THE UNIVERSITY
of EDINBURGH

The Role of Zinc Finger Protein WIZ in the Recruitment of Histone Methylase G9a

THE UNIVERSITY *of* EDINBURGH

Burak Özkan

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

2016

Lay summary

Life of all higher eukaryotes starts with a single fertilized egg cell that contains half of the genome of each parent, and this cell has the ability to give rise to a fully developed organism. The genome of a cell is made of DNA and contains all the information that is required for development. The genome can be represented as a ‘book’ with genes as ‘paragraphs’ and DNA as the ‘letters’.

The DNA sequence does not change as the organism develops, yet cells with special functions can emerge. For example, a brain cell and a kidney cell of an organism have identical genomes but their biological function is completely different. This functional diversity is achieved by turning certain genes off and others on. As cells divide, some of them chemically modify their genome to select which genes to turn on or off. All these chemical modifications that change gene activity without altering the DNA sequence are defined as `epigenetic modifications`.

My project focuses on the epigenetic modifying protein; Euchromatic histone-lysine N-methylase 2 or G9a, which is crucial for development. Mice that lack G9a die before birth with severe growth defects. This indicates that the one of many functions of G9a is to regulate the expression of a number of specific genes that cause embryonic lethality when not controlled. However, G9a is not able to bind DNA or RNA, therefore how G9a can recognize these genes and silence them is still unknown. In this project, I investigated how the DNA binding protein WIZ might recruit G9a to DNA in order for G9a to alter gene expression.

Declaration

I declare that this thesis was composed by myself, the research presented is my own except where otherwise stated and this work has not been submitted for any other degree or professional qualification.

Burak Özkan

August 2016

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr. Irina Stancheva for giving me the opportunity to start my scientific career as a Ph.D. student in her lab and for all of her support and guidance during last four years.

Next, I would like to thank all members of Stancheva Lab; Natalia, Ilaria, Christian and Dani for not only creating such a friendly and supportive working atmosphere but also making the life outside of the lab more fun. I would like to thank specially to Simon for his advice, our scientific and non-scientific discussions. He became a precious friend. I would also like to thank Elisa for being a very hardworking and helpful intern-master student in the lab. It was a great opportunity to supervise a talented student like her.

I wish to thank all my friends at Turkish Society here in Edinburgh for their valuable friendship. Their company prevented me from feeling homesick most of the time during these four years.

No words will be enough to express my love and gratitude to my dear partner in life, Tulin. Your love and patience were my biggest support during these busy and stressful years. We share the enthusiasm, curiosity, and the passion not for life but also for science. Your company made this experience more special and memorable than it should be.

I wish to thank my family back in Turkey. The special thanks go to my mother for letting me go and supporting me to follow my dreams. Your constant support was precious throughout this journey.

Last but not least, I would like to express my gratitude to my examiners; Dr. Elizabeth Bayne and Dr. Pierre-Antoine Defossez. I wish that you derive more pleasure from reading my thesis than I did from writing it.

Table of Contents

Contents

List of Tables and Figures

Abstract

The N-terminal tails of histones are subject to many chemical modifications that are involved in a variety of biological functions. Histone methylation is a major epigenetic modification found in both single and multicellular organisms and is directly involved in the regulation of gene expression. Methylation of lysine 9 of histone 3 (H3K9) has been shown to have diverse functions depending on the number of methyl groups added; H3K9me1 marks the active promoters, while H3K9me2 and H3K9me3 are present within inactive gene promoters and pericentric heterochromatin. G9a, also known as euchromatic histone-lysine N-methyltransferase 2 (Ehmt2), is a histone methylase that catalyses addition of mono- and dimethyl groups to H3K9 in euchromatic regions of the genome to silence genes. Therefore, it is a vital component of the gene expression regulation machinery. In mouse embryonic stem (ES) cells, G9a forms a stable heterodimer with the G9a-like protein (GLP or Ehmt1), which is further stabilised by the C2H2-type zinc finger protein, widely interspaced zinc finger protein (WIZ). These three proteins form the core G9a complex, which is essential for mouse development. Lack of any G9a complex member leads to embryonic lethality at E9.5 with severe growth defects.

The ankyrin repeat domain of G9a/GLP can bind to H3K9me1/2 with high affinity *in vitro* (Collins et al. 2008). This enables the self-recruitment of the G9a complex to sites with H3K9me1/2 and maintenance of the mark. However, the initial recruitment of the G9a complex to sites lacking H3K9me1/2 mark during differentiation is poorly understood. Neither G9a nor GLP has a DNA/RNA binding domain, so recruitment of the G9a

complex to specific sites must be mediated by other binding partners of the G9a complex. Using mass spectrometry, I was able to identify a number of zinc finger proteins as binding partners of G9a. Among these, WIZ was identified in stoichiometric amounts to G9a and GLP, and is a potential DNA binding protein similar to other C2H2-type zinc fingers. The aim of this study was to determine the role of WIZ in the recruitment of the G9a complex to specific sites. I showed that knockdown of WIZ had no significant effect on the chromatin binding of G9a in undifferentiated mouse ES cells, which indicates WIZ is dispensable in the maintenance of H3K9me2. However, I observed a 30% decrease in the G9a levels upon WIZ knockdown, which shows that WIZ might have a role in stabilising G9a. Using recombinant WIZ zinc finger pairs, I was able to show that the 3rd and 4th zinc finger of WIZ bind DNA *in vitro*. Furthermore, using the systematic evolution of ligands exponential enrichment (SELEX) approach I demonstrated that the zinc fingers of WIZ preferentially bind to G-rich double-stranded DNA sequences. Binding site analysis with synthetic DNA indicated that WIZ ZF3-4 require two binding sites that are a certain distance apart from each other for efficient binding. In addition, ZF3-4 binds ssDNA with higher affinity than dsDNA, and binding to ssDNA is sequence-independent.

This study shows for the first time that mouse WIZ zinc finger pairs can bind DNA and RNA *in vitro*. Therefore, sequence-specific recruitment of G9a might be mediated by WIZ during differentiation. Furthermore, DNA binding preference of WIZ might suggest that WIZ-mediated recruitment of G9a to establish H3K9me2 could occur at the R-loops where G-rich DNA forms a hybrid with newly transcribed RNA or at the G-rich repetitive sequences.

Chapter I – Introduction

1.1. Epigenetics

he human genome project has been a big leap forward in deciphering of the human genome and understanding the complexity of our genetic map (Lander et al. 2001). The developing sequencing technology enabled the genome sequencing of many other model organisms. The information gathered from the whole genome sequencing (WGS) projects provided information about the genetic make-up of the organism of interest. For unicellular organisms, WGS projects provide extensive information about the content of their genome, which is useful to understand their lifestyle. However, the complexity of gene regulation and functional differences between cells in multicellular organisms cannot be comprehensively explained by the WGS projects. The same DNA sequence is present in each cell of a multicellular organism, yet the diverse cell types with various functionality emerge as the organism develops. This is because there is a regulation 'above' genetics that controls the fate decisions of cells with the same DNA – Epigenetics. The literal definition of epigenetics is 'over' or 'above' genetics. It refers to the heritable modification of chromatin through which gene expression or phenotype of cells change in a way that DNA sequence remains unaltered (Felsenfeld 2014). Epigenetic regulation utilises DNA-binding transcription factors (TFs) and various chemical modifications of DNA and histones to alter the protein and RNA content of cells, which results in the functional diversity of cells in multicellular organisms. TFs change the gene expression by recruiting activator or repressor complexes to the gene promoters or enhancer elements.

However, TFs usually starts a cascade of events that defines the gene activity, but they cannot always determine the spectrum of gene activity in a cell. This is evident by the low efficiency of the cloning of animals from nuclei of differentiated cells and the generation of induced pluripotent stem cells (iPSCs) by expression of the Yamanaka factors in somatic cells (Wakayama & Yanagimachi 2001; Takahashi & Yamanaka 2006). In 1957, Conrad Waddington illustrated the 'epigenetic landscape' as a ball, representing a cell, ready to roll down on the top of a hill, and each trajectory that the ball takes accounts for a different cell fate decision (Waddington 1953). In the case of cloning of animals, the cytoplasmic content of a totipotent cell is usually not enough to re-shape the epigenetic landscape of the nuclei of a terminally differentiated cell. The epigenetic landscape is formed by chemical modifications on chromatin, which occur either on the DNA itself or on histone proteins that interact closely with the DNA in the nucleus forming nucleosomes. Although all epigenetic modifications are reversible, as 'writer' and 'eraser' proteins exist for each mark, they usually create a stable epigenetic landscape in terminally differentiated cells. Epigenetic modifications are not only necessary for the proper division of labour in cellular function in multicellular organisms but also essential for the passing on traits to offspring. DNA methylation, Polycomb group (PcG) and trithorax group (trxG) derived modifications in *Drosophila* and H3K9 methylation in fission yeast have been shown to be heritable (Bird 2002; Cavalli & Paro 1998; Audergon et al. 2015). These discoveries fuel the heated discussion of 'Nature versus Nurture' in the determination of the development of an organism and its fitness to survive in life.

1.1.1. **Building blocks of the epigenome**

Epigenome refers to all the epigenetic modifications or factors that can modify chromatin to define the function of the genome. The most basic unit of chromatin is a nucleosome, consisting of 146 bp of DNA wrapped around two copies each of the core histones H2A, H2B, H3, and H4 forming an octamer (Richmond et al. 1997). Chromatin is made of repeating nucleosomes like 'beads on a string'. The length of DNA between a pair of nucleosomes, linker DNA, ranges between ~20-90 bp depending on the local chromatin structure within a single cell (van Holde 1988). There are two types of chromatin organisation found in cells: euchromatin and heterochromatin. Euchromatin refers to a more 'open' chromatin structure with exposed linker DNA that is permissive for transcription. On the other hand, heterochromatin refers to a more compact chromatin structure that is usually not permissive to transcription. Dense nucleosome arrangement achieves the compact structure of heterochromatin, which is stabilised by the linker histone. The compact arrangement of nucleosomes was shown to form a zig-zag formation with a diameter of 30 nm ('30 nm fibre') as a structural module within chromosomes *in vitro* (Woodcock et al. 1984; Song et al. 2014). However, a recent study showed that this could be an artefact of *in vitro* conditions as 30 nm fibre was absent in human mitotic chromosomes when they were investigated by cryo-electron microscopy (cryo-EM) and small angle X-ray scattering (SAXS) (Nishino et al. 2012).

By definition, euchromatin refers the all open chromatin in the cell, while heterochromatin comes in two flavours; (1) constitutive heterochromatin and (2) facultative heterochromatin. Constitutive heterochromatin consists of gene-poor, repetitive regions of the genome that can affect the gene expression of nearby genes in the form of position-

effect variegation (Saksouk et al. 2015). On the other hand, facultative heterochromatin contains various chromosomal regions that are enriched in heterochromatic marks upon developmental signals (Trojer & Reinberg 2007). The most common mark of heterochromatin is the hypoacetylation of histones, which is required for the efficient compaction of the chromatin fibre (Gräff & Tsai 2013). Similarly, euchromatin requires histone acetylation an open chromatin formation (Yan & Boyd 2006; Shogren-Knaak et al. 2006). In addition to acetylation, histone methylation plays an essential role in the formation of heterochromatin. Interpretation of histone modifications by cells is more complex when histone methylation considered as the position and the degree of the lysine methylation on a histone affects the interpretation. Namely, H3K9me and H3K27me, along with DNA methylation, are the most common histone modifications found in heterochromatic regions of the genome (Trojer & Reinberg 2007).

1.2. Epigenetic modifications of chromatin

After the discovery of DNA methylation and subsequently histone modifications, many studies focused on the epigenetic modifications to understand the role of chromatin in the regulation of gene expression (Felsenfeld 2014). Recent years have been fruitful in gathering information about the epigenetic modifications, thanks to the pilot project ENCyclopedia of DNA Elements (ENCODE), which provided information about regulatory regions of the genome by interrogating variety of features – DNAse hypersensitive sites, histone modifications, DNA methylation and binding of transcription factors (ENCODE Project Consortium et al. 2007). So far, more than 200 different post-

translational modifications of proteins (PTMs) have been identified (Walsh & Jefferis 2006). DNA methylation and the most common PTMs relevant to epigenetic signalling will be introduced in this section.

1.2.1. **DNA methylation**

The existence of methylated DNA was first discovered in calf thymus (Hotchkiss 1948). Many years later, bacterial DNA methylation was found to be involved in the defence against the foreign DNA in the bacterial genome (Reviewed in Bestor 1990). Bacteria mostly use the adenine methylation, and a number of sequence-specific DNA methyltransferases have been discovered for adenine methylation in bacteria (Kumar & Rao 2013). Apart from its role in immunity, DNA methylation has been implicated in the gene regulation (Low et al. 2001), cell cycle (Reisenauer & Shapiro 2002) and DNA damage repair in bacteria (Bogan & Helmstetter 1996).

The mammalian and plant DNA methylation exclusively occur at the carbon-5 position of cytosine in a CpG context (CpG methylation) (Bird 1987). Additionally, in recent years, non-CpG methylations (or CpH methylation where H is A/C/T) have been shown to make up about 25% of all DNA methylation in adult mouse dentate gyrus (Guo et al. 2013). The abundance of the DNA methylation varies in different tissues but in general, 60-90% of all CpGs in vertebrates are methylated (Ghosh et al. 2010; Tweedie et al. 1997). DNA methylation is involved in a wide range of biological processes in the mammalian genome. Most importantly, most CpG sites overlap with promoters (CpG islands) are almost never methylated except those of the germline genes, genes on the inactive X and in pathological cases such as cancer (Bird 2002). The unmethylated promoters are targeted by a number

of non-methyl CpG-binding effector proteins that can recruit the transcription machinery and actively exclude DNA methyltransferases from these sites (Smith & Meissner 2013). Additionally, CpG-rich regions can recruit these effector proteins regardless of their position in the genome. For example, CXXC finger protein 1 (CFP1), which can bind unmethylated CpGs, can be targeted to an artificial, CpG-rich, promoterless transgene and maintain an unmethylated state (Thomson et al. 2010). On the other hand, DNA methylation at promoters blocks the interaction of these effector proteins with DNA and recruit repressive complexes to silence gene activity (Jones 2012). An important aspect of CpG methylation is that demethylation usually occurs via deamination of 5-methylcytosines into thymidines, which is repaired by base excision repair enzymes (Kress et al. 2001). However, this system is error prone. Therefore, CG-dinucleotides are depleted in the genome, while the unmethylated CGs in the CpG islands remain preserved over evolutionary time. Interestingly, unlike CpG island methylation, the DNA methylation of CG-dinucleotides in gene bodies does not block transcription (Jones 1999). Although the exact function of gene body methylation remains unclear, it might prevent the misplaced start of transcription.

1.2.2. **DNA methyltransferases**

DNA methylation is catalysed by enzymes known as DNA methyltransferases (DNMTs). The catalytic activity of all DNMTs requires S-adenosyl methionine (SAM) as the methyl donor. Catalytically active DNMTs can be categorised into two groups; (1) DNA methyltransferase 1 (DNMT1), which is highly expressed in both somatic and embryonic tissues, is responsible for maintenance of the DNA methylation patterns during DNA

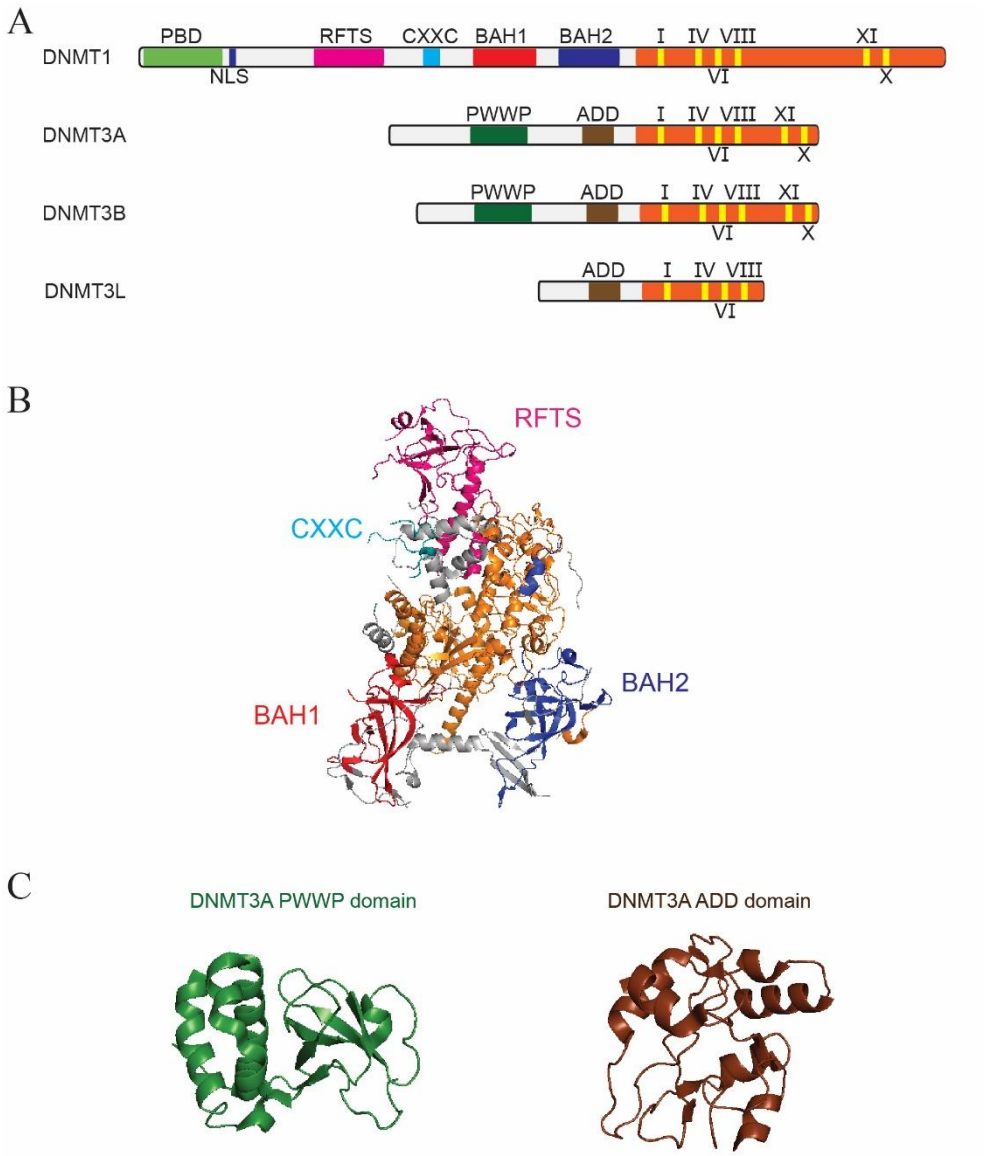


Figure 1. 1 Crystal structure of DNMT1 enzyme and the PWWP and ADD domains of DNMT3A

(**A**) Schematic representation of the domains of DNMT enzymes are shown. DNMT1 consists of a protein binding domain (PBD), a nuclear localization signal (NLS), a replication fork targeting site (RFTS), a CXXC-type zinc finger domain, a polybromo homology domain (BAH1 and BAH2) and a catalytic domain with conserved amino acid motifs. DNMT3A and DNMT3B lacks most N-terminal domains of DNMT1 but contains a PWWP domain, an ADD domain and a functional catalytic domain. DNMTL contains an ADD domain and a inactive catalytic domain. (**B**) Ribbon model of the crystal structure of DNMT1 (291-1620, PDB ID 3av4) is shown. The subunits of DNMT1 is colored according to A. (**C**) PWWP and ADD domain of DNMT3A is shown (PDB ID 3LLR and 3A1A, respectively). All ribbon models are generated by PyMol.

replication, where the unmethylated strand of newly replicated DNA is methylated with the same pattern as its complementary strand (Pradhan et al. 1999). (2) *De-novo* DNA methyltransferases, DNMT3A and DNMT3B, are highly expressed in undifferentiated embryonic stem cells and establish DNA methylation patterns during early development after global DNA demethylation has occurred (Okano et al. 1999). However, in embryonic stem cells, DNMT3A and DNMT3B are also necessary for the proper maintenance of DNA methylation, and DNTM3A/3B deficient cells lose DNA methylation over time (Chen et al. 2003). There are two standard features of catalytically active DNMTs; (1) the ten conserved amino acids motifs in their C-terminal catalytic domain and (2) conserved motifs in the catalytic domain that modulate the catalytic activity, and DNA and cofactor binding ability (Bestor 2000). The crystal structures of catalytic domains of DNMT1 and DNMT3A/B have been solved (Fig 1.1, Song et al. 2011; Jia et al. 2007). The N-terminal part of DNMT1 contains a CXXC zinc finger domain, a DNA replication foci targeting domain and a polybromo homology domain, while these domains are absent in DNMT3A and DNMT3B. Instead, the N-termini of DNMT3A/B contains an ADD (ATRX, DNMT3, DNMT3L) and a PWWP domain (Wu et al. 2011; Otani et al. 2009). DNMT1 preferentially methylates the hemimethylated DNA at the replication foci through its interaction with UHRF1 (Avvakumov et al. 2008; Sharif et al. 2007). DNMT1 can methylate long stretches of DNA with high processivity (Goyal et al. 2006). In the case of *de novo* DNMTs, the co-crystallization of DNMT3A and DNMT3L tetramer (formation: 3L-3A-3A-3L) indicated that the catalytic pockets of two DNMT3As in the tetramer are about 8-10 bp apart. Therefore, the enzymatic activity of *de novo* DNMTs is not processive, but two CG-dinucleotides can be methylated at the same time (Jurkowska et

al. 2008). Moreover, evidence suggests that DNMT3A/B has an intrinsic preference for DNA sequences that contain purine-rich sequences at 5'-, and pyrimidine-rich sequences at 3'-end of CG-dinucleotide (Handa & Jeltsch 2005). Finally, the crosstalk between DNMT3A/B and other repressive complexes occurs via the ADD and PWWP domain, which might target *de novo* DNMTs to specific loci by recognising H3 tail and H3K36me3, respectively (Jurkowska et al. 2011a; Dhayalan et al. 2010a). Although undifferentiated embryonic stem cells lacking DNMTs are viable, the catalytically active DNMTs are required for early development, as the knockout mouse models of DNMT1 and DNMT3B show embryonic lethality, while most DNMT3A-deficient mice die about 4 weeks of age (Li et al. 1992; Okano et al. 1999).

On the other hand, the other two DNMTs – DNMT3L and DNMT2, are not catalytically active. Although it shows homology to catalytically active DNMTs and contains an ADD domain, DNMT3L lacks two important motifs in the catalytic domain (Goll et al. 2006). However, it can still bind DNA and interact with DNMT3A to stimulate its catalytic activity. DNMT3L-DNMT3A interaction is required for proper *de novo* DNA methylation at imprinted loci in germ cells (Jia et al. 2007). Additionally, DNMT3L can interact with unmethylated H3K4 by its ADD domain to facilitate the recruitment of DNMT3A/B to genomic sites depleted with the active H3K4me2/3 mark (Ooi et al. 2007). DNMT2 is the most conserved among the DNMTs as homologues exist in almost all species with very high sequence similarity. However, the catalytic activity of DNMT2 is disrupted by a serine insertion at the critical position in the catalytic domain responsible for DNA methyltransferase activity (Schaefer & Lyko 2010). Instead of DNA, DNMT2 can

methylate cytosine 38 of the tRNA of aspartic acid (Goll et al. 2006). tRNA methylation is essential for folding and stability of tRNAs, although DNMT2-null organisms do not show a lethal phenotype (Okano et al. 1999; Goll et al. 2006; Rai et al. 2007). However, a recent study showed that DNMT2 plays an important role in haematopoiesis (Tuorto et al. 2015).

1.2.3. **Functional role of DNA methylation**

The most commonly observed role of DNA methylation is its effect on gene expression. DNMTs can interact with each other as described earlier. Additionally, DNMTs can interact with many other epigenetic factors, such as histone deacetylases, transcription factors and histone methyltransferases (Jurkowska et al. 2011b). These effector proteins may alter the gene expression in a global manner. In 1983, the first evidence suggesting that DNA methylation might have a role in cancer was discovered. It was shown that the genome of the cancer cells is hypomethylated when compared to the genomes of healthy cells (Feinberg & Vogelstein 1983; Reviewed in Feinberg & Tycko 2004). The hypomethylation of the genome leads to chromosomal instability, reactivation of transposable elements and loss of imprinting (Robertson 2005). In addition to this, cancer cells usually silence tumour suppressor genes by aberrant methylation of their promoters (Jones & Baylin 2002).

There are two individual cases where regulation of gene expression by DNA methylation is well-defined: (1) Genomic imprinting, which refers to the mono-allelic expression of certain genes, occurs by the methylation of one parental allele, which contains a differentially methylated region (DMR) (Shemer et al. 1997). Deletion of DMRs or

DNMTs leads to misexpression of the imprinted genes, which, in some case, causes diseases with growth and behavioural defects (Beckwith–Wiedemann syndrome, Prader–Willi syndrome and Angelman syndrome) and cancer (Wilms tumour) (Reik & Walter 2001). (2) DNA methylation is crucial for the dosage compensation by X-chromosome inactivation in mammals, which refers to complete silencing of one of the X-chromosomes in female mammals (Lyon 1961; Ng et al. 2007). Many rare diseases, such as oral–facial–digital type I (OFDI) and microphthalmia with linear skin-defects (MLS), are associated with the misexpression of X-linked genes, which may happen in the case of a failure in X-inactivation (Franco & Ballabio 2006).

1.2.4. **Histone modifications**

Histone proteins are some of the most conserved proteins among species and are the fundamental building blocks of nucleosomes forming chromatin. There are four core histone proteins: H2A, H2B, H3 and H4. Histones have 20-33 amino acid long N-terminal tails that extend out from nucleosomes and can be modified with different epigenetic marks (Kouzarides 2007). It was shown that globular domain of histones can also be modified (Van Leeuwen et al. 2002). Most commonly seen modifications are methylation, acetylation, ubiquitination, phosphorylation, biotinylation and sumoylation. These modifications can alter the structure of the nucleosome by changing the charge of a histone protein. Chromatin structure which is regulated and organised in this way enables fine tuning of gene activity across the genome. In general, these regulatory actions result in activation of gene expression when they lead to chromatin decondensation by facilitating the access of proteins involved in transcriptional activation (Chambeyron & Bickmore

2004). The modifications that reshape the open chromatin into a more condensed structure reduce accessibility of transcription factors thereby silencing the gene expression (Lelli et al. 2012). For example, acetylation of lysine residues of histones neutralises their positive charge that results in a decrease in the interaction of histone tails and the negatively charged DNA in a nucleosome (Verdone et al. 2006). The downstream interpretation of histone modifications depends on the activity of effector proteins, which can bind to those marks (Becker 2006). Histone tails can be mono-, di- or tri-methylated, and each degree of modification determines the outcome of the interpretation by effector proteins. Different histone modifications may exist on the same locus of chromatin as well as asymmetrically within the same nucleosome (Voigt et al. 2012). Therefore, deciphering the 'histone code' requires an extensive and in-depth knowledge of histone modifications and their effector proteins.

1.2.5. **Distribution of histone marks and their cross-talk with DNA**

A large number of histone and DNA modifications have been described (Bannister & Kouzarides 2011). As these modifications might co-occur in the chromatin context, the combination of two or more modifications on the same nucleosome may occur in thousands of different ways (Margueron & Reinberg 2010). The recent technological development enabled the analysis of the genome-wide distribution of epigenetic marks; some marks have been identified as a landmark of particular genomic sites. For example, H3K4me3 is present at almost 75% of all human active gene promoters (ENCODE Project Consortium 2012). H3K4me3 has a vital role in the regulation of gene expression as the aberrant levels of H3K4me3 leads to misregulated expression of genes in various diseases

(Vashishtha et al. 2013; Mastroeni et al. 2015). H3K4me1 and acetylation of H3K27 (H3K27ac) are commonly found at active gene enhancers (Zhou et al. 2011).

Similarly, H3K9ac is also present at the active gene promoters, while H3K36me2/3 marks the transcribed regions of genome. On the other hand, the H3K9me2/3, H3K27me2/3 and H4K20me3 are usually excluded from active genes. Instead, these modifications often mark silenced genes (Barth et al. 2010). Whether these modifications are the primary reason for gene expression or silencing or just a side-effect of gene activity or inactivity is still heavily debated. A very recent study suggested that targeting an H3K4me3 methylase (PRDM9) to silenced genes can activate gene expression at promoters that lack DNA methylation, but not methylated promoters (Cano-Rodriguez et al. 2016). This study indicates that some histone modifications can play an instructive role in gene regulation. On the other hand, H3K9me2 mark catalysed by histone methylase G9a is present on the methylated allele of imprinted genes. However, a cell with enzymatically inactive G9a can still maintain the DNA methylation at the silenced allele, indicating that H3K9me2 is not causative in the silencing of imprinted genes in ES cells (Zhang et al. 2016). Interestingly, imprinting is maintained by H3K9me2 in the absence of DNA methylation in placenta indicating that the instructive role of histone methylation is context-dependent (Wagschal et al. 2008).

H3K4me3 and H3K27me3 frequently appeared at the same genomic loci in embryonic stem cells. These sites often map to the promoters of genes that are developmentally regulated and stay in a ‘poised’ state until certain developmental signals are received. This unique chromatin state was named as ‘bivalent’ (Bernstein et al. 2006). Heterogeneity in

cultured ES cells is well established by showing fluctuating levels of key pluripotency factors, such as Nanog (Chambers et al. 2007). Since the conventional ChIP-seq experiments use material extracted from millions of cells, therefore, it was argued that the bivalent domains are just a reflection of heterogenic nature of ES cells. Several studies focused on bivalent chromatin were able to confirm the existence of opposing chromatin marks on the same nucleosome (Voigt et al. 2012). Perhaps the debate is over now that bivalency was observed on nucleosomes extracted from various cell types (Shema et al. 2016). The authors of this very recent study developed a technique, in which the single nucleosomes isolated from cells were immobilised on slides coated with polyethylene glycol (PEG). The captured nucleosomes were analysed for various histone marks using specific antibodies. Additionally, the histones can be chemically removed from the slides to perform sequencing of DNA of a particular nucleosome. *Shema et al.* showed that 94% of bivalent nucleosomes are asymmetric while the rest are modified on the same tail. Additionally, bivalent nucleosomes were observed in T-cell acute lymphoblastic leukaemia line, HEK293 cells and an acute lymphoma line with an EZH2 gain of function mutation. On the other hand, significantly less bivalent nucleosomes are observed embryoid bodies, lung fibroblasts and an acute lymphoma line with an EZH2 loss of function mutation (Shema et al. 2016).

The cross-talk between the DNA and histone methylations were suggested by many studies as the three-dimensional organisation of chromatin cannot be considered without a close interaction between the DNA and histones (Ruthenburg et al. 2007). The family of methylated DNA binding proteins (MBPs) involving the methyl-CpG binding domain

(MBD), the Kaiso and the UHRF protein families can recruit HMTs and HDACs to methylated regions of genome (Rottach et al. 2010). It is also known that H3K4 and DNA methylation are mutually exclusive (Okitsu et al. 2010), while H3K9 methylation is frequently associated with DNA methylation (Rose & Klose 2014; Meissner et al. 2008). In the attempt of defining the proteins that are involved in such a cross-talk, a stable isotope labelling with amino acids in cell culture (SILAC) nucleosome affinity purification technique was developed (Bartke et al. 2010). This technique enabled the identification of effector proteins that require only DNA or histone methylations, or both DNA and histone methylation to efficiently bind nucleosomes. For example, MeCP2 needs methylated-DNA in a nucleosomal context for efficient binding, while the TFIIIC complex requires naked DNA indicating that a chromatin remodelling activity is required for its binding.

Acetylation and methylation of histones have been better described and characterised in many studies (Biel et al. 2005). Some of these modifications contribute to the euchromatin formation, which results in the activation of gene expression. For example, trimethylation at H3K4, H3K36 and H3K79 and histone acetylation in general help the formation of open chromatin. On the other hand, histone deacetylation or trimethylation at H3K9, H3K27 and H4K20 result in heterochromatin formation, which silences genes within the condensed chromatin (Kouzarides 2007).

1.2.6. **Histone modifiers**

All histone modifications are reversible. Therefore, every histone mark is made by a ‘writer' enzyme and also removed by an ‘eraser' marker. Well-known histone modifiers will be discussed here.

Modifiers of histone acetylation:

The writers and erasers of histone acetylation are named as Histone acetyltransferases (HATs) and Histone deacetylases (HDACs), respectively. The histone acetylation was first reported in 1964 (ALLFREY et al. 1964). Acetylation of lysine residues takes place at ε-amino group of lysine side chains by HATs, which utilise Acetyl-CoA as a cofactor. This modification reduces the positive charge of the histones, which in return weaken the interaction between the DNA and histones of a nucleosome (Bannister & Kouzarides 2011). Two major classes were defined for HATs; (1) type-B HATs are functional in the cytoplasm, acetylating the free histones. This was shown to be important in the histone deposition (Parthun 2007). (2) type-A HATs are diverse regarding substrate specificity. In general, almost all type-A HATs contains a bromodomain to recognise lysine substrates in a nucleosomal context, and acetylate more than one lysine of a histone protein, including its globular domain (Weaver 2007). There are three subgroups of type-A HATs; (1) Gcn5 related acetyltransferase (GNAT), (2) MYST MYST (for the founding members MOZ, Ybf2/Sas3, Sas2 and Tip60) and (3) p300/CBP. HATs usually co-exist with coactivator complexes and thereby associated with transcriptional activation (Mizzen & Allis 1998).

Four different types of HDACs were defined so far; (1) class I including HDAC1, -2, -3 and -8; (2) class II including HDAC4 to 10 (excluding HDAC8); class IV with HDAC11

and class III with HDACs. Class III-type HDACs depend on Nicotinamide adenine dinucleotide (NAD^+) as a cofactor (Yang & Seto 2007). In general, HDACs do not show a substrate specificity and frequently identified in other protein complexes. For instance, HDAC1 and HDAC2 were detected in both NuRD and CoREST complexes (Denslow & Wade 2007; Grozinger & Schreiber 2002).

Modifiers of histone methylation:

Histone methylation was discovered more than 30 years ago (Murray 1964). The importance of histone methylation started to be recognised after the first histone methyltransferases (HMTase) were discovered and characterised (Rea et al. 2000). Histone methylations can occur either on lysine or arginine amino acids. H3 can be methylated at K4, K9, K27, K36, K79, R2, R17 and R16, while H4 can be methylated at K20 and R3 (Biel et al. 2005). Each lysine residue can be methylated up to three times. The first discoveries of histone methyltransferases started in 2000. Su(Var)3-9 from *Drosophila*, Clr4 from *S. Pombe*, and SUV39H1 and SUV39H2 in mammalian cells were identified as an HMTases containing a SET domain (Su(var)3-9, Enhancer-of-zeste, and Trithorax), and shown to be specific to H3K9 (Rea et al., 2000). Following that, many other SET domain-containing proteins have been identified as the SET domain is highly conserved. These proteins have highly conserved S-adenosyl-L-methionine (SAM) binding site to catalyse the transfer of methyl groups to substrates. Although methyltransferase activity could be predicted by looking for SET domains in proteins, some have been identified as having evolutionary deactivated SET domains (Jenuwein &

Allis 2001). In addition, Dot1 HMTase can methylate H3 at K79 but does not contain a SET domain (Feng et al. 2002).

The SET domains of different HMTases have high homology between species. However, the flanking regions, Pre-SET and Post-SET, display different conformation and structure in various SET domain family members and can determine functional variation. These domains might alter the tertiary structure of the SET domain or interact with other proteins (Min et al. 2002). As another structural aspect, the Pre-SET domain is usually cysteine-rich and has an ability to bind zinc ions, which stabilise the catalytic activity. The Pre-SET domain contains many β-strands that can form β-sheets, and the number of β-sheets varies in different HMTases, whereas there are constant two alpha-helices in the SET domain. These variations are essential for the specific substrate recognition and catalytic activity of these enzymes (Min et al. 2002). Furthermore, the SET domain contains a pocket-like structure that can accommodate histone tails. The amino acid composition of this structure determines the specificity of the enzyme by conferring unique properties such as size or charge to the pocket. Therefore, altering the amino acids in this structure might change the specificity of HMTases (Zhang et al. 2003). In addition to this, a knot-like structure has been observed in the C-terminus of SET7/9 between a β-strand and alpha helix (Jacobs et al. 2002). This structure is essential for the catalytic activity when methyl donor AdoMet and the lysine residue substrate are nearby. A completely conserved tyrosine residue within the knot drives the catalytic activity by deprotonating the e-amino group of lysine. Upon deprotonation, the lysine side chain reacts with the methyl group of AdoMet and the methylated lysine residue occurs as an end product (Jacobs et al. 2002; Trievel et al. 2002).

The discovery of the first histone lysine demethylase was not until 2004 when *Shi et. al.* characterised LSD1 (KDM1A), a nuclear homolog of amine oxidases, which functions as a histone demethylase (Shi et al. 2004). Before this discovery, due to the high thermodynamic stability of histone methylation marks, it was though that lysine methylation was irreversible once established, and could only be replaced by histone exchange (Jenuwein & Allis 2001). After the discovery of LSD1, many other lysine demethylases were identified, and categorised into two classes; (1) KDM1 (Lysine Demethylase 1) class demethylases are FAD-dependent amine oxidases. These enzymes can only remove methylation from mono- and di-methyl lysine. (2) The Jumonji C (JmjC) class demethylases contain a standard JmjC domain and are Fe(II) and 2-oxoglutarate-dependent. Some members of JmjC family, such as Jmjd2b in the case of H3K9me3, can target tri-methyl lysine (Fodor et al. 2006). Therefore, there is a lysine demethylase for every lysine methylation in cells showing that epigenetic regulation is very dynamic (Pedersen & Helin 2010).

1.2.7. **G9a/GLP and H3K9me**

G9a (EHMT2) is a SET domain-containing protein, which acts as a major histone methyltransferase in mammalian cells. G9a adds methyl groups to H3K9 in euchromatic regions of the genome (Tachibana et al. 2001). It has been shown that G9a is responsible for ~90% of H3K9me2 in the genome as well as a small proportion of H3K27 methylation (Yokochi et al. 2009; Trojer et al. 2009). Additionally, G9a interacts with another HMTase, G9a-like protein (GLP), to form a stable heterodimer (Ogawa et al. 2002) which is further stabilised by a zinc finger protein WIZ (Figure 1.3, Ueda et al. 2006; Wu et al.

2010). Furthermore, WIZ depletion by small inhibitory RNAs reduces G9a levels in ES cells. More importantly, structural analysis of G9a and GLP has shown that these proteins bind histone H3 N-terminal peptides mono- and dimethylated at K9 in vitro. This binding occurs via a loop between ankyrin repeats that are present in G9a and GLP (Collins et al. 2008) suggesting that the G9a/GLP complex can read and write histone marks. Another study suggests that G9a/GLP-mediated histone methylation is not dependent on pre-existing methylation marks, and G9a/GLP can methylate non-histone targets (Rathert et al. 2008a). In addition to this, heterodimer formation may allow either G9a or GLP to create methylation marks while the other can recognise these marks via the ankyrin repeat domains.

Apart from being a major histone methyltransferase, the G9a/GLP complex also has other roles. The ability to bind its mark enables it to bind chromatin and protect the methylation mark from further modifications. For example, this ‘gatekeeping' function of G9a suggests that G9a/GLP can restrict transcription activation by any unliganded nuclear receptor (Garcia-Bassets et al. 2007). G9a/GLP complex can also recruit other proteins to reinforce silencing. HP1 (heterochromatin-associated protein 1) is recruited to chromatin by the G9a complex via auto methylation of K239 on G9a, which creates new binding sites for HP1. In addition to this, loss of G9a results in reduced HP1 occupancy at specific promoter sites (Chin et al. 2007; El Gazzar et al. 2008). More importantly, G9a is required for DNA methylation of G9a-target gene loci and can interact with DNMT3A/3B via its ankyrin repeats even in the absence of the catalytic SET domain. This finding suggests that the catalytic activity of G9a is dispensable for DNA methylation, but the presence of G9a is

required to reinforce the recruitment of DNMTs (Epsztejn-Litman et al. 2008; Tachibana et al. 2008). In addition to this, G9a is involved in de novo DNA methylation at specific loci by acting cooperatively with LSH, which is a chromatin remodelling ATPase (Myant et al. 2011). Studies carried out in our lab have shown that G9a deficient mouse embryonic stem cells lose H3K9 methylation and DNA methylation at imprinted loci and a subset of gene promoters (Myant et al. 2011; Auclair et al. 2016; Zhang et al. 2016). Hemimethylated DNA binding protein, UHRF1, can recognize H3K9me2 and H3K9me3 via its plant homeodomain (PHD) and tudor domain (Hashimoto et al. 2008; Karagianni et al. 2008; Rottach et al. 2010). UHRF1 and DNMT1 form a complex during replication (Sharif et al. 2007). Therefore, the involvement of G9a in DNA methylation occurring at specific gene promoters might be explained by the recruitment of UHRF1 to these sites via methylated H3K9 binding.

While the catalytic activity of G9a mainly leads to the repression of genes, some studies identified G9a as an activator. For example, G9a was found in a repressive complex containing H3K4 demethylase Jarid1a and in the activator Mediator complex (Chaturvedi et al. 2012). This study showed that G9a in involved in the activation of β globin β^{maj} gene, while repressing the embryonic globin E^{y} expression. Similarly, G9a was identified in many co-activator complexes (Reviewed in Shankar et al. 2013). Interestingly, the activator action of G9a is usually independent of its catalytic activity, indicating that the functional role of G9a depends on its interacting partners (Purcell et al. 2012).

A growing number of studies have implicated that G9a has a role in cellular proliferation, replication and senescence. One study showed that p21 gene is targeted and repressed by

the recruitment of UHRF1 and G9a to its promoter (Purcell et al. 2012). Because of its role in the cellular proliferation, many cancer types involve upregulation of G9a (Casciello et al. 2015). Therefore, pharmacological inhibition of G9a by small molecules have been tested on various cancer types. As an example, the treatment of a G9a inhibitor, BIX-01294, was able to suppress the cellular proliferation of breast and colon cancer cells by inducing autophagy-associated cell death (Kim et al. 2013).

On the other hand, the major histone methyltransferases for H3K9me1 are PRDM3 and PRDM16 (Pinheiro et al. 2012). It was shown that catalysis of H3K9me1 can occur in the cytoplasm and nucleus of human and mouse cells (Pinheiro et al. 2012). The same study also showed that H3K9me1 is required for the maintenance of H3K9me3 in the pericentric heterochromatin, which is catalysed by Suv39h1/2 and SETDB1 (also known as ESET) enzymes. H3K9me3 marks the pericentric heterochromatin and important for genome stability. Krüppel-associated box (KRAB)-associated protein 1 (KAP1) and SETDB1-mediated H3K9me3 are required for the silencing of retroviral elements in mouse ES cells (Matsui et al. 2010). SETDB1 is also important in the silencing of developmental gene regulators and maintenance of ES cell state (Bilodeau et al. 2009).

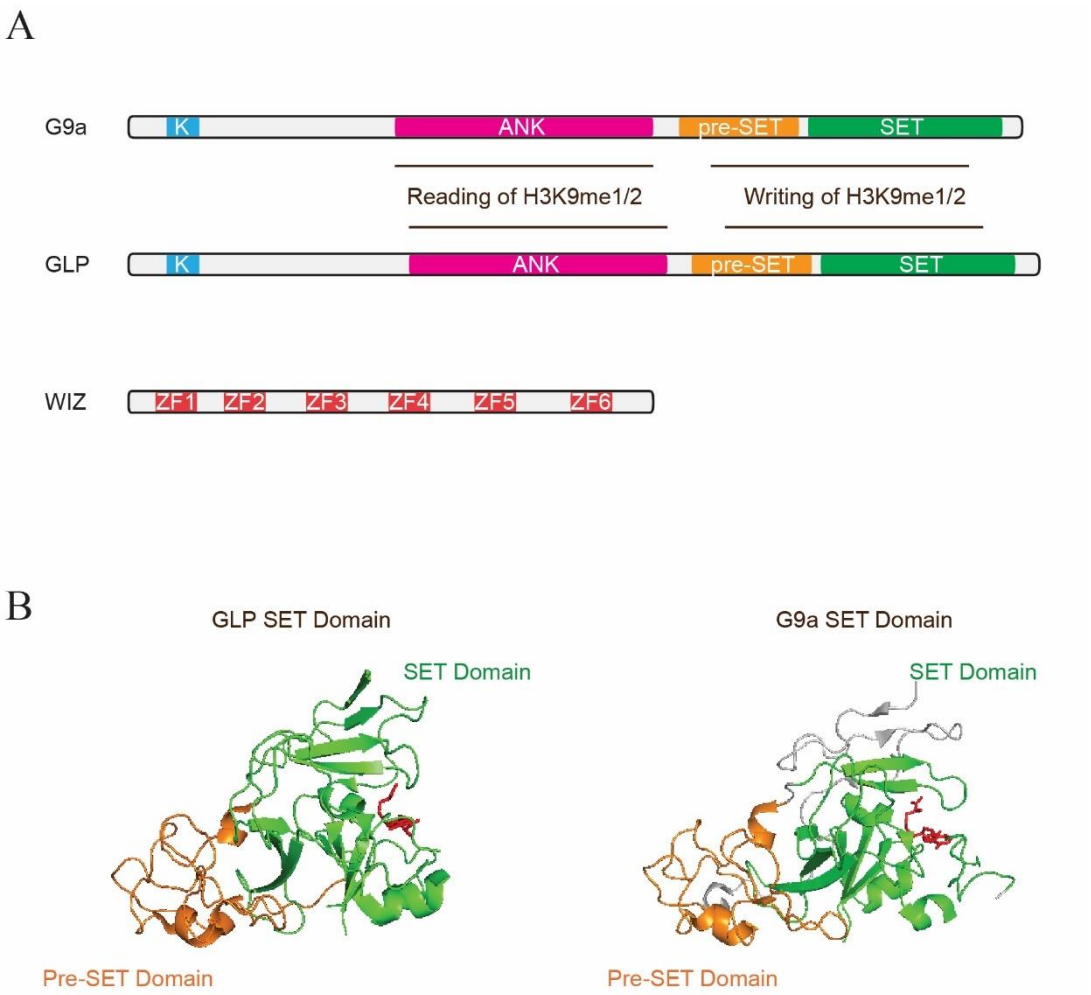


Figure 1. 2 The members of the core G9a complex

(**A**) Schematic representation of the core components of the G9a complex. G9a and GLP form heteromiders via SET domains. Heteromeric G9a/GLP interacts with the 6th zinc finger of WIZ to form the core G9a complex. The SET domains of G9a/GLP can catalyze H3K9me1/2. The ankyrin repeat domain (ANK, shown in magenta) of G9a/GLP can bind to H3K9me1/2. G9a can be automethylated on the N-terminal K239 (shown in teal). (**B**) The crystal structures of G9a/GLP SET domains (G9a PDB ID: 2O8J, GLP PDB ID: 2IGQ). Pre-SET domains are shown in orange. Co-factor S-adenosyl-L-methionine (SAM) was shown in red.

1.3. Zinc finger proteins and their role in epigenome editing

The initial studies of the structure of chromatin were pointing to the direction that some parts of it were active to allow transcription (Klug 1983). An exciting discovery was made when it was shown a 40-kDa protein was required for transcription of 5S RNA genes of *Xenopus laevis*. This protein was the very first transcription factor thereby named as transcription factor IIIA (TFIIIA) (Brown 1984). This study showed that TFIIIA interacts with individual sites on the DNA. Following this study, the zinc finger domains of TFIIIA were identified and their DNA binding was characterised (Miller et al. 1985). Structural studies of TFIIA showed that the zinc finger domains consist of an antiparallel β-sheet that contains a loop formed by the two cysteines, an α-helix containing and two histidines, of which cysteines and histidines are stabilised by a zinc ion (Berg 1988; Lee et al. 1989). Zinc fingers with this conformation are referred as C_2H_2-type zinc fingers, which are commonly found in the proteins that are involved in gene regulation – at least 3% of all human proteins contain a C_2H_2-type zinc finger (Bateman et al. 2002). Initially, the number of and order of zinc binding residues were used to classify zinc fingers (Leon & Roth 2000). However, detailed characterization of DNA-binding zinc finger proteins revealed that there is a rich diversity of zinc fingers in nature, which led to an extended classification system depending on the structural conformation on zinc fingers.

