BirA is composed of three distinct domains. The C- (dark-green) and N-terminal (light-green) domains mediate dimerization and DNA-binding, respectively. The central catalytic domain (light-blue) mediates the formation of activated biotin (biotinyl-AMP, orange) (PDB: 2EWN). B) Two-step biotinylation reaction as catalyzed by BirA. First, BirA binds to biotin and ATP to synthesize biotinyl-AMP. In a second step, this intermediate reacts with the amine group of a lysine residue.

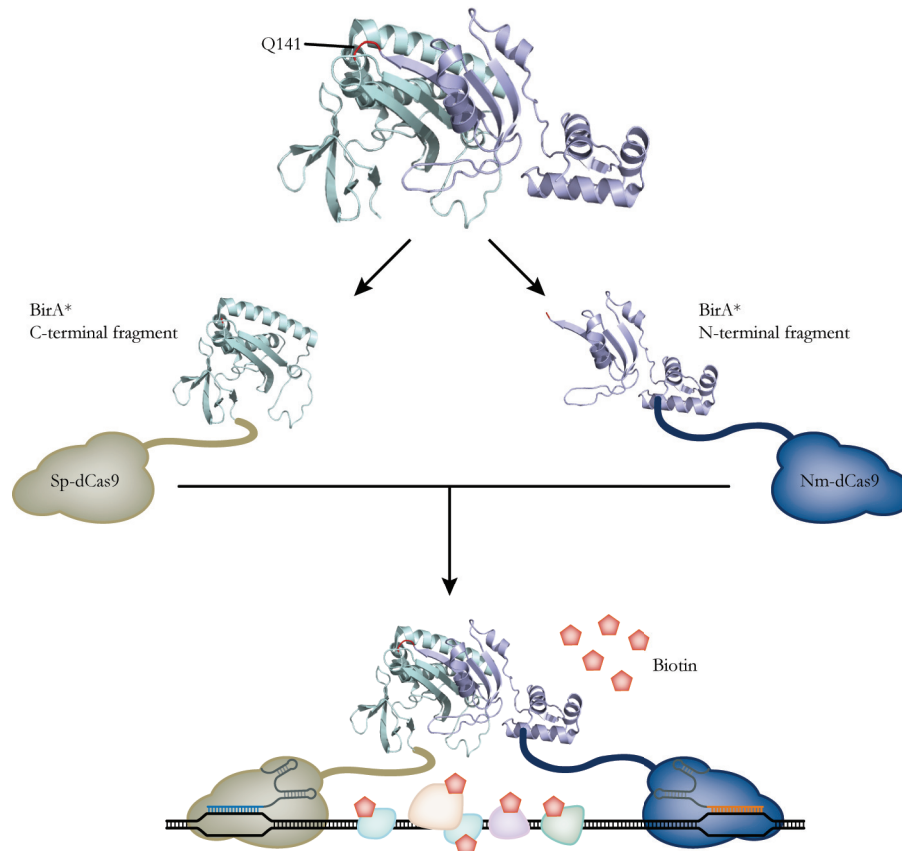
### CRISPR/Cas-mediated biotin identification (CasID)

BioID is a powerful tool to determine dynamic interactions of both DNA-bound and unbound proteins. Yet, the original design of this method renders it challenging to probe the protein composition of a defined genomic locus if interacting proteins are elusive. Hence, we combined the programmable DNA-binding of dCas9 with the promiscuous biotinylation activity of BirA\*.

To test the feasibility of our approach, we first targeted BirA\* to telomeres as they represent a genomic locus, where associated proteins are well-documented [Liu et al., 2004]. While we were able to identify the shelterin components TRF2, TIN2 and ACD, we did not detect additional proteins of the shelterin complex. One possible explanation for this could be that binding of dCas9 to telomeres leads to a sterical clash, which results in selective labeling or eviction of individual shelterin subunits.

We then expanded our method to investigate the local protein composition of MaS sequences. Besides known heterochromatic proteins such as MECP2 (methyl-CpG binding protein 2), SMCHD1 (structural maintenance of chromosomes flexible hinge domain containing 1) and HP1BP3 (heterochromatin protein 1 binding protein 3), we identified the previously uncharacterized protein ZNF512 (zinc finger protein 512). Life-cell imaging of an overexpressed ZNF512-eGFP fusion revealed that this protein remains associated with MaS sequences throughout the cell cycle. Although zinc finger proteins represent a major class of mammalian transcription factors [Mitchell & Tjian, 1989], this strong local confinement of ZNF512 at transcriptionally inactive mitotic chromosomes, suggests a structural role, rather

than transcriptional activation. Additionally, a recent study correlated down-regulation of ZNF512 with a poor survival rate of patients with lung adenocarcinoma [Bao et al., 2016]. Future research, however, is needed to elucidate the function of this protein.



**Figure 13: Principle and design of split-CasID.** Splitting *E. coli* BirA\* at Q141 results in two catalytically inactive fragments (light-green and light-blue). The N- and C-terminally fragment could be fused to Nm-dCas9 and Sp-dCas9, respectively. Target-recognition would result in complementation of the BirA\*-fragments, reconstituting the biotin ligase activity. This approach is likely to drastically reduce background biotinylation (PDB: 2EWN).

Finally, we employed our CasID approach to generate the first dataset that describes the protein environment of minor satellite sequences in myoblasts. Notably, we could detect the known centromere-associated proteins CENPC (centromere protein C) and PCM1 (pericentriolar material 1), reflecting the close proximity of centromeres and minor satellites. Moreover, comparison of minor- and major satellite associated proteins revealed only a small overlap, suggesting a distinct protein composition for either of these repetitive loci.

Future refinements of the CasID technique could benefit from the recent discovery of a biotin ligase from the thermophilic bacterium *Aquifex aeolicus* (BioID2) [Kim et al., 2016]. Although the biotin ligases from both *E. coli* and *A. aeolicus* share a similar overall structure, the latter lacks a DNA-binding domain and thus is substantially smaller. This difference in size was



shown to improve localization of proteins, fused to BioID2. Moreover, it was demonstrated that by modulating the length of the linker, connecting BioID2 and the POI, the radius, within which the biotinylation reaction occurs, can be adjusted. Hence, fine-tuning of CasID experiments is feasible.

Another critical improvement of CasID could be achieved by the utilization of a split-BirA\* variant (Figure 13) [De Munter et al., 2017]. Splitting BirA\* within a surface-loop at position Q141 of the catalytic domain results in two enzymatically inactive fragments. Importantly, it was previously demonstrated that forced heterodimerization of these fragments reconstitutes the biotinylation activity. Therefore, it is conceivable that orthogonal dCas9 proteins, each of them programmed to target consecutive sequences within the same locus and fused to either the N- or C-terminal fragment of split-BirA\*, could mediate biotinylation of target-associated proteins. Although the original design of CasID allows to eliminate background biotinylation from the dataset by subtracting all biotinylated proteins that were also identified in a control sample, high levels of unspecific biotinylation potentially mask signals from low-abundant target proteins. In this context, it is likely that a split-BirA\* drastically reduces unspecific biotinylation, resulting from unbound dCas9-BirA\* molecules.