1.3.1. **Types of zinc finger motifs**

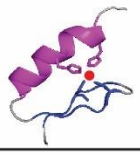
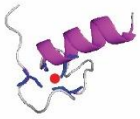
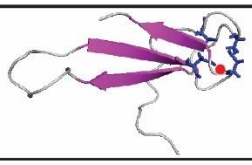
Many zinc finger domain-containing proteins were discovered thus far. Zinc finger motifs in these proteins vary regarding their structure and DNA/RNA or protein binding properties. In general, the structure of zinc finger domains is very stable and does not

undergo a conformational change upon substrate binding. Zinc finger motifs are classified depending on 3D-structure as ‘fold groups’. There are ten different fold groups identified thus far. Most common zinc finger motifs belong to the Cys_2His_2-like, treble clef, and zinc ribbon zinc fingers, which will be introduced in this section (Figure 1.2A, Krishna et al. 2003).

C_2H_2-like:

These are the classical zinc fingers, which hold a zinc ion by two cysteines and histidines residues, which is referred as ‘zinc knuckle'. The alpha helix loop between these residues is responsible for the DNA recognition. A common sequence motif for C2H2-type zinc fingers is $X_{(2)}$-Cys-$X_{(2,4)}$-Cys-$X_{(12)}$-His-$X_{(3,4,5)}$-His. A C2H2-type zinc finger recognises 3-4 bp on a DNA molecule. However, C2H2-type zinc finger proteins often contain more than one zinc finger that is tandemly arrayed, which provides a larger sequence recognition due to overlapping patterns of contact of adjacent zinc fingers. Interestingly, the linker sequence between adjacent zinc fingers, TGEKP, is conserved almost in half of the all zinc finger proteins (Wolfe et al. 2000). Moreover, the linker sequence contributes to the efficient DNA binding of TFIIIA and Wilms' tumour suppressor protein (WT1) by capping the C-terminus of the preceding alpha helix once it binds the correct DNA sequence (Laity et al. 2000). In general, C2H2-type zinc fingers can be classified into three subgroups; (1) triple-C2H2 group contains three adjacent zinc fingers that bind to a single ligand. Zif268 is a well-studied example of such zinc fingers. The crystal structure of Zif268 was solved, and it showed that each zinc finger recognised 3 bp and collectively bound a 9 bp-long DNA sequence (Figure 1.2B, Pavletich & Pabo 1991; Iuchi 2001). The

A

Type	Representative Structure	Name	Reference
C2H2-like		Zinc finger 5 of TFIIIA	Nolte et al. 1998 PDB ID: 1TF6
Treble clef finger		Zinc finger 2 of ESR1	Schwabe et al. 1993 PDB ID: 1HCQ
Zinc ribbon		Zinc finger of TFIIS	Qian et al. 1993 PDB ID: 1TFI

B

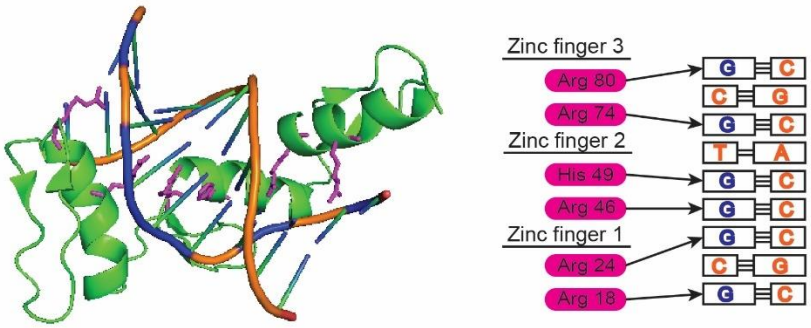


Figure 1. 3 The types of zinc finger motifs and DNA binding of a C2H2-type zinc finger protein

(**A**) Representative crystal structures of the major zinc finger types. C2H2-type zinc fingers contain a zinc ion that interacts with two cysteine and histidine residues. The amino acids in the alpha helix are responsible for the DNA binding. Similar to C2H2-type zinc fingers, four cystein residues interact with a zinc ion in Treble clef finger motif. Zinc ribbon motif contains 3 antiparallel beta strands that form a beta sheet. Zinc ions are shown in red and cystein residues interacting with the zinc ions are shown in blue. All ribbon models are generated by PyMol. (**B**) The crystal structure of tandem Zif268 zinc fingers bound to DNA (PDB ID:1P47). Ribbon model of tandem zinc fingers of Zif268 is shown in green. DNA strands are shown in orange. The amino acids of Zif268 zinc fingers that interact with the DNA bases are shown in magenta and blue, respectively. The diagram on the left shows the one-to-one interaction of individual amino acids and the DNA bases in the crystal structure.

same study also identified the residues on the alpha helix at the position -1, 1, 2 and 6 are crucial in forming a hydrogen bond with the nucleotides on the target DNA. (2) The second group, multiple-C2H2, contains proteins with multiple zinc finger domains that are located close to each other in a specified interval. These proteins usually bind more than one ligand. The most well-known example of this group is TFIIIA, which has 9 zinc finger domains with different binding properties as described earlier (Nolte et al. 1998). (3) The last group, separated-paired-C2H2, contains multiple zinc finger pairs that are widely separated. These type of zinc fingers comprise a flexible linker sequence between the fingers of each pair. Tramtrack (TTK) and RPDII-BF1 (also as Znf40) contains widely interspaced zinc finger pairs that can bind DNA in a similar binding mode as a triple-C2H2 group of zinc fingers (Fairall et al. 1993; Fan & Maniatis 1990).

Treble clef finger:

The zinc finger with treble clef motif is made of a β-hairpin and α-helix at the N- and C-terminus, respectively, which contribute two ligands each for zinc binding. The members of this group are diverse concerning both biological function and the sequence similarity. Therefore, there are many subgroups of treble clef finger proteins. Most notable, Really Interesting New Gene finger-like (RING-like) domain consists of C-X_2-C-$X_{(9-39)}$-C-$X_{(1-3)}$-H-$X_{(2-3)}$-C-X_2-C-$X_{(4-48)}$-C-X_2-C amino acid motif that binds two zinc ions at the same time. This domain contains about 40 to 60 amino acids and proteins containing RING-like domain are usually involved in ubiquitination pathway (Laity et al. 2001). Another significant subgroup includes nuclear receptor-like fingers. Proteins in this subgroup often do not contain any additional secondary structures; however, they contain more than one

treble clef motifs and bind DNA in a very sequence-specific manner. GATA zinc fingers, GATA1, -2 and -3, zinc fingers explicitly recognise DNA sequence of WGATAR (in which W indicates A/T and R indicates A/G) in the regulatory regions of genes. Interestingly, studies on GATA-1 and friend of GATA (GOT) proteins identified that zinc finger domain of GATA-1 can interact with GOT (Liew et al. 2005).

Zinc ribbon:

Zinc ribbon folding group is found in a diverse group of proteins, which do not show much sequence similarity in the DNA-interacting residues. The core structure of zinc ribbon folding group consists of two β-hairpins that form two structurally similar zinc-binding sites. Therefore, the ligands are contributed by two zinc knuckles. The two β-hairpins are in an antiparallel conformation, and usually an additional β-strand forms hydrogen bonds with one of the two β-hairpins. Thus, a typical zinc ribbon finger is made of three antiparallel β-strands that form a β-sheet. The length of the β-hairpins may differ as well distance between two of zinc binding sites (Krishna et al. 2003). Transcription elongation factor SII (TfIIS) and the transcription initiation factor TFIIB are the well-studied examples of zinc ribbon containing proteins (Qian et al. 1993). Additionally, eukaryotic RNA polymerase II subunit 9 (RPB9), which is homologous to TFIIB, contains zinc ribbon motifs and has a function in the transcription start-site selection and termination of elongation (Wang et al. 1998). Apart from the role in transcription, heterotrimeric human ssDNA-binding protein, Replication Protein A (RPA), contains a zinc ribbon motif. RPA was shown to be important in the DNA replication, recombination and repair (Wold 1997; Bochkareva et al. 2000).

1.3.2. **Functional role of zinc finger proteins in epigenetics**

Recruitment by DNA binding proteins:

This type of recruitment requires a DNA-binding domain that mediates the recruitment of histone modifying enzymes to a specific DNA element. In Drosophila, the Polycomb group of proteins (PcG) are required to silence Homeotic genes. The studies have shown that certain DNA elements were responsible for recruitment of the PcG to silence genes nearby, thus named as Polycomb Responsive Elements (PRE) (Ringrose & Paro 2004). PcG proteins are found in protein complexes with distinct roles. One such complex, PHO-repressive complex, contains a zinc finger domain containing protein Pleiohomeotic (PHO) and Sfmbt that can recognise H4K20me1/2. The zinc fingers of PHO can bind DNA in a sequence-specific manner. It has been shown that PHO can target PRC2 complex to chromatin, although PHO alone is not sufficient and efficient recruitment requires other factors (Ringrose et al. 2003).

The transcriptional repression by PcG complexes can be counteracted by the activity of Trithorax group of proteins (TrxG) (Eissenberg & Shilatifard 2010). Trithorax is a methyltransferase responsible for the methylation of H3K4, which is a mark of active promoters. H3K4 and H3K27-methylated nucleosomes are mutually exclusive; however, the PREs and Trithorax recruiting elements (TREs) frequently map to the same sites of the genome (Ringrose & Paro 2007). This indicates a competition between the PcG and TrxG proteins. As PHO recognises the PRE elements to facilitate the recruitment of PcG; GAGA factor, Dorsal switch protein 1 (DSP1) and Zeste can recognise sequences in Trithorax recruitment elements (TREs).

Proteins containing CXXC type zinc fingers are known to recognise CpG sequences. The first evidence of CXXC type zinc finger-mediated recruitment was discovered in the case of MBD1 protein, which can bind to methylated DNA sequences via its MBD domain while CXXC type zinc finger domains bind to unmethylated DNA (Cross et al. 1997; Jørgensen et al. 2004). For example, CXXC domain in KDM2A has been shown to recognise unmethylated CpG sequences and thereby the CXXC domain of KDM2A is involved in the recruitment of KDM2A/B complex (Blackledge et al. 2010). On the other hand, a C2H2 (Cys2His2) type zinc finger protein, Kaiso, and Kaiso-like proteins use all its zinc fingers to bind to both methylated and unmethylated DNA in a sequence-specific manner (Buck-Koehntop et al. 2012).

Due to their sequence-specific DNA-binding ability, zinc finger proteins have been used as a tool in editing genomes. Extensive studies on the DNA binding preferences of zinc finger proteins enabled designing of synthetic zinc fingers peptides that can recognise a sequence of interest with a very high affinity (Choo et al. 1994). Coupling engineered zinc finger motifs with DNA nucleases enable targeting of specific genomic sites to cause DNA damage with high precision. By utilising the DNA damage repair mechanisms, cells can incorporate a donor DNA into the genome of cells by using zinc finger nucleases (Carroll 2008). Following the similar principle, transcription activator–like effectors (TALEs) from the pathogenic bacterium, which is relatively easier to generate when compared to zinc finger nucleases, were used in targeted genome editing (Boch 2011). This technology was later implemented to edit the epigenome of cells.

Engineering zinc finger proteins in transcriptional regulation:

There are a number of examples that showed custom zinc finger proteins could alter chromatin or DNA modifications at a specific location that they are designed to target when coupled with epigenetic factors. One such study used an engineered zinc finger protein that can recognise a 9 bp-motif at the vascular endothelial growth factor A (VEGF-A) gene promoter, which is usually unmethylated (Kungulovski et al. 2015). The engineered zinc fingers were fused to either the catalytic domain of a DNA or histone methyltransferase (DNMT3A or GLP, respectively). Upon ectopic expression of the transgene encoding the custom zinc finger, the VEGF-A promoter could be silenced by either acquiring DNA methylation or H3K9me2 due to the activity of DNMT3A and GLP, respectively. This study shows that the epigenetic modifiers can be targeted by engineered zinc finger to edit locally the chromatin modifications and gene expression in cells. Another study focused on a tumour suppressor gene, *MASPIN*, which is heavily methylated in lung cancer cell lines (Beltran & Blancafort 2011). This study utilised a number of artificial zinc finger proteins that recognise a unique 18 bp-long sites in the *MASPIN* proximal promoter and fused to VP64 transactivator domain. It was shown that two artificial zinc finger proteins were able to re-activate *MASPIN* by inducing DNA demethylation at its promoter, which indicates that reactivation of epigenetically silenced genes is also possible when custom zinc fingers are fused to a transactivator domain. Therefore, custom engineered zinc finger proteins display a promising potential in therapeutic approaches for diseases that are linked to aberrant epigenetic regulation.

Following the zinc finger nucleases, the discovery of TALEs has facilitated the genome/epigenome editing studies. TALEs are consisting of repeating, tandemly-arrayed

and highly similar amino acid repeats. Each repeat recognises a single base-pair thereby the recognition code in TALEs is simple. In addition to this, the neighbour effect in zinc finger nucleases does not apply in the DNA binding of neighbouring amino acid repeats of a TALE. In general, TALEs are simpler to make but lack the high throughput potential as new TALE is required for every target sequence (Jurkowski et al. 2015). As an example to epigenetics editing by TALEs, one study showed that TALEs fused with Lysine-specific demethylase 1 (LSD1) could remove H3K4me2 at enhancer sites targeted by these TALEs (Mendenhall et al. 2013). This study also showed that the removal of active enhancer marks from chromatin causes downregulation of proximal genes. Therefore, TALE-based epigenetics editing potentially can be used for the identification of functional genomic elements.

In recent years, the discovery of CRISPR/Cas9 system changed the gameplay in the area of genome/epigenome editing. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is used as an immune defence mechanism by prokaryotes. Briefly, the short fragments of alien DNA are incorporated into genome between CRISPR repeats. Similar to the antibody production of the adaptive immune system in the animal kingdom, the short foreign DNA fragments are incorporated into CRISPR-RNA, which can guide exonucleases (such as Cas9) to invading genetic elements containing matching sequences (Hsu et al. 2014). The CRISPR-Cas9 system requires the ectopic expression of Cas9 protein and a guide RNA with Cas9-specific protospacer-adjacent motif (PAM) sequence that targets the site of interest. This method could be used to target any number of genomic sites if the same number of guide RNAs were provided (Jurkowski et al. 2015).

Additionally, a catalytically-dead Cas9 (dCas9) protein can be fused to a chromatin modifier or used as bait for recruiting other proteins at a particular genomic site. Such an approach was employed in a study where the dCas9 protein fused with DNMT3A increased the CpG methylation levels at the promoter regions of target loci (Vojta et al. 2016).

1.4. Recruitment of epigenetic modifiers

Histone modifications are highly dynamic such that acetylation, methylation or phosphorylation marks can be made or removed within minutes upon specific stimuli, creating a dynamic chromatin environment. The function of chromatin modifiers depends on being near their substrates. Chromatin modifying enzymes need to possess an affinity towards their substrates enabling efficient binding of these proteins. In other words, epigenetic modifiers have to be recruited to genomic sites where they can function. The recruitment of these enzymes might occur through either direct or indirect mechanism. While direct recruitment involves DNA, RNA or histone mark recognition by domains in histone modifying enzymes, the indirect recruitment depends on the activity of mediator proteins that interact with epigenetic modifiers.

1.4.1. **Recognition of epigenetic modifications**

Post-translational modifications (PTMs) on histones can be recognised by ‘reading' modules that contain a surface suitable to accommodate a particular histone modification.

Structural features of the standard modules that can recognise PTMs will be discussed in this section.

Bromodomain:

Bromodomain modules contain a conserved central hydrophobic pocket that can accommodate an acetylated lysine residue, which can be anchored to a conserved asparagine residue (Ntranos & Casaccia 2016). Most bromodomain-containing proteins have multiple bromodomains that are tandemly arrayed, which increases their affinity to acetylated lysine. Evidence suggests that the cooperative binding of multiple bromodomains of a protein increases its affinity towards a particular combination of PTMs, while individual bromodomains can bind any acetylated lysine with lower affinity (Muller et al. 2011). The function of bromodomain modules depends on the domain composition and the neighbouring domains (such as the PHD) (Sanchez & Zhou 2009). The evidence suggests that bromodomain-containing proteins are important in embryonic development (Shang et al. 2009), normal spermatogenesis and the oocyte-embryo transition (Philipps et al. 2008), and IL-6 signalling pathway (Hou et al. 2008).

Tudor domain:

The Tudor domain was originally identified in Tudor protein of *Drosophila.* The crystal structure of the tudor domain of human survival of motor neurone (SMN) showed that it could bind to di-methylated arginine residues. The crystal structure revealed that Tudor domain forms an anti-parallel β-sheet that is made of 5 β-strands and a barrel-like fold (Sattler et al. 2001). In addition to arginine methylation, tudor domain modules can recognise various lysine methylations as well. For example, the histone demethylase

Jumanji domain containing 2A (JMJD2A) can bind to H3K4me via its double tudor domains (Huang et al. 2006). Tudor domain-containing proteins are involved in RNA metabolism (Liu et al. 2010), splicing (Cheng et al. 2007), RNA interference (RNAi) and related pathways (Pek et al. 2012), and DNA damage response (Botuyan et al. 2006).

Chromodomain:

Chromodomain (also known as chromatin organisation modifier) module usually consists of 40-50 amino acids and is often utilised by proteins involved in chromatin remodelling and regulation of gene expression. Crystal structure of the chromodomain of *Drosophila* HP1 protein was solved, and it showed that 3D-structure of a chromodomain is made of a three-stranded anti-parallel β-sheet that back onto an α-helix (Jacobs et al. 2002). The chromodomain module recognises methylated lysine residues. Some proteins contain a distant relative of chromodomain, chromo shadow domain, which very frequently functions together with the chromodomain (Aasland & Stewart 1995). The chromo shadow domain of HP1 proteins can form a dimer with another chromo-shadow domain, which is essential in the formation of higher-order chromatin structure at heterochromatic regions of the genome (Yamamoto & Sonoda 2003).

14-3-3 proteins:

The 14-3-3 protein family is highly conserved in diverse mammalian species. These proteins can recognise phospho-serine/phospho-threonine and can interact with a number of binding partners that are involved in cellular processes, such as signal transduction, protein trafficking and cell-cycle regulation (Pozuelo Rubio et al. 2004). Lower eukaryotes contain two isoforms of 14-3-3 proteins, while in human, there are seven

different isoforms (β, ε, η, γ, τ, ζ and σ). Almost all 14-3-3 proteins can form homo- and heterodimers. The phospho-serine and phospho-threonine binding motifs of 14-3-3 proteins were identified (Yaffe et al. 1997). In addition to this, it was shown that 14-3-3 proteins can recognise unmethylated serine/threonine (Johnson et al. 2010).

Other histone modification binding modules:

The ankyrin repeat domain consists of repeating 33 amino acids that form two alpha helices separated by loops was first identified in Cdc10 and signalling protein Notch in *Drosophila* (Breeden & Nasmyth 1987). Ankyrin repeat domain-containing proteins are utilised by both eukaryotes and prokaryotes, and ankyrin repeat module is frequently associated with protein-protein interactions. For instance, ankyrin repeat domain of Notch protein can interact with the zinc finger domain of Ying Yang 1 (YY1) (Yeh et al. 2003). As discussed earlier, ankyrin repeats of G9a and GLP mediate the interaction with DNMTs. In addition to this, ankyrin repeats of G9a/GLP can recognise H3K9me1 and H3K9me2 (Collins et al. 2008).

WD40 repeats module is very frequently present in proteins and functions as adaptor module in signal transduction, pre-mRNA processing, cytoskeleton assembly, transcriptional activation and cell cycle control. These repeats consist of conserved tryptophan (W) and aspartic acid (D) residues and a length of 40 amino acids (WD40). WD40 repeats can recognise phosphorylated serine and threonine as well as methylated lysine (Orlicky et al. 2003). An example of WD40 repeats containing protein is WDR5, a WD40-repeat containing histone methylase in *Drosophila*. A study showed that WD40-

repeat protein acts as a module for recognition of H3K4me2 and H3K4me3 marks (Wysocka et al. 2005).

1.4.2. Diverse mechanisms of recruitment of histone-modifying enzymes

The recruitment of histone-modifying enzymes depends on a broad range of molecular mechanisms that include DNA elements, histone modification-recognizing modules and protein-protein interactions. In addition to recruitment via DNA binding factors (mentioned in 1.3.2), other mechanisms for recruitment of epigenetic modifiers have been described and these include recruitment via existing histone modifications and RNA.

Recruitment by the existing histone modifications:

Modifications on chromatin can create an anchoring point for proteins with modification binding modules described earlier. Through additional protein-protein interactions, histone modifying enzymes can be recruited to existing marks via these modification binding modules. However, many modifications, individually or in combination, can be 'read' by more than one 'reader' proteins. For instance, H3K9me2 can be recognised by G9a/GLP, HP1 proteins and EED. Similarly, some 'reader' proteins can interact with both transcriptional activator and repressor complexes (Yun et al. 2011). Chromodomain-containing protein Eaf3 binds to H3K36me2/3, and interacts with both histone acetyltransferase and deacetylase complexes. Therefore, Eaf3 can mediate the recruitment of these complexes to alter gene expression (Keogh et al. 2005; Carrozza et al. 2005). However, the loss of Eaf3 leads to an increase in acetylation in the gene bodies where

H3K36me2/3 is abundant, which indicates that the recruitment of deacetylase complexes depends on other factors. On the other hand, some large protein complexes can recognise a combination of PTMs for efficient binding. The silent information regulator (SIR) complex containing heterodimers of SIR2-3-4 can recognise three nucleosomes at the same time. The interaction occurs through the DNA recognition by SIR4 and unmodified H4 tail binding by SIR3, which is also sensitive to H3K79me3 (Martino et al. 2009). Another well-studied example is preferential binding of DNMT3b to highly compacted by histone methylation and hypoacetylated long nucleosomal array *in vitro* (Kashiwagi et al. 2011). Moreover, the PWWP domain of DNMT3A binds H3K36me3 to target DNMT3A to chromatin carrying that mark (Dhayalan et al. 2010b; Baubec et al. 2015).

Recruitment by RNA:

Involvement of numerous noncoding RNAs in the recruitment of histone modifying enzymes has been reported (Koziol & Rinn 2010). The X chromosome dosage compensation is mostly mediated by noncoding RNAs in mammals and flies. The noncoding RNA *Xist* coats one copy of the X chromosomes in female mammals to recruit Polycomb repressor complexes that eventually lead to compaction of the X chromosome into a Barr body (Chow & Heard 2009; Brown et al. 1991). In Drosophila, male flies have to double gene expression of their X chromosome to compensate for protein levels in the female flies, which have two active X chromosomes. In this case, roX1 and roX2 noncoding RNAs, which can bind to the male X chromosome are transcribed. These RNAs recruit histone H4 acetyltransferase MOF and the dosage compensation complex (DCC) to the male X chromosome increasing the accessibility of transcription factors on

the chromosome to enhance transcription (Smith et al. 2001). The Air noncoding RNA and Kcnq1ot1 have been shown to associate with the G9a complex (Nagano et al. 2008; Pandey et al. 2008). These studies do not indicate a direct recruitment of G9a by these RNAs, and G9a does not contain any RNA recognition motifs. Therefore, the interaction between the noncoding RNAs and G9a is possibly mediated by an RNA binding molecule, which may also interact with the G9a complex.

1.5. Project Aims

Zinc finger protein WIZ, a potential DNA binding protein, interacts with G9a/GLP heterodimer. This interaction is important for the stability of G9a/GLP; however, whether or not it has any role in the recruitment of G9a via DNA/RNA-binding remains unclear.

During my studies, I asked the following two questions; (1) Is WIZ important in the recruitment of G9a to chromatin in a global or sequence-specific manner? (2) Is WIZ a DNA- or RNA-binding protein, and if so, what are the binding preferences of WIZ?

Chapters three and four address the question (1) through analyses of the functional role of WIZ in the G9a complex. The results in these sections suggest that WIZ knockdown reduces the G9a/GLP levels by 30% indicating that WIZ is important in the stability of the G9a complex. However, WIZ knockdown does not affect the catalytic activity of G9a and do not reduce the chromatin binding of G9a without additional factors. Chapter five addresses the question (2) through characterization of DNA-binding zinc fingers of WIZ *in vitro.* The results in this section show that ZF3 and ZF4 of WIZ can bind DNA and

RNA *in vitro*. WIZ ZF3-4 binds G-rich DNA in a sequence-specific manner, while the single-stranded DNA or RNA binding occurs independently of the DNA sequence. Collectively, these results suggest that WIZ is not essential in the maintenance of H3K9me2 by recruitment of G9. However, DNA/RNA-binding of WIZ might still be important for the establishment of H3K9me2 during ES cell differentiation and in vivo during embryonic development. Together, these results suggest a potential new function for WIZ in the recruitment of the G9a complex.

Chapter II – Materials and Methods

2.1. Materials

In this section, the equipment and reagents used in experiments are enlisted. Buffers were stored at room temperature as default unless otherwise stated.

2.1.1. General buffers and bacterial media

Phosphate-buffered saline (PBS): 140 mM NaCl, 3 mM KCl, 2 mM KH2PO4, 10 mM Na2HPO4.

Tris-buffered saline (TBS): 50 mM Tris-HCl pH 8.0, 150 mM NaCl

Tris-borate EDTA (TBE): 89 mM Tris, 89 mM Boric Acid, 2 mM EDTA pH 8.0.

Tris-EDTA (TE): 10mM Tris-HCl pH 7.5, 1mM EDTA pH 8.0

Tris-acetate EDTA (TAE): 40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA pH 8.0.

SOC media - 20 g/L Difco Bacto tryptone, 5 g/L Difco Bacto yeast extract, 10mM NaCl, 2.5mM KCL, 10mM MgCl2, 10mM MgSO4, 20mM glucose.

LB media - 10 g/L Difco Bacto tryptone, 5 g/L Difco Bacto yeast extract, 5 g/L NaCl, pH 7.2.

2.1.2. Recombinant protein purification buffers

Lysis buffer: 50 mM Tris-HCl pH 7.6, 200 mM NaCl, 0.2 mM DTT, 5 mM MgCl2, 1 mg/mL Chicken egg lysozyme, 1X protease inhibitor cocktail.

Wash buffer L: 140 mM NaCl, 3 mM KCl, 2 mM KH2PO4, 10 mM Na2HPO4 and 0.5 mM PMSF.

Wash buffer H: 50mM Tris pH 7.6, 500 mM NaCl, 1mM DTT and 0.5 mM PMSF.

Elution buffer (for GST-tagged proteins): 50 mM Tris-HCl pH 7.6, 500 mM NaCl, 0.2 mM DTT, 100 mM Glutathione, 0.5 mM PMSF and adjust pH to 7.6. It is important to adjust the pH to 7.6 after making the elution buffer.

Elution buffer (for 6xHis-tagged proteins): 50mM Tris pH 7.6, 500 mM NaCl, 1mM DTT and 0.5 mM PMSF, 500 mM Imidazole.

Dialysis buffer: 50mM Tris pH 7.6, 500 mM NaCl, 0.2mM DTT and 0.5 mM PMSF.

Anion buffer A: 50mM Tris pH 7.6, 100mM NaCl, 1mM DTT, 0.2mM PMSF. **Anion buffer B:** 50mM Tris pH 7.6, 1M NaCl, 1mM DTT, 0.2mM PMSF.

Gel filtration buffer: 50mM Tris pH 7.6, 100-1000mM NaCl, 10% Glycerol, 1mM DTT, 0.5mM PMSF.

Dilution buffer: 50 mM Tris-HCl pH 7.6, 500 mM NaCl, 0.2 mM DTT, 0.5 mM PMSF.

2.1.3. **Buffers for protein manipulation**

Nuclear extraction buffer (NE1): 20mM HEPES pH7.5, 10mM KCl, 5mM MgCl2, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol, 0.5mM DTT and Protease Inhibitor Cocktail (1X) (SIGMA-P8340). DTT and 20X complete protease inhibitors stock were added freshly. NE1 buffer aliquots were stored at -20 °C.

4X Sample buffer: 200 mM Tris-HCL pH 6.8, 8% SDS, 20% Glycerol, a tip-full of bromophenol blue. Final DTT concentration was adjusted to 100 mM prior to use.

SDS PAGE resolving gel: 7-20% (w/v) 29:1 acrylamide:bis-acrylamide, 0.1% (w/v) SDS, 375 mM Tris-HCl pH 8.8, 0.1% (v/v) TEMED, 0.1% (w/v) APS. Home-made gels were stored at 4 °C up to one week when necessary.

SDS PAGE stacking gel: 4% (w/v) 29:1 acrylamide:bis-acrylamide, 0.1% (w/v) SDS, 125 mM Tris-HCl pH 6.8, 0.1% (v/v) TEMED, 0.1% (w/v) APS. Home-made gels were stored at 4 °C up to one week when necessary.

Western blot running buffer: 25 mM Tris, 250 mM Glycine, 0.1 % (w/v) SDS.

Western blot transfer buffer: 25 mM Tris, 250 mM Glycine.

Ponceau S staining solution: 1% (v/v) glacial acetic acid, 0.5% (w/v) Ponceau S.

Coomassie A solution: 25 % (v/v) isopropanol, 10% (v/v) glacial acetic acid, 0.05% (w/v) coomassie brilliant blue R-250.

Coomassie D solution: 10% (v/v) glacial acetic acid.

Native Tris-glycine-EDTA (TGE) gel: 50 mM Tris-HCl pH 7.6, 475 mM Glycine, 2 mM EDTA pH 8, 5% (w/v) 29:1 acrylamide:bis-acrylamide, 0.1% (w/v) APS, 0.1% (v/v) TEMED.

Denaturing urea polyacrylamide gel: 89 mM Tris, 89 mM Boric Acid, 2 mM EDTA pH 8.0, 12% (w/v) 29:1 acrylamide:bis-acrylamide, 230 mM Urea (SIGMA-U5378), 0.1% (w/v) APS, 0.1% (v/v) TEMED.

10X Electrophoretic mobility shift assay (EMSA) buffer: 100 mM Tris-HCl pH 7.6, 500 mM KCl, 10 mM DTT. Stored at -20 °C.

EMSA dilution buffer: 50 mM Tris-HCl pH 7.6, 500 mM NaCl, 0.2 mM DTT.

Strep2 wash buffer: 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA.

2.1.4. **Buffers for DNA manipulation**

Orange G loading buffer (6X): 10 mM Tris-HCl (pH 7.6), 0.15% orange G, 60% glycerol, 60 mM EDTA.

PCR buffer IV (10X): 200 mM (NH4)2SO4, 750 mM (Tris-HCl pH 8.8), 0.1% (v/v) Tween® 20, 15 mM MgCl2.

10X Cross-linking buffer: 500 mM Hepes pH 7.9, 1.5 M NaCl, 10 mM EDTA and 5 mM EGTA

ChIP buffer L1: 50 mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40 and 0.25% Triton X-100.

ChIP buffer L2: 10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA.

ChIP buffer L3: 10 mM Tris-HCl pH 8, 1 mM EDTA and 0.5 mM EGTA.

ChIP dilution buffer: 20 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA and 1% Triton X-100.

ChIP wash buffer 1: 20 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA, 1X protease inhibitor complex, 0.1% SDS and 1% Triton X-100.

ChIP wash buffer 2: 20 mM Tris-HCl pH 8, 500 mM NaCl, 2 mM EDTA, 0.1% SDS and 1% Triton X-100.

2.2. Methods

2.2.1. DNA/RNA manipulation experiments

Small scale plasmid DNA preparations: Small scale plasmid DNA preparations were carried out from 2-3 ml confluent bacterial culture following the QIAprep Spin Miniprep Kit (#27106) instructions manual.

Large scale plasmid DNA preparations: Large scale plasmid DNA preparations were carried out from 400 ml confluent bacterial culture following the QIAGEN Plasmid Maxi Kit (10) (#12162) instructions manual.

Polymerase chain reaction (PCR): PCRs using the home-made Taq polymerase (2.5 activity units/μl) were performed in buffer containing 1X Buffer IV, 3 mM MgCl2, 0.4 mM dNTPs, 1 /μl Taq polymerase, 5 μM primers and desired amount of template. The conditions of denature, annealing and elongation were determined empirically for each primer pair. For cloning purposes, Phusion® High-Fidelity was used by following the manufacturer's instructions.

PCR product purification: PCR products were purified by using PureLink® PCR Purification Kit (K3100-01) according to the manufacturer's instructions.

Quantitative PCR (qPCR): qPCR experiments were carried out by using SYBR® Green PCR Master Mix (ThermoFisher Scientific, 4309155). 10 µL of 2X PCR master mix, 1 µL of 2 µM primer mix and 9 µL dH2O was added to 2 µL of template.

Extraction of total RNA from mammalian cells: 80-90% confluent cells were harvested by trypsinization and washed twice by PBS. 1 ml TRIzol® Reagent (15596-026) was added to cell pellet and mixed. The mixture was stored at -80 °C at this stage. Next, 200 µL chloroform was added and the mixture was vortexed for 15 s. After spinning at 13000 rpm at 4 °C for 15 min, the upper phase was transferred to a new 1.5 ml tube. 600 µL isopropanol was added and the samples were kept at room temperature for 10 min. RNA was precipitated by spinning the samples at 13000 rpm at 4 °C for 20 min. The RNA pellet was washed twice with 1 ml ice-cold 70% ethanol. 1 µL RiboLock RNase inhibitor (ThermoFisher Scientific, EO0381) was added directly on the RNA pellet. The pellet was dissolved in desired amount of RNAse-free ddH2O. 1 µL DNAseI (ThermoFisher Scientific, EN0525) was added and samples were incubated at 37 °C for 20 min. DNAseI was inactivated at 95 °C for 3 min. RNA concentration was measured using a Picodrop Spectrophotometer.

Restriction enzyme (RE) digestion: All reaction mixes were prepared using the appropriate reaction buffer for desired RE enzyme. 0.5-10 µL RE was used depending on the amount of DNA to be digested. Incubation time was determined according to manufacturer's instructions. The result of RE digestion was analyzed by agarose gel electrophoresis.

DNA extraction from agarose gel: Samples were run on an agarose gel prepared using UltraPure™ Low Melting Point Agarose (16520-050) with appropriate percentage of agarose. DNA fragments were excised from the gel with a clean scalpel when sufficiently separated. The excised gel pieces were processed using GeneJET Gel Extraction Kit (K0691, Thermo Fisher) according to manufacturer's instructions.

DNA ligation: A vector:insert ratio of 1:3 and 1:5 was used for sticky and blunt end ligation, respectively.

$$\text{Insert mass in ng} = 3 \text{ (or } 5) \times \left[\frac{\text{insert length (bp)}}{\text{vector length (bp)}}\right] \times \text{vector mass in ng}$$

DNA ligation experiments were either carried out by CloneJET PCR Cloning Kit (K1231, Thermo Fisher) T4 DNA ligase (NEB) according the manufacturer's instructions.

Preparation of competent bacteria: For generating competent bacteria, a modified version of the Inoue method was used (Inoue et al. 1990).

Transformation buffer: 55 mM MnCl2, 15 mM CaCl2, 250 mM KCl and 10 mM PIPES (pH 6.7).

A single bacterial colony was picked and grown in 25 mL LB at 37°C. After 6h incubation, three flasks containing 250 ml SOB media were prepared and 10 ml, 4 ml and 2 ml starter culture is transferred to these flasks. Three cultures were incubated overnight at 18°C. Next day, OD was read by a spectrophotometer, the culture with an OD closest to 0.55 was monitored every 45 minutes. When OD reaches to 0.55, the culture flask is put in an ice bath. Cells are harvested by centrifugation and the pellet was resuspended in 80 ml transformation buffer. Cells were mixed by swirling and harvested by centrifugation at 2500g for 10 min at 4°C. Cells were resuspended in 20 ml ice-cold transformation buffer

and 1.5 ml DMSO was mixed to the suspension. After 10 min incubation on ice, the suspension was aliquoted and snap-frozen in liquid nitrogen.

Bacterial Transformation: A frozen competent bacteria aliquot was thawed on ice for 10 min. 1/3 of the aliquot (50 µL) was transferred to a tube. 5 ng of plasmid or a maximum of 10 µL ligation product was mixed with the competent bacteria with gentle pipetting up and down. After 3 min incubation on ice, the mixture was heat shocked at 42°C for 50 seconds in a water bath. Then, the mixture was kept on ice for 3 min. Next, 250 µL SOC media was added on the bacteria and the tube was incubated at 37°C for 1 hour. If a plasmid DNA was transformed, 50 µL of the sample was seeded on agar plates with appropriate antibiotics. If a ligation product was transformed, then the cells were pelleted at 4000 rpm and resuspended in 50 µL LB medium. Then, all bacteria were seeded on agar plates. Next day, individual colonies were picked.

Colony PCR: To verify the correct integration of an insert after a ligation experiment, colony PCR performed. Insert specific primer pairs were added to a standard PCR master mix. Single colonies were picked by a yellow tip and transferred to a replica plate by touching gently. Then the tip was put in a PCR tube containing 10 µl of PCR master mix with primers. PCR amplification was carried out as described above and PCR product was visualized on an agarose gel.

DNA sequencing: PCR or RE digestions products were cleaned by a PureLink® PCR Purification Kit. Plasmids were used directly in the sequencing reaction. 4 µl of DNA (100-500 µg for double stranded DNA), 3 µl of 2.5X Sequencing buffer, 2 µl of BigDye®

(Thermo Fisher Scientific) and 1 µl of primer (3.2 pmol/µl) were mixed well in a PCR tube. Following PCR was carried out:

Temperature (°C)	Time (s)	
96°C	60	
96°C	10	x25 cycles
50°C	5	
60°C	240	
4°C	pause	

All DNA sequencing was carried out by the GenePool service at Edinburgh University, using ABI 3730 capillary sequencers. The sequencing data were analyzed by SeqManPro (DNAstar) software.

Denaturing PAGE purification of DNA fragments: Labelled DNA samples loaded to denaturing PAGE were run at 150 V for 45 min. Gel was scanned by Licor Odyssey and fragments of interest were excised. The gel pieces were put in 500 µL dH2O and boiled for 5 min, then rotated at room temperature overnight. Next day, after eluting the supernatant, purified DNA was ethanol precipitated and re-dissolved in desired amount of dH2O.

Measuring the DNA concentration by Qubit™: 2 µL sample was mixed with 198 µL Qubit™ working solution. After 2 min incubation at room temperature, DNA concentration was measured using Qubit® Fluorometer.

Random DNA library generation: Single-stranded random DNA and 5'-DY-782 (or IR-800) labelled reverse primer was mixed in equimolar amounts. The mixture was incubated at 90 °C for min 3 min and cooled down at room temperature for 5 min. Next, the annealed

double stranded DNA was extended using Klenow Fragment, exo- (Thermo Fisher Scientific, EP0421) following the manufacturer's instructions. The double-stranded random DNA pool was purified by PAGE-purification and DNA concentration was calculated by Qubit™. At least 4 ng of labelled DNA is required for proper detection.

Annealing oligonucleotides: All oligonucleotides were ordered from Eurofins. Lyophized oligonucleotides were resuspended in ddH_2O to a 100 µM stock according to manufacturer's instructions. Equal volumes of complementary oligonucleotides were mixed gently in an Eppendorf® Safe-Lock microcentrifuge tube with equal volumes of 2x TEN buffer (20mM Tris pH 7.5, 100mM NaCl, 2mM EDTA). The mixture was heated to 90°C for 10 minutes in a heat block. Then, the heat block was turned off and blocks removed to room temperature for ~4hrs. The annealed oligonucleotides were analysed via native PAGE after EtBr staining or scanned on the licor machine if IR800 labelled.

Chromatin Immunoprecipitation (ChIP): 80-90% confluent ES cells were harvested, washed twice with PBS and resuspended in 9 mL PBS. 1 mL 10X ChIP cross-linking buffer is added and mixed well. 275 µL of 36% formaldehyde solution (SIGMA-F8775) was added to reach 1% final concentration. Samples were kept in cold room on a rocker platform for 10 min. Cross-linking was stopped by adding 550 µL of freshly prepared 2.5 M glycine. Samples were spun at 1400 rpm for 6 min and supernatant was discarded. Pellet was washed twice with PBS then resuspended in 5 mL ChIP L1 buffer. After 10 min incubation on ice, supernatant was discarded and pellet was resuspended in 5 mL ChIP L2 buffer. After 10 min incubation on ice, supernatant was discarded and pellet was resuspended in 20 µL ChIP L3 buffer for every 1 million cells and final SDS concentration

was adjusted to 0.5%. Samples were sonicated using Bioruptor® Ultrasonicator at high frequency, for 23 cycles (30 seconds on/30 seconds off per cycle). Sonicated sample was spun at 13000 rpm at 4 °C and supernatant was moved to a fresh tube. After adjusting the final SDS concentration to 1%, 50 µL of sonicated chromatin was digested with 5 µL protease K (P4850, Sigma) overnight, at 65 °C. Next day, DNA was purified by using the PureLink® PCR Purification Kit (K3100-01, Thermo Fisher) and analyzed on 1.5% agarose gel. Expected DNA fragment sizes are between 100 bp and 500 bp.

Specific antibody against the protein (see Table 1.1) of interest was mixed with 15 µg of chromatin and incubated at 4 °C for 3 hours. 1.5 µg chromatin was used as input and kept on ice during immunoprecipitation of IP samples. In the meantime, 50 µL Dynabeads® Protein G for Immunoprecipitation (10003D, Thermo Fisher) per IP was washed twice with ChIP dilution buffer. Beads were blocked with 20 µg BSA and sonicated salmon sperm DNA (Sigma) for at least 1 hour. Blocked beads were washed with ChIP dilution buffer and equally distributed to each IP samples. After 2 hours of incubation at 4 °C, the IP samples were washed 5 times with ChIP wash buffer 1, 3 times with ChIP wash buffer 2 and twice with TE, for 5 min each wash. After last wash, beads were resuspended in 100 µL TE. Immunoprecipitated DNA was purified and eluted in 100 µL after overnight protease K digestion as described above. 2 µL DNA was used as template in qPCR.

cDNA synthesis: 4 µg total RNA mixed with 1 µL of 0.5 µM oligo dT in final vlolume of 12 µL was incubated at 65 °C for 5 min. Samples were chilled on ice immediately. 8 µL cDNA synthesis buffer mix, which consists of 4 µL of 5X first strand buffer, 1 µL RiboLock RNase inhibitor, 1 µL of 10 mM dNTPs and 2 µL of 0.1 mM DTT, was added

and samples were incubated at 42 °C for 2 min. 1 µL SuperScript® II Reverse Transcriptase (ThermoFisher Scientific, 18064014) was added and mixed by pipette up and down. Samples were incubated at 42 °C for 2 hours. The reaction was stopped by incubation at 70 °C for 15 min. Stock cDNA was diluted in 1:30 and 2 µL cDNA was used as template in qPCR experiments.

2.2.2. **Protein manipulation experiments**

Extraction of nuclear protein from ES cells: The cell pellet collected from a confluent T175 flask was resuspended in 1 ml ice-cold NE1 buffer. The suspension was transferred to a 2 ml Dounce homogenizer and homogenized by 20 strokes. Next, the sample was spun at 3500 rpm at 4 °C for 5 min. Supernatant was discarded and nuclei was resuspended in 100 µL NE1 buffer. 1 µL Benzonase (Novagen, 70746) was added per 10 million cells and mixed very well. Samples were incubated for 15 min at room temperature first, and then for 30 min at 4 °C. The final concentration of NaCl was adjusted to 300 mM or 500 mM for normal extraction or histone extraction respectively. Samples were rotated in cold room for 1 hour. Nuclear debris was pelleted via cold centrifugation at 13000 rpm and supernatant was kept. Protein concentration was measured with Bicinchoninic acid assay (BCA) method.

Protein quantification by BCA assay: Working solution was prepared by mixing bicinchoninic acid solution (Sigma - B9643) and copper (II) sulfate solution (Sigma - C2284) in 50:1 ratio. 1 µL of the total protein was mixed with 1 mL of working solution. Samples were incubated at 65 °C for 15 min while shaking strongly. After the incubation, samples were chilled on ice briefly and absorbance was measured by using a

spectrophotometer. At wavelength 562 nm, absorbance of a protein sample with concentration of 1 µg/µL is approximately 0.04.

SDS-PAGE gel electrophoresis: First, resolving gel was prepared and poured between glass plates. Some isopropanol was poured on the resolving gel in order to burst bubbles. Once resolving gel sat, the isopropanol was cleaned and stacking gel was poured. Protein samples were prepared by adding 4X sample buffer and DTT to final concentration 1X and 100 mM, respectively. Samples were denatured at 95 °C for 5 min and loaded to the gel.

Western blotting: Nuclear extracts or immunoprecipitated proteins were separated on appropriate SDS-PAGE gel. To transfer proteins, nitrocellulose membrane was preferred for large proteins, and PVDF membrane was used for histones. Membranes were blocked with 4% milk with 0.1% Tween20. Primary antibody detection was done overnight by desired antibodies (see antibody list). Membranes were washed three times 10 min each in PBS + 0.1% Tween and blocked again for 30 min. Desired IR800 or IR700 labelled secondary antibodies (LiCOR) were added and membranes were incubated at room temperature for at least 1 hour. Membranes were washed three times as above (final wash was done with PBS only). Membranes were scanned by Licor Odyssey Scanner.

Step-wise salt extraction: 20 million ES cells were harvested, pelleted and resuspended in 500 µL NE1 buffer. The resuspension was transferred to a 2 ml Dounce Homogenizer. Cells were disrupted by homogenizer pestle with 25 strokes. Homogenized cells were transferred to a 2 ml round bottom Eppendorf tube and spun at 3000 rpm for 4 min at 4 °C. The supernatant, which contains the cytoplasmic fraction was transferred to a fresh

tube. The final salt concentration was adjusted to 500 mM and the sample was kept at -80 °C. Next, the pellet, which contains nuclei, was resuspended in 150 µL NE1 buffer with 0 mM NaCl, incubated on a spinning wheel for 45 min at 4 °C. Nuclei were spun at 3000 rpm for 4 min at 4 °C and supernatant was transferred to a fresh tube. Final salt concentration was adjusted to 500 mM and the nuclear extract was kept at -80 °C. The pellet was resuspended in 150 µL NE1 buffer with 100 mM NaCl and incubated on a spinning wheel for 45 min at 4 °C. Nuclei were spun at 3000 rpm for 4 min at 4 °C and supernatant was transferred to a fresh tube. Final salt concentration was adjusted to 500 mM and the nuclear extract was kept at -80 °C. The pellet was resuspended in 150 µL NE1 buffer with 200 mM NaCl and incubated on a spinning wheel for 45 min at 4 °C. Nuclei were spun at 3000 rpm for 4 min at 4 °C and supernatant was transferred to a fresh tube. Final salt concentration was adjusted to 500 mM and the nuclear extract was kept at -80 °C. The pellet was resuspended in 150 µL NE1 buffer with 300 mM NaCl and incubated on a spinning wheel for 45 min at 4 °C. Nuclei were spun at 3000 rpm for 4 min at 4 °C and supernatant was transferred to a fresh tube. Final salt concentration was adjusted to 500 mM and the nuclear extract was kept at -80 °C. The pellet was resuspended in 150 µL NE1 buffer with 400 mM NaCl and incubated on a spinning wheel for 45 min at 4 °C. Nuclei were spun at 7500 rpm for 5 min for 4 min at 4 °C (nuclear membrane gets lysed at 400 mM NaCl concentration and chromatin becomes like a spider web, which is harder to pellet) and supernatant was transferred to a fresh tube. Final salt concentration was adjusted to 500 mM and the nuclear extract was kept at -80 °C. The pellet was washed twice with 500 µL NE1 buffer with 0 mM NaCl and resuspended in 130 µL NE1 buffer with 0 mM NaCl. 5 µL Benzonase was mixed very well by using a filter tip in order to

digest chromatin. The reaction was incubated at 25 °C for 15 min on bench-top shaker at 650 rpm and for another 30 min on ice. 15 µL of 5M NaCl was added to stop the reaction and samples were incubated on a spinning wheel for 45 min. Samples were spun at 13000 rpm for 10 min at 4 °C and final nuclear extract was collected. The pellet was resuspended in 50 µL NE1 buffer and kept at -80 °C. 10% of each nuclear extract (15-16.5 µL) was loaded to SDS-PAGE gel and protein of interest was detected by Western blotting. It is important to immunoblot an additional protein as reference on the same membrane in order to normalize the signal acquired from the samples.

Immunoprecipitation experiments: FLAG immunoprecipitation was carried out by using ANTI-FLAG® M2 Affinity Gel (Sigma, A2220). Beads were washed twice with 750 µL H2O and once with 750 µL NE1 buffer to calibrate the beads. 50 µL calibrated bead slurry wand 750 µg of total nuclear extract were mixed in a tube, which was incubated for 3 hours at 4° C whilst rotating on a wheel. Beads were pelleted at 2500 rpm and supernatant was kept to detect unbound protein amount. Next, beads were washed 4 times with NE1 buffer containing 200 mM NaCl, for 10 min each. After the last wash, 70 µg Flag peptides was added in 86 µl NE1 buffer containing 150 mM NaCl and 5 mM DTT. Beads were rotated on a wheel at 4° C overnight and supernatant was collected next day.

For endogenous protein immunoprecipitation, 5 µg of antibody against the target protein was mixed to 750 µg total nuclear extract in siliconized microtubes. Tubes were incubated at 4° C for 2 hours to allow antibody binding to proteins. Dynabeads® Protein G beads were pre-cleared and calibrated as described above. 25 µl of beads were added to each

sample and incubation was continued for another 3 hours. Washing step was done as described above. After washing of impurities, samples buffer was directly added to beads and purified proteins were denatured for the western blotting at 90° C for 6 min.

For Strep2-tagged protein purification, 25 µL MagStrep "type 2" Beads (IBA, 2-1612-002) per IP was washed twice with Strep2 wash buffer. 750 µg total nuclear extract was diluted in 1 mL Strep2 wash buffer and transferred to a siliconized microtube. Calibrated beads were mixed to the IP sample and incubated at 4°C for 4 hours. After the binding reaction, IP samples were washed 3 times with 1 mL Strep2 wash buffer. Beads were resuspended in 1X sample buffer containing 100 mM DTT and boiled at 90° C for 6 min. Specific proteins were detected by Western blotting.

Purification of 6xHis-tagged proteins:

IPTG Induction: A single bacterial (BL-21 or ER2566) colony expressing the 6xHis-tagged recombinant protein was grown ON at 37˚C in 50 ml LB containing the appropriate antibiotics. Next day, the overnight culture was diluted 20 times in fresh growth medium. The culture was incubated at 37˚C until the OD600 reached 0.25. In order to calibrate the IPTG induction, the culture was incubated at 25˚C for about 1 hour. When OD600 reached at 0.5, the culture was induced by 1mM IPTG. After 5 hours of culturing at 25˚C, the bacterial pellet was collected by centrifugation at 4000 rmp for 15 min at 4˚C. To collect the bacteria thoroughly, some medium was left in the centrifugation tubes, bacterial pellet was resuspended in 40 mL LB and the pellet was transferred to a 50 mL falcon tube. The culture was spun again and the supernatant was discarded. At this point the pellet was either used directly for protein purification or stored at -80˚C until needed.

Lysis: To lyse the bacteria, 5 mL ice-cold lysis buffer (containing fresh protease inhibitors) was added to resuspend the pellet, which was then incubated on ice for 15 min. If the pellet was frozen prior to lysis, the pellet was warmed at room temperature until the cells can be resuspended by a Pasteur pipette in lysis buffer. The solution should become viscous after successful lysis. Next, the solution was sonicated for 5 min on ice, at least three times for each sample or until the solution became clear. 10% Triton-X was mixed gently to the sonicated sample to achieve a 1% final concentration. The lysate was spun at 13000 rmp for 30 min at 4˚C. The supernatant containing soluble proteins was passed through a 0.45 µm filter (Millipore). The insoluble proteins in the pellet were resuspended in equal volume of buffer to the soluble fraction, and homogenized by a Dounce homogenizer. At this point, the soluble and insoluble fraction were run on SDS-PAGE gel and stained by coomassie to determine the efficiency of IPTG induction, and bacteria lysis.

Purification: For 200 mL culture pellet, 1 mL of Chelating Sepharose Fast Flow (GE Healthcare, 1705750) was used. First, the beads were transferred to a 25 mL plastic column and washed using 1 column volume (CV) of distilled water. Then, 1 CV of 0.5 M NiSO4 was mixed to charge the beads. The activated beads were washed by 2 CV of protein purification wash buffer. In order to reduce the non-specific binding to beads, imidazole was added to the soluble fraction to a final concentration of 20 mM. Then, the soluble fraction was incubated with the beads at 4˚C for 1 hour. After binding, the flowthrough was collected in a falcon tube. The beads were washed with 4 CV of wash buffer L and 1 CV of wash buffer H, and a sample from the first wash was kept to

determine later the binding efficiency. Next, the protein of interest was eluted from the beads with elution buffer for 6xHis purification. 1 mL of elution buffer was mixed with the beads and incubated for 5 min before collecting the eluate. This was repeated 3 more times. The eluted fractions were analyzed on coomassie stained SDS-PAGE and if necessary combined later.

Buffer exchange: The elution pool was concentrated as described below. Next, the concentrated sample was diluted with dialysis buffer at least 10 times and re-concentrated, twice. Glycerol was added to the concentrated sample to 10% final concentration. Finally, the purified protein was aliquoted and kept at -80˚C.

Purification of GST-tagged proteins:

The IPTG induction and lysis steps were performed as described above.

Purification**:** For 200 mL culture pellet, 2 mL of Glutathione Sepharose® 4 Fast Flow (GE Healthcare, 17513201) was used. First, the beads were transferred to a 25 mL plastic column and washed using 1 CV of distilled water to remove residual ethanol. Then beads were washed using 2 CV of protein purification wash buffer. The soluble fraction was incubated with the beads at 4˚C for at least 30 min. After binding, the flowthrough was collected in a falcon tube. The beads were washed with 4 CV of wash buffer L and 1 CV of wash buffer H, and a sample from the first wash was kept for analysis of the binding efficiency. Next, the protein of interest was eluted from the beads by fresh elution buffer for GST purification. 1 mL of elution buffer was mixed with the beads and incubated for 5 min before collecting the eluate. This was repeated 3 more times. The eluted fractions were analyzed on coomassie stained SDS-PAGE and if necessary combined.

Buffer exchange: The concentrated protein was dialyzed overnight in dialysis tubing with 10kDa MWCO (Thermo Fisher) in 0.5L dialysis buffer. The next day, the dialysis buffer was replaced with fresh one twice in 8 hours. The dialyzed sample was concentrated as described below. Glycerol was added to the concentrated sample to a final 10% concentration. Finally, the purified protein was aliquoted and kept at -80˚C.

Gel filtration and anion exchange: Gel filtration was performed using Superdex 200 columns (GE healthcare) in GF buffer. Collected fractions were run on an SDS-PAGE and analyzed by coomassie staining. Anion exchange was carried out using Mono Q column (GE Healthcare) pre-equilibrated with 10 CV of Anion exchange buffer A on an AKTA FPLC system. The program was a 20 CV gradient of 0-100% Anion exchange buffer B with 0.5ml fractions. The flow through and fractions of interest were collected, run on an SDS-PAGE and the gel stained with coomassie.

Concentrating proteins: Proteins were concentrated sing a Vivaspin columns (GE Healthcare) with 10kDa molecular weight cut off (MWCO). Before applying the protein sample, the Vivaspin columns were equilibrated with appropriate protein buffer and centrifuged at 5000 rpm at 4°C.

Storage of recombinant proteins: Once concentrated, all recombinant proteins were aliquoted into flat capped PCR tubes and flash frozen in liquid nitrogen. Aliquots were kept at -80°C and thawed on ice when needed.

***In* vitro DNA binding and Electrophoretic Mobility Shift Assays (EMSA):**

For low-wide electrophoresis gels, glass plates were sealed with 3% agarose before loading of 5% native TGE gel, which was cooled at cold room, in ice-cold 1X TBE buffer for at least an hour. A master mix for EMSA binding reaction was prepared by mixing 2 µL of 10X EMSA buffer, 2 µL of 500 ng/µL Poly(dI-dC) (Sigma, P4929-5UN), 3 µL of 40% glycerol, 1 µL of 5% TWEEN 20, 1 µL of dH2O and desired amount of IR800-labelled DNA. Unless stated otherwise, the master mix was added to protein of interest, which had the 5 µM final concentration in 20 µL reaction volume. The binding reaction was incubated on ice for 30 min. Samples were loaded on a pre-cooled non-denaturing 5% TGE gel and run at 150 V for 45 min in cold room. The gels were scanned on Licor Odyssey scanner.

Systematic evolution of ligands by exponential enrichment (SELEX):

The first round of SELEX was performed as a regular DNA band-shift experiment, which involved the binding of recombinant zinc finger protein to IR800-labelled random-DNA library. After detecting the band-shifts using Licor Odyssey scanner, the DNA-protein complex was cut out of the gel and PAGE-purified as described above. The purified DNA was amplified by PCR and used in the next round of EMSA. SELEX cycles were repeated until the enrichment of band-shifts reached to a maximum, which was determined by the quantification of signal obtained from band shifts. Fully-enriched DNA was cloned to a linear vector using CloneJET PCR Cloning Kit. Single colonies were used as templates to amplify and sequence individual DNA sequences of the enriched DNA library. Only one sequence was kept if two or more sequences shared the exact same DNA sequence.

MEME-suite (http://meme-suite.org/tools/meme) was used to identify shared motives among the unique sequences.

2.2.3. **Cell culture**

Bacterial culture:

All bacteria strains (DH5α, BL21 and ER2566) were grown at 37°C in liquid LB media or on solid LB agar plates. DH5α strain was used for cloning and plasmid amplification purposes, while BL21 and ER2566 were used for recombinant protein production. Final concentrations of antibiotics used were 100 µg/µl and 50 µg/µl for Ampicillin and Kanamycin, respectively, in liquid culture to select positive colonies after transformation.

Mouse embryonic stem cell culture:

Mouse ES cells were culture in MEM media (Thermo Fisher Scientific) containing 10% (v/v) FBS (Sigma), 1% (v/v) non-essential amino acids (Thermo Fisher Scentific), 1% (v/v) PSG (Thermo Fisher Scentific), 1% (v/v) sodium pyruvate (Thermo Fisher Scentific), 0.05% (v/v) 2-mercaptoethanol (Thermo Fisher Scentific) and homemade LIF. Cells were grown on either dishes or flasks, which were coated with 0.1% gelatin in PBS prior to use. The cells were grown in a designated incubator at 37 °C and 5% CO2. Medium was renewed daily if the cells were above 50% confluency. ESCs were eıther harvested or split at 85-95% confluency. Cells were washed twice with warm PBS and then detached by trypsinization using 0.1% trypsin (Thermo Fisher Scentific) in PBS. ESCs were frozen in normal culture media containing 10% DMSO. Usually, at least 3 million cells were frozen per cryotube in order to achieve high viability after thawing.

Initially, cells to be frozen were kept at -80 °C and then moved to liquid nitrogen for long term storage.

G9a/GLP inhibitor treatment and H3K9me2 recovery:

Wild type ESCs were cultured in normal culture media containing either (UNC0638, Sigma, U4885) at final concentration 500 nM or DMSO for 2 days. For H3K9me2 recovery experiments 0.75 million inhibitor treated cells were seeded on 10-cm dishes for each time point. As a control, the same amount of untreated cells were cultured along with the others. The inhibitor was removed by replacing the culture media with normal ES cell media containing no inhibitor at 48 h, 24 h, 12 h, 6 h and 0 h before collecting the cells. Nuclear proteins were extracted from the harvested cells and H3K9me2 levels were analyzed by Western blotting.

WIZ mRNA knockdown by short hairpin RNAs:

Clone ID	Insert Sequence
shRNA1	TGAGCTCTGTGGCCTTTACTT
shRNA2	GGAAGAGTGGGTACGGCATTT
shRNA3	CTCACCTTTGACCAAGAAGTT
shRNA4	GTGACTGAGTGGTCTGTCAAT
shRNAc	ggaatctcattcgatgcatac

Table 2. 1 shRNA against WIZ mRNA

A set of plasmids containing short hairpin RNA against WIZ were ordered from QIAGEN (#336313 KM36796N).

The plasmids were first transformed to DH5α and isolated in large quantity by maxiprep. 100 µg of plasmid was digested by 5 µL SacI enzyme to linearize, and then purified with

Phenol-Chloroform extraction. The pelleted DNA was dissolved in 100 µL ddH2O. 10 million ES cells were harvested and washed with PBS at least twice. The cells were resuspended in 550 µL PBS and transferred to a Gene Pulser®/Micropulser™ electroporation cuvette (#1652081). The linearized plasmid was mixed with cells very well and transfected at 240 V, 500 µF and maximum resistance for 10 ms (Gene Pulser Xcell™). The transfected cells were kept at room temperature for 10-15 min and seeded on at least three gelatinized 10 cm culture dishes. The medium was renewed 24 hours after transfection and 300 µg/mL Geneticin (G418) was added to cell media to select positive cells. Once the positive colonies were grown to the desirable size, the cells were washed with PBS twice and single colonies were picked by 10 µl tips under the microscope. The picked colonies were first incubated in a droplet of trypsin that was put on a dry, non-gelatinized 24-well culture plate. After 2 minutes, the colony was resuspended by pipette up and downs and transferred to ES cell medium containing well.

2.2.4. **Antibodies**

Name	Type	Source, catalog number	Western blot dilution factor	ChIP dilution factor
α-DNMT3a	Mouse monoclonal	Abcam, ab13888	1:1000	
α-DNMT3b	Mouse monoclonal	Abcam, ab13604	1:1000	
α- HDAC1	Rabbit polyclonal	Santa Cruz, sc-7872/D171	1:500	
α- H3K4me3	Rabbot polyclonal	Active motif, 39915		1 µg per 5 µg of chromatin
α- H3K9me2	Mouse monoclonal	Abcam, ab1220	1:2000	1 µg per 5 µg of chromatin
α- H4	Rabbit polyclonal	Milipore, 07-108	1:500	
α- G9a	Mouse monoclonal	Perseus Proteomics, PP-A8620A-00	1:1000	
α- G9a	Rabbit polyclonal	Abcam, ab40542	1:1000	1 µg per 5 µg of chromatin
α- GLP	Mouse monoclonal	Perseus Proteomics, PP-B0422-00	1:1000	
α- LSH	Mouse monoclonal	Santa Cruz, sc-46665/H0205	1:1000	
α- Ran	Mouse monoclonal	Abcam, ab11693	1:5000	
α- Strep2	Mouse monoclonal	IBA, 2-1507-001	1:1000	
α- WIZ	(Rabbit polyclonal	Homemade by Yoichi Shinkai	1:2000	

Table 2. 2 Primary antibodies used in the experiments

Name	Type	Source, catalog number	Western blot dilution factor
IRDye® 680LT	Donkey anti-Rabbit IgG (H + L)	LI-COR, 926-68023	1:10000
IRDye® 800CW	Donkey anti-Mouse IgG (H + L)	LI-COR, 926-32212	1:10000
IRDye® 800CW	Donkey anti-Rabbit IgG (H + L)	LI-COR, 926-32213	1:10000

Table 2. 3 Secondary antibodies used in the experiments

2.2.5. **Primers**

PCR primers used in this thesis are listed below:

Name	Target/Description	Sequence (5' to 3')
ZF2-3_EcoRI_F	mWIZ ZF 2-3	AAAGAATTCATTTATGGTGATGCACAGAGC
ZF2-3_SalI_R	mWIZ ZF 2-3	TTTGTCGACATACCATTCGGTAACGC
ZF3-4_EcoRI_F	mWIZ ZF 3-4	AAAGAATTCTCAGGTCCGGAACCGACC
ZF3-4_SalI_R	mWIZ ZF 3-4	CGGGCGTCGACTGACCATTCTGT
ZF4-5_EcoRI_F	mWIZ ZF 4-5	AAAGAATTCGCACGTGCAGAACCGG
ZF4-5_SalI_R	mWIZ ZF 4-5	TTAGTCGACACACCATTCGGTCAC
G9a_EcoRI_Strep2_F1	G9a cDNA long in pCAG-G9a vector	AAAGAATTCATGTGGTCACATCCACAGTTCGAAAAGGGCGGTGGATCAGGAGGTGGCAG TGGAGGTTCAGCATGGTCTCACCCTCAATTCGAGAAGCGGGGTCTGCCGAGAGGGAG
G9a _R1	G9a cDNA long in pCAG-G9a vector	GAGGAATGGTCACCTCCTGGGCCTCT
G9a_F2	G9a cDNA long in pCAG-G9a vector	TGGAATGCCACCCCGACTTTCGT
G9a_EcoRI _R2	G9a cDNA long in pCAG-G9a vector	TTTGAATTCTCAGGTGTTGATGGGGGG
GFP_EcoRI_Strep2_F1	GFP in pEGFP-N1 vector	AAAGAATTCATGTGGTCACATCCACAGTTCGAAAAGGGCGGTGGATCAGGAGGTGGCAG TGGAGGTTCAGCATGGTCTCACCCTCAATTCGAGAAGGTGAGCAAGGGCGAGGAGCT
GFP_EcoRI _R1	GFP in pEGFP-N1 vector	CCTGAATTCTTACTTGTACAGCTCGTCC
G9a_5' sequencing_F1	G9a cDNA long in pCAG-G9a vector	CTGAACTCTGGTAGCCTGT
G9a_5' sequencing_F2	G9a cDNA long in pCAG-G9a vector	GCAATGCTGCCATCCTTAAG
ZF4-d5-F	pGEX-4T1-ZF4-5 plasmid	AGCACCGAAGCATGATGTGAACTGT
ZF4-d5-R	pGEX-4T1-ZF4-5 plasmid	ACAGTTCACATCATGCTTCGGTGCT
ZFm4-5-F	pGEX-4T1-ZF4-5 plasmid	ATTTTTTGAGAATCGCGCAGGTCTGAG
ZFm4-5-R	pGEX-4T1-ZF4-5 plasmid	CTCAGACCTGCGCGATTCTCAAAAAAT
ZF4-m5-F	pGEX-4T1-ZF4-5 plasmid	AAATCGTGCAGCACTGGCAAGCCA
ZF4-m5-R	pGEX-4T1-ZF4-5 plasmid	TGGCTTGCCAGTGCTGCACGATTT
WIZ qPCR F1	mouse WIZ transcript varian 1, spans ZF6	CTCGTGGTGGTGAGGAAGTC
WIZ qPCR R1	mouse WIZ transcript varian 1, spans ZF6	ACAGAACCTGCACTTGAGGG
WIZ qPCR F2	mouse WIZ transcript varian 1, spans ZF6	CCCTCAAGTGCAGGTTCTGT
WIZ qPCR R2	mouse WIZ transcript varian 1, spans ZF6	GGGGATCTGGAATGCTGTGT
Star-Selex-R2	2N10 SELEX template	TAGTCTTAGGAGATCTC
MfeI_WIZ ZF1-6_F	WIZ cDNA - Transcription variant 1	aaaCAATTGgaggggttactggca
NotI_WIZ ZF1-5_R	WIZ cDNA - Transcription variant 1	aaaGCGGCCGCttaagaggcatctga
NotI_WIZ ZF1-6_R	WIZ cDNA - Transcription variant 1	aaaGCGGCCGCttagggtgcctcta
TEV_ZF3-4_F	GST ZF 3-4 plasmid	aaaGAATTCgaaaacctgtattttcagtctCAATTGTCAGGTCCGGAACCGACC
ZF 3-d4-F	GST ZF 3-4 plasmid	TGATATCCGCTGAGAATTTTGTGGC
ZF 3-d4-R	GST ZF 3-4 plasmid	GCCACAAAATTCTCAGCGGATATCA
ZF d3-4-F	GST ZF 3-4 plasmid	TGATATTCGTTGAGAATTTTGCGGT
ZF d3-4-R	GST ZF 3-4 plasmid	ACCGCAAAATTCTCAACGAATATCA
ZF m3-4-F	GST ZF 3-4 plasmid	TTTGAGAATCGCGCAGGTCTGAGCA
ZF m3-4-R	GST ZF 3-4 plasmid	TGCTCAGACCTGCGCGATTCTCAAA
WIZ-shRes-F1	WIZ cDNA - Transcription variant 1	aaaGTTaaccatgtggtcacatcc
WIZ-shRes-R1	WIZ cDNA - Transcription variant 1	AAATATAGCCCGCATAGTTCGcagcaggcctcggtg
WIZ-shRes-F2	WIZ cDNA - Transcription variant 1	CGAACTATGCGGGCTATATTTtgaaaaccgcaaagccc
WIZ-shRes-R2	WIZ cDNA - Transcription variant 1	tttGCGGCCGCttagggtgcctcta
ChIP-Wfdc15a-F1	Mouse ChIP chromatin	GACGTTTGAATCTGCCGCTG
ChIP-Wfdc15a-R1	Mouse ChIP chromatin	TGTTCCGGGTTGATGCACTT
ChIP-Wfdc15a-F2	Mouse ChIP chromatin	CATGCTGAGGCAAAGGAGGA
ChIP-Wfdc15a-R2	Mouse ChIP chromatin	TCCATGTTCACAGGGCAGTC
ChIP-actinB_in3-F	Mouse ChIP chromatin	AAGTGGCCTTGGAGTGTG
ChIP-actinB_in3-R	Mouse ChIP chromatin	CAAGGAGTGCAAGAACACAGC
ChIP-actinB_pause-F	Mouse ChIP chromatin	AGGTTGGGTGGGTGT
ChIP-actinB_pause-R	Mouse ChIP chromatin	AGACTTCATTAAAGGGCCTGT
ChIP-Magea2 –F	Mouse ChIP chromatin	CTCCAGAACAAAATGGCGCA
ChIP-Magea2 –R	Mouse ChIP chromatin	CCTGAGTTTAACGGAACCTCT
SEQ-WIZ-shRES	pCAG-WIZ-shRES plasmid	GCGCTGCGAAACTTCTTTG
T15-F-FL	WIZ ZF 3-4 protein	TGAGGTTGTTGCGGGTGGGTGTG
T15-R	WIZ ZF 3-4 protein	CACACCCACCCGCAACAACCTCA
T15L-F-FL	WIZ ZF 3-4 protein	TGGTGGGTGTGCGGGTGGGTGTG
T15L-R	WIZ ZF 3-4 protein	CACACCCACCCGCACACCCACCA
T15-R-FL	WIZ ZF 3-4 protein	TGAGGTTGTTGCGGAGGTTGTTG
T15R-R	WIZ ZF 3-4 protein	CAACAACCTCCGCAACAACCTCA
SEQ-WIZ shRES – F1	pCAG-WIZ-shRES plasmid	CGAACTATGCGGGCTATATTT
T15-Shrt-flkIR-F	WIZ ZF 3-4 protein	TCTGGAGGTTGTTGCGGGTGGGTGTGGAGA
T15-Shrt-flnk-R	WIZ ZF 3-4 protein	TCTCCACACCCACCCGCAACAACCTCCAGA
T15-LfullIR-F	WIZ ZF 3-4 protein	GCCTCTGTATGTAGATCTGGAGGTTGTTGCGGGTGGGTGTG
T15-Lfull-R	WIZ ZF 3-4 protein	CACACCCACCCGCAACAACCTCCAGATCTACATACAGAGGC
T15-RfullIR-F	WIZ ZF 3-4 protein	GAGGTTGTTGCGGGTGGGTGTGGAGATCTCCTAAGACT

T15-Lfull-R	WIZ ZF 3-4 protein	AGTCTTAGGAGATCTCCACACCCACCCGCAACAACCTC
WIZ-tgIR-F	WIZ ZF 3-4 protein	GAGTCTCACTCACGCGCATTCCATTCCATCAGATACTAGTACGGTCAG
WIZ target-R	WIZ ZF 3-4 protein	CTGACCGTACTAGTATCTGATGGAATGGAATGCGCGTGAGTGAGACTC
WIZ-shRES-F2	WIZ cDNA - Transcription variant 1	CGAACTATGCGGGCTATATTTgaaaaccgcaaagccc
ChIP-Actin-Pau-F1	Mouse ChIP chromatin	CAGGTTGGGTGGGTGTGGAG
ChIP-Actin-Pau-R1	Mouse ChIP chromatin	ACAGGCCCTTTAATGAAGTCT
ChIP-Actin-Pau-F2	Mouse ChIP chromatin	AGGAAGGTGTGAAGCCTATG
ChIP-Actin-Pau-R2	Mouse ChIP chromatin	GTCATCCTGGAAATCAGCC
T15-FULL-FL-F	WIZ ZF 3-4 protein	TGCCTCTGTATGTAGATCTGGAGGTTGTTGCGGGTGGGTGTGGAGATATCCTAAGACTA
T15-FULL-R	WIZ ZF 3-4 protein	TAGTCTTAGGATATCTCCACACCCACCCGCAACAACCTCCAGATCTACATACAGAGGCA
T15-FULL-FL-R	WIZ ZF 3-4 protein	TAGTCTTAGGATATCTCCACACCCACCCGCAACAACCTCCAGATCTACATACAGAGGCA
T15-FULL-F	WIZ ZF 3-4 protein	TGCCTCTGTATGTAGATCTGGAGGTTGTTGCGGGTGGGTGTGGAGATATCCTAAGACTA
ATR-CON-FL-F	WIZ ZF 3-4 protein	TGACGCTGCATGAATATATAGTCTTTAATTATCGTAGTTCTACATATATCTTAAATCTA
ART-CON-R	WIZ ZF 3-4 protein	TAGATTTAAGATATATGTAGAACTACGATAATTAAAGACTATATATTCATGCAGCGTCA
Seq-WIZ upstream-R	pCAG-WIZ shres plasmid	caggtaatgagaagatggaaag
Seq-WIZ downstream-F	pCAG-WIZ shres plasmid	gtacggcatttacagcggc
Seq-WIZ1-5 downstream-F	pCAG-WIZ 1-5 shres plasmid	gagaaccagggcttgaagct
Seq-LoxP flank-F	RMCE (pLox) plasmid	gtaaaacgacggccagtgaat
Seq-LoxP flank-R	RMCE (pLox) plasmid	caggaaacagctatgaccatg
ChIP-Act Pau-F1	Mouse ChIP chromatin	AGGTTGGGTGGGTGT
ChIP-Act Pau-R1	Mouse ChIP chromatin	AGACTTCATTAAAGGGCCTGT
WIZ-Strep2tag-F	pCAG-WIZ shres plasmid	ccctcaattcgagaaggaggggttactggcaggtg
WIZ-Strep2tag-R	pCAG-WIZ shres plasmid	ctgccagtaacccctccttctcgaattgagggtgaga
Selex-EcoRI-F1	All SELEX DNA	aaagaattcGCCTCTGTATGTAGATCTG
Selex-MfeI-R1	All SELEX DNA	aaacaattgAGTCTTAGGAGATCTC
WIZ target-EcoRI-F1	All SELEX DNA	aaagaattcGAGTCTCACTCACGC
WIZ target-MfeI-R1	All SELEX DNA	aaacaattgCTGACCGTACTAGTATCT
CMV-TET-O-TdT-R2	IPC0654 - TetO-FRT5-Tdt-FRT-Htk-FRT3	tttgaattcttacttgtacagctcgtccat
pJet1.2-F2	pJet linear vector	GGAGCAGGTTCCATTCATTGT
pJet1.2-R2	pJet linear vector	CATGAGAGTCGATTGCCAAG

Table 2. 4 PCR primers used in this thesis

Chapter III – Characterization of the G9a Complex

3.1. Introduction

G9a, also known as EHMT2 and KMT1C, is a SET domain-containing protein, which acts as a major histone methylase for H3K9me2 mark in mammalian cells. G9a adds methyl groups to H3K9 in euchromatic regions of the genome (Tachibana et al. 2001) and is responsible for ~90% of H3K9me2 in the genome, while the remaining H3K9me2 is maintained at pericentric heterochromatin by Suv39h1 (Peters et al. 2003). Aside from its primary product – H3K9me2, which enables HP1 binding to chromatin, G9a can mono- and dimethylate H1.4, (Trojer & Reinberg 2007). Additionally, G9a can form a homodimer as well as a heterodimer by interacting with another histone methylase, G9a-like protein (GLP), to form a heterodimer. Dimeric complexes are set up in a SET domain-dependent manner (Tachibana et al. 2005). The catalytic domains of GLP and G9a are 87% identical, while two proteins share 63% sequence similarity (Ogawa et al. 2002); Tachibana et al. 2005). The heteromeric G9a/GLP shows greater stability, which is further enhanced by a C2H2-type zinc finger protein, widely-interspaced zinc finger protein (WIZ), that interacts with the G9a/GLP heterodimer through its 6^{th} zinc finger (Ueda et al. 2006). The G9a-GLP-WIZ complex occurs by interaction of stoichiometric amounts of each counterpart. Therefore, I will refer it as "the core complex".

The product of the G9a/GLP heterodimer, H3K9me2, is a very abundant modification in mammalian cells. Slightly more than half of the genome is covered with this modification in both mouse embryonic stem cells (ESCs) and neurons derived from in vitro differentiation of these ESCs (Lienert et al. 2011). It was shown that G9a and GLP can bind to H3K9me1 and H3K9me2 via the ankyrin repeat domain (ANK). The dissociation constants of G9a to H3K9me1 and H3K9me2 were found to be 14±3 μM and 6±2 μM, respectively, while ANK domain of GLP had 5±0.4 μM and 7±1 μM for to H3K9me1 and H3K9me2, respectively. On the other hand, affinity to empty histone tails was found to be much lower with a dissociation constant greater that 150 μM (Collins et al. 2008). Being able to "read" H3K9me1 and H3K9me2 via the ANK domain, the G9a/GLP heterodimer has a high affinity for the chromatin that is marked by these modifications. Furthermore, fluorescence recovery after photobleaching (FRAP) experiments indicated the tight chromatin association of G9a (Purcell et al. 2012). Although G9a/GLP exhibit a high affinity to half of the genome due to H3K9me1/2, the initial recruitment of G9a/GLP to genomic sites that lack either H3K9me1 or H3K9me2 may require other proteins, since both G9a and GLP have little affinity to empty histone tails. This type of recruitment might be necessary during ES cell differentiation when the cells need to silence certain active genes that lack the silencing H3K9me1/2. Considering that neither G9a nor GLP contain a DNA/RNA binding domain, binding partners of the core complex might be responsible for the initial recruitment.

Over the years, a number of techniques have been developed to identify binding partners of specific proteins of interest. These methods are described as "binary" or "co-complex",

if the aim is to detect direct interaction between two proteins or an interaction map of a group of proteins, respectively (De Las Rivas & Fontanillo 2010). Most notable example of a binary protein-protein interaction technique is yeast two-hybrid system (Fields & Song 1989). In the yeast two-hybrid assay, the "bait" protein is fused to a DNA-binding domain, while the "prey" protein is fused to a transcription activator domain. A reporter system such as β-galactosidase assay becomes activated when "pray" and "bait" proteins interact in cells.

Co-complex methods are used to identify direct and indirect protein-protein interactions. Mostly, a protein of interest is tagged and co-immunoprecipitated (Co-IP) along with the other proteins in a complex. The purified complex can be analysed by mass spectrometry that can provide great sensitivity for identification of less abundant proteins in purified complexes.

3.1.1. **Aims**

Previous studies have provided detailed information about the functional role of the G9a complex. However, a mechanistic understanding of initial recruitment of the G9a complex to chromatin remains largely unknown. Evidence indicating that G9a and GLP have a low affinity (~150 μM) to non-modified histone, and the lack any DNA or RNA binding motifs of G9a/GLP, suggest that there should be auxiliary factors coordinating the binding of G9a to specific sites in chromatin that lack H3K9me1/2. To investigate these factors, I aimed to purify the G9a/GLP complex using affinity-based methods and identify interacting proteins with DNA/RNA binding properties. Later, the purified complex was

analysed for its content by mass spectrometry to determine the candidate binding partners of the G9a complex.

3.2. Affinity purification of FLAG-tagged G9a from ES cells

The mouse ES cell line expressing FLAG-tagged G9a was generated by Yoichi Shinkai laboratory (Tachibana et al., 2005). These cells are G9a-/- and stably express a FLAG-tagged short isoform of G9a. To confirm the identity of these cells and to investigate the expression levels of tagged G9a I analysed by Western blots G9a and GLP in wild-type (WT), G9a-FLAG and G9a-null cells. The endogenous G9a appeared as two distinct bands corresponding to the large and the short variant, while G9a-FLAG appeared as a single band slightly bigger than the short variant due to FLAG tag (Figure 3.1A, (1) and (2)). As expected, G9a-null cells did not have any detectable G9a and had reduced level of GLP. G9a-FLAG has slightly elevated levels of GLP, which probably due to mild overexpression of G9a in these cells comparing to WT.

Next, to purify the G9a complex, I performed an affinity purification experiment using the nuclear extracts from the G9a-FLAG and WT cells. In this experiment, nuclear extracts were incubated with M2-FLAG affinity beads. After binding of G9a-FLAG to the resin, I performed washes in different stringency conditions, wash buffer contained either 150 mM or 200 mM final NaCl concentration. The G9a complex was eluted from the beads by using FLAG peptides as described in Materials and Methods. The purification efficiency (purified over input) was 0.2% and 0.15% for 150 mM wash and 200 mM wash,

respectively. Nonetheless, the negative control samples did not contain any detectable G9a (Figure 3.1B, compare (3) and (6), (9) and (12)). Next, I wanted to investigate whether the purified G9a complex was good quality for a mass spectrometry experiment or not. I loaded 15% of the eluate from the affinity purification experiment on SDS-PAGE and silver-stained the gel. Although G9a-FLAG was not conspicuous on the gel, there was a number of other proteins present on the same lane (Figure 3C, (1), (3), (5) and (6)). On the other hand, WT lanes were mostly clear indicating that the purification of G9a-FLAG and its interacting proteins was successful (Figure 3C, (2) and (3)).

3.3. Mass spectrometry analysis of the purified G9a complex

With the help of Flavia Alves deLima from Prof. Juri Rappsilber's Laboratory, we performed the mass spectrometry experiment as described in the Materials and Methods section. In this experiment, WT IP samples from the pulldown experiment were used as a negative control to identify possible false positives, and the peptides identified in FLAG-G9a IP samples were matched to candidate binding partners of G9a. All identified protein appearing with more than one peptides in the FLAG-G9a samples were considered as candidate binding partners of G9a. In addition to this, the identified proteins appearing with more than one peptide in the negative control was regarded as false positives (Table 1).

A similar experiment was carried out before I started my project (Table 2). The number of proteins represented by zero or one peptide in the negative control was fewer in this

experiment, which could be explained by the inadequate washing of the beads during the pulldown experiment. Nevertheless, in both datasets, the core complex members – G9a, GLP and WIZ, appeared with high peptide numbers that were comparable to the complex members (Figure 3.2A). This indicated that both experiments were informative, but a quality assessment was required before drawing any conclusions. In the analysis performed by others, one of the candidate binding partners of G9a-FLAG was the Heterogeneous Nuclear Ribonucleoprotein L (HNRPL), which had 10 and 0 peptides in the positive and negative controls, respectively. The same protein appeared as false positive in my mass spectrometry experiment. Therefore, HNRNPL was a good candidate for testing the quality of two independent mass spectrometry analyses of the G9a complex. I used nuclear extracts of G9a-FLAG cells and carried out co-IP experiments to test whether HNRNPL co-purifies with G9a. In addition to that, I performed a reciprocal experiment, in which I used nuclear extracts of WT ES cells and performed an immunoprecipitations using antibodies against G9a, GLP and HNRNPL (Figure 3.2B). G9a co-immunoprecipitated very efficiently with GLP, but not with HNRNPL, in both IP experiments. In FLAG-IP experiment, there was a significant amount of HNRNPL in both G9a-F and WT IP sample. This clearly indicated that HNRNPL could nonspecifically interact with the M2-FLAG beads under the conditions I used for my affinity purification experiments prior to the mass spectrometry analysis. Therefore, HNRNPL is a false positive in the list of candidate G9a-interacting proteins.

As discussed previously, I sought out DNA/RNA-binding proteins that could mediate the recruitment of G9a at specific sites by focusing on the proteins identified as candidate

binding partners from my mass spectrometry data. I identified several zinc finger proteins, other than WIZ, which might be novel candidate binding partners of G9a. Among those, Zinc finger protein 462 (ZFP462) was identified in both mass spectrometry experiments with very high peptide number. It is a protein consisting of 2461 amino acid and 34 C2H2-type zinc finger domains. To date, very little is known about its DNA-binding ability, but some evidence indicates its importance in the maintenance of pluripotency (Massé et al. 2010).

Zinc finger protein 518B (ZNF518B) was another zinc finger protein consisting of 4 C2H2-type zinc fingers that had high peptide score. Using a different mass spectrometry approach, ZNF518B was shown to positively regulate the catalytic activity of G9a (Maier et al. 2015).

Lastly, Zinc finger CCCH domain-containing protein 11A (ZC3H11A) appeared with significant peptide score in my mass spectrometry experiment. ZC3H11A contains three C3H1-type zinc finger, and almost nothing is known about the function of this protein.

Among the four zinc finger proteins identified as candidate binding partners in the mass spectrometry analysis, WIZ was the only protein that was shown to interact with G9a and GLP directly (Ueda et al. 2006). One study revealed that lack of WIZ is embryonic lethal with timing similar to G9a and GLP knockouts (Daxinger et al. 2013). This indicated the crucial role of WIZ, together with G9a, in mouse development. As a zinc finger protein, it would be possible that WIZ could bind specific DNA or RNA sequences to recruit G9a to particular loci during development. However, it is not known whether or not WIZ can

bind DNA or RNA. Being one of the core G9a complex members, and the possibility of possessing a DNA/RNA binding ability gave studying of WIZ a higher priority.

I confirmed the G9a-WIZ interaction by another affinity purification experiment. First, I made cell lines expressing Strep2-tagged proteins. Linear plasmids containing a CAG promoter and the coding sequences of G9a-Strep2 or GFP-Strep2 was transfected to *G9a*$^{-/-}$ cells. Using magnetic Strep2-beads, I was able to purify the Strep2 tagged-proteins with much higher efficiency than the FLAG-affinity purification of G9a (30% of input compared 0.2% of input) (Figure 2C). These experiments confirmed that the G9a-Strep2 interacted with WIZ.

3.4. Downregulation of WIZ mRNA mildly reduces the protein levels of G9a

It was reported that the stability of G9a complex is increased by WIZ and that the knockdown of WIZ significantly reduces the levels of G9a in HeLa cells (Ueda et al., 2006). After confirming that WIZ, indeed, interacts with the G9a complex in ES cells, I wanted to investigate the functional role of WIZ in the G9a complex. To address this, I first stably knocked down (KD) WIZ in WT ES cells. I tested four different plasmids containing shRNAs that specifically target WIZ mRNA and a plasmid containing scrambled shRNA (shCtrl). Transfection of WIZ shRNA1 caused 50% downregulation, while other shRNAs caused at most 25% decrease in WIZ mRNA levels (Figure 3.1A). Thus, I generated clonal ES cell lines from the pool of cells expressing shRNA1. Next, I

screened 14 clonal ES cell lines for the best WIZ KD (Figure 3B). Among those, colony #7, #13 and #16 had ~80% reduction in WIZ mRNA levels (Figure 3.3C). WIZ KD efficiency was slightly better in colony #7 than the other two cell lines, and the result was more reproducible between replicas. Therefore, I used colony #7 as my WIZ KD cell line in the following experiments.

Next, I wanted to investigate the protein levels of the G9a complex members in WIZ KD cell lines (Figure 3.4A). Similar to the reduction in the mRNA levels, WIZ protein levels were also reduced by about 80% in WIZ KD cell lines. The G9a levels were consistently reduced by 30% in all WIZ KD cell lines. However, the GLP levels were not affected in WIZ KD #16, while the reduction was ranging between 20% to 30% in other two WIZ KD cell lines. There was also a small reduction in H3K9me2 levels, which was probably due to the reduction in G9a levels. To firmly confirm that G9a levels are reduced in WIZ KD cells, I analyzed the G9a levels in the nuclear extracts prepared by varying salt concentrations. In all conditions, WIZ KD cells contained 30% less G9a when compared to WT cells (Figure 3.4).

To determine the amount of G9a protein that is lost after WIZ KD, I extracted nuclear proteins from WT and WIZ KD cells using extracting buffers containing high or low salt concentration. As expected, high salt extraction yielded more G9a comparing to low salt extractions. However, in both stringency conditions, WIZ KD cells always had about 30% less G9a when compared to the control shRNA cells. This experiment suggested that the stability of G9a could be enhanced by WIZ in ES cells. However, the mild reduction in the G9a levels after WIZ KD suggested that the formation of G9a/GLP was mostly

independent of WIZ. To test whether the reduction of G9a levels is rescuable by 6th zinc finger of WIZ, I attempted to generate a cell line that ectopically express a Strep2-tagged zinc finger 6 of WIZ (Strep2-ZF6) in WIZ KD cells. Although I was able to make ES cells that transcribe the mRNA of Strep2-ZF6, these cells were unable to produce the recombinant protein (Data not shown). My attempts of full-length WIZ rescue have also failed, suggesting that N-terminal tagging of WIZ causes protein instability.

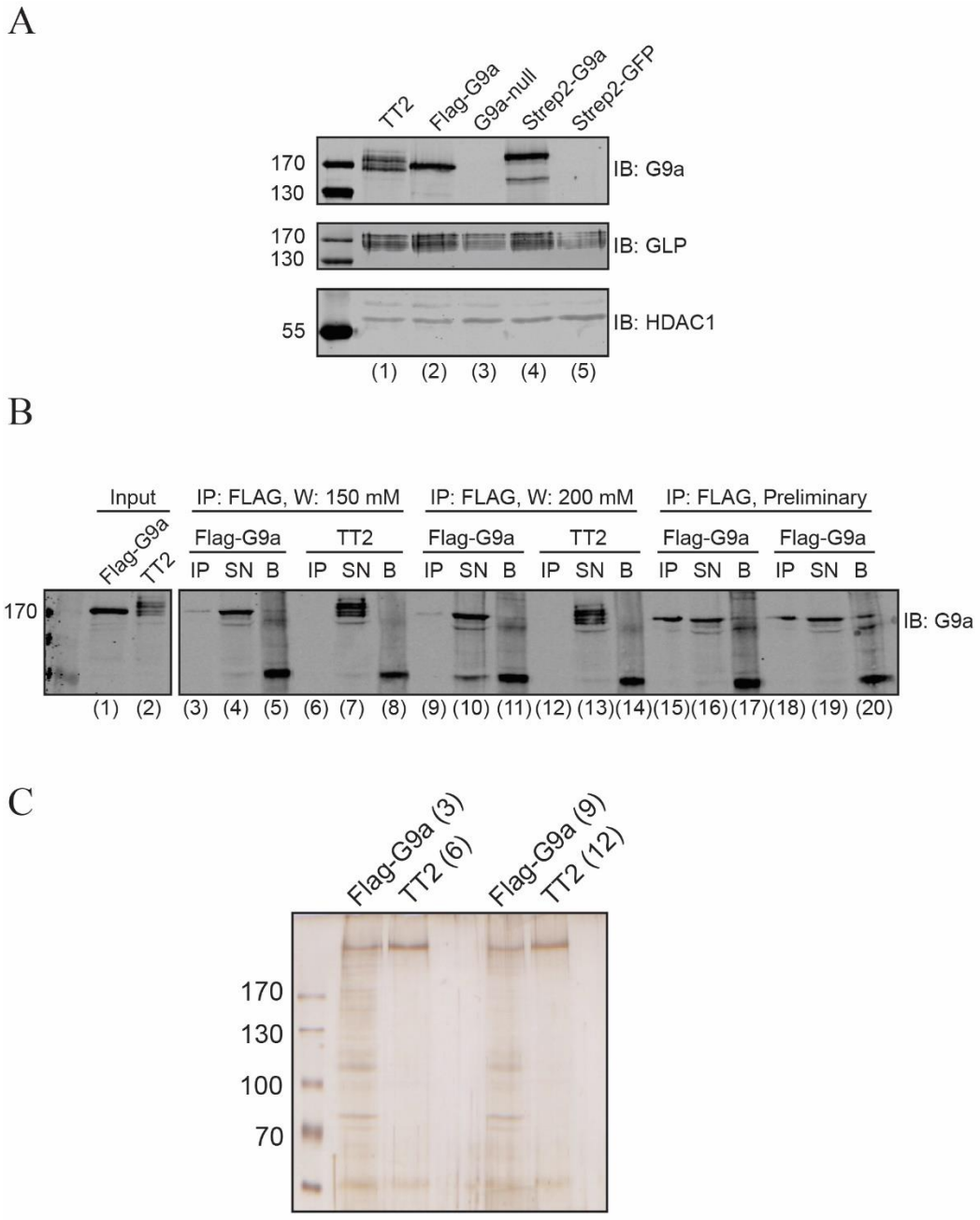


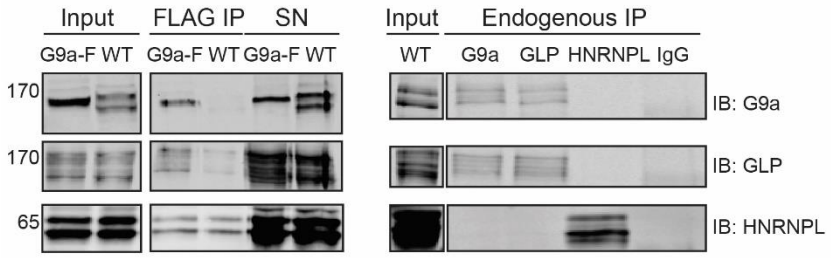
Figure 3. 1 Affinity purification of the G9a complex from ES cells expressing G9a-FLAG

(**A**) G9a and GLP detected by Western blots in the indicated cell lines. HDAC1 is a loading control. (**B**) Efficiency of G9a-FLAG affinity purification from ES nuclear extracts. TT2 cells expressing endogenous G9a are used as negative control. 40 μg total nuclear extract is used in the input lanes. 10% of supernatant after incubation with M2-FLAG beads is used in the "SN" lanes. Proteins bound to the M2-FLAG beads are eluted and 10% of the elution is loaded in "IP" lanes. Beads after the elution are resuspended in NE1 buffer and loaded in lanes labelled as "B". (**C**) Silver stained 8% SDS-PAGE gel of samples with indicated lane numbers from (B) visualises the purified proteins.