---

## 4 ANNEX

### 4.1 References

Aldrup-Macdonald ME, Sullivan BA (2014) The past, present, and future of human centromere genomics. *Genes* **5**: 33-50

Alexandrov IA, Mitkevich SP, Yurov YB (1988) The phylogeny of human chromosome specific alpha satellites. *Chromosoma* **96**: 443-453

Anand R, Marmorstein R (2007) Structure and Mechanism of Lysine-specific Demethylase Enzymes. *Journal of Biological Chemistry* **282**: 35425-35429

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

Antão JM, Mason JM, Déjardin J, Kingston RE (2012) Protein Landscape at *Drosophila melanogaster* Telomere-Associated Sequence Repeats. *Molecular and Cellular Biology* **32**: 2170-2182

Arents G, Moudrianakis EN (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 11170-11174

Avery OT, Macleod CM, McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* **79**: 137-158

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG (2006) Chromatin signatures of pluripotent cell lines. *Nature cell biology* **8**: 532-538

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* **21**: 381-395

Bannister AJ, Schneider R, Kouzarides T (2002) Histone Methylation. *Cell* **109**: 801-806

Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T (2005) Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *The Journal of biological chemistry* **280**: 17732-17736

Bao L, Zhang Y, Wang J, Wang H, Dong N, Su X, Xu M, Wang X (2016) Variations of chromosome 2 gene expressions among patients with lung cancer or non-cancer. *Cell biology and toxicology* **32**: 419-435

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, Guillou F, Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**: 909

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**: 1709

Bernstein BE, Meissner A, Lander ES (2007) The Mammalian Epigenome. *Cell* **128**: 669-681

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326

Bickmore WA, Sumner AT (1989) Mammalian chromosome banding — an expression of genome organization. *Trends in Genetics* **5**: 144-148

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* **41**: 7429-7437

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**: 91-99

Bird AP (1995) Gene number, noise reduction and biological complexity. *Trends in genetics : TIG* **11**: 94-100

Biterge B, Richter F, Mittler G, Schneider R (2014) Methylation of histone H4 at aspartate 24 by protein L-isoaspartate O-methyltransferase (PCMT1) links histone modifications with protein homeostasis. *Scientific reports* **4**: 6674

Blander G, Guarente L (2004) The Sir2 family of protein deacetylases. *Annual review of biochemistry* **73**: 417-435

Blasco MA, Lee H-W, Hande MP, Samper E, Lansdorp PM, DePinho RA, Greider CW (1997) Telomere Shortening and Tumor Formation by Mouse Cells Lacking Telomerase RNA. *Cell* **91**: 25-34

Boch J, Bonas U (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* **48**: 419-436

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**: 1509-1512

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Muller S, Eils R, Cremer C, Speicher MR, Cremer T (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3**: e157

Bornfleth H, Edelmann P, Zink D, Cremer T, Cremer C (1999) Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophysical journal* **77**: 2871-2886

Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. *Annual review of microbiology* **62**: 235-251

Böttger A, Islam MS, Chowdhury R, Schofield CJ, Wolf A (2015) The oxygenase Jmjd6—a case study in conflicting assignments. *The Biochemical journal* **468**: 191-202

Branco MR, Pombo A (2006) Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLoS Biology* **4**: e138

Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsaesser J, Lahaye T, Leonhardt H (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research* **40**: 5368-5377

Buttner D, Bonas U (2010) Regulation and secretion of Xanthomonas virulence factors. *FEMS microbiology reviews* **34**: 107-133

- Byvoet P, Shepherd GR, Hardin JM, Noland BJ (1972) The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Archives of Biochemistry and Biophysics* **148**: 558-567
- Cabantous S, Terwilliger TC, Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature biotechnology* **23**: 102-107
- Caron H, Schaik Bv, Mee Mvd, Baas F, Riggins G, Sluis Pv, Hermus M-C, Asperen Rv, Boon K, Voûte PA, Heisterkamp S, Kampen Av, Versteeg R (2001) The Human Transcriptome Map: Clustering of Highly Expressed Genes in Chromosomal Domains. *Science* **291**: 1289
- Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22**: 3489-3496
- Cavalli G, Misteli T (2013) Functional implications of genome topology. *Nature structural & molecular biology* **20**: 290-299
- Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nature reviews Genetics* **10**: 295-304
- Chakravartty V, Cronan JE (2013) The Wing of a Winged Helix-Turn-Helix Transcription Factor Organizes the Active Site of BirA, a Bifunctional Repressor/Ligase. *The Journal of biological chemistry* **288**: 36029-36039
- Champagne KS, Kutateladze TG (2009) Structural insight into histone recognition by the ING PHD fingers. *Current drug targets* **10**: 432-441
- Chapman-Smith A, Cronan Jr JE (1999) Molecular Biology of Biotin Attachment to Proteins. *The Journal of Nutrition* **129**: 477S-484S
- Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**: 1479-1491
- Chen B, Guan J, Huang B (2016a) Imaging Specific Genomic DNA in Living Cells. *Annual review of biophysics* **45**: 1-23
- Chen B, Hu J, Almeida R, Liu H, Balakrishnan S, Covill-Cooke C, Lim WA, Huang B (2016b) Expanding the CRISPR imaging toolset with *Staphylococcus aureus* Cas9 for simultaneous imaging of multiple genomic loci. *Nucleic acids research* **44**: e75
- Chen Y, Varani G (2013) Engineering RNA-binding proteins for biology. *FEBS Journal* **280**: 3734-3754
- Cheng AW, Jillette N, Lee P, Plaskon D, Fujiwara Y, Wang W, Taghbalout A, Wang H (2016) Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell Res* **26**: 254-257
- Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivalila CS, Dadon DB, Jaenisch R (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res* **23**: 1163-1171
- Cheng X, Blumenthal RM (2008) Mammalian DNA methyltransferases: a structural perspective. *Structure (London, England : 1993)* **16**: 341-350
- Choi-Rhee E, Schulman H, Cronan JE (2004) Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein science : a publication of the Protein Society* **13**: 3043-3050

- Choo KH (1997) *The Centromere*, Oxford, UK: Oxford University Press.
- Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J (2016) CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget* **7**: 46545-46556
- Chow J, Heard E (2009) X inactivation and the complexities of silencing a sex chromosome. *Current Opinion in Cell Biology* **21**: 359-366
- Chuang C-H, Carpenter AE, Fuchsova B, Johnson T, de Lanerolle P, Belmont AS (2006) Long-Range Directional Movement of an Interphase Chromosome Site. *Current Biology* **16**: 825-831
- Chylinski K, Le Rhun A, Charpentier E (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology* **10**: 726-737
- Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annual review of biochemistry* **78**: 273-304
- Collas P (2010) The Current State of Chromatin Immunoprecipitation. *Molecular Biotechnology* **45**: 87-100
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, NY)* **339**: 819-823
- Costantino L, Koshland D (2015) The Yin and Yang of R-loop biology. *Curr Opin Cell Biol* **34**: 39-45
- Cremer M, von Hase J, Volm T, Brero A, Kreth G, Walter J, Fischer C, Solovei I, Cremer C, Cremer T (2001) Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **9**: 541-567
- Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews Genetics* **2**: 292-301
- Cremer T, Cremer C, Baumann H, Luedtke EK, Sperling K, Teuber V, Zorn C (1982) Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human genetics* **60**: 46-56
- Crick F (1970) Central dogma of molecular biology. *Nature* **227**: 561-563
- Crick FH (1958) On protein synthesis. *Symposia of the Society for Experimental Biology* **12**: 138-163
- Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* **145**: 1119-1131
- Cronan JE (2005) Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. *The Journal of nutritional biochemistry* **16**: 416-418
- Dahm R (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human genetics* **122**: 565-581
- Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi SW, Page DC, Jaenisch R (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**: 166-175