A

Description (Name, Gene name)	Number of peptides found in samples				Preliminary data	
	WT (3)	F-G9a (6)	WT (9)	F-G9a (12)	WT	F-G9a
Histone-lysine N-methyltransferase EHMT2, **Ehmt2**	0	11	0	17	0	20
Widely interspaced zinc finger protein, **Wiz**	0	12	0	12	0	18
Histone-lysine N-methyltransferase EHMT1, **Ehmt1**	0	13	0	15	0	16
Protein Zfp462, **Zfp462**	0	23	0	24	0	35
DNA (cytosine-5)-methyltransferase 3-like, **Dnmt3l**	0	8	0	8	2	3
DNA (cytosine-5)-methyltransferase 3A, **Dnmt3a**	1	4	0	2	0	1
DNA (cytosine-5)-methyltransferase 3B, **Dnmt3b**	0	2	0	6	0	0
Lymphocyte-specific helicase (LSH), **Hells**	0	7	0	5	0	0
C-terminal binding protein 2, isoform CRA_b, **Ctbp2**	0	5	0	4	0	3
Zinc finger protein 518B, **ZFP518B**	0	5	0	5	0	0
Heterogeneous nuclear ribonucleoprotein L, **Hnrnpl**	5	0	4	7	0	10

B



C

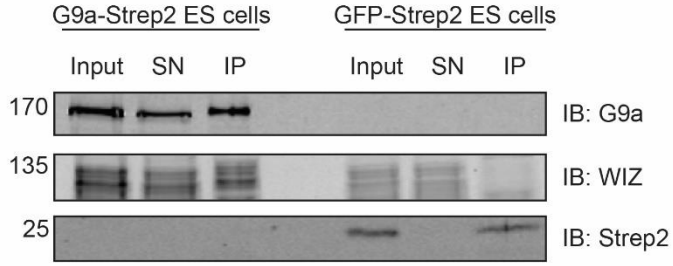


Figure 3. 2 Validation of protein-protein interactions identified by mass spectrometry

(A) Protein identified in three FLAG affinity purifications of G9a. The core complex (green) is present in all experiments. HNRPL (red) is a potential false positive interactor. **(B)** Immunoprecipitations of FLAG-tagged (left) and the endogenous G9a frim nuclear extracts of ES cells expressiong either G9a-FLAG or wild-type (WT) G9a are blotted and the Western blots probed with antibodies against G9a, GLP and HNRNPL. 40 µg of total protein is used as input, and 15% of SN and IP are loaded on to SDS-PAGE for analysis. "IP" indicates immunoprecipitated proteins and "SN" the supernatant after IP. **(C)** G9a-Strep2 and GFP-Strep2 are purified from nuclear extracts of ES cells and analyzed by Western Blots with antibodies against G9a, WIZ and the Strep2 tag. 40 µg of total protein is used as input, and 15% of SN and IP are loaded on to SDS-PAGE for analysis. "

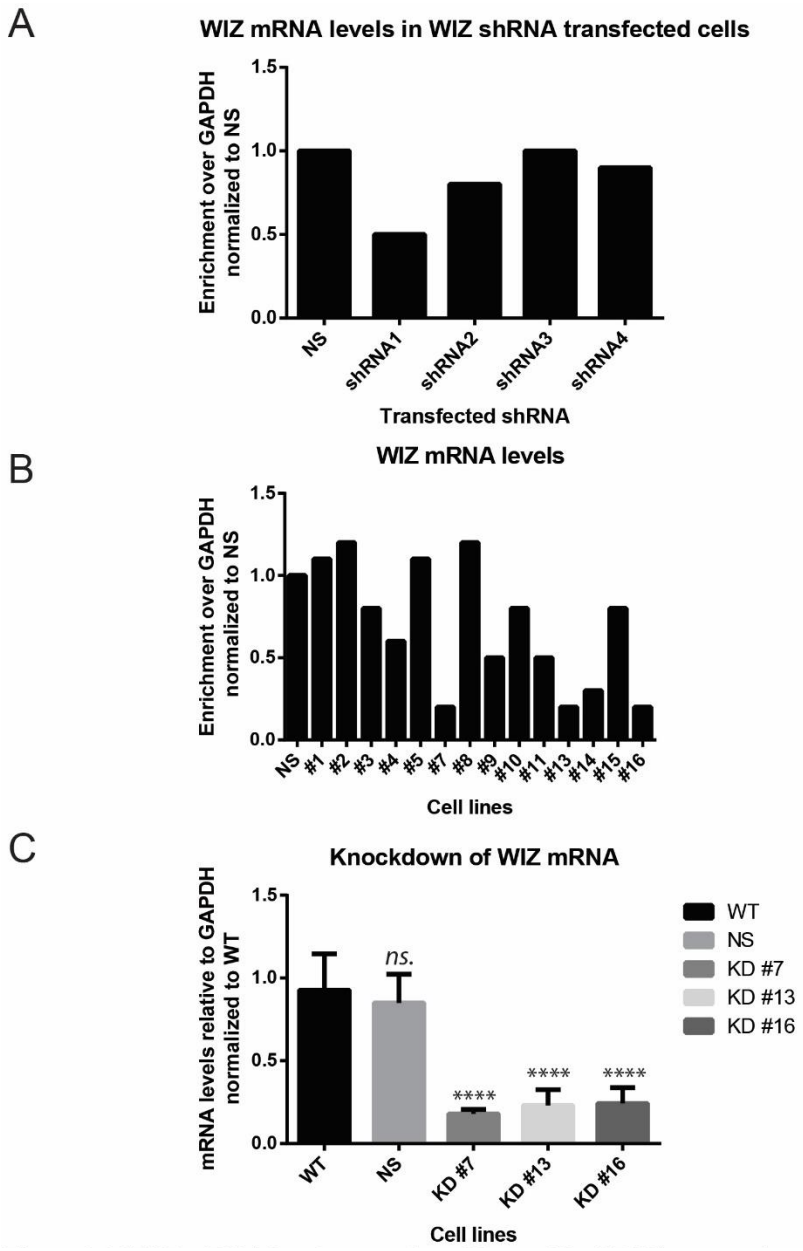


Figure 3. 3 WIZ mRNA levels are reduced by stable shRNA expression

(A) WIZ mRNA levels determined in the indicated cell lines by quatitative PCR. Three stable WIZ knockdown (KD) cell lines and a control (shCtr) cell line are tested and compared to wild-type (WT) cells. **(B)** The cell lines with most WIZ mRNA level reduction. Among 14 cell lines are tested for the best WIZ knockdown. WIZ shRNA1 causes a 80% reduction in WIZ mRNA levels of three cell lines, #7, #13 and #16. **(C)** WIZ knockdown efficency is tested by three biological replicas in the selected cell lines. WIZ mRNA levels in knockdown cell lines are compared to WT WIZ mRNA levels. WIZ KD #7 has the most steady reduction in WIZ mRNA levels among the selected cell lines. The error bars represent the standard deviation from the mean of three technical replicates. "****" indicates $p < 0.0001$. p values determined by one-way ANOVA t test ($n \geq 3$).

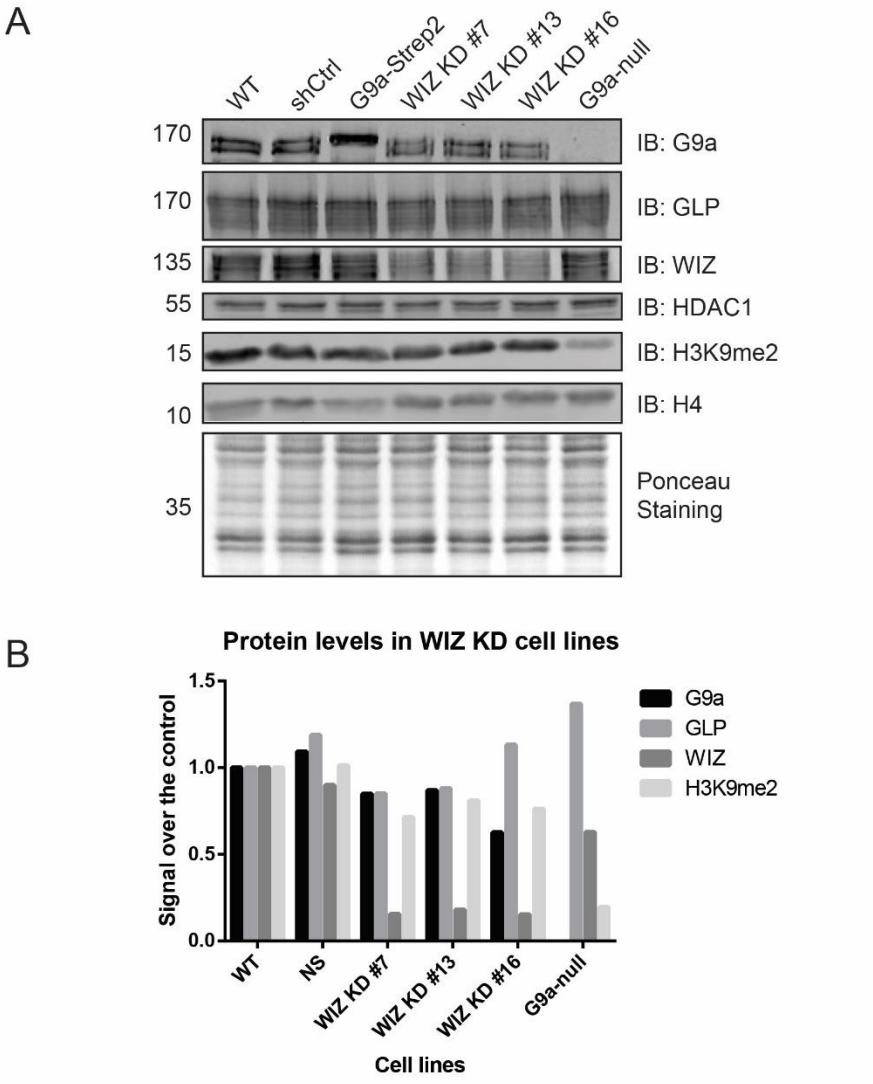


Figure 3. 4 The knockdown of WIZ by shRNA has a limited effect on the stability of the G9a complex

(A) The levels of G9a core complex components, G9a, GLP and WIZ, and H3K9me2 in the indicated cell lines. HDAC1 and H4 are used as loading controls. In addition to loading controls, the Ponceau staining shows equal loading of samples. **(B)** Quantification of the Western blot result in (A). The signal obtained from bands corresponding to G9a, GLP or WIZ is normalized to signal obtained from HDAC1 on the same gel. Similar quantification is applied to H3K9me2 signal by using H4 signal for normalization. The graph shows the normalized signal for indicated proteins.

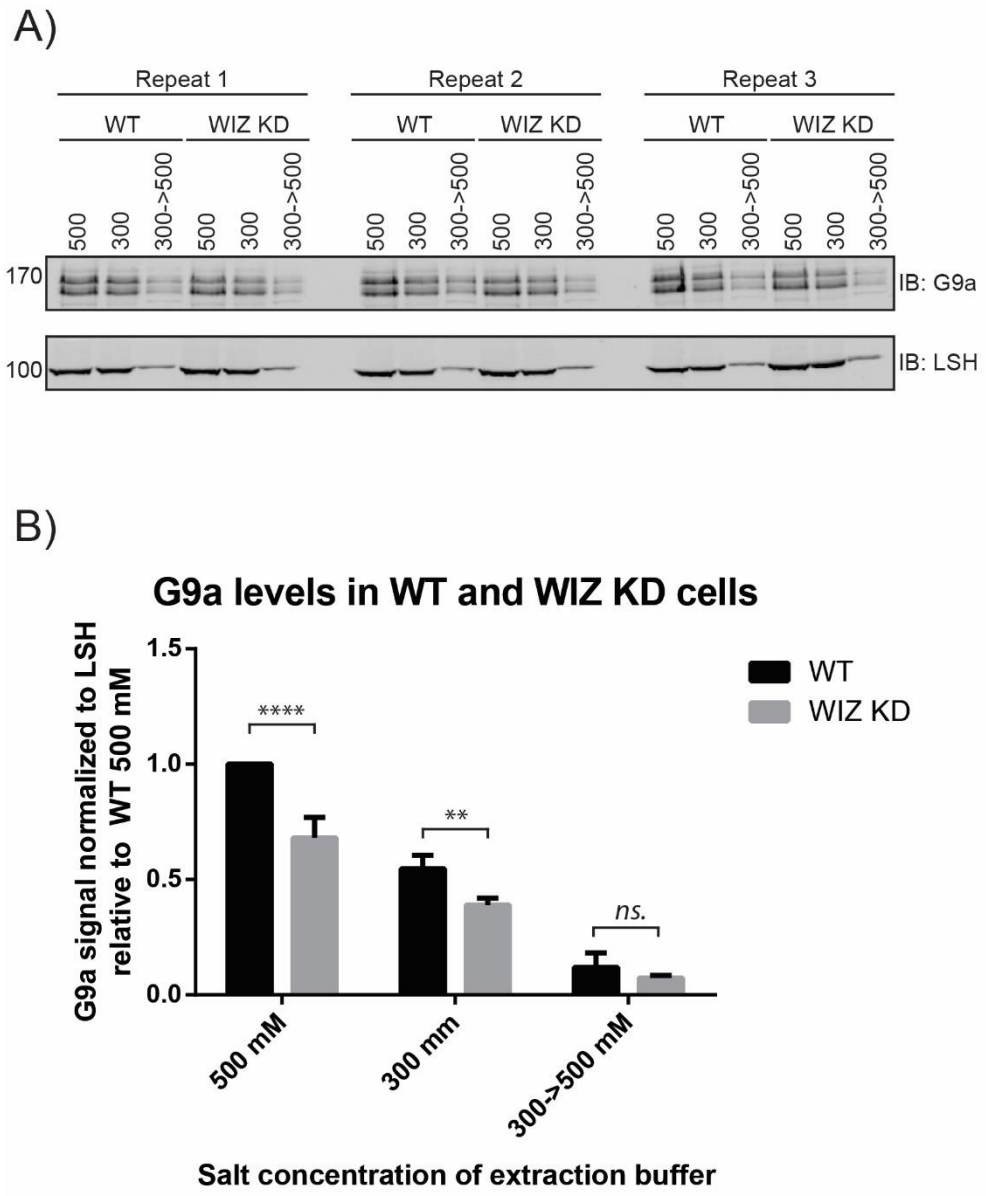


Figure 3. 5 The levels of G9a are reduced in WIZ KD cells

(A) Nuclear proteins extracted with buffer containing (i) high salt (500 mM), (ii) low salt (300 mM) or (iii) first with low salt and then with high salt containing buffer from wild-type (WT) and WIZ knockdown (KD) ES cells analysed on a Western blots with antibodies against G9a and LSH, as a WIZ-independent control. Three separate nuclear extractions are shown. (**B**) Quantification of the G9a signal relative to LSH. The normalized values are plotted relative to the WT 500 mM sample. The error bars represent the standard deviation from the mean of three technical replicates. "**" indicates $p < 0.01$, "****" indicates $p < 0.0001$. p values determined by t test ($n \geq 3$).

3.5. Discussion

By carrying out the mass spectrometry analysis of the purified G9a complex, in addition to the core complex members and some of the already-known binding partners, I was able to show that a number of nuclear proteins are novel binding partners of G9a (Table 1). Earlier studies on G9a have already shown the members of the core G9a complex and reported DNMTs and many zinc finger proteins as binding partners (Shinkai & Tachibana 2011). I did not detect most of these interactors such as Gfi1 (Duan et al. 2005), Gfib (Vassen et al. 2006) or ZNF217 (Banck et al. 2009) because these interactions were identified in other cell types than ES cells. Although I have not confirmed all of the candidates by individual co-IP experiments, using a high precision affinity mass spectrometry approach, some of these potential components of the G9a complex were identified in ES cells by a recent study (Maier et al., 2015). The main discovery of this study was the physical interaction between PRC2 and the G9a complex. The candidate "bridge" proteins to mediate this interaction were identified as ZNF518B, RBBP4, EED, MKI67IP, CBX3 and CBX5. These proteins were found in our mass spectrometry analysis with significant peptide scores as well. In addition to this, I was able to identify DNMT3A, DNMT3B and DNMT3L in our analysis. DNMTs were shown to interact with the G9a complex by previous studies (Chang et al. 2011; Esteve et al. 2006; Epsztejn-Litman et al. 2008). In my mass spectrometry experiment, chromatin remodeler Lymphoid-specific helicase (LSH) and its binding partner Histone deacetylase 1 (HDAC1) were detected as candidate binding partners to G9a. It was shown that LSH directs DNMTs and G9a/GLP

to specific sites to promote DNA methylation and gene silencing during differentiation (Myant et al. 2011). My data supports this study by showing LSH as a novel binding partner of G9a in ES cells. In addition, CHD4, RBBP4 and MTA1 are associated with the nucleosome remodelling and histone deacetylase (NuRD) complex. It was shown that G9a can methylate MTA1, which in turn changes the nucleosome remodelling and transcriptional outcome (Nair et al. 2013). By co-IP experiments, this study showed that G9a can interact with MTA1 and CHD4. Furthermore, Origin recognition complex subunit 1 (ORC1) was also identified as candidate binding partner of G9a; a recent study showed that ORC1 interacts with G9a along with Origin recognition complex-associated protein (ORCA) (Giri et al. 2015). Considering all these together, I could conclude that our mass spectrometry analysis has successfully identified potential members of the G9a complex. These data can be used in future projects to understand the role of G9a in particular actions that are mediated by one of these binding partners.

Several RNA-binding proteins were also identified among the proteins that co-purify with G9a. Some of these proteins were identified with at most three peptides, which is an indication that the interaction of this protein with G9a is probably transient. On the other hand, some of these proteins belong to the heterogeneous nuclear ribonucleoproteins family. The primary function of this protein family is to assist mRNA and miRNA biogenesis, including mRNA translation. All hnRNPs contain at least one RNA-binding domain, and also an auxiliary domain that specifies the function (Geuens et al. 2016). Some of these proteins are very abundant, which was one of the reasons that hnRNP L was identified in one of the FLAG affinity purifications of G9a as a candidate binding

partner. To preserve the protein-protein interactions in the purified complex, nuclear extraction was performed with 200 mM salt containing nuclear extraction buffer. This condition is quite gentle but also susceptible to detection of false positive interactions in mass spectrometry analysis. Abundant proteins, such as hnRNP L, have higher chances of binding the resin or the bound complex non-specifically. Therefore, washing steps in pulldown experiments were repeated for at least six times to eliminate the false positives as much as possible. This is probably the main reason that I was able to show that hnRNP L was a false positive hit.

The primary focus of my project is the recruitment of the G9a complex to chromatin. However, this is a very complex question and quite hard to solve by only one simple answer. As the MS analysis indicated, the G9a complex can interact with a large number of proteins, most of which, if not all, might contribute to the recruitment of the G9a complex. Therefore, it is very probable that some fraction of G9a is recruited to chromatin by RNA-binding proteins. One study has shown an interaction between G9a and Air, which is an imprinted non-coding RNA expressed from intron 2 of Igfr2 gene, by RNA IP experiments (Nagano et al. 2008). Although they were able to confirm this interaction at the Air target locus by ChIP, it is still not clear what mediates the interaction between G9a and Air. Using a similar approach, another study has shown that G9a interacts with Kcnq1ot1, another imprinted non-coding RNA expressed from intron 10 of Kcnq1 gene, in placenta but not in fetal liver (Pandey et al. 2008). This is another indication that the interaction between G9a and the non-coding RNAs is not direct and in some cases the interaction is lost due to the absence of the mediator in certain tissues. The data we

generated could potentially contain some these mediators, and some of them should be investigated further.

Proteins of the CBX family appeared consistently in both replicas of MS analysis. CBX1, CBX3 and CBX5, also known as HP1β, HP1γ and HP1α respectively, are heterochromatin proteins that recognise H3K9me2/3 by their chromo domain (Lachner et al. 2001). It was shown that auto-methylation of G9a on the K239 residue creates a binding site for HP1 proteins (Chin et al. 2007; Sampath et al. 2007). Therefore, it was not surprising to find HP1 proteins in the G9a-FLAG MS dataset. The other CBX proteins (CBX 2/4/6/7/8) form the polycomb subunit of the PRC1 complex (Ma et al. 2014). The interaction between the G9a complex and PRC1 either does not exist, or it is very transient to detect according to my data. On the other hand, core components of PRC2 complex, SUZ12, EED and RBBP4, were identified, but EZH2 was not detected in the MS dataset. It would be expected to see all core components because PRC2 complex is not stable even in the absence of one complex member. Considering that the identified peptide numbers for these three proteins ranged between 2 to 6, and PRC2 complex can be formed by either EZH1 or EZH2; perhaps the peptides from both homologs were very low in quantity and was not detected by mass spectrometry.

Studies have shown that zinc finger proteins might be responsible for the recruitment of epigenetic factors. MLL1 and MLL2 contains a CXXC-type zinc finger domain, which recognises unmethylated CG di-nucleotides and can recruit MLL CpG-rich sequences (Birke et al. 2002). Similarly, CFP1, a CXXC-type zinc finger protein and a component of the SETD1A complex, can recognise the unmethylated CpG sites and is associated with

the recruitment of the PRC2 complex to non-promoter CpG islands to establish H3K4me3 (Thomson et al. 2010). CXXC-type zinc finger domain-containing histone demethylases KDM2A and KDM2B have been shown to be involved in the recruitment of PRC1 complex to CpG islands (Blackledge et al. 2010; Farcas et al. 2012). A similar type of recruitment might be considered for the G9a complex since we were able to identify a number of zinc finger proteins in the MS analysis. As discussed previously, ZF462, ZF518B and ZC3H11A were identified as possible binding partners of G9a; however, information about the structure and function of these zinc fingers are rather limited. On the other hand, WIZ was already shown as an important member of the G9a complex, and I was able to detect WIZ in stoichiometric amounts as compared to G9a in ES cells. Considering that WIZ can bind specific DNA or RNA sequences, recruitment of G9a to specific sites during to establish H3K9me1/2 mark might be dependent on WIZ. Therefore, I decided to focus on WIZ in relation to its importance in the recruitment of the G9a complex.

I was able to confirm that WIZ knockdown causes a 30% reduction in the both G9a and GLP levels, which suggests that the stability of the G9a complex partially depends on WIZ in ES cells. However, the reduction in the G9a levels is significantly higher in the original study, where WIZ was knocked down by siRNA in HeLa cells (Ueda et al. 2006). This can be explained by that in HeLa cells the stability of G9a might depend on fewer factors such as WIZ. In that case, a depletion in WIZ levels could lead to a great reduction in G9a levels. On the other hand, in ES cells G9a can interact with many other proteins

that could enhance the stability of the G9a complex so WIZ depletion might be compensated much effectively.

In addition to this, the original study has indicated that an interaction between CtBP2 and G9a is mediated by WIZ (Ueda et al. 2006). CtBP2 is one of the proteins that I identified in G9a-FLAG purifications, and it is possible that WIZ might mediate the interactions between CtBP2 and other proteins with G9a in ES cells. To understand the role of WIZ, I tried to generate ES cell lines that stably express a tagged-WIZ. I also deleted the 6th zinc finger of WIZ, which is important in the interaction with the G9a complex. Unfortunately, the cell lines I made failed to produce the tagged-WIZ, although the mRNA was detectable. This indicates that the full-length WIZ protein might not fold properly when tagged at the N-terminus. Nonetheless, investing more time to solve this problem by tagging WIZ at the C-terminus might provide crucial information about the function of WIZ in the G9a complex.

Up to this point, I showed a number of possible candidate binding partners of G9a. After narrowing down my focus on WIZ, I was able to show that WIZ is important for the stability of the G9a. To elaborate on the functional role of WIZ in the G9a complex, I wanted to ask whether or not chromatin binding of G9a is dependent on WIZ in ES cells.

Chapter IV – Chromatin binding of the G9a complex

4.1. Introduction

The H3K9me2 catalysed by G9a/GLP is essential for the formation of facultative heterochromatin (Trojer and Reinberg 2007). One study showed that about half of the genome is covered by large organised chromatin K9me2 modifications (LOCKs) to ensure silencing of genes (Wen et al. 2009). The same study suggested that the differentiated liver and brain cells have larger LOCKs when compared to undifferentiated ES cells. Another group argued that there was no significant difference between the levels of H3K9me2 in differentiated cells and ES cells when a different statistical system was applied (Filion & van Steensel 2010). Additionally, one recent study mapped the H3K9me2 on promoters and chr19 in mouse ESC and neurons and showed that the difference in H3K9me2 coverage between two cell types was less than 5% (Lienert et al. 2011). In overall, these studies indicated that the dynamics of H3K9me2 in differentiation might be lineage specific. The LOCKs cover up to 4.9 Mb regions, and their maintenance depends on the G9a/GLP activity. An earlier study has indicated that H3K9me2 mark is enriched at the lamina-associated domains (LADs) (Guelen et al. 2008). Perhaps not surprisingly, 82% of the LOCKs identified in placenta were identified to be at LADs as well. Another study indicated that the catalytic activity of G9a is directly related to the extent of LAD-NL interactions (Kind et al. 2013). In this study, it was shown that G9a knockdown or the G9a inhibitor treatment decreased the LAD-NL interactions indicating

that H3K9me2 mark is required for attachment of LADs to the NL. Conversely, the overexpression of G9a increased the LAD-NL interactions. Together these observations suggest that G9a/GLP complex modifies a significant portion of chromatin to maintain the repressive H3K9me2 mark, which is important for proper gene regulation. Therefore, it is important to understand the chromatin binding capacity of G9a/GLP and the factors contributing to it.

As mentioned earlier, G9a and GLP recognise the H3K9me1/2 by their ankyrin repeat domains and facilitate a feedforward loop for maintenance of the mark. However, it has been reported that there is also a crosstalk between DNA methylation and H3K9me2. For example, reduced DNA methylation due to the loss of DNMT1 was shown to cause a significant reduction in H3K9me2 levels in cancer cells (Espada et al. 2004). In ES cells, maintenance of DNA methylation is not only taken care of by DNMT1 but also DNMT3A/B, which interacts with G9a/GLP via the ankyrin repeat domain. However, studies have shown that the cross-talk between these G9a/GLP and DNMTs has no impact on the maintenance of H3K9 methylation while DNA methylation seems to be dependent on H3K9me for certain genes in ES cells (Tsumura et al. 2006; Matsui et al. 2010). On the other hand, it was shown that the DNA methylation was maintained at the imprinting control regions in ES cells via the interaction between G9a and DNMT3A/B, which is independent of the catalytic activity of G9a (Zhang et al. 2016). Therefore, it is likely that ES cells employ different mechanisms for the maintenance of histone and DNA methylation at different genomic sites (Du et al. 2015).

4.1.1. **Aims**

Although G9a has an affinity for its product, H3K9me1/2, on the chromatin, it is unclear to what extent these marks contribute to the affinity of G9a to chromatin. Additionally, specific DNA sequences, RNA species, interaction with other nuclear proteins or various epigenetic modifications might contribute to the chromatin binding of the G9a complex as well. In this chapter, I aimed to understand the role of H3K9me2 and the G9a/GLP binding partners – WIZ and DNMT3A/B, in the recruitment of G9a to chromatin in ES cells. Mainly, I used a technique called step-wise salt extraction to determine global chromatin binding of G9a in ES cells treated with a small molecule G9a inhibitor (UNC0638), and in cell lines that are deficient in either WIZ or DNMT3A/B.

4.2. Global chromatin binding of G9a in ES cells

A former Ph. D. student in the lab modified the standard nuclear extraction protocol by adding extra extraction steps where the salt concentration in the nuclear extraction buffer gradually increased (Figure 4.1A, see Materials and Methods). Using this technique, it was possible to fractionate any nuclear protein according to its chromatin binding capability. The first two fractions (0 mM and 100 mM) were considered as enriched in proteins that show a very low affinity to chromatin (hereafter referred as a “weak” fraction) since the salt in the nuclear extraction buffer is closest to the physiological condition. The following two fractions (200 mM and 300 mM) were considered as a dynamic fraction because this salt concentration is sufficient to extract nuclear proteins

with low to an intermediate affinity for chromatin, but not high enough to extract proteins tightly associated with chromatin and DNA such as histones. Finally, the last two fractions (400 mM and 500 mM) were considered as enriched with proteins that are very tightly-bound to chromatin and DNA (hereafter referred as a "tight" fraction). It is important to note that prior to the 500 mM extraction, the chromatin was digested with an exonuclease (Benzonase) to solubilize chromatin.

Initially, I investigated that to what extent H3K9me2 contributes to the high-affinity binding of G9a to chromatin in WT ES cells. To this end, I performed the step-wise salt extraction (SWS extraction) on cell pellets collected from WT ES cells grown either in the presence or the absence of G9a and GLP inhibitor (UNC0638). The inhibitor mimics the H3 tail and has been shown to bind specifically to the catalytic pocket of G9a and GLP SET domain thus interfering with the enzymatic activity of these two proteins, but not other SET domain containing enzymes (Vedadi et al. 2011). Adding the G9a inhibitor to the ES cell culture medium reduces the H3K9me2 levels by more than 85% after two days of culturing (Vedadi et al. 2011 and Figure 4.1B). This was confirmed in our lab, and a comparable decrease in the levels of H3K9me3 was detected in inhibitor-treated ES cells (Figure 4.7B, Adapted from Tuo Zhang's thesis). As discussed earlier, G9a/GLP can bind to H3K9me2 mark via the Ankyrin repeat domain. Therefore, it would be expected the high-affinity association of G9a with chromatin to be reduced after diminishing the H3K9me2 levels by the G9a inhibitor treatment. As anticipated, the UNC0638 treatment changed the chromatin binding profile of G9a (Figure 4.1B-C). In wild-type condition, about 70% of the total G9a protein was present in the tight fraction, which indicated that

under normal growth conditions G9a stably associates with chromatin in ES cells. Upon inhibitor treatment, the H3K9me2 levels were reduced by about 85%, which caused a 30% reduction of G9a in the tight fraction. This decrease was reflected by the 10% and 20% increase of G9a in the dynamic and the weak fractions, respectively. It is important to highlight that about 40% of G9a was still tightly associated with chromatin in inhibitor-treated cells. This observation indicates that H3K9me2 mark cannot be the only factor that contributes to the binding of G9a to chromatin.

In comparison to G9a, the Ras-related nuclear protein (RAN), which is involved in nuclear import/export and shuttles between the cytoplasm and the nucleus was predominantly found in the weak fraction, as expected. Another control protein, the Lymphoid-Specific helicase (LSH, or HELLS) is a chromatin remodeler to facilitate methylation by *de novo* DNMTs, and it is important for developmentally DNA methylation (Myant et al. 2011; Myant & Stancheva 2008). LSH was present in the dynamic fraction as well as in the tight fraction. It was not surprising to detect a chromatin remodeler predominantly in the dynamic fraction as remodelling chromatin requires constant association with and dissociation of the remodeler from chromatin. The LSH in the tight fraction might be the proportion of the total protein that interacts with DNMT3B, therefore displaying tighter binding to chromatin. Lastly, histone proteins, for example, Histone 4 (H4), were present exclusively in the very last fraction, as expected. Altogether, the dissociation patterns of the control proteins from chromatin in the salt extraction experiments appeared as predicted. Thus, the chromatin binding profile of G9a should be accurate as well.

To verify that the H3K9me2 mark is directly responsible for the high-affinity association of G9a with chromatin, I used an ES cell line that produces a mutant G9a that lacks the entire ANK repeat domain. These cells stably express shRNAs against the endogenous G9a, while also expressing a shRNA-resistant form of G9a that lacks the Ankyrin repeat domain and carry point mutations that prevent dimerization of G9a with GLP (ANKΔ-DM cells) (Zhang et al. 2016). I applied the SWS extraction protocol to cell pellets collected from ANKΔ-DM ES cells that were either treated with the G9a inhibitor or cultured in medium without inhibitor (Figure 4.2A). The mutant G9a was somewhat unstable in these cells. Therefore, more cells were harvested for this experiment. The G9a KD reduces the endogenous G9a levels by 70%, so the remaining endogenous G9a appeared in the tight fraction. Quantification of the Western blots showed that the inhibitor treatment had no significant effect on the G9a extraction profile in this cell line (Figure 4.2B). This data further supported the previous finding that about 30% of the total G9a proteins relies on binding to the H3K9me2 mark via the ANK-repeat domain for stable association with chromatin.

Next, I wanted to investigate the effect of inhibitor treatment on chromatin binding of G9a in WIZ KD cells. For this purpose, I carried out an SWS extraction on inhibitor treated- or untreated-WIZ KD cells and compared the chromatin binding profiles of G9a (Figure 4.3A). I detected a decrease in G9a in the tight fraction upon inhibitor treatment, but the reduction was not as substantial as observed in the WT cells (20% vs. 30%) (Figure 4.3B). In parallel to earlier results, I was able to show that H3K9me2 mark serves as an “anchoring point” for G9a.

It was surprising to see that in these experiments the dynamic and the weak fractions were almost depleted of G9a in the untreated condition when WIZ KD cells were investigated. Note that WIZ KD caused a 30% reduction in the global G9a levels, which suggests that the lost fraction of G9a in WIZ KD cells forms the weak/dynamic fraction of G9a in ES cells. To recapitulate this, I performed another SWS extraction experiment using WT and WIZ KD ES cells. However, this time, the result did not show a significant difference in the chromatin binding profile of G9a between WT and WIZ KD cells (Figure 4.4A). Quantification of the data showed deviation between replicas (Figure 4.4B). This might be the primary cause of not detecting a small (20%) but significant difference of G9a in the dynamic fraction between the WT and WIZ KD cells.

The crosstalk between H3K9me2 and DNA methylation is evident due to the interaction between G9a and DNMTs. An important question to ask was whether or not DNA methylation has any effect on the chromatin binding of G9a. To investigate the effect of DNA methylation on the recruitment of G9a to chromatin, I used ES cell line deficient in DNMT3A and DNMT3B (DNMT3A/B-null, DKO). This cell line was generated in En Li laboratory, and it was shown that DNA methylation was reduced drastically in ES cells lacking DNMT3A/B (Okano et al. 1999; Meissner et al. 2005). Similar to previous experiments, I performed an SWS extraction to DKO ES cells in the presence or absence of the G9a inhibitor. Surprisingly, the chromatin binding pattern of G9a in DKO cells looked very similar to the pattern of ANKΔ-DM ES cells (Figure 4.5A). Furthermore, quantification of the data showed no significant difference between nuclear extraction profile in the presence or absence of the inhibitor (Figure 4.5B). Considering the fact that

DNMT-G9a interactions occur via the ANK repeat domain of G9a, it was quite intriguing to see that the ANK deletion and the DNMT deficiency lead to a similar effect on the nuclear extraction profile of G9a. Although, this result did not clearly show that DNA methylation is essential for the recruitment of G9a to chromatin, it further supports the idea that the ANK domain of G9a is necessary for the self-recruitment of G9a to chromatin. Additionally, 40% of G9a was still present in the tight fraction in these cells similar to ANKΔ-DM ES cells. These data so far show that high-affinity chromatin binding of G9a partially depends on H3K9me2 and the presence of DNMTs. However, the effect of WIZ in the recruitment of G9a to chromatin in the undifferentiated ES cells seemed rather limited.

4.3. H3K9me2 recovery rate after the G9a inhibitor treatment was not altered in WIZ KD cells

Next, I sought to investigate the importance of WIZ in the G9a complex by another approach. Knowing that WIZ interacts with the G9a/GLP heterodimer via the SET domain, I hypothesised that this interaction could have an effect on the catalytic activity of the complex. Therefore, WIZ could have an indirect effect on the recruitment of G9a by stabilising its catalytic activity and the H3K9me2 mark, which in return recruits G9a to chromatin. To test this hypothesis, I performed an H3K9me2 recovery experiment in G9a inhibitor-treated ES cells. In short, I treated WT or WIZ KD ES cells with the G9a inhibitor for two days, then I removed the inhibitor from the culture media in a time-dependent manner and harvested all the cells at the same time (Figure 4.6A, see Materials

and Methods). Nuclear extracts from these cells were used for Western Blotting to detect the H3K9me2 levels (Figure 4.6B). Quantification of the data showed that H3K9me2 is recovered to the levels comparable to the non-treated control in 48 hours in both WT and WIZ KD ES cells, which indicated that WIZ KD has no effect on the catalytic activity of G9a. Consistent with previous results, I observed about 30% decrease in total H3K9me2 in WIZ KD cells. Interestingly, the reduction was not detectable in the no recovery control and the 6h sample, as it became more significant at later time points. However, this could be related to the sensitivity of the primary or the Licor secondary antibody used in Western blotting. The G9a inhibitor treatment did not reduce the G9a levels over the course of the experiment, indicating that the reduction of H3K9me2 levels was not due to a change in G9a or GLP stability (Figure 4.7A).

The consistent decrease in H3K9me2 levels observed in the WIZ KD cells raised the possibility that the loss of H3K9me2 could occur at specific sites rather than a global reduction of H3K9me2 due to reduced levels of G9a. To this end, I carried out ChIP-qPCR experiments to investigate H3K9me2 levels at a few genomic sites, including Wdfc15a, MageA2, Igfr2a and Peg3, which are marked by G9a-dependent H3K9me2 in wild-type ES cells. I used Actin promoter as a negative control; however, my result showed a low level of H3K9me2 present at this site (Figure 4.8A). No enrichment for H3K9me2 was detected at Peg3 promoter, while the other sites had significant enrichment of H3K9me2 over the Actin control. In general, I obtained more consistent ChIP results for WdFc15a promoter compared to other loci. Therefore, I chose Wdfc15a for further experiments. A recent study has shown that G9a is recruited to the 3’ end of some genes, such as Actin,

where R-loops facilitate the efficient termination of transcription (Skourti-Stathaki et al. 2014). As a candidate DNA-binder, WIZ might have a role in the recruitment of G9a to such sites. Therefore, I investigated the Actin pause site for the presence of H3K9me2 in WIZ KD cells. As an additional control, I also tested the effect of the G9a/GLP inhibitor on H3K9me2 at this site. I carried out a ChIP experiment using either inhibitor-treated or untreated WIZ KD cells. The result showed that there was still a significant H3K9me2 enrichment at action pause site in WIZ KD cells (Figure 4.8B). This enrichment was reduced by about 80% upon the G9a inhibitor treatment. Usually, the inhibitor treatment reduces the H3K9me2 levels by 90%, which meant that the inhibitor worked less efficiently than usual in this particular experiment. This could explain why there is some signal being detected from both Wdfc15a and Actin pause site. Finally, I performed an H3K9me2 ChIP to investigate the presence of H3K9me2 at the Actin Pause site. Comparing results of WT and WIZ KD showed that there is a 30% decrease in the H3K9me2 enrichment in WIZ KD cells at both positive control site and actin pause site. This indicates that the reduction in H3K9me2 is probably a global effect that is caused by the lower levels of G9a protein in WIZ KD cells rather than a lack of direct recruitment of G9a to the Actin Pause site mediated by WIZ, which would have caused a more significant reduction in H3K9me2 at this site in WIZ KD cells. These results show that WIZ does not mediate the site-specific recruitment G9a to the Actin pause site in ES cells.

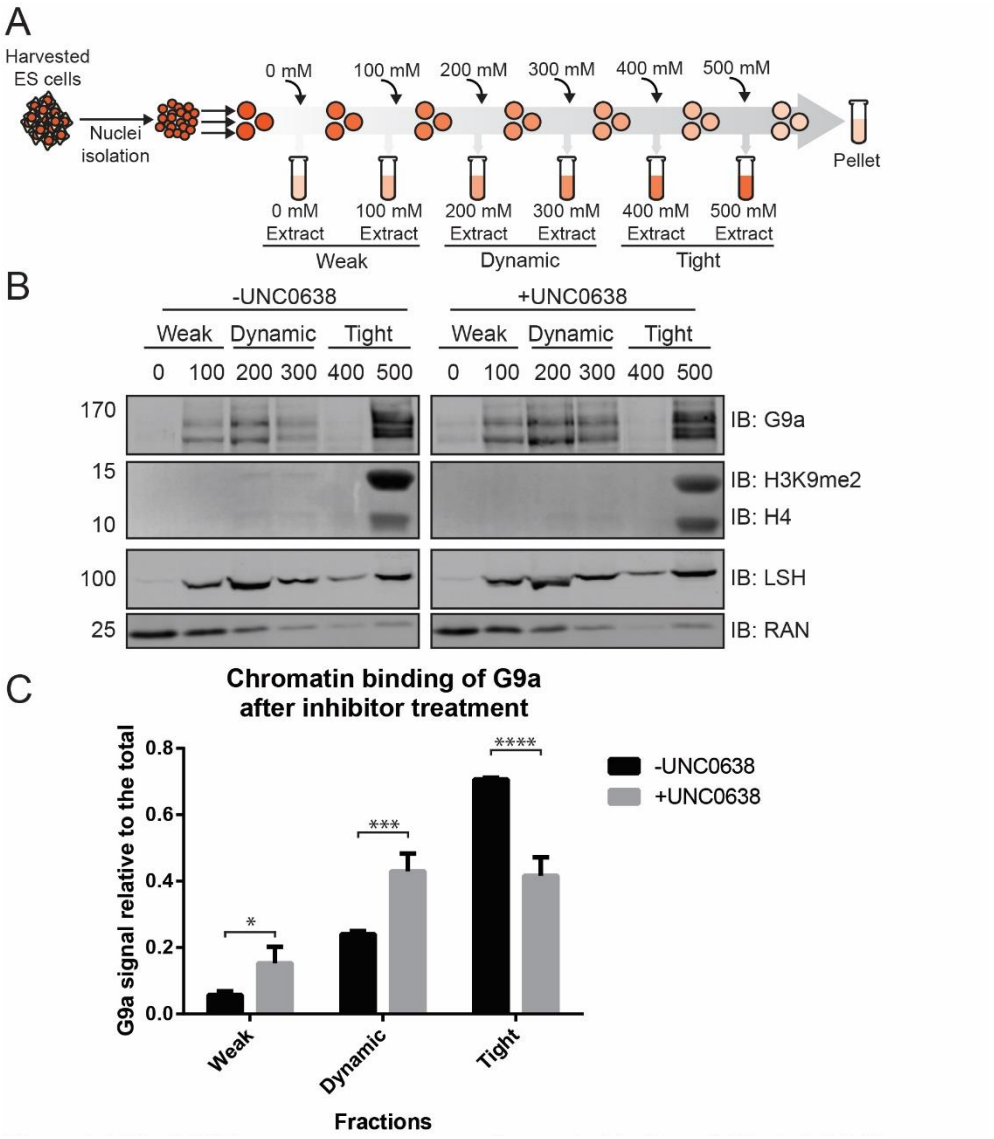


Figure 4. 1 G9a inhibitor treatment reduces chromatin binding of G9a in WT ES cells

(A) Schematic representation of the step-wise salt extraction experiment. Nuclei are isolated from ES cells and split into three aliquots, of which each one is used per technical replicate. Fractions are collected by increasing the salt concentration in the extraction buffer as depicted above. **(B)** Western blot of a step-wise salt extraction experiment carried out in the presence of the G9a/GLP inhibitor (UNC0638). Histone H4, the chromatin remodeller LSH, with moderate affinity for chromatin and a nucleocytoplasmic transporter protein RAN, with low affinity for chromatin are used as controls. **(C)** Quantification of the step-wise salt extraction data. G9a signal in the weak, dynamic and tight fractions is plotted relative to the total G9a signal. The error bars represent the standard deviation from the mean of three technical replicates. "*", "**" and "****" indicate $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. p values determined by t test ($n \geq 3$).

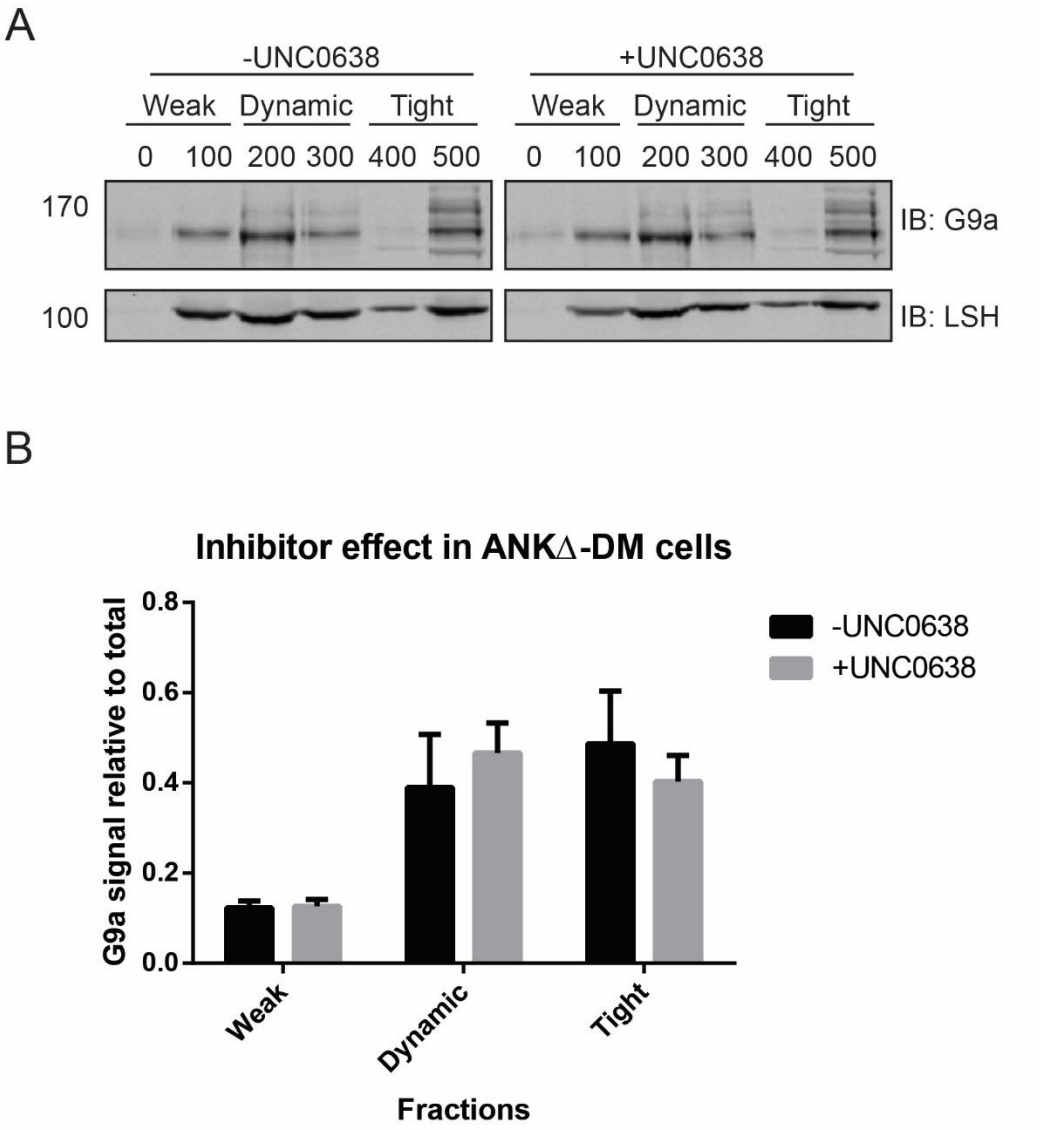


Figure 4. 2 G9a inhibitor treatment does not affect the chromatin binding of G9a in ANKΔ-DM ES cells

(A) Western blot of a step-wise salt extraction experiment carried out in the presence of the G9a/GLP inhibitor (UNC0638). Chromatin remodeller LSH with moderate affinity for chromatin is used as a control. Note that the mutant form of G9a appears as a single band. **(B)** Quantification of the step-wise salt extraction data. The G9a signal in the weak, dynamic and tight fractions is plotted relative to total G9a signal for three technical replicates. The error bars represent the standard deviation from the mean of three technical replicates. Using the *t* test to compare G9a levels between fractions does not reveal any statistically significant differences ($n \geq 3$).

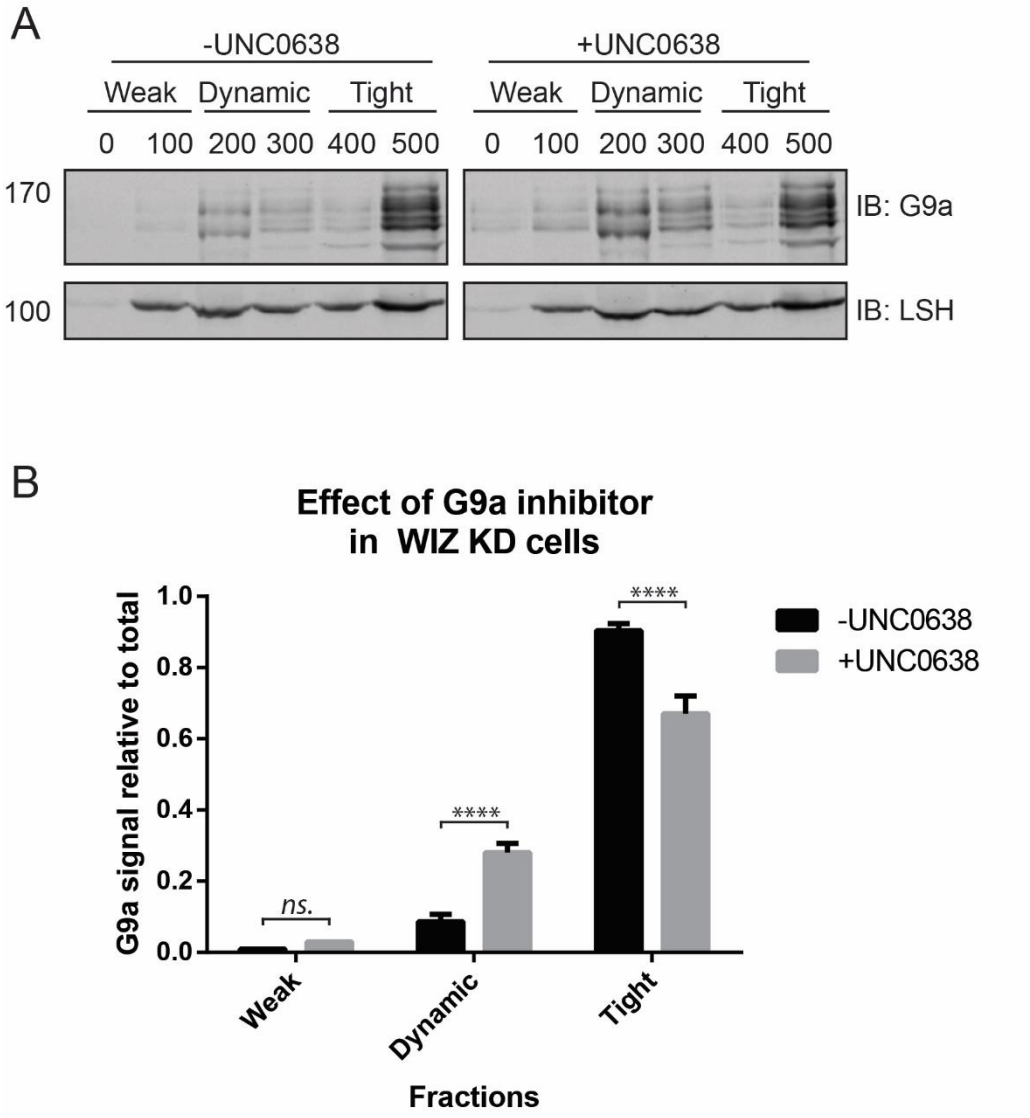


Figure 4. 3 The WIZ knockdown modulates the effect of G9a inhibitor on chromatin binding of G9a

(A) Western blot of a step-wise salt extraction experiment from WIZ knockdown (KD) ES cells grown in the absence and in the presence of the G9a/GLP inhibitor (UNC0638). Chromatin remodeller LSH is used as a control. **(B)** Quantification of the step-wise salt extraction data. The G9a signal in the indicated fractions is plotted relative to total G9a signal for three technical replicates. The error bars represent the standard deviation from the mean of three technical replicates. "****" indicates $p < 0.001$. p values determined by t test ($n \geq 3$).

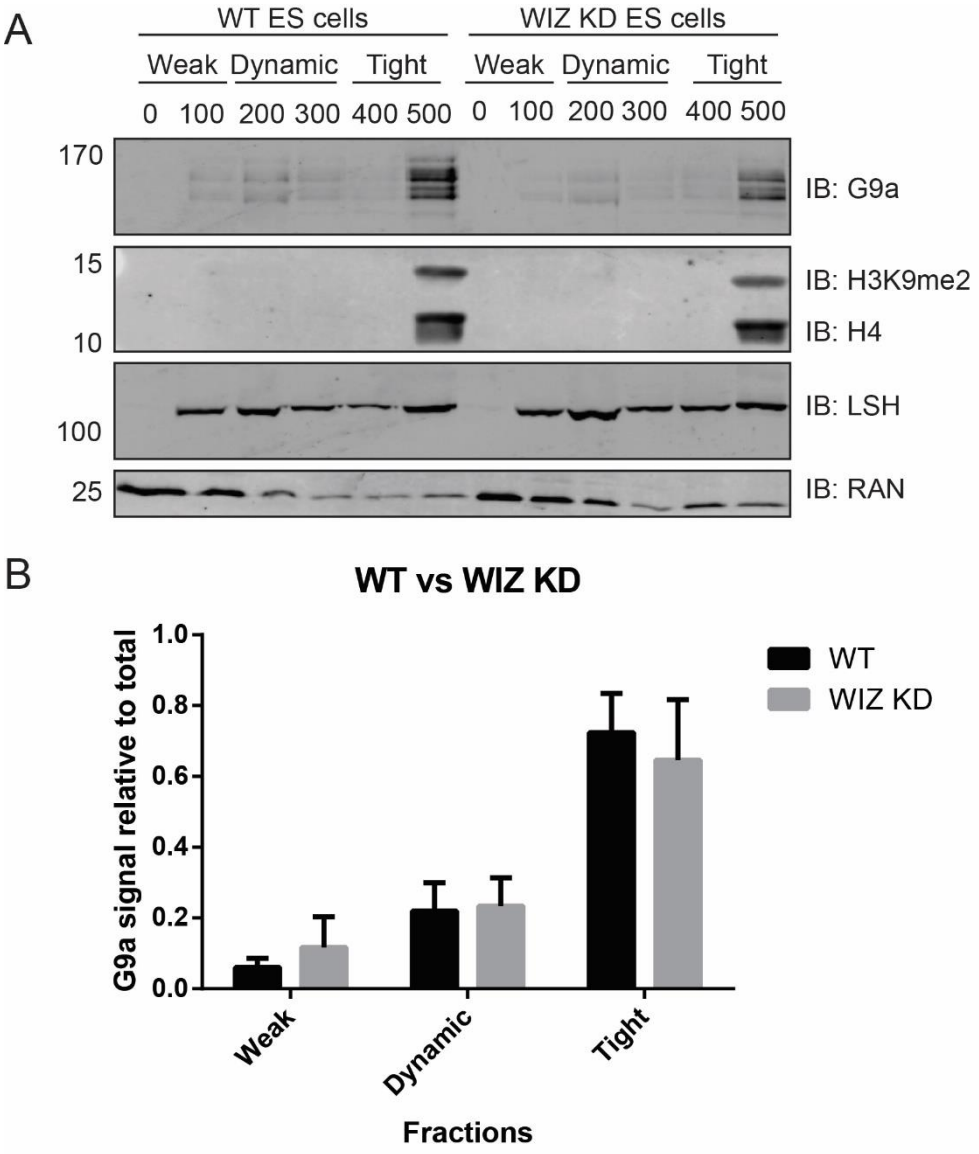


Figure 4. 4 The knockdown of WIZ has no significant effect on chromatin binding of G9a

(A) Western blots of step-wise salt extraction experiment performed on WT and WIZ KD ES cells. Histone H4, the chromatin remodeller LSH and RAN are used as controls. **(B)** Quantification of the step-wise salt extraction data. G9a signal of the indicated fractions relative to total G9a signal is plotted for three technical replicates. The error bars represent the standard deviation from the mean of three technical replicates.Using the *t* test to compare G9a levels between fractions does not reveal any statistically significant differences ($n \geq 3$).

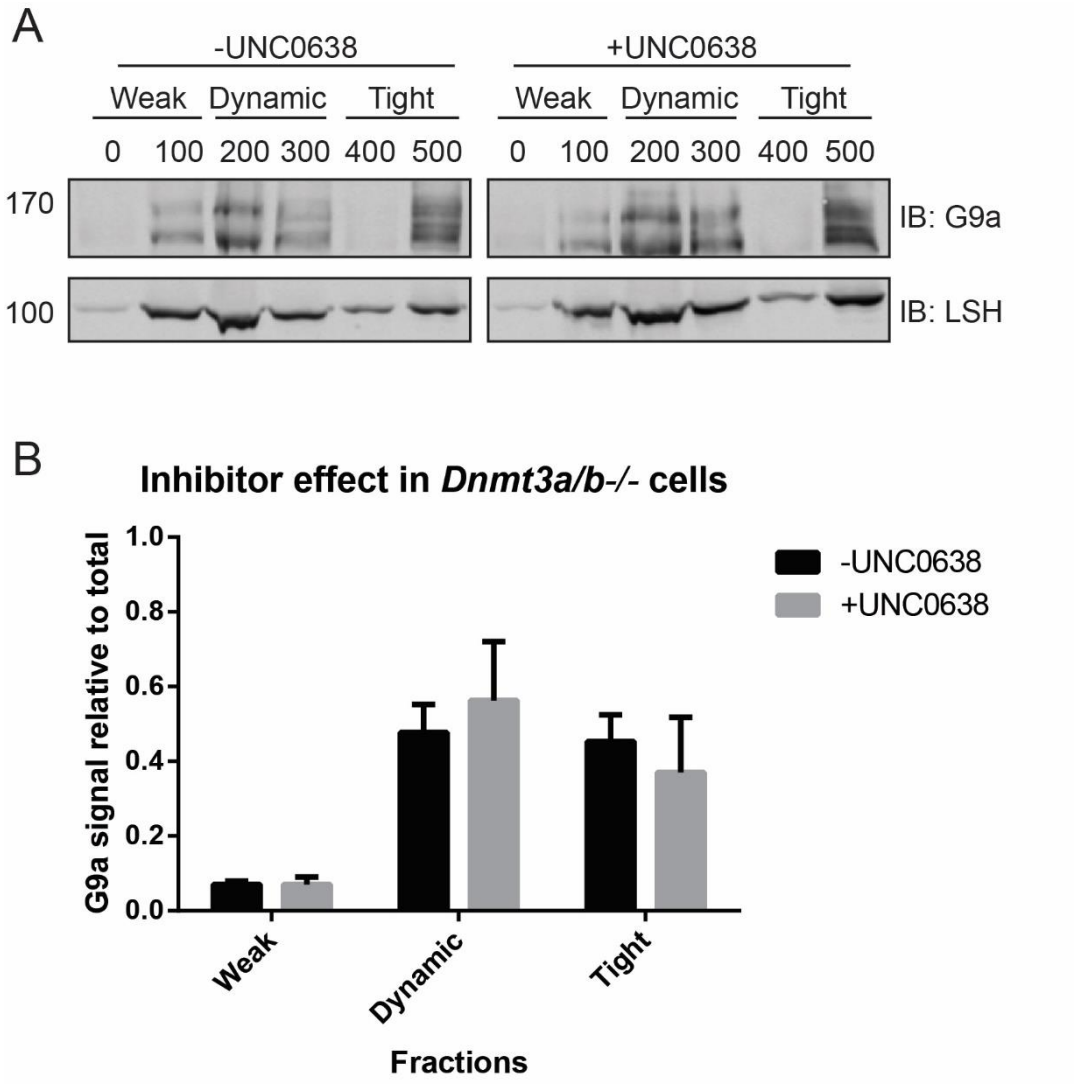


Figure 4. 5 G9a inhibitor has no effect on chromatin binding of G9a in DNMT3A/3B double null ES cells

(A) Western blot result of a step-wise salt extraction experiment carried out in the presence of a G9a inhibitor (UNC0638). Chromatin remodeller LSH is used as a control. **(B)** Quantification of the step-wise salt extraction data for three technical replicates. G9a signal in the indicated fractions is plotted relative to the total G9a signal. The error bars represent the standard deviation from the mean of three technical replicates. Using the *t* test to compare G9a levels between fractions does not reveal any statistically significant differences ($n \geq 3$).

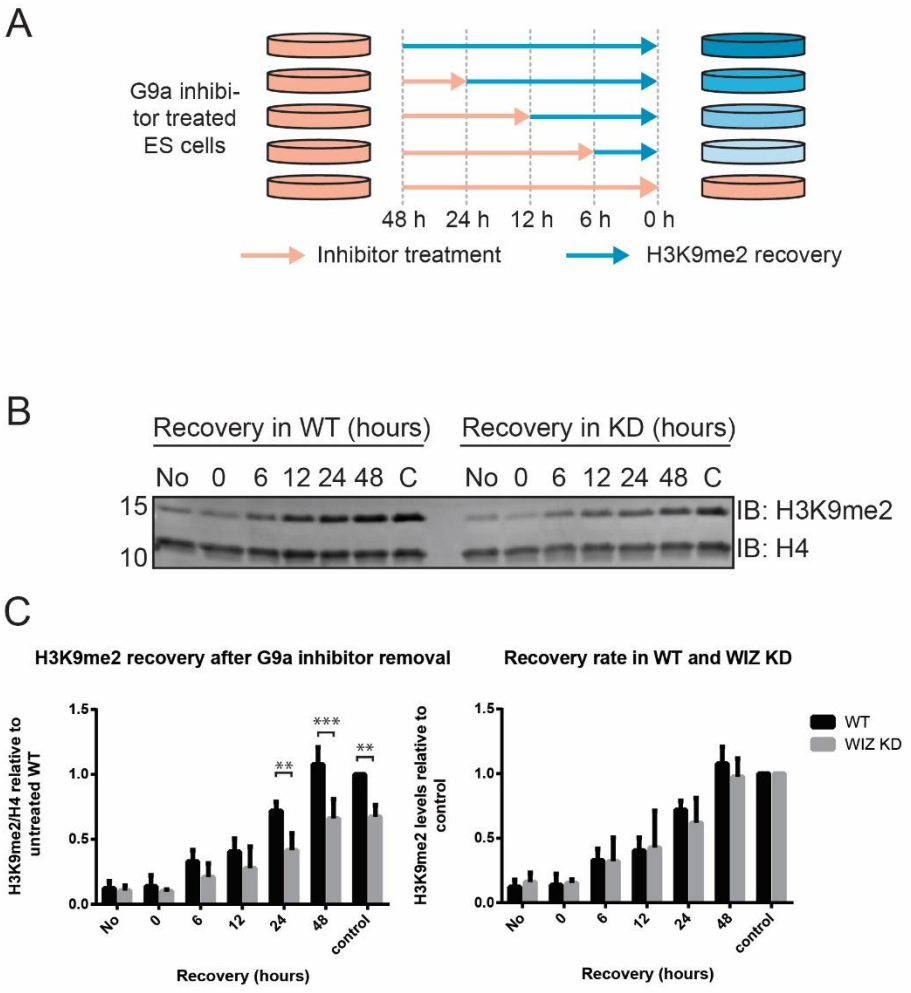


Figure 4. 6 H3K9me2 recovery after G9a inhibitor treatment is not altered in WIZ KD cells

(A) Schematic representation of the inhibitor removal experiment. ES cells are cultured in UNC0638 containing medium for two days. The cells are seeded to be harvested in 48 hours and the G9a inhibitor is removed at the indicated time points. As controls, one dish containes the G9a inhibitor until harvesting (no recovery), another dish is not treated with inhibitor (control). All cells are harvested at the same time. **(B)** Western blot of an inhibitor removal experiment. "No" stands for no recovery control, "C" stands for the control sample representing cells grown without inhibitor. H4 is used as a loading control to quantify the signal obtained from H3K9me2. **(C)** On the left, H3K9me2 signal normalized to H4 to obtain H3K9me2 levels. The average levels are plotted from three technical replicates. On the right, H3K9me2 levels normalized relative to the initial levels of H3K9me2 (non-treated control) to determine the rate of H3K9me2 recovery. The error bars represent the standard deviation from the mean of three technical replicates. "**" and "***" indicate $p < 0.01$ and $p < 0.005$, respectively. p values determined by t test ($n \geq 3$).

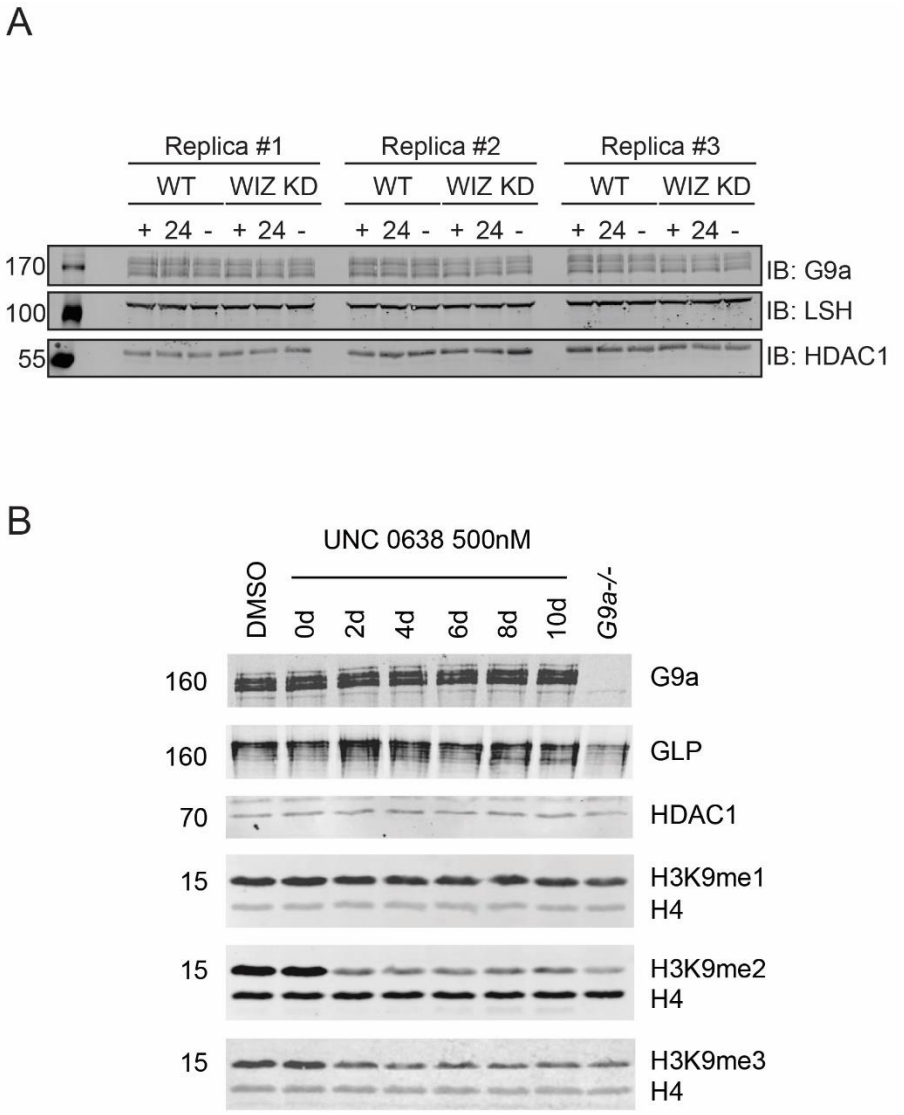


Figure 4. 7 The G9a inhibitor reduces th elevels of H3K9me2/3 after 2 days of treatment but does not reduce the protein levels of G9a or GLP

(A) Western blots showing the G9a levels in cells that are constantly treated with inhibitor (+), cell that recovered H3K9me2 levels for 24 hours (24), and control cells with no inhibitor treatment (-). The G9a levels does not changed over the course of the experiment, indicating that the reduction of H3K9me2 levels is not due to a change in G9a or GLP stability. **(B)** Western blots probed for G9a, GLP and H3K9 mono-, di-, and trimethylation in ES cells treated with UNC0638 for the indicated number of days. DMSO and *G9a-/-* represent WT ES cells treated with DMSO for 10 days and untreated G9a knockout cells, respectively. HDAC1 and H4 are loading controls for G9a and H3K9 methylation, respectively (adapted from Tuo Zhang's PhD thesis).

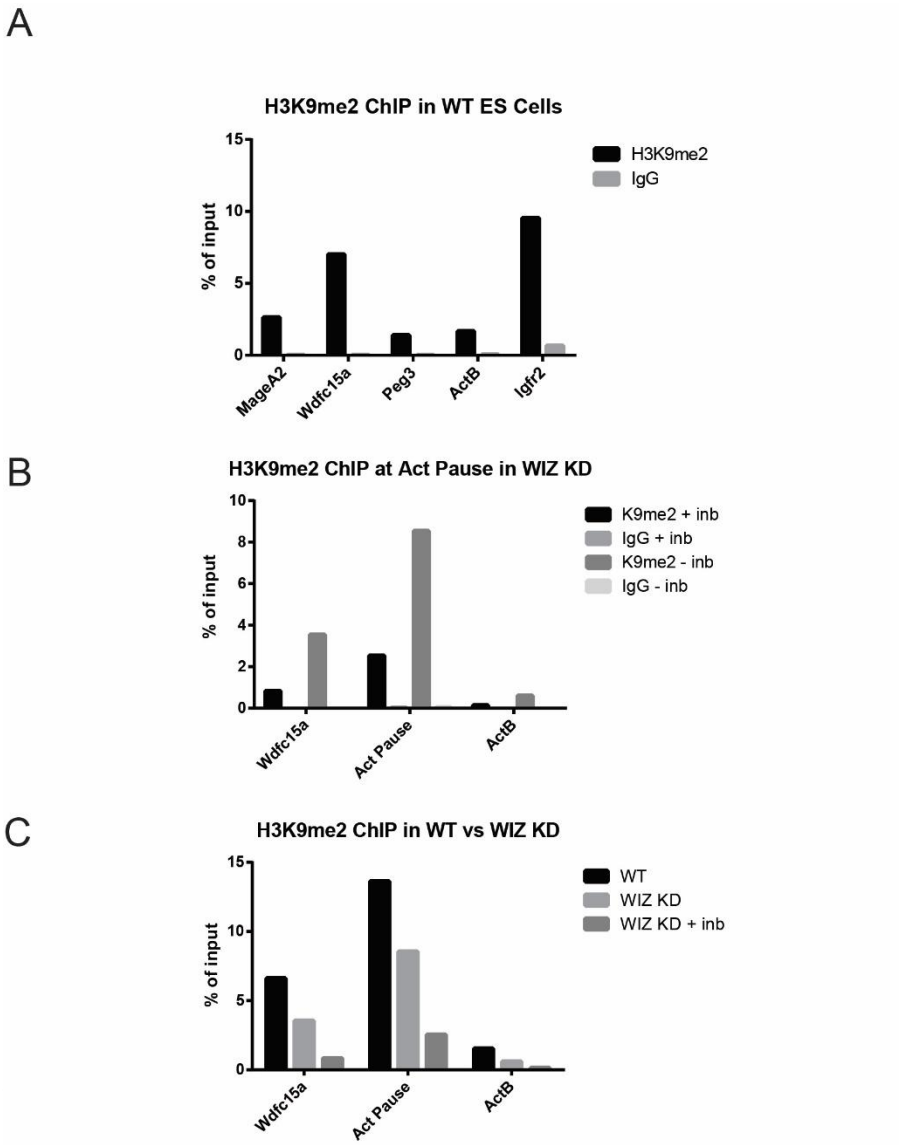


Figure 4. 8 H3K9me2 enrichment is directly affected by the WIZ knockdown

(A) ChIP for H3K9me2 at known G9a-targeted loci. Wdfc15a, Mage-A2, Igf2r and Peg3 are used as positive controls, while ActB is used as a negative control. The Y-axis shows percentage enrichment relative to input DNA. IgG is used as a negative control antibody. **(B)** ChIP for H3K9me2 at Actin pause site in WIZ KD cells. Similarly to (A), Wdfc15a and ActB are used as positive and negative controls, respectively. **(C)** ChIP for H3K9me2 at the Actin pause site in WIZ KD cells compared to WT cells. UNC0638 (G9a inhibitor) treatment reduces H3K9me2 at the Actin pause site in WIZ KD cells.

4.4. Discussion

The SWS extraction experiments performed on WT ES cells revealed that the majority of G9a (70%) is tightly associated with chromatin. This result was expected since it is known that G9a can recognise H3K9me1/2 with high affinity. My experiments quantitatively showed that only 30% of all G9a lost tight chromatin-binding after 80% reduction in H3K9me2 levels. The remaining 40% was observed in "tight" fraction indicating that H3K9me2 is not the only factor important in the recruitment of G9a to chromatin. Moreover, in ANKΔ-DM cells, 40% of G9a remained in the tight fraction. I can conclude here that this fraction of G9a does not rely on H3K9me-binding for stable association with chromatin.

H3K9me1 mark could be another factor that mediates the tight binding of G9a to chromatin. The data generated in our lab indicated that the G9a inhibitor does not reduce H3K9me1 levels in ES cells. This is not surprising because H3K9me1 mark is catalysed by PRDM13 and PRDM16 in the cytoplasm (Pinheiro et al. 2012). H3K9me1 mark is further methylated to H3K9me2 or H3K9me3 in the nucleus by G9a/GLP, Suv39h and SETDB1, respectively. Steady levels of H3K9me1 after the G9a inhibitor treatment suggests that G9a/GLP catalytic activity is mostly required for H3K9me1 to H3K9me2 transition, but not for H3K9me0 to H3K9me1 in the nucleus. Another implication is that demethylase activity on H3K9me1 seems to be rather limited. This might be related to the protection of H3K9me1 mark by proteins, such as G9a/GLP, which can recognise H3K9me1 mark and protect it from demethylation. A weak affinity of lysine demethylase JMJD2E to UNC0638 was detected *in vitro* (Vedadi et al. 2012). One study has shown

that another G9a inhibitor (BIX-01294) can bind to the catalytic pocket of H3K9me2 demethylase JHDM1D (Upadhyay et al. 2012). Therefore, the G9a inhibitor treatment in my experiments might have inactivated the H3K9me1 demethylases as well. However, additional tests are required to confirm this possibility. Since I could not alter the levels of H3K9me1, I cannot directly conclude that H3K9me1 does or does not affect the high-affinity chromatin binding of G9a. My data could be supported by a complementary experiment where ES cells treated with inhibitors for PRDM13 and PRDM16 to reduce the levels of H3K9me1. However, such inhibitors are yet to be discovered.

Interestingly, after the G9a inhibitor treatment, demethylase activity on H3K9me2 seems to be very limited as well. ES cells that I used in my experiments have a doubling time of 16-18 hours. It would be expected to detect about 50% decrease in H3K9me2 levels every time cells divide if they were cultured in the G9a inhibitor containing media. Considering that 48 hours, which is equivalent to three doubling times, treating ES cells for two days would be enough to reduce H3K9me2 levels to 50% after the first doubling, 25% after the second doubling and 12.5% after the third doubling. Experiments that I carried out indicated about 85% reduction in H3K9me2, which clearly shows that reduction of H3K9me2 is not caused by any demethylase activity but a steady dilution by cell division. However, it is important to highlight here that H3K9me2 is a very abundant modification. If the demethylase activity occurs at specific loci, it could still be crucial for gene regulation.

The H3K9me2 mark is not only recognised by G9a/GLP but also HP1 proteins (Lachner et al., 2001). Furthermore, it was found that auto methylation of G9a creates a binding site

for HP1 proteins (Sampath et al. 2007). This study showed that the dissociation constant for HP1-H3K9me3, HP1-G9aK165me2 and HP1-G9aK165me3 is 17 μM, 38 μM and 48 μM, respectively. Although ankyrin repeat domain of G9a displayed a higher affinity (6 μM) for H3K9me2 when the same technique was used (Collins et al. 2008), HP1 proteins are abundant and can form multimeric complexes. As discussed earlier, all HP1 proteins (CBX1, CBX3 and CBX5) have been identified as a binding partner of G9a. It is possible that once the H3K9me2 mark is established by G9a/GLP, the HP1 proteins bind the mark and recruit methylated-G9a for maintenance of H3K9me2.

An interesting observation was that nuclear extraction profile of G9a in DNMT3A/3B DKO ES cells was rather similar to the one in ANKΔ-DM cells. The Ankyrin repeat domain of G9a is known to interact with the methyltransferase domain of DNMT3A/B (Epsztejn-Litman et al. 2008). In fact, the catalytic activity of G9a was found to be dispensable in the G9a-mediated recruitment of DNMTs to Oct4 regulatory regions during ES cell differentiation. Moreover, HP1β was shown to interact directly with DNMT1 (Fuks et al. 2003), and HP1α was shown to interact with DNMT3B in MEFs (Lehnertz et al. 2003). Another study showed that HP1-DNMT1 interaction stimulates the methyltransferase activity of DNMT1 (Smallwood et al. 2007). Therefore, the evidence suggests an interplay between DNA and H3K9me2 methylation, which seems to be mediated by HP1 proteins. My experiments indicate that high-affinity chromatin binding of G9a requires both DNA methylation and H3K9me2 mark, which is also needed for the formation and maintenance of heterochromatin by HP1. For this reason, investigating the

nuclear extraction profile of G9a in HP1-deficient cells would be a right approach to support this hypothesis.

Unlike pluripotent inner cell mass of the mouse blastocyst, *in vitro* cultured mouse ES cells have high levels of DNA methylation when grown in media containing serum. Studies described a ground state of ES cells, in which FGF receptor tyrosine kinases and the ERK cascade are inhibited by two inhibitors (2i) and LIF kept self-renewal of ES cells preserved (Ying et al. 2008). It was also shown that ES cells grown in 2i+LIF condition are hypomethylated and more homogenous while ES cells grown in LIF+Serum condition are hypermethylated and more heterogeneous (Habibi et al. 2013). Considering these facts, it would be interesting to perform an SWS extraction using ES cells cultured in 2i condition and compare the nuclear extraction profile to one obtained from DNMT3A/3B DKO ES cells.

The H3K9me2 recovery after G9a inhibitor treatment experiment suggested that the catalytic activity of G9a is not dependent on the presence of WIZ. Although WIZ is important for the stability of G9a, after G9a inhibitor treatment, H3K9me2 recovery rate in WIZ KD cells was comparable to WT. Interestingly, both cell lines could recover the lost H3K9me2 after removal of the G9a inhibitor in 48 hours. As discussed earlier, this is the time required for passive loss of H3K9me2 by cell division in the absence of the catalytic activity of G9a. One study showed that G9a and DNMT1 localise at the replication fork (Esteve et al. 2006). If the complete recovery requires 48 hours, then it means that catalytic activity of G9a might be connected with the replication machinery. In other words, the unmethylated H3 population might gradually decline as G9a creates

more H3K9me2 mark in each cell cycle. Another possibility is that the UNC0638 binds very tightly to G9a and a protein turnover might be required for proper catalytic activity. To my knowledge, the turnover rate for G9a has not been measured. This information would be instrumental for testing the replication-dependent activity model of G9a. Additionally, since DNMT1 is recruited to newly replicated DNA by UHRF, a replication-dependent mode of action of G9a can be tested in UHRF1 knockout cells as well.

SWS experiments on WIZ KD cells did not show an apparent change in the extraction profile of G9a. However, upon the inhibitor treatment, WIZ seemed to affect the weak and dynamic fraction of G9a. As a technique, SWS extraction was not sensitive enough to detect small changes in extraction profile. Using a technique that identifies the proteins on nascent DNA of replication fork with the help of mass spectrometry (iPOND-MS), WIZ was identified as a protein that facilitates the DNA replication rate in the replication fork (Lopez-Contreras et al. 2013). Since replication forks move along the chromatin to ensure error-free duplication of DNA, it would be likely that proteins regulating the motion are highly dynamic. Therefore, it would not be surprising to observe WIZ as more important for the dynamic fraction of the G9a complex.

As introduced previously, it was shown that G9a is recruited to the 3’ end of genes to create a heterochromatic environment facilitating termination of transcription (Skourti-Stathaki et al. 2014). R-loops formed at these sites consist of a G-rich single-stranded DNA and an RNA-DNA hybrid. My initial intent was to investigate any enrichment of WIZ at the R-loop at the 3’ end of the β-actin gene (Actin pause site). As discussed earlier, I could not generate ES cell line expressing tagged WIZ. The only available WIZ antibody

is not commercially available. Therefore, I investigated the presence of H3K9me2 at Actin pause site. The ChIP-qPCR results for WT and WIZ KD cells suggest that WIZ does not recruit the G9a complex to Actin pause site. However, I cannot conclude that G9a enrichment was not reduced at this site. Although the protein levels of G9a was reduced to 20%, the WIZ KD cells still produce about 20% of endogenous WIZ proteins. Low levels of WIZ might be enough for recruiting G9a to actin pause site. The H3K9me2 mark at this site can be maintained by G9a/GLP heterodimer if remaining WIZ mediates the establishment of the mark. A WIZ knockout ES cell line would have answered above problem very clearly. Another possible way to tackle this is to use a WIZ KD cell line that produces mutant G9a and GLP that are deficient in binding to H3K9me2. In such cell line, WIZ-mediated recruitment of G9a could be traced very clearly by detecting H3K9me2 at any genomic site including R-loops.

In summary, in this chapter, I focused on the factors that contribute to chromatin binding of G9a. H3K9me2 mark and probably DNA methylation facilitate the tight association of G9a with chromatin. However, my data also showed that there must be additional factors. My experiments revealed that WIZ is dispensable for the association of G9a with chromatin. In other words, self-recruitment of G9a to heterochromatin is not mediated by WIZ, but site-specific recruitment of G9a might still be dependent of WIZ in some cases. This hypothesis would be correct only if WIZ is a DNA-binding zinc finger protein. To test this idea, I investigated whether or not the zinc fingers of WIZ bind DNA *in vitro*, and examined the sequence specificity of this binding.

Chapter V – Purification and DNA binding preferences of the WIZ zinc finger pairs

5.1. Introduction

The Mouse WIZ cDNA was isolated for the first time as an off-target of screening the brain of ICR mice cDNA library with probes specific for of *Drosophila eya* gene (Matsumoto et al. 1998). WIZ cDNA shared 53% sequence similarity with *Drosophila eya* at probe-hybridizing region but two genes were not homologous in general. When focused on the WIZ mRNA, the same study showed that mouse WIZ mRNA was alternatively spliced into two isoforms; the short and long isoforms are translated into proteins that are 956 and 1684 amino acids in length, respectively. The long isoform of WIZ contains 11 zinc finger domains, of which the first 5 are missing in the short isoform. The specific function of the two isoforms is unknown. In both isoforms, the last zinc finger, a non-canonical C_2H_2-type zinc finger, is responsible for the interaction between WIZ and G9a/GLP (Ueda et al. 2006). As mentioned before, the knockout of *Wiz* gene in mice leads to early embryonic lethality with timing (E8-9.5) similar to G9a and GLP knockouts (Daxinger et al. 2013). Therefore, understanding the function of WIZ is crucial to determine the reason of embryonic lethality in the absence of G9a complex members.

Recently, it was shown that long isoform of human WIZ protein can bind DNA *in vitro*, and using ChIP-seq, the genomic sites with the presence of WIZ were identified in human embryonic kidney cells (Bian et al. 2015). However, this study did not reveal which WIZ

ZF domains were responsible for the DNA-binding of WIZ. A more recent report using a ChIP-seq approach showed that WIZ co-localizes with CTCF binding sites in mouse cerebellum and suggested a transcriptional activator role for WIZ (Isbel et al. 2016). Interestingly, this study identified 6 different WIZ DNA binding motifs after analyzing the WIZ peaks, which were completely different from the WIZ binding motif suggested by the first study. Considering that WIZ is a core G9a complex component, detecting WIZ at a particular genomic site might be due to recruitment of the G9a complex by other factors. Therefore, more evidence of the DNA binding preference of WIZ zinc fingers is required to identify genomic sites that are directly targeted by WIZ. In this chapter, I will focus on the *in vitro* DNA binding properties of WIZ zinc finger pairs and their DNA binding preferences.

Systematic evolution of ligands by exponential enrichment (SELEX) has been used to identify sequence-binding properties of DNA- and RNA-binding proteins for many years. The technique uses a random-sequence library of nucleic acids, which become enriched in preferred binding sites for the protein of interest through a sequential *in vitro* selection process (see materials and methods). The first study to utilize this technique identified the DNA-binding motif of a eukaryotic transcriptional activator protein GCN4 (Oliphant et al. 1989). The following year, the RNA sequences bound with high affinity by the bacteriophage T4 DNA polymerase were identified using the SELEX protocol (Tuerk & Gold 1990) and as well as the RNA-binding sequences of specific small ligands by SELEX (Ellington & Szostak 1990). In the recent years, the whole-genome sequencing approaches have become more powerful and widely-used. Chromatin

immunoprecipitation sequencing (ChIP-seq) followed by search for specific sequence motifs can be used to determine the binding preferences of DNA- or RNA-interacting proteins. However, an informative ChIP-seq experiment requires good quality antibodies against the protein of interest. Moreover, due to the crosslinking step in the ChIP-seq experiments, genomic sites that are indirectly relevant to the protein of interest cannot be distinguished from the direct targets.

SELEX is usually coupled with electrophoretic mobility shift assay (EMSA), which enables the identification of *in vitro* protein-DNA interactions. In a standard EMSA experiment, the protein of interest and radioactively- or fluorescently- DNA are run on a non-denaturing (native) PAGE. Protein-DNA complexes migrate slower and are reflected as a band shift on the gel, while the unbound DNA is detected at the bottom of the gel. Radioactive labelling of DNA is usually achieved by incorporating a ^{32}P at the either end of the DNA molecules by specific enzymes. Radioactivity provides a better sensitivity in EMSA experiment in the exchange of difficulty in handling of the experiments due to safety measures. On the other hand, fluorescent labelling of DNA is achieved by incorporating an infrared (IR) dye to the oligo of interest, which is commercially available option nowadays. As an advantage, fluorescently labelled DNA is ready to use for further applications, and since there are many IR dyes available to use, more than one DNA species can be tested for *in vitro* DNA binding at the same time.

As described in the original WIZ study, WIZ is highly expressed in the brain of both embryo and adult (Matsumoto et al. 1998). G9a loss of function causes a neurological disorder, Kleefstra syndrome, a human disorder characterized by severe intellectual

disability. A genetic screening in patients with Kleefstra syndrome identified a nonsynonymous variant of WIZ (Balan et al. 2014). Although there is no conclusive evidence, it is thought that mutations of *Wiz* gene might cause Kleefstra syndrome. Moreover, behavioral tests on WIZ heterozygous mice suggested anxiety-like phenotype indicating that reduced WIZ levels might have neurological consequences (Isbel et al. 2016).

5.1.1. **Aims**

The DNA binding of the full-length human WIZ protein was shown. However, the exact DNA-binding properties of WIZ is poorly understood. In this chapter, I aimed to identify the DNA-binding zinc finger(s) and the specific binding sequence of mouse WIZ. To achieve this, I focused on the short isoform of mouse WIZ. I purified recombinant GST- or His-tagged WIZ zinc finger (ZF) pairs expressed in bacteria, and used the purified WIZ ZF pairs in SELEX experiments to identify a binding motif for WIZ ZF pairs.

5.2. Purification of recombinant WIZ zinc finger pairs

Genome editing studies with transcription activator–like (TAL) effector proteins required engineering of zinc fingers proteins with sequence-specific binding properties. In fact, the structure of many ZF proteins has been solved by X-ray crystallography and their binding preferences were extensively studied (Klug 2010). This enabled development of prediction software for zinc finger proteins. I used an online tool to determine which zinc

finger domains of WIZ display a higher probability for DNA-binding. This particular software (http://zf.princeton.edu/index.php) takes the amino acid sequence of any C_2H_2-type zinc finger as an input and predicts the DNA binding specificity by using a set of prediction algorithms (Persikov et al. 2009) (Figure 5.1A). Among the first 5 canonical zinc fingers of WIZ, all but ZF5 had a score greater than 17.7, which is the minimum score for the zinc fingers that possess DNA binding capacity. Additionally, due to its unusual amino acid sequence, WIZ ZF6 had a score of 0.3, suggesting that it is very unlikely to bind DNA. It was an expected result since WIZ ZF6 is involved in the interaction between WIZ and G9a/GLP (Ueda et al. 2006). WIZ ZF3 and ZF4 share the exact same amino acid sequence and had the highest score binding score (23.3) with a predicted binding sequence – "ATCG". WIZ ZF1 and ZF2 had a score of 19.7 and 21.1, respectively. These 4 zinc fingers were considered as the candidate DNA binding fingers of WIZ. On the other hand, WIZ ZF5 had a score of 15.5, which was very close to the threshold. Therefore, it was worthwhile to investigate DNA binding of WIZ ZF5 along with other fingers.

My first aim was to determine which zinc fingers of WIZ could bind DNA *in vitro.* Knowing that many zinc finger proteins bind to DNA via a combination of their zinc finger domains (Iuchi 2001), I decided to perform *in vitro* DNA binding assays using WIZ zinc finger pairs. For this purpose, I designed WIZ zinc finger pairs as shown in Figure 5.1B-C. It is important to highlight that I left 10 amino acids at the either side of each zinc finger and I did not alter the sequence between neighbor fingers. The coding sequence of WIZ ZF pairs were codon optimized for bacteria to improve the yield of protein in the purification experiments.

Initially, using pRSF-DUET expression plasmid, I cloned coding sequence of WIZ ZF pairs in frame with the 6xHis-tag in the plasmid (Figure 5.1B). The cloning was confirmed by restriction enzyme digestion and sequencing. I expressed and purified the His-tagged WIZ zinc finger pairs (Figure 5.2A). The purified protein yield in the initial purification attempts were quite poor. One problem was that at least half of the protein after the IPTG induction was insoluble, which reduced the yield of purified protein (Figure 5.2A, left). Next, I applied the optimized purification protocol to the other zinc finger pairs apart from ZF1-2 (Figure 5.2A, right). Interestingly, His-ZF3-4 had two distinct and strong bands corresponding to 27 kD, which is the expected size of the recombinant protein, and 22 kD. When I carried out a Western blot, I was able to detect both bands by an antibody against 6xHis tag (Figure 5.2B). Moreover, in all purifications, I detected a band at 70 kD, which probably corresponded to a bacterial protein that either interacts with charged nickel beads non-specifically or the WIZ ZF pairs. Considering that band might be a bacterial chaperon, I added 10 µM ATP to the wash buffer during protein purification. However, this did not clean the non-specific band from the eluates. The contaminant protein was not detectable by Western blotting, indicating that the contaminant protein was not a bacterial protein with an endogenous 6xHis-tag.

It is common to increase the stability and solubility of a recombinant protein by changing the tag. Glutathione S-transferase (GST) is particularly effective in this manner (Harper & Speicher 2011). For this reason, I expressed all zinc finger pairs as GST fusion proteins. Overall, using GST-tag increased the stability of recombinant zinc finger pairs. Overnight dialysis did not result in the protein precipitation and there was a marginal increase in

protein yield. I purified all zinc finger pairs for the following *in vitro* DNA experiments (Figure 5.2C). Changing the tag and the beads for purification did not eliminate the presence of the non-specific band I detected in the His-tagged protein purification. To further analyze the possible interaction between the WIZ ZF pairs and this unknown protein, I inserted a TEV cleavage site between the GST tag and the coding sequence of WIZ ZF3-4 (GST-T-ZF3-4). I expressed and purified the GST-T-ZF3-4 protein, and cleaved the GST-tag out by TEV digestion (Figure 5.3A). Tagless ZF3-4 protein was visible in the flowthrough, as well as the unknown protein. The GST and GST-TEV were detected on the beads. This experiment showed that the recombinant zinc finger pairs directly interact with the contaminant bacterial protein.

It was important to show that the contaminant protein does not cause a DNA band shift. To this end, I tried to remove the contaminant protein from the purified protein samples by ACTA system using gel filtration columns. I used ZF4-d5 that is smaller in size than other zinc finger pairs. However, it was still not possible to separate GST-ZF4-d5 and the contaminant protein by gel filtration. Next, I used anion exchange chromatography, which was able to remove the contaminant protein from the pure GST-ZF4-d5 (Figure 5.3B). GST-ZF4-d5 appeared in the void fractions indicating that it did not have a charge at the experimental condition, while the contaminant protein had a positive charge and was eluted at about 100 mM salt concentration. I loaded fractions corresponding to the void sample and the contaminant protein on a 15% SDS-PAGE gel and stained the gel by coomassie. Interestingly, there was an additional contaminant protein at about 25 kD, which was not distinguishable from the degradation products previously. Moreover, there

seemed to be equal amounts of contaminants and the zinc finger protein in D3-D5 fractions. This indicates that two unknown bacterial proteins directly interact with zinc finger pairs in stoichiometric amounts.

5.2.1. *In vitro* DNA binding assays with recombinant WIZ zinc finger pairs

To test the DNA binding of WIZ zinc finger pairs, I used a DNA library consisting of random-sequence DNA. This was essential in the case that WIZ ZF pairs show a sequence specificity in their DNA-binding. The DNA library (2N10) consisted of 59 bp-long oligomers flanked by forward and reverse primer sites. Each unique sequence had a single CG di-nucleotide flanked by 10 bp of random bases (22 bp). After generating the double-stranded 2N10 DNA library, I carried out an *in vitro* DNA-binding assay to determine whether or not the WIZ ZF pairs could bind DNA. I used the GST-tagged zinc fingers because they showed better stability and purity. The EMSA experiment showed that apart from GST-ZF1-2 and the negative control (GST), all zinc finger pairs were able to bind DNA (Figure 5.4A). However, the signal from the shifted bands, as expected, corresponded to only a small fraction of the total DNA. This was an indication that the WIZ ZF pairs might bind DNA in a sequence-specific manner. The EMSA also revealed that WIZ ZF1 and ZF2 cannot bind DNA *in vitro.* However, GST-ZF2-3 was able to shift DNA indicating that ZF3 is the DNA-binding zinc finger. Additionally, the amino acid sequences of ZF3 and ZF4 are identical. Therefore, it could be concluded that WIZ ZF3 and ZF4 were producing the bandshifts observed in other lanes. From this experiment, I could not determine whether or not WIZ ZF5 could bind DNA *in vitro.* However,

according to the zinc finger prediction software, ZF5 had the lowest score. Therefore, it would not be surprising if only ZF3 and ZF4 were the only DNA-binding fingers of WIZ.

To verify the previous experiment with a positive control protein that has a similar size to GST-tagged ZF pairs, I carried out another *in vitro* DNA binding assay using GST-ZF3-4, which was identified as the best DNA-binder among other ZF pairs, and a known DNA binding protein. GST fusion proteins are known to dimerize in non-denaturing condition. Therefore, GST-ZF3-4 (dimer) and LSH (monomer) are both about 100 kD in size. Another lab member purified the chromatin remodeler LSH and he showed that LSH binds DNA *in vitro*. I carried out the EMSA experiment using these proteins and the 2N10 DNA, and was able to confirm that LSH indeed binds DNA as well as GST-ZF3-4. I also used unlabeled 2N10 DNA as a competitor in the DNA-binding reaction. The competition was very effective for GST-ZF3-4 when unlabeled DNA was used 10 times in excess to the labelled one (Figure 5.4B). However, this ratio was insufficient for LSH-labelled DNA binding, which indicates that LSH has much higher affinity for DNA than GST-ZF3-4.

5.2.2. **A point mutation in ZF4 abolishes DNA binding**

My data so far indicated that WIZ ZF3 and ZF4 are the DNA binding fingers of WIZ. In order to firmly confirm that this as the case, I introduced point mutations in the coding sequences of WIZ ZF pairs (Figure 5.1C). First, I introduced a nonsense mutation just before the second zinc finger of GST-ZF3-4 and GST-ZF4-5. Then, I mutated the triplet encoding the lysine residue at the position 1 of the first zinc finger into a triplet encoding an alanine, creating GST-ZFm3-d4 and GST-ZFm4-d5. The residue at position 1 on the ZF1, which does not bind DNA, was an alanine. Therefore, I chose K→A mutation for

the mutant form of both ZF3 and 4 (since they have identical amino acid sequence) to create a loss of function mutations. DNA binding assay using the GST-ZF4-5 and its mutant forms showed that GST-ZF4-5 and GST-ZF4-d5 were able to shift DNA, while GST-ZFm4-d5 did not (Figure 5.5A). This further supports the conclusion that ZF4 in the ZF4-5 pair is the DNA binding finger of WIZ. I was able to confirm this result by performing another EMSA with all other DNA-binder zinc finger pairs (Figure 5.5B). Of note, I tried to mutate the asparagine at position -1 to a threonine, which also mimics the ZF1 amino acid sequence. However, this mutation did not abolish DNA binding. The amino acid sequences of ZF3 and ZF4 are identical. Therefore, I did not test the ZF4 mutations in ZF3 as well but it is safe to conclude that WIZ binds DNA via ZF3 and ZF4.

5.3. Identification of sequences bound by GST-ZF3-4 with high affinity

In vitro DNA binding assays that I performed so far showed that ZF3 and ZF4 are the DNA-binding zinc fingers of WIZ. Next I wanted to determine what kind of sequences were more likely to be bound by GST-ZF3-4. For this purpose, I carried out a SELEX experiment with recombinant GST-ZF3-4 and double stranded 2N10 DNA library. This technique relies on a series of *in vitro* DNA-binding assays and the purification of DNA-protein complexes, which increase the fraction of sequences in the DNA pool that are bound with higher affinity by the protein of interest after each SELEX cycle (Figure 5.6). I will refer to the initial pool of 2N10 as S0 DNA hereafter. I tested a different condition that would affect the DNA binding of WIZ ZF pairs while carrying out the SELEX cycles. I finalized the SELEX experiment after 10 cycles because no further enrichment was

detected at this point (Figure 5.7A). GST-ZF3-4 showed higher affinity to S10 DNA when compared to S0 DNA (Figure 5.7B, left). Moreover, unlabeled S0 DNA failed to compete with S10 DNA, while unlabeled S10 DNA drastically reduced the visible band shift caused by labelled S0 DNA (Figure 5.7B, right). Altogether these results show that the SELEX experiment successfully generated a DNA pool (S10 DNA) that contained sequences that can be bound by GST-ZF3-4 with high affinity.

5.3.1. Identification of sequences bound by GST-ZF3-4 with high affinity

Following the previous experiment, I decided to identify the binding motif of GST-ZF3-4 by analyzing the DNA sequences in S10 DNA. To achieve this, I cloned the S10 DNA and sequenced 105 DNA fragments. Among these, I was able to identify 65 unique DNA sequences (Figure 5.8A). Some sequences were shared by more than one colony, which either indicates that those sequences were more prominent in the S10 pool or originated from the same bacterium. Analyzes of the base pair composition of all unique SELEX sequences revealed that G-C base pairing occurred more frequently than A-T in S10 DNA, when compared to the initial DNA pool represented by 65 sequenced fragments (Figure 5.8B). More interestingly, I observed a strong tendency for G or C skew in the unique S10 sequences (Figure 5.8C, also Figure 5.8A). A skew is the parameter that represents the amount of bias for a nucleotide in a single stranded DNA molecule. In the case of the unique S10 sequences, a skew for G (or C) was calculated as [|(# of G) - (# of C)|] / [(# of G) + (# of C)]. The mean of values of G (or C) skew for unique S0 and S10 sequences were 0.22 and 0.77, respectively.