- 
- De Munter S, Gornemann J, Derua R, Lesage B, Qian J, Heroes E, Waelkens E, Van Eynde A, Beullens M, Bollen M (2017) Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS letters* **591**: 415-424
- DeFrancesco L (2011) Move over ZFNs. *Nat Biotech* **29**: 681-684
- Déjardin J, Kingston RE (2009) Purification of Proteins Associated with Specific Genomic Loci. *Cell* **136**: 175-186
- Dekker J (2014) Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture. *Epigenetics Chromatin* **7**: 25
- Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews Genetics* **14**: 390-403
- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607
- Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**: 720-723
- Dickinson DJ, Goldstein B (2016) CRISPR-Based Methods for *Caenorhabditis elegans* Genome Engineering. *Genetics* **202**: 885-901
- Dillon N, Festenstein R (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends in Genetics* **18**: 252-258
- Dingwall C, Laskey R (1992) The nuclear membrane. *Science* **258**: 942-947
- Dion V, Gasser Susan M (2013) Chromatin Movement in the Maintenance of Genome Stability. *Cell* **152**: 1355-1364
- Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, Diao Y, Liang J, Zhao H, Lobanenko VV, Ecker JR, Thomson JA, Ren B (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**: 331-336
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380
- Dobosy JR, Selker EU (2001) Emerging connections between DNA methylation and histone acetylation. *Cellular and Molecular Life Sciences CMLS* **58**: 721-727
- Duerre JA, Lee CT (1974) In vivo methylation and turnover of rat brain histones. *Journal of neurochemistry* **23**: 541-547
- Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**: 560-561
- Ehrlich M, Gama-Sosa MA, Huang L-H, Midgett RM, Kuo KC, McCune RA, Gehrke C (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic acids research* **10**: 2709-2721
- Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure (London, England : 1993)* **4**: 1171-1180

Epsztejn-Litman S, Feldman N, Abu-Remaileh M, Shufaro Y, Gerson A, Ueda J, Deplus R, Fuks F, Shinkai Y, Cedar H, Bergman Y (2008) De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature structural & molecular biology* **15**: 1176-1183

Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Meth* **10**: 1116-1121

Fatemi M, Hermann A, Gowher H, Jeltsch A (2002) Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. *European Journal of Biochemistry* **269**: 4981-4984

Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature cell biology* **8**: 188-194

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**: 398-402

Finch JT, Lutter LC, Rhodes D, Brown RS, Rushton B, Levitt M, Klug A (1977) Structure of nucleosome core particles of chromatin. *Nature* **269**: 29-36

Fraser P, Bickmore W (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**: 413-417

Fu Y, Rocha PP, Luo VM, Raviram R, Deng Y, Mazzoni EO, Skok JA (2016) CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nat Commun* **7**: 11707

Fuhrmann G, Chung AC, Jackson KJ, Hummelke G, Baniahmad A, Sutter J, Sylvester I, Scholer HR, Cooney AJ (2001) Mouse germline restriction of Oct4 expression by germ cell nuclear factor. *Dev Cell* **1**: 377-387

Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nature cell biology* **6**: 784-791

Gardiner K (1995) Human genome organization. *Current opinion in genetics & development* **5**: 315-322

Garside EL, Schellenberg MJ, Gesner EM, Bonanno JB, Sauder JM, Burley SK, Almo SC, Mehta G, MacMillan AM (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA (New York, NY)* **18**: 2020-2028

Gendrel A-V, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, Tang YA, Huntley D, Montana G, Taylor S, Giannoulatou E, Heard E, Stancheva I, Brockdorff N (2012) Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. *Developmental Cell* **23**: 265-279

Gersbach CA, Perez-Pinera P (2014) Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert Opinion on Therapeutic Targets* **18**: 835-839

Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. *Mol Cell* **49**: 773-782

Gidekel S, Bergman Y (2002) A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element. *The Journal of biological chemistry* **277**: 34521-34530

Gilbert Luke A, Larson Matthew H, Morsut L, Liu Z, Brar Gloria A, Torres Sandra E, Stern-Ginossar N, Brandman O, Whitehead Evan H, Doudna Jennifer A, Lim Wendell A, Weissman Jonathan S, Qi



- Lei S (2013) CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**: 442-451
- Goldman RD, Gruenbaum Y, Moir RD, Shumaker DK, Spann TP (2002) Nuclear lamins: building blocks of nuclear architecture. *Genes Dev* **16**: 533-547
- Gosden JR, Mitchell AR, Buckland RA, Clayton RP, Evans HJ (1975) The location of four human satellite DNAs on human chromosomes. *Experimental cell research* **92**: 148-158
- Grewal SIS, Jia S (2007) Heterochromatin revisited. *Nature reviews Genetics* **8**: 35-46
- Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T (1999) Mammalian Telomeres End in a Large Duplex Loop. *Cell* **97**: 503-514
- Grissa I, Vergnaud G, Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research* **35**: W52-W57
- Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948-951
- Guenatri M, Bailly D, Maison C, Almouzni G (2004) Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *The Journal of Cell Biology* **166**: 493-505
- Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130**: 77-88
- Guo Junjie U, Su Y, Zhong C, Ming G-I, Song H (2011) Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell* **145**: 423-434
- Haeussler M, Concordet J-P (2016) Genome Editing with CRISPR-Cas9: Can It Get Any Better? *Journal of Genetics and Genomics* **43**: 239-250
- Handa V, Jeltsch A (2005) Profound Flanking Sequence Preference of Dnmt3a and Dnmt3b Mammalian DNA Methyltransferases Shape the Human Epigenome. *Journal of Molecular Biology* **348**: 1103-1112
- Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B (2010) Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. *Cell Stem Cell* **6**: 479-491
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**: 1303-1307
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311-318
- Heitz E (1928) Das Heterochromatin der Moose. In *Jahrbücher für wissenschaftliche Botanik*, pp 762-818.
- Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**: 199-202

- Helma J, Cardoso MC, Muyldermans S, Leonhardt H (2015) Nanobodies and recombinant binders in cell biology. *The Journal of Cell Biology* **209**: 633
- Hemmerich P, Schmiedeberg L, Diekmann S (2011) Dynamic as well as stable protein interactions contribute to genome function and maintenance. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **19**: 131-151
- Henao-Mejia J, Williams A, Rongvaux A, Stein J, Hughes C, Flavell RA (2016) Generation of Genetically Modified Mice Using the CRISPR-Cas9 Genome-Editing System. *Cold Spring Harb Protoc* **2016**: pdb.prot090704
- Hermann A, Goyal R, Jeltsch A (2004) The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. *Journal of Biological Chemistry* **279**: 48350-48359
- Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology* **36**: 39-56
- Hewitt SL, High FA, Reiner SL, Fisher AG, Merkenschlager M (2004) Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *European journal of immunology* **34**: 3604-3613
- Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotech* **33**: 510-517
- Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schubeler D, Gilbert DM (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* **6**: e245
- Hirsch JD, Eslamizar L, Filanoski BJ, Malekzadeh N, Haugland RP, Beechem JM, Haugland RP (2002) Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Analytical Biochemistry* **308**: 343-357
- Horlbeck MA, Witkowsky LB, Guglielmi B, Replogle JM, Gilbert LA, Villalta JE, Torigoe SE, Tjian R, Weissman JS (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**: e12677
- Hou C, Li L, Qin ZS, Corces VG (2012) Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**: 471-484
- Housden BE, Perrimon N (2016) Cas9-Mediated Genome Engineering in Drosophila melanogaster. *Cold Spring Harb Protoc* **2016**: pdb.top086843
- Hrle A, Su AA, Ebert J, Benda C, Randau L, Conti E (2013) Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA biology* **10**: 1670-1678
- Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, Yoshida M, Wang X-F, Yao T-P (2002) HDAC6 is a microtubule-associated deacetylase. *Nature* **417**: 455-458
- Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, Lee JC, Hayes DN, Shanmugam KS, Bhattacharjee A, Biondi CA, Kay GF, Hayward NK, Hess JL, Meyerson M (2004) Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* **13**: 587-597
- Ide S, Dejardin J (2015) End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nature Communications* **6**: 6674

- Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J, Mirny LA (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods* **9**: 999-1003
- Isaac RS, Jiang F, Doudna JA, Lim WA, Narlikar GJ, Almeida R (2016) Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* **5**: e13450
- Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology* **169**: 5429-5433
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* **333**: 1300
- Jackson V (1978) Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. *Cell* **15**: 945-954
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33**: 245-254
- Jegou T, Chung I, Heuvelman G, Wachsmuth M, Görisch SM, Greulich-Bode KM, Boukamp P, Lichter P, Rippe K (2009) Dynamics of Telomeres and Promyelocytic Leukemia Nuclear Bodies in a Telomerase-negative Human Cell Line. *Molecular biology of the cell* **20**: 2070-2082
- Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**: 248-251
- Jiang W, Samai P, Marraffini LA (2016) Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**: 710-721
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821
- Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**: 1247997
- Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* **18**: 529-536
- Jurkowska RZ, Jurkowski TP, Jeltsch A (2011) Structure and function of mammalian DNA methyltransferases. *Chembiochem : a European journal of chemical biology* **12**: 206-222
- Kabadi AM, Gersbach CA (2014) Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods (San Diego, Calif)* **69**: 188-197
- Kadauke S, Blobel GA (2013) Mitotic bookmarking by transcription factors. *Epigenetics Chromatin* **6**: 6
- Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature biotechnology* **30**: 90-98
- Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, Ishikawa H, Leonetti MD, Marshall WF, Weissman JS, Huang B (2016) Versatile protein tagging in cells with split fluorescent protein. *Nature Communications* **7**: 11046

- Kanda T, Sullivan KF, Wahl GM (1998) Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Current Biology* **8**: 377-385
- Karytinis A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, Mattevi A (2009) A novel mammalian flavin-dependent histone demethylase. *The Journal of biological chemistry* **284**: 17775-17782
- Kay S, Hahn S, Marois E, Hause G, Bonas U (2007) A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* **318**: 648-651
- Kearns NA, Genga RMJ, Enuameh MS, Garber M, Wolfe SA, Maehr R (2013) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141**: 219
- Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Meth* **12**: 401-403
- Kiefer CM, Hou C, Little JA, Dean A (2008) Epigenetics of beta-globin gene regulation. *Mutation research* **647**: 68-76
- Kim DI, Jensen SC, Noble KA, KC B, Roux KH, Motamedchaboki K, Roux KJ (2016) An improved smaller biotin ligase for BioID proximity labeling. *Molecular biology of the cell* **27**: 1188-1196
- Kim GD, Ni J, Kelesoglu N, Roberts RJ, Pradhan S (2002) Co - operation and communication between the human maintenance and de novo DNA (cytosine - 5) methyltransferases. *The EMBO journal* **21**: 4183
- Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO reports* **7**: 397-403
- Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA, van Steensel B (2013) Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**: 178-192
- Kipling D, Ackford HE, Taylor BA, Cooke HJ (1991) Mouse minor satellite DNA genetically maps to the centromere and is physically linked to the proximal telomere. *Genomics* **11**: 235-241
- Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**: 357-369
- Knight SC, Xie L, Deng W, Guglielmi B, Witkowsky LB, Bosanac L, Zhang ET, El Beheiry M, Masson JB, Dahan M, Liu Z, Doudna JA, Tjian R (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**: 823-826
- Konermann S, Brigham MD, Trevino A, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F (2013) Optical Control of Mammalian Endogenous Transcription and Epigenetic States. *Nature* **500**: 472-476
- Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**: 391-400
- Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H (2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**: 158-162
- Kouzarides T (2007) Chromatin Modifications and Their Function. *Cell* **128**: 693-705
- Krawczyk PM, Borovski T, Stap J, Cijssouw T, Cate Rt, Medema JP, Kanaar R, Franken NAP, Aten JA (2012) Chromatin mobility is increased at sites of DNA double-strand breaks. *Journal of Cell Science* **125**: 2127

- Kupper K, Kolbl A, Biener D, Dittrich S, von Hase J, Thormeyer T, Fiegler H, Carter NP, Speicher MR, Cremer T, Cremer M (2007) Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* **116**: 285-306
- Kwon K, Beckett D (2000) Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein science : a publication of the Protein Society* **9**: 1530-1539
- Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol* **11**: 39-46
- Lanctot C, Cheutin T, Cremer M, Cavalli G, Cremer T (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature reviews Genetics* **8**: 104-115
- Lane Andrew B, Strzelecka M, Ettinger A, Grenfell Andrew W, Wittmann T, Heald R (2015) Enzymatically Generated CRISPR Libraries for Genome Labeling and Screening. *Developmental Cell* **34**: 373-378
- Laurent L, Wong E, Li G, Huynh T, Tsigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL (2010) Dynamic changes in the human methylome during differentiation. *Genome Res* **20**: 320-331
- Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. *Nature* **366**: 362-365
- Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC (1988) Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human genetics* **80**: 224-234
- Lieber MR (2010) The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annual review of biochemistry* **79**: 181-211
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289-293
- Lindhout BI, Fransz P, Tessadori F, Meckel T, Hooykaas PJ, van der Zaal BJ (2007) Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic acids research* **35**: e107
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315-322
- Liu D, O'Connor MS, Qin J, Songyang Z (2004) Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *The Journal of biological chemistry* **279**: 51338-51342
- Liu Q, Segal DJ, Ghiara JB, Barbas CF, 3rd (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 5525-5530
- Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R (2016) Editing DNA Methylation in the Mammalian Genome. *Cell* **167**: 233-247.e217

Lombardi PM, Cole KE, Dowling DP, Christianson DW (2011) Structure, Mechanism, and Inhibition of Histone Deacetylases and Related Metalloenzymes. *Current opinion in structural biology* **21**: 735-743

Lopes Novo C, Rugg-Gunn P (2016) Crosstalk between pluripotency factors and higher-order chromatin organization. *Nucleus* **7**: 447-452

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8[thinsp]Å resolution. *Nature* **389**: 251-260