I also used a motif identification software, Multiple Em for Motif Elicitation (MEME) (http://meme-suite.org/tools/meme), to identify a DNA-binding motif for GST-ZF3-4. Apart from the sequences to be analyzed, MEME-suite required several parameters that could alter the result of the analysis dramatically. One of these parameters was the width of the motif. As mentioned earlier, the amino acids at the position -1, 1, 2 and 6 of a C_2H_2-type zinc finger usually make physical contact to 3 or 4 bases of a DNA molecule. As most C_2H_2-type zinc fingers are tandemly arrayed on a zinc finger protein, protein-DNA interaction might span a longer DNA sequence. The crystal structure of WIZ has not been solved, which means that depending on their 3D conformation, widely interspaced zinc fingers of WIZ might cooperate to bind a longer motif or they can interact with DNA at separate sites (Figure 5.9A). I chose two different widths in my motif analysis: a 6-bp motif would be very accurate if both zinc fingers of GST-ZF3-4 cooperate, while a 21-bp motif would be more realistic if zinc fingers bind DNA with a certain distance between two contact sites. It is important to note that the software cannot efficiently predict non-random motives that are shorter than 6-bp. MEME identified a 6-bp motif in all 65 unique S10 sequences with an e-value of 4.0e-025 (Figure 5.9B). A 21-bp motif appeared in all 65 unique S10 sequences, as well, and had a lower e-value than 6-bp motif. However, the 21-bp motif was less consistent in terms of nucleotide occurrence at indicated positions. The base occurrence was significantly more consistent at the both ends of SELEX DNA. Therefore, it is possible GST-ZF3-4 might require a certain distance between the two putative binding sites due to its widely interspaced zinc fingers structure.

To get a better understanding about the binding motive of GST-ZF3-4, I carried out a SELEX experiment using GST-ZF3-d4. Interestingly, 7 cycles were sufficient to reach to the maximum SELEX enrichment (Figure 5.10A). I sequenced more than 70 colonies and obtained 30 unique DNA fragments. Analyzes of base pair frequency revealed that GST-ZF3-d4 also preferred G-rich DNA (Figure 5.10B). Moreover, there was a significant G (or C) skew tendency in this SELEX pool as well (Figure 5.10C). However, the MEME analysis of the GST-ZF3-d4 SELEX pool produced a 6-bp motif for GST-ZF3-d4 which was not identical to the 6-bp motif of GST-ZF3-4. These motives shared a common starting 'G' nucleotide and were rich in G and T (Figure 5.9B). More importantly, in 21 bp-motive of GST-ZF3-d4, there was only one putative binding sites. This was most likely due to the deletion of the second zinc finger. The data so far suggested that WIZ ZF 3-4 prefers DNA sequences with a G (or C) skew and probably requires two binding sites that are certain distance apart.

5.4. Further analysis of S10 DNA pool

The binding motifs of WIZ ZF 3-4 were not 'clear-cut' that can be searched for in the mouse genome for possible WIZ-target sites. It was possible that SELEX cycles failed to eliminate low affinity sequences in the pool during the SELEX cycles. My ultimate aim was to test the WIZ-mediated recruitment of G9a, for which I needed to identify sequences bound by WIZ with high affinity. Then it would be possible to insert this sequence into the genome and ask whether it is sufficient to recruit G9a/GLP. For this purpose, I aimed to identify such DNA sequences for GST-ZF3-4 in the sequenced S10 DNA pool. To do

so, I divided the sequenced S10 DNA pool into six subgroups, M1 to M6 (Figure 5.8A) and used these pools in EMSA experiments to compare the affinity of ZF3-4 to these sequence pools and S10 DNA as a positive control (Figure 5.11A). GST-ZF3-4 shifted DNA from each subgroup mix as efficiently as S10 DNA, which indicated that there was a least one sequence bound by GST-ZF3-4 with high affinity in each subgroup. I also investigated the affinity of these subgroups to the other DNA binder zinc finger pairs and found that all DNA-binder zinc finger pairs were able to shift the M2 subgroup DNA (Figure 5.11B). Interestingly, GST-ZF2-3 and GST-ZF4-5 were able to shift more labelled M2 DNA than S10 DNA. This was not detectable in the case of GST-ZF3-4, which, again, indicated that S10 DNA was enriched specifically for sequences bound by GST-ZF3-4 with high affinity. Unexpectedly, when the individual sequences of the M2 subgroup were mixed after labelling by PCR, the band shift intensity was significantly reduced in GST-ZF3-4 lane. In the case of GST-ZF2-3 and GST-ZF4-5, the bandshift at the expected location was almost abolished and a non-specific shift occurred at a higher position. This phenomenon could be explained by that the DNA bound by GST-ZF3-4 with high affinity in the template mix might be amplified more efficiently due to a PCR bias. This would increase the concentration of such DNA in M2 subgroup, while mixing the labelled DNA after PCR would dilute the concentration of those DNA in M2* subgroup.

Following this, I investigated the affinity of GST-ZF3-4 to individual DNA sequences in the M2 subgroup (Figure 5.12A). Surprisingly, GST-ZF3-4 was able to shift only T22, T24, T25 and T46 fragments while for the other DNA fragments the shift was either undetectable or very weak. Moreover, none of the band shifts were as strong as the S10

shift. This suggests that the SELEX cycles failed to eliminate DNA bound by GST ZF3-4 with low affinity. Performing a series of EMSA experiments, I analyzed more than 40 individual sequences, and identified DNA sequences that GST-ZF3-4 binds to with moderate to high affinity (Figure 5.12B). In parallel to earlier observations, none of these sequences caused a bandshift as strong as the S10 DNA. When I focused on T25 sequence for further analysis, I realized that it was actually a mix of several DNA fragments. I isolated the DNA bound by GST-ZF3-4 with the highest affinity in T25 mix (Figure 5.13A). I analyzed other DNA with similar sequence to T15 in S10 DNA pool. GST-ZF3-4 showed moderate affinity for a subset of these DNA but none of these DNA caused a bandshift as strong as T15 (Figure 5.13B). In conclusion, I identified the DNA fragment (T15) bound by GST-ZF3-4 with highest affinity among more than 40 individual S10 DNA sequences.

5.5. Narrowing down the possible binding site for ZF 3-4

To identify the GST-ZF3-4 binding sites on T15 DNA, I focused on the two putative binding sites identified by MEME (indicated with red lines in 5.13B). To investigate these in more detail I used synthetic variants of T15 DNA (Figure 5.14A), which differ in size or composition of putative binding sites. The EMSA with T15 variants and His-ZF3-4 (Figure 5.14B) showed that the T15short sequence was sufficient to cause a bandshift indicating that the flanking primer sites did not contribute to DNA binding. Intriguingly, His-ZF3-4 failed to bind T15LL and T15RR despite the expectation that they contained two possible binding sites. This clearly indicates that there must be a certain spacing

between the two binding sites for the most efficient DNA binding of ZF3-4. Interestingly, in the absence of only one primer site (T15Lfull and T15Rfull), the DNA binding of ZF3-4 was dramatically reduced. This could be due to a possible secondary structure occurring between the primer site and the closest ZF3-4 binding site. Such a secondary structure might block the binding of ZF3-4 to the DNA, which supports the idea that ZF3-4 requires two binding sites on a DNA molecule for the most efficient DNA binding.

5.6. WIZ ZF3-4 binds single-stranded DNA and RNA *in vitro* better than double stranded DNA

The experiments described so far showed that WIZ zinc finger pairs can bind double-stranded DNA in a sequence-specific manner *in vitro*. As mentioned earlier, WIZ was shown to be present on newly replicated DNA as well as DNMT1 and UHRF1 (Lopez-Contreras et al. 2013). This could be due to the interaction between DNMT1 and G9a at replication foci. However, it is not known whether or not WIZ can promote the recruitment of G9a to replication foci by binding to single-stranded DNA. To investigate this, I used the single and double stranded DNA in EMSA experiments and compared the affinity of His-ZF3-4 to these DNA fragments. As controls, I used an AT-rich DNA and the human WIZ-target sequence identified by others (Bian et al. 2015). His-ZF3-4 was able to bind all single-stranded DNA used in the experiment (Figure 5.15A). More importantly, His-ZF3-4 shifted the single-stranded DNA more efficiently when compared to double-stranded DNA. Additionally, His-ZF3-4 could bind to T15 ssDNA variants without sequence specificity. Interestingly, the His-ZF 3-4 showed a very low affinity towards the

double-stranded human WIZ-target sequence, while it showed a moderate affinity to double-stranded AT-rich sequence. To recapitulate that His-ZF3-4 had better affinity for ssDNA, I performed EMSA experiments using unlabeled single- or double-stranded competitors (Figure 5.15B, left). I observed that the unlabeled T15 ssDNA reduced the band shift signal for T15 dsDNA better than the double-stranded competitor tested against single stranded T15 (Figure 5.15B, right).

In some cases, DNA-binding zinc finger proteins were shown to bind RNA as well (Caricasole et al. 1996; Hall 2005). As discussed earlier, Air and Kcnq1ot1 non-coding RNAs can interact with G9a (Nagano et al. 2008, Pandey et al. 2008). Considering that these interactions might be mediated by WIZ, I investigated whether or not WIZ ZF pairs could bind RNA as well. An EMSA experiment carried out with single-stranded T15 short and 25-mer RNA with sequence different from T15 showed that His-ZF3-4 could bind RNA *in vitro* (Figure 5.16A). I also tested the binding of His-ZF3-4 to dsRNA and found that His-ZF3-4 was able to bind to dsRNA, but the efficiency of the band shift was lower than the band shift for the ssRNA (Figure 5.16B). These results showed that WIZ can bind RNA as well as DNA *in vitro.*

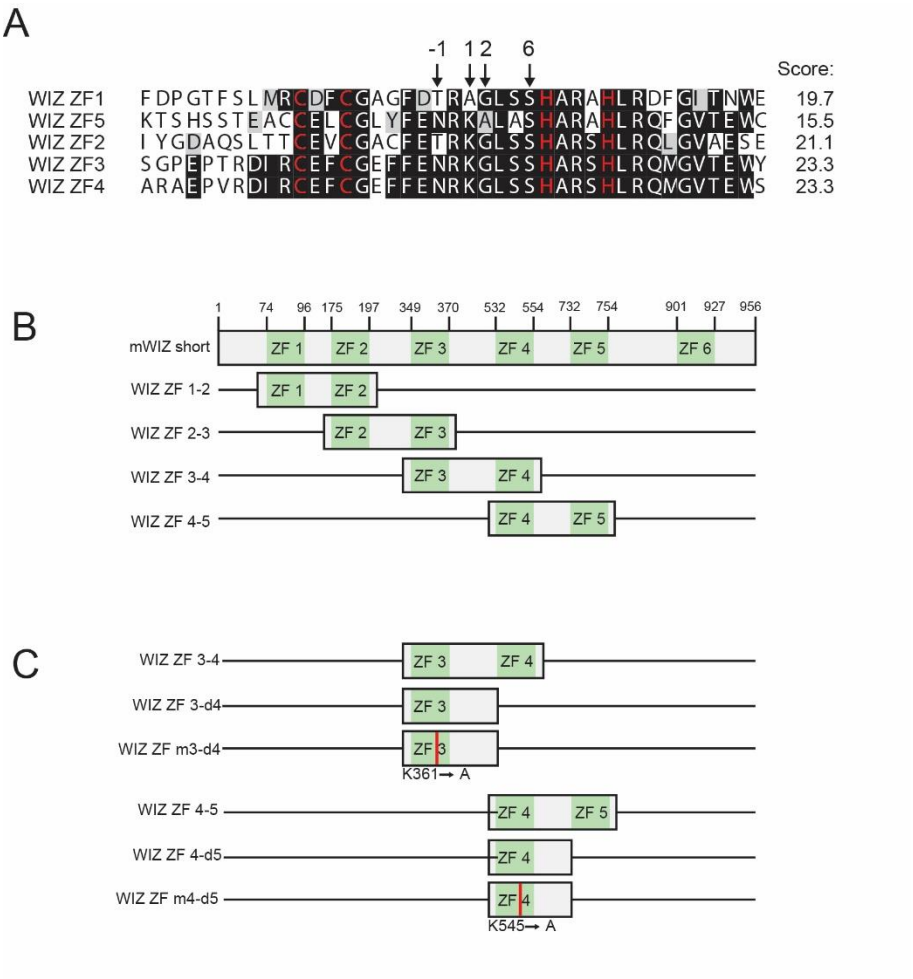


Figure 5. 1 Design of recombinant proteins containing the WIZ zinc finger pairs

(A) Alignment of amino acid sequences of WIZ zinc fingers. Two cystines and histidines of each zinc finger is shown in red. BoxShade software was used to visualise the alignment in color code. The black boxes indicate the perfect alignment, while the grey boxes indicate the amino acids with similar charge. Arrows indicate the position (-1, 1, 2 and 6 from right to left) of amino acids that are normally in direct contact with DNA in C2H2-type zinc fingers. The DNA binding score predicted by a software (http://zf.princeton.edu/index.php) is shown. **(B)** The full-lenght short isoform of WIZ protein and its zinc finger domains. The zinc finger domains of WIZ are indicated with green boxes. The codon optimized sequences of WIZ zinc finger pairs were ordered from GeneArt. PCR products later cloned in frame with either GST or 6xHis tag coding sequence in an expression vector. **(C)** The mutant versions of WIZ ZF 3-4 and WIZ ZF 4-5. Deletions are made by adding a stop codon right before the ZF 4 and ZF 5 in WIZ ZF 3-d4 and WIZ ZF 4-d5, respectively. Point mutations (K mutated to A) were created at position "1" of ZF3 and ZF4 as decribed in (A).

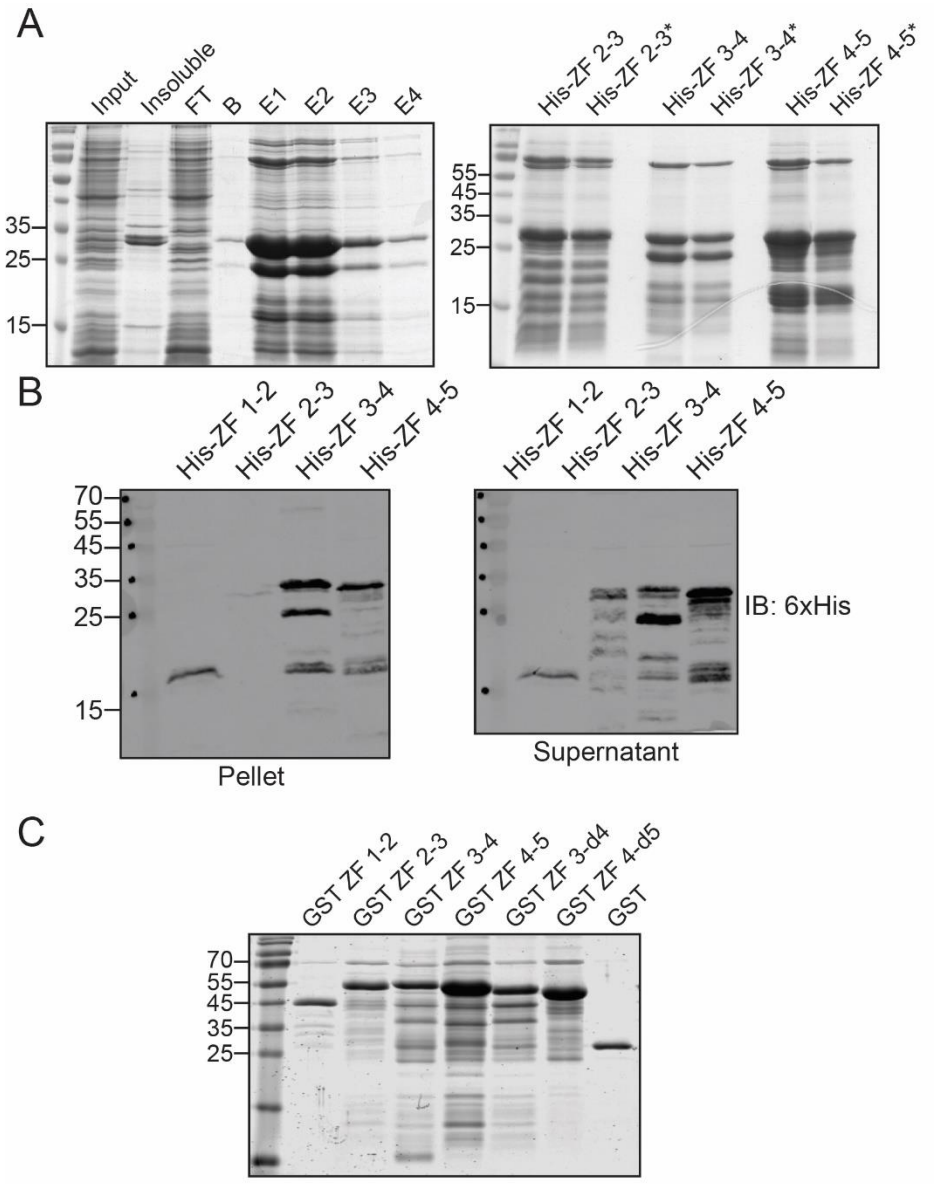


Figure 5. 2 Purification of the recombinant WIZ zinc finger pairs

(A) Purification of His-ZF3-4 by using an optimized protocol. On the left, His-ZF3-4 is purified using chelating agarose charged with nickel ions. The expression of the recombinant protein in BL21 expression strain is low, therefore the protein is purified from 2L bacterial culture. On the right, all DNA-binding WIZ zinc fingers are purified, buffer exchanged with EMSA buffer (shown with *) and concentrated. FT: Flow thorugh after binding, E: eluted fractions, B: the beads after elution. **(B)** His-tagged WIZ ZF pairs analyzed by Western blotting using an antibody against 6xHis. Equal amounts of recombinant protein are loaded on a 15% SDS-PAGE. **(C)** Purified GST-tagged WIZ zinc finger pairs. Mutant forms of GST-ZF3-4 and GST-ZF4-5 are created as described in Figure 5.1A.

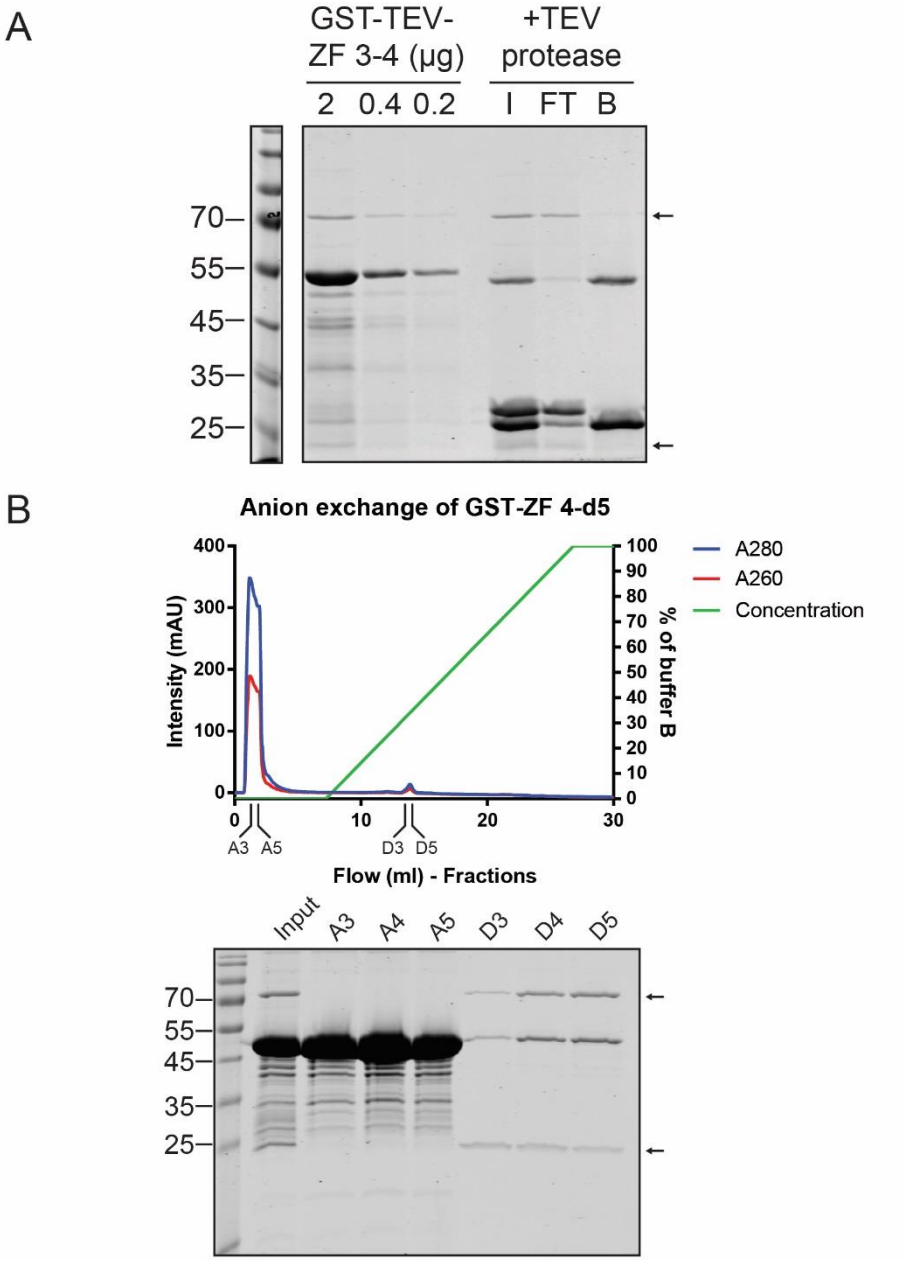


Figure 5. 3 The contaminants directly interact with recombinant ZF pairs

(A) The GST-tag of GST-T-ZF3-4 cleaved by TEV protease. GST-T-ZF3-4 before the cleavage is shown on the left. After TEV cleavage, the sample (I:Input) is bound to the Glutathione Sepharose 4 beads. Flowthrough (FT) contained the tagless ZF3-4 and the contaminant proteins, while beads contained the GST tag and GST-tagged TEV protease. The contaminants at 70 kD and 25 kD are indicated with black arrows. (B) GST-ZF4-d5 purified by anion exchange chromatography. A subset of the fractions collected are run on SDS-PAGE and stained with coomassie. The arrows indicate the contaminant proteins.

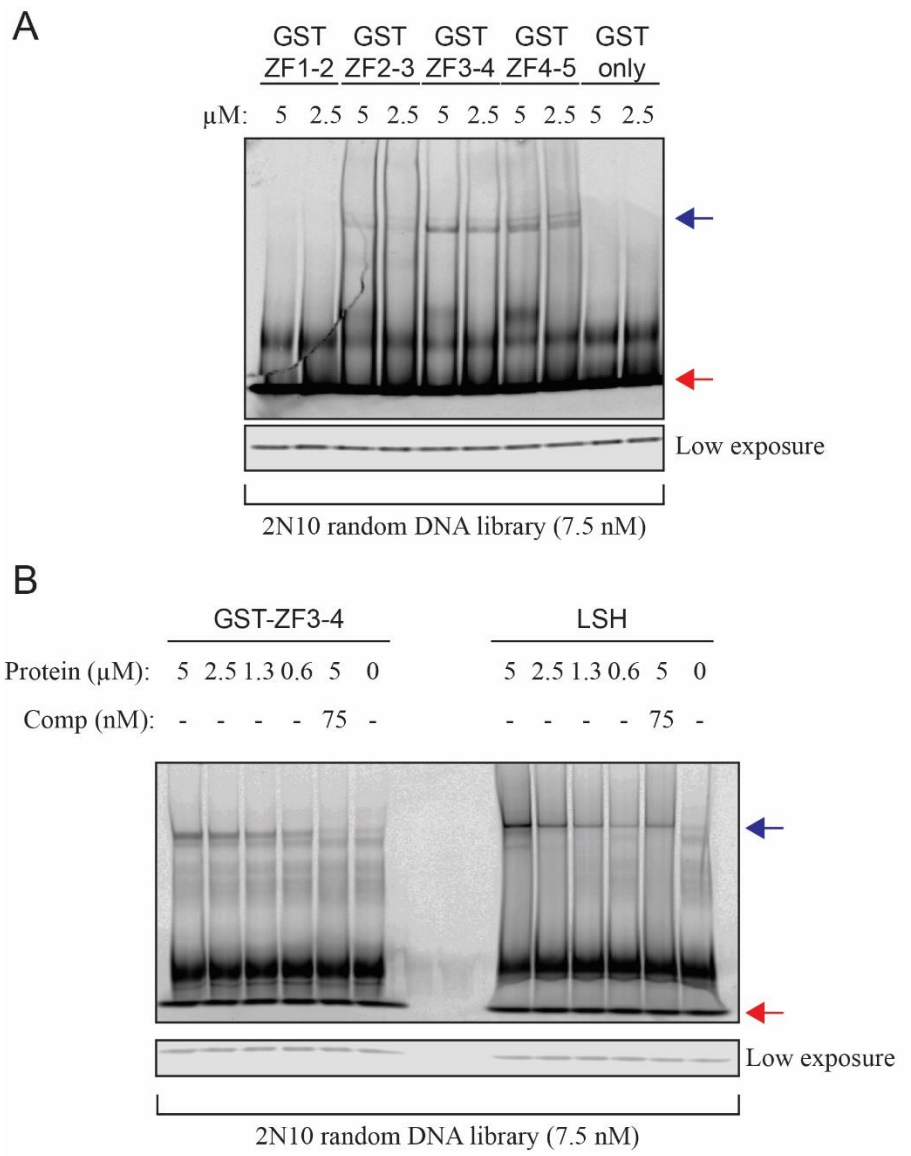


Figure 5. 4 WIZ ZF 3 and ZF 4 can bind DNA *in vitro*

(A) *In vitro* DNA binding assay of all GST-tagged WIZ ZF pairs. 5 or 2.5 μM recombinant protein is mixed with 7.5 nM of IR700 fluorescently labelled 2N10 DNA library. GST is used as a negative control. After 30 min of binding, the samples are loaded to a 5% non-denaturing PAGE. The gel is scanned on Licor Odyssey Scanner. Low exposure of the unbound DNA is shown at the bottom. The protein-DNA complexes and unbound DNA are indicated by blue and red arrows, respectively. **(B)** *In vitro* DNA binding assay of GST-ZF3-4 and chromatin remodeller LSH. Recombinant LSH protein is used as a positive control for DNA binding. Indicated amount of labelled DNA is used as a competitor to the labelled one.

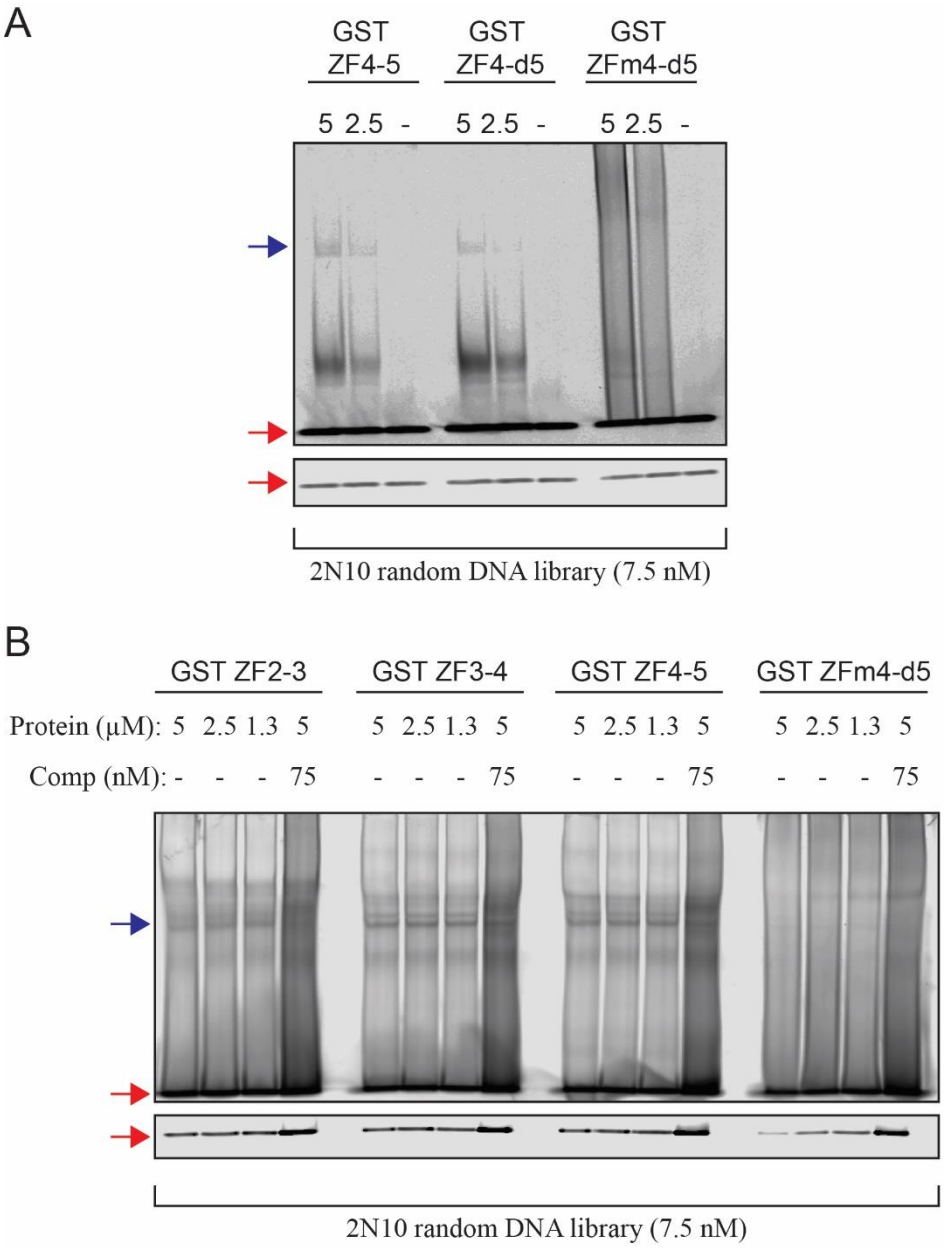


Figure 5. 5 A point mutation in ZF4 abolishes DNA binding

(A) *In vitro* DNA binding assay with GST-ZF4-5 and its mutant forms. 5 or 2.5 µM recombinant protein is mixed with 7.5 nM labelled DNA. The band shifts are visualized by Licor Odyssey Scanner. The specific protein:DNA complexes and unbound DNA are indicated by blue and red arrows, respectively. Low exposure of the unbound DNA is also shown. **(B)** The DNA-binding WIZ ZF pairs and the mutant WIZ ZFm4-d5 assessed by EMSA. All wild type WIZ ZF pairs shift DNA except ZFm4-d5. The unlabelled 2N10 DNA library was used as a competitor.

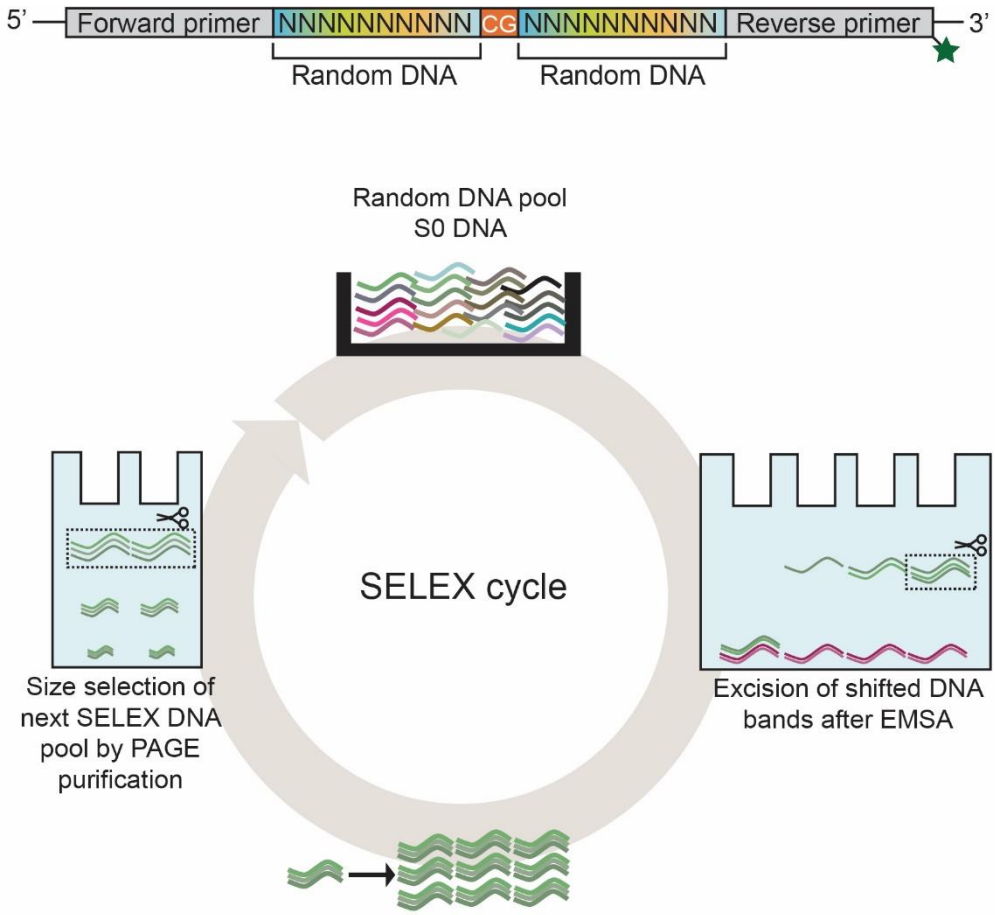


Figure 5. 6 Schematic representation of systematic evolution of ligands by exponential enrichment (SELEX) experiment

A schematic representation of the workflow of a SELEX experiment. A random DNA library (top) is used as initial DNA pool (2N10 DNA library). The SELEX experiment involves 3 major steps per cycle. First, an *in vitro* DNA binding is conducted by incubating the random-sequence DNA library and the protein of interest. DNA-protein complexes are detected by EMSA experiments. Shifted DNA bands containing DNA-protein complexes are excised from the gel and DNA eluted. In the next step, the eluted DNA is PCR amplified. In the final step, amplified DNA is run on a 12% denaturing PAGE. PCR products with the correct size are excised, purified and used as a DNA pool for the next SELEX cycle. The cycles are repeated until the intensity of the shifted band does not increase in subsequent SELEX cycles.

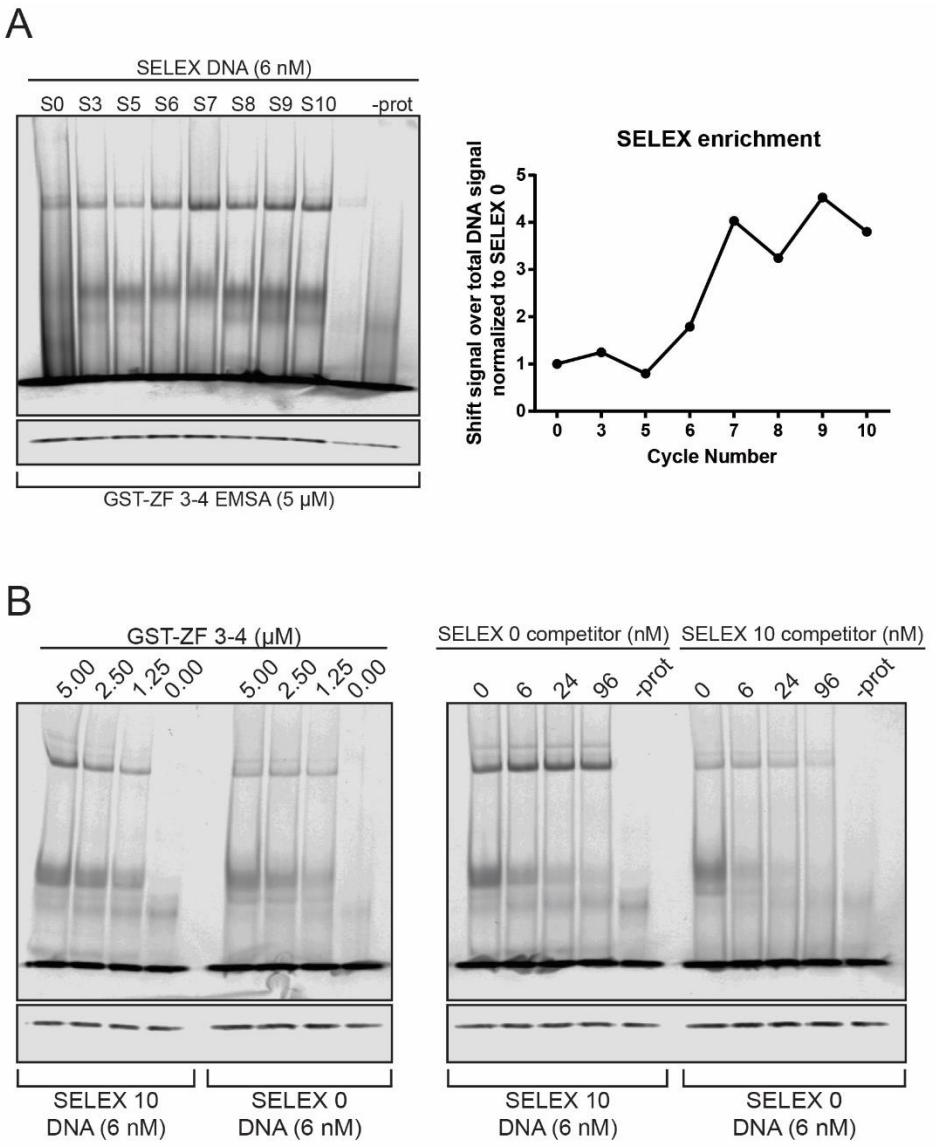


Figure 5. 7 Sequences bound by GST-ZF 3-4 with high affinity are fully enriched in SELEX 10 DNA

(A) EMSA GST-ZF3-4 and SELEX DNA pools generated during the SELEX cycles (left). For each SELEX DNA, the fluorescence signal from the band shift is divided by the total signal (band shift + unbound). This value is normalized to the value obtanied from S0 DNA to quantifiy the SELEX enrichment for each SELEX DNA pool. After SELEX cycle 7, the SELEX enrichment does not increase consistently indicating that no further SELEX enrichment is achieved. **(B)** EMSA affinity of GST-ZF3-4 to SELEX 10 and SELEX 0 DNA (left). Competition experiment is performed by adding unlabelled SELEX DNA at indicated amounts to the labeled DNA in the *in vitro* DNA binding reaction mix (right).

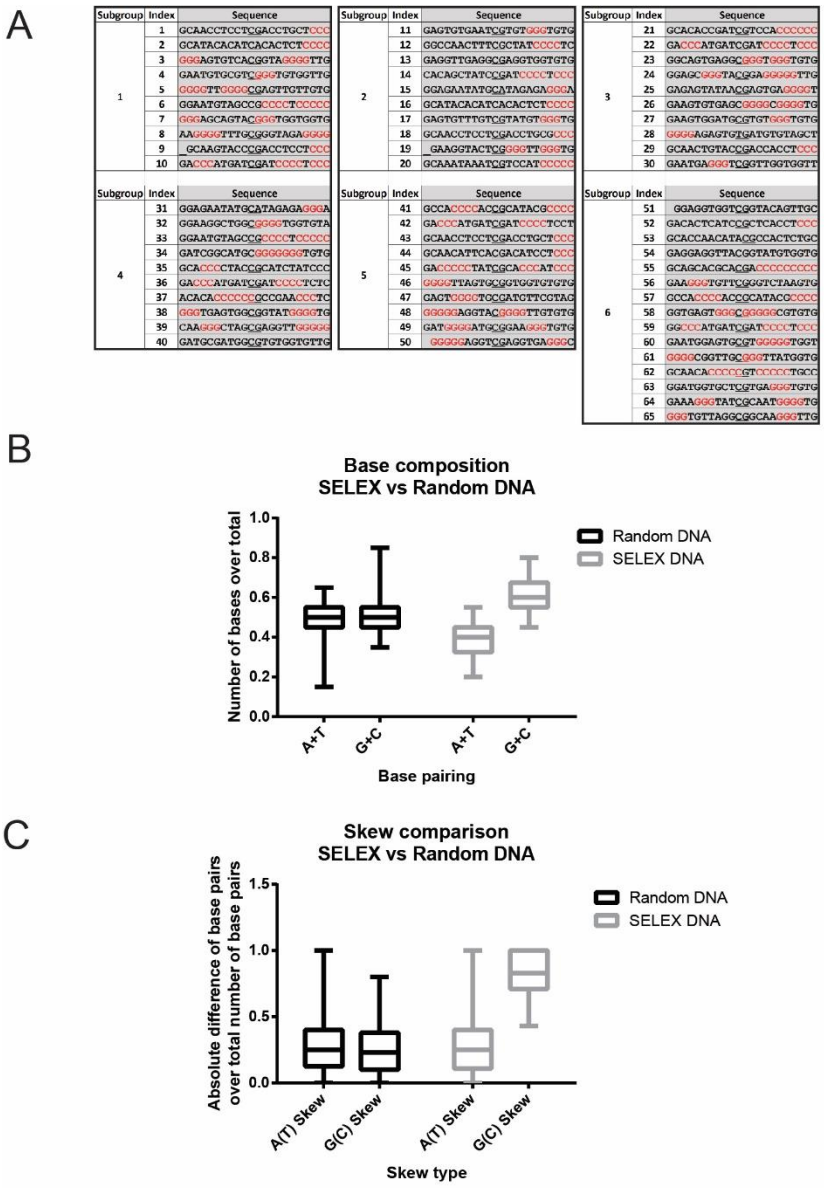


Figure 5. 8 WIZ ZF3-4 tends to bind G-rich DNA sequences

(A) A list of individual unique sequences obtained from the SELEX 10 DNA pool. The back to back G or C bases are shown in red if the string is longer than 3 bases. All unique sequences were randomly divided into 6 subgroups as indicated **(B)** The base composition of the sequenced SELEX 10 DNA compared to 65 random-sequences from the SELEX 0 pool carried out by the following online tool:
(http://www.bioinformatics.org/sms2/random_dna.html)
(C) Skew analysis of sequenced S10 and S0 DNA. To calculate skew, the absolute value of base number difference was divided by the total number of base pairs.
(for G(C) skew: [|(# of G) - (# of C)|] / [(# of G) + (# of C)] was calculated.)

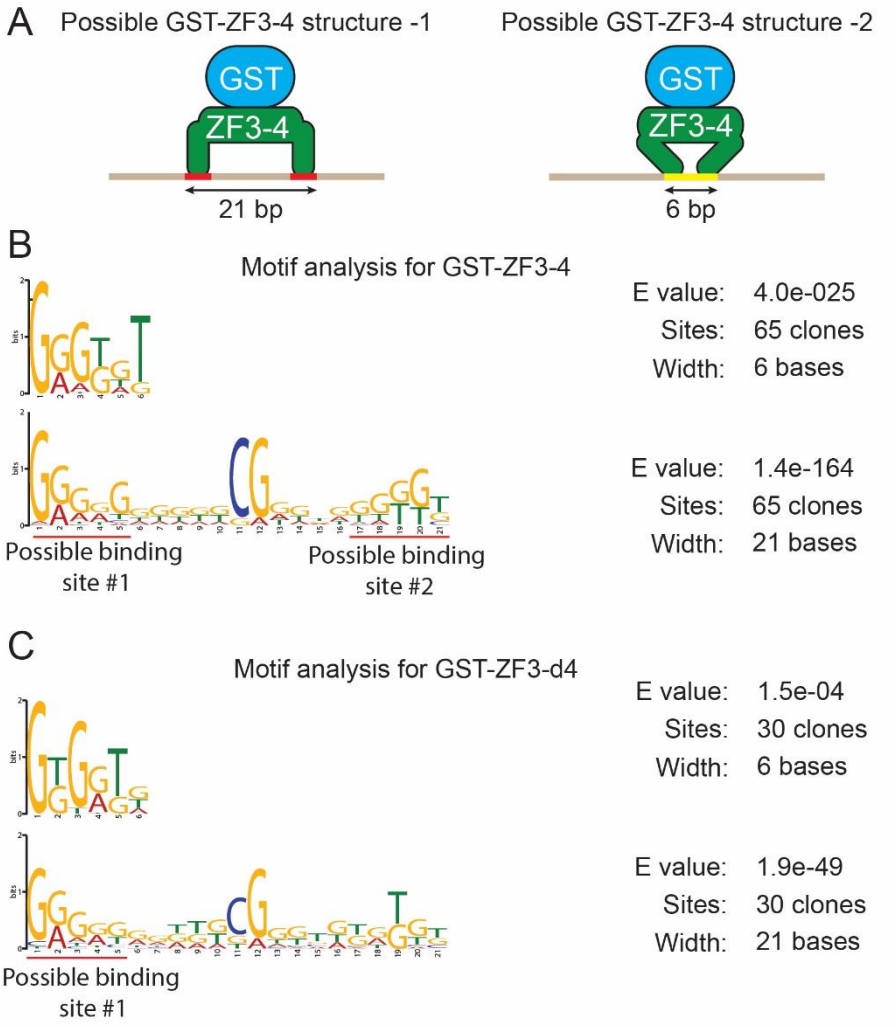


Figure 5. 9 Motif analysis for GST-ZF3-4 and GST-ZF3-d4 SELEX [A DNA fragments

(A) Two possible 3D-structures of GST-ZF 3-4. On the left, zinc fingers are widely seperated and there are one binding site (3-4 base pairs) per zinc finger on a target DNA molecule. On the right, the zinc fingers are close enough to contact with DNA at a single binding site. **(B)** The DNA binding motifs for GST-ZF3-4 predicted by the Multiple Em for Motif Elicitation (MEME) software. Occurence of only one motif with the indicated witdh size is looked for per sequence. E-value indicates the likelyhood of observing the identified motif randomly. Sites indicate the number of occurence of the motif in the input DNA. The scale (bits) are a unit of measure with a precise thermodynamic relationship to energy. **(B)** DNA binding motifs for GST-ZF3-d4 predicted by MEME in the sequenced SELEX DNA pool containing 30 unique sequences.

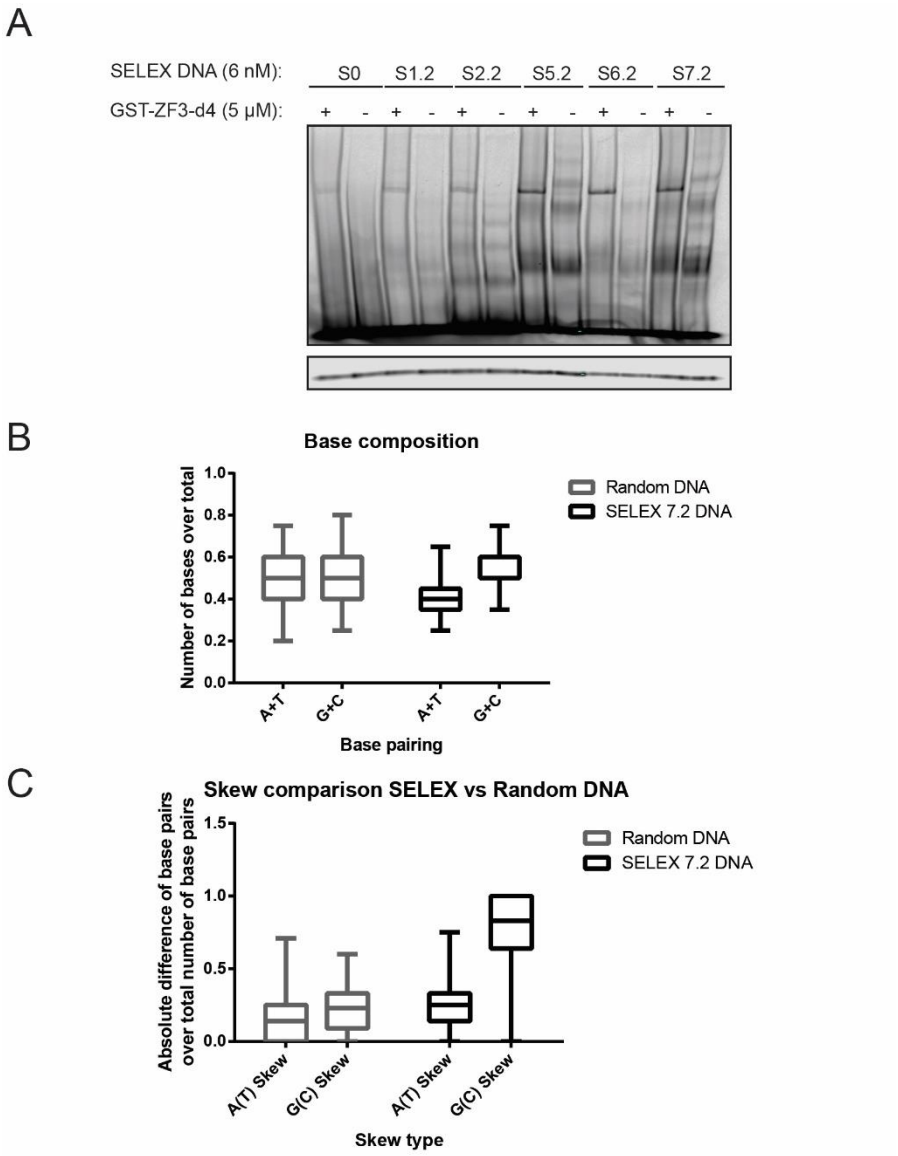


Figure 5. 10 Sequences bound by GST-ZF 3-d4 with high affinity are fully enriched in SELEX 7.2 DNA

(A) EMSA results of GST-ZF3-d4 and SELEX DNA pools generated during the SELEX cycles. The SELEX enrichment for GST-ZF3-d4 reached ~~at~~ a maximum after SELEX cycle 6. **(B)** The base compisition of sequenced S7.2 DNA. As control, 30 random-sequences created by the following online tool:
(http://www.bioinformatics.org/sms2/random_dna.html).
As expected this tool generates equal number of A/T and G/C base pairs. **(C)** Skew analysis of sequenced S7.2 DNA. To calculate skew, the absolute value of base number difference is divided by the total number of base pairs.
(for G(C) skew: [|(# of G) - (# of C)|] / [(# of G) + (# of C)] is calculated.)