Lunyak VV, Prefontaine GG, Nunez E, Cramer T, Ju BG, Ohgi KA, Hutt K, Roy R, Garcia-Diaz A, Zhu X, Yung Y, Montoliu L, Glass CK, Rosenfeld MG (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**: 248-251

Luo J, Su F, Chen D, Shiloh A, Gu W (2000) Deacetylation of p53 modulates its effect on cell growth and apoptosis. *Nature* **408**: 377-381

Luo RX, Dean DC (1999) Chromatin Remodeling and Transcriptional Regulation. *Journal of the National Cancer Institute* **91**: 1288-1294

Ma H, Naseri A, Reyes-Gutierrez P, Wolfe SA, Zhang S, Pederson T (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 3002-3007

Ma H, Reyes-Gutierrez P, Pederson T (2013) Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 21048-21053

Mahfouz MM, Li L, Piatek M, Fang X, Mansour H, Bangarusamy DK, Zhu J-K (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Molecular Biology* **78**: 311-321

Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**: 716-719

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826

Manders EM, Kimura H, Cook PR (1999) Direct imaging of DNA in living cells reveals the dynamics of chromosome formation. *J Cell Biol* **144**: 813-821

Manuelidis L (1978) Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma* **66**: 23-32

Martin RM, Leonhardt H, Cardoso MC (2005) DNA labeling in living cells. *Cytometry Part A : the journal of the International Society for Analytical Cytology* **67**: 45-52

Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP (2003) The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in biochemical sciences* **28**: 69-74

Mayer R, Brero A, von Hase J, Schroeder T, Cremer T, Dietzel S (2005) Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC cell biology* **6**: 44

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000) Embryogenesis: Demethylation of the zygotic paternal genome. *Nature* **403**: 501-502

- McClintock B (1938) The fusion of broken ends of sister half-chromatids following chromatid breakage at meiotic anaphase. *Mo Agric Exp Stn Res Bull* **290**: 1-48
- McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR, Challen GA (2016) Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biology Open* **5**: 866
- McGhee JD, Ginder GD (1979) Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* **280**: 419-420
- Meaburn KJ, Misteli T (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *The Journal of Cell Biology* **180**: 39
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766-770
- Meshorer E, Misteli T (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews Molecular cell biology* **7**: 540-546
- Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, Reinders M, Wessels L, van Steensel B (2013) Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23**: 270-280
- Meyne J, Ratliff RL, Moyzis RK (1989) Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 7049-7053
- Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *The EMBO journal* **4**: 1609-1614
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotech* **29**: 143-148
- Misteli T (2007) Beyond the Sequence: Cellular Organization of Genome Function. *Cell* **128**: 787-800
- Mitchell AR, Gosden JR, Miller DA (1985) A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma* **92**: 369-377
- Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371-378
- Miyanari Y, Ziegler-Birling C, Torres-Padilla M-E (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nature structural & molecular biology* **20**: 1321-1324
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**: 755-766
- Morbitzer R, Elsaesser J, Hausner J, Lahaye T (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic acids research* **39**: 5790-5799
- Morocho AM, Karamyshev V, Shcherbinina O, Polushin N (2005) Biotin-labeled oligonucleotides with extraordinarily long tethering arms. *Methods in molecular biology* **288**: 225-240

- Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* **326**: 1501
- Mulholland CB, Smets M, Schmidtman E, Leidescher S, Markaki Y, Hofweber M, Qin W, Manzo M, Kremmer E, Thanisch K, Bauer C, Rombaut P, Herzog F, Leonhardt H, Bultmann S (2015) A modular open platform for systematic functional studies under physiological conditions. *Nucleic acids research* **43**: e112
- Müller U, Bauer C, Siegl M, Rottach A, Leonhardt H (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic acids research* **42**: 8592-8604
- Musselman CA, Lalonde M-E, Côté J, Kutateladze TG (2012) Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology* **19**: 1218-1227
- Mussolino C, Morbitzer R, Lütge F, Dannemann N, Lahaye T, Cathomen T (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research* **39**: 9283-9293
- Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, Tanaka S, Shiota K, Nakano T (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature cell biology* **9**: 64-71
- Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T (2012) PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**: 415-419
- Nam KH, Haitjema C, Liu X, Ding F, Wang H, DeLisa MP, Ke A (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure (London, England : 1993)* **20**: 1574-1584
- Nelles David A, Fang Mark Y, O'Connell Mitchell R, Xu Jia L, Markmiller Sebastian J, Doudna Jennifer A, Yeo Gene W (2016) Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* **165**: 488-496
- Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, Solovei I, Cremer T, Dopazo J, Langst G (2010) Initial genomics of the human nucleolus. *PLoS genetics* **6**: e1000889
- Nishida H, Suzuki T, Kondo S, Miura H, Fujimura Y, Hayashizaki Y (2006) Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **14**: 203-211
- Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata S, Dohmae N, Ishitani R, Zhang F, Nureki O (2014) Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**: 935-949
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Bluthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381-385
- Norris DP, Brockdorff N, Rastan S (1991) Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mammalian genome : official journal of the International Mammalian Genome Society* **1**: 78-83
- Nunez JK, Lee AS, Engelman A, Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**: 193-198



- 
- O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**: 263-266
- O'Sullivan RJ, Karlseder J (2010) Telomeres: protecting chromosomes against genome instability. *Nature reviews Molecular cell biology* **11**: 171-181
- Okano M, Bell DW, Haber DA, Li E (1999) DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**: 247-257
- Olins AL, Olins DE (1974) Spheroid Chromatin Units ( $\nu$  Bodies). *Science* **183**: 330
- Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**: 714-717
- Oudet P, Gross-Bellard M, Chambon P (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**: 281-300
- Ozsolak F, Song JS, Liu XS, Fisher DE (2007) High-throughput mapping of the chromatin structure of human promoters. *Nat Biotech* **25**: 244-248
- Parthun MR (2007) Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**: 5319-5328
- Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**: 394-397
- Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**: 809-817
- Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* **10**: 973-976
- Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Graf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**: 603-613
- Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie (International ed in English)* **50**: 7008-7012
- Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**: 1281-1295
- Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B (2006) Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nat Genet* **38**: 1005-1014
- Pietras DF, Bennett KL, Siracusa LD, Woodworth-Gutai M, Chapman VM, Gross KW, Kane-Haas C, Hastie ND (1983) Construction of a small *Mus musculus* repetitive DNA library: identification of a new satellite sequence in *Mus musculus*. *Nucleic acids research* **11**: 6965-6983

- Plath K, Talbot D, Hamer KM, Otte AP, Yang TP, Jaenisch R, Panning B (2004) Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *J Cell Biol* **167**: 1025-1035
- Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC (1995) The Centromere: Hub of Chromosomal Activities. *Science* **270**: 1591
- Prokocimer M, Davidovich M, Nissim-Rafinia M, Wiesel-Motiuk N, Bar DZ, Barkan R, Meshorer E, Gruenbaum Y (2009) Nuclear lamins: key regulators of nuclear structure and activities. *Journal of cellular and molecular medicine* **13**: 1059-1085
- Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173-1183
- Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, Haussler D, Kamakaka RT (2011) Human tRNA genes function as chromatin insulators. *The EMBO journal* **31**: 330
- Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA, Cathomen T, Voytas DF, Joung JK (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature methods* **5**: 374-375
- Ratner HK, Sampson TR, Weiss DS (2016) Overview of CRISPR–Cas9 Biology. *Cold Spring Harbor Protocols* **2016**: pdb.top088849
- 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) Visualization of aging-associated chromatin alterations with an engineered TALE system. *Cell Res*
- Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology* **30**: 460-465
- Richman R, Chicoine LG, Collini MP, Cook RG, Allis CD (1988) Micronuclei and the cytoplasm of growing Tetrahymena contain a histone acetylase activity which is highly specific for free histone H4. *J Cell Biol* **106**: 1017-1026
- Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* **423**: 145-150
- Ringrose L, Paro R (2004) Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annual review of genetics* **38**: 413-443
- Robinett CC, Straight A, Li G, Wilhelm C, Sudlow G, Murray A, Belmont AS (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol* **135**: 1685-1700
- Romer T, Leonhardt H, Rothbauer U (2011) Engineering antibodies and proteins for molecular in vivo imaging. *Current opinion in biotechnology* **22**: 882-887
- Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, Backmann N, Conrath K, Muyltermans S, Cardoso MC, Leonhardt H (2006) Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nature methods* **3**: 887-889
- Rottach A, Leonhardt H, Spada F (2009) DNA methylation-mediated epigenetic control. *Journal of cellular biochemistry* **108**: 43-51
- Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T (2013) Spatial Dynamics of Chromosome Translocations in Living Cells. *Science* **341**: 660

- Roux KJ, Kim DI, Raida M, Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology* **196**: 801
- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LS, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJ, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vossball LB, Zhang J, Zhao Q, Zheng XH, Lewis S (2000) Comparative genomics of the eukaryotes. *Science* **287**: 2204-2215
- Ruthenburg AJ, Li H, Patel DJ, Allis CD (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nature reviews Molecular cell biology* **8**: 983-994
- Saksouk N, Barth Teresa K, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla M-E, Imhof A, Déjardin J (2014) Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Molecular Cell* **56**: 580-594
- Saksouk N, Simboeck E, Déjardin J (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin* **8**: 3
- Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA (2015) Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**: 1164-1174
- Samson JE, Magadan AH, Sabri M, Moineau S (2013) Revenge of the phages: defeating bacterial defences. *Nat Rev Micro* **11**: 675-687
- Sanchez R, Zhou M-M (2009) The role of human bromodomains in chromatin biology and gene transcription. *Current opinion in drug discovery & development* **12**: 659-665
- Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences* **103**: 1412-1417
- Schalch T, Duda S, Sargent DF, Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**: 138-141
- Schär P, Fritsch O (2011) DNA repair and the control of DNA methylation. *Progress in drug research Fortschritte der Arzneimittelforschung Progres des recherches pharmaceutiques* **67**: 51-68
- Schermelleh L, Solovei I, Zink D, Cremer T (2001) Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **9**: 77-80
- Scheuermann MO, Tajbakhsh J, Kurz A, Saracoglu K, Eils R, Lichter P (2004) Topology of genes and nontranscribed sequences in human interphase nuclei. *Experimental cell research* **301**: 266-279
- Schmid-Burgk JL, Schmidt T, Kaiser V, Honing K, Hornung V (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotech* **31**: 76-81
- Scholze H, Boch J (2010) TAL effector-DNA specificity. *Virulence* **1**: 428-432

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development* **18**: 1251-1262

Schueler MG, Dunn JM, Bird CP, Ross MT, Viggiano L, Rocchi M, Willard HF, Green ED (2005) Progressive proximal expansion of the primate X chromosome centromere. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 10563-10568

Schwartz YB, Pirrotta V (2008) Polycomb complexes and epigenetic states. *Current Opinion in Cell Biology* **20**: 266-273

Segal DJ, Dreier B, Beerli RR, Barbas CF, 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 2758-2763

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**: 458-472

Shelby RD, Hahn KM, Sullivan KF (1996) Dynamic elastic behavior of alpha-satellite DNA domains visualized in situ in living human cells. *The Journal of Cell Biology* **135**: 545

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941-953

Shiama N (1997) The p300/CBP family: integrating signals with transcription factors and chromatin. *Trends in Cell Biology* **7**: 230-236

Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74-79

Simon M, North JA, Shimko JC, Forties RA, Ferdinand MB, Manohar M, Zhang M, Fishel R, Ottesen JJ, Poirier MG (2011) Histone fold modifications control nucleosome unwrapping and disassembly. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 12711-12716

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**: 1348-1354

Soutoglou E, Misteli T (2007) Mobility and immobility of chromatin in transcription and genome stability. *Current opinion in genetics & development* **17**: 435-442

Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJ, Yonekura K, van der Oost J, Shinkai A (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* **52**: 135-145

Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K, Sakamoto K, Suzuki T, Dohmae N, Yokoyama S, Schaap PJ, Urlaub H, Heck AJ, Nogales E, Doudna JA, Shinkai A, van der Oost J (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell* **56**: 518-530

- Stansel RM, de Lange T, Griffith JD (2001) T - loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *The EMBO journal* **20**: 5532-5540
- Stern A, Keren L, Wurtzel O, Amitai G, Sorek R (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in genetics : TIG* **26**: 335-340
- Sternberg SH, LaFrance B, Kaplan M, Doudna JA (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**: 110-113
- Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**: 62-67
- Sterner R, Vidali G, Allfrey VG (1979) Studies of acetylation and deacetylation in high mobility group proteins. Identification of the sites of acetylation in HMG-1. *Journal of Biological Chemistry* **254**: 11577-11583
- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403**: 41-45
- Streaker ED, Beckett D (1998) A map of the biotin repressor-biotin operator interface: binding of a winged helix-turn-helix protein dimer to a forty base-pair site. *J Mol Biol* **278**: 787-800
- Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SI (2005) RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 152-157
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews Genetics* **9**: 465-476
- Suzuki MM, Kerr AR, De Sousa D, Bird A (2007) CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* **17**: 625-631
- Sved J, Bird A (1990) The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 4692-4696
- Szurek B, Rossier O, Hause G, Bonas U (2002) Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. *Molecular microbiology* **46**: 13-23
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935
- Tamaru H, Zhang X, McMillen D, Singh PB, Nakayama J, Grewal SI, Allis CD, Cheng X, Selker EU (2003) Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nat Genet* **34**: 75-79
- Tanenbaum Marvin E, Gilbert Luke A, Qi Lei S, Weissman Jonathan S, Vale Ronald D (2014) A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**: 635-646
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology* **14**: 1025-1040
- Taylor GC, Eskeland R, Hekimoglu-Balkan B, Pradeepa MM, Bickmore WA (2013) H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res* **23**: 2053-2065