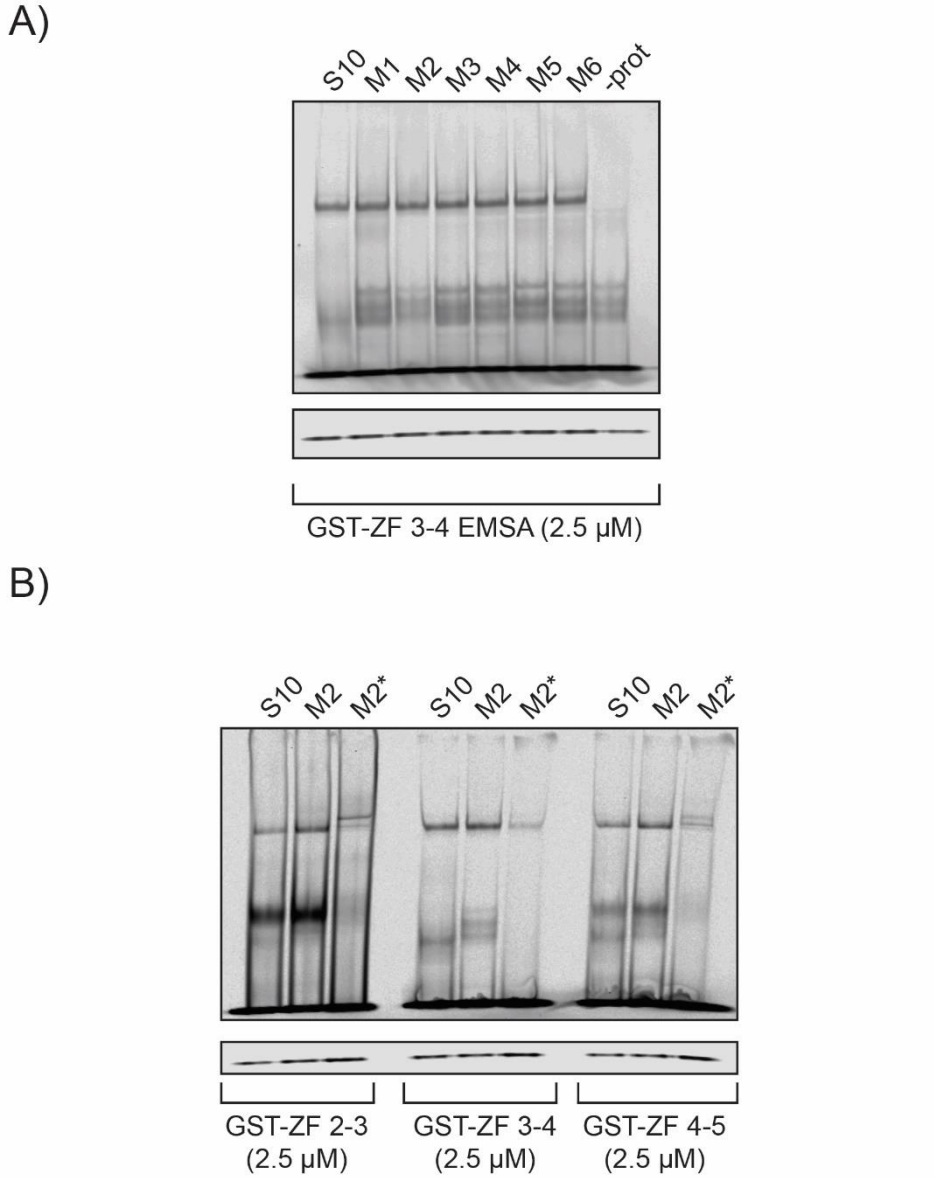


Figure 5. 11 All DNA-binding WIZ ZFs show high affinity to subgroups of sequenced S10 DNA

(A) EMSA with GST-ZF3-4 and the subgroups of sequenced S10 DNA. The mixed templates are amplified by PCR using labelled primer pairs, which created labelled mixed-subgroups M1 to M6. 6 nM DNA is used in binding reaction. No protein control sample containes the mix of all DNA used in the experiment. **(B)** All DNA-binding WIZ ZF pairs bind to indicated labelled DNA. The binding efficiency is significantly reduced when indicidual sequences in a subgroup are amplified separately and mixed later to form mixed-subgroups (indicated with `*`). 6 nM DNA is used in each binding reaction..

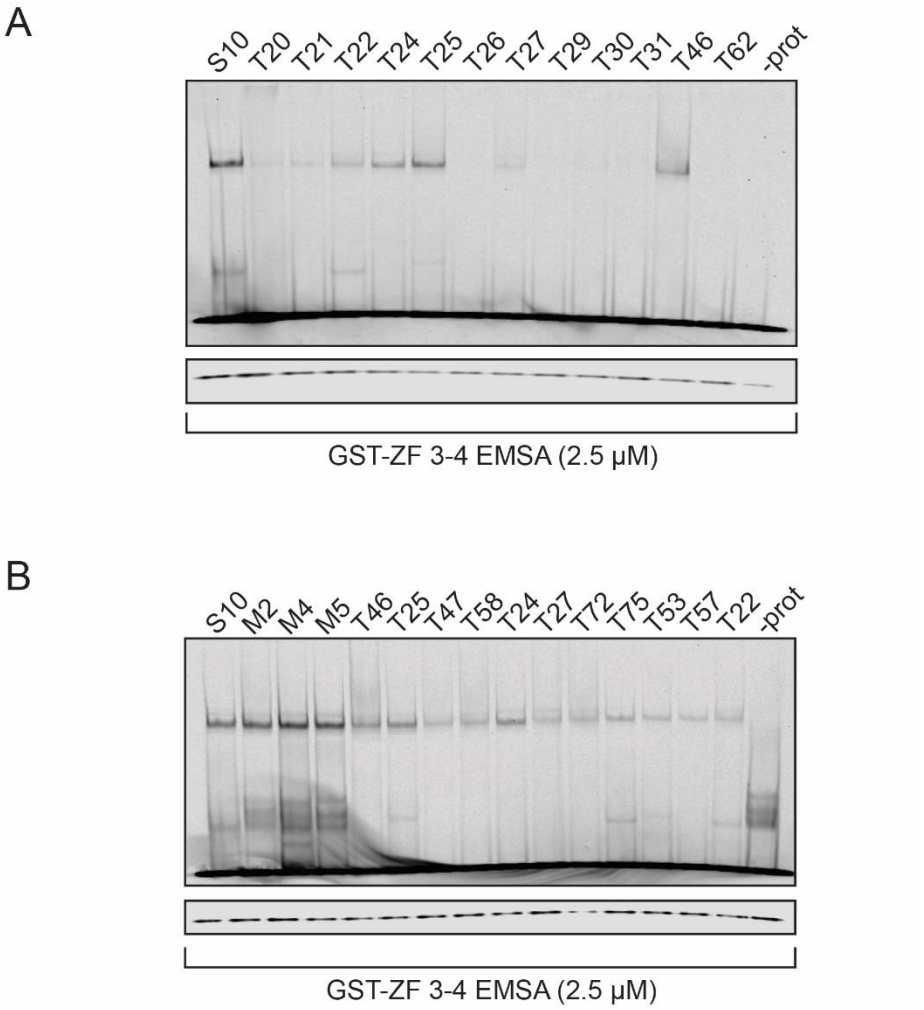


Figure 5. 12 Identification of DNA bound by GST-ZF 3-4 with high affinity

(A) An EMSA experiment using 6 nM M2 subgroup DNA and GST-ZF 3-4 protein. Most sequences show very low binding affinity for GST-ZF3-4, only some had comparable affinity to S10 DNA. **(B)** DNA sequences from M2, M3 and M4 subgroups are tested in EMSA experiments with GST-ZF3-4. Lane 2, 3 and 4 contained DNA that is pre-mixed and PCR amplified. The strongest bandshifts are detected with T24 and T25 DNA sequences. No protein control sample containes the mix of all DNA used in the experiment and no GST-ZF3-4.

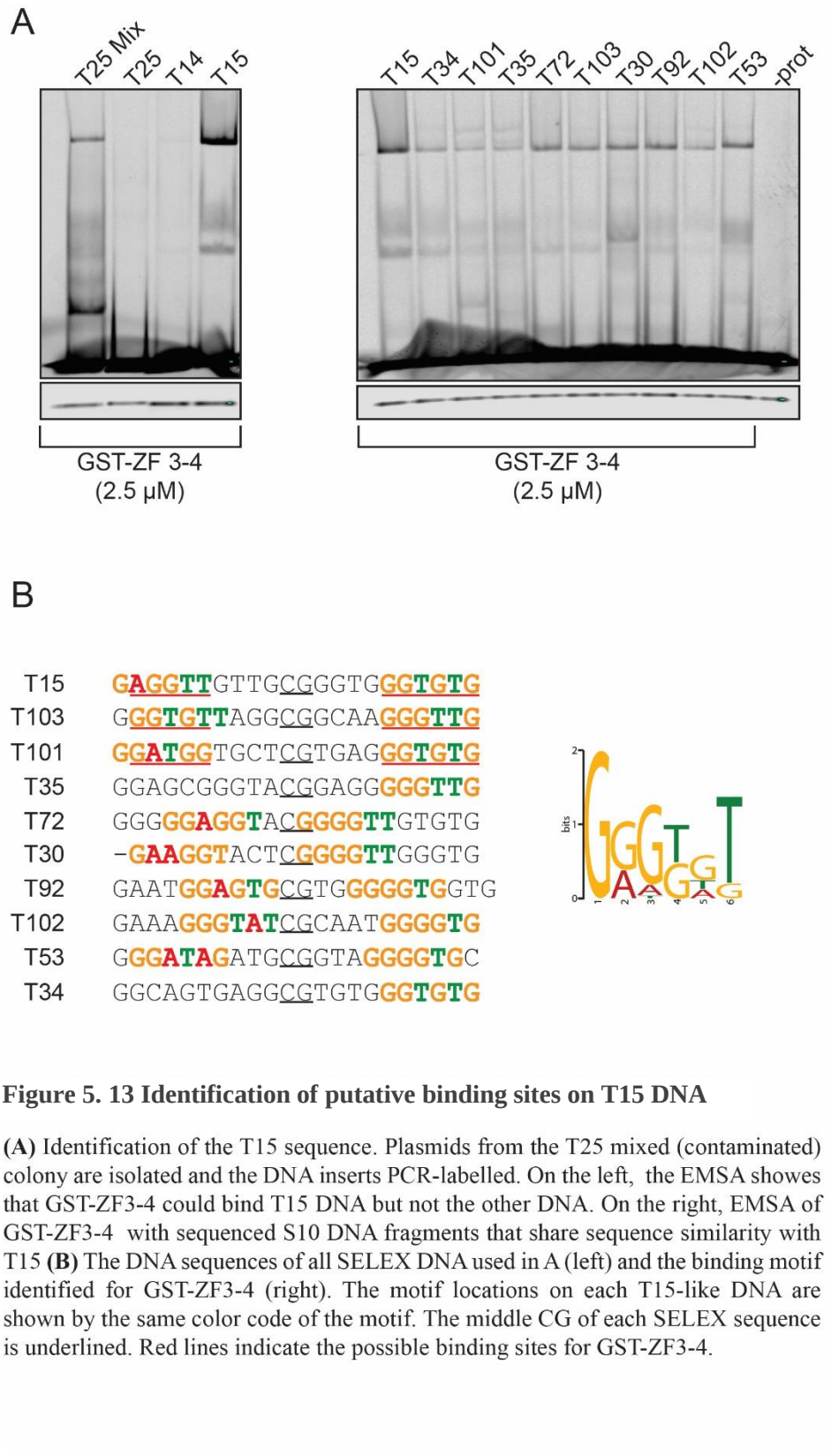


Figure 5. 13 Identification of putative binding sites on T15 DNA

(A) Identification of the T15 sequence. Plasmids from the T25 mixed (contaminated) colony are isolated and the DNA inserts PCR-labelled. On the left, the EMSA showes that GST-ZF3-4 could bind T15 DNA but not the other DNA. On the right, EMSA of GST-ZF3-4 with sequenced S10 DNA fragments that share sequence similarity with T15 **(B)** The DNA sequences of all SELEX DNA used in A (left) and the binding motif identified for GST-ZF3-4 (right). The motif locations on each T15-like DNA are shown by the same color code of the motif. The middle CG of each SELEX sequence is underlined. Red lines indicate the possible binding sites for GST-ZF3-4.

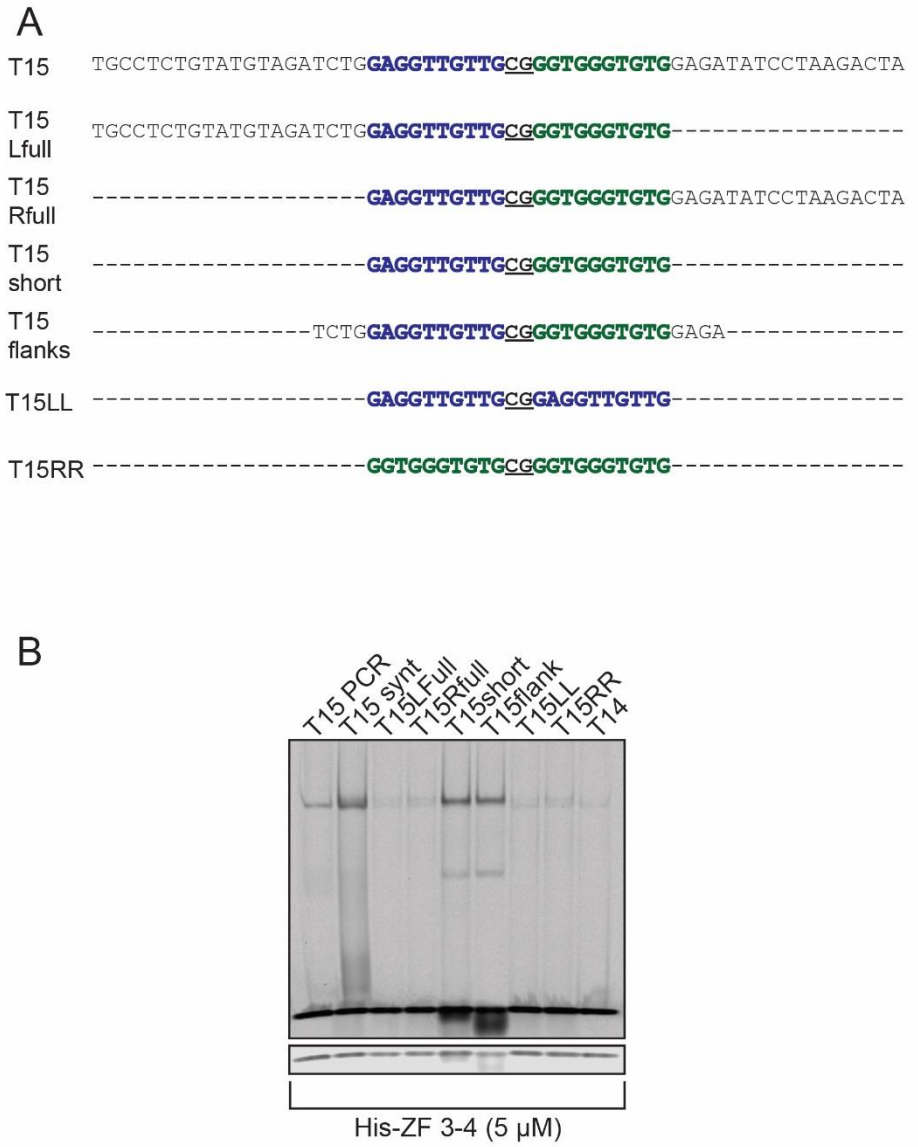


Figure 5. 14 Narrowing down the possible binding sites for ZF3-4 on T15 DNA

(A) The tested T15-variant DNA sequences. The CG di-nucleotide in the middle is underlined. The sequences on the left (L) and right (R) hand side of the CG are shown in blue and green, respectively. Single-stranded sense and antisense strand of T15 variants are annealed to generate double-stranded DNA. Either the sense or the antisense oligo is labelled with IR-800. The size of the double-stranded labelled DNA is confirmed by Licor Odyssey Scanner prior to EMSA experiments. **(B)** EMSA of His-ZF3-4 and double-stranded IR-800 labelled T15 DNA variants. PCR amplified T15 and T14 DNA are used as positive and negative controls, respectively. The shortest variant, T15short DNA, is sufficient for efficient ZF3-4 binding.

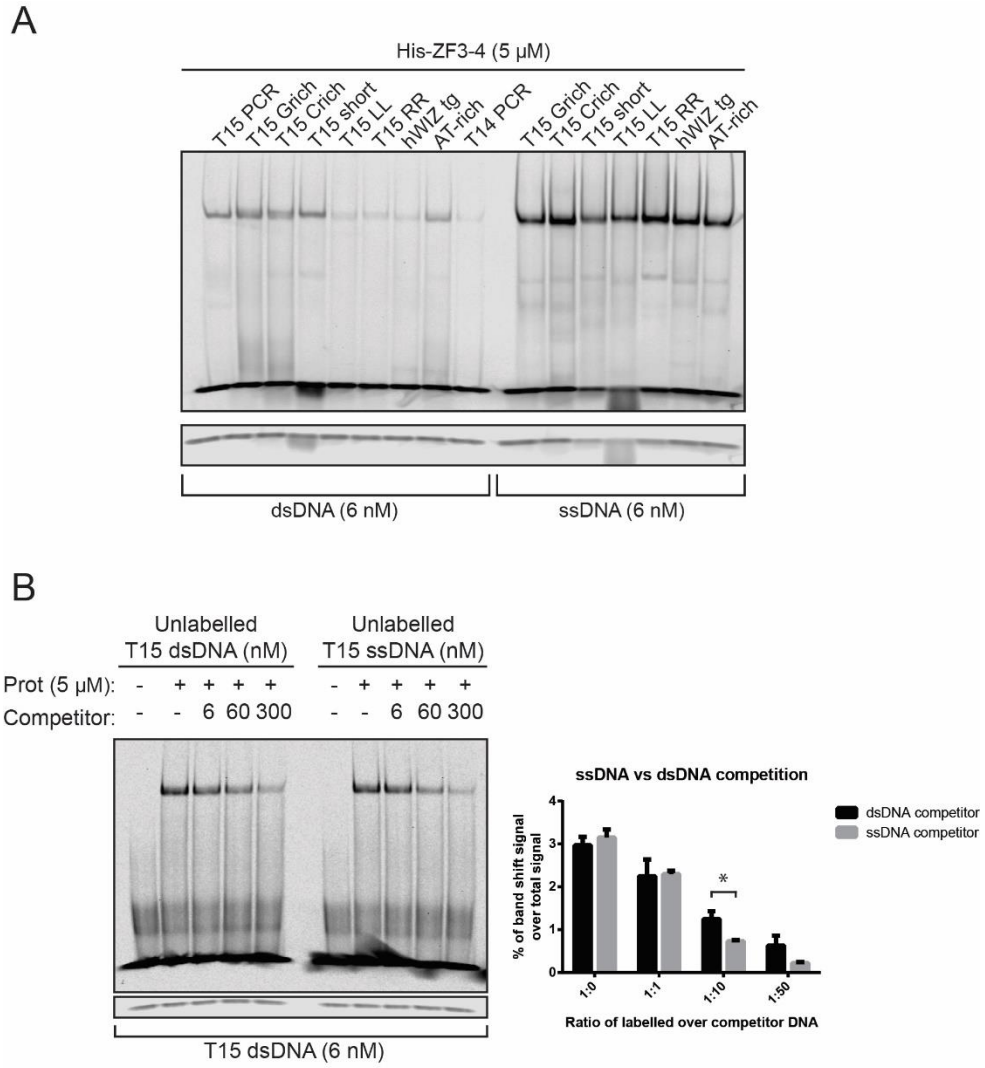


Figure 5. 15 His-ZF 3-4 displays a higher affinity for single-stranded DNA than dsDNA *in vitro*

(A) An EMSA experiment with His-ZF3-4 and single-stranded T15 variants. Only the IR-800 labelled strand of T15 variants is used in the band shift experiments. His-ZF3-4 binds to all T15 ssDNA variants with higher affinity comparing to T15 dsDNA. **(B)** On the left, unlabelled T15 either dsDNA or ssDNA are used as a competitors against labelled T15 dsDNA. On the right, quantification of the average signal obtained from band shifts over the total (shifted + unbound DNA) signal from three independent EMSA experiments. The error bars represent standard deviation from the mean. The error bars represent the standard deviation from the mean of three technical replicates. "*" indicates $p < 0.05$. p values determined by t test ($n \geq 3$).

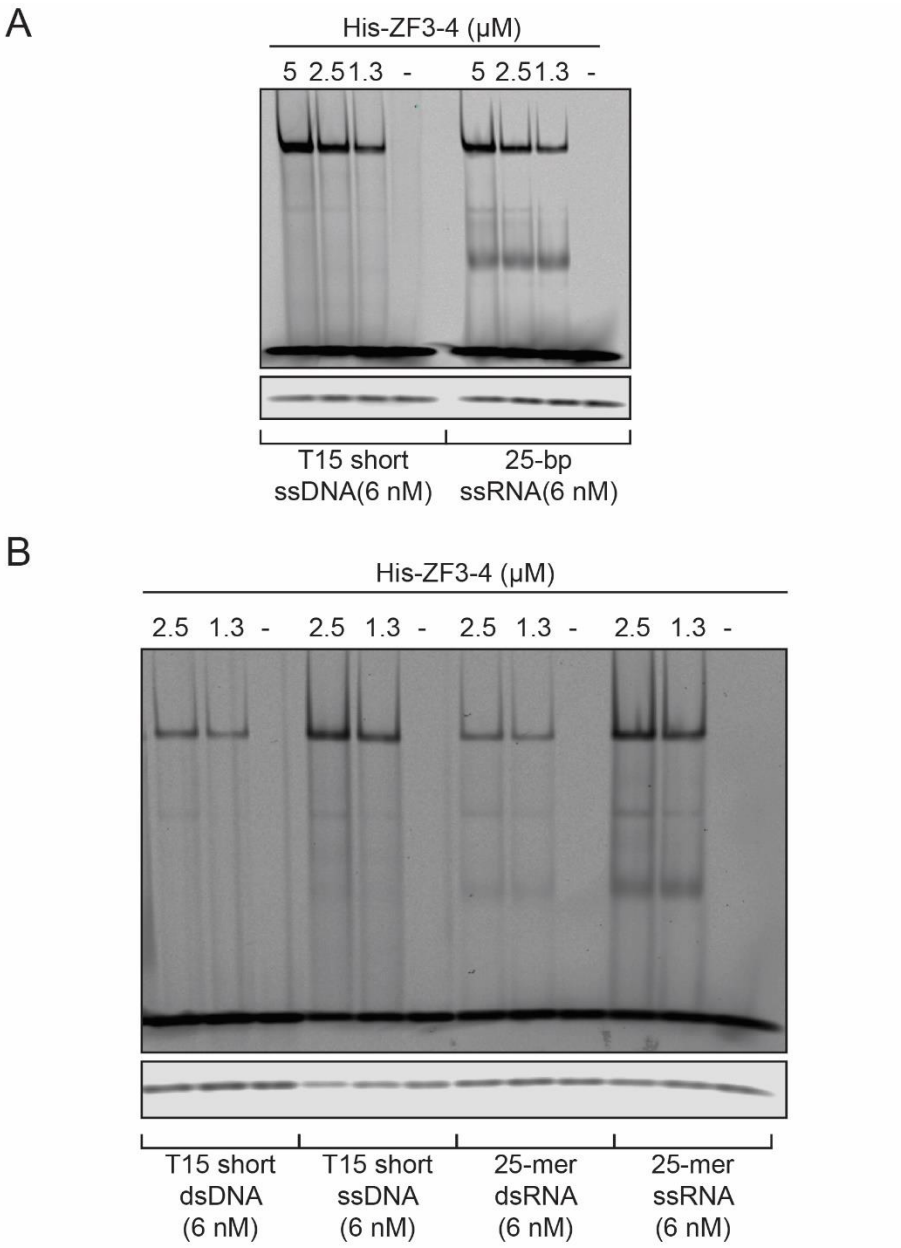


Figure 5. 16 His-ZF 3-4 binds single- and double-stranded RNA in vitro

(A) EMSA of His-ZF3-4 with IR-800 labelled, 22-bp T15 ssDNA and IR-700 labelled, 25-bp single-stranded RNA. His-ZF3-4 can bind to both ssRNA and ssDNA *in vitro*. **(B)** EMSA of His-ZF3-4 with single- or double-stranded DNA and RNA. Double-stranded RNA is generated by annealing the complementary strands of 25-mer IR-700 labelled RNA.

5.7. Discussion

The zinc finger domains of WIZ are widely separated by at least 50 amino acids. This is rather unusual compared to other C2H2-type zinc finger proteins, most of which contain only several linker amino acids between tandemly arrayed zinc finger domains (Schuh et al. 1986). A number of zinc finger proteins, such as Su(Var) 3-7 in *Drosophila* and ZNF292 in mammals, with widely-separated zinc finger domains have been shown to interact with DNA (Cléard & Spierer 2001; Lipkin et al. 1993). Similar to these zinc fingers, my data clearly showed that WIZ can bind DNA *in vitro*. The strategy I followed was to create zinc finger pairs of WIZ. This was essential for the identification of DNA binding fingers of WIZ individually.

The 2N10 library contained a CG di-nucleotide in the middle of a 20 bp-long random DNA sequence. I chose to use this library, in case WIZ was a methylated-DNA binding zinc finger. I did not find any direct evidence for methylated-DNA binding of WIZ. On the contrary, the putative binding motif for WIZ ZF3-4 appeared far from the middle CG, and it did not contain any CG. My conclusion here was that WIZ ZF pairs have no detectable affinity towards unmethylated CG di-nucleotides. Therefore, it is unlikely that WIZ could bind to CpG islands, which are rich in unmethylated CG-dinucleotide.

By SELEX experiments, I showed that WIZ ZF3 and ZF4 prefers G- or C-rich sequences and identified a sequence (T15short) bound by GST-ZF3-4 with high affinity. My data indicated that for efficient DNA binding His-ZF3-4 requires two binding sites that are a certain distance apart. Finally, I showed that WIZ zinc finger pairs can bind ssDNA and RNA without sequence specificity. Repetitive regions of genome usually contain G-rich

DNA. These include; (1) telomeric repeats with contain runs of Gs (Parkinson et al. 2002), (2) non-template DNA strand of rDNA repeats (Hanakahi et al. 1999), (3) the immunoglobulin heavy chain switch (S) regions (Arakawa et al. 1993) and human repetitive minisatellites (Weitzmann et al. 1997). The G-rich repeats from a four-stranded structure that is stabilized by 4 guanines, which is called G-quadruplex or G4 DNA (Gellert et al. 1962). The G-rich DNA preference of WIZ zinc finger pairs indicates that WIZ might have an affinity to these repetitive regions. I have not performed any experiment that would directly show WIZ binding to G4 DNA. However, it is known that G-rich DNA sequences might readily form G-quadruplex *in vitro* (Gellert et al. 1962). Therefore, it is likely that G4 DNA was present in S10 DNA. Nowadays, G4 DNA can be detected in both mouse and human cells by specific antibodies (Henderson et al. 2014). It would be possible to test whether or not WIZ can localize at the G4 DNA sites by ChIP experiments.

I identified two possible binding motifs for GST-ZF3-4 by SELEX sequencing but these motives are not `clear-cut` motifs. There are three reasons for this outcome; (1) SELEX cycles do not eliminate the DNA bound by ZF pairs with low affinity (Figure 5.12). The sequences of these DNA were included to the sequence set in the MEME analysis, which lowered the prediction power of the tool. (2) Although my data indicated that there must be a certain distance between two zinc fingers for efficient DNA binding, I did not determine the exact distance between two binding sites. Therefore, width of the actual binding motif might be different from the ones I used in my analysis (Figure 5.9). (3) I used only 65 and 30 unique sequences in the MEME analysis for GST-ZF3-4 and GST-

ZF3-d4, respectively. These numbers might not be adequate for MEME analysis to provide a more `clear-cut` motif. The most recent WIZ study mapped the localization of WIZ in the genome of mouse brain cells (Isbel et al. 2016). In this study, six different WIZ binding motifs were identified by MEME from ~40000 WIZ binding sites. The most common motif that was present in 70% of all WIZ binding site was also not a `clear-cut` motif. Nonetheless, all the motifs identified in this study and the SELEX sequencing were G-rich. This might suggest that DNA binding of WIZ is influenced more by the base composition rather than the actual DNA sequence target site. To elaborate more on this hypothesis, the crystal structure of WIZ has to be solved. It would also make sense to purify the full-length WIZ from insect cells, and test its DNA binding with the sequences that I identified from the SELEX experiments. Nonetheless, it is clear that WIZ prefers G-rich sequences *in vivo* and *in vitro.*

It is also important to note that the WIZ ChIP-seq experiment performed on human embryonic kidney cells suggested a different binding motif for WIZ (Bian et al. 2015). This motif was discovered by analyzing only a small fraction (458 out of 11853 binding sites) of WIZ peaks, which might introduce a bias. Additionally, the motif did not display a G-skew. When I used the human WIZ-target DNA in my EMSA experiments, I observed a weak DNA band shift (Figure 5.15A). The amino acid sequence of WIZ ZF3 and ZF4 in human and mouse are identical. Therefore, it is likely that the full-length human WIZ protein binds to hWIZ-target DNA by the first 5 zinc fingers that are absent in the short WIZ isoform.

In my experiments, affinity of WIZ to ssDNA was much higher than its affinity to dsDNA. I was not able to observe a complete shift of labelled dsDNA even though I used a 1000-fold excess of recombinant WIZ ZFs (5 µM protein compared to 6 nM labelled DNA). Therefore, I cannot determine the kD of WIZ-DNA interaction from the EMSA experiments. Nonetheless, in my EMSA experiments, 5 µM WIZ ZF3-4 was able to shift at most 3% of 6 nM labelled dsDNA (Figure 5.15A). The same amount of WIZ ZF3-4 shifted about 30% of all labelled ssDNA (Figure 5.16A). Fluorescence polarization (FP) could be used to measure the kD of WIZ for single- or double-stranded DNA. This technique requires a pure protein and a fluorescently-labelled DNA, which contains Alexa and Oregon Green dye instead of the IR-800 that I used in my experiments (Rusinova et al. 2002). Nonetheless, it can be concluded here that WIZ has a higher affinity for single-stranded DNA.

As mentioned earlier, it has been shown that G9a is recruited to R-loops, which occur at the G-rich terminator elements of genes but not the R-loops at CpG islands (Skourti-Stathaki et al. 2014). In the previous chapter, I tested whether or not the recruitment of G9a to the R-loop of Actin gene was dependent on WIZ, and could not detect a clear indication that showed WIZ mediated the recruitment of G9a (Figure 4.8). However, I have shown the single-stranded DNA binding of WIZ ZF3-4 and it is possible that WIZ might be important in the recruitment of G9a to at least a fraction of R-loops in mouse genome. Since WIZ ZF3-4 can bind single- and double-stranded RNA, it might bind a DNA-RNA hybrid as well. This can easily be tested *in vitro*, and such an interaction would reinforce the idea that WIZ is important in the recruitment of G9a to R-loops. In addition

to this, according to ChIP-seq data of Isbel et al., 31% and 13% of WIZ peaks mapped to the active and inactive promoter sites, respectively. Similar to transcription termination sites, R-loops are formed at the transcription start (Santos-Pereira & Aguilera 2015). Therefore, WIZ presence at the active promoters might be related to its ability to bind single-stranded DNA.

Chapter VI – Discussion

Histone methylation plays an essential role in the formation of chromatin domains and the regulation of gene expression (Jenuwein 2006). Histone lysine methylation states at H3K9 are associated with gene silencing in eukaryotes by formation of heterochromatin (Grewal & Jia 2007), maintenance of genome stability by silencing repetitive elements by H3K9me2 and creating a stringent high-order structure at pericentric heterochromatin by H3K9me3 (Peters et al. 2001; Peng & Karpen 2009), and X chromosome inactivation in female mammals (Rougeulle et al. 2004). A number of histone methylases are responsible for the catalysis of the individual lysine methylation states at H3K9 due to their particular SET-domain structure, which provides the substrate specificity (Pinheiro et al. 2012; Tachibana et al. 2001; Ait-Si-Ali et al. 2004; Rathert et al. 2008b; Zhang et al. 2003). The importance of G9a and GLP is not limited to gene silencing by the catalysis of H3K9me2 in euchromatic regions of the genome because G9a/GLP can interact with DNMTs to direct de novo DNA methylation during ES cell differentiation (Epsztejn-Litman et al. 2008). In parallel, G9a/GLP plays an essential role in the maintenance of imprinted DNA methylation by recruiting DNMTs to imprinting control regions in ES cells (Zhang et al. 2016). Moreover, G9a/GLP can methylate non-histone proteins, such as CDYL1, WIZ, ACINUS and itself (Rathert et al. 2008b). Despite the functional characterization of G9a/GLP, little is known about the factors involved in the recruitment mechanisms of G9a/GLP to chromatin. Structural analysis on G9a/GLP revealed that the ANK repeat domain binds H3K9me1/2 *in vitro* (Collins et al. 2008). This suggests that the product of G9a/GLP might serve as an `anchoring point` for the G9a complex to maintain the

H3K9me1/2 mark. However, the molecular basis of the de novo establishment of the silencing H3K9me2 mark by G9a/GLP is still poorly understood. The binding partners of G9a/GLP may mediate the chromatin binding of G9a/GLP by recognition of specific DNA or RNA sequences. The zinc finger protein WIZ interacts with G9a/GLP in stoichiometric amounts and is a core component of the G9a complex (Ueda et al. 2006). The aim of this study was to identify the factors that control the chromatin binding of G9a and characterise DNA-binding properties of WIZ to understand its possible role in the recruitment of the G9a complex to specific loci.

6.1. Binding partners of G9a in ES cells

To identify the binding partners of G9a, I used ES cells expressing FLAG-tagged G9a to co-purify G9a and its binding partners. Using mass spectrometry, I identified a number of nuclear proteins, which were candidate binding partners of G9a. My data correlated well with another study that used a similar approach (Meier et al. 2015) and contained already-known binding partners of G9a, such as DNMT3A and DNMT3B (Esteve et al. 2006, Epsztejn-Litman et al. 2008). The list of identified candidate binding partners included four zinc finger proteins – WIZ, ZNF518B, ZF462 and ZC3H11A. These proteins are potential DNA-binding zinc fingers that could recruit the G9a complex in a sequence-specific manner. Identification of the DNA binding preferences of these zinc finger proteins might reveal possible ways of de novo recruitment of the G9a complex.

Focusing on the function of WIZ in the G9a complex, I knocked down WIZ and found that the WIZ knockdown in ES cells led to reduced levels of G9a and GLP. The depletion of WIZ in HeLa cells leads to a more substantial decrease in the G9a but not GLP levels (Ueda et al. 2006). It is clear that G9a and GLP can form homo- or heterodimers, but they require accessory proteins for enhanced protein stability (Tachibana et al. 2005). Then, the question arises, why does WIZ deficiency differentially affect the stability of the G9a complex in different cell lines? A possible explanation is that G9a and GLP are ubiquitously expressed while most binding partners of the G9a complex are not. Therefore, it is very likely that, in various cell types, the stability of G9a/GLP might depend on different binding partners. Supporting evidence for this idea is that G9a was shown to interact with various proteins, or the same protein with different affinity depending on the cellular state or external stimuli in macrophages (Liu et al. 2014). The same study also identified G9a in a number of protein complexes including NuRD, CoREST, SWI/SNF and CtBP complex. It was shown that WIZ interacts with CtBP1 and CtBP2 (Ueda et al. 2006), and G9a can interact with CoREST and NRSF/REST in HeLa and HEK cells, respectively (Roopra et al. 2004). Moreover, G9a on euchromatin can display a transcription activator role, which is independent of its enzymatic function (Lee et al. 2006). Recently, the transcriptional regulatory network mediated by NKX3.1-G9a-UTY was shown to regulate prostate differentiation *in vivo* (Dutta et al. 2016). This study indicated that G9a acts as a coregulator of NKX3.1, which is associated with both activation and repression of NKX3.1 target genes. These interactions link G9a to both activating and repressive complexes but whether or not G9a/GLP directly interact with these complexes is still elusive. Identification of the proteins that act as a bridge between

G9a/GLP and the other repressive or activator protein complexes is crucial to identify the role of individual proteins in the G9a interactome. Since WIZ is one of the known core components of the G9a complex, it would be informative to identify proteins that interact with WIZ directly by mass spectrometry. This approach would shed light on the function of WIZ within and out with the G9a complex. Additionally, a catalytic activity-based purification (as described by (Liu et al. 2014)) of the G9a complex followed by mass spectrometry in different stages of ES cell differentiation would be very informative for identification of additional roles of the G9a complex in cell fate decisions.

6.2. Possible role of WIZ and the ankyrin repeat domain of G9a/GLP in the spreading of H3K9me2

Spreading of heterochromatin along chromatin fibers is an important concept, which is the primary cause of position-effect variegation. Three main mechanisms were proposed for the spreading of heterochromatin; (1) 'looping' model is related to the structural change of chromatin fiber which enables the interaction of distant regions, (2) 'oozing' model suggests the spreading of a heterochromatic mark is initiated at a "nucleation site" and constant recruitment of chromatin modifiers to the adjacent sites causes the spread of heterochromatin and (3) 'sliding' model suggests a transcription- or replication-coupled recruitment of chromatin modifiers by RNA or DNA polymerase, respectively, to create heterochromatin over large genomic region (Talbert & Henikoff 2006). Maintenance of the heterochromatic marks is carried out by protein complexes utilising the above mechanisms individually or in combination. H3K9me2 mark is maintained by G9a/GLP,

as the deficiency in either of these enzyme leads to 90% decrease in H3K9me2 (Tachibana et al. 2005). Step-wise salt extraction experiments on WT ES cells in the presence or absence of the G9a inhibitor showed that the ankyrin repeat domain of G9a, which can recognise H3K9me2, is critical for the chromatin binding of G9a. This indicates that G9a might be using an 'oozing' model to spread H3K9me2 on nearby nucleosomes, since inhibiting its catalytic activity reduces the affinity of G9a to chromatin due to the lack of H3K9me2. This hypothetical model would require specific mediator proteins to establish H3K9me2 at the nucleation site, and the ankyrin repeat domains of G9a/GLP to spread the mark along nearby nucleosomes. To clearly show that G9a can spread H3K9me2 by an 'oozing' model, a cell line expressing only mutant G9a/GLP, which cannot bind two H3K9me2 due to a point mutation in the ankyrin repeat domains of both proteins should be generated. The ANKΔ-DM ES cells that I used in my experiments have been a good control to test the effect of the inhibitor treatment as they did not respond to the presence of the G9a inhibitor. However, these cells show a significant reduction in H3K9me2 levels due to destabilisation of the G9a/GLP complex. Thus, to test the above hypothesis, the spreading of H3K9me2 should be investigated in a cell line expression ANKMutant G9a and GLP. If the 'oozing' model is valid for G9a/GLP, it will be possible to identify nucleation sites for G9a by focusing on the presence of H3K9me2, which could not be spread over large chromatin domains. Additionally, the DNA sequence of the nucleation sites identified this way can be used to generate nucleosomes *in vitro* to identify novel binding partners of G9a that can mediate the establishment of H3K9me2 by a SILAC nucleosome affinity purification (SNAP) (Bartke & Kouzarides 2011).

My SWS extraction data showed that the knockdown of WIZ has no significant effect on chromatin binding by G9a. Therefore, it is unlikely that WIZ has a role in the maintenance of already-established H3K9me2. However, it is still possible that WIZ might be essential in the recruitment of G9a to certain nucleation sites during differentiation. To test this possibility, a WIZ-knockout ES cell line can be differentiated into various cell types and global distribution of H3K9me2 can be tested. On the other hand, WIZ might be involved in the spreading of H3K9me2 at certain genomic sites by the other two models. It has been shown that G9a interacts with *Kcnq1ot1* and *Air* noncoding RNAs, whose promoters map to the differentially methylated imprinting control regions (ICRs) (Mancini-Dinardo et al. 2006; Sleutels et al. 2002; Pandey et al. 2008; Nagano et al. 2008). However, the study on *Kcnq1ot1* also showed that the interaction between *Kcnq1ot1* and G9a is specific to placenta. This indicates that if a 'sliding' model existed for G9a, it would require lineage-specific RNA-binding proteins to mediate the interaction. WIZ, as a potential RNA-binding protein, or other zinc finger proteins identified as binding partners of G9a might be substantial in this type of recruitment. Considering altogether, I suggest a hypothetical WIZ-mediated recruitment model for G9a as described in Figure 6.1.

6.3. DNA binding preferences of WIZ

In my studies, I showed that WIZ could bind G-rich double-stranded DNA, as well as any single stranded DNA. G-rich sequences are commonly found in repetitive elements. Similarly, regardless of the organism, H3K9me2/me3 marks are also enriched at the repetitive regions (Black et al. 2012). G9a deficiency causes loss of DNA methylation at

the repetitive elements, indicating the role of G9a in the regulation of epigenetic marks in these regions (Dong et al. 2008). The above evidence clearly shows the presence of G9a at the repetitive elements but what mediates the recruitment of G9a remains unclear. My data suggests that WIZ might play a role in the recruitment of G9a to repetitive elements by recognition of G-rich sequences. Although further evidence is required, it would be interesting to analyse DNA methylation levels at the repetitive regions in ES cells expressing a mutant WIZ that cannot interact with G9a/GLP.

TFIIIA is an example to C_2H_2-type zinc finger-containing proteins that can bind RNA and DNA (Theunissen et al. 1992). An extensive study has been done on the zinc fingers of TFIIIA. Among the total nine zinc fingers of TFIIIA, ZF 1-3, 5 and 7-9 can interact with DNA in the major groove, while ZF 4-6 bind loop structures of 5 S RNA specifically (Searles et al. 2000; Clemens et al. 1993). Zinc finger 5 of TFIIIA was found to be in contact with both DNA and RNA, indicating that certain ZF structures allow binding of RNA and DNA. My data showed that WIZ ZF3-4 could bind both DNA and RNA. However, this has to be tested in cells to exclude the possibility that RNA-binding of WIZ is an artefact of *in vitro* experiments. Cross-linking immunoprecipitation (CLIP) experiments are commonly used to identify protein-RNA interactions in cells (Ule et al. 2003). These experiments use UV-light at a particular wavelength to crosslink proteins to RNA. Such an approach could validate the RNA-binding of WIZ.

My experiments showed that WIZ does not have a preference for unmethylated CG di-nucleotide, as it was specifically under-represented in S10 DNA. However, I did not perform experiments, which would test methylated DNA-binding of WIZ. WIZ might

prefer to bind methylated CG only. *Isbel et al*. mapped WIZ binding sites at both active and silenced promoters in adult mouse cerebellum (Isbel et al. 2016). Binding of WIZ to active promoters could be explained by the putative single-stranded DNA binding of WIZ at R-loops, which occurs naturally near the transcription start site. These R-loops contain GC-rich single stranded- DNA, for which WIZ might display a high affinity. However, localisation of WIZ to silent promoters occur via a different mechanism. One possibility is that WIZ binds methylated CG containing sequences, which requires further experimental evidence. Secondly, the repression by H3K9me2 or DNA methylation at these sites should be maintained by G9a and DNMT1, which is a binding partner of G9a, respectively. Therefore, the localisation of WIZ to silent promoters might be independent of DNA-binding of WIZ.

6.4. Cross-talk between H3K9 and DNA methylation

H3K9 and DNA methylation are highly conserved epigenetic marks associated with gene silencing in vertebrates (Grewal & Jia 2007; Jones 2012). There is a great deal of evidence indicating that these two epigenetic marks are co-regulated (Du et al. 2015). In *Arabidopsis*, DNA methylation is maintained by three different mechanisms, one of which involves chromomethylase 2 and 3 (CMT2 and CMT3, respectively). These enzymes are DNA methylases with a pair of chromodomains that can bind to H3K9me2, which is required for the recruitment of CMT2/3 to heterochromatin. A methylated-DNA binding protein KRYPTONITE recognises the DNA methylation catalysed by CMT2/3, and recruits H3K9me-specific methylases - SUVH5 and SUVH6. This feedforward loop is

necessary for the maintenance of the DNA and H3K9me2 methylation at the transposable elements and pericentric heterochromatin. Similarly, the UHRF1 acts as an adaptor protein which binds to H3K9me3 to mediate the recruitment of DNMT1 to hemi-methylated CpG sites (Nady et al. 2011), and G9a/GLP interacts with DNMTs indicating a cross-talk of H3K9me2 and DNA methylation in mammalian cells (Epsztejn-Litman et al., 2008, Estève et al., 2006). The step-wise salt extractions I performed on DNMT-DKO cells suggest that lack of either DNA methylation or DNMT3A/B has an impact on the high-affinity chromatin binding of G9a. More interestingly, the catalytic activity of G9a (presence or absence of the G9a inhibitor) did not affect the chromatin binding of G9a in DKO ES cells. Since it is not possible to distinguish the loss of DNA methylation from the loss of DNMTs in DKO ES cells, my result indicates one of two possibilities; (1) the efficient chromatin binding of G9a might primarily depends on methylated-DNA binding proteins, (2) high-affinity chromatin binding of ~30% of all G9a is dependent on the direct recruitment by DNMT3A/B in ES cells. The first possibility requires identification of methyl-DNA binding proteins that interact with G9a/GLP and the second one contradicts the evidence, which showed G9a recruits DNMT3A/B to silence Oct4 during the differentiation of ES cells (Epsztejn-Litman et al. 2008; Athanasiadou et al. 2010). However, this could be due to different mechanisms employed in the establishment and maintenance of DNA methylation. CRISPR/Cas9 system could be used in a more systematic approach that involves generation of a number of G9a/GLP- or DNMT-mutants to identify the order of events.