- Tessarz P, Santos-Rosa H, Robson SC, Sylvestersen KB, Nelson CJ, Nielsen ML, Kouzarides T (2014) Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**: 564-568
- Thanisch K, Schneider K, Morbitzer R, Solovei I, Lahaye T, Bultmann S, Leonhardt H (2014) Targeting and tracing of specific DNA sequences with dTALEs in living cells. *Nucleic acids research* **42**: e38-e38
- Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082-1086
- Tremethick DJ (2007) Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell* **128**: 651-654
- Trojer P, Reinberg D (2007) Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* **28**: 1-13
- Tsirigos A, Rigoutsos I (2009) Alu and B1 Repeats Have Been Selectively Retained in the Upstream and Intronic Regions of Genes of Specific Functional Classes. *PLOS Computational Biology* **5**: e1000610
- Tumbar T, Belmont AS (2001) Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. *Nature cell biology* **3**: 134-139
- Tupler R, Perini G, Green MR (2001) Expressing the human genome. *Nature* **409**: 832-833
- Van den Ackerveken G, Marois E, Bonas U (1996) Recognition of the Bacterial Avirulence Protein AvrBs3 Occurs inside the Host Plant Cell. *Cell* **87**: 1307-1316
- van Koningsbruggen S, Gierlinski M, Schofield P, Martin D, Barton GJ, Ariyurek Y, den Dunnen JT, Lamond AI (2010) High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Molecular biology of the cell* **21**: 3735-3748
- Vaquero A, Scher M, Lee D, Erdjument-Bromage H, Tempst P, Reinberg D (2004) Human SirT1 Interacts with Histone H1 and Promotes Formation of Facultative Heterochromatin. *Molecular Cell* **16**: 93-105
- Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ, Myers RM (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* **23**: 555-567
- Vassetzky NS, Ten OA, Kramerov DA (2003) B1 and related SINEs in mammalian genomes. *Gene* **319**: 149-160
- Vazquez J, Belmont AS, Sedat JW (2001) Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Current biology : CB* **11**: 1227-1239
- Vejnar CE, Moreno-Mateos MA, Cifuentes D, Bazzini AA, Giraldez AJ (2016) Optimized CRISPR-Cas9 System for Genome Editing in Zebrafish. *Cold Spring Harb Protoc* **2016**: pdb.prot086850
- Vester B, Wengel J (2004) LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary RNA and DNA. *Biochemistry* **43**: 13233-13241
- Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**: 871-874

- Vissel B, Choo KH (1989) Mouse major ( $\gamma$ ) satellite DNA is highly conserved and organized into extremely long tandem arrays: Implications for recombination between nonhomologous chromosomes. *Genomics* **5**: 407-414
- Vojta A, Dobrinić P, Tadić V, Bočkor L, Korać P, Julg B, Klasić M, Zoldoš V (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic acids research* **44**: 5615-5628
- Voss TC, Hager GL (2008) Visualizing chromatin dynamics in intact cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783**: 2044-2051
- Waddington CH (1942) The Epigenotype. *Endeavour* **1**: 18-20
- Waddington CH (1957) *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin
- Wallrath LL, Elgin SC (1995) Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes & Development* **9**: 1263-1277
- Walport LJ, Hopkinson RJ, Chowdhury R, Schiller R, Ge W, Kawamura A, Schofield CJ (2016) Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat Commun* **7**: 11974
- Walter J, Schermelleh L, Cremer M, Tashiro S, Cremer T (2003) Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J Cell Biol* **160**: 685-697
- Wang S, Su JH, Zhang F, Zhuang X (2016) An RNA-aptamer-based two-color CRISPR labeling system. *Scientific reports* **6**: 26857
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**: 897-903
- Watson JD, Crick FH (1953a) Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**: 964-967
- Watson JD, Crick FH (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**: 737-738
- Weber AR, Krawczyk C, Robertson AB, Kuśnierczyk A, Vågbo CB, Schuermann D, Klungland A, Schär P (2016) Biochemical reconstitution of TET1 - TDG - BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications* **7**: 10806
- Wei Y, Terns RM, Terns MP (2015) Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev* **29**: 356-361
- Weiss M (2008) Probing the interior of living cells with fluorescence correlation spectroscopy. *Annals of the New York Academy of Sciences* **1130**: 21-27
- Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJ, van der Oost J (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**: 595-605

- Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M, Shi Y (2006) Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125**: 467-481
- Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338
- Wilbanks EG, Facciotti MT (2010) Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS one* **5**: e11471
- Willard HF (1985) Chromosome-specific organization of human alpha satellite DNA. *American Journal of Human Genetics* **37**: 524-532
- Williams RRE, Azuara V, Perry P, Sauer S, Dvorkina M, Jørgensen H, Roix J, McQueen P, Misteli T, Merkenschlager M, Fisher AG (2005) Neural induction promotes large-scale chromatin reorganization of the *Mash1* locus. *Journal of Cell Science* **119**: 132
- Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes & Development* **28**: 2778-2791
- Wolf SF, Jolly DJ, Lunnen KD, Friedmann T, Migeon BR (1984) Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proceedings of the National Academy of Sciences of the United States of America* **81**: 2806-2810
- Wong AKC, Rattner JB (1988) Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic acids research* **16**: 11645-11661
- Woodcock CL (2005) A milestone in the odyssey of higher-order chromatin structure. *Nature structural & molecular biology* **12**: 639-640
- Woodcock CL, Dimitrov S (2001) Higher-order structure of chromatin and chromosomes. *Current opinion in genetics & development* **11**: 130-135
- Woodcock CL, Ghosh RP (2010) Chromatin Higher-order Structure and Dynamics. *Cold Spring Harbor Perspectives in Biology* **2**
- Woodcock DM, Lawler CB, Linsenmeyer ME, Doherty JP, Warren WD (1997) Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *The Journal of biological chemistry* **272**: 7810-7816
- Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell S, Taylor IA, Blackburn GM, Gambelin SJ (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**: 652-656
- Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, Ruan K, Wang F, Xu G-l, Hu R (2016) A CRISPR-based approach for targeted DNA demethylation. *Cell Discovery* **2**: 16009
- Yang X, Han H, De Carvalho Daniel D, Lay Fides D, Jones Peter A, Liang G (2014) Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell* **26**: 577-590
- Yang X-J, Seto E (2003) Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Current opinion in genetics & development* **13**: 143-153
- Yang X-J, Seto E (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nature reviews Molecular cell biology* **9**: 206-218



- Yang XJ, Seto E (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* **26**: 5310-5318
- Yang Y, Lu Y, Espejo A, Wu J, Xu W, Liang S, Bedford MT (2010) TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell* **40**: 1016-1023
- Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, Kieft R, Christodoulou E, Perrakis A, Simmons JM, Hausinger RP, van Luenen HG, Rigden DJ, Sabatini R, Borst P (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic acids research* **35**: 2107-2115
- Yuan K, Shermoen AW, O'Farrell PH (2014) Illuminating DNA replication during *Drosophila* development using TALE-lights. *Current biology : CB* **24**: R144-145
- Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou M-M (2010) Mechanism and Regulation of Acetylated Histone Binding by the Tandem PHD Finger of DPF3b. *Nature* **466**: 258-262
- Zhang F, Cong L, Lodato S, Kosuri S, Church G, Arlotta P (2011) Programmable Sequence-Specific Transcriptional Regulation of Mammalian Genome Using Designer TAL Effectors. *Nature biotechnology* **29**: 149-153
- Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C (2012a) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature chemical biology* **8**: 328-330
- Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J (2012b) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**: 908-921
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750-756
- Zhu W, Yang B, Chittoor JM, Johnson LB, White FF (1998) AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Molecular plant-microbe interactions : MPMI* **11**: 824-832
- Zink D, Amaral MD, Englmann A, Lang S, Clarke LA, Rudolph C, Alt F, Luther K, Braz C, Sadoni N, Rosenecker J, Schindelbauer D (2004) Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *The Journal of Cell Biology* **166**: 815