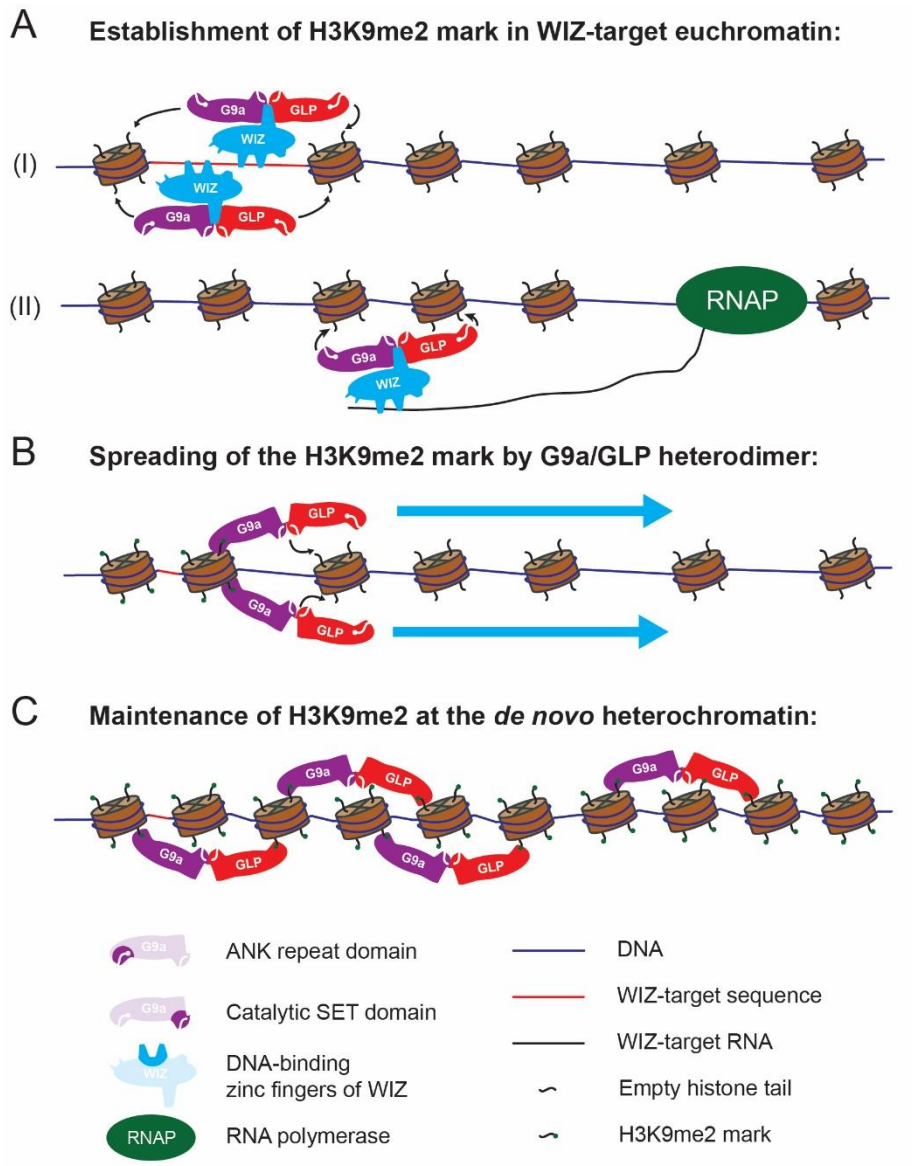


Figure 6. 1 A model of WIZ-dependent recruitment of G9a/GLP heterodimer at a euchromatic region

The representation of a model of WIZ-dependent G9a recruitment to a specific euchromatin locus. **(A)** (I) Sequence specific recruitment of the G9a complex to a 'nucleation site' is mediated by DNA binding fingers of WIZ, which leads to establishment of H3K9me2 mark at the target euchromatic site. RNA binding of WIZ mediates the recruitment of G9a/GLP to establish H3K9me2. **(B)** Ability to bind its own modification enables G9a/GLP heterodimer to spread the mark along the nearby nucleosomes, which creates a heterochromatic environment. **(C)** Maintenance of the H3K9me2 mark is achieved by G9a/GLP heterodimer during cell cycle. G9a/GLP binding to H3K9me2 modified histone tails could also protect the mark against histone demethylation.

6.5. Future directions

Two ChIP-seq of WIZ provided a great deal of information about WIZ localisation despite the fact that the identified WIZ binding motifs in these studies were completely different (Bian et al. 2015 and Isbel et al. 2016). However, these studies failed to differentiate between the direct and indirect localisation of WIZ at these sites. To understand the role of WIZ in the G9a complex, I generated plasmids for N and C-terminally FLAG-tagged WIZ and FLAG-tagged WIZ that is missing the sixth zinc finger. The stable expression of these proteins in WIZ KD ES cells followed by a mass spectrometry might provide valuable information about the interaction network of the G9a complex members. WIZ is a large protein (135 kD large and 100 kD small isoforms, respectively) and, apart from its DNA-binding, might play a scaffolding role in the interaction of the G9a complex members. More importantly, a comparison between ChIP-seq of FLAG-WIZ and FLAG-WIZΔZF6 would identify the genomic sites, which are directly targeted by WIZ but not by G9a/GLP.

The DNA binding of WIZ has become well-documented by studies mentioned above and by my data. However, it is still unclear whether WIZ plays a role in the establishment of the H3K9me2 by directly recruiting G9a to a specific genomic locus. To elaborate on this topic, I have generated a reporter plasmid containing a TET-responsive element upstream of a CMV promoter, a GFP or RFP gene, three tandem repeats of T15short sequence and a polyadenylation signal sequence, which are flanked by two loxP sites. Recombination-mediated cassette exchange (RMCE) or a CRISPR/Cas9 approach can be used to insert

the reporter gene into the genome of ES or another type of cells at a location lacking H3K9me2. Then, by performing ChIP experiments on this locus, it can be tested whether or not T15 sites bind WIZ *in vivo* and whether WIZ-recruits G9a to establish H3K9me2 and silence transcription. This strategy would allow investigation of WIZ-mediated G9a recruitment at a heterochromatic or a euchromatic locus, and in the presence and the absence of transcription. For this purpose, I knocked down WIZ in a cell line with loxP sites integrated into the beta-globin locus (RMCE cell line; Lienert et al. 2011). Similarly, other RMCE cell lines containing loxP sites that are at an open chromatin locus such as, ROSA26 can be used. These cell lines will be useful to study the mechanisms of WIZ-mediated G9a recruitment.

6.6. Concluding Remarks

In summary, the results obtained during my studies have provided novel information about the DNA binding preferences of WIZ. The mass spectrometry data has added additional names to the binding partner catalogue of G9a. Finally, my data suggest that the chromatin binding of G9a is dependent on H3K9me2, indicating the existence of a potential 'oozing' model in the spreading of H3K9me2. Additionally, I showed that the presence of DNMTs is required for high-affinity chromatin binding of G9a. Future studies should be focusing on the following issues; (1) What are the DNA- or RNA-targets of WIZ? (2) Is WIZ is essential for the de novo establishment of H3K9me2? It is important to note that the other zinc finger proteins identified in the MS data might have a similar role as WIZ in the recruitment of G9a. (3) What is the importance of the WIZ-mediated G9a recruitment

during embryonic development? Addressing these questions is crucial to understand the reason behind the embryonic lethality observed in mice lacking G9a/GLP or WIZ.

Reference List

Aasland, R. & Stewart, A.F., 1995. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. *Nucleic acids research*, 23(16), pp.3168–73.

Ait-Si-Ali, S. et al., 2004. A Suv39h-dependent mechanism for silencing S-phase genes in differentiating but not in cycling cells. *The EMBO journal*, 23(3), pp.605–15.

Allfrey, V.G., Faulkner, R. & Mirsky, A.E., 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 51, pp.786–94.

Arakawa, H. et al., 1993. The complete murine immunoglobulin class switch region of the alpha heavy chain gene-hierarchic repetitive structure and recombination breakpoints. *The Journal of biological chemistry*, 268(7), pp.4651–5.

Athanasiadou, R. et al., 2010. Targeting of De Novo DNA Methylation Throughout the Oct-4 Gene Regulatory Region in Differentiating Embryonic Stem Cells R. Feil, ed. *PLoS ONE*, 5(4), p.e9937.

Auclair, G. et al., 2016. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome research*, 26(2), pp.192–202.

Audergon, P.N.C.B. et al., 2015. Epigenetics. Restricted epigenetic inheritance of H3K9 methylation. *Science (New York, N.Y.)*, 348(6230), pp.132–5.

Avvakumov, G. V et al., 2008. Structural basis for recognition of hemi-methylated DNA

by the SRA domain of human UHRF1. *Nature*, 455(7214), pp.822–5.

Balan, S. et al., 2014. Exon resequencing of H3K9 methyltransferase complex genes, EHMT1, EHTM2 and WIZ, in Japanese autism subjects. *Molecular autism*, 5(1), p.49.

Banck, M.S. et al., 2009. The ZNF217 oncogene is a candidate organizer of repressive histone modifiers. *Epigenetics*, 4(2), pp.100–106.

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–95.

Barth, T.K. et al., 2010. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences*, 35(11), pp.618–626.

Bartke, T. et al., 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, 143(3), pp.470–84.

Bartke, T. & Kouzarides, T., 2011. Decoding the chromatin modification landscape. *Cell Cycle*, 10(2), pp.182–182.

Bateman, A. et al., 2002. The Pfam Protein Families Database. *Nucleic Acids Research*, 30(1), pp.276–280.

Baubec, T. et al., 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, 520(7546), pp.243–7.

Becker, P.B., 2006. Gene regulation: A finger on the mark. *Nature*, 442(7098), pp.31–32.

Beltran, A.S. & Blancafort, P., 2011. Reactivation of MASPIN in non-small cell lung

carcinoma (NSCLC) cells by artificial transcription factors (ATFs). *Epigenetics*, 6(2), pp.224–35.

Berg, J.M., 1988. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 85(1), pp.99–102.

Bernstein, B.E. et al., 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, 125(2), pp.315–326.

Bestor, T.H., 1990. DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 326(1235), pp.179–87.

Bestor, T.H., 2000. The DNA methyltransferases of mammals. *Human molecular genetics*, 9(16), pp.2395–402.

Bian, C., Chen, Q. & Yu, X., 2015. The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *eLife*, 4.

Biel, M., Wascholowski, V. & Giannis, A., 2005. Epigenetics--an epicenter of gene regulation: histones and histone-modifying enzymes. *Angewandte Chemie (International ed. in English)*, 44(21), pp.3186–216.

Bilodeau, S. et al., 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes & Development*, 23(21), pp.2484–2489.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1), pp.6–21.

Bird, A.P., 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics*, 3, pp.342–347.

Birke, M. et al., 2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic acids research*, 30(4), pp.958–65.

Black, J.C., Van Rechem, C. & Whetstine, J.R., 2012. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell*, 48(4), pp.491–507.

Blackledge, N.P. et al., 2010. CpG islands recruit a histone H3 lysine 36 demethylase. *Molecular cell*, 38(2), pp.179–90.

Boch, J., 2011. TALEs of genome targeting. *Nature Biotechnology*, 29(2), pp.135–136.

Bochkareva, E., Korolev, S. & Bochkarev, A., 2000. The role for zinc in replication protein A. *The Journal of biological chemistry*, 275(35), pp.27332–8.

Bogan, J.A. & Helmstetter, C.E., 1996. mioC transcription, initiation of replication, and the eclipse in Escherichia coli. *Journal of bacteriology*, 178(11), pp.3201–6.

Botuyan, M.V. et al., 2006. Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell*, 127(7), pp.1361–1373.

Breeden, L. & Nasmyth, K., 1987. Similarity between cell-cycle genes of budding yeast

and fission yeast and the Notch gene of Drosophila. *Nature*, 329(6140), pp.651–654.

Brown, C.J. et al., 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349(6304), pp.38–44.

Brown, D.D., 1984. The role of stable complexes that repress and activate eucaryotic genes. *Cell*, 37(2), pp.359–365.

Buck-Koehntop, B. a et al., 2012. Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proceedings of the National Academy of Sciences of the United States of America*, 109(38), pp.15229–34.

Cano-Rodriguez, D. et al., 2016. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nature Communications*, 7, p.12284.

Caricasole, A. et al., 1996. RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 93(15), pp.7562–6.

Carroll, D., 2008. Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Therapy*, 15(22), pp.1463–1468.

Carrozza, M.J. et al., 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 123(4), pp.581–92.

Casciello, F. et al., 2015. Functional role of G9a histone methyltransferase in cancer.

Frontiers in Immunology, 6(SEP), pp.3–9.

Cavalli, G. & Paro, R., 1998. The Drosophila Fab-7 Chromosomal Element Conveys Epigenetic Inheritance during Mitosis and Meiosis. *Cell*, 93(4), pp.505–518.

Chambers, I. et al., 2007. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173), pp.1230–4.

Chambeyron, S. & Bickmore, W.A., 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. , pp.1119–1130.

Chang, Y. et al., 2011. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nature communications*, 2, p.533.

Chaturvedi, C.-P. et al., 2012. Maintenance of gene silencing by the coordinate action of the H3K9 methyltransferase G9a/KMT1C and the H3K4 demethylase Jarid1a/KDM5A. *Proceedings of the National Academy of Sciences of the United States of America*, 109(46), pp.18845–50.

Chen, T. et al., 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and cellular biology*, 23(16), pp.5594–605.

Chin, H.G. et al., 2007. Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. *Nucleic acids research*, 35(21), pp.7313–23.

Choo, Y., Sánchez-García, I. & Klug, A., 1994. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature*, 372(6507), pp.642–645.

Chow, J. & Heard, E., 2009. X inactivation and the complexities of silencing a sex chromosome. *Current Opinion in Cell Biology*, 21(3), pp.359–366.

Cléard, F. & Spierer, P., 2001. Position-effect variegation in Drosophila: the modifier Su(var)3-7 is a modular DNA-binding protein. *EMBO reports*, 2(12), pp.1095–100.

Clemens, K.R. et al., 1993. Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science (New York, N.Y.)*, 260(5107), pp.530–3.

Collins, R.E. et al., 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol*, 15(3), pp.245–250.

Cross, S.H. et al., 1997. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature genetics*, 16(3), pp.256–9.

Daxinger, L. et al., 2013. An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biology*, 14(9), p.R96.

Denslow, S.A. & Wade, P.A., 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene*, 26(37), pp.5433–8.

Dhayalan, A. et al., 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *The Journal of biological chemistry*, 285(34), pp.26114–20.

Dong, K.B. et al., 2008. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *The EMBO Journal*, 27(20),

pp.2691–2701.

Du, J. et al., 2015. DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology*, 16(9), pp.519–532.

Duan, Z. et al., 2005. Gfi1 Coordinates Epigenetic Repression of p21Cip/WAF1 by Recruitment of Histone Lysine Methyltransferase G9a and Histone Deacetylase 1. *Molecular and Cellular Biology*, 25(23), pp.10338–10351.

Eissenberg, J.C. & Shilatifard, A., 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental biology*, 339(2), pp.240–9.

Ellington, A.D. & Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), pp.818–822.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

ENCODE Project Consortium et al., 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), pp.799–816.

Epsztejn-Litman, S. et al., 2008. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural & Molecular Biology*, 15(11), pp.1176–1183.

Espada, J. et al., 2004. Human DNA Methyltransferase 1 Is Required for Maintenance of the Histone H3 Modification Pattern. *Journal of Biological Chemistry*, 279(35), pp.37175–37184.

Esteve, P.-O. et al., 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & Development*, 20(22), pp.3089–3103.

Fairall, L. et al., 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature*, 366(6454), pp.483–487.

Fan, C.M. & Maniatis, T., 1990. A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. *Genes & development*, 4(1), pp.29–42.

Farcas, A.M. et al., 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife*, 1, p.e00205.

Feinberg, A.P. & Tycko, B., 2004. The history of cancer epigenetics. *Nature Reviews Cancer*, 4(2), pp.143–153.

Feinberg, A.P. & Vogelstein, B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301(5895), pp.89–92.

Felsenfeld, G., 2014. A brief history of epigenetics. *Cold Spring Harbor perspectives in biology*, 6(1).

Feng, Q. et al., 2002. Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain University of North Carolina at Chapel Hill. , 12(2), pp.1052–1058.

Fields, S. & Song, O., 1989. A novel genetic system to detect protein-protein interactions.

Nature, 340(6230), pp.245–6.

Filion, G.J. & van Steensel, B., 2010. Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. *Nature genetics*, 42(1), p.4; author reply 5-6.

Fodor, B.D. et al., 2006. Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. *Genes & development*, 20(12), pp.1557–62.

Franco, B. & Ballabio, A., 2006. X-inactivation and human disease: X-linked dominant male-lethal disorders. *Current Opinion in Genetics & Development*, 16(3), pp.254–259.

Fuks, F. et al., 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research*, 31(9), pp.2305–12.

Garcia-Bassets, I. et al., 2007. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell*, 128(3), pp.505–18.

El Gazzar, M. et al., 2008. G9a and HP1 couple histone and DNA methylation to TNFalpha transcription silencing during endotoxin tolerance. *The Journal of biological chemistry*, 283(47), pp.32198–208.

Gellert, M., Lipsett, M.N. & Davies, D.R., 1962. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*, 48(12), pp.2013–8.

Geuens, T., Bouhy, D. & Timmerman, V., 2016. The hnRNP family: insights into their role in health and disease. *Human Genetics*, 135(8), pp.851–867.

Ghosh, S. et al., 2010. Tissue specific DNA methylation of CpG islands in normal human

adult somatic tissues distinguishes neural from non-neural tissues. *Epigenetics*, 5(6), pp.527–38.

Giri, S. et al., 2015. The preRC protein ORCA organizes heterochromatin by assembling histone H3 lysine 9 methyltransferases on chromatin. *eLife*, 4.

Goll, M.G. et al., 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science (New York, N.Y.)*, 311(5759), pp.395–8.

Goyal, R., Reinhardt, R. & Jeltsch, A., 2006. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic acids research*, 34(4), pp.1182–8.

Gräff, J. & Tsai, L.-H., 2013. Histone acetylation: molecular mnemonics on the chromatin. *Nature reviews. Neuroscience*, 14(2), pp.97–111.

Grewal, S.I.S. & Jia, S., 2007. Heterochromatin revisited. *Nature Reviews Genetics*, 8(1), pp.35–46.

Grozinger, C.M. & Schreiber, S.L., 2002. Deacetylase Enzymes: Biological Functions and the Use of Small-Molecule Inhibitors. *Chemistry & Biology*, 9(1), pp.3–16.

Guelen, L. et al., 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197), pp.948–51.

Guo, J.U. et al., 2013. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, 17(2), pp.215–222.

Habibi, E. et al., 2013. *Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells*,

Hall, T.M.T., 2005. Multiple modes of RNA recognition by zinc finger proteins. *Current Opinion in Structural Biology*, 15(3), pp.367–373.

Hanakahi, L.A., Sun, H. & Maizels, N., 1999. High Affinity Interactions of Nucleolin with G-G-paired rDNA. *Journal of Biological Chemistry*, 274(22), pp.15908–15912.

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(5), pp.1103–1112.

Harper, S. & Speicher, D.W., 2011. Purification of proteins fused to glutathione S-transferase. *Methods in molecular biology (Clifton, N.J.)*, 681, pp.259–80.

Hashimoto, H. et al., 2008. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, 455(7214), pp.826–9.

Henderson, A. et al., 2014. Detection of G-quadruplex DNA in mammalian cells. *Nucleic acids research*, 42(2), pp.860–9.

Hotchkiss, R.D., 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry*, 175(1), pp.315–32.

Hou, T. et al., 2008. The STAT3 NH2-terminal domain stabilizes enhanceosome assembly by interacting with the p300 bromodomain. *The Journal of biological chemistry*, 283(45), pp.30725–34.

Hsu, P.D. et al., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), pp.1262–1278.

Huang, Y. et al., 2006. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science (New York, N.Y.)*, 312(5774), pp.748–51.

Isbel, L. et al., 2016. Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. *eLife*, 5, p.e15082.

Iuchi, S., 2001. Three classes of C2H2 zinc finger proteins. *Cellular and Molecular Life Sciences*, 58(4), pp.625–635.

Jacobs, S.A. et al., 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science (New York, N.Y.)*, 295(5562), pp.2080–3.

Jenuwein, T., 2006. The epigenetic magic of histone lysine methylation. *FEBS Journal*, 273(14), pp.3121–3135.

Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science (New York, N.Y.)*, 293(5532), pp.1074–80.

Jia, D. et al., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, 449(7159), pp.248–51.

Johnson, C. et al., 2010. Bioinformatic and experimental survey of 14-3-3-binding sites. *The Biochemical journal*, 427(1), pp.69–78.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7), pp.484–492.

Jones, P.A., 1999. The DNA methylation paradox. *Trends in Genetics*, 15(1), pp.34–37.

Jones, P.A. & Baylin, S.B., 2002. The fundamental role of epigenetic events in cancer.

Nature reviews. Genetics, 3(6), pp.415–28.

Jørgensen, H.F., Ben-Porath, I. & Bird, A.P., 2004. Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Molecular and cellular biology*, 24(8), pp.3387–95.

Jurkowska, R.Z. et al., 2008. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Research*, 36(21), pp.6656–6663.

Jurkowska, R.Z., Jurkowski, T.P. & Jeltsch, A., 2011. Structure and function of mammalian DNA methyltransferases. *Chembiochem : a European journal of chemical biology*, 12(2), pp.206–22.

Jurkowski, T.P. et al., 2015. Synthetic epigenetics—towards intelligent control of epigenetic states and cell identity. *Clinical Epigenetics*, 7(1), p.18.

Karagianni, P. et al., 2008. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Molecular and cellular biology*, 28(2), pp.705–17.

Kashiwagi, K. et al., 2011. DNA methyltransferase 3b preferentially associates with condensed chromatin. *Nucleic Acids Research*, 39(3), pp.874–888.

Keogh, M.-C. et al., 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*, 123(4), pp.593–605.

Kim, Y. et al., 2013. BIX-01294 induces autophagy-associated cell death via EHMT2/G9a dysfunction and intracellular reactive oxygen species production. *Autophagy*, 9(12),

pp.2126–2139.

Kind, J. et al., 2013. Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell*, 153(1), pp.178–192.

Klug, A., 1983. From macromolecules to biological assemblies. Nobel Lecture, 8 December 1982. *Bioscience reports*, 3(5), pp.395–430.

Klug, A., 2010. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry*, 79(1), pp.213–231.

Kouzarides, T., 2007. Chromatin modifications and their function. *Cell*, 128(4), pp.693–705.

Koziol, M.J. & Rinn, J.L., 2010. RNA traffic control of chromatin complexes. *Current opinion in genetics & development*, 20(2), pp.142–8.

Kress, C., Thomassin, H. & Grange, T., 2001. Local DNA demethylation in vertebrates: how could it be performed and targeted? *FEBS Letters*, 494(3), pp.135–140.

Krishna, S.S., Majumdar, I. & Grishin, N. V, 2003. Structural classification of zinc fingers: survey and summary. *Nucleic acids research*, 31(2), pp.532–50.

Kumar, R. & Rao, D.N., 2013. Role of DNA Methyltransferases in Epigenetic Regulation in Bacteria. In Springer Netherlands, pp. 81–102.

Kungulovski, G. et al., 2015. Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenetics & Chromatin*, 8(1), p.12.

Lachner, M. et al., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1

proteins. *Nature*, 410(6824), pp.116–120.

Laity, J.H., Dyson, H.J. & Wright, P.E., 2000. *DNA-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in Cys2-His2 zinc fingers*,

Laity, J.H., Lee, B.M. & Wright, P.E., 2001. Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*, 11(1), pp.39–46.

Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860–921.

De Las Rivas, J. & Fontanillo, C., 2010. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks F. Lewitter, ed. *PLoS Computational Biology*, 6(6), p.e1000807.

Lee, D.Y. et al., 2006. Histone H3 Lysine 9 Methyltransferase G9a Is a Transcriptional Coactivator for Nuclear Receptors. *Journal of Biological Chemistry*, 281(13), pp.8476–8485.

Lee, M.S. et al., 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science (New York, N.Y.)*, 245(4918), pp.635–7.

Van Leeuwen, F., Gafken, P.R. & Gottschling, D.E., 2002. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell*, 109(6), pp.745–756.

Lehnertz, B. et al., 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current biology : CB*, 13(14), pp.1192–200.

Lelli, K.M., Slattery, M. & Mann, R.S., 2012. Disentangling the many layers of eukaryotic transcriptional regulation. *Annual review of genetics*, 46, pp.43–68.

Leon, O. & Roth, M., 2000. Zinc fingers: DNA binding and protein-protein interactions. *Biological Research*, 33(1), pp.21–30.

Li, E. et al., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6), pp.915–926.

Lienert, F. et al., 2011. Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS Genet*, 7(6), p.e1002090.

Liew, C.K. et al., 2005. Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3), pp.583–8.

Lipkin, S.M. et al., 1993. Identification of a novel zinc finger protein binding a conserved element critical for Pit-1-dependent growth hormone gene expression. *Genes & development*, 7(9), pp.1674–87.

Liu, C. et al., 2014. A chromatin activity-based chemoproteomic approach reveals a transcriptional repressome for gene-specific silencing. *Nature Communications*, 5, p.5733.

Liu, K. et al., 2010. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proceedings of the National Academy of Sciences*, 107(43), pp.18398–18403.

Lopez-Contreras, A.J. et al., 2013. A proteomic characterization of factors enriched at nascent DNA molecules. *Cell reports*, 3(4), pp.1105–16.

Low, D.A., Weyand, N.J. & Mahan, M.J., 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infection and immunity*, 69(12), pp.7197–204.

Lyon, M.F., 1961. Gene Action in the X-chromosome of the Mouse (Mus musculus L.). *Nature*, 190(4773), pp.372–373.

Ma, R. et al., 2014. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *Journal of Zhejiang University. Science. B*, 15(5), pp.412–28.

Maier, V.K. et al., 2015. Functional Proteomic Analysis of Repressive Histone Methyltransferase Complexes Reveals ZNF518B as a G9A Regulator. *Molecular & cellular proteomics : MCP*, 14(6), pp.1435–46.

Mancini-Dinardo, D. et al., 2006. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes & development*, 20(10), pp.1268–82.

Margueron, R. & Reinberg, D., 2010. Chromatin structure and the inheritance of epigenetic information. *Nature reviews. Genetics*, 11(4), pp.285–96.

Martino, F. et al., 2009. Reconstitution of Yeast Silent Chromatin: Multiple Contact Sites and O-AADPR Binding Load SIR Complexes onto Nucleosomes In Vitro. *Molecular Cell*, 33(3), pp.323–334.

Massé, J. et al., 2010. Involvement of ZFPIP/Zfp462 in chromatin integrity and survival

of P19 pluripotent cells. *Experimental cell research*, 316(7), pp.1190–201.

Mastroeni, D. et al., 2015. Aberrant intracellular localization of H3k4me3 demonstrates an early epigenetic phenomenon in Alzheimer's disease. *Neurobiology of aging*, 36(12), pp.3121–9.

Matsui, T. et al., 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*, 464(7290), pp.927–31.

Matsumoto, K. et al., 1998. Molecular cloning and distinct developmental expression pattern of spliced forms of a novel zinc finger gene wiz in the mouse cerebellum. *Molecular Brain Research*, 61(1), pp.179–189.

Meissner, A. et al., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), pp.766–70.

Meissner, A. et al., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research*, 33(18), pp.5868–77.

Mendenhall, E.M. et al., 2013. Locus-specific editing of histone modifications at endogenous enhancers. *Nature biotechnology*, 31(12), pp.1133–6.

Miller, J., McLachlan, A.D. & Klug, A., 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO journal*, 4(6), pp.1609–14.

Min, J. et al., 2002. Structure of the SET domain histone lysine methyltransferase Clr4. *Nature structural biology*, 9(11), pp.828–32.

Mizzen, C.A. & Allis, C.D., 1998. Linking histone acetylation to transcriptional regulation. *Cellular and Molecular Life Sciences (CMLS)*, 54(1), pp.6–20.

Muller, S., Filippakopoulos, P. & Knapp, S., 2011. Bromodomains as therapeutic targets. *Expert reviews in molecular medicine*, 13, p.e29.

Murray, K., 1964. The occurrence of epsilon-N-methyl lysine in histones. *Biochemistry*, 3, pp.10–5.

Myant, K. et al., 2011. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res*, 21(1), pp.83–94.

Myant, K. & Stancheva, I., 2008. LSH cooperates with DNA methyltransferases to repress transcription. *Molecular and cellular biology*, 28(1), pp.215–26.

Nady, N. et al., 2011. Recognition of Multivalent Histone States Associated with Heterochromatin by UHRF1 Protein. *Journal of Biological Chemistry*, 286(27), pp.24300–24311.

Nagano, T. et al., 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*, 322(5908), pp.1717–1720.

Nair, S.S., Li, D.-Q. & Kumar, R., 2013. A core chromatin remodeling factor instructs global chromatin signaling through multivalent reading of nucleosome codes. *Molecular cell*, 49(4), pp.704–18.

Ng, K. et al., 2007. Xist and the order of silencing. *EMBO reports*, 8(1), pp.34–9.

Nishino, Y. et al., 2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal*,

31(7), pp.1644–53.

Nolte, R.T. et al., 1998. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), pp.2938–43.

Ntranos, A. & Casaccia, P., 2016. Bromodomains: Translating the words of lysine acetylation into myelin injury and repair. *Neuroscience Letters*, 625, pp.4–10.

Ogawa, H. et al., 2002. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science*, 296(5570), pp.1132–1136.

Okano, M. et al., 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. , 99, pp.247–257.

Okitsu, C.Y., Hsieh, J.C.F. & Hsieh, C.L., 2010. Transcriptional Activity Affects the H3K4me3 Level and Distribution in the Coding Region. *Molecular and Cellular Biology*, 30(12), pp.2933–2946.

Oliphant, A.R., Brandl, C.J. & Struhl, K., 1989. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Molecular and Cellular Biology*, 9(7), pp.2944–2949.

Ooi, S.K.T. et al., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154), pp.714–7.

Orlicky, S. et al., 2003. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell*, 112(2), pp.243–56.

Otani, J. et al., 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep*, 10(11), pp.1235–1241.

Pandey, R.R. et al., 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell*, 32(2), pp.232–246.

Parkinson, G.N., Lee, M.P.H. & Neidle, S., 2002. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, 417(6891), pp.876–880.

Parthun, M.R., 2007. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene*, 26(37), pp.5319–28.

Pavletich, N.P. & Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science (New York, N.Y.)*, 252(5007), pp.809–17.

Pedersen, M.T. & Helin, K., 2010. Histone demethylases in development and disease. *Trends in cell biology*, 20(11), pp.662–71.

Pek, J.W. et al., 2012. Tudor domain proteins in development. *Development (Cambridge, England)*, 139(13), pp.2255–66.

Peng, J.C. & Karpen, G.H., 2009. Heterochromatic Genome Stability Requires Regulators of Histone H3 K9 Methylation R. S. Hawley, ed. *PLoS Genetics*, 5(3), p.e1000435.

Persikov, A. V., Osada, R. & Singh, M., 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics*, 25(1), pp.22–29.

Peters, A.H.F.M. et al., 2001. Loss of the Suv39h Histone Methyltransferases Impairs

Mammalian Heterochromatin and Genome Stability. *Cell*, 107(3), pp.323–337.

Peters, A.H.F.M. et al., 2003. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Molecular Cell*, 12(6), pp.1577–1589.

Philipps, D.L. et al., 2008. The dual bromodomain and WD repeat-containing mouse protein BRWD1 is required for normal spermiogenesis and the oocyte-embryo transition. *Developmental biology*, 317(1), pp.72–82.

Pinheiro, I. et al., 2012. Prdm3 and Prdm16 are H3K9me1 Methyltransferases Required for Mammalian Heterochromatin Integrity. *Cell*, 150(5), pp.948–960.

Pozuelo Rubio, M. et al., 2004. 14-3-3-affinity purification of over 200 human phosphoproteins reveals new links to regulation of cellular metabolism, proliferation and trafficking. *The Biochemical journal*, 379(Pt 2), pp.395–408.

Pradhan, S. et al., 1999. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *The Journal of biological chemistry*, 274(46), pp.33002–10.

Purcell, D.J. et al., 2012. Recruitment of coregulator G9a by Runx2 for selective enhancement or suppression of transcription. *Journal of cellular biochemistry*, 113(7), pp.2406–14.

Qian, X. et al., 1993. Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry*, 32(38), pp.9944–59.

Rai, K. et al., 2007. Dnmt2 functions in the cytoplasm to promote liver, brain, and retina

development in zebrafish. *Genes & development*, 21(3), pp.261–6.

Rathert, P. et al., 2008a. Protein lysine methyltransferase G9a acts on non-histone targets. *Nature chemical biology*, 4(6), pp.344–6.

Rathert, P. et al., 2008b. Protein lysine methyltransferase G9a acts on non-histone targets. *Nature chemical biology*, 4(6), pp.344–6.

Rea, S. et al., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), pp.593–9.

Reik, W. & Walter, J., 2001. Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics*, 2(1), pp.21–32.

Reisenauer, A. & Shapiro, L., 2002. DNA methylation affects the cell cycle transcription of the CtrA global regulator in Caulobacter. *The EMBO journal*, 21(18), pp.4969–77.

Richmond, T.J. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 A| resolution. *Nature*, 389(6648), pp.251–260.

Ringrose, L. et al., 2003. Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. *Developmental cell*, 5(5), pp.759–71.

Ringrose, L. & Paro, R., 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual review of genetics*, 38, pp.413–43.

Ringrose, L. & Paro, R., 2007. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development (Cambridge, England)*, 134(2), pp.223–32.

Robertson, K.D., 2005. DNA methylation and human disease. *Nature Reviews Genetics*, 6(8), pp.597–610.

Roopra, A. et al., 2004. Localized Domains of G9a-Mediated Histone Methylation Are Required for Silencing of Neuronal Genes. *Molecular Cell*, 14(6), pp.727–738.

Rose, N.R. & Klose, R.J., 2014. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(12), pp.1362–1372.

Rottach, A. et al., 2010. The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic acids research*, 38(6), pp.1796–804.

Rougeulle, C. et al., 2004. Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Molecular and cellular biology*, 24(12), pp.5475–84.

Rusinova, E. et al., 2002. Alexa and Oregon Green dyes as fluorescence anisotropy probes for measuring protein–protein and protein–nucleic acid interactions. *Analytical Biochemistry*, 308(1), pp.18–25.

Ruthenburg, A.J. et al., 2007. Multivalent engagement of chromatin modifications by linked binding modules. *Nature reviews. Molecular cell biology*, 8(12), pp.983–94.

Saksouk, N. et al., 2015. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin*, 8(1), p.3.

Sampath, S.C. et al., 2007. Methylation of a Histone Mimic within the Histone Methyltransferase G9a Regulates Protein Complex Assembly. *Molecular Cell*, 27(4), pp.596–608.

Sanchez, R. & Zhou, M.-M., 2009. The role of human bromodomains in chromatin biology and gene transcription. *Current opinion in drug discovery & development*, 12(5), pp.659–65.

Santos-Pereira, J.M. & Aguilera, A., 2015. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics*, 16(10), pp.583–597.

Sattler, M. et al., 2001. SMN Tudor domain structure and its interaction with the Sm proteins. *Nature Structural Biology*, 8(1), pp.27–31.

Schaefer, M. & Lyko, F., 2010. Solving the Dnmt2 enigma. *Chromosoma*, 119(1), pp.35–40.

Schuh, R. et al., 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Krüppel, a Drosophila segmentation gene. *Cell*, 47(6), pp.1025–1032.

Searles, M.A., Lu, D. & Klug, A., 2000. The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA. *Journal of molecular biology*, 301(1), pp.47–60.

Shang, E. et al., 2009. Double bromodomain-containing gene Brd2 is essential for embryonic development in mouse. *Developmental dynamics : an official publication of the American Association of Anatomists*, 238(4), pp.908–17.

Shankar, S.R. et al., 2013. G9a, a multipotent regulator of gene expression. *Epigenetics*, 8(1), pp.16–22.

Sharif, J. et al., 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 450(7171), pp.908–12.

Shema, E. et al., 2016. Single-molecule decoding of combinatorially modified nucleosomes. *Science (New York, N.Y.)*, 352(6286), pp.717–21.

Shemer, R. et al., 1997. Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. *Proceedings of the National Academy of Sciences of the United States of America*, 94(19), pp.10267–72.

Shi, Y. et al., 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119(7), pp.941–53.

Shinkai, Y. & Tachibana, M., 2011. H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev*, 25(8), pp.781–788.

Shogren-Knaak, M. et al., 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (New York, N.Y.)*, 311(5762), pp.844–7.

Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N.J., 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, 516(7531), pp.436–9.

Sleutels, F., Zwart, R. & Barlow, D.P., 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, 415(6873), pp.810–3.

Smallwood, A. et al., 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes & development*, 21(10), pp.1169–78.

Smith, E.R., Allis, C.D. & Lucchesi, J.C., 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in Drosophila males. *The Journal of biological chemistry*, 276(34), pp.31483–6.

Smith, Z.D. & Meissner, A., 2013. DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, 14(3), pp.204–220.

Song, C.-X. et al., 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature biotechnology*, 29(1), pp.68–72.

Song, F. et al., 2014. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science*, 344(6182).

Tachibana, M. et al., 2008. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *The EMBO journal*, 27(20), pp.2681–90.

Tachibana, M. et al., 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. , 9, pp.815–826.

Tachibana, M. et al., 2001. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *The Journal of biological chemistry*, 276(27), pp.25309–17.

Takahashi, K. & Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4), pp.663–76.

Talbert, P.B. & Henikoff, S., 2006. Spreading of silent chromatin: inaction at a distance. *Nature Reviews Genetics*, 7(10), pp.793–803.

Theunissen, O. et al., 1992. RNA and DNA binding zinc fingers in Xenopus TFIIIA. *Cell*, 71(4), pp.679–690.

Thomson, J.P. et al., 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, 464(7291), pp.1082–6.

Trievel, R.C. et al., 2002. Structure and Catalytic Mechanism of a SET Domain Protein Methyltransferase. *Cell*, 111(1), pp.91–103.

Trojer, P. et al., 2009. Dynamic Histone H1 Isotype 4 Methylation and Demethylation by Histone Lysine Methyltransferase G9a/KMT1C and the Jumonji Domain-containing JMJD2/KDM4 Proteins. *The Journal of biological chemistry*, 284(13), pp.8395–405.

Trojer, P. & Reinberg, D., 2007. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell*, 28(1), pp.1–13.

Tsumura, A. et al., 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to cells : devoted to molecular & cellular mechanisms*, 11(7), pp.805–14.

Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, 249(4968), pp.505–10.

Tuorto, F. et al., 2015. The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *The EMBO Journal*, 34(18), pp.2350–2362.

Tweedie, S. et al., 1997. Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology*, 17(3), pp.1469–1475.

Ueda, J. et al., 2006. Zinc finger protein Wiz links G9a/GLP histone methyltransferases

to the co-repressor molecule CtBP. *The Journal of biological chemistry*, 281(29), pp.20120–8.

Ule, J. et al., 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science (New York, N.Y.)*, 302(5648), pp.1212–5.

Upadhyay, A.K. et al., 2012. An analog of BIX-01294 selectively inhibits a family of histone H3 lysine 9 Jumonji demethylases. *Journal of molecular biology*, 416(3), pp.319–27.

Vashishtha, M. et al., 2013. Targeting H3K4 trimethylation in Huntington disease. *Proceedings of the National Academy of Sciences*, 110(32), pp.E3027–E3036. Vassen, L. et al., 2006. Gfi1b alters histone methylation at target gene promoters and sites of γ-satellite containing heterochromatin. *The EMBO Journal*, 25(11), pp.2409–2419.

Vedadi, M. et al., 2011. A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nature chemical biology*, 7(8), pp.566–74.

Verdone, L. et al., 2006. Histone acetylation in gene regulation. *Briefings in functional genomics & proteomics*, 5(3), pp.209–21.

Voigt, P. et al., 2012. Asymmetrically Modified Nucleosomes. *Cell*, 151(1), pp.181–193.

Vojta, A. et al., 2016. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research*, 44(12), pp.5615–5628.

Waddington, C.H., 1953. Genetic Assimilation of an Acquired Character. *Evolution*, 7(2), p.118.

Wakayama, T. & Yanagimachi, R., 2001. Mouse cloning with nucleus donor cells of different age and type. *Molecular Reproduction and Development*, 58(4), pp.376–383.

Walsh, G. & Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. *Nature biotechnology*, 24(10), pp.1241–52.

Wang, B. et al., 1998. High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure (London, England : 1993)*, 6(5), pp.555–69.

Weaver, I.C.G., 2007. Epigenetic programming by maternal behavior and pharmacological intervention. Nature versus nurture: let's call the whole thing off. *Epigenetics*, 2(1), pp.22–8.

Weitzmann, M.N., Woodford, K.J. & Usdin, K., 1997. DNA secondary structures and the evolution of hypervariable tandem arrays. *The Journal of biological chemistry*, 272(14), pp.9517–23.

Wen, B. et al., 2009. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature genetics*, 41(2), pp.246–50.

Wold, M.S., 1997. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annual Review of Biochemistry*, 66(1), pp.61–92.

Wolfe, S.A., Nekludova, L. & Pabo, C.O., 2000. DNA Recognition by Cys $_2$ His $_2$ Zinc Finger Proteins. *Annual Review of Biophysics and Biomolecular Structure*, 29(1),

pp.183–212.

Woodcock, C.L., Frado, L.L. & Rattner, J.B., 1984. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *The Journal of cell biology*, 99(1 Pt 1), pp.42–52.

Wu, H. et al., 2011. Structural and histone binding ability characterizations of human PWWP domains. *PLoS ONE*, 6(6).

Wu, H. et al., 2010. Structural biology of human H3K9 methyltransferases. *PLoS ONE*, 5(1).

Wysocka, J. et al., 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, 121(6), pp.859–72.

Yaffe, M.B. et al., 1997. The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell*, 91(7), pp.961–71.

Yamamoto, K. & Sonoda, M., 2003. Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. *Biochemical and Biophysical Research Communications*, 301(2), pp.287–292.

Yan, C. & Boyd, D.D., 2006. Histone H3 acetylation and H3 K4 methylation define distinct chromatin regions permissive for transgene expression. *Molecular and cellular biology*, 26(17), pp.6357–71.

Yang, X.-J. & Seto, E., 2007. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, 26(37), pp.5310–5318.

Ying, Q.-L. et al., 2008. The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194), pp.519–23. Yokochi, T. et al., 2009. G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proceedings of the National Academy of Sciences*, 106(46), pp.19363–19368.

Yun, M. et al., 2011. Readers of histone modifications. *Cell Research*, 21(4), pp.564–578.

Zhang, T. et al., 2016. G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells. *Cell Reports*, 15(1), pp.77–85.

Zhang, X. et al., 2003. Structural basis for the product specificity of histone lysine methyltransferases. *Molecular cell*, 12(1), pp.177–85.

Zhou, V.W., Goren, A. & Bernstein, B.E., 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, 12(1), pp.7–18.

Appendices

Description (Name, Gene Name)	Number of peptides found in samples			
	WT (3)	F-G9a (6)	WT (9)	F-G9a (12)
Histone-lysine N-methyltransferase, **Ehmt2 (G9a)**		11		17
Widely interspaced zinc finger protein, **Wiz**		12		12
Histone-lysine N-methyltransferase EHMT1, **Ehmt1 (GLP)**		13		15
Protein Zfp462, **Zfp462**		23		24
DEAD (Asp-Glu-Ala-Asp) box polypeptide 17, **Ddx17**		12		10
U5 small nuclear ribonucleoprotein 200 kDa helicase, **Snrnp200**		24		18
DNA (cytosine-5)-methyltransferase 3A, **Dnmt3a**	1	4		2
DNA (cytosine-5)-methyltransferase 3B, **Dnmt3b**		2		6
DNA (cytosine-5)-methyltransferase 3-like, **Dnmt3l**		8		8
C-terminal binding protein 2, isoform CRA_b, **Ctbp2**		5		4
Lymphocyte-specific helicase, **Hells**		7		5
Chromodomain-helicase-DNA-binding protein 4, **Chd4**		9		8
Histone deacetylase 1, **Hdac1**		7		5
Putative ribosomal RNA methyltransferase NOP2, **Nop2**		7		4
Polycomb protein EED, **Eed**		4		6
Zinc finger protein 518B, **Znf518b**		5		5
Histone-binding protein RBBP4, **Rbbp4**		4		4
Zinc finger CCCH domain-containing protein 11A, **Zc3h11a**		3		3
Polycomb protein Suz12, **Suz12**		2		5
Protein Mki67, **Mki67**		7		5
Protein Supt16, **Supt16**	1	15		14
Pre-mRNA-processing factor 19, **Prpf19**	1	7		10
Importin subunit alpha-2, **Kpna2**		6		4
DNA mismatch repair protein Msh6, **Msh6**		8		8
RuvB-like 1, **Ruvbl1**		8		6
WD40 repeat-containing protein SMU1, **Smu1**		6		7
Telomere-associated protein RIF1, **Rif1**		9		9
Myb-binding protein 1A, **Mybbp1a**		13		16
Nucleolar protein 56, **Nop56**		17		11
H/ACA ribonucleoprotein complex subunit 4, **Dkc1**		6		7
Tight junction protein ZO-2, **Tjp2**		10		16
E3 ubiquitin-protein ligase TRIP12, **Trip12**		9		5
Transcriptional regulator ATRX, **Atrx**		5		5
Origin recognition complex subunit 1, **Orc1**		6		5
E3 ubiquitin-protein ligase RING2, **Rnf2**		2		2
WD repeat-containing protein 76, **Wdr76**		3		2
Tight junction protein ZO-1, **Tjp1**		4		4
Scaffold attachment factor B1, **Safb**		3		4
SWI/SNF complex subunit SMARCC1, **Smarcc1**		4		2
Polycomb protein Suz12, **Suz12**		2		5
E3 ubiquitin-protein ligase TRIM33, **Trim33**		2		2
Bcl-2-associated transcription factor 1, **Bclaf1**		3		4
Transformer-2 protein homolog beta, **Tra2b**		3		3
Origin recognition complex subunit 4, **Orc4**		4		2
DNA mismatch repair protein Msh2, **Msh2**		4		3
F-actin-capping protein subunit alpha-2, **Capza2**		3		2
Chromobox protein homolog 1, **Cbx1**		5		4
Pogo transposable element with ZNF domain, **Pogz**		2		2

Protein Srsf6, **Srsf6**	3	3
Testis-expressed sequence 10 protein, **Tex10**	3	2
Kinesin-like protein KIF22, **Kif22**	4	4
WD repeat-containing protein 18, **Wdr18**	2	2
Chromobox protein homolog 5, **Cbx5**	2	3
Serine/arginine-rich splicing factor 1, **Srsf1**	4	5
Serine/arginine-rich splicing factor 7, **Srsf7**	2	3
Actin-binding protein anillin, **Anln**	3	4
Chromosome alignment-maintaining phosphoprotein 1, **Champ1**	2	3
ATP-dependent RNA helicase DDX39A, **Ddx39a**	4	5
U4/U6 small nuclear ribonucleoprotein Prp3, **Prpf3**	2	3
N-acylneuraminate cytidylyltransferase, **Cmas**	2	2
E3 ubiquitin-protein ligase RING2, **Rnf2**	2	2
Interleukin enhancer-binding factor 2, **Ilf2**	2	4
Proline-, glutamic acid- and leucine-rich protein 1, **Pelp1**	3	4
Protein phosphatase 1 regulatory subunit 12A, **Ppp1r12a**	4	5
Cbx3 protein, **Cbx3**	3	6
Ankycorbin, **Rai14**	3	6
Cleavage and polyadenylation specificity factor subunit 1, **Cpsf1**	2	4
LIM domain and actin-binding protein 1, **Lima1**	3	6
Metastasis-associated protein MTA2, **Mta2**	2	4

Table S1. 1 Proteins identified by mass spectrometry of purified G9a-FLAG

The full list of proteins identified in the mass spectrometry experiment. The G9a core complex members are colored in green. Proteins that were also identified in the preliminary mass spectrometry experiment are colored in blue. Proteins that were identified as a G9a-interacting protein by others are colored in yellow. Only proteins with more than 2 peptides in G9a-F and less than 2 peptides in WT control are included in the list.

Description (Name, Gene name)	Number of peptides found in samples				Irina's data	
	WT (3)	F-G9a (6)	WT (9)	F-G9a (12)	WT	F-G9a
Histone-lysine N-methyltransferase EHMT2, **Ehmt2**	0	11	0	17	0	20
Widely interspaced zinc finger protein, **Wiz**	0	12	0	12	0	18
Histone-lysine N-methyltransferase EHMT1, **Ehmt1**	0	13	0	15	0	16
Protein Zfp462, **Zfp462**	0	23	0	24	0	35
C-terminal binding protein 2, isoform CRA_b, **Ctbp2**	0	5	0	4	0	3
Pogo transposable element with ZNF domain, **Pogz**	0	2	0	2	0	2
DNA (cytosine-5)-methyltransferase 3A, **Dnmt3a**	1	4	0	2	0	1
Chromobox protein homolog 1, **Cbx1**	0	5	0	4	1	5
Metastasis-associated protein MTA2, **Mta2**	0	2	0	4	1	7
U5 small nuclear ribonucleoprotein 200 kDa helicase, **Snrnp200**	0	24	0	18	1	2
Interleukin enhancer-binding factor 2, **Ilf2**	0	2	0	4	2	2
Cbx3 protein, **Cbx3**	0	3	0	6	2	9
DNA (cytosine-5)-methyltransferase 3-like, **Dnmt3l**	0	8	0	8	2	3
Heterogeneous nuclear ribonucleoprotein L, **Hnrnpl**	5	0	4	7	0	10

Table S1. 2 Proteins identified in both mass spectrometry experiments

The G9a core complex members are colored in green.. Only proteins with more than 1 peptides in G9a-F and less than 3 peptides in WT control are included in the list. HNRNP L was included as to show a false positive protein.