## 4.2 Abbreviations

2-OG	2-oxoglutarate
3C	Chromosome conformation capture
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ac	Acetylation
AID	Activation-induced deaminase
APOBEC	Apolipoprotein B mRNA-editing enzyme complex
ATP	Adenosine triphosphate
BAT	Biotin acceptor tag
BER	Base excision repair
BioID	Proximity-dependent biotin identification
BirA	Bifunctional ligase/repressor
bp	Base pair
Cas	CRISPR associated
CasID	CRISPR/Cas-mediated biotin identification
CC	Chromocenter
CD	Chromosomal domain
CGI	CpG island
cHC	Constitutive heterochromatin
ChIP	Chromatin immunoprecipitation
cLAD	Constitutive LAD
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
CT	Chromosome territory
Dam	DNA adenine methyltransferase
dCas9	Catalytically dead Cas9
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DSB	DNA double-strand break
eGFP	Enhanced GFP
ESC	Embryonic stem cell

---

FCS	Fluorescence correlation spectroscopy
fHC	Facultative heterochromatin
FISH	Fluorescence <i>in situ</i> hybridization
fLAD	Facultative LAD
FLIP	Fluorescence loss in photobleaching
FP	Fluorescent protein
FRAP	Fluorescence recovery after photobleaching
GBP	GFP-binding protein
GFP	Green fluorescent protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDR	Homology-directed repair
HKMT	Histone lysine methyltransferase
HMT	Histone methyltransferase
HOR	Higher order repeat
HP1	Heterochromatin protein 1
KDM	Lysine demethylase
LAD	Lamina-associated domain
LINE	Long interspersed nuclear element
MaS	Major Satellite
Mb	Megabase
me	Methylation
MiS	Minor satellite
MS	Mass spectrometry
NAD	Nucleolar-associated domain
NE	Nuclear envelope
NHEJ	Non-homologous end joining
NL	Nuclear lamina
NLS	Nuclear localization signal
NUC	Nuclease lobe
PAM	Protospacer-adjacent motif
PBS	PUF-binding site
PI	PAM-interacting domain
PiCh	Proteomics of isolated chromatin segments
PRC	Polycomb repressive complex

PRMT	Histone arginine methyltransferase
PTM	Posttranslational modification
PUF	Pumilio
PZF	Polydactyl ZnF protein
REC	Recognition lobe
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
RVD	Repeat variable diresidue
SAM	S-adenosyl-L-methionine
scRNA	Scaffold RNA
sfGFP	Super-folder GFP
sgRNA	Single guide RNA
SINE	Short interspersed nuclear element
SunTag	SUPerNova tag
TAD	Topologically-associated domain
TALE	Transcription activator-like effector
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TF	Transcription factor
tracrRNA	Trans-activating crRNA
TRF1	Telomeric repeat-binding factor 1
TRF2	Telomeric repeat-binding factor 2
TRX	Thioredoxin
TSS	Transcriptional start site
Xi	Inactive X chromosome
XIC	X inactivation center
ZnF	Zinc finger

### 4.3 Statutory declaration and statement

#### **Eidesstattliche Erklärung**

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde.

München, den 21.03.2017

\_\_\_\_\_  
Tobias Anton

#### **Erklärung**

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig ohne Erfolg einer Doktorprüfung unterzogen habe.

München, den 21.03.2017

\_\_\_\_\_  
Tobias Anton

## 4.4 Curriculum vitae

### Personal details

---

Name: Tobias Anton  
Date of birth: 18.08.1986  
Place of birth: Dachau, Germany  
Nationality: German

### Education

---

01.2013 – onwards: Dissertation (Ph.D.)  
Ludwig-Maximilians-University Munich  
Faculty of Biology, Human Biology and BioImaging  
Group of Prof. Heinrich Leonhardt

10.2010 – 09.2012: Master of Science (M.Sc.)  
Ludwig-Maximilians-University Munich  
Faculty of Biology, Human Biology and BioImaging  
Group of Prof. Heinrich Leonhardt

10.2007 – 08.2010: Bachelor of Science (B.Sc.)  
Ludwig-Maximilians-University Munich  
Faculty of Biology, Anthropology and Human Genomics  
Group of Prof. Gisela Grupe

09.1997 – 06.2006: Final High School Examination (Abitur)  
Rainer-Maria-Rilke-Gymnasium Icking, Germany

### Publications

---

**Anton T**, Bultmann S, Leonhardt H and Markaki Y (2014). Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus*

**Anton T**, Leonhardt H and Markaki Y (2016). Visualization of Genomic Loci in Living Cells with a Fluorescent CRISPR/Cas9 System. *Methods in Molecular Biology*

Schmidtman E\*, **Anton T**\*, Rombaut P, Herzog F and Leonhardt H (2016). Determination of Local Chromatin Composition by CasID. *Nucleus*

**Anton T** and Bultmann S (2017). Site-specific Recruitment of Epigenetic Factors with a Modular CRISPR/Cas System. *Nucleus (in press)*

*Manuscripts in preparation:*

Natale F, Rapp A, Yu W, Maiser A, Harz H, Scholl A, Grulich S, **Anton T**, Hörl D, Chen W, Durante M, Taucher-Scholz G, Leonhardt H and Cardoso C. Identification of the Elementary Structural Units of the DNA Damage Response. *In preparation*

**Anton T**, Schmidtman E, Leonhardt H and Bultmann S. Applications of the CRISPR/Cas System Beyond Genome Engineering. *In preparation*

---

\* These authors contributed equally

## 4.5 Contributions

*Declaration of contributions to “Visualization of specific DNA sequences with a programmable fluorescent CRISPR/Cas system”*

Prof. Dr. Heinrich Leonhardt, Dr. Sebastian Bultmann and I conceived the project. I cloned all expression constructs and tested the localization of dCas9-eGFP. I also performed FISH experiments (under supervision of Dr. Yolanda Markaki). Additionally, I contributed to all figures and wrote the manuscript together with Prof. Dr. Heinrich Leonhardt and Dr. Yolanda Markaki.

*Declaration of contributions to “Visualization of genomic loci in living cells with a fluorescent CRISPR/Cas system”*

I wrote the manuscript and the corresponding protocols with the help of Dr. Yolanda Markaki. I also prepared all figures.

*Declaration of contributions to “Site-specific recruitment of epigenetic factors with a modular CRISPR/Cas system”*

Prof. Dr. Heinrich Leonhardt, Dr. Sebastian Bultmann and I conceived the project. I cloned all dCas9 expression constructs, performed the experiments and prepared all figures. I wrote the manuscript with help of Dr. Sebastian Bultmann.

*Declaration of contributions to “Determination of local chromatin composition by CasID”*

Prof. Dr. Heinrich Leonhardt, Elisabeth Schmidtman and I conceived this project. I cloned the BirA\*-dCas9-eGFP and ZNF512-eGFP constructs and performed the immunofluorescence experiments. I prepared figures 1 and 2 and contributed to figure 3. Additionally, I wrote corresponding parts of the manuscript. Sample preparation for mass spectrometry, data analysis and preparation of figure 3 were performed by Elisabeth Schmidtman (co-first author).

---

Prof. Dr. Heinrich Leonhardt

---

Tobias Anton

---

Elisabeth Schmidtman